ADAPTIVE BLOCK BANDING PRECISION MATRIX ESTIMATION FOR MULTIVARIATE LONGITUDINAL DATA

Chunhui Liang[1], Wenqing Ma[2] and Yanyuan Ma[3]

[1] *Tianjin University of Commerce*

[2] *Capital Normal University*

[3] *The Pennsylvania State University*

Supplementary Material

In this supplement, we provide the detailed proofs of lemmas and theorems omitted from the body of this paper, and some simulation and real data studies with figures.

S1 Dual problem: Proof of Theorem 1.

Proof. Note that $(\sum_{m=1}^{l}\mathbf{\Psi}_{jm}^2)^{1/2} = \|(\mathbf{\Psi}_j)_{1:l}\|_2 = \max\{\langle \mathbf{A}_{\cdot,l}, \mathbf{\Psi}_j\rangle,\ \text{s.t. } \|\mathbf{A}_{1:l,l}\|_2 \leq 1, \mathbf{A}_{l+1:j,l} = \mathbf{0}\}$. The optimization problem (2.8) can be written as

$$\begin{aligned}
&\min_{\mathbf{\Psi}_j}\{\frac{\rho}{2}\|\mathbf{\Psi}_j - \mathbf{Z}_j\|_2^2 + \lambda_2\sum_{l=1}^{j-1}\|(\mathbf{\Psi}_j)_{1:l}\|_2\}\\
=&\min_{\mathbf{\Psi}_j}\{\max_{\mathbf{A}}\{\frac{\rho}{2}\|\mathbf{\Psi}_j - \mathbf{Z}_j\|_2^2 + \lambda_2\sum_{l=1}^{j-1}\langle \mathbf{A}_{\cdot,l}, \mathbf{\Psi}_j\rangle\}\},\ \text{s.t. } \|\mathbf{A}_{1:l,l}\|_2 \leq 1, \mathbf{A}_{l+1:j,l} = \mathbf{0},\\
=&\max_{\mathbf{A}}\{\min_{\mathbf{\Psi}_j}\{\frac{\rho}{2}\|\mathbf{\Psi}_j - \mathbf{Z}_j\|_2^2 + \lambda_2\sum_{l=1}^{j-1}\langle \mathbf{A}_{\cdot,l}, \mathbf{\Psi}_j\rangle\}\},\ \text{s.t. } \|\mathbf{A}_{1:l,l}\|_2 \leq 1, \mathbf{A}_{l+1:j,l} = \mathbf{0},
\end{aligned} \tag{S1.1}$$

where the last equality is due to Bien et al. (2016). Now the inner minimization in (S1.1) is a quadratic optimization problem and has explicit solution $\mathbf{\Psi}_j = \mathbf{Z}_j - \lambda_2 \sum_{l=1}^{j-1} \mathbf{A}_{\cdot,l}/\rho$. Inserting this into (S1.1) leads to

$$\max_{\mathbf{A}}\{\frac{\rho}{2}\|\frac{\lambda_2}{\rho}\sum_{l=1}^{j-1}\mathbf{A}_{\cdot,l}\|_2^2 + \lambda_2 \sum_{l=1}^{j-1}\mathbf{Z}_j^{\mathrm{T}}\mathbf{A}_{\cdot,l} - \frac{\lambda_2^2}{\rho}\|\sum_{l=1}^{j-1}\mathbf{A}_{\cdot,l}\|_2^2,$$
$$\text{s.t. } \|(\mathbf{A}_{1:l,l})\|_2 \le 1, \mathbf{A}_{l+1:j,l} = \mathbf{0}\}, \text{ for } l = 1, \dots, j-1,$$

which is equivalent to

$$\min_{\mathbf{A}} \|\mathbf{Z}_j - \frac{\lambda_2}{\rho}\sum_{l=1}^{j-1}\mathbf{A}_{\cdot,l}\|_2^2, \text{ s.t. } \|(\mathbf{A}_{1:l,l})\|_2 \le 1, \mathbf{A}_{l+1:j,l} = \mathbf{0}, \text{ for } l = 1, \dots, j-1,$$

that is the dual function of (2.8). □

S2 Convergence in Frobenius norm: Proof of Theorem 2.

To show this, we first establish two lemmas, which can be found in Lam and Fan (2009). Lemma S2.1 provides inequalities involving the operator and the Frobenius norms, and Lemma S2.2 provides the approximation error rate between sample variance and true variance.

Lemma S2.1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be real matrices such that the product* $\mathbf{AB}$ *is defined. Then*

$$\sigma_{\min}(\mathbf{A})\|B\|_F \le \|\mathbf{AB}\|_F \le \sigma_{\max}(\mathbf{A})\|B\|_F.$$

Specifically, if $\mathbf{A} = (a_{ij})$*, then* $|a_{ij}| \le \sigma_{\max}(\mathbf{A})$ *for each* i, j*.*

Lemma S2.2. *Let* $\mathbf{S}$ *be a sample covariance matrix based on a random sample*

$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})^{\mathrm{T}}$, $i = 1, \ldots, n$, *where* $\mathrm{E}(\mathbf{Y}_i) = \mathbf{0}$, $\mathrm{cov}(\mathbf{Y}_i) = \mathbf{\Sigma}$. *Let* $Y_{ij} \sim F_j$, *where* F_j *is the cumulative distribution function (cdf) of* Y_{ij}. *Let* G_j *be the cdf of* Y_{ij}^2 *and it satisfies*

$$\max_{1 \le j \le m} \int_0^\infty \exp(\psi t) dG_j(t) < \infty, \text{ for all } \psi \in (0, \psi_0),$$

where $\psi_0 > 0$ *is a constant. Assume that* $\log m/n = o(1)$ *and that* $\mathbf{\Sigma}$ *has its eigenvalues uniformly bounded when* $n \to \infty$. *Then, for constant matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *with bounded* $\sigma_{\max}(\mathbf{A})$, $\sigma_{\max}(\mathbf{B})$, *we have that* $\max_{i,j} |\{\mathbf{A}(\mathbf{S} - \mathbf{\Sigma})\mathbf{B}\}_{ij}| = O_p(\sqrt{\log m/n})$.

Remark S2.1. The conditions on Y_{ij} in Lemma S2.2 are the same as those used in Bickel and Levina (2008) for relaxing the normality assumption. Proofs of Lemma S2.1 and Lemma S2.2 can be found in Lam and Fan (2009).

Proof. **Part 1.** We first prove the first half of the theorem. The objective function (2.3) is a biconvex function of $\mathbf{W}^{-1}$ and $\mathbf{L}$ at any fixed $\boldsymbol{\lambda}$. At a fixed $\boldsymbol{\lambda}$, let $(\widehat{\mathbf{W}}^{-1}, \widehat{\mathbf{L}})$ be a local minimum of $Q(\mathbf{W}^{-1}, \mathbf{L}, \boldsymbol{\lambda})$. Define

$$\mathcal{A} \equiv \{\Delta : \Delta \in \mathcal{R}^{K \times K}, \Delta = \Delta^{\mathrm{T}}, \|\Delta\|_F = M r_{1n}\},$$

and

$$\mathcal{B} \equiv \{\Delta : \Delta \in \mathcal{R}^{J \times J}, \Delta_{j,j'} = 0, \text{ for } j' > j, (\Delta + \mathbf{L})_{j,j} > 0, \|\Delta_{j,1:j}\|_2 = N J^{-1/2} r_{2n}\},$$

where $M, N > 0$ are sufficiently large constants. For any $\Delta \in \mathcal{B}$, let $\Delta_j \equiv \Delta_{j,1:j}$. Then $\|\Delta\|_F^2 = \sum_{j=1}^J \|\Delta_j\|_2^2 = N^2 r_{2n}^2$, so $\|\Delta\|_F = N r_{2n}$ for any $\Delta \in \mathcal{B}$. Define $\eta_1 = \eta_2 \equiv (\log m/n)^{1/2}$, $r_{1n} \equiv \eta_1 \sqrt{K}$, $r_{2n} \equiv \eta_2 \sqrt{J}$. Note that $r_{1n} \to 0$, $r_{2n} \to 0$

under the assumptions of Theorem 2. Let $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{B}$, for each fixed $\Delta_1 \in \mathcal{A}$, and $\Delta_2 \in \mathcal{B}$, denote $\mathcal{C}_{\Delta_1} \equiv \{\Delta_2 \in \mathcal{B} | (\Delta_1, \Delta_2) \in \mathcal{C}\}$ and $\mathcal{C}_{\Delta_2} \equiv \{\Delta_1 \in \mathcal{A} | (\Delta_1, \Delta_2) \in \mathcal{C}\}$. Below, we will show that

$$P(\inf_{\Delta_1 \in \mathcal{C}_{\Delta_2}} \{Q(\mathbf{W}^{-1} + \Delta_1, \mathbf{L} + \Delta_2, \boldsymbol{\lambda})\} > Q(\mathbf{W}^{-1}, \mathbf{L} + \Delta_2, \boldsymbol{\lambda})) \to 1, \tag{S2.1}$$

and

$$P(\inf_{\Delta_2 \in \mathcal{C}_{\Delta_1}} \{Q(\mathbf{W}^{-1} + \Delta_1, \mathbf{L} + \Delta_2, \boldsymbol{\lambda})\} > Q(\mathbf{W}^{-1} + \Delta_1, \mathbf{L}, \boldsymbol{\lambda})) \to 1. \tag{S2.2}$$

We consider proving (S2.1) first. The objective function (2.3) leads to $(KJ)^{-1}\{Q(\mathbf{W}^{-1} + \Delta_1, \mathbf{L} + \Delta_2, \boldsymbol{\lambda}) - Q(\mathbf{W}^{-1}, \mathbf{L} + \Delta_2, \boldsymbol{\lambda})\} = T_1 + T_2 + T_3$, where

$$T_1 \equiv -\frac{1}{K}(\log|\mathbf{W}^{-1} + \Delta_1| - \log|\mathbf{W}^{-1}|) + \frac{1}{K}\mathrm{tr}(\mathbf{W}\Delta_1),$$

$$T_2 \equiv \frac{\lambda_1}{K}\{\|\mathrm{vec}(\mathbf{W}^{-1} + \Delta_1)\|_1 - \|\mathrm{vec}(\mathbf{W}^{-1})\|_1\},$$

$$T_3 \equiv \frac{1}{nKJ}\sum_{i=1}^{n} \mathrm{tr}\{\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i(\mathbf{L} + \Delta_2)^{\mathrm{T}}(\mathbf{L} + \Delta_2)\} - \frac{1}{K}\mathrm{tr}(\mathbf{W}\Delta_1).$$

Let $f(s) \equiv \log|\mathbf{W}^{-1} + s\Delta_1|$. Appendix I of Yu and Bien (2017) showed that the first and second derivatives of $f(s)$ are respectively $f'(s) = \mathrm{tr}\{(\mathbf{W}^{-1} + s\Delta_1)^{-1}\Delta_1\}$, and $f''(s) = -\mathrm{vec}(\Delta_1)^{\mathrm{T}}\{(\mathbf{W}^{-1} + s\Delta_1)^{-1} \otimes (\mathbf{W}^{-1} + s\Delta_1)^{-1}\}\mathrm{vec}(\Delta_1)$. By Taylor's expansion at $s = 0$, we can obtain that $\log|\mathbf{W}^{-1} + \Delta_1| - \log|\mathbf{W}^{-1}| = \mathrm{tr}(\mathbf{W}\Delta_1) - \mathrm{vec}(\Delta_1)^{\mathrm{T}}\left\{\int_0^1 (1-s)(\mathbf{W}^{-1} + s\Delta_1)^{-1} \otimes (\mathbf{W}^{-1} + s\Delta_1)^{-1} ds\right\}\mathrm{vec}(\Delta_1)$. Thus,

$$
\begin{aligned}
T_1 &= \frac{1}{K}\mathrm{vec}(\Delta_1)^{\mathrm{T}}\left\{\int_0^1 (1-s)(\mathbf{W}^{-1}+s\Delta_1)^{-1}\otimes(\mathbf{W}^{-1}+s\Delta_1)^{-1}ds\right\}\mathrm{vec}(\Delta_1)\\
&\geq \frac{1}{K}\|\mathrm{vec}(\Delta_1)\|_2^2\lambda_{\min}\left\{\int_0^1 (1-s)(\mathbf{W}^{-1}+s\Delta_1)^{-1}\otimes(\mathbf{W}^{-1}+s\Delta_1)^{-1}ds\right\}\\
&\geq \frac{1}{K}\|\mathrm{vec}(\Delta_1)\|_2^2\int_0^1 (1-s)\min_{0\leq s\leq 1}\lambda_{\min}^2\{(\mathbf{W}^{-1}+s\Delta_1)^{-1}\}ds\\
&\geq \frac{1}{2K}\|\mathrm{vec}(\Delta_1)\|_2^2\min_{0\leq s\leq 1}\lambda_{\min}^2\{(\mathbf{W}^{-1}+s\Delta_1)^{-1}\}\\
&\geq \frac{1}{2K}\|\mathrm{vec}(\Delta_1)\|_2^2\min\{\lambda_{\max}^{-2}(\mathbf{W}^{-1}+\widetilde{\Delta}):\|\widetilde{\Delta}\|_F\leq Mr_{1n},\widetilde{\Delta}=\widetilde{\Delta}^{\mathrm{T}}\},
\end{aligned}
$$

where $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote the minimum and maximum eigenvalues of a generic real square matrix $\mathbf{A}$. Note that $\mathbf{W}^{-1}$ is positive definite and $\|\widetilde{\Delta}\|_F = o(1)$, hence $\mathbf{W}^{-1}+\widetilde{\Delta}$ is also positive definite. Therefore, following Lemma S2.1, $\lambda_{\max}^{-2}(\mathbf{W}^{-1}+\widetilde{\Delta}) \geq \{\sigma_{\max}(\mathbf{W}^{-1})+\sigma_{\max}(\widetilde{\Delta})\}^{-2} \geq 1/\{2\sigma_{\max}^2(\mathbf{W}^{-1})\}$ for sufficiently large n, hence

$$
T_1 \geq \frac{1}{4K}\|\Delta_1\|_F^2\sigma_{\min}^2(\mathbf{W}) = \frac{1}{4K}M^2r_{1n}^2\sigma_{\min}^2(\mathbf{W}). \tag{S2.3}
$$

We define $\mathcal{S}_{\mathbf{W}^{-1}} \equiv \{(l,l') : \mathbf{W}_{l,l'}^{-1} \neq 0, l,l' \in \{1,\cdots,K\}\}$ and $\mathcal{S}_{\mathbf{W}^{-1}}^c \equiv \{(l,l') : \mathbf{W}_{l,l'}^{-1} = 0, l,l' \in \{1,\cdots,K\}\}$. Let the cardinality of $\mathcal{S}_{\mathbf{W}^{-1}}$ be $|\mathcal{S}_{\mathbf{W}^{-1}}| = w+K$. Because $\|\mathrm{vec}(\mathbf{W}^{-1})\|_1 = \|\mathrm{vec}\{(\mathbf{W}^{-1})_{\mathcal{S}_{\mathbf{W}^{-1}}}\}\|_1$, and $\|\mathrm{vec}(\mathbf{W}^{-1}+\Delta_1)\|_1 = \|\mathrm{vec}\{(\mathbf{W}^{-1}+\Delta_1)_{\mathcal{S}_\Omega}\}\|_1 + \|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}^{(c)}_{\mathbf{W}^{-1}}}\}\|_1$, following the triangle inequality, and $w = O(K)$, we get

$$
\begin{aligned}
T_2 =& \frac{\lambda_1}{K}[\|\mathrm{vec}\{(\mathbf{W}^{-1}+\Delta_1)_{\mathcal{S}_\Omega}\}\|_1 + \|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}^{(c)}_{\mathbf{W}^{-1}}}\}\|_1 - \|\mathrm{vec}\{(\mathbf{W}^{-1})_{\mathcal{S}_{\mathbf{W}^{-1}}}\}\|_1] \\
\geq & \frac{\lambda_1}{K}[\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}^c_{\mathbf{W}^{-1}}}\}\|_1 - \|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_{\mathbf{W}^{-1}}}\}\|_1] \\
\geq & \frac{\lambda_1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}^c_{\mathbf{W}^{-1}}}\}\|_1 - \frac{\lambda_1}{K}(w+K)^{1/2}\|\Delta_1\|_F \\
\geq & \frac{\lambda_1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}^c_{\mathbf{W}^{-1}}}\}\|_1 - \frac{\lambda_1}{\sqrt{K}}\|\Delta_1\|_F. \qquad \text{(S2.4)}
\end{aligned}
$$

Note that the above inequality of T_2 always holds for any $\lambda_1 > 0$. Write $T_3 = A + B + C$, where

$$
\begin{aligned}
A &\equiv \frac{1}{nKJ}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}) - \frac{1}{K}\mathrm{tr}(\mathbf{W}\Delta_1), \\
B &\equiv \frac{1}{nKJ}\sum_{i=1}^{n}\{\mathrm{tr}(\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\Delta_2) + \mathrm{tr}(\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i\Delta_2^{\mathrm{T}}\mathbf{L})\}, \\
C &\equiv \frac{1}{nKJ}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i\Delta_2^{\mathrm{T}}\Delta_2).
\end{aligned}
$$

Using $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}^{\mathrm{T}})$ and $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ for any square matrices $\mathbf{A}$ and $\mathbf{B}$, then we get $\mathrm{tr}(\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\Delta_2) = \mathrm{tr}\{(\Delta_2^{\mathrm{T}}\mathbf{L})(\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i)\} = \mathrm{tr}(\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i\Delta_2^{\mathrm{T}}\mathbf{L})$. Hence we obtain

$$
B = 2\sum_{i=1}^{n}\mathrm{tr}(\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\Delta_2)/(nKJ).
$$

From $\mathrm{tr}(\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\Delta_2) = \mathrm{vec}(\mathbf{Y}_i)^{\mathrm{T}}\mathrm{vec}(\Delta_1\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\Delta_2) = \mathrm{vec}(\mathbf{Y}_i)^{\mathrm{T}}(\Delta_2^{\mathrm{T}}\mathbf{L}\otimes\Delta_1)\mathrm{vec}(\mathbf{Y}_i) = \mathrm{tr}\{(\Delta_2^{\mathrm{T}}\mathbf{L}\otimes\Delta_1)\mathrm{vec}(\mathbf{Y}_i)\mathrm{vec}(\mathbf{Y}_i)^{\mathrm{T}}\}$, and $\mathrm{E}\{\mathrm{vec}(\mathbf{Y}_i)\mathrm{vec}(\mathbf{Y}_i)^{\mathrm{T}}\} = \mathrm{cov}\{\mathrm{vec}(\mathbf{Y}_i)\} = \mathbf{R}\otimes\mathbf{W}$, we have $\mathrm{E}\{\mathrm{tr}(\mathbf{Y}_i^{\mathrm{T}}\Delta_1\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\Delta_2)\} = \mathrm{tr}\{(\Delta_2^{\mathrm{T}}\mathbf{L}\otimes\Delta_1)(\mathbf{R}\otimes\mathbf{W})\} = \mathrm{tr}(\Delta_2^{\mathrm{T}}\mathbf{LR})\mathrm{tr}(\Delta_1\mathbf{W})$.

By law of large numbers, we get

$$\frac{B}{2} = \frac{1}{KJ}\mathrm{tr}(\Delta_2^{\mathrm{T}}\mathbf{L}\mathbf{R})\mathrm{tr}(\Delta_1\mathbf{W})\{1 + o_p(1)\}.$$

Noting that $\mathbf{W}$ is positive definite, we get $|\mathrm{tr}(\Delta_1\mathbf{W})| \leq \lambda_{\max}(\mathbf{W})\sqrt{K}\|\Delta_1\|_F$. Similarly, we also have

$$|\mathrm{tr}(\Delta_2^{\mathrm{T}}\mathbf{L}\mathbf{R})| \leq \lambda_{\max}(\mathbf{R})\sqrt{J}\|\Delta_2^{\mathrm{T}}\mathbf{L}\|_F \leq \lambda_{\max}(\mathbf{R})\sqrt{J}\sigma_{\max}(\mathbf{L})\|\Delta_2\|_F,$$

where the last inequality is due to Lemma S2.1. Note that $\Delta_2 \in \mathcal{B}$, so $\|\Delta_2\|_F = Nr_{2n}$. Thus,

$$\begin{aligned}
\frac{|B|}{2} \leq &(KJ)^{-1/2}\sigma_{\max}(\mathbf{L})\lambda_{\max}(\mathbf{W})\lambda_{\max}(\mathbf{R})\|\Delta_1\|_F\|\Delta_2\|_F\{1 + o_p(1)\} \\
= &(KJ)^{-1/2}\sigma_{\max}(\mathbf{L})\lambda_{\max}(\mathbf{W})\lambda_{\max}(\mathbf{R})\|\Delta_1\|_F Nr_{2n}\{1 + o_p(1)\} \\
= &O_p\{(KJ)^{-1/2}\|\Delta_1\|_F\eta_2\sqrt{J}\} \\
= &O_p(\eta_1\|\Delta_1\|_F/\sqrt{K}),
\end{aligned} \tag{S2.5}$$

where the last equality is due to Assumption C1 and $\eta_1 \asymp \eta_2$. Similarly, replacing $\mathbf{L}$ with Δ_2 in treating B, we get

$$
\begin{aligned}
|C| \leq & \frac{1}{KJ}\lambda_{\max}(\mathbf{W})\sqrt{K}Mr_{1n}\lambda_{\max}(\mathbf{R})\sqrt{J}\sigma_{\max}(\Delta_2)\|\Delta_2\|_F\{1+o_p(1)\} \\
\leq & \frac{1}{KJ}\lambda_{\max}(\mathbf{W})\sqrt{K}Mr_{1n}\lambda_{\max}(\mathbf{R})\sqrt{J}\|\Delta_2\|_F^2\{1+o_p(1)\} \\
\leq & (KJ)^{-1/2}\lambda_{\max}(\mathbf{W})\lambda_{\max}(\mathbf{R})MN^2 r_{1n}r_{2n}^2\{1+o_p(1)\} \\
= & O_p\{(KJ)^{-1/2}\|\Delta_1\|_F r_{2n}\eta_2\sqrt{J}\} \\
= & o_p(\eta_1\|\Delta_1\|_F/\sqrt{K}),
\end{aligned} \tag{S2.6}
$$

where the last equality is due to Assumption C1, $\eta_1 \asymp \eta_2$ and $r_{2n} = o(1)$. Finally,

$$
\begin{aligned}
|A| = & |\frac{1}{K}\mathrm{tr}\{(\frac{1}{nJ}\sum_{i=1}^{n}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{Y}_i^{\mathrm{T}} - \mathbf{W})\Delta_1\}| \\
\leq & \frac{1}{K}|\sum_{(s,t)\in\mathcal{S}_{\mathbf{W}^{-1}}}(\frac{1}{nJ}\sum_{i=1}^{n}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{Y}_i^{\mathrm{T}} - \mathbf{W})_{s,t}(\Delta_1)_{t,s}| \\
+ & \frac{1}{K}|\sum_{(s,t)\in\mathcal{S}^c_{\mathbf{W}^{-1}}}(\frac{1}{nJ}\sum_{i=1}^{n}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{Y}_i^{\mathrm{T}} - \mathbf{W})_{s,t}(\Delta_1)_{t,s}| \\
\leq & \frac{1}{K}\max_{s,t}|(\frac{1}{nJ}\sum_{i=1}^{n}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{Y}_i^{\mathrm{T}} - \mathbf{W})_{s,t}|[\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_{\mathbf{W}^{-1}}}\}\|_1 + \|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}^c_{\mathbf{W}^{-1}}}\}\|_1] \\
\leq & \frac{1}{K}\max_{s,t}|(\frac{1}{nJ}\sum_{i=1}^{n}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{Y}_i^{\mathrm{T}} - \mathbf{W})_{s,t}|[(w+K)^{1/2}\|\Delta_1\|_F + \|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}^c_{\mathbf{W}^{-1}}}\}\|_1] \\
\leq & \max_{s,t}|(\frac{1}{nJ}\sum_{i=1}^{n}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{Y}_i^{\mathrm{T}} - \mathbf{W})_{s,t}|[\frac{1}{\sqrt{K}}\|\Delta_1\|_F + \frac{1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}^c_{\mathbf{W}^{-1}}}\}\|_1]
\end{aligned} \tag{S2.7}
$$

for sufficiently large n, where the lastly equality is due to $w = O(K)$ in Assumption C1. Note that $\mathrm{E}(\mathbf{Y}_i) = \mathbf{0}$, $\mathrm{cov}\{\mathrm{vec}(\mathbf{Y}_i)\} = \mathbf{R}\otimes\mathbf{W}$, and $\sum_{i=1}^{n}\mathrm{vec}(\mathbf{Y}_i)\mathrm{vec}(\mathbf{Y}_i)^{\mathrm{T}}/n$ is the sample covariance matrix of $\mathrm{vec}(\mathbf{Y}_i)$. Under Assumption C3, using Lemma

S2.2, we can obtain that

$$\max_{s,t} |\{\sum_{i=1}^{n} \mathrm{vec}(\mathbf{Y}_i)\mathrm{vec}(\mathbf{Y}_i)^{\mathrm{T}}/n - \mathbf{R}\otimes\mathbf{W}\}_{s,t}| = O_p\{(\log m/n)^{1/2}\}.$$

In Lemma S2.2, letting $\mathbf{A} = \mathbf{B} = \mathbf{L}\otimes\mathbf{I}_K$, following Assumption C2, we know that $\sigma_{\max}(\mathbf{L}\otimes\mathbf{I}_K)$ is bounded, then, $(\mathbf{L}\otimes\mathbf{I}_K)\{\sum_{i=1}^{n} \mathrm{vec}(\mathbf{Y}_i)\mathrm{vec}(\mathbf{Y}_i)^{\mathrm{T}}/n - \mathbf{R}\otimes\mathbf{W}\}(\mathbf{L}^{\mathrm{T}}\otimes\mathbf{I}_K) = \sum_{i=1}^{n} \mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}})^{\mathrm{T}}/n - \mathbf{I}_J\otimes\mathbf{W}$, and

$$\max_{s,t} |\{\sum_{i=1}^{n} \mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}})^{\mathrm{T}}/n - \mathbf{I}_J\otimes\mathbf{W}\}_{s,t}| = O_p\{(\log m/n)^{1/2}\}.$$

In addition, $\sum_{i=1}^{n} \mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{Y}_i^{\mathrm{T}}/n - J\mathbf{W}$ is the sum of the diagonal $K\times K$ blocks of $\sum_{i=1}^{n} \mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}})^{\mathrm{T}}/n - \mathbf{I}_J\otimes\mathbf{W}$, therefore, we have

$$\begin{aligned} & \max_{s,t} |\{\sum_{i=1}^{n} \mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{Y}_i^{\mathrm{T}}/(nJ) - \mathbf{W}\}_{s,t}| \\ \leq \ & \max_{s,t} |\{\sum_{i=1}^{n} \mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}})^{\mathrm{T}}/n - \mathbf{I}_J\otimes\mathbf{W}\}_{s,t}| = O_p\{(\log m/n)^{1/2}\}. \end{aligned}$$

Combining these results, we obtain

$$\begin{aligned} |A| & \leq O_p\left(\{\log K/(nK)\}^{1/2}\|\Delta_1\|_F + \frac{1}{K}(\log K/n)^{1/2}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_\Omega^c}\}\|_1\right) \\ & = O_p\left(\frac{\eta_1}{\sqrt{K}}\|\Delta_1\|_F + \frac{\eta_1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_\Omega^c}\}\|_1\right). \end{aligned}$$

Noting that $\eta_1 \asymp \eta_2$, thus, there exists a constant C_1 so that

$$|T_3| = |A+B+C| \leq C_1\left(\frac{\eta_1}{\sqrt{K}}\|\Delta_1\|_F + \frac{\eta_1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_\Omega^c}\}\|_1\right)$$

with probability tending to 1. Now setting $\lambda_1 = C_1\eta_1$, we have

$$\begin{aligned} T_1 + T_2 - |T_3| \ &\geq \ \frac{1}{4K}M^2 r_{1n}^2 \sigma_{\min}^2(\mathbf{W}) + \frac{\lambda_1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_\Omega^c}\}\|_1 - \frac{\lambda_1}{\sqrt{K}}\|\Delta_1\|_F \\ &\quad -\frac{C_1\eta_1}{\sqrt{K}}\|\Delta_1\|_F - \frac{C_1\eta_1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_\Omega^c}\}\|_1 \\ &\geq \ \left\{\frac{1}{4K}\sigma_{\min}^2(\mathbf{W})M - \frac{2C_1}{K}\right\} M r_{1n}^2 > 0 \end{aligned}$$

with probabiity tending to 1, where the last equality holds as long as the constant M satisfies $M > 8C_1/\sigma_{\min}^2(\mathbf{W})$. This proves (S2.1).

Next, we consider (S2.2). Define $p(\mathbf{L}) \equiv \sum_{j=2}^J p(\mathbf{L}_j)$. Noting that $\mathbf{L}$ and Δ_2 are lower triangular matrices, hence $\log|\mathbf{L} + s\Delta_2| = \sum_{j=1}^J \log(L_{j,j} + s(\Delta_2)_{j,j})$. By Taylor's expansion of $\log|\mathbf{L} + s\Delta_2|$ as a function of s at $s = 0$, we obtain

$$\log|\mathbf{L} + s\Delta_2| - \log|\mathbf{L}| = \sum_{j=1}^J \frac{(\Delta_2)_{j,j}}{L_{j,j}} s - \sum_{j=1}^J \frac{(\Delta_2)_{j,j}^2}{2\{L_{j,j} + \widetilde{s}(\Delta_2)_{j,j}\}^2} s^2,$$

where $\widetilde{s} \in (0, s)$. Then, we have

$$\log|\mathbf{L} + \Delta_2| - \log|\mathbf{L}| = \mathrm{tr}(\mathbf{L}^{-1}\Delta_2) - \sum_{j=1}^J (\Delta_2)_{j,j}^2/[2\{L_{j,j} + \widetilde{s}(\Delta_2)_{j,j}\}^2].$$

Thus, (2.3) leads to $(KJ)^{-1}\{Q(\mathbf{W}^{-1} + \Delta_1, \mathbf{L} + \Delta_2, \boldsymbol{\lambda}) - Q(\mathbf{W}^{-1} + \Delta_1, \mathbf{L}, \boldsymbol{\lambda})\} = I_1 + B + I_2 + I_3 + I_4$, where B is defined as same as in T_3, and

$$\begin{aligned} I_1 \ &\equiv \ \frac{1}{J}\sum_{j=1}^J (\Delta_2)_{j,j}^2/\{L_{j,j} + \widetilde{s}(\Delta_2)_{j,j}\}^2, \\ I_2 \ &\equiv \ \frac{2}{J}\mathrm{tr}(\frac{1}{nK}\sum_{i=1}^n \mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\Delta_2 - \mathbf{L}^{-1}\Delta_2), \end{aligned}$$

$$
\begin{aligned}
I_3 &\equiv \frac{1}{nKJ}\sum_{i=1}^{n}\mathrm{tr}\{\mathbf{Y}_i^{\mathrm{T}}(\mathbf{W}^{-1}+\Delta_1)\mathbf{Y}_i\Delta_2^{\mathrm{T}}\Delta_2\},\\
I_4 &\equiv \frac{\lambda_2}{J}\{p(\mathbf{L}+\Delta_2)-p(\mathbf{L})\}.
\end{aligned}
$$

Due to $r_{2n}=o(1)$ and Assumption condition C2, for any $j\in\{1,\cdots,J\}$ and any $\widetilde{s}\in(0,1)$, we have $|(\mathbf{L}+\widetilde{s}\Delta_2)_{j,j}|\le\sigma_{\max}(\mathbf{L}+\widetilde{s}\Delta_2)\le 2\sigma_{\max}(\mathbf{L})\le 2\tau_1$ for sufficiently large n. Thus,

$$
I_1=\frac{1}{J}\sum_{j=1}^{J}\frac{(\Delta_2)_{j,j}^2}{\{L_{j,j}+\widetilde{s}(\Delta_2)_{j,j}\}^2}\ge\frac{\sum_{j=1}^{J}(\Delta_2)_{j,j}^2}{4J\tau_1^2}. \tag{S2.8}
$$

Following similar derivation of (S2.5), we have

$$
\begin{aligned}
|B| \le& 2(KJ)^{-1/2}\sigma_{\max}(\mathbf{L})\lambda_{\max}(\mathbf{W})\lambda_{\max}(\mathbf{R})\|\Delta_1\|_F\|\Delta_2\|_F\{1+o_p(1)\}\\
=& 2(KJ)^{-1/2}\sigma_{\max}(\mathbf{L})\lambda_{\max}(\mathbf{W})\lambda_{\max}(\mathbf{R})\|\Delta_2\|_F M r_{1n}\{1+o_p(1)\}\\
=& O_p\{(KJ)^{-1/2}\|\Delta_2\|_F\eta_1\sqrt{K}\}\\
=& O_p\{\eta_2\|\Delta_2\|_F/\sqrt{J}\}.
\end{aligned} \tag{S2.9}
$$

Next, we consider I_2. Let $\mathcal{S}_{\mathbf{L}}\equiv\{(l,l'):L_{l,l'}\neq 0, l,l'\in\{1,\cdots,J\}\}$, $\mathcal{S}_{\mathbf{L}}^{(c)}\equiv\{(l,l'):L_{l,l'}=0, l>l', l,l'\in\{1,\cdots,J\}\}$, and let $v\equiv|\mathcal{S}_{\mathbf{L}}|-J$, where $|\mathcal{S}_{\mathbf{L}}|$ is the cardinality of $\mathcal{S}_{\mathbf{L}}$. We have

$$
\begin{aligned}
|I_2| =& \frac{2}{J}|\mathrm{tr}\{(\frac{1}{nK}\sum_{i=1}^{n}\mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}} - \mathbf{L}^{-1})\Delta_2\}| \\
=& \frac{2}{J}|\sum_{t=1}^{J}\sum_{j=1}^{t}(\frac{1}{nK}\sum_{i=1}^{n}\mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}} - \mathbf{L}^{-1})_{t,j}(\Delta_2)_{j,t}| \\
\leq& \frac{2}{J}\max_{t,j}|(\frac{1}{nK}\sum_{i=1}^{n}\mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}} - \mathbf{L}^{-1})_{t,j}|\{\sum_{(l,l')\in\mathcal{S}_{\mathbf{L}}}|(\Delta_2)_{l,l'}| + \sum_{(l,l')\in\mathcal{S}_{\mathbf{L}}^c}|(\Delta_2)_{l,l'}|\}.
\end{aligned}
\tag{S2.10}
$$

Recall that $\mathrm{E}(\mathbf{Y}_i) = \mathbf{0}$, and $\mathrm{cov}\{\mathrm{vec}(\mathbf{Y}_i)\} = \mathbf{R}\otimes\mathbf{W}$, hence $\mathrm{cov}\{\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})\} = \mathbf{W}\otimes\mathbf{R}$, and we know that $\sum_{i=1}^{n}\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})^{\mathrm{T}}/n$ is the sample covariance matrix of $\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})$. Under Assumption C3, using Lemma S2.2, we can obtain that $\max_{s,t}|\{\sum_{i=1}^{n}\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})^{\mathrm{T}}/n - \mathbf{W}\otimes\mathbf{R}\}_{s,t}| = O_p\{(\log m/n)^{1/2}\}$. Let $\mathbf{A}\equiv\mathbf{W}^{-1}\otimes\mathbf{I}_J$ and $\mathbf{B}\equiv\mathbf{I}_K\otimes\mathbf{L}^{\mathrm{T}}$, under Assumption C2, $\sigma_{\max}(\mathbf{A})$ and $\sigma_{\max}(\mathbf{B})$ both are bounded. Following Lemma S2.2, we can obtain that $\max_{t,j}|\{\mathbf{A}(\sum_{i=1}^{n}\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})^{\mathrm{T}}/n - \mathbf{W}\otimes\mathbf{R})\mathbf{B}\}_{t,j}| = O_p\{(\log m/n)^{1/2}\}$. In addition,

$$
\mathbf{A}(\sum_{i=1}^{n}\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})^{\mathrm{T}}/n - \mathbf{W}\otimes\mathbf{R})\mathbf{B} = \sum_{i=1}^{n}\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1})\mathrm{vec}(\mathbf{L}\mathbf{Y}_i^{\mathrm{T}})^{\mathrm{T}}/n - \mathbf{I}_K\otimes\mathbf{L}^{-1},
$$

which has its K diagonal $J\times J$ blocks summing to $\sum_{i=1}^{n}\mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}/n - K\mathbf{L}^{-1}$, thus, $\max_{t,j}|(\frac{1}{nK}\sum_{i=1}^{n}\mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}} - \mathbf{L}^{-1})_{t,j}| \leq \max_{t,j}|\{\mathbf{A}(\sum_{i=1}^{n}\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})^{\mathrm{T}}/n - \mathbf{W}\otimes\mathbf{R})\mathbf{B}\}_{t,j}| = O_p\{(\log m/n)^{1/2}\}$. Incorporating $|\mathcal{S}_{\mathbf{L}}| = v + J = O(J)$ in Assumption C1, we have

$$
|I_2| \leq O_p\left(\frac{\eta_2}{\sqrt{J}}\|\Delta_2\|_F + \frac{\eta_2}{J}\|\mathrm{vec}\{(\Delta_2)_{\mathcal{S}_{\mathbf{L}}^c}\}\|_1\right). \tag{S2.11}
$$

Hence

$$\begin{aligned}|B|+|I_2| =& O_p\left(\frac{\eta_2}{\sqrt{J}}\|\Delta_2\|_F\right)+O_p\left(\frac{\eta_2}{\sqrt{J}}\|\Delta_2\|_F+\frac{\eta_2}{J}\|\mathrm{vec}\{(\Delta_2)_{\mathcal{S}_{\mathbf{L}}^c}\}\|_1\right)\\ \le& C_2\left(\frac{\eta_2}{\sqrt{J}}\|\Delta_2\|_F+\frac{\eta_2}{J}\|\mathrm{vec}\{(\Delta_2)_{\mathcal{S}_{\mathbf{L}}^c}\}\|_1\right) \quad \text{(S2.12)}\end{aligned}$$

for some constant C_2 with probability tending to 1. We next inspect I_3. Because $\|\Delta_1\|_F = o(1)$, $\mathbf{W}^{-1}$ is positive definite, $\mathbf{W}^{-1}+\Delta_1$ is positive definite for sufficiently large n. Following Lemma S2.1, we have

$$I_3 = \frac{1}{nKJ}\sum_{i=1}^n \|(\mathbf{W}^{-1}+\Delta_1)^{1/2}\mathbf{Y}_i\Delta_2^{\mathrm{T}}\|_F^2 \ge \lambda_{\min}(\mathbf{W}^{-1}+\Delta_1)\frac{1}{nKJ}\sum_{i=1}^n \|\mathbf{Y}_i\Delta_2^{\mathrm{T}}\|_F^2.$$

Because the relation $\mathrm{tr}(\mathbf{Y}_i\Delta_2^{\mathrm{T}}\Delta_2\mathbf{Y}_i^{\mathrm{T}}) = \mathrm{vec}(\mathbf{Y}_i)^{\mathrm{T}}(\Delta_2^{\mathrm{T}}\Delta_2\otimes\mathbf{I}_K)\mathrm{vec}(\mathbf{Y}_i) = \mathrm{tr}\{(\Delta_2^{\mathrm{T}}\Delta_2\otimes \mathbf{I}_K)\mathrm{vec}(\mathbf{Y}_i)\mathrm{vec}(\mathbf{Y}_i)^{\mathrm{T}}\}$, by law of large numbers,

$$\begin{aligned}\frac{1}{nKJ}\sum_{i=1}^n \|\mathbf{Y}_i\Delta_2^{\mathrm{T}}\|_F^2 &= \mathrm{tr}\{(\Delta_2^{\mathrm{T}}\Delta_2\otimes\mathbf{I}_K)(\mathbf{R}\otimes\mathbf{W})\}\\ &= \frac{1}{KJ}\mathrm{tr}(\Delta_2^{\mathrm{T}}\Delta_2\mathbf{R})\mathrm{tr}(\mathbf{W})\{1+o_p(1)\}.\end{aligned}$$

Note further that $\mathrm{tr}(\Delta_2^{\mathrm{T}}\Delta_2\mathbf{R}) = \|\Delta_2\mathbf{L}^{-1}\|_F^2 \ge \sigma_{\min}^2(\mathbf{L}^{-1})\|\Delta_2\|_F^2$, thus,

$$I_3 \ge J^{-1}\lambda_{\min}(\mathbf{W}^{-1}+\Delta_1)\lambda_{\min}(\mathbf{W})\sigma_{\min}^2(\mathbf{L}^{-1})\|\Delta_2\|_F^2 = c\sigma_{\min}^2(L^{-1})\|\Delta_2\|_F^2/J, \quad \text{(S2.13)}$$

where $c \equiv \lambda_{\min}(\mathbf{W}^{-1}+\Delta_1)\lambda_{\min}(\mathbf{W}) > 0$. Lastly, we consider I_4. Recall that the jth row of $\mathbf{L}$ has bandwidth d_j, hence $L_{j,l}=0$ for $1\le l<j-d_j$. Then, we have $p(\mathbf{L}) = \sum_{j=2}^J\sum_{l=1}^{j-1}(\sum_{m=1}^l L_{j,m}^2)^{1/2} = \sum_{j=2}^J\sum_{l=j-d_j}^{j-1}(\sum_{m=j-d_j}^l L_{j,m}^2)^{1/2}$.

Thus,

$$
\begin{aligned}
& p(\mathbf{L}+\Delta_2)-p(\mathbf{L}) = \sum_{j=2}^{J}\sum_{l=1}^{j-1}(\sum_{m=1}^{l}(\mathbf{L}+\Delta_2)_{j,m}^2)^{1/2}-p(\mathbf{L}) \\
&= \sum_{j=2}^{J}\sum_{l=1}^{j-d_j-1}(\sum_{m=1}^{l}(\mathbf{L}+\Delta_2)_{j,m}^2)^{1/2}+\sum_{j=2}^{J}\sum_{l=j-d_j}^{j-1}(\sum_{m=1}^{l}(\mathbf{L}+\Delta_2)_{j,m}^2)^{1/2}-p(\mathbf{L}) \\
&\geq \sum_{j=2}^{J}\sum_{l=1}^{j-d_j-1}|(\mathbf{L}+\Delta_2)_{j,l}|+\sum_{j=2}^{J}\sum_{l=j-d_j}^{j-1}\{\sum_{m=j-d_j}^{l}(\mathbf{L}+\Delta_2)_{j,m}^2\}^{1/2}-\sum_{j=2}^{J}\sum_{l=j-d_j}^{j-1}(\sum_{m=j-d_j}^{l}L_{j,m}^2)^{1/2} \\
&\geq \sum_{j=2}^{J}\sum_{l=1}^{j-d_j-1}|(\Delta_2)_{j,l}|-\sum_{j=2}^{J}\sum_{l=j-d_j}^{j-1}\{\sum_{m=j-d_j}^{l}(\Delta_2)_{j,m}^2\}^{1/2}, \qquad \text{(S2.14)}
\end{aligned}
$$

where the last inequality comes from triangle inequality. Using Cauchy's inequality, we have

$$
\begin{aligned}
2\lambda_2\sum_{l=j-d_j}^{j-1}\{\sum_{m=j-d_j}^{l}(\Delta_2)_{j,m}^2\}^{1/2} &\leq d_j\lambda_2^2 a+\sum_{l=j-d_j}^{j-1}\sum_{m=j-d_j}^{l}(\Delta_2)_{j,m}^2/a \\
&= d_j\lambda_2^2 a+\sum_{m=j-d_j}^{j-1}(j-m)(\Delta_2)_{j,m}^2/a \\
&\leq d_j\lambda_2^2 a+d_j\sum_{m=j-d_j}^{j-1}(\Delta_2)_{j,m}^2/a.
\end{aligned}
$$

Letting $a=\max_j d_j/\{c\sigma_{\min}^2(\mathbf{L}^{-1})\}$, recalling that $v=\sum_{j=2}^{J}d_j$, we get

$$
\begin{aligned}
\lambda_2\sum_{j=2}^{J}\sum_{l=j-d_j}^{j-1}\{\sum_{m=j-d_j}^{l}(\Delta_2)_{j,m}^2\}^{1/2} &\leq \tfrac{a}{2}\lambda_2^2 v+\max_j(d_j)\sum_{j=2}^{J}\sum_{m=j-d_j}^{j-1}\tfrac{(\Delta_2)_{j,m}^2}{2a} \\
&= \tfrac{a}{2}\lambda_2^2 v+\tfrac{c\sigma_{\min}^2(\mathbf{L}^{-1})}{2}\|(\Delta_2)\|_F^2.
\end{aligned}
$$

Combining with (S2.14), noting that $\mathcal{S}_{\mathbf{L}}^c=\{(j,l):2\leq j\leq J,1\leq l\leq j-d_j-1\}$, we have

$$I_4 \geq \frac{\lambda_2}{J}\left[\sum_{j=2}^{J}\sum_{l=1}^{j-d_j-1}|(\Delta_2)_{j,l}| - \sum_{j=2}^{J}\sum_{l=j-d_j}^{j-1}\{\sum_{m=j-d_j}^{l}(\Delta_2)_{j,m}^2\}^{1/2}\right]$$
$$\geq \frac{\lambda_2}{J}\|\mathrm{vec}\{(\Delta_2)_{\mathcal{S}_{\mathbf{L}}^c}\}\|_1 - \frac{1}{J}(\frac{a}{2}\lambda_2^2 v + \frac{c\sigma_{\min}^2(\mathbf{L}^{-1})}{2}\|\Delta_2\|_F^2). \quad \text{(S2.15)}$$

Note that this inequality holds for any $\lambda_2 > 0$. Let $\lambda_2 \equiv C_2\eta_2$. Collecting the results in (S2.8),(S2.12),(S2.13) and (S2.15), we get

$$\begin{aligned}
& \frac{1}{KJ}\{Q(\mathbf{W}^{-1}+\Delta_1, \mathbf{L}+\Delta_2, \boldsymbol{\lambda}) - Q(\mathbf{W}^{-1}+\Delta_1, \mathbf{L}, \boldsymbol{\lambda})\} \\
\geq\ & \frac{1}{4J\tau_1^2}\sum_{j=1}^{J}(\Delta_2)_{j,j}^2 - \frac{C_2\eta_2}{\sqrt{J}}\|\Delta_2\|_F - \frac{C_2\eta_2}{J}\|\mathrm{vec}\{(\Delta_2)_{\mathcal{S}_{\mathbf{L}}^c}\}\|_1 \\
& + \frac{c\sigma_{\min}^2(\mathbf{L}^{-1})}{J}\|\Delta_2\|_F^2 + \frac{\lambda_2}{J}\|\mathrm{vec}\{(\Delta_2)_{\mathcal{S}_{\mathbf{L}}^c}\}\|_1 - \frac{a}{2J}\lambda_2^2 v - \frac{c\sigma_{\min}^2(\mathbf{L}^{-1})}{2J}\|\Delta_2\|_F^2 \\
=\ & \frac{1}{4J\tau_1^2}\sum_{j=1}^{J}(\Delta_2)_{j,j}^2 - \frac{C_2\eta_2}{\sqrt{J}}Nr_{2n} + \frac{c\sigma_{\min}^2(\mathbf{L}^{-1})}{2J}N^2 r_{2n}^2 - \frac{av}{2J}C_2^2\eta_2^2 \\
\geq\ & \{\frac{c}{2J}\sigma_{\min}^2(\mathbf{L}^{-1})N^2 - \frac{C_2}{J}N - \frac{avC_2^2}{2J^2}\}r_{2n}^2 > 0
\end{aligned}$$

with probability tending to 1, where the lastly equality holds for sufficiently large constant N. Thus (S2.2) is shown. Given (S2.1) and (S2.2), since $Q(\mathbf{W}^{-1}, \mathbf{L}, \boldsymbol{\lambda})$ is biconvex in $\mathbf{W}^{-1}$ and $\mathbf{L}$, for sufficiently large constants M and N, there exists a local minimum $(\widehat{\mathbf{W}}^{-1}, \widehat{\mathbf{L}})$ in the set $\{\mathbf{W}^{-1}+\Delta_1 : \|\Delta_1\|_F \leq Mr_{1n}\} \times \{\mathbf{L}+\Delta_2 : (\Delta_2)_{i,j} = 0, \text{ for } j > i, (\Delta_2)_{j,j} > 0, \|(\Delta_2)_j\|_2 \leq NJ^{-1/2}r_{2n}\}$. Hence, $\|\widehat{\mathbf{W}}^{-1} - \mathbf{W}^{-1}\|_F = O_p(r_{1n})$, $\|\widehat{\mathbf{L}}_j - \mathbf{L}_j\|_2 = O_p(J^{-1/2}r_{2n})$ and $\|\widehat{\mathbf{L}} - \mathbf{L}\|_F = O_p(r_{2n})$. Further, there exists a constant D_1 so that

$$
\begin{aligned}
\|\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}\|_F &= \|\widehat{\mathbf{L}}^{\mathrm{T}}\widehat{\mathbf{L}} - \mathbf{L}^{\mathrm{T}}\mathbf{L}\|_F \\
&\leq \|(\widehat{\mathbf{L}}^{\mathrm{T}} - \mathbf{L}^{\mathrm{T}})\widehat{\mathbf{L}}\|_F + \|\mathbf{L}^{\mathrm{T}}(\widehat{\mathbf{L}} - \mathbf{L})\|_F \\
&\leq \sigma_{\max}(\widehat{\mathbf{L}})\|\widehat{\mathbf{L}}^{\mathrm{T}} - \mathbf{L}^{\mathrm{T}}\|_F + \sigma_{\max}(\mathbf{L}^{\mathrm{T}})\|\widehat{\mathbf{L}} - \mathbf{L}\|_F \\
&\leq D_1\|\widehat{\mathbf{L}} - \mathbf{L}\|_F.
\end{aligned}
$$

The last inequality is based on the following analysis. We can write $\widehat{\mathbf{L}} = \mathbf{L} + \Delta_2$, where Δ_2 is lower triangular and $\|\Delta_2\|_F \leq Nr_{2n} = o(1)$. Then there exists $\mathbf{x}$ with $\|\mathbf{x}\| = 1$, so that

$$
\begin{aligned}
\sigma^2_{\max}(\widehat{\mathbf{L}}) &= \sigma^2_{\max}(\mathbf{L} + \Delta_2) = \mathbf{x}^{\mathrm{T}}(\mathbf{L} + \Delta_2)^{\mathrm{T}}(\mathbf{L} + \Delta_2)\mathbf{x} \\
&= \mathbf{x}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\Delta_2\mathbf{x} + \mathbf{x}^{\mathrm{T}}\Delta_2^{\mathrm{T}}\mathbf{L}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\Delta_2^{\mathrm{T}}\Delta_2\mathbf{x} \\
&\leq \sigma^2_{\max}(\mathbf{L}) + \sigma^2_{\max}(\Delta_2) + 2\{(\mathbf{x}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{x})(\mathbf{x}^{\mathrm{T}}\Delta_2^{\mathrm{T}}\Delta_2\mathbf{x})\}^{1/2} \\
&\leq \sigma^2_{\max}(\mathbf{L}) + \sigma^2_{\max}(\Delta_2) + 2\sigma_{\max}(\mathbf{L})\sigma_{\max}(\Delta_2) \\
&\leq \sigma^2_{\max}(\mathbf{L}) + \|\Delta_2\|_F^2 + 2\sigma_{\max}(\mathbf{L})\|\Delta_2)\|_F \\
&= \sigma^2_{\max}(\mathbf{L}) + o(1) \leq 2\sigma^2_{\max}(\mathbf{L})
\end{aligned}
$$

for sufficiently large n. Thus, we have $\|\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}\|_F = O_p(r_{2n})$. In addition,

$$
\begin{aligned}
& \|\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\|_F \\
= \ & \|(\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}) \otimes (\widehat{\mathbf{W}}^{-1} - \mathbf{W}^{-1}) + (\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}) \otimes \mathbf{W}^{-1} + \mathbf{R}^{-1} \otimes (\widehat{\mathbf{W}}^{-1} - \mathbf{W}^{-1})\|_F \\
\leq \ & \|(\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}) \otimes (\widehat{\mathbf{W}}^{-1} - \mathbf{W}^{-1})\|_F + \|(\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}) \otimes \mathbf{W}^{-1}\|_F + \|\mathbf{R}^{-1} \otimes (\widehat{\mathbf{W}}^{-1} - \mathbf{W}^{-1})\|_F \\
= \ & \|\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}\|_F \|\widehat{\mathbf{W}}^{-1} - \mathbf{W}^{-1}\|_F + \|\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}\|_F \|\mathbf{W}^{-1}\|_F + \|\mathbf{R}^{-1}\|_F \|\widehat{\mathbf{W}}^{-1} - \mathbf{W}^{-1}\|_F.
\end{aligned}
$$

So that, we have

$$
\begin{aligned}
& (KJ)^{-1/2} \|\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\|_F \\
\leq \ & (KJ)^{-1/2} \{O_p(r_{1n} r_{2n}) + O_p(r_{2n}) \sqrt{K} \sigma_{max}(\mathbf{W}^{-1}) + O_p(r_{1n}) \sqrt{J} \sigma_{max}(\mathbf{R}^{-1})\} \\
\leq \ & \sigma_{\max}(\mathbf{W}^{-1}) O_p(r_{2n}/\sqrt{J}) + \sigma_{\max}(\mathbf{R}^{-1}) O_p(r_{1n}/\sqrt{K}) \\
= \ & O_p\{(\log m/n)^{1/2}\}.
\end{aligned}
$$

Thus, $\|\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\|_F = O_p\{(m \log m/n)^{1/2}\}$.

Part 2. We now prove the second part of the theorem, where $\text{vec}(\mathbf{Y}_i) \sim \mathbf{N}(\mathbf{0}, \mathbf{R} \otimes \mathbf{W})$. We set $\eta_1 \equiv \{\log K/(nJ)\}^{1/2}$, $\eta_2 \equiv \{\log J/(nK)\}^{1/2}$. Obviously, $\eta_1 \asymp \eta_2$, and $r_{1n} \equiv \eta_1 \sqrt{K}$, $r_{2n} \equiv \eta_2 \sqrt{J}$ satisfy $r_{1n} \to 0$, $r_{2n} \to 0$ under the assumption in the second half of the theorem. At a fixed $\boldsymbol{\lambda}$, let $(\widehat{\mathbf{W}}^{-1}, \widehat{\mathbf{L}})$ be a local minimum of $Q(\mathbf{W}^{-1}, \mathbf{L}, \boldsymbol{\lambda})$. Define $\mathcal{A}$ and $\mathcal{B}$ as **Part 1**. Let $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{B}$, for each fixed $\Delta_1 \in \mathcal{A}$, and $\Delta_2 \in \mathcal{B}$, denote $\mathcal{C}_{\Delta_1} \equiv \{\Delta_2 \in \mathcal{B} | (\Delta_1, \Delta_2) \in \mathcal{C}\}$ and $\mathcal{C}_{\Delta_2} \equiv \{\Delta_1 \in \mathcal{A} | (\Delta_1, \Delta_2) \in \mathcal{C}\}$. To show the result, we need to show that (S2.1) and (S2.2) hold. Firstly, we consider proving (S2.1). Following the same derivation as in **Part 1**, we have the same decomposition $(KJ)^{-1}\{Q(\mathbf{W}^{-1} + \Delta_1, \mathbf{L} + \Delta_2, \boldsymbol{\lambda}) - Q(\mathbf{W}^{-1}, \mathbf{L} + \Delta_2, \boldsymbol{\lambda})\} =$

$T_1+T_2+T_3=T_1+T_2+A+B+C$, where T_1, T_2 have the same lower bounds as in (S2.3), (S2.4), and $|A|, |B|, |C|$ have the same upper bounds as in (S2.7), (S2.5), (S2.6). Since $\mathrm{vec}(\mathbf{Y}_i) \sim N(\mathbf{0}, \mathbf{R}\otimes\mathbf{W})$ and $\mathbf{R}^{-1}=\mathbf{L}^{\mathrm{T}}\mathbf{L}$, we have $\mathrm{cov}\{\mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}})\} = \mathbf{I}_J\otimes\mathbf{W}$, and $\mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}) \sim N(\mathbf{0}, \mathbf{I}_J\otimes\mathbf{W})$, which implies that the J columns of $\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}$ are independent with each other and identically distributed normal random vectors with covariance matrix $\mathbf{W}$. Noting that $\sum_{i=1}^n \mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{Y}_i^{\mathrm{T}}/(nJ)$ is the sample covariance matrix of an aribitary column of $\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}$, under Assumption C3, using Lemma S2.2, we get $\max_{s,t}|\{\sum_{i=1}^n \mathbf{Y}_i\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{Y}_i^{\mathrm{T}}/(nJ)-\mathbf{W}\}_{s,t}| = O_p[\{\log K/(nJ)\}^{1/2}]$. Inserting these results to (S2.7), we obtain

$$|A| \le O_p\left(\frac{\eta_1}{\sqrt{K}}\|\Delta_1\|_F + \frac{\eta_1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_\Omega^c}\}\|_1\right).$$

Thus, there exists a constant C_1 so that

$$|T_3| = |A+B+C| \le C_1\left[\frac{\eta_1}{\sqrt{K}}\|\Delta_1\|_F + \frac{\eta_1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_\Omega^c}\}\|_1\right]$$

with probability tending to 1. Now set $\lambda_1 = C_1\eta_1$, we have

$$\begin{aligned} T_1+T_2-|T_3| \ &\ge\ \frac{1}{4K}M^2 r_{1n}^2\sigma_{\min}^2(\mathbf{W}) + \frac{\lambda_1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_\Omega^c}\}\|_1 - \frac{\lambda_1}{\sqrt{K}}\|\Delta_1\|_F \\ &\quad -\frac{C_1\eta_1}{\sqrt{K}}\|\Delta_1\|_F - \frac{C_1\eta_1}{K}\|\mathrm{vec}\{(\Delta_1)_{\mathcal{S}_\Omega^c}\}\|_1 \\ &\ge\ \left\{\frac{1}{4K}\sigma_{\min}^2(\mathbf{W})M - \frac{2C_1}{K}\right\}Mr_{1n}^2 > 0 \end{aligned}$$

with probabiity tending to 1, where the last equality holds as long as the constant M satisfies $M > 8C_1/\sigma_{\min}^2(\mathbf{W})$. This proves (S2.1).

Next, we consider (S2.2). The same derivation as in **Part 1** leads to the same decomposition $(KJ)^{-1}\{Q(\mathbf{W}^{-1}+\Delta_1,\mathbf{L}+\Delta_2,\boldsymbol{\lambda})-Q(\mathbf{W}^{-1}+\Delta_1,\mathbf{L},\boldsymbol{\lambda})\} = I_1+B+I_2+I_3+I_4$, where I_1, I_3, I_4 have the same lower bounds as in (S2.8), (S2.13), (S2.15), and $|B|$ has the same upper bound as in (S2.9). Next, we consider I_2. Recall that $\text{vec}(\mathbf{Y}_i)\sim N(\mathbf{0},\mathbf{R}\otimes\mathbf{W})$, and $\mathbf{W}^{-1}=\mathbf{W}^{-1}$, we have $\text{cov}[\text{vec}\{(\mathbf{W}^{-1})^{1/2}\mathbf{Y}_i\}] = \mathbf{R}\otimes\mathbf{I}_K$, and $\text{vec}\{(\mathbf{W}^{-1})^{1/2}\mathbf{Y}_i\}\sim N(\mathbf{0},\mathbf{R}\otimes\mathbf{I}_K)$. which implies that the K rows of $(\mathbf{W}^{-1})^{1/2}\mathbf{Y}_i$ are independent with each other and identically distributed normal random vectors with covariance matrix $\mathbf{R}$. Further, $\sum_{i=1}^n \mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i/(nK)$ is the sample covariance matrix of each row of $(\mathbf{W}^{-1})^{1/2}\mathbf{Y}_i$. Under Assumption C3, using Lemma S2.2, we can obtain that $\max_{s,t}|\{\sum_{i=1}^n \mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i/(nK)-\mathbf{R}\}_{s,t}| = O_p[\{\log J/(nK)\}^{1/2}]$. Further setting $\mathbf{A}=\mathbf{I}_J$ and $\mathbf{B}=\mathbf{L}^{\mathrm{T}}$ in Lemma S2.2 leads to $\max_{s,t}|\{\sum_{i=1}^n \mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i\mathbf{L}^{\mathrm{T}}/(nK)-\mathbf{L}^{-1}\}_{s,t}| = O_p[\{\log J/(nK)\}^{1/2}]$. Incorporating this result into (S2.10) leads to the same upper bounded as (S2.11). Similar as in **Part 1**, we now let $\lambda_2\equiv C_2\eta_2$ to obtain $(KJ)^{-1}\{Q(\mathbf{W}^{-1}+\Delta_1,\mathbf{L}+\Delta_2,\boldsymbol{\lambda})-Q(\mathbf{W}^{-1}+\Delta_1,\mathbf{L},\boldsymbol{\lambda})\}>0$ with probability tending to 1, where the lastly equality holds for sufficiently large constant N. Thus (S2.2) is shown. Given (S2.1) and (S2.2), since $Q(\mathbf{W}^{-1},\mathbf{L},\boldsymbol{\lambda})$ is biconvex in $\mathbf{W}^{-1}$ and $\mathbf{L}$, for sufficiently large constants M and N, there exists a local minimum $(\widehat{\mathbf{W}}^{-1},\widehat{\mathbf{L}})$ in the set $\{\mathbf{W}^{-1}+\Delta_1:\|\Delta_1\|_F\le Mr_{1n}\}\times\{\mathbf{L}+\Delta_2:(\Delta_2)_{i,j}=0,\text{ for } j>i,(\Delta_2)_{j,j}>0,\|(\Delta_2)_j\|_2\le NJ^{-1/2}r_{2n}\}$. Hence, $\|\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1}\|_F=O_p(r_{1n})$, $\|\widehat{\mathbf{L}}_j-\mathbf{L}_j\|_2=O_p(J^{-1/2}r_{2n})=O_p[\{\log J/(nK)\}^{1/2}]$ and $\|\widehat{\mathbf{L}}-\mathbf{L}\|_F=O_p(r_{2n})$. Further, the same derivation as in **Part 1** leads to $\|\widehat{\mathbf{R}}^{-1}-\mathbf{R}^{-1}\|_F=O_p(r_{2n})$ and

$$
\begin{aligned}
(KJ)^{-1/2}\|\widehat{\mathbf{\Sigma}}^{-1}-\mathbf{\Sigma}^{-1}\|_F &\leq \sigma_{\max}(\mathbf{W}^{-1})O_p(r_{2n}/\sqrt{J})+\sigma_{\max}(\mathbf{R}^{-1})O_p(r_{1n}/\sqrt{K}) \\
&= O_p[\max\{\{\log J/(nK)\}^{1/2},\{\log K/(nJ)\}^{1/2}\}].
\end{aligned}
$$

Thus, $\|\widehat{\mathbf{\Sigma}}^{-1}-\mathbf{\Sigma}^{-1}\|_F = O_p[\max\{(J\log J/n)^{1/2},(K\log K/n)^{1/2}\}]$. □

S3 Uniqueness of the precision matrix estimator BKS

Proof of Lemma 1. Define

$$
L(\tau,\mathbf{V}_j,\mathbf{L}_j;\vartheta,\boldsymbol{\phi},\mathbf{A}) \equiv -2\log\tau+\frac{1}{nK}\|\mathbf{V}_j\|_2^2+\vartheta(\tau-L_{jj})+\frac{1}{nK}\langle\boldsymbol{\phi},\mathbf{V}_j-\mathbf{Y}^*_{\cdot,1:j}\mathbf{L}_j\rangle+\lambda_2\sum_{l=1}^{j-1}\langle\mathbf{A}_{\cdot,l},\mathbf{L}_j\rangle.
$$

Note that $\min_{L_{jj}>0,\mathbf{L}_j\in\mathcal{R}^j} f(\mathbf{L}_j)$ can be equivalently written as the minimization problem

$$
\min_{\tau,\mathbf{V}_j,\mathbf{L}_j}[\max_{\vartheta,\boldsymbol{\phi},\mathbf{A}}\{L(\tau,\mathbf{V}_j,\mathbf{L}_j;\vartheta,\boldsymbol{\phi},\mathbf{A})\}]. \tag{S3.1}
$$

To solve (S3.1), consider the dual function

$$
\begin{aligned}
g(\vartheta,\boldsymbol{\phi},\mathbf{A}) &\equiv \min_{\tau,\mathbf{V}_j,\mathbf{L}_j} L(\tau,\mathbf{V}_j,\mathbf{L}_j;\vartheta,\boldsymbol{\phi},\mathbf{A}) \\
&= \min_{\tau}\{-2\log\tau+\vartheta\tau\}+\min_{\mathbf{V}_j}\{\frac{1}{nK}\|\mathbf{V}_j\|_2^2+\frac{1}{nK}\langle\boldsymbol{\phi},\mathbf{V}_j\rangle\} \\
&+ \min_{\mathbf{L}_j}\{-\vartheta L_{jj}-\frac{1}{nK}\langle\boldsymbol{\phi},\mathbf{Y}^*_{\cdot,1:j}\mathbf{L}_j\rangle+\lambda_2\sum_{l=1}^{j-1}\langle\mathbf{A}_{\cdot,l},\mathbf{L}_j\rangle\} \\
&= I(\vartheta>0)\{2\log(\vartheta/2)+2\}-I(\vartheta\leq 0)\infty-\frac{1}{4nK}\|\boldsymbol{\phi}\|_2^2 \\
&\quad -I(-\vartheta\mathbf{e}_j-\frac{1}{nK}\mathbf{Y}^{*\mathrm{T}}_{\cdot,1:j}\boldsymbol{\phi}+\lambda_2\sum_{l=1}^{j-1}\mathbf{A}_{\cdot,l}\neq\mathbf{0})\infty,
\end{aligned}
$$

where $I(\cdot)$ is the indicator function. Then, (S3.1) is equivalent to the dual problem $\max_{\vartheta,\boldsymbol{\phi},\mathbf{A}}\{g(\vartheta,\boldsymbol{\phi},\mathbf{A})\}$, which is equivalent to

$$\min_{\vartheta,\boldsymbol{\phi},\mathbf{A}}\{-2\log\vartheta+\frac{1}{4nK}\|\boldsymbol{\phi}\|_2^2\} \tag{S3.2}$$

under the constraints $\vartheta>0$, $\|\mathbf{A}_{1:l,l}\|_2\le 1, \mathbf{A}_{l+1:j,l}=\mathbf{0}$, and $\vartheta\mathbf{e}_j+\mathbf{Y}^{*\mathrm{T}}_{\cdot,1:j}\boldsymbol{\phi}/(nK)=\lambda_2\sum_{l=1}^{j-1}\mathbf{A}_{\cdot,l}$ for any $l=1,\ldots,j-1$. Note that the primal-dual relation requires the optimal solution to satisfy

$$\widehat{L}_{jj}=\widehat{\tau}=2/\widehat{\vartheta},\quad \widehat{\boldsymbol{\phi}}=-2\widehat{\mathbf{V}}_j=-2\mathbf{Y}^*_{\cdot,1:j}\widehat{\mathbf{L}}_j. \tag{S3.3}$$

Thus, the above constraints are equivalently written as

$$-\frac{2}{\widehat{L}_{jj}}\mathbf{e}_j+\frac{2}{nK}\mathbf{Y}^{*\mathrm{T}}_{\cdot,1:j}\mathbf{Y}^*_{\cdot,1:j}\widehat{\mathbf{L}}_j+\lambda_2\sum_{l=1}^{j-1}\widehat{\mathbf{A}}_{\cdot,l}=0,\quad \|\widehat{\mathbf{A}}_{1:l,l}\|_2\le 1,\widehat{\mathbf{A}}_{l+1:j,l}=\mathbf{0}.$$

Following the combination of $f(\widehat{\mathbf{L}}_j)=L(\widehat{\tau},\widehat{V},\widehat{\mathbf{L}}_j;\widehat{\vartheta},\widehat{\boldsymbol{\phi}},\widehat{\mathbf{A}})$ and the primal-dual relations, we can obtain that

$$p(\widehat{\mathbf{L}}_j)=\sum_{l=1}^{j-1}\langle\widehat{\mathbf{A}}_{\cdot,l},\widehat{\mathbf{L}}_j\rangle. \tag{S3.4}$$

Assume that there exists l, where $(\widehat{\mathbf{L}}_j)_{1:l}\ne\mathbf{0}$, but $\widehat{\mathbf{A}}_{1:l,l}\ne(\widehat{\mathbf{L}}_j)_{1:l}/\|(\widehat{\mathbf{L}}_j)_{1:l}\|_2$, then $\langle\widehat{\mathbf{L}}_j,\widehat{\mathbf{A}}_{\cdot,l}\rangle<\|(\widehat{\mathbf{L}}_j)_{1:l}\|_2$. For all $l'\ne l$, the Cauchy-Schwartz inequality leads to $\langle\widehat{\mathbf{L}}_j,\widehat{\mathbf{A}}_{\cdot,l'}\rangle\le\|(\widehat{\mathbf{L}}_j)_{1:l'}\|_2$. Thus,

$$p(\widehat{\mathbf{L}}_j)=\sum_{l=1}^{j-1}\|(\widehat{\mathbf{L}}_j)_{1:l}\|_2>\sum_{l=1}^{j-1}\langle\widehat{\mathbf{L}}_j,\widehat{\mathbf{A}}_{\cdot,l}\rangle,$$

which is a contradiction. Therefore, $\widehat{\mathbf{A}}_{1:l,l} = (\widehat{\mathbf{L}}_j)_{1:l}/\|(\widehat{\mathbf{L}}_j)_{1:l}\|_2$ for $(\widehat{\mathbf{L}}_j)_{1:l} \neq \mathbf{0}$. Of course we still have $\|\widehat{\mathbf{A}}_{1:l,l}\| \leq 1$ as required originally. Thus, we have shown that solving (S3.2) under the original constraints is equivalent to solving (S3.2) under the constraints stated in Lemma 1. We now see that (S3.2) is also irrelevant to these constrains. Thus, we can ignore (S3.2) and directly obtain the equivalence stated in Lemma 1.

In the above analysis, we have shown that given $\widehat{\mathbf{W}}^{-1}$, let $\widehat{\mathbf{L}}_j$ be a solution of the optimal problem $\min_{L_{jj}>0, \mathbf{L}_j \in \mathcal{R}^j} f(\mathbf{L}_j)$, then there exists some $\widehat{\mathbf{A}}$, so that $\widehat{\mathbf{L}}_j$ and $\widehat{\mathbf{A}}$ satisfy (3.2) and (3.3). We now construct one specific $\widehat{\mathbf{A}}$ as follows. For $l = j - \widehat{d}_j, \cdots, j-1$, let $\widehat{\mathbf{A}}_{1:l,l} \equiv (\widehat{\mathbf{L}}_j)_{1:l}/\|(\widehat{\mathbf{L}}_j)_{1:l}\|_2$, $\widehat{\mathbf{A}}_{l+1:j,l} \equiv \mathbf{0}$. For $l = 1, \cdots, j - \widehat{d}_j - 1$, let $\widehat{\mathbf{A}}_{l,l} \equiv -2\mathbf{Y}^{*\mathrm{T}}_{\cdot,l}\mathbf{Y}^*_{,1:j}\widehat{\mathbf{L}}_j/nK\lambda_2$, and let all other components of $\widehat{\mathbf{A}}_{\cdot,l}$ be 0. Next, we will show that when λ_2 is sufficiently large, $\widehat{\mathbf{A}}$ satisfies $\|\widehat{\mathbf{A}}_{1:l,l}\|_2 < 1$ for $l = 1, \cdots, j-1-\widehat{d}_j$.

$$
\begin{aligned}
|\lambda_2 \widehat{\mathbf{A}}_{l,l}/2| = &\left|\frac{1}{nK}\mathbf{Y}^{*\mathrm{T}}_{\cdot,l}\mathbf{Y}^*_{\cdot,1:j}\widehat{\mathbf{L}}_j\right| = |\{\frac{1}{nK}\sum_{i=1}^n (\mathbf{Y}_i)^{\mathrm{T}}_{\cdot,1:j}\widehat{\mathbf{W}}^{-1}(\mathbf{Y}_i)_{\cdot,1:j}(\widehat{\mathbf{L}}_j - \mathbf{L}_j + \mathbf{L}_j)\}_l| \\
=&|\frac{1}{nK}\sum_{i=1}^n \mathbf{e}_l^{\mathrm{T}}(\mathbf{Y}_i)^{\mathrm{T}}_{\cdot,1:j}(\widehat{\mathbf{W}}^{-1} - \mathbf{W}^{-1} + \mathbf{W}^{-1})(\mathbf{Y}_i)_{\cdot,1:j}(\widehat{\mathbf{L}}_j - \mathbf{L}_j + \mathbf{L}_j)| \\
\leq&|\frac{1}{nK}\sum_{i=1}^n \mathbf{e}_l^{\mathrm{T}}(\mathbf{Y}_i)^{\mathrm{T}}_{\cdot,1:j}(\widehat{\mathbf{W}}^{-1} - \mathbf{W}^{-1})(\mathbf{Y}_i)_{\cdot,1:j}(\widehat{\mathbf{L}}_j - \mathbf{L}_j)| \\
+&|\frac{1}{nK}\sum_{i=1}^n \mathbf{e}_l^{\mathrm{T}}(\mathbf{Y}_i)^{\mathrm{T}}_{\cdot,1:j}(\widehat{\mathbf{W}}^{-1} - \mathbf{W}^{-1})(\mathbf{Y}_i)_{\cdot,1:j}\mathbf{L}_j| \\
+&|\frac{1}{nK}\sum_{i=1}^n \mathbf{e}_l^{\mathrm{T}}(\mathbf{Y}_i)^{\mathrm{T}}_{\cdot,1:j}\mathbf{W}^{-1}(\mathbf{Y}_i)_{\cdot,1:j}(\widehat{\mathbf{L}}_j - \mathbf{L}_j)| + |\frac{1}{nK}\sum_{i=1}^n \mathbf{e}_l^{\mathrm{T}}(\mathbf{Y}_i)^{\mathrm{T}}_{\cdot,1:j}\mathbf{W}^{-1}(\mathbf{Y}_i)_{\cdot,1:j}\mathbf{L}_j|,
\end{aligned}
\tag{S3.5}
$$

where $\mathbf{e}_l \in \mathcal{R}^j$, with its lth element 1 and all other elements 0. We inspect the first term in (S3.5) first. Following the law of large numbers,

$$
\begin{aligned}
&(n)^{-1}\sum_{i=1}^{n}\mathrm{tr}\{\mathbf{e}_l^{\mathrm{T}}(\mathbf{Y}_i)_{\cdot,1:j}^{\mathrm{T}}(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})(\mathbf{Y}_i)_{\cdot,1:j}(\widehat{\mathbf{L}}_j-\mathbf{L}_j)\}\\
=&(n)^{-1}\sum_{i=1}^{n}\mathrm{vec}\{(\mathbf{Y}_i)_{\cdot,1:j}\}^{\mathrm{T}}\mathrm{vec}\{(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})(\mathbf{Y}_i)_{\cdot,1:j}(\widehat{\mathbf{L}}_j-\mathbf{L}_j)\mathbf{e}_l^{\mathrm{T}}\}\\
=&(n)^{-1}\sum_{i=1}^{n}\mathrm{vec}\{(\mathbf{Y}_i)_{\cdot,1:j}\}^{\mathrm{T}}[\{(\widehat{\mathbf{L}}_j-\mathbf{L}_j)\mathbf{e}_l^{\mathrm{T}}\}^{\mathrm{T}}\otimes(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})]\mathrm{vec}\{(\mathbf{Y}_i)_{\cdot,1:j}\}\\
=&(n)^{-1}\sum_{i=1}^{n}\mathrm{tr}([\{(\widehat{\mathbf{L}}_j-\mathbf{L}_j)\mathbf{e}_l^{\mathrm{T}}\}^{\mathrm{T}}\otimes(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})]\mathrm{vec}\{(\mathbf{Y}_i)_{\cdot,1:j}\}\mathrm{vec}\{(\mathbf{Y}_i)_{\cdot,1:j}\}^{\mathrm{T}})\\
=&\mathrm{tr}[\{(\widehat{\mathbf{L}}_j-\mathbf{L}_j)\mathbf{e}_l^{\mathrm{T}}\}^{\mathrm{T}}\mathbf{R}_{1:j,1:j}\otimes(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})\mathbf{W}]\{1+o_p(1)\}\\
=&\mathrm{tr}[\{(\widehat{\mathbf{L}}_j-\mathbf{L}_j)\mathbf{e}_l^{\mathrm{T}}\}^{\mathrm{T}}\mathbf{R}_{1:j,1:j}]\mathrm{tr}\{(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})\mathbf{W}\}\{1+o_p(1)\}.
\end{aligned}
\tag{S3.6}
$$

Following Theorem 2, we have $\|\widehat{\mathbf{L}}_j-\mathbf{L}_j\|_2=O_p\{(\log m/n)^{1/2}\}$ and $\|\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1}\|_F=O_p\{(K\log m/n)^{1/2}\}$, so we have $|\mathrm{tr}[\{(\widehat{\mathbf{L}}_j-\mathbf{L}_j)\mathbf{e}_l^{\mathrm{T}}\}^{\mathrm{T}}\mathbf{R}_{1:j,1:j}]|=|(\widehat{\mathbf{L}}_j-\mathbf{L}_j)^{\mathrm{T}}\mathbf{R}_{1:j,1:j}\mathbf{e}|\leq\lambda_{\max}(\mathbf{R}_{1:j,1:j})\|(\widehat{\mathbf{L}}_j-\mathbf{L}_j)\|_2=O_p\{(\log m/n)^{1/2}\}$, where the last equation is due to $\|\widehat{\mathbf{L}}_j-\mathbf{L}_j\|_2=O_p\{(\log m/n)^{1/2}\}$, and $|\mathrm{tr}\{(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})\mathbf{W}\}|\leq\lambda_{\max}(\mathbf{W})\sqrt{K}\|\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1}\|_F=O_p\{K(\log m/n)^{1/2}\}$. Thus,

$$
|(nK)^{-1}\sum_{i=1}^{n}\mathbf{e}_l^{\mathrm{T}}(\mathbf{Y}_i)_{\cdot,1:j}^{\mathrm{T}}(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})(\mathbf{Y}_i)_{\cdot,1:j}(\widehat{\mathbf{L}}_j-\mathbf{L}_j)|=O_p(\log m/n). \tag{S3.7}
$$

We now treat the second term in (S3.5). Note that $\mathbf{e}_l^{\mathrm{T}}\mathbf{R}_{1:j,1:j}\mathbf{L}_j=(\mathbf{R}\mathbf{L}^{\mathrm{T}})_{l,j}=(\mathbf{L}^{-1})_{l,j}=0$ for $l<j$. Thus, following the same derivation as in (S3.6), for $l<j$,

we get

$$
\begin{aligned}
&|\frac{1}{nK}\sum_{i=1}^{n}\mathbf{e}_l^{\mathrm{T}}(\mathbf{Y}_i)_{\cdot,1:j}^{\mathrm{T}}(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})(\mathbf{Y}_i)_{\cdot,1:j}\mathbf{L}_j|\\
=&\mathrm{tr}[\{\mathbf{L}_j\mathbf{e}_l^{\mathrm{T}}\}^{\mathrm{T}}\mathbf{R}_{1:j,1:j}]\mathrm{tr}\{(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})\mathbf{W}\}\{1+o_p(1)\}\\
=&\mathrm{tr}(\mathbf{e}_l^{\mathrm{T}}\mathbf{L}_j^{\mathrm{T}}\mathbf{R}_{1:j,1:j})\mathrm{tr}\{(\widehat{\mathbf{W}}^{-1}-\mathbf{W}^{-1})\mathbf{W}\}\{1+o_p(1)\}=0. \qquad \text{(S3.8)}
\end{aligned}
$$

We now consider the third term in (S3.5). Because $\mathrm{E}(\mathbf{Y}_i)=\mathbf{0}$, and $\mathrm{Var}\{\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})\}=\mathbf{W}\otimes\mathbf{R}$, we have $\mathrm{Var}\{\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}}(\mathbf{W}^{-1})^{1/2})\}=\mathbf{I}\otimes\mathbf{R}$, Lemma S2.2 leads to

$$
\max_{s,t}|\{\sum_{i=1}^{n}\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})^{\mathrm{T}}/n-\mathbf{W}\otimes\mathbf{R}\}_{s,t}|=O_p\{(\log m/n)^{1/2}\}.
$$

For $\mathbf{A}\equiv\mathbf{W}^{-1}\otimes\mathbf{I}_J$ and $\mathbf{B}\equiv\mathbf{I}_{KJ}$, we have $\mathbf{A}\{\sum_{i=1}^{n}\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})^{\mathrm{T}}/n-\mathbf{W}\otimes\mathbf{R}\}\mathbf{B}=n^{-1}\sum_{i=1}^{n}\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1})\mathrm{vec}(\mathbf{Y}_i^{\mathrm{T}})^{\mathrm{T}}-\mathbf{I}_K\otimes\mathbf{R}$, with its K size $J\times J$ diagonal blocks summing to $n^{-1}\sum_{i=1}^{n}\mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i-K\mathbf{R}$. Thus, Lemma S2.2 also yields $\max_{s,t}|\{\sum_{i=1}^{n}\mathbf{Y}_i^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{Y}_i)/(nK)-\mathbf{R}\}_{s,t}|=O_p\{(\log m/n)^{1/2}\}$. Theorem 2 ensures $\|\widehat{\mathbf{L}}_j-\mathbf{L}_j\|_2=O_p\{(\log m/n)^{1/2}\}$, hence $|\mathbf{e}_l^{\mathrm{T}}\mathbf{R}_{1:j,1:j}(\widehat{\mathbf{L}}_j-\mathbf{L}_j)|\le\sigma_{\max}(\mathbf{R})\|\widehat{\mathbf{L}}_j-\mathbf{L}_j\|_2=O_p\{(\log m/n)^{1/2}\}$. Combining these results, we get

$$
\begin{aligned}
&|\frac{1}{nK}\sum_{i=1}^{n}\mathbf{e}_l^{\mathrm{T}}(\mathbf{Y}_i)_{\cdot,1:j}^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{Y}_i)_{\cdot,1:j}(\widehat{\mathbf{L}}_j-\mathbf{L}_j)|\\
\le&|\mathbf{e}_l^{\mathrm{T}}\{\frac{1}{nK}\sum_{i=1}^{n}(\mathbf{Y}_i)_{\cdot,1:j}^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{Y}_i)_{\cdot,1:j}-\mathbf{R}_{1:j,1:j}\}(\widehat{\mathbf{L}}_j-\mathbf{L}_j)|+|\mathbf{e}_l^{\mathrm{T}}\mathbf{R}_{1:j,1:j}(\widehat{\mathbf{L}}_j-\mathbf{L}_j)|\\
=&O_p(j^{1/2}\log m/n)+O_p\{(\log m/n)^{1/2}\}. \qquad \text{(S3.9)}
\end{aligned}
$$

For the last term in (S3.5),

$$
\begin{aligned}
&|(nK)^{-1}\sum_{i=1}^{n}\mathbf{e}_l^{\mathrm{T}}(\mathbf{Y}_i)_{\cdot,1:j}^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{Y}_i)_{\cdot,1:j}\mathbf{L}_j|\\
\leq&|\mathbf{e}_l^{\mathrm{T}}\{(nK)^{-1}\sum_{i=1}^{n}(\mathbf{Y}_i)_{\cdot,1:j}^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{Y}_i)_{\cdot,1:j}-\mathbf{R}_{1:j,1:j}\}\mathbf{L}_j|+|\mathbf{e}_l^{\mathrm{T}}\mathbf{R}_{1:j,1:j}\mathbf{L}_j|\\
\leq&\max_{s,t}|\{(nK)^{-1}\sum_{i=1}^{n}(\mathbf{Y}_i)_{\cdot,1:j}^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{Y}_i)_{\cdot,1:j}-\mathbf{R}_{1:j,1:j}\}_{s,t}|\|\mathbf{L}_j\|_1\\
=&O_p\{(d_j+1)(\log m/n)^{1/2}\},
\end{aligned}
\tag{S3.10}
$$

where the last inequality used $\mathbf{e}_l^{\mathrm{T}}\mathbf{R}_{1:j,1:j}\mathbf{L}_j=0$, and the last equality used $\|\mathbf{L}_j\|_0=d_j+1$ and $|L_{j,l}|\leq\sigma_{\max}(\mathbf{L})=O(1)$ for any $j-d_j\leq l\leq j$. Inserting (S3.7), (S3.8), (S3.9) and (S3.10) into (S3.5), noting that $\max_j d_j=O(1)$ in Assumption C1, under the condition in Theorem 2 that $J\log m/n=o(1)$, we obtain $|\lambda_2\widehat{A}_{l,l}/2|=O_p\{j^{1/2}\log m/n+(\log m/n)^{1/2}\}=O_p\{(\log m/n)^{1/2}\}$. Now for $\lambda_2=C(\log m/n)^{1/2}$, where C is a sufficiently large constant, we have $|\widehat{A}_{l,l}|<1$ for sufficiently large n. Recall that $\widehat{\mathbf{A}}_{1:l-1,l}=\mathbf{0}$, so $\|\widehat{\mathbf{A}}_{1:l,l}\|_2<1$ for sufficiently large n. □

Proof of Lemma 2. Consider the same objective function $L(\tau,\mathbf{V}_j,\mathbf{L}_j;\vartheta,\boldsymbol{\phi},\mathbf{A})$ as in Lemma 1, which is jointly convex at $(\tau,\mathbf{V}_j,\mathbf{L}_j)$, and it is strictly convex at τ and $\mathbf{V}_j$. Therefore, the solution $\widehat{\tau},\widehat{\mathbf{V}}_j$ of minimizing the objective function is unique. Thus, $\widehat{L}_{jj}$ and $\mathbf{Y}^*_{\cdot,1:j}\widehat{\mathbf{L}}_j$ are also unique due to (S3.3). Assume that $\widehat{\mathbf{L}}_j$ and $\widetilde{\mathbf{L}}_j$ are two solutions of minimizing the objective function $f(\mathbf{L}_j)$, then $f(\widehat{\mathbf{L}}_j)=f(\widetilde{\mathbf{L}}_j)$, hence $p(\widehat{\mathbf{L}}_j)=p(\widetilde{\mathbf{L}}_j)$. Note that (S3.4) leads to

$$
\sum_{l=1}^{j-1}\langle\widehat{\mathbf{A}}_{\cdot,l},\widehat{\mathbf{L}}_j\rangle=p(\widehat{\mathbf{L}}_j)=p(\widetilde{\mathbf{L}}_j)=\sum_{l=1}^{j-1}\langle\widehat{\mathbf{A}}_{\cdot,l},\widetilde{\mathbf{L}}_j\rangle=\sum_{l=1}^{j-1}\|(\widetilde{\mathbf{L}}_j)_{1:l}\|_2. \tag{S3.11}
$$

Since for any $l \leq j - \widehat{d}_j - 1$, $\widehat{L}_{jl} = 0$ and $\|\widehat{\mathbf{A}}_{1:l,l}\|_2 < 1$ by the assumption of the lemma, and $\|\widehat{\mathbf{A}}_{1:l,l}\|_2 \leq 1$ in general, for (S3.11) to hold, we must have $(\widetilde{\mathbf{L}}_j)_{1:l} = \mathbf{0}$ for all $l \leq j - \widehat{d}_j - 1$. This means $\widetilde{d}_j \leq \widehat{d}_j$. □

Proof of Theorem 3. Let $\widetilde{\mathbf{L}}_j$ be another solution to $\min_{L_{jj}>0, \mathbf{L}_j \in \mathcal{R}^j} f(\mathbf{L}_j)$ with bandwidth $\widetilde{d}_j$. Following Lemma 1, when $\lambda_2 = C(\log m/n)^{1/2}$ and C is sufficiently large, $\|\widehat{\mathbf{A}}_{1:l,l}\|_2 < 1$ for $l < j - \widehat{d}_j - 1$. By Lemma 2, $(\widetilde{\mathbf{L}}_j)_{1:j-\widehat{d}_j-1} = \mathbf{0}$. Therefore, we can write $\widetilde{\mathbf{L}}_j$ as $\widetilde{\mathbf{L}}_j = (\mathbf{0}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{\gamma}$ is a $\widehat{d}_j + 1$-dimensional parameter vector. Subsequently, the objective function $f(\mathbf{L}_j)$ can be equivalently written as

$$\min_{\boldsymbol{\gamma} \in \mathcal{R}^{\widehat{d}_j+1}, \gamma_{\widehat{d}_j+1}>0} \{-2\log \gamma_{\widehat{d}_j+1} + \frac{1}{nK}\|\mathbf{Y}^*_{\cdot,\widehat{D}}\boldsymbol{\gamma}\|_2^2 + \lambda_2 \sum_{l=1}^{\widehat{d}_j} \|(\boldsymbol{\gamma})_{1:l}\|_2\}.$$

Detailed calculation reveals that the Hessian matrices of the first and third terms are non-negative definite, while the Hessian matrix of second term is $2\mathbf{Y}^{*\mathrm{T}}_{\cdot,\widehat{D}}\mathbf{Y}^*_{\cdot,\widehat{D}}/nK$, which is strictly positive definite since $\mathbf{Y}^*_{\cdot,\widehat{D}}$ has full column rank. Thus, we minimize a strictly convex function of $\boldsymbol{\gamma}$ hence the minimum $\widehat{\boldsymbol{\gamma}}$ is unique. This indicates that $\widetilde{\mathbf{L}}_j = \widehat{\mathbf{L}}_j$, hence $\widehat{\mathbf{L}}_j$ is unique. □

S4 Bandwidth recovery: Proof of Theorem 4.

Proof. Recall that d_j is the true bandwidth of the jth row for the lower triangle matrix $\mathbf{L}$ and $j - d_j - 1$ is the number of zero elements in $\mathbf{L}_j$. Let $\widehat{\mathbf{L}}$ be the solution for the optimization problem in (2.5) and let $\widehat{\mathbf{L}}_j = \widehat{\mathbf{L}}^{\mathrm{T}}_{j,1:j}$. Let $\widetilde{\mathbf{L}}$ and $\widetilde{\mathbf{L}}_j$ be the corresponding constrained solution of (2.5) under the true bandwidth $d_j, j = 1, \ldots, J$. Because minimizing (2.5) is equivalent to the minimization problems in

Lemma 1, so $\widehat{\mathbf{L}}_j$ and $\widetilde{\mathbf{L}}_j$ both are the solutions described in Lemma 1. Following Lemma 1, there exist $\widehat{\mathbf{A}} \in \mathcal{R}^{j\times(j-1)}$ and $\widetilde{\mathbf{A}} \in \mathcal{R}^{j\times(j-1)}$, so that $(\widehat{\mathbf{L}}_j, \widehat{\mathbf{A}})$ and $(\widetilde{\mathbf{L}}_j, \widetilde{\mathbf{A}})$ both satisfy (3.2).

Next, we construct one specific $\widetilde{\mathbf{A}}$. We can verify that $\widetilde{\mathbf{L}}_j = (\mathbf{0}_{j-d_j-1}^{\mathrm{T}}, \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}})^{\mathrm{T}}$, where $\widetilde{\boldsymbol{\gamma}} \in \mathcal{R}^{d_j+1}$ and is given as

$$\widetilde{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}\in\mathcal{R}^{d_j+1}}{\operatorname{argmin}}\{-2\log\boldsymbol{\gamma}_{d_j+1} + \frac{1}{nK}\|\mathbf{Y}^*_{\cdot,j-d_j:j}\boldsymbol{\gamma}\|_2^2 + \lambda_2 p(\boldsymbol{\gamma})\}.$$

There exists $\widetilde{\mathbf{B}} \in \mathcal{R}^{(d_j+1)\times d_j}$, such that $\widetilde{\mathbf{B}}_{(l+1):(d_j+1),l} = \mathbf{0}$, $\widetilde{\mathbf{B}}_{1:l,l} = (\widetilde{\boldsymbol{\gamma}})_{1:l}/\|(\widetilde{\boldsymbol{\gamma}})_{1:l}\|_2$ if $\|(\widetilde{\boldsymbol{\gamma}})_{1:l}\|_2 \neq 0$, $\|(\widetilde{\boldsymbol{\gamma}})_{1:l}\|_2 \leq 1$, for $1 \leq l \leq d_j$, and

$$-\frac{2}{\widetilde{\boldsymbol{\gamma}}_{d_j+1}}\mathbf{e}_{d_j+1} + \frac{2}{nK}\mathbf{Y}^{*\mathrm{T}}_{\cdot,j-d_j:j}\mathbf{Y}^*_{\cdot,j-d_j:j}\widetilde{\boldsymbol{\gamma}} + \lambda_2\sum_{l=1}^{d_j}\widetilde{\mathbf{B}}_{\cdot,l} = \mathbf{0}. \tag{S3.12}$$

For $l = 1, \cdots, j-d_j-1$, we set $\widetilde{\mathbf{A}}_{l',l} \equiv \mathbf{0}$, for any $l \neq l'$ and $\widetilde{\mathbf{A}}_{l,l} \equiv -2(\mathbf{Y}^{*\mathrm{T}}_{,l}\mathbf{Y}^*_{,1:j}\widetilde{\mathbf{L}}_j)/(nK\lambda_2)$. Further, let $\widetilde{\mathbf{A}}_{\cdot,j-d_j:j-1} \equiv (\mathbf{0}^{\mathrm{T}}, \widetilde{\mathbf{B}}^{\mathrm{T}})^{\mathrm{T}} \in \mathcal{R}^{j\times d_j}$. We can verify that for $l = j - d_j, \cdots, j-1$, $\widetilde{\mathbf{A}}_{l+1:j,l} = \mathbf{0}$, $\|\widetilde{\mathbf{A}}_{1:l,l}\|_2 \leq 1$, and $\widetilde{\mathbf{A}}_{1:l,l} = (\widetilde{\mathbf{L}}_j)_{1:l,l}/\|(\widetilde{\mathbf{L}}_j)_{1:l,l}\|_2$ when $(\widetilde{\mathbf{L}}_j)_{1:l} \neq \mathbf{0}$. Following from (S3.12), the constructed $\widetilde{\mathbf{A}}$ and the constrained solution $\widetilde{\mathbf{L}}_j$ satisfy (3.2). Now taking $\lambda_2 = C(\log m/n)^{1/2}$ for a sufficiently large constant C, we obtain $\|\widetilde{\mathbf{A}}_{l,l}\|_2 < 1$ for $l = 1, \ldots, j-d_j-1$ following Lemma 1. Further, Theorem 3 ensures that $\widetilde{\mathbf{L}}_j$ is unique. Because $\widetilde{\mathbf{L}}_j, \widetilde{\mathbf{A}}$ and $\widehat{\mathbf{L}}_j, \widehat{\mathbf{A}}$ are both solutions defined in Lemma 1, and $\|\widetilde{\mathbf{A}}_{1:l,l}\|_2 < 1$ for $l = 1, \ldots, j-d_j-1$, by Lemma 2, $\widehat{d}_j \leq d_j$. We next prove that $\widehat{d}_j = d_j$ for all $j = 2, \ldots, J$ by contradiction. Without loss of generality, assume $\widehat{d}_j < d_j$. We have $\|\widehat{\mathbf{L}}_j - \mathbf{L}_j\|_2 = O_p\{(\log m/n)^{1/2}\}$ by Theorem 2 and

$$\|\widehat{\mathbf{L}}_j - \mathbf{L}_j\|_2^2 = \sum_{l=j-d_j}^{j-\widehat{d}_j-1} L_{jl}^2 + \sum_{l=j-\widehat{d}_j}^{j} |\widehat{L}_{jl} - L_{jl}|^2$$

by construction. Since $\min_{j\in\{2,\cdots,J\}} \min_{l\geq j-d_j} |L_{jl}| > \lambda_2$, we have

$$\begin{aligned}
\sum_{l=j-\widehat{d}_j}^{j} |\widehat{L}_{jl} - L_{jl}|^2 &= \|\widehat{\mathbf{L}}_j - \mathbf{L}_j\|_2^2 - \sum_{l=j-d_j}^{j-\widehat{d}_j-1} |L_{jl}|^2 \leq \|\widehat{\mathbf{L}}_j - \mathbf{L}_j\|_2^2 - \lambda_2^2(d_j - \widehat{d}_j) \\
&\leq O_p\left(\frac{\log m}{n}\right) - C^2(d_j - \widehat{d}_j)\frac{\log m}{n} < 0
\end{aligned}$$

for sufficiently large constant C, which is a contradiction. Thus, $\mathrm{pr}(\sup_j |\widehat{d}_j - d_j| = 0) \to 1$. □

S5 Examples.

EEG example. In brain imaging studies, it is common practice to utilize electroencephalography (EEG) on individuals. In this particular study, we used a public EEG dataset from the UCI machine learning repository to illustrate the proposed model. Each individual in the study had 64 electrodes placed on their scalp, and measurements were collected at a frequency of 256 Hz (3.9ms epoch) per second for a specified duration. Consequently, the observations for each individual were represented by a sequence of $K = 64$ vectors, each of length $J = 256$ time points. The EEG measurements of the ith individual at time j can be denoted as $\mathbf{Y}_{ij\cdot} \equiv (Y_{ij1}, \ldots Y_{ijK})^{\mathrm{T}}$, representing a K-dimensional random vector. Here, i ranges from 1 to n, and j ranges from 1 to J. Let $\mathbf{Y}_i$ denote the $K \times J$ random matrix associated with individual i over all time points. We assume that $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent and identically distributed (iid), which is commonly observed in EEG data research. We now vectorize each $\mathbf{Y}_i$ by forming $\text{vec}(\mathbf{Y}_i) \equiv (\mathbf{Y}_{i1\cdot}^{\mathrm{T}}, \ldots, \mathbf{Y}_{iJ\cdot}^{\mathrm{T}})^{\mathrm{T}} \in \mathcal{R}^{KJ}$. In EEG data, for $\mathbf{\Sigma}$, when the time points are closely spaced, it can be reasonable to assume that the correlation matrix remains unchanged for the K positions, and only the absolute signal strengths vary over time. Thus, we can express $\mathbf{\Sigma}^{(jl)}$ as $r_{jl}\mathbf{W}$ for all j, l, where r_{jl} represents a constant indicating the signal amplication at different time points. For convenience, we define $\mathbf{W}$ as $\mathbf{\Sigma}^{(11)}/\mathbf{\Sigma}_{1,1}$, where $\mathbf{\Sigma}_{1,1}$ represents the entry at the $(1,1)$ position of $\mathbf{\Sigma}$. Based on the previous analysis, it follows that $\mathbf{\Sigma}$ can be expressed as $\mathbf{R} \otimes \mathbf{W}$, where both $\mathbf{R}$ and $\mathbf{W}$ are positive definite matrices. Consequently, the precision matrix $\mathbf{\Sigma}^{-1}$ can be expressed as $\mathbf{R}^{-1} \otimes \mathbf{W}^{-1}$. In EEG research, the precision matrix is often of significant interest, as it captures the conditional correlation structure. For instance, the entry (j, h) of $\mathbf{\Sigma}^{-1}$ pertains to the correlation between the jth and the hth components of the random vector while conditioned on all other components. If the (j, h) entry is zero, it implies conditional uncorrelation of the jth and the hth variables when the others taken into account. Furthermore, an element of $\mathbf{W}^{-1}$ represents the correlation between two brain regions while considering the influence of other regions. When K is large, it is commonly observed that only a few of these conditional correlations are non-zero. Therefore, it is assumed that $\mathbf{W}^{-1}$ exhibits sparsity. On the other hand, the (j, l) element of $\mathbf{R}^{-1}$ represents the correlation between the jth and the lth time points while considering the effects from all other observed time points. When the distance between j and l, denotes as $|j - l|$, is large, the conditional correlation tends to approach zero. Consequently, we assume $\mathbf{R}_{j,l}^{-1} = 0$, for all $|j - l| > d$, where

d represents a constant that limits the time difference and is typically much smaller than J. As a result, $\mathbf{R}^{-1}$ forms a $J \times J$ banded matrix with a bandwidth of d.

Multivariate Time Series example. In the field of multivariate time series analysis, a response vector of interest, with a dimension of K, is measured repeatedly at various time points. For each $i = 1, \ldots, n$, let $Y_{ij\cdot}$ denote a K-dimensional random vector at the jth time point for $j = 1, \ldots, J$. Assuming an autoregressive process, we have $L_{1,1}\mathbf{y}_{i1\cdot} = \boldsymbol{\varepsilon}_{i1\cdot}$, and

$$L_{j,j}\mathbf{y}_{ij\cdot} = -\sum_{l=1}^{j-1} L_{jl}\mathbf{y}_{il\cdot} + \boldsymbol{\varepsilon}_{ij\cdot}, \quad j = 2, \ldots, J. \tag{S3.13}$$

Here $\boldsymbol{\varepsilon}_{ij\cdot} = (\boldsymbol{\varepsilon}_{ij1}, \ldots, \boldsymbol{\varepsilon}_{ijK})^{\mathrm{T}}$ represents the mean zero error vector, and we assume $\mathrm{cov}(\boldsymbol{\varepsilon}_{ij\cdot}) = \mathbf{W}$. Let $\boldsymbol{\varepsilon}_i \equiv (\boldsymbol{\varepsilon}_{i1\cdot}, \ldots, \boldsymbol{\varepsilon}_{iJ\cdot})$ represent a $K \times J$ matrix. We assume that $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ are independent and identically distributed (iid). Consequently, the expected value of $\mathrm{vec}(\boldsymbol{\varepsilon}_i)$ is zero and the covariance of $\mathrm{vec}(\boldsymbol{\varepsilon}_i)$ is represented by the Kronecker product $\mathbf{I}_J \otimes \mathbf{W}$. Furthermore, the autoregressive expression (S3.13) can be rewritten in matrix form as $\mathbf{Y}_i\mathbf{L}^{\mathrm{T}} = \boldsymbol{\varepsilon}_i$, where $\mathbf{Y}_i = (\mathbf{Y}_{i1\cdot}, \ldots, \mathbf{Y}_{iJ\cdot}) \in \mathcal{R}^{K\times J}$, and $\mathbf{L}$ is a lower triangle matrix with its (j,l) entry denotes as $L_{j,l}$ for all $1 \leq l \leq j \leq J$. Let $\boldsymbol{\Sigma} \equiv \mathrm{cov}\{\mathrm{vec}(\mathbf{Y}_i)\}$, it can be shown that $(\mathbf{L} \otimes \mathbf{I}_K)\boldsymbol{\Sigma}(\mathbf{L}^{\mathrm{T}} \otimes \mathbf{I}_K) = \mathrm{cov}\{(\mathbf{L} \otimes \mathbf{I}_K)\mathrm{vec}(\mathbf{Y}_i)\} = \mathrm{cov}\{\mathrm{vec}(\mathbf{Y}_i\mathbf{L}^{\mathrm{T}})\} = \mathrm{cov}\{\mathrm{vec}(\boldsymbol{\varepsilon}_i)\} = \mathbf{I}_J \otimes \mathbf{W}$. Consequently, we can deduce that $\boldsymbol{\Sigma} = (\mathbf{L}^{\mathrm{T}}\mathbf{L})^{-1} \otimes \mathbf{W}$. Let us consider the precision matrix $\boldsymbol{\Sigma}^{-1}$ of interest, where $\boldsymbol{\Sigma}^{-1}$ is defined as $\mathbf{R}^{-1} \otimes \mathbf{W}^{-1}$, with $\mathbf{R}^{-1}$ denoting $\mathbf{L}^{\mathrm{T}}\mathbf{L}$. In case where the autoregression process in (S3.13) has an order of d, the lower triangle matrix $\mathbf{L}$ satisfies $L_{j,l} = 0$ for $j - l > d$. Notably, d is significantly smaller than J, which implies that $\mathbf{R}^{-1}$ is a banded matrix with bandwidth d. Furthermore, it is commonly assumed that $\boldsymbol{\varepsilon}_i$ consists of white noise, resulting in $\mathbf{W}^{-1}$ being a diagonal matrix. However, for increased flexibility, we allow $\mathbf{W}^{-1}$ to have a small number of nonzero off-diagonal elements, thus making $\mathbf{W}^{-1}$ sparse.

S6 Simulation and real data studies.

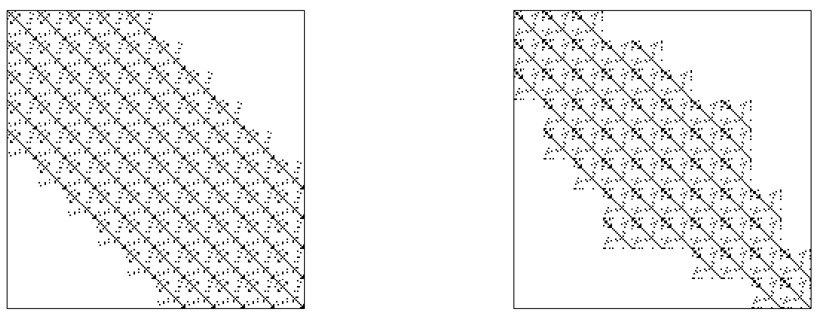

Figure 1: The true precision matrix for $K = 20, J = 10$ in Case 1 (left) and Case 2(right).

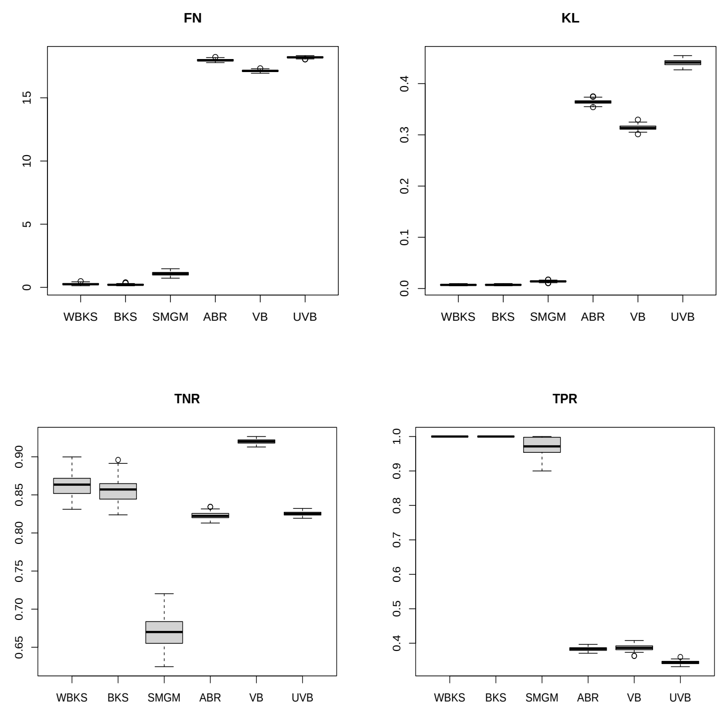

Figure 2: The boxplots of FN, KL, TNR and TPR values for six estimators at $K = 20, J = 10, n = 100$ in Case 1.

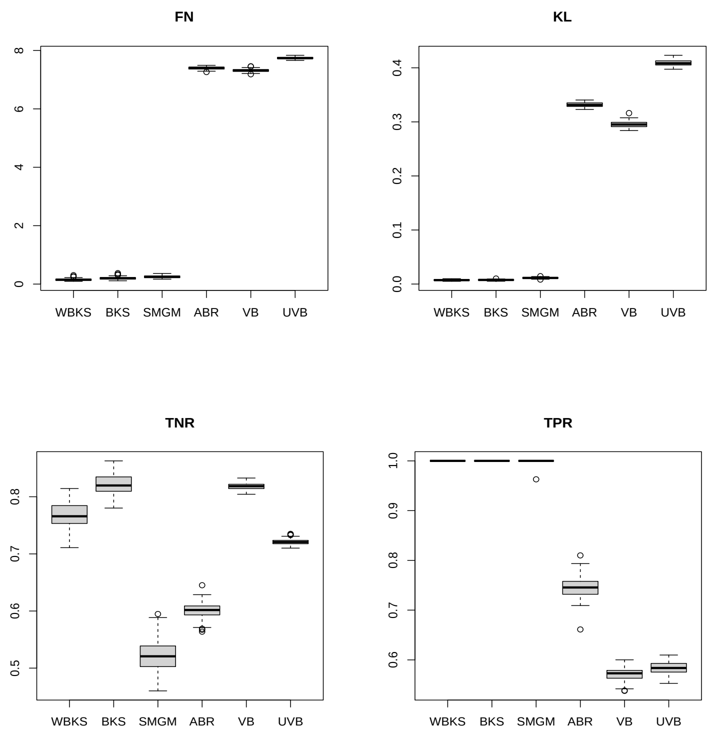

Figure 3: The boxplots of FN, KL, TNR and TPR values for six estimators at $K = 20, J = 10, n = 100$ in Case 2.

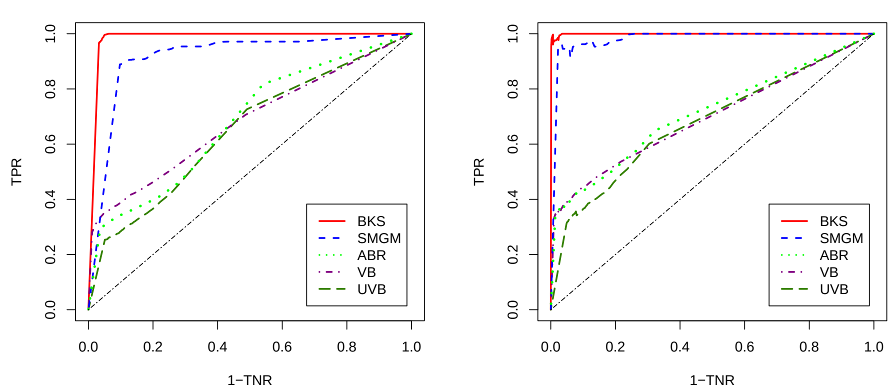

Figure 4: The ROC for six estimators at $K = 20, J = 10, n = 100$ in Case 1 (left) and Case 2 (right).

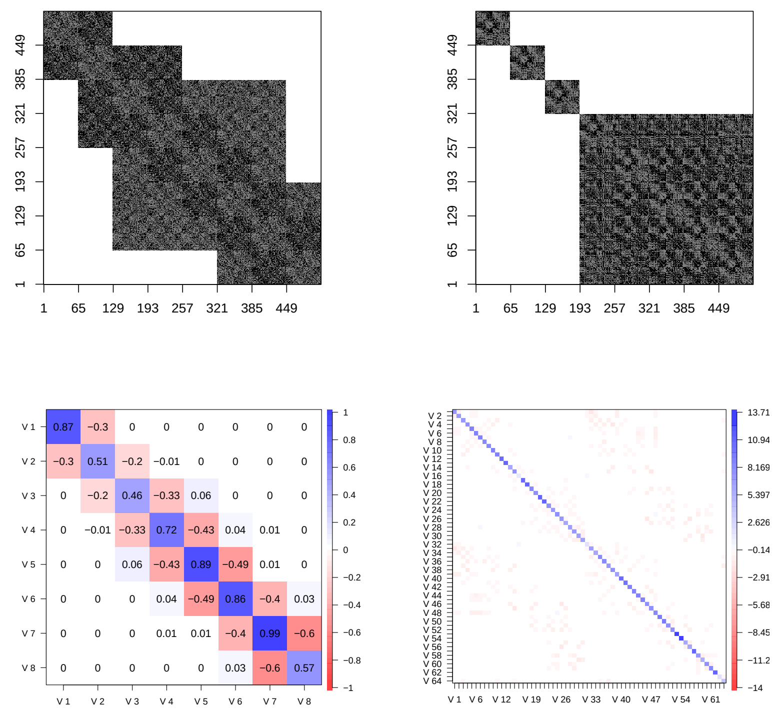

Figure 5: Estimated precision matrix based on WBKS, BKS (upper left) and SMGM (upper right). $\widehat{\mathbf{R}}^{-1}$ (lower left) and $\widehat{\mathbf{W}}^{-1}$ (lower right) by BKS.

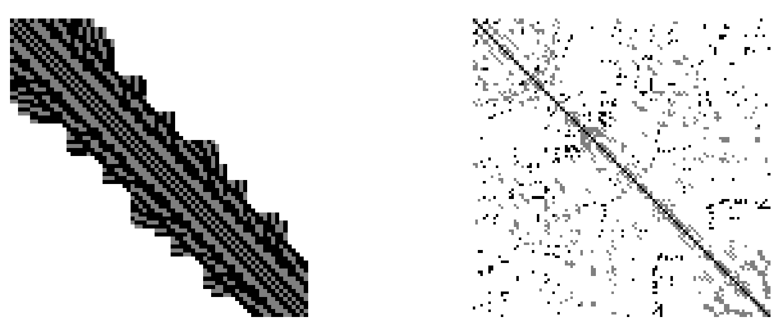

Figure 6: The estimation of precision matrix in the temporal dimension (J=74) (left) and precision matrix in the variable dimension (K=116) (right) obtained by BKS estimation.

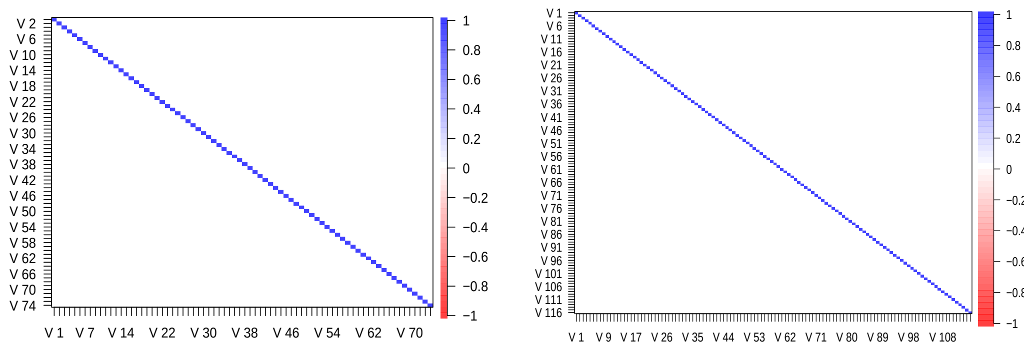

Figure 7: The estimation of precision matrix in the time dimension (J=74) (left) and precision matrix in the variable dimension (K=116) (right) obtained by SMGM estimation.

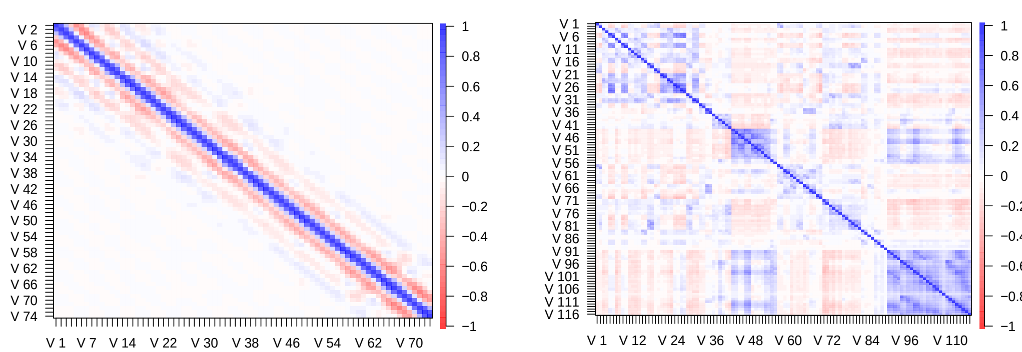

Figure 8: The estimation of correlation matrix in the time dimension (J=74) (left) and correlation matrix in the variable dimension (K=116) (right) obtained by BKS estimation.

References

Bickel, P. J. and E. Levina (2008). Regularized estimation of large covariance matrices. *Annals of Statistics 36*(1), 199–227.

Bien, J., F. Bunea, and L. Xiao (2016). Convex banding of the covariance matrix. *Journal of the American Statistical Association 111*(514), 834–845.

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

Yu, G. and J. Bien (2017). Learning local dependence in ordered data. *Journal of Machine Learning Research 18*(42), 1–60.