Biometrics 0000, 1–28 DOI: 000
00 0000

Supporting Information for "Gaussian Graphical Model-based Heterogeneity Analysis via Penalized Fusion" by Mingyang Ren, Sanguo Zhang, Qingzhao Zhang, Shuangge Ma

This paper has been submitted for consideration for publication in *Biometrics*

A. Details of the computational algorithm

Recall that the objective function is:

$$\mathcal{L}(\boldsymbol{\Omega}, \boldsymbol{\pi}|\boldsymbol{X}) := \frac{1}{n}\sum_{i=1}^{n}\log\left(\sum_{k=1}^{K}\pi_k f_k\left(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Theta}_k^{-1}\right)\right) - \mathcal{P}(\boldsymbol{\Omega}), \tag{A.1}$$

where $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_K))^\top$, $\boldsymbol{\Theta}_k = \boldsymbol{\Sigma}_k^{-1}$ is the k-th precision matrix with the ij-th entry θ_{kij}, $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_1^\top, \cdots, \boldsymbol{\Omega}_K^\top)^\top$, $\boldsymbol{\Omega}_k = \mathrm{vec}\,(\boldsymbol{\mu}_k, \boldsymbol{\Theta}_k) = (\mu_{k1}, \ldots, \mu_{kp}, \theta_{k11}, \ldots, \theta_{kp1}, \ldots, \theta_{k1p}, \ldots, \theta_{kpp}) \in \mathbb{R}^{p^2+p}$,

$$\begin{aligned}\mathcal{P}(\boldsymbol{\Omega}) =& \sum_{k=1}^{K}\sum_{j=1}^{p} p(|\mu_{kj}|, \lambda_1) + \sum_{k=1}^{K}\sum_{i\neq j} p(|\theta_{kij}|, \lambda_2) \\ &+ \sum_{k<k'} p\left(\left(\|\boldsymbol{\mu}_k - \boldsymbol{\mu}_{k'}\|_2^2 + \|\boldsymbol{\Theta}_k - \boldsymbol{\Theta}_{k'}\|_F^2\right)^{1/2}, \lambda_3\right),\end{aligned} \tag{A.2}$$

and $p(,\lambda)$ is the MCP function with tuning parameter $\lambda > 0$.

In the t-th step of the EM algorithm, the following function needs to be maximized:

$$E_{\boldsymbol{\gamma}|\boldsymbol{X},\boldsymbol{\Omega}^{(t-1)}}[\mathcal{L}(\boldsymbol{\Omega}|\boldsymbol{X}, \boldsymbol{\gamma})] = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\gamma_{ik}^{(t)}\left[\log\pi_k + \log f_k\left(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Theta}_k^{-1}\right)\right] - \mathcal{P}(\boldsymbol{\Omega}), \tag{A.3}$$

where $\mathcal{P}(\boldsymbol{\Omega})$ is defined in (A.2), and $\gamma_{ik}^{(t)}$ can be computed based on $\pi_k^{(t-1)}$, $\boldsymbol{\mu}_k^{(t-1)}$, and $\boldsymbol{\Theta}_k^{(t-1)}$ obtained in the previous iteration. More specifically,

$$\gamma_{ik}^{(t)} = \frac{\pi_k^{(t-1)} f_k\left(\boldsymbol{x}_i; \boldsymbol{\mu}_k^{(t-1)}, \left(\boldsymbol{\Theta}_k^{(t-1)}\right)^{-1}\right)}{\sum_{k=1}^{K}\pi_k^{(t-1)} f_k\left(\boldsymbol{x}_i; \boldsymbol{\mu}_k^{(t-1)}, \left(\boldsymbol{\Theta}_k^{(t-1)}\right)^{-1}\right)}. \tag{A.4}$$

Maximizing (A.3) with respect to $\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Theta}_k$ yields the update of parameters. More specifically, the update of π_k is given by:

$$\pi_k^{(t)} = \frac{1}{n}\sum_{i=1}^{n}\gamma_{ik}^{(t)}. \tag{A.5}$$

For $\boldsymbol{\mu}_k$, maximizing (A.3) with respect to $\{\boldsymbol{\mu}\} = \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$ is equivalent to solving:

$$\{\boldsymbol{\mu}^{(t)}\} = \underset{\{\boldsymbol{\mu}\}}{\mathrm{argmin}}\left(\frac{1}{2n}\sum_{i=1}^{n}\sum_{k=1}^{K}\gamma_{ik}^{(t)}\left\{(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^\top\boldsymbol{\Theta}_k^{(t-1)}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} + \mathcal{P}(\boldsymbol{\Omega})\right). \tag{A.6}$$

For this problem, the local quadratic approximation can be adopted, which can lead to an explicit solution at each iteration. Details are provided in Section A.1.

Maximizing (A.3) with respect to $\{\boldsymbol{\Theta}\}$ is equivalent to solving:

$$\{\boldsymbol{\Theta}_k^{(t)}, k=1,\ldots,K\} = \underset{\{\boldsymbol{\Theta}\}}{\operatorname{argmax}} \left(\sum_{k=1}^{K} n_k \left[\log\{\det(\boldsymbol{\Theta}_k)\} - \operatorname{tr}\left(\widetilde{\boldsymbol{S}}_k \boldsymbol{\Theta}_k\right) \right] - \mathcal{P}(\{\boldsymbol{\Theta}\}) \right), \quad \text{(A.7)}$$

where $n_k = \sum_{i=1}^{n} \gamma_{ik}^{(t)}$, $\widetilde{\boldsymbol{S}}_k$ is the pseudo sample covariance matrix defined by:

$$\widetilde{\boldsymbol{S}}_k = \frac{\sum_{i=1}^{n} \gamma_{ik}^{(t)} \left(\boldsymbol{x}_i - \boldsymbol{\mu}_k^{(t)}\right)\left(\boldsymbol{x}_i - \boldsymbol{\mu}_k^{(t)}\right)^{\top}}{\sum_{i=1}^{n} \gamma_{ik}^{(t)}},$$

and $\mathcal{P}(\{\boldsymbol{\Theta}\}) = \sum_{k=1}^{K} \sum_{i \neq j} p(|\theta_{kij}|, \lambda_2) + \sum_{k<k'} p\left((\|\boldsymbol{\mu}_k^{(t)} - \boldsymbol{\mu}_{k'}^{(t)}\|_2^2 + \|\boldsymbol{\Theta}_k - \boldsymbol{\Theta}_{k'}\|_F^2)^{1/2}, \lambda_3 \right)$.

The solution for (A.7) can be effectively obtained using the ADMM technique. More details are provided in Section A.2. Overall, we propose the algorithm summarized in Algorithm S1.

Algorithm S1 for maximizing (A.1)

Input: $\boldsymbol{x}_i, i = 1, \cdots, n$, tuning parameters $\lambda_1, \lambda_2, \lambda_3$, and K.

Output: The estimated mean vectors and precision matrices.

Initialization: Mean vectors $\boldsymbol{\mu}_k^{(0)}$, positive-definite precision matrices $\boldsymbol{\Theta}_k^{(0)}$, and $\pi_k^{(0)}$ obtained using the K-means method, for $k = 1, \cdots, K$.

Repeat for $t = 1, 2, 3, \ldots$ **as follows**:

(1) E-step: Update the subpopulation assignment $\gamma_{ik}^{(t)}$ by (A.4).

(2) M-step: Given $\gamma_{ik}^{(t)}$, update $\pi_k^{(t)}$, $\boldsymbol{\mu}_k^{(t)}$, and $\boldsymbol{\Theta}_k^{(t)}$ by (A.5), (A.6), and (A.7) respectively.

Until: $\sum_{k=1}^{K} \left\{ \frac{\left\|\boldsymbol{\mu}_k^{(t)} - \boldsymbol{\mu}_k^{(t-1)}\right\|_2}{\left\|\boldsymbol{\mu}_k^{(t-1)}\right\|_2} + \frac{\left\|\boldsymbol{\Theta}_k^{(t)} - \boldsymbol{\Theta}_k^{(t-1)}\right\|_F}{\left\|\boldsymbol{\Theta}_k^{(t-1)}\right\|_F} \right\}$ $<$ a pre-specified cutoff (taken as 0.01 in our numerical study).

Return: The estimate of $\{\boldsymbol{\mu}_k^{(t)}, \boldsymbol{\Theta}_k^{(t)}, \pi_k^{(t)}, k = 1, \ldots, K\}$ at convergence.

A.1 *Update of* $\{\boldsymbol{\mu}\}$ *in the EM algorithm*

Recall that maximizing (A.3) with respect to $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K\}$ is equivalent to solving:

$$\{\boldsymbol{\mu}^{(t)}\} = \underset{\{\boldsymbol{\mu}\}}{\operatorname{argmin}} \left(\frac{1}{2n} \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik}^{(t)} \left[(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^{\top} \boldsymbol{\Theta}_k^{(t-1)} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right] + \mathcal{P}(\boldsymbol{\Omega}) \right). \quad \text{(A.8)}$$

We adopt the local quadratic approximation technique and solve for:

$$\{\boldsymbol{\mu}^{(t)}\} = \underset{\{\boldsymbol{\mu}\}}{\operatorname{argmin}} \left(\frac{1}{2n} \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik}^{(t)} \left[(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Theta}_k^{(t-1)} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right] + \sum_{k=1}^{K} \sum_{j=1}^{p} \frac{1}{2} \frac{p'(|\mu_{kj}^{(t-1)}|, \lambda_1)}{|\mu_{kj}^{(t-1)}|} \mu_{kj}^2 + \sum_{k<k'} \frac{1}{2} \frac{p'(\tau_{k,k'}^{(t-1,t-1)}, \lambda_3)}{\tau_{k,k'}^{(t-1,t-1)}} \|\boldsymbol{\mu}_k - \boldsymbol{\mu}_{k'}\|_2^2 \right). \tag{A.9}$$

The update of $\boldsymbol{\mu}_k^{(t)}$ is as follows. For $j = 1, \ldots, p$,

$$\mu_{kj}^{(t)} = \begin{cases} \frac{h_j^{(t-1)} + n\widetilde{v}_k^{(t-1)} - np'(|\mu_{kj}^{(t-1)}|, \lambda_1) \operatorname{sign}\left(\mu_{kj}^{(t-1)}\right)}{n\widehat{v}_k^{(t-1)} + n_k \theta_{kjj}^{(t-1)}}, & \text{if } |h_j^{(t-1)} + n\widetilde{v}_k^{(t-1)}| > np'(|\mu_{kj}^{(t-1)}|, \lambda_1), \\ 0, & \text{otherwise}, \end{cases} \tag{A.10}$$

where

$$\begin{aligned} h_j^{(t-1)} &= \sum_{i=1}^{n} \gamma_{ik}^{(t)} \left[\sum_{l \neq j}^{p} \theta_{kjl}^{(t-1)} \left(x_{il} - \mu_{kl}^{(t-1)} I(l > j) - \mu_{kl}^{(t)} I(l < j) \right) + \theta_{kjj}^{(t-1)} x_{ij} \right], \\ \widetilde{v}_k^{(t-1)} &= \sum_{1 \leqslant k' < k} \frac{p'(\widetilde{\tau}_{k,k'}^{(t-1)}, \lambda_3)}{\widetilde{\tau}_{k,k'}^{(t-1)}} \mu_{k'j}^{(t)} + \sum_{k < k' \leqslant K} \frac{p'(\widetilde{\tau}_{k,k'}^{(t-1)}, \lambda_3)}{\widetilde{\tau}_{k,k'}^{(t-1)}} \mu_{k'j}^{(t-1)}, \\ \widehat{v}_k^{(t-1)} &= \sum_{k \neq k'} \frac{p'(\widetilde{\tau}_{k,k'}^{(t-1)}, \lambda_3)}{\widetilde{\tau}_{k,k'}^{(t-1)}}, \\ \widetilde{\tau}_{k,k'}^{(t-1)} &= \tau_{k,k'}^{(t-1,t)} I(1 \leqslant k' < k) + \tau_{k,k'}^{(t-1,t-1)} I(k < k' \leqslant K), \\ \tau_{k,k'}^{(r_1,r_2)} &= \left(\|\boldsymbol{\mu}_k^{(r_1)} - \boldsymbol{\mu}_{k'}^{(r_2)}\|_2^2 + \|\boldsymbol{\Theta}_k^{(t-1)} - \boldsymbol{\Theta}_{k'}^{(t-1)}\|_F^2 \right)^{1/2}, \end{aligned}$$

and $\boldsymbol{\Theta}_k^{(t-1)}$, $\boldsymbol{\Theta}_{k'}^{(t-1)}$ are estimators from the $(t-1)$th iteration.

A.2 *Update of* $\{\boldsymbol{\Theta}\}$ *in the EM algorithm*

Recall that maximizing (A.3) with respect to $\boldsymbol{\Theta}$ is equivalent to solving:

$$\{\boldsymbol{\Theta}_k^{(t)}, k = 1, \ldots, K\} = \underset{\{\boldsymbol{\Theta}\}}{\operatorname{argmax}} \left(\sum_{k=1}^{K} n_k \left[\log\{\det(\boldsymbol{\Theta}_k)\} - \operatorname{tr}\left(\widetilde{\boldsymbol{S}}_k \boldsymbol{\Theta}_k \right) \right] - \mathcal{P}(\{\boldsymbol{\Theta}\}) \right). \tag{A.11}$$

This can be efficiently achieved using the ADMM algorithm by modifying the joint graphical lasso algorithm in Danaher et al. (2014). More specifically, this optimization can be reformulated as:

$$\underset{\{\boldsymbol{\Theta}, \boldsymbol{\Xi}\}}{\operatorname{argmin}} \left(-\sum_{k=1}^{K} n_k \left[\log\{\det(\boldsymbol{\Theta}_k)\} - \operatorname{tr}\left(\widetilde{\boldsymbol{S}}_k \boldsymbol{\Theta}_k \right) \right] + \mathcal{P}(\{\boldsymbol{\Xi}\}) \right), \tag{A.12}$$

subject to the constraint that $\boldsymbol{\Xi}_k = \boldsymbol{\Theta}_k, k = 1, \cdots, K$ as well as the positive definiteness constraint, where $\{\boldsymbol{\Xi}\} = \boldsymbol{\Xi}_1, \ldots, \boldsymbol{\Xi}_K$, and $\boldsymbol{\Xi}_k = (\xi_{kij})_{1 \leqslant i,j \leqslant p}$. The scaled augmented Lagrangian form for this problem is given by:

$$\begin{aligned} \mathcal{Q}_\kappa(\{\boldsymbol{\Theta}\}, \{\boldsymbol{\Xi}\}, \{\boldsymbol{\Psi}\}) = &- \sum_{k=1}^{K} n_k \left[\log\left\{\det\left(\boldsymbol{\Theta}_k\right)\right\} - \operatorname{tr}\left(\widetilde{\boldsymbol{S}}_k \boldsymbol{\Theta}_k\right)\right] + \mathcal{P}(\{\boldsymbol{\Xi}\}) \\ &+ \frac{\kappa}{2} \sum_{k=1}^{K} \left\|\boldsymbol{\Theta}_k - \boldsymbol{\Xi}_k + \boldsymbol{\Psi}_k\right\|_{\mathrm{F}}^2 - \frac{\kappa}{2} \sum_{k=1}^{K} \left\|\boldsymbol{\Psi}_k\right\|_{\mathrm{F}}^2, \end{aligned} \tag{A.13}$$

where $\{\boldsymbol{\Psi}\} = \boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_K$ are dual variables, and κ is the penalty parameter. Throughout this study, we set $\kappa = 1$. The ADMM algorithm for solving (A.12) is summarized in Algorithm S2.

It is noted that the positive definiteness of the estimated precision matrices is naturally enforced by the update in Step 1 of Algorithm S2. The solution $\{\boldsymbol{\Theta}_1^{(t)}, \ldots, \boldsymbol{\Theta}_K^{(t)}\}$ for (A.11) obtained by Algorithm S2 is not necessarily symmetric in general. They can be symmetrized using the strategy of Cai et al. (2016) and Hao et al. (2018):

$$\theta_{kij}^{(t)} = \theta_{kij}^{(t)} I(|\theta_{kij}^{(t)}| \leqslant |\theta_{kji}^{(t)}|) + \theta_{kji}^{(t)} I(|\theta_{kij}^{(t)}| > |\theta_{kji}^{(t)}|). \tag{A.14}$$

This step does not affect the convergence rate of the final estimator (Cai et al., 2016).

Next, for solving (A.15), the efficient sparse alternating minimization algorithm (S-AMA, Wang et al. (2018)) can be used. To implement the S-AMA, (A.15) can be rewritten as:

$$\begin{aligned} &\min_{\{\boldsymbol{\Xi}\}} \frac{\kappa}{2} \sum_{j=1}^{p^2} \left\|\boldsymbol{\xi}_{(j)} - \boldsymbol{z}_{(j)}\right\|_2^2 + \sum_{r \in \mathcal{E}} p\left((\eta_r^{(t)} + \|\boldsymbol{v}_r\|_2^2)^{1/2}, \lambda_3\right) + \sum_{j=1}^{p^2} \sum_{k=1}^{K} p(|\xi_{kj}|, \lambda_2) \cdot I(j \in \mathcal{O}), \\ &\text{s.t. } \operatorname{vec}\boldsymbol{\Xi}_k - \operatorname{vec}\boldsymbol{\Xi}_{k'} - \mathbf{v}_r = 0, \end{aligned}$$

where $\mathcal{E} = \{(k, k') : 1 \leqslant k, k' \leqslant K\}$, $\eta_r^{(t)} = \|\boldsymbol{\mu}_k^{(t)} - \boldsymbol{\mu}_{k'}^{(t)}\|_2^2$, $\boldsymbol{\xi}_{(j)}, \boldsymbol{z}_{(j)} \in \mathbb{R}^K$ are the j-th columns of $(\operatorname{vec}\boldsymbol{\Xi}_1, \cdots, \operatorname{vec}\boldsymbol{\Xi}_K)^\top$ and $(\operatorname{vec}\boldsymbol{\mathcal{Z}}_1, \cdots, \operatorname{vec}\boldsymbol{\mathcal{Z}}_K)^\top$, respectively, $j = 1, \cdots, p^2$, ξ_{kj} is the k-th element of $\boldsymbol{\xi}_{(j)}$, and $\mathcal{O} = \{j : j \neq d(p+1) + 1, d = 0, 1, \cdots, p-1\}$ is the index set of the off-diagonal components of the precision matrices. It is equivalent to minimizing the

Algorithm S2 ADMM algorithm for solving (A.12)

Input: The pseudo sample covariance matrices $\widetilde{\boldsymbol{S}}_k, k = 1, \cdots, K$, tuning parameters λ_2, λ_3, and penalty parameter κ.

Output: The estimated precision matrices $\{\boldsymbol{\Theta}_k, k = 1, \ldots, K\}$.

Initialization: $\boldsymbol{\Theta}_k^{(0)} = \boldsymbol{I}, \boldsymbol{\Xi}_k^{(0)} = \boldsymbol{0}, \boldsymbol{\Psi}_k^{(0)} = \boldsymbol{0}$, for $k = 1, \cdots, K$.

Repeat for $m = 1, 2, 3, \ldots$:

(1) For $k = 1, \ldots, K$, update $\boldsymbol{\Theta}_k^{(m)}$ by solving

$$\underset{\{\boldsymbol{\Theta}\}}{\operatorname{argmin}} \left(-n_k \left[\log \left\{ \det \left(\boldsymbol{\Theta}_k \right) \right\} - \operatorname{tr} \left(\widetilde{\boldsymbol{S}}_k \boldsymbol{\Theta}_k \right) \right] + \frac{\kappa}{2} \left\| \boldsymbol{\Theta}_k - \boldsymbol{\Xi}_k^{(m-1)} + \boldsymbol{\Psi}_k^{(m-1)} \right\|_{\mathrm{F}}^2 \right).$$

Following Witten and Tibshirani (2009), the solution is given by:

$$\boldsymbol{\Theta}_k^{(m)} = \boldsymbol{U} \widetilde{\boldsymbol{D}} \boldsymbol{U}^\top,$$

where $\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$ is the eigendecomposition of $\widetilde{\boldsymbol{S}}_k - \kappa \boldsymbol{\Xi}_k^{(m-1)}/n_k + \kappa \boldsymbol{\Psi}_k^{(m-1)}/n_k$, $\widetilde{\boldsymbol{D}}$ is a diagonal matrix with the jth diagonal element $\frac{n_k}{2\kappa} \left[-D_{jj} + \left(D_{jj}^2 + 4\kappa/n_k \right)^{1/2} \right]$, and D_{jj} is the jth diagonal element of $\boldsymbol{D}$.

(2) Update $\{\boldsymbol{\Xi}^{(m)}\}$ by solving:

$$\underset{\{\boldsymbol{\Xi}\}}{\operatorname{argmin}} \left(\frac{\kappa}{2} \sum_{k=1}^{K} \| \boldsymbol{\Xi}_k - \boldsymbol{\mathcal{Z}}_k \|_{\mathrm{F}}^2 + \mathcal{P}(\{\boldsymbol{\Xi}\}) \right) \tag{A.15}$$

using the S-AMA algorithm (Wang et al., 2018), where $\boldsymbol{\mathcal{Z}}_k = \boldsymbol{\Theta}_k^{(m)} + \boldsymbol{\Psi}_k^{(m-1)}$.

(3) Update $\{\boldsymbol{\Psi}^{(m)}\}$ by $\boldsymbol{\Psi}_k^{(m)} = \boldsymbol{\Psi}_k^{(m-1)} + \boldsymbol{\Theta}_k^{(m)} - \boldsymbol{\Xi}_k^{(m)}$, for $k = 1, \ldots, K$.

Until: $\sum_{k=1}^{K} \frac{\left\| \boldsymbol{\Theta}_k^{(m)} - \boldsymbol{\Theta}_k^{(m-1)} \right\|_F}{\left\| \boldsymbol{\Theta}_k^{(m-1)} \right\|_F} < 10^{-2}$.

Return: The estimate of $\{\boldsymbol{\Theta}_k^{(m)}, k = 1, \ldots, K\}$ at convergence.

following augmented Lagrangian function:

$$\begin{aligned}
\mathcal{Q}_{\kappa'}(\{\boldsymbol{\Xi}\}, \boldsymbol{V}, \boldsymbol{\Delta}) =& \frac{\kappa}{2} \sum_{j=1}^{p^2} \left\| \boldsymbol{\xi}_{(j)} - \boldsymbol{z}_{(j)} \right\|_2^2 + \sum_{r \in \mathcal{E}} p\left((\eta_r^{(t)} + \| \mathbf{v}_r \|_2^2)^{1/2}, \lambda_3 \right) \\
&+ \sum_{j=1}^{p^2} \sum_{k=1}^{K} p(|\xi_{jk}|, \lambda_2) \cdot I(j \in \mathcal{O}) + \sum_{r \in \mathcal{E}} \langle \boldsymbol{\delta}_r, \mathbf{v}_r - \mathrm{vec}\boldsymbol{\Xi}_k + \mathrm{vec}\boldsymbol{\Xi}_{k'} \rangle \\
&+ \frac{\kappa'}{2} \sum_{r \in \mathcal{E}} \| \mathbf{v}_r - \mathrm{vec}\boldsymbol{\Xi}_k + \mathrm{vec}\boldsymbol{\Xi}_{k'} \|_2^2,
\end{aligned}$$

where κ' is a small penalty parameter. In this study, we set $\kappa' = 1$. $\boldsymbol{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_{|\mathcal{E}|})$, and $\boldsymbol{\Delta} = (\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_{|\mathcal{E}|})$. S-AMA minimizes the augmented Lagrangian problem by alternatively solving one block of variables at a time:

$$\{\boldsymbol{\Xi}^{(s+1)}\} = \underset{\{\boldsymbol{\Xi}\}}{\operatorname{argmin}} \mathcal{Q}_0(\{\boldsymbol{\Xi}\}, \boldsymbol{V}^{(s)}, \boldsymbol{\Delta}^{(s)}),$$

$$\boldsymbol{V}^{(s+1)} = \underset{\boldsymbol{V}}{\operatorname{argmin}} \mathcal{Q}_{\kappa'}(\{\boldsymbol{\Xi}^{(s+1)}\}, \boldsymbol{V}, \boldsymbol{\Delta}^{(s)}),$$

$$\boldsymbol{\delta}_r^{(s+1)} = \boldsymbol{\delta}_r^{(s)} + \kappa'(\mathbf{v}_r^{(s+1)} - \mathrm{vec}\boldsymbol{\Xi}_k^{(s+1)} + \mathrm{vec}\boldsymbol{\Xi}_{k'}^{(s+1)}) \cdot I(\|\mathbf{v}_r^{(s+1)}\|_2 > 0), r \in \mathcal{E}.$$

It is noted that AMA differs from ADMM in the update of $\{\boldsymbol{\Xi}\}$. Specifically, AMA solves $\{\boldsymbol{\Xi}\}$ by treating $\kappa' = 0$. The detailed updating implementations for $\{\boldsymbol{\Xi}\}$ and $\boldsymbol{V}$ are as follows.

Following Wang et al. (2018), when $\kappa' = 0$, updating $\{\boldsymbol{\Xi}\}$ requires solving p^2 individual penalization problems:

$$\min_{\boldsymbol{\xi}_{(j)}} \frac{\kappa}{2} \left\|\boldsymbol{\xi}_{(j)} - \boldsymbol{u}_{(j)}\right\|_2^2 + \sum_{k=1}^{K} p(|\xi_{kj}|, \lambda_2) \cdot I(j \in \mathcal{O}), \tag{A.16}$$

where $\boldsymbol{u}_{(j)} = \boldsymbol{z}_{(j)} + \sum_{r\in\mathcal{E}} \delta_{jr}(\boldsymbol{e}_k - \boldsymbol{e}_{k'})$, δ_{jr} is the jth element of $\boldsymbol{\delta}_r$, and $\boldsymbol{e}_k$ is a K-dimensional vector with the k-th component being 1 and the other components being 0. For (A.16), the MCP estimators have the following closed form:

$$\hat{\xi}_{kj} = I(j \in \mathcal{O}) \cdot \begin{cases} \frac{S(u_{kj}, \lambda_2/\kappa)}{1-1/(a\kappa)} & \text{if } |u_{kj}| \leqslant a\lambda_2 \\ u_{kj} & \text{if } |u_{kj}| > a\lambda_2 \end{cases} + I(j \notin \mathcal{O}) \cdot u_{kj},$$

where $S(t, \lambda) = (|t| - \lambda)_+ \cdot \mathrm{sign}(t)$.

For $\boldsymbol{V}$, $\mathbf{v}_r$'s can be updated separately. That is,

$$\hat{\mathbf{v}}_r = \underset{\mathbf{v}_r}{\operatorname{argmin}} \frac{\kappa'}{2}\|\mathbf{v}_r - \boldsymbol{\omega}_r\|_2^2 + p\left((\eta_r^{(t)} + \|\mathbf{v}_r\|_2^2)^{1/2}, \lambda_3\right), \tag{A.17}$$

where $\boldsymbol{\omega}_r = \mathrm{vec}\boldsymbol{\Xi}_k - \mathrm{vec}\boldsymbol{\Xi}_{k'} - \boldsymbol{\delta}_r/\kappa'$. By the Karush–Kuhn–Tucker (KKT) conditions of the group MCP problem (Breheny and Huang, 2015), the solution to (A.17) has a closed form:

$$\hat{\mathbf{v}}_r = \frac{\boldsymbol{\omega}_r}{(\eta_r^{(t)} + \|\boldsymbol{\omega}_r\|_2^2)^{1/2}} \cdot \begin{cases} \frac{S((\eta_r^{(t)} + \|\boldsymbol{\omega}_r\|_2^2)^{1/2}, \lambda_3/\kappa')}{1-1/(a\kappa')} & \text{if } (\eta_r^{(t)} + \|\boldsymbol{\omega}_r\|_2^2)^{1/2} \leqslant a\lambda_3, \\ (\eta_r^{(t)} + \|\boldsymbol{\omega}_r\|_2^2)^{1/2} & \text{if } (\eta_r^{(t)} + \|\boldsymbol{\omega}_r\|_2^2)^{1/2} > a\lambda_3. \end{cases}$$

B. Proof of Theorem 1

B.1 *Notations and Conditions*

Denote the true values of parameters as $\boldsymbol{\Omega}^* = (\boldsymbol{\Omega}_1^{*\top}, \cdots, \boldsymbol{\Omega}_K^{*\top})^\top$, $\boldsymbol{\Omega}_k^* = \text{vec}\,(\boldsymbol{\mu}_k^*, \boldsymbol{\Theta}_k^*) \in \mathbb{R}^{p^2+p}$. Let $\{\boldsymbol{\Upsilon}_1^*, , \cdots, \boldsymbol{\Upsilon}_{K_0}^*\}$ be the distinct values of $\boldsymbol{\Omega}^*$, and $\mathcal{T}_l^* = \{j : \boldsymbol{\Omega}_j^* \equiv \boldsymbol{\Upsilon}_l^*\}, 1 \leqslant l \leqslant K_0$. Then $\{\mathcal{T}_1^*, \cdots, \mathcal{T}_{K_0}^*\}$ constitutes a partition of $\{1, \cdots, K\}$. Define $\boldsymbol{\Upsilon}^* = (\boldsymbol{\Upsilon}_1^{*\top}, \cdots, \boldsymbol{\Upsilon}_{K_0}^{*\top})^\top$. Write $|\mathcal{T}_{\min}| = \min_{1\leqslant l\leqslant K_0} |\mathcal{T}_l^*|$, and $|\mathcal{T}_{\max}| = \max_{1\leqslant l\leqslant K_0} |\mathcal{T}_l^*|$, where $|\cdot|$ is the cardinality of the set. Define:

$$\Lambda_{\mathcal{T}^*} = \{\boldsymbol{\Omega} \in \mathbb{R}^{K(p+p^2)} : \boldsymbol{\Omega}_k \equiv \boldsymbol{\Omega}_{k'}, \text{ for any } k, k' \in \mathcal{T}_l^*, 1 \leqslant l \leqslant K_0\}.$$

Let $b = \min_{k_1\in\mathcal{T}_{l_1}^*, k_2\in\mathcal{T}_{l_2}^*, 1\leqslant l_1\neq l_2\leqslant K_0} \left\|\boldsymbol{\Omega}_{k_1}^* - \boldsymbol{\Omega}_{k_2}^*\right\|_2 = \min_{1\leqslant l_1\neq l_2\leqslant K_0} \left\|\boldsymbol{\Upsilon}_{l_1}^* - \boldsymbol{\Upsilon}_{l_2}^*\right\|_2$. We define the oracle estimator for $\boldsymbol{\Omega}$, under which the true subgrouping memberships $\{\mathcal{T}_1^*, \cdots, \mathcal{T}_{K_0}^*\}$ are known, as:

$$\begin{aligned}(\widehat{\boldsymbol{\Omega}}^o, \widehat{\boldsymbol{\pi}}^o) = \underset{\boldsymbol{\Omega}\in\Lambda_{\mathcal{T}^*}}{\text{argmax}} \Bigg\{ & \frac{1}{n}\sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} \pi_k f_k\left(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Theta}_k^{-1}\right)\right) \\ & - \sum_{k=1}^{K}\sum_{j=1}^{p} p(|\mu_{kj}|, \lambda_1) - \sum_{k=1}^{K}\sum_{i\neq j} p(|\theta_{kij}|, \lambda_2) \Bigg\}.\end{aligned} \tag{B.1}$$

From $\widehat{\boldsymbol{\Omega}}^o$, we can obtain the distinct values of $\widehat{\boldsymbol{\Omega}}^o$ similarly, that is, $\{\widehat{\boldsymbol{\Upsilon}}_1^o, , \cdots, \widehat{\boldsymbol{\Upsilon}}_{K_0}^o\}$, and define $\widehat{\boldsymbol{\Upsilon}}^o = (\widehat{\boldsymbol{\Upsilon}}_1^{o\top}, \cdots, \widehat{\boldsymbol{\Upsilon}}_{K_0}^{o\top})^\top$. For $l = 1, \cdots, K_0$, define $\mathcal{S}_l = \left\{(i,j) : \theta_{lij}^* \neq 0, 1 \leqslant i \neq j \leqslant p\right\}$, $\mathcal{D}_l = \left\{j : \mu_{lj}^* \neq 0, 1 \leqslant j \leqslant p\right\}$, and the sparsity parameters $s = \max\{|\mathcal{S}_l|, l = 1, \ldots, K_0\}$, $d = \max\{|\mathcal{D}_l|, l = 1, \ldots, K_0\}$.

Denote the index set of the diagonal components of the K_0 precision matrices by:

$$\mathcal{G} = \bigcup_{l=1}^{K_0} \mathcal{G}_l, \ \mathcal{G}_l = \left(l(p+1), l(2p+2), \ldots, l\left(p^2+p\right)\right). \tag{B.2}$$

Then $\boldsymbol{\Upsilon}_{\mathcal{G}} = (\theta_{111}, \ldots, \theta_{1pp}, \ldots, \theta_{K_0 11}, \ldots, \theta_{K_0 pp}) \in \mathbb{R}^{K_0 p}$. In addition, define the index set of

the nonzero components of $\boldsymbol{\Upsilon}^*$ as:

$$\begin{aligned}
&\mathcal{M} = \bigcup_{l=1}^{K_0} \mathcal{M}_l,\ \mathcal{M}_l = \mathcal{M}_{\mathcal{D}_l} \cup \mathcal{M}_{\mathcal{S}_l} \cup \mathcal{G}_l, \\
&\mathcal{M}_{\mathcal{D}_l} = \left\{(l-1)(p+p^2) + j : j \in \mathcal{D}_l\right\}, \\
&\mathcal{M}_{\mathcal{S}_l} = \left\{(l-1)(p+p^2) + ip + j : (i,j) \in \mathcal{S}_l\right\}.
\end{aligned} \tag{B.3}$$

Let $\mathcal{G}^C$ and $\mathcal{M}^C$ be the complement of $\mathcal{G}$ and $\mathcal{M}$, respectively.

Denote $x \lesssim y$ if $x \leqslant Dy$, and $x \simeq y$ if $x = Dy$, for some positive constant D. Define the 2-norm of the q-dimensional vector $\boldsymbol{z} = (z_1, \cdots, z_q)^\top$ as $\|\boldsymbol{z}\|_2 = \sqrt{\sum_{j=1}^q z_j^2}$. To establish statistical properties of recovering the true subgroups, the following conditions are needed:

CONDITION B.1: $0 < \beta_1 < \min_{l=1,\ldots,K_0} \psi_{\min}^{(l)} < \max_{l=1,\ldots,K_0} \psi_{\max}^{(l)} < \beta_2$, for some positive constants β_1, β_2, where $\psi_{\min}^{(l)}$ and $\psi_{\max}^{(l)}$ are the smallest and largest eigenvalues of $\boldsymbol{\Theta}_l^*$, respectively.

CONDITION B.2: $\|\boldsymbol{\mu}^*\|_\infty := \max_{l=1,\ldots,K_0} \|\boldsymbol{\mu}_l^*\|_\infty$ and $\|\boldsymbol{\Theta}^*\|_\infty := \max_{l=1,\ldots,K_0} \|\boldsymbol{\Theta}_l^*\|_\infty$ are bounded, where $\|\boldsymbol{\mu}^*\|_\infty$ and $\|\boldsymbol{\Theta}^*\|_\infty$ denote the maximum absolute value of the elements of vector $\boldsymbol{\mu}^*$ and maximum absolute value of the row sums of matrix $\boldsymbol{\Theta}^*$, respectively.

CONDITION B.3: The K_0 clusters are sufficiently separable such that, for any $\boldsymbol{\Upsilon}$ in a α_0-neighborhood of $\boldsymbol{\Upsilon}^*$, that is, $\boldsymbol{\Upsilon} \in \mathcal{B}_{\alpha_0}(\boldsymbol{\Upsilon}^*) := \{\boldsymbol{\Upsilon} : \|\boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}^*\|_2 \leqslant \alpha_0\}$, and each pair $\{(l_1, l_2), 1 \leqslant l_1 \neq l_2 \leqslant K_0\}$,

$$\mathrm{pr}(\boldsymbol{x} \in \mathcal{A}_{l_1} | \boldsymbol{\Upsilon}) \cdot \mathrm{pr}(\boldsymbol{x} \in \mathcal{A}_{l_2} | \boldsymbol{\Upsilon}) \leqslant \frac{\varrho}{24(K_0 - 1)\sqrt{\max\{W, W', W''\}}},$$

where $\mathcal{A}_{l_1}$ is the l_1-th subgroup, $\varrho = c \cdot \min\{\beta_1, 0.5(\beta_2 + 2\alpha_0)^{-2}\}$ for a constant c, and the definitions of W, W', W'' are as in Condition 6 and Lemma S.1 of Hao et al. (2018).

CONDITION B.4: $\rho(t) = \lambda^{-1} p(t, \lambda)$ is concave in $t \in [0, \infty)$ with a continuous derivative $\rho'(t)$ satisfying $\rho(0+) = 1$, and $\rho'(0+)$ is independent of λ. There exists a constant $0 < a < \infty$ such that $\rho(t)$ is constant for all $t \geqslant a\lambda$.

CONDITION B.5: $K_0^2 = o\left(p(\log n)^{-1}\right)$, and $\min_{l=1,\ldots,K_0} \pi_l^* = O\left(\max_{l=1,\ldots,K_0} \pi_l^*\right)$.

Conditions B.1 and B.2 have been commonly assumed in relevant literature. Condition B.3 is the so-called sufficiently separable condition, which requires that if a sample belongs to a subgroup with probability close to 1, then it belongs to any other subgroup with probability close to 0. See Hao et al. (2018) for further discussions. The SCAD (Fan and Li, 2001) and MCP (Zhang, 2010) penalties both satisfy Condition B.4. Condition B.5 requires that the subgroups are not too imbalanced. K_0 is allowed to grow with the sample size n and dimension p, at a rate slower than $p(\log n)^{-1}$, which is also considered in Hao et al. (2018).

To prove Theorem 1, it is equivalent to establishing the following Results 1 and 2.

RESULT 1: *Suppose that Conditions B.1 - B.5 hold. Assume* $\left(K^2K_0^4(s+p)\log p\right)/n = o(1)$, $\lambda_1 \simeq \lambda_2 \simeq K_0^2\sqrt{\frac{(s+p)\log p}{n}}$, *and the minimal signal in the true parameters* $\boldsymbol{\Upsilon}^*$ *is larger than* $a \cdot max\{\lambda_1, \lambda_2\}$, *where* a *is defined in Condition B.4. Then, with probability tending to 1:*

1. The oracle estimator $\widehat{\boldsymbol{\Upsilon}}^o$ *satisfies:*

$$\|\widehat{\boldsymbol{\Upsilon}}^o - \boldsymbol{\Upsilon}^*\|_2 = O_p\left(K_0^2\left[\sqrt{d\log p/n} + \sqrt{(s+p)\log p/n}\right]\right).$$

2. Denote the set of the nonzero off-diagonal elements of $\widehat{\boldsymbol{\Theta}}_l^o$ *as* $\widehat{\mathcal{S}}_l^o$, *and the set of the nonzero elements of* $\widehat{\boldsymbol{\mu}}_l^o$ *as* $\widehat{\mathcal{D}}_l^o$. $\widehat{\mathcal{S}}_l^o = \mathcal{S}_l$, *and* $\widehat{\mathcal{D}}_l^o = \mathcal{D}_l$ *for* $l = 1, \ldots, K_0$.

RESULT 2: *Assume that* $\lambda_3 \gg KK_0^2\sqrt{\frac{(s+p)\log p}{n}}$, $b > a\lambda_3$, *and the conditions in Result 1 hold. Then there exists* $\widehat{\boldsymbol{\Omega}}$, *a local maximum of* $\mathcal{L}(\boldsymbol{\Omega}|\boldsymbol{X})$ *defined in (A.1), that satisfies*

$$P\left(\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Omega}}^o\right) \to 1.$$

We prove that $\widehat{\boldsymbol{\Upsilon}}^o$ obtained by the EM algorithm satisfies Result 1 in Section B.2. The proof of Result 2 is given in Section B.3.

B.2 *Proof of Result 1 in Theorem 1*

The objective function in (B.1) can be rewritten as:

$$\mathcal{Q}(\boldsymbol{\Upsilon}) = \frac{1}{n}\sum_{i=1}^{n}\log\left(\sum_{l=1}^{K_0}\pi_l f_l\left(\boldsymbol{x}_i;\boldsymbol{\Upsilon}_l\right)\right) - \lambda_1\mathcal{P}_1(\boldsymbol{\Upsilon}) - \lambda_2\mathcal{P}_2(\boldsymbol{\Upsilon}), \tag{B.4}$$

where π_l is the sum of the mixture probabilities labeled by $\mathcal{T}_l^*$, $\boldsymbol{\Upsilon} = (\boldsymbol{\Upsilon}_1^\top, \cdots, \boldsymbol{\Upsilon}_{K_0}^\top)^\top$, $\boldsymbol{\Upsilon}_l = \mathrm{vec}\left(\boldsymbol{\mu}_l, \boldsymbol{\Theta}_l\right) = (\mu_{l1}, \ldots, \mu_{lp}, \theta_{l11}, \ldots, \theta_{lp1}, \ldots, \theta_{l1p}, \ldots, \theta_{lpp}) \in \mathbb{R}^{p^2+p}, 1 \leqslant l \leqslant K_0$,

$$\mathcal{P}_1(\boldsymbol{\Upsilon}) = \sum_{k=1}^{K_0}\sum_{j=1}^{p}|\mathcal{T}_l^*|\rho(|\mu_{kj}|,\lambda_1), \mathcal{P}_2(\boldsymbol{\Upsilon}) = \sum_{k=1}^{K_0}\sum_{i\neq j}|\mathcal{T}_l^*|\rho(|\theta_{kij}|,\lambda_2),$$

and $\rho(t) = \lambda^{-1}p(t,\lambda)$ is defined in Condition B.4.

Denote $\boldsymbol{\gamma}' = (\gamma_{il})_{n\times K_0}$, where $\gamma_{il} = I(\boldsymbol{x}_i \in \mathcal{A}_l)$ is the latent indicator variable showing the component membership of the ith observation in the mixture. If γ_{ik} is available, the penalized log-likelihood function (B.4) for the complete data can be written as:

$$\mathcal{Q}(\boldsymbol{\Upsilon}|\boldsymbol{X},\boldsymbol{\gamma}') := \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{K_0}\gamma_{il}\left[\log\pi_l + \log f_l\left(\boldsymbol{x}_i;\boldsymbol{\Upsilon}_l\right)\right] - \lambda_1\mathcal{P}_1(\boldsymbol{\Upsilon}) - \lambda_2\mathcal{P}_2(\boldsymbol{\Upsilon}). \tag{B.5}$$

In the t-th expectation step of the EM algorithm, the conditional expectation of (B.5) is computed as:

$$E_{\boldsymbol{\gamma}'|\boldsymbol{X},\boldsymbol{\Upsilon}^{(t-1)}}[\mathcal{Q}(\boldsymbol{\Upsilon}|\boldsymbol{X},\boldsymbol{\gamma}')] = \mathcal{H}_n(\boldsymbol{\Upsilon}|\boldsymbol{\Upsilon}^{(t-1)}) - \lambda_1\mathcal{P}_1(\boldsymbol{\Upsilon}) - \lambda_2\mathcal{P}_2(\boldsymbol{\Upsilon}),$$

where

$$\mathcal{H}_n(\boldsymbol{\Upsilon}|\boldsymbol{\Upsilon}^{(t-1)}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{K_0}\gamma_{il}^{(t)}(\boldsymbol{\Upsilon}_l)\left[\log\pi_l + \log f_l\left(\boldsymbol{x}_i;\boldsymbol{\Upsilon}_l\right)\right]. \tag{B.6}$$

$\gamma_{il}^{(t)}(\boldsymbol{\Upsilon}_l)$ can be computed based on $\pi_l^{(t-1)}$, $\boldsymbol{\mu}_l^{(t-1)}$, and $\boldsymbol{\Theta}_l^{(t-1)}$ obtained in the previous iteration. To prove Result 1, a corresponding population version of $\mathcal{H}_n$ in (B.6) needs to be defined as:

$$\mathcal{H}(\boldsymbol{\Upsilon}|\boldsymbol{\Upsilon}') = E\left[\sum_{l=1}^{K_0}\gamma_{il}(\boldsymbol{\Upsilon}')\left[\log\pi_l + \log f_l\left(\boldsymbol{x}_i;\boldsymbol{\Upsilon}_l\right)\right]\right]. \tag{B.7}$$

Define the function evaluating the error between the iterative estimator $\boldsymbol{\Upsilon}$ and true

parameter $\boldsymbol{\Upsilon}^*$ as:

$$\begin{aligned} q(\boldsymbol{\upsilon}) =& \mathcal{H}_n(\boldsymbol{\Upsilon}^* + \boldsymbol{\upsilon}|\boldsymbol{\Upsilon}^{(t-1)}) - \mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) \\ & - \lambda_1[\mathcal{P}_1(\boldsymbol{\Upsilon}^* + \boldsymbol{\upsilon}) - \mathcal{P}_1(\boldsymbol{\Upsilon}^*)] - \lambda_2[\mathcal{P}_2(\boldsymbol{\Upsilon}^* + \boldsymbol{\upsilon}) - \mathcal{P}_2(\boldsymbol{\Upsilon}^*)]. \end{aligned}$$

First, we show that, if the estimate obtained in the $(t-1)$-th iteration of the EM algorithm $\boldsymbol{\Upsilon}^{(t-1)} \in \mathcal{B}_\alpha(\boldsymbol{\Upsilon}^*)$, where we set $\alpha = O\left(K_0^3(s+p)\sqrt{\log p/n}\right)$, then there is a conditional local maximizer $\boldsymbol{\Upsilon}^{(t)}$ in $\{\boldsymbol{\Upsilon} : \|\boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}^*\|_2 \leqslant \chi\}$, where $\chi = \frac{4\epsilon}{\varrho} + \iota\|\boldsymbol{\Upsilon}^{(t-1)} - \boldsymbol{\Upsilon}^*\|_2$, $\epsilon = CK_0^2\left(\sqrt{\frac{d\log p}{n}} + \sqrt{\frac{(s+p)\log p}{n}}\right)$ for a positive constant C, ϱ is defined in Condition B.3, and $1/6 \leqslant \iota < 1$ is a positive constant. It is important to note that $\alpha \gg \epsilon$, so the selection of α is reasonable. It suffices to show that $P\left(\sup_{\boldsymbol{\upsilon}\in\mathcal{C}(\chi)} q(\boldsymbol{\upsilon}) < 0\right) \to 1$, where $\mathcal{C}(\chi) := \{\boldsymbol{\upsilon} : \|\boldsymbol{\upsilon}\|_2 = \chi\}$.

Then we establish an upper bound for $q(\boldsymbol{\upsilon})$ over the set $\mathcal{C}(\chi)$. For a sufficiently large n, $4\epsilon/\varrho \leqslant (2-\iota)\alpha$. Note that $\|\boldsymbol{\Upsilon}^{(t-1)} - \boldsymbol{\Upsilon}^*\|_2 \leqslant \alpha$, so $\chi = \frac{4\epsilon}{\varrho} + \iota\|\boldsymbol{\Upsilon}^{(t-1)} - \boldsymbol{\Upsilon}^*\|_2 \leqslant 2\alpha$. So $\mathcal{C}(\chi) \subseteq \{\boldsymbol{\upsilon} : \|\boldsymbol{\upsilon}\|_2 \leqslant 2\alpha\}$. Then, according to Lemma 9 and Lemma S.1 of Hao et al. (2018):

$$\mathcal{H}_n(\boldsymbol{\Upsilon}^* + \boldsymbol{\upsilon}|\boldsymbol{\Upsilon}^{(t-1)}) - \mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) \leqslant \langle \nabla\mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}), \boldsymbol{\upsilon}\rangle - \frac{\varrho}{2}\|\boldsymbol{\upsilon}\|_2^2,$$

with probability at least $1 - 1/p$. Then

$$\begin{aligned} q(\boldsymbol{\upsilon}) \leqslant & \langle \nabla\mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}), \boldsymbol{\upsilon}\rangle - \lambda_1[\mathcal{P}_1(\boldsymbol{\Upsilon}^* + \boldsymbol{\upsilon}) - \mathcal{P}_1(\boldsymbol{\Upsilon}^*)] \\ & - \lambda_2[\mathcal{P}_2(\boldsymbol{\Upsilon}^* + \boldsymbol{\upsilon}) - \mathcal{P}_2(\boldsymbol{\Upsilon}^*)] - \frac{\varrho}{2}\|\boldsymbol{\upsilon}\|_2^2 \\ = & \langle \nabla\mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla\mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) + \nabla\mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla\mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^*), \boldsymbol{\upsilon}\rangle \\ & - \lambda_1[\mathcal{P}_1(\boldsymbol{\Upsilon}^* + \boldsymbol{\upsilon}) - \mathcal{P}_1(\boldsymbol{\Upsilon}^*)] - \lambda_2[\mathcal{P}_2(\boldsymbol{\Upsilon}^* + \boldsymbol{\upsilon}) - \mathcal{P}_2(\boldsymbol{\Upsilon}^*)] - \frac{\varrho}{2}\|\boldsymbol{\upsilon}\|_2^2 \\ = & q_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C}, \boldsymbol{\upsilon}_{\mathcal{M}^C}) + q_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}}, \boldsymbol{\upsilon}_{\mathcal{M}}) + q_2(\boldsymbol{\Upsilon}^*, \boldsymbol{\upsilon}), \end{aligned} \tag{B.8}$$

where the first equality is from the self-consistency property of the population version of the objective function (McLachlan and Krishnan, 2007), that is $\boldsymbol{\Upsilon}^* = \text{argmax}_{\boldsymbol{\Upsilon}'}\mathcal{H}(\boldsymbol{\Upsilon}'|\boldsymbol{\Upsilon}^*)$. By

the definition of $\mathcal{M}$ and $\mathcal{M}^C$ in (B.3),

$$q_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C}, \boldsymbol{\upsilon}_{\mathcal{M}^C}) = \langle \nabla_{\boldsymbol{\Upsilon}^*_{\mathcal{M}^C}} \mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla_{\boldsymbol{\Upsilon}^*_{\mathcal{M}^C}} \mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}), \boldsymbol{\upsilon}_{\mathcal{M}^C} \rangle$$
$$- \lambda_1[\mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C} + \boldsymbol{\upsilon}_{\mathcal{M}^C}) - \mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C})] - \lambda_2[\mathcal{P}_2(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C} + \boldsymbol{\upsilon}_{\mathcal{M}^C}) - \mathcal{P}_2(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C})],$$
$$q_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}}, \boldsymbol{\upsilon}_{\mathcal{M}}) = \langle \nabla_{\boldsymbol{\Upsilon}^*_{\mathcal{M}}} \mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla_{\boldsymbol{\Upsilon}^*_{\mathcal{M}}} \mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}), \boldsymbol{\upsilon}_{\mathcal{M}} \rangle$$
$$- \lambda_1[\mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}} + \boldsymbol{\upsilon}_{\mathcal{M}}) - \mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}})] - \lambda_2[\mathcal{P}_2(\boldsymbol{\Upsilon}^*_{\mathcal{M}} + \boldsymbol{\upsilon}_{\mathcal{M}}) - \mathcal{P}_2(\boldsymbol{\Upsilon}^*_{\mathcal{M}})],$$
$$q_2(\boldsymbol{\Upsilon}^*, \boldsymbol{\upsilon}) = \langle \nabla \mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla \mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^*), \boldsymbol{\upsilon} \rangle - \frac{\varrho}{2} \|\boldsymbol{\upsilon}\|_2^2.$$

Consider $q_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C}, \boldsymbol{\upsilon}_{\mathcal{M}^C})$. Note that

$$\lambda_1[\mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C} + \boldsymbol{\upsilon}_{\mathcal{M}^C}) - \mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C})] = \lambda_1 \mathcal{P}_1(\boldsymbol{\upsilon}_{\mathcal{M}^C}) = \sum_{l=1}^{K_0} \sum_{j \in \mathcal{M}^C_{\mathcal{D}_l}} \lambda_1 |\mathcal{T}^*_l| \rho(|\upsilon_{lj}|, \lambda_1).$$

By Condition B.2, if $|\upsilon_{lj}| \leqslant a\lambda_1$, we have $\lambda_1|\mathcal{T}^*_l|\rho(|\upsilon_{lj}|, \lambda_1) \geqslant \lambda_1 C_{\lambda_1}|\mathcal{T}_{\min}||\upsilon_{lj}|$ for a constant $C_{\lambda_1} > 0$, and if $|\upsilon_{lj}| > a\lambda_1$, $\lambda_1|\mathcal{T}^*_l|\rho(|\upsilon_{lj}|, \lambda_1) \geqslant \frac{1}{2}a\lambda_1^2|\mathcal{T}_{\min}|$. The inequalities are similar for $|\upsilon_{lj}| \leqslant a\lambda_2$ and $|\upsilon_{lj}| > a\lambda_2$. So

$$\begin{aligned}
&q_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C}, \boldsymbol{\upsilon}_{\mathcal{M}^C}) \\
&\leqslant \sum_{l=1}^{K_0} \sum_{j \in \mathcal{M}^C_{\mathcal{D}_l}} \left(\|\nabla \mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla \mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)})\|_\infty - \lambda_1 C_{\lambda_1}|\mathcal{T}_{\min}| \right) |\upsilon_{lj}| \cdot I(|\upsilon_{lj}| \leqslant a\lambda_1) \\
&+ \sum_{l=1}^{K_0} \sum_{j \in \mathcal{M}^C_{\mathcal{D}_l}} \left(\|\nabla \mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla \mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)})\|_\infty \|\upsilon\|_\infty - \frac{1}{2}a\lambda_1^2|\mathcal{T}_{\min}| \right) \cdot I(|\upsilon_{lj}| > a\lambda_1) \\
&+ \sum_{l=1}^{K_0} \sum_{j \in \mathcal{M}^C_{\mathcal{S}_l}} \left(\|\nabla \mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla \mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)})\|_\infty - \lambda_2 C_{\lambda_2}|\mathcal{T}_{\min}| \right) |\upsilon_{lj}| \cdot I(|\upsilon_{lj}| \leqslant a\lambda_2) \\
&+ \sum_{l=1}^{K_0} \sum_{j \in \mathcal{M}^C_{\mathcal{S}_l}} \left(\|\nabla \mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla \mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)})\|_\infty \|\upsilon\|_\infty - \frac{1}{2}a\lambda_2^2|\mathcal{T}_{\min}| \right) \cdot I(|\upsilon_{lj}| > a\lambda_2).
\end{aligned}$$

According to the proof of Lemma S.1 in Hao et al. (2018) and Condition B.2, we have $\|\nabla \mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla \mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)})\|_\infty = O_p\left(K_0\sqrt{\log p/n}\right)$. Note that $\|\upsilon\|_\infty \leqslant \frac{4\epsilon}{\varrho} + \iota\alpha$, $\alpha = O\left(K_0^3(s+p)\sqrt{\log p/n}\right)$, and $\lambda_1 \simeq \lambda_2 \simeq K_0^2\sqrt{(s+p)\log p/n}$. Therefore,

$$q_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}^C}, \boldsymbol{\upsilon}_{\mathcal{M}^C}) < 0. \tag{B.9}$$

As for $q_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}}, \boldsymbol{v}_{\mathcal{M}})$, denote $u_{\mathcal{M}}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) = \nabla_{\boldsymbol{\Upsilon}^*_{\mathcal{M}}}\mathcal{H}_n(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla_{\boldsymbol{\Upsilon}^*_{\mathcal{M}}}\mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)})$ for simplicity.

Then by the definition of $\mathcal{G}$ and $\mathcal{G}^{\mathcal{C}}$ in (B.2), $\langle u_{\mathcal{M}}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}), \boldsymbol{v}_{\mathcal{M}}\rangle$ can be further decomposed:

$$\begin{aligned}\langle u_{\mathcal{M}}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}), \boldsymbol{v}_{\mathcal{M}}\rangle &\leqslant |\langle u_{\mathcal{M}\cap\mathcal{G}^C}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}), \boldsymbol{v}_{\mathcal{M}\cap\mathcal{G}^C}\rangle| + |\langle u_{\mathcal{M}\cap\mathcal{G}}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}), \boldsymbol{v}_{\mathcal{M}\cap\mathcal{G}}\rangle| \\ &\leqslant \|u_{\mathcal{M}\cap\mathcal{G}^C}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)})\|_\infty\|\boldsymbol{v}_{\mathcal{M}\cap\mathcal{G}^C}\|_1 + \|u_{\mathcal{M}\cap\mathcal{G}}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)})\|_2\|\boldsymbol{v}_{\mathcal{M}\cap\mathcal{G}}\|_2 \\ &\leqslant \epsilon_1\sqrt{K_0(d+s)}\|\boldsymbol{v}_{\mathcal{M}\cap\mathcal{G}^C}\|_2 + \epsilon_2\|\boldsymbol{v}_{\mathcal{M}\cap\mathcal{G}}\|_2 \\ &\leqslant \left(\epsilon_1\sqrt{K_0(d+s)} + \epsilon_2\right)\|\boldsymbol{v}\|_2,\end{aligned} \tag{B.10}$$

with probability at least $1-(26K_0^2+8K_0)/p$, where $\epsilon_1 = C_\epsilon\sqrt{K_0^3\log p/n}$, $\epsilon_2 = C_\epsilon\sqrt{K_0^3 p\log p/n}$, for a positive constant C_ϵ. The second inequality is from the Holder's inequality. The third inequality is from Lemma S.1 in Hao et al. (2018), Conditions B.2, B.5, and the Cauchy-Schwarz inequality, that is, $\|u_{\mathcal{G}^C}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)})\|_\infty \leqslant \epsilon_1$, $\|u_{\mathcal{G}}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)})\|_2 \leqslant \epsilon_2$, and $\|\boldsymbol{v}_{\mathcal{M}\cap\mathcal{G}^C}\|_1 \leqslant \sqrt{K_0(d+s)}\|\boldsymbol{v}_{\mathcal{M}\cap\mathcal{G}^C}\|_2$. In addition,

$$\begin{aligned}-\lambda_1[\mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}} + \boldsymbol{v}_{\mathcal{M}}) - \mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}})] &\leqslant \lambda_1|\mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}} + \boldsymbol{v}_{\mathcal{M}}) - \mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}})| \\ &\leqslant \lambda_1 C'_{\mathcal{P}_1}|\nabla\mathcal{P}_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}})^\top\boldsymbol{v}_{\mathcal{M}}| = 0,\end{aligned} \tag{B.11}$$

for a positive constant $C'_{\mathcal{P}_1}$. The last equality is from the minimal signal condition of the true parameters and Condition B.4. Similarly,

$$-\lambda_2[\mathcal{P}_2(\boldsymbol{\Upsilon}^*_{\mathcal{M}} + \boldsymbol{v}_{\mathcal{M}}) - \mathcal{P}_2(\boldsymbol{\Upsilon}^*_{\mathcal{M}})] \leqslant 0. \tag{B.12}$$

By (B.10), (B.11), and (B.12), we have

$$q_1(\boldsymbol{\Upsilon}^*_{\mathcal{M}}, \boldsymbol{v}_{\mathcal{M}}) \leqslant \left(\epsilon_1\sqrt{K_0(d+s)} + \epsilon_2\right)\|\boldsymbol{v}\|_2. \tag{B.13}$$

As for $q_2(\boldsymbol{\Upsilon}^*, \boldsymbol{v})$, note that $\boldsymbol{\Upsilon}^{(t-1)} \in \mathcal{B}_\alpha(\boldsymbol{\Upsilon}^*)$. So according to Lemma 7 of Hao et al. (2018) and Condition B.3,

$$\begin{aligned}|\langle\nabla\mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla\mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^*), \boldsymbol{v}\rangle| &\leqslant \|\nabla\mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^{(t-1)}) - \nabla\mathcal{H}(\boldsymbol{\Upsilon}^*|\boldsymbol{\Upsilon}^*)\|_2\|\boldsymbol{v}\|_2 \\ &\leqslant \tau\cdot\|\boldsymbol{\Upsilon}^{(t-1)} - \boldsymbol{\Upsilon}^*\|_2\|\boldsymbol{v}\|_2,\end{aligned}$$

where $\tau \leqslant \varrho/12$. Then,

$$q_2(\boldsymbol{\Upsilon}^*, \boldsymbol{\upsilon}) \leqslant -\frac{\varrho}{2}\|\boldsymbol{\upsilon}\|_2^2 + \tau \cdot \|\boldsymbol{\Upsilon}^{(t-1)} - \boldsymbol{\Upsilon}^*\|_2 \|\boldsymbol{\upsilon}\|_2. \tag{B.14}$$

By (B.8), (B.9), (B.13), and (B.14), an upper bound of $q(\boldsymbol{\upsilon})$ can be established as:

$$q(\boldsymbol{\upsilon}) < -\frac{\varrho}{2}\|\boldsymbol{\upsilon}\|_2^2 + \left(\epsilon_1\sqrt{K_0(d+s)} + \epsilon_2 + \tau \cdot \|\boldsymbol{\Upsilon}^{(t-1)} - \boldsymbol{\Upsilon}^*\|_2\right)\|\boldsymbol{\upsilon}\|_2,$$

with probability at least $1 - (26K_0^2 + 8K_0 + 1)/p$.

Note that $\epsilon = CK_0^2\left(\sqrt{\frac{d\log p}{n}} + \sqrt{\frac{(s+p)\log p}{n}}\right)$. So for a properly chosen positive constant C, we have

$$\epsilon_1\sqrt{K_0(d+s)} + \epsilon_2 \leqslant \epsilon.$$

Then, an upper bound of $q(\boldsymbol{\upsilon})$ can be obtained as:

$$q(\boldsymbol{\upsilon}) < -\frac{\varrho}{2}\|\boldsymbol{\upsilon}\|_2^2 + \left(\epsilon + \tau \cdot \|\boldsymbol{\Upsilon}^{(t-1)} - \boldsymbol{\Upsilon}^*\|_2\right)\|\boldsymbol{\upsilon}\|_2,$$

with probability at least $1 - (26K_0^2 + 8K_0 + 1)/p$.

Taking advantage of properties of quadratic functions, when $\|\boldsymbol{\upsilon}\|_2 > \frac{2\epsilon}{\varrho} + \frac{2\tau}{\varrho}\|\boldsymbol{\Upsilon}^{(t-1)} - \boldsymbol{\Upsilon}^*\|_2$, $q(\boldsymbol{\upsilon}) < 0$. Note that $\|\boldsymbol{\upsilon}\|_2 = \chi = \frac{4\epsilon}{\varrho} + \iota\|\boldsymbol{\Upsilon}^{(t-1)} - \boldsymbol{\Upsilon}^*\|_2$, $1/6 \leqslant \iota < 1$, and $\tau \leqslant \varrho/12$. Therefore, there is a conditional local maximizer $\boldsymbol{\Upsilon}^{(t)}$ that follows $\boldsymbol{\Upsilon}^{(t-1)}$ and satisfies:

$$\text{if } \boldsymbol{\Upsilon}^{(t-1)} \in \mathcal{B}_\alpha\left(\boldsymbol{\Upsilon}^*\right), \text{then } \|\boldsymbol{\Upsilon}^{(t)} - \boldsymbol{\Upsilon}^*\|_2 \leqslant \chi, \tag{B.15}$$

with probability at least $1 - (26K_0^2 + 8K_0 + 1)/p$.

In addition, for a sufficiently large n, $\epsilon \leqslant (1-\iota)\varrho\alpha/4$. And note that $\boldsymbol{\Upsilon}^{(t-1)} \in \mathcal{B}_\alpha\left(\boldsymbol{\Upsilon}^*\right)$. Then we have $\|\boldsymbol{\Upsilon}^{(t)} - \boldsymbol{\Upsilon}^*\|_2 \leqslant \chi \leqslant (1-\iota)\alpha + \iota\alpha = \alpha$. That is:

$$\boldsymbol{\Upsilon}^{(t)} \in \mathcal{B}_\alpha\left(\boldsymbol{\Upsilon}^*\right). \tag{B.16}$$

With (B.15) and (B.16), it can be obtained that if $\boldsymbol{\Upsilon}^{(0)} \in \mathcal{B}_\alpha\left(\boldsymbol{\Upsilon}^*\right)$, for any $t \geqslant 1$,

$$\|\boldsymbol{\Upsilon}^{(t)} - \boldsymbol{\Upsilon}^*\|_2 \leqslant \frac{1-\iota^t}{1-\iota}\frac{4\epsilon}{\varrho} + \iota^t\|\boldsymbol{\Upsilon}^{(0)} - \boldsymbol{\Upsilon}^*\|_2 \leqslant \frac{8\epsilon}{\varrho} + \iota^t\|\boldsymbol{\Upsilon}^{(0)} - \boldsymbol{\Upsilon}^*\|_2, \tag{B.17}$$

with probability at least $1 - t(26K_0^2 + 8K_0 + 1)/p$.

Note that, when $t \geqslant \hat{t} := \log_{1/\iota}\left(\frac{\varrho\|\boldsymbol{\Upsilon}^{(0)} - \boldsymbol{\Upsilon}^*\|_2}{8\epsilon}\right)$, $\iota^t\|\boldsymbol{\Upsilon}^{(0)} - \boldsymbol{\Upsilon}^*\|_2$ is dominated by $8\epsilon/\varrho$. So the

final error of $\widehat{\boldsymbol{\Upsilon}}$ can be bounded as:

$$\|\widehat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}^*\|_2 \leqslant 16\epsilon/\varrho = O\left(K_0^2\sqrt{\frac{d\log p}{n}} + K_0^2\sqrt{\frac{(s+p)\log p}{n}}\right), \tag{B.18}$$

with probability at least $1 - \hat{t}(26K_0^2 + 8K_0 + 1)/p$. And note that $\hat{t}(26K_0^2 + 8K_0 + 1)/p$ goes to 0 as p and n diverge by Condition B.5. This concludes the proof of part 1 of Result 1.

Next, we prove the variable selection consistency of the final precision matrix estimators. In Step 1, $\widehat{\mathcal{S}}_l^o \supset \mathcal{S}_l$ is established. Then, it is sufficient to show that for any $(i,j) \in \mathcal{S}_l$ with $l = 1, \ldots, K_0$, $\hat{\theta}_{lij} \neq 0$. We have:

$$\left|\hat{\theta}_{lij}\right| \geqslant \left|\theta^*_{lij}\right| - \left|\hat{\theta}_{lij} - \theta^*_{lij}\right| \geqslant \left|\theta^*_{lij}\right| - \sqrt{\sum_{1\leqslant i,j\leqslant p}(\hat{\theta}_{lij} - \theta^*_{lij})^2} \geqslant \left|\theta^*_{lij}\right| - \|\widehat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}^*\|_2.$$

According to the minimal signal condition of the true parameters and (B.18), we have $|\hat{\theta}_{lij}| > 0$, which implies $\widehat{\mathcal{S}}_l^o \supset \mathcal{S}_l$.

In Step 2, $\widehat{\mathcal{S}}_l^o \subset \mathcal{S}_l$ is established. It is sufficient to show that for any $(j,m) \in \mathcal{S}_l^C$, $\hat{\theta}_{ljm} = 0$. Inspired by Lam and Fan (2009), we consider a local maximum point $\widehat{\boldsymbol{\Upsilon}}$ that satisfies (B.18) and optimizes the objective function (B.4). The derivative of $\mathcal{Q}(\widehat{\boldsymbol{\Upsilon}})$ with respect to θ_{ljm} for $(j,m) \in \mathcal{S}_l^C, l = 1, \cdots, K_0$ is:

$$\frac{\partial \mathcal{Q}(\widehat{\boldsymbol{\Upsilon}})}{\partial \theta_{ljm}} = \mathcal{R}(\widehat{\boldsymbol{\Upsilon}}_l) - \lambda_2 \rho'(|\hat{\theta}_{ljm}|)\text{sgn}\ (\hat{\theta}_{ljm}),$$

where $\mathcal{R}(\widehat{\boldsymbol{\Upsilon}}_l) = \frac{1}{2n}\sum_{i=1}^n r(\boldsymbol{x}_i; \widehat{\boldsymbol{\Upsilon}}_l)[\hat{\sigma}_{ljm} - (x_{ij} - \hat{\mu}_{lj})(x_{im} - \hat{\mu}_{lm})]$, $r(\boldsymbol{x}_i; \widehat{\boldsymbol{\Upsilon}}_l) = \frac{\hat{\pi}_l f_l(\boldsymbol{x}_i; \widehat{\boldsymbol{\Upsilon}}_l)}{\sum_{l=1}^{K_0}\hat{\pi}_l f_l(\boldsymbol{x}_i; \widehat{\boldsymbol{\Upsilon}}_l)}$, $\hat{\sigma}_{ljm}$ is the (j,m)th element of $\widehat{\boldsymbol{\Theta}}_l^{-1}$, and sgn $(\hat{\theta}_{ljm})$ denotes the sign of $\hat{\theta}_{ljm}$. We need to prove that $\hat{\theta}_{ljm} = 0$ for all $(j,m) \in \mathcal{S}_l^C, l = 1, \cdots, K_0$ with probability tending to 1, for which it suffices to show that the sign of $\frac{\partial \mathcal{Q}(\widehat{\boldsymbol{\Upsilon}})}{\partial \theta_{ljm}}$ only depends on sgn $(\hat{\theta}_{ljm})$ with probability tending to 1, if $\hat{\theta}_{ljm}$ lies in a small neighborhood of 0. It is necessary to estimate the upper bound of the order of $\mathcal{R}(\widehat{\boldsymbol{\Upsilon}}_l)$ independent of l.

It can be decomposed that:

$$\mathcal{R}(\widehat{\boldsymbol{\Upsilon}}_l) \leqslant |\hat{\sigma}_{ljm} - \sigma^*_{ljm}| + \mathcal{R}^*(\widehat{\boldsymbol{\Upsilon}}_l), \tag{B.19}$$

where σ^*_{ljm} denotes the true value of σ_{ljm}, and $\mathcal{R}^*(\widehat{\boldsymbol{\Upsilon}}_l) = \frac{1}{2n}\sum_{i=1}^n r(\boldsymbol{x}_i; \widehat{\boldsymbol{\Upsilon}}_l)[\sigma^*_{ljm} - (x_{ij} -$

$\hat{\mu}_{lj})(x_{im} - \hat{\mu}_{lm})]$. First, we consider

$$\begin{aligned} |\hat{\sigma}_{ljm} - \sigma^*_{ljm}| &\leqslant \|\widehat{\boldsymbol{\Theta}}_l^{-1} - (\boldsymbol{\Theta}^*)_l^{-1}\|_2 \leqslant \|\widehat{\boldsymbol{\Theta}}_l^{-1}(\boldsymbol{\Theta}^*_l - \widehat{\boldsymbol{\Theta}}_l)(\boldsymbol{\Theta}^*_l)^{-1}\|_2 \\ &\leqslant \|\widehat{\boldsymbol{\Theta}}_l^{-1}\|_2 \cdot \|\boldsymbol{\Theta}^*_l - \widehat{\boldsymbol{\Theta}}_l\|_2 \cdot \|(\boldsymbol{\Theta}^*_l)^{-1}\|_2, \\ &\leqslant \|\boldsymbol{\Theta}^*_l - \widehat{\boldsymbol{\Theta}}_l\|_2, \end{aligned} \tag{B.20}$$

where the 2-norm of a matrix is defined as $\|A\|_2 = \psi^{1/2}_{\max}(A^\top A)$, $\psi(A)$ denotes the eigenvalue of matrix A, and the last inequality is from $\|(\boldsymbol{\Theta}^*_l)^{-1}\|_2 = O(1)$, and $\|\widehat{\boldsymbol{\Theta}}_l^{-1}\|_2 = \psi^{-1}_{\min}(\widehat{\boldsymbol{\Theta}}_l) \leqslant \left(\psi_{\min}(\boldsymbol{\Theta}^*_l) + \psi_{\min}(\widehat{\boldsymbol{\Theta}}_l - \boldsymbol{\Theta}^*_l)\right)^{-1} = O(1)$ according to Condition B.1 and (B.18). Note that $\|\boldsymbol{\Theta}^*_l - \widehat{\boldsymbol{\Theta}}_l\|_2 \leqslant \|\widehat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}^*\|_2$, so

$$|\hat{\sigma}_{ljm} - \sigma^*_{ljm}| \lesssim K_0^2\sqrt{(s+p)\log p/n}. \tag{B.21}$$

Next, we examine the upper bound of the order of $\mathcal{R}^*(\widehat{\boldsymbol{\Upsilon}}_l)$:

$$|\mathcal{R}^*(\widehat{\boldsymbol{\Upsilon}}_l)| \leqslant \frac{1}{2}|\sigma^*_{ljm} - \mathcal{R}^*_1| + \frac{1}{2}|\mathcal{R}^*_2|, \tag{B.22}$$

where

$$\begin{aligned} \mathcal{R}^*_1 &= \frac{1}{n}\sum_{i=1}^n r(\boldsymbol{x}_i; \boldsymbol{\Upsilon}^*_l)[(x_{ij} - \mu^*_{lj})(x_{im} - \mu^*_{lm})], \\ \mathcal{R}^*_2 &= \frac{1}{n}\sum_{i=1}^n [r(\boldsymbol{x}_i; \widehat{\boldsymbol{\Upsilon}}_l) - r(\boldsymbol{x}_i; \boldsymbol{\Upsilon}^*_l)][(x_{ij} - \hat{\mu}_{lj})(x_{im} - \hat{\mu}_{lm})] \\ &+ \frac{1}{n}\sum_{i=1}^n r(\boldsymbol{x}_i; \boldsymbol{\Upsilon}^*_l)[(x_{ij} - \mu^*_{lj})(\mu^*_{lm} - \hat{\mu}_{lm}) + (x_{im} - \mu^*_{lm})(\mu^*_{lj} - \hat{\mu}_{lj}) + (\mu^*_{lm} - \hat{\mu}_{lm})(\mu^*_{lj} - \hat{\mu}_{lj})]. \end{aligned}$$

Note that $r(\boldsymbol{x}_i; \cdot)$ is a continuous function. Then,

$$\begin{aligned} \mathcal{R}^*_2 &= O\left(\sup_{i,l}\{r(\boldsymbol{x}_i; \widehat{\boldsymbol{\Upsilon}}_l) - r(\boldsymbol{x}_i; \boldsymbol{\Upsilon}^*_l)\} + \sup_{l,j}\{|\mu^*_{lj} - \hat{\mu}_{lj}| + \sup_{l,j,m}\{|\mu^*_{lj} - \hat{\mu}_{lj}| \cdot |\mu^*_{lm} - \hat{\mu}_{lm}|\}\right) \\ &\lesssim \|\widehat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}^*\|_2 \lesssim K_0^2\sqrt{(s+p)\log p/n}. \end{aligned} \tag{B.23}$$

As for $\mathcal{R}^*_1$, note that $r(\boldsymbol{x}_i; \boldsymbol{\Upsilon}^*_l) = \text{pr}(\boldsymbol{x}_i \in \mathcal{A}_l | \boldsymbol{\Upsilon}^*) = E[I(\boldsymbol{x}_i \in \mathcal{A}_l)]$, where $\mathcal{A}_l$ is the l-th subgroup. So $\sigma^*_{ljm} - \mathcal{R}^*_1$ can be decomposed as:

$$\begin{aligned} \sigma^*_{ljm} - \mathcal{R}^*_1 &= \sigma^*_{ljm} - \frac{1}{n}\sum_{i=1}^n I(\boldsymbol{x}_i \in \mathcal{A}_l)[(x_{ij} - \mu^*_{lj})(x_{im} - \mu^*_{lm})] \\ &+ \frac{1}{n}\sum_{i=1}^n \left(I(\boldsymbol{x}_i \in \mathcal{A}_l) - E[I(\boldsymbol{x}_i \in \mathcal{A}_l)]\right)[(x_{ij} - \mu^*_{lj})(x_{im} - \mu^*_{lm})]. \end{aligned} \tag{B.24}$$

Note that $\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{x}_i \in \mathcal{A}_l)[(x_{ij}-\mu_{lj}^*)(x_{im}-\mu_{lm}^*)]$ is the estimated covariance when the mean is known for the l-th subgroup. So according to Lemma 2 of Lam and Fan (2009), we have:

$$\max_{l,j,m}\left|\sigma_{ljm}^* - \frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{x}_i \in \mathcal{A}_l)[(x_{ij}-\mu_{lj}^*)(x_{im}-\mu_{lm}^*)]\right| = O_p(\sqrt{\frac{\log p}{n}}). \tag{B.25}$$

In addition, from the Hoeffding's inequality,

$$\text{pr}\left(\frac{1}{n}\sum_{i=1}^{n}|I(\boldsymbol{x}_i \in \mathcal{A}_l) - E[I(\boldsymbol{x}_i \in \mathcal{A}_l)]| > \sqrt{\frac{\log p}{n}}\right) \leqslant \frac{2}{p\exp(2)}. \tag{B.26}$$

Combining (B.24), (B.25), (B.26), we have:

$$|\sigma_{ljm}^* - \mathcal{R}_1^*| \lesssim \sqrt{\frac{\log p}{n}}, \tag{B.27}$$

with probability at least $1-\frac{2}{p\exp(2)}$.

Then, with (B.19), (B.21), (B.22), (B.23), and (B.27), it can be derived that with probability tending to 1,

$$|\mathcal{R}(\widehat{\boldsymbol{\Upsilon}}_l)| \lesssim K_0^2\sqrt{(s+p)\log p/n}.$$

By Condition B.4, we have:

$$\lambda_2\rho'(|\hat{\theta}_{ljm}|) = C_\lambda K_0^2\sqrt{(s+p)\log p/n},$$

for $\hat{\theta}_{ljm}$ in a small neighborhood of 0 and some positive constant C_λ. Therefore, if $\hat{\theta}_{ljm}$ lies in a small neighborhood of 0, the sign of $\frac{\partial\mathcal{Q}(\widehat{\boldsymbol{\Upsilon}})}{\partial\theta_{ljm}}$ only depends on sgn $(\hat{\theta}_{ljm})$, independent of l, with probability tending to 1. Then, we can obtain the variable selection consistency of the final precision matrix estimators.

The variable selection consistency of the final mean vector estimators is similar. This concludes the proof of part 2 of Result 1.

B.3 *Proofs of Result 2 in Theorem 1*

Define

$$\mathcal{L}(\boldsymbol{\Omega}) = \mathcal{Q}(\boldsymbol{\Omega}) + \mathcal{P}_3(\boldsymbol{\Omega}), \mathcal{L}^{\mathcal{T}}(\boldsymbol{\Omega}) = \mathcal{Q}^{\mathcal{T}}(\boldsymbol{\Omega}) + \mathcal{P}_3^{\mathcal{T}}(\boldsymbol{\Omega}),$$

where

$$\mathcal{Q}(\mathbf{\Omega}) = \frac{1}{n}\sum_{i=1}^{n}\log\left(\sum_{k=1}^{K}\pi_k f_k\left(\boldsymbol{x}_i;\mathbf{\Omega}_k\right)\right) - \sum_{k=1}^{K}\sum_{j=1}^{p}p(|\mu_{kj}|,\lambda_1) - \sum_{k=1}^{K}\sum_{i\neq j}p(|\theta_{kij}|,\lambda_2),$$

$$\mathcal{P}_3(\mathbf{\Omega}) = \sum_{1\leqslant k<k'\leqslant K} p\left(\|\mathbf{\Omega}_k - \mathbf{\Omega}_{k'}\|_2,\lambda_3\right),$$

$$\mathcal{Q}^{\mathcal{T}}(\mathbf{\Omega}) = \frac{1}{n}\sum_{i=1}^{n}\log\left(\sum_{k=1}^{K}\pi_k f_k\left(\boldsymbol{x}_i;\mathbf{\Omega}_k\right)\right) - \sum_{k=1}^{K}\sum_{j=1}^{p}p(|\mu_{kj}|,\lambda_1) - \sum_{k=1}^{K}\sum_{i\neq j}p(|\theta_{kij}|,\lambda_2),\ \text{s.t. } \mathbf{\Omega}\in\Lambda_{\mathcal{T}^*},$$

$$\mathcal{P}_3^{\mathcal{T}}(\mathbf{\Omega}) = \sum_{1\leqslant k<k'\leqslant K} p\left(\|\mathbf{\Omega}_k - \mathbf{\Omega}_{k'}\|_2,\lambda_3\right),\ \text{s.t. } \mathbf{\Omega}\in\Lambda_{\mathcal{T}^*}.$$

Let $G:\Lambda_{\mathcal{T}^*}\to\mathbb{R}^{K_0(p+p^2)}$ be the mapping such that $G(\mathbf{\Omega})$ is the $K_0(p+p^2)\times 1$ vector consisting of K_0 vectors with dimension $p+p^2$, and its lth vector component equals the common value of $\mathbf{\Omega}_k$ for $k\in\mathcal{T}_l^*$. Let $\breve{G}:\mathbb{R}^{K(p+p^2)}\to\mathbb{R}^{K_0(p+p^2)}$ be the mapping such that $\breve{G}(\mathbf{\Omega}) = \{|\mathcal{T}_l^*|^{-1}\sum_{k\in\mathcal{T}_l^*}\mathbf{\Omega}_k^\top, l=1,\cdots,K_0\}^\top$. For any $\mathbf{\Omega}\in\mathbb{R}^{K(p+p^2)}$, denote $\breve{\mathbf{\Omega}} = G^{-1}\left(\breve{G}(\mathbf{\Omega})\right)$. Clearly, $\breve{\mathbf{\Omega}}\in\Lambda_{\mathcal{T}^*}$, and

$$\mathcal{L}(\breve{\mathbf{\Omega}}) = \mathcal{L}^{\mathcal{T}}(\breve{\mathbf{\Omega}}). \tag{B.28}$$

The locations of the parameters labeled by $\left\{j:\mu_{kj}^*\neq 0, 1\leqslant j\leqslant p\right\}$ in the true parameters of the k-th subgroup $\mathbf{\Omega}_k^*$ are indexed by $\mathcal{U}_k$. The locations of the parameters labeled by $\left\{(i,j):\theta_{kij}^*\neq 0, 1\leqslant i,j\leqslant p\right\}$ in $\mathbf{\Omega}_k^*$ are indexed by $\mathcal{V}_k$, $k=1,\cdots,K$. Define the index set of the nonzero components of $\mathbf{\Omega}_k^*$ as $\mathcal{W}_k = \{j:j\in\mathcal{U}_k\}\cup\{ip+j:(i,j)\in\mathcal{V}_k\}$. And note that there are estimated parameters for K subgroups in $\widehat{\mathbf{\Omega}}^o$, which correspond to K_0 subgroups in $\widehat{\mathbf{\Upsilon}}$. So

$$\|\widehat{\mathbf{\Omega}}^o - \mathbf{\Omega}^*\|_2 = O_p\left(K\|\widehat{\mathbf{\Upsilon}}^o - \mathbf{\Upsilon}^*\|_2\right) = O_p\left(\epsilon_n\right), \tag{B.29}$$

where $\epsilon_n = KK_0^2\left[\sqrt{d\log p/n}+\sqrt{(s+p)\log p/n}\right]$.

Consider the neighborhood of $\mathbf{\Omega}^*$:

$$\mathcal{C} = \{\mathbf{\Omega}\in\mathbb{R}^{K(p+p^2)}:\sup_k\|\mathbf{\Omega}_k - \mathbf{\Omega}_k^*\|_2\leqslant\epsilon_n, \text{and } \Omega_{kj}=0,\ \text{for } j\notin\mathcal{W}_k, k=1,\cdots,K\}.$$

By Result 1 and (B.29), there exists an event E_1 in which:

$$\sup_k \|\widehat{\mathbf{\Omega}}_k^o - \mathbf{\Omega}_k^*\|_2 \leqslant \epsilon_n, \text{and } \widehat{\Omega}_{kj}^o = 0, \text{ for } j \notin \mathcal{W}_k, k = 1, \cdots, K,$$

and $P(E_1^C) \to 0$. Thus $\widehat{\mathbf{\Omega}}^o \in \mathcal{C}$. Then, with the following two steps, we show that $\widehat{\mathbf{\Omega}}^o$ is a strict local maximum point of objective function (A.1) with probability converging to 1:

(i) On event E_1, for any $\breve{\mathbf{\Omega}} \in \mathcal{C}$ and $\breve{\mathbf{\Omega}} \neq \widehat{\mathbf{\Omega}}^o$, $\mathcal{L}(\breve{\mathbf{\Omega}}) < \mathcal{L}(\widehat{\mathbf{\Omega}}^o)$.

(ii) On event E_1, there is a neighborhood of $\widehat{\mathbf{\Omega}}^o$, denoted by $\mathcal{C}_n$, such that $\mathcal{L}(\mathbf{\Omega}) \leqslant \mathcal{L}(\breve{\mathbf{\Omega}})$, for any $\mathbf{\Omega} \in \mathcal{C} \cap \mathcal{C}_n$ and a sufficiently large n.

With the results in (i) and (ii), for any $\mathbf{\Omega} \in \mathcal{C} \cap \mathcal{C}_n$ and $\mathbf{\Omega} \neq \widehat{\mathbf{\Omega}}^o$ on $\mathcal{C} \cap \mathcal{C}_n$, we have $\mathcal{L}(\mathbf{\Omega}) \leqslant \mathcal{L}(\widehat{\mathbf{\Omega}}^o)$. So $\widehat{\mathbf{\Omega}}^o$ is a strict local maximum point of (A.1) on event $\mathcal{C} \cap \mathcal{C}_n$ with $P(\mathcal{C} \cap \mathcal{C}_n) \to 1$ for a sufficiently large n.

First, we prove the result in (i). Let $\breve{G}(\mathbf{\Omega}) = \mathbf{\Upsilon} = (\mathbf{\Upsilon}_1^\top, \cdots, \mathbf{\Upsilon}_{K0}^\top)^\top$. Then $\|\mathbf{\Upsilon}_l - \mathbf{\Upsilon}_{l'}\|_2 \geqslant \|\mathbf{\Upsilon}_l^* - \mathbf{\Upsilon}_{l'}^*\|_2 - 2\sup_l \|\mathbf{\Upsilon}_l - \mathbf{\Upsilon}_l^*\|_2$, and for any $\mathbf{\Omega} \in \mathcal{C}$,

$$\begin{aligned}
\sup_l \|\mathbf{\Upsilon}_l - \mathbf{\Upsilon}_l^*\|_2^2 &= \sup_l \| \, |\mathcal{T}_l^*|^{-1} \sum_{k \in \mathcal{T}_l^*} \mathbf{\Omega}_k - \mathbf{\Upsilon}_l^* \|_2^2 \\
&= \sup_l \| \, |\mathcal{T}_l^*|^{-1} \sum_{k \in \mathcal{T}_l^*} (\mathbf{\Omega}_k - \mathbf{\Omega}_k^*) \, \|_2^2 \\
&= \sup_l |\mathcal{T}_l^*|^{-2} \| \sum_{k \in \mathcal{T}_l^*} (\mathbf{\Omega}_k - \mathbf{\Omega}_k^*) \, \|_2^2 \qquad \text{(B.30)} \\
&\leqslant \sup_l |\mathcal{T}_l^*|^{-1} \sum_{k \in \mathcal{T}_l^*} \|\mathbf{\Omega}_k - \mathbf{\Omega}_k^*\|_2^2 \\
&\leqslant \sup_k \|\mathbf{\Omega}_k - \mathbf{\Omega}_k^*\|_2^2 \leqslant \epsilon_n^2.
\end{aligned}$$

Note that $b = \min_{1 \leqslant l \neq l' \leqslant K_0} \|\mathbf{\Upsilon}_l^* - \mathbf{\Upsilon}_{l'}^*\|_2$ is sufficiently large such that $b > a\lambda_3$, and $\lambda_3 \gg \epsilon_n$. Then for any l and l', $\|\mathbf{\Upsilon}_l - \mathbf{\Upsilon}_{l'}\|_2 \geqslant b - 2\epsilon_n > a\lambda_3$. So $p\left(\|\mathbf{\Upsilon}_l - \mathbf{\Upsilon}_{l'}\|_2, \lambda_3\right)$ is a constant. Note that for any $\mathbf{\Omega} \in \Lambda_{\mathcal{T}^*}$,

$$\mathcal{P}_3^{\mathcal{T}}(\mathbf{\Omega}) = \sum_{1 \leqslant k < k' \leqslant K} p\left(\|\mathbf{\Omega}_k - \mathbf{\Omega}_{k'}\|_2, \lambda_3\right) = \sum_{1 \leqslant l < l' \leqslant K_0} |\mathcal{T}_l^*| \, |\mathcal{T}_{l'}^*| \, p\left(\|\mathbf{\Upsilon}_l - \mathbf{\Upsilon}_{l'}\|_2, \lambda_3\right).$$

So $\mathcal{P}_3^{\mathcal{T}}(\mathbf{\Omega})$ is a constant that does not depend on $\mathbf{\Omega}$, for any $\mathbf{\Omega} \in \mathcal{C} \cap \Lambda_{\mathcal{T}^*}$.

Since $\widehat{\mathbf{\Omega}}^o$ is the unique maximum point of $\mathcal{Q}^{\mathcal{T}}(\mathbf{\Omega})$, $\mathcal{Q}^{\mathcal{T}}(\breve{\mathbf{\Omega}}) < \mathcal{Q}^{\mathcal{T}}(\widehat{\mathbf{\Omega}}^o)$ for any $\breve{\mathbf{\Omega}} \in \mathcal{C}$, and

$\breve{\boldsymbol{\Omega}} \neq \widehat{\boldsymbol{\Omega}}^o$. Note that $\breve{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Omega}}^o \in \Lambda_{\mathcal{T}^*}$. So $\mathcal{P}_3^{\mathcal{T}}(\breve{\boldsymbol{\Omega}}) = \mathcal{P}_3^{\mathcal{T}}(\widehat{\boldsymbol{\Omega}}^o)$, and we have $\mathcal{L}^{\mathcal{T}}(\breve{\boldsymbol{\Omega}}) < \mathcal{L}^{\mathcal{T}}(\widehat{\boldsymbol{\Omega}}^o)$. By (B.28), $\mathcal{L}(\breve{\boldsymbol{\Omega}}) < \mathcal{L}(\widehat{\boldsymbol{\Omega}}^o)$. Therefore, the result in (i) is proved.

Next, we prove the result in (ii). Given a positive sequence ϕ_n, consider $\mathcal{C}_n = \{\boldsymbol{\Omega} \in \mathbb{R}^{K(p+p^2)} : \sup_k \|\boldsymbol{\Omega}_k - \widehat{\boldsymbol{\Omega}}_k^o\|_\infty \leqslant \phi_n/\sqrt{d+s}\}$. For $\boldsymbol{\Omega} \in \mathcal{C} \cap \mathcal{C}_n$ and $\breve{\boldsymbol{\Omega}} \in \mathcal{C}$, by Taylor's expansion (only for $(k,j) : \Omega_{kj}^* \neq 0,$ that is, for $j \in \mathcal{W}_k, k = 1, \cdots, K$),

$$\mathcal{L}(\boldsymbol{\Omega}) - \mathcal{L}(\breve{\boldsymbol{\Omega}}) = \ell_1 - \ell_2,$$

where

$$\begin{aligned}
\ell_1 &= \sum_{k=1}^K \boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \breve{\boldsymbol{\Omega}}_k), \\
\ell_2 &= \sum_{k=1}^K \frac{\partial \mathcal{P}_3(\widetilde{\boldsymbol{\Omega}})}{\partial \boldsymbol{\Omega}_k^\top} (\boldsymbol{\Omega}_k - \breve{\boldsymbol{\Omega}}_k), \\
\boldsymbol{D}_k^\top &= \left(\boldsymbol{D}_{1k}^\top, \boldsymbol{D}_{2k}^\top\right)^\top \boldsymbol{\mathcal{I}}_{\mathcal{W}_k}, \\
\boldsymbol{D}_{1k} &= \frac{1}{n} \sum_{i=1}^n r(\boldsymbol{x}_i; \widetilde{\boldsymbol{\Omega}}_k) \widetilde{\boldsymbol{\Theta}}_k (\boldsymbol{x}_i - \tilde{\boldsymbol{\mu}}_k) - \lambda_1 \boldsymbol{s}_1(\tilde{\boldsymbol{\mu}}_k), \\
\boldsymbol{D}_{2k} &= \frac{1}{n} \sum_{i=1}^n r(\boldsymbol{x}_i; \widetilde{\boldsymbol{\Omega}}_k) \frac{1}{2} [\mathrm{vec}(\widetilde{\boldsymbol{\Theta}}_k)^{-1} - \mathrm{vec}(\boldsymbol{x}_i - \tilde{\boldsymbol{\mu}}_k)(\boldsymbol{x}_i - \tilde{\boldsymbol{\mu}}_k)^\top] - \lambda_2 \boldsymbol{s}_2(\widetilde{\boldsymbol{\Theta}}_k),
\end{aligned}$$

$\boldsymbol{\mathcal{I}}_{\mathcal{W}_k}$ is a diagonal matrix with the jth diagonal element $I(j \in \mathcal{W}_k)$, $j = 1, \ldots, p + p^2$, $r(\boldsymbol{x}_i; \widetilde{\boldsymbol{\Omega}}_k) = \frac{\pi_k f_k(\boldsymbol{x}_i; \widetilde{\boldsymbol{\Omega}}_k)}{\sum_{k=1}^K \pi_k f_k(\boldsymbol{x}_i; \widetilde{\boldsymbol{\Omega}}_k)}$, $\boldsymbol{s}_1(\tilde{\boldsymbol{\mu}}_k) = (\rho'(|\tilde{\mu}_{k1}| \text{ sgn } (\tilde{\mu}_{k1})), \cdots, \rho'(|\tilde{\mu}_{kp}|) \text{ sgn } (\tilde{\mu}_{kp}))^\top \in \mathbb{R}^p$, $\boldsymbol{s}_2(\widetilde{\boldsymbol{\Theta}}_k) \in \mathbb{R}^{p^2}$ such that the $((l-1)p + j)$th element of $\boldsymbol{s}_2(\widetilde{\boldsymbol{\Theta}}_k)$ is $\rho'(|\tilde{\theta}_{kjl}|) \text{ sgn } (\tilde{\theta}_{kjl}) I(j \neq l)$, and $\widetilde{\boldsymbol{\Omega}} = \varsigma \boldsymbol{\Omega} + (1 - \varsigma) \breve{\boldsymbol{\Omega}}$ for some constant $\varsigma \in (0, 1)$.

First, we consider ℓ_2. It can be derived that:

$$\begin{aligned}
\ell_2 &= \lambda_3 \sum_{k=1}^{K} \sum_{k \neq k'} \rho'(\|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2) \|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2^{-1} (\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'})^\top (\mathbf{\Omega}_k - \breve{\mathbf{\Omega}}_k) \\
&= \lambda_3 \sum_{k<k'} \rho'(\|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2) \|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2^{-1} (\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'})^\top (\mathbf{\Omega}_k - \breve{\mathbf{\Omega}}_k) \\
&\quad + \lambda_3 \sum_{k>k'} \rho'(\|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2) \|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2^{-1} (\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'})^\top (\mathbf{\Omega}_k - \breve{\mathbf{\Omega}}_k) \\
&= \lambda_3 \sum_{k<k'} \rho'(\|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2) \|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2^{-1} (\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'})^\top (\mathbf{\Omega}_k - \breve{\mathbf{\Omega}}_k) \\
&\quad + \lambda_3 \sum_{k'>k} \rho'(\|\widetilde{\mathbf{\Omega}}_{k'} - \widetilde{\mathbf{\Omega}}_k\|_2) \|\widetilde{\mathbf{\Omega}}_{k'} - \widetilde{\mathbf{\Omega}}_k\|_2^{-1} (\widetilde{\mathbf{\Omega}}_{k'} - \widetilde{\mathbf{\Omega}}_k)^\top (\mathbf{\Omega}_{k'} - \breve{\mathbf{\Omega}}_{k'}) \\
&= \lambda_3 \sum_{k<k'} \rho'(\|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2) \|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2^{-1} (\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'})^\top [(\mathbf{\Omega}_k - \breve{\mathbf{\Omega}}_k) - (\mathbf{\Omega}_{k'} - \breve{\mathbf{\Omega}}_{k'})].
\end{aligned}$$

If $k, k' \in \mathcal{T}_l^*$, $\breve{\mathbf{\Omega}}_k = \breve{\mathbf{\Omega}}_{k'}$. So

$$\begin{aligned}
\ell_2 = \lambda_3 \sum_{l=1}^{K_0} \sum_{\left\{k,k' \in \mathcal{T}_l^*, k<k'\right\}} & \rho'(\|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2) \|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2^{-1} (\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'})^\top (\mathbf{\Omega}_k - \mathbf{\Omega}_{k'}) \\
+ \lambda_3 \sum_{l<l'} \sum_{\left\{k \in \mathcal{T}_l^*, k' \in \mathcal{T}_{l'}^*\right\}} & \rho'(\|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2) \|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2^{-1} (\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'})^\top [(\mathbf{\Omega}_k - \breve{\mathbf{\Omega}}_k) - (\mathbf{\Omega}_{k'} - \breve{\mathbf{\Omega}}_{k'})].
\end{aligned}$$

Note that $\sup_k \|\breve{\mathbf{\Omega}}_k - \mathbf{\Omega}_k^*\|_2^2 = \sup_l \|\mathbf{\Upsilon}_l - \mathbf{\Upsilon}_l^*\|_2^2 \leqslant \epsilon_n^2$, following from (B.30). And then

$$\sup_k \|\widetilde{\mathbf{\Omega}}_k - \mathbf{\Omega}_k^*\|_2 \leqslant \varsigma \sup_k \|\mathbf{\Omega}_k - \mathbf{\Omega}_k^*\|_2 + (1-\varsigma) \sup_k \|\breve{\mathbf{\Omega}}_k - \mathbf{\Omega}_k^*\|_2 \leqslant \varsigma\epsilon_n + (1-\varsigma)\epsilon_n = \epsilon_n. \quad \text{(B.31)}$$

Hence for $k \in \mathcal{T}_l^*, k' \in \mathcal{T}_{l'}^*, l \neq l'$, $\|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2 \geqslant \min_{k \in \mathcal{T}_l^*, k' \in \mathcal{T}_{l'}^*} \|\mathbf{\Omega}_k^* - \mathbf{\Omega}_{k'}^*\|_2 - 2\sup_k \|\widetilde{\mathbf{\Omega}}_k - \mathbf{\Omega}_k^*\|_2 \geqslant b - 2\epsilon_n > a\lambda_3$, and so $\rho'(\|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2) = 0$. Noting that $\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'} = \varsigma(\mathbf{\Omega}_k - \mathbf{\Omega}_{k'})$ for $k, k' \in \mathcal{T}_l^*$, we have

$$\ell_2 = \lambda_3 \sum_{l=1}^{K_0} \sum_{\left\{k,k' \in \mathcal{T}_l^*, k<k'\right\}} \rho'(\|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2) \|\mathbf{\Omega}_k - \mathbf{\Omega}_{k'}\|_2.$$

Similar to (B.30), $\sup_k \|\breve{\mathbf{\Omega}}_k - \widehat{\mathbf{\Omega}}_k^o\|_2 \leqslant \sup_k \|\mathbf{\Omega}_k - \widehat{\mathbf{\Omega}}_k^o\|_2$. And note that $\breve{\mathbf{\Omega}}_k = \breve{\mathbf{\Omega}}_{k'}$ for $k, k' \in \mathcal{T}_l^*$. Then it can be shown that:

$$\begin{aligned}
\sup_k \|\widetilde{\mathbf{\Omega}}_k - \widetilde{\mathbf{\Omega}}_{k'}\|_2 &= \sup_k \|\widetilde{\mathbf{\Omega}}_k - \breve{\mathbf{\Omega}}_k + \breve{\mathbf{\Omega}}_{k'} - \widetilde{\mathbf{\Omega}}_{k'}\|_2 \leqslant 2\sup_k \|\widetilde{\mathbf{\Omega}}_k - \breve{\mathbf{\Omega}}_k\|_2 \leqslant 2\sup_k \|\mathbf{\Omega}_k - \breve{\mathbf{\Omega}}_k\|_2 \\
&\leqslant 2(\sup_k \|\mathbf{\Omega}_k - \widehat{\mathbf{\Omega}}_k^o\|_2 + \|\breve{\mathbf{\Omega}}_k - \widehat{\mathbf{\Omega}}_k^o\|_2) \leqslant 4\sup_k \|\mathbf{\Omega}_k - \widehat{\mathbf{\Omega}}_k^o\|_2 \leqslant 4\phi_n.
\end{aligned}$$

Therefore, $\rho'(\|\widetilde{\boldsymbol{\Omega}}_k - \widetilde{\boldsymbol{\Omega}}_{k'}\|_2) \geqslant \rho'(4\phi_n)$ by the concavity of $\rho(\cdot)$. So we have:

$$\ell_2 \geqslant \sum_{l=1}^{K_0} \sum_{\left\{k,k' \in \mathcal{T}_l^*, k<k'\right\}} \lambda_3 \rho'(4\phi_n) \|\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'}\|_2. \tag{B.32}$$

Next, ℓ_1 can be rewritten as:

$$\begin{aligned}
\ell_1 &= \sum_{k=1}^{K} \boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \breve{\boldsymbol{\Omega}}_k) = \sum_{l=1}^{K_0} \sum_{k \in \mathcal{T}_l^*} \boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \breve{\boldsymbol{\Omega}}_k) \\
&= \sum_{l=1}^{K_0} \sum_{k \in \mathcal{T}_l^*} \sum_{k' \in \mathcal{T}_l^*} \frac{\boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \breve{\boldsymbol{\Omega}}_k)}{|\mathcal{T}_l^*|} = \sum_{l=1}^{K_0} \sum_{k \in \mathcal{T}_l^*} \sum_{k' \in \mathcal{T}_l^*} \frac{\boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'} + \boldsymbol{\Omega}_{k'} - \breve{\boldsymbol{\Omega}}_{k'})}{|\mathcal{T}_l^*|} \\
&= \sum_{l=1}^{K_0} \sum_{k \in \mathcal{T}_l^*} \sum_{k' \in \mathcal{T}_l^*} \left[\frac{\boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'})}{|\mathcal{T}_l^*|} + \frac{\boldsymbol{D}_k^\top (\boldsymbol{\Omega}_{k'} - \breve{\boldsymbol{\Omega}}_{k'})}{|\mathcal{T}_l^*|} \right] \\
&= \sum_{l=1}^{K_0} \sum_{\{k,k' \in \mathcal{T}_l^*\}} \frac{\boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'})}{|\mathcal{T}_l^*|} + \sum_{l=1}^{K_0} \sum_{k \in \mathcal{T}_l^*} \boldsymbol{D}_k^\top \left(\sum_{k' \in \mathcal{T}_l^*} \frac{\boldsymbol{\Omega}_{k'}}{|\mathcal{T}_l^*|} - \breve{\boldsymbol{\Omega}}_{k'} \right).
\end{aligned}$$

Note that $\sum_{k' \in \mathcal{T}_l^*} \frac{\boldsymbol{\Omega}_{k'}}{|\mathcal{T}_l^*|} = \breve{\boldsymbol{\Omega}}_{k'}$. So

$$\begin{aligned}
\ell_1 &= \sum_{l=1}^{K_0} \sum_{\{k,k' \in \mathcal{T}_l^*\}} \frac{\boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'})}{|\mathcal{T}_l^*|} \\
&= \sum_{l=1}^{K_0} \sum_{\{k,k' \in \mathcal{T}_l^*\}} \frac{\boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'})}{2|\mathcal{T}_l^*|} + \sum_{l=1}^{K_0} \sum_{\{k,k' \in \mathcal{T}_l^*\}} \frac{\boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'})}{2|\mathcal{T}_l^*|} \\
&= \sum_{l=1}^{K_0} \sum_{\{k,k' \in \mathcal{T}_l^*\}} \frac{\boldsymbol{D}_k^\top (\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'})}{2|\mathcal{T}_l^*|} + \sum_{l=1}^{K_0} \sum_{\{k,k' \in \mathcal{T}_l^*\}} \frac{\boldsymbol{D}_{k'}^\top (\boldsymbol{\Omega}_{k'} - \boldsymbol{\Omega}_k)}{2|\mathcal{T}_l^*|} \\
&= \sum_{l=1}^{K_0} \sum_{\{k,k' \in \mathcal{T}_l^*\}} \frac{(\boldsymbol{D}_k - \boldsymbol{D}_{k'})^\top (\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'})}{2|\mathcal{T}_l^*|} = \sum_{l=1}^{K_0} \sum_{\{k,k' \in \mathcal{T}_l^*, k<k'\}} \frac{(\boldsymbol{D}_k - \boldsymbol{D}_{k'})^\top (\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'})}{|\mathcal{T}_l^*|} \\
&\leqslant \sum_{l=1}^{K_0} \sum_{\{k,k' \in \mathcal{T}_l^*, k<k'\}} |\mathcal{T}_{\min}|^{-1} \sup_k \|\boldsymbol{D}_k - \boldsymbol{D}_{k'}\|_2 \|\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_{k'}\|_2.
\end{aligned} \tag{B.33}$$

It can be shown that:

$$\sup_k \|\boldsymbol{D}_k - \boldsymbol{D}_{k'}\|_2 \leqslant \sqrt{d \cdot \sup_k \|\boldsymbol{D}_{1k} - \boldsymbol{D}_{1k'}\|_\infty^2 + s \cdot \sup_k \|\boldsymbol{D}_{2k} - \boldsymbol{D}_{2k'}\|_\infty^2}. \tag{B.34}$$

Consider $\sup_k \|\boldsymbol{D}_{2k} - \boldsymbol{D}_{2k'}\|_\infty^2$. With (B.31), the minimal signal condition of the true parameters, and Condition B.4, we have $\left(\boldsymbol{s}_1(\tilde{\boldsymbol{\mu}}_k)^\top, \boldsymbol{s}_2(\widetilde{\boldsymbol{\Theta}}_k)^\top\right)^\top \boldsymbol{\mathcal{I}}_{\mathcal{W}_k} = \mathbf{0}^\top$. So $\boldsymbol{D}_{1k}$ and $\boldsymbol{D}_{2k}$ do

not contain terms related to λ_1 and λ_2. Note that

$$\begin{aligned}
&\|\boldsymbol{D}_{2k}-\boldsymbol{D}_{2k'}\|_\infty \leqslant \sup_i\|r(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)-r(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_{k'})m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_{k'})\|_\infty\\
&=\sup_i\|r(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)[m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)-m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_{k'})]+[r(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)-r(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_{k'})]m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_{k'})\|_\infty\\
&\leqslant \sup_i\{r(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)\}\sup_i\|m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)-m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_{k'})\|_\infty\\
&\quad+\sup_i\{r(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)-r(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_{k'})\}\sup_i\|m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_{k'})\|_\infty,
\end{aligned}\tag{B.35}$$

where $m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)=\frac{1}{2}[\mathrm{vec}(\widetilde{\boldsymbol{\Theta}}_k)^{-1}-\mathrm{vec}(\boldsymbol{x}_i-\tilde{\boldsymbol{\mu}}_k)(\boldsymbol{x}_i-\tilde{\boldsymbol{\mu}}_k)^\top]$. In addition, $r(\boldsymbol{x}_i;\cdot)$ and $m(\boldsymbol{x}_i;\cdot)$ are continuous functions. Then, for a constant $C'>0$,

$$\begin{aligned}
\sup_i\|m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)-m(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_{k'})\|_\infty &\leqslant C'\|\widetilde{\boldsymbol{\Omega}}_k-\widetilde{\boldsymbol{\Omega}}_{k'}\|_\infty,\\
\sup_i\{r(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_k)-r(\boldsymbol{x}_i;\widetilde{\boldsymbol{\Omega}}_{k'})\} &\leqslant C'\|\widetilde{\boldsymbol{\Omega}}_k-\widetilde{\boldsymbol{\Omega}}_{k'}\|_\infty.
\end{aligned}\tag{B.36}$$

It is also noted that $\widetilde{\boldsymbol{\Omega}}_k=\varsigma\boldsymbol{\Omega}_k+(1-\varsigma)\breve{\boldsymbol{\Omega}}_k$, $\widetilde{\boldsymbol{\Omega}}_{k'}=\varsigma\boldsymbol{\Omega}_{k'}+(1-\varsigma)\breve{\boldsymbol{\Omega}}_{k'}$, $\breve{\boldsymbol{\Omega}}_k=\breve{\boldsymbol{\Omega}}_{k'}$ for $k,k'\in\mathcal{T}_l^*$, and $\boldsymbol{\Omega}_k,\boldsymbol{\Omega}_{k'}\in\mathcal{C}_n$. Then,

$$\|\widetilde{\boldsymbol{\Omega}}_k-\widetilde{\boldsymbol{\Omega}}_{k'}\|_\infty=\varsigma\|\boldsymbol{\Omega}_k-\boldsymbol{\Omega}_{k'}\|_\infty\leqslant 2\phi_n/\sqrt{d+s}.\tag{B.37}$$

By (B.35), (B.36), and (B.37), we have

$$\sup_k\|\boldsymbol{D}_{2k}-\boldsymbol{D}_{2k'}\|_\infty=O(\phi_n/\sqrt{d+s}).\tag{B.38}$$

Similarly,

$$\sup_k\|\boldsymbol{D}_{1k}-\boldsymbol{D}_{1k'}\|_\infty=O(\phi_n/\sqrt{d+s}).\tag{B.39}$$

By (B.32) and (B.33), we have:

$$\mathcal{L}(\boldsymbol{\Omega})-\mathcal{L}(\breve{\boldsymbol{\Omega}})\leqslant\sum_{l=1}^{K_0}\sum_{\{k,k'\in\mathcal{T}_l^*,k<k'\}}\left[|\mathcal{T}_{\min}|^{-1}\sup_k\|\boldsymbol{D}_k-\boldsymbol{D}_{k'}\|_2-\lambda_3\rho'(4\phi_n)\right]\|\boldsymbol{\Omega}_k-\boldsymbol{\Omega}_{k'}\|_2.$$

Let $\phi_n=\epsilon_n$. Then $\rho'(4\phi_n)\to 1$. Note that $\lambda_3\gg\epsilon_n$, and $|\mathcal{T}_{\min}|\geqslant 1$. Therefore, $|\mathcal{T}_{\min}|\lambda_3\gg\sup_k\|\boldsymbol{D}_k-\boldsymbol{D}_{k'}\|_2$. And then, $\mathcal{L}(\boldsymbol{\Omega})-\mathcal{L}(\breve{\boldsymbol{\Omega}})\leqslant 0$. So the result in (ii) is proved. This concludes the proof of Result 2.

C. Additional discussions on sparsistency

Recall that the derivative of the objective function $\mathcal{Q}(\widehat{\boldsymbol{\Upsilon}})$ at a local maximum point $\widehat{\boldsymbol{\Upsilon}}$ that satisfies (B.18) with respect to θ_{ljm}, for $(j,m) \in \mathcal{S}_l^C, l = 1, \cdots, K_0$, is:

$$\frac{\partial \mathcal{Q}(\widehat{\boldsymbol{\Upsilon}})}{\partial \theta_{ljm}} = \mathcal{R}(\widehat{\boldsymbol{\Upsilon}}_l) - \lambda_2 \rho'(|\hat{\theta}_{ljm}|)\ \text{sgn}\ (\hat{\theta}_{ljm}),\ \text{and}$$

$$\mathcal{R}(\widehat{\boldsymbol{\Upsilon}}) \leqslant \|\boldsymbol{\Theta}_l^* - \widehat{\boldsymbol{\Theta}}_l\|_2 + \mathcal{R}^*(\widehat{\boldsymbol{\Upsilon}}_l).$$

To compare the orders of the sparsity parameter s under the L_1 and MCP penalties, we further consider the range of the upper bound of the order of $\mathcal{R}(\widehat{\boldsymbol{\Upsilon}}_l)$. Obviously, $\|\widehat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}^*\|_2^2/(K_0 p) \leqslant \|\boldsymbol{\Theta}_l^* - \widehat{\boldsymbol{\Theta}}_l\|_2^2 \leqslant \|\widehat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}^*\|_2^2$. That is,

$$K_0^3(1 + s/p)\frac{\log p}{n} \lesssim \|\boldsymbol{\Theta}_l^* - \widehat{\boldsymbol{\Theta}}_l\|_2^2 \lesssim K_0^4(s+p)\text{log}p/n.$$

As for $\mathcal{R}^*(\widehat{\boldsymbol{\Upsilon}}_l)$, with (B.22), (B.23), and (B.27),

$$\frac{\log p}{n} \lesssim [\mathcal{R}^*(\widehat{\boldsymbol{\Upsilon}}_l)]^2 \lesssim K_0^4(s+p)\text{log}p/n.$$

To establish the variable selection consistency of the final precision matrix estimators, according to the proof of part 2 of Result 1, we need $\|\boldsymbol{\Theta}_l^* - \widehat{\boldsymbol{\Theta}}_l\|_2^2 + [\mathcal{R}^*(\widehat{\boldsymbol{\Upsilon}}_l)]^2 \lesssim \lambda_2$. Note that $\lambda_2^2 = O\left(\frac{K_0^3 \log p}{n} + \frac{K_0^3 p \log p}{n(d+s)}\right)$ under the L_1 penalty (Hao et al., 2018). We have the following observations on the orders of the sparsity parameter s under the L_1 and MCP penalties:

- In the worst case scenario with $\|\boldsymbol{\Theta}_l^* - \widehat{\boldsymbol{\Theta}}_l\|_2^2 \simeq [\mathcal{R}^*(\widehat{\boldsymbol{\Upsilon}}_l)]^2 \simeq K_0^4(s+p)\text{log}p/n$, for the L_1 penalty, the following needs to be satisfied:

$$\frac{K_0^4(s+p)\log p}{n} \lesssim K_0^3\left[\frac{\log p}{n} + \frac{p \log p}{n(d+s)}\right],$$

  so that $s = O(1)$, and $K_0 = O(1)$.

- In the optimistic scenario with $\|\boldsymbol{\Theta}_l^* - \widehat{\boldsymbol{\Theta}}_l\|_2^2 = K_0^2(1+s/p)\frac{\log p}{n}$ and $[\mathcal{R}^*(\widehat{\boldsymbol{\Upsilon}}_l)]^2 = \frac{\log p}{n}$, for the L_1 penalty,

$$K_0^3(1+s/p)\frac{\log p}{n} + \frac{\log p}{n} \lesssim K_0^3\left[\frac{\log p}{n} + \frac{p \log p}{n(d+s)}\right],$$

  so that $s = O(p)$.

For the MCP, the order of s can be larger than $O(p)$ under both scenarios.

D. Additional numerical results

Heterogeneity analysis of regulatory T cells in non-small-cell lung cancer. In Figure 1 of the main text, we have shown that several immune related genes express differently across the three subgroups identified by the proposed approach. Here we provide brief information on those genes, particularly their relevance to lung cancer. Published literature suggests that the expression of FOXP3 characterizes Tregs that engages in the maintenance of immunological self-tolerance and immune homeostasis, and that it can be used as a prognostic marker for non-small-cell lung cancer (Tao et al., 2012). Post-translational modifications of FOXO1 regulate epidermal growth factor receptor tyrosine kinase inhibitor that has resistance for non-small cell lung cancer cells. FOXO1 is down regulated in human non-small cell lung cancer cells, and silencing of FOXO1 is associated with the invasive stage of tumor progression (Gao et al., 2018). Among the genes that are over-expressed in the second subgroup, it has been confirmed using fluorescence-activated cell sorting that TIGIT is expressed by a large percentage of non-small-cell lung tumor-infiltrating T cells (Johnston et al., 2014). Among the genes that are over-expressed in the third subgroup, it has been demonstrated that IRF4 serves as a tumor promoter in human non-small-cell lung cancer cells through activating the Notch-Akt signaling pathway, and that the knockdown of IRF4 can significantly decrease the non-small-cell lung cancer cell proliferation rate, colony formation, and expression levels of phosphorylated protein kinase B (Qian et al., 2017). It has been reported that CTLA-4 is the main inhibitory molecules of the immune system, whose physiologic role is the maintenance of peripheral tolerance and termination of immune responses. Treg cells and CTLA-4 co-inhibitory molecule act as the main arms of immunity to attenuate immune responses (Erfani et al., 2012).

Data is also analyzed using the two alternatives considered in simulation by setting $K = 3$ (for better comparability). The three subgroups identified by the K-means+JGL method

have 322, 270, and 310 samples, respectively. The estimated network structures are shown in Figure S1. They have 120, 75, and 90 edges, respectively, with 8, 2, and 14 overlapping with those under the proposed analysis. The three subgroups identified by SCAN have 387, 188, and 327 samples, respectively. The estimated network structures are shown in Figure S2. They have 80, 109, and 111 edges, respectively, with 5, 5, and 25 overlapping with those under the proposed analysis. The sample overlaps of the subgroups identified by different methods are summarized in Table S4. The proposed analysis has a higher degree of agreement with SCAN.

LUAD heterogeneity analysis using histopathological imaging data. The histopathological imaging data is also analyzed using the two alternatives by setting $K = 2$ (for better comparability). The two subgroups identified by the K-means+JGL method have 244 and 63 samples, respectively. The estimated network structures are shown in Figure S3. They have 243 and 247 edges respectively, with 189 and 169 overlapping with those under the proposed analysis. The two subgroups identified by SCAN have 265 and 42 samples, respectively. The estimated network structures are shown in Figure S4. They have 285 and 238 edges, respectively, with 222 and 156 overlapping with those under the proposed analysis. The sample overlaps of the subgroups identified by different methods are shown in Table S5. The proposed approach agrees with the K-means+JGL method to a higher extent.

References

Breheny, P. and Huang, J. (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and computing*, **25**, 173–187.

Cai, T., Li, H., Liu, W. and Xie, J. (2016). Joint estimation of multiple high-dimensional precision matrices. *Statistica Sinica*, **26**, 445–464.

Danaher, P., Wang, P. and Witten, D. M. (2014). The joint graphical lasso for inverse

covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 373–397.

Erfani, N., Mehrabadi, S. M., Ghayumi, M. A., Haghshenas, M. R., Mojtahedi, Z., Ghaderi, A. et al. (2012). Increase of regulatory T cells in metastatic stage and CTLA-4 over expression in lymphocytes of patients with non-small cell lung cancer (NSCLC). *Lung cancer*, **77**, 306–311.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348-1360.

Gao, Z., Liu, R., Ye, N., Liu, C., Li, X., Guo, X., et al. (2018). FOXO1 inhibits tumor cell migration via regulating cell surface morphology in non-small cell lung cancer cells. *Cellular Physiology and Biochemistry*, **48**, 138–148.

Gao, C., Zhu, Y., Shen, X. and Pan, W. (2016). Estimation of multiple networks in gaussian mixture models. *Electronic journal of statistics*, **10**, 1133-1154.

Hao, B., Sun, W., Liu, Y. and Cheng, G. (2018). Simultaneous clustering and estimation of heterogeneous graphical models. *The Journal of Machine Learning Research*, **18**, 7981–8038.

Johnston, R. J., Comps-Agrar, L., Hackney, J., Yu, X., Huseni, M., Yang, Y., et al. (2014). The immunoreceptor TIGIT regulates antitumor and antiviral CD8+ T cell effector function. *Cancer cell*, **26**, 923–937.

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

McLachlan, G. J. and Krishnan, T. (2007). The EM algorithm and extensions. *Wiley Series in Probability and Statistics*.

Qian, Y., Du, Z., Xing, Y., Zhou, T., Chen, T. and Shi, M. (2017). Interferon regulatory factor 4 (IRF4) is overexpressed in human non-small cell lung cancer (NSCLC) and

activates the Notch signaling pathway. *Molecular medicine reports*, **16**, 6034–6040.

Tao, H., Mimura, Y., Aoe, K., Kobayashi, S., Yamamoto, H., Matsuda, E., et al. (2012). Prognostic potential of FOXP3 expression in non-small cell lung cancer cells combined with tumor-infiltrating regulatory T cells. *Lung cancer*, **75**, 95–101.

Wang, B., Zhang, Y., Sun, W. W. and Fang, Y. (2018). Sparse convex clustering. *Journal of Computational and Graphical Statistics*, **27**, 393–403.

Witten, D. M. and Tibshirani, R. (2009). Covariance-regularized regression and classification for high dimensional problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 615–636.

Zhang, C. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, **38**, 894-942.

[Table 1 about here.]

[Table 2 about here.]

[Figure 1 about here.]

[Figure 2 about here.]

[Table 3 about here.]

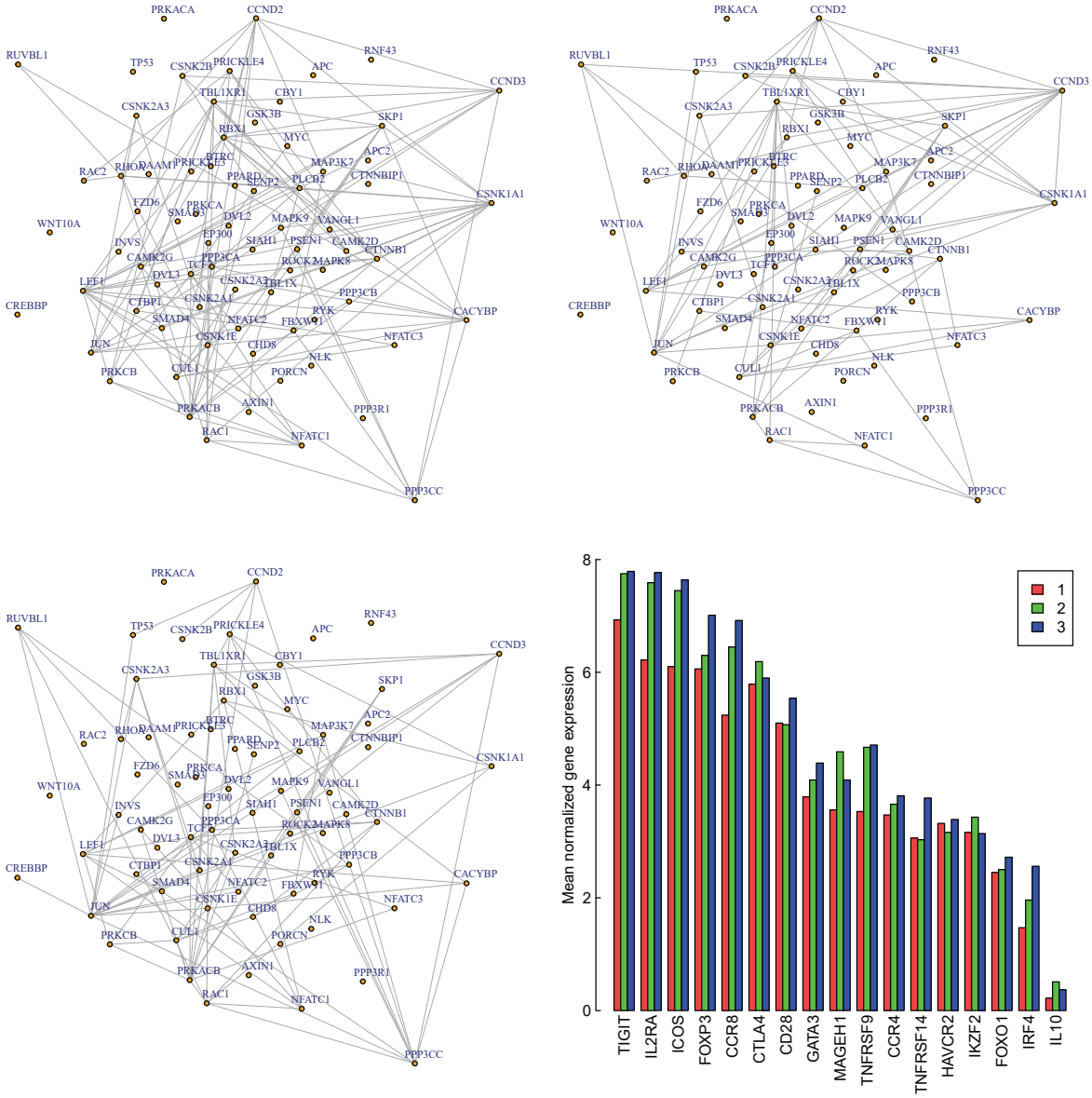

Figure S1. Analysis of regulatory T cells: identified subgrouping network structures using the K-means+JGL method and mean normalized expressions of 17 immune related genes in the subgroups.

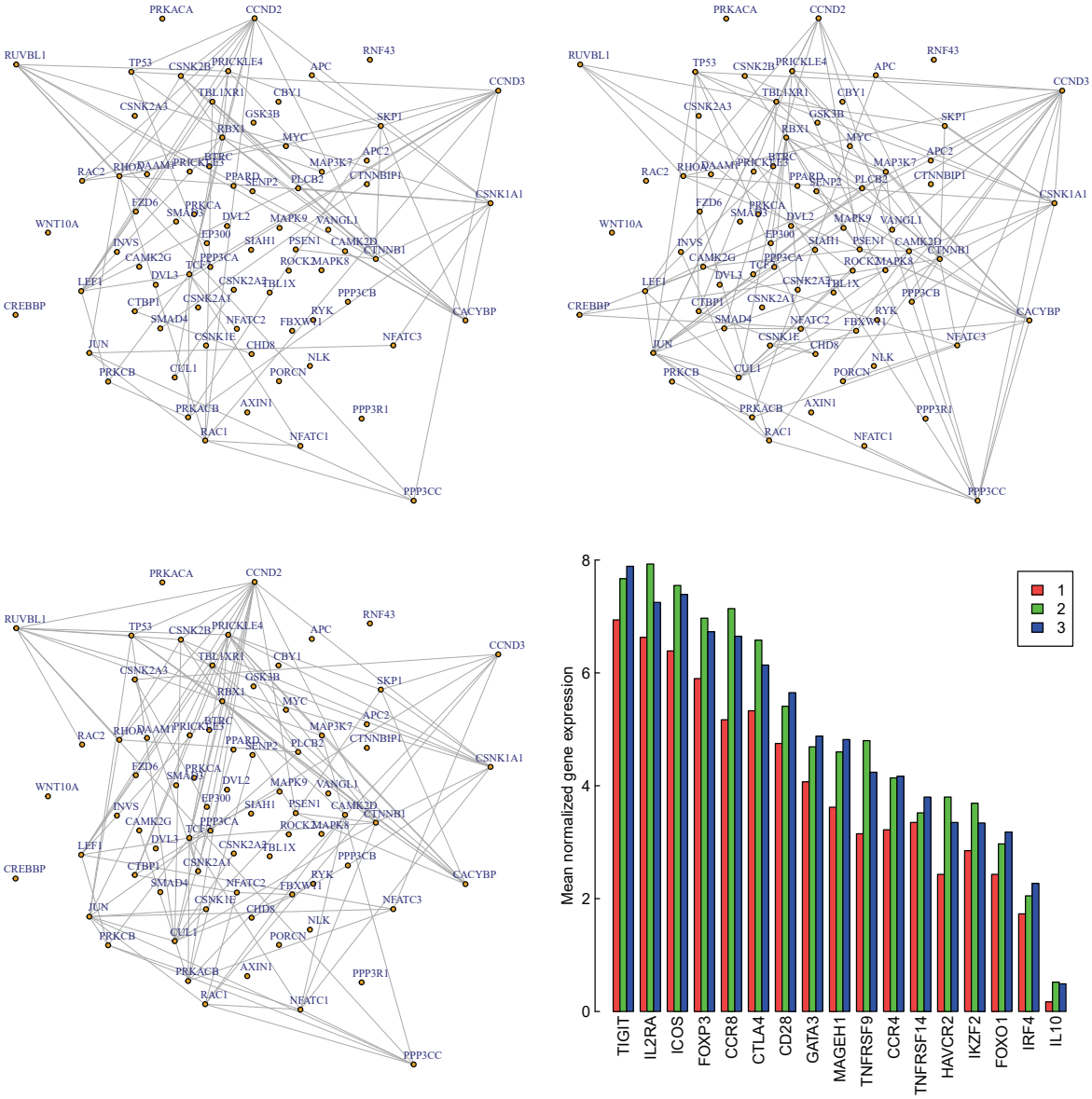

Figure S2. Analysis of regulatory T cells: identified subgrouping network structures using the SCAN method and mean normalized expressions of 17 immune related genes in the subgroups.

Table S1

Simulation results under the balanced design and $\mu = 2$: mean (sd).

Network	Method	per	$\widehat{K}_0$	CE	MSE($\boldsymbol{\mu}$)	MSE($\boldsymbol{\Theta}$)	TPR	FPR
tridiagonal	K-means	0	2(0)	0.296(0.103)	2.036(0.003)	—	—	—
	K-means+JGL	0	2(0)	0.296(0.103)	2.036(0.003)	1.487(0.003)	1.000(0.000)	0.132(0.015)
	SCAN($K = 2$)	0	2(0)	0.295(0.033)	2.030(0.002)	1.485(0.010)	1.000(0.000)	0.133(0.014)
	SCAN($K = 3$)	1	3(0)	0.000(0.000)	0.444(0.024)	1.380(0.032)	0.952(0.017)	0.101(0.014)
	SCAN($K = 4$)	0	4(0)	0.107(0.020)	0.709(0.004)	1.383(0.004)	0.952(0.013)	0.115(0.010)
	SCAN($K = 6$)	0	6(0)	0.201(0.012)	0.601(0.016)	1.421(0.002)	0.867(0.021)	0.090(0.005)
	FGGM	1	3(0)	0.000(0.000)	0.129(0.022)	1.279(0.044)	0.970(0.010)	0.112(0.012)
nearest-neighbor	K-means	1	3(0)	0.002(0.003)	0.495(0.004)	—	—	—
	K-means+JGL	1	3(0)	0.002(0.003)	0.495(0.004)	2.052(0.005)	0.848(0.004)	0.122(0.002)
	SCAN($K = 2$)	0	2(0)	0.286(0.093)	2.192(0.007)	2.290(0.073)	0.891(0.011)	0.159(0.013)
	SCAN($K = 3$)	1	3(0)	0.004(0.004)	0.462(0.029)	1.956(0.048)	0.858(0.021)	0.128(0.019)
	SCAN($K = 4$)	0	4(0)	0.104(0.021)	0.823(0.066)	2.050(0.041)	0.773(0.024)	0.100(0.008)
	SCAN($K = 6$)	0	6(0)	0.241(0.063)	0.711(0.051)	1.748(0.038)	0.750(0.022)	0.121(0.070)
	FGGM	1	3(0)	0.003(0.003)	0.140(0.019)	1.743(0.034)	0.877(0.017)	0.143(0.020)
power-law	K-means	0.99	2.99(0.10)	0.003(0.089)	0.544(0.149)	—	—	—
	K-means+JGL	0.99	2.99(0.10)	0.003(0.089)	0.544(0.149)	2.187(0.038)	0.846(0.006)	0.120(0.008)
	SCAN($K = 2$)	0	2(0)	0.197(0.095)	2.106(0.012)	2.518(0.060)	0.886(0.003)	0.167(0.009)
	SCAN($K = 3$)	1	3(0)	0.002(0.003)	0.532(0.029)	1.899(0.052)	0.861(0.017)	0.140(0.020)
	SCAN($K = 4$)	0	4(0)	0.096(0.012)	0.884(0.017)	1.953(0.037)	0.825(0.010)	0.171(0.009)
	SCAN($K = 6$)	0	6(0)	0.234(0.016)	0.653(0.046)	1.884(0.022)	0.683(0.011)	0.097(0.011)
	FGGM	1	3(0)	0.001(0.002)	0.140(0.022)	1.791(0.045)	0.872(0.015)	0.133(0.019)

Table S2

Simulation results under the imbalanced design and $\mu = 2$: mean (sd).

Network	Method	per	$\widehat{K}_0$	CE	MSE($\boldsymbol{\mu}$)	MSE($\boldsymbol{\Theta}$)	TPR	FPR
tridiagonal	K-means	0	2(0)	0.313(0.013)	1.779(0.026)	—	—	—
	K-means+JGL	0	2(0)	0.313(0.013)	1.779(0.026)	1.769(0.049)	0.999(0.002)	0.116(0.015)
	SCAN($K = 2$)	0	2(0)	0.304(0.024)	2.025(0.004)	1.470(0.034)	0.999(0.002)	0.108(0.011)
	SCAN($K = 3$)	1	3(0)	0.000(0.000)	0.490(0.027)	1.438(0.031)	0.924(0.021)	0.084(0.022)
	SCAN($K = 4$)	0	4(0)	0.118(0.020)	0.626(0.032)	1.369(0.056)	0.935(0.026)	0.115(0.023)
	SCAN($K = 6$)	0	6(0)	0.168(0.016)	0.765(0.031)	1.486(0.017)	0.884(0.010)	0.092(0.002)
	FGGM	0.96	3.12(0.59)	0.031(0.150)	0.130(0.022)	1.421(0.034)	0.945(0.022)	0.112(0.049)
nearest-neighbor	K-means	0	2(0)	0.314(0.030)	1.660(0.086)	—	—	—
	K-means+JGL	0	2(0)	0.314(0.030)	1.660(0.086)	2.454(0.063)	0.885(0.014)	0.139(0.021)
	SCAN($K = 2$)	0	2(0)	0.301(0.053)	2.189(0.121)	2.324(0.114)	0.890(0.020)	0.147(0.026)
	SCAN($K = 3$)	1	3(0)	0.003(0.003)	0.522(0.055)	2.037(0.056)	0.859(0.016)	0.144(0.020)
	SCAN($K = 4$)	0	4(0)	0.118(0.035)	0.821(0.038)	1.920(0.035)	0.823(0.016)	0.144(0.017)
	SCAN($K = 6$)	0	6(0)	0.235(0.034)	0.730(0.023)	1.788(0.024)	0.733(0.009)	0.125(0.002)
	FGGM	1	3(0)	0.002(0.003)	0.138(0.022)	1.809(0.038)	0.873(0.016)	0.146(0.021)
power-law	K-means	0	2(0)	0.312(0.018)	1.741(0.053)	—	—	—
	K-means+JGL	0	2(0)	0.312(0.018)	1.741(0.053)	2.491(0.081)	0.894(0.015)	0.162(0.021)
	SCAN($K = 2$)	0	2(0)	0.302(0.047)	2.090(0.021)	2.403(0.090)	0.904(0.015)	0.170(0.021)
	SCAN($K = 3$)	1	3(0)	0.002(0.002)	0.587(0.039)	2.092(0.057)	0.874(0.016)	0.163(0.018)
	SCAN($K = 4$)	0	4(0)	0.115(0.022)	0.787(0.050)	1.985(0.046)	0.829(0.020)	0.149(0.016)
	SCAN($K = 6$)	0	6(0)	0.235(0.035)	0.616(0.030)	1.883(0.011)	0.692(0.011)	0.107(0.003)
	FGGM	1	3(0)	0.001(0.002)	0.137(0.024)	1.835(0.036)	0.871(0.016)	0.141(0.022)

Table S3

Simulation results of the principal components based K-means+JGL method under the balanced design and $\mu = 1.5$: mean (sd).

Network	Method	per	$\widehat{K}_0$	CE	MSE($\boldsymbol{\mu}$)	MSE($\boldsymbol{\Theta}$)	TPR	FPR
tridiagonal	K-means+JGL	0	2(0)	0.225(0.002)	1.602(0.026)	1.529(0.041)	1.000(0.001)	0.124(0.017)
	80%PC+K-means+JGL	0	2(0)	0.225(0.002)	1.602(0.025)	1.528(0.041)	1.000(0.001)	0.125(0.017)
	90%PC+K-means+JGL	0	2(0)	0.225(0.002)	1.601(0.026)	1.526(0.040)	1.000(0.001)	0.125(0.017)
	95%PC+K-means+JGL	0	2(0)	0.225(0.002)	1.601(0.026)	1.526(0.040)	1.000(0.001)	0.125(0.017)
nearest-neighbor	K-means+JGL	0	2(0)	0.268(0.012)	1.401(0.048)	2.480(0.066)	0.881(0.019)	0.146(0.021)
	80%PC+K-means+JGL	0	2(0)	0.268(0.011)	1.403(0.051)	2.488(0.068)	0.878(0.018)	0.143(0.022)
	90%PC+K-means+JGL	0	2(0)	0.268(0.011)	1.401(0.050)	2.487(0.062)	0.879(0.016)	0.143(0.020)
	95%PC+K-means+JGL	0	2(0)	0.268(0.011)	1.402(0.052)	2.480(0.065)	0.880(0.018)	0.145(0.020)
power-law	K-means+JGL	0	2(0)	0.252(0.014)	1.494(0.063)	2.556(0.085)	0.893(0.016)	0.168(0.024)
	80%PC+K-means+JGL	0	2(0)	0.251(0.012)	1.497(0.055)	2.562(0.073)	0.893(0.015)	0.166(0.018)
	90%PC+K-means+JGL	0	2(0)	0.251(0.013)	1.496(0.059)	2.562(0.086)	0.893(0.016)	0.167(0.024)
	95%PC+K-means+JGL	0	2(0)	0.252(0.013)	1.494(0.065)	2.562(0.081)	0.893(0.016)	0.166(0.020)

Table S4

Sample overlaps of three subgroups identified by different methods in the analysis of regulatory T cells.

		SCAN			K-means+JGL		
		1	2	3	1	2	3
proposed	1	322	46	41	212	127	70
	2	57	122	137	75	99	142
	3	8	20	149	35	44	98

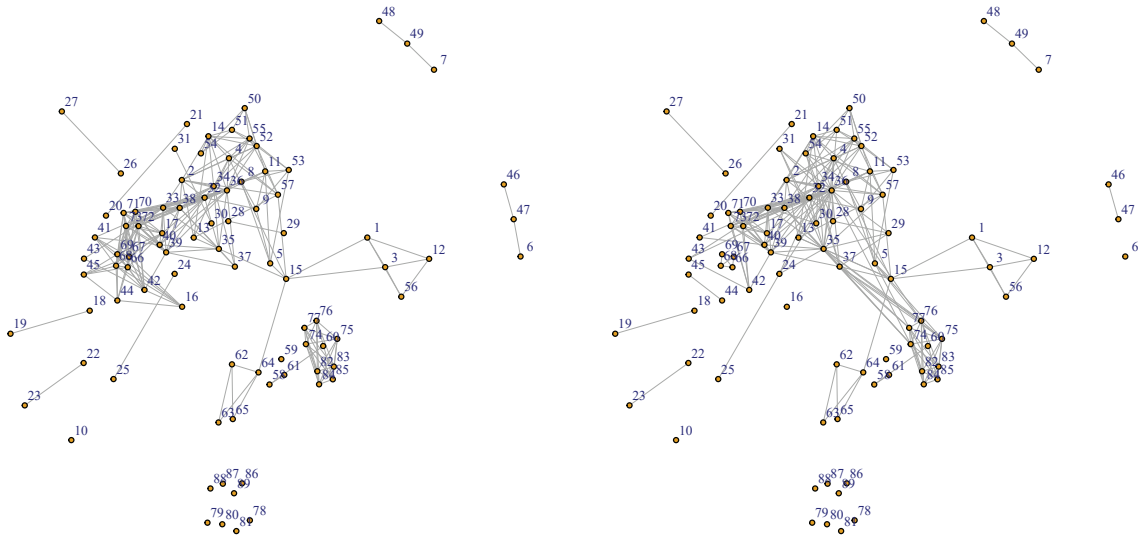

Figure S3. Analysis of LUAD histopathological imaging data: identified subgrouping network structures using the K-means+JGL method.

Table S5

Sample overlaps of two subgroups identified by different methods in the analysis of LUAD histopathological imaging data.

		SCAN		K-means+JGL	
		1	2	1	2
proposed	1	237	9	241	5
	2	28	33	3	58

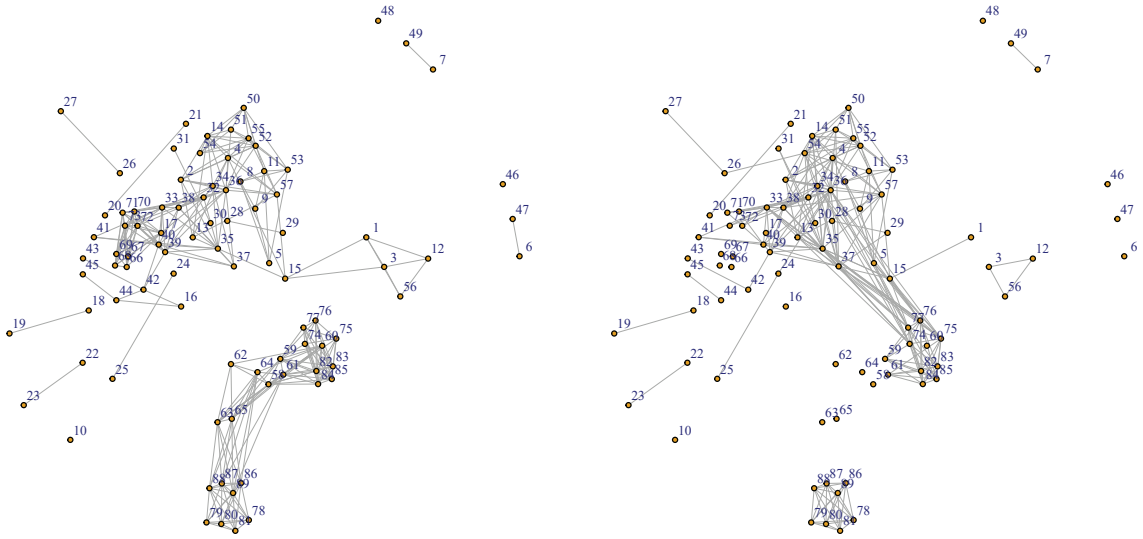

Figure S4. Analysis of LUAD histopathological imaging data: identified subgrouping network structures using the SCAN method.

Table S6

Imaging feature numbers and corresponding names in the LUAD data analysis.

Feature number	Image feature names	Feature number	Image feature names
1	AreaOccupied-AreaOccupied-Identifyeosinprimarycytoplasm	46	Location-Center-X
2	AreaOccupied-AreaOccupied-identifyhemaprimarynuclei	47	Location-Center-X.1
3	AreaOccupied-Perimeter-Identifyeosinprimarycytoplasm	48	Location-Center-Y
4	AreaOccupied-Perimeter-identifyhemaprimarynuclei	49	Location-Center-Y.1
5	AreaShape-Area	50	Neighbors-AngleBetweenNeighbors-Adjacent
6	AreaShape-Center-X	51	Neighbors-FirstClosestDistance-Adjacent
7	AreaShape-Center-Y	52	Neighbors-FirstClosestObjectNumber-Adjacent
8	AreaShape-MajorAxisLength	53	Neighbors-PercentTouching-Adjacent
9	AreaShape-MaxFeretDiameter	54	Neighbors-SecondClosestDistance-Adjacent
10	AreaShape-Orientation	55	Neighbors-SecondClosestObjectNumber-Adjacent
11	AreaShape-Perimeter	56	ObjectNumber
12	Count-Identifyeosinprimarycytoplasm	57	ObjectNumber.1
13	Count-identifyhemaprimarynuclei	58	Texture-Contrast-ImageAfterMath-3-00
14	Count-Identifyhemasub2	59	Texture-Contrast-ImageAfterMath-3-01
15	Count-identifytissueregion	60	Texture-Contrast-ImageAfterMath-3-02
16	Granularity-1-ImageAfterMath	61	Texture-Contrast-ImageAfterMath-3-03
17	Granularity-1-ImageAfterMath.1	62	Texture-Contrast-maskosingray-3-00
18	Granularity-10-ImageAfterMath	63	Texture-Contrast-maskosingray-3-01
19	Granularity-10-ImageAfterMath.1	64	Texture-Contrast-maskosingray-3-02
20	Granularity-11-ImageAfterMath	65	Texture-Contrast-maskosingray-3-03
21	Granularity-11-ImageAfterMath.1	66	Texture-SumAverage-ImageAfterMath-3-00
22	Granularity-12-ImageAfterMath	67	Texture-SumAverage-ImageAfterMath-3-01
23	Granularity-12-ImageAfterMath.1	68	Texture-SumAverage-ImageAfterMath-3-02
24	Granularity-13-ImageAfterMath	69	Texture-SumAverage-ImageAfterMath-3-03
25	Granularity-13-ImageAfterMath.1	70	Texture-SumAverage-maskosingray-3-00
26	Granularity-14-ImageAfterMath	71	Texture-SumAverage-maskosingray-3-01
27	Granularity-14-ImageAfterMath.1	72	Texture-SumAverage-maskosingray-3-02
28	Granularity-15-ImageAfterMath	73	Texture-SumAverage-maskosingray-3-03
29	Granularity-15-ImageAfterMath.1	74	Texture-SumVariance-ImageAfterMath-3-00
30	Granularity-16-ImageAfterMath	75	Texture-SumVariance-ImageAfterMath-3-01
31	Granularity-16-ImageAfterMath.1	76	Texture-SumVariance-ImageAfterMath-3-02
32	Granularity-2-ImageAfterMath	77	Texture-SumVariance-ImageAfterMath-3-03
33	Granularity-2-ImageAfterMath.1	78	Texture-SumVariance-maskosingray-3-00
34	Granularity-3-ImageAfterMath	79	Texture-SumVariance-maskosingray-3-01
35	Granularity-3-ImageAfterMath.1	80	Texture-SumVariance-maskosingray-3-02
36	Granularity-4-ImageAfterMath	81	Texture-SumVariance-maskosingray-3-03
37	Granularity-4-ImageAfterMath.1	82	Texture-Variance-ImageAfterMath-3-00
38	Granularity-5-ImageAfterMath	83	Texture-Variance-ImageAfterMath-3-01
39	Granularity-6-ImageAfterMath	84	Texture-Variance-ImageAfterMath-3-02
40	Granularity-7-ImageAfterMath	85	Texture-Variance-ImageAfterMath-3-03
41	Granularity-7-ImageAfterMath.1	86	Texture-Variance-maskosingray-3-00
42	Granularity-8-ImageAfterMath	87	Texture-Variance-maskosingray-3-01
43	Granularity-8-ImageAfterMath.1	88	Texture-Variance-maskosingray-3-02
44	Granularity-9-ImageAfterMath	89	Texture-Variance-maskosingray-3-03
45	Granularity-9-ImageAfterMath.1		