SUPPLEMENTARY MATERIAL

Appendix A: Proofs of the Main Theorems

For convenience, we introduce the following additional notation that will be used throughout the Appendix.

i. Let $\tilde{W}$ denote the difference between the sample covariance matrix S and the true covariance matrix $\Sigma^0 = (\Theta^0)^{-1}$ and Δ the difference between an estimate $\tilde{\Theta}$ and the true precision matrix Θ^0. That is,

$$\begin{aligned} \tilde{W} &= S - \Sigma^0 \\ \Delta &= \tilde{\Theta} - \Theta^0. \end{aligned}$$

ii. Let $R(\Delta)$ denote the difference between $n\tilde{\Theta}^{-1}/2$, the gradient of $n \log \det(\tilde{\Theta})/2$, and its first-order Taylor expansion at Θ^0:

$$R(\Delta) = \frac{n}{2}\left(\tilde{\Theta}^{-1} - \Sigma^0 + \Sigma^0 \Delta \Sigma^0\right).$$

iii. Recall our objective function

$$L(\Theta) = \frac{n}{2}\Big(\mathrm{tr}(S\Theta) - \log\det(\Theta)\Big) + \frac{1}{2}\sum_{i,j} \mathrm{pen}_{SS}(\theta_{ij}) + \sum_i \mathrm{pen}_1(\theta_{ii}),$$

where

$$\mathrm{pen}_{SS}(\theta_{ij}) = -\log\left[\left(\frac{\eta}{2v_1}\right)e^{-\frac{|\theta_{ij}|}{v_1}} + \left(\frac{1-\eta}{2v_0}\right)e^{-\frac{|\theta_{ij}|}{v_0}}\right], \text{ and } \ \mathrm{pen}_1(\theta_{ii}) = \tau|\theta_{ii}|$$

denote the penalty terms on θ_{ij} $(i \neq j)$ and θ_{ii}, respectively.

Let Z_{ij} denote the subgradient of the penalty term with respect to θ_{ij}:

$$Z_{ij} = Z_{ij}(\theta_{ij}) = \begin{cases} \tau & \text{if } i = j \\ \frac{1}{2}\mathrm{pen}'_{SS}(\theta_{ij}) & \text{if } i \neq j, \quad \theta_{ij} \neq 0 \\ [-1,1] \times \dfrac{\frac{\eta}{2v_1^2} + \frac{1-\eta}{2v_0^2}}{\frac{\eta}{v_1} + \frac{1-\eta}{v_0}} & \text{if } i \neq j, \quad \theta_{ij} = 0 \end{cases}$$

where

$$\mathrm{pen}'_{SS}(\theta_{ij}) = \frac{\frac{\eta}{2v_1^2}e^{-\frac{|\theta_{ij}|}{v_1}} + \frac{1-\eta}{2v_0^2}e^{-\frac{|\theta_{ij}|}{v_0}}}{\frac{\eta}{2v_1}e^{-\frac{|\theta_{ij}|}{v_1}} + \frac{1-\eta}{2v_0}e^{-\frac{|\theta_{ij}|}{v_0}}}\mathrm{sign}(\theta_{ij}).$$

Let $Z = [Z_{ij}]$, then the subgradient of the objective function $L(\Theta)$ is

$$\partial L(\Theta) = \frac{n}{2}\left(S - \Theta^{-1}\right) + Z.$$

iv. We denote the index set of diagonal entries as $\mathcal{D} := \{(i,j) : i = j\}$. For any subset $\mathcal{S}$ of $\{(i,j) : 1 \le i, j \le p\}$ and $p \times p$ matrix A, we use $A_{\mathcal{S}}$ to denote the submatrix of A with entries indexed by $\mathcal{S}$.

In this Appendix, we first prove the following main result.

Theorem A. *Assume condition (A1) and $\|\tilde{W}\|_\infty = \max_{ij}|s_{ij} - \sigma^0_{ij}| \le C_1\sqrt{\log p/n}$. If (i) the prior hyper-parameters v_0, v_1, η and τ satisfy:*

$$\begin{cases} \frac{1}{nv_1} = C_3\sqrt{\frac{\log p}{n}}(1-\varepsilon_1), \ \textit{where } C_3 < C_2, \varepsilon_1 > 0, \\ \frac{1}{nv_0} > C_4\sqrt{\frac{\log p}{n}}, \\ \frac{v_1^2(1-\eta)}{v_0^2\eta} \le \varepsilon_1 p^{2(C_2-C_3)M_{\Gamma^0}[C_4-C_3]}, \\ \tau \le C_3\frac{n}{2}\sqrt{\frac{\log p}{n}}, \end{cases} \tag{1}$$

where $C_4 = (C_1 + M_{\Sigma^0}^2 2(C_1+C_3)M_{\Gamma^0} + 6(C_1+C_3)^2 d M_{\Gamma^0}^2 M_{\Sigma^0}^3/M)$,
(ii) the spectral norm B satisfies $1/k_1 + 2d(C_1+C_3)M_{\Gamma^0}\sqrt{\log p/n} < B < (2nv_0)^{\frac{1}{2}}$, and
(iii) the sample size n satisfies $\sqrt{n} \ge M\sqrt{\log p}$, where

$$M = \max\left\{2d(C_1+C_3)M_{\Gamma^0}\max\left(3M_{\Sigma^0}, 3M_{\Gamma^0}M_{\Sigma^0}^3, 2/k_1^2\right), 2C_3\varepsilon_1/k_1^2\right\},$$

then the MAP estimator $\tilde{\Theta}$ satisfies

$$\|\tilde{\Theta} - \Theta^0\|_\infty < 2(C_1+C_3)M_{\Gamma^0}\sqrt{\frac{\log p}{n}}.$$

Before presenting our proof, we list two preliminary results as lemmas and list some properties of the penalty function $\text{pen}_{SS}(\delta)$, which will be useful. Proofs of these lemmas are in Appendix B.

Lemma 1. *Define* $r := \max\left\{2M_{\Gamma^0}\left(\|\tilde{W}\|_\infty + \frac{2}{n}\max(\frac{1}{2}pen'_{SS}(\delta), \tau)\right), 2(C_1+C_3)M_{\Gamma^0}\sqrt{\frac{\log p}{n}}\right\}$, *and* $\mathcal{A} := \left\{\Theta : \frac{n}{2}\left(S - \Theta^{-1}\right)_{\mathcal{B}} + Z_{\mathcal{B}} = 0, \Theta \succ 0, \|\Theta\|_2 \le B\right\}$ *with* $\mathcal{B} = \{(i,j) : |\theta^0_{ij}| > 2(C_1 + C_3)M_{\Gamma^0}\sqrt{\log p/n}\} \cup \mathcal{D}$. *If parameters r and B satisfy:*

$$\begin{cases} r \le \min\left\{\frac{1}{3M_{\Sigma^0}d}, \frac{1}{3dM_{\Gamma^0}M^3_{\Sigma^0}}\right\}, \\ \min|\theta^0_{\mathcal{B}\cap\mathcal{D}^c}| \ge r + \delta, \\ 1/k_1 + dr < B, \end{cases}$$

for some $\delta > 0$, where k_1 is the lower bound on $\lambda_{\min}(\Sigma^0)$, then the set $\mathcal{A}$ is non-empty. Moreover, there exists a $\tilde{\Theta} \in \mathcal{A}$ such that $\|\Delta\|_\infty := \|\tilde{\Theta} - \Theta^0\|_\infty \le r$.

Lemma 2. *Suppose that $\|\tilde{\Theta} - \Theta^0\|_\infty \le r$, then*

$$\|\tilde{\Theta} - \Theta^0\|_F \le r\sqrt{p+s}, \tag{2}$$

$$\left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert \tilde{\Theta} - \Theta^0 \right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_\infty, \|\tilde{\Theta} - \Theta^0\|_2 \le r\min\{d, \sqrt{p+s}\}, \ \textit{and} \tag{3}$$

$$\|\tilde{\Theta}^{-1} - \Sigma^0\|_\infty \le M^2_{\Sigma^0}r + \frac{3}{2}dM^3_{\Sigma^0}r^2. \tag{4}$$

Properties of pen$_{SS}(\delta)$

We now provide some useful results on the penalty function $\text{pen}_{SS}(\delta)$.

- Bound on the magnitude of the first derivative of $\text{pen}_{SS}(\delta)$:

$$\begin{aligned} \frac{1}{n}|\text{pen}'_{SS}(\delta)| &= \frac{\frac{\eta}{2v_1^2}e^{-\frac{|\delta|}{v_1}} + \frac{1-\eta}{2v_0^2}e^{-\frac{|\delta|}{v_0}}}{n\left(\frac{\eta}{2v_1}e^{-\frac{|\delta|}{v_1}} + \frac{1-\eta}{2v_0}e^{-\frac{|\delta|}{v_0}}\right)} \\ &= \frac{1}{nv_1} + \frac{\frac{1}{n}(\frac{1}{v_0} - \frac{1}{v_1})}{\frac{\eta v_0}{(1-\eta)v_1}e^{\frac{|\delta|}{v_0} - \frac{|\delta|}{v_1}} + 1} \\ &< \frac{1}{nv_1}\left(1 + \frac{\frac{v_1^2(1-\eta)}{v_0^2\eta}}{e^{\frac{|\delta|}{v_0} - \frac{|\delta|}{v_1}}}\right). \end{aligned} \tag{5}$$

Choose $1/(nv_0) > C_4\sqrt{\log p/n}$ and $1/(nv_1) < C_3\sqrt{\log p/n}$ as in Theorem A, and if further let $v_1^2(1-\eta)/(v_0^2\eta) = \xi p^{\psi[C_4-C_3]}$, when $\delta \geq \psi\sqrt{\log p/n}$, then we have

$$\frac{\frac{v_1^2(1-\eta)}{v_0^2\eta}}{e^{\frac{|\delta|}{v_0}-\frac{|\delta|}{v_1}}} \leq \frac{\xi p^{\psi[C_4-C_3]}}{p^{\psi[C_4-C_3]}} \leq \xi. \tag{6}$$

Let ξ to be sufficiently small, i.e., $\xi < \varepsilon_1$, then we have

$$\frac{1}{n}|\mathrm{pen}'_{SS}(\delta)| < C_3\sqrt{\frac{\log p}{n}}.$$

- Bound on the magnitude of the second derivative of $\mathrm{pen}_{SS}(\delta)$:

  With the same choice of v_0 and v_1 as in Theorem A, when $\delta \geq \psi\sqrt{\log p/n}$, we have

$$\begin{aligned}
\frac{1}{2n}|\mathrm{pen}''_{SS}(\delta)| &= \frac{\left(\frac{1}{v_0}-\frac{1}{v_1}\right)\frac{\eta v_0}{(1-\eta)v_1}e^{\frac{\delta}{v_0}-\frac{\delta}{v_1}}}{2n\left(\frac{\eta v_0}{(1-\eta)v_1}e^{\frac{\delta}{v_0}-\frac{\delta}{v_1}}+1\right)^2} \\
&< \frac{\left(\frac{1}{v_0}-\frac{1}{v_1}\right)}{2n\left(\frac{\eta v_0}{(1-\eta)v_1}e^{\frac{\delta}{v_0}-\frac{\delta}{v_1}}+1\right)} \\
&< \frac{(1-\eta)v_1}{2nv_0^2\eta e^{\frac{\delta}{v_0}-\frac{\delta}{v_1}}} < \frac{\xi}{2nv_1} && (7) \\
&< \frac{C_3}{2}\xi\sqrt{\frac{\log p}{n}} < \frac{C_3}{2}\varepsilon_1\sqrt{\frac{\log p}{n}}, && (8)
\end{aligned}$$

  where (7) is due to (6). In addition, when n satisfies the condition (iii) in Theorem A, (8) is always upper bounded by $\frac{1}{4}k_1^2$.

Proof of Theorem A. Our proof is inspired by the techniques from Rothman et al. (2008) and Ravikumar et al. (2011).

Here is the outline of the proof.

- *Step 1:* Construct a solution set $\mathcal{A}$ for the constraint problem:

$$\arg\min_{\Theta\succ 0, \|\Theta\|_2\leq B, \Theta_{\mathcal{B}^c}=0} L(\Theta),$$

  by defining

$$\mathcal{A} = \left\{\Theta : \frac{n}{2}\left(S-\Theta^{-1}\right)_{\mathcal{B}} + Z_{\mathcal{B}} = 0, \Theta\succ 0, \|\Theta\|_2 \leq B\right\},$$

where $\mathcal{B} = \{(i,j) : |\theta_{ij}^0| > 2(C_1 + C_3)M_{\Gamma^0}\sqrt{\log p/n}\} \cup \mathcal{D}$. For $\theta_{ij}^0 \in \mathcal{B} \cap \mathcal{D}^c$ and , define $\min\left(|\theta_{ij}^0|\right)$ as $2(C_1 + C_2)M_{\Gamma^0}\sqrt{\log p/n}$. We then have $|\theta_{ij}^0| \geq 2(C_1 + C_2)M_{\Gamma^0}\sqrt{\log p/n}$ when $\theta_{ij}^0 \in \mathcal{B} \cap \mathcal{D}^c$ and $|\theta_{ij}^0| \leq 2(C_1 + C_3)M_{\Gamma^0}\sqrt{\log p/n}$ when $\theta_{ij}^0 \in \mathcal{B}^c \cap \mathcal{D}^c$.

- *Step 2:* Prove $\mathcal{A}$ is not empty and further show that there exists $\tilde{\Theta} \in \mathcal{A}$ satisifying $\|\tilde{\Theta} - \Theta^0\|_\infty = O_p\Big(\sqrt{\log p/n}\Big)$.

- *Step 3:* Finally prove that $\tilde{\Theta}$, which is positive definite by construction, is a local minimizer of the loss function $L(\Theta)$ by showing $L(\Theta) \geq L(\tilde{\Theta})$ for any Θ in a small neighborhood of $\tilde{\Theta}$. Since $L(\Theta)$ is strictly convex when $B < (2nv_0)^{\frac{1}{2}}$, we then conclude that $\tilde{\Theta}$ is the unique minimizer such that $\|\tilde{\Theta} - \Theta^0\|_\infty = O_p\left(\sqrt{\log p/n}\right)$.

At *Step 2*, we apply Lemma 1. First we check its conditions.

1. Consider $r = 2(C_1 + C_3)M_{\Gamma^0}\sqrt{\log p/n}$. For $\theta_{ij}^0 \in \mathcal{B} \cap \mathcal{D}^c$, we have $\theta_{ij}^0 \geq r + 2(C_2 - C_3)M_{\Gamma^0}\sqrt{\log p/n}$. That is, the δ defined in Lemma 1 is greater or equal to $2(C_2 - C_3)M_{\Gamma^0}\sqrt{\log p/n}$.

2. Recall the properties of $\mathrm{pen}_{SS}(\delta)$. We have $|\mathrm{pen}'{}_{SS}(\delta)|/n < C_3\sqrt{\log p/n}$. With the bound of $\|\tilde{W}\|_\infty$ and the condition on sample size n, we have

$$
\begin{aligned}
2M_{\Gamma^0}\left(\|W\|_\infty + \max\left(\frac{1}{n}\mathrm{pen}'_{SS}(\delta), \frac{2}{n}\tau\right)\right) &\leq 2(C_1 + C_3)M_{\Gamma^0}\sqrt{\frac{\log p}{n}} \\
&\leq \min\left\{\frac{1}{3M_{\Sigma^0}d}, \frac{1}{\frac{3}{2}dM_{\Gamma^0}M_{\Sigma^0}^3}\right\}.
\end{aligned}
$$

Thus, conditions for Lemma 1 are all satisfied. By Lemma 1, we conclude that there exists a solution $\tilde{\Theta} \in \mathcal{A}$ satisfying

$$
\|\tilde{\Theta} - \Theta^0\|_\infty = \|\Delta\|_\infty \leq 2(C_1 + C_3)M_{\Gamma^0}\sqrt{\frac{\log p}{n}}.
$$

That is, the solution $\tilde{\Theta}$ we constructed is $O_p\left(\sqrt{\log p/n}\right)$ from the truth in entrywise l_∞ norm.

At *Step 3*, we need to show that the solution $\tilde{\Theta}$ we constructed is indeed a local minimizer of the objective function $L(\Theta)$. It suffices to show that

$$
G(\Delta_1) = L(\tilde{\Theta} + \Delta_1) - L(\tilde{\Theta}) \geq 0
$$

for any Δ_1 with $\|\Delta_1\|_\infty \leq \epsilon$. Re-organize $G(\Delta_1)$ as follows:

$$
\begin{aligned}
G(\Delta_1) = \; & \frac{n}{2}\Big(tr\left(\Delta_1\left(S-\tilde{\Theta}^{-1}\right)\right) - \left(\log|\tilde{\Theta}+\Delta_1| - \log|\tilde{\Theta}|\right) + tr\left(\Delta_1\tilde{\Theta}^{-1}\right)\Big) \\
& - \sum_{i<j}\log\left(\frac{\eta}{2v_1}e^{-\frac{|\tilde{\theta}_{ij}+\Delta_{1ij}|}{v_1}} + \frac{1-\eta}{2v_0}e^{-\frac{|\tilde{\theta}_{ij}+\Delta_{1ij}|}{v_0}}\right) \\
& + \sum_{i<j}\log\left(\frac{\eta}{2v_1}e^{-\frac{|\tilde{\theta}_{ij}|}{v_1}} + \frac{1-\eta}{2v_0}e^{-\frac{|\tilde{\theta}_{ij}|}{v_0}}\right) + \tau\sum_i\left(\tilde{\theta}_{ii}+\Delta_1-\tilde{\theta}_{ii}\right) \\
& = \text{(I)} + \text{(II)} + \text{(III)},
\end{aligned}
$$

where

$$
\begin{aligned}
\text{(I)} &= \frac{n}{2}\Big(tr\left(\Delta_1\left(S-\tilde{\Theta}^{-1}\right)\right) - \left(\log|\tilde{\Theta}+\Delta_1| - \log|\tilde{\Theta}|\right) + tr\left(\Delta_1\tilde{\Theta}^{-1}\right)\Big), \\
\text{(II)} &= -\frac{1}{2}\sum_{i<j}\log\left(\frac{\eta}{2v_1}e^{-\frac{|\tilde{\theta}_{ij}+\Delta_{1ij}|}{v_1}} + \frac{1-\eta}{2v_0}e^{-\frac{|\tilde{\theta}_{ij}+\Delta_{1ij}|}{v_0}}\right) \\
& \quad +\frac{1}{2}\sum_{i<j}\log\left(\frac{\eta}{2v_1}e^{-\frac{|\tilde{\theta}_{ij}|}{v_1}} + \frac{1-\eta}{2v_0}e^{-\frac{|\tilde{\theta}_{ij}|}{v_0}}\right), \\
\text{(III)} &= \tau\sum_i\left(\tilde{\theta}_{ii}+\Delta_{1ii}-\tilde{\theta}_{ii}\right) = \tau\Delta_{1ii}.
\end{aligned}
$$

Bound (I) as follows.

$$
\begin{aligned}
& \log|\tilde{\Theta}+\Delta_1| - \log|\tilde{\Theta}| \\
= \; & \mathrm{tr}\left(\Delta_1\tilde{\Theta}^{-1}\right) - \mathrm{vec}\left(\Delta_1\right)^T\int_0^1(1-v)\Big((\tilde{\Theta}^{-1}+v\Delta_1)^{-1}\otimes(\tilde{\Theta}^{-1}+v\Delta_1)^{-1}dv\Big)\mathrm{vec}(\Delta_1) \\
\leq \; & \mathrm{tr}\left(\Delta_1\tilde{\Theta}^{-1}\right) - \frac{1}{4}k_1^2\|\Delta_1\|_F^2.
\end{aligned}
$$

where the last inequality can be shown with the same proof for Theorem 1 in Rothman et al. (2008) with $\sqrt{n} \geq 4(C_1+C_3)dM_{\Gamma^0}/k_1^2\sqrt{\log p}$. Thus,

$$
\begin{aligned}
\text{(I)} \; &\geq \; \frac{n}{2}\Big(tr\left(\Delta_1\left(S-\tilde{\Theta}^{-1}\right)\right) + \frac{1}{4}k_1^2\|\Delta_1\|_F^2\Big) \\
&= \; \frac{n}{2}\Big(\sum_{i,j}\left(\Delta_{1ij}\left(s_{ij}-\tilde{\Theta}_{ij}^{-1}\right)\right) + \frac{1}{4}k_1^2\|\Delta_1\|_F^2\Big).
\end{aligned}
$$

Next consider (II). For any $(i,j) \notin \mathcal{B}$, $\tilde{\theta}_{ij} = 0$, $|\tilde{\theta}_{ij} + \Delta_{1ij}| = |\Delta_{1ij}|$, and therefore

$$
\begin{aligned}
& -\log\left(\frac{\eta}{2v_1} e^{-\frac{|\tilde{\theta}_{ij}+\Delta_{1ij}|}{v_1}} + \frac{1-\eta}{2v_0} e^{-\frac{|\tilde{\theta}_{ij}+\Delta_{1ij}|}{v_0}}\right) + \log\left(\frac{\eta}{2v_1} e^{-\frac{|\tilde{\theta}_{ij}|}{v_1}} + \frac{1-\eta}{2v_0} e^{-\frac{|\tilde{\theta}_{ij}|}{v_0}}\right) \\
= \; & \log \frac{\left(\frac{\eta}{2v_1} e^{-\frac{|0|}{v_1}}\right) + \left(\frac{1-\eta}{2v_0} e^{-\frac{|0|}{v_0}}\right)}{\left(\frac{\eta}{2v_1} e^{-\frac{|\Delta_{1ij}|}{v_1}}\right) + \left(\frac{1-\eta}{2v_0} e^{-\frac{|\Delta_{1ij}|}{v_0}}\right)} \\
= \; & \frac{|\Delta_{1ij}|}{v_0} - \log\Big(\frac{v_0 \eta e^{\frac{|\Delta_{1ij}|}{v_0} - \frac{|\Delta_{1ij}|}{v_1}} + v_1(1-\eta)}{v_0\eta + v_1(1-\eta)}\Big).
\end{aligned}
$$

For any $(i,j) \in \mathcal{B}$ and $i \neq j$, applying Taylor expansion, for some $v \in (0,1)$, we have

$$
\begin{aligned}
& -\log\left(\frac{\eta}{2v_1} e^{-\frac{|\tilde{\theta}_{ij}+\Delta_{1ij}|}{v_1}} + \frac{1-\eta}{2v_0} e^{-\frac{|\tilde{\theta}_{ij}+\Delta_{1ij}|}{v_0}}\right) + \log\left(\frac{\eta}{2v_1} e^{-\frac{|\tilde{\theta}_{ij}|}{v_1}} + \frac{1-\eta}{2v_0} e^{-\frac{|\tilde{\theta}_{ij}|}{v_0}}\right) \\
& = \text{pen}_{SS}{}'(\tilde{\theta}_{ij})\Delta_{1ij} + \frac{1}{2}\text{pen}_{SS}{}''\left(\tilde{\theta}_{ij} + v\Delta_{1ij}\right)\Delta_{1ij}^2.
\end{aligned}
$$

Combining the results above, we have

$$
\begin{aligned}
G(\Delta_1) \geq & \frac{n}{2}\Big(\sum_{i,j}\left(\Delta_{1ij}(s_{ij} - \tilde{\Theta}_{ij}^{-1})\right) + \frac{1}{4}k_1^2\|\Delta_1\|_F^2\Big) + \sum_{i=j,\in\mathcal{B}} \tau\Delta_{1ii} \\
& + \frac{1}{2}\sum_{i\neq j,\in\mathcal{B}}\left(\text{pen}_{SS}'(\tilde{\theta}_{ij})\Delta_{1ij} + \frac{1}{2}\text{pen}_{SS}''(\tilde{\theta}_{ij} + v\Delta_{1ij})\Delta_{1ij}^2\right) \\
& - \frac{1}{2}\sum_{\notin\mathcal{B}}\left(-\frac{|\Delta_{1ij}|}{v_0} + \log\Big(\frac{v_0\eta e^{\frac{|\Delta_{1ij}|}{v_0} - \frac{|\Delta_{1ij}|}{v_1}} + v_1(1-\eta)}{v_0\eta + v_1(1-\eta)}\Big)\right) \\
& = \text{(A)} + \text{(B)} + \text{(C)},
\end{aligned}
$$

where

$$
\begin{aligned}
\text{(A)} = & \frac{n}{2}\Big(\sum_{(i,j)\in\mathcal{B}} \Delta_{1ij}(s_{ij} - \tilde{\Theta}_{ij}^{-1} + \frac{2}{n}Z_{ij})\Big), \\
\text{(B)} = & \frac{n}{2}\left(\sum_{(i,j)\notin\mathcal{B}}\left(\Delta_{1ij}(s_{ij} - \tilde{\Theta}_{ij}^{-1}) - \frac{1}{n}\Big(-\frac{|\Delta_{1ij}|}{v_0} + \log\frac{v_0\eta e^{\frac{|\Delta_{1ij}|}{v_0} - \frac{|\Delta_{1ij}|}{v_1}} + v_1(1-\eta)}{v_0\eta + v_1(1-\eta)}\Big)\right)\right), \\
\text{(C)} = & \frac{n}{8}k_1^2\|\Delta_1\|_F^2 + \sum_{i\neq j,\in\mathcal{B}} \frac{1}{4}\text{pen}_{SS}{}''\left(\tilde{\theta}_{ij} + v\Delta_{1ij}\right)\Delta_{1ij}^2.
\end{aligned}
$$

Next, we show that all three terms, (A), (B), and (C), are non-negative.

- (A) = 0 because of the way $\tilde{\Theta}$ is constructed.

- (C) ≥ 0 by the property of $\text{pen}_{SS}''(\delta)$ stated before.

- For term (B), we will first bound $s_{ij} - \tilde{\Theta}^{-1}_{ij}$:

$$
\begin{aligned}
|s_{ij} - \tilde{\Theta}^{-1}_{ij}| &\leq |s_{ij} - \sigma^0_{ij}| + |\tilde{\Theta}^{-1}_{ij} - \sigma^0_{ij}| \\
&\leq C_1\sqrt{\frac{\log p}{n}} + M^2_{\Sigma^0} 2\left(C_1 + C_3\right) M_{\Gamma^0}\sqrt{\frac{\log p}{n}} + \frac{3}{2} d M^3_{\Sigma^0}\left(2(C_1 + C_3) M_{\Gamma^0}\sqrt{\frac{\log p}{n}}\right)^2 \\
&\leq \left(C_1 + M^2_{\Sigma^0} 2\left(C_1 + C_3\right) M_{\Gamma^0} + 6\left(C_1 + C_3\right)^2 d M^2_{\Gamma^0} M^3_{\Sigma^0}/M\right)\sqrt{\frac{\log p}{n}},
\end{aligned}
$$

  where the second line is due to Lemma 2.

  Next, we bound the fraction after the log function in (B). For simplicity, denote it by $f(\Delta_{1ij})$. Since $1/v_0 - 1/v_1 > 0$, $f(\Delta_{1ij})$ is a monotone function of Δ_{1ij} and $f(\Delta_{1ij})$ goes to 1 as Δ_{1ij} goes to 0. That is, $f(\Delta_{1ij})$ can be arbitrary close to 0, when Δ_{1ij} is sufficiently small. Therefore the second term after summation can be arbitrary close to $\Delta_{1ij}/(nv_0)$.

  So if choosing $1/(nv_0) > C_1 + M^2_{\Sigma^0} 2(C_1 + C_3) M_{\Gamma^0} + 6(C_1 + C_3)^2 d M^2_{\Gamma^0} M^3_{\Sigma^0}/M$ and $\epsilon > 0$ sufficiently small, we have (B)>0 when $\|\Delta_1\|_\infty \leq \epsilon$.

Combining the results above, we have shown that there always exists a small $\epsilon > 0$, such that $G(\Delta_1) \geq 0$ for any $\|\Delta_1\|_\infty \leq \epsilon$. That is, $\tilde{\Theta}$ is a local minimizer. So we have proved Theorem A. □

Proof of Theorem 2. Cai et al. (2011) have shown that the sample noise $\tilde{W}$ can be bounded by $\sqrt{\frac{\log p}{n}}$ times a constant with high probability for both exponential tail and polynomial tail (see the proofs of their Theorem 1 and 4). That is,

- When condition (C1) holds,

$$
\|\tilde{W}\|_\infty \leq \eta_1^{-1}(2 + \tau_0 + \eta_1^{-1} K^2)\sqrt{\frac{\log p}{n}}
$$

  with probability greater than $1 - 2p^{-\tau_0}$.

- When condition (C2) holds,

$$\|\tilde{W}\|_\infty \leq \sqrt{(\theta^0_{max}+1)(4+\tau_0)\frac{\log p}{n}}, \quad \theta^0_{max} = \max_{ij} \theta_{ij},$$

  with probability greater than $1 - O(n^{-\delta_0/8} + p^{-\tau_0/2})$.

With the results above on $\|\tilde{W}\|_\infty$ and Theorem A, we have proven Theorem 2. $\square$

Appendix B: Other Proofs

Proof of Lemma 1. Show both $|\Delta_{\mathcal{B}}\|_\infty$ and $\|\Delta_{\mathcal{B}^c}\|_\infty$ are bounded by r. Thus, $\|\Delta\|_\infty \leq r$.

1. By construction,

$$\|\Delta_{\mathcal{B}^c}\|_\infty \leq 2(C_1 + C_3) M_{\Gamma^0} \sqrt{\log p/n} \leq r.$$

2. The proof for $\|\Delta_{\mathcal{B}}\|_\infty \leq r$ is inspired by Ravikumar et al. (2011). Define $G(\Theta_{\mathcal{B}}) = n\left(-\Theta_{\mathcal{B}}^{-1} + S_{\mathcal{B}}\right)/2 + Z_{\mathcal{B}}$. By definition, the set of $\Theta_{\mathcal{B}}$ that satisfies $G(\Theta_{\mathcal{B}}) = 0$ is the set $\mathcal{A}$. Consider a mapping F from $\mathbb{R}^{|\mathcal{B}|} \to \mathbb{R}^{|\mathcal{B}|}$:

$$F\left(\text{vec}\left(\Delta_{\mathcal{B}}\right)\right) = \frac{2}{n}\Big(- {\Gamma^0}^{-1}_{\mathcal{B}\mathcal{B}} \text{vec}\left(G\left(\Theta^0_{\mathcal{B}} + \Delta_{\mathcal{B}}\right)\right)\Big) + \text{vec}\left(\Delta_{\mathcal{B}}\right). \tag{9}$$

   By construction, $F\left(\text{vec}\left(\Delta_{\mathcal{B}}\right)\right) = \text{vec}\left(\Delta_{\mathcal{B}}\right)$ if and only if $G\left(\Theta^0_{\mathcal{B}} + \Delta_{\mathcal{B}}\right) = G\left(\Theta_{\mathcal{B}}\right) = 0$.

   Let $\mathbb{B}\left(r\right)$ denote the ℓ_∞ ball in $\mathbb{R}^{|\mathcal{B}|}$. If we could show that $F\left(\mathbb{B}\left(r\right)\right) \subseteq \mathbb{B}\left(r\right)$, then because F is continuous and $\mathbb{B}\left(r\right)$ is convex and compact, by Brouwer's fixed point theorem, there exists a fixed point $\text{vec}(\Delta_{\mathcal{B}}) \in \mathbb{B}(r)$. Thus $\|\Delta_{\mathcal{B}}\|_\infty \leq r$.

   Let $\Delta \in \mathbb{R}^{p\times p}$ denote the zero-padded matrix, equal to $\Delta_{\mathcal{B}}$ on $\mathcal{B}$ and zero on $\mathcal{B}^c$.

$$\begin{aligned}
F\left(\text{vec}\left(\Delta_{\mathcal{B}}\right)\right) &= \frac{2}{n}\Big(- {\Gamma^0}^{-1}_{\mathcal{B}\mathcal{B}} \text{vec}\left(G\left(\Theta^0_{\mathcal{B}} + \Delta_{\mathcal{B}}\right)\right)\Big) + \text{vec}\left(\Delta_{\mathcal{B}}\right)\\
&= -{\Gamma^0}^{-1}_{\mathcal{B}\mathcal{B}}\Big(\left(-(\Theta^0 + \Delta)^{-1}_{\mathcal{B}} + S_{\mathcal{B}}\right) + \frac{2}{n} Z_{\mathcal{B}}\Big) + \text{vec}\left(\Delta_{\mathcal{B}}\right)\\
&= -{\Gamma^0}^{-1}_{\mathcal{B}\mathcal{B}}\Big(- \left(\Theta^0 + \Delta\right)^{-1}_{\mathcal{B}} + {\Theta^0_{\mathcal{B}}}^{-1} - {\Theta^0_{\mathcal{B}}}^{-1} + S_{\mathcal{B}}) + \frac{2}{n} Z_{\mathcal{B}}\Big) + \text{vec}\left(\Delta_{\mathcal{B}}\right)\\
&= {\Gamma^0}^{-1}_{\mathcal{B}\mathcal{B}} \text{vec}\left({\Theta^0}^{-1} \Delta {\Theta^0}^{-1} \Delta J {\Theta^0}^{-1}\right)_{\mathcal{B}} - {\Gamma^0}^{-1}_{\mathcal{B}\mathcal{B}} \left(\text{vec}\left(W_{\mathcal{B}} + \frac{2}{n} Z_{\mathcal{B}}\right)\right).
\end{aligned}$$

Denote

$$\begin{aligned}\mathbf{I} &= {\Gamma^0}^{-1}_{\mathcal{B}\mathcal{B}}\mathrm{vec}\left(\Theta^{0^{-1}}\Delta\Theta^{0^{-1}}\Delta J\Theta^{0^{-1}}\right)_{\mathcal{B}}\\ \mathbf{II} &= {\Gamma^0}^{-1}_{\mathcal{B}\mathcal{B}}\left(\mathrm{vec}\left(W_{\mathcal{B}}+\frac{2}{n}Z_{\mathcal{B}}\right)\right).\end{aligned}$$

Then $F\left(\mathrm{vec}\left(\Delta_{\mathcal{B}}\right)\right) \leq \|\mathbf{I}\|_\infty + \|\mathbf{II}\|_\infty$. So it suffices to show $\|\mathbf{I}\|_\infty + \|\mathbf{II}\|_\infty \leq r$.

For the first relationship, we have

$$\begin{aligned}\|\mathbf{I}\|_\infty &\leq \left|\left|\left|{\Gamma^0}^{-1}_{\mathcal{B}\mathcal{B}}\right|\right|\right|_\infty \|\mathrm{vec}(\Theta^{0^{-1}}\Delta\Theta^{0^{-1}}\Delta J\Theta^{0^{-1}})_{\mathcal{B}}\|_\infty\\ &\leq M_{\Gamma^0}\|R(\Delta)\|_\infty\\ &\leq \frac{3}{2}dM_{\Gamma^0}M^3_{\Sigma^0}\|\Delta\|^2_\infty,\end{aligned}$$

where the last inequality is due to $\|\Delta\|_\infty \leq r \leq 1/(3M_{\Sigma^0}d)$ and Lemma 5 from Ravikumar et al. (2011). Since $r \leq 1/(3dM_{\Gamma^0}M^3_{\Sigma^0})$, we further have $\|\mathbf{I}\|_\infty \leq r/2$.

By assumption, $\min|\theta^0_{\mathcal{B}\cap\mathcal{D}^c}| \geq r+\delta$, thus when $\|\Delta\|_\infty \leq r$, $\min|\theta_{\mathcal{B}\cap\mathcal{D}^c}| \geq \delta$, since $\mathrm{pen}'_{SS}(|\theta|)$ is monotonic decreasing, we have $\|Z_{\mathcal{B}\cap\mathcal{D}^c}\|_\infty \leq \frac{1}{2}\mathrm{pen}'_{SS}(\delta)$. Thus, for the second relationship, we have

$$\begin{aligned}\|\mathbf{II}\|_\infty &\leq {\Gamma^0}^{-1}_{\mathcal{B}\mathcal{B}}\left(\|W\|_\infty + \frac{2}{n}\max\left(\frac{1}{2}\mathrm{pen}'_{SS}(\delta),\tau\right)\right)\\ &\leq M_{\Gamma^0}\left(\|W\|_\infty + \frac{2}{n}\max\left(\frac{1}{2}\mathrm{pen}'_{SS}(\delta),\tau\right)\right) \leq r/2\end{aligned}$$

by assumption.

Thus, there exists a point $\tilde{\Theta}$ such that $\|\tilde{\Theta}-\Theta^0\|_\infty \leq r$.

Because $\|\tilde{\Theta}\|_2 \leq \|\tilde{\Theta}-\Theta^0\|_2 + \|\Theta^0\|_2$ and $\|\tilde{\Theta}-\Theta^0\|_2 \leq \left|\left|\left|\tilde{\Theta}-\Theta^0\right|\right|\right|_\infty \leq dr$, we have $\|\tilde{\Theta}\|_2 \leq 1/k_1 + dr < B$. Because $dr < \frac{1}{3M_{\Sigma 0}} < \frac{1}{3}\lambda_{\min}(\Theta^0)$, we have $\lambda_{\min}(\tilde{\Theta}) > 0$.So it is inside $\mathcal{A}$ by assumption. That is, $\mathcal{A}$ is non empty. □

Proof of Lemma 2. Since there are only $p+s$ nonzero entries, we prove (2):

$$\|\tilde{\Theta}-\Theta^0\|_F = \sqrt{\sum_{(i,j)\in S_g}(\tilde{\theta}_{ij}-\theta^0_{ij})^2} \leq r\sqrt{p+s}.$$

Since there are at most d nonzero entries in each column of Θ and Θ is symmetric,

$$\|\tilde{\Theta} - \Theta^0\|_2 \le \left\|\left\|\tilde{\Theta} - \Theta^0\right\|\right\|_\infty \le rd.$$

In addition, since the ℓ_∞/ℓ_∞ operator norm is bounded by Frobenius norm, we prove (3). We skip the proof for (4), which is nearly identical to Corollary 4 in Ravikumar et al. (2008). $\square$

Proof of Theorem 4. (Selection consistency)
Recall

$$\begin{aligned}\log\frac{p_{ij}}{1-p_{ij}} &= \Big(\log\frac{v_0\eta}{v_1(1-\eta)} - \frac{|\tilde{\theta}_{ij}|}{v_1} + \frac{|\tilde{\theta}_{ij}|}{v_0}\Big)\\ &= \Big(-\log\frac{v_1(1-\eta)}{v_0\eta} - \frac{|\tilde{\theta}_{ij}|}{v_1} + \frac{|\tilde{\theta}_{ij}|}{v_0}\Big).\end{aligned} \tag{10}$$

- When $\theta^0_{ij} = 0$, by constructor, $\tilde{\theta}_{ij} = 0$. Then with our choice of $v_1(1-\eta)/\left(v_0\eta\right)$,

$$\log\frac{p_{ij}}{1-p_{ij}} \to -\infty.$$

- When $\theta^0_{ij} \ne 0$, we have

$$\begin{aligned}\log\frac{p_{ij}}{1-p_{ij}} &= \Big(\log\frac{v_0\eta}{v_1(1-\eta)} - \frac{|\theta_{ij}|}{v_1} + \frac{|\theta_{ij}|}{v_0}\Big)\\ &\ge \left(-\log\frac{v_1(1-\eta)}{v_0\eta} + \left(\frac{1}{v_0} - \frac{1}{v_1}\right)\left(|\theta^0_{ij}| - |\theta^0_{ij} - \theta_{ij}|\right)\right)\\ &\ge -\log\frac{v_1(1-\eta)}{v_0\eta} + (C_4 - C_3)\left(K_0 - 2(C_1 + C_3)M_{\Gamma^0}\right)\log p.\end{aligned} \tag{11}$$

  Then with our choice of $v_1(1-\eta)/(v_0\eta)$,

$$\log\frac{p_{ij}}{1-p_{ij}} \to +\infty.$$

$\square$

Proof of Theorem 5. The estimate of the precision matrix is symmetric due to construction.

Next we show that the estimate is ensured to be positive definite. Assume $\Theta^{(t)}$, the t-th update of the estimate is positive definite. Apparently, this assumption is satisfied with $t = 0$ since the initial estimate $\Theta^{(0)}$ is positive definite.

Then it suffices to show that $\det(\Theta^{(t+1)}) \succ 0$. WLOG, assume we update the last column of Θ in the $(t+1)$-th iteration. Using Schur complements, we have

$$\det\left(\Theta^{(t+1)}\right) = \det\left(\Theta_{11}^{(t)}\right)\left(\theta_{22}^{(t+1)} - \theta_{12}^{(t+1)^T}\Theta_{11}^{(t)^{-1}}\theta_{12}^{(t+1)}\right).$$

Because $\det(\Theta^{(t)}) \succ 0$, we have $\det\left(\Theta_{11}^{(t)}\right) > 0$. Further, the updating rule of our algorithm ensures that

$$\left(\theta_{22}^{(t+1)} - \theta_{12}^{(t+1)^T}\Theta_{11}^{(t)^{-1}}\theta_{12}^{(t+1)}\right) = \frac{1}{w_{22}^{(t+1)}} > 0.$$

Thus, $\det\left(\Theta^{(t+1)}\right) > 0$. □

Appendix C: Checking $\|\Theta\|_2 \leq B$.

Algorithm 1 involves checking the spectral norm constraint $\|\Theta\|_2 \leq B$ after every column update of Θ. Computing $\|\Theta\|_2$ can be computationally intensive, however, since we only change one column (and corresponding one row) at a time, the constraint can be checked without calculating $\|\Theta\|_2$ every time. Suppose we know $\|\Theta^{(t)}\|_2$ (or an upper bound) at the previous step, and denote $\Delta^{(t)} := \Theta^{(t+1)} - \Theta^{(t)}$ to be the difference between the estimates after one column update. In order to check the bound, it is sufficient to make sure that $\|\Theta^{(t)}\|_2 + \|\Delta^{(t)}\|_2 < B$. It is easy to check this constraint because $\|\Delta^{(t)}\|_2$ is a rank two matrix with its maximum eigenvalue available in closed form. Only when $\|\Theta^{(t)}\|_2 + \|\Delta\|_2$ exceeds B, we will need to recalculate $\|\Theta^{(t+1)}\|_2$ again.

References

Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2008). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *arXiv preprint arXiv:0811.3628*.

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980.

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