SHARP OPTIMALITY FOR HIGH DIMENSIONAL COVARIANCE TESTING UNDER SPARSE SIGNALS

BY SONG XI CHEN[1,a*], YUMOU QIU[2,b†] AND SHUYI ZHANG[3,c]

[1]*Guanghua School of Management and Center for Statistical Science, Peking University,* [a]*songxichen@pku.edu.cn*

[2]*School of Mathematical Sciences and Center for Statistical Science, Peking University,* [b]*qiuyumou@math.pku.edu.cn*

[3]*KLATASDS-MoE, School of Statistics, Academy of Statistics and Interdisciplinary Sciences, East China Normal University,* [c]*syzhang@fem.ecnu.edu.cn*

This paper considers one-sample testing of a high dimensional covariance matrix by deriving the detection boundary as a function of the signal sparsity and signal strength under the sparse alternative hypotheses. It first shows that the optimal detection boundary for testing sparse means is the minimax detection lower boundary for testing the covariance matrix. A multi-level thresholding test is proposed and is shown to be able to attain the detection lower boundary over a substantial range of the sparsity parameter, implying that the multi-level thresholding test is sharp optimal in the minimax sense over the range. The asymptotic distribution of the multi-level thresholding statistic for covariance matrices is derived under both Gaussian and non-Gaussian distributions by developing a novel U-statistic decomposition in conjunction with the matrix blocking and the coupling techniques to handle the complex dependence among the elements of the sample covariance matrix. The superiority in the detection boundary of the multi-level thresholding test over the existing tests is also demonstrated.

1. Introduction. As part of high dimensional statistical inference, testing for high dimensional covariances has been an active area of statistical research in the last two decades. Early high dimensional tests (Ledoit and Wolf, 2002; Jiang, 2004; Schott, 2005; Chen et al., 2010) were largely formulated by modifying the classical fixed dimensional tests (Nagao, 1973; Anderson, 2003), while more general banded covariance structures were considered in Cai and Jiang (2011) and Qiu and Chen (2012).

The existing formulations of one-sample high dimensional covariance tests are generally based on two types of distance measures between the sample and the hypothesized covariance matrices, namely the sum-of-square (L_2) and the maximum ($L_{\max}$) statistics as represented, respectively, by Chen et al. (2010); Qiu and Chen (2012) for the L_2-type and Jiang (2004); Cai and Jiang (2011) for the $L_{\max}$-type. Cai and Ma (2013) and Cai et al. (2013) studied the minimax power of separating the alternative from the null hypotheses under different signal regimes. They showed that the L_2 and $L_{\max}$-tests are minimax rate optimal under the dense signal regime and the sparse and strong signal regime, respectively. In addition to these results, Arias-Castro et al. (2012) investigated near optimal tests for detecting nonzero correlations in a one-sample setting for Gaussian data with clustered signals.

In particular, Cai and Ma (2013) considered the minimax power of the one-sample testing problem of a covariance $\mathbf{\Sigma}$ being identity under the Gaussian distribution. Let $\mathcal{W}_{1,\alpha}$ be the collection of all α-level test procedures for the null hypothesis $H_0 : \mathbf{\Sigma} = \mathbf{I}_p$, where $\mathbf{I}_p$ is the p-dimensional identity matrix. Let $\mathcal{U}_1(b) = \{\mathbf{\Sigma} : \|\mathbf{\Sigma} - \mathbf{I}_p\|_F \geq b(p/n)^{1/2}\}$ be the covariances

*Authors are in the alphabetical order

†Corresponding author

Keywords and phrases: Covariance matrix, Detection boundary, High dimensionality, Minimax, Rare and faint signal, Thresholding.

at least $b(p/n)^{1/2}$ apart from the identity matrix in the Frobenius norm, which constitutes the alternative hypothesis. They showed that there exist positive constants b_0 and $\omega \in (0,1)$ such that

$$\sup_{W \in \mathcal{W}_{1,\alpha}} \inf_{\mathbf{\Sigma} \in \mathcal{U}_1(b_0)} \mathbb{P}(W=1) \leq 1-\omega, \tag{1.1}$$

as $n, p \to \infty$. This indicates that $(p/n)^{1/2}$ is the minimum rate for the Frobenius distance between $\mathbf{\Sigma}$ and $\mathbf{I}_p$ such that the minimax risk for testing $\mathbf{\Sigma} = \mathbf{I}_p$ may diminish to 0. Cai and Ma (2013) also shows that the L_2-norm based test proposed by Chen et al. (2010) is rate optimal under the class $\mathcal{U}_1(b)$ with power tending to 1 if $b \to \infty$ as $n, p \to \infty$.

The above minimax results provide the rate of the minimum signals that can be detected. However, the results are not sharp as the expressions of b_0 in $\mathcal{U}_1(b_0)$ are unknown. It is of both theoretical and practical value to derive the sharp optimality result which shows how the constant depends on signal strength and sparsity. It is noted that although the aforementioned L_2-type and the $L_{\max}$-type tests are rate optimal under the dense signal and the strong signal regimes, respectively, they may not be sharp optimal, and the minimax results for the most challenging sparse and weak signal regime has not been studied for high dimensional covariance matrix testing.

This study aims at deriving the tight minimax detection boundary for testing a diagonal covariance matrix as a function of the signal sparsity and strength. Our main findings are the followings.

1. We show that for Gaussian data, the optimal detection boundary $\mathrm{DB}(\beta)$ defined in (2.9) for testing mean vectors (Ingster, 1997; Donoho and Jin, 2004) is the minimax detection lower boundary for testing a covariance being diagonal against the sparse and weak alternatives as shown in Theorem 1.
2. A multi-level thresholding (MT) test is proposed for Gaussian data, which is shown to have proper size control (Theorems 2-3) and be able to achieve the minimax detection lower boundary $\mathrm{DB}(\beta)$ over at least 75% of the sparsity range (Corollary 1), and hence, it is sharp optimal over the range.
3. Extend the MT test to sub-Gaussian data (Theorems 4-5, Proposition 5) and establish its detection boundary which coincides with that for the Gaussian case over at least 50% of the sparsity range.

The minimax detection lower boundary for a hypothesis testing problem prescribes a region of signals in terms of the signal sparsity and strength parameters such that no test can distinguish the null and the alternative hypotheses if the signal strength is below the boundary. Here, a test is said to be able to distinguish the null and alternative hypotheses of a testing problem if the sum of the probabilities of committing the types I and II errors diminishes to zero asymptotically. If a test procedure can attain a minimax detection lower boundary as its detection boundary in the sense that the test can distinguish the null and the alternative hypotheses for any combination of signal strength and sparsity above the detection lower boundary, then we say the minimax detection lower boundary is optimal or tight and the test procedure is sharp optimal for the testing problem.

Although Donoho and Jin (2004, 2015) and Qiu et al. (2018) have established the tight detection boundary for testing high dimensional means and regression coefficients, attaining similar results for covariance matrices is more challenging due to the more complex dependence among the entries of the sample covariances. We propose a new method to construct the least favorable prior on the sparse and faint non-diagonal covariances, which leads to the minimax detection lower boundary under the least favorable prior. We develop a test that achieves the detection lower boundary by conducting multi-level thresholding that removes the non-signal bearing entries of the covariance matrix from the test statistic formulation, and reduces the overall noise (variance) of the test statistic. Since the proposed MT test is able to attain the detection lower boundary over a large portion of the sparsity range (depending on the relationship between n and p), the derived minimax detection lower boundary is tight and the proposed test is sharp optimal over this range. Furthermore, we demonstrate that both the L_2 test and $L_{\max}$ test can not achieve this detection boundary.

The paper is organized as follows. We present the sparse and weak signal setting for the high dimensional covariance testing problem, and the minimax detection lower boundary for the Gaussian data in Section 2. Section 3 introduces the multi-level thresholding test after providing the asymptotic distribution of the test statistic. Section 4 reports a power analysis of the MT test and shows the MT test can attain the minimax detection lower boundary over at least 75% of the sparsity range. Extension to the sub-Gaussian data is made in Section 5. Results of simulation studies are reported in Section 6. Technical proofs and other results are given either in Appendix or the supplementary material (SM).

2. Hypotheses and minimax detection boundary. We start with Gaussian data while the extension to non-Gaussian data is made in Section 5. Suppose there is a random sample of independent and identically distributed (IID) random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ drawn from a p-dimensional multivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^{\mathrm{T}}$ and covariance $\boldsymbol{\Sigma} = (\sigma_{j_1 j_2})_{p\times p}$, where $\mathbf{X}_k = (X_{k1}, \dots, X_{kp})^{\mathrm{T}}$ for $k = 1, \dots, n$. Let $\boldsymbol{\Psi} = (\rho_{j_1 j_2})_{p\times p}$ be the correlation matrix of $\boldsymbol{\Sigma}$. We consider testing

$$H_0 : \sigma_{j_1 j_2} = 0 \ \text{ for all } \ j_1 \neq j_2 \quad \text{vs.} \quad H_a : \sigma_{j_1 j_2} \neq 0 \ \text{ for some } \ j_1 \neq j_2 \tag{2.1}$$

under a high dimensional setting where $p \gg n$, as well as a moderate dimensional setting where $p^{\xi} \sim n$ for $\xi \in [1, 2]$. Here, for two real sequences $\{a_n\}$ and $\{b_n\}$, $a_n \sim b_n$ means that there are two positive constants c_1, c_2 such that $c_1 \le a_n/b_n \le c_2$ for all n.

2.1. *Sparse and weak signals.* We want to derive the tight detection boundary for testing the hypotheses (2.1), which is a function of the signal strength and sparsity that separates the signals into the testable and un-testable regions; see Donoho and Jin (2004) and Hall and Jin (2010) for the detection boundary of testing high dimensional means. Intuitively, if the nonzero covariances (signals) are too few and too faint relative to those prescribed by the boundary, no test can distinguish the null and alternative hypotheses of (2.1).

To this end, we construct the least favorable prior on a subset of the alternatives that constitutes the challenging setting with the number of nonzero $\sigma_{j_1 j_2}$ being rare and their magnitude being faint. Let $q = p(p-1)/2$ be the number of upper-diagonal elements in $\boldsymbol{\Sigma}$. Suppose that there are $m_a = \lfloor q^{(1-\beta)} \rfloor$ nonzero $\sigma_{j_1 j_2}$ with $j_1 < j_2$, where $\beta \in (1/2, 1)$ is the sparsity parameter and $\lfloor \cdot \rfloor$ denotes the floor function. We note that $\beta \in (1/2, 1)$ represents the sparse case of signal detection, while that $\beta \in (0, 1/2)$ constitutes the dense signal case where the number of signals m_a is much larger than p. Throughout the paper, we call $(1/2, 1)$ as the sparsity range of the signals. For this sparse case, we consider the strength of signals as

$$|\sigma_{j_1 j_2}| = \sqrt{2 r_{j_1 j_2} \log(q)/n} = \sqrt{4 r_{j_1 j_2} \log(p)/n}\{1 + o(1)\} \ \text{ if } \ \sigma_{j_1 j_2} \neq 0 \tag{2.2}$$

for $r_{j_1 j_2} > 0$. Here, the nonzero covariances can be either negative or positive. Note that, under the covariance class $\mathcal{U}_1(b) = \{\boldsymbol{\Sigma} : \|\boldsymbol{\Sigma} - \mathbf{I}_p\|_F \ge b(p/n)^{1/2}\}$ considered in Cai and Ma (2013), the signal strength for each nonzero covariance is at least at the order $\{p/(nm_a)\}^{1/2} = n^{-1/2} p^{\beta - 1/2}$ if all the signals have the same strength. Under this case, the minimum signal strength under the class $\mathcal{U}_1(b)$ is much larger than $n^{-1/2}$ if $\beta > 1/2$ for the sparse case. Meanwhile, it is smaller than $n^{-1/2}$ if $\beta < 1/2$ for the dense case under $\mathcal{U}_1(b)$. Our setting reflects signals that are both sparse and weak, which are difficult to be tested.

Let $t_p = \max_{1\le j_1\le p}\sum_{j_2\ne j_1}\mathbb{I}(\sigma_{j_1j_2}\ne 0)$ be the maximum number of nonzero off-diagonal values in the rows of $\mathbf{\Sigma}$, where $\mathbb{I}(\cdot)$ is the indicator function. We consider non-structured signals such that any component of the random vector is not correlated with many other components. This implies $t_p \le C_1$ uniformly for all p and a positive constant C_1. For positive constants C_1 and C_2, we consider a class of covariance matrices

$$\mathcal{U}(\beta, r_0, \tau) = \big\{\mathbf{\Sigma}:\ \text{there are } m_a = \lfloor q^{(1-\beta)}\rfloor \text{ nonzero } \sigma_{j_1j_2} \text{ as (2.2) with } r_{j_1j_2}\ge r_0 \\ \text{for } j_1<j_2, \max_{1\le j\le p}\sigma_{jj}\le\tau, \min_{1\le j\le p}\sigma_{jj}\ge C_2 \text{ and } t_p\le C_1\big\}. \tag{2.3}$$

While the hypotheses in (2.1) offer general alternatives against a diagonal covariance, $\mathcal{U}(\beta, r_0, \tau)$ in (2.3) restricts on the number and magnitude of signals under the alternative hypothesis. Here, the signal strength parameters $\{r_{j_1j_2}\}$ together with the sparsity parameter $\beta\in(1/2,1)$ constitute the rare and faint signal setting. Comparing to the class $\mathcal{U}_1(b)$ analyzed in (1.1) based on the Frobenius distance under alternative hypotheses, $\mathcal{U}(\beta, r_0, \tau)$ is more structured that specifies the level of signal strength to facilitate our analysis.

For $\mathbf{\Sigma} = (\sigma_{j_1j_2})\in\mathcal{U}(\beta, r_0, \tau)$, the standardized signal strength to detect nonzero $\sigma_{j_1j_2}$ is

$$\tilde{r}_{j_1j_2} = r_{j_1j_2}/(\sigma_{j_1j_1}\sigma_{j_2j_2}) \text{ for } \sigma_{j_1j_2}\ne 0, \tag{2.4}$$

by recognizing that the numerator is the signal strength level in $|\sigma_{j_1j_2}|$ and the denominator is the main order term of the variance of $\sqrt{n}(\hat{\sigma}_{j_1j_2}-\sigma_{j_1j_2})$ under the Gaussian distribution. Let $\mathcal{A}_1 = \{(j_1, j_2): j_1<j_2, \sigma_{j_1j_2}\ne 0\}$. Define the maximal and minimal standardized signal strength of $\mathbf{\Sigma}$ as

$$\bar{r} = \max_{(j_1,j_2)\in\mathcal{A}_1}\tilde{r}_{j_1j_2} \text{ and } \underline{r} = \min_{(j_1,j_2)\in\mathcal{A}_1}\tilde{r}_{j_1j_2}. \tag{2.5}$$

Since all $r_{j_1j_2}$ are larger than r_0 and τ is the upper bound of $\{\sigma_{jj}\}$ in the class $\mathcal{U}(\beta, r_0, \tau)$, the minimal standardized signal strength over $\mathcal{U}(\beta, r_0, \tau)$ is $\min_{\mathbf{\Sigma}\in\mathcal{U}(\beta,r_0,\tau)}\underline{r} = r_0\tau^{-2}$. Therefore, the minimax detection boundary for testing the hypotheses in (2.1) over $\mathcal{U}(\beta, r_0, \tau)$ is characterized by $r_0\tau^{-2}$.

2.2. *Detection lower boundary.* The key to evaluating the minimax risk is to derive the Bayes risk under the least favorable prior (Lehmann, 1959; Berger, 1985). For that purpose, we introduce graph notations to construct the least favorable prior. Let $[p] = \{1,\ldots,p\}$ be the index set of variables and $\mathcal{G} = (A, U)$ be a graph built on the nonzero covariances (signals) of $\mathbf{\Sigma}$, where the vertex set $A = \{j_1 : j_1\in[p] \text{ and } \sigma_{j_1j_2}\ne 0 \text{ for some } j_2\in[p] \text{ and } j_2\ne j_1\}$ and the edge set $U = \{(j_1, j_2): j_1, j_2\in[p], j_1\ne j_2 \text{ and } \sigma_{j_1j_2}\ne 0\}$ contain the variable indices and pair of variables with nonzero covariances, respectively. The least favorable prior can be specified as follows:

(2.6) the node set A is composed by $2m_a$ vertices selected from $[p]$ uniformly at random,

(2.7) the edge set U is formed by a perfect matching on A uniformly at random,

(2.8) $\sigma_{jj} = \tau$ for all $j\in[p]$ and $\sigma_{j_1j_2} = \sqrt{4r_0\log(p)/n}\,\mathbb{I}\{(j_1, j_2)\in U\}$.

where $\sqrt{4r_0\log(p)/n}$ is the universal minimum signal strength on U under (2.2) and (2.3). Here, a perfect matching means each vertex in A is incident to one and only one edge in U. Random perfect matching means all the vertices in A are randomly paired with each other to form the edges, and no two edges share common vertices. Under (2.6)–(2.8), the covariance matrices have diagonal elements being τ, at most one nonzero off-diagonal element in each row and each column of $\mathbf{\Sigma}$, and random allocations of these nonzero $\sigma_{j_1j_2}$. Note that the least favorable prior is constructed on the covariance matrices with positive off-diagonal values.

Let $\mathcal{C}_{m_a}$ be the collection of all the sets that choose $2m_a$ distinct indices from $[p]$. Let $\mathcal{D}(A)$ be the collections of all edge sets formed by distinct perfect matching of the node set A. It can be shown that the size of $\mathcal{C}_{m_a}$ is $\binom{p}{2m_a}$, and the number of distinct perfect matching of a given node set A is $N_0 = |\mathcal{D}(A)| = (2m_a)!/(m_a!2^{m_a})$. Let $\mathcal{M} = \{\mathcal{G} : \mathcal{G} = (A, U), A \in \mathcal{C}_{m_a}, U \in \mathcal{D}(A)\}$ be the collection of all graphs that satisfy (2.6) and (2.7). Then, the probability of a graph $\mathcal{G}$ being chosen uniformly at random from $\mathcal{M}$ is

$$\mathbb{P}(\mathcal{G} \in \mathcal{M}) = \binom{p}{2m_a}^{-1} N_0^{-1}.$$

Given a graph $\mathcal{G}$, let $Q_{\mathcal{G}}$ be the joint distribution of the data $\{\mathbf{X}_k\}_{k=1}^n$ from the normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}_{\mathcal{G}}$ under the alternative hypothesis, where $\Sigma_{\mathcal{G}}$ is generated by the graph $\mathcal{G}$ and (2.8). Let

$$Q_a = \binom{p}{2m_a}^{-1} N_0^{-1} \sum_{\mathcal{G} \in \mathcal{M}} Q_{\mathcal{G}}$$

be the average measure of $Q_{\mathcal{G}}$ over all graphs in $\mathcal{M}$ with the uniform prior. Essentially, this is the data distribution under the least favorable prior on the covariance class $\mathcal{U}(\beta, r_0, \tau)$, constructed by the random perfect matching of $2m_a$ randomly selected components from $\{1, \dots, p\}$.

Let Q_0 denote the distribution of the data $\{\mathbf{X}_k\}_{k=1}^n$ from the normal distribution with the diagonal covariance $\sigma_{jj} = \tau$ for $j = 1, \dots, p$ under the null hypothesis of (2.1). The essence of deriving the detection lower boundary of (2.1) is to evaluate the Hellinger distance between Q_a and Q_0. This amounts to studying the likelihood ratio as a function of β and r_0. Due to involving the inverse of non-diagonal covariances in Q_a, the analysis of the Hellinger distance in our case is much more challenging than that for the means.

Let $\mathcal{W}_\alpha$ be the collection of all α level tests for the hypotheses in (2.1) under the Gaussian distribution. For any $W \in \mathcal{W}_\alpha$, $W = 1$ stands for the rejection of the null hypothesis. Let

$$\mathrm{DB}(\beta) = \begin{cases} \beta - 1/2 & \text{if } 1/2 < \beta \le 3/4, \\ (1 - \sqrt{1-\beta})^2 & \text{if } 3/4 < \beta < 1, \end{cases} \tag{2.9}$$

which is the optimal detection boundary for testing means (Ingster, 1997; Donoho and Jin, 2004). The following theorem shows that $\mathrm{DB}(\beta)$ is also the detection lower boundary for testing (2.1) in terms of the minimax power of all α level tests over the covariance class $\mathcal{U}(\beta, r_0, \tau)$ in (2.3).

THEOREM 1. *For $\log p = o(n^{1/3})$ and Gaussian distributed data, if $r_0\tau^{-2} < \mathrm{DB}(\beta)$ and $\beta \in (1/2, 1)$,*

$$\sup_{W \in \mathcal{W}_\alpha} \inf_{\boldsymbol{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)} \mathbb{P}(W = 1) \le 1 - \omega \tag{2.10}$$

for any $\omega \in (0, 1 - \alpha)$ as $n, p \to \infty$.

It is noted that Theorem 1 is for the sparse signal setting that $\beta > 1/2$. This theorem indicates that no test can consistently separate H_a from H_0 for the hypotheses (2.1) if $r_0\tau^{-2}$ falls below the detection lower boundary $\mathrm{DB}(\beta)$ at a sparsity level β. In the next section, we construct a sharp optimal test that attains $\mathrm{DB}(\beta)$ as its detection boundary for a major portion of the sparsity range $\beta \in (1/2, 1)$ such that the sum of type I and type II error probabilities of the test will converge to 0 for any $\boldsymbol{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)$ if $r_0\tau^{-2}$ is above the boundary.

3. Thresholding Test. Let $\bar{\mathbf{X}} = \sum_{k=1}^{n} \mathbf{X}_k/n = (\bar{X}_1, \ldots, \bar{X}_p)^{\mathrm{T}}$ be the sample mean and

$$\widehat{\mathbf{\Sigma}} = (\hat{\sigma}_{j_1j_2})_{p\times p} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^{\mathrm{T}}$$

be the sample covariance matrix. Let $\theta_{j_1j_2} = \mathrm{Var}\{(X_{kj_1} - \mu_{j_1})(X_{kj_2} - \mu_{j_2})\} = \sigma_{j_1j_1}\sigma_{j_2j_2} + \sigma_{j_1j_2}^2$. As $\hat{\sigma}_{j_1j_1}\hat{\sigma}_{j_2j_2}/n$ is a consistent estimator for the variance of $\hat{\sigma}_{j_1j_2}$ under the null hypothesis that prescribes $\sigma_{j_1j_2} = 0$, we define a standardized statistic of $\hat{\sigma}_{j_1j_2}$ as

$$M_{j_1j_2} = n\hat{\rho}_{j_1j_2}^2,\ 1 \le j_1 < j_2 \le p, \tag{3.1}$$

where $\hat{\rho}_{j_1j_2} = \hat{\sigma}_{j_1j_2}(\hat{\sigma}_{j_1j_1}\hat{\sigma}_{j_2j_2})^{-1/2}$ is the sample correlation between X_{kj_1} and X_{kj_2}. As the squared sample covariance is used, it can detect both the positive and negative values of the nonzero $\sigma_{j_1j_2}$.

We first conduct thresholding on $\{\hat{\rho}_{j_1j_2}\}$ to filter out the potential signals and then to aggregate the filtered sample correlations in an L_2 formulation. Then, a multi-level thresholding (MT) procedure is applied to enhance the power of the test by utilizing different threshold levels. We will show that the proposed MT test attains the detection lower boundary in Theorem 1 over at least 75% of the sparsity range $\beta \in (1/2, 1)$, and is more powerful than both the L_2 and $L_{\max}$-type tests when the signals are rare and faint.

3.1. *Single level thresholding statistic.* We first introduce the thresholding statistic with a single threshold level. By the moderate deviation result in Lemma S2 of the SM, under Assumptions 1A (or 1B), 2 and H_0 of (2.1), $\mathbb{P}\big\{\max_{1\le j_1<j_2\le p} M_{j_1j_2} > 4\log(p)\big\} \to 0$ as $n, p \to \infty$. This implies that a threshold level of $4\log(p)$ is asymptotically too large under the null hypothesis, and suggests applying a smaller threshold $\lambda_p(s) = 4s\log(p)$ for a thresholding parameter $s \in (0, 1)$. This leads to a thresholding statistic

$$T_n(s) = \sum_{1\le j_1<j_2\le p} M_{j_1j_2}\mathbb{I}\{M_{j_1j_2} \ge \lambda_p(s)\}, \tag{3.2}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Compared with the L_2-statistic of Chen et al. (2010), $T_n(s)$ keeps only large $M_{j_1j_2}$ after filtering out the potentially insignificant sample correlations. By removing those smaller $M_{j_1j_2}$, the variance of $T_n(s)$ is much reduced which translates to a higher power.

We make the following assumptions in our analysis.

ASSUMPTION 1A. As $n \to \infty$, $p \to \infty$, $\log p \sim n^{\varpi}$ for a $\varpi \in (0, 1/5)$.

ASSUMPTION 1B. As $n \to \infty$, $p \to \infty$, $n \sim p^{\xi}$ for a $\xi \in (0, 2]$.

ASSUMPTION 2. Variances are bounded from above and below such that $\sigma_{jj} \le \tau$ and $\sigma_{jj} \ge C_2$ for all $j = 1, \ldots, p$ and positive constants τ and C_2.

Assumptions 1A and 1B specify the exponential and polynomial growth rates of p relative to n, respectively. Assumption 2 prescribes the marginal variances uniformly bounded away from zero and infinity. These two assumptions are common in high dimensional literature. Note that $\xi \in [1, 2]$ under Assumption 1B prescribes a moderate dimensional case where $p \to \infty$ but p is at the same order or smaller than n. The reason to include this setting is to reflect the impact of dimensionality on the detection boundary of the proposed multi-thresholding test in (3.7). The detection boundary of this test matches the minimax detection boundary derived in Theorem 1 over the entire range $\beta \in (1/2, 1)$ of sparse signals under $\xi = 2$. But, it increases with the decrease of ξ as shown in Proposition 3.

Let $\mathrm{E}_0(\cdot)$ and $\mathrm{Var}_0(\cdot)$ denote the mean and variance of the thresholding statistic $T_n(s)$ under H_0, and $\mathrm{E}_a(\cdot)$ and $\mathrm{Var}_a(\cdot)$ denote the mean and variance under the alternative hypothesis in (2.1), respectively. Notice that

$$\mathrm{E}_0\{T_n(s)\} = q\mathrm{E}\big[M_{j_1 j_2}\mathbb{I}\{M_{j_1 j_2} \geq \lambda_p(s)\}\big] = 2nq \int_{\sqrt{\frac{\lambda_p(s)}{n}}}^{1} r^2 f_\rho(r) dr, \tag{3.3}$$

where $f_\rho(r) = (1-r^2)^{\frac{n-4}{2}}\{B(\frac{1}{2}, \frac{n-2}{2})\}^{-1}$ is the density function of $\hat{\rho}_{j_1 j_2}$ under the normal distribution with zero $\sigma_{j_1 j_2}$ (Hotelling, 1953), and $B(\cdot,\cdot)$ is the beta function. Let $\phi(\cdot)$ and $\bar{\Phi}(\cdot)$ be the density and survival functions of $N(0,1)$, respectively. The following proposition provides expansions of $\mathrm{E}_0\{T_n(s)\}$ and $\mathrm{Var}_0\{T_n(s)\}$.

PROPOSITION 1. *Under Assumptions 1A or 1B and Assumption 2 and for Gaussian distributed data and $q = p(p-1)/2$, $\mathrm{E}_0\{T_n(s)\} = \mu(s,p)\{1 + O(\lambda_p^{3/2}(s) n^{-1/2})\}$ where*

$$\mu(s,p) = q\{2\lambda_p^{1/2}(s)\phi(\lambda_p^{1/2}(s)) + 2\bar{\Phi}(\lambda_p^{1/2}(s))\}.$$

In addition, under either (i) Assumption 1A with $s > 1/2$ or (ii) Assumption 1B with $s > 1/2 - \xi/4$, $\mathrm{Var}_0\{T_n(s)\} = \sigma^2(s,p)\{1 + o(1)\}$ where $\sigma^2(s,p) = q[2\{\lambda_p^{3/2}(s) + 3\lambda_p^{1/2}(s)\}\phi(\lambda_p^{1/2}(s)) + 6\bar{\Phi}(\lambda_p^{1/2}(s))]$.

Proposition 1 suggests that the main order of $\mathrm{Var}_0\{T_n(s)\}$ is known and solely determined by p and s, and hence can be used to estimate the null variance of $T_n(s)$. The following theorem shows the asymptotic distribution of $T_n(s)$ at a given s under H_0.

THEOREM 2. *Under the H_0 of (2.1), Assumption 2, the Gaussian distribution and either (i) Assumption 1A with $s > 1/2$ or (ii) Assumption 1B with $s > 1/2 - \xi/4$, we have*

$$\mathrm{Var}_0^{-1/2}\{T_n(s)\}[T_n(s) - \mathrm{E}_0\{T_n(s)\}] \xrightarrow{d} N(0,1) \quad \text{as } n, p \to \infty.$$

Comparing with the thresholding statistics on means (Zhong et al., 2013), thresholding on the sample covariance involves a more complex dependency structure. Although the variables $\{X_{kj}\}_{j=1}^{p}$ are independent under the null hypothesis and Gaussianity, the statistics $\{M_{j_1 j_2}\}$ are dependent and display a kind of circular pattern of dependence due to the nature of sample correlations. For each pair (j_1, j_2), the number of sample correlations that are dependent with $\hat{\rho}_{j_1 j_2}$ is of order p under the null hypothesis. As the covariance between $M_{j_1 j_2}\mathbb{I}\{M_{j_1 j_2} \geq \lambda_p(s)\}$ and $M_{j_1 j_3}\mathbb{I}\{M_{j_1 j_3} \geq \lambda_p(s)\}$ decreases as the threshold level s increases for $j_2 \neq j_3$, to control the dependence, we require a lower bound restriction on the threshold level s in Proposition 1 and Theorem 2. This restriction on s would affect the detection boundary of the thresholding test as discussed in Section 4. Under Assumption 1B that prescribes the polynomial growth of p, the minimum threshold level that guarantees the Gaussian limit of $T_n(s)$ can be chosen as close to 0 as ξ approaches 2.

The circular dependence among $\{M_{j_1 j_2}\}$ also makes the conventional data blocking and the coupling approach (Berbee, 1979) commonly used for analyzing weakly dependent data insufficient to establish the asymptotic distribution of $T_n(s)$. To tackle the issue, we view $T_n(s)$ as a U-statistic with a kernel function of each pair of variables, and derive its asymptotic normality by the martingale central limit theorem. Different from the classical approach on the U-statistics built on the permutation of observations, the martingale is constructed on the σ-field filtration of the variables. The essence is to establish and use the conditional moderate deviation results of $M_{j_1 j_2}$ given either the j_1th or j_2th variable. Details of this approach are provided in the proof of Theorem 2 in the SM.

3.2. *Multi-level thresholding test.* From Theorem 2, a single level thresholding test rejects H_0 of (2.1) if

$$T_n(s) > \mathrm{E}_0\{T_n(s)\} + z_\alpha \sigma(s,p), \tag{3.4}$$

where z_α is the upper α quantile of $N(0,1)$. Clearly, the power of such a test relies on the tuning parameter s. In the proof of Proposition 3, it is shown that the test (3.4) can achieve the minimax detection boundary $\mathrm{DB}(\beta)$ in (2.9) for a range of β at certain threshold level s that depends on the unknown signal strength and sparsity. This means that the single level thresholding test is not universally powerful for all signals with different strength and sparsity under the covariance class $\mathcal{U}(\beta, r_0, \tau)$. This motivates us to conduct the testing procedure with a sequence of threshold levels to suit the situations with unknown signal strength and sparsity levels. Indeed, utilizing many thresholding levels may capture the sparse and faint signals as shown in the tests for means (Donoho and Jin, 2004; Hall and Jin, 2010; Zhong et al., 2013).

Specifically, let $\mathcal{T}_n(s) = [T_n(s) - \mathrm{E}_0\{T_n(s)\}]\sigma^{-1}(s,p)$ be the standardization of $T_n(s)$. For a threshold lower bound s_0 and an arbitrarily small positive constant η, we construct a multi-level thresholding (MT) statistic

$$\mathcal{V}_n(s_0) = \max_{s \in (s_0, 1-\eta]} \mathcal{T}_n(s)$$

by maximizing $\mathcal{T}_n(s)$ over $s \in (s_0, 1-\eta]$. From Theorem 2, a choice of s_0 is either $1/2$ or $1/2 - \xi/4$ depending on p having the exponential or polynomial growth with respect to n.

Let $M_{(1)} \leq \ldots \leq M_{(q)}$ be the ordered $\{M_{j_1 j_2} : 1 \leq j_1 < j_2 \leq p\}$. Note that $\mathrm{E}_0\{T_n(s)\}$ and $\sigma(s,p)$ are decreasing functions of s. For any two consecutive thresholds $s_{(i)} = M_{(i)}/(4\log p)$ and $s_{(i+1)} = M_{(i+1)}/(4\log p)$, $T_n(s)$ remains constant for $s \in (s_{(i)}, s_{(i+1)}]$. Specifically, if $\mathcal{T}_n(s_{(i+1)}) > 0$, we have $\mathcal{T}_n(s) \leq \mathcal{T}_n(s_{(i+1)})$ over $s \in (s_{(i)}, s_{(i+1)}]$; and if $\mathcal{T}_n(s_{(i+1)}) \leq 0$, we have $\mathcal{T}_n(s) \leq 0$ for $s \in (s_{(i)}, s_{(i+1)}]$. This means that the maximum of $\mathcal{T}_n(s)$ over the interval $(s_0, 1-\eta]$ can be attained on a discrete set of thresholds determined by $\{M_{j_1 j_2}\}$. Let

$$\begin{aligned} \mathcal{S}_n(s_0) = \{s_{j_1 j_2} : s_{j_1 j_2} = M_{j_1 j_2}/(4\log p) \ \text{ for all } \ 1 \leq j_1 < j_2 \leq p \\ \text{satisfying } \ s_0 < s_{j_1 j_2} < (1-\eta)\} \cup \{1-\eta\}. \end{aligned} \tag{3.5}$$

The multi-level thresholding statistic $\mathcal{V}_n(s_0)$ can be expressed as

$$\mathcal{V}_n(s_0) = \max_{s \in \mathcal{S}_n(s_0)} \mathcal{T}_n(s). \tag{3.6}$$

The asymptotic distribution of $\mathcal{V}_n(s_0)$ is given in the following theorem.

THEOREM 3. *Under the conditions of Theorem 2 and H_0 of (2.1),*

$$\mathbb{P}\{a(\log(p))\mathcal{V}_n(s_0) - b(\log(p), s_0, \eta) \leq x\} \to \exp(-e^{-x}),$$

where $a(y) = (2\log(y))^{1/2}$ and $b(y, s_0, \eta) = 2\log(y) + 2^{-1}\log\log(y) - 2^{-1}\log(\pi)$ $+ \log(1 - s_0 - \eta)$.

This leads to an asymptotic α-level multi-thresholding test (MT test) that rejects H_0 if

$$\mathcal{V}_n(s_0) > \{q_\alpha + b(\log(p), s_0, \eta)\}/a(\log(p)), \tag{3.7}$$

where q_α is the upper α quantile of the Gumbel distribution. It is noted that the proposed MT test in (3.7) is equivalent to conducting a sequence of the single level thresholding tests with thresholds from $\mathcal{S}_n(s_0)$ and using the minimum p-value of these tests to make a decision for the hypotheses (2.1). Comparing to the normal quantile z_α used in the single level thresholding test in (3.4), a higher rejection cut-off value at the order $\{\log(\log p)\}^{1/2}$ is needed for the MT test as prescribed from Theorem 3.

4. Power analysis and detection boundary of MT test. We evaluate the power performance of the proposed MT test in (3.7), which leads to obtaining its detection boundary under the covariance class $\mathcal{U}(\beta, r_0, \tau)$ in (2.3). Using the detection boundary result, we demonstrate the superiority of the proposed test over the L_2 and $L_{\max}$ type tests.

Under the alternative hypothesis in (2.1), the power of the proposed MT test is

$$\text{Power}_n(\mathbf{\Sigma}) = P\big[\mathcal{V}_n(s_0) > \{q_\alpha + b(\log(p), s_0, \eta)\}/a(\log(p))|\mathbf{\Sigma}\big].$$

Recall that $\text{E}_a(\cdot)$ and $\text{Var}_a(\cdot)$ denote the mean and variance of $T_n(s)$ under H_a in (2.1). Let

$$\text{SNR}(s) = \frac{\text{E}_a\{T_n(s)\} - \text{E}_0\{T_n(s)\}}{\text{Var}_a^{1/2}\{T_n(s)\}}$$

be the signal to noise ratio of $T_n(s)$ under the alternative hypothesis. Since

$$\mathcal{V}_n(s_0) = \max_{s \in \mathcal{S}_n(s_0)} \frac{\text{Var}_a^{1/2}\{T_n(s)\}}{\sigma(s,p)} \left\{ \frac{T_n(s) - \text{E}_a\{T_n(s)\}}{\text{Var}_a^{1/2}\{T_n(s)\}} + \text{SNR}(s) \right\},$$

the power of the MT test is critically determined by $\text{SNR}(s)$. Let $L_p = \bar{c}_1 \log^{\bar{c}_2}(p)$ be a multi-$\log(p)$ term for two positive constants $\bar{c}_1$ and $\bar{c}_2$ whose values may change in the context. Recall that $\tilde{r}_{j_1 j_2}$ is the standardized signal strength of a nonzero $\sigma_{j_1 j_2}$, and $\bar{r}$ and $\underline{r}$ are the maximal and minimal standardized signal strength of $\mathbf{\Sigma}$, defined in (2.4) and (2.5). The following proposition gives the mean and variance of $T_n(s)$ for $\mathbf{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)$ under the alternative hypothesis.

PROPOSITION 2. *Under the Gaussian distribution, Assumptions 1A or 1B, for $\mathbf{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)$ and $\beta \in (1/2, 1)$, $\text{E}_a\{T_n(s)\} = \text{E}_0\{T_n(s)\} + \Delta(s)$ where*

$$\Delta(s) = L_p \sum_{(j_1,j_2)\in\mathcal{A}_1} \mathbb{I}(s < \tilde{r}_{j_1 j_2}) + L_p \sum_{(j_1,j_2)\in\mathcal{A}_1} p^{-2(\sqrt{s}-\sqrt{\tilde{r}_{j_1 j_2}})^2} \mathbb{I}(s > \tilde{r}_{j_1 j_2}).$$

In addition, under either (i) Assumption 1A with $s > 1/2$ or (ii) Assumption 1B with $s > 1/2 - \xi/4$, $\text{Var}_a\{T_n(s)\} = L_p q p^{-2s} + L_p \sum_{(j_1,j_2)\in\mathcal{A}_1} p^{-2(\sqrt{s}-\sqrt{\tilde{r}_{j_1 j_2}})^2} \mathbb{I}(s > \tilde{r}_{j_1 j_2}) + L_p \sum_{(j_1,j_2)\in\mathcal{A}_1} \mathbb{I}(s < \tilde{r}_{j_1 j_2})$.

It is shown in Section S3 of the SM that the power of the proposed MT test can be bounded from below (above) via replacing all $\tilde{r}_{j_1 j_2}$ by the minimum (maximum) standardized signal strength $\underline{r}$ ($\bar{r}$). Together with Proposition 2, we obtain the detection boundary of the proposed MT test in Propositions 3 and 4 below. Define a family of phase transition functions indexed by $\xi \in [0, 2]$ that connects p and n via $n \sim p^\xi$:

$$\text{DB}^*(\beta, \xi) = \begin{cases} \frac{(\sqrt{4-2\xi}-\sqrt{6-8\beta-\xi})^2}{8}, & 1/2 < \beta \leq 5/8 - \xi/16; \\ \beta - 1/2, & 5/8 - \xi/16 < \beta \leq 3/4; \\ (1-\sqrt{1-\beta})^2, & 3/4 < \beta < 1. \end{cases} \tag{4.1}$$

Comparing with the minimax boundary $\text{DB}(\beta)$ given in (2.9), $\text{DB}^*(\beta, \xi)$ is slightly elevated over the sub-interval $\beta \in (1/2, 5/8 - \xi/16)$, which accounts at most 25% of the sparsity range of $\beta \in (1/2, 1)$. This is due to the restriction on the threshold levels that $s > s_0$, which is needed to control the dependence among the entries of the sample covariance matrix so that the asymptotic distribution of the test statistic can be established and the thresholding test can properly control its size.

The following proposition considers the case of $n \sim p^\xi$ for $\xi \in (0, 2]$ as prescribed in Assumption 1B, which mirrors a case considered in Delaigle et al. (2011) for testing means.

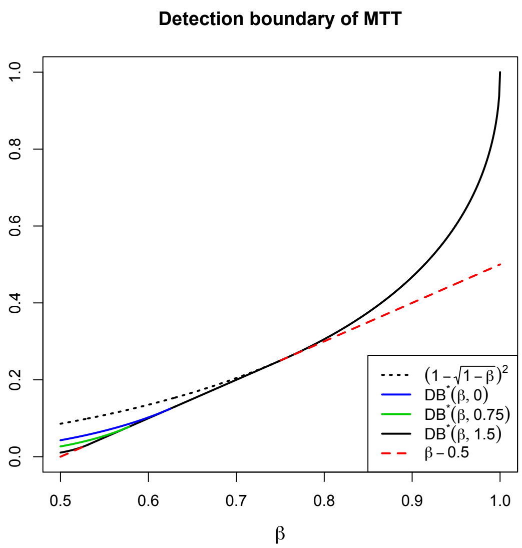


Fig 1: The detection boundary $\mathrm{DB}^*(\beta,\xi)$ in (5.3) of the proposed multi-level thresholding test with $s_0 = 1/2 - \xi/4$ for $\xi = 0, 0.75, 1.5$ and $n = p^{\xi}$.

PROPOSITION 3. *Under Assumptions 1B with the Gaussian data, for* $\mathbf{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)$*,* $\beta \in (1/2, 1)$*,* $s_0 = 1/2 - \xi/4$*, an arbitrarily small* $\epsilon > 0$ *and a series of nominal sizes* $\alpha_n = \bar{\Phi}((\log p)^{\epsilon}) \to 0$*, we have as* $n, p \to \infty$*,*
(i) if $\underline{r} > \mathrm{DB}^*(\beta,\xi)$*, the power of the MT test* $\mathrm{Power}_n(\mathbf{\Sigma})$ *defined in (3.7) converges to 1;*
(ii) if $\bar{r} < \mathrm{DB}^*(\beta,\xi)$*, the power of the MT test* $\mathrm{Power}_n(\mathbf{\Sigma})$ *converges to 0.*

Proposition 3 shows that the power of the proposed MT test is characterized by the signal sparsity β, and the minimum and maximum standardized signal strength. More importantly, $\mathrm{DB}^*(\beta,\xi)$ is the detection boundary of the MT test. The power converges to 1 if $\underline{r}$ is above this boundary, and diminishes to 0 if $\bar{r}$ is below it, while the size of the test approaches zero. The detection boundaries $\mathrm{DB}^*(\beta,\xi)$ are displayed in Figure 1 for three ξ values. Note that $\mathrm{DB}(\beta)$ coincides with $\mathrm{DB}^*(\beta,2)$, the detection boundary of the MT test with $\xi = 2$ (namely, $n \sim p^2$) that corresponds to $s_0 = 0$. Restricting $s > s_0$ elevates the detection boundary $\mathrm{DB}^*(\beta,\xi)$ of the proposed MT test beyond $\mathrm{DB}(\beta)$ over the interval $\beta \in (1/2, 5/8 - \xi/16)$ as a price for controlling the size of the test.

The following proposition shows that $\mathrm{DB}^*(\beta,0)$ is the detection boundary when dimension p grows exponentially fast with n, which can be viewed as a degenerated polynomial growth case with $\xi = 0$.

PROPOSITION 4. *Under the Gaussian distribution and Assumption 1A, for* $\mathbf{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)$*,* $\beta \in (1/2, 1)$*,* $s_0 = 1/2$*, an arbitrarily small* $\epsilon > 0$*, and a series of nominal sizes* $\alpha_n = \bar{\Phi}((\log p)^{\epsilon}) \to 0$*, as* $n, p \to \infty$*, we have*
(i) if $\underline{r} > \mathrm{DB}^*(\beta,0)$*, the power of the MT test* $\mathrm{Power}_n(\mathbf{\Sigma}) \to 1$*;*
(ii) if $\bar{r} < \mathrm{DB}^*(\beta,0)$*, the power of the MT test* $\mathrm{Power}_n(\mathbf{\Sigma}) \to 0$.

As $\mathrm{DB}^*(\beta,0) \geq \mathrm{DB}^*(\beta,\xi)$ for any $\xi \in (0,2]$, the result also shows that a higher growth rate of p leads to a higher detection boundary that may be viewed as a sacrifice of the power due to the higher dimensionality.

Propositions 3 and 4 provide the detection boundary of the MT test for $\mathbf{\Sigma}$ in the class $\mathcal{U}(\beta, r_0, \tau)$. As discussed after (2.5), $\min\{\underline{r} : \mathbf{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)\} = r_0\tau^{-2}$, implying that $r_0\tau^{-2}$

is the minimum standardized signal strength of $\mathbf{\Sigma}$ over all covariances in the class $\mathcal{U}(\beta, r_0, \tau)$. Combining the results in Theorem 1 and Propositions 3 and 4, we have the following corollary showing that the MT test can achieve the minimax detection lower boundary in the sub-interval $\beta \in (5/8 - \xi/16, 1)$. Recall that $\mathcal{W}_\alpha$ is the collection of all α level tests for the hypotheses (2.1) under the Gaussian distribution.

COROLLARY 1. *Under the Gaussian distribution and either (i) Assumption 1A with* $5/8 < \beta < 1$ *or (ii) Assumption 1B with* $5/8 - \xi/16 < \beta < 1$, *as* $n, p \to \infty$,

(i) if $r_0\tau^{-2} > \mathrm{DB}(\beta)$, $\inf_{\mathbf{\Sigma}\in\mathcal{U}(\beta,r_0,\tau)} \mathrm{Power}_n(\mathbf{\Sigma}) \to 1$;

(ii) if $r_0\tau^{-2} < \mathrm{DB}(\beta)$, $\sup_{W\in\mathcal{W}_\alpha} \inf_{\mathbf{\Sigma}\in\mathcal{U}(\beta,r_0,\tau)} \mathbb{P}(W = 1) \le 1-\omega$ *for any* $\omega \in (0, 1-\alpha)$.

Corollary 1 suggests that the proposed MT test attains $\mathrm{DB}(\beta)$ as its detection boundary over the sparsity range $\beta \in (5/8 - \xi/16, 1)$ for both the polynomial and the exponential growth of the dimension p as ξ can be assigned 0 for the exponential growth. As the detection boundary of the MT test reaches the minimax detection lower boundary derived in Theorem 1 for $\beta \in (5/8 - \xi/16, 1)$, the MT test is sharp optimal for testing hypotheses (2.1) against the sparse and weak signals over this range. There is a sparsity range $(1/2, 5/8 - \xi/16)$ of at most width $1/8$ where the sharp optimality of the MT test can not be established for the same reason discussed when we introduce $\mathrm{DB}^*(\beta, \xi)$ in (4.1).

Comparing with the existing L_2-type tests, it is noted that $\sum_{j_1\neq j_2} \sigma^2_{j_1j_2} = c\log(p)q^{1-\beta}/n$ for $\mathbf{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)$ and a constant c. Thus, as shown in Cai and Ma (2013), the L_2-type test do not have power beyond the size of the test for $\beta > 1/2$. For the $L_{\max}$-type test, Cai et al. (2013) showed that the power of the maximum test for the equivalence of two covariances converges to 1 if the standardized signal strength is over 4, which is equivalent to $r_0\tau^{-2} > 4$ in our context. Hence, the signal strength required by the $L_{\max}$ test is much stronger than that established in Propositions 3 and 4 as $\mathrm{DB}^*(\beta, \xi) \in (0, 1)$.

5. Non-Gaussian data. In this section, we extend the MT test to non-Gaussian data. Recall that $\theta_{j_1j_2} = \mathrm{Var}\{(X_{kj_1} - \mu_{j_1})(X_{kj_2} - \mu_{j_2})\}$. As $\theta_{j_1j_2}$ may not equal to $\sigma_{j_1j_1}\sigma_{j_2j_2} + \sigma^2_{j_1j_2}$ for non-Gaussian data, we estimate $\theta_{j_1j_2}$ by

$$\hat{\theta}_{j_1j_2} = \frac{1}{n}\sum_{k=1}^{n}\{(X_{kj_1} - \bar{X}_{j_1})(X_{kj_2} - \bar{X}_{j_2}) - \hat{\sigma}_{j_1j_2}\}^2.$$

Let $\tilde{M}_{j_1j_2} = n\hat{\sigma}^2_{j_1j_2}\hat{\theta}^{-1}_{j_1j_2}$, where $\sqrt{n}\hat{\sigma}_{j_1j_2}\hat{\theta}^{-1/2}_{j_1j_2}$ is the standardization of $\hat{\sigma}_{j_1j_2}$ by $\hat{\theta}_{j_1j_2}$ under H_0 of (2.1). Similar to the thresholding statistic $T_n(s)$ in (3.2), let $\tilde{T}_n(s) = \sum_{1\le j_1<j_2\le p} \tilde{M}_{j_1j_2}\mathbb{I}\{\tilde{M}_{j_1j_2} \ge \lambda_p(s)\}$. We first show the asymptotic normality of $\tilde{T}_n(s)$.

To counter the complex dependence among sample covariances under non-Gaussian data, we assume the variables are weakly dependent after a certain permutation. Let $\{\pi_{\ell,p}\}_{\ell=1}^{p!}$ denote all possible permutations of $\{1, \dots, p\}$ and $\mathbf{X}_k(\pi_{\ell,p})$ be the reordering of $\mathbf{X}_k$ corresponding to a permutation $\pi_{\ell,p}$. We assume that there is a permutation $\pi_{\ell_*,p}$ such that $\mathbf{X}_k(\pi_{\ell_*,p})$ is β-mixing (Bradley, 2005). As the thresholding statistic $\tilde{T}_n(s)$ is invariant to permutations of variables, there is no need to know $\pi_{\ell_*,p}$. It is noted that this condition is not needed for Gaussian data.

Let $\{\mathbf{X}_k\} = \{\mathbf{X}_k(\pi_{\ell_*,p})\}$ to simplify notation, and $\mathcal{F}^{m_b}_{m_a}(\mathbf{X}_k) = \sigma\{X_{kj} : m_a \le j \le m_b\}$ be the σ-fields generated by $\{\mathbf{X}_k\}$ for $1 \le m_a \le m_b \le p$. The β-mixing coefficients are $\zeta_p(h) = \sup_{1\le m\le p-h} \zeta\{\mathcal{F}^m_1(\mathbf{X}_k), \mathcal{F}^p_{m+h}(\mathbf{X}_k)\}$, where for two σ-fields $\mathcal{A}$ and $\mathcal{B}$,

$$\zeta(\mathcal{A}, \mathcal{B}) = \frac{1}{2}\sup\sum_{l_1=1}^{u_1}\sum_{l_2=1}^{u_2}\left|P(A_{l_1} \cap B_{l_2}) - P(A_{l_1})P(B_{l_2})\right|.$$

Here, the supremum is taken over all finite partitions $\{A_{l_1} \in \mathcal{A}\}_{l_1=1}^{u_1}$ and $\{B_{l_2} \in \mathcal{B}\}_{l_2=1}^{u_2}$ of the sample space, and $u_1, u_2 \in \mathbb{Z}^+$, the set of positive integers. Let $\rho_{j_1j_2,j_3j_4} = \mathrm{Cor}\{(X_{kj_1} - \mu_{j_1})(X_{kj_2} - \mu_{j_2}), (X_{kj_3} - \mu_{j_3})(X_{kj_4} - \mu_{j_4})\}$. We make the following assumptions to study thresholding statistics on sample covariances for non-Gaussian data.

ASSUMPTION 3. There exists a small positive constant $\rho_0 < 1$ such that $\frac{\theta_{j_1j_2}}{\sigma_{j_1j_1}\sigma_{j_2j_2}} > \rho_0$ for any $j_1 \neq j_2$, and $|\rho_{j_1j_2,j_3j_4}| < 1 - \rho_0$ for any $(j_1, j_2) \neq (j_3, j_4)$.

ASSUMPTION 4. There exist positive constants η and C such that for all $|t| < \eta$,

$$\mathrm{E}[\exp\{t(X_{kj} - \mu_j)^2\}] \leq C \text{ for } j = 1, \ldots, p.$$

ASSUMPTION 5. There is a permutation $(\pi_{\ell_*,p})$ of the data sequences $\{X_{kj}\}_{j=1}^p$ such that the permuted sequences are β-mixing with the mixing coefficients satisfying $\zeta_p(h) \leq C\gamma^h$ for a constant $\gamma \in (0, 1)$, any $p \in \mathbb{Z}^+$ and positive integer $h \leq p - 1$.

Assumption 3 is a technical condition that prescribes $\theta_{j_1j_2}$ being bounded away from zero, and implies the correlations among $\{\sqrt{n}\hat{\sigma}_{j_1j_2}\hat{\theta}_{j_1j_2}^{-1/2}\}$ being bounded away from 1. Assumption 4 assumes the distributions of $\{X_{kj}\}_{j=1}^p$ are sub-Gaussian which is often assumed in the high dimensional literature (Bickel and Levina, 2008a; Xue et al., 2012; Cai et al., 2013). The β-mixing in Assumption 5 is made for the unknown variable permutation $\pi_{\ell_*,p}$. Similar mixing conditions for the column-wise dependence were made in Delaigle et al. (2011) and Qiu et al. (2018) for thresholding tests of means and regression coefficients. If $\{X_{kj}\}_{j=1}^p$ is a Markov chain (the vector sequence under the variable permutation), Theorem 3.3 in Bradley (2005) provides conditions for the processes being β-mixing. If $\{X_{kj}\}_{j=1}^p$ is a linear process with IID innovations, which include the ARMA process as a special case, then it is β-mixing provided the innovation process is absolutely continuous (Mokkadem, 1988). The β-mixing coefficients are assumed to decay at an exponential rate in Assumption 5 to simplify proofs, while arithmetic rates can be entertained at the expense of more technical details.

As discussed after Theorem 2, even though the data vector is β-mixing under a permutation, the vectorization of $(\tilde{M}_{ij})_{p\times p}$ is not necessarily a mixing sequence due to the circular dependence feature of sample covariances. Thresholding tests for covariances involve a more complex dependency structure than those for time series and spatial data. As non-Gaussian distributed random variables may not be independent under the null hypothesis, the martingale approach used to prove Theorem 2 can not be directly applied here. To tackle these challenges, we first use a combination of matrix blocking, as illustrated in Figure S1 in the SM, and the coupling method, which creates independent blocks of variables. Then, a novel U-statistic representation of the thresholding statistic $\tilde{T}_n(s)$ is constructed based on the independent blocks. This allows the use of the martingale central limit theorem on the U-statistic representation to attain the asymptotic normality of $\tilde{T}_n(s)$.

The following theorem is the non-Gaussian counterpart of Theorem 2.

THEOREM 4. *Suppose Assumptions 2-5 are satisfied. Then, under H_0 of (2.1) and either (i) Assumption 1A with $s > 1/2$ or (ii) Assumption 1B with $s > 1/2 - \xi/4$, we have*

$$[\mathrm{Var}_0\{\tilde{T}_n(s)\}]^{-1/2}[\tilde{T}_n(s) - \mathrm{E}_0\{\tilde{T}_n(s)\}] \xrightarrow{d} N(0, 1) \quad \text{as } n, p \to \infty.$$

It is shown in Section S4.2 of the SM that Proposition 1 still holds under the conditions of Theorem 4, and $\mu(s, p)$ and $\sigma^2(s, p)$ in Proposition 1 are still the main order terms of

$\mathrm{E}_0\{\tilde{T}_n(s)\}$ and $\mathrm{Var}_0\{\tilde{T}_n(s)\}$, respectively. Similar to $\mathcal{V}_n(s_0)$ in (3.6), this implies constructing the analogous multi-level thresholding statistic

$$\tilde{\mathcal{V}}_n(\tilde{s}_0) = \max_{s\in\mathcal{S}_n(\tilde{s}_0)} \tilde{\mathcal{T}}_n(s) \ \text{ for } \ \tilde{\mathcal{T}}_n(s) = \sigma^{-1}(s,p)\{\tilde{T}_n(s) - \mu(s,p)\}, \tag{5.1}$$

where $\tilde{s}_0 = 1 - \xi/2$. According to the expansion for the mean of the thresholding statistic in Proposition S1 of the SM, the lower threshold level $\tilde{s}_0$ is chosen to guarantee

$$\frac{\mathrm{E}_0\{\tilde{T}_n(s)\} - \mu(s,p)}{\sigma(s,p)} = O\{\lambda_p^{5/4}(s)p^{1-s}n^{-1/2}\}, \tag{5.2}$$

which converges to zero under Assumption 1B for $s > 1 - \xi/2$. The threshold bound $\tilde{s}_0$ can be lowered to $1/2 - \xi/4$ in the Gaussian case if a more accurate estimator $\hat{\mathrm{E}}_0\{\tilde{T}_n(s)\}$ can be obtained so that $\hat{\mathrm{E}}_0\{\tilde{T}_n(s)\} - \mathrm{E}_0\{\tilde{T}_n(s)\} = o_p\{\sigma(s,p)\}$. Bootstrap estimators could be constructed which correct the bias in the expansions of $\mathrm{E}_0\{\tilde{T}_n(s)\}$; see Delaigle et al. (2011) which used bootstrap to provide a more accurate estimator for the expectation of the thresholded t-statistics for means.

The following theorem gives the asymptotic distribution of $\tilde{\mathcal{V}}_n(\tilde{s}_0)$ under the sub-Gaussian distribution, which is a counterpart of Theorem 3 for the Gaussian data.

THEOREM 5. *Suppose Assumptions 1B, 2-5 are satisfied, then under H_0 of (2.1), we have*

$$P\{a(\log(p))\tilde{\mathcal{V}}_n(\tilde{s}_0) - b(\log(p), \tilde{s}_0, \eta) \le x\} \to \exp(-e^{-x}).$$

where $\tilde{s}_0 = 1 - \xi/2$, and $a(y)$ and $b(y, \tilde{s}_0, \eta)$ are given in Theorem 3.

Similar to (3.7), the MT test based on $\tilde{\mathcal{V}}_n(\tilde{s}_0)$ rejects H_0 of (2.1) if $\tilde{\mathcal{V}}_n(\tilde{s}_0) > \{q_\alpha + b(\log(p), \tilde{s}_0, \eta)\}/a(\log(p))$. Theorem 5 indicates that the MT test can asymptotically control the size for testing (2.1) for the sub-Gaussian distributed data.

Let $\bar{r}_1 = \max_{(j_1,j_2)\in\mathcal{A}_1}\{r_{j_1j_2}/\theta_{j_1j_2}\}$ and $\underline{r}_1 = \min_{(j_1,j_2)\in\mathcal{A}_1}\{r_{j_1j_2}/\theta_{j_1j_2}\}$ be the maximal and minimal standardized signal strength respectively, where $\mathcal{A}_1 = \{(j_1,j_2) : j_1 < j_2, \sigma_{j_1j_2} \neq 0\}$. The sub-Gaussian counterpart of the phase transition function $\mathrm{DB}^*(\beta,\xi)$ is

$$\mathrm{DB}_1^*(\beta,\xi) = \begin{cases} (\sqrt{1-\xi/2} - \sqrt{1-\beta-\xi/4})^2, & 1/2 < \beta \le 3/4 - \xi/8; \\ \beta - 1/2, & 3/4 - \xi/8 < \beta \le 3/4; \\ (1-\sqrt{1-\beta})^2, & 3/4 < \beta < 1. \end{cases} \tag{5.3}$$

It differs from the minimax detection lower boundary $\mathrm{DB}(\beta)$ over a wider interval $\beta \in (1/2, 3/4 - \xi/8)$ than $(1/2, 5/8 - \xi/16)$ of $\mathrm{DB}^*(\beta,\xi)$ for the Gaussian case, and $\mathrm{DB}_1^*(\beta,\xi) > \mathrm{DB}^*(\beta,\xi)$ for $\beta \in (1/2, 3/4 - \xi/8)$.

The following proposition mirrors Proposition 3 and provides the detection boundary of the MT test based on $\tilde{\mathcal{V}}_n(\tilde{s}_0)$ for the sub-Gaussian data.

PROPOSITION 5. *Under Assumptions 1B, 2-5, for $\mathbf{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)$, $\beta \in (1/2, 1)$ $\tilde{s}_0 = 1 - \xi/2$, an arbitrarily small $\epsilon > 0$, and a series of nominal sizes $\alpha_n = \bar{\Phi}((\log p)^\epsilon) \to 0$, as $n, p \to \infty$, we have*

(i) if $\underline{r}_1 > \mathrm{DB}_1^*(\beta,\xi)$, the power of the MT test converges to 1;

(ii) if $\bar{r}_1 < \mathrm{DB}_1^*(\beta,\xi)$, the power of the MT test converges to 0.

Proposition 5 shows that $\mathrm{DB}_1^*(\beta,\xi)$ is the detection boundary of the MT test based on $\tilde{\mathcal{V}}_n(\tilde{s}_0)$. It is shown in the proof of Proposition 5 that $\mathrm{DB}^*(\beta,\xi) < \mathrm{DB}_1^*(\beta,\xi)$ for $\beta < 3/4 - \xi/8$, which indicates the detection boundary of the MT test under the sub-Gaussian distribution is higher than that under the Gaussian distribution. This is due to the higher threshold bound $\tilde{s}_0 = 1 - \xi/2$ to ensure (5.2) for the size control under the sub-Gaussian distribution. It also reflects a price paid in power when the exact expression of $\mathrm{E}_0\{\tilde{T}_n(s)\}$ is unknown and is estimated by its main order term $\mu(s,p)$ under the sub-Gaussian distribution.

6. Simulation Results. We evaluate the empirical size and power of the proposed MT test and compare it with the L_2 test in Qiu and Chen (2012) (QC) and the $L_{\max}$ test in Cai and Jiang (2011) (CJ) by simulation experiments under high dimensionality.

We generated IID data from a p-dimensional distribution with mean zero and covariance matrix $\mathbf{\Sigma} = (\sigma_{j_1 j_2})_{p\times p}$. We set $\mathbf{\Sigma} = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{jj})$ under the null hypothesis of (2.1). Three settings of $\mathbf{\Sigma}$ were designed under the null hypothesis: (i) $\sigma_{jj} = 1$ for all $j = 1, \ldots, p$, (ii) $\sigma_{jj} \sim \mathrm{Uniform}(1,5)$, and (iii) $\sigma_{jj} \sim \mathrm{Uniform}(1,10)$. Setting (i) was the homogeneous case, while settings (ii) and (iii) were the heterogeneous case where the variances were generated from a super uniform distribution with the minimum value being 1 and the maximum values being 5 and 10, respectively. Once a covariance was generated, it was kept fixed through the simulation. We considered two data generation distributions. One was the multivariate normal distribution where $\mathbf{X}_k \sim N(\mathbf{0}, \mathbf{\Sigma})$ for $k = 1, \ldots, n$. The other one is the uniform distribution where $\{X_{0,kj}\} \overset{\mathrm{IID}}{\sim} \mathrm{Uniform}(-\sqrt{3}, \sqrt{3})$ for $k = 1, \ldots, n$ and $j = 1, \ldots, p$, and $X_{kj} = \sqrt{\sigma_{jj}} X_{0,kj}$. The $\mathrm{Uniform}(-\sqrt{3}, \sqrt{3})$ distribution was chosen so that $\mathrm{E}(X_{0,kj}) = 0$ and $\mathrm{Var}(X_{0,kj}) = 1$. We set $n = 100, 150$ and $p = 500, 1000, 1500$ for size evaluation. According to Theorems 2 and 3, the lower threshold bound s_0 in the MT test was set as 0.5, and η was set to be 0.01. The simulation experiments were replicated 1000 times.

TABLE 1
The empirical sizes of the MT, QC and CJ tests with a nominal level of significance 0.05 for Gaussian and uniform data and the combinations of the sample size n, the dimension p and the covariance matrix $\mathbf{\Sigma}$.

Gaussian distributed data										
		$\sigma_{jj}=1$			$\sigma_{jj} \sim \mathrm{Uniform}(1,5)$			$\sigma_{jj} \sim \mathrm{Uniform}(1,10)$		
n	p	MT	QC	CJ	MT	QC	CJ	MT	QC	CJ
100	500	0.027	0.077	0.005	0.034	0.058	0.005	0.038	0.034	0.010
	1000	0.016	0.041	0.006	0.013	0.056	0.002	0.024	0.057	0.008
	1500	0.018	0.072	0.004	0.013	0.050	0.003	0.011	0.052	0
150	500	0.065	0.060	0.021	0.054	0.041	0.016	0.045	0.038	0.011
	1000	0.049	0.044	0.020	0.044	0.068	0.015	0.046	0.078	0.009
	1500	0.033	0.051	0.014	0.038	0.055	0.016	0.036	0.051	0.012
Uniform distributed data										
		$\sigma_{jj}=1$			$\sigma_{jj} \sim \mathrm{Uniform}(1,5)$			$\sigma_{jj} \sim \mathrm{Uniform}(1,10)$		
n	p	MT	QC	CJ	MT	QC	CJ	MT	QC	CJ
100	500	0.024	0.055	0.017	0.029	0.067	0.015	0.026	0.048	0.020
	1000	0.027	0.054	0.014	0.018	0.064	0.007	0.024	0.072	0.021
	1500	0.019	0.050	0.014	0.018	0.058	0.010	0.020	0.034	0.014
150	500	0.058	0.060	0.026	0.044	0.055	0.027	0.028	0.047	0.015
	1000	0.025	0.051	0.006	0.041	0.054	0.021	0.026	0.046	0.010
	1500	0.028	0.055	0.015	0.019	0.054	0.007	0.029	0.044	0.017

Table 1 reports the empirical sizes of the proposed MT test, and the L_2 QC test and the $L_{\max}$ CJ test at the nominal level 0.05 under different combinations of n, p, $\mathbf{\Sigma}$ and data generation distribution. From the table, we observe that the MT test could control the size around the nominal level in all the cases. It had a consistently good performance under both the Gaussian and uniform distributions and the three covariance structures considered.

Although the size of the MT test was a little conservative with the increase of p, it became closer to 0.05 for the larger sample size $n = 150$. This lent empirical support for the theoretic results in Theorems 2–5 under the null hypothesis. At the same time, the QC test also had reasonable sizes, but was slightly liberal. The maximum CJ test was rather conservative, exhibiting some size distortion even for the larger sample size.

To evaluate and compare powers under the alternative hypothesis, we set $\rho_{jj+1} = \rho_1$ and $\sigma_{jj+1} = \rho_1(\sigma_{jj}\sigma_{j+1j+1})^{1/2}$ for $j = 1, \dots, m_a$ and kept all other values of $\mathbf{\Sigma}$ the same as those under H_0, where m_a denotes the number of nonzero off-diagonal elements of $\mathbf{\Sigma}$. We considered three settings $m_a = p/2, p/5$ and $p/10$ for the number of signals. Note that $q^{(1-\beta)} \approx p/\sqrt{2}$ for $\beta = 0.5$, where $q = p(p-1)/2$. The settings of m_a corresponded to $\beta > 0.5$, representing the sparse signal regime for power evaluation as discussed in Sections 2 and 4. The correlation ρ_1 serves as a measure for signal strength, which ranged from 0.15 to 0.3 by an increment 0.03. Under this setting, there were m_a nonzero correlations with common strength ρ_1 on the first off-diagonal of $\mathbf{\Sigma}$. Although signals could appear on different off-diagonals under the alternative hypothesis, only adding signals on the first off-diagonal easily guarantees the positive definiteness of $\mathbf{\Sigma}$. Due to the limited space, we only report the empirical powers of the three tests under the setting $\sigma_{jj} \sim \text{Uniform}(1, 5)$ for the diagonal values of the covariance matrix.

Figure 2 displays the empirical powers of the three tests with respect to the correlation ρ_1 for signal strength under three levels of signal sparsity. The results in Figure 2 suggest that the MT test had the best power among the three tests under most of the settings, and had superior power over the maximal test CJ in all cases. Under $m_a = p/2$ which was the setting with the most signals, the L_2 test QC had a higher power than the MT test for the smaller correlations $\rho_1 \le 0.18$ under $n = 100$. But, the power of the MT test surpassed that of QC with the increase of ρ_1, and quickly reached levels close to 1. Except for the smallest correlation $\rho_1 = 0.15$, the MT test also became more powerful when n was increased to 150. The MT test was more powerful than the QC test for the sparser signal settings of $m_a = p/5$ and $p/10$. Under those two cases, although the powers of the three tests were similar when ρ_1 was small, the MT's power quickly surpassed the other two when $\rho_1 \ge 0.27$ under $n = 100$ and $\rho_1 \ge 0.21$ under $n = 150$. Compared to QC, the extra power advantage of the MT test became larger as the correlation ρ_1 was increased. Meanwhile, the maximal test CJ had much subdued power for small signal strength when $\rho_1 \le 0.24$ under $n = 100$. Under $n = 150$, its power became comparable to that of the MT test for the cases of larger signal strength with $\rho_1 = 0.27$ and 0.3 under $m_a = p/2$ and $\rho_1 = 0.3$ under $m_a = p/5$ and $p/10$.

Comparing the three settings of m_a, QC's power reached 1 for the denser signal setting of $m_a = p/2$. However, its power declined quickly for $m_a = p/5$ and $p/10$, which confirmed that the L_2 test has reduced power for sparse signals due to incorporating all the non-informative components in its formulation that lowered the signal to noise ratio. It is noted that, under $m_a = p/10$, there were only 50, 100 and 150 nonzero entries among a total of about 124 thousands, half million and one million unique off-diagonal entries of $\mathbf{\Sigma}$ for $p = 500$, 1000 and 1500, respectively. The corresponding proportions of signals in $\mathbf{\Sigma}$ were fairly small, around 0.04%, 0.02% and 0.013% for the three covariance matrices. The power of QC was better than that of CJ when the correlation ρ_1 was small. However, for larger ρ_1, CJ gradually became more powerful than QC. This reflected the fact that the L_2 and $L_{\max}$ tests work under different types of signal regimes.

We also compared the power performance of the three tests under the dense signal regime in Figure 3, where $m_a = 2p$ which corresponds to the sparsity parameter β less than 0.5 and the correlation ρ_1 for signal strength ranged from 0.09 to 0.21 by an increment 0.03. Since the number of signals was larger than p under this case, besides the first off-diagonal, those nonzero covariances were allocated in the second and third off-diagonals of $\mathbf{\Sigma}$. From Figure

Fig 2: Empirical powers of the proposed multi-level thresholding (MT) test (red circle line), the L_2 test of Qiu and Chen (2012) (QC, green triangle line) and the $L_{\max}$ test of Cai and Jiang (2011) (CJ, blue square line) with respect to the correlation ρ_1 (as a measure of signal strength) under three settings $m_a = p/2, p/5$ and $p/10$ for the number of signals, normal (upper two panels) and uniform (lower two panels) distributed data, the sample size $n = 100, 150$, the dimension $p = 500, 1000, 1500$ (reflected by different line types) and $\sigma_{jj} \sim \text{Uniform}(1, 5)$. The black dashed lines indicate the nominal level 0.05.

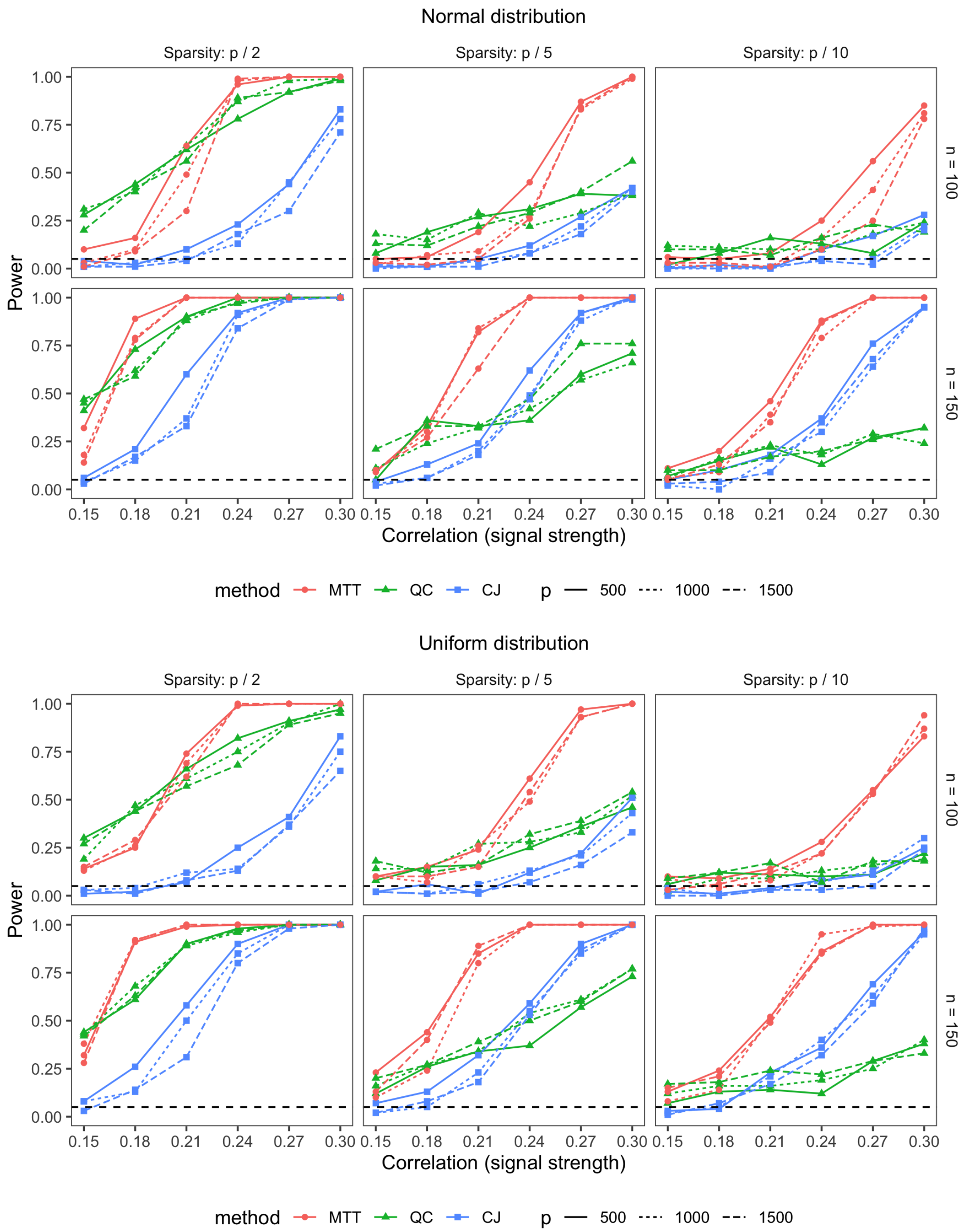
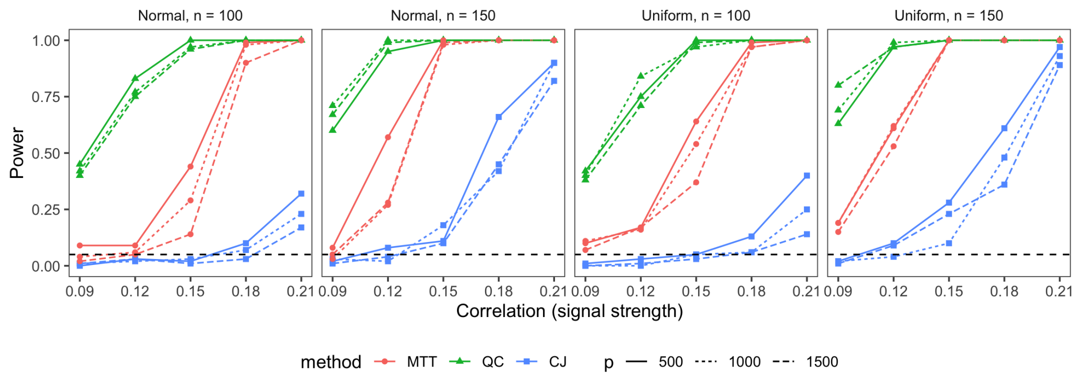


Fig 3: Empirical powers of the proposed multi-level thresholding test (MT test, red circle line), the L_2 test of Qiu and Chen (2012) (QC, green triangle line) and the $L_{\max}$ test of Cai and Jiang (2011) (CJ, blue square line) with respect to the correlation ρ_1 (as a measure of signal strength) under a dense signal setting $m_a = 2p$ for the number of signals, normal and uniform distributed data, the sample size $n = 100, 150$, the dimension $p = 500, 1000, 1500$ (reflected by different line types) and $\sigma_{jj} \sim$ Uniform$(1, 5)$. The black dashed lines indicate the nominal size of 0.05.



3, the QC test had the highest power and the CJ test had the lowest power under this case. The power of the MT test reached to levels close to 1 when $\rho_1 \geq 0.18$ under $n = 100$ and when $\rho_1 \geq 0.15$ under $n = 150$.

In summary, the MT test showed superior performance under different signal sparsity and strength in the sparse and weak signal regime. The simulation results supported the theoretical result that the proposed test has attractive power properties and attains an attractive detection boundary.

APPENDIX A: APPENDIX

In this Appendix, we provide the proof of Theorem 1 for the minimax detection lower boundary. The theoretical proofs for all other propositions and theorems are relegated to the supplementary material. Without loss of generality, we assume $\boldsymbol{\mu} = \mathrm{E}(\mathbf{X}_1) = 0$. Let C be a positive constant and $L_p = \bar{c}_1 \log^{\bar{c}_2}(p)$ be a multi-log(p) term, which may change from case to case, where $\bar{c}_1$ and $\bar{c}_2$ are two positive constants. In the proofs, to simplify notations, we use subscripts (i, j) to denote the pair made by the ith and jth variables.

Proof of Theorem 1. Recall that $\mathcal{G} = (A, U)$ is the graph of the nonzero covariances, $\mathcal{M}$ is the set of all graphs satisfying (2.6) and (2.7), $Q_{\mathcal{G}}$ is the data distribution under the graph $\mathcal{G}$ and (2.8), and Q_0 is the data distribution corresponding to the diagonal covariance with $\sigma_{jj} = \tau$ for all j. Recall that $\mathcal{W}_\alpha$ is the collection of the rejection functions of all α level tests for the hypotheses (2.1), and $\mathbf{D} = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$ is the diagonal matrix of $\boldsymbol{\Sigma}$, where for $W \in \mathcal{W}_\alpha$, $W = 1$ stands for the rejection of the null hypothesis of (2.1). Let $\mathrm{E}_{\boldsymbol{\Sigma}}$ be the expectation with respect to the data under the normal distribution with mean zero and covariance $\boldsymbol{\Sigma}$. Let E_{Q_0} and $\mathrm{E}_{Q_{\mathcal{G}}}$ be the expectation with respect to the data distributions Q_0 and $Q_{\mathcal{G}}$ under the null and alternative hypotheses of (2.1), respectively. Note that

$$1 + \alpha - \sup_{W \in \mathcal{W}_\alpha} \inf_{\boldsymbol{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)} \mathrm{E}_{\boldsymbol{\Sigma}}(W) \geq \inf_{W \in \mathcal{W}_\alpha} \sup_{\boldsymbol{\Sigma} \in \mathcal{U}(\beta, r_0, \tau)} \{\mathrm{E}_{\mathbf{D}} W + \mathrm{E}_{\boldsymbol{\Sigma}}(1 - W)\}$$

$$\geq \inf_{W \in \mathcal{W}_\alpha} \sup_{\mathcal{G} \in \mathcal{M}} \{\mathrm{E}_{Q_0} W + \mathrm{E}_{Q_{\mathcal{G}}}(1 - W)\},$$

and

$$\inf_{W\in\mathcal{W}_\alpha}\binom{p}{2m_a}^{-1}N_0^{-1}\sum_{\mathcal{G}\in\mathcal{M}}\{\mathrm{E}_{Q_0}W+\mathrm{E}_{Q_\mathcal{G}}(1-W)\}=\inf_{W\in\mathcal{W}_\alpha}\{\mathrm{E}_{Q_0}W+\mathrm{E}_{Q_a}(1-W)\}.$$

Thus, the minimax risk is larger than the Bayesian risk with the least favorable prior Q_a.

Let B be a rejection region on the sample space. The optimal sum of Type I and Type II error probabilities under Q_0 and Q_a is defined as

$$\mathcal{E}(Q_0,Q_a)=\inf_B\{Q_0(B)+Q_a(B^c)\},$$

which can be expressed in terms of the total variation distance as $\mathcal{E}(Q_0,Q_a)=1-\mathrm{TV}(Q_0,Q_a)$, where

$$\mathrm{TV}(Q_0,Q_a)=\sup_B|Q_0(B)-Q_a(B)|=\frac{1}{2}\int|dQ_0-dQ_a|.$$

Therefore, we have

$$1+\alpha-\sup_{W\in\mathcal{W}_\alpha}\inf_{\mathbf{\Sigma}\in\mathcal{U}(\beta,r_0,\tau)}\mathrm{E}_{\mathbf{\Sigma}}(W)\geq\inf_{W\in\mathcal{W}_\alpha}\{\mathrm{E}_{Q_0}W+\mathrm{E}_{Q_a}(1-W)\}\geq1-\mathrm{TV}(Q_0,Q_a),$$

which implies

$$\sup_{W\in\mathcal{W}_\alpha}\inf_{\mathbf{\Sigma}\in\mathcal{U}(\beta,r_0,\tau)}P(W=1)\leq\alpha+\mathrm{TV}(Q_0,Q_a).$$

To prove Theorem 1, it suffices to show $\mathrm{TV}(Q_0,Q_a)\to0$ as $n,p\to\infty$, if $r_0\tau^{-2}<\mathrm{DB}(\beta)$.

It is known that the total variation distance is connected with the Hellinger distance between Q_0 and Q_a,

$$H^2(Q_0,Q_a)=\int(\sqrt{dQ_0}-\sqrt{dQ_a})^2=\int\left(\sqrt{\frac{dQ_a}{dQ_0}}-1\right)^2dQ_0. \tag{A.1}$$

Their relationship is expressed via the following inequality

$$H^2(Q_0,Q_a)/2\leq\mathrm{TV}(Q_0,Q_a)\leq H(Q_0,Q_a)\sqrt{1-H^2(Q_0,Q_a)/4},$$

from which we see that $\mathrm{TV}(Q_0,Q_a)\to0$ if and only if $H^2(Q_0,Q_a)\to0$, meaning that no test can distinguish all the covariances in the class $\mathcal{U}(\beta,r_0,\tau)$ under the alternative hypothesis from the covariance under the null hypothesis of (2.1) if the Hellinger distance between Q_0 and Q_a diminishes to zero.

Let $\mathbf{Z}_k=\mathbf{D}^{-1/2}(\mathbf{X}_k-\boldsymbol{\mu})$ be a standardization of $\mathbf{X}_k$, where $\mathbf{Z}_k=(Z_{k1},\ldots,Z_{kp})^{\mathrm{T}}$. Let $\tilde{Q}_a$ be the distribution of the standardized data $\{\mathbf{Z}_k\}_{k=1}^n$ under Q_a, and $\tilde{Q}_0$ be the distribution of n independent $N(\mathbf{0},\mathbf{I}_p)$. By a change of the probability measure, it can be shown that

$$H^2(Q_0,Q_a)=\int\left(\sqrt{\frac{d\tilde{Q}_a}{d\tilde{Q}_0}}-1\right)^2d\tilde{Q}_0=\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}(\sqrt{L}-1)^2, \tag{A.2}$$

where $L=d\tilde{Q}_a/d\tilde{Q}_0$ is the likelihood ratio comparing the distribution $\tilde{Q}_a$ with np independent standard normal distributions, and $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}(\cdot)$ denotes the expectation under $N(\mathbf{0},\mathbf{I}_p)$. Let $\delta_a=\sqrt{4r_0\tau^{-2}\log(p)/n}$. Note that, for a given graph $\mathcal{G}=(A,U)$, $\mathrm{Cov}(Z_{ki},Z_{kj})=\delta_a$ if $(i,j)\in U$; otherwise, $\mathrm{Cov}(Z_{ki},Z_{kj})=0$ under (2.8). Therefore, the standardized data $\mathbf{Z}_k$ under the graph $\mathcal{G}=(A,U)$ and (2.8) can be represented by a mixture of the standard normal distributions as

$$\begin{gathered}Z_{ki_h}=\sqrt{\delta_a}V_{kh}+\sqrt{1-\delta_a}\tilde{V}_{k\,i_h},\ Z_{kj_h}=\sqrt{\delta_a}V_{kh}+\sqrt{1-\delta_a}\tilde{V}_{k\,j_h},h=1,\ldots,m_a,\\ Z_{kj}=\tilde{V}_{kj}\ \text{ for }\ j\notin A,\end{gathered} \tag{A.3}$$

where $\{(i_1, j_1), \dots, (i_{m_a}, j_{m_a})\}$ are all edges in U, and $V_{k1}, \dots, V_{km_a}, \tilde{V}_{k1}, \dots, \tilde{V}_{kp}$ are independent $N(0,1)$ random variables.

To evaluate the likelihood ratio L, we utilize the representation (A.3) for $\mathbf{Z}_k$. Let $f(\mathbf{z}; \mathbf{I}_p)$ be the density function of $N(\mathbf{0}, \mathbf{I}_p)$, and $f_a(\{\mathbf{Z}_k\}; \mathcal{G})$ and $f_a(\{\mathbf{Z}_k\}|\{V_{kh}\}; \mathcal{G})$ be the density function of $\{\mathbf{Z}_k\}_{k=1}^n$ and the conditional density function of $\{\mathbf{Z}_k\}_{k=1}^n$ given $\{V_{kh}\}$ in (A.3) under the graph $\mathcal{G}$, respectively. Let $f_0(\{\mathbf{Z}_k\}) = \prod_{k=1}^n f(\mathbf{Z}_k; \mathbf{I}_p)$ be the density function of $\{\mathbf{Z}_k\}_{k=1}^n$ under the $N(\mathbf{0}, \mathbf{I}_p)$ distribution. Let $L(\mathcal{G}) = f_a(\{\mathbf{Z}_k\}; \mathcal{G}) / f_0(\{\mathbf{Z}_k\})$. Since

$$f_a(\{\mathbf{Z}_k\}|\{V_{kh}\}; \mathcal{G}) = \prod_{i \in A^c} (2\pi)^{-\frac{n}{2}} \exp\biggl(-\frac{1}{2} \sum_{k=1}^n Z_{ki}^2 \biggr) \prod_{h=1}^{m_a} \{2\pi(1-\delta_a)\}^{-n}$$
$$\exp\biggl\{ -\frac{\sum_{k=1}^n (Z_{ki_h} - \sqrt{\delta_a} V_{kh})^2 + (Z_{kj_h} - \sqrt{\delta_a} V_{kh})^2}{2(1-\delta_a)} \biggr\},$$

then

$$\frac{f_a(\{\mathbf{Z}_k\}|\{V_{kh}\}; \mathcal{G})}{f_0(\{\mathbf{Z}_k\})} = \prod_{h=1}^{m_a} (1-\delta_a)^{-n} \exp\biggl\{ -\frac{\delta_a}{2(1-\delta_a)} \sum_{k=1}^n (Z_{ki_h}^2 + Z_{kj_h}^2) - \frac{\delta_a}{1-\delta_a} \sum_{k=1}^n V_{kh}^2$$
$$+ \frac{\sqrt{\delta_a}}{1-\delta_a} \sum_{k=1}^n (Z_{ki_h} + Z_{kj_h}) V_{kh} \biggr\}.$$

Let $\mathbf{V}_h = (V_{1h}, \dots, V_{nh})^{\mathrm{T}}$. Through the derivation, we note the fact that

$$\int (2\pi)^{\frac{n}{2}} \exp\biggl\{ -\frac{\delta_a \sum_{k=1}^n V_{kh}^2}{1-\delta_a} - \frac{\sum_{k=1}^n V_{kh}^2}{2} + \frac{\sqrt{\delta_a}}{1-\delta_a} \sum_{k=1}^n (Z_{ki_h} + Z_{kj_h}) V_{kh} \biggr\} d\mathbf{V}_h$$
$$= \biggl(\frac{1-\delta_a}{1+\delta_a} \biggr)^{\frac{n}{2}} \exp\biggl\{ \frac{\delta_a \sum_{k=1}^n (Z_{ki_h} + Z_{kj_h})^2}{2(1-\delta_a)(1+\delta_a)} \biggr\},$$

which implies the unconditional likelihood ratio function $L(\mathcal{G})$ given $\mathcal{G} = (A, U)$ as

$$L(\mathcal{G}) = \exp\biggl\{ -\sum_{k=1}^n \sum_{(i,j)\in U} \frac{\delta_a(\delta_a Z_{ki}^2 + \delta_a Z_{kj}^2 - 2 Z_{ki} Z_{kj})}{2(1-\delta_a)(1+\delta_a)} \biggr\} (1-\delta_a^2)^{-\frac{n m_a}{2}}$$

$$= \exp\biggl\{ \widetilde{C} \sum_{(i,j)\in U} \widetilde{R}_{ij} - m_a \widetilde{B} \biggr\}, \tag{A.4}$$

where, for each pair of indices $i, j \in [p]$ and $i \neq j$,

$$\widetilde{R}_{ij} = -\frac{\sum_{k=1}^n (\delta_a Z_{ki}^2 + \delta_a Z_{kj}^2 - 2 Z_{ki} Z_{kj} - 2\delta_a)}{\sqrt{4n(1+\delta_a^2)}},$$

$$\widetilde{B} = \frac{n\delta_a^2}{1-\delta_a^2} + \frac{n}{2} \log(1-\delta_a^2) \ \text{ and } \ \widetilde{C} = \frac{\sqrt{n}\delta_a \sqrt{1+\delta_a^2}}{1-\delta_a^2}.$$

Using (A.4), the likelihood ratio $L = d\tilde{Q}_a / d\tilde{Q}_0$ can be expressed as

$$L = \binom{p}{2m_a}^{-1} N_0^{-1} \sum_{\mathcal{G}\in\mathcal{M}} L(\mathcal{G}) = \binom{p}{2m_a}^{-1} N_0^{-1} \sum_{\mathcal{G}\in\mathcal{M}} \exp\biggl\{ \widetilde{C} \sum_{(i,j)\in U} \widetilde{R}_{ij} - m_a \widetilde{B} \biggr\}. \tag{A.5}$$

Let $H^2 = H^2(Q_0, Q_a) = \mathrm{E}_{N(\mathbf{0}, \mathbf{I}_p)} (\sqrt{L} - 1)^2$, and $W = W(Q_0, Q_a) = \mathrm{E}_{N(\mathbf{0}, \mathbf{I}_p)}(\sqrt{L})$ be the Hellinger affinity. It is clear to see that $H^2 = 2(1-W)$, which means that $H^2 \to 0$ is equivalent to $W \to 1$ as $n, p \to \infty$. Directly calculating the first and second moments of L can only show that the Hellinger distance converges to 0 if $r_0\tau^{-2} < \beta - 1/2$ for $\beta \in (1/2, 1)$. Note that this result is suboptimal since $\beta - 1/2 < (1 - \sqrt{1-\beta})^2$ when $3/4 < \beta < 1$. The suboptimality is because the function $\exp(2\widetilde{C}\widetilde{R}_{ij})$ in L^2 diverges exponentially fast when $\widetilde{R}_{ij}$ is large. See the derivations in the paragraph after (A.13) for details.

Therefore, to show $H^2 \to 0$ if $r_0\tau^{-2} < \mathrm{DB}(\beta)$, we consider the moments of a truncated likelihood ratio. Let $D_p = \{\widetilde{R}_{ij} \le \sqrt{4\log p}$ for all pairs of $i, j \in [p]$ and $i < j\}$. Since there are $q = p(p-1)/2$ distinct pairs of variables, by the moderate deviation results on $\widetilde{R}_{ij}$, it can be shown that $P(D_p^c) = o(1)$, and by Cauchy-Schwarz inequality, $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L^{1/2}\mathbb{I}(D_p^c)\} = o(1)$. Let $\widetilde{W} = \mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{\sqrt{L}\mathbb{I}(D_p)\}$, we have

$$\widetilde{W} = W - \mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{\sqrt{L}\mathbb{I}(D_p^c)\} = W + o(1).$$

Now, to prove Theorem 1, it suffices to show $\widetilde{W} \to 1$ if $r_0\tau^{-2} < \mathrm{DB}(\beta)$. Since

$$|\sqrt{L}\mathbb{I}(D_p) - 1| \le |L\mathbb{I}(D_p) - 1|,$$

by Cauchy-Schwarz inequality, we have that

$$(\widetilde{W} - 1)^2 \le \mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L\mathbb{I}(D_p) - 1\}^2 = \mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L^2\mathbb{I}(D_p)\} - 2\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L\mathbb{I}(D_p)\} + 1.$$

Therefore, to show $\widetilde{W} \to 1$ if $r_0\tau^{-2} < \mathrm{DB}(\beta)$, it is sufficient to prove that

$$\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L\mathbb{I}(D_p)\} = 1 + o(1) \text{ and } \mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L^2\mathbb{I}(D_p)\} = 1 + o(1). \tag{A.6}$$

From Lemma S12 in the SM, we have

$$\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)} \exp\big(\widetilde{C}\widetilde{R}_{ij} - \widetilde{B}\big) = 1 \text{ and } \mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)} \exp\big(2\widetilde{C}\widetilde{R}_{ij} - 2\widetilde{B}\big) = (1 - \delta_a^2)^{-n} \tag{A.7}$$

for any $1 \le i, j \le p$ and $i \ne j$. To show the first claim of (A.6), noting that $\mathbb{I}(D_p^c) \le \sum_{i<j}\mathbb{I}(\widetilde{R}_{ij} > \sqrt{4\log p})$, given each graph $\mathcal{G} \in \mathcal{M}$ and $\mathcal{G} = (A, U)$, we have that

$$\begin{aligned} &\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L(\mathcal{G})\mathbb{I}(D_p^c)\} \\ \text{(A.8)}\qquad &\le \sum_{i_1<j_1} \mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\Big[\exp\Big\{\widetilde{C}\sum_{(i_2,j_2)\in U}\widetilde{R}_{i_2j_2} - m_a\widetilde{B}\Big\}\mathbb{I}(\widetilde{R}_{i_1j_1} > \sqrt{4\log p})\Big]. \end{aligned}$$

Let $\tilde{r}_0 = r_0\tau^{-2}$, then $\delta_a = \sqrt{4\tilde{r}_0\log(p)/n}$. Through some calculation, it can be shown that $\widetilde{B} = 2\tilde{r}_0\log(p)\{1 + o(1)\}$ and $\widetilde{C} = 2\sqrt{\tilde{r}_0\log(p)}\{1 + o(1)\}$. By Lemma S14 in the SM and Lemma A.10 in Hall and Jin (2010), we have

$$\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\big\{\exp\big(\widetilde{C}\widetilde{R}_{i_1j_1} - \widetilde{B}\big)\mathbb{I}(\widetilde{R}_{i_1j_1} > \sqrt{4\log p})\big\} \le Cp^{-2(1-\sqrt{\tilde{r}_0})^2} \tag{A.9}$$

for a positive constant C. Notice that for mutually distinct i_1, i_2, j_1, j_2,

$$\begin{aligned} &\mathrm{Cov}(\widetilde{R}_{i_1j_2}, \widetilde{R}_{i_1j_1}) = \delta_a^2(2 + 2\delta_a^2)^{-1} = O\{\log(p)/n\} \text{ and} \\ &\mathrm{Cov}(\widetilde{R}_{i_1j_2} + \widetilde{R}_{i_2j_1}, \widetilde{R}_{i_1j_1}) = O\{\log(p)/n\}. \end{aligned}$$

By Lemma S15 in the SM and Lemma A.10 in Hall and Jin (2010), it can be shown that for mutually distinct i_1, i_2, j_1, j_2,

$$\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\big\{\exp\big(\widetilde{C}\widetilde{R}_{i_1j_2} - \widetilde{B}\big)\mathbb{I}(\widetilde{R}_{i_1j_1} > \sqrt{4\log p})\big\} \le Cp^{-2(1-L_p/n)^2} \text{ and} \tag{A.10}$$

$$\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\big[\exp\big\{\widetilde{C}(\widetilde{R}_{i_1j_2} + \widetilde{R}_{i_2j_1}) - 2\widetilde{B}\big\}\mathbb{I}(\widetilde{R}_{i_1j_1} > \sqrt{4\log p})\big] \le Cp^{-2(1-L_p/n)^2} \tag{A.11}$$

where C is a positive constant and L_p is a multi-$\log p$ term.

Based on the inequalities (A.9)–(A.11), the summation over (i_1, j_1) on the right side of (A.8) can be decomposed into four cases: (a) neither i_1 and j_1 are in the vertex set A; (b) only one i_1 or j_1 is in A; (c) both i_1 and j_1 are in A, and they belong to the edge set U such that $(i_1, j_1) \in U$; and (d) both i_1 and j_1 are in A, but $(i_1, j_1) \notin U$. The number of terms in case (a) is at the order $(p - 2m_a)^2$. By (A.7) and the independence between the pairs within the graph $\mathcal{G}$ and outside $\mathcal{G}$, each of the terms in case (a) is bounded by $P(\widetilde{R}_{i_1 j_1} > \sqrt{4 \log p}) = o(p^{-2})$. There are $2m_a(p - 2m_a)$ terms in case (b), and each term is smaller than $Cp^{-2(1-L_p/n)^2} \approx Cp^{-2}$ by (A.10). Case (c) has m_a terms, and each of them is bounded by the order $p^{-2(1-\sqrt{\tilde{r}_0})^2}$ according to (A.9). The number of terms in the last case (d) is less than $4m_a^2$, and by (A.11), each of those terms is also bounded by $Cp^{-2(1-L_p/n)^2}$ as those in case (b). Summing all the terms in the four cases together, we have that

$$\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L(\mathcal{G})\mathbb{I}(D_p^c)\} \le Cp^{2\{1-\beta-(1-\sqrt{\tilde{r}_0})^2\}} + o(1)$$

for all $\mathcal{G} \in \mathcal{M}$, which implies $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L\mathbb{I}(D_p^c)\} \le Cp^{2\{1-\beta-(1-\sqrt{\tilde{r}_0})^2\}} + o(1)$. When $\tilde{r}_0 < \mathrm{DB}(\beta)$, it can be shown that $1 - \beta - (1 - \sqrt{\tilde{r}_0})^2 < 0$, and hence, $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L\mathbb{I}(D_p^c)\} = o(1)$. Since $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}(L) = 1$, this implies that $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L\mathbb{I}(D_p)\} = 1 + o(1)$.

For the second claim of (A.6), using (A.5), L^2 can be expressed as

$$L^2 = \binom{p}{2m_a}^{-2} N_0^{-2} \sum_{\mathcal{G}_1=(A_1,U_1)\in\mathcal{M}} \sum_{\mathcal{G}_2=(A_2,U_2)\in\mathcal{M}}$$

$$\exp\biggl\{\widetilde{C}\biggl(\sum_{(i_1,j_1)\in U_1} \widetilde{R}_{i_1 j_1} + \sum_{(i_2,j_2)\in U_2} \widetilde{R}_{i_2 j_2}\biggr) - 2m_a \widetilde{B}\biggr\}. \tag{A.12}$$

The key to evaluating $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L^2\mathbb{I}(D_p)\}$ is to identify the common vertices and edges shared by any pair of graphs $\mathcal{G}_1 = (A_1, U_1), \mathcal{G}_2 = (A_2, U_2) \in \mathcal{M}$. Let $\mathcal{G}_{12} = (A_1 \cup A_2, U_1 \cup U_2)$ be the joint graph of $\mathcal{G}_1$ and $\mathcal{G}_2$. Since both G_1 and G_2 are graphs of perfect matching, there is no vertex in $\mathcal{G}_{12}$ that connects to more than two other vertices. Let $\mathcal{G}_{12}^*$ be the sub-graph of $\mathcal{G}_{12}$ excluding the common edges $U_1 \cap U_2$ of $\mathcal{G}_1$ and $\mathcal{G}_2$. Define $(i_1, i_2, \dots, i_{d_0})$ as a maximal path of $\mathcal{G}_{12}^*$ if $(i_\ell, i_{\ell+1}) \in U_{12}$ for $\ell = 1, \dots, d_0 - 1$ and there is no additional vertex connecting to the vertices on this path, where $U_{12} = (U_1 \cup U_2) \cap (U_1 \cap U_2)^c$. Let $\mathcal{P}_{12}$ be the collection of all maximal paths of $\mathcal{G}_{12}^*$, which forms a partition for the edge set U_{12}. Note that any two paths in $\mathcal{P}_{12}$ do not share common vertices. Let $\mathcal{P}_{12,c}$ and $\mathcal{P}_{12,ac}$ be the sets of all cyclic and acyclic paths in $\mathcal{P}_{12}$, such that $\mathcal{P}_{12} = \mathcal{P}_{12,c} \cup \mathcal{P}_{12,ac}$. For example, if $(i_1, i_2), (i_3, i_4) \in U_1$ and $(i_2, i_3), (i_1, i_4) \in U_2$, then the path $(i_1, i_2, i_3, i_4, i_1)$ is cyclic. If $(i_2, i_3), (i_4, i_5) \in U_2$ and $i_5 \notin A_1$, the path $(i_1, i_2, i_3, i_4, i_5)$ is acyclic.

For a given graph $\mathcal{G}_1$ and $0 \le h_0 \le m_a$, let $\mathcal{M}(\mathcal{G}_1, h_0) = \{\mathcal{G}_2 \in \mathcal{M} : \mathcal{G}_2 = (A_2, U_2), |U_1 \cap U_2| = h_0\}$ be the sub-collection of all graphs from $\mathcal{M}$ that share exactly h_0 edges with $\mathcal{G}_1$. Let $\widetilde{\mathcal{M}}(\mathcal{G}_1, h_0) = \{\mathcal{G}_2 \in \mathcal{M} : \mathcal{G}_2 = (A_2, U_2), |U_1 \cap U_2| \ge h_0\}$ be the sub-collection of graphs sharing at least h_0 edges with $\mathcal{G}_1$. Let $N_{h_0}(\mathcal{G}_1) = |\mathcal{M}(\mathcal{G}_1, h_0)|$ and $\widetilde{N}_{h_0}(\mathcal{G}_1) = |\widetilde{\mathcal{M}}(\mathcal{G}_1, h_0)|$. By direct calculation, it can be shown that

$$\widetilde{N}_{h_0}(\mathcal{G}_1) = \binom{m_a}{h_0}\binom{p - 2h_0}{2m_a - 2h_0}\frac{\{2(m_a - h_0)\}!}{(m_a - h_0)!2^{m_a - h_0}} = \frac{m_a!(p - 2h_0)!2^{h_0 - m_a}}{\{(m_a - h_0)!\}^2(p - 2m_a)!h_0!}$$

for all $\mathcal{G}_1 \in \mathcal{M}$. Note that $N_{h_0}(\mathcal{G}_1) = \widetilde{N}_{h_0}(\mathcal{G}_1) - \widetilde{N}_{h_0+1}(\mathcal{G}_1)$ and

$$\frac{\widetilde{N}_{h_0}(\mathcal{G}_1)}{\binom{p}{2m_a}N_0} = \left\{\frac{m_a!}{(m_a - h_0)!}\right\}^2 \frac{2^{h_0}}{h_0!} \prod_{j=1}^{2h_0} \frac{1}{p - j + 1}.$$

Since $m_a = o(p)$ for $\beta > 1/2$, $\widetilde{N}_{h_0+1}(\mathcal{G}_1) = o(\widetilde{N}_{h_0}(\mathcal{G}_1))$ for $0 \le h_0 < m_a$. Therefore, $N_{h_0}(\mathcal{G}_1)$ can be approximated by $\widetilde{N}_{h_0}(\mathcal{G}_1)$, and its proportion to the total size of $\mathcal{M}$ is bounded by

$$\frac{N_{h_0}(\mathcal{G}_1)}{\binom{p}{2m_a} N_0} \le \left(\frac{Cm_a^2}{p^2}\right)^{h_0} \frac{1}{h_0!} \tag{A.13}$$

for all $\mathcal{G}_1 \in \mathcal{M}$ and a positive constant C.

We further decompose the graphs in $\mathcal{M}(\mathcal{G}_1, h_0)$ based on the number of edges that form cyclic paths with $\mathcal{G}_1$. Let

$$\mathcal{M}(\mathcal{G}_1, h_0, k_0) = \{\mathcal{G}_2 \in \mathcal{M}(\mathcal{G}_1, h_0) : k_0 \text{ edges in } \mathcal{G}_2 \text{ form cyclic paths with } \mathcal{G}_1\},$$

and $N_{h_0,k_0}(\mathcal{G}_1) = |\mathcal{M}(\mathcal{G}_1, h_0, k_0)|$. It is clear that $\mathcal{M}(\mathcal{G}_1, h_0) = \cup_{k_0=0}^{m_a-h_0} \mathcal{M}(\mathcal{G}_1, h_0, k_0)$. Given k_0 selected edges from $\mathcal{G}_1$, there are at most $k_0! 3^{k_0}$ different ways of constructing the set $\mathcal{P}_{12,c}$ of cyclic paths, allowing multiple disjoint paths in $\mathcal{P}_{12,c}$. Therefore,

$$\begin{aligned}
N_{h_0,k_0}(\mathcal{G}_1) &\le \binom{m_a}{h_0}\binom{m_a-h_0}{k_0}\binom{p-2h_0-2k_0}{2m_a-2h_0-2k_0} \frac{\{2(m_a-h_0-k_0)\}! k_0! 3^{k_0}}{(m_a-h_0-k_0)! 2^{m_a-h_0-k_0}} \\
&= \frac{m_a!(p-2h_0-2k_0)! 2^{h_0+k_0-m_a} 3^{k_0}}{\{(m_a-h_0-k_0)!\}^2 (p-2m_a)! h_0!},
\end{aligned}$$

and it follows that

$$\frac{N_{h_0,k_0}(\mathcal{G}_1)}{N_{h_0}(\mathcal{G}_1)} \le \frac{\{(m_a-h_0)!\}^2}{\{(m_a-h_0-k_0)!\}^2} \frac{(p-2h_0-2k_0)!}{(p-2h_0)!} 6^{k_0} \le \frac{m_a^{2k_0} 6^{k_0}}{p^{2k_0}}$$

for all $\mathcal{G}_1 \in \mathcal{M}$.

For any pair of graphs $\mathcal{G}_1 = (A_1, U_1), \mathcal{G}_2 = (A_2, U_2) \in \mathcal{M}(\mathcal{G}_1, h_0)$, we group the terms in the summation $\sum_{(i_1,j_1)\in U_1} \widetilde{R}_{i_1j_1} + \sum_{(i_2,j_2)\in U_2} \widetilde{R}_{i_2j_2}$ by the common edges of U_1 and U_2 and the different paths in $\mathcal{P}_{12,ac}$ and $\mathcal{P}_{12,c}$. Therefore,

$$\begin{aligned}
L(\mathcal{G}_1)L(\mathcal{G}_2)\mathbb{I}(D_p) &= \exp\left\{\widetilde{C}\Big(\sum_{(i_1,j_1)\in U_1} \widetilde{R}_{i_1j_1} + \sum_{(i_2,j_2)\in U_2} \widetilde{R}_{i_2j_2}\Big) - 2m_a\widetilde{B}\right\}\mathbb{I}(D_p) \\
&= \prod_{(i,j)\in U_1\cap U_2} \exp(2\widetilde{C}\widetilde{R}_{ij} - 2\widetilde{B})\mathbb{I}(D_p) \\
&\quad \prod_{(i_1,i_2,\ldots,i_{d_0})\in\mathcal{P}_{12,ac}} \exp\left(\widetilde{C}\sum_{\ell=1}^{d_0-1} \widetilde{R}_{i_\ell i_{\ell+1}} - (d_0-1)\widetilde{B}\right) \\
&\quad \prod_{(i_1,\ldots,i_{\tilde{d}_0},i_1)\in\mathcal{P}_{12,c}} \exp\left(\widetilde{C}\sum_{\ell=1}^{\tilde{d}_0-1} \widetilde{R}_{i_\ell i_{\ell+1}} + \widetilde{C}\widetilde{R}_{i_{\tilde{d}_0} i_1} - \tilde{d}_0\widetilde{B}\right)\mathbb{I}(D_p).
\end{aligned}$$

To calculate $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L(\mathcal{G}_1)L(\mathcal{G}_2)\mathbb{I}(D_p)\}$, we first consider the h_0 common edges in $U_1 \cap U_2$. Since $\mathbb{I}(D_p) \le \mathbb{I}\{\widetilde{R}_{ij} \le \sqrt{4\log p} \text{ for } (i,j) \in U_1 \cap U_2\}$, it is easy to see

$$\begin{aligned}
&\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\Big\{\prod_{(i,j)\in U_1\cap U_2} \exp(2\widetilde{C}\widetilde{R}_{ij} - 2\widetilde{B})\mathbb{I}(D_p)\Big\} \\
&\le \prod_{(i,j)\in U_1\cap U_2} \mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\big\{\exp\big(2\widetilde{C}\widetilde{R}_{ij} - 2\widetilde{B}\big)\mathbb{I}(\widetilde{R}_{ij} \le \sqrt{4\log p})\big\}
\end{aligned}$$

Similar as (A.9), by Lemma S14 in the SM, it can be shown that

$$\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\big\{\exp(2\widetilde{C}\widetilde{R}_{ij}-2\widetilde{B})\mathbb{I}(\widetilde{R}_{ij}\le\sqrt{4\log p})\big\}\le Cp^{2d(\tilde{r}_0)}$$

for $(i,j)\in U_1\cap U_2$, where $d(r)=2r\mathbb{I}(r\le 1/4)+\{1-2(1-\sqrt{r})^2\}\mathbb{I}(r>1/4)$. Note that $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\big\{\exp(2\widetilde{C}\widetilde{R}_{ij}-2\widetilde{B})\big\}=p^{4\tilde{r}_0}(1+o(1))$. Since $2\tilde{r}_0>d(\tilde{r}_0)$ when $\tilde{r}_0>1/4$, the truncation by $\mathbb{I}(D_p)$ leads to a tighter bound on the second moment of the likelihood ratio. Hence,

$$\prod_{(i,j)\in U_1\cap U_2}\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\big\{\exp\big(2\widetilde{C}\widetilde{R}_{ij}-2\widetilde{B}\big)\mathbb{I}(\widetilde{R}_{ij}\le\sqrt{4\log p})\big\}\le C^{h_0}p^{2h_0 d(\tilde{r}_0)}.$$

Next, we consider the acyclic paths. For any path $(i,j)\in\mathcal{P}_{12,ac}$ with length 1, note that $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\exp\big(\widetilde{C}\widetilde{R}_{ij}-\widetilde{B}\big)=1$ from (A.7). For paths $(i_1,i_2,\ldots,i_{d_0})\in\mathcal{P}_{12,ac}$ with length longer than 1, we have

$$\widetilde{C}\sum_{\ell=1}^{d_0-1}\widetilde{R}_{i_\ell i_{\ell+1}}-(d_0-1)\widetilde{B}=-\frac{\sum_{k=1}^n\widetilde{Z}_{k,(i_1,\ldots,i_{d_0})}}{2(1-\delta_a^2)}-\frac{n(d_0-1)}{2}\log(1-\delta_a^2),$$

where $\{\widetilde{Z}_{k,(i_1,\ldots,i_{d_0})}\}_{k=1}^n$ is a sequence of IID random variables with

$$\widetilde{Z}_{k,(i_1,\ldots,i_{d_0})}=\delta_a^2Z_{ki_1}^2+\delta_a^2Z_{ki_{d_0}}^2+2\delta_a^2\sum_{\ell=2}^{d_0-1}Z_{ki_\ell}^2-2\delta_a\sum_{\ell=1}^{d_0-1}Z_{ki_\ell}Z_{ki_{\ell+1}}.$$

By taking the conditional expectation of the end nodes i_1 and i_{d_0} given the middle nodes $i_2,\ldots,i_{d_0-1}$, and recursively applying Lemma S13 in the SM given the middle nodes, we have

$$\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\exp\left\{\frac{-\widetilde{Z}_{k,(i_1,\ldots,i_{d_0})}}{2(1-\delta_a^2)}\right\}=(1-\delta_a^2)^{(d_0-1)/2},$$

which leads to

$$\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\left\{\exp\left(\widetilde{C}\sum_{\ell=1}^{d_0-1}\widetilde{R}_{i_\ell i_{\ell+1}}-(d_0-1)\widetilde{B}\right)\right\}=1.$$

Lastly, for a cyclic path $(i_1,\ldots,i_{\tilde{d}_0},i_1)\in\mathcal{P}_{12,c}$, $\tilde{d}_0$ has to be an even number. We calculate $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{\exp(\widetilde{C}\sum\widetilde{R}_{i_\ell i_{\ell+1}}+\widetilde{C}\widetilde{R}_{i_{\tilde{d}_0}i_1}-\tilde{d}_0\widetilde{B})\mathbb{I}(D_p)\}$ by its conditional expectation with respect to the i_1th variable given all other variables on this cyclic path. By Lemma S16 in the SM, we have

$$\begin{aligned}&\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\big\{\exp\big(\widetilde{C}\widetilde{R}_{i_1i_2}+\widetilde{C}\widetilde{R}_{i_{\tilde{d}_0}i_1}-2\widetilde{B}\big)\mathbb{I}(D_p)\big|Z_{ki_2},Z_{ki_{\tilde{d}_0}},k=1,\ldots,n\big\}\\ \le&\prod_{k=1}^n\mathrm{E}\left[\exp\left\{\frac{2\delta_aZ_{ki_1}Z_{ki_2}+2\delta_aZ_{ki_1}Z_{ki_{\tilde{d}_0}}-2\delta_a^2Z_{ki_1}^2-\delta_a^2Z_{ki_2}^2-\delta_a^2Z_{ki_{\tilde{d}_0}}^2}{2(1-\delta_a^2)}\right\}\middle|Z_{i_2},Z_{i_{\tilde{d}_0}}\right]\\ &(1-\delta_a^2)^{-n}\mathbb{I}(\widetilde{R}_{i_2i_{\tilde{d}_0}}\le\sqrt{4\log p})\\ =&(1-\delta_a^4)^{-n/2}\exp\left\{\delta_a^2\frac{\sum_{k=1}^n(2Z_{ki_2}Z_{ki_{\tilde{d}_0}}-\delta_a^2Z_{ki_2}^2-\delta_a^2Z_{ki_{\tilde{d}_0}}^2)}{2(1-\delta_a^4)}\right\}\mathbb{I}(\widetilde{R}_{i_2i_{\tilde{d}_0}}\le\sqrt{4\log p}),\end{aligned}$$

which is equal to $1+O(n^{-1/2}(\log p)^{3/2})$ for all values of $\{Z_{k_2},Z_{ki_{\tilde{d}_0}}\}_{k=1}^n$ under the event D_p. Therefore, $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{\exp(\widetilde{C}\sum_{\ell=1}^{\tilde{d}_0-1}\widetilde{R}_{i_\ell i_{\ell+1}}+\widetilde{C}\widetilde{R}_{i_{\tilde{d}_0}i_1}-\tilde{d}_0\widetilde{B})\mathbb{I}(D_p)\}$ is bounded by

$$\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\left\{\exp\left(\widetilde{C}\sum_{\ell=2}^{\tilde{d}_0-1}\widetilde{R}_{i_\ell i_{\ell+1}}-(\tilde{d}_0-2)\widetilde{B}\right)\mathbb{I}(D_p)\right\}\exp\left\{\frac{(\log p)^{3/2}}{n^{1/2}}\right\}=\exp\left\{\frac{(\log p)^{3/2}}{n^{1/2}}\right\}.$$

Since different paths in $\mathcal{P}_{12}$ do not share common vertices, all the common edges in $U_1 \cap U_2$ and the different paths in $\mathcal{P}_{12}$ are mutually independent. Based on the above results, for all $\mathcal{G}_1 \in \mathcal{M}$ and $\mathcal{G}_2 \in \mathcal{M}(\mathcal{G}_1, h_0, k_0)$, we have

$$\text{(A.14)} \qquad \mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L(\mathcal{G}_1)L(\mathcal{G}_2)\mathbb{I}(D_p)\} \le C^{h_0} p^{2h_0 d(\tilde{r}_0)} \exp(k_0 L_p n^{-1/2}).$$

Combining (A.12), (A.13) and (A.14), it follows that

$$\begin{aligned}
&\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L^2\mathbb{I}(D_p)\} \\
&\le \binom{p}{2m_a}^{-2} N_0^{-2} \sum_{\mathcal{G}_1 \in \mathcal{M}} \sum_{h_0=0}^{m_a} \sum_{\mathcal{G}_2 \in \mathcal{M}(\mathcal{G}_1, h_0)} \mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L(\mathcal{G}_1)L(\mathcal{G}_2)\mathbb{I}(D_p)\} \\
&\le \binom{p}{2m_a}^{-1} N_0^{-1} \sum_{\mathcal{G}_1 \in \mathcal{M}} \sum_{h_0=0}^{m_a} \sum_{k_0=0}^{m_a-h_0} \left(\frac{Cm_a^2}{p^2}\right)^{h_0} \frac{p^{2h_0 d(\tilde{r}_0)}}{h_0!} \frac{m_a^{2k_0} 6^{k_0}}{p^{2k_0}} \exp(k_0 L_p n^{-1/2}) \\
&\le \exp\{Cm_a^2 p^{2d(\tilde{r}_0)-2}\},
\end{aligned}$$

where the last inequality is due to $\sum_{k_0=0}^{m_a-h_0}\{6m_a^2 p^{-2}\exp(L_p n^{-1/2})\}^{k_0} = 1 + o(1)$ since $m_a = o(p)$ for $\beta > 1/2$. Note that

$$\exp\{Cm_a^2 p^{2d(\tilde{r}_0)-2}\} = \exp\{C(p^2)^{2(1-\beta)-1+d(\tilde{r}_0)}\},$$

which converges to 1 if $1 - 2\beta + d(\tilde{r}_0) < 0$. Note that $1 - 2\beta + d(\tilde{r}_0) < 0$ is equivalent to $\tilde{r}_0 < \mathrm{DB}(\beta)$. Therefore, $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L^2\mathbb{I}(D_p)\} \le 1 + o(1)$ if $\tilde{r}_0 < \mathrm{DB}(\beta)$. Since $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L^2\mathbb{I}(D_p)\} \ge \mathrm{E}^2_{N(\mathbf{0},\mathbf{I}_p)}\{L\mathbb{I}(D_p)\}$, we have $\mathrm{E}_{N(\mathbf{0},\mathbf{I}_p)}\{L^2\mathbb{I}(D_p)\} = 1 + o(1)$ if $\tilde{r}_0 < \mathrm{DB}(\beta)$. This finishes the proof of (A.6), and the conclusion of Theorem 1 follows from (A.6). □

Acknowledgments. The authors would like to thank the anonymous referees, the Associate Editor and the Editor for thoughtful comments and suggestions which have improved the presentation of the paper. Part of the work reported in the paper was conducted while Qiu was affiliated with Iowa State University. He thanks ISU's statistics department for providing various support during the course of the project.

Funding. Chen was partially supported by National Natural Science Foundation of China Grants 12292983 and 12071013. Zhang was partially supported by National Key R&D Program of China grant 2021YFA1000101, National Natural Science Foundation of China grants 71931004 and 72201101, and Shanghai Pujiang Program (21PJC034).

SUPPLEMENTARY MATERIAL

Online Supplement

The online supplement provides the proofs of Propositions 1–5 and Theorems 2–5.

Code and Reproducibility

The codes for the simulation study can be found at https://github.com/yumouqiu/Optimal-Test-Cov.

REFERENCES

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed.), New York: John Wiley & Sons.

Arias-Castro, E., Bubeck, S. and Lugosi, G. (2012). Detection of correlations. *The Annals of Statistics*, **40**, 412–435.

Berbee, H. (1979). *Random Walks with Stationary Increments and Renewal Theory*, Amsterdam: Mathematical Centre.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.), New York: Springer.

Bickel, P. and Levina, E. (2008a). Regularized estimation of large covariance matrices. *The Annals of Statistics*, **36**, 199-227.

Bradley, R. (2005). Basic properties of strong mixing conditions: a survey and some open questions. *Probability Surveys*, **2**, 107-144.

Cai, T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *The Annals of Statistics*, **39**, 1496-1525.

Cai, T., Liu, W. D. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the Americain Statistical Association*, **108**, 265-277.

Cai, T. and Ma, Z. M. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, **19**, 2359–2388.

Chen, S. X., Zhang, L. X. and Zhong, P. S. (2010). Tests for high-dimensional covariance matrices. *Journal of the Americain Statistical Association*, **105**, 810-819.

Delaigle, A., Hall, P. and Jin, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**, 283-301.

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, **32**, 962 - 994.

Donoho, D. and Jin, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science*, **30**, 1–25.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*, Academic Press.

Hall, P. and Jin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics*, **38**, 1686-1732.

Hotelling, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **15**, 193–232.

Ingster, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Mathematical Methods of Statistics*, **6**, 47-69.

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Annals of Applied Probability*, **14**, 865-880.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, **30**, 1081-1102.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, New York: John Wiley & Sons.

Lin, Z. Y. and Bai, Z. D. (2010). *Probability Inequalities*, Beijing: Science Press.

Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic processes and their applications*, **29**, 309–315.

Nagao, H. (1973). On some test criteria for covariance matrix. *The Annals of Statistics*, **1**, 700-709.

Qiu, Y. and Chen, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *The Annals of Statistics*, **40**, 1285-1314.

Qiu, Y., Chen, S. X. and Nettleton, D. (2018). Detecting rare and faint signals via thresholding maximum likelihood estimators. *The Annals of Statistics*, **46**, 895-923.

Saulis, L. and Statulevičius, V. A. (1991). *Limit Theorems for Large Deviations,* Dordrecht: Kluwer Academic.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, **92**, 951-956.

Xue, L. Z., Ma, S. Q. and Zou, H. (2012). Positive-definite ℓ_1-penalized estimation of large covariance matrices. *Journal of the Americain Statistical Association*, **107**, 1480-1491.

Zhong, P. S., Chen, S. X. and Xu, M. Y. (2013). Tests alternative to higher criticism for high dimensional means under sparsity and column-wise dependence. *The Annals of Statistics*, **41**, 2820-2851.