
Supplementary Material for Bayesian Joint Estimation of Multiple Graphical Models

1 Proofs of the Main Results

In this section, we prove main results in the main paper. For simplicity, we assume sample sizes of the K classes are the same: $n_1 = \dots = n_K = n$.

Recall our objective function is

$$L(\Theta) = \sum_{k=1}^K -\ell(\Theta_k) + \alpha \sum_{i=1}^p \sum_{k=1}^K \text{pen}_1(\theta_{k,ii}) + \frac{\alpha}{2} \sum_{i \neq j} \text{pen}_2(\theta_{ij}), \quad (1.1)$$

where

$$\begin{aligned} \ell(\Theta_k) &= -\frac{n}{2} (\text{tr}(S_k \Theta_k) - \log \det(\Theta_k)), \\ \text{pen}_1(\theta_{k,ii}) &= \tau \theta_{k,ii}, \\ \text{pen}_2(\theta_{ij}) &= -\log \left(\frac{p_1}{(2v_1)^K} e^{-\|\theta_{ij}\|_1/v_1} + \frac{1-p_1}{(2v_0)^K} e^{-\|\theta_{ij}\|_1/v_0} \right) + C, \end{aligned}$$

for a constant C that makes $\text{pen}_2(\mathbf{0}) = 0$.

1.1 Proof of Theorem 1

In this subsection, we show $L(\Theta)$ is strictly convex, under the constraints that $\Theta_k > 0$, $\|\Theta_k\|_2 \leq B$, $k = 1, \dots, K$.

Decompose $L(\Theta)$ into the following two parts $L_1(\Theta) + L_2(\Theta)$:

$$\begin{aligned} L_1(\Theta) &= \sum_k L_1(\Theta_k) = \sum_k \left[-\ell(\Theta_k) - \frac{\alpha K}{8v_0^2} \|\Theta_k\|_F^2 \right], \\ L_2(\Theta) &= \alpha \sum_i \sum_k \text{pen}_1(\theta_{k,ii}) + \frac{\alpha}{2} \sum_{i \neq j} \text{pen}_2(\theta_{ij}) + \frac{\alpha K}{8v_0^2} \sum_k \|\Theta_k\|_F^2. \end{aligned}$$

We show $L_1(\Theta)$ and $L_2(\Theta)$ are both strictly convex, so that $L(\Theta)$ is strictly convex.

For $L_1(\Theta) = \sum_k L_1(\Theta_k)$, the Hessian matrix of $L_1(\Theta_k)$ is as follows:

$$\nabla^2 L_1(\Theta_k) = \frac{n}{2} (\Theta_k \otimes \Theta_k)^{-1} - \frac{\alpha K}{4v_0^2} I_{p^2 \times p^2}.$$

The minimum eigenvalue of the Hessian matrix above can be bounded by

$$\lambda_{\min}(\nabla^2 L_1(\Theta_k)) = \frac{n}{2} \lambda_{\max}^{-1}(\Theta_k \otimes \Theta_k) - \frac{\alpha K}{4v_0^2}$$

$$\begin{aligned}
&= \frac{n}{2} \lambda_{\max}^{-2}(\Theta_k) - \frac{\alpha K}{4v_0^2} \\
&\geq \frac{n}{2} \|\Theta_k\|_2^{-2} - \frac{\alpha K}{4v_0^2} \\
&\geq \frac{n}{2} B^{-2} - \frac{\alpha K}{4v_0^2} > 0.
\end{aligned}$$

Therefore, $L_1(\Theta_k)$, $k = 1, \dots, K$ and $L_1(\Theta) = \sum_k L_1(\Theta_k)$ are all strictly convex.

For $L_2(\Theta)$, let $\text{pen}(\Theta) = \sum_i \sum_k \text{pen}_1(\theta_{k,ii}) + \frac{1}{2} \sum_{i \neq j} \text{pen}_2(\theta_{ij})$. We now quantify the magnitude of its second-order subgradients. For any $k, k' \in \{1, \dots, K\}$ and $i \neq j$, we have

$$\begin{cases} \left| \nabla_{\theta_{k,ij}, \theta_{k',ij}}^2 \text{pen}(\Theta) \right| = \left(\frac{1}{v_0} - \frac{1}{v_1} \right)^2 \frac{\frac{p_1}{1-p_1} \left(\frac{v_0}{v_1} \right)^K e^{\left(\frac{1}{v_0} - \frac{1}{v_1} \right) \|\theta_{ij}\|_1}}{\left[1 + \frac{p_1}{1-p_1} \left(\frac{v_0}{v_1} \right)^K e^{\left(\frac{1}{v_0} - \frac{1}{v_1} \right) \|\theta_{ij}\|_1} \right]^2} \leq \frac{1}{4v_0^2} \\ \nabla_{\theta_{k,ii}, \theta_{k',ii}}^2 \text{pen}(\Theta) = 0. \end{cases}$$

Therefore, for any (i, j) pair, we have

$$\lambda_{\min}(\nabla_{\theta_{ij}, \theta_{ij}}^2 \text{pen}(\Theta)) > -\frac{K}{4v_0^2}.$$

Hence, $\nabla_{\theta_{ij}, \theta_{ij}}^2 L_2(\Theta)$ is positive definite for any (i, j) pair considering

$$\begin{aligned}
\lambda_{\min}(\nabla_{\theta_{ij}, \theta_{ij}}^2 L_2(\Theta)) &= \alpha \lambda_{\min}(\nabla_{\theta_{ij}, \theta_{ij}}^2 \text{pen}(\Theta)) + \frac{\alpha K}{4v_0^2} \\
&> -\frac{\alpha K}{4v_0^2} + \frac{\alpha K}{4v_0^2} = 0,
\end{aligned}$$

Since $\nabla_{\theta_{ij}, \theta_{i'j'}}^2 L_2(\Theta) = \mathbf{0}$ for any different pairs $(i, j) \neq (i', j')$, we have $\nabla^2 L_2(\Theta)$ is positive definite. Therefore, $L_2(\Theta)$ is strictly convex.

1.2 Proof of Theorem 2

In this subsection, we provide the proof of Theorem 2 of estimation consistency in the main paper.

We first introduce some notations that we use in the proof. Let Σ_k^0 denote the k 'th true covariance matrix and $W_k = S_k - \Sigma_k^0$ denote the difference between the k 'th sample covariance matrix and Σ_k^0 . Use $\Sigma^0 = (\Sigma_1^0, \dots, \Sigma_K^0)$ and $\mathbf{W} = (W_1, \dots, W_K)$ to denote the collection of Σ_k^0 's and W_k 's, respectively. For any subset \mathcal{M} of $\{(i, j) : 1 \leq i, j \leq p\}$ and a $p \times p$ matrix Θ_k , let $(\Theta_k)_{\mathcal{M}}$ denote the submatrix of Θ_k with entries indexed by \mathcal{M} ; for $\Theta = (\Theta_1, \dots, \Theta_K)$, let $\Theta_{\mathcal{M}} = ((\Theta_1)_{\mathcal{M}}, \dots, (\Theta_K)_{\mathcal{M}})$. We use $\Gamma_{\mathcal{M}\mathcal{M}}^0$ to denote the Hessian matrix $\nabla_{\Theta_{\mathcal{M}}, \Theta_{\mathcal{M}}}^2 (-\sum_k \log \det \Theta_k)$ evaluated at Θ^0 , and $(\Gamma_k^0)_{\mathcal{M}\mathcal{M}}$ to denote the Hessian matrix $\nabla_{(\Theta_k)_{\mathcal{M}}, (\Theta_k)_{\mathcal{M}}}^2 (-\log \det \Theta_k)$ evaluated at Θ_k^0 .

Our proof is motivated by the proof techniques from [6] and [2] and we prove Theorem 2 through three steps:

- Step 1. Construct a solution set \mathcal{A} for the following constrained minimization problem:

$$\min_{\Theta \in \Omega_1} L(\Theta),$$

where $\Omega_1 := \Omega \cap \{\Theta : \Theta_{\mathcal{B}^c} = 0\}$ with $\mathcal{B} = \{(i, j) : i = j \text{ or } |\theta_{k,ij}^0| \geq C_5 \sqrt{\log p/n} \text{ for some } k\}$ and $\Omega = \{\Theta : \Theta_k > 0, \|\Theta_k\|_2 \leq B, k = 1, \dots, K\}$ is the original constraint space.

- Step 2. Show that there exists $\tilde{\Theta} \in \mathcal{A}$ such that $\|\tilde{\Theta} - \Theta^0\|_{\infty} < r_e = C_5 \sqrt{\frac{\log p}{n}}$.
- Step 3. Show that the constructed $\tilde{\Theta}$ in Step 2 is a local minimizer of $L(\Theta)$. Then, from the convexity of the constrained minimization problem, $\tilde{\Theta}$ must be the global minimizer.

Let $M_{\Gamma^0} = \|(\Gamma_{\mathcal{B}\mathcal{B}}^0)^{-1}\|_\infty = \sup_k \|(\Gamma_k^0)_{\mathcal{B}\mathcal{B}}^{-1}\|_\infty$ and $M_{\Sigma^0} = \sup_k \|\Sigma_k^0\|_\infty$ where $\|\cdot\|_\infty$ denotes the matrix maximum absolute row sum norm. We first present a more general theorem below. Theorem 2 in the main paper directly follows from the following theorem by checking its condition (v) with tail conditions and standard concentration inequalities (c.f., (the proof of) Theorem 1 of [1]).

Theorem 4 *Under the following conditions:*

(i) *rates of the hyperparameters v_1 , v_0 , p_1 , and τ :*

$$\begin{aligned} \frac{\alpha}{nv_1} &< C_3/(1 + \epsilon_1)\sqrt{\log p/n}, \\ \frac{\alpha}{nv_0} &> C_4\sqrt{\log p/n}, \\ \epsilon_2 &< \frac{v_1^K(1 - p_1)}{v_0^K p_1} < \frac{v_1^{K+2}(1 - p_1)}{v_0^{K+2} p_1} \leq \epsilon_1 p^{\epsilon_0/\alpha}, \\ \frac{\alpha\tau}{n} &\leq \frac{C_3}{2}\sqrt{\log p/n}, \end{aligned}$$

where $\epsilon_0 = (C_2 - C_3)M_{\Gamma^0}(C_4 - C_3)$ with $C_2 > C_3 > 0$, $C_4 > (1 + \epsilon_2)/\epsilon_2[C_1 + 2(C_1 + C_2)M_{\Gamma^0}M_{\Sigma^0}^2 + 6(C_1 + C_2)^2dM_{\Gamma^0}^2M_{\Sigma^0}^3/M]$, and $\epsilon_1 > \epsilon_2 > 0$,

(ii) *the eigenvalues of the true precision matrices:*

$$1/\xi_0 < \lambda_{\min}(\Theta_k^0) < \lambda_{\max}(\Theta_k^0) \leq 1/\xi_1 < \sqrt{\frac{2\alpha p^{\epsilon_0/\alpha}}{KC_3^2 \log p}} \quad \text{for } k = 1, \dots, K,$$

(iii) *the sample size n : $\sqrt{n} \geq M\sqrt{\log p}$ with*

$$M = 2d(C_1 + C_2)M_{\Gamma^0} \max\{2M_{\Gamma^0}(2M_{\Sigma^0}(M_{\Sigma^0}^2 + \frac{3}{2}dM_{\Sigma^0}^3) + (M_{\Sigma^0}^2 + \frac{3}{2}M_{\Sigma^0}^3)^2), 1, 2/\xi_1^2, \xi_0\},$$

(iv) *the bounds on the spectral norms of the estimated precision matrices:*

$$1/\xi_1 + 2d(C_1 + C_2)M_{\Gamma^0}\sqrt{\log p/n} < B < \left(\frac{2nv_0^2}{\alpha K}\right)^{1/2},$$

(v) *difference between the sample covariance matrices and the true covariance matrices: $\|\mathbf{W}\|_\infty \leq C_1\sqrt{\log p/n}$,*

we have

$$\|\hat{\Theta} - \Theta^0\|_\infty < 2(C_1 + C_2)M_{\Gamma^0}\sqrt{\log p/n},$$

where C_2 is chosen such that $2(C_1 + C_2)M_{\Gamma^0} = C_5$.

Proof of Theorem 4

• Step 1. Construct a solution set for minimizing the objective function (1.1) as follows:

$$\mathcal{A} = \{\Theta : \mathbf{0} \in G(\Theta)_{\mathcal{B}}, \Theta_{\mathcal{B}^c} = \mathbf{0}\} \cap \Omega,$$

where $\mathcal{B} = \{(i, j) : i = j \text{ or } |\theta_{k,ij}^0| \geq r_e \text{ for some } k\}$ for $r_e = 2(C_1 + C_2)M_{\Gamma^0}\sqrt{\log p/n}$ and $G(\Theta) = \frac{n}{2}(\mathbf{S} - \Theta^{-1} + \frac{2\alpha}{n}\mathbf{Z}(\Theta))$ is the subgradient of the objective function $L(\Theta)$ with

$$Z_{k,ij}(\Theta) = Z_{k,ij}(\theta_{ij}) = \begin{cases} \tau & \text{if } i = j \\ \frac{1}{2} \frac{\partial}{\partial |\theta_{k,ij}|} \text{pen}_2(\theta_{ij}) \text{sign}(\theta_{k,ij}) & \text{if } i \neq j \end{cases}$$

where

$$\frac{\partial}{\partial |\theta_{k,ij}|} \text{pen}_2(\theta_{ij}) = w(\theta_{ij}) \frac{1}{v_1} + (1 - w(\theta_{ij})) \frac{1}{v_0},$$

and

$$\text{sign}(\theta_{k,ij}) = \begin{cases} 1 & \text{when } \theta_{k,ij} > 0, \\ -1 & \text{when } \theta_{k,ij} < 0, \\ [-1, 1] & \text{when } \theta_{k,ij} = 0. \end{cases}$$

- Step 2. Show that there exists some $\tilde{\Theta} \in \mathcal{A}$ such that $\|\tilde{\Theta} - \Theta^0\|_\infty < r_e$.

We only need to show for the entries of Θ indexed by \mathcal{B} , because $\|(\tilde{\Theta} - \Theta^0)_{\mathcal{B}^c}\|_\infty < r_e$ by the way that \mathcal{A} is constructed.

Define the following mapping from $\mathbb{R}^{K|\mathcal{B}|}$ to $\mathbb{R}^{K|\mathcal{B}|}$:

$$F(\text{vec}(\Delta_{\mathcal{B}})) = -\frac{2}{n}(\Gamma_{\mathcal{B}\mathcal{B}}^0)^{-1}\text{vec}\left(\tilde{G}(\Theta^0 + \Delta)_{\mathcal{B}}\right) + \text{vec}(\Delta_{\mathcal{B}}),$$

where $\tilde{G}(\Theta) = \frac{n}{2}(\mathcal{S} - \Theta^{-1} + \frac{2\alpha}{n}\tilde{\mathcal{Z}})$ with $\tilde{\mathcal{Z}}$ to be a member of the subgradient of the unique minimizer of (1.1) under the constraint set Ω , and Δ satisfies $(\Theta^0 + \Delta)_{\mathcal{B}^c} = \mathbf{0}$, that is, $\Delta_{\mathcal{B}^c} = -\Theta_{\mathcal{B}^c}^0$. We will first show $F(\mathbb{B}(r)) \subseteq \mathbb{B}(r)$ for the ℓ_∞ ball $\mathbb{B}(r)$ in $\mathbb{R}^{K|\mathcal{B}|}$ with $r = 2(C_1 + C_3)M_{\Gamma^0}\sqrt{\log p/n} < r_e$. Then, by Brouwer's fixed point theorem [4], there exists a fixed point $\text{vec}(\tilde{\Delta}_{\mathcal{B}}) \in \mathbb{B}(r)$ such that $F(\text{vec}(\tilde{\Delta}_{\mathcal{B}})) = \text{vec}(\tilde{\Delta}_{\mathcal{B}})$. Equivalently, we have a fixed point $\text{vec}(\tilde{\Delta}_{\mathcal{B}}) \in \mathbb{B}(r)$ such that $\tilde{G}(\Theta^0 + \tilde{\Delta})_{\mathcal{B}} = \mathbf{0}$ or $\mathbf{0} \in G(\Theta^0 + \tilde{\Delta})_{\mathcal{B}}$.

For any $\text{vec}(\Delta_{\mathcal{B}}) \in \mathbb{B}(r)$, we have

$$\begin{aligned} F(\text{vec}(\Delta_{\mathcal{B}})) &= -(\Gamma_{\mathcal{B}\mathcal{B}}^0)^{-1}\text{vec}\left(\left(\mathcal{S} - (\Theta^0 + \Delta)^{-1} + \frac{2\alpha}{n}\tilde{\mathcal{Z}}\right)_{\mathcal{B}}\right) + \text{vec}(\Delta_{\mathcal{B}}) \\ &= -(\Gamma_{\mathcal{B}\mathcal{B}}^0)^{-1}\left(\text{vec}\left((\Sigma^0 - (\Theta^0 + \Delta)^{-1})_{\mathcal{B}}\right) - (\Gamma_{\mathcal{B}\mathcal{B}}^0)\text{vec}(\Delta_{\mathcal{B}})\right) \end{aligned} \quad (1.2)$$

$$- (\Gamma_{\mathcal{B}\mathcal{B}}^0)^{-1}\text{vec}\left(\mathbf{W}_{\mathcal{B}} + \frac{2\alpha}{n}\tilde{\mathcal{Z}}_{\mathcal{B}}\right) \quad (1.3)$$

where $(\Theta^0 + \Delta)^{-1} = ((\Theta_1^0 + \Delta_1)^{-1}, \dots, (\Theta_K^0 + \Delta_K)^{-1})$.

For (1.2), we have

$$\begin{aligned} &\left\| -(\Gamma_{\mathcal{B}\mathcal{B}}^0)^{-1}\left(\text{vec}\left((\Sigma^0 - (\Theta^0 + \Delta)^{-1})_{\mathcal{B}}\right) - (\Gamma_{\mathcal{B}\mathcal{B}}^0)\text{vec}(\Delta_{\mathcal{B}})\right) \right\|_\infty \\ &\leq \sup_k \left\| -(\Gamma_k^0)_{\mathcal{B}\mathcal{B}}^{-1}\left(\text{vec}\left((\Sigma_k^0 - (\Theta_k^0 + \Delta_k)^{-1})_{\mathcal{B}}\right) - (\Gamma_k^0)_{\mathcal{B}\mathcal{B}}\text{vec}((\Delta_k)_{\mathcal{B}})\right) \right\|_\infty \\ &\leq \sup_k \left\| (\Gamma_k^0)_{\mathcal{B}\mathcal{B}}^{-1} \right\|_\infty \left\| \text{vec}\left((\Sigma_k^0 - (\Theta_k^0 + \Delta_k)^{-1})_{\mathcal{B}}\right) - (\Gamma_k^0)_{\mathcal{B}\mathcal{B}}\text{vec}((\Delta_k)_{\mathcal{B}}) \right\|_\infty \\ &\leq M_{\Gamma^0} \sup_k \left\| \int_0^1 \{(\Theta_k^0 + t\Delta_k)^{-1} \otimes (\Theta_k^0 + t\Delta_k)^{-1} - (\Theta_k^0)^{-1} \otimes (\Theta_k^0)^{-1}\} dt \right\|_\infty \left\| \text{vec}((\Delta_k)_{\mathcal{B}}) \right\|_\infty \\ &\leq M_{\Gamma^0} r \sup_k \int_0^1 \left\| (\Theta_k^0 + t\Delta_k)^{-1} \otimes (\Theta_k^0 + t\Delta_k)^{-1} - (\Theta_k^0)^{-1} \otimes (\Theta_k^0)^{-1} \right\|_\infty dt \\ &\leq M_{\Gamma^0} (2M_{\Sigma^0}(M_{\Sigma^0}^2 + \frac{3}{2}dM_{\Sigma^0}^3)dr + (M_{\Sigma^0}^2 + \frac{3}{2}M_{\Sigma^0}^3)^2 d^2 r^2) r \\ &\leq \frac{r}{2}, \end{aligned} \quad (1.4)$$

where the second inequality is because $\|AB\|_\infty \leq \|A\|_\infty \|B\|_\infty$, the third inequality is because $\Gamma_k^0 = (\Theta_k^0)^{-1} \otimes (\Theta_k^0)^{-1}$, the fifth inequality is because $\|(\Theta_k^0 + \Delta_k)^{-1} - (\Theta_k^0)^{-1}\|_\infty \leq (M_{\Sigma}^2 + \frac{3}{2}M_{\Sigma}^3)r$ from Corollary 4 of [5], $\|(\Theta_k^0 + \Delta_k)^{-1} - (\Theta_k^0)^{-1}\|_\infty \leq d\|(\Theta_k^0 + \Delta_k)^{-1} - (\Theta_k^0)^{-1}\|_\infty$ and Lemma 13 of [3], and last inequality is by $M_{\Gamma^0}(2M_{\Sigma^0}(M_{\Sigma^0}^2 + \frac{3}{2}dM_{\Sigma^0}^3) + (M_{\Sigma^0}^2 + \frac{3}{2}M_{\Sigma^0}^3)^2)d^2r < \frac{1}{2}$ and $r < \frac{1}{d}$ from condition (iii).

For (1.3), we have

$$\begin{aligned} \left\| (\Gamma_{\mathcal{B}\mathcal{B}}^0)^{-1}\text{vec}\left(\mathbf{W}_{\mathcal{B}} + \frac{2\alpha}{n}\tilde{\mathcal{Z}}_{\mathcal{B}}\right) \right\|_\infty &\leq \sup_k \left\| ((\Gamma_k^0)_{\mathcal{B}\mathcal{B}})^{-1}\text{vec}\left((\mathbf{W}_k)_{\mathcal{B}} + \frac{2\alpha}{n}(\tilde{\mathcal{Z}}_k)_{\mathcal{B}}\right) \right\|_\infty \\ &\leq \sup_k M_{\Gamma^0} \left(\|\mathbf{W}_k\|_\infty + \left\| \frac{2\alpha}{n}(\tilde{\mathcal{Z}}_k)_{\mathcal{B}} \right\|_\infty \right) \\ &\leq M_{\Gamma^0} \left(C_1\sqrt{\log p/n} + C_3\sqrt{\log p/n} \right) = \frac{r}{2}, \end{aligned} \quad (1.5)$$

where the third inequality is because of the upper bound on the magnitude of the first derivatives from Lemma 5 and $\|\mathbf{W}_k\|_\infty \leq \|\mathbf{W}\|_\infty \leq C_1\sqrt{\log p/n}$ from condition (v).

Thus, combining (1.3), (1.4), and (1.5), we have $\|F(\text{vec}(\Delta_B))\|_\infty \leq r$, that is, $F(\text{vec}(\Delta_B)) \subseteq \mathbb{B}(r)$ for any $\text{vec}(\Delta_B) \in \mathbb{B}(r)$. Therefore, we have $F(\mathbb{B}(r)) \subseteq \mathbb{B}(r)$. By Brouwer's fixed point theorem [4], there exists a fixed point $\text{vec}(\tilde{\Delta}_B) \in \mathbb{B}(r)$ such that $F(\text{vec}(\tilde{\Delta}_B)) = \text{vec}(\tilde{\Delta}_B)$, which is equivalent to $\tilde{G}(\Theta^0 + \tilde{\Delta})_B = \mathbf{0}$ or $\mathbf{0} \in G(\Theta^0 + \tilde{\Delta})_B$. Let $\tilde{\Theta} = \Theta^0 + \tilde{\Delta}$, we have

$$\begin{cases} \tilde{\Theta}_B = \Theta_B^0 + \tilde{\Delta}_B \\ \tilde{\Theta}_{B^c} = \Theta_{B^c}^0 + \tilde{\Delta}_{B^c} = \mathbf{0}. \end{cases}$$

In addition, we have this estimate $\tilde{\Theta}$ satisfies $\mathbf{0} \in G(\tilde{\Theta})_B$, $\tilde{\Theta}_{B^c} = \mathbf{0}$, and $\|(\tilde{\Theta} - \Theta^0)_B\|_\infty \leq r < r_e$. As long as $\tilde{\Theta}$ is in the parameter space of interest Ω , that is, $\|\tilde{\Theta}_k\|_2 < B$ and $\tilde{\Theta}_k > 0$, we have the statement for Step 2 as desired.

Indeed, for any k , by condition (iv), we have

$$\begin{aligned} \|\tilde{\Theta}_k\|_2 &\leq \|\tilde{\Theta}_k - \Theta_k^0\|_2 + \|\Theta_k^0\|_2 \\ &\leq \|\tilde{\Delta}_k\|_\infty + \frac{1}{\xi_1} \\ &\leq dr_e + \frac{1}{\xi_1} < B, \end{aligned}$$

and, by conditions (ii) and (iii), we have

$$\lambda_{\min}(\tilde{\Theta}_k) \geq \lambda_{\min}(\Theta_k^0) - \|\tilde{\Delta}_k\|_2 \geq \lambda_{\min}(\Theta_k^0) - dr_e > 0.$$

Therefore, we have $\tilde{\Theta} \in \mathcal{A}$ and $\|\tilde{\Theta} - \Theta^0\|_\infty < r_e$.

- Step 3. Show that the constructed $\tilde{\Theta}$ in Step 2 is a local minimizer of $L(\Theta)$.

It suffices to show that there exists some $\epsilon > 0$ such that $H(\Delta) = L(\tilde{\Theta} + \Delta) - L(\tilde{\Theta}) \geq 0$ for any Δ with $\|\Delta\|_\infty < \epsilon$. We have

$$\begin{aligned} H(\Delta) &= \sum_{k=1}^K \frac{n}{2} \left\{ \text{tr}(\Delta_k(S_k - \tilde{\Theta}_k^{-1})) + \text{tr}(\Delta_k \tilde{\Theta}_k^{-1}) - \left[\log \det(\tilde{\Theta}_k + \Delta_k) - \log \det(\tilde{\Theta}_k) \right] \right\} \\ &\quad + \alpha \sum_{k=1}^K \sum_{i=1}^p \tau \Delta_{k,ii} + \frac{\alpha}{2} \sum_{i \neq j} \left[\text{pen}_2(\tilde{\theta}_{ij} + \Delta_{ij}) - \text{pen}_2(\tilde{\theta}_{ij}) \right]. \end{aligned}$$

Under the condition $\sqrt{n} \geq 4(C_1 + C_2)M_{\Gamma^0} d \sqrt{\log p} / \xi_1^2$ and with the same proof for Theorem 1 in [7], we have the following upper bound on $\log \det(\tilde{\Theta}_k + \Delta_k) - \log \det(\tilde{\Theta}_k)$:

$$\log \det(\tilde{\Theta}_k + \Delta_k) - \log \det(\tilde{\Theta}_k) \leq \text{tr}(\Delta_k \tilde{\Theta}_k^{-1}) - \frac{1}{4} \xi_1^2 \|\Delta_k\|_F^2, \quad k = 1, \dots, K. \quad (1.6)$$

For any (i, j) pair in B^c with $i \neq j$, we have $\tilde{\theta}_{ij} = \mathbf{0}$ and therefore, for some u_1 in $(0, 1)$,

$$\text{pen}_2(\tilde{\theta}_{ij} + \Delta_{ij}) - \text{pen}_2(\tilde{\theta}_{ij}) = \nabla_{\theta_{ij}} \text{pen}_2(u_1 \Delta_{ij}) \Delta_{ij}. \quad (1.7)$$

For any (i, j) pair in B with $i \neq j$, for some u_2 in $(0, 1)$, we have

$$\begin{aligned} \text{pen}_2(\tilde{\theta}_{ij} + \Delta_{ij}) - \text{pen}_2(\tilde{\theta}_{ij}) &= \sum_{k: \tilde{\theta}_{k,ij}=0} \frac{\partial}{\partial |\theta_{k,ij}|} \text{pen}_2(\tilde{\theta}_{ij}) |\Delta_{k,ij}| + \sum_{k: \tilde{\theta}_{k,ij} \neq 0} \frac{\partial}{\partial \theta_{k,ij}} \text{pen}_2(\tilde{\theta}_{ij}) \Delta_{k,ij} \\ &\quad + \frac{1}{2} \Delta_{ij}^T \nabla_{\theta_{ij}, \theta_{ij}}^2 \text{pen}_2(\tilde{\theta}_{ij} + u_2 \Delta_{ij}) \Delta_{ij}. \end{aligned} \quad (1.8)$$

Combining (1.7) and (1.8), we have the following lower bound for $H(\Delta)$:

$$\begin{aligned} H(\Delta) &\geq \sum_{k=1}^K \frac{n}{2} \left(\sum_{i,j} \Delta_{k,ij} (s_{k,ij} - \tilde{\Theta}_{k,ij}^{-1}) + \frac{1}{4} \xi_1^2 \|\Delta_k\|_F^2 \right) + \alpha \sum_{k=1}^K \sum_{i=1}^p \tau \Delta_{k,ii} \\ &\quad + \frac{\alpha}{2} \sum_{i \neq j, (i,j) \in B^c} \sum_{k=1}^K \frac{\partial}{\partial |\theta_{k,ij}|} \text{pen}_2(u_1 \Delta_{ij}) |\Delta_{k,ij}| \end{aligned}$$

$$\begin{aligned}
& + \frac{\alpha}{2} \sum_{i \neq j, (i,j) \in \mathcal{B}} \left[\sum_{k: \tilde{\theta}_{k,ij}=0} \frac{\partial}{\partial |\theta_{k,ij}|} \text{pen}_2(\tilde{\theta}_{ij}) |\Delta_{k,ij}| + \sum_{k: \tilde{\theta}_{k,ij} \neq 0} \frac{\partial}{\partial \theta_{k,ij}} \text{pen}_2(\tilde{\theta}_{ij}) \Delta_{k,ij} \right. \\
& \quad \left. + \frac{1}{2} \Delta_{ij}^T \nabla_{\tilde{\theta}_{ij}, \theta_{ij}}^2 \text{pen}_2(\tilde{\theta}_{ij} + u_2 \Delta_{ij}) \Delta_{ij} \right]
\end{aligned}$$

$$= (\text{I}) + (\text{II}) + (\text{III}),$$

where

$$\begin{aligned}
(\text{I}) &= \frac{n}{2} \sum_{k=1}^K \sum_{\tilde{\theta}_{k,ij} \neq 0, (i,j) \in \mathcal{B}} \Delta_{k,ij} \left(s_{k,ij} - \tilde{\Theta}_{k,ij}^{-1} + \frac{2\alpha}{n} Z_{k,ij}(\tilde{\theta}_{ij}) \right) \\
& \quad + \frac{n}{2} \sum_{k=1}^K \sum_{\tilde{\theta}_{k,ij}=0, (i,j) \in \mathcal{B}} \left[\Delta_{k,ij} \left(s_{k,ij} - \tilde{\Theta}_{k,ij}^{-1} \right) + |\Delta_{k,ij}| \frac{\alpha}{n} \frac{\partial}{\partial |\theta_{k,ij}|} \text{pen}_2(\tilde{\theta}_{ij}) \right], \\
(\text{II}) &= \sum_{k=1}^K \frac{n}{8} \xi_1^2 \|\Delta_k\|_F^2 + \sum_{i \neq j, (i,j) \in \mathcal{B}} \frac{\alpha}{4} \Delta_{ij}^T \nabla_{\tilde{\theta}_{ij}, \theta_{ij}}^2 \text{pen}_2(\tilde{\theta}_{ij} + u_2 \Delta_{ij}) \Delta_{ij}, \\
(\text{III}) &= \frac{n}{2} \sum_{i \neq j, (i,j) \in \mathcal{B}^c} \sum_{k=1}^K \left[\Delta_{k,ij} (s_{k,ij} - \tilde{\Theta}_{k,ij}^{-1}) + |\Delta_{k,ij}| \frac{\alpha}{n} \frac{\partial}{\partial |\theta_{k,ij}|} \text{pen}_2(u_1 \Delta_{ij}) \right].
\end{aligned}$$

Due to the construction of $\tilde{\Theta}$, we have $(\text{I}) \geq 0$.

For (II) , by condition (ii) and the upper bound on the magnitude of the second derivatives of $\text{pen}_2(\theta_{ij})$ in Lemma 5, we have

$$\begin{aligned}
(\text{II}) &\geq \sum_{i,j} \sum_{k=1}^K \frac{n}{8} \xi_1^2 \Delta_{k,ij}^2 - \sum_{i \neq j, (i,j) \in \mathcal{B}} \frac{\frac{n C_3^2 \log p}{4 \alpha p^{\epsilon_0/\alpha}}}{4} \left(\sum_{k=1}^K \Delta_{k,ij} \right)^2 \\
&\geq \sum_{i \neq j, (i,j) \in \mathcal{B}^c} n \left(\frac{1}{8K} \xi_1^2 - \frac{C_3^2 \log p}{16 \alpha p^{\epsilon_0/\alpha}} \right) \left(\sum_{k=1}^K \Delta_{k,ij} \right)^2 \geq 0.
\end{aligned}$$

In (III) , we have the following upper bound on the magnitude of $s_{k,ij} - \tilde{\Theta}_{k,ij}^{-1}$:

$$\begin{aligned}
|s_{k,ij} - \tilde{\Theta}_{k,ij}^{-1}| &\leq |s_{k,ij} - \sigma_{k,ij}^0| + |\tilde{\Theta}_{k,ij}^{-1} - \sigma_{k,ij}^0| \\
&\leq C_1 \sqrt{\frac{\log p}{n}} + M_{\Sigma_k^0}^2 r_e + \frac{3}{2} d_k M_{\Sigma_k^0}^3 r_e^2,
\end{aligned}$$

where $|\tilde{\Theta}_{k,ij}^{-1} - \sigma_{k,ij}^0| < M_{\Sigma_k^0}^2 r_e + \frac{3}{2} d_k M_{\Sigma_k^0}^3 r_e^2$ is by Corollary 4 in [5].

For the rest of the first derivative inside the inner summation in (III) , we have

$$\begin{aligned}
\lim_{\Delta_{ij} \rightarrow \mathbf{0}} \frac{\alpha}{n} \frac{\partial}{\partial |\theta_{k,ij}|} \text{pen}_2(u_1 \Delta_{ij}) &= \frac{\alpha}{n} \left[w(\mathbf{0}) \frac{1}{v_1} + (1 - w(\mathbf{0})) \frac{1}{v_0} \right] \\
&> (1 - w(\mathbf{0})) \frac{\alpha}{n v_0} \\
&> \frac{\epsilon_2}{1 + \epsilon_2} \frac{\alpha}{n v_0} \\
&> C_1 \sqrt{\frac{\log p}{n}} + M_{\Sigma_k^0}^2 r_e + \frac{3}{2} d_k M_{\Sigma_k^0}^3 r_e^2 \\
&\geq |s_{k,ij} - \tilde{\Theta}_{k,ij}^{-1}|,
\end{aligned}$$

where the second inequality is due to $\epsilon_2 < v_1^K (1 - p_1) / (v_0^K p_1)$. Therefore, there exists some small enough $\epsilon > 0$ such that $(\text{III}) \geq 0$.

Hence, there exists some $\epsilon > 0$ such that $H(\Delta) \geq (\text{I}) + (\text{II}) + (\text{III}) \geq 0$. Due to the strict convexity of our problem, we conclude that the constructed $\tilde{\Theta}$ is equal to the unique global solution $\hat{\Theta}$.

■

1.3 Proof of Theorem 3

In this subsection, we provide the proof of Theorem 3 of selection consistency in the main paper.

- When $\theta_{ij}^0 = \mathbf{0}$, we have the estimate $\hat{\theta}_{ij} = \mathbf{0}$. Therefore,

$$w(\hat{\theta}_{ij}) = w(\mathbf{0}) = \frac{1}{1 + \frac{1-p_1}{p_1} \left(\frac{v_1}{v_0}\right)^K} \leq t.$$

- On the other hand, when $\theta_{ij}^0 \neq \mathbf{0}$, by the minimal signal strength assumption in Theorem 3, we have $\|\hat{\theta}_{ij}\|_1 > (L_0 - C_5)\sqrt{\log p/n}$ with probability going to one. Thus,

$$\begin{aligned} w(\hat{\theta}_{ij}) &= \frac{1}{1 + \frac{1-p_1}{p_1} \left(\frac{v_1}{v_0}\right)^K \exp \left\{ -\left(\frac{1}{v_0} - \frac{1}{v_1}\right) \|\hat{\theta}_{ij}\|_1 \right\}} \\ &> \frac{1}{1 + \frac{1-p_1}{p_1} \left(\frac{v_1}{v_0}\right)^K p^{-(C_4-C_3)(L_0-C_5)/\alpha}} \\ &> t. \end{aligned}$$

Therefore, we have

$$\mathbb{P}(\hat{\mathcal{S}} = \mathcal{S}^0) \rightarrow 1.$$

2 Other Proof

The following lemma gives the upper bounds on the magnitudes of the first derivative and second derivative of $\text{pen}_2(\theta_{ij})$ when θ_{ij} is $O(\sqrt{\log p/n})$ away from the zero point $\mathbf{0}$.

Lemma 5 *Under condition (i) in Theorem 4, if $\|\theta_{ij}\|_1 \geq 2(C_2 - C_3)M_{\Gamma^0}\sqrt{\log p/n}$, then for any k and k' , we have*

$$\frac{\alpha}{n} \left| \frac{\partial}{\partial \theta_{k,ij}} \text{pen}_2(\theta_{ij}) \right| < C_3 \sqrt{\log p/n}.$$

and

$$\frac{\alpha}{n} \left| \frac{\partial^2}{\partial \theta_{k,ij} \partial \theta_{k',ij}} \text{pen}_2(\theta_{ij}) \right| < \frac{C_3^2 \log p}{4\alpha p^{\epsilon_0/\alpha}}$$

Proof

- For the first derivative, we have

$$\begin{aligned} \frac{1}{n} \left| \frac{\partial}{\partial \theta_{k,ij}} \text{pen}_2(\theta_{ij}) \right| &= \frac{1}{n} \frac{\frac{p_1}{v_1^{K+1}} e^{-\|\theta_{ij}\|_1/v_1} + \frac{1-p_1}{v_0^{K+1}} e^{-\|\theta_{ij}\|_1/v_0}}{\frac{p_1}{v_1^K} e^{-\|\theta_{ij}\|_1/v_1} + \frac{1-p_1}{v_0^K} e^{-\|\theta_{ij}\|_1/v_0}} \\ &= \frac{1}{nv_1} + \frac{\frac{1}{v_0} - \frac{1}{v_1}}{n \left[\frac{p_1 v_0^K}{(1-p_1)v_1^K} e^{\|\theta_{ij}\|_1(1/v_0-1/v_1)} + 1 \right]} \\ &\leq \frac{1}{nv_1} \left(1 + \frac{\frac{(1-p_1)v_1^{K+1}}{p_1 v_0^{K+1}}}{e^{\|\theta_{ij}\|_1(1/v_0-1/v_1)}} \right) \\ &\leq \frac{1}{nv_1} (1 + \epsilon_1), \end{aligned}$$

where the last line is because

$$\begin{cases} \frac{(1-p_1)v_1^{K+1}}{p_1 v_0^{K+1}} < \epsilon_1 p^{\epsilon_0/\alpha}, \\ e^{\|\boldsymbol{\theta}_{ij}\|_1(1/v_0-1/v_1)} > p^{2\epsilon_0/\alpha}. \end{cases}$$

Therefore,

$$\begin{aligned} \frac{\alpha}{n} \left| \frac{\partial}{\partial \theta_{k,ij}} \text{pen}_2(\boldsymbol{\theta}_{ij}) \right| &\leq \frac{\alpha}{nv_1} (1 + \epsilon_1) \\ &\leq C_3 \sqrt{\log p/n}. \end{aligned}$$

- For the second derivative, we have

$$\begin{aligned} \frac{1}{n} \left| \frac{\partial^2}{\partial \theta_{k,ij} \partial \theta_{k',ij}} \text{pen}_2(\boldsymbol{\theta}_{ij}) \right| &= \frac{\left(\frac{1}{v_0} - \frac{1}{v_1} \right)^2 \frac{p_1 v_0^K}{(1-p_1)v_1^K} e^{-\|\boldsymbol{\theta}_{ij}\|_1(1/v_0-1/v_1)}}{n \left[\frac{p_1 v_0^K}{(1-p_1)v_1^K} e^{-\|\boldsymbol{\theta}_{ij}\|_1(1/v_0-1/v_1)} + 1 \right]^2} \\ &\leq \frac{\left(\frac{1}{v_0} - \frac{1}{v_1} \right)^2}{n \left[\frac{p_1 v_0^K}{(1-p_1)v_1^K} e^{\|\boldsymbol{\theta}_{ij}\|_1(1/v_0-1/v_1)} + 1 \right]} \\ &\leq \frac{1}{nv_1^2} \frac{\frac{(1-p_1)v_1^{K+2}}{p_1 v_0^{K+2}}}{e^{\|\boldsymbol{\theta}_{ij}\|_1(1/v_0-1/v_1)}} \\ &\leq \frac{1}{nv_1^2} \frac{\epsilon_1}{p^{\epsilon_0/\alpha}}. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\alpha}{n} \left| \frac{\partial^2}{\partial \theta_{k,ij} \partial \theta_{k',ij}} \text{pen}_2(\boldsymbol{\theta}_{ij}) \right| &\leq \frac{\alpha}{nv_1^2} \frac{\epsilon_1}{p^{\epsilon_0/\alpha}} \\ &\leq \frac{\alpha n^2}{n \alpha^2} \frac{C_3^2}{(1 + \epsilon_1)^2} \frac{\log p}{n} \frac{\epsilon_1}{p^{\epsilon_0/\alpha}} \\ &\leq \frac{C_3^2 \log p}{4\alpha p^{\epsilon_0/\alpha}}. \end{aligned}$$

■

3 Computation: an EM algorithm

In this section, we present an EM algorithm for the computation. We treat $\Gamma = (\gamma_{ij})$ as latent variables and estimate $\boldsymbol{\Theta}$ by applying the Expectation step (E-step) and Maximization step (M-step) iteratively.

3.1 E-step

In the E-step, we compute the Q function defined as the expectation of the full log likelihood with respect to the joint posterior distribution of the latent variables, Γ , given \mathbf{Y} and $\boldsymbol{\Theta}^{(t)}$, the current estimate of $\boldsymbol{\Theta}$:

$$\begin{aligned} Q(\boldsymbol{\Theta} \mid \boldsymbol{\Theta}^{(t)}) &= \mathbb{E}_{\Gamma \mid \boldsymbol{\Theta}^{(t)}, \mathbf{Y}} \log \left(p(\mathbf{Y} \mid \boldsymbol{\Theta}, \Gamma)^{\frac{1}{\alpha}} p(\boldsymbol{\Theta}, \Gamma) \right) \\ &= \mathbb{E}_{\Gamma \mid \boldsymbol{\Theta}^{(t)}, \mathbf{Y}} \left(- \sum_{i < j} \sum_{k=1}^K \left[r_{k,ij} \frac{|\theta_{k,ij}|}{v_1} + (1 - r_{k,ij}) \frac{|\theta_{k,ij}|}{v_0} \right] \right) \end{aligned}$$

$$+ \frac{1}{\alpha} \log p(\mathbf{Y} \mid \boldsymbol{\Theta}) - \sum_{i=1}^p \sum_{k=1}^K \tau \theta_{k,ii} + C,$$

where C is some constant with respect to $\boldsymbol{\Theta}$.

For any $i < j$, we have

$$p(\gamma_{ij}, \mid \boldsymbol{\Theta}^{(t)}, Y) \propto \left[p_1 \prod_{k=1}^K \text{LP}(\theta_{k,ij}^{(t)}, v_1) \right]^{\gamma_{ij}} \left[(1 - p_1) \prod_{k=1}^K \text{LP}(\theta_{k,ij}^{(t)}, v_0) \right]^{1-\gamma_{ij}},$$

and therefore,

$$\mathbb{E}_{\Gamma \mid \boldsymbol{\Theta}^{(t)}, Y}(\gamma_{ij}) = \frac{1}{1 + \frac{1-p_1}{p_1} \left(\frac{v_1}{v_0}\right)^K \exp \left\{ -\left(\frac{1}{v_0} - \frac{1}{v_1}\right) \|\boldsymbol{\theta}_{ij}^{(t)}\|_1 \right\}} = w(\boldsymbol{\theta}_{ij}^{(t)}). \quad (3.1)$$

Thus, we have the Q function as

$$\begin{aligned} Q(\boldsymbol{\Theta} \mid \boldsymbol{\Theta}^{(t)}) = & \sum_{k=1}^K \left\{ \frac{n_k}{2\alpha} \log \det(\Theta_k) - \frac{n_k}{2\alpha} \text{tr}(S_k \Theta_k) - \sum_{i=1}^p \tau \theta_{k,ii} \right. \\ & \left. - \sum_{i < j} \left[w(\boldsymbol{\theta}_{ij}^{(t)}) \frac{|\theta_{k,ij}|}{v_1} + \left(1 - w(\boldsymbol{\theta}_{ij}^{(t)})\right) \frac{|\theta_{k,ij}|}{v_0} \right] \right\} + C, \end{aligned} \quad (3.2)$$

where C is some constant with respect to $\boldsymbol{\Theta}$.

3.2 M-step

In the M-step, we update the estimate of $\boldsymbol{\Theta}$ by maximizing the Q function (3.2) under the constraints that $\Theta_k > 0$ and $\|\Theta_k\|_2 \leq B$ for $k = 1, \dots, K$. It is worth noting that $Q(\boldsymbol{\Theta} \mid \boldsymbol{\Theta}^{(t)})$ is a summation of K terms with each to be a function of Θ_k only. Therefore, our constrained optimization problem can be paralleled by maximizing

$$\frac{n_k}{2\alpha} \log \det(\Theta_k) - \frac{n_k}{2\alpha} \text{tr}(S_k \Theta_k) + \sum_{i=1}^p \tau \theta_{k,ii} - \sum_{i < j} \left[\frac{1}{v_0} + \left(\frac{1}{v_1} - \frac{1}{v_0}\right) w(\boldsymbol{\theta}_{ij}^{(t)}) \right] |\theta_{k,ij}|, \quad (3.3)$$

under the constraints of $\Theta_k > 0$ and $\|\Theta_k\|_2 \leq B$ for $k = 1, \dots, K$. For each sub-problem of maximizing (3.3), we adopt the BAGUS algorithm [2] and update Θ_k sequentially in the same column by column fashion. The time complexity of our algorithm is thus $\mathcal{O}(p^3)$, the same as the state-of-the-art algorithms for graphical Lasso problems. In addition, our final estimates $\hat{\Theta}_k$'s are all symmetric and positive definite as guaranteed in [2].

The maximization of (3.3) can be interpreted as a graphical Lasso problem with adaptive shrinkage where the weight of the penalty is a decreasing function of $\|\boldsymbol{\theta}_{ij}^{(t)}\|_1$. Therefore, for a group with large signals, every entry within the group would not be shrunk heavily. This adaptive shrinkage is advantageous compared to individual estimation methods, because it takes into account the group information to adjust the adaptive weights of penalization for each entry. For example, when small signals are presented in a group with large signals, their estimates will get shrunk less compared to individual estimation methods.

3.3 Algorithm

The algorithm is summarized as Algorithm 1. The outputs from the algorithm are the posterior inclusion probability matrix of common structures P and the estimates of the K precision matrices Θ_k 's. For each entry of P , we can threshold a pre-specified level, e.g., 0.5, to get an estimate of the sparsity structure.

Algorithm 1 Joint estimation of multiple graphical models.

```
repeat
  Calculate  $P = (w(\boldsymbol{\theta}_{ij}^{(t)}))$  using (3.1)
  for  $k = 1$  to  $K$  do
    Update  $\Theta_k$  by maximizing (3.3)
  end for
until Convergence
Return  $P$  and  $\Theta_k, k = 1, \dots, K$ 
```

References

- [1] 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.
- [2] Gan, L., Narisetty, N. N., and Liang, F. (2019). Bayesian regularization for graphical models with unequal shrinkage. *Journal of the American Statistical Association*, 114(527):1218–1231.
- [3] Loh, P.-L. and Wainwright, M. J. (2017). Support recovery without incoherence: A case for nonconvex regularization. *The Annals of Statistics*, 45(6):2455–2482.
- [4] Ortega, J. M. and Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables*, volume 30. Siam.
- [5] 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*.
- [6] 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.
- [7] Rothman, A. J., Bickel, P. J., Levina, E., and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:494–515.