ESTIMATING LARGE PRECISION MATRICES VIA MODIFIED CHOLESKY DECOMPOSITION

Inha University and Seoul National University

Supplementary Material

S1 Notation

We introduce some notation here which will be used in the proofs. For any $(j-1)$-dimensional vector $a = (a_1, \ldots, a_{j-1})^T \in \mathbb{R}^{j-1}$, $B_{k,j}(a)$ is defined by $B_{k,j}(a) := (b_i = a_{j-1-k+i}, 1 \le i \le k)$. Let $\Omega_{0,n}$ be the true precision matrix and $\Omega_{0,n} = (I_p - A_{0,n})^T D_{0,n}^{-1}(I_p - A_{0,n})$ be its modified Cholesky decomposition with $A_{0,n} = (a_{0,jl})$ and $D_{0,n} = diag(d_{0,j})$. It is easy to check that the explicit forms of $a_{0,j} = (a_{0,j1}, \ldots, a_{0,j,j-1})^T$ and $d_{0,j}$ are

$$\begin{aligned} a_{0,j} &= \mathrm{Var}(X_{1,1:(j-1)})^{-1}\mathrm{Cov}(X_{1,1:(j-1)}, X_{1j}), \\ d_{0,j} &= \mathrm{Var}(X_{1j}) - \mathrm{Cov}(X_{1j}, X_{1,1:(j-1)})\mathrm{Var}(X_{1,1:(j-1)})^{-1}\mathrm{Cov}(X_{1,1:(j-1)}, X_{1j}), \end{aligned} \quad \text{(S1.1)}$$

where $X_{1,a:b} = (X_{1a}, \ldots, X_{1b})^T$ denotes the sub-vector of the first observation $X_1 = (X_{11}, \ldots, X_{1p})^T \in \mathbb{R}^p$ for any positive integers $1 \le a \le b \le p$. Since we assume $X_1, \ldots, X_n \overset{iid}{\sim} N_p(0, \Omega_{0,n}^{-1})$, $X_{1,a:b}$ can be replaced by $X_{i,a:b}$

for any $i = 2, \ldots, n$. For more details on the above expression (S1.1), refer to Bickel and Levina (2008) (pages 202 and 221).

For a given k, we denote

$$\begin{aligned} a_{0,j}^{(k)} &= \mathrm{Var}(X_{1,(j-k):(j-1)})^{-1}\mathrm{Cov}(X_{1,(j-k):(j-1)}, X_{1j}), \\ d_{0,jk} &= \mathrm{Var}(X_{1j}) - \mathrm{Cov}(X_{1j}, X_{1,(j-k):(j-1)})\mathrm{Var}(X_{1,(j-k):(j-1)})^{-1}\mathrm{Cov}(X_{1,(j-k):(j-1)}, X_{1j}). \end{aligned}$$

We denote the empirical estimators by $\widehat{\mathrm{Var}}(X_{1,(j-k):(j-1)}) = n^{-1}X_{\cdot,(j-k):(j-1)}^T X_{\cdot,(j-k):(j-1)}$ and $\widehat{\mathrm{Cov}}(X_{1,(j-k):(j-1)}, X_{1j}) = n^{-1}X_{\cdot,(j-k):(j-1)}^T X_{\cdot,j}$ for any $j = 2, \ldots, p$. For any $j = 1, \ldots, p$, we define $\widehat{\mathrm{Var}}(X_{1j}) = n^{-1}\|X_{\cdot,j}\|_2^2$.

S2 Proofs

S2.1 Proof of Proposition 1

Proof. First, we prove only the exponentially decreasing case, $\gamma(k) = Ce^{-\beta k}$ for some $\beta > 0$ and $C > 0$, because the proposition trivially holds for the exact banding case.

Suppose $\Omega_{0,n} = (\omega_{0,ij}) \in \mathcal{U}(\epsilon_0, \gamma)$ and let $\Omega_{0,n} = (I_p - A_{0,n})^T D_{0,n}^{-1}(I_p -$

$A_{0,n}$) where $A_{0,n} = (a_{0,ij})$ and $D_{0,n} = diag(d_{0,j})$. Note that

$$\begin{aligned} \|D_{0,n}^{-1}\| &= \max_j d_{0,j}^{-1} \\ &= \max_j \left\| \mathrm{Var}^{1/2}(X_{1,1:j}) \begin{pmatrix} -a_{0,j} \\ 1 \end{pmatrix} \right\|_2^{-2} \\ &\le \max_j \lambda_{\min}(\mathrm{Var}(X_{1,1:j}))^{-1} \cdot \left\| \begin{pmatrix} -a_{0,j} \\ 1 \end{pmatrix} \right\|_2^{-2} \\ &\le \max_j \lambda_{\min}(\mathrm{Var}(X_{1,1:j}))^{-1} \\ &\le \lambda_{\min}(\Omega_{0,n})^{-1} \ \le \ \epsilon_0^{-1} \end{aligned}$$

and

$$\begin{aligned} \|A_{0,n}\|_{\max} &\le \max_j \|a_{0,j}\|_2 = \max_j \|\mathrm{Var}(X_{1,1:(j-1)})^{-1}\mathrm{Cov}(X_{1,1:(j-1)}, X_{1j})\|_2 \\ &\le \max_j \|\mathrm{Var}(X_{1,1:(j-1)})^{-1}\| \|\mathrm{Cov}(X_{1,1:(j-1)}, X_{1j})\|_2 \\ &\le \max_j \|\mathrm{Var}(X_{1,1:(j-1)})^{-1}\| \|\mathrm{Var}(X_{1,1:j})\| \ \le \ \epsilon_0^{-2} \end{aligned}$$

by the assumption $\epsilon_0 \le \lambda_{\min}(\Omega_{0,n}) \le \lambda_{\max}(\Omega_{0,n}) \le \epsilon_0^{-1}$.

Furthermore,

$$\|A_{0,n} - B_k(A_{0,n})\|_\infty = \max_i \sum_{j<i-k} |a_{0,ij}| \ \le \ \gamma(k), \tag{S2.2}$$

$$\|A_{0,n} - B_k(A_{0,n})\|_1 = \max_j \sum_{i>j+k} |a_{0,ij}| \ \le \ \sum_{m=k}^{\infty} \gamma(m) \ \le \ C'\gamma(k), \tag{S2.3}$$

for some $C' > 1$ because $\gamma(k) = Ce^{-\beta k}$. Note that $\omega_{0,pp} = d_{0,p}^{-1}$ and

$$\omega_{0,ij} = \ -d_{0,j}^{-1} a_{0,ji} + \sum_{l=j+1}^{p} d_{0,l}^{-1} a_{0,li} a_{0,lj} \quad \text{for any } 1 \le i < j \le p. \tag{S2.4}$$

Then for $1 \le i < p$, define k so that $i = p - k - 1$. Then, $k \ge 0$ and

$$\begin{aligned} |\omega_{0,ip}| &= d_{0,p}^{-1}|a_{0,pi}| \\ &\le \epsilon_0^{-1}\gamma(k) \end{aligned}$$

by (S2.2). On the other hand, for $1 \le i < j \le p-1$, define k so that $j - i = k + 1$. Then, $k \ge 0$ and

$$\begin{aligned} |\omega_{0,ij}| &= |-d_{0,j}a_{0,ji} + \sum_{l=j+1}^{p} d_{0,l}^{-1} a_{0,li}a_{0,lj}| \\ &\le d_{0,j}^{-1}|a_{0,ji}| + \sum_{l=j+1}^{p} d_{0,l}^{-1}|a_{0,li}a_{0,lj}| \\ &\le \epsilon_0^{-3}\left(|a_{0,ji}| + \sum_{l=j+1}^{p}|a_{0,li}|\right) \\ &= \epsilon_0^{-3}\sum_{l=j}^{p}|a_{0,li}| \ \le\ \epsilon_0^{-3}C'\gamma(k) \end{aligned}$$

by (S2.3). Thus, we have

$$\begin{aligned} \|\Omega_{0,n} - B_k(\Omega_{0,n})\|_\infty &= \max_i \sum_{j:|i-j|>k}|\omega_{0,ij}| \\ &\le \max_i \sum_{j>i+k}|\omega_{0,ij}| + \max_i \sum_{j<i-k}|\omega_{0,ji}| \\ &\le 2\epsilon_0^{-3}C'\sum_{m=k}^{\infty}\gamma(m) \ \le\ C''\gamma(k) \end{aligned}$$

for some constant $C'' > 0$. This proves the first inequality.

Suppose $\Omega_{0,n} \in \mathcal{U}^*(\epsilon_0, \gamma)$. We need to prove that $\max_i \sum_{j<i-k}|a_{0,ij}| = \max_i \sum_{j=1}^{i-k-1}|a_{0,ij}| \le C\gamma(k)$ for some constant $C > 0$. Note that from

(S2.4), we have

$$d_{0,p}^{-1}\sum_{j=1}^{p-k-1}|a_{0,pj}| \;=\; \sum_{j=1}^{p-k-1}|\omega_{0,jp}| \;\le\; \gamma(k), \tag{S2.5}$$

for any $0 \le k \le p-2$. We will show that

$$d_{0,p-t}^{-1}\sum_{j=1}^{p-t-k-1}|a_{0,p-t,j}| \;\le\; \gamma(k)+\epsilon_0^{-2}\sum_{m=1}^{t}(1+\epsilon_0^{-2})^{m-1}\gamma(k+m) \tag{S2.6}$$

for any $1 \le t \le p-k-2$ for some $0 \le k \le p-3$. Then, (S2.5) and (S2.6) imply $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, C'\gamma)$ for some $C' > 0$ because $\max_j d_{0,j} \le \max_j \mathrm{Var}(X_{1j}) \le \epsilon_0^{-1}$ and we assume that $\gamma(k) = Ce^{-\beta k}$ and $\beta > \log(\epsilon_0^{-2}+1)$.

By (S2.4) and the assumption $\Omega_{0,n} \in \mathcal{U}^*(\epsilon_0,\gamma)$,

$$\sum_{j=1}^{p-k-2}|-d_{0,p-1}^{-1}a_{0,p-1,j}+d_{0,p}^{-1}a_{0,pj}a_{0,p,p-1}| \;=\; \sum_{j=1}^{p-k-2}|\omega_{0,j,p-1}| \;\le\; \gamma(k) \tag{S2.7}$$

for any $0 \le k \le p-3$. Thus, (S2.5) and (S2.7) imply that

$$\begin{aligned} d_{0,p-1}^{-1}\sum_{j=1}^{p-k-2}|a_{0,p-1,j}| \;&\le\; \sum_{j=1}^{p-k-2}|-d_{0,p-1}^{-1}a_{0,p-1,j}+d_{0,p}^{-1}a_{0,pj}a_{0,p,p-1}| + \sum_{j=1}^{p-k-2}|d_{0,p}^{-1}a_{0,pj}a_{0,p,p-1}| \\ &\le\; \gamma(k)+\epsilon_0^{-2}\gamma(k+1) \end{aligned}$$

because $\Omega_{0,n} \in \mathcal{U}^*(\epsilon_0,\gamma)$ means $|a_{0,p,p-1}| \le \|A_{0,n}\|_{\max} \le \epsilon_0^{-2}$. Thus, (S2.6) holds for $t=1$. Now assume that (S2.6) holds for $t-1$ and consider for the case of t. Note that

$$\begin{aligned} \gamma(k) \;&\ge\; \sum_{j=1}^{p-t-k-1}|\omega_{0,j,p-t}| \\ &=\; \sum_{j=1}^{p-t-k-1}\left|-d_{0,p-t}^{-1}a_{0,p-t,j}+\sum_{l=p-t+1}^{p}d_{0,l}^{-1}a_{0,lj}a_{0,l,p-t}\right|, \end{aligned}$$

which implies that

$$
\begin{aligned}
& d_{0,p-t}^{-1} \sum_{j=1}^{p-t-k-1} |a_{0,p-t,j}| \\
& \leq \gamma(k) + \sum_{j=1}^{p-t-k-1} \sum_{l=p-t+1}^{p} d_{0,l}^{-1} |a_{0,lj} a_{0,l,p-t}| \\
& \leq \gamma(k) + \epsilon_0^{-2} \sum_{l=p-t+1}^{p} d_{0,l}^{-1} \sum_{j=1}^{p-t-k-1} |a_{0,lj}| \\
& = \gamma(k) + \epsilon_0^{-2} \sum_{t_1=0}^{t-1} d_{0,p-t_1}^{-1} \sum_{j=1}^{p-t_1-(k+t-t_1)-1} |a_{0,p-t_1,j}| \\
& \leq \gamma(k) + \epsilon_0^{-2} \gamma(k+t) \\
& + \epsilon_0^{-2} \sum_{t_1=1}^{t-1} \left(\gamma(k+t-t_1) + \epsilon_0^{-2} \sum_{m=1}^{t_1} (1+\epsilon_0^{-2})^{m-1} \gamma(k+t-t_1+m) \right).
\end{aligned} \tag{S2.8}
$$

In (S2.8), one can check that the coefficient of $\gamma(k+t-t')$ is

$$
\epsilon_0^{-2} + \epsilon_0^{-4} \sum_{m=1}^{t-t'-1} (1+\epsilon_0^{-2})^{m-1} \quad = \quad \epsilon_0^{-2}(1+\epsilon_0^{-2})^{t-t'-1}
$$

for $0 \leq t' \leq t-1$, and the coefficient of $\gamma(k)$ is 1. Thus,

$$
d_{0,p-t}^{-1} \sum_{j=1}^{p-t-k-1} |a_{0,p-t,j}| \quad \leq \quad \gamma(k) + \epsilon_0^{-2} \sum_{m=1}^{t} (1+\epsilon_0^{-2})^{m-1} \gamma(k+m).
$$

This completes the proof by induction.

Now suppose that $\gamma(k) = Ck^{-\alpha}$ for some constant $C > 0$ and $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$. We will show that $\Omega_{0,n} \in \mathcal{U}^*(\epsilon_0, \gamma')$, where $\gamma'(k) = C'k^{1-\alpha}$ for some constant $C' > 0$. Let $Q = D_{0,n}^{-1/2}(I_p - A_{0,n})$, then by the proof of

Lemma 2 in Liu and Ren (2017),

$$\begin{aligned} & \|\Omega_{0,n} - B_k(\Omega_{0,n})\|_\infty \\ \leq \;& \|(Q - B_k(Q))^T Q\|_\infty + \|Q^T(Q - B_k(Q))\|_\infty + \|(Q - B_k(Q))^T(Q - B_k(Q))\|_\infty \quad \text{(S2.9)} \\ + \;& \|B_k\big[(Q - B_k(Q))^T B_k(Q)\big]\|_\infty + \|B_k\big[B_k(Q)^T(Q - B_k(Q))\big]\|_\infty \quad \text{(S2.10)} \\ + \;& \|B_k\big[(Q - B_k(Q))^T(Q - B_k(Q))\big]\|_\infty. \quad \text{(S2.11)} \end{aligned}$$

The two terms in (S2.10) are bounded above by $C'k^{1-2\alpha}$ for some constant $C' > 0$ by the proof of Lemma 2 in Liu and Ren (2017). With a slightly modified version of Lemma 24 and Lemma 25 in Liu and Ren (2017) by considering $\|\cdot\|_1$ and $\|\cdot\|_\infty$ instead of $\|\cdot\|$, one can show that the term (S2.11) is also bounded above by $C'k^{1-2\alpha}$. Three terms in (S2.9) are bounded above by $C'k^{1-\alpha}$ by the modified version of Lemma 24 in Liu and Ren (2017) and Lemma 8. It completes the proof. □

S2.2 Proof of Minimax Lower Bounds: Theorem 1 and Theorem 3

Proof of Theorem 1. We follow closely the line of a proof in Cai et al. (2010). Consider the polynomially decreasing case, $\gamma(k) = Ck^{-\alpha}$, first. Two parameter classes are considered depending on the relation between p and n.

For $\exp(n^{1/(2\alpha+1)}) \geq p$ case, we show that

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n}\in\mathcal{U}_{11}} \mathbb{E}_{0n}\|\widehat{\Omega}_n - \Omega_{0,n}\| \ \gtrsim \ \min\left(n^{-\alpha/(2\alpha+1)}, \left(\frac{p}{n}\right)^{1/2}\right), \quad (S2.12)$$

and for $\exp(n^{1/(2\alpha+1)}) \leq p$ case, we show that

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n}\in\mathcal{U}_{12}} \mathbb{E}_{0n}\|\widehat{\Omega}_n - \Omega_{0,n}\| \ \gtrsim \ \left(\frac{\log p}{n}\right)^{1/2} \quad (S2.13)$$

for some $\mathcal{U}_{11} \cup \mathcal{U}_{12} \subset \mathcal{U}(\epsilon_0, \gamma)$. Then, it gives a lower bound for the parameter space $\mathcal{U}(\epsilon_0, \gamma)$,

$$\begin{aligned}
\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n}\in\mathcal{U}(\epsilon_0,\gamma)} \mathbb{E}_{0n}\|\widehat{\Omega}_n - \Omega_{0,n}\| &\geq \inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n}\in\mathcal{U}_{11}\cup\mathcal{U}_{12}} \mathbb{E}_{0n}\|\widehat{\Omega}_n - \Omega_{0,n}\| \\
&\geq \inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n}\in\mathcal{U}_{11}} \mathbb{E}_{0n}\|\widehat{\Omega}_n - \Omega_{0,n}\| I\big(\exp(n^{1/(2\alpha+1)}) \geq p\big) \\
&\quad + \inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n}\in\mathcal{U}_{12}} \mathbb{E}_{0n}\|\widehat{\Omega}_n - \Omega_{0,n}\| I\big(\exp(n^{1/(2\alpha+1)}) < p\big) \\
&\gtrsim \min\left\{\left(\frac{\log p}{n}\right)^{1/2} + n^{-\alpha/(2\alpha+1)}, \left(\frac{p}{n}\right)^{1/2}\right\},
\end{aligned}$$

which is the desired result.

Consider $\exp(n^{1/(2\alpha+1)}) \geq p$ case first. Without loss of generality, we assume $k = \min(n^{1/(2\alpha+1)}, p)$ is an even number, and define a class of precision matrices

$$\begin{aligned}
\mathcal{U}_{11} &= \Big\{\Omega(\theta) \in \mathbb{R}^{p\times p} : \Omega(\theta) = (I_p - A(\theta))^T(I_p - A(\theta)), \\
&\qquad A(\theta) = -\tau a \sum_{m=1}^{k/2} \theta_m B(m,k), \theta = (\theta_m, 1 \leq m \leq k/2) \in \{0,1\}^{k/2}\Big\}
\end{aligned}$$

where $B(m,k) = (b_{ij} = I(i = m+1, \ldots, k \text{ and } j = m), 1 \leq i,j \leq p)$ is a $p \times p$ matrix and $a = (nk)^{-1/2}$. If we choose sufficiently small constant

$\tau > 0$, it is easy to check that for any $\Omega(\theta) \in \mathcal{U}_{11}$, $\epsilon_0 \le \lambda_{\min}(\Omega(\theta)) \le \lambda_{\max}(\Omega(\theta)) \le \epsilon_0^{-1}$ and $\|A(\theta) - B_{k_1}(A(\theta))\|_\infty \le C k_1^{-\alpha}$ for any $k_1 > 0$, so that $\mathcal{U}_{11} \subset \mathcal{U}(\epsilon_0, \gamma)$ for all sufficiently large n.

We use the Assouad's lemma (Assouad, 1983)

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega(\theta)\in\mathcal{U}_{11}} 2\mathbb{E}_\theta \|\widehat{\Omega}_n - \Omega(\theta)\| \;\ge\; \min_{H(\theta,\theta')\ge 1} \frac{\|\Omega(\theta) - \Omega(\theta')\|}{H(\theta,\theta')} \frac{k/2}{2} \min_{H(\theta,\theta')=1} \|\mathbb{P}_\theta \wedge \mathbb{P}_{\theta'}\|$$

where $H(\theta,\theta') = \sum_{m=1}^{k/2} |\theta_m - \theta'_m|$, $\|\mathbb{P}_\theta \wedge \mathbb{P}_{\theta'}\| = \int p_\theta \wedge p_{\theta'} d\mu$, and $\mathbb{P}_\theta$ and p_θ are the joint distribution function and density function, with respect to a dominating measure μ, of observation $X_1, \ldots, X_n \overset{iid}{\sim} N_p(0, \Omega(\theta)^{-1})$, respectively. If we show that

$$\min_{H(\theta,\theta')\ge 1} \frac{\|\Omega(\theta) - \Omega(\theta')\|}{H(\theta,\theta')} \;\gtrsim\; a \tag{S2.14}$$

and

$$\min_{H(\theta,\theta')=1} \|\mathbb{P}_\theta \wedge \mathbb{P}_{\theta'}\| \;\ge\; c \tag{S2.15}$$

for some constant $c > 0$, it will complete the proof. To show (S2.14), define a p-dimensional vector $v = (I(k/2 \le i \le k), 1 \le i \le p)$. By the construction of $\Omega(\theta)$ and v, one can check that

$$\big((\Omega(\theta) - \Omega(\theta'))v\big)_l = \begin{cases} (\tau a)^2 \frac{k}{2}(\theta_1\theta_{k/2} - \theta'_1\theta'_{k/2}) + \tau a(\frac{k}{2}+1)(\theta_1 - \theta'_1) & \text{if } 1 \le l \le \frac{k}{2} - 1 \\ (\tau a)^2 \frac{k}{2}(\theta_{k/2} - \theta'_{k/2}) + \tau a \frac{k}{2}(\theta_{k/2} - \theta'_{k/2}) & \text{if } l = \frac{k}{2} \\ \tau a\big(\theta_{k+1-l} - \theta'_{k+1-l}\big) & \text{if } \frac{k}{2} + 1 \le l \le k \\ 0 & \text{if } l \ge k+1. \end{cases}$$

Then, we have $\|(\Omega(\theta)-\Omega(\theta'))v\|_2^2 \geq (\tau a)^2 (k/2)^2 H(\theta,\theta')$ and

$$
\begin{aligned}
\|\Omega(\theta)-\Omega(\theta')\| &\geq \frac{\|(\Omega(\theta)-\Omega(\theta'))v\|_2}{\|v\|_2} \\
&= \frac{\|(\Omega(\theta)-\Omega(\theta'))v\|_2}{\sqrt{k/2}} \\
&\geq \left(\frac{(k/2\ \times \tau a)^2 H(\theta,\theta')}{k/2}\right)^{1/2} \\
&= \left(\frac{k/2}{H(\theta,\theta')}\right)^{1/2} \tau a H(\theta,\theta') \\
&\geq \tau a H(\theta,\theta').
\end{aligned}
$$

Thus, we have shown the first part.

To show (S2.15), note that

$$
\begin{aligned}
\|\mathbb{P}_\theta \wedge \mathbb{P}_{\theta'}\| &= \int_{p_\theta > p_{\theta'}} p_{\theta'} d\mu + \int_{p_\theta \leq p_{\theta'}} p_\theta d\mu \\
&= \Big(\frac{1}{2} - \frac{1}{2}\int_{p_\theta \leq p_{\theta'}} p_{\theta'} d\mu + \frac{1}{2}\int_{p_\theta > p_{\theta'}} p_{\theta'} d\mu\Big) + \Big(\frac{1}{2} - \frac{1}{2}\int_{p_\theta > p_{\theta'}} p_\theta d\mu + \frac{1}{2}\int_{p_\theta \leq p_{\theta'}} p_\theta d\mu\Big) \\
&= 1 - \frac{1}{2}\int_{p_\theta > p_{\theta'}} (p_\theta - p_{\theta'}) d\mu - \frac{1}{2}\int_{p_\theta \leq p_{\theta'}} (p_{\theta'} - p_\theta) d\mu \\
&= 1 - \frac{1}{2}\int |p_\theta - p_{\theta'}| d\mu.
\end{aligned}
$$

Let $\|\mathbb{P}_\theta - \mathbb{P}_{\theta'}\|_1 = \int |p_\theta - p_{\theta'}| d\mu$. Thus, it suffices to show that $\|\mathbb{P}_\theta - \mathbb{P}_{\theta'}\|_1^2 \leq$

1/2 when $H(\theta, \theta') = 1$. Also note that

$$
\begin{aligned}
\|\mathbb{P}_\theta - \mathbb{P}_{\theta'}\|_1 &\leq 2K(\mathbb{P}_{\theta'} \mid \mathbb{P}_\theta) \\
&= n\left[tr(\Omega(\theta')^{-1}\Omega(\theta)) - \log\det(\Omega(\theta')^{-1}\Omega(\theta)) - p\right] \\
&= n\left[tr(\Omega(\theta')^{-1}D_1) - \log\det(\Omega(\theta')^{-1}D_1 + I_p)\right] \\
&= n\left[tr(\Omega(\theta')^{-1/2}D_1\Omega(\theta')^{-1/2}) - \log\det(\Omega(\theta')^{-1/2}D_1\Omega(\theta')^{-1/2} + I_p)\right]
\end{aligned}
$$

where $K(\mathbb{P}_{\theta'} \mid \mathbb{P}_\theta) = \int \log(\frac{p_{\theta'}}{p_\theta})p_{\theta'}d\mu$ is the Kullback-Leibler divergence and $D_1 = \Omega(\theta) - \Omega(\theta')$. Let $\Omega(\theta')^{-1} = UVU^T$ be the diagonalization of $\Omega(\theta')^{-1}$. U is a orthogonal matrix whose columns are the eigenvectors of $\Omega(\theta')^{-1}$, and V is a diagonal matrix whose ith diagonal element is the eigenvalue of $\Omega(\theta')^{-1}$ corresponding to the ith column of U. It is easy to check that

$$
\begin{aligned}
\|\Omega(\theta')^{-1/2}D_1\Omega(\theta')^{-1/2}\|_F^2 &= \|UV^{1/2}U^TD_1UV^{1/2}U^T\|_F^2 \\
&= \|V^{1/2}U^TD_1UV^{1/2}\|_F^2 \\
&\leq \|V\|^2\|U^TD_1U\|_F^2 \\
&= \|\Omega(\theta')^{-1}\|^2\|D_1\|_F^2 \\
&\leq Ck(\tau a)^2
\end{aligned}
$$

for some constant $C > 0$ because $\|\Omega(\theta')^{-1}\| \le \epsilon_0^{-1}$ and

$$\big(\Omega(\theta)\big)_{(i,j)} = \begin{cases} 1 + (\tau a)^2\theta_i(k-i) & \text{if } 1 \le i = j \le \frac{k}{2} \\ (\tau a)\theta_i + (\tau a)^2\theta_i\theta_j(k-j) & \text{if } 1 \le i \ne j \le \frac{k}{2} \\ \tau a\theta_i & \text{if } 1 \le i \le \frac{k}{2}, \frac{k}{2}+1 \le j \le k \\ \tau a\theta_j & \text{if } \frac{k}{2}+1 \le i \le k, 1 \le j \le \frac{k}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Also note that, if $\lambda_1(\theta,\theta') \le \cdots \le \lambda_p(\theta,\theta')$ are the eigenvalues of $\Omega(\theta')^{-1/2}D_1\Omega(\theta')^{-1/2}$, we have $\sum_{j=1}^p \lambda_j(\theta,\theta')^2 = \|\Omega(\theta')^{-1/2}D_1\Omega(\theta')^{-1/2}\|_F^2 \le Ck(\tau a)^2 = C\tau^2/n$, which implies $|\lambda_j(\theta,\theta')| \le \sqrt{C}\tau/\sqrt{n}$ for all $1 \le j \le p$. Thus, $\Omega(\theta')^{-1/2}D_1\Omega(\theta')^{-1/2} + I_p$ is a positive definite matrix for all large n. Since $\|\Omega(\theta')^{-1/2}D_1\Omega(\theta')^{-1/2}\|_F^2$ is small, by Lemma C.2 in Lee and Lee (2018),

$$\|\mathbb{P}_\theta - \mathbb{P}_{\theta'}\|_1 \;\; \le \;\; nR_n$$

where $R_n \le C\|\Omega(\theta')^{-1/2}D_1\Omega(\theta')^{-1/2}\|_F^2$ for some constant $C > 0$. Thus, we have $\|\mathbb{P}_\theta - \mathbb{P}_{\theta'}\|_1 \le 1/2$ for some small $\tau > 0$ because $nka^2 = 1$.

Now consider $\exp(n^{1/(2\alpha+1)}) \le p$ case. To show (S2.13), define a class of diagonal precision matrices

$$\mathcal{U}_{12} \;\; = \;\; \left\{ \Omega_m \in \mathbb{R}^{p\times p} : \Omega_m = I_p + \tau\left(\frac{\log p}{n}\right)^{1/2}\Big(I(i=j=m)\Big),\ 0 \le m \le p \right\}$$

for some small $\tau > 0$. Since $p \le \exp(cn)$ for some constant $c > 0$, $\mathcal{U}_{12} \subset \mathcal{U}(\epsilon_0,\gamma)$ holds trivially. Let $r_{\min} = \inf_{1\le m\le p}\|\Omega_0 - \Omega_m\|$. We use the Le

Cam's lemma (LeCam, 1973)

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_m \in \mathcal{U}_{12}} \mathbb{E}_m \|\widehat{\Omega}_n - \Omega_m\| \geq \frac{1}{2} r_{\min} \|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\|$$

where $\bar{\mathbb{P}} = p^{-1} \sum_{m=1}^{p} \mathbb{P}_m$ and $\mathbb{P}_m$ is the distribution function of $N_p(0, \Omega_m^{-1})$ with observation $\mathbf{X}_n$. Note that $r_{\min} = \tau(\log p/n)^{1/2}$. We only need to show that $\|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\| \geq c$ for some constant $c > 0$. By the same argument with Cai et al. (2010) (page 2129), it suffices to show that

$$\int \frac{(p^{-1} \sum_{m=1}^{p} f_m)^2}{f_0} d\mu - 1 \longrightarrow 0, \tag{S2.16}$$

as $n \to \infty$ where f_m is the density function of $\mathbb{P}_m$ with respect to a σ-finite measure μ. Note that

$$\int \frac{(p^{-1} \sum_{m=1}^{p} f_m)^2}{f_0} d\mu - 1 = \frac{1}{p^2} \sum_{m=1}^{p} \int \frac{f_m^2}{f_0} d\mu + \frac{1}{p^2} \sum_{m \neq j} \int \frac{f_m f_j}{f_0} d\mu - 1$$

and $\int f_m f_j / f_0 d\mu = 1$ for any $m \neq j$. Also note that

$$\begin{aligned} \int \frac{f_m^2}{f_0} d\mu &= (1+b)^{n/2} \left(1 - \frac{b}{1+2b}\right)^{n/2} \\ &\leq e^{nb^2/(1+2b)} \\ &\leq e^{nb^2} = e^{\tau^2 \log p} \end{aligned}$$

where $b = \tau(\log p/n)^{1/2}$. Thus, (S2.16) holds for some small $\tau > 0$. It completes the proof for the case of polynomially decreasing $\gamma(k)$.

For the case of exponentially decreasing $\gamma(k) = Ce^{-\beta k}$, consider $k =$

$\min(\log n, p)$ for $\mathcal{U}_{11}$ instead of $k = \min(n^{1/(2\alpha+1)}, p)$. Then, similar arguments for the lower bounds of $\mathcal{U}_{11}$ and $\mathcal{U}_{12}$ give the desired result.

For the exact banding $\gamma(k)$, consider $\mathcal{U}_{11}$ with $k = k_0$ and $a = (\log p/n)^{1/2}$, then it completes the proof. $\square$

Proof of Theorem 3. We follow closely the line of a proof in Cai and Zhou (2012). Consider the polynomially decreasing case, $\gamma(k) = Ck^{-\alpha}$, first. Two parameter classes are considered depending on the relation between p and n. For $\exp(n^{1/(2\alpha+2)}) \geq p$ case, we show that

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n} \in \mathcal{G}_{11}} \mathbb{E}_{0n} \|\widehat{\Omega}_n - \Omega_{0,n}\|_\infty \gtrsim \min\left(n^{-\alpha/(2\alpha+2)}, \frac{p}{n^{1/2}}\right), \quad \text{(S2.17)}$$

and for $\exp(n^{1/(2\alpha+2)}) \leq p$ case, we show that

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_{0,n} \in \mathcal{G}_{12}} \mathbb{E}_{0n} \|\widehat{\Omega}_n - \Omega_{0,n}\|_\infty \gtrsim \left(\frac{\log p}{n}\right)^{\alpha/(2\alpha+1)} \quad \text{(S2.18)}$$

for some $\mathcal{G}_{11} \cup \mathcal{G}_{12} \subset \mathcal{U}(\epsilon_0, \gamma)$.

Consider $\exp(n^{1/(2\alpha+2)}) \geq p$ case first. Define a class of precision matrices

$$\begin{aligned} \mathcal{G}_{11} &= \Big\{ \Omega(\theta) \in \mathbb{R}^{p \times p} : \Omega(\theta) = (I_p - A(\theta))^T (I_p - A(\theta)), \\ &\qquad A(\theta) = -\tau a \sum_{s=2}^{k} \theta_{s-1} G_s, \theta = (\theta_s) \in \{0,1\}^{k-1} \Big\} \end{aligned}$$

where $G_s = (I(i = s, j = 1))$ is a $p \times p$ matrix and $a = n^{-1/2}$ and $k = \min(n^{1/(2\alpha+2)}, p)$. It is easy to show that $\mathcal{G}_{11} \subset \mathcal{U}(\epsilon_0, \gamma)$ for some small constant $\tau > 0$ and all sufficiently large n.

We use the Assouad's lemma,

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega(\theta)\in\mathcal{G}_{11}} 2\mathbb{E}_\theta \|\widehat{\Omega}_n - \Omega(\theta)\|_\infty \;\geq\; \min_{H(\theta,\theta')\geq 1} \frac{\|\Omega(\theta)-\Omega(\theta')\|_\infty}{H(\theta,\theta')} \frac{k-1}{2} \min_{H(\theta,\theta')=1} \|\mathbb{P}_\theta \wedge \mathbb{P}_{\theta'}\|.$$

It is easy to see that

$$\min_{H(\theta,\theta')\geq 1} \frac{\|\Omega(\theta)-\Omega(\theta')\|_\infty}{H(\theta,\theta')} \;\geq\; \tau a.$$

To show $\min_{H(\theta,\theta')=1} \|\mathbb{P}_\theta \wedge \mathbb{P}_{\theta'}\| \geq c$ for some $c > 0$, it suffices to prove that $\|\mathbb{P}_\theta - \mathbb{P}_{\theta'}\|_1 \leq 1$. Note that

$$\begin{aligned} \|\mathbb{P}_\theta - \mathbb{P}_{\theta'}\|_1^2 &\leq 2K(\mathbb{P}_{\theta'} \mid \mathbb{P}_\theta) \\ &\leq Cn\|\Omega(\theta')^{-1/2} D_1 \Omega(\theta')^{-1/2}\|_F^2 \end{aligned}$$

for some constant $C > 0$ where $D_1 = \Omega(\theta) - \Omega(\theta')$. By the same argument used in the proof of Theorem 1, one can show that $\|\mathbb{P}_\theta - \mathbb{P}_{\theta'}\|_1^2 \leq C'n(\tau a)^2$ for some constant $C' > 0$, and it is smaller than 1 for some small constant $\tau > 0$. Thus, we have proved the (S2.17) part.

Now consider $\exp(n^{1/(2\alpha+2)}) \leq p$ case. To show (S2.18) part, define a class of precision matrices

$$\mathcal{G}_{12} = \left\{ \Omega_m \in \mathbb{R}^{p\times p} : \Omega_m = (I_p - A_m)^T (I_p - A_m),\ A_m = -\tau B_m \left(\frac{\log p}{nk}\right)^{1/2},\ 1 \leq m \leq m_* \right\}$$

where $B_m = (I(m+1 \leq i \leq m+k-1,\ j = m))$ is a $p\times p$ matrix, $m_* = p/k-1$ and $k = (n/\log p)^{1/(2\alpha+1)}$. Without loss of generality, we assume that p can be divided by k. By the definition of $\mathcal{G}_{12}$, tedious calculations yield that $\mathcal{G}_{12} \subset \mathcal{U}(\epsilon_0, \gamma)$.

Let $\Omega_0 = I_p$ and $\mathbb{P}_m$ be the distribution function of $N(0, \Omega_m^{-1})$ with observation $\mathbf{X}_n$. It is easy to check that for any $0 \le m \neq m' \le m_*$,

$$\|\Omega_m - \Omega_{m'}\|_\infty \;\; \ge \;\; \tau \left(\frac{k \log p}{n} \right)^{1/2} \;\; = \;\; \tau \left(\frac{\log p}{n} \right)^{\alpha/(2\alpha+1)}$$

by the definition of $\mathcal{G}_{12}$ and k. Since $k^2 \le p$, for any $1 \le m \le m_*$,

$$\begin{aligned} K(\mathbb{P}_m \mid \mathbb{P}_0) \;\; &\le \;\; Cn \|\Omega_{m'}^{-1/2} D_1 \Omega_{m'}^{-1/2}\|_F^2 \\ &\le \;\; C' \tau^2 \log p \\ &\le \;\; c \log m_* \end{aligned}$$

for some constants $C, C' > 0, 0 < c < 1/8$ and small $\tau > 0$, which implies that for any $1 \le m \le m_*$,

$$\frac{1}{m_*} \sum_{m=1}^{m_*} K(\mathbb{P}_m \mid \mathbb{P}_0) \;\; \le \;\; c \log m_*$$

for some $0 < c < 1/8$, so we can use Fano's lemma,

$$\inf_{\widehat{\Omega}_n} \sup_{\Omega_m \in \mathcal{G}_{12}} \mathbb{E}_m \|\widehat{\Omega}_n - \Omega_m\|_\infty \;\; \ge \;\; \min_{0 \le m \neq m' \le m_*} \frac{\|\Omega_m - \Omega_{m'}\|_\infty}{4} \frac{{m_*}^{1/2}}{1 + {m_*}^{1/2}} \left(1 - 2c - \left(\frac{2c}{\log m_*} \right)^{1/2} \right).$$

It completes the proof. For more details about Fano's lemma, see Tsybakov (2008).

For the case of exponentially decreasing $\gamma(k) = Ce^{-\beta k}$, consider $k = \min([\log n \, \log p]^{1/2}, p)$ for $\mathcal{G}_{11}$ instead of $k = \min(n^{1/(2\alpha+2)}, p)$. Then, similar arguments for the lower bound of $\mathcal{G}_{11}$ give the desired result.

For the exact banding $\gamma(k)$, consider $\mathcal{G}_{11}$ with $k = k_0$ and $a = (\log p/n)^{1/2}$, then it completes the proof. □

S2.3 Proof of the P-loss Convergence Rates: Theorem 2 and Theorem 4

Lemma 1–5 are used to prove the main theorems.

Lemma 1. *Let $X_1, \ldots, X_n \overset{iid}{\sim} N_p(0, \Omega_{0,n}^{-1})$ with $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$ defined at (2.8),*

$$
\begin{aligned}
N_{1n} &= \left\{ \mathbf{X}_n : \max_j \left\| \widehat{\mathrm{Var}}(X_{1,(j-k):j}) \right\| \le C_1 \right\}, \\
N_{2n} &= \left\{ \mathbf{X}_n : \max_j \left\| \widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j}) \right\| \le C_2 \right\}, \\
N_{3n} &= \left\{ \mathbf{X}_n : \max_j \left\| \widehat{\mathrm{Var}}(X_{1,(j-k):j}) - \mathrm{Var}(X_{1,(j-k):j}) \right\| \le \Big(C_3(k + \log(n \vee p))/n \Big)^{1/2} \right\}, \\
N_{4n} &= \left\{ \mathbf{X}_n : \max_j \left\| \widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j}) - \mathrm{Var}^{-1}(X_{1,(j-k):j}) \right\| \le \Big(C_4(k + \log(n \vee p))/n \Big)^{1/2} \right\},
\end{aligned}
$$

where $C_1 = \epsilon_0^{-1}(2 + \big((k+1)/n\big)^{1/2})^2$, $C_2 = 4\epsilon_0^{-1}(1 - \big((k+1)/n\big)^{1/2})^{-2}$, $C_4 = C_3 C_2^2 \epsilon_0^{-2}$ and $N_n = \bigcap_{j=1}^4 N_{jn}$. If $k + \log p = o(n)$ and $1 \le k \le p-1$, then for any large constant C_3, there exists a positive constant C_5 such that

$$
\mathbb{P}_{0n}\big(\mathbf{X}_n \in N_n^c\big) \le 6pe^{-n(1-((k+1)/n)^{1/2})^2/8} + 4 \times 5^k e^{-C_3 C_5 \epsilon_0^2 (\log(n\vee p)+k)},
$$

for all sufficiently large n. Here, C_5 does not depend on C_3.

Proof of Lemma 1. We will show that for any large constant C_3,

$$\mathbb{P}_{0n}\big(\mathbf{X}_n \in N_{1n}^c\big) \;\le\; 2pe^{-n/2}, \tag{S2.19}$$

$$\mathbb{P}_{0n}\big(\mathbf{X}_n \in N_{2n}^c\big) \;\le\; 2pe^{-n(1-((k+1)/n)^{1/2})^2/8}, \tag{S2.20}$$

$$\mathbb{P}_{0n}\big(\mathbf{X}_n \in N_{3n}^c\big) \;\le\; 2\times 5^k e^{-C_3C_5\epsilon_0^2(k+\log(n\vee p))}, \tag{S2.21}$$

$$\mathbb{P}_{0n}\big(\mathbf{X}_n \in N_{4n}^c\big) \;\le\; 2\times 5^k e^{-C_3C_5\epsilon_0^2(k+\log(n\vee p))} + 2pe^{-n(1-\sqrt{(k+1)/n})^2/8}, \tag{S2.22}$$

for some positive constants C_4 and C_5. The inequalities (S2.19) and (S2.20) follow from Lemma B.7 in Lee and Lee (2018). Note that for any large constant $C_3 > 0$,

$$\mathbb{P}_{0n}\big(\mathbf{X}_n \in N_{3n}^c\big) \;\le\; p\,5^{k+1}\Big(e^{-C_3C_6\epsilon_0^2(k+\log(n\vee p))} + e^{-C_3^{1/2}C_7\epsilon_0\{n(k+\log(n\vee p))\}^{1/2}}\Big) \tag{S2.23}$$

for all sufficiently large n and some positive constants C_6 and C_7 by Lemma B.6 in Lee and Lee (2018). If we take $C_5 = C_6/2$, the right hand side (RHS) of (S2.23) is bounded by $2\times 5^k \exp\{-C_3C_5\epsilon_0^2(k+\log(n\vee p))\}$ for any constant $C_3 > 0$ and all sufficiently large n because $k + \log(n \vee p) = o(n)$. Similarly,

$$\begin{aligned}
&\mathbb{P}_{0n}\big(\mathbf{X}_n \in N_{4n}^c\big)\\
\le\;& \mathbb{P}_{0n}\big(\mathbf{X}_n \in N_{4n}^c \cap N_{2n}\big) + \mathbb{P}_{0n}\big(\mathbf{X}_n \in N_{2n}^c\big)\\
\le\;& \mathbb{P}_{0n}\left(\max_j \big\|\widehat{\mathrm{Var}}(X_{1,(j-k):j}) - \mathrm{Var}(X_{1,(j-k):j})\big\| \ge C_2^{-1}\epsilon_0\Big(C_4\frac{k+\log(n\vee p)}{n}\Big)^{1/2}\right)\\
&+ 2pe^{-n(1-((k+1)/n)^{1/2})^2/8}\\
\le\;& 2\times 5^k e^{-C_3C_5\epsilon_0^2(k+\log(n\vee p))} + 2pe^{-n(1-((k+1)/n)^{1/2})^2/8}
\end{aligned}$$

for $C_4 = C_3C_2^2\epsilon_0^{-2}$ and all sufficiently large n. Since the inequalities (S2.21) and (S2.22) also hold, this completes the proof. □

Lemma 2. *Consider model* $X_1, \ldots, X_n \overset{iid}{\sim} N_p(0, \Omega_{0,n}^{-1})$ *with* $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$ *defined at (2.8) and* $\sum_{m=1}^{\infty} \gamma(m) < \infty$. *Denote* $\widehat{\Omega}_{nk} = (I_p - \widehat{A}_{nk})^T \widehat{D}_{nk}^{-1}(I_p - \widehat{A}_{nk})$, $\widehat{D}_{nk} = diag(\widehat{d}_{jk})$ *and* $\widehat{A}_{nk} = (\widehat{a}_{jl}^{(k)})$ *for* $1 \le k \le p-1$, *where* $(\widehat{a}_{j,j-k}^{(k)}, \ldots, \widehat{a}_{j,j-1}^{(k)})^T = \widehat{a}_j^{(k)}$, *and* $\widehat{a}_{jl}^{(k)} = 0$ *if* $1 \le j \le l \le p$ *or* $|j-l| > k$. $\widehat{a}_j^{(k)}$ *and* $\widehat{d}_{jk}$ *are defined at (2.6). If* $k^{3/2}(k + \log(n \vee p)) = O(n)$, *then*

$$\mathbb{E}_{0n}\left[\|\widehat{\Omega}_{nk} - \Omega_{0,n}\| I(\mathbf{X}_n \in N_n)\right] \;\lesssim\; k^{3/4}\left[\left(\frac{k + \log(n \vee p)}{n}\right)^{1/2} + \gamma(k)\right],$$

and if $k(k + \log(n \vee p)) = O(n)$, *then*

$$\mathbb{E}_{0n}\left[\|\widehat{\Omega}_{nk} - \Omega_{0,n}\|_\infty I(\mathbf{X}_n \in N_n)\right] \;\lesssim\; k\left[\left(\frac{k + \log(n \vee p)}{n}\right)^{1/2} + \gamma(k)\right],$$

where the set N_n *is defined at Lemma 1.*

Proof of Lemma 2. Let

$$A_{0,nk} = (a_{0,jl}^{(k)}) \quad \text{and} \quad D_{0,nk} = diag(d_{0,jk}), \tag{S2.24}$$

where $(a_{0,j,j-k}^{(k)}, \ldots, a_{0,j,j-1}^{(k)})^T = a_{0,j}^{(k)}$, and $a_{0,jl}^{(k)} = 0$ if $1 \le j \le l \le p$ or

$|j-l|>k$. Define $\Omega_{0,nk}=(I_p-A_{0,nk})^T D_{0,nk}^{-1}(I_p-A_{0,nk})$. Note that

$$\begin{aligned}
&\mathbb{E}_{0n}\Big[\|\widehat{\Omega}_{nk}-\Omega_{0,n}\| I(\mathbf{X}_n\in N_n)\Big]\\
&\le \mathbb{E}_{0n}\Big[\|\widehat{\Omega}_{nk}-\Omega_{0,nk}\| I(\mathbf{X}_n\in N_n)\Big]+\|\Omega_{0,nk}-\Omega_{0,n}\|\\
&\le \mathbb{E}_{0n}\Big[\|\widehat{A}_{nk}^T-A_{0,nk}^T\|\,\|D_{0,nk}^{-1}\|\,\|I_p-A_{0,nk}\| I(\mathbf{X}_n\in N_n)\Big]\\
&+\mathbb{E}_{0n}\Big[\|\widehat{D}_{nk}^{-1}-D_{0,nk}^{-1}\|\,\|I_p-A_{0,nk}^T\|\,\|I_p-A_{0,nk}\| I(\mathbf{X}_n\in N_n)\Big]\\
&+\mathbb{E}_{0n}\Big[\|\widehat{A}_{nk}-A_{0,nk}\|\,\|D_{0,nk}^{-1}\|\,\|I_p-A_{0,nk}^T\| I(\mathbf{X}_n\in N_n)\Big]\\
&+\mathbb{E}_{0n}\Big[\|I_p-A_{0,nk}^T\|\,\|\widehat{D}_{nk}^{-1}-D_{0,nk}^{-1}\|\,\|\widehat{A}_{nk}-A_{0,nk}\| I(\mathbf{X}_n\in N_n)\Big]\\
&+\mathbb{E}_{0n}\Big[\|D_{0,nk}^{-1}\|\,\|\widehat{A}_{nk}^T-A_{0,nk}^T\|\,\|\widehat{A}_{nk}-A_{0,nk}\| I(\mathbf{X}_n\in N_n)\Big]\\
&+\mathbb{E}_{0n}\Big[\|I_p-A_{0,nk}\|\,\|\widehat{A}_{nk}^T-A_{0,nk}^T\|\,\|\widehat{D}_{nk}^{-1}-D_{0,nk}^{-1}\| I(\mathbf{X}_n\in N_n)\Big]\\
&+\mathbb{E}_{0n}\Big[\|\widehat{A}_{nk}^T-A_{0,nk}^T\|\,\|\widehat{D}_{nk}^{-1}-D_{0,nk}^{-1}\|\,\|\widehat{A}_{nk}-A_{0,nk}\| I(\mathbf{X}_n\in N_n)\Big]\\
&+\|\Omega_{0,nk}-\Omega_{0,n}\|
\end{aligned}\tag{S2.25}$$

by the triangle inequality (See page 223 of Bickel and Levina (2008)). Also note that

$$\begin{aligned}
\|I_p-A_{0,nk}\|_\infty &\le 1+\|A_{0,nk}-A_{0,n}\|_\infty+\|A_{0,n}\|_\infty\\
&\le 1+C(k^{1/2}\gamma(k)+1),\\
\|I_p-A_{0,nk}\|_1 &\le 1+\|A_{0,nk}-A_{0,n}\|_\infty+\|A_{0,n}\|_1\\
&\le 1+Ck\gamma(k)+\sum_{m=1}^{\infty}\gamma(m),
\end{aligned}$$

for some constant $C>0$ by Lemma 10, and $\|D_{0,nk}^{-1}\|\le \max_j \|\mathrm{Var}^{-1}(X_{1,(j-k):j})\|\le$

ϵ_0^{-1} using the similar argument to (S3.42). If we show that, on $(\mathbf{X}_n \in N_n)$,

$$\|\widehat{A}_{nk} - A_{0,nk}\|_\infty \lesssim k^{1/2}\left(\frac{k+\log(n\vee p)}{n}\right)^{1/2}, \tag{S2.26}$$

$$\|\widehat{A}_{nk} - A_{0,nk}\|_1 \lesssim k\left(\frac{k+\log(n\vee p)}{n}\right)^{1/2}, \tag{S2.27}$$

$$\|\widehat{D}_{nk}^{-1} - D_{0,nk}^{-1}\|_\infty \lesssim \left(\frac{k+\log(n\vee p)}{n}\right)^{1/2}, \tag{S2.28}$$

$\|\Omega_{0,nk} - \Omega_{0,n}\| \lesssim k^{3/4}\gamma(k)$ and $\|\Omega_{0,nk} - \Omega_{0,n}\|_\infty \lesssim k\gamma(k)$, the proof is completed by (S2.25).

To show (S2.26), note that

$$\begin{aligned}
&\|\widehat{A}_{nk} - A_{0,nk}\|_\infty \\
=\ & \max_j \|\widehat{a}_j^{(k)} - a_{0,j}^{(k)}\|_1 \\
\le\ & k^{1/2}\max_j \|\widehat{a}_j^{(k)} - a_{0,j}^{(k)}\|_2 \\
=\ & k^{1/2}\max_j \Big\|\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):(j-1)})\widehat{\mathrm{Cov}}(X_{1,(j-k):(j-1)}, X_{1j}) \\
& -\mathrm{Var}^{-1}(X_{1,(j-k):(j-1)})\mathrm{Cov}(X_{1,(j-k):(j-1)}, X_{1j})\Big\|_2 \\
\le\ & k^{1/2}\Big\{\max_j \Big\|\mathrm{Var}^{-1}(X_{1,(j-k):(j-1)})\left(\widehat{\mathrm{Cov}}(X_{1,(j-k):(j-1)}, X_{1j}) - \mathrm{Cov}(X_{1,(j-k):(j-1)}, X_{1j})\right)\Big\|_2 \\
& +\max_j \Big\|(\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):(j-1)}) - \mathrm{Var}^{-1}(X_{1,(j-k):(j-1)}))\widehat{\mathrm{Cov}}(X_{1,(j-k):(j-1)}, X_{1j})\Big\|_2\Big\}.
\end{aligned}$$

The first part of the last line can be bounded above by

$$\begin{aligned}
& k^{1/2} \max_j \Big\| \mathrm{Var}^{-1}(X_{1,(j-k):(j-1)}) \, (\widehat{\mathrm{Cov}}(X_{1,(j-k):(j-1)}, X_{1j}) - \mathrm{Cov}(X_{1,(j-k):(j-1)}, X_{1j})) \Big\|_2 \\
\le\ & k^{1/2} \max_j \big\| \mathrm{Var}^{-1}(X_{1,(j-k):(j-1)}) \big\| \, \Big\| \widehat{\mathrm{Cov}}(X_{1,(j-k):(j-1)}, X_{1j}) - \mathrm{Cov}(X_{1,(j-k):(j-1)}, X_{1j}) \Big\|_2 \\
\le\ & k^{1/2} \max_j \big\| \mathrm{Var}^{-1}(X_{1,(j-k):(j-1)}) \big\| \, \Big\| \widehat{\mathrm{Var}}(X_{1,(j-k):j}) - \mathrm{Var}(X_{1,(j-k):j}) \Big\| \\
\lesssim\ & k^{1/2} \left(\frac{k + \log(n \vee p)}{n} \right)^{1/2} \qquad \text{on } (\mathbf{X}_n \in N_n).
\end{aligned}$$

The first inequality holds by the definition of the spectral norm, and the second inequality holds because the spectral norm of a matrix is larger than a ℓ_2 norm of any columns. The second part can be bounded similarly

$$\begin{aligned}
& k^{1/2} \max_j \Big\| (\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):(j-1)}) - \mathrm{Var}^{-1}(X_{1,(j-k):(j-1)})) \widehat{\mathrm{Cov}}(X_{1,(j-k):(j-1)}, X_{1j}) \Big\|_2 \\
\le\ & k^{1/2} \max_j \Big\| \widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):(j-1)}) - \mathrm{Var}^{-1}(X_{1,(j-k):(j-1)}) \Big\| \, \Big\| \widehat{\mathrm{Var}}(X_{1,(j-k):j}) \Big\| \\
\lesssim\ & k^{1/2} \left(\frac{k + \log(n \vee p)}{n} \right)^{1/2} \qquad \text{on } (\mathbf{X}_n \in N_n).
\end{aligned}$$

By similar arguments, we can show that the inequality (S2.27) holds:

$$\begin{aligned}
\|\widehat{A}_{nk} - A_{0,nk}\|_1 &\le k \max_j \|\widehat{a}_j^{(k)} - a_{0,j}^{(k)}\|_{\max} \\
&\le k \max_j \|\widehat{a}_j^{(k)} - a_{0,j}^{(k)}\|_2 \\
&\lesssim k \left(\frac{k + \log(n \vee p)}{n} \right)^{1/2} \qquad \text{on } (\mathbf{X}_n \in N_n).
\end{aligned}$$

To show (S2.28), note that

$$\|\widehat{D}_{nk}^{-1} - D_{0,nk}^{-1}\|_\infty \le \|\widehat{D}_{nk}^{-1}\|_\infty \|D_{0,nk}^{-1}\|_\infty \|\widehat{D}_{nk} - D_{0,nk}\|_\infty$$

and $\|\widehat{D}_{nk}^{-1}\|_\infty \cdot \|D_{0,nk}^{-1}\|_\infty \le \max_j \|\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j})\| \cdot \epsilon_0^{-1} \le C_2\epsilon_0^{-1}$ on $(\mathbf{X}_n \in N_n)$ for $C_2 > 0$ used in set N_{2n}, by the similar argument to (S3.42). The rest part is easily bounded above as follows:

$$\begin{aligned}
\|\widehat{D}_{nk} - D_{0,nk}\|_\infty &= \max_j |\widehat{d}_{jk} - d_{0,jk}| \\
&\le \max_j \left|\widehat{\mathrm{Var}}(X_{1j}) - \mathrm{Var}(X_{1j})\right| \\
&+ \max_j \left|\widehat{\mathrm{Cov}}(X_{1j}, X_{1,(j-k):(j-1)})\, \widehat{a}_j^{(k)} - \mathrm{Cov}(X_{1j}, X_{1,(j-k):(j-1)})\, a_{0,j}^{(k)}\right| \\
&\le \max_j \left|\widehat{\mathrm{Var}}(X_{1j}) - \mathrm{Var}(X_{1j})\right| + \max_j \left|\widehat{\mathrm{Cov}}(X_{1j}, X_{1,(j-k):(j-1)})\, \left(\widehat{a}_j^{(k)} - a_{0,j}^{(k)}\right)\right| \\
&+ \max_j \left|\left(\widehat{\mathrm{Cov}}(X_{1j}, X_{1,(j-k):(j-1)}) - \mathrm{Cov}(X_{1j}, X_{1,(j-k):(j-1)})\right) a_{0,j}^{(k)}\right| \\
&\lesssim \left(\frac{k + \log(n \vee p)}{n}\right)^{1/2} \quad \text{on } (\mathbf{X}_n \in N_n).
\end{aligned}$$

Hence, by (S2.25), we have shown that

$$\mathbb{E}_{0n}\left[\|\widehat{\Omega}_{nk} - \Omega_{0,nk}\| I(\mathbf{X}_n \in N_n)\right] \lesssim k^{3/4}\left(\frac{k + \log(n \vee p)}{n}\right)^{1/2} + \|\Omega_{0,nk} - \Omega_{0,n}\|$$

when $k^{3/2}(k + \log(n \vee p)) = O(n)$, and

$$\mathbb{E}_{0n}\left[\|\widehat{\Omega}_{nk} - \Omega_{0,nk}\|_\infty I(\mathbf{X}_n \in N_n)\right] \lesssim k\left(\frac{k + \log(n \vee p)}{n}\right)^{1/2} + \|\Omega_{0,nk} - \Omega_{0,n}\|_\infty$$

when $k(k + \log(n \vee p)) = O(n)$. The conditions $k^{3/2}(k + \log(n \vee p)) = O(n)$ and $k(k + \log(n \vee p)) = O(n)$ are required due to the term

$$\mathbb{E}_{0n}\left[\|D_{0,nk}^{-1}\|\, \|\widehat{A}_{nk}^T - A_{0,nk}^T\|\, \|\widehat{A}_{nk} - A_{0,nk}\| I(\mathbf{X}_n \in N_n)\right]$$

in (S2.25).

If we show that $\|\Omega_{0,nk}-\Omega_{0,n}\| \lesssim k^{3/4}\gamma(k)$ and $\|\Omega_{0,nk}-\Omega_{0,n}\|_\infty \lesssim k\gamma(k)$, this completes the proof. By Lemma 10, we have $\|A_{0,nk}-A_{0,n}\|_\infty \lesssim k^{1/2}\gamma(k)$ and $\|A_{0,nk}-A_{0,n}\|_1 \lesssim k\gamma(k)$. Note that

$$\begin{aligned}
\|D_{0,nk}-D_{0,n}\|_\infty &= \max_j \left| a_{0,j}^{(k)T}\mathrm{Var}(X_{1,(j-k):(j-1)})a_{0,j}^{(k)} - a_{0,j}^T \mathrm{Var}(X_{1,1:(j-1)})a_{0,j} \right| \\
&= \max_j \left| ((0^T, a_{0,j}^{(k)T}) - a_{0,j}^T)\mathrm{Var}(X_{1,1:(j-1)}) \left(\begin{pmatrix} 0 \\ a_{0,j}^{(k)} \end{pmatrix} + a_{0,j} \right) \right| \\
&\le \|A_{0,nk}-A_{0,n}\|_\infty \max_j \left(\|a_{0,j}^{(k)}\|_2 + \|a_{0,j}\|_2 \right) \left\| \mathrm{Var}(X_{1,1:(j-1)}) \right\| \\
&\lesssim k^{1/2}\gamma(k).
\end{aligned}$$

Thus, it is easy to show that $\|\Omega_{0,nk}-\Omega_{0,n}\| \lesssim k^{3/4}\gamma(k)$ and $\|\Omega_{0,nk}-\Omega_{0,n}\|_\infty \lesssim k\gamma(k)$ by the triangle inequality in (S2.25). □

Lemma 3. *Consider model $X_1,\ldots,X_n \overset{iid}{\sim} N_p(0,\Omega_{0,n}^{-1})$ and the k-BC prior. Assume that $\Omega_{0,n} \in \mathcal{U}(\epsilon_0,\gamma)$ defined at (2.8) and $\sum_{m=1}^\infty \gamma(m) < \infty$. Let*

$$\begin{aligned}
\pi(d_j \mid \mathbf{X}_n) &= IG\left(d_j \mid \frac{n_j}{2}, \frac{n}{2}\widehat{d}_{jk}, d_j \le M\right), \\
\widetilde{\pi}(d_j \mid \mathbf{X}_n) &= IG\left(d_j \mid \frac{n_j}{2}, \frac{n}{2}\widehat{d}_{jk}\right),
\end{aligned}$$

for $j=1,\ldots,p$, where $\widehat{d}_{jk}$ defined at (2.6). If $M \ge 9\epsilon_0^{-1}$, $\nu_0 = o(n)$, $k+\log p = o(n)$ and $1 \le k \le p-1$, then on $(\mathbf{X}_n \in N_n)$,

$$\begin{aligned}
\pi(A_n, D_n \mid \mathbf{X}_n) &= \pi(d_1 \mid \mathbf{X}_n)\prod_{j=2}^p \pi(a_j \mid d_j, \mathbf{X}_n)\pi(d_j \mid \mathbf{X}_n) \\
&\lesssim \widetilde{\pi}(d_1 \mid \mathbf{X}_n)\prod_{j=2}^p \pi(a_j \mid d_j, \mathbf{X}_n)\widetilde{\pi}(d_j \mid \mathbf{X}_n)
\end{aligned} \tag{S2.29}$$

for all sufficiently large n, where the set N_n is defined at Lemma 1.

Proof of Lemma 3. We have

$$\pi(d_j \mid \mathbf{X}_n) = \frac{IG\left(d_j \mid n_j/2, n\widehat{d}_{jk}/2\right) I(d_j \le M)}{\int_0^M IG\left(d_j' \mid n_j/2, n\widehat{d}_{jk}/2\right) dd_j'}$$

for $j = 1, \ldots, p$. To show (S2.29), it suffices to prove, on $(\mathbf{X}_n \in N_n)$,

$$\left[\min_j \widetilde{\pi}(d_j \le M \mid \mathbf{X}_n)\right]^{-p} \le C$$

for some constant $C > 0$. Note that on $(\mathbf{X}_n \in N_n)$, $C_1^{-1} \le \widehat{d}_{jk}^{-1} \le C_2$ and

$$\begin{aligned}
\widetilde{\pi}(d_j \le M \mid \mathbf{X}_n) &= \widetilde{\pi}(M^{-1} \le d_j^{-1} \mid \mathbf{X}_n) \\
&= \widetilde{\pi}\left(M^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1} \le d_j^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1} \mid \mathbf{X}_n\right) \\
&= 1 - \widetilde{\pi}\left(d_j^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1} < M^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1} \mid \mathbf{X}_n\right).
\end{aligned}$$

By page 29 of Boucheron et al. (2013), if X is a sub-gamma random variable with variance factor ν and scale parameter c,

$$\max\left[P(X > (2\nu t)^{1/2} + ct), P(X < -(2\nu t)^{1/2} - ct)\right] \le e^{-t} \quad \text{(S2.30)}$$

for all $t > 0$. Since a centered $Gamma(a, b)$ random variable is a sub-gamma random variable with $\nu = a/b^2$ and $c = 1/b$, applying $t = nt'$ with $t' = (M - 2C_1)^2/(8M)^2 < 1$ to the inequality (S2.30),

$$\begin{aligned}
e^{-nt'} &\ge \widetilde{\pi}\left(d_j^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1} < -2\left(\frac{n_j}{n}\right)^{1/2}\widehat{d}_{jk}^{-1}(t')^{1/2} - 2\widehat{d}_{jk}^{-1}t' \mid \mathbf{X}_n\right) \\
&\ge \widetilde{\pi}\left(d_j^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1} < -4\widehat{d}_{jk}^{-1}(t')^{1/2} \mid \mathbf{X}_n\right) \\
&\ge \widetilde{\pi}\left(d_j^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1} < M^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1} \mid \mathbf{X}_n\right)
\end{aligned}$$

because $M \geq 9\epsilon_0^{-1} > 2C_1$ for all sufficiently large n and $\nu_0 = o(n)$. Thus, for some constant $C > 0$, on $(\mathbf{X}_n \in N_n)$,

$$\widetilde{\pi}(d_j \leq M \mid \mathbf{X}_n) \;\geq\; 1 - e^{-Cn}, \tag{S2.31}$$

and

$$\begin{aligned}\left[\min_j \widetilde{\pi}(d_j \leq M \mid \mathbf{X}_n)\right]^{-p} &\leq (1 - e^{-Cn})^{-p} \\ &= (1 - e^{-Cn})^{-e^{Cn} \times p/e^{Cn}} \\ &\leq (C')^{p/e^{Cn}} \longrightarrow 1\end{aligned}$$

as $n \to \infty$ for some constant $C' > 0$. □

Lemma 4. *Consider the model* $X_1, \ldots, X_n \overset{iid}{\sim} N_p(0, \Omega_{0,n}^{-1})$ *and the k-BC prior. Assume that* $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$ *defined at (2.8) and* $\sum_{m=1}^{\infty} \gamma(m) < \infty$. *If* $M \geq 9\epsilon_0^{-1}$, $\nu_0 = o(n)$, $k + \log p = o(n)$ *and* $1 \leq k \leq p-1$, *then*

$$\begin{aligned}\mathbb{E}^{\pi}\left(\|A_n - \widehat{A}_{nk}\|_\infty^2 \mid \mathbf{X}_n\right) &\leq Ck\left(\frac{k + \log p}{n}\right) \quad \text{on } (\mathbf{X}_n \in N_n), \\ \mathbb{E}^{\pi}\left(\|A_n - \widehat{A}_{nk}\|_1^2 \mid \mathbf{X}_n\right) &\leq Ck\left(\frac{k + \log p}{n}\right) \quad \text{on } (\mathbf{X}_n \in N_n),\end{aligned}$$

for some constant $C > 0$ *and all sufficiently large* n, *where* $\widehat{A}_{nk}$ *is defined at Lemma 2.*

Proof of Lemma 4. Let $\mathbb{E}^{\widetilde{\pi}}(\cdot \mid \mathbf{X}_n)$ denote the expectation with respect to $\widetilde{\pi}(d_1 \mid \mathbf{X}_n) \prod_{j=2}^p \pi(a_j \mid d_j, \mathbf{X}_n)\widetilde{\pi}(d_j \mid \mathbf{X}_n)$ in Lemma 3. Note that on $(\mathbf{X}_n \in$

N_n),

$$
\begin{aligned}
& \mathbb{E}^{\pi}\left(\|A_n - \widehat{A}_{nk}\|_\infty^2 \mid \mathbf{X}_n\right) \\
\leq\ & k\,\mathbb{E}^{\pi}\left(\max_j \|a_j - \widehat{a}_j^{(k)}\|_2^2 \mid \mathbf{X}_n\right) \\
\leq\ & k\,\mathbb{E}^{\pi}\left(\max_j \frac{d_j}{n}\left\|\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):(j-1)})\right\| \left\|\left(\frac{n}{d_j}\right)^{1/2}\widehat{\mathrm{Var}}^{1/2}(X_{1,(j-k):(j-1)})\,(a_j - \widehat{a}_j^{(k)})\right\|_2^2 \mid \mathbf{X}_n\right) \\
\leq\ & \frac{kMC_2}{n}\,\mathbb{E}^{\pi}\left(\max_j \left\|\left(\frac{n}{d_j}\right)^{1/2}\widehat{\mathrm{Var}}^{1/2}(X_{1,(j-k):(j-1)})\,(a_j - \widehat{a}_j^{(k)})\right\|_2^2 \mid \mathbf{X}_n\right) \\
\lesssim\ & \frac{k}{n}\,\mathbb{E}^{\widetilde{\pi}}\left(\max_j \left\|\left(\frac{n}{d_j}\right)^{1/2}\widehat{\mathrm{Var}}^{1/2}(X_{1,(j-k):(j-1)})\,(a_j - \widehat{a}_j^{(k)})\right\|_2^2 \mid \mathbf{X}_n\right) \\
=\ & \frac{k}{n}\,\mathbb{E}\left(\max_j \chi_{jk}^2\right)
\end{aligned}
$$

by Lemma 3. χ_{jk}^2 is a chi-square random variable with $k_j = \min(j-1, k)$ degree of freedom. By the maximal inequality for chi-square random variables (Example 2.7 in Boucheron et al. (2013)),

$$
\begin{aligned}
\mathbb{E}\left(\max_j \chi_{jk}^2\right) &= k_j + \mathbb{E}\left(\max_j \chi_{jk}^2 - k_j\right) \\
&\leq C\,(k + \log p)
\end{aligned}
$$

for some constant $C > 0$. Thus, we have

$$
\mathbb{E}^{\pi}\left(\|A_n - \widehat{A}_{nk}\|_\infty^2 \mid \mathbf{X}_n\right) \leq Ck\left(\frac{k + \log p}{n}\right)
$$

on $(\mathbf{X}_n \in N_n)$, for some constant $C > 0$.

Let $a_{c_j} = (a_{j+1,j}, \ldots, a_{\min(j+k,p),j})^T$ be the nonzero column vector of A_n. Since the posterior distributions for a_{c_j}'s are the independent multivariate normal distributions with finite variances whose rate is $1/n$ on $(\mathbf{X}_n \in N_n)$, it is easy to show that

$$\mathbb{E}^{\pi}\left(\|A_n - \widehat{A}_{nk}\|_1^2 \mid \mathbf{X}_n\right) \;\; \le \;\; Ck\left(\frac{k+\log p}{n}\right)$$

on $(\mathbf{X}_n \in N_n)$, for some constant $C > 0$ using similar arguments. □

Lemma 5. *Consider the model* $X_1, \ldots, X_n \overset{iid}{\sim} N_p(0, \Omega_{0,n}^{-1})$ *and the* k*-BC prior. Assume that* $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$ *defined at (2.8) and* $\sum_{m=1}^{\infty} \gamma(m) < \infty$. *If* $M \ge 9\epsilon_0^{-1}$, $\nu_0 = o(n)$, $k + \log p = o(n)$, $1 \le k \le p-1$ *and* $k^2 = O(n\log p)$, *then*

$$\mathbb{E}^{\pi}\left(\|D_n^{-1} - \widehat{D}_{nk}^{-1}\|_\infty \mid \mathbf{X}_n\right) \;\; \le \;\; C\left(\frac{\log p}{n}\right)^{1/2} \qquad \text{on } (\mathbf{X}_n \in N_n)$$

for some constant $C > 0$ *and all sufficiently large* n*, where* $\widehat{D}_{nk}$ *is defined at Lemma 2.*

Proof of Lemma 5. By Lemma 3, on $(\mathbf{X}_n \in N_n)$,

$$\mathbb{E}^{\pi}\left(\|D_n^{-1} - \widehat{D}_{nk}^{-1}\|_\infty \mid \mathbf{X}_n\right) \;\; \le \;\; C\mathbb{E}^{\widetilde{\pi}}\left(\|D_n^{-1} - \widehat{D}_{nk}^{-1}\|_\infty \mid \mathbf{X}_n\right)$$

for some constant $C > 0$. It is easy to show that

$$\begin{aligned}
\mathbb{E}^{\widetilde{\pi}}\left(\|D_n^{-1} - \widehat{D}_{nk}^{-1}\|_\infty \mid \mathbf{X}_n\right) \;\; &\le \;\; \mathbb{E}^{\widetilde{\pi}}\left(\max_j \left| d_j^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1}\right| \mid \mathbf{X}_n\right) + \max_j \left|\frac{n-n_j}{n}\widehat{d}_{jk}^{-1}\right| \\
&\le \;\; \frac{1}{\lambda}\log\exp \mathbb{E}^{\widetilde{\pi}}\left(\lambda \max_j \left| d_j^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1}\right| \mid \mathbf{X}_n\right) + \frac{2k}{n}C_2 \\
&\le \;\; \frac{1}{\lambda}\log \mathbb{E}^{\widetilde{\pi}}\left(\max_j e^{\lambda|d_j^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1}|} \mid \mathbf{X}_n\right) + \frac{2k}{n}C_2 \\
&\le \;\; \frac{1}{\lambda}\log\left[p\,\max_j \mathbb{E}^{\widetilde{\pi}}\left(e^{\lambda|d_j^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1}|} \mid \mathbf{X}_n\right)\right] + \frac{2k}{n}C_2
\end{aligned}$$

for any $\lambda > 0$, on $(\mathbf{X}_n \in N_n)$. Let $\lambda < n\widehat{d}_{jk}/2$. Note that the upper bound for the moment generating function of $|d_j^{-1} - n_j\widehat{d}_{jk}^{-1}/n|$ is

$$\begin{aligned}
\mathbb{E}^{\widetilde{\pi}}\left(e^{\lambda|d_j^{-1}-\frac{n_j}{n}\widehat{d}_{jk}^{-1}|} \mid \mathbf{X}_n\right) &= \int_0^\infty e^{\lambda|d_j^{-1}-\frac{n_j}{n}\widehat{d}_{jk}^{-1}|} Gamma\left(d_j^{-1} \mid \frac{n_j}{2}, \frac{n}{2}\widehat{d}_{jk}\right) dd_j^{-1} \\
&\leq \int_0^{n_j\widehat{d}_{jk}^{-1}/n} e^{\lambda(\frac{n_j}{n}\widehat{d}_{jk}^{-1}-d_j^{-1})} Gamma\left(d_j^{-1} \mid \frac{n_j}{2}, \frac{n}{2}\widehat{d}_{jk}\right) dd_j^{-1} \\
&+ \mathbb{E}^{\widetilde{\pi}}\left(e^{\lambda(d_j^{-1}-\frac{n_j}{n}\widehat{d}_{jk}^{-1})} \mid \mathbf{X}_n\right) \\
&\leq e^{\lambda\frac{n_j}{n}\widehat{d}_{jk}^{-1}} \int_0^\infty e^{-\lambda d_j^{-1}} Gamma\left(d_j^{-1} \mid \frac{n_j}{2}, \frac{n}{2}\widehat{d}_{jk}\right) dd_j^{-1} \\
&+ \exp\left(\frac{n_j\lambda^2}{n\widehat{d}_{jk}(n\widehat{d}_{jk}-2\lambda)}\right) \\
&\leq e^{\lambda\frac{n_j}{n}\widehat{d}_{jk}^{-1}} \left(\frac{n\widehat{d}_{jk}}{n\widehat{d}_{jk}+2\lambda}\right)^{n_j/2} + \exp\left(\frac{n_j\lambda^2}{n\widehat{d}_{jk}(n\widehat{d}_{jk}-2\lambda)}\right).
\end{aligned}$$

The second inequality follow from page 28 of Boucheron et al. (2013). Since $\lambda < n\widehat{d}_{jk}/2$,

$$\begin{aligned}
e^{\lambda\frac{n_j}{n}\widehat{d}_{jk}^{-1}} \left(\frac{n\widehat{d}_{jk}}{n\widehat{d}_{jk}+2\lambda}\right)^{n_j/2} &= e^{\lambda n_j/(n\widehat{d}_{jk})} \left(1 + \frac{2\lambda}{n\widehat{d}_{jk}}\right)^{-n_j/2} \\
&\leq \left(1 + \frac{2\lambda}{n\widehat{d}_{jk}}\right)^{\lambda n_j/(2n\widehat{d}_{jk})} \\
&= \left(1 + \frac{2\lambda}{n\widehat{d}_{jk}}\right)^{n\widehat{d}_{jk}/(2\lambda)\,\lambda^2 n_j/(n^2\widehat{d}_{jk}^2)} \\
&\leq \exp\left(\frac{\lambda^2 n_j}{n^2\widehat{d}_{jk}^2}\right),
\end{aligned}$$

where the first inequality follows from Lemma 7. Thus, on $(\mathbf{X}_n \in N_n)$,

$$\begin{aligned}
\mathbb{E}^{\widetilde{\pi}}\left(\|D_n^{-1} - \widehat{D}_{nk}^{-1}\|_\infty \mid \mathbf{X}_n\right) &\le \frac{1}{\lambda}\log\left[p\,\max_j \mathbb{E}^{\widetilde{\pi}}\left(e^{\lambda|d_j^{-1} - \frac{n_j}{n}\widehat{d}_{jk}^{-1}|} \mid \mathbf{X}_n\right)\right] + \frac{2k}{n}C_2 \\
&\le \frac{\log p}{\lambda} + \frac{1}{\lambda}\max_j \log\left[\exp\left(\frac{\lambda^2 n_j}{n^2\widehat{d}_{jk}^2}\right) + \exp\left(\frac{n_j\lambda^2}{n\widehat{d}_{jk}(n\widehat{d}_{jk} - 2\lambda)}\right)\right] \\
&+ \frac{2k}{n}C_2 \\
&\le \frac{\log p}{\lambda} + \frac{2\log 2}{\lambda} + \max_j\left(\frac{\lambda n_j}{n^2\widehat{d}_{jk}^2} + \frac{n_j\lambda}{n\widehat{d}_{jk}(n\widehat{d}_{jk} - 2\lambda)}\right) + \frac{2k}{n}C_2 \\
&\le \frac{\log p}{\lambda} + \frac{2\log 2}{\lambda} + \frac{\lambda C_2^2}{n} + \frac{\lambda C_2}{(nC_2^{-1} - 2\lambda)} + \frac{2k}{n}C_2 \\
&\le C\left(\frac{\log p}{n}\right)^{1/2}
\end{aligned}$$

for some constant $C > 0$ if we choose $\lambda \asymp (n\log p)^{1/2}$. □

Proof of Theorem 2. Note that

$$\begin{aligned}
&\mathbb{E}_{0n}\mathbb{E}^{\pi}\left(\|\Omega_n - \Omega_{0,n}\| \mid \mathbf{X}_n\right) \\
\le\ & \mathbb{E}_{0n}\left[\mathbb{E}^{\pi}\left(\|\Omega_n - \Omega_{0,n}\| \mid \mathbf{X}_n\right) I(\mathbf{X}_n \in N_n)\right] \qquad \text{(S2.32)} \\
+\ & \mathbb{E}_{0n}\left[\mathbb{E}^{\pi}\left(\|\Omega_n - \Omega_{0,n}\| \mid \mathbf{X}_n\right) I(\mathbf{X}_n \in N_n^c)\right] \qquad \text{(S2.33)}
\end{aligned}$$

where the set N_n is defined at Lemma 1. The term (S2.33) is bounded above by

$$
\begin{aligned}
& \mathbb{E}_{0n}\left[\left(\mathbb{E}^{\pi}(\|\Omega_n\| \mid \mathbf{X}_n) + \|\Omega_{0,n}\|\right) I(\mathbf{X}_n \in N_n^c)\right] \\
\leq\ & \mathbb{E}_{0n}\Big[\Big(\mathbb{E}^{\pi}(\|I_p - A_n\|_1 \|I_p - A_n\|_\infty \|D_n^{-1}\| \mid \mathbf{X}_n) + \|\Omega_{0,n}\|\Big) I(\mathbf{X}_n \in N_n^c)\Big] \\
\leq\ & \Big\{\mathbb{E}_{0n}\left[\mathbb{E}^{\pi}(\|I_p - A_n\|_1 \|I_p - A_n\|_\infty \|D_n^{-1}\| \mid \mathbf{X}_n)\right]^2\Big\}^{1/2} \mathbb{P}_{0n}(\mathbf{X}_n \in N_n^c)^{1/2} \\
+\ & \|\Omega_{0,n}\|_\infty \mathbb{P}_{0n}(\mathbf{X}_n \in N_n^c) \\
\leq\ & p^{\kappa}\, \mathbb{P}_{0n}(\mathbf{X}_n \in N_n^c)^{1/2} + \|\Omega_{0,n}\|_\infty \mathbb{P}_{0n}(\mathbf{X}_n \in N_n^c) \\
\leq\ & (p^{\kappa} + C)\ \Big(6pe^{-n(1-((k+1)/n)^{1/2})^2/8} + 4 \times 5^k e^{-C_3 C_5 \epsilon_0^2 (k + \log(n \vee p))}\Big)^{1/2} \\
\lesssim\ & n^{-1}
\end{aligned}
$$

for all sufficiently large n and some positive constants κ, C_3 and C_5. The fourth inequality follows from Lemmas 1 and 8. The third inequality holds because

$$
\begin{aligned}
& \Big[\mathbb{E}^{\pi}(\|I_p - A_n\|_1 \|I_p - A_n\|_\infty \|D_n^{-1}\| \mid \mathbf{X}_n)\Big]^2 \\
\leq\ & \Big[\mathbb{E}^{\pi}(p^3 \max_{j,l} \|I_p - A_n\|_{\max}^2 \cdot \max_j \|D_n^{-1}\|_{\max} \mid \mathbf{X}_n)\Big]^2 \\
\leq\ & p^6 \Big[\mathbb{E}^{\pi}\Big((1 + \sum_{j,l} a_{jl})^2 \cdot \sum_j d_j^{-1} \mid \mathbf{X}_n\Big)\Big]^2 \\
\leq\ & 4p^6 \Big[\sum_j \mathbb{E}^{\pi}\Big(d_j^{-1} \mid \mathbf{X}_n\Big) + \mathbb{E}^{\pi}\Big((\sum_{j,l} a_{jl})^2 \cdot \sum_j d_j^{-1} \mid \mathbf{X}_n\Big)\Big]^2 \\
\leq\ & 4p^6 \Big[p \max_j \mathbb{E}^{\pi}\Big(d_j^{-1} \mid \mathbf{X}_n\Big) + p^5\, \mathbb{E}^{\pi}\Big(\max_{j,j',l} a_{jl}^2 d_{j'}^{-1} \mid \mathbf{X}_n\Big)\Big]^2 \\
\leq\ & 4p^6 \Big[p \max_j \mathbb{E}^{\tilde{\pi}}\Big(d_j^{-1} \mid \mathbf{X}_n\Big) + p^8 \max_{j,j',l} \mathbb{E}^{\pi}\Big(a_{jl}^2 d_{j'}^{-1} \mid \mathbf{X}_n\Big)\Big]^2 \\
\leq\ & 4p^6 \Big[p \max_j \frac{n_j}{n} \widehat{d}_{jk}^{-1} + p^8 \max_{j,j',l} \Big((\widehat{a}_{jl}^{(k)})^2 + M[(n\widehat{\mathrm{Var}}(X_{1,(j-k):(j-1)}))^{-1}]_{(l-j+k+1,l-j+k+1)}\Big) \frac{n_{j'}}{n} \widehat{d}_{j'k}^{-1}\Big]^2,
\end{aligned}
$$

whose expectation is bounded above by p^c for some constant $c > 0$ by Lemma 6 and its proof, where the fifth and sixth inequalities follow from Lemma 3.

We decompose the term (S2.32) as follows:

$$\begin{aligned}
& \mathbb{E}_{0n}\left[\mathbb{E}^{\pi}\left(\|\Omega_n - \Omega_{0,n}\| \mid \mathbf{X}_n\right) I(\mathbf{X}_n \in N_n)\right] \\
\leq \; & \mathbb{E}_{0n}\left[\mathbb{E}^{\pi}\left(\|\Omega_n - \widehat{\Omega}_{nk}\| \mid \mathbf{X}_n\right) I(\mathbf{X}_n \in N_n)\right] \qquad \text{(S2.34)} \\
+ \; & \mathbb{E}_{0n}\left[\|\widehat{\Omega}_{nk} - \Omega_{0,n}\| I(\mathbf{X}_n \in N_n)\right], \qquad \text{(S2.35)}
\end{aligned}$$

where $\widehat{\Omega}_{nk}$ is defined at Lemma 2. By Lemma 2, the upper bound for (S2.35) is $Ck^{3/4}[((k + \log(n \vee p))/n)^{1/2} + \gamma(k)]$ for some constant $C > 0$ because we assume that $k^{3/2}(k + \log(n \vee p)) = O(n)$. Note that the term (S2.34) can be decomposed as (S2.25) and

$$\begin{aligned}
\|I_p - \widehat{A}_{nk}\|_1 & \leq \|I_p - A_{0,nk}\|_1 + \|\widehat{A}_{nk} - A_{0,nk}\|_1 \\
& \leq 1 + \sum_{m=1}^{\infty} \gamma(m) + Ck\gamma(k) + Ck\left(\frac{k + \log(n \vee p)}{n}\right)^{1/2}, \\
\|I_p - \widehat{A}_{nk}\|_\infty & \leq \|I_p - A_{0,nk}\|_\infty + \|\widehat{A}_{nk} - A_{0,nk}\|_\infty \\
& \leq 1 + \gamma(1) + Ck^{1/2}\gamma(k) + Ck^{1/2}\left(\frac{k + \log(n \vee p)}{n}\right)^{1/2}, \\
\|I_p - \widehat{A}_{nk}\| & \leq \|I_p - A_{0,nk}\| + \|\widehat{A}_{nk} - A_{0,nk}\| \\
& \leq 1 + \sum_{m=1}^{\infty} \gamma(m) + Ck^{3/4}\gamma(k) + Ck^{3/4}\left(\frac{k + \log(n \vee p)}{n}\right)^{1/2}
\end{aligned}$$

and $\|\widehat{D}_{nk}^{-1}\| \leq C_2$ on $(\mathbf{X}_n \in N_n)$ for some constant $C > 0$. By Lemma 4 and Lemma 5, it is easy to show that the upper bound for (S2.34) is

$Ck^{1/2}((k+\log(n\vee p))/n)^{1/2}$ for some constant $C>0$ because we assume that $k^{3/2}(k+\log(n\vee p))=O(n)$. □

Proof of Theorem 4. We can use the same arguments used in the proof of Theorem 2. It suffices to prove that

$$\|I_p-\widehat{A}_{nk}\|_1 \;\lesssim\; k^{1/2} \text{ on } (\mathbf{X}_n\in N_n).$$

It trivially holds because we assume that $k(k+\log(n\vee p))=O(n)$. □

S2.4 Proof of Corollary 1

Lemma 6 is used to prove Corollary 1.

Lemma 6. *Consider the model* $X_1,\ldots,X_n \overset{iid}{\sim} N_p(0,\Omega_{0,n}^{-1})$ *and* $\Omega_{0,n}\in\mathcal{U}(\epsilon_0,\gamma)$ *defined at (2.8). If* $k=o(n)$*, then for given positive integer* m*,*

$$\begin{aligned}\mathbb{E}_{0n}(\widehat{d}_{jk}^{-m}) &\;\lesssim\; (k+1)^{m+1},\\ \mathbb{E}_{0n}((\widehat{a}_{ji}^{(k)})^m) &\;\lesssim\; (k+1)^{2m+1},\end{aligned}$$

where $\widehat{d}_{jk}$ *and* $\widehat{a}_{ji}^{(k)}$ *be defined at (2.6).*

Proof. Note that

$$\begin{aligned}\mathbb{E}_{0n}(\widehat{d}_{jk}^{-m}) &\le \mathbb{E}_{0n}\|\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j})\|^m\\ &\le \mathbb{E}_{0n}\left[tr\left(\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j})\right)\right]^m\\ &\le (k+1)^m\sum_{l=1}^{k+1}\mathbb{E}_{0n}\left[\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j})_{(l)}\right]^m\end{aligned}$$

where for any $p \times p$ matrix A, $A_{(i)}$ is the (i,i) component of A. Also note that $[\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j})]_{(l)}$ is a inverse-gamma distribution $IG((n-k)/2, n[\mathrm{Var}^{-1}(X_{1,(j-k):j})]_{(l)}/2)$ because diagonal elements of a inverse-Wishart matrix are inverse-gamma random variables (Huang and Wand, 2013). Since $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$,

$$
\begin{aligned}
(k+1)^m \sum_l \mathbb{E}_{0n}\left[\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j})_{(l)}\right]^m &\leq (k+1)^{m+1}\left(\frac{n\epsilon_0^{-1}}{n-k-2m}\right)^m \\
&\lesssim (k+1)^{m+1}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathbb{E}_{0n}((\widehat{a}_{ji}^{(k)})^m) &\leq \mathbb{E}_{0n}\left[\|\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j})\|^m \|\widehat{\mathrm{Var}}(X_{1,(j-k):j})\|^m\right] \\
&\leq \mathbb{E}_{0n}\left\{\left[tr\left(\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j})\right)\right]^m \left[tr\left(\widehat{\mathrm{Var}}(X_{1,(j-k):j})\right)\right]^m\right\} \\
&\leq \left\{\mathbb{E}_{0n}\left[tr\left(\widehat{\mathrm{Var}}^{-1}(X_{1,(j-k):j})\right)\right]^{2m} \mathbb{E}_{0n}\left[tr\left(\widehat{\mathrm{Var}}(X_{1,(j-k):j})\right)\right]^{2m}\right\}^{1/2} \\
&\lesssim (k+1)^{2m+1}
\end{aligned}
$$

because diagonal elements of a Wishart matrix are gamma random variables (Rao, 2009), i.e. $[\widehat{\mathrm{Var}}(X_{1,(j-k):j})]_{(l)} \sim Gamma(n/2, n[\mathrm{Var}(X_{1,(j-k):j})]_{(l)}^{-1}/2)$.

□

Proof of Corollary 1. Since

$$
\begin{aligned}
\mathbb{E}_{0n}\|\widehat{\Omega}_{nk}^{LL} - \Omega_{0,n}\| &\leq \mathbb{E}_{0n}\|\mathbb{E}^\pi(\Omega_n \mid \mathbf{X}_n) - \Omega_{0,n}\| + \mathbb{E}_{0n}\|\mathbb{E}^\pi(\Omega_n \mid \mathbf{X}_n) - \widehat{\Omega}_{nk}^{LL}\| \\
&\leq \mathbb{E}_{0n}\mathbb{E}^\pi(\|\Omega_n - \Omega_{0,n}\| \mid \mathbf{X}_n) + \mathbb{E}_{0n}\|\mathbb{E}^\pi(\Omega_n \mid \mathbf{X}_n) - \widehat{\Omega}_{nk}^{LL}\|,
\end{aligned}
$$

it suffices to prove

$$\begin{aligned}\mathbb{E}_{0n}\|\mathbb{E}^{\pi}(\Omega_n \mid \mathbf{X}_n) - \widehat{\Omega}_{nk}^{LL}\|_\infty &\le \frac{Ck^2}{n} \\ &\le k^{3/4}\left[\left(\frac{k+\log(n\vee p)}{n}\right)^{1/2} + \gamma(k)\right]\end{aligned}$$

for some constant $C > 0$ because of the assumption $k(k+\log(n\vee p)) = O(n)$.

Let $\Omega_n = (\omega_{ij})$ and $\widehat{\Omega}_{nk}^{LL} = (\widehat{\omega}_{ij}^{LL})$, then for $i < j \le i+k$,

$$\begin{aligned}&\mathbb{E}_{0n}\left|\mathbb{E}^{\pi}(\omega_{ij} \mid \mathbf{X}_n) - \widehat{\omega}_{ij}^{LL}\right| \\ \le\ & \mathbb{E}_{0n}\left|\mathbb{E}^{\pi}(d_j^{-1}a_{ji} \mid \mathbf{X}_n) - \frac{n_j}{n}\widehat{d}_{jk}^{-1}\widehat{a}_{ji}^{(k)}\right| \qquad \text{(S2.36)} \\ +\ & \sum_{l=j+1}^{i+k}\mathbb{E}_{0n}\left|\mathbb{E}^{\pi}(d_l^{-1}a_{li}a_{lj} \mid \mathbf{X}_n) - \frac{n_l}{n}\widehat{d}_{lk}^{-1}\widehat{a}_{li}^{(k)}\widehat{a}_{lj}^{(k)}\right| \qquad \text{(S2.37)}\end{aligned}$$

by (S2.4). The (S2.36) term can be decomposed by

$$\begin{aligned}&\mathbb{E}_{0n}\left|\left(\mathbb{E}^{\pi}(d_j^{-1}a_{ji} \mid \mathbf{X}_n) - \frac{n_j}{n}\widehat{d}_{jk}^{-1}\widehat{a}_{ji}^{(k)}\right) I(\mathbf{X}_n \in N_n)\right| \qquad \text{(S2.38)} \\ +\ & \mathbb{E}_{0n}\left|\left(\mathbb{E}^{\pi}(d_j^{-1}a_{ji} \mid \mathbf{X}_n) - \frac{n_j}{n}\widehat{d}_{jk}^{-1}\widehat{a}_{ji}^{(k)}\right) I(\mathbf{X}_n \in N_n^c)\right|. \qquad \text{(S2.39)}\end{aligned}$$

To deal with the above terms, we need to compute the expectation of truncated distributions. When Y is a truncated gamma distribution $Y \sim Gamma^{Tr}(\alpha, \beta, c_1 \le Y \le c_2)$, the expectation of Y is

$$\mathbb{E}Y = \frac{\alpha}{\beta}\frac{\int_{c_1}^{c_2} Gamma(y \mid \alpha+1, \beta)dy}{\int_{c_1}^{c_2} Gamma(y \mid \alpha, \beta)dy}$$

(Coffey and Muller, 2000). Thus, one can show that (S2.38) is bounded above by

$$
\begin{aligned}
& \mathbb{E}_{0n} \left| \frac{n_j}{n} \widehat{d}_{jk}^{-1} \widehat{a}_{ji}^{(k)} \left(\frac{\int_0^M Gamma(d_j^{-1} \mid \frac{n_j}{2}+1, \frac{n}{2}\widehat{d}_{jk}) dd_j^{-1}}{\int_0^M Gamma(d_j^{-1} \mid \frac{n_j}{2}, \frac{n}{2}\widehat{d}_{jk}) dd_j^{-1}} - 1 \right) I(\mathbf{X}_n \in N_n) \right| \\
\leq \; & C_1 C_2^2 e^{-cn}
\end{aligned}
$$

for all sufficiently large n and some positive constant c by the same argument with (S2.31). On the other hand, (S2.39) is bounded above by

$$
\begin{aligned}
& C \left[\mathbb{E}_{0n} (\widehat{d}_{jk}^{-2} (\widehat{a}_{ji}^{(k)})^2) \right]^{1/2} \mathbb{P}_{0n}(\mathbf{X}_n \in N_n^c) \\
\lesssim \; & (k+1)^{7/2} \mathbb{P}_{0n}(\mathbf{X}_n \in N_n^c) \\
\leq \; & \frac{1}{n^2}
\end{aligned}
$$

for some constant $C > 0$ and all sufficiently large n by Lemma 1, Lemma 6 and the choice of large C_3 in the set N_n.

The (S2.37) can be decomposed by

$$
\sum_{l=j+1}^{i+k} \mathbb{E}_{0n} \left| \left(\mathbb{E}^{\pi}(d_l^{-1} a_{li} a_{lj} \mid \mathbf{X}_n) - \frac{n_l}{n} \widehat{d}_{lk}^{-1} \widehat{a}_{li}^{(k)} \widehat{a}_{lj}^{(k)} \right) I(\mathbf{X}_n \in N_n) \right| \quad \text{(S2.40)}
$$

$$
+ \sum_{l=j+1}^{i+k} \mathbb{E}_{0n} \left| \left(\mathbb{E}^{\pi}(d_l^{-1} a_{li} a_{lj} \mid \mathbf{X}_n) - \frac{n_l}{n} \widehat{d}_{lk}^{-1} \widehat{a}_{li}^{(k)} \widehat{a}_{lj}^{(k)} \right) I(\mathbf{X}_n \in N_n^c) \right| \quad \text{(S2.41)}
$$

Note that in (S2.40),

$$
\begin{aligned}
\mathbb{E}^{\pi}(d_l^{-1} a_{li} a_{lj} \mid \mathbf{X}_n) &= \mathbb{E}^{\pi}(d_l^{-1} \mathbb{E}^{\pi}(a_{li} a_{lj} \mid d_l, \mathbf{X}_n) \mid \mathbf{X}_n) \\
&= \mathbb{E}^{\pi}(d_l^{-1} \mathbb{E}^{\pi}(a_{li} \mid d_l, \mathbf{X}_n) \mathbb{E}^{\pi}(a_{lj} \mid d_l, \mathbf{X}_n) \mid \mathbf{X}_n) \\
&+ \mathbb{E}^{\pi}(d_l^{-1} \mathrm{Cov}^{\pi}(a_{li}, a_{lj} \mid d_l, \mathbf{X}_n) \mid \mathbf{X}_n).
\end{aligned}
$$

If we prove that $\sum_{l=j+1}^{i+k} \mathbb{E}_{0n} |\mathbb{E}^\pi(d_l^{-1} \mathrm{Cov}^\pi(a_{li}, a_{lj} \mid d_l, \mathbf{X}_n) \mid \mathbf{X}_n) I(\mathbf{X}_n \in N_n)| \lesssim k/n$, (S2.40) is bounded above by Ck/n for some constant $C > 0$ by the similar arguments used in (S2.38). It is easy to show that

$$\begin{aligned}
& \sum_{l=j+1}^{i+k} \mathbb{E}_{0n} \left| \mathbb{E}^\pi(d_l^{-1} \mathrm{Cov}^\pi(a_{li}, a_{lj} | d_l, \mathbf{X}_n) \mid \mathbf{X}_n) I(\mathbf{X}_n \in N_n) \right| \\
\leq & \sum_{l=j+1}^{i+k} \mathbb{E}_{0n} \left[\mathbb{E}^\pi \left(d_l^{-1} \left| \mathrm{Cov}^\pi(a_{li}, a_{lj} | d_l, \mathbf{X}_n) \right| \,\middle|\, \mathbf{X}_n \right) I(\mathbf{X}_n \in N_n) \right] \\
\leq & \sum_{l=j+1}^{i+k} \mathbb{E}_{0n} \left(\mathbb{E}^\pi \left(d_l^{-1} \left[\mathrm{Var}^\pi(a_{li} | d_l, \mathbf{X}_n) \mathrm{Var}^\pi(a_{lj} | d_l, \mathbf{X}_n) \right]^{1/2} \,\middle|\, \mathbf{X}_n \right) I(\mathbf{X}_n \in N_n) \right) \\
\lesssim & \frac{k}{n}.
\end{aligned}$$

Similar to (S2.39), (S2.41) is bounded above by C/n^2 for some constant $C > 0$. Thus, we have shown

$$\mathbb{E}_{0n} \left| \mathbb{E}^\pi(\omega_{ij} \mid \mathbf{X}_n) - \widehat{\omega}_{ij}^{LL} \right| \lesssim \frac{k}{n}$$

for any $i < j \leq i + k$. Since $\omega_{ii} = d_i^{-1} + \sum_{l=i+1}^{i+k} d_l^{-1} a_{li}^2$ for $i < p$ and $\omega_{pp} = d_p^{-1}$,

$$\mathbb{E}_{0n} \left| \mathbb{E}^\pi(\Omega_{n,ii} \mid \mathbf{X}_n) - \widehat{\Omega}_{ii}^{LL} \right| \lesssim \frac{k}{n}$$

can be shown easily for $1 \leq i \leq p$ by similar arguments. Thus, it implies

$$\mathbb{E}_{0n} \| \mathbb{E}^\pi(\Omega_n \mid \mathbf{X}_n) - \widehat{\Omega}_{nk}^{LL} \|_\infty \lesssim \frac{k^2}{n}.$$

□

S3 Auxiliary results

Lemma 7. *For any $x, n > 0$,*

$$e^x \;\le\; \Big(1 + \frac{x}{n}\Big)^{n+x/2} .$$

The proof can be obtained by a simple algebra.

Lemma 8. *If we assume that $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$ (defined at (2.8)) and $\sum_{k=1}^{\infty} \gamma(k) < \infty$, then*

$$\|\Omega_{0,n}\|_\infty \;<\; C$$

for some $C > 0$ not depending on p.

Proof. Let $\Omega_{0,n} = (I_p - A_{0,n})^T D_{0,n}^{-1}(I_p - A_{0,n})$ be the modified Cholesky decomposition of $\Omega_{0,n}$. Since $\|\Omega_{0,n}\|_\infty \le \|I_p - A_{0,n}\|_1 \|D_{0,n}^{-1}\|_\infty \|I_p - A_{0,n}\|_\infty$ and

$$\begin{aligned}
\|I_p - A_{0,n}\|_\infty &\le 1 + \|A_{0,n}\|_\infty \le 1 + \gamma(1), \\
\|D_{0,n}^{-1}\|_\infty &= \max_j d_{0,j}^{-1} \\
&= \max_j \left\| \mathrm{Var}^{1/2}(X_{1,1:j}) \begin{pmatrix} -a_{0,j} \\ 1 \end{pmatrix} \right\|_2^{-2} \\
&\le \max_j \lambda_{\min}\left(\mathrm{Var}(X_{1,1:j})\right)^{-1} = \max_j \left\|\mathrm{Var}^{-1}(X_{1,1:j})\right\| \le \epsilon_0^{-1},
\end{aligned} \tag{S3.42}$$

we only need to prove $\|A_{0,n}\|_1 \le C$ for some $C > 0$. By the definition of

$\mathcal{U}(\epsilon_0, \gamma)$, it is easy to show $|a_{0,ij}| \leq \gamma(i-j)$ for all $i > j$. Thus,

$$\begin{aligned} \|A_{0,n}\|_1 &= \max_j \sum_{i=j+1}^{p} |a_{0,ij}| \\ &\leq \max_j \sum_{i=j+1}^{p} \gamma(i-j) \\ &\leq \sum_{m=1}^{\infty} \gamma(m) \;<\; \infty. \end{aligned}$$

□

Lemma 9. *For any positive integers p_1 and p_2, let A_{11}, A_{12} and A_{22} be a $p_1 \times p_1, p_1 \times p_2$ and $p_2 \times p_2$ matrix,*

$$\|A_{12}\| \leq \left\| \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix} \right\|,$$

where $||\cdot||$ is the matrix L_2 norm.

Proof. Note

$$\begin{aligned} \left\| \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix} \right\| &= \sup_{\|x\|_2=1} \left\| \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix} x \right\|_2 \\ &= \sup_{\|x\|_2=1} \left\| \begin{pmatrix} A_{11}x_1 + A_{12}x_2 \\ A_{22}x_2 + A_{12}^T x_1 \end{pmatrix} \right\|_2 \\ &\geq \sup_{\|x_2\|_2=1} \left\| \begin{pmatrix} A_{12}x_2 \\ A_{22}x_2 \end{pmatrix} \right\|_2 \geq \sup_{\|x_2\|_2=1} \|A_{12}x_2\|_2 = \|A_{12}\| \end{aligned}$$

where $x = (x_1^T, x_2^T)^T$ and $x_1 \in \mathbb{R}^{p_1}, x_2 \in \mathbb{R}^{p_2}$. This completes the proof. □

Lemma 10. *If we assume that $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$, which is defined at (2.8), then*

$$\|A_{0,nk} - A_{0,n}\|_\infty \;\leq\; Ck^{1/2}\gamma(k),$$
$$\|A_{0,nk} - A_{0,n}\|_1 \;\leq\; Ck\gamma(k)$$

for some $C > 0$, where $A_{0,nk}$ is defined at (S2.24).

Proof of Lemma 10. We only consider $k < j-1$ case because $A_{0,nk} = A_{0,n}$ trivially holds when $k \geq j-1$. Note first that

$$\|A_{0,nk} - A_{0,n}\|_\infty \;\leq\; \|A_{0,nk} - B_k(A_{0,n})\|_\infty + \|B_k(A_{0,n}) - A_{0,n}\|_\infty.$$

The second term is bounded above by $\gamma(k)$ by the definition of $\mathcal{U}(\epsilon_0, \gamma)$. Denote

$$\mathrm{Var}^{-1}(X_{1,1:(j-1)}) \;=\; \begin{pmatrix} \Omega_{11,j} & \Omega_{12,j} \\ \Omega_{21,j} & \Omega_{22,j} \end{pmatrix},$$
$$\mathrm{Cov}(X_{1,1:(j-1)}, X_{1j}) \;=\; \begin{pmatrix} \Sigma_{1j} \\ \Sigma_{2j} \end{pmatrix},$$

where $\Omega_{11,j}$ is a $(j-k-1) \times (j-k-1)$ matrix, $\Omega_{22,j}$ is a $k \times k$ matrix and $\Sigma_{2j} = \mathrm{Cov}(X_{1,(j-k):(j-1)}, X_{1j})$ is a k-dimensional vector. Since $\max_j \|a_{0,j} - B_{k-1,j}(a_{0,j})\|_1 \leq \gamma(k)$ by assumption, it directly implies

$$\max_j \|\Omega_{11,j}\Sigma_{1j} + \Omega_{12,j}\Sigma_{2j}\|_1 \;\leq\; \gamma(k). \tag{S3.43}$$

Also note that $\text{Var}^{-1}(X_{1,(j-k):(j-1)}) = \Omega_{22,j} - \Omega_{21,j}\Omega_{11,j}^{-1}\Omega_{12,j}$ by the inversion of partitioned matrix. With this fact, we have the following upper bound for $\|A_{0,nk} - B_k(A_{0,n})\|_\infty$,

$$\begin{aligned}
\|A_{0,nk} - B_k(A_{0,n})\|_\infty &= \max_j \|a_{0,j}^{(k)} - B_{k-1,j}(a_{0,j})\|_1 \\
&= \max_j \|\Omega_{21,j}\Sigma_{1j} + \Omega_{21,j}\Omega_{11,j}^{-1}\Omega_{12,j}\Sigma_{2j})\|_1 \\
&= \max_j \|\Omega_{21,j}\Omega_{11,j}^{-1}(\Omega_{11,j}\Sigma_{1j} + \Omega_{12,j}\Sigma_{2j})\|_1 \\
&\le \max_j \|\Omega_{21,j}\Omega_{11,j}^{-1}\|_1 \|\Omega_{11,j}\Sigma_{1j} + \Omega_{12,j}\Sigma_{2j}\|_1 \\
&\le \max_j k^{1/2}\|\Omega_{21,j}\Omega_{11,j}^{-1}\| \|\Omega_{11,j}\Sigma_{1j} + \Omega_{12,j}\Sigma_{2j}\|_1 \\
&\le \max_j k^{1/2}\|\Omega_{21,j}\| \|\Omega_{11,j}^{-1}\| \cdot \gamma(k) \\
&\le \epsilon_0^{-2} k^{1/2}\gamma(k).
\end{aligned}$$

The second inequality holds because $\|A\|_1 \le {p_1}^{1/2}\|A\|$ for any $p_1 \times p_2$ matrix A (Horn and Johnson, 1990). The third inequality follows from the Cauchy-Schwarz inequality and (S3.43). The last inequality holds because $\|\Omega_{21,j}\| \le \|\text{Var}^{-1}(X_{1,1:(j-1)})\| = \lambda_{\min}(\text{Var}(X_{1,1:(j-1)}))^{-1} \le \lambda_{\min}(\Omega_{0,n})^{-1} \le \epsilon_0^{-1}$ and $\|\Omega_{11,j}^{-1}\| = \lambda_{\min}(\Omega_{11,j})^{-1} \le \lambda_{\min}(\text{Var}^{-1}(X_{1,1:(j-1)}))^{-1} = \lambda_{\max}(\text{Var}(X_{1,1:(j-1)})) \le \lambda_{\max}(\Omega_{0,n}) \le \epsilon_0^{-1}$ by Lemma 9 and $\Omega_{0,n} \in \mathcal{U}(\epsilon_0, \gamma)$. It proves the first part of Lemma 10.

To show the second argument of Lemma 10, note that

$$\|A_{0,nk} - A_{0,n}\|_1 \le \|A_{0,nk} - B_k(A_{0,n})\|_1 + \|B_k(A_{0,n}) - A_{0,n}\|_1.$$

The first term is bounded above by

$$\begin{aligned}
\|A_{0,nk} - B_k(A_{0,n})\|_1 &\leq k \max_j \|a_{0,j}^{(k)} - B_{k-1,j}(a_{0,j})\|_{\max} \\
&\leq k \max_j \|a_{0,j}^{(k)} - B_{k-1,j}(a_{0,j})\| \\
&= k \max_j \|\Omega_{21,j}\Omega_{11,j}^{-1}(\Omega_{11,j}\Sigma_{1j} + \Omega_{12,j}\Sigma_{2j})\|_2 \\
&\leq k \max_j \|\Omega_{21,j}\Omega_{11,j}^{-1}\|\|\Omega_{11,j}\Sigma_{1j} + \Omega_{12,j}\Sigma_{2j}\|_2 \\
&\leq \epsilon_0^{-2} k\gamma(k)
\end{aligned}$$

by the similar arguments used in the previous paragraph. Also note that

$$\begin{aligned}
\|B_k(A_{0,n}) - A_{0,n}\|_1 &= \sum_{i=j+k}^{p} |a_{0,ij}| \\
&\leq \sum_{i=j+k}^{p} \sum_{j'=1}^{j} |a_{0,ij'}| \\
&\leq \sum_{i=j+k}^{p} \gamma(i-j) \\
&\leq \sum_{m=k}^{\infty} \gamma(m).
\end{aligned}$$

If we assume the polynomially decreasing $\gamma(k) = Ck^{-\alpha}$, we have $\sum_{m=k}^{\infty} \gamma(m) \leq C' k\gamma(k)$ for some constant $C' > 0$. If we assume the exact band or exponentially decreasing $\gamma(k) = Ce^{-\beta k}$, it is easy to show that $\sum_{m=k}^{\infty} \gamma(m) \leq C'\gamma(k)$ for some constant $C' > 0$. Thus, $\|A_{0,nk} - A_{0,n}\|_1$ is bounded above by $C'' k\gamma(k)$ for some constant $C'' > 0$. □

References

Assouad, P. (1983). Deux remarques sur l'estimation. *Comptes rendus des séances de l'Académie des sciences. Série 1 Mathématique 296*(23), 1021–1024.

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

Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford.

Cai, T. T., C.-H. Zhang, and H. H. Zhou (2010). Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics 38*(4), 2118–2144.

Cai, T. T. and H. H. Zhou (2012). Minimax estimation of large covariance matrices under ℓ_1-norm. *Statistica Sinica 22*(4), 1319–1349.

Coffey, C. S. and K. E. Muller (2000). Properties of doubly-truncated gamma variables. *Communications in Statistics-Theory and Methods 29*(4), 851–857.

Horn, R. and C. Johnson (1990). *Matrix Analysis*. Cambridge University Press.

Huang, A. and M. P. Wand (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis 8*(2), 439–452.

LeCam, L. (1973). Convergence of estimates under dimensionality restrictions. *The Annals of Statistics 1*(1), 38–53.

Lee, K. and J. Lee (2018). Optimal Bayesian Minimax Rates for Unconstrained Large Covariance

Matrices. *Bayesian Analysis 13*(4), 1211–1229.

Liu, Y. and Z. Ren (2017, December). Minimax Estimation of Large Precision Matrices with Bandable Cholesky Factor. *ArXiv e-prints*.

Rao, C. (2009). *Linear Statistical Inference and its Applications.* Wiley Series in Probability and Statistics. Wiley.

Tsybakov, A. (2008). *Introduction to Nonparametric Estimation.* Springer Series in Statistics. Springer New York.