

# A unified theory of confidence intervals for high-dimensional precision matrix<sup>\*</sup>

WANG Yue, LI Yang<sup>†</sup>, ZHENG Zemin

(School of Management, University of Science and Technology of China, Hefei 230026, China)

(Received 2 August 2019; Revised 21 October 2019)

Wang Y, Li Y, Zheng Z M. A unified theory of confidence intervals for high-dimensional precision matrix [J]. Journal of University of Chinese Academy of Sciences, 2021, 38(2): 171-180.

**Abstract** Precision matrix inference is of fundamental importance nowadays in high-dimensional data analysis for measuring conditional dependence. Despite the fast growing literature, developing approaches to make simultaneous inference for precision matrix with low computational cost is still in urgent need. In this paper, we apply bootstrap-assisted procedure to conduct simultaneous inference for high-dimensional precision matrix based on the recent de-biased nodewise Lasso estimator, which does not require the irrepresentability condition and is easy to implement with low computational cost. Furthermore, we summary a unified framework to perform simultaneous confidence intervals for high-dimensional precision matrix under the sub-Gaussian case. We show that as long as some precision matrix estimation effects are satisfied, our procedure can focus on different precision matrix estimation methods which owns great flexibility. Besides, distinct from earlier Bonferroni-Holm procedure, this bootstrap method is asymptotically nonconservative. Both numerical results confirm the theoretical results and computational advantage of our method.

**Keywords** precision matrix; high dimensionality; bootstrap-assisted; confidence intervals; simultaneous inference; de-biased

**CLC number:** O212 **Document code:** A **doi:** 10. 7523/j. issn. 2095-6134. 2021. 02. 003

## 基于高维精度矩阵的置信区间的一致性理论

王月, 李阳, 郑泽敏

(中国科学技术大学管理学院, 合肥 230026)

**摘 要** 随着高维数据的不断发展, 精度矩阵作为衡量变量间条件相依性的有效工具引起广泛关注。尽管已有大量文献研究精度矩阵, 但如何发展一种低计算成本的方法构造高维精度矩阵的同时推断变得尤为迫切。基于 nodewise Lasso 估计量, 利用 bootstrap assisted 策略构造同时置信区间。与现有方法相比, 该方法在理论上不需要不可解释性条件且计算成本非常低。进一步, 总结出在次高斯情形下, 精度矩阵同时置信区间的一致性理论, 即只要精度矩阵某些

<sup>\*</sup> Supported by National Natural Science Foundation of China (11601501, 11671374, and 71731010), Anhui Provincial Natural Science Foundation (1708085QA02), and Fundamental Research Funds for the Central Universities (WK2040160028)

<sup>†</sup> Corresponding author, E-mail: tjly@mail.ustc.edu.cn

估计性质满足,该方法可以基于不同的精度矩阵估计方法进行推断。此外,不同于传统的 Bonferroni-Holm,该方法是渐近非保守的。模拟结果验证了该方法的优势。

**关键词** 精度矩阵;高维;bootstrap-assisted;置信区间;同时推断;纠偏

Nowadays, high-dimensional data which are referred to as small  $n$  large  $p$  data, develop extremely rapidly. Graphical models have been extensively used as a solid tool to measure conditional dependence structure between different variables, ranging from genetics, proteins and brain networks to social networks, online marketing and portfolio optimization. It is well known that the edges of Gaussian graphical model (GGM) are encoded by the corresponding entries of the precision matrix<sup>[1]</sup>. While most of the existing work concentrates on the estimation and individual inference of precision matrix, simultaneous inference methods are generally reckoned to be more useful in practical applications because of the valid reliability assurance. Therefore, it is in urgent need to develop approaches to make inference for groups of entries of the precision matrix.

Making individual inference for the precision matrix has been widely studied in the literature. Ref. [2] first advocated multiple testing for conditional dependence in GGM with false discovery rates control. It's a pity that this method can not be applied to construct confidence intervals directly. To address this issue, based on the so-called de-biased or de-sparsified procedure, Refs. [3-4] designed to remove the bias term of the initial Lasso-type penalized estimators and achieved asymptotically normal distribution for each entry of the precision matrix. Difference lies in that Ref. [3] adopted graphical Lasso as initial Lasso-type penalized estimator but Ref. [4] focused on nodewise Lasso. They both followed the way of Refs. [5-8] which proposed de-biased steps for inference in high-dimensional linear models.

While most recent studies have focused on the individual inference in high-dimensional regime, the simultaneous inference remains largely unexplored. Refs. [9-11] creatively proposed multiplier bootstrap method. Based on the individual confidence

interval, Ref. [12] proposed simultaneous confidence intervals via applying bootstrap scheme to high-dimensional linear models. Distinct from earlier Bonferroni-Holm procedure, this bootstrap method is asymptotically nonconservative because it considers the correlation among the test statistics. More recently, Ref. [13] considered combinatorial inference aiming at testing the global structure of the graph at the cost of heavy computation and only limited to the Gaussian case.

Motivated by these concerns, we develop a bootstrap-assisted procedure to conduct simultaneous inference for high-dimensional precision matrix, based on the de-biased nodewise Lasso estimator. Moreover, we summary a unified framework to perform simultaneous inference for high-dimensional precision matrix. Our method imitates Ref. [12] but generalizes bootstrap-assisted scheme to graphical models and we conclude general theory that our method is applicative as long as precision matrix estimation satisfies some common conditions. The major contributions of this paper are threefold. First of all, we develop a bootstrap-assisted procedure to conduct simultaneous inference for high-dimensional precision matrix, which is adaptive to the dimension of the concerned component and considers the dependence within the de-biased nodewise Lasso estimators while Bonferroni-Holm procedure cannot attain. Second, our method is easy to implement and enjoy nice computational efficiency without loss of accuracy. Last, we provide theoretical guarantees for constructing simultaneous confidence intervals of the precision matrix under a unified framework. We prove that our simultaneous testing procedure asymptotically achieves the preassigned significance level even when the model is sub Gaussian and the dimension is exponentially larger than sample size.

Notations. For a vector  $\mathbf{x} = (x_1, \cdots, x_p)^T$ , denote  $\|\mathbf{x}\|_q = (\sum_{j=1}^p |x_j|^q)^{1/q}$  the  $l_q$ -norm for  $q \in$

$(0, \infty)$ ,  $\|\mathbf{x}\|_0 = |\{j: x_j \neq 0\}|$ ,  $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq p} |x_i|$ ,  $(\mathbf{x})_i$  denote the  $i$ th row of  $\mathbf{x}$  and  $\mathbf{x}_{-j}$  denote the sub-vector without  $j$ th component. For a matrix  $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{p \times q}$ , denote  $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq q} \sum_{i=1}^p |a_{ij}|$ ,  $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq p} \sum_{j=1}^q |a_{ij}|$ ,  $\|\mathbf{A}\|_{\max} = \max_{i,j} |a_{ij}|$ ,  $\|\mathbf{A}\|_2 = \lambda_{\max}(\mathbf{A}^T \mathbf{A})^{1/2}$  the matrix  $l_1$ -norm,  $l_\infty$ -norm, elementwise maximum norm, and spectral norm, respectively, where  $\lambda_{\min}(\mathbf{A})$  and  $\lambda_{\max}(\mathbf{A})$  denote the minimum and maximum eigenvalues of the given matrix  $\mathbf{A}$ .  $A_{ij}$  denote the  $(i, j)$ -entry of  $\mathbf{A}$ ,  $\mathbf{A}_j$  denote the  $j$ th column of  $\mathbf{A}$ ,  $\mathbf{A}_{(i)}$  denote the  $i$ th row of  $\mathbf{A}$ ,  $\mathbf{A}_{-j}$  denote the sub-matrix of  $\mathbf{A}$  without the  $j$ th column,  $\mathbf{A}_{-i,j}$  denote the  $j$ th column of  $\mathbf{A}$  without its  $i$ th entry and  $\mathbf{A}_{-i,-j}$  denote the sub-matrix of  $\mathbf{A}$  without the  $i$ th row and  $j$ th column. For two real sequences  $\{f_n\}$  and  $\{g_n\}$ , we write  $f_n = O(g_n)$  if there exists a constant  $C$  such that  $|f_n| \leq C |g_n|$  holds for all  $n$ ,  $f_n = o(g_n)$  if  $\lim_{n \rightarrow \infty} f_n/g_n = 0$  and  $f_n \asymp g_n$  if  $f_n = O(g_n)$  and  $g_n = O(f_n)$ . The sub-gaussian norm of a random variable  $Z$ , denoted by  $\|Z\|_{\psi_2}$ , is defined as  $\|Z\|_{\psi_2} = \sup_{q \geq 1} q^{-1/2} (\mathbb{E} Z^q)^{1/q}$ . The sub-Gaussian norm of a random vector  $\mathbf{Z}$  is defined as  $\|\mathbf{Z}\|_{\psi_2} = \sup_{\|\mathbf{x}\|_2=1} \langle \mathbf{Z}, \mathbf{x} \rangle$ . Finally, the sub-exponential norm of random variable  $Z$  is defined as  $\|Z\|_{\psi_1} = \sup_{q \geq 1} q^{-1} (\mathbb{E} Z^q)^{1/q}$ .  $c_i$  and  $C_i$  be some constants independent of  $n$  and  $p$ .

## 1 Methodology

### 1.1 Model setting

Under the graphical model framework, denote by  $\mathbf{X}$  an  $n \times p$  random design matrix with  $p$  covariates. Assume that  $\mathbf{X}$  has independent sub-Gaussian rows  $\mathbf{X}_{(i)}$ , that is, there exists constant  $K$  such that

$$\sup_{\alpha \in \mathbb{R}^p: \|\alpha\|_2 \leq 1} \mathbb{E} \exp(|\alpha^T \mathbf{X}_{(i)}|^2 / K^2) \leq 2. \quad (1)$$

### 1.2 De-biased nodewise Lasso

Characterizing the distribution of Lasso-type estimator for precision matrix is difficult because Lasso-type estimator is biased due to the  $l_1$  penalization. To address this problem, Refs. [3-4] adopted de-biasing idea which is to start with graphical Lasso or nodewise Lasso estimator and then

remove its bias. This results in de-biased estimator generally taking the form

$$\tilde{\boldsymbol{\Theta}} = \hat{\boldsymbol{\Theta}} - \hat{\boldsymbol{\Theta}}^T (\hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\Theta}} - \mathbf{I}_p).$$

Then we have the following estimation error decomposition

$$\begin{aligned} \sqrt{n}(\tilde{\boldsymbol{\Theta}}_{jk} - \boldsymbol{\Theta}_{jk}) &= -\sqrt{n} \boldsymbol{\Theta}_j^T (\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \boldsymbol{\Theta}_k + \Delta_{jk} \\ &= -\sum_{i=1}^n (\boldsymbol{\Theta}_j^T \mathbf{X}_{(i)} \mathbf{X}_{(i)}^T \boldsymbol{\Theta}_k / \sqrt{n} - \boldsymbol{\Theta}_{jk}) + \Delta_{jk}, \end{aligned}$$

where  $\hat{\boldsymbol{\Sigma}} = \mathbf{X}^T \mathbf{X} / n$  is the empirical covariance. The first term is the main term whose covariance works out as  $\text{Var}(\boldsymbol{\Theta}_j^T \mathbf{X}_{(i)} \mathbf{X}_{(i)}^T \boldsymbol{\Theta}_k)$ , while the second term is the bias term and is controlled by  $O_p(\text{slogp} / (\sqrt{n}))$ .

Various estimations of  $\hat{\boldsymbol{\Theta}}$  have been offered by different works, but the basic insight lies in that  $\hat{\boldsymbol{\Theta}}$  should act as a good initial estimator. In this paper, we adopt the de-biased nodewise Lasso proposed in Ref. [4] to illustrate our simultaneous inference approach due to the appealing property that de-biased nodewise Lasso does not require the irrepresentability condition and is easy to implement with low computational cost. More general choices of  $\hat{\boldsymbol{\Theta}}$  will be introduced in Section 4. Following Ref. [4], for each  $j = 1, \dots, p$ , define the vector  $\hat{\boldsymbol{\gamma}}_j = \{\hat{\boldsymbol{\gamma}}_{j,k}, k = 1, \dots, p, j \neq k\}$  as

$$\hat{\boldsymbol{\gamma}}_j = \arg \min_{\boldsymbol{\gamma} \in \mathbb{R}^{p-1}} \left\{ \frac{1}{n} \|\mathbf{X}_j - \mathbf{X}_{-j} \boldsymbol{\gamma}\|_2^2 + 2\lambda_j \|\boldsymbol{\gamma}\|_1 \right\}.$$

Further we let

$$\begin{aligned} \hat{\Gamma}_j &= (-\hat{\boldsymbol{\gamma}}_{j,1}, \dots, -\hat{\boldsymbol{\gamma}}_{j,j-1}, 1, -\hat{\boldsymbol{\gamma}}_{j,j+1}, \dots, -\hat{\boldsymbol{\gamma}}_{j,p})^T, \\ \hat{\tau}_j^2 &= \|\mathbf{X}_j - \mathbf{X}_{-j} \hat{\boldsymbol{\gamma}}_j\|_2^2 / n + \lambda_j \|\hat{\boldsymbol{\gamma}}_j\|_1. \end{aligned}$$

Then the  $j$ th column of the nodewise Lasso estimator  $\hat{\boldsymbol{\Theta}}$  is defined as

$$\hat{\boldsymbol{\Theta}}_j = \hat{\Gamma}_j / \hat{\tau}_j.$$

However, we realize that  $\hat{\boldsymbol{\Theta}}$  is not unique. It only needs to be a good estimator of  $\boldsymbol{\Sigma}^{-1}$  in order for the de-biasing procedure to work. As a consequence, our approach applies to general precision matrix inference. By contrast, earlier approaches applied only to de-biased graphical Lasso, as in Ref. [3], or de-biased nodewise Lasso as in Ref. [4]. The only assumptions we make on  $\hat{\boldsymbol{\Theta}}$  are the event  $H$  containing three common conditions described in Section 1.4.

### 1.3 Simultaneous confidence intervals

We extend the idea of de-biased nodewise Lasso estimator to construct confidence intervals for any subsets of the entries of the precision matrix. Specifically, we are interested in deriving the distribution of  $\max_{(j,k) \in E} \sqrt{n}(\check{\Theta}_{jk} - \Theta_{jk})$ . For convenience, define statistic

$$T_E =: \max_{(j,k) \in E} \sqrt{n}(\check{\Theta}_{jk} - \Theta_{jk}). \quad (2)$$

Following the idea of Refs. [9-10, 12], we use a bootstrap-assisted scheme to make simultaneous inference for graphical model. Let  $\{e_i\}_{i=1}^n$  be an i. i. d. sequence of  $N(0,1)$  random variables independent of  $X_i$ . Define the multiplier bootstrap statistic as follows

$$W_E =: \max_{(j,k) \in E} \left| \sum_{i=1}^n \hat{Z}_{ijk} e_i / \sqrt{n} \right|, \quad (3)$$

where  $\hat{Z}_{ijk} = \hat{\Theta}_j^T X_{(i)} X_{(i)}^T \hat{\Theta}_k - \hat{\Theta}_{jk}$  independent of  $\{e_i\}_{i=1}^n$  and  $E$  to be an arbitrary subset of  $[p] \times [p]$ . The bootstrap critical value is determined by

$$c_{1-\alpha,E} = \inf\{t \in \mathbb{R} : \mathbb{P}_e(W_E \leq t) \leq 1 - \alpha\}, \quad (4)$$

denoting the  $(1 - \alpha)$ -quantile of the statistic  $W_E$ , where  $\mathbb{P}_e$  is the probability measure induced by the variables  $\{e_i\}_{i=1}^n$  with  $X$  fixed, that is,  $\mathbb{P}_e(W_E \leq t) = \mathbb{P}(W_E \leq t | X)$ . The core point lies in that we want to use the quantile of the statistic  $W_E$  to asymptotically estimate the quantile of the statistic  $T_E$ . We will provide the strict theoretical proof in Section 3.

**Remark 1.1** Bonferroni-Holm adjustment states that if an experimenter is testing  $p$  hypotheses on a set of data, then the statistical significance level for each independent hypothesis separately is  $1/p$  times what it would be if only one hypothesis were tested. However, the bootstrap uses the quantile of the multiplier bootstrap statistic to asymptotically estimate the quantile of the target statistic and takes dependence among the test statistics into account. Thus the original method with Bonferroni-Holm is on the conservative side, while the bootstrap is closer to the preassigned significance level.

### 1.4 A unified theory for confidence intervals

Define the parameter set

$$\mathcal{M}(s) = \{\Theta \in \mathbb{R}^{p \times p} : 1/L \leq \lambda_{\min}(\Theta) \leq \lambda_{\max}(\Theta) \leq L,$$

$$\max_{j \in [p]} \|\Theta_j\|_0 \leq s, \|\Theta\|_1 \leq C\}$$

for some  $1 \leq L \leq C$ . We can extend the above conclusions to more general regime. Let  $\hat{\Theta}$  be any estimator of the precision matrix  $\Theta$  satisfying the following event  $H$ :

$$\|\hat{\Theta} - \Theta\|_{\max} = O_p(\sqrt{\log p/n}),$$

$$\|\hat{\Theta} - \Theta\|_1 = O_p(s\sqrt{\log p/n}),$$

$$\|\hat{\Sigma}\hat{\Theta} - I_p\|_{\max} = O_p(\sqrt{\log p/n}),$$

with probability tending to one uniformly over the parameter space  $\mathcal{M}(s)$ . In fact, event  $H$  contains very mild conditions. As we all know, many precision matrix estimation methods satisfy the above conditions, such as graphical Lasso, CLIME<sup>[14]</sup>, SCIO<sup>[15]</sup> and so on. The de-biased graphical Lasso inference has already been proposed in Ref. [3]. Different from them, we conclude a more general theory as long as it satisfies the event  $H$ , we can obtain simultaneous confidence intervals honestly. Besides, our theory applies to sub-Gaussian case. Theoretical properties will be provided in the following section.

## 2 Theoretical properties

Before giving the theoretical properties, we list two technical conditions.

(A1) Assume that  $s\log p/\sqrt{n} = o(1)$ .

(A2) Assume that  $B_n^2(\log(pn))^7/n \leq C_1 n^{-c_1}$ , where  $B_n \geq 1$  be a sequence of constants and  $\lim_{n \rightarrow \infty} B_n = \infty$ , and  $c_1$  and  $C_1$  are some positive constants.

**Proposition 2.1** (Lemma 1 of Ref. [4])

Consider the sub-Gaussian model and let  $\hat{\Theta}$  be the nodewise Lasso estimator with  $\lambda_j \asymp \sqrt{\log p/n}$  uniformly in  $j$ . Then for any  $(j,k) \in [p] \times [p]$ , we have

$$\sqrt{n}(\check{\Theta}_{jk} - \Theta_{jk}) = -\check{\Theta}_j^T(\hat{\Sigma} - \Sigma)\Theta_k + \Delta_{jk},$$

where  $\Delta_{jk} = -\sqrt{n}\{(\hat{\Theta}_j - \Theta_j)^T(\hat{\Sigma}\hat{\Theta}_k - e_k) + (\hat{\Theta}_k - \Theta_k)^T(\hat{\Sigma}\Theta_j - e_j)\}$  and

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\Theta} \in \mathcal{M}(s)} \mathbb{P} \left\{ \max_{(j,k) \in [p] \times [p]} |\Delta_{jk}| \geq O \left( \frac{\text{slogp}}{\sqrt{n}} \right) \right\} = 0.$$

Furthermore, define the variance  $\sigma_{jk}^2 = \text{Var}(\boldsymbol{\Theta}_j^T \mathbf{X}_1 \mathbf{X}_1^T \boldsymbol{\Theta}_k)$ . If condition (A1) holds, we have

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\Theta} \in \mathcal{M}(s)} |\mathbb{P}(\sqrt{n}(\check{\boldsymbol{\Theta}}_{jk} - \boldsymbol{\Theta}_{jk})/\hat{\sigma}_{jk} \leq z) - \Phi(z)| = 0,$$

where  $\hat{\sigma}_{jk}^2 = \sum_{i=1}^n (\hat{\boldsymbol{\Theta}}_j^T \mathbf{X}_{(i)} \mathbf{X}_{(i)}^T \hat{\boldsymbol{\Theta}}_k)^2/n - \hat{\boldsymbol{\Theta}}_{jk}^2$  is a consistent estimator of  $\sigma_{jk}^2$ .

Based on the asymptotic normality properties established in Proposition 2.1, we have the following simultaneous confidence intervals for multiple entries  $\boldsymbol{\Theta}_{jk}$ .

**Theorem 2.1** Assume that conditions (A1)-(A2) hold. Then for any  $E \subseteq [p] \times [p]$ , we have

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\Theta} \in \mathcal{M}(s)} \sup_{\alpha \in (0,1)} |\mathbb{P}(\sqrt{n} \|\check{\boldsymbol{\Theta}}_E - \boldsymbol{\Theta}_E\|_{\max} \leq c_{1-\alpha,E}) - (1-\alpha)| = 0,$$

where  $\boldsymbol{\Theta}_E$  denotes the entries of  $\boldsymbol{\Theta}$  with indices in  $E$ .

Theorem 2.1 states that we can approximate the  $(1-\alpha)$ -quantile of  $\max_{(j,k) \in E} \sqrt{n} |\check{\boldsymbol{\Theta}}_{jk} - \boldsymbol{\Theta}_{jk}|$  by  $(1-\alpha)$ -quantile of the multiplier bootstrap statistic  $W_E$  consistently. Note that although we cannot compute the quantile of  $W_E$  analytically, we can obtain it via Monte Carlo simulations in practice. This result derives the simultaneous confidence intervals for  $\boldsymbol{\Theta}_{jk}$  as:  $\check{\boldsymbol{\Theta}}_{jk} \pm c_{1-\alpha,E}/\sqrt{n}$  where  $(j,k) \in E$ . As discussed in Ref. [9], our method applies to high-dimensional situation. In contrast with the earlier Bonferroni simultaneous confidence intervals, our bootstrap-assisted scheme is asymptotically nonconservative and can work for arbitrary subset  $E$ .

Next, we extend the above theory to more general case and conclude the unified theory for precision matrix inference.

**Theorem 2.2** Assume that event  $H$  holds. Then we have

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\Theta} \in \mathcal{M}(s)} \sup_{\alpha \in (0,1)} |\mathbb{P}(\sqrt{n} \|\check{\boldsymbol{\Theta}}_E - \boldsymbol{\Theta}_E\|_{\max} \leq c_{1-\alpha,E}) - (1-\alpha)| = 0,$$

for any  $E \subseteq [p] \times [p]$ , where  $\boldsymbol{\Theta}_E$  denotes the entries of  $\boldsymbol{\Theta}$  with indices in  $E$ .

Next, we extend the above theory to more general case and conclude the unified theory for

precision matrix inference.

**Theorem 2.3** Assume that event  $H$  holds.

Then we have

(A) (Individual inference)

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\Theta} \in \mathcal{M}(s)} |\mathbb{P}(\sqrt{n}(\check{\boldsymbol{\Theta}}_{jk} - \boldsymbol{\Theta}_{jk})/\hat{\sigma}_{jk} \leq z) - \Phi(z)| = 0,$$

where  $\hat{\sigma}_{jk}^2 = \sum_{i=1}^n (\hat{\boldsymbol{\Theta}}_j^T \mathbf{X}_{(i)} \mathbf{X}_{(i)}^T \hat{\boldsymbol{\Theta}}_k)^2/n - \hat{\boldsymbol{\Theta}}_{jk}^2$  is a consistent estimator of  $\sigma_{jk}^2$ .

(B) (Simultaneous inference)

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\Theta} \in \mathcal{M}(s)} \sup_{\alpha \in (0,1)} |\mathbb{P}(\sqrt{n} \|\check{\boldsymbol{\Theta}}_E - \boldsymbol{\Theta}_E\|_{\max} \leq c_{1-\alpha,E}) - (1-\alpha)| = 0,$$

for any  $E \subseteq [p] \times [p]$ , where  $\boldsymbol{\Theta}_E$  denotes the entries of  $\boldsymbol{\Theta}$  with indices in  $E$ .

Theorem 2.3 presents general conclusions for both individual and simultaneous confidence intervals. That is, our inferential procedures work for any estimation methods for precision matrix as long as the estimation effect satisfies event  $H$ .

### 3 Numerical studies

In this section, we investigate the finite sample performance of the methods proposed in Section 3 and provide a comparison to simultaneous confidence interval for de-biased graphical Lasso, denoted by S-NL and S-GL, respectively. We now present two numerical examples and evaluate the methods by estimated average coverage probabilities (avgcov) and average confidence interval lengths (avglen) over two cases: support set  $S$  and its complement  $S^c$ . For convenience, we only consider Gaussian setting. The implementation for de-biased nodewise Lasso and de-biased graphical Lasso are suggested by Ref. [4]. Throughout the simulation, the level of significance is set at  $\alpha = 0.05$  and the coverage probabilities and interval lengths calculated by averaging over 100 simulation runs and 500 Monte Carlo replications. For extra comparison, we also record individual confidence intervals for de-biased nodewise Lasso and de-biased graphical Lasso, denoted by I-NL and I-GL, respectively.

#### 3.1 Numerical example 1: band structure

We start with a numerical example which has

the similar setting as that in Ref. [3]. We consider the precision matrix  $\boldsymbol{\Theta}$  with the band structure, where  $\Theta_{jj} = 1, \Theta_{j,j+1} = \Theta_{j+1,j} = \rho$  for  $j = 1, 2, \dots, p - 1$ , and zero otherwise. We sample the rows of the  $n \times p$  data matrix  $\boldsymbol{X}$  as i. i. d. copies from the multivariate Gaussian distribution  $N(\boldsymbol{0}, \boldsymbol{\Sigma})$  where  $\boldsymbol{\Sigma} = \boldsymbol{\Theta}^{-1}$ . We fix the sample size  $n = 100$  and consider a range of dimensionality  $p = 300, 500$  and link strength  $\rho = 0.2, 0.3, 0.4, 0.5$ , respectively. The results are summarized in Table 1.

In terms of avgcov and avglen, it is clear that our proposed S-NL method outperforms other

alternative methods with higher avgcov and shorter avglen in most settings. Although the avglen over  $S^c$  may be a little longer in some cases, it is amazing that the coverage probabilities in  $S$  approach the nominal coverage 95%. On the other hand, the advantage becomes more evident as  $p$  and  $\rho$  increase. Compared with individual confidence intervals, simultaneous confidence intervals have longer lengths and lower coverage probabilities. This is reasonable because multiplicity adjustment damages partial accuracy which is inevitable.

Table 1 Averaged coverage probabilities and lengths over the support set  $S$  and its complement  $S^c$  in Section 3.1

$\rho$	Method	$p = 300$				$p = 500$			
		$S$		$S^c$		$S$		$S^c$	
		Avgcov	Avglen	Avgcov	Avglen	Avgcov	Avglen	Avgcov	Avglen
0.2	S-NL	0.932	0.824	0.931	0.793	0.925	0.743	0.945	0.841
	S-GL	0.788	0.798	0.942	0.812	0.893	0.751	0.953	0.847
	I-NL	0.955	0.424	0.952	0.375	0.954	0.428	0.958	0.373
	I-GL	0.956	0.425	0.976	0.372	0.953	0.427	0.981	0.375
0.3	S-NL	0.919	0.726	0.935	0.712	0.916	0.722	0.943	0.833
	S-GL	0.752	0.740	0.938	0.802	0.846	0.745	0.944	0.816
	I-NL	0.931	0.399	0.963	0.349	0.927	0.396	0.962	0.346
	I-GL	0.899	0.400	0.987	0.349	0.846	0.402	0.992	0.351
0.4	S-NL	0.912	0.713	0.934	0.804	0.873	0.699	0.939	0.821
	S-GL	0.725	0.721	0.936	0.798	0.623	0.704	0.936	0.813
	I-NL	0.932	0.399	0.963	0.349	0.895	0.364	0.977	0.315
	I-GL	0.807	0.388	0.995	0.333	0.696	0.393	0.997	0.338
0.5	S-NL	0.790	0.498	0.864	0.458	0.731	0.256	0.792	0.305
	S-GL	0.598	0.782	0.931	0.672	0.425	0.902	0.931	0.683
	I-NL	0.811	0.234	0.890	0.214	0.796	0.134	0.834	0.153
	I-GL	0.642	0.422	0.997	0.353	0.538	0.454	0.990	0.381

### 3.2 Numerical example 2: nonband structure

For the second numerical example, we use the same setup as simulation example 1 in Ref. [16] to test the performance of S-NL in more general cases. We generate the precision matrix in two steps. First, we create a band matrix  $\boldsymbol{\Theta}_0$  the same as that in Section 3.1. Second, we randomly permute the rows and columns of  $\boldsymbol{\Theta}_0$  to obtain the precision matrix  $\boldsymbol{\Theta}$ . The final precision matrix  $\boldsymbol{\Theta}$  no longer has the band structure. Then we sample the rows of the  $n \times p$  data matrix  $\boldsymbol{X}$  as i. i. d. copies from the multivariate Gaussian distribution  $N(\boldsymbol{0}, \boldsymbol{\Sigma})$  where  $\boldsymbol{\Sigma} = \boldsymbol{\Theta}^{-1}$ . Throughout this simulation, we fix the sample size  $n = 200$ , dimensionality  $p = 1\,000$  and consider a

range of  $\rho = 0.2, 0.3, 0.4, 0.5$ .

Simulation results summarized in Table 2 also illustrate that our method can achieve the preassigned significance level asymptotically and behaves better than others in most cases. Moreover, we can see our method is very robust especially in large  $\rho$ .

## 4 Discussions

In this paper, we apply bootstrap-assisted procedure to make valid simultaneous inference for high-dimensional precision matrix based on the recent de-biased nodewise Lasso estimator. In addition, we summary a unified framework to perform simultaneous confidence intervals for high-



**Table 2** Averaged coverage probabilities and lengths over the support set  $S$  and its complement  $S^c$  in Section 3. 2

$\rho$	Method	$S$		$S^c$	
		Avgcov	Avglen	Avgcov	Avglen
0. 2	S-NL	0. 936	0. 458	0. 939	0. 615
	S-GL	0. 925	0. 463	0. 935	0. 623
	I-NL	0. 954	0. 291	0. 969	0. 252
	I-GL	0. 941	0. 292	0. 988	0. 254
0. 3	S-NL	0. 910	0. 602	0. 936	0. 611
	S-GL	0. 831	0. 601	0. 932	0. 613
	I-NL	0. 936	0. 283	0. 964	0. 249
	I-GL	0. 859	0. 284	0. 991	0. 249
0. 4	S-NL	0. 908	0. 587	0. 942	0. 601
	S-GL	0. 458	0. 598	0. 943	0. 597
	I-NL	0. 933	0. 264	0. 975	0. 232
	I-GL	0. 692	0. 279	0. 996	0. 246
0. 5	S-NL	0. 825	0. 324	0. 856	0. 258
	S-GL	0. 227	0. 269	0. 883	0. 236
	I-NL	0. 864	0. 152	0. 899	0. 134
	I-GL	0. 321	0. 129	0. 991	0. 101

dimensional precision matrix under the sub-Gaussian case. As long as some estimation effects are satisfied, our procedure can focus on different precision matrix estimation methods which owns great flexibility. Further, this method can be expended to more general settings, such as functional graphical model where the samples are consisted of functional data. We leave this problem for further investigations.

References

[ 1 ]

Lauritzen S L. Graphical Models [ M ]. New York: Oxford University Press, 1996.

[ 2 ]

Liu W. Gaussian graphical model estimation with false discovery rate control [ J ]. Ann Statist, 2013, 41 ( 6 ) : 2948-2978.

[ 3 ]

Janková J, van de Geer S. Confidence intervals for high-dimensional inverse covariance estimation [ J ]. Electron J Statist, 2015, 9: 1205-1229.

[ 4 ]

Janková J, Van de Geer S. Honest confidence regions and optimality in high-dimensional precision matrix estimation

[ J ]. Test, 2017, 26 ( 1 ) : 143-162.

[ 5 ]

Bühlmann P. Statistical significance in high-dimensional linear models [ J ]. Bernoulli, 2013, 19 ( 4 ) : 1212-1242.

[ 6 ]

Javanmard A, Montanari A. Confidence intervals and hypothesis testing for high-dimensional regression [ J ]. J Mach Learn Res, 2014, 15: 2869-2909.

[ 7 ]

van de Geer S, Bühlmann P, Ritov Y, et al. On asymptotically optimal confidence regions and tests for high-dimensional models [ J ]. Ann Statist, 2014, 42 ( 3 ) : 1166-1202.

[ 8 ]

Zhang C, Zhang S. Confidence intervals for low dimensional parameters in high dimensional linear models [ J ]. J R Stat Soc Ser B Stat Methodol, 2014, 76 ( 1 ) : 217-242.

[ 9 ]

Chernozhukov V, Chetverikov D, Kato K. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors [ J ]. Ann Statist, 2013, 41 ( 6 ) : 2786-2819.

[ 10 ]

Chernozhukov V, Chetverikov D, Kato K. Comparison and anti-concentration bounds for maxima of Gaussian random vectors [ J ]. Probability Theory and Related Fields, 2015, 162: 47-70.

[ 11 ]

Chernozhukov V, Chetverikov D, Kato K. Central limit theorems and bootstrap in high dimensions [ J ]. Annals of Probability, 2016, 45 ( 4 ) : 2309-2352.

[ 12 ]

Zhang X, Cheng G. Simultaneous inference for high-dimensional linear models [ J ]. J Amer Statist Assoc, 2016, 112 ( 2 ) : 757-768.

[ 13 ]

Neykov M, Lu J, Liu H. Combinatorial inference for graphical models [ J ]. Ann Statist, 2019, 47 ( 6 ) : 795-827.

[ 14 ]

Cai T, Liu W, Luo X. A constrained  $l_1$  minimization approach to sparse precision matrix estimation [ J ]. J Amer Statist Assoc, 2011, 106 ( 494 ) : 594-607.

[ 15 ]

Liu W D, Luo X. Fast and adaptive sparse precision matrix estimation in high dimensions [ J ]. Journal of Multivariate Analysis, 2015, 135 ( 4 ) : 153-162.

[ 16 ]

Fan Y, Lv J. Innovated scalable efficient estimation in ultra-large gaussian graphical models [ J ]. Ann Statist, 2016, 44 ( 5 ) : 2098-2126.

[ 17 ]

Vershynin R. Introduction to the non-asymptotic analysis of random matrices [ M ] // Eldar Y C, Kutyniok G. Compressed sensing: theory and applications. Cambridge: Cambridge University Press, 2012.

Appendix

A. 1 Preliminaries

We first provide a brief overview of the results for the nodewise Lasso and Gaussian approximation in the following Propositions.

**Proposition A. 1** (Theorem 1 of Ref. [ 4 ], Asymptotic normality ). Suppose that  $\hat{\boldsymbol{\theta}}$  is the nodewise Lasso estimator with regularization parameters  $\lambda_j \asymp \sqrt{\frac{\log p}{n}}$  uniformly in  $j$ . Then, for every  $(i, j) \in E$  and  $z \in \mathbb{R}$ , it holds

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \mathcal{M}(s)} |\mathbb{P}(\sqrt{n}(\check{\boldsymbol{\theta}}_{ij} - \boldsymbol{\theta}_{ij})/\sigma_{ij} \leq z) - \Phi(z)| = 0.$$

**Proposition A.2** (Lemma 2 of Ref. [4], Variance estimation). Suppose that  $\lim_{n \rightarrow \infty} \log^4(p \vee n)/n^{1-\epsilon} = 0$  for some  $\epsilon > 0$ . Let  $\hat{\boldsymbol{\theta}}$  be the nodewise Lasso estimator and let  $\lambda_j \geq c\tau \sqrt{\log p/n}$  uniformly in  $j$  for some  $\tau, c > 0$ . Let  $\hat{\sigma}_{ij}^2 = \sum_{k=1}^n (\hat{\boldsymbol{\theta}}_i^T \mathbf{X}_{(k)} \mathbf{X}_{(k)}^T \hat{\boldsymbol{\theta}}_j)^2/n - \hat{\boldsymbol{\theta}}_{ij}^2$ . Then for all  $\eta > 0$

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \mathcal{M}(s)} \mathbb{P}(\max_{i,j=1,\dots,p} |\hat{\sigma}_{ij}^2 - \sigma_{ij}^2| \geq \eta) = 0.$$

**Proposition A.3** (Corollary 2.1 of Ref. [9], Gaussian approximation). Let  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T \in \mathbb{R}^p$ ,  $\mathbf{X} := (X_1, \dots, X_p)' := \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i$ ,  $T_0 := \max_{1 \leq j \leq p} X_j$  and  $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})^T \in \mathbb{R}^p$ ,  $\mathbf{Y} := (Y_1, \dots, Y_p)'$  is the Gaussian analog of  $X$  in the sense of sharing the same mean and covariance matrix, namely  $E[\mathbf{X}] = E[\mathbf{Y}] = \mathbf{0}$  and

$E[\mathbf{X}\mathbf{X}'] = E[\mathbf{Y}\mathbf{Y}'] = n^{-1} \sum_{i=1}^n E[\mathbf{x}_i \mathbf{x}_i']$ . Define the Gaussian analog  $Z_0$  of  $T_0$  as the maximum coordinate of vector  $\mathbf{Y}$ :  $Z_0 := \max_{1 \leq j \leq n} Y_j$ . Let  $B_n \geq 1$  be a sequence of constants and  $\lim_{n \rightarrow \infty} B_n = \infty$ . Suppose that there exist constants  $c_1 > 0$ ,  $C_1 > 0$ ,  $c_2 > 0$  and  $C_2 > 0$  such that the following condition (E.1) is satisfied:

$$c_1 \leq \sum_{i=1}^n \mathbb{E} x_{ij}^2/n \leq C_1,$$

$$\max_{r=1,2} \sum_{i=1}^n \mathbb{E}(|x_{ij}|^{2+r}/B_n^r)/n + \mathbb{E}(\exp(|x_{ij}|/B_n)) \leq 4,$$

and  $B_n^4(\log(pn))^7/n \leq C_2 n^{-c_2}$ . Then there exist constants  $c > 0$  and  $C > 0$  depending only on  $c_1, C_1, c_2$ , and  $C_2$  such that

$$\rho := \sup_{t \in \mathbb{R}} |P(T_0 \leq t) - P(Z_0 \leq t)| \leq C n^{-c} \rightarrow 0.$$

## A.2 Proof of Theorem 2.1

Without loss of generality, we set  $E = [p] \times [p]$ . For any  $(j, k) \in E$ , define

$$T_E := \max_{(j,k) \in E} \sqrt{n}(\check{\boldsymbol{\theta}}_{jk} - \boldsymbol{\theta}_{jk}), W_E := \max_{(j,k) \in E} \sum_{i=1}^n \hat{Z}_{ijk} e_i / \sqrt{n},$$

$$T_0 := \max_{(j,k) \in E} \sum_{i=1}^n Z_{ijk} / \sqrt{n}, W_0 := \max_{(j,k) \in E} \sum_{i=1}^n Z_{ijk} e_i / \sqrt{n},$$

where  $\hat{Z}_{ijk} = \hat{\boldsymbol{\theta}}_j^T \mathbf{X}_{(i)} \mathbf{X}_{(i)}^T \hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{jk}$  and  $Z_{ijk} = \boldsymbol{\theta}_j^T \mathbf{X}_{(i)} \mathbf{X}_{(i)}^T \boldsymbol{\theta}_k$ .

$\boldsymbol{\theta}_k - \boldsymbol{\theta}_{jk}$ . By following the same arguments in the proof of Lemma 3 and Lemma 4, we have

$$\mathbb{P}(c_{1-\alpha, W} \leq c_{(1-\alpha+\xi_2), W_0} + \xi_1) \geq 1 - \xi_2,$$

$$\mathbb{P}(c_{1-\alpha, W_0} \leq c_{1-\alpha+\pi(\nu), W}) \geq 1 - \mathbb{P}(\Gamma > \nu).$$

Let  $\pi(\nu) := C_2 \nu^{1/3} (1 \vee \log(|E|/\nu))^{2/3}$ ,  $\kappa_1(\nu) = c_{1-\alpha-\xi_2-\pi(\nu), Y_0}$  and  $\kappa_2(\nu) = c_{1-\alpha+\xi_2+\pi(\nu), Y_0}$ . Then by Lemma A.1, A.3 and invoking the similar proof of Corollary 3.1 in Ref. [9], we have  $\sup_{\alpha \in (0,1)} |\mathbb{P}(T_E < c_{1-\alpha, W_E}) - (1-\alpha)| \leq \sup_{\alpha \in (0,1)} \mathbb{P}((T_E < c_{1-\alpha, W_E}) \ominus (T_0 < c_{1-\alpha, Y_0})) \leq \mathbb{P}(\kappa_1(\nu) - 2\xi_1 < T_0 \leq \kappa_2(\nu) + 2\xi_1) + 2\mathbb{P}(\Gamma > \nu) + 3\xi_2 \leq \mathbb{P}(\kappa_1(\nu) - 2\xi_1 < Y_0 \leq \kappa_2(\nu) + 2\xi_1) + 2\mathbb{P}(\Gamma > \nu) + 3\xi_2 + 2Cn^{-c} \leq 2\pi(\nu) + 2\mathbb{P}(\Gamma > \nu) + 2Cn^{-c} + C_2\xi_1 \sqrt{1 \vee \log(|E|/\xi_1)} + 5\xi_2 \leq o(1)$ , where  $A \ominus B$  denotes their symmetric difference, that is,  $A \ominus B = (A \setminus B) \cup (B \setminus A)$ . Thus we get

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \mathcal{M}(s)} \sup_{\alpha \in (0,1)} |(\mathbb{P}\sqrt{n} \|\check{\boldsymbol{\theta}}_E - \boldsymbol{\theta}_E\|_{\max} \leq c_{1-\alpha, E}) - (1-\alpha)| = 0,$$

which conclude the proof.

## A.3 Proof of Theorem 2.3

To enhance the readability, we split the proof into three steps by providing the bound on bias term, establishing asymptotic normality and verifying the variance consistency.

### Step 1

$$\begin{aligned} \check{\boldsymbol{\theta}} - \boldsymbol{\theta} &= \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^T (\hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\theta}} - \mathbf{I}_p) - \boldsymbol{\theta} \\ &= -\boldsymbol{\theta} (\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \boldsymbol{\theta} - (\boldsymbol{\theta} \hat{\boldsymbol{\Sigma}} - \mathbf{I}_p) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T (\hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\theta}} - \mathbf{I}_p) \\ &:= \mathbf{Z} + \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2. \end{aligned}$$

For the first part bias, we have  $\|\boldsymbol{\Delta}_1\|_{\max} = \|(\boldsymbol{\theta} \hat{\boldsymbol{\Sigma}} - \mathbf{I}_p) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\|_{\max} \leq \|\boldsymbol{\theta} \hat{\boldsymbol{\Sigma}} - \mathbf{I}_p\|_{\max} \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_1 = O_p(\text{slogp}/n)$ , which is the result of event  $H$  and  $\max_{k \in [p]}$

$\|\hat{\boldsymbol{\Sigma}} \boldsymbol{\theta}_k - \mathbf{e}_k\|_{\infty} \leq \|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max} \|\boldsymbol{\theta}_k\|_1 = O_p(\sqrt{\log p/n})$  by Lemma 6. For the second part bias, we have  $\|\boldsymbol{\Delta}_2\|_{\max} = \|(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T (\hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\theta}} - \mathbf{I}_p)\|_{\max} \leq \|\hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\theta}} - \mathbf{I}_p\|_{\max} \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_1 = O_p(\text{slogp}/n)$ . Therefore, by condition A2, we have bias term  $\|\boldsymbol{\Delta}\|_{\max} = \|\sqrt{n}(\boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2)\|_{\max} \leq \sqrt{n}(\|\boldsymbol{\Delta}_1\|_{\max} + \|\boldsymbol{\Delta}_2\|_{\max}) = O_p(\text{slogp}/\sqrt{n}) = o_p(1)$ , which complete the first step proof.

**Step 2** The proof is the direct conclusion of Theorem 1 of Ref. [4].



**Step 3** The proof is the direct conclusion of Lemma 2 of Ref. [4].

The subsequent proof is similar to Theorem 2.1, thus we omit the details.

#### A.4 Lemmas and their proofs

The following lemmas will be used in the proof of the main theorem.

**Lemma A.1** Assume that conditions (A1) - (A4) hold. Then for any  $E \subseteq [p] \times [p]$  we have

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\max_{(j,k) \in E} \sum_{i=1}^n Z_{ijk} / \sqrt{n} \leq t\right) - \mathbb{P}\left(\max_{(j,k) \in E} \sum_{i=1}^n Y_{ijk} / \sqrt{n} \leq t\right) \right| \leq C_0 n^{-c_0},$$

where  $\{Y_{ijk}\}_{(j,k) \in E}$  are Gaussian analogs of  $\{Z_{ijk}\}_{(j,k) \in E}$  in the sense of sharing the same mean and covariance for  $i = 1, 2, \dots, n$ .

**Proof** The proof is based upon verifying conditions from Corollary 2.1 of Ref. [9]. To be concrete, we require to prove the following condition (E.1).

$$c_1 \leq \sum_{i=1}^n \mathbb{E} Z_{ijk}^2 / n \leq C_1, \\ \max_{r=1,2} \sum_{i=1}^n \mathbb{E}(|Z_{ijk}|^{2+r} / B_n^r) / n + \mathbb{E}(\exp(|Z_{ijk}| / B_n)) \leq 4,$$

where  $B_n \geq 1$  be a sequence of constants and  $\lim_{n \rightarrow \infty} B_n = \infty$ . By sub-Gaussian setting and condition A1, we have  $\|\boldsymbol{\Theta}\|_2 \leq \lambda_{\max}(\boldsymbol{\Theta}) \leq L$  and  $K = O(1)$ . Then invoking Lemma 8, we have a moment bound for  $r \geq 2$ ,  $\mathbb{E}|Z_{ijk}|^r \leq (2L^2 K^2)^r r! / 2$ . Then for  $r = 2$ , there exist  $c_1$  and  $C_1$  such that  $c_1 \leq \mathbb{E} Z_{ijk}^2 \leq C_1$ . On

the other hand,  $\mathbb{E} \exp(|Z_{ijk}| / B_n) = 1 + \sum_{r=1}^{\infty} \frac{\mathbb{E}|Z_{ijk}|^r}{B_n^r r!} \leq 1 + \sum_{r=1}^{\infty} \frac{(C')^r r!}{2B_n^r r!} \leq 1 + \sum_{r=1}^{\infty} \left(\frac{C'}{B_n}\right)^r < \infty$ , where

we use the fact that  $B_n$  is some sufficiently large constant. Thus we have

$$\max_{r=1,2} \mathbb{E}|Z_{ijk}|^{2+r} / B_n^r + \mathbb{E} \exp(|Z_{ijk}| / B_n) \leq 4,$$

which conclude the proof.

**Lemma A.2** Let  $\mathbf{V}$  and  $\mathbf{Y}$  be centered Gaussian random vectors in  $\mathbb{R}^p$  with covariance matrices  $\boldsymbol{\Sigma}^{\mathbf{V}}$  and  $\boldsymbol{\Sigma}^{\mathbf{Y}}$  respectively. Suppose that there are some constants  $0 < c_1 < C_1$  such that  $c_1 \leq \boldsymbol{\Sigma}_{jj}^{\mathbf{V}} \leq C_1$  for all  $1 \leq j \leq p$ . Then there exists a constant  $C >$

0 depending only on  $c_1$  and  $C_1$  such that

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(\max_{1 \leq j \leq p} V_j \leq t) - \mathbb{P}(\max_{1 \leq j \leq p} Y_j \leq t) \right| \leq C \Delta_0^{1/3} (1 \vee \log(p / \Delta_0))^{2/3},$$

where  $\Delta_0 := \max_{1 \leq j, k \leq p} |\boldsymbol{\Sigma}_{jk}^{\mathbf{V}} - \boldsymbol{\Sigma}_{jk}^{\mathbf{Y}}|$ .

**Proof** The proof is the same as Lemma 3.1 of Ref. [9]

**Lemma A.3** Suppose that there are some constants  $0 < c_1 < C_1$  such that  $c_1 \leq \sum_{i=1}^n \mathbb{E} Z_{ijk}^2 / n \leq C_1$  for all  $(j, k) \in E$ . Let  $Y_0 = \max_{(j,k) \in E} \sum_{i=1}^n Y_{ijk} / \sqrt{n}$ , where  $Y_{ijk}$  is the same with that in Lemma 1. Then for every  $\alpha \in (0, 1)$ ,

$$\mathbb{P}(c_{1-\alpha, W_0} \leq c_{1-\alpha+\pi(\nu), Y_0}) \geq 1 - \mathbb{P}(\Gamma > \nu),$$

$$\mathbb{P}(c_{1-\alpha, Y_0} \leq c_{1-\alpha+\pi(\nu), W_0}) \geq 1 - \mathbb{P}(\Gamma > \nu),$$

where  $\pi(\nu) := C_2 \nu^{1/3} (1 \vee \log(|E| / \nu))^{2/3}$ .

**Proof** Recall that  $\Gamma = \max_{(j,k) \in E} \sum_{i=1}^n Z_{ijk} Z_{ijk'} / n -$

$\sum_{i=1}^n \mathbb{E}(Z_{ijk} Z_{ijk'}) / n$ . By Lemma A.2, on the event  $\{X: \Gamma \leq \nu\}$ , we have  $|\mathbb{P}(Y_0 \leq t) - \mathbb{P}(W_0 \leq t | X)| \leq \pi(\nu)$ , for all  $t \in \mathbb{R}$ . Thus condition on this event we have  $\mathbb{P}(W_0 \leq c_{1-\alpha+\pi(\nu), Y_0}) \geq \mathbb{P}(c_{1-\alpha+\pi(\nu), Y_0} - \pi(\nu) \geq \alpha + \pi(\nu) - \pi(\nu) = \alpha)$ , which conclude the first proof. The second claim follows similarly.

**Lemma A.4** Assume that conditions (A1) - (A4) hold. Then for any  $(j, k) \in E$  we have

$$\mathbb{P}(|T_E - T_0| > \xi_1) < \xi_2,$$

$$\mathbb{P}(\mathbb{P}_e(|W_E - W_0| > \xi_1) > \xi_2) < \xi_2,$$

where  $\xi_1 = o(1)$ ,  $\xi_2 = o(1)$  and  $\xi_1 \sqrt{\log p / n} + \xi_2 \leq C_2 n^{-c_2}$ .

**Proof**

**Bounds for  $|T_E - T_0|$ :** Recall that

$$|T_E - T_0| \leq \max_{(j,k) \in E} |\Delta_{jk}|$$

It follows from Theorem 2.1 that

$$\mathbb{P}\left\{\max_{(j,k) \in E} |\Delta_{jk}| \geq O\left(\frac{\log p}{n}\right)\right\} \leq o(1),$$

where  $\xi_1 = O(\log p / n) = o(1)$  and  $\xi_2 = o(1)$ .

**Bounds for  $|W_E - W_0|$ :**

$$|W_E - W_0| \leq \max_{(j,k) \in E} \left| \sum_{i=1}^n (\hat{Z}_{ijk} - Z_{ijk}) e_i / \sqrt{n} \right|.$$

Let  $A_n = \sum_{i=1}^n (\hat{Z}_{ijk} - Z_{ijk}) e_i / \sqrt{n}$ , we have

$$\mathbb{E}(A_n) \leq \mathbb{E}_X \sqrt{\mathbb{E}_e \left[ \sum_{i=1}^n (\hat{Z}_{ijk} - Z_{ijk}) e_i / \sqrt{n} \right]^2} \leq$$

$$\mathbb{E}_X \sqrt{\sum_{i=1}^n (\hat{Z}_{ijk} - Z_{ijk})^2 / n} \leq \sqrt{\mathbb{E}_X \left[ \sum_{i=1}^n (\hat{Z}_{ijk} - Z_{ijk})^2 / n \right]}.$$

By Lemma A.5,  $\sum_{i=1}^n (\hat{Z}_{ijk} - Z_{ijk})^2 / n = o(1)$ , we directly get  $\mathbb{E}(A_n) \leq o(1)$ . Thus, we have  $A_n = o(1)$  by applying Markov's inequality. Further, there exist  $\xi_1 = o(1)$  and  $\xi_2 = o(1)$  such that  $\mathbb{P}(|W_E - W_0| > \xi_1) < \xi_2$ . By Markov's inequality, we have

$$\begin{aligned} \mathbb{P}(\mathbb{P}_e(|W_E - W_0| > \xi_1) > \xi_2) &\leq \\ \mathbb{E}[\mathbb{P}_e(|W_E - W_0| > \xi_1)] / \xi_2 &= \\ \mathbb{P}(|W_E - W_0| > \xi_1) / \xi_2 &\leq \xi_2^2 / \xi_2 = \xi_2, \end{aligned}$$

which conclude the proof.

#### Lemma A.5

$$\max_{(j,k) \in E} \sum_{i=1}^n (\hat{Z}_{ijk} - Z_{ijk})^2 / n = o_p(1).$$

#### Proof

Since  $(a-b)^2 \leq 2(a^2 + b^2)$ , we have  $\max_{(j,k) \in E} \sum_{i=1}^n (\hat{Z}_{ijk} - Z_{ijk})^2 / n = \max_{(j,k) \in E} \sum_{i=1}^n (\hat{\Theta}_j^T X_{(i)} X_{(i)}^T \hat{\Theta}_k - \hat{\Theta}_{jk} - \Theta_j^T X_{(i)} X_{(i)}^T \Theta_k + \Theta_{jk})^2 / n \leq 2 \max_{(j,k) \in E} \left[ \sum_{i=1}^n (\hat{\Theta}_j^T X_{(i)} X_{(i)}^T \hat{\Theta}_k - \Theta_j^T X_{(i)} X_{(i)}^T \Theta_k)^2 \right] + 2 \max_{(j,k) \in E} (\hat{\Theta}_{jk} - \Theta_{jk})^2 =: 2I_1 + 2I_2$ .

For the first part, it follows from triangle inequality that  $I_1 \leq \max_{(j,k) \in E} \sum_{i=1}^n (\hat{\Theta}_j^T X_{(i)} X_{(i)}^T (\hat{\Theta}_k - \Theta_k))^2 / n + \max_{(j,k) \in E} \sum_{i=1}^n ((\hat{\Theta}_j - \Theta_j)^T X_{(i)} X_{(i)}^T \Theta_k)^2 / n$ . Since  $\left\| \sum_{i=1}^n X_{(i)} X_{(i)}^T \hat{\Theta}_j \hat{\Theta}_j^T X_{(i)} X_{(i)}^T / n \right\|_{\max} \leq \max_{i \in [n]} \|X_{(i)} X_{(i)}^T\|_{\max}$   $\max_{j \in [p]} \sum_{i=1}^n \hat{\Theta}_j^T X_{(i)} X_{(i)}^T \hat{\Theta}_j / n \leq O_p(\log(np))$   $\max_{j \in [p]} \|\hat{\Theta}_j\|_1 \|\hat{\Sigma}\|_{\infty} \leq O_p(\log(np))$   $\max_{j \in [p]} (\|\Theta_j\|_1 + \|\hat{\Theta}_j - \Theta_j\|_1) (1 + O_p(\sqrt{\log p/n})) = O_p(\log(np))$ , we have  $\max_{(j,k) \in E} \sum_{i=1}^n$

$$\begin{aligned} (\hat{\Theta}_j^T X_{(i)} X_{(i)}^T (\hat{\Theta}_k - \Theta_k))^2 / n &= \max_{(j,k) \in E} (\hat{\Theta}_k - \Theta_k)^T \sum_{i=1}^n X_{(i)} X_{(i)}^T \\ \hat{\Theta}_j \hat{\Theta}_j^T X_{(i)} X_{(i)}^T (\hat{\Theta}_k - \Theta_k) / n &\leq \max_{k \in [p]} \|\hat{\Theta}_k - \Theta_k\|_2^2 O_p(\log(np)) = O_p(s^2 \log \log(np) / n) = o_p(1). \end{aligned}$$

For the second part, it is obvious that

$$|I_2| \leq O_p(\log p / n),$$

which is a direct result of event  $H$ .

Combining them together, we conclude  $= O_p(s^2 \log \log(np) / n) = o_p(1)$ .

**Lemma A.6** Assume that conditions (A1) -

(A4) hold. Let

$$\|\hat{\Sigma} - \Sigma\|_{\max} \leq O_p(\sqrt{\log p / n}).$$

**Proof** The proof is the same as Lemma L.3 of Ref. [13], which follows by invoking the inequality

$$\|X_i X_j\|_{\psi_1} \leq 2 \|X_i\|_{\psi_2} \|X_j\|_{\psi_2} \leq 2c^{-2}$$

and Proposition 5.16 in Ref. [17] and the union bound.

**Lemma A.7** Let  $\{X_i\}_{i=1}^n$  be identically  $p$ -dimensional sub-Gaussian vectors with  $\max_{i \in [n], j \in [p]} \|X_{ij}\|_{\psi_2} = C$ . Then we have

$$\max_{i \in [n]} \|X_{(i)} X_{(i)}^T\|_{\max} < O_p(\log(np)).$$

**Proof** The proof is the same as Lemma L.5 of Ref. [13]. It follows from the fact  $\max_{i \in [n]} \|X_{(i)} X_{(i)}^T\|_{\max} = \max_{i \in [n]} \|X_i\|_{\infty}^2$  and (5.10) in Ref. [17] with the union bound.

**Lemma A.8** Let  $\alpha, \beta \in \mathbb{R}^p$  such that  $\|\alpha\|_2 \leq M, \|\beta\|_2 \leq M$ . Let  $X_i$  satisfy the sub-Gaussian setting with a positive constant  $K$ . Then for any  $r \geq 2$ , we have

$$\begin{aligned} \mathbb{E} |\alpha^T X_{(i)} X_{(i)}^T \beta - \mathbb{E} \alpha^T X_{(i)} X_{(i)}^T \beta|^r / \\ (2M^2 K^2)^r \leq r! / 2. \end{aligned}$$

**Proof** The proof is the same as Lemma 5 of Ref. [3]. Since  $\|\alpha\|_2 \leq M, \|\beta\|_2 \leq M$  and sub-Gaussian assumption with a constant  $K$ , we obtain

$$\mathbb{E} e^{|\alpha^T X_{(i)} X_{(i)}^T \beta|^2 / (MK)^2} \leq 2 \text{ and } \mathbb{E} e^{|\alpha^T X_{(i)} X_{(i)}^T \beta|^2 / (MK)^2} \leq 2.$$

By the inequality  $ab \leq a^2/2 + b^2/2$  (for any  $a, b \in \mathbb{R}$ ) and Cauchy-Schwarz inequality we have

$$\begin{aligned} \mathbb{E} e^{|\alpha^T X_{(i)} X_{(i)}^T \beta|^r / (MK)^2} &\leq \mathbb{E} e^{|\alpha^T X_{(i)} X_{(i)}^T \beta|^2 / (MK)^2 / 2} \mathbb{E} e^{|\alpha^T X_{(i)} X_{(i)}^T \beta|^2 / (MK)^2 / 2} \leq \\ \{\mathbb{E} e^{|\alpha^T X_{(i)} X_{(i)}^T \beta|^2 / (MK)^2}\}^{1/2} \{\mathbb{E} e^{|\alpha^T X_{(i)} X_{(i)}^T \beta|^2 / (MK)^2}\}^{1/2} &\leq 2. \end{aligned}$$

By the Taylor expansion, we have the inequality

$$\begin{aligned} 1 + \frac{1}{r!} \mathbb{E} |\alpha^T X_{(i)} X_{(i)}^T \beta|^r / (MK)^{2r} &\leq \\ \mathbb{E} e^{|\alpha^T X_{(i)} X_{(i)}^T \beta|^r / (MK)^2} &. \end{aligned}$$

Next it follows

$$\begin{aligned} \mathbb{E} |\alpha^T X_{(i)} X_{(i)}^T \beta - \mathbb{E} \alpha^T X_{(i)} X_{(i)}^T \beta|^r / (MK)^{2r} &\leq \\ 2^{r-1} \mathbb{E} |\alpha^T X_{(i)} X_{(i)}^T \beta|^r / (MK)^{2r} &\leq \\ 2^{r-1} r! (\mathbb{E} e^{|\alpha^T X_{(i)} X_{(i)}^T \beta|^r / (MK)^2} - 1) &= 2^{r-1} r! = \frac{r!}{2} 2^r. \end{aligned}$$

Therefore, we have

$$\mathbb{E} |\alpha^T X_{(i)} X_{(i)}^T \beta - \mathbb{E} \alpha^T X_{(i)} X_{(i)}^T \beta|^r / (2M^2 K^2)^r \leq \frac{r!}{2}.$$