High-dimensional two-sample precision matrices test: an adaptive approach through multiplier bootstrap

MINGJUAN ZHANG, YONG HE[*,†], CHENG ZHOU, AND XINSHENG ZHANG[‡]

Precision matrix, which is the inverse of covariance matrix, plays an important role in statistics, as it captures the partial correlation between variables. Testing the equality of two precision matrices in high dimensional setting is a very challenging but meaningful problem, especially in the differential network modelling. To our best knowledge, existing test is only powerful for sparse alternative patterns where two precision matrices differ in a small number of elements. In this paper we propose a data-adaptive test which is powerful against either dense or sparse alternatives. Multiplier bootstrap approach is utilized to approximate the limiting distribution of the test statistic. Theoretical properties including asymptotic size and power of the test are investigated. Simulation study verifies that the data-adaptive test performs well under various alternative scenarios. The practical usefulness of the test is illustrated by applying it to a gene expression data set associated with lung cancer.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 62H15; secondary 62P10.
KEYWORDS AND PHRASES: Differential network, High-dimensional, Precision matrix, Multiplier bootstrap.

1. INTRODUCTION

In recent years, Gaussian graphical model has been an important tool to capture the conditional dependency structure among variables. The edges of the Gaussian graphical network are characterized by the inverse covariances for each pair of nodes. To be more specific, for Gaussian graphical model, the joint distribution of p random variables $(X_1, \ldots, X_p)^\top$ is assumed to be multivariate Gaussian $N(\mathbf{0}, \mathbf{\Omega}^{-1})$, where $\mathbf{\Omega}$ is the inverse of the covariance matrix and is called precision matrix. It is known that for Gaussian graphical model, the conditional dependency structure is completely encoded in the precision matrix, i.e., for each pair of nodes X_a and X_b, they are conditionally independent given all other variables if and only if the (a, b)-th entry of $\mathbf{\Omega}$ is equal to zero. A growing number of literature has focused on the support recovery and link strength estimation of Gaussian graphical model in high-dimensional setting, see, for example, [16, 25, 9, 24, 1, 3, 14], among many others. For more detailed discussions and comparisons of these methods, we refer to [18] and [8]. The works mentioned above focus on analyzing one particular Gaussian graph. However, in some cases, it is of greater interest to investigate how the network of connected node pairs change from one state to another. For example, in genomic studies, it is more meaningful to investigate how the network of connected gene pairs change from different experimental condition, which provides deeper insights on an underlying biological process, e.g., identification of pathways that correspond to the condition change. Indeed, differential networking modeling has drawn much attention as an important tool to analyze a set of changes in graph structure. The differential network is typically modeled as the difference of two precision matrices and this type of model has been used by [13, 11, 7, 26, 23, 22]. To investigate the differential network, in the first step, we need to identify whether there exists any network change, which is equivalent to test the equality of two precision matrices:,

$$(1.1) \qquad \mathbf{H}_0 : \mathbf{\Omega}_1 = \mathbf{\Omega}_2.$$

Although the equality of two precision matrices is equivalent to the equality of two covariance matrices from mathematical view, the test problem could be very different due to the fundamental difference between conditional and unconditional dependencies. Literatures on testing equality of two covariance matrices in high-dimensional setting mainly falls into two categories, sum-of-square type testing and maximum type testing. The sum-of-square type testing are particular powerful under dense alternative where the two covariance matrices differ in a large number of entries [20, 21, 12] while the maximum type testing are particular powerful under sparse alternative where the two covariance matrices differ only in a small number of entries [2].

Literature [28] proposed a unified framework for developing tests based on U-statistics, which includes testing the equality of two covariance matrices as a special case. The tests are powerful against a large variety of alternative scenarios. This research area is very active, and as a result, this list of references is illustrative rather than comprehensive. In contrast, literatures on testing equality of two precision matrices rarely exists. Literature [23] proposed a maximum-type testing which is powerful against alternative where $\boldsymbol{\Delta} = \boldsymbol{\Omega}_1 - \boldsymbol{\Omega}_2$ is sparse. As far as we know, this is the unique existing work on testing the equality of two precision matrices in the high-dimensional setting. In other word, a powerful testing for hypothesis (1.1) under dense alternative still don't exist, which urges us to consider such a problem.

In this paper, we propose a testing procedure for hypothesis (1.1) which is powerful against a large variety of alternative scenarios in high dimensions. Both theoretical results and numerical simulation show the advantage of proposed test against existing methods. The rest of the paper is organized as follows. In Section 2, we introduce some notations and briefly review the test statistic proposed by [23]. In Section 3 we present our test statistic and the multiplier bootstrap procedure to obtain the critical value or p-value of the test. Section 4 gives the theoretical analysis of the test. In Section 5, we conduct thorough numerical simulation to investigate the empirical performance of the test. A real gene expression data set is analyzed to illustrate the usefulness of the test. At last we discuss possible future directions in the last section.

2. BACKGROUND

2.1 Notation

For a vector $\mathbf{v} = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$, let $\|\mathbf{v}\|_p = \left(\sum_{j=1}^d |v_j|^p \right)^{1/p}$ as the L_p-norm. As $p = \infty$, we set $\|\mathbf{v}\|_\infty = \max_{1 \le j \le d} |v_j|$. As $p = 0$, we set $\|\mathbf{v}\|_0 = \sum_{j=1}^d I\{v_j \ne 0\}$. We use $v^{(1)}, v^{(2)}, \ldots, v^{(d)}$ to denote the order statistics of the absolute value of $\mathbf{v}$'s entries with $v^{(1)} \le v^{(2)} \le \ldots \le v^{(d)}$. Apparently, we have $v^{(j)} \ge 0$ for $j = 1, \ldots, d$. We define the (s_0, p)-norm of $\mathbf{v}$ as $\|\mathbf{v}\|_{(s_0, p)} = \left(\sum_{j=d-s_0+1}^d (v^{(j)})^p \right)^{1/p}$. As $p = \infty$, we set $\|\mathbf{v}\|_{(s_0, p)} = \|\mathbf{v}\|_\infty = v^{(d)}$ for any s_0. We denote $\mathbb{S}^{d-1} := \{\mathbf{v} \in \mathbb{R}^d : \|\mathbf{v}\|_2 = 1\}$ as the spherical surface in $\mathbb{R}^d$. For any vector $\boldsymbol{\mu}_m \in \mathbb{R}^d$, let $\boldsymbol{\mu}_{m,-i}$ denote the $(d-1) \times 1$ vector by removing the i-th entry from $\boldsymbol{\mu}_m$. For a data matrix $\mathbf{U} = (\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n)^\top \in \mathbb{R}^{n \times d}$, let $\mathbf{U}_{\cdot,-i} = (\boldsymbol{U}_{1,-i}, \ldots, \boldsymbol{U}_{n,-i})^\top$ with dimension $(n \times (d-1))$, $\overline{\boldsymbol{U}}_{\cdot,-i} = n^{-1} \sum_{k=1}^n \boldsymbol{U}_{k,-i}$ with dimension $(d-1) \times 1$, $\boldsymbol{U}_{(i)} = (U_{1,i}, \ldots, U_{n,i})^\top$ with dimension $n \times 1$, $\overline{\boldsymbol{U}}_{(i)} = (\overline{U}_i, \ldots, \overline{U}_i)^\top$ with dimension $n \times 1$ where $\overline{U}_i = n^{-1} \sum_{k=1}^n U_{k,i}$ and $\overline{\boldsymbol{U}}_{(\cdot,-i)} = (\overline{\boldsymbol{U}}_{\cdot,-i}, \ldots, \overline{\boldsymbol{U}}_{\cdot,-i})^\top$ with dimension $n \times (d-1)$. For tuning parameter λ, let $\lambda_{n_m,i,m}$ represent the i-th tuning parameter for binary trait m, which depends on the sample size n_m.

For a matrix $\boldsymbol{A} = [a_{i,j}] \in \mathbb{R}^{d \times d}$, we denote the matrix ℓ_1 norm, the matrix element-wise infinity norm and the matrix element-wise ℓ_1-norm by $\|\boldsymbol{A}\| = \max_{1 \le j \le d} \sum_{i=1}^d |a_{i,j}|$, $|\boldsymbol{A}|_\infty = \max_{i,j} |a_{i,j}|$ and $|\boldsymbol{A}|_1 = \sum_{i=1}^d \sum_{j=1}^d |a_{i,j}|$ respectively. $\boldsymbol{A}_{i,-j}$ denote the i-th row of $\boldsymbol{A}$ with its j-th entry removed and $\boldsymbol{A}_{-i,j}$ denote the j-th column of $\boldsymbol{A}$ with its i-th entry removed. $\boldsymbol{A}_{-i,-j}$ denotes a $(d-1) \times (d-1)$ matrix obtained by removing the i-th row and j-th column of $\boldsymbol{A}$. We say $\boldsymbol{A}$ is k-sparse if each row/column has at most k nonzero entries. For a symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{d \times d}$, we use $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ to denote the smallest and largest eigenvalues of $\boldsymbol{A}$ respectively. Besides, we define a $d(d-1)/2$-dimension vector

$$\text{trivec}(\boldsymbol{A}) = (a_{21}, \ldots, a_{d1}, a_{32}, \ldots, a_{3d}, \ldots, a_{(d-1)d})^\top$$

which is obtained by concatenating the lower triangular part of $\boldsymbol{A}$ column by column. We use $a_{i_s j_s}$ to denote the s-th entry of $\text{trivec}(\boldsymbol{A})$.

For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ if there exists a constant C such that $|a_n| \le C|b_n|$ holds for all n, write $a_n = o(b_n)$ if $\lim_{n \to \infty} a_n / b_n = 0$, and write $a_n \asymp b_n$ if there exist constants c and C such that $c \le a_n / b_n \le C$ for all n. For a sequence of random variables $\{\xi_1, \xi_2, \ldots\}$, we use $\lim_{n \to \infty} \xi_n = \xi$ to denote that the sequence $\{\xi_n\}$ converges in probability towards ξ as $n \to \infty$. For simplicity, we also use $\xi_n = o_p(1)$ to denote $\lim_{n \to \infty} \xi_n = 0$. For random variables ξ and η, we use $\text{Cov}(\xi, \eta)$ and $\text{Corr}(\xi, \eta)$ to denote the covariance and correlation coefficients between ξ and η. Let $\text{Var}(\xi)$ be the variance of random variable ξ. For a set $\mathcal{H}$, denote by $\#\{\mathcal{H}\}$ the cardinality of $\mathcal{H}$.

2.2 General setup

Let $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$ be two d-dimensional random vectors independent of each other. $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1}$ are independent and identically distributed (i.i.d.) random samples from $\boldsymbol{X} \sim \boldsymbol{N}(\boldsymbol{u}_1, \boldsymbol{\Sigma}_1)$ with $\boldsymbol{X}_k = (X_{k,1}, X_{k,2}, \ldots, X_{k,d})^\top$. Similarly, $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n_2}$ are i.i.d. random samples from $\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{u}_2, \boldsymbol{\Sigma}_2)$ with $\boldsymbol{Y}_k = (Y_{k,1}, Y_{k,2}, \ldots, Y_{k,d})^\top$. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1})^\top$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n_1})^\top$ denote the data matrices. Let $\boldsymbol{\Sigma}_m = (\sigma_{i,j,m})$ and $\boldsymbol{\Omega}_m = (\omega_{i,j,m}) = \boldsymbol{\Sigma}_m^{-1}$ for $m = 1, 2$. Let $\boldsymbol{\beta}_{i,1} = (\beta_{1,i,1}, \ldots, \beta_{d-1,i,1})^\top$ denote the regression coefficients of $X_{k,i}$ regressed on the rest of the entries of $\boldsymbol{X}_k$ and let $\boldsymbol{\beta}_{i,2} = (\beta_{1,i,2}, \ldots, \beta_{d-1,i,2})^\top$ denote the regression coefficients of $Y_{k,i}$ regressed on the rest of the entries of $\boldsymbol{Y}_k$.

In the Gaussian setting, the precision matrix can be described in terms of regression models. Specifically:

$$(2.1) \quad \begin{aligned} X_{k,i} &= \alpha_{i,1} + \boldsymbol{X}_{k,-i}^\top \boldsymbol{\beta}_{i,1} + \epsilon_{k,i,1}, \\ Y_{k,i} &= \alpha_{i,2} + \boldsymbol{Y}_{k,-i}^\top \boldsymbol{\beta}_{i,2} + \epsilon_{k,i,2}, \end{aligned}$$

where the error terms $\epsilon_{k,i,m}$ follow normal distribution with mean zero and variance $\{\sigma_{i,i,m} -$

$\Sigma_{i,-i,m}(\Sigma_{-i,-i,m})^{-1}\Sigma_{-i,i,m}\}$ and $\epsilon_{k,i,1}$, $\epsilon_{k,i,2}$ are independent of $\boldsymbol{X}_{k,-i}$ and $\boldsymbol{Y}_{k,-i}$ respectively. Besides, we have $\alpha_{i,m} = \mu_{i,m} - \Sigma_{i,-i,m}\Sigma_{-i,-i,m}^{-1}\boldsymbol{\mu}_{-i,m}$. The regression coefficient vectors $\boldsymbol{\beta}_{i,m}$ and error terms $\epsilon_{k,i,m}$ satisfy

$$\boldsymbol{\beta}_{i,m} = -\omega_{i,i,m}^{-1}\boldsymbol{\Omega}_{-i,i,m},$$
$$r_{i,j,m} = \text{Cov}(\epsilon_{k,i,m},\epsilon_{k,j,m}) = \frac{\omega_{i,j,m}}{\omega_{i,i,m}\omega_{j,j,m}}.$$

We aim to test the null hypothesis:

$$\mathbf{H}_0 : \boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_2 \ \text{ or equivalentlly } \ \boldsymbol{\Delta} = \boldsymbol{\Omega}_1 - \boldsymbol{\Omega}_2 = 0.$$

Let $\widehat{\boldsymbol{\beta}}_{i,m} = (\widehat{\beta}_{1,i,m},\ldots,\widehat{\beta}_{d-1,i,m})^\top$ be estimators of $\boldsymbol{\beta}_{i,m}$ by Lasso or Dantzig selector satisfying

$$(2.2) \quad \begin{aligned} \max_{1\le i\le d}\|\widehat{\boldsymbol{\beta}}_{i,m} - \boldsymbol{\beta}_{i,m}\|_1 &= \boldsymbol{o}_p\{(\log(d))^{-1}\}, \\ \max_{1\le i\le d}\|\widehat{\boldsymbol{\beta}}_{i,m} - \boldsymbol{\beta}_{i,m}\|_2 &= \boldsymbol{o}_p\{(n_m\log(d))^{-1/4}\}. \end{aligned}$$

Under the sparsity conditions $\max_{1\le i\le d}|\beta_i|_0 = \boldsymbol{o}(n^{1/2}/(\log(d))^{3/2})$, together with with Assumption **(B)** in Section 4.1, both the Lasso and Dantzig selector estimators satisfy the condition in (2.2) according to the Proposition 4.1 in [14].

With the $\widehat{\boldsymbol{\beta}}_{i,m}$, define the residuals by

$$(2.3) \quad \begin{aligned} \widehat{\epsilon}_{k,i,1} &= X_{k,i} - \bar{X}_i - (\boldsymbol{X}_{k,-i} - \bar{\boldsymbol{X}}_{\cdot,-i})^\top\widehat{\boldsymbol{\beta}}_{i,1}, \\ \widehat{\epsilon}_{k,i,2} &= Y_{k,i} - \bar{Y}_i - (\boldsymbol{Y}_{k,-i} - \bar{\boldsymbol{Y}}_{\cdot,-i})^\top\widehat{\boldsymbol{\beta}}_{i,2}. \end{aligned}$$

Let $\widetilde{r}_{i,j,m} = (1/n_m)\sum_{k=1}^{n_m}\widehat{\epsilon}_{k,i,m}\widehat{\epsilon}_{k,j,m}$ be the empirical covariance between $\{\widehat{\epsilon}_{k,i,m} : k = 1,\ldots,n_m\}$ and $\{\widehat{\epsilon}_{k,j,m} : k = 1,\ldots,n_m\}$. Similarly, let $\widetilde{R}_{i,j,m} = (1/n_m)\sum_{k=1}^{n_m}(\epsilon_{k,i,m} - \bar{\epsilon}_{i,m})(\epsilon_{k,j,m} - \bar{\epsilon}_{j,m})$ be the empirical covariance between $\{\epsilon_{k,i,m} : k = 1,\ldots,n_m\}$ and $\{\epsilon_{k,j,m} : k = 1,\ldots,n_m\}$. Lemma 2 in [23] shows that

$$(2.4) \quad \begin{aligned} \widetilde{r}_{i,j,m} =&\widetilde{R}_{i,j,m} - \widetilde{r}_{i,i,m}(\widehat{\beta}_{i,j,m} - \beta_{i,j,m}) \\ & - \widetilde{r}_{j,j,m}(\widehat{\beta}_{j-1,i,m} - \beta_{j-1,i,m}) \\ & + \boldsymbol{o}_p\{(n_m\log(d))^{-1/2}\}. \end{aligned}$$

For $1 \le i < j \le d$, it can be shown that

$$\beta_{i,j,m} = -\omega_{i,j,m}/\omega_{j,j,m}, \quad \beta_{j-1,i,m} = -\omega_{i,j,m}/\omega_{i,i,m}.$$

A bias-corrected estimator of $r_{i,j,m}(1 \le i < j \le d)$ is initially proposed by [14]:

$$(2.5) \quad \widehat{r}_{i,j,m} = -(\widetilde{r}_{i,j,m} + \widetilde{r}_{i,i,m}\widehat{\beta}_{i,j,m} + \widetilde{r}_{j,j,m}\widehat{\beta}_{j-1,i,m}),$$

For $i = j$, the Lemma 2 of [23] showed that

$$(2.6) \quad \max_{1\le i\le d}|\widetilde{r}_{i,i,m} - r_{i,i,m}| = \boldsymbol{O}_p\{(\log(d)/n_m)^{1/2}\},$$

which implies that $\widehat{r}_{i,i,m} = \widetilde{r}_{i,i,m}$ is a nearly unbiased estimator of $r_{i,i,m}$. Thus one can naturally estimate $\omega_{i,j,m}$ by

$$(2.7) \quad T_{i,j,m} = \frac{\widehat{r}_{i,j,m}}{\widehat{r}_{i,i,m}\widehat{r}_{j,j,m}}, 1 \le i \le j \le d,$$

and test $\mathbf{H}_0 : \boldsymbol{\Delta} = 0$ based on the estimators $\mathcal{T} = \{T_{i,j,1} - T_{i,j,2}, 1 \le i \le j \le d\}$.

Considering the heteroscedasticity of estimators in $\mathcal{T}$, Literature [23] proposed the following test statistic for the null hypothesis $\mathbf{H}_0$:

$$M_n = \max_{1\le i\le j\le d} W_{i,j}^2 = \max_{1\le i\le j\le d} \frac{(T_{i,j,1} - T_{i,j,2})^2}{\widehat{\theta}_{i,j,1} + \widehat{\theta}_{i,j,2}},$$

where

$$(2.8) \quad \begin{aligned} W_{i,j} &= \frac{T_{i,j,1} - T_{i,j,2}}{(\widehat{\theta}_{i,j,1} + \widehat{\theta}_{i,j,2})^{1/2}}, \\ \widehat{\theta}_{i,j,m} &= \widehat{\text{Var}}(T_{i,j,m}) = \frac{1 + \widehat{\beta}_{i,j,m}^2\widehat{r}_{i,i,m}/\widehat{r}_{j,j,m}}{n_m\widehat{r}_{i,i,m}\widehat{r}_{j,j,m}}. \end{aligned}$$

Literature [23] obtained the asymptotic null distribution of M_n under suitable conditions, which is type I extreme value distribution. However, this limiting distribution of maximum-type statistic based approach has two fatal limitations. Firstly, the convergence rate of extreme-value statistics is notoriously slow and the process of getting the limiting distribution ignores the correlation between coordinates. Secondly, the maximum-type statistic is particularly powerful against large and sparse signal alternatives, however, it is powerless against small and dense signal alternatives.

In this paper we develop new tests for hypothesis (1.1), which are adaptive to a large variety of alternative scenarios in high dimensions. We utilize the multiplier bootstrap method to approximate the asymptotic distribution of the proposed test statistics and thus overcomes the limitation of the extreme-value-type statistic M_n.

3. METHODOLOGY

As the extreme-value-type statistic is only powerful against the sparse large alternatives, we aim to provide a data-driven adaptive test for the hypothesis (1.1) in this section. In Section 3.1, a family of tests based on (s_0, p)-norm are proposed. The (s_0, p)-norm was first introduced in [28]. The tests based on different p have different powers under different alternative scenarios. For example, (s_0, ∞)-norm based test are sensitive to large perturbations on a small number of entries of $\boldsymbol{\Omega}_1 - \boldsymbol{\Omega}_2$. Moreover, $(s_0, 2)$-norm are sensitive to small perturbations on a large number of entries of $\boldsymbol{\Omega}_1 - \boldsymbol{\Omega}_2$. By combining a family of (s_0, p)-norm based tests with various p, we present our adaptive test in Section 3.2

3.1 The (s_0, p)-norm based test statistics

In this section, we provide some (s_0, p)-norm based tests. Recall that we have defined the $W_{i,j}$ in (2.8). Based on the statistics in $\mathcal{W} = \{W_{i,j}, 1 \leq i \leq j \leq d\}$, define $\mathbf{W} = (W_{i,j})_{d \times d}$. we then propose our test statistic based on (s_0, p)-norm of the vector trivec$(\mathbf{W})$. Specifically, we propose the (s_0, p)-norm based test statistic is

$$(3.1) \qquad N_{(s_0,p)} = \big\| \text{trivec}(\mathbf{W}) \big\|_{(s_0,p)}.$$

With the proposed test statistic, we still need to obtain the critical value or P-value to test (1.1). To this end, we develop a multiplier bootstrap method to approximate the limiting distribution of the test statistic $N_{(s_0,p)}$.

In the high dimensional setting, [4] introduced the multiplier bootstrap method for the sum of independent random vectors. In detail, let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be independent zero mean random vectors in $\mathbb{R}^d$ with $\mathbf{Z}_k = (Z_{k1}, \ldots, Z_{kd})^\top$. The bootstrap sample for the sample mean $n^{-1} \sum_{k=1}^n \mathbf{Z}_k$ then becomes $n^{-1} \sum_{k=1}^n \varepsilon_k \mathbf{Z}_k$, where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent standard normal random variables. Inspired by the multiplier bootstrap method in [4], we propose a specific multiplier bootstrap procedure for the problem here. In detail, we generate independent samples $\eta_{1,1}^b, \ldots, \eta_{1,n_1}^b$ and $\eta_{2,1}^b, \ldots, \eta_{2,n_2}^b$ from $\eta \sim N(0,1)$ for $b = 1, \ldots, B$. Similarly, we set the b-th multiplier bootstrap sample for $\widetilde{r}_{i,j,m}, 1 \leq i \leq j \leq d$ as

$$(3.2) \qquad \widetilde{r}_{i,j,m}^b = \frac{1}{n_m} \sum_{k=1}^{n_m} \eta_{m,k}^b (\widehat{\epsilon}_{k,i,m} \widehat{\epsilon}_{k,j,m} - \widetilde{r}_{i,j,m}).$$

Considering the definitions of $\widehat{r}_{i,j,m}$ in (2.5), we set its b-th bootstrap sample as

$$\widehat{r}_{i,j,m}^b = -\big(\widetilde{r}_{i,j,m}^b + \widetilde{r}_{i,i,m}^b \widehat{\beta}_{i,j,m} + \widetilde{r}_{j,j,m}^b \widehat{\beta}_{j-1,i,m} \big)$$

for $1 \leq i < j \leq d$ and $\widehat{r}_{i,i,m}^b = \widetilde{r}_{i,i,m}^b$.

Further, by the definitions of $T_{i,j,m}$ and $W_{i,j}$ in (2.7) and (2.8) respectively, we then get the b-th bootstrap sample of $T_{i,j,m}$ and $W_{i,j}$ as

$$(3.3) \qquad \begin{aligned} T_{i,j,m}^b &= \frac{\widehat{r}_{i,j,m}^b}{\widehat{r}_{i,i,m} \widehat{r}_{j,j,m}}, 1 \leq i \leq j \leq d, \\ W_{i,j}^b &= \frac{T_{i,j,1}^b - T_{i,j,2}^b}{(\widehat{\theta}_{i,j,1} + \widehat{\theta}_{i,j,2})^{1/2}}, 1 \leq i \leq j \leq d. \end{aligned}$$

With $W_{i,j}^b$, we set $\mathbf{W}^b = (W_{i,j}^b)_{d \times d}$ and finally obtain the bootstrap samples of $N_{(s_0,p)}$ as

$$(3.4) \qquad N_{(s_0,p)}^b = \| \text{trivec}(\mathbf{W}^b) \|_{(s_0,p)}, \qquad b = 1 \ldots B.$$

Given the significance level α, we use $t_{\alpha,(s_0,p)}^N$ to denote the oracle critical values of $N_{(s_0,p)}$. Given the bootstrap samples,

we then estimate $t_{\alpha,(s_0,p)}^N$ by

$$(3.5)$$
$$\widehat{t}_{\alpha,(s_0,p)}^N = \inf \Big\{ t \in \mathbb{R} : \frac{1}{B} \sum_{b=1}^B \mathbb{1}\{N_{(s_0,p)}^b \leq t\} > 1 - \alpha \Big\}.$$

Therefore, we obtain the (s_0, p)-norm based tests for (1.1) as

$$(3.6) \qquad T_{(s_0,p)}^N = \mathbb{1}\big\{ N_{(s_0,p)} \geq \widehat{t}_{\alpha,(s_0,p)}^N \big\}.$$

We reject $\mathbf{H}_0$ of (1.1) if and only if $T_{(s_0,p)}^N = 1$. Accordingly, we estimate $N_{(s_0,p)}$'s oracle P-values $P_{(s_0,p)}^N$ by

$$(3.7) \qquad \widehat{P}_{(s_0,p)}^N = \frac{\sum_{b=1}^B \mathbb{1}\{N_{(s_0,p)}^b > N_{(s_0,p)}\}}{B + 1}.$$

Therefore, given a significance level α, we reject $\mathbf{H}_0$ of (1.1) if and only if $\widehat{P}_{(s_0,p)}^N \leq \alpha$.

3.2 Data adaptive combined test

After providing the (s_0, p)-norm based tests for each individual p, we propose a data-driven adaptive test by combining a group of the (s_0, p)-norm based tests in this section.

Set $\mathcal{P} = \{p_1, p_2, \cdots\}$ as a finite set of positive numbers, and set the size of $\mathcal{P}$ as a finite fixed constant. Then we combine the (s_0, p)-norm based test with $p \in \mathcal{P}$ by taking the minimum P-value of these tests. Specifically, we set the data-adaptive test statistic N_{ad} as

$$(3.8) \qquad N_{\mathrm{ad}} = \min_{p \in \mathcal{P}} \widehat{P}_{(s_0,p)}^N.$$

The detail process of getting N_{ad} is in Algorithm 1. The set $\{\mathcal{P}\}$ can be chosen by users with prior information about the alternative patterns. If one knows the alternative pattern, then he/she can choose the set $\mathcal{P}$ accordingly to improve the power performance of the data adaptive test. For example, let $\mathcal{P}$ consists of large values of p with prior information that the alternative pattern is sparse. If one knows nothing about the alternative pattern, then a balanced set $\mathcal{P}$ containing both large and small p is recommended. For example, one may choose the set $\mathcal{P}$ to be $\{1, 2, 3, 4, 5, \infty\}$.

For the data adaptive test, we need to get the P-value. It's difficult to get the limiting distribution for the (s_0, p)-norm based statistics, not to mention for the data adaptive test statistic. Hence, the intuitive way is to do a double loop bootstrap procedure to get the empirical distribution for our data adaptive test. But this way is too costly for computation. As is shown by Algorithm 1, in addition to the data adaptive statistic N_{ad}, we also obtain the bootstrap samples for $N_{(s_0,p)}$, i.e, $\{N_{(s_0,p)}^1, \ldots, N_{(s_0,p)}^B\}$. Therefore, we can recycle the bootstrap samples to accelerate our computation speed. Specifically, for $b = 1, \ldots, B$ and $p \in \mathcal{P}$, we set

$$\widehat{P}_{(s_0,p)}^{b,N} = \frac{\sum_{b_1 \neq b} \mathbb{1}\{N_{(s_0,p)}^{b_1} > N_{(s_0,p)}^b\}}{B}.$$

Algorithm 1 A bootstrap procedure to obtain N_{ad}
___
Input: $\mathcal{X}$.
      Output: $N_{(s_0,p)}^1, \ldots, N_{(s_0,p)}^B$ with $p \in \mathcal{P}$, and N_{ad}.

1: **procedure**
2:    $N_{(s_0,p)} = \|\mathrm{trivec}(\mathbf{W})\|_{(s_0,p)}$ with $\mathbf{W} = (W_{i,j})_{d \times d}^\top$ and $W_{i,j} = (T_{i,j,1} - T_{i,j,2})/(\widehat{\theta}_{i,j,1} + \widehat{\theta}_{i,j,2})^{1/2}$.
3:    **for** $b \leftarrow 1$ **to** B **do**
4:       Sample independent standard normal random variables $\{\eta_{1,1}^b, \ldots, \eta_{1,n_m}^b\}$ for $m = 1, 2$.
5:       For $1 \leq i \leq j \leq d$, set $\widetilde{r}_{i,j,m}^b = (1/n_m) \sum_{k=1}^{n_m} \eta_{m,k}^b (\widehat{\epsilon}_{k,i,m} \widehat{\epsilon}_{k,j,m} - \widetilde{r}_{i,j,m})$.
6:       Set $\widehat{r}_{i,i,m}^b = \widetilde{r}_{i,i,m}^b$, and set $\widehat{r}_{i,j,m}^b = -(\widetilde{r}_{i,j,m}^b + \widetilde{r}_{i,i,m}^b \widehat{\beta}_{i,j,m} + \widetilde{r}_{j,j,m}^b \widehat{\beta}_{j-1,i,m})$ for $1 \leq i < j \leq d$.
7:       Set $T_{i,j,m}^b = \widehat{r}_{i,j,m}^b/(\widehat{r}_{i,i,m} \widehat{r}_{j,j,m})$, $1 \leq i \leq j \leq d$.
8:       Set $W_{i,j}^b = (T_{i,j,1}^b - T_{i,j,2}^b)/(\widehat{\theta}_{i,j,1} + \widehat{\theta}_{i,j,2})^{1/2}$, $1 \leq i \leq j \leq d$ and $\mathbf{W}^b = (W_{i,j}^b)_{d \times d}$.
9:       **for** p in $\mathcal{P}$ **do**
10:         $N_{(s_0,p)}^b = \|\mathrm{trivec}(\mathbf{W}^b)\|_{(s_0,p)}$ with $\mathbf{W}^b = (W_{i,j}^b)_{d \times d}$.
11:       **end for**
12:    **end for**
13:    $\widehat{P}_{(s_0,p)}^N = \sum_{b=1}^B \mathbb{1}\{N_{(s_0,p)}^b > N_{(s_0,p)}\}/(B+1)$ for $p \in \mathcal{P}$.
14:    $N_{\mathrm{ad}} = \min_{p \in \mathcal{P}} \widehat{P}_{(s_0,p)}^N$.
15: **end procedure**

___

Algorithm 2 A low-cost bootstrap procedure
___
      Input: $\mathcal{X}$ and $N_{(s_0,p)}^1, \ldots, N_{(s_0,p)}^B$ for $p \in \mathcal{P}$.
Output: $N_{\mathrm{ad}}^1, \ldots, N_{\mathrm{ad}}^B$.

1: **procedure**
2:    **for** $b \leftarrow 1$ **to** B **do**
3:       **for** p in $\mathcal{P}$ **do**
4:         $\widehat{P}_{(s_0,p)}^{b,N} = \sum_{b_1 \neq b} \mathbb{1}\{N_{(s_0,p)}^{b_1} > N_{(s_0,p)}^b\}/B$.
5:       **end for**
6:       $N_{\mathrm{ad}}^b = \min_{p \in \mathcal{P}} \widehat{P}_{(s_0,p)}^{b,N}$.
7:    **end for**
8: **end procedure**

___

We use $N_{\mathrm{ad}}^b = \min_{p \in \mathcal{P}} \widehat{P}_{(s_0,p)}^{b,N}$ as the bootstrap sample for N_{ad}. We then estimate the oracle P-value of N_{ad} by

$$(3.9) \qquad \widehat{P}_{\mathrm{ad}}^N = \frac{\left(\sum_{b=1}^B \mathbb{1}\{N_{\mathrm{ad}}^b \leq N_{\mathrm{ad}}\}\right) + 1}{B + 1}.$$

For more details, see Algorithm 2. The samples $N_{\mathrm{ad}}^1, \ldots, N_{\mathrm{ad}}^B$ are nonindependent. But as $n, B \to \infty$, they are asymptotically independent. Hence, it doesn't affect the consistency of $\widehat{P}_{\mathrm{ad}}^N$. After getting the estimated P-values of the data-adaptive tests N_{ad}, given the significance level α, we set

$$(3.10) \qquad T_{\mathrm{ad}}^N = \mathbb{1}\{\widehat{P}_{\mathrm{ad}}^N \leq \alpha\}.$$

Therefore, we reject $\mathbf{H}_0$ of (1.1) if and only if $T_{\mathrm{ad}}^N = 1$.

4. THEORETICAL PROPERTIES

In this section, we investigate the theoretical properties of our proposed test. Firstly, some assumptions are introduced in Section 4.1. In Section 4.2, we verify the validity of multiplier bootstrap which is used in Section 3 and then analyze the theoretical properties of the proposed test.

4.1 Assumptions

In this section, we introduce some assumptions that are commonly used in high-dimensional analysis.

(A) Set $n = \max(n_1, n_2)$, there exists some $0 < \delta < 1/7$ such that $s_0^2 = o(\log(d))$ and $s_0^2 \log(d) = \boldsymbol{o}(n^\delta)$ hold, where $n_1 \asymp n_2 \asymp n$.

Assumption **(A)** allows s_0 and d go to infinity as long as $s_0^2 = o(\log(d))$ and $s_0^2 \log(d) = \boldsymbol{o}(n^\delta)$ hold. These are technical conditions that are critical to the validity of the multiplier bootstrap procedure. By using the multiplier bootstrap to get the critical values for our tests, we need some more assumptions compared with [23]. Other than the Assumption **(A)**, we also introduce a more strong Assumption **(A)**′ to state the scaling of s_0, d and n. Before stating the next assumption, we need some additional notations. For $1 \leq i, j \leq d$ and $m = 1, 2$, let $U_{k,i,j,m} = \epsilon_{k,i,m} \epsilon_{k,j,m} - \mathbb{E}(\epsilon_{k,i,m} \epsilon_{k,j,m})$ and define $\widetilde{U}_{k,i,j,m} = (r_{i,j,m} - U_{k,i,j,m})/r_{i,i,m} r_{j,j,m}$ with $k = 1, \ldots, n_m$. Define $\widetilde{\mathbf{U}}_{k,m} = (\widetilde{U}_{k,i,j,m})$ as a square matrix of order d and denote the covariance matrix of $\mathrm{trivec}(\widetilde{\mathbf{U}}_{k,m})$ as $\boldsymbol{\Sigma}_m^{\widetilde{U}} = (\sigma_{s,t,m}^{\widetilde{U}})_{1 \leq s,t \leq d(d-1)/2}$, where

$$\sigma_{s,t,m}^{\widetilde{U}} = \begin{cases} \dfrac{1 + \beta_{i,j,m}^2 r_{i,i,m}/r_{j,j,m}}{r_{i,i,m} r_{j,j,m}}, & s = t, \\[2ex] \dfrac{r_{i_1,i_2,m} r_{j_1,j_2,m} + r_{i_1,j_2,m} r_{i_2,j_1,m}}{r_{i_1,i_1,m} r_{j_1,j_1,m} r_{i_2,i_2,m} r_{j_2,j_2,m}}, & s \neq t, \end{cases}$$

with $1 \leq i < j \leq d$, $1 \leq i_1 < j_1 \leq d$, $1 \leq i_2 < j_2 \leq d$, $i_1 \neq i_2$, $j_1 \neq j_2$, $m = 1, 2$. Furthermore, the variance for the sample mean of $\widetilde{U}_{k,i,j,m}$ is $\theta_{i,j,m} = \sigma_{s,s,m}^{\widetilde{U}}/n_m$.

Let $\boldsymbol{G}$ be a Gaussian random vector in $\mathbb{R}^{d(d-1)/2}$ with mean zero and covariance matrix $\mathbf{R}_{12}^{\widetilde{U}}$, where $\mathbf{R}_{12}^{\widetilde{U}} = (\mathbf{D}_{12}^{\widetilde{U}})^{-1/2} \boldsymbol{\Sigma}_{12}^{\widetilde{U}} (\mathbf{D}_{12}^{\widetilde{U}})^{-1/2}$ with $\boldsymbol{\Sigma}_{12}^{\widetilde{U}} = \boldsymbol{\Sigma}_1^{\widetilde{U}}/n_1 + \boldsymbol{\Sigma}_2^{\widetilde{U}}/n_2$ and $\mathbf{D}_{12}^{\widetilde{U}} = \mathrm{Diag}(\boldsymbol{\Sigma}_{12}^{\widetilde{U}})$. Set the probability density function (PDF) and the α-quantile of $\|\boldsymbol{G}\|_{(s_0,p)}$ as $f_{\boldsymbol{G},(s_0,p)}$ and $c_{(s_0,p)}(\alpha)$ respectively. We then define $h_T(z)$ as

$$h_T(z) = \max_{p \in \mathcal{P}} \max_{x \in C_{(s_0,p)}(z)} f_{\boldsymbol{G},(s_0,p)}^{-1}(x)$$

$$\text{with} \quad C_{(s_0,p)}(z) = [c_{(s_0,p)}(z), c_{(s_0,p)}(1-z)].$$

With these new notations, we introduce the following assumption.

(A)′ Define $n = \max(n_1, n_2)$. We assume that $s_0^2 = o(\log(d))$ and $h_T^{0.6}(z) s_0^2 \log(d) = \boldsymbol{o}(n^{1/10})$ holds for any $0 < z < 1$ as $n, d \to \infty$ and $n_1 \asymp n_2 \asymp n$.

Assumption $(\mathbf{A})'$ is more stringent. It is critical to guarantee the uniform convergence of the distribution functions and the corresponding quantile functions of the test statistics $N_{(s_0,p)}$ for any $p \in \mathcal{P}$. The next two mild assumptions are often used in high dimensional setting, especially when the inference for covariance matrix and precision matrix are involved.

(B) There exist some positive constants $C_0 < C_1$, such that $\lambda_{\min}(\mathbf{\Omega}_m) \geq C_0$ and $\lambda_{\max}(\mathbf{\Omega}_m) \leq C_1$, with $m = 1, 2$. There exists some $\tau > 0$ such that $|A_\tau| = \boldsymbol{o}(d^{1/16})$ where $A_\tau = \{(i,j) : |w_{i,j,m}| \geq (\log(d))^{-2-\tau}, 1 \leq i < j \leq d$, for $m = 1$ or $2\}$.

(C) Let D_m be the diagonal of $\mathbf{\Omega}_m$ and let $(\eta_{i,j,m}) = D_m^{-1/2}\mathbf{\Omega}_m D_m^{-1/2}$, for $m = 1, 2$. Assume that $\max_{1 \leq i < j \leq d} |\eta_{i,j,m}| \leq \eta_m \leq c$, where $0 < c < 1$ is a constant.

(D) Suppose $\max_{1 \leq i \leq d} s_{i,m} = \boldsymbol{o}(n^{1/2}/(\log(d))^{3/2})$, where $s_{i,m}$ is sparsity for the i-th row or column of $\mathbf{\Omega}_m$ for $m = 1, 2$.

Note that $\boldsymbol{\beta}_{i,m} = -\omega_{i,i,m}^{-1}\mathbf{\Omega}_{-i,i,m}$, then the sparsity conditions of the Proposition 4.1 in [14] are automatically satisfied under Assumption **(D)**.

4.2 Theoretical analysis

After introducing some needing assumptions, we analyze the theoretical properties of our test. Due to the complicated structure of our test statistics, we use the multiplier bootstrap to get the critical values for our test in Section 3. But this procedure is different from [28]. Specifically, other than the testing statistics cannot be rewritten as a sum of independent random variables, there are also some bias correction terms. Hence, we need to justify the validity of this multiplier bootstrap.

Theorem 4.1. *Suppose Assumptions $(\mathbf{A})$-$(\mathbf{D})$ hold. Under the null hypothesis $\mathbf{H}_0$ of (1.1), we have as $n, d \to \infty$,*

(4.1)
$$\sup_{z \in (0,\infty)} \left| \mathbb{P}(N_{(s_0,p)} \leq z) - \mathbb{P}(N_{(s_0,p)}^b \leq z | \mathcal{X}, \mathcal{Y}) \right| = \boldsymbol{o}_p(1).$$

Under the Gaussian distribution setting, it is easily to check that the sub-exponential distribution assumption and the moment condition in [4] are satisfied. Hence, under milder conditions, Theorem 4.1 verifies the validity of the multiplier bootstrap method. The proof of Theorem 4.1 is in the Appendix.

By Theorem (4.1) hold, it's easy to prove that the size of $T_{(s_0,p)}^N$ asymptotically coverges to pre-specified significance level α.

Corollary 1. *Suppose Assumptions $(\mathbf{A})$-$(\mathbf{D})$ hold. Under the null hypothesis $\mathbf{H}_0$ of (1.1), we have*

$$\mathbb{P}_{\mathbf{H}_0}(T_{(d_0,p)}^N = 1) \to \alpha,$$

as $n, B \to \infty$.

With Theorem 4.1, we then show the theoretical properties of our data-adaptive test T_{ad}^N. By the definition of $\widehat{P}_{\mathrm{ad}}^N$ and T_{ad}^N in (3.9) and (3.10), it can be seen that T_{ad}^N relies on the estimated P-values of $N_{(s_0,p)}$. Therefore, we suppose the more stringent Assumption $(\mathbf{A})'$ holds, which guarantees the uniform convergence of the distribution functions and the corresponding quantile functions of the test statistics $N_{(s_0,p)}$ for any $p \in \mathcal{P}$. Under this condition, we show that the empirical size of the data-adaptive test T_{ad}^N approximates to the pre-specified level α.

Theorem 4.2. *Suppose Assumptions $(\mathbf{A})'$, $(\mathbf{B})$-$(\mathbf{D})$ hold. Under the null hypothesis $\mathbf{H}_0$ of (1.1), we have*

$$\mathbb{P}_{\mathbf{H}_0}(T_{\mathrm{ad}}^N = 1) \to \alpha,$$

as $n, B \to \infty$.

After analyzing the asymptotic sizes of the tests in Section 3, we summarize the asymptotic power properties in the following theorem. To analyze the power performance of T_{ad}^N, we need to introduce some other notations. Define $\mathbf{W}^* = (W_{i,j}^*)_{d \times d}$ with

$$W_{i,j}^* = \left| \widetilde{U}_{i,j,1} - \widetilde{U}_{i,j,2} \right| \Big/ \sqrt{\theta_{i,j,1}/n_1 + \theta_{i,j,2}/n_2},$$

where $\theta_{i,j,m}$ are the diagonal elements of $\mathbf{\Sigma}_m^{\widetilde{U}}$, $m = 1, 2$. We then introduce the following theorem to characterize the asymptotic power properties of $T_{(s_0,p)}^N$ and T_{ad}^N.

Theorem 4.3. *Suppose Assumptions $(\mathbf{B})$-$(\mathbf{D})$ hold and assume $\varepsilon_n = \boldsymbol{o}(1)$, $\varepsilon_n \sqrt{\log(d)} \to \infty$ as $n, d \to \infty$.*

(a) As $n, d \to \infty$, there exists some $\delta_1 > 0$ such that $\log(d) = \boldsymbol{o}(n^{1/3})$ and $n = \boldsymbol{O}(d^{2\delta_1})$. Assume $s_0 = \boldsymbol{O}\big((\log(d))^{\delta_2}\big)$ for some postive constant δ_2. Under the alternative hypothesis $\mathbf{H}_1$ of (1.1) and with

$$\left\| \mathrm{trivec}(\mathbf{W}^*) \right\|_{(s_0,p)} \geq s_0(1 + \varepsilon_n)\big(\sqrt{2\log(d(d-1)/2)} + \sqrt{2\log(1/\alpha)}\big),$$

hold, we have $\mathbb{P}_{\mathbf{H}_1}\big(T_{(s_0,p)}^N = 1\big) \to 1$ as $n, d, B \to \infty$.

(b) With Assumption $(\mathbf{A})'$ hold, and under the alternative hypothesis $\mathbf{H}_1$ of (1.1) and suppose

$$\left\| \mathrm{trivec}(\mathbf{W}^*) \right\|_{(s_0,p)} \geq s_0(1 + \varepsilon_n)\big(\sqrt{2\log(d(d-1)/2)} + \sqrt{2\log(\#\{\mathcal{P}\}/\alpha)}\big),$$

hold, we have $\mathbb{P}_{\mathbf{H}_1}\big(T_{\mathrm{ad}}^N = 1\big) \to 1$ as $n, d, B \to \infty$.

By Theorem 4.3, we show that the asymptotic powers of $T_{(s_0,p)}^N$ and T_{ad}^N converge to 1 under the minimum signal condition on $\|\mathrm{trivec}(\mathbf{W}^*)\|_{(s_0,p)}$.

5. EXPERIMENTS

5.1 Simulation study

In this section, we conduct simulation study to investigate the empirical size and power of the proposed test. To show the adaptivity of our method, we compare it with recently developed method proposed by [23] under various model settings. We denote the test proposed by [23] as T_{CX} for simplifing notations. To distinguish the adaptive test with different s_0, we denote the adaptive test with any fixed s_0 as $T_{s_0,\mathrm{ad}}^N$.

In the simulation study, the sample sizes are set to be $n_1 = n_2 = 200$, while the dimension $d = 100$. Although the dimension d seems to be small compared with the sample size, the parameters in the precision matrix which we are interested in are already much larger than the sample size $(d(d-1)/2)$. In all simulations, the bootstrap sample sizes B are set to be 1000 and all the simulation results are based on 1000 replications. Under the null hypothesis $\mathbf{H}_0$, we set $\boldsymbol{\Omega}_2 = \boldsymbol{\Omega}_1 = \boldsymbol{\Omega}$. Under the alternative hypothesis $\mathbf{H}_1$, we set $\boldsymbol{\Omega}_1 = \boldsymbol{\Omega} + \delta\boldsymbol{I}$ and $\boldsymbol{\Omega}_2 = \boldsymbol{\Omega} + \boldsymbol{\Gamma} + \delta\boldsymbol{I}$, where $\boldsymbol{\Gamma} = (\gamma_{i,j})_{d\times d}$ is a nonzero matrix and $\delta = |\lambda_{\min}(\boldsymbol{\Omega}+\boldsymbol{\Gamma})| + 0.05$. Suppose there are m_t nonzero entries of $\boldsymbol{\Gamma}$. Specially, we random sample $m_t/2$ locations in the upper triangle of $\boldsymbol{\Gamma}$ and set each with a magnitude r. By the symmetric requirement of $\boldsymbol{\Omega}_2$, the location and magnitude of the other $m_t/2$ nonzero entries of $\boldsymbol{\Gamma}$ can be determined by its upper triangle. To show that our test is adaptive to various alternative patterns, we set the nonzero entries of the $\boldsymbol{\Gamma}$ as $m_t = 20, 200, 1000, 2000$. The $m_t = 20, 200$ are to illustrate the sparse alternatives and $m_t = 1000, 2000$ are to represent the dense alternatives. For all the simulations, simulation data are generated from multivariate Gaussian distributions with mean $\mathbf{0}$ and covariance matrices $\boldsymbol{\Sigma}_1 = (\boldsymbol{\Omega}_1)^{-1}$ and $\boldsymbol{\Sigma}_2 = (\boldsymbol{\Omega}_2)^{-1}$. The nominal significance level for all the tests are set to be $\alpha = 0.05$. To study the empirical performance of the test, the following three models of $\boldsymbol{\Omega}$ are considered.

Model 1: $\boldsymbol{\Omega}^* = (\omega_{i,j}^*)$ where $\omega_{i,i}^* = 1$, $\omega_{i,j}^* = 0.5 \times$ Bernoulli$(1, 0.5)$ for $i < j$ and $\omega_{j,i}^* = \omega_{i,j}^*$. $\boldsymbol{\Omega} = (\boldsymbol{\Omega}^* + \delta\boldsymbol{I})/(1+\delta)$ with $\delta = |\lambda_{\min}(\boldsymbol{\Omega}^*)| + 0.05$.

Model 2: $\boldsymbol{\Sigma}^* = (\sigma_{i,j}^{*(1)})$ where $\sigma_{i,i}^{*(1)} = 1$, $\sigma_{i,j}^{*(1)} = 0.5$ for $2(k-1)+1 \leq i \neq j \leq 2k$, where $k = 1, \ldots, [p/2]$ and $\sigma_{i,j}^{*(1)} = 0$ otherwise. $\boldsymbol{\Omega} = \{(\boldsymbol{\Sigma}^* + \delta\boldsymbol{I})/(1+\delta)\}^{-1}$ with $\delta = |\lambda_{\min}(\boldsymbol{\Sigma}_1^*)| + 0.05$.

Model 3: $\boldsymbol{\Omega}^* = (\omega_{i,j}^{*(1)})$ where $\omega_{i,i}^{*(1)} = 1$, $\omega_{i,j}^{*(1)} = 0.5 \times$ Bernoulli$(1, 0.3)$ for $i < j$ and $\omega_{j,i}^{*(1)} = \omega_{i,j}^1$. Other than that, we set $\omega_{i,j}^{*(1)} = \omega_{j,i}^{*(1)} = 0.5$ for $i = 20(k-1)+1$ and $20(k-1)+2 \leq j \leq 20(k-1)+20$, $1 \leq k \leq p/20$. $\omega_{i,j}^{*(1)} = 0$ otherwise. $\boldsymbol{\Omega} = (\boldsymbol{\Omega}^* + \delta\boldsymbol{I})/(1+\delta)$ with $\delta = |\lambda_{\min}(\boldsymbol{\Omega}^*)| + 0.05$.

The performances of the test methods under various alternative patterns for Model 1 are shown in Figure 1. In Figure 1, the orange line with circles represents the adaptive test $T_{10,\mathrm{ad}}^N$, the blue line with triangles represents the adaptive test $T_{100,\mathrm{ad}}^N$, the red line with crosses represents the adaptive test $T_{500,\mathrm{ad}}^N$, the green line with diamonds represents the adaptive test $T_{1000,\mathrm{ad}}^N$, the black line with stars represents the T_{CX} test proposed by [23]. The horizontal axis represents magnitude r in the upper triangle of $\boldsymbol{\Gamma}$, a larger value of r indicates a stronger signal. The vertical axis represents the empirical powers of different tests, while $r = 0$ corresponds to the empirical sizes.

From Figure 1, we can see that all the empirical sizes of different tests are under control. Under sparse alternative setting with $m_t = 20$ (corresponds to the upper left panel in Figure 1 with 20 non-equal signals), the empirical powers of maximum-norm based test T_{CX} are the highest and the empirical powers of the adaptive test with $s_0 = 10$ are still comparable though a little bit lower than CX test. Besides, with s_0 decreasing, the adaptive test tends to more powerful under sparse alternative setting. As the non-equal number m_t becomes bigger, the empirical powers of the adaptive test with larger s_0 are getting better and better. Under the dense alternative (corresponds to the upper right panel and lower panels in Figure 1 with more than 200 non-equal signals), the empirical powers of the adaptive tests are greater than those of the maximum-norm based CX tests with the magnitude r larger than certain threshold. Although the empirical powers of adaptive test with different s_0 have some difference, the empirical powers of the adaptive test show some robustness for the small changes of s_0. From Figure 1, we can also see that the empirical power of $T_{500,\mathrm{ad}}^N$ and $T_{1000,\mathrm{ad}}^N$ are almost equal to each other.

At last, we point out that the influence of the parameter s_0 on the power performance is more complicated. However, by choosing s_0 close to half of the true number of nonzero signals $m_t/2$, the tests enjoy good performance. In practice, we can determine s_0 by the prior information or some empirical information.

The empirical results for Model 2 and Model 3 are similar as for Model 1 and thus are presented in the supplementary materials (http://intlpress.com/site/pub/files/_supp/sii/2020/0013/0001/SII-2020-0013-0001-s004.pdf) for saving space here.

5.2 Real data analysis

In this section, we apply our adaptive test method to a gene expression data set which is associated with lung cancer. The data set is publicly available from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) at accession number GDS2771. The data set is made up of 22,283 microarray-derived gene expression measurements from large airway epithelial cells sampled from 97 patients with lung cancer, and 90 control patients. [15] showed that the Wnt pathway associated with lung cancer and many other lung diseases such as interstitial lung disease (ILD)
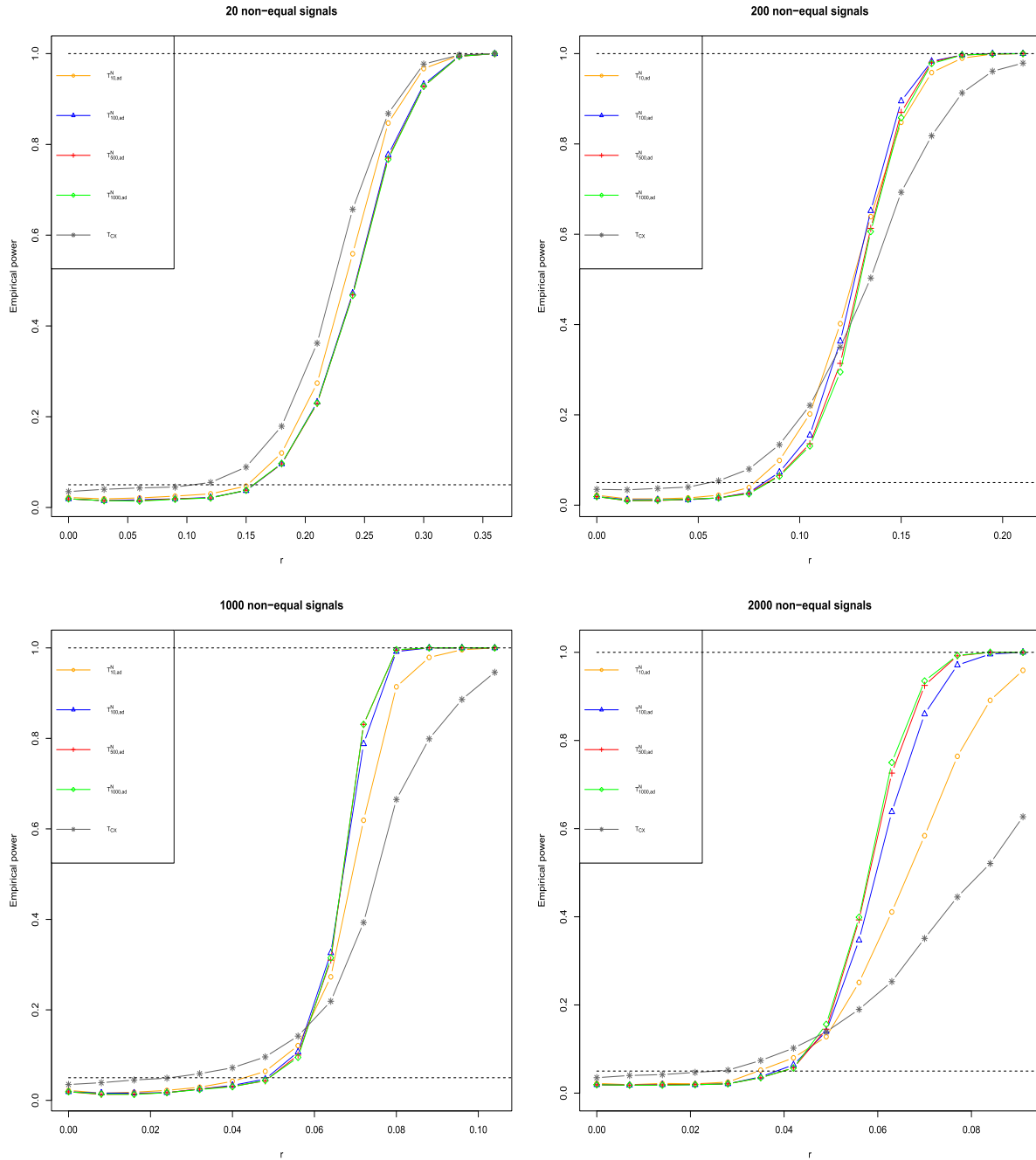
*Figure 1. Empirical powers of various tests for **Model 1**. The orange line with circles represents the adaptive test $T^N_{10,\mathrm{ad}}$, the blue line with triangles represents the adaptive test $T^N_{100,\mathrm{ad}}$, the red line with crosses represents the adaptive test $T^N_{500,\mathrm{ad}}$, the green line with diamonds represents the adaptive test $T^N_{1000,\mathrm{ad}}$, the black line with stars represents the T_{CX} test.*

and asthma. The Wnt pathway is also implicated in the development of several types of cancers, such as gastric cancer [5], breast cancer [10].

Hence, in this paper, we focus our analysis on the 188 genes in the Wnt signaling pathway, with 97 patients with lung cancer and 90 control patients. Gene expression levels were analyzed on a logarithmic scale and each gene feature was standardized within each group. Although the true conditional dependence relationships are unknown, we believe that there exists some specific links among genes in the Wnt signaling pathway of the patients with lung cancer. Hence, we use our method to test whether the underlying precision matrices of the patients with cancer or not are equal to each other. In the real example, we know nothing about the un-
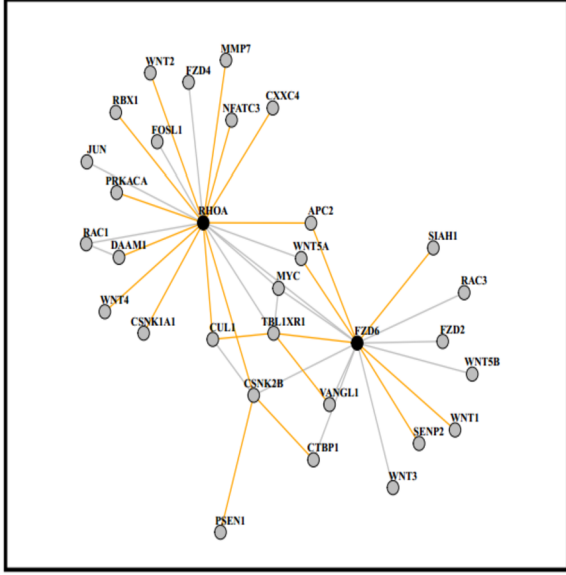
Figure 2. The differential networks estimated for the Wnt signaling pathway. Orange edges show an increase in conditional dependency from control group to lung cancer patient group; grey edges show a decrease.

derlying alternative patterns. Hence, as to the choice of s_0, we propose to tune s_0 in a finite set $\mathcal{S}$. Specifically, we set the doubly tuned data-adaptive test statistic as

$$(5.1) \qquad TN_{\mathrm{ad}} = \min_{p \in \mathcal{P}, s_0 \in \mathcal{S}} \widehat{P}^N_{(s_0, p)}.$$

As long as the cardinality of the set $\mathcal{S}$ is fixed, all the theoretical properties for the adaptive test N_{ad} with fixed s_0 still hold for TN_{ad}. Furthermore, the simulation study in Section 5.1 showed that the empirical powers of the adaptive tests show robustness to different s_0. Hence, assuming the cardinality of the set $\mathcal{S}$ to be finite is reasonable. Specifically, we choose $\mathcal{S} = \{10, 50, 100, 500, 1000, 2000\}$ here. By the adaptive test TN_{ad}, we reject the null hypothesis and think that there are difference for the underlying conditional dependence relationships. Hence, we use the differential network estimation approach in [27] to estimate the differential network between the control group and the patient group. In detail, we choose the tuning parameter by the Bayesian information criterion (BIC) using the element-wise L_1 loss function. The differential network structure is given in Figure 2, in which the black edges represents the conditional correlations in lung cancer group are stronger compared with those in the control group, and the gray edges the other way around. From Figure 2, many potentially important genes for lung cancer are detected, such as WNT1, WNT2, WNT5A etc, see [15]. By Figure 2, we see that RHOA and FZD6 are two hubs in this graph. Hence, we may conclude that the connections of these two genes to other genes are important for identifying the lung cancer.

Actually the importance of FDZ and RHOA can be referred to [6], [17].

6. DISCUSSION

In this paper we propose an adaptive approach for testing the equality of two precision matrices, i.e. to investigate whether the network of connected node pairs change from one state to another. In the Gaussian setting, the precision matrix can be described in terms of regression models and the elements of the precision matrix have a direct correspondence connection with the correlations of the error term. By Lasso or Dantzig selector, the regression coefficient estimator and the corresponding estimated regression errors are obtained. Based on the bias corrected estimator of the correlations of the error terms, we propose to construct a family of (s_0, p)-norm based test statistics with different p. By taking the minimum P-value of these tests, we construct an adaptive test statistics. We utilize multiplier bootstrap method to approximate the limiting distribution of the test statistic. Theoretical guarantees are provided for the proposed procedure and numerical study illustrates its good empirical performance under various alternatives.

The current work relies heavily on the Gaussian graph assumption which is sometimes restrictive in real application. In the future, we will consider the adaptive test of more general graphical models.

APPENDIX A. TECHNICAL LEMMAS

Before proving the main results, we introduce some technical lemmas which is useful in proving the main theorems. Let $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ be independent centered random vectors in $\mathbb{R}^d$ with $\boldsymbol{Z}_k = (Z_{k1}, \ldots, Z_{kd})^\top$, for $k = 1, \ldots, n$ and $\boldsymbol{G}_k$ $(k = 1, \ldots, n)$ be independent Gaussian random vectors in $\mathbb{R}^d$ with the same mean vector and covariance matrix as $\boldsymbol{Z}_k$, and assume the following conditions hold:

(M1) $n^{-1} \sum_{k=1}^n \mathbb{E}\big[(\mathbf{v}^\top \boldsymbol{Z}_k)^2\big] \geq b > 0$ for any $\mathbf{v} \in \mathcal{V}_{s_0}$ with $\mathcal{V}_{s_0} := \{\mathbf{v} \in \mathbb{S}^{d-1} : \|\mathbf{v}\|_0 \leq s_0\}$.

(M2) $n^{-1} \sum_{k=1}^n \mathbb{E}\big[|Z_{ki}|^{2+\ell}\big] \leq Q^\ell$ for $\ell = 1, 2$ and $i = 1, \ldots, d$.

(M3) $\mathbb{E}\big[\exp(|Z_{ki}|/Q)\big] \leq 2$ for $i = 1, \ldots, d$ and $k = 1, \ldots, n$.

Lemma A.1. *(Lemma A.1 in [28]) Assume $s_0^2 \log(dn) = O(n^\zeta)$ with $0 < \zeta < 1/7$ and $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ satisfy **(M1)**, **(M2)**, and **(M3)**. For $1 \leq p \leq \infty$ and sufficiently large n, there is a constant $\zeta_0 > 0$ such that*

$$\sup_{z \in (0, \infty)} \Big| \mathbb{P}\Big(\|S_n^{\boldsymbol{Z}}\|_{(s_0, p)} \leq z\Big) \\ - \mathbb{P}\Big(\|S_n^{\boldsymbol{G}}\|_{(s_0, p)} \leq z\Big)\Big| \leq n^{-\zeta_0},$$

where $S_n^{\boldsymbol{Z}} = n^{-1/2} \sum_{k=1}^n \boldsymbol{Z}_k$, $S_n^{\boldsymbol{G}} = n^{-1/2} \sum_{k=1}^n \boldsymbol{G}_k$ and C depends on b and Q.

Lemma A.2. *(Product of sub-Gaussian is sub-exponential, Lemma 2.7.7 in [19]) Let X and Y be sub-Gaussian random variables, we have XY is sub-exponential. Moreover,*

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2},$$

where the ψ_α-norm ($\alpha \geq 1$) of X is defined by

$$\|X\|_{\psi_\alpha} := \inf\Big(c > 0 : \mathbb{E}\big(\exp(|X|^\alpha/c^\alpha)\big) \leq 2\Big).$$

We then introduce the following lemma to get the bound for $U_{i,j,m}$ and the bias of $\widehat{r}_{i,j,m}$ uniformly in $1 \leq i < j \leq d$.

Lemma A.3. *Let $\overline{U}_{i,j,m} = 1/n_m \sum_{k=1}^{n_m} U_{k,i,j,m}$. With Assumption (B)-(D) hold, and for $\log(d) = o\{n^{1/2}\}$, we have*

$$(A.1) \qquad \max_{\substack{1 \leq i < j \leq d \\ m=1,2}} |\overline{U}_{i,j,m}| = O_p\left\{\big(\log(d)/n\big)^{1/2}\right\},$$

and

$$(A.2) \qquad \max_{\substack{1 \leq i < j \leq d \\ m=1,2}} |r_{i,j,m} - \widehat{r}_{i,j,m}| = O_p\left\{\big(\log(d)/n\big)^{1/2}\right\},$$

for sufficient large n.

The proof of this lemma is in Section 1.2 of the supplementary materials.

Corresponding to the definition of $\boldsymbol{\Sigma}_m^{\widetilde{U}}$ in Section 4.1, we introduce its plug-in covariance matrix estimator as $\widehat{\boldsymbol{\Sigma}}_m^{\widetilde{U}} = (\widehat{\sigma}_{s,t,m}^{\widetilde{U}})_{1 \leq s,t \leq d(d-1)/2}$, where

$$(A.3)$$

$$\widehat{\sigma}_{s,t,m}^{\widetilde{U}} = \begin{cases} \dfrac{1 + \widehat{\beta}_{i,j,m}^2 \widehat{r}_{i,i,m}/\widehat{r}_{j,j,m}}{\widehat{r}_{i,i,m}\widehat{r}_{j,j,m}}, & s = t, \\[3mm] \dfrac{\widehat{r}_{i_1,i_2,m}\widehat{r}_{j_1,j_2,m} + \widehat{r}_{i_1,j_2,m}\widehat{r}_{i_2,j_1,m}}{\widehat{r}_{i_1,i_1,m}\widehat{r}_{j_1,j_1,m}\widehat{r}_{i_2,i_2,m}\widehat{r}_{j_2,j_2,m}}, & s \neq t, \end{cases}$$

with $1 \leq i < j \leq d$, $1 \leq i_1 < j_1 \leq d$, $1 \leq i_2 < j_2 \leq d$, $i_1 \neq i_2$, $j_1 \neq j_2$, $m = 1,2$. Hence, for the variance, we have $\widehat{\theta}_{i,j,m} = \widehat{\sigma}_{s,s,m}^{\widetilde{U}}/n_m$.

Set the correlation matrix of $\mathrm{trivec}(\widetilde{\mathbf{U}}_m)$ as $\mathbf{R}_m^{\widetilde{U}} = (r_{s,t,m}^{\widetilde{U}})_{1 \leq s,t \leq d(d-1)/2}$, and its plug-in estimator as $\widehat{\mathbf{R}}_m^{\widetilde{U}} = (\widehat{r}_{s,t,m}^{\widetilde{U}})_{1 \leq s,t \leq d(d-1)/2}$. By the definition of correlation, we have

$$r_{s,t,m}^{\widetilde{U}} = \frac{\sigma_{s,t,m}^{\widetilde{U}}}{\sqrt{\sigma_{s,s,m}^{\widetilde{U}}\sigma_{t,t,m}^{\widetilde{U}}}}, \quad \widehat{r}_{s,t,m}^{\widetilde{U}} = \frac{\widehat{\sigma}_{s,t,m}^{\widetilde{U}}}{\sqrt{\widehat{\sigma}_{s,s,m}^{\widetilde{U}}\widehat{\sigma}_{t,t,m}^{\widetilde{U}}}}.$$

With the Lemma A.3 holding, we then introduce the following lemma to bound the estimation error of the plug-in estimator $\widehat{\sigma}_{s,t,m}^{\widetilde{U}}$ and $\widehat{r}_{s,t}$.

Lemma A.4. *With Assumptions (B)-(D) hold, and for $\log(d) = o\{n^{1/2}\}$, we have*

$$(A.4) \quad \begin{aligned} &\max_{\substack{1 \leq s,t \leq d(d-1)/2 \\ m=1,2}} \left(|\widehat{\sigma}_{s,t,m}^{\widetilde{U}} - \sigma_{s,t,m}^{\widetilde{U}}|, |\widehat{r}_{s,t,m}^{\widetilde{U}} - r_{s,t,m}^{\widetilde{U}}|\right) \\ &= o_p\left\{\big(n\log(d)\big)^{-1/4}\right\}, \end{aligned}$$

for sufficient large n.

The proof of Lemma A.4 is in Section 1.3 of the supplementary materials.

Lemma A.4 bound the estimation error of the plug-in estimator, it is important to get the approximated distribution of $\|\mathrm{trivec}(\mathbf{W})\|_{(s_0,p)}$, i.e. the proof of Theorem 4.1. In addition, to prove Theorem 4.1, the approximated distribution for $\|\mathrm{trivec}(\mathbf{W}^b)\|_{(s_0,p)}$ is also needed. By the definition of $T_{i,j,m}^b$ and $W_{i,j}^b$ in (3.3), and as $\eta_{m,k}^b, m = 1, 2, k = 1,\ldots,n$ are independent standard normal random variables, we have $T_{i,j,m}^b|\mathcal{X},\mathcal{Y} \sim \boldsymbol{N}(0,\widetilde{\theta}_{i,j,m})$ with $1 \leq i,j \leq d$. Set $\mathbf{T}_m^b|\mathcal{X},\mathcal{Y} = (T_{i,j,m}^b|\mathcal{X},\mathcal{Y})_{d \times d}^\top$, we then have $\mathrm{trivec}(\mathbf{T}_m^b|\mathcal{X},\mathcal{Y}) \sim \boldsymbol{N}(0,\widehat{\boldsymbol{\Sigma}}_m^{\mathbf{T}^b})$, where $\widehat{\boldsymbol{\Sigma}}_m^{\mathbf{T}^b} = (\widehat{\sigma}_{s,t,m}^{\mathbf{T}^b})_{1 \leq s,t \leq d(d-1)/2}$ and $\widehat{\sigma}_{s,t,m}^{\mathbf{T}^b} = \widehat{\theta}_{i,j,m}$ for $s = t$. Actually, it also can be seen as the sample estimator for $\boldsymbol{\Sigma}_m^{\widetilde{U}}/n_m$. Similarly, set the corresponding correlation estimator as $\widehat{\mathbf{R}}_m^{\mathbf{T}^b} = (\widehat{r}_{s,t,m}^{\mathbf{T}^b})_{1 \leq s,t \leq d(d-1)/2}$, with $\widehat{r}_{s,t,m}^{\mathbf{T}^b} = \widehat{\sigma}_{s,t,m}^{\mathbf{T}^b}/\sqrt{\widehat{\sigma}_{s,s,m}^{\mathbf{T}^b}\widehat{\sigma}_{t,t,m}^{\mathbf{T}^b}}$. We then provide the following lemma to bound the estimation errors.

Lemma A.5. *With Assumption (B)-(D) hold, and for $\log(d) = o\{n^{1/2}\}$, we have*

$$(A.5)$$

$$\begin{aligned} &\max_{\substack{1 \leq s,t \leq d(d-1)/2 \\ m=1,2}} \left(|n_m\widehat{\sigma}_{s,t,m}^{\mathbf{T}^b} - \sigma_{s,t,m}^{\widetilde{U}}|, |\widehat{r}_{s,t,m}^{\mathbf{T}^b} - r_{s,t,m}^{\widetilde{U}}|\right) \\ &= o_p\left\{\big(\log(d)\big)^{-1/4}\right\}. \end{aligned}$$

The proof of Lemma A.5 is in the supplementary materials.

APPENDIX B. PROOF OF THEOREMS

By the results in Lemma A.4, Lemma A.5, and Theorem 4.1, actually the proof of Theorem 4.2 is similar to the proof of Theorem 3.6 in [28], the proof of Theorem 4.3 is similar to the proof of Theorem 3.3 and Theorem 3.7 in [28]. Hence, we omit these proofs and only show the detailed proof of Theorem 4.1.

B.1 Proof of Theorem 4.1

Proof. We provide the detail proof of (4.1) in two steps.

Step (i). In this step, we establish the approximated distribution of $\mathrm{trivec}(\mathbf{W})$. To this end, we introduce another intermediate variable. Define $\mathbf{H} \in \mathbb{R}^{d \times d}$ with

$$H_{i,j} = (\overline{\widetilde{U}}_{i,j,1} - \overline{\widetilde{U}}_{i,j,2})/\sqrt{\theta_{i,j,1} + \theta_{i,j,2}}, \quad 1 \leq i,j \leq d,$$

where $\overline{\widetilde{U}}_{i,j,m} = 1/n_m \sum_{k=1}^{n_m} \widetilde{U}_{k,i,j,m}$, $\theta_{i,j,m} = \sigma_{s,s,m}^{\widetilde{U}}/n_m$, $m = 1, 2$.

The following lemma establish that trivec($\mathbf{H}$) is a good approximation of trivec($\mathbf{W}$) under (s_0, p)-norm.

Lemma B.1. *We assume that Assumption **(A)-(D)** hold. Under $\mathbf{H}_0$ of (1.1), we have that there is a constant $C > 0$ such that*

$$\mathbb{P}\left(\|\text{trivec}(\mathbf{W}) - \text{trivec}(\mathbf{H})\|_{(s_0,p)} > \epsilon\right) = o(1),$$

as $n \to \infty$, where $\epsilon = o\left\{s_0\big(\log(d)\big)^{-1/2}\right\}$.

The proof for this lemma is in the supplementary materials. By the definition of $\widetilde{U}_{i,j,m}$ and by Lemma A.2, the Assumptions **(M1)-(M3)** are hold. As trivec($\mathbf{H}$) is a sum of independent random vectors with mean zero and covariance matrix $\mathbf{R}_{12}^{\widetilde{U}}$, where $\mathbf{R}_{12}^{\widetilde{U}} = (D_{12}^{\widetilde{U}})^{-1/2}\mathbf{\Sigma}_{12}^{\widetilde{U}}(D_{12}^{\widetilde{U}})^{-1/2}$ with $\mathbf{\Sigma}_{12}^{\widetilde{U}} = \mathbf{\Sigma}_1^{\widetilde{U}}/n_1 + \mathbf{\Sigma}_2^{\widetilde{U}}/n_2$ and $D_{12}^{\widetilde{U}} = \text{Diag}(\mathbf{\Sigma}_{12}^{\widetilde{U}})$. we use a Gaussian random vector with the same covariance matrices as its approximation. Let $\boldsymbol{G} \in \mathbb{R}^{d(d-1)/2}$ be a Gaussian random vector with mean zero and covariance matrix $\mathbf{R}_{12}^{\widetilde{U}}$. By Lemma A.1, we have

$$\sup_{z-\epsilon>0} \left|\mathbb{P}(\|\text{trivec}(\mathbf{H})\|_{(s_0,p)} > z - \epsilon) - \mathbb{P}(\|\boldsymbol{G}\|_{(s_0,p)} > z - \epsilon)\right| \leq Cn^{-\zeta_0}.$$

and

$$\sup_{z+\epsilon>0} \left|\mathbb{P}(\|\text{trivec}(\mathbf{H})\|_{(s_0,p)} > z + \epsilon) - \mathbb{P}(\|\boldsymbol{G}\|_{(s_0,p)} > z + \epsilon)\right| \leq Cn^{-\zeta_0}.$$

Further, by the triangle inequality, we have

(B.1)
$$\begin{aligned} \mathbb{P}(\|\text{trivec}(\mathbf{W})\|_{(s_0,p)} > z) &\leq \\ \mathbb{P}(\|\text{trivec}(\mathbf{H})\|_{(s_0,p)} > z - \epsilon) & \\ + \mathbb{P}(\|\text{trivec}(\mathbf{W}) - \text{trivec}(\mathbf{H})\|_{(s_0,p)} > \epsilon) \end{aligned}$$

(B.2)
$$\begin{aligned} \mathbb{P}(\|\text{trivec}(\mathbf{W})\|_{(s_0,p)} > z) &\geq \\ \mathbb{P}(\|\text{trivec}(\mathbf{H})\|_{(s_0,p)} > z + \epsilon) & \\ - \mathbb{P}(\|\text{trivec}(\mathbf{W}) - \text{trivec}(\mathbf{H})\|_{(s_0,p)} > \epsilon). \end{aligned}$$

Thus, by the triangle inequality and Lemma B.1, and combining Equation (B.1) and (B.2), we have

(B.3)
$$\begin{aligned} &\mathbb{P}\left(\|\boldsymbol{G}\|_{(s_0,p)} > z + \epsilon\right) - o(1) \\ \leq &\mathbb{P}\left(\|\text{trivec}(\mathbf{W})\|_{(s_0,p)} > z\right) \\ \leq &\mathbb{P}\left(\|\boldsymbol{G}\|_{(s_0,p)} > z - \epsilon\right) + o(1). \end{aligned}$$

Lemma B.2. *(Lemma B.2 in [28]) Assumptions **(A)** hold. For any $z > 0$ and $\varepsilon = O\big(s_0 \log^2(dn)n^{-1/2}\big)$, we have $\mathbb{P}\big(z - \varepsilon < \|\boldsymbol{G}\|_{(s_0,p)} \leq z\big) = o(1)$ as $n \to \infty$.*

By Lemma B.2 and combining (B.3), as $n \to \infty$, we have

(B.4)
$$\begin{aligned} \sup_{z>0} \Big|&\mathbb{P}\big(\|\text{trivec}(\mathbf{W})\|_{(s_0,p)} > z\big) \\ &- \mathbb{P}\big(\|\boldsymbol{G}\|_{(s_0,p)} > z\big)\Big| = o(1). \end{aligned}$$

Step (ii). In this step, we aim to obtain the distribution of trivec($\mathbf{W}^b$) given $\mathcal{X}$ and $\mathcal{Y}$. By the definition of $T_{i,j,m}^b$ in (3.3), as $\eta_{m,k}^b, m = 1, 2, k = 1, \ldots, n$ are independent standard normal random variables, we have $T_{i,j,m}^b|\mathcal{X},\mathcal{Y} \sim N(0, \widetilde{\theta}_{i,j,m})$ with $1 \leq i, j \leq d$, and trivec($\mathbf{T}_m^b|\mathcal{X},\mathcal{Y}) \sim N(\mathbf{0}, \widehat{\mathbf{\Sigma}}_m^{\boldsymbol{T}^b})$ with $\mathbf{T}_m^b|\mathcal{X},\mathcal{Y} = (T_{i,j,m}^b|\mathcal{X},\mathcal{Y})_{d \times d}^\top$. By setting $\widehat{\mathbf{\Sigma}}_{12}^{\boldsymbol{T}^b} = \widehat{\mathbf{\Sigma}}_1^{\boldsymbol{T}^b} + \widehat{\mathbf{\Sigma}}_2^{\boldsymbol{T}^b}$ and $\widehat{\mathbf{D}}_{12}^{\boldsymbol{T}^b} = \text{Diag}(\widehat{\mathbf{\Sigma}}_{12}^{\boldsymbol{T}^b})$, and conditional on $\mathcal{X}$ and $\mathcal{Y}$, we have

$$\text{trivec}(\mathbf{W}^b) \sim N(\mathbf{0}, \widehat{\mathbf{R}}_{12}^{\boldsymbol{T}^b}),$$

where $\widehat{\mathbf{R}}_{12}^{\boldsymbol{T}^b} := (\widehat{\mathbf{D}}_{12}^{\boldsymbol{T}^b})^{-1/2}\widehat{\mathbf{\Sigma}}_{12}^{\boldsymbol{T}^b}(\widehat{\mathbf{D}}_{12}^{\boldsymbol{T}^b})^{-1/2}$.

Recall that in last step, we have $\boldsymbol{G} \sim N(\mathbf{0}, \boldsymbol{R}_{12}^{\widetilde{U}})$. By Lemma A.5 and similar argument in Lemma B.3 of [28], we get the following lemma to establish the upper bound for the approximation error between $\|\boldsymbol{T}^{*b}\|_{(s_0,p)}$ and $\|\boldsymbol{G}\|_{(s_0,p)}$.

Lemma B.3. *Under Assumptions **(A)-(D)**, with probability at least $1 - Cn^{-1}$, we have*

$$\begin{aligned} \sup_{z>0} \Big|&\mathbb{P}\big(\|\boldsymbol{G}\|_{(s_0,p)} > z\big) \\ &- \mathbb{P}\big(\|\text{trivec}(\boldsymbol{W}^b)\|_{(s_0,p)} > z|\mathcal{X},\mathcal{Y}\big)\Big| = o_p(1). \end{aligned}$$

By the triangle inequality, we have

$$\begin{aligned} &\Big|\mathbb{P}(N_{(s_0,p)} > z) - \mathbb{P}(N_{(s_0,p)}^b > z|\mathcal{X},\mathcal{Y})\Big| \leq \\ &\Big|\mathbb{P}(\|\text{trivec}(\boldsymbol{W})\|_{(s_0,p)} > z) - \mathbb{P}(\|\boldsymbol{G}\|_{(s_0,p)} > z)\Big| + \\ &\Big|\mathbb{P}(\|\boldsymbol{G}\|_{(s_0,p)} > z) - \mathbb{P}(\|\text{trivec}(\boldsymbol{W}^b)\|_{(s_0,p)} > z|\mathcal{X},\mathcal{Y})\Big|. \end{aligned}$$

Hence, combining Equation (B.4) and the result in Lemma B.3, we have

$$\sup_{z \in (0,\infty)} \Big|\mathbb{P}\big(N_{(s_0,p)} > z\big) - \mathbb{P}\big(N_{(s_0,p)}^b > z|\mathcal{X},\mathcal{Y}\big)\Big| = o_p(1),$$

which finishes the proof of Theorem 4.1. □

REFERENCES

[1] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* **106** 594–607. MR2847973

[2] CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association* **108** 265–277. MR3174618

[3] Cai, T. T., Liu, W. and Zhou, H. H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Annals of Statistics* **44** 455–488. MR3476606

[4] Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability* **45** 2309-2352. MR3693963

[5] Clements, W. M., Wang, J., Sarnaik, A., Kim, O. J., Macdonald, J., Fenogliopreiser, C., Groden, J. and Lowy, A. M. (2002). beta-Catenin mutation is a frequent cause of Wnt pathway activation in gastric cancer. *Cancer Research* **62** 3503-6.

[6] Corda, G. and Sala, A. (2017). Non-canonical WNT/PCP signalling in cancer: Fzd6 takes centre stage:. *Oncogenesis* **6** e364.

[7] Danaher, P., Wang, P. and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society* **76** 373–397. MR3164871

[8] Fan, Y. and Lv, J. (2016). Innovated scalable efficient estimation in ultra-large Gaussian graphical models. *Annals of Statistics* **44** 2098-2126. MR3546445

[9] Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical LASSO. *Biostatistics* **9** 432–441.

[10] Howe, L. R. and Brown, A. M. (2004). Wnt signaling and breast cancer. *Cancer Biology & Therapy* **3** 36-41.

[11] Ideker, T. and Krogan, N. J. (2012). Differential network biology. *Molecular systems biology* **8** 565.

[12] Li, J. and Chen, S. (2012). Two sample tests for high-dimensional covariance matrices. *Annals of Statistics* **40** 908-940. MR2985938

[13] Li, K. C., Palotie, A., Yuan, S., Bronnikov, D., Chen, D., Wei, X., Choi, O. W., Saarela, J. and Peltonen, L. (2007). Finding disease candidate genes by liquid association. *Genome Biology* **8** R205.

[14] Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Annals of Statistics* **41** 2948-2978. MR3161453

[15] Mazieres, J., He, B., You, L., Xu, Z. and Jablons, D. M. (2005). Wnt signaling in lung cancer. *Cancer Letters* **222** 1–10.

[16] Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the LASSO. *Annals of Statistics* **34** 1436–1462. MR2278363

[17] Rapp, J., Jaromi, L., Kvell, K., Miskei, G. and Pongracz, J. E. (2017). WNT signaling – lung cancer is no exception. *Respiratory Research* **18** 167.

[18] Ren, Z., Sun, T., Zhang, C. H. and Zhou, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Annals of Statistics* **43** 991–1026. MR3346695

[19] Roman, V. (2017). *High-dimensional probability an introduction with applications in data science.* https://www.math.uci.edu/~rvershyn/. MR3837109

[20] Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics and Data Analysis* **51** 6535-6542. MR2408613

[21] Srivastava, M. S. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis* **101** 1319-1329. MR2609494

[22] Tian, D., Gu, Q. and Jian, M. (2016). Identifying gene regulatory network rewiring using latent differential graphical models. *Nucleic Acids Research* **44** e140-e140.

[23] Xia, Y., Cai, T. and Cai, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika* **102**. MR3371002

[24] Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research* **11** 2261–2286. MR2719856

[25] Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824

[26] Zhao, S. D., Cai, T. T. and Li, H. (2014). Direct estimation of differential networks. *Biometrika* **101** 253–268. MR3215346

[27] Zhao, S. D., Cai, T. T. and Li, H. (2015). Direct estimation of differential networks. *Biometrika* **2** 253-268. MR3215346

[28] Zhou, C., Zhang, X., Zhou, W. and Liu, H. (2018). A unified framework for testing high dimensional parameters: a data-adaptive approach. *arXiv preprint arXiv:1808.02648*.

Mingjuan Zhang
School of Statistics and Mathematics
Shanghai Lixin University of Accounting and Finance
Shanghai 201209
China
E-mail address: 12110690006@fudan.edu.cn

Yong He
School of Statistics
Shandong University of Finance and Economics
250014 Jinan
China
E-mail address: heyong@sdufe.edu.cn

Cheng Zhou
School of Management
Fudan University
Shanghai 200433
China
E-mail address: chengzhmike@gmail.com

Xinsheng Zhang
School of Management
Fudan University
Shanghai 200433
China
E-mail address: xszhang@fudan.edu.cn