Supplement Materials: On Block Cholesky Decomposition for Sparse Inverse Covariance Estimation

Xiaoning Kang[†], Jiayi Lian[‡] and Xinwei Deng[‡]

[‡]*Institute of Supply Chain Analytics and International Business College, Dongbei University of Finance and Economics, Dalian, China*

[‡]*Department of Statistics, Virginia Tech, Blacksburg, VA*

Supplementary Material

In the supplementary materials, we provide all the technical proofs for the main results of the paper. Before proving theorems, we present several lemmas.

Lemma 1. *Denote a squared block diagonal matrix by* $\boldsymbol{D} = diag(\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_M)$*. Suppose* $\boldsymbol{D}_i$ *have eigenvalues* $\mathcal{C}_{\lambda_i} = \{\lambda_{ip_1}, \lambda_{ip_2}, \ldots, \lambda_{ip_i}\}, i = 1, 2, \ldots, M$*, then the eigenvalues of matrix* $\boldsymbol{D}$ *are* $\mathcal{C}_{\lambda_1}, \mathcal{C}_{\lambda_2}, \ldots, \mathcal{C}_{\lambda_M}$*.*

Proof. Let $\boldsymbol{D}_i = \boldsymbol{P}_i \boldsymbol{\Lambda}_i \boldsymbol{P}_i^{-1}$ be the eigenvalue decomposition, where $\boldsymbol{\Lambda}_i = \text{diag}(\lambda_{ip_1}, \lambda_{ip_2}, \ldots, \lambda_{ip_i})$, and $\boldsymbol{P}_i$ is composed of the corresponding eigenvectors. Define $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \ldots, \boldsymbol{\Lambda}_M)$ and $\boldsymbol{P} = \text{diag}(\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_M)$. Then we have

$$\boldsymbol{D}\boldsymbol{P} = \text{diag}(\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_M)\text{diag}(\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_M)$$

$$
\begin{aligned}
&= \mathrm{diag}(\boldsymbol{D}_1\boldsymbol{P}_1, \boldsymbol{D}_2\boldsymbol{P}_2, \ldots, \boldsymbol{D}_M\boldsymbol{P}_M) \\
&= \mathrm{diag}(\boldsymbol{P}_1\boldsymbol{\Lambda}_1, \boldsymbol{P}_2\boldsymbol{\Lambda}_2, \ldots, \boldsymbol{P}_M\boldsymbol{\Lambda}_M) \\
&= \mathrm{diag}(\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_M)\mathrm{diag}(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \ldots, \boldsymbol{\Lambda}_M) \\
&= \boldsymbol{P}\boldsymbol{\Lambda},
\end{aligned}
$$

which indicates $\boldsymbol{D} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}$, and establishes the lemma. □

Lemma 1 describes a property of eigenvalues for the block diagonal matrix. The following Lemma 2 is from Theorem A.10 in Bai and Silverstein (2010). It demonstrates the property of matrix singular values. Its result is stated here for completeness.

Lemma 2. *Let $\boldsymbol{B}$ and $\boldsymbol{C}$ be two matrices of order $m_1 \times m_2$ and $m_2 \times m_3$. For any $i, j \geq 0$, we have*

$$\varphi_{i+j+1}(\boldsymbol{B}\boldsymbol{C}) \leq \varphi_{i+1}(\boldsymbol{B})\varphi_{j+1}(\boldsymbol{C}).$$

Based on the results of Lemmas 1 and 2, we present the following Lemma 3, which provides a relationship between matrix $\boldsymbol{\Omega}$ and its block Cholesky factor matrices $(\boldsymbol{T}^{-1}, \boldsymbol{D}^{-1})$ in terms of their singular values.

Lemma 3. *Let $\boldsymbol{\Omega} = \boldsymbol{T}'\boldsymbol{D}^{-1}\boldsymbol{T}$ be the block MCD of the inverse covariance matrix.*

If the condition (3.8) is satisfied, that is, there exists a constant $\theta > 0$ *such that* $1/\theta < \varphi_p(\mathbf{\Omega}) \le \varphi_1(\mathbf{\Omega}) < \theta$*, then there exist constants* h_1 *and* h_2 *such that*

$$0 < h_1 < \varphi_p(\boldsymbol{T}^{-1}) \le \varphi_1(\boldsymbol{T}^{-1}) < h_2 < \infty,$$

and

$$0 < h_1 < \varphi_p(\boldsymbol{D}^{-1}) \le \varphi_1(\boldsymbol{D}^{-1}) < h_2 < \infty.$$

Proof. By the decomposition (2.1), we partition $\mathbf{\Omega}$ into blocks according to the variable groups $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots, \boldsymbol{X}^{(M)}$ such that its diagonal blocks are $\mathbf{\Omega}_{ii}$ of order $p_i \times p_i, i = 1, 2, \ldots, M$, and $\sum_{i=1}^{M} p_i = p$. Write $\mathbf{\Omega} = \boldsymbol{T}'\boldsymbol{D}^{-1}\boldsymbol{T} = \boldsymbol{T}'\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{T} = \boldsymbol{R}'\boldsymbol{R}$, where

$$\boldsymbol{R} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{T} = \begin{pmatrix} \boldsymbol{R}_{11} & \mathbf{0} & \ldots & \mathbf{0} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{R}_{M1} & \boldsymbol{R}_{M2} & \ldots & \boldsymbol{R}_{MM} \end{pmatrix}$$

with $\boldsymbol{R}_{ii} = \boldsymbol{D}_i^{-\frac{1}{2}}$. Note that $\boldsymbol{R}_{ii}$ is a symmetric matrix due to the symmetry of $\boldsymbol{D}_i$. In addition, it is obvious to have $\mathbf{\Omega}_{ii} = \sum_{i \ge k} \boldsymbol{R}'_{ik}\boldsymbol{R}_{ik}$, implying that $\mathbf{\Omega}_{ii} - \boldsymbol{R}'_{ii}\boldsymbol{R}_{ii} =$

$\boldsymbol{\Omega}_{ii} - \boldsymbol{D}_i^{-1}$ is semi-positive definite. Consequently we have

$$\varphi_p(\boldsymbol{D}_i^{-1}) \leq \varphi_1(\boldsymbol{D}_i^{-1}) \leq \varphi_1(\boldsymbol{\Omega}_{ii}) \leq \theta. \tag{S0.1}$$

Taking determinant on both sides of $\boldsymbol{\Omega} = \boldsymbol{T}'\boldsymbol{D}^{-1}\boldsymbol{T}$ yields

$$\varphi_p(\boldsymbol{\Omega}) \cdots \varphi_1(\boldsymbol{\Omega}) = \varphi_p(\boldsymbol{D}^{-1}) \cdots \varphi_1(\boldsymbol{D}^{-1}).$$

By $\varphi_1(\boldsymbol{D}_i^{-1}) \leq \theta$ for each $i = 1, 2, \ldots, M$ via (S0.1), together with Lemma 1, it is easy to see $\varphi_1(\boldsymbol{D}^{-1}) \leq \theta$. We hence have

$$(\frac{1}{\theta})^p \leq \varphi_p^p(\boldsymbol{\Omega}) \leq \prod_{i=1}^{p} \varphi_i(\boldsymbol{\Omega}) = \prod_{i=1}^{p} \varphi_i(\boldsymbol{D}^{-1}) \leq \theta^{p-1}\varphi_p(\boldsymbol{D}^{-1}),$$

which gives $\varphi_p(\boldsymbol{D}^{-1}) \geq (\frac{1}{\theta})^{2p-1}$. As a result,

$$0 < (\frac{1}{\theta})^{2p-1} \leq \varphi_p(\boldsymbol{D}^{-1}) \leq \varphi_1(\boldsymbol{D}^{-1}) \leq \theta < \infty.$$

To bound singular values of matrix $\boldsymbol{T}^{-1}$, on one hand, we use Lemma 2 to obtain $\varphi_p(\boldsymbol{\Omega}) = \varphi_p(\boldsymbol{T}'\boldsymbol{D}^{-1}\boldsymbol{T}) = \varphi_p(\boldsymbol{T}\boldsymbol{T}'\boldsymbol{D}^{-1}) \leq \varphi_p(\boldsymbol{T}\boldsymbol{T}')\varphi_1(\boldsymbol{D}^{-1}) = \varphi_p(\boldsymbol{T}')\varphi_p(\boldsymbol{T})\varphi_1(\boldsymbol{D}^{-1})$, indicating

$$\varphi_p(\boldsymbol{T}) \geq \sqrt{\varphi_p(\boldsymbol{\Omega})/\varphi_1(\boldsymbol{D}^{-1})} \geq \sqrt{1/\theta^2} = \frac{1}{\theta}.$$

On the other hand, applying Lemma 2 again for $\boldsymbol{D}^{-1} = \boldsymbol{T}'^{-1}\boldsymbol{\Omega}\boldsymbol{T}^{-1}$ yields $\varphi_p(\boldsymbol{D}^{-1}) \leq \varphi_p^2(\boldsymbol{T}^{-1})\varphi_1(\boldsymbol{\Omega}) = \varphi_1(\boldsymbol{\Omega})/\varphi_1^2(\boldsymbol{T})$, implying

$$\varphi_1(\boldsymbol{T}) \leq \sqrt{\frac{\varphi_1(\boldsymbol{\Omega})}{\varphi_p(\boldsymbol{D}^{-1})}} \leq \sqrt{\frac{\theta}{1/\theta^{2p-1}}} = \theta^p.$$

As a result,

$$0 < \frac{1}{\theta} \leq \varphi_p(\boldsymbol{T}) \leq \varphi_1(\boldsymbol{T}) \leq \theta^p < \infty$$
$$0 < (\frac{1}{\theta})^p \leq \varphi_p(\boldsymbol{T}^{-1}) \leq \varphi_1(\boldsymbol{T}^{-1}) \leq \theta < \infty.$$

Taking $h_1 = \min(\theta^{1-2p}, \theta^{-p})$ and $h_2 = \theta$ establishes the lemma. □

It is seen from Lemma 3 that the singular values of the matrices $\boldsymbol{T}^{-1}$ and $\boldsymbol{D}^{-1}$ are bounded if the singular values of the inverse covariance matrix $\boldsymbol{\Omega}$ are bounded. Now we give the proofs of Theorems.

Proof. **Proof of Theorem 1**.

From the negative log-likelihood (2.4), we have

$$L(\boldsymbol{T}, \boldsymbol{D}) = -\sum_{j=1}^{M} \log |\boldsymbol{D}_j^{-1}| + \sum_{j=1}^{M} \operatorname{tr}\left[\boldsymbol{S}_{\epsilon_j}\boldsymbol{D}_j^{-1}\right]$$

$$= \sum_{j=1}^{M} \log|\boldsymbol{D}_j| + \mathrm{tr}\begin{pmatrix} \boldsymbol{S}_{\epsilon_1}\boldsymbol{D}_1^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_{\epsilon_2}\boldsymbol{D}_2^{-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{S}_{\epsilon_M}\boldsymbol{D}_M^{-1} \end{pmatrix}$$

$$= \log|\boldsymbol{D}| + \mathrm{tr}\begin{pmatrix} \boldsymbol{S}_{\epsilon_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_{\epsilon_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{S}_{\epsilon_M} \end{pmatrix}\boldsymbol{D}^{-1}.$$

By the notation $\boldsymbol{S}_{\epsilon_j} = \frac{1}{n}(\mathbb{X}^{(j)} - \mathbb{Z}^{(j)}\boldsymbol{A}_j')'(\mathbb{X}^{(j)} - \mathbb{Z}^{(j)}\boldsymbol{A}_j')$, it is easy to see

$$L(\boldsymbol{T}, \boldsymbol{D}) = \log|\boldsymbol{D}| + \frac{1}{n}\mathrm{tr}\begin{pmatrix} (\mathbb{X}^{(1)})' \\ (\mathbb{X}^{(2)} - \mathbb{Z}^{(2)}\boldsymbol{A}_2')' \\ \vdots \\ (\mathbb{X}^{(M)} - \mathbb{Z}^{(M)}\boldsymbol{A}_M')' \end{pmatrix}$$

$$\left(\mathbb{X}^{(1)}, \mathbb{X}^{(2)} - \mathbb{Z}^{(2)}\boldsymbol{A}_2', \dots, \mathbb{X}^{(M)} - \mathbb{Z}^{(M)}\boldsymbol{A}_M'\right)\boldsymbol{D}^{-1}$$

$$= \log|\boldsymbol{D}| + \frac{1}{n}\mathrm{tr}\left[\boldsymbol{T}\mathbb{X}'\mathbb{X}\boldsymbol{T}'\boldsymbol{D}^{-1}\right]$$

$$= \log|\boldsymbol{D}| + \mathrm{tr}\left[\boldsymbol{T}'\boldsymbol{D}^{-1}\boldsymbol{T}\boldsymbol{S}\right],$$

where $\boldsymbol{S} = \frac{1}{n}\mathbb{X}'\mathbb{X}$. Consequently, $L_\lambda(\boldsymbol{T}, \boldsymbol{D})$ can be written as

$$\begin{aligned}
L_\lambda(\boldsymbol{T}, \boldsymbol{D}) &= \log|\boldsymbol{D}| + \text{tr}\left[\boldsymbol{T}'\boldsymbol{D}^{-1}\boldsymbol{T}\boldsymbol{S}\right] + \lambda_1\|\boldsymbol{A}\|_1 + \lambda_2\|\boldsymbol{D}^{-1}\|_1^- \\
&= \log|\boldsymbol{D}| + \text{tr}\left[\boldsymbol{T}'\boldsymbol{D}^{-1}\boldsymbol{T}\boldsymbol{S}\right] + \lambda_1\|\boldsymbol{T}\|_1 + \lambda_2\|\boldsymbol{D}^{-1}\|_1^- \\
&= \log|\boldsymbol{D}| + \text{tr}\left[\boldsymbol{T}'\boldsymbol{D}^{-1}\boldsymbol{T}\boldsymbol{S}\right] + \lambda_1\sum_{i>k}|t_{ik}| + \lambda_2\sum_{i\neq k}|\psi_{ik}|,
\end{aligned}$$

where t_{ik} and ψ_{ik} are the (i,k)th elements of matrices $\boldsymbol{T}$ and $\boldsymbol{D}^{-1}$, respectively.

For part (a), we define $G_1(\Delta_T) = L_\lambda(\boldsymbol{T}_0 + \Delta_T|\boldsymbol{D}_*) - L_\lambda(\boldsymbol{T}_0|\boldsymbol{D}_*)$. Let $\mathcal{A}_{U_j} = \{\Delta_{T_j} : \Delta_{T_j} = \Delta'_{T_j}, \|\Delta_{T_j}\|_F^2 \leq U_j^2 s_{T_j}\log(\sum_{k=1}^j p_k)/n\}$ for $j = 1, 2, \ldots, M$, where U_j are positive constants. We will show that for $\Delta_{T_j} \in \partial\mathcal{A}_{U_j}$, probability $\Pr(G_1(\Delta_T)) > 0$ is tending to 1 as $n \to \infty$ for sufficiently large U_j, where $\partial\mathcal{A}_{U_j}$ are the boundaries of $\mathcal{A}_{U_j}$. Additionally, since $G_1(\Delta_T) = 0$ when $\Delta_{T_j} = 0$, the minimum point of $G_1(\Delta_T)$ is achieved when $\Delta_{T_j} \in \mathcal{A}_{U_j}$. That is $\|\Delta_{T_j}\|_F^2 = O_p(s_{T_j}\log(\sum_{k=1}^j p_k)/n)$.

Assume $\|\Delta_{T_j}\|_F^2 = U_j^2 s_{T_j}\log(\sum_{k=1}^j p_k)/n$. Write $\boldsymbol{T} = \boldsymbol{T}_0 + \Delta_T$, then we decompose $G_1(\Delta_T)$ as

$$\begin{aligned}
G_1(\Delta_T) &= L_\lambda(\boldsymbol{T}_0 + \Delta_T|\boldsymbol{D}_*) - L_\lambda(\boldsymbol{T}_0|\boldsymbol{D}_*) \\
&= \text{tr}\left[\boldsymbol{T}'\boldsymbol{D}_*^{-1}\boldsymbol{T}\boldsymbol{S}\right] - \text{tr}\left[\boldsymbol{T}_0'\boldsymbol{D}_*^{-1}\boldsymbol{T}_0\boldsymbol{S}\right] + \lambda_1\sum|t_{ik}| - \lambda_1\sum|t_{0ik}| \\
&= M_1 + M_2 + M_3,
\end{aligned}$$

where

$$M_1 = \mathrm{tr}[\boldsymbol{D}_*^{-1}(\boldsymbol{T}(\boldsymbol{S}-\boldsymbol{\Sigma}_0)\boldsymbol{T}' - \boldsymbol{T}_0(\boldsymbol{S}-\boldsymbol{\Sigma}_0)\boldsymbol{T}_0')],$$

$$M_2 = \mathrm{tr}[\boldsymbol{D}_*^{-1}(\boldsymbol{T}\boldsymbol{\Sigma}_0\boldsymbol{T}' - \boldsymbol{T}_0\boldsymbol{\Sigma}_0\boldsymbol{T}_0')],$$

$$M_3 = \lambda_1 \sum |t_{ik}| - \lambda_1 \sum |t_{0ik}|.$$

The above decomposition of $G_1(\Delta_T)$ into M_1 to M_3 is very similar to that in the proof of Lemma 3 of Kang and Deng (2021); hence it is omitted here. Now we bound each component respectively. Note that $\|\Delta_T\|_F^2 = \|\boldsymbol{T}-\boldsymbol{T}_0\|_F^2 = \sum_{j=1}^M \|\Delta_{T_j}\|_F^2$. Therefore, based on the proof of Theorem 3.1 in Jiang (2012), for any $\epsilon > 0$, there exists a constant $V_1 > 0$ such that with probability greater than $1-\epsilon$, we have

$$\begin{aligned}
&M_2 - |M_1| \\
&> \frac{\|\Delta_T\|_F^2}{h^4} - V_1 \sum_{j=1}^{M} \left(||\boldsymbol{T}_j - \boldsymbol{T}_{j0}||_1 \sqrt{\log(\sum_{k=1}^{j} p_k)/n} \right) \\
&= \frac{\sum_{j=1}^M \|\Delta_{T_j}\|_F^2}{h^4} - V_1 \sum_{j=1}^{M} \left(\sqrt{\log(\sum_{k=1}^{j} p_k)/n} \sum_{(i,k)\in \mathcal{Z}_{T_j}^c} |t_{ik}| \right) \\
&\qquad\qquad - V_1 \sum_{j=1}^{M} \left(\sqrt{\log(\sum_{k=1}^{j} p_k)/n} \sum_{(i,k)\in \mathcal{Z}_{T_j}} |t_{ik} - t_{0ik}| \right) \\
&\geq \frac{\sum_{j=1}^M \|\Delta_{T_j}\|_F^2}{h^4} - V_1 \sqrt{\log(p)/n} \sum_{(i,k)\in \bigcup_{j=1}^M \mathcal{Z}_{T_j}^c} |t_{ik}| - V_1 \sum_{j=1}^{M} \sqrt{s_{T_j} \log(\sum_{k=1}^{j} p_k)/n \|\Delta_{T_j}\|_F^2}
\end{aligned}$$

$$= \frac{1}{h^4}\sum_{j=1}^{M} U_j^2 s_{T_j} \log(\sum_{k=1}^{j} p_k)/n - V_1\sqrt{\log(p)/n} \sum_{(i,k)\in\bigcup_{j=1}^{M}\mathcal{Z}_{T_j}^c} |t_{ik}|$$

$$- V_1 \sum_{j=1}^{M} U_j s_{T_j} \log(\sum_{k=1}^{j} p_k)/n$$

$$\geq \frac{1}{nh^4}\sum_{j=1}^{M} U_j^2 s_{T_j}(\log\gamma_j + C_j \log p) - V_1\sqrt{\log(p)/n} \sum_{(i,k)\in\bigcup_{j=1}^{M}\mathcal{Z}_{T_j}^c} |t_{ik}| - V_1 \frac{\log p}{n}\sum_{j=1}^{M} U_j s_{T_j}$$

$$\geq \frac{1}{nh^4}\sum_{j=1}^{M} U_j^2 s_{T_j}\log\gamma_j + \frac{1}{h^4}\frac{\log p}{n}\tau_c\tau_u \sum_{j=1}^{M} U_j s_{T_j} - V_1\sqrt{\log(p)/n} \sum_{(i,k)\in\bigcup_{j=1}^{M}\mathcal{Z}_{T_j}^c} |t_{ik}|$$

$$- V_1 \frac{\log p}{n}\sum_{j=1}^{M} U_j s_{T_j},$$

where τ_u is a positive constant satisfying $\tau_u \leq U_j, j = 1, 2, \ldots, M$. Next, for the penalty term corresponding to λ_1,

$$M_3 = \lambda_1 \sum_{(i,k)\in\bigcup_{j=1}^{M}\mathcal{Z}_{T_j}^c} |t_{ik}| + \lambda_1 \sum_{(i,k)\in\bigcup_{j=1}^{M}\mathcal{Z}_{T_j}} (|t_{ik}| - |t_{0ik}|) = M_3^{(1)} + M_3^{(2)},$$

where $M_3^{(1)} = \lambda_1 \sum_{(i,k)\in\bigcup_{j=1}^{M}\mathcal{Z}_{T_j}^c} |t_{ik}|$, and

$$|M_3^{(2)}| = |\lambda_1 \sum_{(i,k)\in\bigcup_{j=1}^{M}\mathcal{Z}_{T_j}} (|t_{ik}| - |t_{0ik}|)| \leq \lambda_1 \sum_{(i,k)\in\bigcup_{j=1}^{M}\mathcal{Z}_{T_j}} |t_{ik} - t_{0ik}|$$

$$\leq \lambda_1 \sum_{j=1}^{M} \|\Delta_{T_j}\|_F \sqrt{s_{T_j}}$$

$$\leq \lambda_1 \sqrt{\frac{\log p}{n}} \sum_{j=1}^{M} U_j s_{T_j}.$$

Combine all the terms above together, with probability greater than $1-\epsilon$, we have

$$\begin{aligned}
&G_1(\Delta_T) \geq M_2 - |M_1| + M_3^{(1)} - |M_3^{(2)}| \\
&\geq \frac{1}{nh^4} \sum_{j=1}^{M} U_j^2 s_{T_j} \log \gamma_j + \frac{1}{h^4} \frac{\log p}{n} \tau_c \tau_u \sum_{j=1}^{M} U_j s_{T_j} - V_1 \sqrt{\log(p)/n} \sum_{(i,k) \in \bigcup_{j=1}^{M} \mathcal{Z}_{T_j}^c} |t_{ik}| \\
&\quad - V_1 \frac{\log p}{n} \sum_{j=1}^{M} U_j s_{T_j} + \lambda_1 \sum_{(i,k) \in \bigcup_{j=1}^{M} \mathcal{Z}_{T_j}^c} |t_{ik}| - \lambda_1 \sqrt{\frac{\log p}{n}} \sum_{j=1}^{M} U_j s_{T_j} \\
&= \frac{1}{nh^4} \sum_{j=1}^{M} U_j^2 s_{T_j} \log \gamma_j + \frac{\log p \sum_{j=1}^{M} U_j s_{T_j}}{n} \left(\frac{\tau_c \tau_u}{h^4} - V_1 - \frac{\lambda_1}{\sqrt{\log(p)/n}} \right) \\
&\quad + (\lambda_1 - V_1 \sqrt{\log(p)/n}) \sum_{(i,k) \in \bigcup_{j=1}^{M} \mathcal{Z}_{T_j}^c} |t_{ik}|.
\end{aligned}$$

Here V_1 is only related to the sample size n and ϵ. Assume $\lambda_1 = K_1 \sqrt{\log(p)/n}$ where $K_1 > V_1$, and choose $\tau_u > h^4(K_1 + V_1)/\tau_c$, then $G_1(\Delta_T) > 0$. Therefore, we prove $\|\Delta_{T_j}\|_F^2 = O_p(s_{T_j} \log(\sum_{k=1}^{j} p_k)/n)$.

The proof of part (b) follows the same principle as that for part (a). Similarly, define $G_2(\Delta_D) = L_\lambda(\boldsymbol{D}_0 + \Delta_D | \boldsymbol{T}_*) - L_\lambda(\boldsymbol{D}_0 | \boldsymbol{T}_*)$. Let $\mathcal{B}_{W_j} = \{\Delta_{D_j} : \Delta_{D_j} = \Delta_{D_j}', \|\Delta_{D_j}\|_F^2 \leq W_j^2 (s_{D_j} + p_j) \log(p_j)/n\}$ for $j = 1, 2, \ldots, M$, where W_j are positive constants. We only need to show that for $\Delta_{D_j} \in \partial \mathcal{B}_{W_j}$, probability $P(G_2(\Delta_D) > 0)$ is tending to 1 as $n \to \infty$ for sufficiently large W_j, where $\partial \mathcal{B}_{W_j}$ are the boundaries of

$\mathcal{B}_{W_j}$.

Assume $\|\Delta_{D_j}\|_F^2 = W_j^2(s_{D_j} + p_j)\log(p_j)/n$. Write $\boldsymbol{D} = \boldsymbol{D}_0 + \Delta_D$, then we decompose $G_2(\Delta_D)$ as

$$
\begin{aligned}
G_2(\Delta_D) &= L_\lambda(\boldsymbol{D}_0 + \Delta_D|\boldsymbol{T}_*) - L_\lambda(\boldsymbol{D}_0|\boldsymbol{T}_*)\\
&= \log|\boldsymbol{D}| - \log|\boldsymbol{D}_0| + \operatorname{tr}\left[\boldsymbol{T}_*'\boldsymbol{D}^{-1}\boldsymbol{T}_*\boldsymbol{S} - \boldsymbol{T}_*'\boldsymbol{D}_0^{-1}\boldsymbol{T}_*\boldsymbol{S}\right] + \lambda_2\sum_{i\neq k}|\psi_{ik}| - \lambda_2\sum_{i\neq k}|\psi_{0ik}|\\
&= M_4 + M_5 + M_6,
\end{aligned}
$$

where

$$
\begin{aligned}
M_4 &= \log|\boldsymbol{D}| - \log|\boldsymbol{D}_0| + \operatorname{tr}[(\boldsymbol{D}^{-1} - \boldsymbol{D}_0^{-1})\boldsymbol{D}_0],\\
M_5 &= \operatorname{tr}(\boldsymbol{D}^{-1} - \boldsymbol{D}_0^{-1})[\boldsymbol{T}_*(\boldsymbol{S} - \boldsymbol{\Sigma}_0)\boldsymbol{T}_*'],\\
M_6 &= \lambda_2\sum_{i\neq k}|\psi_{ik}| - \lambda_2\sum_{i\neq k}|\psi_{0ik}|.
\end{aligned}
$$

The above decomposition of $G_2(\Delta_D)$ into M_4 to M_6 is similar to that in the proof of Lemma 3 of Kang and Deng (2021). Next we bound each component respectively. Note that $\|\Delta_D\|_F^2 = \|\boldsymbol{D} - \boldsymbol{D}_0\|_F^2 = \sum_{j=1}^M \|\boldsymbol{D}_j - \boldsymbol{D}_{j0}\|_F^2 = \sum_{j=1}^M \|\Delta_{D_j}\|_F^2$. Therefore, based on the proof of Theorem 3.1 in Jiang (2012) together with Lemma 3, we can have the following two results (I) and (II).

(I) Let τ_w be a positive constant satisfying $\tau_w \leq W_j, j = 1, 2, \ldots, M$, and note that $(1/h^2)\|\Delta_D\|_F^2 \leq \|\boldsymbol{D}^{-1} - \boldsymbol{D}_0^{-1}\|_F^2 \leq h^2\|\Delta_D\|_F^2$, then

$$
\begin{aligned}
M_4 &\geq \frac{1}{8h^2}\|\boldsymbol{D}^{-1} - \boldsymbol{D}_0^{-1}\|_F^2 \geq \frac{1}{8h^4}\|\Delta_D\|_F^2 = \frac{1}{8h^4}\sum_{j=1}^{M}\|\Delta_{D_j}\|_F^2 \\
&= \frac{1}{8h^4}\sum_{j=1}^{M} W_j^2(s_{D_j} + p_j)\log(p_j)/n \\
&= \frac{1}{8h^4}\sum_{j=1}^{M} W_j^2(s_{D_j} + p_j)(\log\gamma_j + C_j\log p)/n \\
&\geq \frac{1}{8nh^4}\sum_{j=1}^{M} W_j^2(s_{D_j} + p_j)\log\gamma_j + \frac{1}{8h^4}\frac{\log p}{n}\tau_c\tau_w\sum_{j=1}^{M} W_j(s_{D_j} + p_j).
\end{aligned}
$$

(II) For any $\epsilon > 0$, there exists a constant $V_2 > 0$ such that with probability greater than $1 - \epsilon$, we have

$$
\begin{aligned}
|M_5| = |\mathrm{tr}(\boldsymbol{D}^{-1} - \boldsymbol{D}_0^{-1})[\boldsymbol{T}_*(\boldsymbol{S} - \boldsymbol{\Sigma}_0)\boldsymbol{T}_*']| &\leq \max|\xi_{ik}|\sum_{j=1}^{M}||\boldsymbol{D}_j - \boldsymbol{D}_{j0}||_1 \\
&\leq V_2\sqrt{\frac{\log p}{n}}\sum_{j=1}^{M}\sqrt{(s_{D_j} + p_j)\|\Delta_{D_j}\|_F^2} \\
&\leq V_2\frac{\log p}{n}\sum_{j=1}^{M} W_j(s_{D_j} + p_j),
\end{aligned}
$$

where ξ_{ik} is the (i,k)th element of matrix $\boldsymbol{T}_*(\boldsymbol{S} - \boldsymbol{\Sigma}_0)\boldsymbol{T}_*'$, and the second inequality applies Lemma 3 of Lam and Fan (2009).

Next, we decompose $M_6 = M_6^{(1)} + M_6^{(2)}$, where $M_6^{(1)} = \lambda_2\sum_{(i,k)\in\bigcup_{j=1}^{M}\mathcal{Z}_{D_j}^c}|\psi_{ik}|$, and

$$
\begin{aligned}
|M_6^{(2)}| \le |\lambda_2 \sum_{(i,k)\in\bigcup_{j=1}^M \mathcal{Z}_{D_j}} (|\psi_{ik}| - |\psi_{0ik}|)| &\le \lambda_2 \sum_{(i,k)\in\bigcup_{j=1}^M \mathcal{Z}_{D_j}} |\psi_{ik} - \psi_{0ik}| \\
&\le \lambda_2 \sum_{j=1}^M \|\Delta_{D_j}\|_F \sqrt{s_{D_j} + p_j} \\
&\le \lambda_2 \sqrt{\frac{\log p}{n}} \sum_{j=1}^M W_j (s_{D_j} + p_j).
\end{aligned}
$$

Combine all the terms above together, with probability greater than $1-\epsilon$, we have

$$
\begin{aligned}
& G_2(\Delta_D) \ge M_4 - |M_5| + M_6^{(1)} - |M_6^{(2)}| \\
\ge & \frac{1}{8nh^4} \sum_{j=1}^M W_j^2 (s_{D_j} + p_j) \log \gamma_j + \frac{1}{8h^4} \frac{\log p}{n} \tau_c \tau_w \sum_{j=1}^M W_j (s_{D_j} + p_j) \\
& - V_2 \frac{\log p}{n} \sum_{j=1}^M W_j (s_{D_j} + p_j) + \lambda_2 \sum_{(i,k)\in\bigcup_{j=1}^M \mathcal{Z}_{D_j}^c} |\psi_{ik}| - \lambda_2 \sqrt{\frac{\log p}{n}} \sum_{j=1}^M W_j (s_{D_j} + p_j) \\
= & \frac{1}{8nh^4} \sum_{j=1}^M W_j^2 (s_{D_j} + p_j) \log \gamma_j + \frac{\log p \sum_{j=1}^M W_j (s_{D_j} + p_j)}{n} \left(\frac{\tau_c \tau_w}{8h^4} - V_2 - \frac{\lambda_2}{\sqrt{\log(p)/n}}\right) \\
& + \lambda_2 \sum_{(i,k)\in\bigcup_{j=1}^M \mathcal{Z}_{D_j}^c} |\psi_{ik}|.
\end{aligned}
$$

Here V_2 is only related to the sample size n and ϵ. Assume $\lambda_2 = K_2 \sqrt{\log(p)/n}$ where $K_2 > 0$, and choose $\tau_w > 8h^4(K_2 + V_1)/\tau_c$, then $G_2(\Delta_D) > 0$. Therefore, we prove $\|\Delta_{D_j}\|_F^2 = W_j^2 (s_{D_j} + p_j) \log(p_j)/n$. □

Proof. **Proof of Theorem 2**.

Let $\hat{\boldsymbol{A}}_j$ and $\hat{\boldsymbol{D}}_j$ be the estimates obtained from Step 3 in Algorithm 1. We first prove the consistent rates under Frobenius norm of $\hat{\boldsymbol{T}}_j = -\hat{\boldsymbol{A}}_j$ and $\hat{\boldsymbol{D}}_j$ are $s_{T_j} \log(\sum_{k=1}^{j} p_k)/n$ and $(s_{D_j} + p_j)\log(p_j)/n$, respectively.

At the first iteration of Step 1 in Algorithm 1, the estimate $\hat{\boldsymbol{A}}_{j;1}$ found by minimizing $\ell_\lambda(\boldsymbol{A}_j|\boldsymbol{I})$ is $s_{T_j} \log(\sum_{k=1}^{j} p_k)/n$ consistent according to part (a) of Theorem 1. In Step 2, the estimate $\hat{\boldsymbol{D}}_{j;1}$ found by minimizing $\ell_\lambda(\boldsymbol{D}_j^{-1}|\hat{\boldsymbol{A}}_{j;1})$ is $(s_{D_j}+p_j)\log(p_j)/n$ consistent according to part (b) of Theorem 1. Next, an estimate $\hat{\boldsymbol{A}}_{j;2}$ obtained by minimizing $\ell_\lambda(\boldsymbol{A}_j|\hat{\boldsymbol{D}}_{j;1})$ is $s_{T_j} \log(\sum_{k=1}^{j} p_k)/n$ consistent, and $\hat{\boldsymbol{D}}_{j;2}$ which minimizes $\ell_\lambda(\boldsymbol{D}_j^{-1}|\hat{\boldsymbol{A}}_{j;2})$ is $(s_{D_j} + p_j)\log(p_j)/n$ consistent. Following this, we hence have that $\hat{\boldsymbol{A}}_j$ (or equivalently $\hat{\boldsymbol{T}}_j$) and $\hat{\boldsymbol{D}}_j$ are $s_{T_j} \log(\sum_{k=1}^{j} p_k)/n$ and $(s_{D_j} + p_j)\log(p_j)/n$ consistent. This implies

$$\|\hat{\boldsymbol{T}} - \boldsymbol{T}_0\|_F^2 = \sum_{j=1}^{M} ||\hat{\boldsymbol{T}}_j - \boldsymbol{T}_0||_F^2 = \sum_{j=1}^{M} O_p(s_{T_j} \log(\sum_{k=1}^{j} p_k)/n) \le O_p(s_T \log(p)/n)$$

and

$$\|\hat{\boldsymbol{D}} - \boldsymbol{D}_0\|_F^2 = \sum_{j=1}^{M} O_p\left((s_{D_j} + p_j)\log(p_j)/n\right) = O_p(\sum_{j=1}^{M}(s_{D_j} + p_j)\log(p_j)/n).$$

Next, we derive of consistent rate of the estimate $\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{T}}'\hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{T}}$. Let $\boldsymbol{\Delta}_T = \hat{\boldsymbol{T}} - \boldsymbol{T}_0$

and $\boldsymbol{\Delta}_D = \hat{\boldsymbol{D}} - \boldsymbol{D}_0$, then we decompose $\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_F^2$ as

$$\begin{aligned}
&\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_F^2 \\
&= \|\hat{\boldsymbol{T}}'\hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{T}} - \boldsymbol{T}_0'\boldsymbol{D}_0^{-1}\boldsymbol{T}_0\|_F^2 \\
&= \|(\boldsymbol{\Delta}_T' + \boldsymbol{T}_0')\hat{\boldsymbol{D}}^{-1}(\boldsymbol{\Delta}_T + \boldsymbol{T}_0) - \boldsymbol{T}_0'\boldsymbol{D}_0^{-1}\boldsymbol{T}_0\|_F^2 \\
&\leq \|\boldsymbol{\Delta}_T'\hat{\boldsymbol{D}}^{-1}\boldsymbol{T}_0\|_F^2 + \|\boldsymbol{T}_0'\hat{\boldsymbol{D}}^{-1}\boldsymbol{\Delta}_T\|_F^2 + \|\boldsymbol{\Delta}_T'\hat{\boldsymbol{D}}^{-1}\boldsymbol{\Delta}_T\|_F^2 + \|\boldsymbol{T}_0'(\hat{\boldsymbol{D}}^{-1} - \boldsymbol{D}_0^{-1})\boldsymbol{T}_0\|_F^2.
\end{aligned}$$

Now we bound four terms separately. Use the symbol $\|\boldsymbol{A}\|$ to represent the spectral norm of matrix $\boldsymbol{A}$. Since $\|\boldsymbol{T}_0\| = O(1)$ and $\|\boldsymbol{D}_0\| = O(1)$ by Lemma 3, it is obvious that $\|\hat{\boldsymbol{D}}\| = \|\hat{\boldsymbol{D}} - \boldsymbol{D}_0 + \boldsymbol{D}_0\| \leq \|\boldsymbol{\Delta}_D\| + \|\boldsymbol{D}_0\| \leq \|\boldsymbol{\Delta}_D\|_F + \|\boldsymbol{D}_0\| = O_p(1)$. In addition, the single values of $\boldsymbol{\Omega}^{-1}$ are bounded since the single values of $\boldsymbol{\Omega}$ are bounded, which together with Lemma 3 leads to $\|\boldsymbol{D}_0^{-1}\| = O(1)$, hence similarly $\|\hat{\boldsymbol{D}}^{-1}\| = O_p(1)$. As a result, it is easy to obtain

$$\|\boldsymbol{\Delta}_T'\hat{\boldsymbol{D}}^{-1}\boldsymbol{T}_0\|_F^2 \leq \|\boldsymbol{\Delta}_T'\|_F^2\|\hat{\boldsymbol{D}}^{-1}\|\|\boldsymbol{T}_0\| = O_p(\|\boldsymbol{\Delta}_T\|_F^2),$$

and the second term $\|\boldsymbol{T}_0'\hat{\boldsymbol{D}}^{-1}\boldsymbol{\Delta}_T\|_F^2 = \|\boldsymbol{\Delta}_T'\hat{\boldsymbol{D}}^{-1}\boldsymbol{T}_0\|_F^2 = O_p(\|\boldsymbol{\Delta}_T\|_F^2)$. For the third term,

$$\|\boldsymbol{\Delta}_T'\hat{\boldsymbol{D}}^{-1}\boldsymbol{\Delta}_T\|_F^2 \leq \|\boldsymbol{\Delta}_T'\|_F^2\|\hat{\boldsymbol{D}}^{-1}\|\|\boldsymbol{\Delta}_T\|_F^2 = o_p(\|\boldsymbol{\Delta}_T\|_F^2).$$

For the fourth term,

$$\|\boldsymbol{T}_0'(\hat{\boldsymbol{D}}^{-1}-\boldsymbol{D}_0^{-1})\boldsymbol{T}_0\|_F^2 \leq \|\boldsymbol{T}_0'\|\|\hat{\boldsymbol{D}}^{-1}-\boldsymbol{D}_0^{-1}\|_F^2\|\boldsymbol{T}_0\| = O_p(\|\hat{\boldsymbol{D}}-\boldsymbol{D}_0\|_F^2).$$

Consequently, by the convergence rates of $\hat{\boldsymbol{T}}-\boldsymbol{T}_0$ and $\hat{\boldsymbol{D}}-\boldsymbol{D}_0$ from Theorem 1, we reach the conclusion

$$\begin{aligned}\|\hat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}_0\|_F^2 &= O_p(\|\hat{\boldsymbol{T}}-\boldsymbol{T}_0\|_F^2)+O_p(\|\hat{\boldsymbol{D}}-\boldsymbol{D}_0\|_F^2)\\ &= O_p\left(\frac{s_T\log p+\sum_{j=1}^M(s_{D_j}+p_j)\log p_j}{n}\right).\end{aligned}$$

□

References

Bai, Z., and Silverstein, J. (2010), *Spectral analysis of large dimensional random matrices* Springer Series in Statistics. Springer, New York.

Jiang, X. (2012), "Joint estimation of covariance matrix via Cholesky Decomposition," *Ph.D Dissertation*, Department of Statistics and Applied Probability, National University of Singapore.

Kang, X., and Deng, X. (2021), "On variable ordination of Cholesky-based estimation for a sparse covariance matrix," *Canadian Journal of Statistics*, 49(2), 283–310.

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