Journal of the Royal Statistical Society Series B: Statistical Methodology, 2024, **00**, 1–22
https://doi.org/10.1093/jrsssb/qkae029

Royal Statistical Society — Data | Evidence | Decisions
Journal of the Statistics Society
Series B
Statistical Methodology

Original Article

Minimax detection boundary and sharp optimal test for Gaussian graphical models

Yumou Qiu[1] and Bin Guo[2]

[1]School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing, People's Republic of China
[2]Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, People's Republic of China

Address for correspondence: Yumou Qiu, School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing 100871, People's Republic of China. Email: qiuyumou@math.pku.edu.cn

Abstract

In this article, we derive the minimax detection boundary for testing a sub-block of variables in a precision matrix under the Gaussian distribution. Compared to the results on the minimum rate of signals for testing precision matrices in literature, our result gives the exact minimum signal strength in a precision matrix that can be detected. We propose a thresholding test that is able to achieve the minimax detection boundary under certain cases by adaptively choosing the threshold level. The asymptotic distribution of the thresholding statistic for precision matrices is derived. Power analysis is conducted to show the proposed test is powerful against sparse and weak signals, which cannot be detected by the existing L_{max} and L_2 tests. Simulation studies show the proposed test has an accurate size around the nominal level and is more powerful than the existing tests for detecting sparse and weak signals in precision matrices. Real data analysis on brain imaging data is carried out to illustrate the utility of the proposed test in practice, which reveals functional connectivity between brain regions for Alzheimer's disease patients and normal healthy people.

Keywords: detection boundary, high dimensionality, minimax, precision matrix, sparse and weak signal, thresholding

1 Introduction

Precision matrix, which is the inverse of the covariance matrix, plays an important role in multivariate analysis, for example, Hotelling's mean test and Fisher's discriminant analysis (Anderson, 2003). It provides a measure for conditional dependence among variables. Under the multivariate Gaussian distribution, the precision coefficient between two variables being zero is equivalent to conditional independence between those variables. Therefore, for Gaussian graphical models, the support of the precision matrix provides a graphical representation of the conditional dependence network. See applications of precision matrices in genomics to identify interactions between genes and pathways (Wang & Huang, 2014), and in neuroscience to study brain connectivity (Huang et al., 2010).

Classical estimation of precision matrices is by inverting sample covariance matrices. However, when the sample size n is smaller than the dimension p, this classical estimator no longer exists as the sample covariance matrix is not invertible. Penalization methods have been developed for estimating a high-dimensional precision matrix by utilizing its sparsity structure. Meinshausen and Bühlmann (2006) considered to estimate nonzero coefficients in nodewise regressions. Yuan and Lin (2007) and Friedman et al. (2008) proposed the graphical lasso estimation via a penalized likelihood method. Yuan (2010) and Cai et al. (2011) developed constraint L_1 minimization approaches. Also see Liu et al. (2012), Lafferty et al. (2012), and Xue and Zou (2012) for penalized estimation of precision matrices under the nonparanormal model.

Received: August 19, 2022. **Revised:** February 26, 2024. **Accepted:** March 1, 2024

Several recent methods have also been proposed for the inference of each element in a high-dimensional precision matrix. Liu (2013) constructed a multiple testing procedure for detecting nonzero precision coefficients based on residuals from nodewise regressions. Ren et al. (2015) developed a confidence interval for each precision coefficient based on residuals from pairwise regressions. Ren et al. (2019) considered the inference problem of multiple precision matrices from several groups. Qiu and Zhou (2020, 2022) proposed multiple testing procedures for partial correlations with application to brain connectivity studies under independent and time series data, respectively.

For the global tests of precision matrices, Xia et al. (2015) proposed a maximum type test ($L_{\max}$-test) for the equality of two precision matrices together with a multiple testing procedure to identify different precision coefficients between the two samples. Chang et al. (2018) developed the Gaussian approximation result for the maximum statistic on precision matrices under time-dependent data. Xia et al. (2018) proposed an L_2 type test for sub-matrices with finite size being zero in a high-dimensional precision matrix. Particularly, Xia et al. (2015) showed that $\{(\log p)/n\}^{1/2}$ is the minimum rate of signals that can be detected, where n and p denote the sample size and dimension, respectively. Namely, there exists a positive constant $\tilde{c}_0$ such that no test is uniformly powerful over all precision matrices with minimum signal larger than $\tilde{c}_0\{(\log p)/n\}^{1/2}$. However, this result only provides the minimum rate of signals that can be detected. The expression of the constant $\tilde{c}_0$ and its relationship with signal sparsity are unknown. For testing means, Donoho and Jin (2004) showed that the L_2 and $L_{\max}$ type tests may have no power in detecting sparse and weak signals. More powerful tests for precision matrices need to be developed.

In this article, we consider the problem of testing sub-blocks of a high-dimensional precision matrix being zero. Under the Gaussian distribution, the null hypothesis implies conditional independence among the variables in this sub-block. We first derive the minimax detection boundary for testing precision matrices that gives the explicit expression of the signal strength as a function of signal sparsity that can possibly be detected by a testing procedure. No test can be uniformly powerful if the signal strength and sparsity parameters fall under the detection boundary. Compared to the results on minimum rate of signals for testing precision matrices (Xia et al., 2015), our result gives the exact minimum signal strength in a precision matrix that can be detected. Second, we construct a thresholding test for precision matrices with an adaptively chosen threshold level to maximize its power. Essentially, the proposed test utilizes multiple threshold levels and chooses the one that is the most significant. This is in the same spirit as the Higher Criticism (HC) tests for means (Donoho & Jin, 2004; Hall & Jin, 2010; Zhong et al., 2013) and fixed-dimensional regression coefficients (Qiu et al., 2018). We show that under certain constraints on the sample size, dimension and the size of the sub-block to be tested, the power of the proposed thresholding test converges to 1 as long as the signal strength and sparsity are above the detection boundary. In this sense, the derived minimax detection boundary is tight, and the proposed test is sharp optimal. Simulation studies show the proposed test has an accurate size around the nominal level under various settings, and it is more powerful than the $L_{\max}$ and L_2 tests for detecting sparse and weak signals in precision matrices. The proposed method is applied to a brain imaging data set, which reveals functional connectivity between brain regions for Alzheimer's disease patients and normal healthy people.

Two main contributions of this work are to derive the minimax detection boundary for signal detection in high-dimensional precision matrices and to show that the proposed multi-level thresholding test for precision matrices can attain this detection boundary. Both results are new in the literature and imply the phrase transition property of testing precision matrices. Those two problems are quite challenging and are not direct extensions of the existing detection boundary and higher criticism results for testing means and parameters in fixed-dimensional models (Donoho & Jin, 2004, 2015; Hall & Jin, 2010; Qiu et al., 2018; Zhong et al., 2013). First, compared to the results on the minimax rate of signal detection, deriving the detection boundary for precision matrices requires a more delicate analysis and tight bounds for the moments of the likelihood ratio function. To our knowledge, there is no related work in literature. We have developed a new technique based on counting the cyclic and acyclic paths in the graphs formed by perfect matching. Second, compared to the existing works, analysing the asymptotic properties of the thresholding statistic on the debiased estimates of precision coefficients is much more difficult than that on sample means. It requires careful control of small order terms, moderate deviation results of the debiased estimates, and handling the complex dependence among the estimated entries in the same row and column of the precision matrix. Note that the estimates in the same row or the same column share common variables. We have successfully tackled those challenges and derived the asymptotic properties and detection boundary of the proposed test.

The article is organized as follows. Section 2 introduces the background, the hypotheses of interest and the application examples. Section 3 establishes the minimax detection boundary for testing precision matrices. Section 4 constructs the proposed multi-level thresholding test. Section 5 states the asymptotic results for the proposed test statistic under the null hypotheses. Section 6 studies the power of the proposed test, and shows it can achieve the minimax detection boundary derived in Section 3 under certain cases. Simulation studies and real data analysis on brain connectivity are reported in Sections 7 and 8, respectively. Discussion and extension of the proposed method are provided in Section 9. All the technical proofs are relegated to the online supplementary material.

2 Hypotheses and applications

Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ be independent and identically distributed (i.i.d.) random vectors from a p-dimensional Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^{\mathrm{T}}$ and $\boldsymbol{\Sigma} = (\sigma_{j_1 j_2})$. Let $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} = (\omega_{j_1 j_2})$ be the precision matrix. Under the Gaussian distribution, $\omega_{j_1 j_2} = 0$ is equivalent to the conditional independence between the j_1th and j_2th variables. The Gaussian graphical model (GGM) $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ is constructed based on the nonzero elements of $\boldsymbol{\Omega}$, where $\mathcal{N} = \{1, \ldots, p\}$ is the node set, and $\mathcal{E} = (e_{j_1 j_2})$ denotes the edges among the nodes with $e_{j_1 j_2} = \mathbb{I}(\omega_{j_1 j_2} \neq 0)$. Here, $e_{j_1 j_2} = 1$ stands for an edge between the j_1th and j_2th variables, which represents the two variables being conditionally dependent under the GGM.

Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be a partition of $\mathcal{N}$ such that $\mathcal{A}_1 \cup \mathcal{A}_2 = \mathcal{N}$ and $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$. We are interested in testing for the conditional independence for all the variables in the set $\mathcal{A}_1$. Namely, we consider the hypotheses

$$\begin{aligned} &H_0: \omega_{j_1 j_2} = 0 \quad \text{for all } j_1, j_2 \in \mathcal{A}_1 \text{ and } j_1 \neq j_2 \text{ vs.} \\ &H_a: \omega_{j_1 j_2} \neq 0 \quad \text{for some } j_1, j_2 \in \mathcal{A}_1 \text{ and } j_1 \neq j_2. \end{aligned} \tag{2.1}$$

Let p_1 and p_2 be the cardinality of the sets $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively. Note that $q = p_1(p_1 - 1)/2$ is the number of free off-diagonal parameters of interest in the hypotheses (2.1). In general, the null hypothesis of (2.1) is not equivalent to the corresponding covariances $\sigma_{j_1 j_2}$ being zero for $j_1, j_2 \in \mathcal{A}_1$ and $j_1 \neq j_2$. An exemption is the special case that $\mathcal{A}_1 = \{1, \ldots, p\}$, where testing the precision matrix $\boldsymbol{\Omega}$ being diagonal is equivalent to testing $\boldsymbol{\Sigma}$ being diagonal. However, even under this special case, testing for the precision matrix may result in a higher power if the nonzero off-diagonal values in $\boldsymbol{\Omega}$ are stronger than those in $\boldsymbol{\Sigma}$ under the alternative hypothesis.

In the following, we provide two examples that require testing for the conditional independence structure among variables.

Example 1 (Testing for concliques in Gaussian Markov Random Field). Consider data observed with replications on a regular lattice of size $d \times d$. Let $\mathbf{s}_j = (u_{1j}, u_{2j})$ denote the jth geo-referenced location in a lattice for a horizontal coordinate u_{1j} and a vertical coordinate u_{2j}. Let $X_{i\mathbf{s}_j}$ be the data observed at the location $\mathbf{s}_j$ for the ith replication, and $F_{ij}(\cdot)$ be its conditional cumulative distribution function given all the variables observed at other locations. Let $N(\mathbf{s}_j) = \{\mathbf{s}_k : \mathbf{s}_k \neq \mathbf{s}_j$ and $F_{ij}(\cdot)$ depends functionally on $X_{i\mathbf{s}_k}\}$ be the neighbourhood set of $\mathbf{s}_j$. In a Gaussian Markov Random Field (GMRF) model (Besag, 1974), $F_{ij}(\cdot)$ follows the normal distribution with the conditional mean specified by $\theta_j + \sum_{\mathbf{s}_k \in N(\mathbf{s}_j)} \theta_{jk}(X_{i\mathbf{s}_k} - \theta_k)$. Under this model, $X_{i\mathbf{s}_j}$ are conditional dependent with $X_{i\mathbf{s}_k}$ if and only if $\mathbf{s}_k$ is in the neighbourhood of $\mathbf{s}_j$. For spatial data analysis, the key is to specify a correct neighbourhood structure. Therefore, it is important to test for conditional independence among variables at different locations (Kaiser & Nordman, 2012).

Cliques and concliques are commonly used in graph theory to describe the connections of a network (Skiena, 1991). A clique of a graph $\mathcal{G}$ is a complete subgraph with pairwise connections between all nodes, while a conclique is a

Table 1. Two conclique sets in a 4-nearest neighbourhood model, indexed by the letters ‘a’ and ‘b’

a	b	a	b	a	b	a	b	a	b	a	b
b	a	b	a	b	a	b	a	b	a	b	a
a	b	a	b	a	b	a	b	a	b	a	b
b	a	b	a	b	a	b	a	b	a	b	a

subgraph with no connections between any pair of nodes. In a GMRF model with the four-nearest neighbourhood structure, a clique takes the form $\{(u_1, u_2), (u_1 + 1, u_2)\}$, $\{(u_1, u_2), (u_1 - 1, u_2)\}$, $\{(u_1, u_2), (u_1, u_2 + 1)\}$, or $\{(u_1, u_2), (u_1, u_2 - 1)\}$, and there is no clique with more than 2 nodes. Meanwhile, the locations can be partitioned into two concliques $\mathcal{A}_1$ and $\mathcal{A}_2$ as shown in Table 1, where $\mathcal{A}_1$ and $\mathcal{A}_2$ include all the locations marked by the letter ‘a’ and ‘b’, respectively. Testing conditional independence for both $\mathcal{A}_1$ and $\mathcal{A}_2$ via the hypotheses (2.1) can be used to check the validity of the four-nearest neighbourhood structure.

Example 2 (Testing for conditional independence of variables from omics data). With the advance of high-throughput measurement technologies, we are able to collect different types of data on the same observational unit from different platforms. For example, in plant science research, gene expression levels and metabolites of plants, soil microbiomes, and plant traits can be collected from various systems. Using the omics data, it is of interest to study whether the expression of genes in a certain pathway is conditionally independent with a family of microbes in the soil given all other factors, which can be formulated as a hypotheses testing problem for the precision matrix in (2.1).

3 Minimax detection boundary

In this section, we derive the detection boundary for the hypotheses (2.1) in terms of the minimax power against sparse and weak signals. Here, the nonzero off-diagonal precision coefficients $\omega_{j_1j_2}$ for $j_1, j_2 \in \mathcal{A}_1$ are considered as signals under H_a of (2.1). The detection boundary is a function of signal strength and sparsity which provides the undetectable region of signals. No test procedure is able to distinguish the null and alternative hypotheses if the signal strength is lower than the boundary at each sparsity level of the signals.

Let $\lfloor\cdot\rfloor$ and $\mathbb{I}(\cdot)$ denote the floor function and the indicator function, respectively. Let $\mathbf{\Omega}_{\mathcal{A}_1}$ be the sub-matrix of $\mathbf{\Omega}$ with rows and columns in $\mathcal{A}_1$. Recall that $q = p_1(p_1 - 1)/2$ and p_1 is the cardinality of the set $\mathcal{A}_1$. Let $m_a = \sum_{(j_1,j_2)\in\mathcal{A}_1, j_1<j_2} \mathbb{I}(\omega_{j_1j_2} \neq 0)$ be the total number of nonzero upper-diagonal elements in $\mathbf{\Omega}_{\mathcal{A}_1}$, and $t_{\mathcal{A}_1} = \max_{j_1\in\mathcal{A}_1} \sum_{j_2\in\mathcal{A}_1} \mathbb{I}(\omega_{j_1j_2} \neq 0)$ be the maximal number of nonzero elements in each row of $\mathbf{\Omega}_{\mathcal{A}_1}$. For a sparsity parameter $\beta \in (1/2, 1)$, the rarity of signals under the alternative hypothesis of (2.1) is characterized by $q^{(1-\beta)}$ such that $m_a \geq \lfloor q^{(1-\beta)}\rfloor$. Note that $m_a = \lfloor q^{(1-\beta)}\rfloor$ with $\beta > 1/2$ represents the sparse signal regime which implies that less than one nonzero off-diagonal element in each row of $\mathbf{\Omega}_{\mathcal{A}_1}$ on average as $m_a < p_1$. We consider the signal strength at the order $\{\log(q)/n\}^{1/2}$ such that

$$|\omega_{j_1j_2}| = \sqrt{2r_{j_1j_2}\log(q)/n} \quad \text{for } r_{j_1j_2} > 0, \tag{3.1}$$

if $\omega_{j_1j_2} \neq 0$, $(j_1, j_2) \in \mathcal{A}_1$ and $j_1 \neq j_2$. Similar settings of signal sparsity and strength are considered for testing means (Donoho & Jin, 2004; Hall & Jin, 2010), regression coefficients (Qiu et al., 2018) and covariance matrices (Chen et al., 2023).

Let $\lambda_{\min}(\mathbf{\Sigma})$ be the minimum eigenvalue of $\mathbf{\Sigma}$, and $s = \max_{1\leq j_1\leq p} \sum_{j_2=1}^{p} \mathbb{I}(\omega_{j_1j_2} \neq 0)$ be the maximal number of nonzero elements in each row of $\mathbf{\Omega}$. For a positive constant C, we derive the minimax power for testing the hypotheses (2.1) under the following class of precision matrices:

$$\mathcal{U}(\beta, r_0, \tau) = \{\boldsymbol{\Omega} : m_a \geq \lfloor q^{(1-\beta)} \rfloor \text{ nonzero } \omega_{j_1 j_2} \text{ specified by (3.1) in the upper-triangle of } \boldsymbol{\Omega}_{\mathcal{A}_1} \text{ with } r_{j_1 j_2} \geq r_0,\ \max_{j \in \mathcal{A}_1} \omega_{jj} \leq \tau,\ \max_{1 \leq j \leq p} \sigma_{jj} \leq C, \quad (3.2)$$
$$\lambda_{\min}(\boldsymbol{\Sigma}) \geq C^{-1},\ s \leq Cn^{\nu} \text{ for } \nu \in [0, 1/2) \text{ and } t_{\mathcal{A}_1} \leq C\}.$$

Under this class, the parameters β and r_0 specify the minimum signal number and strength for the hypotheses (2.1), respectively, which constitute the sparse and weak signal regime. Clearly, the most difficult cases for separating the null and alternative hypotheses of (2.1) under the class $\mathcal{U}(\beta, r_0, \tau)$ is that $m_a = \lfloor q^{(1-\beta)} \rfloor$ and $r_{j_1 j_2} = r_0$ for all nonzero off-diagonal values in $\boldsymbol{\Omega}_{\mathcal{A}_1}$. The class $\mathcal{U}(\beta, r_0, \tau)$ does not have restriction on the sub-blocks $\mathcal{A}_2 \times \mathcal{A}_2$ and $\mathcal{A}_1 \times \mathcal{A}_2$ of $\boldsymbol{\Omega}$ other than the sparsity condition $s \leq Cn^{\nu}$. Testing for sparse and weak signals is the most challenging case in signal detection, which has not been studied for precision matrices.

The conditions $\lambda_{\min}(\boldsymbol{\Sigma}) \geq C^{-1}$, $s \leq Cn^{\nu}$ for $\nu \in [0, 1/2)$ and $t_{\mathcal{A}_1} \leq C$ in the last row of (3.2) and the condition of bounded variances are required for the asymptotics of the proposed test. Note that we do not require the maximum eigenvalue of $\boldsymbol{\Sigma}$ to be bounded from above. See Conditions 2 and 3 in Section 5. Those conditions are included in the class $\mathcal{U}(\beta, r_0, \tau)$ so that the minimax detection lower boundary in Theorem 1 is consistent with the detection upper boundary of the proposed test derived in Theorem 4 under a common class of precision matrices. Particularly, under the constraint $t_{\mathcal{A}_1} \leq C$, we consider the precision matrices with non-clustered signals under the alternative hypotheses of (2.1) such that no row/column of $\boldsymbol{\Omega}_{\mathcal{A}_1}$ has a diverging number of nonzero $\omega_{j_1 j_2}$. For instance, in Example 1 of testing the four-nearest neighbourhood structure of a GMRF model (Kaiser et al., 2012; Kaiser & Nordman, 2012), we are interested in the power of the test under the alternative hypothesis that a few locations have local spatial dependence beyond the four-nearest neighbours. Since $\beta > 1/2$, $q^{(1-\beta)} = o(p_1)$. This means that the majority of variables in $\mathcal{A}_1$ have zero precision coefficients in $\boldsymbol{\Omega}_{\mathcal{A}_1}$ and the average number of signals in each row of $\boldsymbol{\Omega}_{\mathcal{A}_1}$ is less than 1. For the class of precision matrices with unbounded $t_{\mathcal{A}_1}$, the signal regime for minimax results could be different as the signals are clustered within a few rows of $\boldsymbol{\Omega}_{\mathcal{A}_1}$. More discussions on this point are provided at the end of Section 6.

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

$$\mathrm{DB}(\beta) = \begin{cases} \beta - 1/2 & \text{if } 1/2 < \beta \leq 3/4, \\ (1 - \sqrt{1-\beta})^2 & \text{if } 3/4 < \beta < 1, \end{cases} \tag{3.3}$$

which is the optimal detection boundary for testing means under independent normal distributions with unit variance (Donoho & Jin, 2004; Ingster, 1997). For two real sequences $\{a_n\}$ and $\{b_n\}$, let $a_n \asymp b_n$ denote the two sequences being at the same order. Namely, there are two positive constants c_1, c_2 such that $c_1 \leq a_n/b_n \leq c_2$ for all n. The following theorem shows that $\mathrm{DB}(\beta)$ is also the minimax detection lower boundary for testing the hypotheses (2.1) over the precision matrix class $\mathcal{U}(\beta, r_0, \tau)$ in (3.2) with $\beta \in (\max\{1/2, (3 - \xi\gamma^{-1})/4\}, 1)$ under the Gaussian distribution, where $\xi > 0$ specifies the growth rate of p such that $n \asymp p^{\xi}$ and $\gamma \in (0, 1]$ specifies the number p_1 of variables in the testing set $\mathcal{A}_1$ such that $p_1 \asymp p^{\gamma}$.

Theorem 1 Under $n \asymp p^{\xi}$ for $\xi > 0$, $p_1 \asymp p^{\gamma}$ for $\gamma \in (0, 1]$ and Gaussian distributed data, if $r_0 \tau^{-2} < \mathrm{DB}(\beta)$ and $\max\{1/2, (3 - \xi\gamma^{-1})/4\} < \beta < 1$, we have

$$\sup_{W \in \mathcal{W}_\alpha} \inf_{\boldsymbol{\Omega} \in \mathcal{U}(\beta, r_0, \tau)} \mathbb{P}(W = 1) \leq 1 - \omega \tag{3.4}$$

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

This theorem shows that the power of any test will not converge to 1 uniformly over the class $\mathcal{U}(\beta, r_0, \tau)$ if the standardized signal strength $r_0 \tau^{-2}$ falls below the detection lower boundary $\mathrm{DB}(\beta)$ for $\beta \in (\max\{1/2, (3 - \xi\gamma^{-1})/4\}, 1)$. This provides the region of signals where the null hypothesis of (2.1) and the alternative hypotheses from the precision matrix class $\mathcal{U}(\beta, r_0, \tau)$ are theoretically non-separable. Note that the conclusion of Theorem 1 is valid under a broader class of $\mathcal{U}(\beta, r_0, \tau)$ without the constraints $\lambda_{\min}(\boldsymbol{\Sigma}) \geq C^{-1}$, $s \leq Cn^{\nu}$ for $\nu \in [0, 1/2)$ and $t_{\mathcal{A}_1} \leq C$. But the detection upper boundary could probably be inconsistent with the detection lower boundary under such a broader class.

Compared to the detection boundary for testing means, our result requires $\beta > (3 - \xi\gamma^{-1})/4$ due to the dependence among the pairwise product $X_{ij_1}X_{ij_2}$ of variables in the second moment of the likelihood ratio function in (S.4) in the online supplementary material. To derive the minimax lower boundary, we consider the edges (nonzero $\omega_{j_1j_2}$) in the sub-matrix $\boldsymbol{\Omega}_{\mathcal{A}_1}$ formed by a random perfect matching on m_a randomly selected variables from $\mathcal{A}_1$ under the alternative hypothesis. To control the second moment of the likelihood ratio, we need to control the summation of the term $\mathbb{E}\exp\{-\delta_a\sum_{i=1}^n(Z_{ij_1}Z_{ij_2} + Z_{ij_2}Z_{ij_3} + \ldots + Z_{ij_\ell}Z_{ij_{\ell+1}})\}(1 - \delta_a^2)^{\frac{\ell n}{2}} = \exp(\ell L_q/n)$ over all pairs of perfect matching graphs with randomly selected nodes, where $Z_{ij_1}, \ldots, Z_{ij_{\ell+1}} \overset{i.i.d.}{\sim} N(0, 1)$, $\delta_a = \sqrt{2r_0\tau^{-2}\log(q)/n}$ and $\ell < 2m_a$ is the number of overlapped nodes with different edges in two graphs. This requires $m_a^2/(np_1) \asymp p_1^{3-\xi\gamma^{-1}-4\beta} \to 0$. See the derivation from (S.19) to (S.24) in the online supplementary material. Note that those pairwise products $X_{ij_1}X_{ij_2}$ are formed from the nonzero values in $\boldsymbol{\Omega}_{\mathcal{A}_1}$, which are not present in the likelihood ratio for testing means of independent Gaussian variables.

Also notice that $(3 - \xi\gamma^{-1})/4 \leq 1/2$ when $\xi \geq \gamma$, meaning the derived minimax lower boundary covers the whole sparse signal regime $\beta \in (1/2, 1)$ if p_1 grows slower than or at the same rate of n. This also shows the impact of dimensionality on testing precision matrices. When $\xi < \gamma$, the minimax result for the range $\beta \in (1/2, (3 - \xi\gamma^{-1})/4)$ is unknown, which corresponds to denser signals compared with the range $\beta \in ((3 - \xi\gamma^{-1})/4, 1)$. Note that $(3 - \xi\gamma^{-1})/4 < (3 - \xi)/4$, and this range $(1/2, (3 - \xi\gamma^{-1})/4)$ with unknown minimax boundary becomes wider as the growth rate of p_1 increases.

In the next section, we propose a thresholding test that is generally applicable for the hypotheses in (2.1) and is powerful against rare and faint signals in $\boldsymbol{\Omega}_{\mathcal{A}_1}$. Particularly, in Section 6, we show that the proposed test is able to attain the minimax detection lower boundary derived in Theorem 1 so that the power of the test would converge to 1 uniformly over all $\boldsymbol{\Omega} \in \mathcal{U}(\beta, r_0, \tau)$ if $r_0\tau^{-2} > \mathrm{DB}(\beta)$ and $(\xi/\gamma)(1/2 - \nu) \geq 1$. Hence, the derived lower boundary $\mathrm{DB}(\beta)$ is tight for testing the hypotheses in (2.1) under this case.

4 Multi-level thresholding test

To construct a test that is able to achieve the minimax detection boundary, we first estimate the precision matrix $\boldsymbol{\Omega}$ via the nodewise regression approach (Chang et al., 2018; Liu, 2013). Consider the regression of X_{ij_1} on all the other variables:

$$X_{ij_1} - \mu_{j_1} = \sum_{j_2 \neq j_1} \beta_{j_1j_2}(X_{ij_2} - \mu_{j_2}) + \epsilon_{ij_1} \tag{4.1}$$

for $j_1 = 1, \ldots, p$. From Lemma 1 in Peng et al. (2009), the regression error ϵ_{ij_1} is uncorrelated with the covariates X_{ij_2} for $j_2 \neq j_1$ if and only if $\beta_{j_1j_2} = -\omega_{j_1j_2}/\omega_{j_1j_1}$ for all $j_2 \neq j_1$, and under such regression coefficients, we have $\mathrm{Cov}(\epsilon_{ij_1}, \epsilon_{ij_2}) = \omega_{j_1j_2}/(\omega_{j_1j_1}\omega_{j_2j_2})$. Therefore, an estimator of $\omega_{j_1j_2}$ can be constructed by the residuals from the nodewise regressions in (4.1).

Let $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^{\mathrm{T}}$, and $\mathbf{V} = \mathrm{Cov}(\boldsymbol{\epsilon}_i) = (v_{j_1j_2})$. Let $\boldsymbol{\Psi} = (\rho_{j_1j_2})$ be the correlation matrix of $\boldsymbol{\epsilon}_i$, where $\rho_{j_1j_2} = v_{j_1j_2}(v_{j_1j_1}v_{j_2j_2})^{-1/2}$ is the correlation between ϵ_{ij_1} and ϵ_{ij_2}. Let $\bar{\mathbf{X}} = (\bar{X}_1, \ldots, \bar{X}_p)^{\mathrm{T}} = \sum_{i=1}^n \mathbf{X}_i/n$ and $\widehat{\boldsymbol{\Sigma}} = (\widehat{\sigma}_{j_1j_2})$ be the sample mean and sample covariance matrix, respectively. We fit the high-dimensional nodewise regression in (4.1) by lasso (Tibshirani, 1996). Other regularized estimation methods can be applied as well, such as Dantzig selector (Candes & Tao, 2007) or scaled lasso (Sun & Zhang, 2012). Let $\boldsymbol{\beta}_j =$

$(\beta_{j1}, \ldots, \beta_{jp})^{\mathrm{T}}$ with $\beta_{jj} = -1$, and

$$\widehat{\boldsymbol{\beta}}_j = \underset{\boldsymbol{\beta}_j \in \mathbb{R}^p, \beta_{jj}=-1}{\operatorname{argmin}} \left[\frac{1}{2n} \sum_{i=1}^{n} \{\boldsymbol{\beta}_j^{\mathrm{T}}(\mathbf{X}_i - \bar{\mathbf{X}})\}^2 + \lambda_j \sum_{k \neq j} \widehat{\sigma}_{kk}^{1/2} |\beta_{jk}| \right] \tag{4.2}$$

be its lasso estimator, where λ_j is the penalty parameter. For each $i = 1, \ldots, n$, let $\widehat{\boldsymbol{\epsilon}}_i = (\widehat{\epsilon}_{i1}, \ldots, \widehat{\epsilon}_{ip})^{\mathrm{T}}$ be the residuals of fitting (4.1), where $\widehat{\epsilon}_{ij} = -\widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}(\mathbf{X}_i - \bar{\mathbf{X}})$. Following Liu (2013) and Chang et al. (2018), the bias corrected estimator for $\omega_{j_1 j_2}$ is

$$\widehat{\omega}_{j_1 j_2} = \frac{\widehat{v}_{j_1 j_2}}{\widehat{v}_{j_1 j_1} \widehat{v}_{j_2 j_2}} \text{ where } \widehat{v}_{j_1 j_2} = -\frac{1}{n} \sum_{i=1}^{n} (\widehat{\epsilon}_{ij_1} \widehat{\epsilon}_{ij_2} + \widehat{\beta}_{j_1 j_2} \widehat{\epsilon}_{ij_2}^2 + \widehat{\beta}_{j_2 j_1} \widehat{\epsilon}_{ij_1}^2). \tag{4.3}$$

Note that $\widehat{v}_{jj} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\epsilon}_{ij}^2$ since $\widehat{\beta}_{jj}$ is set as -1 for all $j = 1, \ldots, p$.

Based on the asymptotic expansion of $\widehat{\omega}_{j_1 j_2}$ provided in Lemma 1, the main order term of $\mathrm{Var}\{\sqrt{n}(\widehat{\omega}_{j_1 j_2} - \omega_{j_1 j_2})\}$ is $\theta_{j_1 j_2} = \mathrm{Var}(\epsilon_{ij_1} \epsilon_{ij_2}) / (v_{j_1 j_1}^2 v_{j_2 j_2}^2) = (1 + \rho_{j_1 j_2}^2)(v_{j_1 j_1} v_{j_2 j_2})^{-1}$, which can be estimated as $\widehat{\theta}_{j_1 j_2} = (\widehat{v}_{j_1 j_1} \widehat{v}_{j_2 j_2})^{-1}$ under the null hypothesis $\omega_{j_1 j_2} = 0$. Let $\widehat{\rho}_{j_1 j_2} = \widehat{v}_{j_1 j_2} (\widehat{v}_{j_1 j_1} \widehat{v}_{j_2 j_2})^{-1/2}$. Note that $-\widehat{\rho}_{j_1 j_2}$ is the estimated partial correlation between the j_1th and j_2th variables while controlling all other variables (Qiu & Zhou, 2020). The standardized statistic for testing $\omega_{j_1 j_2} = 0$ can be constructed as

$$V_{j_1 j_2} = n \widehat{\omega}_{j_1 j_2}^2 \widehat{\theta}_{j_1 j_2}^{-1} = n \widehat{\rho}_{j_1 j_2}^2. \tag{4.4}$$

To construct a test for the global hypotheses in (2.1) that is powerful against sparse and weak signals, the key is to screen out the non-signal components in $\boldsymbol{\Omega}_{\mathcal{A}_1}$. Therefore, we consider to aggregate the informative $V_{j_1 j_2}$ with nonzero values of $\omega_{j_1 j_2}$ together. Let

$$T(t) = \sum_{j_1, j_2 \in \mathcal{A}_1, j_1 < j_2} V_{j_1 j_2} \mathbb{I}\{V_{j_1 j_2} \geq \lambda_q(t)\} \tag{4.5}$$

be a thresholding statistic with a threshold level $t \in (0, 1)$, where $\lambda_q(t) = 2t \log q$. Xia et al. (2018) proposed an L_2 type statistic for testing (2.1) with a finite $\mathcal{A}_1$, which sums all $V_{j_1 j_2}$ over $j_1, j_2 \in \mathcal{A}_1$. Comparing to the L_2 statistic, $T(t)$ can be viewed as a regularized summation of $V_{j_1 j_2}$ by thresholding, which removes small values of $V_{j_1 j_2}$ that are unlikely to carry signals. The thresholding component is able to reduce the variance of $T(t)$, and hence, to increase its signal-to-noise ratio to detect sparse signals. A similar formulation of the thresholding statistic has been used for testing means (Zhong et al., 2013) and regression coefficients (Qiu et al., 2018). Furthermore, we allow the cardinality of the test set $\mathcal{A}_1$ to diverge, while the L_2 test requires a finite size of $\mathcal{A}_1$. This is due to the L_2 statistic suffers from error accumulation as $\widehat{\omega}_{j_1 j_2}^2$ is not an unbiased estimator of $\omega_{j_1 j_2}^2$. However, this problem can be avoided by our thresholding procedure. Comparing to the $L_{\max}$ type test (Xia et al., 2015), the thresholding test uses all $V_{j_1 j_2}$ over a threshold. This increases signals in the proposed statistic and results in higher power in detecting weak signals. Therefore, the thresholding statistic is more powerful than the L_2 and $L_{\max}$ type tests to detect signals that are both sparse and weak.

As $\sqrt{n} \widehat{\rho}_{j_1 j_2}$ is asymptotic normal distributed if $\rho_{j_1 j_2} = 0$, by the large deviation result established in Lemma S1 in the online supplementary material, the tail distribution of $V_{j_1 j_2}$ can be approximated by that of the Chi-square distribution $\mathcal{X}_1^2$ with 1 degree of freedom. It follows that

$$\mathbb{P}\left(\max_{j_1, j_2 \in \mathcal{A}_1, j_1 < j_2} V_{j_1 j_2} \geq 2 \log q \mid H_0 \right) = o(1). \tag{4.6}$$

Therefore, the constraint of the threshold level less than 1 prevents zero value for the thresholding statistic $T(t)$ under the null hypothesis.

Let $\phi(\cdot)$ and $\Phi(\cdot)$ be the density and cumulative distribution functions of the standard normal distribution, respectively, and let $\bar{\Phi}(x) = 1 - \Phi(x)$. As the tail probability of $V_{j_1 j_2}$ can be approximated by that of the Chi-square distribution $\mathcal{X}_1^2$ from Lemma S1 in the online supplementary material, the mean and variance of $V_{j_1 j_2}$ can be estimated by those of $\mathcal{X}_1^2 \mathbb{I}\{\mathcal{X}_1^2 > \lambda_q(t)\}$. Let

$$\begin{aligned} \tilde{\mu}_0(t) &= q[2\lambda_q^{1/2}(t)\phi\{\lambda_q^{1/2}(t)\} + 2\bar{\Phi}\{\lambda_q^{1/2}(t)\}] \text{ and} \\ \tilde{\sigma}_0^2(t) &= q[2\{\lambda_q^{3/2}(t) + 3\lambda_q^{1/2}(t)\}\phi\{\lambda_q^{1/2}(t)\} + 6\bar{\Phi}\{\lambda_q^{1/2}(t)\}]. \end{aligned} \tag{4.7}$$

It can be shown that $\tilde{\mu}_0(t)$ and $\tilde{\sigma}_0^2(t)$ are the main order terms of $\mathbb{E}\{T(t)\}$ and $\mathrm{Var}\{T(t)\}$ under the null hypothesis in (2.1), respectively. Note that $\tilde{\mu}_0(t)/q$ and $\tilde{\sigma}_0^2(t)/q$ are the mean and variance of $\mathcal{X}_1^2 \mathbb{I}\{\mathcal{X}_1^2 > \lambda_q(t)\}$, respectively. Theorem 2 in Section 5 establishes the asymptotic normality of $\{T(t) - \tilde{\mu}_0(t)\}\tilde{\sigma}_0^{-1}(t)$ under the null hypothesis, which implies a single-level thresholding test with the p-value $\mathrm{pv}(t) = \mathbb{P}[N(0, 1) > \{T(t) - \tilde{\mu}_0(t)\}\tilde{\sigma}_0^{-1}(t)]$. The null hypothesis of (2.1) is rejected if $\mathrm{pv}(t) < \alpha$.

Clearly, both the statistic $T(t)$ and its associated p-value $\mathrm{pv}(t)$ depend on the threshold level t. To choose the threshold that maximizes the power for testing (2.1), we consider minimizing the p-values $\{\mathrm{pv}(t)\}$ of the single-level thresholding test over t. As the p-value is determined by the standardized thresholding statistic $\{T(t) - \tilde{\mu}_0(t)\}\tilde{\sigma}_0^{-1}(t)$, minimizing the p-values is equivalent to maximizing the standardized values of $T(t)$ over the threshold level t. This motivates us to construct a multi-level thresholding statistic

$$\mathcal{M}(t_0) = \sup_{t_0 < t \le 1-\eta} \frac{T(t) - \tilde{\mu}_0(t)}{\tilde{\sigma}_0(t)}, \tag{4.8}$$

where t_0 is the threshold lower bound and η is an arbitrarily small positive constant, which prevents the threshold reaches to 1 for the reason explained in (4.6). By maximizing over t, the statistic $\mathcal{M}(t_0)$ is a more critical measure for the null hypothesis than the single thresholding statistic $T(t)$. This is in the same spirit as the HC tests for means (Donoho & Jin, 2004; Hall & Jin, 2010). Here, the lower bound t_0 avoids small threshold values and regulates the dependence among $\{V_{j_1 j_2}\}$ in $T(t)$. More discussions on the threshold lower bound and its impact on the power of the test are provided after Theorems 2 and 4.

Let $\mathcal{S}(t_0) = \{t_{j_1 j_2} : t_{j_1 j_2} = V_{j_1 j_2}/(2\log q) \text{ and } t_0 < t_{j_1 j_2} < (1-\eta)\}$ be the set of threshold levels at which the value of $T(t)$ changes. Since both $\tilde{\mu}_0(t)$ and $\tilde{\sigma}_0(t)$ are monotone functions of t, it follows that $\mathcal{M}(t_0)$ can be equivalently expressed as

$$\mathcal{M}(t_0) = \sup_{t \in \mathcal{S}(t_0)} \frac{T(t) - \tilde{\mu}_0(t)}{\tilde{\sigma}_0(t)}, \tag{4.9}$$

which avoids calculation of the standardized statistic $\{T(t) - \tilde{\mu}_0(t)\}\tilde{\sigma}_0^{-1}(t)$ for all values of t in the interval $(t_0, 1-\eta]$. It is shown in Theorem 2 that $\mathcal{M}(t_0)$ converges to a Gumbel distribution as $n, p_1 \to \infty$. Based on this result, we construct the multi-level thresholding test (MTT) that rejects H_0 in (2.1) at the significance level α if

$$\mathcal{M}(t_0) > [q_\alpha + b\{\log(q), t_0, \eta\}]/a\{\log(q)\}, \tag{4.10}$$

where $q_\alpha = -\log\log\{(1-\alpha)^{-1}\}$ is the upper α quantile of the Gumbel distribution, $a(y) = \{2\log(y)\}^{1/2}$ and $b(y, t_0, \eta) = 2\log(y) + 2^{-1}\log\log(y) - 2^{-1}\log(\pi) + \log\{(1 - t_0 - \eta)/2\}$.

From the proof of Theorem 4, if the minimum standardized signal strength $r_0\tau^{-2}$ is above the detection boundary, defined in (6.2), of the proposed multi-level thresholding test at the sparsity level β, there exists a threshold level $t^*(\beta, r_0\tau^{-2})$ such that the power of the single-level thresholding test with the threshold level $t^*(\beta, r_0\tau^{-2})$ converges to 1 uniformly over $\mathbf{\Omega} \in \mathcal{U}(\beta, r_0, \tau)$. However, this threshold level $t^*(\beta, r_0\tau^{-2})$ is unknown in practice and changing with respect to β and $r_0\tau^{-2}$. The logic of multi-level thresholding is to conduct the single-level thresholding test for a sequence of threshold levels and to choose the most significant one. This makes the multi-level thresholding test more powerful and adaptive to signals with different sparsity levels and strengths. In this way, the multi-level thresholding procedure makes the proposed test in (4.10) optimal in terms of achieving the minimax detection boundary $\mathrm{DB}(\beta)$ under the precision matrix class $\mathcal{U}(\beta, r_0, \tau)$.

5 Asymptotics under null hypothesis

To derive the theoretical properties of the proposed thresholding statistics, we make the following regularity conditions. Recall that $\lambda_{\min}(\boldsymbol{\Sigma})$ is the minimum eigenvalue of $\boldsymbol{\Sigma}$, and $s = \max_{1 \le j_1 \le p} \sum_{j_2=1}^{p} \mathbb{I}(\omega_{j_1 j_2} \neq 0)$. Without loss of generality, we assume $\mathbb{E}(\mathbf{X}_i) = \mathbf{0}$. Let C be a positive constant which may change from case to case.

Condition 1 The dimension $p \to \infty$ as $n \to \infty$ such that $n \asymp p^{\xi}$ for $\xi > 0$.

Condition 2 For a positive constant C, $\max_{1 \le j \le p} \sigma_{jj} \le C$ and $\lambda_{\min}(\boldsymbol{\Sigma}) \ge C^{-1}$.

Condition 3 Suppose that $s \le Cn^{\nu}$ for $\nu \in [0, 1/2)$.

Condition 1 prescribes the relationship between n and p, where p can grow at a polynomial rate of n. Our theoretical results could allow an exponential growth rate of p, but this requires a more restrictive condition on the sparsity of $\boldsymbol{\Omega}$ such that $s \le Cn^{\nu}/\log p$. Condition 2 assumes the minimum eigenvalue of $\boldsymbol{\Sigma}$ is bounded away from zero, which is used to guarantee the compatibility condition for all p nodewise regressions (Bühlmann & van de Geer, 2011). Since $\omega_{jj}\sigma_{jj} \ge 1$ for all j, this condition implies $C^{-1} \le \min\{\omega_{jj}, \sigma_{jj} : j = 1, \ldots, p\} \le \max\{\omega_{jj}, \sigma_{jj} : j = 1, \ldots, p\} \le C$. Since $\theta_{j_1 j_2} = (1 + \rho^2_{j_1 j_2})\omega_{j_1 j_1}\omega_{j_2 j_2}$, Condition 2 also leads to bounded $\{\theta_{j_1 j_2}\}$ away from zero and ∞, which avoids the singularity issue in the standardization of $\widehat{\omega}_{j_1 j_2}$. The maximum eigenvalue $\lambda_{\max}(\boldsymbol{\Sigma})$ of $\boldsymbol{\Sigma}$ is not necessarily bounded from above under Condition 2. Note that the condition of $\lambda_{\max}(\boldsymbol{\Sigma}) \le C$ is not needed for the de-sparsified lasso estimator in van de Geer et al. (2014) and the nodewise regression approach for estimating precision matrices in Liu (2013).

Condition 3 imposes sparse structure on the precision matrix $\boldsymbol{\Omega}$ such that the nonzero values in each row of $\boldsymbol{\Omega}$ do not exceed the order n^{ν} for $\nu \in [0, 1/2)$. Under the null hypothesis of (2.1), $\boldsymbol{\Omega}_{\mathcal{A}_1}$ is a diagonal matrix, but we allow the elements of $\boldsymbol{\Omega}$ in the sub-blocks $\mathcal{A}_1 \times \mathcal{A}_2$ and $\mathcal{A}_2 \times \mathcal{A}_2$ to be nonzero. Note that for high-dimensional regression, a weaker condition that the number of nonzero regression coefficients being at a smaller order of $n/\log(p)$ is sufficient for the convergence of the lasso type estimators in the prediction and ℓ_2 norms under the compatibility or restricted eigenvalue conditions on the design matrix (Bühlmann & van de Geer, 2011). However, the statistical inference of de-biased estimators of regularized estimation typically requires a stronger sparsity condition. Particularly, for estimating a sparse precision matrix, Ren et al. (2015) showed that $s\log(p)/n + n^{-1/2}$ is the rate of the minimax lower bound for the error of estimating $\omega_{j_1 j_2}$, and the condition $s = o\{n^{1/2}(\log p)^{-1}\}$ is needed to derive the asymptotic distribution of $\sqrt{n}(\widehat{\omega}_{j_1 j_2} - \omega_{j_1 j_2})$. The same sparsity condition is also required for the de-sparsified lasso estimator for the inference of regression coefficients (van de Geer et al., 2014; Zhang & Zhang, 2014). Comparing to the sparsity conditions $s = o\{n^{1/2}(\log p)^{-1}\}$ (Ren et al., 2015) and $s = o\{n^{1/2}(\log p)^{-3/2}\}$ (Chang et al., 2018; Liu, 2013; Xia et al., 2015) on precision matrices in literature, our Condition 3 is slightly stronger. This is because we need to uniformly control the small order terms in the expansion of $\widehat{\omega}_{j_1 j_2}$ over all $j_1, j_2 \in \mathcal{A}_1$ and derive its Cramér type large deviation result with an explicit error bound.

The following lemma gives the asymptotic expansion of the estimated precision coefficient $\widehat{\omega}_{j_1 j_2}$, which prepares for the large deviation results of $V_{j_1 j_2}$ and the theoretical properties of $T(t)$. Recall that $a_n \asymp b_n$ means two sequences $\{a_n\}$ and $\{b_n\}$ being at the same order, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^{\mathrm{T}}$ is the nodewise regression errors in (4.1).

Lemma 1 Under Conditions 1–3, $\lambda_j = \psi\omega_{jj}^{-1/2}\{\log(p)/n\}^{1/2}$ for any constant $\psi > 2$ and $j = 1, \ldots, p$, it holds that

$$\widehat{\omega}_{j_1 j_2} - \omega_{j_1 j_2} = -(\widetilde{\upsilon}_{j_1 j_2} - \upsilon_{j_1 j_2})/(\upsilon_{j_1 j_1}\upsilon_{j_2 j_2}) + O_p\{s\log(p)/n\}, \tag{5.1}$$

where $\widetilde{\upsilon}_{j_1 j_2} = n^{-1}\sum_{i=1}^{n} \epsilon_{ij_1}\epsilon_{ij_2}$ and $O_p\{s\log(p)/n\}$ is a higher order term uniformly for all j_1 and j_2. Furthermore, $\max_{j_1, j_2 \in \mathcal{A}_1} |\widehat{\theta}_{j_1 j_2}/\theta_{j_1 j_2} - 1| = O_p\{n^{-1/2}(\log p)^{1/2}\}$ under H_0 of (2.1).

Lemma 1 requires the penalty parameters $\{\lambda_j\}$ not too small. Since $\omega_{jj}^{-1} \le \sigma_{jj}$, an initial estimate $\widehat{\omega}_{jj,(1)}$ of ω_{jj} can be obtained by choosing the penalty parameter in (4.2) as $2\{\widehat{\sigma}_{jj}\log(p)/n\}^{1/2}$. Then, the updated penalty parameter $\lambda_j = \psi\widehat{\omega}_{jj,(1)}^{-1/2}\{\log(p)/n\}^{1/2}$ is used to re-estimate the nodewise regression in (4.2), where ψ is set as 2.1, a constant slightly larger than 2. Cross-validation selection of ψ that restricts $\psi > 2$ can be implemented as well. Both fixed and cross-validated ψ are implemented in the simulation study.

Since $s\log p = o(\sqrt{n})$ under Condition 3, Lemma 1 implies the asymptotic normality of the estimator $\widehat{\omega}_{j_1j_2}$ for each entry of $\mathbf{\Omega}$. Similar result on the expansion of $\widehat{\omega}_{j_1j_2}$ is given in Chang et al. (2018). Here, we additionally provide the exact rate $s\log(p)/n$ of the small order term in $\widehat{\omega}_{j_1j_2}$, which is needed to derive the Cramér type large deviation result for $\widehat{\omega}_{j_1j_2}$, and to control the small order terms in the thresholding statistic $T(t)$.

Specifically, let $V^*_{j_1j_2} = n\widetilde{v}^2_{j_1j_2}/(\theta_{j_1j_2}v^2_{j_1j_1}v^2_{j_2j_2}) = n\widetilde{v}^2_{j_1j_2}/\mathrm{Var}(\epsilon_{ij_1}\epsilon_{ij_2})$. From Lemma 1, since $\widehat{\omega}_{j_1j_2} = -\widetilde{v}_{j_1j_2}/(v_{j_1j_1}v_{j_2j_2}) + O_p\{s\log(p)/n\}$ under H_0 of (2.1), we have

$$V_{j_1j_2} = n\widehat{\omega}^2_{j_1j_2}\widehat{\theta}^{-1}_{j_1j_2} = V^*_{j_1j_2} + O_p\{s\log^{3/2}(p)/\sqrt{n}\}$$

for $\omega_{j_1j_2} = 0$. Although the error terms $O_p\{s\log^{3/2}(p)/\sqrt{n}\}$ accumulates in the proposed statistic $T(t)$ by the summation over $j_1, j_2 \in \mathcal{A}_1$, due to its thresholding component $\mathbb{I}\{V_{j_1j_2} \ge \lambda_q(t)\}$ to remove noisy non-signal components in $T(t)$, those small order terms are negligible in $T(t)$ by choosing a proper range for the threshold level t. This range of t is given in the following theorem, which establishes the asymptotic normality of $\{T(t) - \tilde{\mu}_0(t)\}/\tilde{\sigma}_0(t)$, where $\tilde{\mu}_0(t)$ and $\tilde{\sigma}^2_0(t)$ are defined in (4.7).

Theorem 2 Under Conditions 1–3, H_0 of (2.1), $\lambda_j = \psi\omega_{jj}^{-1/2}\{\log(p)/n\}^{1/2}$ for any constant $\psi > 2$ and $j = 1, \ldots, p$, $p_1 \asymp p^\gamma$ for $\gamma \in (0, 1]$, and the threshold level t that satisfies $q^{1/2-t/2}n^{\nu-1/2} = o(1)$, we have

$$\tilde{\sigma}_0^{-1}(t)\{T(t) - \tilde{\mu}_0(t)\}d \to N(0, 1) \text{ as } n, p, q \to \infty.$$

Theorem 2 shows that the single-level thresholding test which rejects the null hypothesis in (2.1) if $T(t) > \tilde{\mu}_0(t) + z_\alpha\tilde{\sigma}_0(t)$ is able to control the size at α asymptotically, where z_α denotes the upper α quantile of the standard normal distribution. The condition $q^{1/2-t/2}n^{\nu-1/2} = o(1)$ imposes a lower bound on the threshold level t. This threshold lower bound is needed to derive the asymptotic properties of $T(t)$ for three reasons. First, to prove Theorem 2, we approximate the statistic $V_{j_1j_2}$ by $V^*_{j_1j_2} = n\widetilde{v}^2_{j_1j_2}/\mathrm{Var}(\epsilon_{ij_1}\epsilon_{ij_2})$ using the true errors of the nodewise regressions. A larger threshold level is needed to control the accumulated estimation error of $\widehat{\omega}_{j_1j_2}$ in $T(t)$ due to the lasso estimation. Second, the result of Theorem 2 requires the bias for estimating the mean of $T(t)$ being sufficiently small such that $|\mathbb{E}\{T(t)|H_0\} - \tilde{\mu}_0(t)| = o\{\tilde{\sigma}_0(t)\}$, where $|\mathbb{E}\{T(t)|H_0\} - \tilde{\mu}_0(t)|/\tilde{\sigma}_0(t) \asymp s(\log p)^{3/2} q^{(1-t)/2}n^{-1/2}$. The bias of $\tilde{\mu}_0(t)$ is due to the error term $O\{s(\log p)^{3/2}n^{-1/2}\}$ in the Cramér-type tail probability of $V_{j_1j_2}$. Note that this error term still exists in the large deviation result of $\widetilde{v}_{j_1j_2}$ based on the true regression errors, but it is at a smaller order $O\{(\log p)^{3/2}n^{-1/2}\}$. The extra factor s is due to the ℓ_1 estimation error of the nodewise regression coefficients. Please see Lemmas S1 and S2 in the online supplementary material. Due to a similar reason, threshold lower bounds are required for thresholding statistics built on sample means under non-Gaussian data (Zhong et al., 2013), component-wise t-statistics (Delaigle et al., 2011) and maximum likelihood estimators (Qiu et al., 2018). Third, as the covariances among $\{V_{j_1j_2}\mathbb{I}\{V_{j_1j_2} \ge \lambda_q(t)\}\}$ decrease with the increase of the threshold level t, a larger value of t is also needed to mitigate the dependence in the thresholding statistic $T(t)$.

Note that, when $p_1 \asymp p^\gamma$ for $\gamma \in (0, 1]$, and equivalently $q \asymp p^{2\gamma}$, the condition $q^{1/2-t/2}n^{\nu-1/2} = o(1)$ implies $t > \max\{1 - (\xi/\gamma)(1/2 - \nu), 0\}$. When $\gamma \ge 1/2$ meaning the number of the testing parameters in $\mathbf{\Omega}_{\mathcal{A}_1}$ is at least at the order p, since $(1/2 - \nu)/\gamma < 1$, our required lower bound of t is more restrictive than the requirement $t > \max\{1 - \xi, 0\}$ in testing means (Zhong et al., 2013) and fixed-dimensional regression coefficients (Qiu et al., 2018). This is because estimating a high-dimensional precision matrix is much more difficult than estimating means and parameters in a fixed-dimensional model. From the nodewise regression in (4.1) and Lemma 1, estimation of

$\omega_{j_1j_2}$ requires estimating a high-dimensional set of nuisance parameters $\{\beta_{j_1j_2}\}$, which leads to a larger estimation error at the order $O_p\{s\log(p)/n\}$ for the estimator $\widehat{\omega}_{j_1j_2}$.

Based on the result in Theorem 2, we choose $t_0 = \max\{1-(\xi/\gamma)(1/2-\nu), 0\}$ for the multi-level thresholding statistic $\mathcal{M}(t_0)$ under the specification $p_1 \asymp p^\gamma$ for the size of $\mathcal{A}_1$. Recall that $a(y) = \{2\log(y)\}^{1/2}$ and $b(y, t_0, \eta) = 2\log(y) + 2^{-1}\log\log(y) - 2^{-1}\log(\pi) + \log\{(1-t_0-\eta)/2\}$. The following theorem shows the asymptotic distribution of $\mathcal{M}(t_0)$.

Theorem 3 Under Conditions 1–3, H_0 of (2.1), $\lambda_j = \psi\omega_{jj}^{-1/2}\{\log(p)/n\}^{1/2}$ for any constant $\psi > 2$ and $j = 1, \ldots, p$, $p_1 \asymp p^\gamma$ for $\gamma \in (0, 1]$, and $t_0 = \max\{1-(\xi/\gamma)(1/2-\nu), 0\}$, we have $\mathbb{P}[a\{\log(q)\}\mathcal{M}(t_0) - b\{\log(q), t_0, \eta\} \le x] \to \exp(-e^{-x})$.

Theorem 3 shows the multi-level thresholding statistic converges to an extreme value distribution. It verifies that the proposed multi-level thresholding test in (4.10) is able to control its size at α for testing the hypotheses in (2.1) when n, p, and q are large.

6 Power analysis

In this section, we study the power of the proposed multi-level thresholding test. We derive its detection boundary under the class $\mathcal{U}(\beta, r_0, \tau)$ in (3.2), which means the power of the proposed test would converge to 1 universally over $\mathcal{U}(\beta, r_0, \tau)$ for any signal above this boundary and show this boundary matches with the minimax detection lower boundary derived in Theorem 1 under certain conditions.

From Lemma 1, $\sqrt{n}(\widehat{\omega}_{j_1j_2} - \omega_{j_1j_2})\widehat{\theta}_{j_1j_2}^{-1/2}$ is asymptotic standard normal distributed if $s\log p = o(\sqrt{n})$. Therefore, $\sqrt{n}|\omega_{j_1j_2}|\theta_{j_1j_2}^{-1/2}$ is the standardized signal strength for testing $\omega_{j_1j_2}$. For $|\omega_{j_1j_2}| = \{2r_{j_1j_2}\log(q)/n\}^{1/2}$, its standardized signal strength is equal to $\{2(r_{j_1j_2}/\theta_{j_1j_2})\log(q)\}^{1/2}$, which is characterized by $r_{j_1j_2}/\theta_{j_1j_2}$. Here, $r_{j_1j_2}$ represents the strength of the nonzero $\omega_{j_1j_2}$, and $\theta_{j_1j_2}$ measures the variation of the estimator $\widehat{\omega}_{j_1j_2}$.

For a precision matrix $\mathbf{\Omega} \in \mathcal{U}(\beta, r_0, \tau)$, the power of the proposed test is

$$\mathrm{Power}_n(\mathbf{\Omega}) = \mathbb{P}(\mathcal{M}(t_0) > [q_\alpha + b\{\log(q), t_0, \eta\}]/a\{\log(q)\} \mid \mathbf{\Omega}).$$

We consider the minimal power over the class $\mathcal{U}(\beta, r_0, \tau)$. Note that $\{2r_0\tau^{-2}\log q\}^{1/2}$ is the minimal standardized signal strength for all precision matrices from $\mathcal{U}(\beta, r_0, \tau)$, where $r_0\tau^{-2} = \min\{r_{j_1j_2}/\theta_{j_1j_2} : \mathbf{\Omega} \in \mathcal{U}(\beta, r_0, \tau)\}$. Let $\mathbf{\Omega}_0$ denote a precision matrix from $\mathcal{U}(\beta, r_0, \tau)$ with $q^{1-\beta}$ nonzero elements in the upper-triangle of $\mathbf{\Omega}_{0,\mathcal{A}_1}$, all nonzero values being $|\omega_{j_1j_2}| = \{2r_0\log(q)/n\}^{1/2}$, and $\omega_{jj} = \tau$ for all $j \in \mathcal{A}_1$. Then, the standardized signal strength for all nonzero elements in $\mathbf{\Omega}_{0,\mathcal{A}_1}$ are universally equal to the minimal value $\{2r_0\tau^{-2}\log q\}^{1/2}$. Intuitively, $\mathbf{\Omega}_0$ is the most difficult case in the class $\mathcal{U}(\beta, r_0, \tau)$ to separate the null and alternative hypotheses of (2.1). In the proof of Theorem 4 in the online supplementary material, we have shown that $\mathrm{Power}_n(\mathbf{\Omega}_0) = \min_{\mathbf{\Omega}\in\mathcal{U}(\beta,r_0,\tau)}\{\mathrm{Power}_n(\mathbf{\Omega})\}$. Therefore, it suffices to consider the power for testing $\mathbf{\Omega}_0$ to derive the detection boundary of the proposed test.

Let $L_p = c_1\log^{c_2}(p)$ denote a multi-$\log p$ term for some constants $c_1 > 0$ and c_2 which may change from case to case. It is shown in Section S3 of the online supplementary material that the main orders for the mean and variance of the thresholding statistic $T(t)$ under the precision matrix $\mathbf{\Omega}_0$ are $\tilde{\mu}_0(t) + \tilde{\mu}_s(t)$ and $\tilde{\sigma}_a^2(t) = \tilde{\sigma}_0^2(t) + \tilde{\sigma}_s^2(t)$, respectively, where $\tilde{\mu}_0(t)$ and $\tilde{\sigma}_0^2(t)$, given in (4.7), are the mean and variance of $T(t)$ under the null hypothesis, and the expressions of both $\tilde{\mu}_s(t)$ and $\tilde{\sigma}_s^2(t)$ are in the form

$$L_p q^{(1-\beta)}\mathbb{I}(t < r_0\tau^{-2}) + L_p q^{(1-\beta)-(\sqrt{t}-\sqrt{r_0\tau^{-2}})^2}\mathbb{I}(t > r_0\tau^{-2}). \tag{6.1}$$

Here, the multi-$\log p$ term L_p is different for $\tilde{\mu}_s(t)$ and $\tilde{\sigma}_s^2(t)$. Note that

$$\mathcal{M}(t_0) = \sup_{t_0<t\le 1-\eta}\frac{T(t)-\tilde{\mu}_0(t)}{\tilde{\sigma}_0(t)} = \sup_{t_0<t\le 1-\eta}\frac{\tilde{\sigma}_a(t)}{\tilde{\sigma}_0(t)}\left\{\frac{T(t)-\tilde{\mu}_0(t)-\tilde{\mu}_s(t)}{\tilde{\sigma}_a(t)} + \mathrm{SNR}(t)\right\},$$

where $\mathrm{SNR}(t) = \tilde{\mu}_s(t)\tilde{\sigma}_a^{-1}(t)$ is the signal-to-noise ratio for the single-level thresholding test with threshold level t. In the expression of $\mathrm{SNR}(t)$, the 'signal' is reflected by $\tilde{\mu}_s(t)$, which is the difference of $\mathbb{E}\{T(t)\}$ between the null and alternative hypothesis, and $\tilde{\sigma}_a(t)$ is the 'noise' part, which is the standard deviation of $T(t)$ under the alternative hypothesis.

As the rejection value for $\mathcal{M}(t_0)$ is a multi-log p term and $\{T(t) - \tilde{\mu}_0(t) - \tilde{\mu}_s(t)\}\tilde{\sigma}_a^{-1}(t) = O_p(1)$ under the alternative hypothesis, the power of the proposed multi-level thresholding test converges to 1 if the maximal signal-to-noise ratio $\sup_{t_0<t\le 1-\eta}\mathrm{SNR}(t)$ diverges to infinity at a polynomial rate of p. The expression of $\sup_{t_0<t\le 1-\eta}\mathrm{SNR}(t)$ can be derived based on the expressions of $\tilde{\sigma}_0^2(t)$, $\tilde{\mu}_s(t)$, and $\tilde{\sigma}_s^2(t)$ given in (4.7) and (6.1), which depends on the signal sparsity parameter β, the minimal standardized signal strength $r_0\tau^{-2}$ under $\mathcal{U}(\beta, r_0, \tau)$, and the threshold lower bound $t_0 = \max\{1 - (\xi/\gamma)(1/2 - \nu), 0\}$.

By deriving the range of β and $r_0\tau^{-2}$ which makes $\sup_{t_0<t\le 1-\eta}\mathrm{SNR}(t) \to \infty$ at a polynomial rate of p, our study establishes the detection boundary $\mathrm{DB}(\beta, \xi/\gamma, \nu)$ for the power of the proposed test, defined as

$$\mathrm{DB}(\beta, \xi/\gamma, \nu) = \begin{cases} \dfrac{\{\sqrt{8 - 8c(\xi/\gamma, \nu)} - \sqrt{8 - 8\beta - 4c(\xi/\gamma, \nu)}\}^2}{8}, & 1/2 < \beta \le 3/4 - c(\xi/\gamma, \nu)/4, \\ \beta - 1/2, & 3/4 - c(\xi/\gamma, \nu)/4 < \beta \le 3/4, \\ (1 - \sqrt{1-\beta})^2, & 3/4 < \beta < 1, \end{cases} \tag{6.2}$$

where $c(\xi/\gamma, \nu) = \min\{(\xi/\gamma)(1/2 - \nu), 1\}$. The following theorem shows that the power of the proposed test would converge to 1 uniformly over the class $\mathcal{U}(\beta, r_0, \tau)$ if $r_0\tau^{-2} > \mathrm{DB}(\beta, \xi/\gamma, \nu)$.

Theorem 4 Under Condition 1, the precision matrix class $\mathcal{U}(\beta, r_0, \tau)$ under the alternative hypothesis of (2.1), $p_1 \asymp p^\gamma$ for $\gamma \in (0, 1]$, $\lambda_j = \psi\omega_{jj}^{-1/2}\{\log(p)/n\}^{1/2}$ for any constant $\psi > 2$ and $j = 1, \ldots, p$, and $t_0 = \max\{1 - (\xi/\gamma)(1/2 - \nu), 0\}$, if $r_0\tau^{-2} > \mathrm{DB}(\beta, \xi/\gamma, \nu)$, we have $\min_{\boldsymbol{\Omega}\in\mathcal{U}(\beta,r_0,\tau)} \mathrm{Power}_n(\boldsymbol{\Omega}) \to 1$ as $n, p, q \to \infty$.

Theorem 4 provides a sufficient condition for the power of the proposed test converging to 1 over the class $\mathcal{U}(\beta, r_0, \tau)$ under the alternative hypothesis of (2.1). Comparing to the detection boundary $\mathrm{DB}(\beta)$ in (3.3) for testing means under independent normal distributions with unit variance (Ingster, 1997), the detection boundary $\mathrm{DB}(\beta, \xi/\gamma, \nu)$ of the proposed test is different from $\mathrm{DB}(\beta)$ when $1/2 < \beta \le \{3 - c(\xi/\gamma, \nu)\}/4$. From the expression in (6.2), it can be shown that $\mathrm{DB}(\beta, \xi/\gamma, \nu) > \mathrm{DB}(\beta)$ in this interval. Note that $\max\{1/2, (3 - \xi\gamma^{-1})/4\} \le \{3 - c(\xi/\gamma, \nu)\}/4$. The proposed test cannot attain the minimax detection lower boundary in the range $\max\{1/2, (3 - \xi\gamma^{-1})/4\} \le \beta \le \{3 - c(\xi/\gamma, \nu)\}/4$ if $(\xi/\gamma)(1/2 - \nu) < 1$. The elevated detection boundary of the proposed test is due to the imposed threshold lower bound t_0, which is used to control the aggregation of the small order terms in $\widehat{\omega}_{j_1j_2}$, the bias of $\tilde{\mu}_0(t)$ for estimating $\mathbb{E}\{T(t)|H_0\}$ and the dependence among $\{V_{j_1j_2}\mathbb{I}\{V_{j_1j_2} \ge \lambda_q(t)\}\}$ in the statistic $T(t)$, as discussed after Theorem 2. Note that such a threshold lower bound is not needed for testing means under independent Gaussian distributions as sample means are normally distributed. Testing for precision matrices is more challenging, as both the estimation of $\omega_{j_1j_2}$ and the dependence among the estimates $\widehat{\omega}_{j_1j_2}$ are much more involved and complicated than sample means.

If $(\xi/\gamma)(1/2 - \nu) \ge 1$, we have $\max\{1/2, (3 - \xi\gamma^{-1})/4\} = \{3 - c(\xi/\gamma, \nu)\}/4 = 1/2$, and hence, $\mathrm{DB}(\beta, \xi/\gamma, \nu) = \mathrm{DB}(\beta)$ for the whole range of sparse signal $\beta \in (1/2, 1)$. The results in Theorems 1 and 4 imply that under this case, $\mathrm{DB}(\beta)$ is the optimal detection boundary and the proposed test is sharp optimal in the sense that no test is able to separate the null and alternative hypotheses of (2.1) uniformly over $\mathcal{U}(\beta, r_0, \tau)$ if $r_0\tau^{-2}$ is lower than $\mathrm{DB}(\beta)$, and the power of the proposed test approaches to 1 if $r_0\tau^{-2}$ is above this boundary. The condition $(\xi/\gamma)(1/2 - \nu) \ge 1$ means that, given the sparsity level $s = O(n^\nu)$ of $\boldsymbol{\Omega}$ for $\nu \in [0, 1/2)$, the proposed test is able to achieve the whole minimax detection boundary over $\beta \in (1/2, 1)$ if the size of the testing set $\mathcal{A}_1$ is sufficiently small or the sample size is sufficiently large.

The minimax undetectable region is plotted in Figure 1 together with the detection boundary $\mathrm{DB}(\beta, \xi/\gamma, \nu)$ of the proposed multi-level thresholding test. Figure 1 shows the minimax detection boundary $\mathrm{DB}(\beta)$ covers the whole sparse regime $\beta \in (1/2, 1)$ when $\xi = 1, 0.8$ and $\gamma = 0.4, 0.8$. For the case $\xi = 1, \gamma = 0.4$ and $\nu = 0.1$ in the left panel, the proposed test is able to achieve the minimax

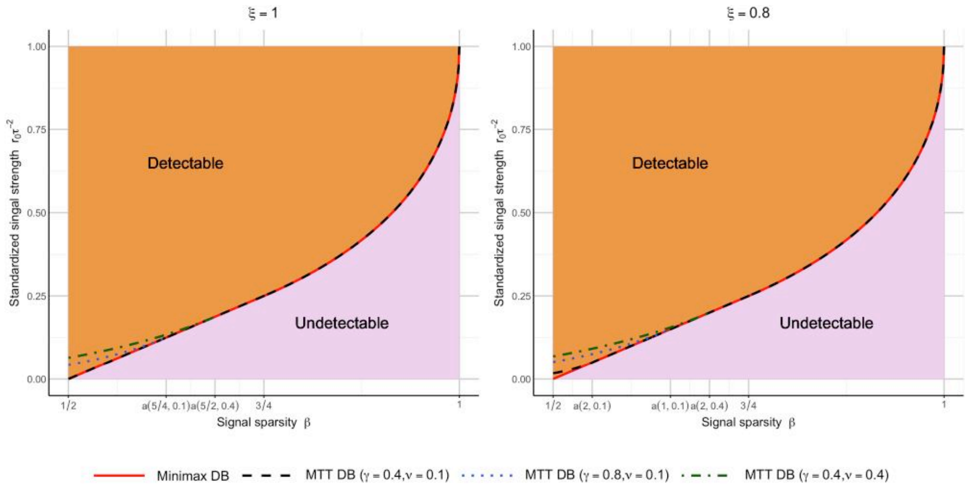

Figure 1. The minimax detection boundary DB(β) and the detection boundary DB(β, ξ/γ, ν) of the proposed multi-level thresholding test for $\xi = 1$ (left panel) and 0.8 (right panel), where $a(\xi/\gamma, \nu) = 3/4 - c(\xi/\gamma, \nu)/4$ on the x-axis indicating the point where DB(β, ξ/γ, ν) deviates from DB(β).

detection boundary DB(β) for $\beta \in (1/2, 1)$. However, for the other combinations of ξ, γ and ν considered in Figure 1, the proposed test is only able to achieve DB(β) for β larger than $3/4 - c(\xi/\gamma, \nu)/4$, denoted as $a(\xi/\gamma, \nu)$ in Figure 1. It has an elevated detection boundary for $1/2 < \beta < 3/4 - c(\xi/\gamma, \nu)/4$.

Let $\widetilde{\mathcal{U}}(\beta, r_0, \tau)$ be the class $\mathcal{U}(\beta, r_0, \tau)$ of precision matrices without the constraint $t_{\mathcal{A}_1} \leq C$. Now, we discuss the power of the proposed test under the class $\widetilde{\mathcal{U}}(\beta, r_0, \tau)$ that allows a diverging number of signals in the rows of $\boldsymbol{\Omega}$. One case is that all the nonzero precision coefficients locate in only one row of $\boldsymbol{\Omega}$, say $\mathcal{A}_1 = \{1, \ldots, p_1\}$ and $\omega_{12}, \ldots, \omega_{1m_a+1} \neq 0$. Under this case, the covariances among $\{V_{j_1j_2}\mathbb{I}\{V_{j_1j_2} \geq \lambda_q(t)\}\}$ could dominate the variance of $T(t)$ due to the clustered signal structure, which would lower the signal-to-noise ratio of the proposed test. The following corollary studies the power of the proposed test without the constraint $t_{\mathcal{A}_1} \leq C$.

Corollary 1 Under Condition 1, the precision matrix class $\widetilde{\mathcal{U}}(\beta, r_0, \tau)$, $p_1 \asymp p^\gamma$ for $\gamma \in (0, 1]$, $\lambda_j = \psi\omega_{jj}^{-1/2}\{\log(p)/n\}^{1/2}$ for any constant $\psi > 2$ and $j = 1, \ldots, p$, and $t_0 = \max\{1 - (\xi/\gamma)(1/2 - \nu), 0\}$, if $r_0\tau^{-2} > \text{DB}(\beta, \xi/\gamma, \nu)$ and $r_0\tau^{-2} > t_0$, we have $\min_{\boldsymbol{\Omega}\in\widetilde{\mathcal{U}}(\beta, r_0, \tau)} \text{Power}_n(\boldsymbol{\Omega}) \to 1$ as $n, p, q \to \infty$.

Corollary 1 shows that the power of MTT can still converge to 1 under $\widetilde{\mathcal{U}}(\beta, r_0, \tau)$, but it requires a stronger condition $r_0\tau^{-2} > t_0$ in addition to $r_0\tau^{-2} > \text{DB}(\beta, \xi/\gamma, \nu)$ on the signal strength, compared to the power analysis under $\mathcal{U}(\beta, r_0, \tau)$ in Theorem 4. Note that the condition $r_0\tau^{-2} > t_0$ trivially holds if $\xi/\gamma \geq (1/2 - \nu)^{-1}$. Meanwhile, if the signals are clustered in at least one row of $\boldsymbol{\Omega}$ under H_a of (2.1), the test statistic $R(t) = \max_{j_1\in\mathcal{A}_1}\sum_{j_2\in\mathcal{A}_1} V_{j_1j_2}\mathbb{I}\{V_{j_1j_2} \geq \lambda_q(t)\}$ that takes the maximal accumulative signals per row could be more powerful than $T(t)$.

7 Numerical study

In this section, we evaluate the performance of the proposed multi-level thresholding test, and compare it with the $L_{\max}$ type test in Xia et al. (2015) (denoted as $L_{\max}$-test) and the L_2 type test in Xia et al. (2018) (denoted as L_2-test). We consider both the overall test for $\boldsymbol{\Omega}$ being diagonal and the test for precision coefficients of a subset of variables. Note that the L_2-test was constructed for testing a finite number of precision coefficients, which is not applicable for testing diverging sizes of sub-blocks or the whole precision matrix. Meanwhile, the $L_{\max}$-test was originally designed for testing two precision matrices being the same by utilizing the maximum difference between the two-sample estimates. We modify it for the one-sample hypotheses in (2.1) by the maximum of $|\widehat{\omega}_{j_1j_2}|$ over $\mathcal{A}_1$.

Table 2. Empirical sizes of the proposed MTT with fixed $\psi = 2.1$ ($\text{MTT}_{2.1}$) and cross-validated ψ (MTT_{cv}) for the lasso penalty parameter λ_j in (4.2), and the $L_{\max}$-test for the hypotheses $H_0 : \omega_{j_1 j_2} = 0$ for all $j_1 \neq j_2$, under the homogeneous case ($\omega_{jj} = 1$ for all j) and two heterogeneity cases (ω_{jj} generated from the uniform distribution $U(a, b)$)

		$\omega_{jj} = 1$			$\omega_{jj} \sim U(0.5, 2.5)$			$\omega_{jj} \sim U(0.1, 5)$		
n	p	$\text{MTT}_{2.1}$	MTT_{cv}	$L_{\max}$	$\text{MTT}_{2.1}$	MTT_{cv}	$L_{\max}$	$\text{MTT}_{2.1}$	MTT_{cv}	$L_{\max}$
100	250	0.059	0.060	0.012	0.061	0.064	0.010	0.062	0.063	0.012
	350	0.062	0.062	0.014	0.057	0.059	0.015	0.064	0.066	0.010
	450	0.066	0.064	0.011	0.064	0.057	0.012	0.062	0.067	0.013
125	250	0.048	0.041	0.016	0.044	0.048	0.015	0.055	0.046	0.012
	350	0.041	0.055	0.018	0.050	0.047	0.015	0.051	0.045	0.022
	450	0.048	0.041	0.018	0.051	0.043	0.014	0.058	0.054	0.019
150	250	0.045	0.045	0.021	0.044	0.049	0.026	0.058	0.057	0.019
	350	0.045	0.052	0.015	0.052	0.058	0.020	0.057	0.049	0.025
	450	0.044	0.045	0.021	0.046	0.051	0.025	0.055	0.047	0.017

Note. The nominal level is 0.05.

We set $n = 100, 125, 150$ and $p = 250, 350, 450$, and determined the threshold lower bound as $t_0 = \max\{1 - (\xi/\gamma)(1/2 - \nu), 0\}$ from Theorem 3, where ξ and ν were treated as 1 and 0 in the simulation, respectively, and $\gamma = 1$ for testing the whole precision matrix $\mathbf{\Omega}$. This led to $t_0 = 0.5$ in the simulation. We still kept $t_0 = 0.5$ for testing sub-blocks of $\mathbf{\Omega}$, although it can be set smaller as $\gamma < 1$ for those cases. Note that $\nu = 0$ stands for at most finite number of nonzero components in each row of $\mathbf{\Omega}$. The technical constant η was set as 0.05. As discussed after Lemma 1, the lasso penalty parameter λ_j in (4.2) for the nodewise regression was chosen as $\lambda_j = \psi\widehat{\omega}_{jj,(1)}^{-1/2}\{\log(p)/n\}^{1/2}$, where $\widehat{\omega}_{jj,(1)}$ is the initial estimate of ω_{jj} using the penalty value $2\{\widehat{\sigma}_{jj}\log(p)/n\}^{1/2}$ for the nodewise regression. We used both fixed penalty $\psi = 2.1$ (denoted as $\text{MTT}_{2.1}$) and cross-validation (CV) selected ψ from the set $\{2 + k/10 : k = 1, \ldots, 30\}$ (denoted as MTT_{cv}), where the lower bound 2 was imposed on ψ to satisfy the condition of Lemma 1. The nominal size was taken as 0.05 and all simulation settings were carried out with 1,000 replications.

For testing $\mathbf{\Omega} = (\omega_{j_1 j_2})$ being diagonal, we considered the homogeneous variance case where $\omega_{jj} = 1$ for all $j = 1, \ldots, p$, and the heterogeneity case where ω_{jj} was randomly generated from the uniform distribution $U(a, b)$ for two positive constants $a \leq b$. We took the distribution $U(0.5, 2.5)$ for the moderate heterogeneity case and $U(0.1, 5)$ for the high heterogeneity case, where the largest value is 50 times larger than the smallest value on the diagonal of $\mathbf{\Omega}$. We set $\omega_{j_1 j_2} = 0$ for $j_1 \neq j_2$ under the null hypothesis. To evaluate power under sparse and weak signals, we randomly selected $m_a = \lfloor q^{(1-\beta)} \rfloor$ elements from the upper-triangle of $\mathbf{\Omega}$, and set those selected components as $\omega_{j_1 j_2} = \{2r\omega_{j_1 j_1}\omega_{j_2 j_2}(\log q)/n\}^{1/2}$ under the alternative hypothesis, where $q = p(p-1)/2$. The sparsity parameter β and the signal strength r were ranged from $\beta \in [0.5, 0.9]$ and $r \in (0, 0.5]$, respectively. If the precision matrix $\mathbf{\Omega}$ was not positively definite, we modified it as $\mathbf{\Omega} + \delta\mathbf{I}_p$, where $\delta = |\lambda_{\min}(\mathbf{\Omega})| + 0.01$ and $\lambda_{\min}(\mathbf{\Omega})$ is the minimal eigenvalue of $\mathbf{\Omega}$. Once generated, the precision matrix $\mathbf{\Omega}$ was set fixed over 1,000 repetitions. The data $\mathbf{X}_1, \ldots, \mathbf{X}_n$ were generated independently from $N(0, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = \mathbf{\Omega}^{-1}$.

Table 2 reports the empirical sizes of the proposed test ($\text{MTT}_{2.1}$ and MTT_{cv}) and the maximum test ($L_{\max}$-test) modified from Xia et al. (2015) for the hypothesis $H_0 : \omega_{j_1 j_2} = 0$ for all $j_1 \neq j_2$. Note that the L_2 type test is not applicable to this hypothesis of the whole precision matrix. From Table 2, we notice that the proposed test with both fixed and CV selected ψ had an accurate size around the nominal level 5%. This verifies the asymptotic results of the proposed thresholding statistics under the null hypothesis in Section 5. However, the $L_{\max}$-test was conservative with empirical sizes around 1%–2%.

Figure 2 shows the powers of $\text{MTT}_{2.1}$ with fixed ψ and the $L_{\max}$-test under $n = 125$ and the three cases of heterogeneity considered in Table 2. The results of MTT_{cv} by CV selection of ψ were

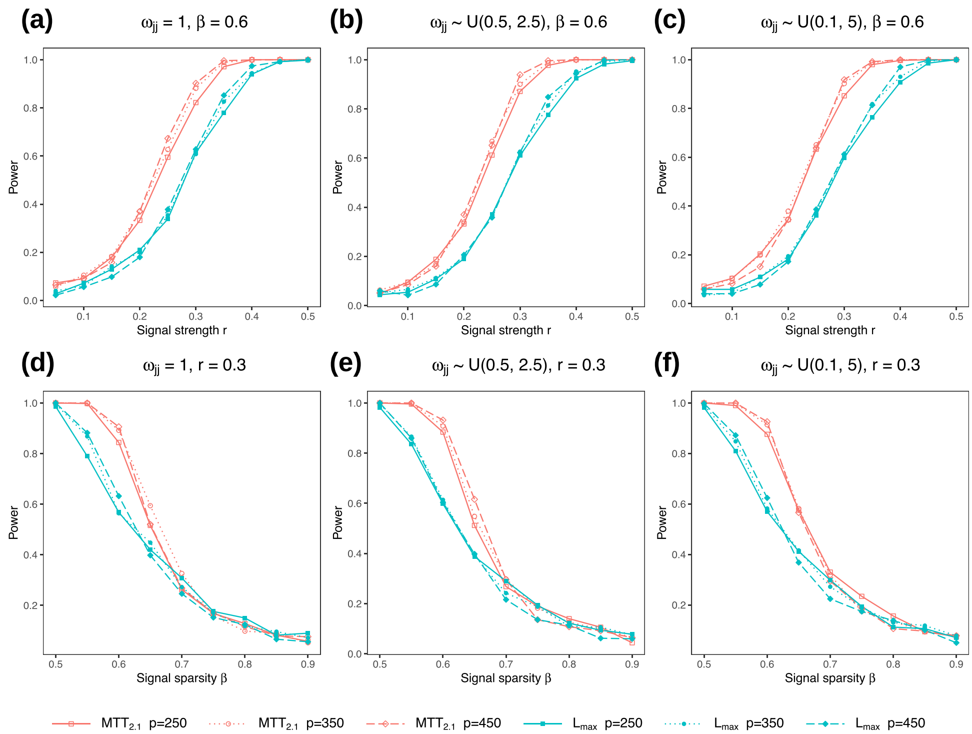

Figure 2. Empirical powers of the proposed MTT with fixed $\psi = 2.1$ ($\text{MTT}_{2.1}$) for the lasso penalty parameter λ_j in (4.2) and the $L_{\max}$-test for the hypotheses $H_0: \omega_{j_1 j_2} = 0$ for all $j_1 \neq j_2$ under $n = 125$, $p = 250$, 350, 450, three settings of $\{\omega_{jj}\}$ and the nominal level 0.05. The top panels (a)–(c) are for fixed $\beta = 0.6$ and varying r from 0.1 to 0.5; the bottom panels (d)–(f) are for fixed $r = 0.3$ and varying β from 0.5 to 0.9.

similar to $\text{MTT}_{2.1}$, which are reported in Figure S2 in the online supplementary material. The results in Table 2 and Figure S2 online supplementary material indicate that the proposed test with fixed $\psi = 2.1$ and CV selected ψ worked equally well. The results under $n = 100$ and 150 were similar, which were omitted here. As the $L_{\max}$-test was more conservative, for a fair comparison, we adjusted the rejection criteria of the two tests such that their empirical sizes were 5%. To evaluate the performance of tests under different combinations of signal sparsity and strength, we considered two scenarios. The top three panels of Figure 2 report the cases for the sparsity parameter β being fixed at 0.6 and the signal strength r changing from 0.1 to 0.5 with the increment 0.05. While the bottom panels report the cases for $r = 0.3$ and β increasing from 0.5 to 0.9. From Figure 2, we notice that the power of the proposed test was at least comparable to that of the $L_{\max}$-test. In most cases, MTT was more powerful. This verifies the power advantage of the proposed test as compared with the maximum type tests. The powers of the two tests increased (decreased) as the increase of r (β). This is because larger r leads to stronger signals, but larger β makes signals fewer.

For testing the hypotheses (2.1) for a subset of variables, we set the target variables as $\mathcal{A}_1 = \{1, \ldots, p_1\}$ and the ancillary variables as $\mathcal{A}_2 = \{p_1 + 1, \ldots, p\}$. We considered two designs for the precision matrix $\mathbf{\Omega}$ under the null hypothesis of (2.1):

Design A: the identity matrix $\mathbf{\Omega} = \mathbf{I}_p$;
Design B: $\omega_{jj} = 1$ for all $j = 1, \ldots, p$, and $\omega_{jj+p_1} = 0.25$ for $j = 1, \ldots, p - p_1$.

In both designs, $\mathbf{\Omega}_{\mathcal{A}_1} = \mathbf{I}_{p_1}$ under the null hypotheses, but there are nonzero partial correlations between $\mathcal{A}_1$ and $\mathcal{A}_2$ in Design B. We set the size of $\mathcal{A}_1$ as $p_1 = 10$, 20 and 30. To evaluate power, we randomly selected $\lfloor q_1^{(1-\beta)} \rfloor$ components from the upper-diagonal of the sub-matrix $\mathbf{\Omega}_{\mathcal{A}_1}$, and set them as $\{2r(\log q_1)/n\}^{1/2}$, where $q_1 = p_1(p_1 - 1)/2$ is the number of free parameters in $\mathbf{\Omega}_{\mathcal{A}_1}$.

Table 3. Empirical sizes of the proposed MTT with fixed $\psi = 2.1$ ($\mathrm{MTT}_{2.1}$) and cross-validated ψ ($\mathrm{MTT}_{\mathrm{cv}}$) for the lasso penalty parameter λ_j in (4.2), the $L_{\max}$-test and the L_2-test for the hypotheses (2.1) with the size of $\mathcal{A}_1$ being $p_1 = 10, 20, 30$, under designs A and B, $n = 100, 125, 150$, $p = 250, 350, 450$, and the nominal level 0.05

n	p	$p_1 = 10$				$p_1 = 20$				$p_1 = 30$			
		$\mathrm{MTT}_{2.1}$	$\mathrm{MTT}_{\mathrm{cv}}$	$L_{\max}$	L_2	$\mathrm{MTT}_{2.1}$	$\mathrm{MTT}_{\mathrm{cv}}$	$L_{\max}$	L_2	$\mathrm{MTT}_{2.1}$	$\mathrm{MTT}_{\mathrm{cv}}$	$L_{\max}$	L_2
Design A													
100	250	0.054	0.056	0.013	0.067	0.054	0.043	0.015	0.068	0.044	0.048	0.011	0.065
	350	0.050	0.051	0.020	0.093	0.052	0.049	0.014	0.066	0.054	0.052	0.012	0.050
	450	0.053	0.059	0.017	0.077	0.060	0.058	0.010	0.067	0.058	0.060	0.012	0.070
125	250	0.059	0.053	0.019	0.065	0.061	0.053	0.015	0.085	0.052	0.058	0.020	0.063
	350	0.060	0.059	0.023	0.070	0.042	0.048	0.011	0.067	0.049	0.048	0.018	0.076
	450	0.058	0.056	0.019	0.063	0.058	0.049	0.010	0.070	0.054	0.052	0.017	0.058
150	250	0.061	0.056	0.015	0.085	0.051	0.059	0.012	0.068	0.053	0.054	0.016	0.080
	350	0.054	0.052	0.013	0.067	0.052	0.053	0.011	0.089	0.057	0.045	0.012	0.073
	450	0.055	0.059	0.015	0.079	0.058	0.057	0.020	0.063	0.049	0.059	0.013	0.062
Design B													
100	250	0.046	0.058	0.017	0.062	0.056	0.056	0.012	0.064	0.052	0.052	0.010	0.057
	350	0.052	0.061	0.010	0.056	0.046	0.057	0.012	0.058	0.055	0.059	0.016	0.056
	450	0.052	0.053	0.017	0.067	0.052	0.057	0.016	0.070	0.059	0.062	0.015	0.063
125	250	0.047	0.058	0.010	0.072	0.044	0.055	0.012	0.063	0.058	0.058	0.011	0.067
	350	0.049	0.059	0.014	0.071	0.049	0.055	0.010	0.075	0.058	0.054	0.010	0.066
	450	0.055	0.058	0.020	0.066	0.055	0.058	0.018	0.059	0.050	0.048	0.012	0.065
150	250	0.053	0.042	0.013	0.073	0.058	0.059	0.017	0.066	0.053	0.052	0.018	0.065
	350	0.053	0.055	0.011	0.064	0.047	0.043	0.016	0.066	0.063	0.050	0.023	0.068
	450	0.055	0.055	0.019	0.064	0.047	0.059	0.020	0.064	0.051	0.047	0.014	0.071

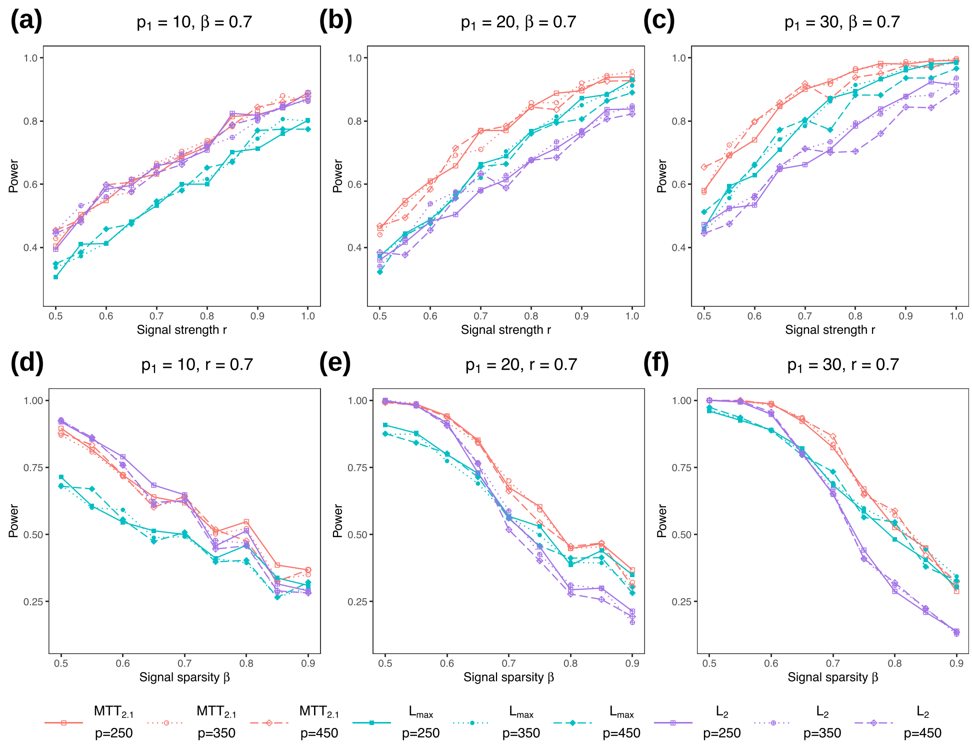

Figure 3. Empirical powers of the proposed MTT with fixed $\psi = 2.1$ ($MTT_{2.1}$) for the lasso penalty parameter λ_j in (4.2), the L_{max}-test and the L_2-test for the hypotheses (2.1) with the size of $\mathcal{A}_1$ being $p_1 = 10, 20, 30$, under Design A, $n = 125$, $p = 250, 350, 450$ and the nominal level 0.05. The top panels (a)–(c) are for fixed $\beta = 0.7$ and varying r from 0.5 to 1; the bottom panels (d)–(f) are for fixed $r = 0.7$ and varying β from 0.5 to 0.9.

Similar as the case of testing $\mathbf{\Omega}$ being diagonal, if $\mathbf{\Omega}$ was not positively definite, we modified it as $\mathbf{\Omega} + (|\lambda_{\min}(\mathbf{\Omega})| + 0.01)\mathbf{I}_p$.

Table 3 reports the empirical sizes of the proposed MTT, L_{max}, and L_2 tests for the hypotheses (2.1) under Designs A and B, different combinations of n and p, and three settings of the target set $\mathcal{A}_1$. From Table 3, we see that the proposed methods $MTT_{2.1}$ and MTT_{cv} were similar and had the most accurate size around 5% for almost all the cases. The L_{max} type test was conservative as in the previous setting of testing $\mathbf{\Omega}$ being diagonal. While the L_2 type test was a little liberal.

To evaluate the powers of those tests, similar to testing $\mathbf{\Omega}$ being diagonal, we adjusted the rejection criteria such that the empirical sizes of all tests were 0.05. The power curves of those tests with respect to different signal strength r and sparsity level β are shown in Figures 3 and 4. Here, we only report the power of $MTT_{2.1}$ as the performance of MTT_{cv} was quite similar to $MTT_{2.1}$, which is reported in Figures S3 and S4 in the online supplementary material. We considered the case of fixed $\beta = 0.7$ and varying r, and the case of fixed $r = 0.7$ and varying β. Due to limited space, we only displayed the results under $n = 125$. The results under $n = 100$ and 150 were similar and were omitted.

From Figures 3 and 4, the power of the proposed MTT was higher than that of the L_{max}-test under almost all the cases. For the high sparsity scenarios where $\beta \geq 0.8$, those tests had similar powers since the number of signals under the alternative hypothesis of (2.1) are very few. This coincides with the findings in Figure 2. The powers of MTT and the L_2-test were comparable for $p_1 = 10$, which represents small sub-blocks of $\mathbf{\Omega}$ to be tested. However, when p_1 was increased to 20 and 30, the proposed test had superior powers over the L_2-test, especially for the sparsity level β larger than 0.6 as shown in panels (c) and (f). This is because the L_2-test takes the sum of squares for all the estimated precision coefficients in the testing set $\mathcal{A}_1$, which would lose power if the size of $\mathcal{A}_1$ (p_1) is large and the signals are sparse (large β), as it includes too many non-informative components. The overall performance of the proposed MTT was better than both

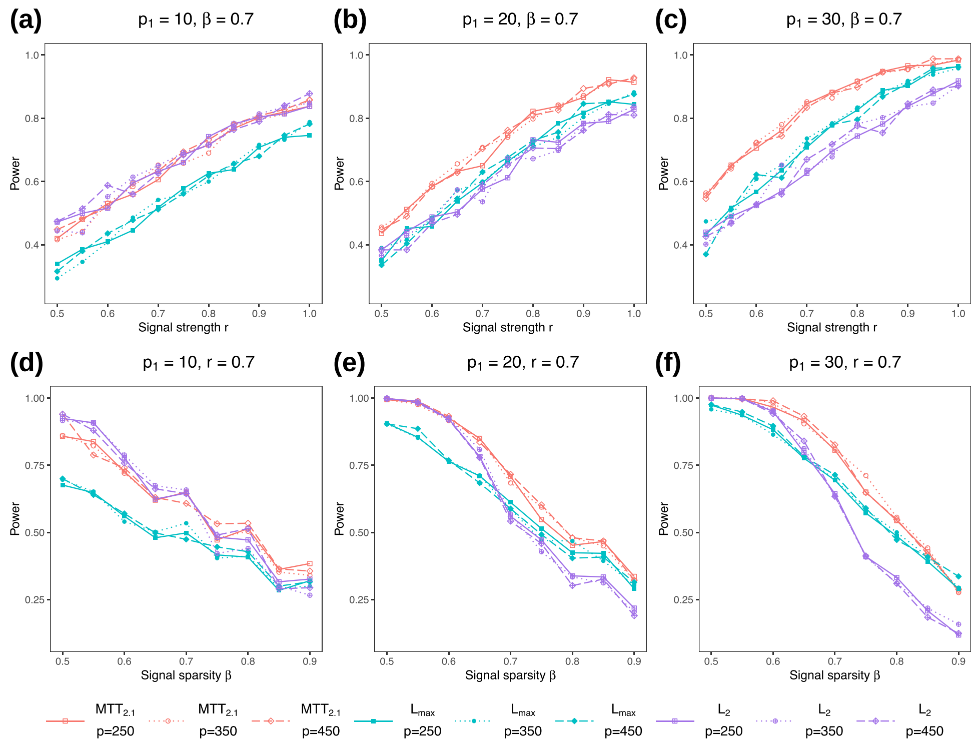

Figure 4. Empirical powers of the proposed MTT with fixed $\psi = 2.1$ ($\text{MTT}_{2.1}$) for the lasso penalty parameter λ_j in (4.2), the $L_{\max}$-test and the L_2-test for the hypotheses (2.1) with the size of $\mathcal{A}_1$ being $p_1 = 10, 20, 30$, under Design B, $n = 125$, $p = 250, 350, 450$ and the nominal level 0.05. The top panels (a)–(c) are for fixed $\beta = 0.7$ and varying r from 0.5 to 1; the bottom panels (d)–(f) are for fixed $r = 0.7$ and varying β from 0.5 to 0.9.

the $L_{\max}$-test and the L_2-test, in terms of more accurate size control under the null hypotheses and higher power for detecting sparse and weak signals in precision matrices.

8 Real data analysis

It is well known that different brain regions work together for everyday activities. Neurodegenerative disorders, such as Alzheimer's disease (AD), alter brain connectivity, which result in impairment of memory and cognitive decline. In this section, we used an FDG-PET brain imaging data set to study brain connectivity for healthy individuals and AD patients. The data contain the readings of 42 brain anatomical volumes of interest (AVOIs) from 49 AD patients and 67 normal control (NC) subjects. Those AVOIs are distributed in four brain regions: prefrontal, parietal, occipital, and temporal lobes, which contain 12, 8, 6, and 16 AVOIs, respectively.

Using this data set, Huang et al. (2010) and Qiu and Zhou (2020) studied the brain functional connectivity by estimating the precision coefficients among those 42 AVOIs and treated two AVOIs being connected if their precision coefficient is nonzero. In those studies, the precision coefficient shows the conditional association between two AVOIs after removing the linear effects of other AVOIs, which serves as a measure for conditional dependence between brain AVOIs. Let $\mathbf{\Omega}_{\text{AD}} = (\omega_{\text{AD},j_1j_2})$ and $\mathbf{\Omega}_{\text{NC}} = (\omega_{\text{NC},j_1j_2})$ be the precision matrices of the 42 AVOIs for the AD and NC groups, respectively. Let $\mathcal{A}_1, \ldots, \mathcal{A}_4$ denote the index sets of AVOIs corresponding to the four brain regions. Let $\mathbf{\Omega}_{\text{AD},k_1k_2}$ and $\mathbf{\Omega}_{\text{NC},k_1k_2}$ be the sub-matrices of $\mathbf{\Omega}_{\text{AD}}$ and $\mathbf{\Omega}_{\text{NC}}$ with elements from $\mathcal{A}_{k_1} \times \mathcal{A}_{k_2}$. For the AD group, we were interested in testing the hypotheses

$$H_{k,0} : \omega_{\text{AD},j_1j_2} = 0 \quad \text{for all } j_1, j_2 \in \mathcal{A}_k \text{ and } j_1 \neq j_2 \text{ vs. } H_{k,a} : \text{ not } H_{k,0} \tag{8.1}$$

for $k_1 = k_2 = k \in \{1, \ldots, 4\}$, and the hypotheses

Table 4. *P*-values of the hypotheses (8.1) and (8.2) by the proposed MTT with fixed $\psi = 2.1$ (MTT$_{2.1}$) for the lasso penalty parameter λ_j in (4.2), the $L_{\max}$-test and the L_2-test for the Alzheimer's disease and normal control groups

	Alzheimer's disease (AD)			Normal control (NC)		
Brain regions	**MTT$_{2.1}$**	**$L_{\max}$-test**	**L_2-test**	**MTT$_{2.1}$**	**$L_{\max}$-test**	**L_2-test**
Prefrontal–prefrontal	0	0	0	0	0	0
Prefrontal–parietal	4.49×10^{-9}	0.073	1.32×10^{-8}	0	5.51×10^{-4}	3.99×10^{-8}
Prefrontal–occipital	2.33×10^{-15}	2.30×10^{-3}	1.75×10^{-10}	0	2.42×10^{-4}	3.85×10^{-9}
Prefrontal–temporal lobes	9.58×10^{-5}	9.70×10^{-3}	1.21×10^{-14}	3.01×10^{-14}	0.020	1.67×10^{-15}
Parietal–parietal	0	1.08×10^{-10}	0	0	1.11×10^{-16}	0
Parietal–occipital	0	2.84×10^{-6}	8.19×10^{-14}	3.42×10^{-14}	1.73×10^{-4}	2.72×10^{-5}
Parietal–temporal lobes	1.19×10^{-4}	1.90×10^{-3}	5.43×10^{-14}	0.481	0.035	1.05×10^{-8}
Occipital–occipital	0	2.22×10^{-16}	0	0	0	0
Occipital–temporal lobes	0	1.19×10^{-6}	1.51×10^{-9}	0	2.92×10^{-7}	4.11×10^{-15}
Temporal lobes–temporal lobes	0	1.67×10^{-15}	0	0	1.22×10^{-15}	0

$$H_{k_1k_2,0}:\boldsymbol{\Omega}_{\mathrm{AD},k_1k_2}=0 \text{ vs. } H_{k_1k_2,a}:\boldsymbol{\Omega}_{\mathrm{AD},k_1k_2}\neq 0 \tag{8.2}$$

for $k_1 \neq k_2 \in \{1, \ldots, 4\}$. Similar hypotheses were considered for the NC group as well. Note that the proposed thresholding procedure can be applied for testing the non-diagonal sub-blocks $\boldsymbol{\Omega}_{\mathrm{AD},k_1k_2}$ for $k_1 \neq k_2$. See the discussion in the first paragraph of Section 9 for details. Those two hypotheses were used to investigate whether any connection within one brain region or between two brain regions existed.

We applied three methods, MTT with fixed $\psi = 2.1$, the $L_{\max}$-test and L_2-test, for the hypotheses (8.1) and (8.2). The p-values of those three tests are reported in Table 4, where 'prefrontal–prefrontal' represents the hypotheses (8.1) for the prefrontal region, and 'prefrontal–parietal' represents the hypotheses (8.2) for testing the precision coefficients between the prefrontal and parietal regions. As there were 10 hypotheses, to control the family-wise error rate at 0.05, we rejected each of the hypotheses if the corresponding p-value was less than 0.005 by Bonferroni correction. From Table 4, MTT and the L_2-test rejected almost all the hypotheses for both the AD and NC groups, which implies that there existed at least one connection between different AVOIS within each brain region and between two brain regions. This is expected as brain regions should function together for healthy people. Although Alzheimer's disease may affect brain connectivity, our results indicate brain connections among different regions still exist for AD patients. However, the $L_{\max}$-test failed to reject the hypotheses 'prefrontal–parietal' and 'prefrontal–temporal lobes' for the AD group, and the hypotheses 'prefrontal–temporal lobes' for the NC group. This may be due to the power loss of the $L_{\max}$-test as revealed by the simulation study.

9 Discussion

This article proposes a minimax optimal test for the hypotheses (2.1) for precision coefficients of a sub-group of variables. It can be applied for testing cross-blocks of precision coefficients being zero, namely, $H_0:\omega_{j_1j_2}=0$ for all $j_1 \in \mathcal{A}_1$ and $j_2 \in \mathcal{A}_2$, where $\mathcal{A}_1$ and $\mathcal{A}_2$ are two non-overlapping subsets of $\mathcal{N}=\{1, \ldots, p\}$. The same multi-level thresholding test in (4.10) can be used. The only differences are to change $q=|\mathcal{A}_1||\mathcal{A}_2|$ and the thresholding statistic $T(t)$ in (4.5) to be $T(t)=\sum_{j_1\in\mathcal{A}_1, j_2\in\mathcal{A}_2} V_{j_1j_2}\mathbb{I}\{V_{j_1j_2}\geq\lambda_q(t)\}$.

Under sub-Gaussian distributed data, the expansion of $\widehat{\omega}_{j_1j_2}$ in (5.1) is still valid. However, the main order variance $\theta_{j_1j_2}=\mathrm{Var}(\epsilon_{ij_1}\epsilon_{ij_2})/(\upsilon_{j_1j_1}^2\upsilon_{j_2j_2}^2)$ of $\sqrt{n}(\widehat{\omega}_{j_1j_2}-\omega_{j_1j_2})$ is no longer equal to $(1+\rho_{j_1j_2}^2)(\upsilon_{j_1j_1}\upsilon_{j_2j_2})^{-1}$. For non-Gaussian data, $\theta_{j_1j_2}$ can be estimated by

$$\widetilde{\theta}_{j_1j_2} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{\epsilon}_{ij_1}\widehat{\epsilon}_{ij_2} - \widehat{\upsilon}_{j_1j_2})^2/(\widehat{\upsilon}^2_{j_1j_1}\widehat{\upsilon}^2_{j_2j_2}),$$

and the thresholding statistic is $T_{\text{NG}}(t) = \sum_{j_1,j_2\in\mathcal{A}_1, j_1<j_2} V^{\text{NG}}_{j_1j_2}\mathbb{I}\{V^{\text{NG}}_{j_1j_2} \geq \lambda_q(t)\}$ where $V^{\text{NG}}_{j_1j_2} = n\widehat{\omega}^2_{j_1j_2}\widetilde{\theta}^{-1}_{j_1j_2}$ is the standardization of $\widehat{\omega}_{j_1j_2}$ using $\widetilde{\theta}_{j_1j_2}$. Lemma S5 in the online supplementary material shows that $\max_{j_1,j_2}|\widetilde{\theta}_{j_1j_2}/\theta_{j_1j_2} - 1| = O_p\{s(\log p)^{1/2}n^{-1/2}\}$, and the large deviation result of $V^{\text{NG}}_{j_1j_2}$ under sub-Gaussian distributed data is the same as that of $V_{j_1j_2}$ under Gaussian data. Therefore, under suitable conditions, it might be shown that the asymptotic distribution of $T_{\text{NG}}(t)$ is the same as that of $T(t)$ under the Gaussian distribution. The multi-level thresholding statistic can be constructed similarly as $\mathcal{M}(t_0)$ based on $T_{\text{NG}}(t)$, and the rejection criterion of MTT under sub-Gaussian distributed data should be the same as that in (4.10).

The proposed procedure can also be extended to test the equivalence of two precision matrices or their sub-blocks between two populations. Let $\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_2}$ be i.i.d. samples from two p-dimensional Gaussian distributions with means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and covariances $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively, where $\boldsymbol{\Omega}_k = (\omega_{k,j_1j_2}) = \boldsymbol{\Sigma}_k^{-1}$ for $k = 1, 2$. For a subset $\mathcal{A}_1$ of variables, we are interested in testing whether $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are the same in the sub-block $\mathcal{A}_1 \times \mathcal{A}_1$, namely, the two-sample hypotheses

$$\begin{aligned} H_0: \omega_{1,j_1j_2} = \omega_{2,j_1j_2} \quad &\text{for all } j_1, j_2 \in \mathcal{A}_1 \text{ vs.} \\ H_a: \omega_{1,j_1j_2} \neq \omega_{2,j_1j_2} \quad &\text{for some } j_1, j_2 \in \mathcal{A}_1. \end{aligned} \tag{9.1}$$

The nodewise regression estimator in (4.3) can be applied to obtain the estimate $\widehat{\omega}_{k,j_1j_2}$ of ω_{k,j_1j_2}, and the corresponding estimate $\widehat{\rho}_{k,j_1j_2}$ of the partial correlation for $k = 1, 2$. Let $\widehat{\theta}_{k,j_1j_2} = (1 + \widehat{\rho}^2_{k,j_1j_2})(\widehat{\omega}_{k,j_1j_1}\widehat{\omega}_{k,j_2j_2})$ be the estimated variance of $\sqrt{n}_k(\widehat{\omega}_{k,j_1j_2} - \omega_{k,j_1j_2})$. The standardized difference between $\widehat{\omega}_{1,j_1j_2}$ and $\widehat{\omega}_{2,j_1j_2}$ is

$$V_{2s,j_1j_2} = \frac{(\widehat{\omega}_{1,j_1j_2} - \widehat{\omega}_{2,j_1j_2})^2}{\widehat{\theta}_{1,j_1j_2}/n_1 + \widehat{\theta}_{2,j_1j_2}/n_2}.$$

Let $\lambda_{2s}(t) = 2t\log\tilde{q}$ for $\tilde{q} = p_1(p_1 + 1)/2$, where $p_1 = |\mathcal{A}_1|$. Similar as (4.5) and (4.8), we construct the single-level thresholding statistic for the hypotheses (9.1) as

$$T_{2s}(t) = \sum_{(j_1,j_2)\in\mathcal{A}_1, j_1\leq j_2} V_{2s,j_1j_2}\mathbb{I}\{V_{2s,j_1j_2} > \lambda_{2s}(t)\},$$

and the multi-level thresholding statistic based on $T_{2s}(t)$ as

$$\mathcal{M}_{2s}(\tilde{t}_0) = \sup_{t\in(\tilde{t}_0, 1-\eta]} \frac{T_{2s}(t) - \tilde{\mu}_{2s,0}(t)}{\tilde{\sigma}_{2s,0}(t)}, \tag{9.2}$$

where $\tilde{t}_0$ is the threshold lower bound, η is an arbitrarily small positive constant, and $\tilde{\mu}_{2s,0}(t) = \tilde{q}[2\lambda^{1/2}_{2s}(t)\phi\{\lambda^{1/2}_{2s}(t)\} + 2\bar{\Phi}\{\lambda^{1/2}_{2s}(t)\}]$ and $\tilde{\sigma}^2_{2s,0}(t) = \tilde{q}[2\{\lambda^{3/2}_{2s}(t) + 3\lambda^{1/2}_{2s}(t)\}\phi\{\lambda^{1/2}_{2s}(t)\} + 6\bar{\Phi}\{\lambda^{1/2}_{2s}(t)\}]$ are the main order terms of the mean and variance of $T_{2s}(t)$ under the null hypothesis in (9.1). Similar as (4.10) for the one sample case, we reject the null hypotheses of (9.1) if $\mathcal{M}_{2s}(\tilde{t}_0) > [q_\alpha + b\{\log(\tilde{q}), \tilde{t}_0, \eta\}]/a\{\log(\tilde{q})\}$.

Acknowledgments

We would like to thank the Editor, the Associate Editor, and two anonymous reviewers for thoughtful comments and suggestions which have improved the content and presentation of the article. Part of the work reported in the article 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.

Conflicts of interest: None declared.

Funding

Y.Q. was partially supported by the National Natural Science Foundation of China (92358303). B.G. was partially supported by the National Natural Science Foundation of China (11971390) and Sichuan Science and Technology Program (2023NSFSC1355).

Data availability

The data that support the findings in this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu/). The pre-processed data are available from Huang et al. (2010). Restrictions may apply to the availability of these data.

Supplementary material

Supplementary material is available at *Journal of the Royal Statistical Society: Series B* online.

References

Anderson T. W. (2003). *An introduction to multivariate statistical analysis*. Wiley.

Besag J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society*, *36*, 192–236. https://doi.org/10.1111/j.2517-6161.1974.tb00999.x

Bühlmann P., & van de Geer S (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Springer.

Cai T. T., Liu W., & Luo X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, *106*(494), 594–607. https://doi.org/10.1198/jasa.2011.tm10155

Candes E., & Tao T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, *35*(6), 2313–2351. https://doi.org/10.1214/009053606000001523

Chang J., Qiu Y., Yao Q., & Zou T. (2018). Statistical inference for large precision matrices with dependent data and their applications. *Journal of Econometrics*, *206*(1), 57–82. https://doi.org/10.1016/j.jeconom.2018.03.020

Chen S. X., Qiu Y, & Zhang S. (2023). Sharp optimality for high-dimensional covariance testing under sparse signals. *The Annals of Statistics*, *51*(5), 1921–1945. https://doi.org/10.1214/23-AOS2310

Delaigle A., Hall P., & 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*(3), 283–301. https://doi.org/10.1111/j.1467-9868.2010.00761.x

Donoho D., & Jin J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, *32*(3), 962–994. https://doi.org/10.1214/009053604000000265

Donoho D., & Jin J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science*, *30*(1), 1–25. https://doi.org/10.1214/14-STS506

Friedman J., Hastie T., & Tibshirani R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Hall P., & Jin J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics*, *38*, 1686–1732. https://doi.org/10.1214/09-AOS764

Huang S., Li J., Sun L., Liu J., Wu T., Chen K., & Reiman E. (2010). Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. *Neuroimage*, *50*(3), 935–949. https://doi.org/10.1016/j.neuroimage.2009.12.120

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

Kaiser M., Lahiri S., & Nordman D. (2012). Goodness of fit tests for a class of Markov random field models. *The Annals of Statistics*, *40* (1), 104–130. https://doi.org/10.1214/11-AOS948

Kaiser M., & Nordman D. (2012). Blockwise empirical likelihood for spatial Markov model assessment. *Statistics and Its Interface*, *5*(3), 303–318. https://doi.org/10.4310/SII.2012.v5.n3.a3

Lafferty J., Liu H., & Wasserman L. (2012). Sparse nonparametric graphical models. *Statistical Science*, *27*(4), 519–537. https://doi.org/10.1214/12-STS391

Liu H., Han F., Yuan M., Lafferty J., & Wasserman L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics*, *40*(4), 2293–2326. https://doi.org/10.1214/12-AOS1037

Liu W. (2013). Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics*, *41*(6), 2948–2978. https://doi.org/10.1214/13-AOS1169

Meinshausen N., & Bühlmann P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics*, *34*(3), 1436–1462. https://doi.org/10.1214/009053606000000281

Peng J., Wang P., Zhou N., & Zhu J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, *104*(486), 735–746. https://doi.org/10.1198/jasa.2009.0126

Qiu Y., Chen S. X., & Nettleton D. (2018). Detecting rare and faint signals via thresholding maximum likelihood estimators. *The Annals of Statistics*, *46*(2), 895–923. https://doi.org/10.1214/17-AOS1574

Qiu Y., & Zhou X. H. (2020). Estimating *c*-level partial correlation graphs with application to brain imaging. *Biostatistics*, *21*(4), 641–658. https://doi.org/10.1093/biostatistics/kxy076

Qiu Y., & Zhou X. H. (2022). Inference on multi-level partial correlations based on multi-subject time series data. *Journal of the American Statistical Association*, *117*(540), 2268–2282. https://doi.org/10.1080/01621459.2021.1917417

Ren Z., Kang Y., Fan Y., & Lv J. (2019). Tuning-free heterogeneous inference in massive networks. *Journal of the American Statistical Association*, *114*(528), 1908–1925. https://doi.org/10.1080/01621459.2018.1537920

Ren Z., Sun T., Zhang C. H., & Zhou H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *The Annals of Statistics*, *43*(3), 991–1026. https://doi.org/10.1214/14-AOS1286

Skiena S. (1991). *Implementing discrete mathematics: Combinatorics and graph theory with mathematica*. Addison-Wesley.

Sun T., & Zhang C. H. (2012). Scaled sparse linear regression. *Biometrika*, *99*(4), 879–898. https://doi.org/10.1093/biomet/ass043

Tibshirani R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

van de Geer S., Bühlmann P., Ritov Y., & Dezeure R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, *42*(3), 1166–1202. https://doi.org/10.1214/14-AOS1221

Wang Y. X., & Huang H. (2014). Review on statistical methods for gene network reconstruction using expression data. *Journal of Theoretical Biology*, *362*, 53–61. https://doi.org/10.1016/j.jtbi.2014.03.040

Xia Y., Cai T., & Cai T. T. (2015). Testing differential networks with applications to detecting gene-by-gene interactions. *Biometrika*, *102*(2), 247–266. https://doi.org/10.1093/biomet/asu074

Xia Y., Cai T., & Cai T. T. (2018). Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions. *Journal of the American Statistical Association*, *113*(521), 328–339. https://doi.org/10.1080/01621459.2016.1251930

Xue L., & Zou H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, *40*(5), 2541–2571. https://doi.org/10.1214/12-AOS1041

Yuan M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*, *11*, 2261–2286.

Yuan M., & Lin Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, *94*(1), 19–35. https://doi.org/10.1093/biomet/asm018

Zhang C. H., & Zhang S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *76*(1), 217–242. https://doi.org/10.1111/rssb.12026

Zhong P. S., Chen S. X., & Xu M. (2013). Tests alternative to higher criticism for high dimensional means under sparsity and column-wise dependence. *The Annals of Statistics*, *41*(6), 2820–2851. https://doi.org/10.1214/13-AOS1168