JUSTC

Article

http://justc.ustc.edu.cn

Received: July 28, 2022; Accepted: October 23, 2022

# Gaussian graphical model estimation with measurement error

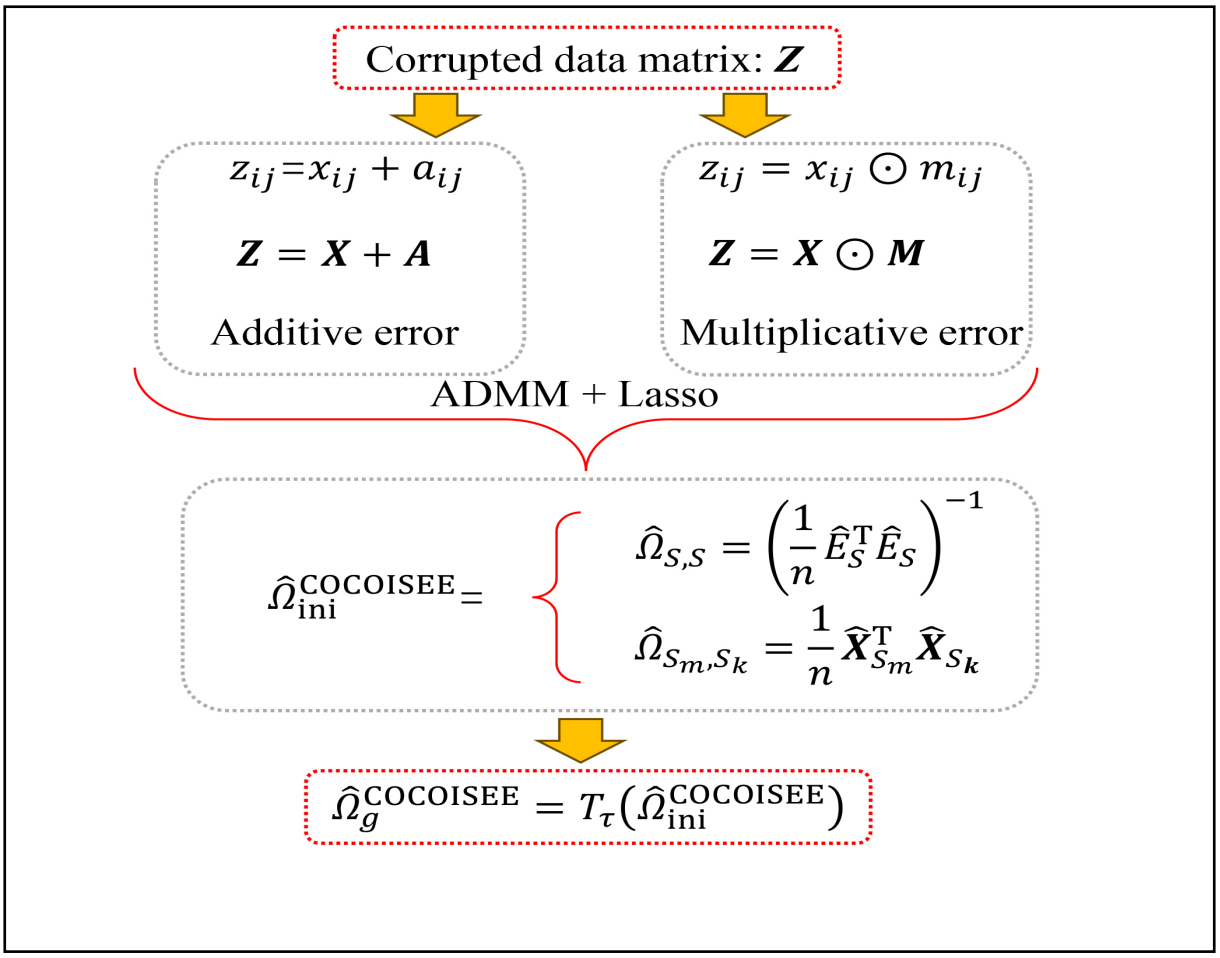
Xianglu Wang ✉

*Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China*

✉Correspondence: Xianglu Wang, E-mail: wz124517@mail.ustc.edu.cn



## Graphical abstract




*The process of recovering the precision matrix in the presence of additive and multiplicative measurement errors.*

## Public summary

- We propose a new methodology COCOISEE to achieve scalable and interpretable estimation for Gaussian graphical model under both additive and multiplicative measurement errors.
- The method is stepwise convex, computationally stable, efficient and scalable.
- Both theoretical and simulation results verify the feasibility of our method.

**Citation:** Wang X L. Gaussian graphical model estimation with measurement error. *JUSTC*, **2023**, 53(11): 1105. DOI: 10.52396/JUSTC-2022-0108

JUSTC

Article

http://justc.ustc.edu.cn

Received: July 28, 2022; Accepted: October 23, 2022

# Gaussian graphical model estimation with measurement error

Xianglu Wang ✉

*Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China*

✉Correspondence: Xianglu Wang, E-mail: wz124517@mail.ustc.edu.cn



Cite This: *JUSTC*, **2023**, 53(11): 1105 (11pp)

Read Online

**Abstract:** It is well known that regression methods designed for clean data will lead to erroneous results if directly applied to corrupted data. Despite the recent methodological and algorithmic advances in Gaussian graphical model estimation, how to achieve efficient and scalable estimation under contaminated covariates is unclear. Here a new methodology called convex conditioned innovative scalable efficient estimation (COCOISEE) for Gaussian graphical models under both additive and multiplicative measurement errors is developed. It combines the strengths of the innovative scalable efficient estimation in the Gaussian graphical model and the nearest positive semidefinite matrix projection, thus enjoying stepwise convexity and scalability. Comprehensive theoretical guarantees are provided and the effectiveness of the proposed methodology is demonstrated through numerical studies.

**Keywords:** Gaussian graphical model; measurement error; scalability; stepwise convexity

**CLC number:** O212.4 **Document code:** A

**2020 Mathematics Subject Classification:** 62J07

## 1 Introduction

The estimation precision (inverse covariance) matrix has become a basic problem in modern multivariate statistical analysis and has a very wide range of practical applications. For example, in genome-wide association studies, genomic data are used to estimate the precision matrix and obtain the genetic correlation between traits from the nonzero terms. It is not only crucial for understanding gene regulatory pathways and gene functions but also contributes to a deeper understanding of the genetic basis of quantitative variation in complex traits[1]. Similarly, in social network analysis[2, 3], a precision matrix can be used to explore the correlation between users and is also widely used in economics and finance[4–6], health science[7], high-dimensional discriminant analysis[8], and complex data visualization[9].

The estimation of the precision matrix has attracted many scholars to study in recent years. In terms of the form of research methods, the currently proposed methods can be roughly divided into two classes: penalized likelihood methods and column-by-column estimation methods. The former class includes, for example, Ref. [10], which has laid down a general framework for the penalized likelihood method, and Refs. [11–15], which used different penalty algorithms to select zero elements in the precision matrix, and these methods all impose a penalty term on the negative log-likelihood function to estimate the precision matrix. The theoretical properties of these methods have been thoroughly studied by Refs. [16, 17]. The latter class includes, for instance, Refs. [8, 14, 18–21]. Such methods convert the problem of precision matrix estimation into a nodewise or pairwise regression, or optimization problem and then apply the technique of high-dimensional regularization using the Lasso or Dantzig selector type methods. However, most of these existing methods are designed for clean data, while corrupted data are often encountered in various fields. Naively applying the aforementioned methods to analyze the corrupted data can lead to inconsistent and unstable estimates, thus leading to misleading conclusions. Therefore, it is urgent to develop an efficient method for large precision matrix estimation under measurement errors.

Various statistical methods have been proposed to mitigate the effects of measurement errors. Specifically, there are a series of works to address measurement errors in univariate response linear regression models, which can be traced back to Ref. [22] and its extensions include Refs. [23, 24]. In recent years, great progress has also been made in high-dimensional error-in-variables regression. For instance, Ref. [25] developed a $l_1$-regularized likelihood approach to handle missing data by solving a negative log-likelihood optimization problem via the EM algorithm. Similarly, Ref. [26] developed a Lasso-type estimator by replacing the corrupted Gram matrix with unbiased estimates for corrupted data. Furthermore, Ref. [27] proposed a Dantzig selector-type estimator based on the compensated matrix uncertainty method. However, after adjusting for corrupted data, negative likelihood functions are usually not convex and may depend on some crucial hidden parameters. To address this difficulty, Ref. [28] proposed the convex conditioned Lasso (CoCoLasso) method by replacing the unbiased covariance matrix estimate with the nearest positive semidefinite matrix, thus enjoying stepwise convexity in high-dimensional measurement error regression.

In this work, motivated by the study of the above scholars,



DOI: 10.52396/JUSTC-2022-0108

*JUSTC*, **2023**, 53(11): 1105

we develop a new approach called convex conditioned innovative scalable efficient estimation (COCOISEE) to address Gaussian graphical models under measurement errors by combining the strengths of the recently developed innovative scalable efficient estimation[21] and the nearest positive semi-definite matrix projection[28], thus enjoying stepwise convexity and scalability.

The rest of the paper is organized as follows. Section 2 presents the model setting and the new methodology. Theoretical properties including consistency in estimation are established in Section 3. We provide simulation examples in Section 4. Section 5 discusses possible extensions of our method and possible future work. All technical details are relegated to Appendix.

**Notation.** We briefly summarize some symbols to be used throughout the paper. We use $\boldsymbol{x}$, $X$, and $\boldsymbol{X}$ to denote vectors, elements, and matrices, respectively. For any matrix $\boldsymbol{X}=(X_{ij})$, let $\|\boldsymbol{X}\|_\infty=\max_i\sum_j|X_{ij}|$ denote the matrix $l_\infty$ norm, whereas $\|\boldsymbol{X}\|_{\max}=\max_{i,j}|X_{ij}|$ denotes the elementwise maximum norm. Similarly, for any vector $\boldsymbol{x}=(X_1,...,X_p)^{\mathrm{T}}$, let $\|\boldsymbol{x}\|_q=\left(\sum_i|X_i|^q\right)^{1/q}$ be the $l_q$ norm of vector $\boldsymbol{x}$, and $l_\infty$ vector norm is $\|\boldsymbol{x}\|_\infty=\max_i|X_i|$. For any subsets $A,B\subset\{1,...,p\}$, denote by $\boldsymbol{x}_A$ a subvector of $\boldsymbol{x}$ formed by its components with indices in $A$, denote by $\boldsymbol{x}_{-A}$ a subvector of $\boldsymbol{x}$ formed by its components without indices in $A$, and $\boldsymbol{\Omega}_{A,B}=(w_{ij})_{i\in A,j\in B}$ a submatrix of $\boldsymbol{\Omega}$ with rows in $A$ and columns in $B$. Let $(S_l)_{1\leqslant l\leqslant L}$ be a partition of the index set $\{1,...,p\}$. $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the smallest and largest eigenvalues of a given symmetric matrix, respectively. We use $\Delta$ to denote a fixed sequence going to zero.

## 2 Gaussian graphical model under measurement errors

### 2.1 Model setting

Consider an undirected Gaussian graphical model $G=(V,E)$ for a $p$-dimensional Gaussian random vector

$$\boldsymbol{x}=(X_1,...,X_p)^{\mathrm{T}}\sim N(\mu,\Sigma),$$

where $G$ is an undirected graph related to vector $\boldsymbol{x}$, $V=\{X_1,...,X_p\}$ is a node set, $E=(e_{ij})$ is an edge matrix with each entry $e_{ij}$ representing the edge between $X_i$ and $X_j$, $\mu=(\mu_i)_{1\leqslant i\leqslant p}$ is a mean vector and $\Sigma=(\sigma_{ij})$ is a covariance matrix. Assume that $\{\boldsymbol{x}_i\}_{1\leqslant i\leqslant n}$ is an independently identically distributed (i.i.d.) sample from the Gaussian distribution. Without loss of generality, suppose that the mean vector $\mu=0$ throughout the paper. Denote by $\boldsymbol{\Omega}=(\omega_{ij})$ the precision matrix, that is, the inverse $\Sigma^{-1}$ of the covariance matrix $\Sigma$. It is well known that in Gaussian graphical model theory, an edge $(i,j)$ exists (or $e_{ij}\neq0$) between $X_i$ and $X_j$ if and only if the corresponding entry $\omega_{ij}$ of the precision matrix is nonzero. The above characterization of the precision matrix shows that the problem of estimating Gaussian graphical model $G$ can be equivalent to estimating the support:

$$\mathrm{supp}(\boldsymbol{\Omega})=\{(i,j):\omega_{ij}\neq0\}.$$

To motivate our new method, we consider a linear transformation of the precision matrix similar to Ref. [21]. Specifically, based on the transformation of $\boldsymbol{x}$, we have the following structure:

$$\tilde{\boldsymbol{x}}=\boldsymbol{\Omega}\boldsymbol{x},\quad \tilde{\boldsymbol{x}}\sim N(0,\boldsymbol{\Omega}). \tag{1}$$

Therefore, from the distribution of $\tilde{\boldsymbol{x}}$, we can see that if estimator $\tilde{\boldsymbol{x}}$ is obtained, the precision matrix $\boldsymbol{\Omega}$ can be achieved by estimating the covariance matrix of $\tilde{\boldsymbol{x}}$.

By the definition of $\tilde{\boldsymbol{x}}$, we can write the subvector $\tilde{\boldsymbol{x}}_S$ in following form:

$$\tilde{\boldsymbol{x}}_S=\boldsymbol{\Omega}_{S,S}(\boldsymbol{x}_S+\boldsymbol{\Omega}_{S,S}^{-1}\boldsymbol{\Omega}_{S,-S}\boldsymbol{x}_{-S})=\boldsymbol{\Omega}_{S,S}\boldsymbol{\eta}_S. \tag{2}$$

It is well known that in the Gaussian graphical model, the condition distribution about the vector $\boldsymbol{x}_S$ can be written as

$$\boldsymbol{x}_S|\boldsymbol{x}_{-S}\sim N(-\boldsymbol{\Omega}_{S,S}^{-1}\boldsymbol{\Omega}_{S,-S}\boldsymbol{x}_{-S},\boldsymbol{\Omega}_{S,S}^{-1}). \tag{3}$$

Formula (3) can be expressed as the following multivariate linear regression model:

$$\boldsymbol{x}_S=\Theta_S^{\mathrm{T}}\boldsymbol{x}_{-S}+\boldsymbol{\eta}_S, \tag{4}$$

where $\Theta_S=-\boldsymbol{\Omega}_{-S,S}\boldsymbol{\Omega}_{S,S}^{-1}$ is a $n\times(p-|S|)$ dimensional regression coefficients matrix, and $\boldsymbol{\eta}_S=\boldsymbol{x}_S+\boldsymbol{\Omega}_{S,S}^{-1}\boldsymbol{\Omega}_{S,-S}\boldsymbol{x}_{-S}$ is the matrix of model errors which has a multivariate Gaussian distribution $N(0,\boldsymbol{\Omega}_{S,S}^{-1})$ and independent of $\boldsymbol{x}_{-S}$. Thus, the estimates of the two terms on the right side of Eq. (2) are correlated and can be efficiently achieved by linear regression techniques, i.e., Lasso[29].

The representation of subvector $\boldsymbol{x}_S$ shows that the unknown subvector $\tilde{\boldsymbol{x}}_S$ can be estimated by the regression technique. To see this, let $\hat{\boldsymbol{\eta}}_S$ be the residual vector obtained by fitting model (4) using Lasso. Then the unknown matrix $\boldsymbol{\Omega}_{S,S}$ can be estimated as the inverse of the sample covariance matrix of the model residual vector $\hat{\boldsymbol{\eta}}_S$. Then we can estimate the subvector $\tilde{\boldsymbol{x}}_S$ in Eq. (2) as $\hat{\boldsymbol{x}}_S=\hat{\boldsymbol{\Omega}}_{S,S}\hat{\boldsymbol{\eta}}_S$.

The ISEE[21] repeated the above procedure for each $S_l$ with $1\leqslant l\leqslant p$ to obtain the estimated subvectors $\hat{\boldsymbol{x}}_{S_l}$'s, and then stacked all these subvectors together to form an estimate $\hat{\boldsymbol{x}}$ of the oracle innovated vector $\tilde{\boldsymbol{x}}$. Thus, the problem of estimating the precision matrix based on the original vector $\boldsymbol{x}$ reduces to that of estimating the covariance matrix based on the estimated transformed vector $\tilde{\boldsymbol{x}}$.

However, in many practical applications, the data we collect may contain unobvious measurement errors, resulting in inaccurate Gaussian graphical model estimation. In this paper, we consider two kinds of measurement errors associated with matrix $\boldsymbol{X}$, as shown below:

- Additive error. We assume that the observed elements of the data matrix $\boldsymbol{Z}$ are affected by additive measurement errors $z_{ij}=x_{ij}+a_{ij}$, and write it as a matrix form $\boldsymbol{Z}=\boldsymbol{X}+\boldsymbol{A}$, where $\boldsymbol{A}=(a_{ij})_{p\times p}$ is the measurement error matrix. We


DOI: 10.52396/JUSTC-2022-0108
JUSTC, 2023, 53(11): 1105

assume that the rows of $\boldsymbol{A}$ are i.i.d. with mean $\mathbf{0}$ and finite covariance matrix $\Sigma^A$.

- Multiplicative error. We assume that the observed elements of the data matrix $\boldsymbol{Z}$ are affected by multiplicative measurement errors $z_{ij} = x_{ij} \odot m_{ij}$, and write it as a matrix form $\boldsymbol{Z} = \boldsymbol{X} \odot \boldsymbol{M}$, where $\odot$ denotes the Hadamard product and $\boldsymbol{M} = (m_{ij})_{p\times p}$ is the measurement error matrix. We also assume that the rows of $\boldsymbol{M}$ are i.i.d. with mean $\boldsymbol{\mu}^M$ and finite covariance matrix $\Sigma^M$. Particularly, missing data can be regarded as a special case of the multiplicative measurement error, where $m_{ij} = I(x_{ij})$, and $I(\cdot)$ represents the indicator function.

## 2.2 Scalable estimation by COCOISEE

The data matrix is denote by $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)^{\mathrm{T}} \in \mathrm{R}^{n\times p}$. Then, the transformed data matrix is defined as $\tilde{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{\Omega}$. Thus, Eq. (4) can be rewritten as

$$\boldsymbol{X}_S = \boldsymbol{X}_{-S}\Theta_S + E_S, \tag{5}$$

where $E_S$ is an error matrix with rows being i.i.d. copies from $\boldsymbol{\eta}_S^{\mathrm{T}}$. Then, $\tilde{\boldsymbol{X}}_S$ can be written as

$$\begin{aligned}\tilde{\boldsymbol{X}}_S =&(\boldsymbol{X}\boldsymbol{\Omega})_S = \boldsymbol{X}_S\boldsymbol{\Omega}_{S,S} + \boldsymbol{X}_{-S}\boldsymbol{\Omega}_{-S,S} = \\ &(\boldsymbol{X}_S + \boldsymbol{X}_{-S}\boldsymbol{\Omega}_{-S,S}\boldsymbol{\Omega}_{S,S}^{-1})\boldsymbol{\Omega}_{S,S} = E_S\boldsymbol{\Omega}_{S,S}.\end{aligned} \tag{6}$$

The representation in Eq. (6) provides the basis for the estimation of matrix $\tilde{\boldsymbol{X}}$.

To fit the Gaussian linear regression model (5), we suggest univariate regression methods. Specifically, for each node $i$ in the index set $S$, we consider the univariate linear regression model for $\boldsymbol{X}_i$:

$$\boldsymbol{X}_i = \boldsymbol{X}_{-S}\theta_i + E_i. \tag{7}$$

In general, we can use some high-dimensional regression methods to fit Eq. (7), such as SCAD[30], Lasso[29], elastic net[31], adaptive Lasso[32], and Dantzig selector[33]. However, we observe the data matrix $\boldsymbol{Z}$ instead of the matrix $\boldsymbol{X}$ in linear models with error-in-variables. Thus, directly applying Lasso to this equation by minimizing

$$\frac{1}{2}\|\boldsymbol{Z}_i - \boldsymbol{Z}_{-S}\theta_i\|_2^2 + \lambda_i\|\theta_i\|_1 \Longleftrightarrow \frac{1}{2n}\theta_i^{\mathrm{T}}\Sigma_{-S,-S}\theta_i - \theta_i^{\mathrm{T}}\Sigma_{-S,i} + \lambda_i\|\theta_i\|_1$$

is often erroneous, where $\Sigma_{-S,-S} = \frac{1}{n}\boldsymbol{Z}_{-S}^{\mathrm{T}}\boldsymbol{Z}_{-S}$ and $\Sigma_{-S,i} = \frac{1}{n}\boldsymbol{Z}_{-S}^{\mathrm{T}}\boldsymbol{Z}_i$. Based on the corrupted data matrix $\boldsymbol{Z}$, Ref. [28] suggested constructing unbiased surrogates $\hat{\Sigma}_{-S,-S}$ and $\tilde{\Sigma}_{-S,i}$ for $\Sigma_{-S,-S}$ and $\Sigma_{-S,i}$ that cannot be observed to mitigate the effects of measurement errors. For the additive error setting, the unbiased surrogates are defined as

$$\begin{aligned}\hat{\Sigma}_{-S,-S}^{\text{add}} &= \frac{1}{n}\boldsymbol{Z}_{-S}^{\mathrm{T}}\boldsymbol{Z}_{-S} - \Sigma_{-S,-S}^A, \\ \tilde{\Sigma}_{-S,i}^{\text{add}} &= \frac{1}{n}\boldsymbol{Z}_{-S}^{\mathrm{T}}\boldsymbol{Z}_i.\end{aligned} \tag{8}$$

And for the multiplicative error setting, the unbiased surrogates are defined as

$$\begin{aligned}\hat{\Sigma}_{-S,-S}^{\text{mult}} &= \frac{1}{n}\boldsymbol{Z}_{-S}^{\mathrm{T}}\boldsymbol{Z}_{-S} \oslash (\Sigma_{-S,-S}^M + \mu_{-S}^M(\mu_{-S}^M)^{\mathrm{T}}), \\ \tilde{\Sigma}_{-S,i}^{\text{mult}} &= \frac{1}{n}\boldsymbol{Z}_{-S}^{\mathrm{T}}\boldsymbol{Z}_i \oslash \mu_{-S}^M,\end{aligned} \tag{9}$$

where $\oslash$ represents the element division operator of vectors and matrices. Since $\Sigma^A$ or $(\Sigma^M, \mu^M)$ can be obtained in practice through repeated measurements or field experience, we assume that they are known for the model.

However, the estimator $\hat{\Sigma}$ is usually not positive semidefinite in high-dimensional models, resulting in the related optimization problems that may no longer be convex. To overcome this problem, we draw on the idea of Ref. [28] to obtain the sparse regression coefficients $\theta_i$ by minimizing

$$\hat{\theta}_i := \underset{\theta_i\in R^{p-|S|}}{\operatorname{argmin}}\left(\frac{1}{2}\theta_i^{\mathrm{T}}\tilde{\Sigma}_{-S,-S}\theta_i - \theta_i^{\mathrm{T}}\tilde{\Sigma}_{-S,i} + \lambda_i\|\theta_i\|_1\right), \tag{10}$$

where $\tilde{\Sigma}_{-S,-S}$ is the principal submatrix of $\tilde{\Sigma}$, and $\tilde{\Sigma} = (\hat{\Sigma})_+ = \underset{\Sigma\succcurlyeq 0}{\operatorname{argmin}}\|\Sigma - \hat{\Sigma}\|_{\max}$ is the nearest positive semidefinite matrix. From the definition of $\tilde{\Sigma}$, we can solve it efficiently by an alternating direction method of multipliers (ADMM) and have

$$\left\|\tilde{\Sigma} - \Sigma\right\|_{\max} \leqslant \left\|\tilde{\Sigma} - \hat{\Sigma}\right\|_{\max} + \left\|\hat{\Sigma} - \Sigma\right\|_{\max} \leqslant 2\left\|\hat{\Sigma} - \Sigma\right\|_{\max}. \tag{11}$$

Eq. (11) is obtained by the triangle inequality and the definition of $\tilde{\Sigma}$ ensures $\tilde{\Sigma}$ approximates $\Sigma$ as well as the initial surrogate $\hat{\Sigma}$.

Based on the regression step, for each node $i$ in the index set $S$ we cannot directly replace $\boldsymbol{Z}$ with $\boldsymbol{X}$ in Eq. (5) to calculate $\hat{E}_S$. Instead, we can use $\hat{E}_S$ as an intermediate variable to calculate $\hat{\boldsymbol{\Omega}}_{S,S}$. Then the estimator $\hat{\boldsymbol{\Omega}}_{S,S}$ can be written as

$$\hat{\boldsymbol{\Omega}}_{S,S} = \left(\frac{1}{n}\hat{E}_S^{\mathrm{T}}\hat{E}_S\right)^{-1} = (\hat{\Sigma}_{S,S} - 2\hat{\Sigma}_{S,-S}\hat{\boldsymbol{\Theta}}_S + \hat{\boldsymbol{\Theta}}_S^{\mathrm{T}}\hat{\Sigma}_{-S,-S}\hat{\boldsymbol{\Theta}}_S)^{-1}, \tag{12}$$

where $\hat{\boldsymbol{\Theta}}_S = (\hat{\theta}_i)_{i\in S}$. These observations suggest a simple plug-in estimator $\hat{E}_S\hat{\boldsymbol{\Omega}}_{S,S}$ for the unobservable submatrix $\hat{\boldsymbol{X}}_S$ in Eq. (6). Since the error matrix $\hat{E}_S$ is unknown, we cannot directly calculate $\hat{\boldsymbol{X}}_S$. Similarly, for each $1 \leqslant k \neq m \leqslant L$, we can use $\hat{\boldsymbol{X}}_{S_k}$ and $\hat{\boldsymbol{X}}_{S_m}$ as intermediate variables to calculate $\hat{\boldsymbol{\Omega}}_{S_m,S_k}$. Then the estimator $\hat{\boldsymbol{\Omega}}_{S_m,S_k}$ can be written as

$$\begin{aligned}\hat{\boldsymbol{\Omega}}_{S_m,S_k} =&\frac{1}{n}\hat{\boldsymbol{X}}_{S_m}^{\mathrm{T}}\hat{\boldsymbol{X}}_{S_k} = \\ &\frac{1}{n}(\hat{E}_{S_m}\hat{\boldsymbol{\Omega}}_{S_m,S_m})^{\mathrm{T}}(\hat{E}_{S_k}\hat{\boldsymbol{\Omega}}_{S_k,S_k}) = \\ &\hat{\boldsymbol{\Omega}}_{S_m,S_m}^{\mathrm{T}}\left(\frac{1}{n}\hat{E}_{S_m}^{\mathrm{T}}\hat{E}_{S_k}\right)\hat{\boldsymbol{\Omega}}_{S_k,S_k} = \\ &\hat{\boldsymbol{\Omega}}_{S_m,S_m}^{\mathrm{T}}(\hat{\Sigma}_{S_m,S_k} - \hat{\Sigma}_{S_m,-S_k}\hat{\boldsymbol{\Theta}}_{S_k} - \\ &\hat{\boldsymbol{\Theta}}_{S_m}^{\mathrm{T}}\hat{\Sigma}_{-S_m,S_k} + \hat{\boldsymbol{\Theta}}_{S_m}^{\mathrm{T}}\hat{\Sigma}_{-S_m,-S_k}\hat{\boldsymbol{\Theta}}_{S_k})\hat{\boldsymbol{\Omega}}_{S_k,S_k}.\end{aligned} \tag{13}$$

By doing so, the initial COCOISEE estimator is


DOI: 10.52396/JUSTC-2022-0108
JUSTC, 2023, 53(11): 1105

$$\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}} = \begin{cases} (\hat{\Sigma}_{S,S} - 2\hat{\Sigma}_{S,-S}\hat{\Theta}_S + \hat{\Theta}_S^T\hat{\Sigma}_{-S,-S}\hat{\Theta}_S)^{-1}, \\ \quad k = m; \\ \hat{\boldsymbol{\Omega}}_{S_m,S_m}^{\mathrm{T}}(\hat{\Sigma}_{S_m,S_k} - \hat{\Sigma}_{S_m,-S_k}\hat{\Theta}_{S_k} - \hat{\Theta}_{S_m}^{\mathrm{T}}\hat{\Sigma}_{-S_m,S_k} + \\ \hat{\Theta}_{S_m}^{\mathrm{T}}\hat{\Sigma}_{-S_m,-S_k}\hat{\Theta}_{S_k})\hat{\boldsymbol{\Omega}}_{S_k,S_k}, \\ \quad k \neq m. \end{cases} \tag{14}$$

Then for a given threshold $\tau \geqslant 0$, define

$$\hat{\boldsymbol{\Omega}}_g^{\text{COCOISEE}} = T_\tau(\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}), \tag{15}$$

where $T_\tau(\boldsymbol{X}) = (x_{ij}I_{|x_{ij}|\geqslant\tau})$ denotes the matrix $\boldsymbol{X}$ thresholded at $\tau$. Thus, the set of links of graphical structure is $\text{supp}(\hat{\boldsymbol{\Omega}}_g^{\text{COCOISEE}})$. The choice of the threshold $\tau$ in Eq. (15) is important for practical implementation. We use the cross-validation method for large precision matrix estimation. Specifically, we randomly split the sample of $n$ rows of the sample matrix $\boldsymbol{Z}$ into two subsamples of sizes $n_1$ and $n_2$, and repeat this $R$ times. Denote by $\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}(n_1,i)$ and $\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}(n_2,i)$ the precision matrices as defined in Eq. (14) based on these two subsamples, respectively, for the $i$th split. Thus, the threshold $\tau$ can be chosen to minimize

$$\tau := \frac{1}{R}\sum_{i=1}^{R}\left\|T_\tau(\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}(n_1,i)) - \hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}(n_2,i)\right\|^2. \tag{16}$$

The implementation of COCOISEE is summarized in Algorithm 1.

## 3 Theoretical properties

In this section, we will present several technical conditions and then analyze the theoretical properties of COCOISEE.

### 3.1 Technical conditions

**Condition 1.** (Closeness condition) Assume that the distributions of $\hat{\Sigma}_{-S,-S}$ and $\tilde{\Sigma}_{-S,j}$ are identified by a set of parameters $\theta$. Then there exist universal constants $C$ and $c$, and positive functions $\zeta$ and $\epsilon_0$ depending on $\theta$ and $\sigma^2$ such that for any $\varepsilon \leqslant \varepsilon_0$, $\hat{\Sigma}_{-S,-S}$ and $\tilde{\Sigma}_{-S,j}$ satisfy the following probability statements:

$$\begin{aligned} &\Pr(|(\hat{\Sigma}_{-S,-S})_{ij} - (\Sigma_{-S,-S})_{ij}| \geqslant \varepsilon) \leqslant C\exp(-cn\varepsilon^2\zeta^{-1}), \\ &\forall i,j \in (\{1,...,p\} - S). \end{aligned} \tag{17}$$

$$\begin{aligned} &\Pr(|(\tilde{\Sigma}_{-S,j})_i - (\Sigma_{-S,j})_i| \geqslant \varepsilon) \leqslant C\exp(-cn\varepsilon^2K^{-2}\zeta^{-1}), \\ &\forall i \in (\{1,...,p\} - S), j = \in \{1,...,p\}. \end{aligned} \tag{18}$$

**Condition 2.** (Restricted eigenvalue condition) The restricted eigenvalue of the Gram matrix $\Sigma = \frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ satisfies:

$$0 < \boldsymbol{\Omega} = \min_{\boldsymbol{x}\neq 0, \|\boldsymbol{x}_{-\kappa}\|_1 \leqslant 3\|\boldsymbol{x}_\kappa\|_1} \frac{\boldsymbol{x}^{\mathrm{T}}\Sigma\boldsymbol{x}}{\|\boldsymbol{x}\|_2^2}. \tag{19}$$

**Condition 3.** (Irrepresentable condition)

$$\|\Sigma_{-\kappa,\kappa}\Sigma_{\kappa,\kappa}^{-1}\|_\infty < 1, \tag{20}$$

where $\kappa = \{1,...,K\}$ is the true support set of the regression coefficient vector $\theta^*$.

**Condition 4.** The entries of measurement error matrices $A$ and $M$ are all independent and identically distributed sub-Gaussian random variables.

**Condition 5.** The random vectors $\boldsymbol{\eta}_i$ obey a zero-mean sub-Gaussian distribution with a finite variance parameter, i.e., $\sigma_i^2 \leqslant C$.

**Condition 6.** The Gaussian graph model we focus on satisfies the following conditions.

$$\begin{aligned} &\mathcal{G}(M,K) = \\ &\{\boldsymbol{\Omega} : \text{each row has at most } K \\ &\text{nonzero off-diagonal entries and} \\ &M^{-1} \leqslant \lambda_{\min}(\boldsymbol{\Omega}) \leqslant \lambda_{\max}(\boldsymbol{\Omega}) \leqslant M\}. \end{aligned} \tag{21}$$

Condition 1 ensures that the surrogates $\hat{\Sigma}_{-S,-S}$ (and hence $\tilde{\Sigma}_{-S,-S}$) and $\hat{\Sigma}_{-S,j}$ are close to $\Sigma_{-S,-S}$ and $\Sigma_{-S,j}$ respectively in terms of the elementwise maximum norm. Condition 2 is the restricted eigenvalue condition proposed in Ref. [34], and is widely used in Lasso related articles. It imposes a lower bound on eigenvalues of the Gram matrix $\Sigma$ to constrain the correlations between relatively small numbers of predictors in the design matrix $\boldsymbol{X}$. Condition 3 is the irrepresentable condition proposed in Refs. [35, 36], and was obtained to prove a model selection consistency result for Gaussian graphical model selection using the Lasso. Conditions 1–3 are necessary conditions for proving Lemma A.1 introduced in Appendix.

Condition 4 imposes a mild assumption on measurement error matrices since sub-Gaussians are a natural kind of random variable for which the properties of Gaussians can be extended[37]. Condition 5 is standard and has been widely used in high-dimensional linear regression models. Condition 6 guarantees that the number of links for each node is bounded by $K$ from above and that the precision matrix has a bounded spectrum[21].

**Algorithm 1.** COCOISEE

**Input:**

Sample matrix $\boldsymbol{Z}$.

Additive error: covariance matrix $\Sigma^A$; **OR** Multiplicative errors: mean vector $\mu^M$ and covariance matrix $\Sigma^M$.

**Output:** An estimate of the inverse covariance matrix $\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}$ and $\hat{\boldsymbol{\Omega}}_g^{\text{COCOISEE}}$.

**Construct** $\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}$

**for** $S \subset \{1,...,p\}$ and $i \in S$ **do**

$\hat{\Sigma}_{-S,-S}^{\text{add}} = \frac{1}{n}\boldsymbol{Z}_{-S}^{\mathrm{T}}\boldsymbol{Z}_{-S} - \Sigma_{-S,-S}^A$ or

$\hat{\Sigma}_{-S,-S}^{\text{mult}} = \frac{1}{n}\boldsymbol{Z}_{-S}^{\mathrm{T}}\boldsymbol{Z}_{-S} \oslash (\Sigma_{-S,-S}^M + \mu_{-S}^M(\mu_{-S}^M)^{\mathrm{T}})$,

$\tilde{\Sigma}_{-S,-S} = (\hat{\Sigma}_{-S,-S})_+$,

$\tilde{\Sigma}_{-S,i}^{\text{add}} = \frac{1}{n}\boldsymbol{Z}_{-S}^{\mathrm{T}}\boldsymbol{Z}_i$ or $\tilde{\Sigma}_{-S,i}^{\text{mult}} = \frac{1}{n}\boldsymbol{Z}_{-S}^{\mathrm{T}}\boldsymbol{Z}_i \oslash \mu_{-S}^M$,

$\hat{\theta}_i := \underset{\theta_i\in R^{p-|S|}}{\text{argmin}}(\frac{1}{2}\theta_i^{\mathrm{T}}\tilde{\Sigma}_{-S,-S}\theta_i - \theta_i^{\mathrm{T}}\tilde{\Sigma}_{-S,i} + \lambda_i\|\theta_i\|_1)$,

$\hat{\Theta}_S = (\hat{\theta}_i)_{i\in S}$.

**end for**

Construct the block-diagonal entries $\hat{\boldsymbol{\Omega}}_{S,S}$ according to Eq. (12).

Construct block-diagonal entries $\hat{\boldsymbol{\Omega}}_{S_m,S_k}$ according to Eq. (13).

Construct the initial COCOISEE estimator $\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}$.

**Construct** $\hat{\boldsymbol{\Omega}}_g^{\text{COCOISEE}}$

$\hat{\boldsymbol{\Omega}}_g^{\text{COCOISEE}} = T_\tau(\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}})$.


DOI: 10.52396/JUSTC-2022-0108
JUSTC, 2023, 53(11): 1105

### 3.2 Main results

**Theorem 1.** Assume that Conditions 1–6 hold and $\lambda K \to o(1)$. Then with probability $1-\Delta$ tending to one the initial COCOISEE estimator $\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}$ in Eq. (14) satisfies that

$$\|\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}} - \boldsymbol{\Omega}\|_{\infty} = O(\lambda). \tag{22}$$

Theorem 1 establishes the entrywise infinity norm estimation bound for the initial COCOISEE estimator. From the proof of Theorem 1 in Appendix, we see that the rate of convergence for the initial COCOISEE estimator is the maximum of two components $\lambda^2 K$ and $\lambda$ for both block-diagonal and off-block-diagonal entries of $\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}$. Since we assume that $\lambda K \to o(1)$, the rate of convergence $O(\lambda)$ dominates that of $O(\lambda^2 K)$, meaning that the rate of convergence for the initial COCOISEE estimator is change into $O(\lambda)$.

**Theorem 2.** Assume that the conditions of Theorem 1 hold and $\omega_0 = \min\{|\omega_{jk}| : (j,k) \in \text{supp}(\boldsymbol{\Omega})\} \geqslant \omega_0^* = C\lambda$ with $C > 0$ and some sufficiently large constants. Then, with probability $1-\Delta$ tending to one, and for any $\tau \in [c\omega_0^*, \omega_0 - c\omega_0^*]$ with $c \in (0, 1/2)$, we have

$$\text{supp}(\hat{\boldsymbol{\Omega}}_{g}^{\text{COCOISEE}}) = \text{supp}(\boldsymbol{\Omega}). \tag{23}$$

Theorem 2 shows that the sparse precision matrix estimator $\hat{\boldsymbol{\Omega}}_{g}^{\text{COCOISEE}}$ enjoy good asymptotic properties.

## 4 Simulation studies

In this section, we simulated the finite-sample performance of the proposed COCOISEE method. To illustrate the impact of ignoring measurement errors, we compared two methods designed for clean data sets: Glasso[12] and ISEE[21]. The three methods were implemented as follows. Glasso was implemented by the R package *glasso* and had a tuning parameter that was chosen using five-fold cross-validation. ISEE[21] selected the tuning parameters in scaled Lasso[38] following the suggestion of Ref. [39] and adapted the cross-validation method proposed in Refs. [40, 41] for the threshold parameters. For the implementation of our COCOISEE approach, we used 5-fold corrected cross-validation, similar to Ref. [29]. Specifically, we used 90% of the sample to calculate $\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}(n_1, i)$ and the remained 10% to calculate $\hat{\boldsymbol{\Omega}}_{\text{ini}}^{\text{COCOISEE}}(n_2, i)$. Then, we choose $\tau$ from 20 values with the same interval by minimizing criterion (16) with the number of random splits set to $R = 5$.

We generated 50 data sets. For each data set, we generate the precision matrix $\boldsymbol{\Omega}$ in two steps. First, we produce a band matrix $\boldsymbol{\Omega}_0$ with diagonal entries being one, $\boldsymbol{\Omega}_0(i, i+1) = \boldsymbol{\Omega}_0(i+1, i) = 0.5$ for $i = 1, ..., p$, and all other entries being zero. Second, we randomly permute the rows and columns of $\boldsymbol{\Omega}_0$ to obtain the precision matrix $\boldsymbol{\Omega}$. Thus, in general, the final precision matrix $\boldsymbol{\Omega}$ no longer has the band structure. We next sample the rows of the $n \times p$ data matrix $\boldsymbol{X}$ as i.i.d. vectors copied from the multivariate Gaussian distribution $N(0, \boldsymbol{\Omega}^{-1})$. Then, based on different types of measurement errors, the corrupted covariate matrices $\boldsymbol{Z}$ can be provided respectively as follows.

- Additive errors case. The observed covariates $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{A}$, where the rows of $\boldsymbol{A}$ are i.i.d. vectors copied from $N(0, \tau^2)$ with $\tau = 0.2$.
- Multiplicative errors case. The observed covariates are $\boldsymbol{Z} = \boldsymbol{X} \odot \boldsymbol{M}$, where the elements of $\boldsymbol{M}$ follow a log-normal distribution, meaning that $\log(m_{ij})$'s are i.i.d. variables from $N(0, \tau^2)$ with $\tau = 0.2$.
- Missing data case. The observed covariates are defined as $z_{ij} = x_{ij} m_{ij}$, where $m_{ij} = I(x_{ij}$ is not missing), so that the covariates are missing at random with probability 0.1.

Due to high computational cost, we consider settings with $n = 100$, whose dimensionality $p$ varies in $\{50, 100, 150, 200\}$.

To compare the aforementioned methods, we consider the same performance measures as suggested in Ref. [21]. The first two measures are the true positive rate (TPR) and the false positive rate (FPR), defined as

$$\text{TPR} = \frac{\#\text{ of correctly identified edges}}{\#\text{ of identified edges in total}},$$

$$\text{FPR} = \frac{\#\text{ of falsely identified edges}}{\#\text{ of identified nonedges in total}},$$

respectively. We use the Frobenius norm to calculate the normalized estimation error (NEE) defined as

$$\text{NEE}(\hat{\boldsymbol{\Omega}}) = \frac{\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\|_F}{\|\boldsymbol{\Omega}\|_F}.$$

Tables 1–3 summarize the simulation results of these three performance measures for additive, multiplicative error and missing data cases. In view of TPR, FPR, and NEE in these tables, it is clear that the NEE of COCOISEE is the best, and both TPR and FPR are also doing very well.

## 5 Conclusions

In this paper, we have introduced a new methodology COCOISEE to achieve scalable and interpretable estimation for a Gaussian graphical model under both additive and multiplicative measurement errors. It takes full advantage of recently developed innovative scalable efficient estimation and the nearest positive semidefinite matrix projection, thus enjoying stepwise convexity and scalability in Gaussian graphical model estimation. The suggested method is ideal for parallel and distributed computing and cloud computing and has been shown to enjoy appealing theoretical properties. Both the established theoretical properties and numerical performances demonstrate that the proposed method enjoys good estimation, recovery accuracy, and high scalability under both additive and multiplicative measurement errors. Similar theoretical conclusions can also be extended to random sub-Gaussian settings. It would be very interesting to study several extensions of COCOISEE to more general model settings such as the time series model, the generalized linear model and the large nonparanormal graphical model, which are beyond the scope of the current paper and demand future studies.

There are three main contributions in this paper. First, the proposed COCOISEE method has scalability and high efficiency, because the main calculation cost comes from the estimation of the regression coefficient in the case of measurement error, and the other steps adopt matrix operation, which


DOI: 10.52396/JUSTC-2022-0108
*JUSTC*, **2023**, 53(11): 1105

**Table 1.** Additive errors case.

| $p$ | Algorithm | TPR ($\times10^{-2}$) | SE(TPR) ($\times10^{-2}$) | FPR ($\times10^{-2}$) | SE(FPR) ($\times10^{-2}$) | NEE ($\times10^{-2}$) | SE(NEE) ($\times10^{-2}$) |
|---|---|---|---|---|---|---|---|
| 50 | Glasso | 100.00 | 0.00 | 23.36 | 3.90 | 19.49 | 3.73 |
| | ISEE | 99.97 | 0.19 | 9.01 | 1.27 | 15.67 | 1.54 |
| | COCOISEE | 93.92 | 2.16 | 4.26 | 1.26 | 14.92 | 1.71 |
| 100 | Glasso | 100.00 | 0.00 | 14.07 | 1.25 | 28.59 | 2.76 |
| | ISEE | 100.00 | 0.00 | 8.83 | 0.85 | 29.06 | 2.23 |
| | COCOISEE | 91.14 | 2.64 | 3.27 | 0.31 | 21.70 | 2.55 |
| 150 | Glasso | 99.99 | 0.06 | 9.89 | 0.85 | 37.34 | 2.77 |
| | ISEE | 100.00 | 0.00 | 9.34 | 0.75 | 45.00 | 3.42 |
| | COCOISEE | 89.11 | 1.33 | 2.60 | 0.09 | 25.35 | 1.83 |
| 200 | Glasso | 99.97 | 0.10 | 7.67 | 0.66 | 44.08 | 3.01 |
| | ISEE | 99.99 | 0.05 | 9.59 | 0.62 | 60.49 | 3.53 |
| | COCOISEE | 89.10 | 3.49 | 9.31 | 9.63 | 33.66 | 2.14 |

**Table 2.** Multiplicative errors case.

| $p$ | Algorithm | TPR ($\times10^{-2}$) | SE(TPR) ($\times10^{-2}$) | FPR ($\times10^{-2}$) | SE(FPR) ($\times10^{-2}$) | NEE ($\times10^{-2}$) | SE(NEE) ($\times10^{-2}$) |
|---|---|---|---|---|---|---|---|
| 50 | Glasso | 98.30 | 1.63 | 24.83 | 8.60 | 71.92 | 4.08 |
| | ISEE | 87.73 | 3.44 | 11.73 | 1.47 | 63.59 | 1.71 |
| | COCOISEE | 80.74 | 3.25 | 6.47 | 0.91 | 46.11 | 3.49 |
| 100 | Glasso | 95.79 | 1.34 | 17.92 | 6.18 | 82.88 | 2.67 |
| | ISEE | 80.44 | 2.96 | 10.79 | 0.89 | 76.92 | 1.10 |
| | COCOISEE | 79.46 | 2.44 | 7.56 | 2.27 | 57.75 | 1.54 |
| 150 | Glasso | 92.49 | 1.83 | 13.73 | 4.39 | 87.71 | 1.74 |
| | ISEE | 75.07 | 2.30 | 9.87 | 0.78 | 82.61 | 0.89 |
| | COCOISEE | 85.04 | 2.59 | 12.01 | 4.17 | 65.66 | 0.87 |
| 200 | Glasso | 90.24 | 1.82 | 9.98 | 3.69 | 91.01 | 1.83 |
| | ISEE | 72.15 | 2.25 | 9.77 | 0.72 | 86.02 | 0.73 |
| | COCOISEE | 82.24 | 2.33 | 8.65 | 3.31 | 67.35 | 1.42 |

**Table 3.** Missing data case.

| $p$ | Algorithm | TPR ($\times10^{-2}$) | SE(TPR) ($\times10^{-2}$) | FPR ($\times10^{-2}$) | SE(FPR) ($\times10^{-2}$) | NEE ($\times10^{-2}$) | SE(NEE) ($\times10^{-2}$) |
|---|---|---|---|---|---|---|---|
| 50 | Glasso | 99.62 | 0.72 | 24.57 | 9.21 | 53.61 | 7.18 |
| | ISEE | 91.70 | 2.63 | 11.53 | 1.25 | 43.93 | 2.29 |
| | COCOISEE | 81.49 | 3.26 | 3.79 | 0.34 | 39.02 | 2.69 |
| 100 | Glasso | 98.70 | 0.80 | 16.40 | 5.82 | 69.26 | 4.17 |
| | ISEE | 87.03 | 2.17 | 9.75 | 0.67 | 59.78 | 1.89 |
| | COCOISEE | 80.60 | 4.67 | 5.42 | 1.01 | 54.60 | 1.74 |
| 150 | Glasso | 97.29 | 1.07 | 12.69 | 4.72 | 77.40 | 3.79 |
| | ISEE | 81.96 | 2.19 | 9.03 | 0.59 | 68.88 | 1.63 |
| | COCOISEE | 82.28 | 2.87 | 8.74 | 2.49 | 61.20 | 1.57 |
| 200 | Glasso | 96.23 | 1.04 | 10.75 | 3.65 | 81.53 | 3.04 |
| | ISEE | 79.15 | 2.39 | 8.65 | 0.55 | 73.94 | 1.30 |
| | COCOISEE | 84.85 | 2.68 | 13.40 | 3.99 | 65.01 | 2.05 |


DOI: 10.52396/JUSTC-2022-0108
*JUSTC*, **2023**, 53(11): 1105

has relatively high calculation efficiency. Second, we use a recently developed semipositive definite matrix projection technique to mitigate the effects of measurement errors when recovering high-dimensional coefficient matrices from column-by-column regression. Therefore, COCOISEE enjoys progressive convexity and computational stability. Finally, we provide comprehensive theoretical properties to the proposed method by establishing the consistency of the estimator. Numerical results show the effectiveness of the proposed method.

## Conflict of interest

The author declares that he has no conflict of interest.

## Biography

**Xianglu Wang** is currently a master student at the School of Management, University of Science and Technology of China (USTC). He received his B.S. degree from USTC in 2019. His research mainly focuses on high-dimensional variable selection and inference.

## References


[1] Baselmans B M, Jansen R, Ip H F, et al. Multivariate genome-wide analyses of the well-being spectrum. *Nature Genetics,* **2019**, *51* (3): 445–451.

[2] Yang K, Lee L F. Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *Journal of Econometrics,* **2017**, *196* (1): 196–214.

[3] Zhu X, Huang D, Pan R, et al. Multivariate spatial autoregressive model for large scale social networks. *Journal of Econometrics,* **2020**, *215* (2): 591–606.

[4] Han F, Liu H. Optimal rates of convergence for latent generalized correlation matrix estimation in transelliptical distribution. arXiv: 1305.6916, **2013**.

[5] Rubinstein M. Markowitz's "portfolio selection": A fifty-year retrospective. *The Journal of Finance,* **2002**, *57* (3): 1041–1045.

[6] Wegkamp M, Zhao Y. Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas. *Bernoulli,* **2016**, *22* (2): 1184–1226.

[7] Fan J, Han F, Liu H. Challenges of big data analysis. *National Science Review,* **2014**, *1* (2): 293–314.

[8] Cai T, Liu W, Luo X. A constrained $\ell_1$ minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association,* **2011**, *106* (494): 594–607.

[9] Tokuda T, Goodrich B, van Mechelen I, et al. Visualizing distributions of covariance matrices. New York: Columbia University, **2011**.

[10] Fan J, Peng H. Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics,* **2004**, *32* (3): 928–961.

[11] Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model. *Biometrika,* **2007**, *94* (1): 19–35.

[12] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics,* **2008**, *9* (3): 432–441.

[13] Banerjee O, El Ghaoui L, d'Aspremont A. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *The Journal of Machine Learning Research,* **2008**, *9*: 485–516.

[14] Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics,* **2006**, *34* (3): 1436–1462.

[15] Wille A, Zimmermann P, Vranova E, et al. Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. *Genome Biology,* **2004**, *5* (11): R92.

[16] Rothman A J, Bickel P J, Levina E, et al. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics,* **2008**, *2*: 494–515.

[17] Lam C, Fan J. Sparsistency and rates of convergence in large covariance matrix estimation. *The Annals of Statistics,* **2009**, *37* (6B): 4254–4278.

[18] Yuan M. High dimensional inverse covariance matrix estimation via linear programming. *The Journal of Machine Learning Research,* **2010**, *11*: 2261–2286.

[19] Liu W, Luo X. High-dimensional sparse precision matrix estimation via sparse column inverse operator. arXiv: 1203.3896, **2012**.

[20] Sun T, Zhang C H. Sparse matrix inversion with scaled Lasso. *The Journal of Machine Learning Research,* **2013**, *14* (1): 3385–3418.

[21] Fan Y, Lv J. Innovated scalable efficient estimation in ultra-large Gaussian graphical models. *The Annals of Statistics,* **2016**, *44* (5): 2098–2126.

[22] Bickel P, Ritov Y. Efficient estimation in the errors in variables model. *The Annals of Statistics,* **1987**, *15* (2): 513–540.

[23] Ma Y, Li R. Variable selection in measurement error models. *Bernoulli,* **2010**, *16* (1): 274–300.

[24] Liang H, Li R. Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association,* **2009**, *104* (485): 234–248.

[25] Städler N, Bühlmann P. Missing values: Sparse inverse covariance estimation and an extension to sparse regression. *Statistics and Computing,* **2012**, *22* (1): 219–235.

[26] Loh P L, Wainwright M J. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Advances in Neural Information Processing Systems,* **2012**, *40* (3): 1637–1664.

[27] Belloni A, Rosenbaum M, Tsybakov A B. Linear and conic programming estimators in high dimensional errors-in-variables models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* **2017**, *79* (3): 939–956.

[28] Datta A, Zou H. Cocolasso for high-dimensional error-in-variables regression. *The Annals of Statistics,* **2017**, *45* (6): 2400–2426.

[29] Tibshirani R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological),* **1996**, *58* (1): 267–288.

[30] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association,* **2001**, *96* (456): 1348–1360.

[31] Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* **2005**, *67* (2): 301–320.

[32] Zou H. The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association,* **2006**, *101* (476): 1418–1429.

[33] Candes E, Tao T. The Dantzig selector: Statistical estimation when $p$ is much larger than $n$. *The Annals of Statistics,* **2007**, *35* (6): 2313–2351.

[34] Bickel P J, Ritov Y, Tsybakov A B. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics,* **2009**, *37* (4): 1705–1732.

[35] Zhao P, Yu B. On model selection consistency of Lasso. *The Journal of Machine Learning Research,* **2006**, *7*: 2541–2563.

[36] Wainwright M J. Sharp thresholds for high-dimensional and noisy sparsity recovery using $l_1$-constrained quadratic programming (Lasso). *IEEE Transactions on Information Theory,* **2009**, *55* (5): 2183–2202.

[37] Buldygin V V, Kozachenko Yu V. Metric Characterization of Random Variables and Random Processes. Providence, RI: American Mathematical Society, **2000**.

[38] Sun T, Zhang C H. Scaled sparse linear regression. *Biometrika,* **2012**, *99* (4): 879–898.




DOI: 10.52396/JUSTC-2022-0108
*JUSTC*, **2023**, 53(11): 1105

[39] Ren Z, Sun T, Zhang C H, et al. Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *The Annals of Statistics*, **2015**, *43* (3): 991–1026.
[40] Bickel P J, Levina E. Regularized estimation of large covariance matrices. *The Annals of Statistics*, **2008**, *36* (1): 199–227.
[41] Bickel P J, Levina E. Covariance regularization by thresholding. *The Annals of Statistics*, **2008**, *36* (6): 2577–2604.

## Appendix

In this section, we will show proofs of Theorems 1 and 2. For convenience of representation we denote $\boldsymbol{\Omega}_{S_l,S_l}$ as $\boldsymbol{\Omega}_{S_l}$. Lemma A.1 introduced below is fundamental to the proofs.

**Lemma A.1.** Under assumption conditions 1–6, for a suitable choice of penalty parameter $\lambda$ and for the convex optimization problem in (10), we have

$$\left\|\hat{\theta}_{i,l}-\theta^*_{i,l}\right\|_1 = O(\lambda K), \tag{A1}$$

$$\left\|\hat{\theta}_{i,l}-\theta^*_{i,l}\right\|_2^2 = O(\lambda^2 K), \tag{A2}$$

$$\frac{1}{n}\left\|\boldsymbol{X}_{-S_l}(\hat{\theta}_{i,l}-\theta^*_{i,l})\right\|_2^2 = O(\lambda^2 K), \tag{A3}$$

$$\left\|\frac{1}{n}\boldsymbol{X}^{\mathrm{T}}_{-S_l}E_{i,l}\right\|_\infty = O(\lambda), \tag{A4}$$

$$\left\|\hat{\theta}_{i,l}-\theta^*_{i,l}\right\|_\infty = O(\lambda) \tag{A5}$$

with the probability $1-\Delta$ tending to one and it holds uniformly over all nodes $i$ in the index sets $S_l$ with $1\leqslant l\leqslant L$.

**Proof.** The conclusion about the convergence rate in Lemma A.1 has been proven in many studies except Eq. (A3), such as Ref. [28]. Their proofs are similar and we omit them.

We now prove Eq. (A3). Using the application of the triangular inequality and inequality (23) in Ref. [38], which is derived from the Karush–Kuhn–Tucker condition, yields

$$\frac{1}{n}\|\boldsymbol{X}_{-S_l}(\hat{\theta}_{i,l}-\theta^*_{i,l})\|_2^2 \leqslant O(\lambda K)(\|\theta^*_{i,l}\|_1-\|\hat{\theta}^*_{i,l}\|_1)+\frac{1}{n}\|\boldsymbol{X}^{\mathrm{T}}_{-S_l}E_{i,l}\|_\infty\cdot\|\hat{\theta}_{i,l}-\theta^*_{i,l}\|_1.$$

Combining (A1) and (A4), we have

$$\frac{1}{n}\left\|\boldsymbol{X}_{-S_l}(\hat{\theta}_{i,l}-\theta^*_{i,l})\right\|_2^2 \leqslant O(\lambda^2 K).$$

**Proof of Theorem 1.** According to the estimation process of $\hat{\boldsymbol{\Omega}}$, we divide the proof process into two steps: The first step proves the block-diagonal part, and the second step proves the off-block-diagonal part.

**Step 1:** Deriving the uniform bounds on block-diagonal entries of $\hat{\boldsymbol{\Omega}}$. According to (5), the residual matrix is decomposed as follows:

$$\hat{E}_{S_l}=\boldsymbol{X}_{S_l}-\boldsymbol{X}_{-S_l}\hat{\Theta}_{S_l}=E_{S_l}-\boldsymbol{X}_{-S_l}\Theta_{S_l}-\boldsymbol{X}_{-S_l}\hat{\Theta}_{S_l}=E_{S_l}-\boldsymbol{X}_{-S_l}(\Theta_{S_l}-\hat{\Theta}_{S_l}). \tag{A6}$$

Combining (12) yields

$$\hat{\boldsymbol{\Omega}}^{-1}_{S_l}-\boldsymbol{\Omega}^{-1}_{S_l}=\frac{1}{n}\hat{E}^T_{S_l}\hat{E}_{S_l}-\boldsymbol{\Omega}^{-1}_{S_l}=\underbrace{\frac{1}{n}E^{\mathrm{T}}_{S_l}E_{S_l}-\boldsymbol{\Omega}^{-1}_{S_l}}_{\iota_1}\underbrace{-\frac{2}{n}E^{\mathrm{T}}_{S_l}\boldsymbol{X}_{-S_l}(\hat{\Theta}_{S_l}-\Theta_{S_l})}_{\iota_2}+\underbrace{\frac{1}{n}(\hat{\Theta}_{S_l}-\Theta_{S_l})^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}_{-S_l}\boldsymbol{X}_{-S_l}(\hat{\Theta}_{S_l}-\Theta_{S_l})}_{\iota_3}. \tag{A7}$$

For the $\iota_1$ part: According to Condition 6, the spectrum of the precision matrix $\boldsymbol{\Omega}$ is bounded between $M^{-1}$ and $M$, so the spectrum of its principal submatrix $\boldsymbol{\Omega}_{S_l}$ is also between $M^{-1}$ and $M$, its inverse $\boldsymbol{\Omega}^{-1}_{S_l}$ is same. Since $\frac{1}{n}E^{\mathrm{T}}_{S_l}E_{S_l}l$ is the oracle sample covariance matrix estimator for $\boldsymbol{\Omega}^{-1}_{S_l}$, using the concentration bounds in Refs. [29] and [40], Bonferroni's inequality, and $\max_l|S_l|=O(1)$ lead to

$$\max_{1\leqslant l\leqslant L}\left\|\frac{1}{n}E^T_{S_l}E_{S_l}-\boldsymbol{\Omega}^{-1}_{S_l}\right\|_\infty = O(\lambda), \tag{A8}$$

with the probability $1-\Delta$ tending to one.

For the $\iota_2$ and $\iota_3$ parts: In Lemma A.1, bounds (A1) and (A4) control the maximum columnwise $l_1$ norm of matrix $\hat{\Theta}_{S_l}-\Theta_{S_l}$


DOI: 10.52396/JUSTC-2022-0108
*JUSTC*, **2023**, 53(11): 1105

and maximum rowwise $l_\infty$ norm of matrix $\frac{1}{n}E_{S_l}^{\mathrm{T}}\boldsymbol{X}_{-S_l}$ respectively, which yields

$$\|\iota_2\|_\infty \leqslant \left\|\hat{\Theta}_{S_l}-\Theta_{S_l}\right\|_1 \cdot \left\|\frac{1}{n}E_{S_l}^{\mathrm{T}}\boldsymbol{X}_{-S_l}\right\|_\infty = O(\lambda^2 K). \tag{A9}$$

Applying the Cauchy–Schwartz inequality to the bound (A3), we obtain a result for the $\iota_3$ part:

$$\|\iota_3\|_\infty = O(\lambda^2 K). \tag{A10}$$

Thus, combining (A8)–(A10) leads to

$$\max_{1\leqslant l\leqslant L}\left\|\hat{\boldsymbol{\Omega}}_{S_l}^{-1}-\boldsymbol{\Omega}_{S_l}^{-1}\right\|_\infty = O\{\max(\lambda^2 K,\lambda)\}. \tag{A11}$$

Next, we will derive the bounds for $\hat{\boldsymbol{\Omega}}_{S_l}$. Based on (A11) and $|S_l| = O(1)$, we have the Frobenius norm about $\hat{\boldsymbol{\Omega}}_{S_l}^{-1}-\boldsymbol{\Omega}_{S_l}^{-1}$ is

$$\left\|\hat{\boldsymbol{\Omega}}_{S_l}^{-1}-\boldsymbol{\Omega}_{S_l}^{-1}\right\|_F = O\{\max(\lambda^2 K,\lambda)\}. \tag{A12}$$

Then using matrix perturbation theory and large enough $n$ leads to

$$\lambda_{\max}(\hat{\boldsymbol{\Omega}}_{S_l}) = (\lambda_{\min}(\hat{\boldsymbol{\Omega}}_{S_l}^{-1}))^{-1} \leqslant (\lambda_{\min}(\Omega_{S_l}^{-1}) - \|\hat{\boldsymbol{\Omega}}_{S_l}^{-1}-\boldsymbol{\Omega}_{S_l}^{-1}\|_F)^{-1} \leqslant (M^{-1}-O\{\max(\lambda^2 K,\lambda)\})^{-1} \leqslant 2M = O(1). \tag{A13}$$

Note that the entrywise $l_\infty$ norm of any symmetric positive definite matrix is bounded above by its largest eigenvalue, and we have

$$\begin{aligned}\|\boldsymbol{\Omega}_{S_l}\|_\infty &\leqslant \lambda_{\max}(\boldsymbol{\Omega}_{S_l}) = O(1),\\ \|\hat{\boldsymbol{\Omega}}_{S_l}\|_\infty &\leqslant \lambda_{\max}(\hat{\boldsymbol{\Omega}}_{S_l}) = O(1).\end{aligned} \tag{A14}$$

Combining (A11), (A14), and $\max_l|S_l| = O(1)$ leads to

$$\left\|\hat{\boldsymbol{\Omega}}_{S_l}-\boldsymbol{\Omega}_{S_l}\right\|_\infty = \left\|\boldsymbol{\Omega}_{S_l}(\hat{\boldsymbol{\Omega}}_{S_l}^{-1}-\boldsymbol{\Omega}_{S_l}^{-1})\hat{\boldsymbol{\Omega}}_{S_l}\right\|_\infty = O\{\max(\lambda^2 K,\lambda)\}, \tag{A15}$$

with the probability $1-\Delta$ tending to one.

**Step 2:** Deriving the uniform bounds on off-block-diagonal entries of $\hat{\boldsymbol{\Omega}}$. Note that by (13) and (A7), we have the following decomposition of the $\hat{\boldsymbol{X}}_{S_l}$ matrix:

$$\hat{\boldsymbol{X}}_{S_l} = \hat{E}_{S_l}\hat{\boldsymbol{\Omega}}_{S_l} = E_{S_l}\boldsymbol{\Omega}_{S_l} + E_{S_l}(\hat{\boldsymbol{\Omega}}_{S_l}-\boldsymbol{\Omega}_{S_l}) - \boldsymbol{X}_{-S_l}(\hat{\Theta}_{S_l}-\Theta_{S_l})\hat{\boldsymbol{\Omega}}_{S_l} = \widetilde{\boldsymbol{X}}_{S_l} + E_{S_l}(\hat{\boldsymbol{\Omega}}_{S_l}-\boldsymbol{\Omega}_{S_l}) - \boldsymbol{X}_{-S_l}(\hat{\Theta}_{S_l}-\Theta_{S_l})\hat{\boldsymbol{\Omega}}_{S_l}. \tag{A16}$$

Combining (13) and (A16) yields

$$\frac{1}{n}\hat{\boldsymbol{X}}_{S_l}^{\mathrm{T}}\hat{\boldsymbol{X}}_{S_m} = \iota_1+\iota_2+\iota_3+\iota_4, \tag{A17}$$

where

$$\iota_1 = \frac{1}{n}\widetilde{\boldsymbol{X}}_{S_l}^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{S_m}, \tag{A18}$$

$$\iota_2 = \frac{1}{n}\widetilde{\boldsymbol{X}}_{S_l}^{\mathrm{T}}(E_{S_m}(\hat{\boldsymbol{\Omega}}_{S_m}-\boldsymbol{\Omega}_{S_m}) - \boldsymbol{X}_{-S_m}(\hat{\Theta}_{S_m}-\Theta_{S_m})\hat{\boldsymbol{\Omega}}_{S_m}), \tag{A19}$$

$$\iota_3 = \frac{1}{n}(E_{S_l}(\hat{\boldsymbol{\Omega}}_{S_l}-\boldsymbol{\Omega}_{S_l}) - \boldsymbol{X}_{-S_l}(\hat{\Theta}_{S_l}-\Theta_{S_l})\hat{\boldsymbol{\Omega}}_{S_l})^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{S_m}, \tag{A20}$$

$$\iota_4 = \frac{1}{n}(E_{S_l}(\hat{\boldsymbol{\Omega}}_{S_l}-\boldsymbol{\Omega}_{S_l}) - \boldsymbol{X}_{-S_l}(\hat{\Theta}_{S_l}-\Theta_{S_l})\hat{\boldsymbol{\Omega}}_{S_l})^{\mathrm{T}}(E_{S_m}(\hat{\boldsymbol{\Omega}}_{S_m}-\boldsymbol{\Omega}_{S_m}) - \boldsymbol{X}_{-S_m}(\hat{\Theta}_{S_m}-\Theta_{S_m})). \tag{A21}$$

Part $\iota_1$: Since $\frac{1}{n}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}$ is the oracle sample covariance matrix estimator for the precision matrix $\boldsymbol{\Omega}$. Similar to (A8), we have

$$\left\|\frac{1}{n}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}-\boldsymbol{\Omega}\right\|_\infty \leqslant O(\lambda), \tag{A22}$$

with the probability $1-\Delta$ tending to one and provides a uniform bound on $\iota_1$.

Part $\iota_2$: Using (6), we can rewrite $\iota_2$ as


DOI: 10.52396/JUSTC-2022-0108
*JUSTC*, **2023**, 53(11): 1105

$$\begin{aligned}\iota_2 =& \frac{1}{n}\widetilde{\boldsymbol{X}}_{S_l}^{\mathrm{T}}(E_{S_m}(\hat{\boldsymbol{\Omega}}_{S_m}-\boldsymbol{\Omega}_{S_m})-\boldsymbol{X}_{-S_m}(\hat{\Theta}_{S_m}-\Theta_{S_m})\hat{\boldsymbol{\Omega}}_{S_m})=\\ &\frac{1}{n}\boldsymbol{\Omega}_{S_l}^{\mathrm{T}}E_{S_l}^{\mathrm{T}}(E_{S_m}(\hat{\boldsymbol{\Omega}}_{S_m}-\boldsymbol{\Omega}_{S_m})-\boldsymbol{X}_{-S_m}(\hat{\Theta}_{S_m}-\Theta_{S_m})\hat{\boldsymbol{\Omega}}_{S_m})=\\ &\boldsymbol{\Omega}_{S_l}^{\mathrm{T}}(\frac{1}{n}E_{S_l}^{\mathrm{T}}E_{S_m})(\hat{\boldsymbol{\Omega}}_{S_m}-\boldsymbol{\Omega}_{S_m})-\boldsymbol{\Omega}_{S_l}^{\mathrm{T}}(\frac{1}{n}\boldsymbol{X}_{-S_m}^{\mathrm{T}}E_{S_l})^{\mathrm{T}}(\hat{\Theta}_{S_m}-\Theta_{S_m})\hat{\boldsymbol{\Omega}}_{S_m}=\\ &\Xi_1-\Xi_2. \end{aligned} \tag{A23}$$

For the first term $\Xi_1$. Since the mean of $\frac{1}{n}E_{S_l}^{\mathrm{T}}E_{S_m}$ is 0, using the concentration bounds, Bonferroni's inequality over $1\leqslant l\neq m\leqslant L$ yields

$$\max_{1\leqslant l\neq m\leqslant L}\left\|\frac{1}{n}E_{S_l}^{\mathrm{T}}E_{S_m}\right\|_{\infty}=O(\lambda), \tag{A24}$$

with the probability $1-\Delta$ tending to one. Combining $\|\boldsymbol{\Omega}_{S_l}\|_{\infty}=O(1)$ and the result of (A15), we have

$$\|\Xi_1\|_{\infty}=O\{\max(\lambda^3K,\lambda^2)\}. \tag{A25}$$

For the second term $\Xi_2$, we consider the intermediate matrix part $\xi_1$:

$$\xi_1=(\frac{1}{n}\boldsymbol{X}_{-S_m}^{\mathrm{T}}E_{S_l})^{\mathrm{T}}(\hat{\Theta}_{S_m}-\Theta_{S_m})=\xi_2+\xi_3, \tag{A26}$$

where $\xi_2$ and $\xi_3$ are defined through matrix multiplication by taking the rows of $\frac{1}{n}\boldsymbol{X}_{-S_m}^{\mathrm{T}}E_{S_l}$ and $\hat{\Theta}_{S_m}-\Theta_{S_m}$ from nodes in index sets $(-S_m)\cap(-S_l)$ and $S_l$, respectively. According to (A1) and (A4) in Lemma A.1, we have

$$\|\xi_2\|_{\infty}=O(\lambda^2K). \tag{A27}$$

We also define $\xi_4$ as the submatrix of $\hat{\Theta}_{S_m}-\Theta_{S_m}$ given by rows corresponding to nodes in $S_l$. In view of (A5), we have

$$\max_{i\in S_l,1\leqslant l\leqslant L}\left\|\hat{\theta}_{i,l}-\theta_{i,l}\right\|_{\infty}=O(\lambda), \tag{A28}$$

with the probability $1-\Delta$ tending to one. Using similar arguments to those for proving (A24) yields

$$\max_{1\leqslant l\leqslant L}\left\|\frac{1}{n}\boldsymbol{X}_{S_l}^{\mathrm{T}}E_{S_l}-\boldsymbol{\Omega}_{S_l}^{-1}\right\|_{\infty}=O(\lambda), \tag{A29}$$

and combining the fact of $\|\boldsymbol{\Omega}_{S_l}^{-1}\|_{\infty}=O(1)$ leads to $\left\|\frac{1}{n}\boldsymbol{X}_{S_l}^{\mathrm{T}}E_{S_l}\right\|_{\infty}=O(1)$. Therefore

$$\|\xi_3\|_{\infty}=\left\|(\frac{1}{n}\boldsymbol{X}_{S_l}^{\mathrm{T}}E_{S_l})^{\mathrm{T}}\xi_4\right\|_{\infty}=O(\lambda). \tag{A30}$$

Combining (A15), (A26), (A27), and (A30), we have

$$\|\Xi_2\|_{\infty}=O\{\max(\lambda^2K,\lambda)\}. \tag{A31}$$

Since $\iota_3$ shares the same form as $\iota_2$, putting (A23), (A25), and (A31) together leads to

$$\max_{1\leqslant l\neq m\leqslant L}\max(\|\iota_2\|_{\infty},\|\iota_3\|_{\infty})=O\{\max(\lambda^2K,\lambda)\}, \tag{A32}$$

with the probability $1-\Delta$ tending to one.

Part $\iota_4$: We expand the product form of (A21) into four terms:

$$\iota_4=\Gamma_1-\Gamma_2-\Gamma_3+\Gamma_4, \tag{A33}$$

where

$$\Gamma_1=\frac{1}{n}(\hat{\boldsymbol{\Omega}}_{S_l}-\boldsymbol{\Omega}_{S_l})E_{S_l}^{\mathrm{T}}E_{S_m}(\hat{\boldsymbol{\Omega}}_{S_m}-\boldsymbol{\Omega}_{S_m}), \tag{A34}$$

$$\Gamma_2=\frac{1}{n}(\hat{\boldsymbol{\Omega}}_{S_l}-\boldsymbol{\Omega}_{S_l})E_{S_l}^{\mathrm{T}}\boldsymbol{X}_{-S_m}(\hat{\Theta}_{S_m}-\Theta_{S_m})\hat{\boldsymbol{\Omega}}_{S_m}, \tag{A35}$$


DOI: 10.52396/JUSTC-2022-0108
*JUSTC*, **2023**, 53(11): 1105

$$\Gamma_3 = \frac{1}{n}\hat{\boldsymbol{\Omega}}_{S_l}(\hat{\Theta}_{S_l} - \Theta_{S_l})^{\mathrm{T}}\boldsymbol{X}_{-S_l}^{\mathrm{T}}E_{S_m}(\hat{\boldsymbol{\Omega}}_{S_m} - \boldsymbol{\Omega}_{S_m}), \tag{A36}$$

$$\Gamma_4 = \frac{1}{n}\hat{\boldsymbol{\Omega}}_{S_l}(\hat{\Theta}_{S_l} - \Theta_{S_l})^{\mathrm{T}}\boldsymbol{X}_{-S_l}^{\mathrm{T}}\boldsymbol{X}_{-S_m}(\hat{\Theta}_{S_m} - \Theta_{S_m})\hat{\boldsymbol{\Omega}}_{S_m}. \tag{A37}$$

For term $\Gamma_1$:

$$\|\Gamma_1\|_\infty \leqslant O\{\max(\lambda^2 K, \lambda)\} \cdot O(\lambda) \cdot O\{\max(\lambda^2 K, \lambda)\} = O\{\lambda(\max(\lambda^2 K, \lambda))^2\}. \tag{A38}$$

For term $\Gamma_2$ and term $\Gamma_3$: In view of (A26), we have

$$\Gamma_2 = (\hat{\boldsymbol{\Omega}}_{S_l} - \boldsymbol{\Omega}_{S_l})\xi_1\hat{\boldsymbol{\Omega}}_{S_m}, \tag{A39}$$

and $\Gamma_3^{\mathrm{T}}$ shares the same form as $\Gamma_2$. Combining (A15) and (A31) along with the fact of $\|\hat{\boldsymbol{\Omega}}_{S_m}\|_\infty = O(1)$ leads to

$$\max(\|\Gamma_2\|_\infty, \|\Gamma_3\|_\infty) \leqslant O\{\max(\lambda^2 K, \lambda) \cdot \max(\lambda^2 K, \lambda)\} = O\{(\max(\lambda^2 K, \lambda))^2\}. \tag{A40}$$

For the last term $\Gamma_4$: Applying the Cauchy–Schwartz inequality to the bound (A3), we get

$$\left\|\frac{1}{n}\hat{\boldsymbol{\Omega}}_{S_l}(\hat{\Theta}_{S_l} - \Theta_{S_l})^{\mathrm{T}}\boldsymbol{X}_{-S_l}^{\mathrm{T}}\boldsymbol{X}_{-S_m}(\hat{\Theta}_{S_m} - \Theta_{S_m})\right\|_\infty \leqslant O(\lambda^2 K). \tag{A41}$$

Combining (A41) and the facts of $\left\|\hat{\boldsymbol{\Omega}}_{S_l}\right\|_\infty = O(1)$ leads to

$$\|\Gamma_4\|_\infty \leqslant O(\lambda^2 K). \tag{A42}$$

Therefore, combining (A38)–(A42) yields

$$\max_{1\leqslant l\neq m\leqslant L} \|\iota_4\|_\infty = O\{\max(\lambda^2 K, \lambda)\}, \tag{A43}$$

with the probability $1-\Delta$ tending to one. Thus, combining (A22), (A32), and (A43) gives

$$\left\|\frac{1}{n}\hat{\boldsymbol{X}}_{S_l}^{\mathrm{T}}\hat{\boldsymbol{X}}_{S_m}\right\|_\infty = O\{\max(\lambda^2 K, \lambda)\}, \tag{A44}$$

with the probability $1-\Delta$ tending to one.

By doing so, combining steps 1 and 2 with probability $1-\Delta$ tending to one, $\hat{\boldsymbol{\Omega}}$ satisfies that

$$\left\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\right\|_\infty = O\{\max(\lambda^2 K, \lambda)\}.$$

Since we assume that $\lambda K \to o(1)$, the rate of convergence $O(\lambda)$ dominates that of $O(\lambda^2 K)$. Thus,

$$\left\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\right\|_\infty = O(\lambda).$$

**Proof of Theorem 2.** According to the assumption $\omega_0^* = C\lambda$ with $C > 0$ some sufficiently large constants, we have

$$\left\|\hat{\boldsymbol{\Omega}}_{\mathrm{ini}}^{\mathrm{COCOISEE}} - \boldsymbol{\Omega}\right\|_\infty \leqslant c\omega_0^*,$$

where $c \in (0, \frac{1}{2})$ is some positive constants, similar to Ref. [21]. Thus,

- If $\tau \geqslant c\omega_0^*$, $\mathrm{supp}(\hat{\boldsymbol{\Omega}}_g^{\mathrm{COCOISEE}}) \subset \mathrm{supp}(\boldsymbol{\Omega})$.
- If $\tau \leqslant \omega_0 - c\omega_0^*$, $\mathrm{supp}(\boldsymbol{\Omega}) \subset \mathrm{supp}(\hat{\boldsymbol{\Omega}}_g^{\mathrm{COCOISEE}})$.

Combining these two parts leads to

$$\mathrm{supp}(\boldsymbol{\Omega}) = \mathrm{supp}(\hat{\boldsymbol{\Omega}}_g^{\mathrm{COCOISEE}}), \quad \forall \tau \in [c\omega_0^*, \omega_0 - c\omega_0^*].$$


DOI: 10.52396/JUSTC-2022-0108
*JUSTC*, **2023**, 53(11): 1105