Meta Learning for Support Recovery in High-dimensional Precision Matrix Estimation: Supplementary Material

A. Proof of Lemma 1

Define $\mathcal{S}_{++}^{N} := \{A \in \mathbb{R}^{N \times N} | A \succ 0\}$. We first prove the following result:

Lemma 2. *For $\ell(\Omega)$ defined in (4), if $\Omega \in \mathcal{S}_{++}^{N}$, then $\ell(\Omega)$ is strictly convex.*

Proof. The gradient of $\ell(\Omega)$ is:

$$\nabla \ell\left(\Omega\right) = \sum_{k=1}^{K} T^{(k)} \left(\hat{\Sigma}^{(k)} - \Omega^{-1}\right) \tag{17}$$

The Hessian of $\ell(\Omega)$ is:

$$\nabla^2 \ell\left(\Omega\right) = T\Gamma\left(\Omega\right)$$

where $\Gamma\left(\Omega\right) = \Omega^{-1} \otimes \Omega^{-1} \in \mathbb{R}^{N^2 \times N^2}$.

Since $\Omega \in \mathcal{S}_{++}^{N}$, we have $\Omega \succ 0$ and thus $\Omega^{-1} \succ 0$. According to Theorem 4.2.12 in (Horn et al., 1994), any eigenvalue of $\Gamma\left(\Omega\right) = \Omega^{-1} \otimes \Omega^{-1}$ is the product of two eigenvalues of Ω^{-1}, hence positive. Therefore,

$$\Gamma\left(\Omega\right) \succ 0$$

$$\nabla^2 \ell\left(\Omega\right) \succ 0$$

$\ell(\Omega)$ is strictly convex. □

Now consider $\ell(\Omega) + \lambda \|\Omega\|_1$. Since $\lambda > 0$, by Lemma 2, we know $\ell(\Omega) + \lambda\|\Omega\|_1$ is strictly convex for $\Omega \in \mathcal{S}_{++}^{N}$. Therefore, the problem in (5) is strict convex and has a unique solution $\hat{\Omega}$.

For $\hat{\Omega}^{(K+1)}$ in (6), we have

$$\nabla \ell^{(K+1)}(\Omega) = \hat{\Sigma}^{(K+1)} - \Omega^{-1}$$

and

$$\nabla^2 \ell^{(K+1)}(\Omega) = \Gamma(\Omega) = \Omega^{-1} \otimes \Omega^{-1}$$

Thus according to the proof of Lemma 2, we know $\ell^{(K+1)}(\Omega)$ is strictly convex. Then $\ell^{(K+1)}(\Omega) + \lambda\|\Omega\|_1$ is strictly convex for $\lambda > 0$ on $\mathcal{S}_{++}^{N}$. Notice that the constraints $\mathrm{supp}(\Omega) \subseteq \mathrm{supp}(\hat{\Omega})$ and $\mathrm{diag}(\Omega) = \mathrm{diag}(\hat{\Omega})$ in (6) can be expressed as $\Omega_{ij} = 0$ for $(i,j) \notin S$ and $\Omega_{ii} = \hat{\Omega}_{ii}$ for $i \in \{1, \ldots, n\}$. Therefore the constraints are linear. Furthermore, (6) is strictly convex for $\lambda > 0$ on $\mathcal{S}_{++}^{N}$.

B. Proof of Theorem 1

Our proof follows the primal-dual witness approach (Ravikumar et al., 2011) which uses Karush-Kuhn Tucker conditions (from optimization) together with concentration inequalities (from statistical learning theory).

B.1. Preliminaries

Before the formal proof, we first introduce two inequalities with respect to the matrix ℓ_∞-operator-norm $|||\cdot|||_\infty$:

Lemma 3. *For a pair of matrices $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$ and a vector $x \in \mathbb{R}^n$, we have:*

$$\|Ax\|_\infty \leq |||A|||_\infty \|x\|_\infty \tag{18}$$

$$|||AB|||_\infty \leq |||A|||_\infty |||B|||_\infty \tag{19}$$

Proof. Note that

$$\begin{aligned}\|Ax\|_\infty &= \max_{1\le i\le m} |\langle a_i, x\rangle| \\ &\le \max_{1\le i\le m} \|a_i\|_1 \|x\|_\infty \\ &= |||A|||_\infty \|x\|_\infty\end{aligned}$$

where a_i is the vector corresponding to the i-th row of A and $\langle\cdot,\cdot\rangle$ is the inner product. Similarly, we have

$$\begin{aligned}\|AB\|_\infty &= \max_{1\le i\le m} \|a_i B\|_1 \\ &= \max_{1\le i\le m} \sum_{k=1}^{q} \Big| \sum_{j=1}^{n} A_{ij} B_{jk} \Big| \\ &\le \max_{1\le i\le m} \sum_{j=1}^{n} |A_{ij}| \sum_{k=1}^{q} |B_{jk}| \\ &\le \max_{1\le i\le m} \sum_{j=1}^{n} |A_{ij}| \max_{1\le l\le n} \sum_{k=1}^{q} |B_{lk}| \\ &= \max_{1\le i\le m} \sum_{j=1}^{n} |A_{ij}| |||B|||_\infty \\ &= |||A|||_\infty |||B|||_\infty\end{aligned}$$

□

Then we prove Theorem 1 with the five steps in the primal-dual witness approach.

B.2. Step 1

Let $(\Omega_S, 0)$ denote the $N \times N$ matrix such that $\Omega_{S^c} = 0$. For any $\Omega = (\Omega_S, 0) \in \mathcal{S}_{++}^N$, we need to verify that $\left[\nabla^2 \ell\left((\Omega_S, 0)\right)\right]_{SS} \succ 0$.

According to Lemma 2, since $(\Omega_S, 0) \in \mathcal{S}_{++}^N$, we have

$$\nabla^2 \ell\left((\Omega_S, 0)\right) \succ 0 \tag{20}$$

Denote the vectorization of a matrix A with vec(A) or $\overrightarrow{A}$. We use $|S|$ to denote the number of elements in S. Then we have $\left[\nabla^2 \ell\left((\Omega_S, 0)\right)\right]_{SS} \in \mathbb{R}^{|S|\times|S|}$. For $\forall x \in \mathbb{R}^{|S|}$, $v \neq 0$, there exists a matrix $A \in \mathbb{R}^{N\times N}$, $A \neq 0$, such that $\overrightarrow{A_S} = x$. Thus we have

$$\begin{aligned}x^{\mathrm{T}} \left[\nabla^2 \ell\left((\Omega_S, 0)\right)\right]_{SS} x &= \left[\overrightarrow{A_S}\right]^{\mathrm{T}} \left[\nabla^2 \ell\left((\Omega_S, 0)\right)\right]_{SS} \overrightarrow{A_S} \\ &= \left[\overrightarrow{(A_S, 0)}\right]^{\mathrm{T}} \nabla^2 \ell\left((\Omega_S, 0)\right) \overrightarrow{(A_S, 0)} \\ &> 0\end{aligned}$$

where the inequality follows from (20). Hence $\left[\nabla^2 \ell\left((\Omega_S, 0)\right)\right]_{SS} \succ 0$. Thus the step 1 in primal-dual witness is verified.

B.3. Step 2

Construct the primal variable $\tilde{\Omega}$ by making $\tilde{\Omega}_{S^c} = 0$ and solving the restricted problem:

$$\tilde{\Omega}_S = \arg\min_{(\Omega_S, 0)\in\mathcal{S}_{++}^N} \ell\left((\Omega_S, 0)\right) + \lambda \|\Omega_S\|_1 \tag{21}$$

B.4. Step 3

Choose the dual variable $\tilde{Z}$ in order to fulfill the complementary slackness condition of (5):

$$\begin{cases} \tilde{Z}_{ij} = 1, \text{ if } \tilde{\Omega}_{ij} > 0 \\ \tilde{Z}_{ij} = -1, \text{ if } \tilde{\Omega}_{ij} < 0 \\ \tilde{Z}_{ij} \in [-1, 1], \text{ if } \tilde{\Omega}_{ij} = 0 \end{cases} \tag{22}$$

Therefore we have

$$\|\tilde{Z}\|_\infty \leq 1 \tag{23}$$

B.5. Step 4

$\tilde{Z}$ is the subgradient of $\|\tilde{\Omega}\|_1$. Solve for the dual variable $\tilde{Z}_{S^c}$ in order that $(\tilde{\Omega}, \tilde{Z})$ fulfills the stationarity condition of (5):

$$\left[\nabla \ell \left(\left(\tilde{\Omega}_S, 0\right)\right)\right]_S + \lambda \tilde{Z}_S = 0 \tag{24}$$

$$\left[\nabla \ell \left(\left(\tilde{\Omega}_S, 0\right)\right)\right]_{S^c} + \lambda \tilde{Z}_{S^c} = 0 \tag{25}$$

B.6. Step 5

Now we need to verify that the dual variable solved by Step 4 satisfied the strict dual feasibility condition:

$$\|\tilde{Z}_{S^c}\|_\infty < 1 \tag{26}$$

which, according to the stationarity condition, is equivalent to

$$\frac{1}{\lambda} \| \left[\nabla \ell \left(\left(\tilde{\Omega}_S, 0\right)\right)\right]_{S^c} \|_\infty < 1 \tag{27}$$

This is the crucial part in the primal-dual witness approach. If we can show the strict dual feasibility condition holds, we can claim that the solution in (21) is equal to the solution in (5), i.e., $\tilde{\Omega} = \hat{\Omega}$. Thus we will have

$$\text{supp}\left(\hat{\Omega}\right) = \text{supp}\left(\tilde{\Omega}\right) \subseteq S = \text{supp}\left(\bar{\Omega}\right)$$

B.7. Proof of the Strict Dual Feasibility Condition

Plug the gradient of loss function (17) in the stationarity condition of (5), we have

$$\sum_{k=1}^{K} T^{(k)} \left(\hat{\Sigma}^{(k)} - \tilde{\Omega}^{-1}\right) + \lambda \tilde{Z} = 0 \tag{28}$$

Define $\bar{\Sigma} = \bar{\Omega}^{-1}$, $W^{(k)} := \hat{\Sigma}^{(k)} - \bar{\Sigma}$, $\Psi := \tilde{\Omega} - \bar{\Omega}$, $R(\Psi) := \tilde{\Omega}^{-1} - \bar{\Sigma} + \bar{\Omega}^{-1}\Psi\bar{\Omega}^{-1}$. Then we can rewrite (28) as

$$\sum_k T^{(k)} W^{(k)} + T\left(\bar{\Omega}^{-1}\Psi\bar{\Omega}^{-1} - R(\Psi)\right) + \lambda \tilde{Z} = 0 \tag{29}$$

From vectorization of product of matrices, we have:

$$\overrightarrow{\bar{\Omega}^{-1}\Psi\bar{\Omega}^{-1}} = \bar{\Gamma}\overrightarrow{\Psi} \tag{30}$$

where $\bar{\Gamma} := \bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1}$. Then vectorize both sides of (29) and we can get:

$$T\left(\bar{\Gamma}_{SS}\overrightarrow{\Psi_S} - \overrightarrow{R_S}\right) + \sum_{k=1}^{K} T^{(k)} \overrightarrow{W_S^{(k)}} + \lambda \overrightarrow{\tilde{Z}_S} = 0 \tag{31}$$

$$T\left(\bar{\Gamma}_{S^cS}\overrightarrow{\Psi_S} - \overrightarrow{R_{S^c}}\right) + \sum_{k=1}^{K} T^{(k)}\overrightarrow{W_{S^c}^{(k)}} + \lambda\overrightarrow{\tilde{Z}_{S^c}} = 0 \tag{32}$$

where we write $R(\Psi)$ as R for simplicity. By solving (31) for $\overrightarrow{\Psi_S}$, we get:

$$\overrightarrow{\Psi_S} = \frac{1}{T}\bar{\Gamma}_{SS}^{-1}\left(T\overrightarrow{R_S} - \sum_{k=1}^{K} T^{(k)}\overrightarrow{W_S^{(k)}} - \lambda\overrightarrow{\tilde{Z}_S}\right) \tag{33}$$

where we write $(\bar{\Gamma}_{SS})^{-1}$ as $\bar{\Gamma}_{SS}^{-1}$ for simplicity. Plug (33) in (32) to solve for $\overrightarrow{\tilde{Z}_{S^c}}$:

$$\begin{aligned}
\overrightarrow{\tilde{Z}_{S^c}} &= -\frac{1}{\lambda}T\bar{\Gamma}_{S^cS}\overrightarrow{\Psi_S} + \frac{1}{\lambda}T\overrightarrow{R_{S^c}} - \frac{1}{\lambda}\sum_{k=1}^{K} T^{(k)}\overrightarrow{W_{S^c}^{(k)}} \\
&= -\frac{1}{\lambda}\bar{\Gamma}_{S^cS}\bar{\Gamma}_{SS}^{-1}\left(T\overrightarrow{R_S} - \sum_{k=1}^{K} T^{(k)}\overrightarrow{W_S^{(k)}} - \lambda\overrightarrow{\tilde{Z}_S}\right) + \frac{1}{\lambda}T\overrightarrow{R_{S^c}} - \frac{1}{\lambda}\sum_{k=1}^{K} T^{(k)}\overrightarrow{W_{S^c}^{(k)}} \\
&= -\frac{1}{\lambda}\bar{\Gamma}_{S^cS}\bar{\Gamma}_{SS}^{-1}\left(T\overrightarrow{R_S} - \sum_{k=1}^{K} T^{(k)}\overrightarrow{W_S^{(k)}}\right) + \bar{\Gamma}_{S^cS}\bar{\Gamma}_{SS}^{-1}\overrightarrow{\tilde{Z}_S} + \frac{1}{\lambda}\left(T\overrightarrow{R_{S^c}} - \sum_{k=1}^{K} T^{(k)}\overrightarrow{W_{S^c}^{(k)}}\right)
\end{aligned}$$

According to (18) and the expression above, we have:

$$\begin{aligned}
\|\overrightarrow{\tilde{Z}_{S^c}}\|_\infty \le& \frac{1}{\lambda}\|\bar{\Gamma}_{S^cS}\bar{\Gamma}_{SS}^{-1}\left(T\overrightarrow{R_S} - \sum_{k=1}^{K} T^{(k)}\overrightarrow{W_S^{(k)}}\right)\|_\infty + \|\bar{\Gamma}_{S^cS}\bar{\Gamma}_{SS}^{-1}\overrightarrow{\tilde{Z}_S}\|_\infty \\
&+ \frac{1}{\lambda}\left(T\|\overrightarrow{R_{S^c}}\|_\infty + \|\sum_{k=1}^{K} T^{(k)}\overrightarrow{W_S^{(k)}}\|_\infty\right) \\
\le& \frac{1}{\lambda}|||\bar{\Gamma}_{S^cS}\bar{\Gamma}_{SS}^{-1}|||_\infty\left(T\|\overrightarrow{R_S}\|_\infty + \|\sum_{k=1}^{K} T^{(k)}\overrightarrow{W_S^{(k)}}\|_\infty\right) \\
&+ |||\bar{\Gamma}_{S^cS}\bar{\Gamma}_{SS}^{-1}|||_\infty + \frac{1}{\lambda}\left(T\|\overrightarrow{R_{S^c}}\|_\infty + \|\sum_{k=1}^{K} T^{(k)}\overrightarrow{W_S^{(k)}}\|_\infty\right)
\end{aligned}$$

where we have used $\|\overrightarrow{\tilde{Z}_S}\|_\infty \le 1$ by (23).

Therefore under Assumption 1, we have:

$$\|\tilde{Z}_{S^c}\|_\infty = \|\overrightarrow{\tilde{Z}_{S^c}}\|_\infty \le \frac{2-\alpha}{\lambda}\left(T\|\overrightarrow{R}\|_\infty + \|\sum_{k=1}^{K} T^{(k)}\overrightarrow{W^{(k)}}\|_\infty\right) + 1 - \alpha$$

If we can bound the two terms: $T\|\overrightarrow{R}\|_\infty, \|\sum_{k=1}^{K} T^{(k)}\overrightarrow{W^{(k)}}\|_\infty \le \frac{\alpha\lambda}{8}$, then we will have:

$$\|\tilde{Z}_{S^c}\|_\infty \le 1 - \frac{\alpha}{2} < 1$$

From all the reasoning so far, we have the following Lemma:

Lemma 4. *If we have* $T\|\overrightarrow{R(\Psi)}\|_\infty, \|\sum_{k=1}^{K} T^{(k)}\overrightarrow{W^{(k)}}\|_\infty \le \frac{\alpha\lambda}{8}$, *then*

$$\|\tilde{Z}_{S^c}\|_\infty < 1,$$

i.e., the strict-dual feasibility condition is fulfilled.

Thus the key step is to bound $T\|\overrightarrow{R}\|_\infty$ and $\|\sum_{k=1}^K T^{(k)}\overrightarrow{W^{(k)}}\|_\infty$ by $\frac{\alpha\lambda}{8}$. We will first consider $T\|\overrightarrow{R}\|_\infty$.

We have the following Lemma in (Ravikumar et al., 2011) (Lemma 5):

Lemma 5. *For any $\rho \in \mathbb{R}^{N\times N}$, If we have $\|\rho\|_\infty \leq \frac{1}{3}\kappa_{\bar{\Sigma}} d$, then the matrix $J(\rho) := \sum_{k=0}^\infty (-1)^k(\bar{\Omega}^{-1}\rho)^k$ will satisfy $\left|\!\left|\!\left| J^T \right|\!\right|\!\right|_\infty \leq \frac{3}{2}$ and the matrix $R(\rho) := (\bar{\Omega}+\rho)^{-1} - \bar{\Omega}^{-1} + \bar{\Omega}^{-1}\rho\bar{\Omega}^{-1}$ will satisfy:*

$$R(\rho) = \bar{\Omega}^{-1}\rho\bar{\Omega}^{-1}\rho J(\rho)\bar{\Omega}^{-1} \tag{34}$$

and

$$\|R(\rho)\|_\infty \leq \frac{3}{2} d\|\rho\|_\infty^2 \kappa_{\bar{\Sigma}}^3 \tag{35}$$

Here $\kappa_{\bar{\Sigma}} := \left|\!\left|\!\left|\bar{\Sigma}\right|\!\right|\!\right|_\infty = \left|\!\left|\!\left|\bar{\Omega}^{-1}\right|\!\right|\!\right|_\infty$, $d := \max_{1\leq i\leq N} \#\left\{j : 1 \leq j \leq N, \bar{\Omega}_{ij} \neq 0\right\}$.

For $R(\rho)$ defined in the above Lemma, we vectorize $R(\rho)_S$ and then we have

$$\begin{aligned}\overrightarrow{R(\rho)_S} =&\text{vec}\left(\left[(\bar{\Omega}+\rho)^{-1} - \bar{\Omega}^{-1}\right]_S\right) + \text{vec}\left([\bar{\Omega}^{-1}\rho\bar{\Omega}^{-1}]_S\right)\\ =&\text{vec}\left(\left[(\bar{\Omega}+\rho)^{-1}\right]_S - \left[\bar{\Omega}^{-1}\right]_S\right) + \bar{\Gamma}_{SS}\overrightarrow{\rho_S}\end{aligned} \tag{36}$$

where the first line follows from the definition of $R(\rho)$ in Lemma 5 and the second line follows from (30)

Define $\kappa_{\bar{\Gamma}} := \left|\!\left|\!\left|\bar{\Gamma}_{SS}^{-1}\right|\!\right|\!\right|_\infty$. For $\Omega \in \mathbb{R}^{N\times N}$, define the subgradient of (21) as $G(\Omega_S)$, i.e., $G(\Omega_S) := -T[\Omega^{-1}]_S + \sum_{k=1}^K T^{(k)}\hat{\Sigma}_S^{(k)} + \lambda\tilde{Z}_S$. Since we have proved in Step 1 that ℓ is strictly convex, $\tilde{\Omega}_S$ is the only solution of the restricted problem of (21). Therefore $\tilde{\Omega}_S$ is the only solution that satisfies the stationary condition $G(\Omega_S) = 0$.

Next for $\rho \in \mathbb{R}^{N\times N}$, define $F(\overrightarrow{\rho_S}) = -\frac{1}{T}\bar{\Gamma}_{SS}^{-1}\overrightarrow{G}(\bar{\Omega}_S + \rho_S) + \overrightarrow{\rho_S}$. Then:

$$F(\overrightarrow{\rho_S}) = \overrightarrow{\rho_S} \Leftrightarrow G(\bar{\Omega}_S + \rho_S) = 0 \Leftrightarrow \bar{\Omega}_S + \rho_S = \tilde{\Omega}_S$$

Thus the fixed point of $F(\cdot)$ is $\Psi_S = \tilde{\Omega}_S - \bar{\Omega}_S$ and it is unique.

Now define $r := 2\kappa_{\bar{\Gamma}}\left(\frac{\lambda}{T} + \|\sum_{k=1}^K \frac{T^{(k)}}{T}W^{(k)}\|_\infty\right)$. Suppose $r \leq \min\left\{\frac{1}{3\kappa_{\bar{\Sigma}}d}, \frac{1}{3\kappa_{\bar{\Sigma}}^3\kappa_{\bar{\Gamma}}d}\right\}$. Define the ℓ_∞ radius-r ball $\mathbb{B}(r) := \{\rho_S : \|\rho_S\|_\infty \leq r\}$. For $\forall\rho_S \in \mathbb{B}(r)$, define $\rho = (\rho_S, 0)$, i.e., $[\rho]_S = \rho_S$ and $[\rho]_{S^c} = 0$. We have:

$$G(\bar{\Omega}_S + \rho_S) = T\left(-[(\bar{\Omega}+\rho)^{-1}]_S + [\bar{\Omega}^{-1}]_S\right) + \sum_{k=1}^K T^{(k)}W_S^{(k)} + \lambda\tilde{Z}_S$$

Then,

$$\begin{aligned}F(\overrightarrow{\rho_S}) =& -\frac{1}{T}\bar{\Gamma}_{SS}^{-1}\text{vec}\left(T\left(-[(\bar{\Omega}+\rho)^{-1}]_S + [\bar{\Omega}^{-1}]_S\right) + \sum_{k=1}^K T^{(k)}W_S^{(k)} + \lambda\tilde{Z}_S\right) + \overrightarrow{\rho_S}\\ =&\bar{\Gamma}_{SS}^{-1}\left\{\text{vec}\left([(\bar{\Omega}+\rho)^{-1}]_S - [\bar{\Omega}^{-1}]_S\right) + \bar{\Gamma}_{SS}\overrightarrow{\rho_S}\right\} - \frac{1}{T}\bar{\Gamma}_{SS}^{-1}\text{vec}\left(\sum_{k=1}^K T^{(k)}W_S^{(k)} + \lambda\tilde{Z}_S\right)\\ =&\underbrace{\bar{\Gamma}_{SS}^{-1}\overrightarrow{R(\rho)_S}}_{V_1} - \underbrace{\frac{1}{T}\bar{\Gamma}_{SS}^{-1}\left(\sum_{k=1}^K T^{(k)}\overrightarrow{W_S^{(k)}} + \lambda\overrightarrow{\tilde{Z}_S}\right)}_{V_2}\end{aligned} \tag{37}$$

where the third line follows from (36). For V_2 defined above we have:

$$\begin{aligned}\|V_2\|_\infty \leq&\left|\!\left|\!\left|\bar{\Gamma}_{SS}^{-1}\right|\!\right|\!\right|_\infty \|\frac{\lambda}{T}\overrightarrow{\tilde{Z}_S} + \sum_{k=1}^K \frac{T^{(k)}}{T}W^{(k)}\|_\infty\\ \leq&\kappa_{\bar{\Gamma}}\left(\frac{\lambda}{T} + \|\sum_{k=1}^K \frac{T^{(k)}}{T}W^{(k)}\|_\infty\right)\\ =&\frac{r}{2}\end{aligned}$$

where the first inequality follows from (18), the second inequality follows from (23) and the third line follows from the definition of r.

For V_1 defined in (37) we have:

$$\begin{aligned}\|V_1\|_\infty &\le \left|\left|\left|\bar{\Gamma}_{SS}^{-1}\right|\right|\right|_\infty \|R(\rho)_S\|_\infty \\ &\le \kappa_{\bar{\Gamma}}\|R(\rho)\|_\infty \\ &\le \kappa_{\bar{\Gamma}}\left(\frac{3}{2}d\kappa_{\bar{\Sigma}}^3\right)\|\rho\|_\infty^2 \\ &\le \frac{3}{2}d\kappa_{\bar{\Sigma}}^3\kappa_{\bar{\Gamma}}r^2 \\ &\le \frac{r}{2}\end{aligned} \tag{38}$$

where the first inequality is due to (18) and the second inequality is due to Lemma 5 and $\|\rho\|_\infty = \|\rho_S\|_\infty \le r$.

Thus $\|F(\overrightarrow{\rho_S})\|_\infty \le r$, $F(\overrightarrow{\rho_S}) \in \mathbb{B}(r)$, which indicates $F(\mathbb{B}(r)) \subset \mathbb{B}(r)$. By Brouwer's fixed point theorem (see e.g., (Ortega & Rheinboldt, 2000)), there exists some fixed point of $F(\cdot)$ in $\mathbb{B}(r)$. We have proved that the fixed point of $F(\cdot)$ is Ψ_S and it is unique, therefore $\Psi_S \in \mathbb{B}(r)$, i.e., $\|\Psi\|_\infty = \|\Psi_S\|_\infty \le r$. Thus by Lemma 5, $\|R(\Psi)\|_\infty \le \frac{3}{2}d\|\Psi\|_\infty^2\kappa_{\bar{\Sigma}}^3$.

From all the reasoning so far, we have the following Lemma:

Lemma 6. *If* $r = 2\kappa_{\bar{\Gamma}}\left(\frac{\lambda}{T} + \|\sum_{k=1}^K \frac{T^{(k)}}{T}W^{(k)}\|_\infty\right) \le \min\left\{\frac{1}{3\kappa_{\bar{\Sigma}}d}, \frac{1}{3\kappa_{\bar{\Sigma}}^3\kappa_{\bar{\Gamma}}d}\right\}$*, then*

$$\|\Psi\|_\infty \le r$$

and

$$\|R(\Psi)\|_\infty \le \frac{3}{2}d\|\Psi\|_\infty^2\kappa_{\bar{\Sigma}}^3$$

If $\|\sum_{k=1}^K \frac{T^{(k)}}{T}W^{(k)}\|_\infty \le \xi$ with $\xi > 0$, then choosing $\lambda = \frac{8T\xi}{\alpha}$, we will have

$$\|\sum_{k=1}^K T^{(k)}W^{(k)}\|_\infty \le \frac{\alpha\lambda}{8}$$

as well as

$$r = 2\kappa_{\bar{\Gamma}}\left(\frac{\lambda}{T} + \|\sum_{k=1}^K \frac{T^{(k)}}{T}W^{(k)}\|_\infty\right) \le 2\kappa_{\bar{\Gamma}}\left(\frac{8}{\alpha}+1\right)\xi$$

For $\xi \le \delta^* := \frac{\alpha^2}{2\kappa_{\bar{\Gamma}}(\alpha+8)^2}\min\left\{\frac{1}{3\kappa_{\bar{\Sigma}}d}, \frac{1}{3\kappa_{\bar{\Sigma}}^3\kappa_{\bar{\Gamma}}d}\right\}$, we have $r \le \min\left\{\frac{1}{3\kappa_{\bar{\Sigma}}d}, \frac{1}{3\kappa_{\bar{\Sigma}}^3\kappa_{\bar{\Gamma}}d}\right\}$. Thus according to Lemma 6, we have

$$\|\Psi\|_\infty = \|\Psi_S\|_\infty \le r \le 2\kappa_{\bar{\Gamma}}\left(\frac{8}{\alpha}+1\right)\xi$$

Therefore,

$$\begin{aligned}\|R(\Psi)\|_\infty \le& \frac{3}{2}d\|\Psi\|_\infty^2\kappa_{\bar{\Sigma}}^3 \\ &\le 6d\kappa_{\bar{\Sigma}}^3\kappa_{\bar{\Gamma}}^2\left(\frac{8}{\alpha}+1\right)^2\delta^2 \\ &= \left(6d\kappa_{\bar{\Sigma}}^3\kappa_{\bar{\Gamma}}^2\left(\frac{8}{\alpha}+1\right)^2\xi\right)\frac{\alpha\lambda}{8T} \\ &\le \frac{\alpha\lambda}{8T}\end{aligned}$$

Then by Lemma 4, $\|\tilde{Z}_{S^c}\|_\infty < 1$ and the strict dual feasibility condition is fulfilled. According to the primal-dual witness approach, $\text{supp}(\hat{\Omega}) = \text{supp}(\tilde{\Omega}) \subseteq \text{supp}\left(\bar{\Omega}\right)$.

From all the reasoning so far, we can state the following lemma.

Lemma 7. *If* $\|\sum_{k=1}^{K} \frac{T^{(k)}}{T} W^{(k)}\|_{\infty} \leq \xi$ *with* $\xi \in (0, \delta^*]$*, then choosing* $\lambda = \frac{8T\xi}{\alpha}$*, we have* $\hat{\Omega} = \tilde{\Omega}$*, supp*$(\hat{\Omega}) \subseteq$ *supp* $(\bar{\Omega})$ *and*

$$\|\hat{\Omega} - \bar{\Omega}\|_{\infty} = \|\Psi\|_{\infty} \leq 2\kappa_{\bar{\Gamma}} \left(\frac{8}{\alpha} + 1\right) \xi$$

For the next step, we need to prove the tail condition of $\sum_{k=1}^{K} \frac{T^{(k)}}{T} W^{(k)}$, that is, for $\xi > 0$, $\|\sum_{k=1}^{K} \frac{T^{(k)}}{T} W^{(k)}\|_{\infty} \leq \xi$ with high probability.

B.8. Proof of the Tail Condition

Note that for $k = 1, \dots, K$,

$$W^{(k)} = \hat{\Sigma}^{(k)} - \bar{\Sigma} = \hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} + \bar{\Sigma}^{(k)} - \bar{\Sigma} = \hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} + \left(\bar{\Omega} + \Delta^{(k)}\right)^{-1} - \bar{\Sigma} \tag{39}$$

Here $\{\Delta^{(k)}\}_{k=1}^{K}$ are i.i.d. random matrices following the distribution P specified in Definition 3. To achieve the tail condition of $\sum_{k=1}^{K} \frac{T^{(k)}}{T} W^{(k)}$, we can bound the random terms with respect to $\{\Delta^{(k)}\}_{k=1}^{K}$ and the random terms with respect to the empirical sample covariance matrices $\{\hat{\Sigma}^{(k)}\}_{k=1}^{K}$ separately.

We have assumed that the sample size is the same for all tasks, i.e., there are n samples for each of the K tasks and $T^{(k)}/T = 1/K$. For the sample covariance matrices, we have the following lemma:

Lemma 8. *For* $\{X_t^{(k)}\}_{1 \leq t \leq n, 1 \leq k \leq K}$ *following a family of random* N*-dimensional multivariate sub-Gaussian distributions of size* K *with parameter* σ *described in Definition 3, we have*

$$\mathbb{P}\left[\left|\sum_{k=1}^{K} \frac{1}{K}\left(\hat{\Sigma}_{ij}^{(k)} - \bar{\Sigma}_{ij}^{(k)}\right)\right| > \nu\right] \leq \exp\left\{-\frac{nK\nu^2}{128\left(1 + 4\sigma^2\right)^2 \gamma^2}\right\} \tag{40}$$

and

$$\mathbb{P}\left[\|\sum_{k=1}^{K} \frac{1}{K}\left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)}\right)\|_{\infty} > \nu\right] \leq 2N(N+1)\exp\left\{-\frac{nK\nu^2}{128\left(1 + 4\sigma^2\right)^2 \gamma^2}\right\} \tag{41}$$

for $\hat{\Sigma}^{(k)} = \frac{1}{n}\sum_{t=1}^{n} X_t^{(k)} (X_t^{(k)})^T$*,* $1 \leq i, j \leq N$*, and* $0 \leq \nu \leq 8\left(1 + 4\sigma^2\right)\gamma$*.*

The proof of this lemma is in Section G.

For $\{\Delta^{(1)}\}_{k=1}^{K}$, we have the following lemma

Lemma 9. *For* $\{\Delta^{(k)}\}_{k=1}^{K}$ *in a family of random* N*-dimensional multivariate sub-Gaussian distributions of size* K *with parameter* σ *described in Definition 3, define*

$$H(\Delta^{(1)}, \dots, \Delta^{(K)}) := \frac{1}{K}\sum_{k=1}^{K} \bar{\Sigma}^{(k)} = \frac{1}{K}\sum_{k=1}^{K}\left(\bar{\Omega} + \Delta^{(k)}\right)^{-1} \tag{42}$$

Then we have

$$\mathbb{P}\left[|||H - \mathbb{E}[H]|||_2 > t\right] \leq 2N\exp\left\{-\frac{\lambda_{\min}^4 K t^2}{128 c_{\max}^2}\right\} \tag{43}$$

for $t \geq 0$ *and* $\lambda_{\min} = \lambda_{\min}(\bar{\Omega})$*.*

The proof of this lemma is in Section H.

Our goal is to find a probability upper bound for $\|\sum_{k=1}^{K} \frac{T^{(k)}}{T} W^{(k)}\|_{\infty} > \xi$ with $0 < \xi \leq \delta^*$. According to (39) and the condition $\beta \leq \delta^*/2$, we have

$$
\begin{aligned}
\| \sum_{k=1}^{K} \frac{T^{(k)}}{T} W^{(k)} \|_\infty =& \| \sum_{k=1}^{K} \frac{1}{K} W^{(k)} \|_\infty \\
\leq& \| \sum_{k=1}^{K} \frac{1}{K} \left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} \right) \|_\infty + \| \frac{1}{K} \sum_{k=1}^{K} \bar{\Sigma}^{(k)} - \bar{\Sigma} \|_\infty \\
=& \| \sum_{k=1}^{K} \frac{1}{K} \left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} \right) \|_\infty + \| H - \mathbb{E}[H] + \mathbb{E}[H] - \bar{\Sigma} \|_\infty \\
=& \| \sum_{k=1}^{K} \frac{1}{K} \left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} \right) \|_\infty + \| H - \mathbb{E}[H] \|_2 + \| \mathbb{E}[H] - \bar{\Sigma} \|_\infty \\
\leq& \| \sum_{k=1}^{K} \frac{1}{K} \left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} \right) \|_\infty + \| H - \mathbb{E}[H] \|_2 + \beta
\end{aligned}
\tag{44}
$$

where we have used the property that $\| A \|_2 \geq \| A \|_\infty$ for any matrix A (see e.g., (Horn & Johnson, 2012)).

Now for $\delta \in (0, \delta^*/2]$, consider

$$
\xi = \delta + \delta^*/2 \tag{45}
$$

then $0 < \xi \leq \delta^*$, $\delta + \tau \leq \xi$ and $\lambda = \frac{8T\xi}{\alpha} = \frac{8\delta + 4\delta^*}{\alpha}$.

According to the condition $\beta \leq \delta^*/2$, we know that $\delta^*/2 - \beta \geq 0$. Set $t = \delta^*/2 - \beta$ in (43). Then,

$$
\mathbb{P}\left[\| H - \mathbb{E}[H] \|_2 > \delta^*/2 - \beta \right] \leq 2N \exp\left(-\frac{\lambda_{\min}^4 K}{128 c_{\max}^2} \left(\frac{\delta^*}{2} - \beta \right)^2 \right) \tag{46}
$$

By (44) and (45), we have

$$
\left\{ \| \sum_{k=1}^{K} \frac{1}{K} \left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} \right) \|_\infty \leq \delta \text{ and } \| H - \mathbb{E}[H] \|_2 \leq \frac{\delta^*}{2} - \beta \right\} \Rightarrow \left\{ \| \sum_{k=1}^{K} \frac{T^{(k)}}{T} W^{(k)} \|_\infty \leq \xi \right\}
$$

and thus

$$
\begin{aligned}
\mathbb{P}\left[\| \sum_{k=1}^{K} \frac{T^{(k)}}{T} W^{(k)} \|_\infty \leq \xi \right] \geq& \mathbb{P}\left[\| \sum_{k=1}^{K} \frac{1}{K} \left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} \right) \|_\infty \leq \delta \text{ and } \| H - \mathbb{E}[H] \|_2 \leq \frac{\delta^*}{2} - \beta \right] \\
=& 1 - \mathbb{P}\left[\| \sum_{k=1}^{K} \frac{1}{K} \left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} \right) \|_\infty > \delta \text{ or } \| H - \mathbb{E}[H] \|_2 > \frac{\delta^*}{2} - \beta \right] \\
\geq& 1 - \left(\mathbb{P}\left[\| \sum_{k=1}^{K} \frac{1}{K} \left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} \right) \|_\infty > \delta \right] + \mathbb{P}\left[\| H - \mathbb{E}[H] \|_2 > \frac{\delta^*}{2} - \beta \right] \right) \\
=& 1 - \mathbb{P}\left[\| \sum_{k=1}^{K} \frac{1}{K} \left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} \right) \|_\infty > \delta \right] - 2N \exp\left(-\frac{\lambda_{\min}^4 K}{128 c_{\max}^2} \left(\frac{\delta^*}{2} - \beta \right)^2 \right)
\end{aligned}
\tag{47}
$$

where we have applied (46) for the last step.

When $0 < \delta < 8\left(1 + 4\sigma^2\right)\gamma$, we can let $\nu = \delta$ in (41) to get

$$
\mathbb{P}\left[\| \sum_{k=1}^{K} \frac{1}{K} \left(\hat{\Sigma}^{(k)} - \bar{\Sigma}^{(k)} \right) \|_\infty > \delta \right] \leq 1 - 2N(N+1) \exp\left\{ -\frac{nK\delta^2}{128\left(1 + 4\sigma^2\right)^2 \gamma^2} \right\} \tag{48}
$$

When $\delta \geq 8\left(1+4\sigma^2\right)\gamma$, we set $\nu = 8\left(1+4\sigma^2\right)\gamma$ in (41) to get

$$\begin{aligned}\mathbb{P}\left[\|\sum_{k=1}^{K}\frac{1}{K}\left(\hat{\Sigma}^{(k)}-\bar{\Sigma}^{(k)}\right)\|_\infty > \delta\right] \leq & \mathbb{P}\left[\|\sum_{k=1}^{K}\frac{1}{K}\left(\hat{\Sigma}^{(k)}-\bar{\Sigma}^{(k)}\right)\|_\infty > 8\left(1+4\sigma^2\right)\gamma\right] \\ \leq & 2N(N+1)\exp\left\{-\frac{nK(8\left(1+4\sigma^2\right)\gamma)^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\} \\ = & 2N(N+1)\exp\left\{-\frac{nK}{2}\right\}\end{aligned} \tag{49}$$

Consider the maximum value of the two upper bounds in (48) and (49). We can get

$$\begin{aligned}\mathbb{P}\left[\|\sum_{k=1}^{K}\frac{1}{K}\left(\hat{\Sigma}^{(k)}-\bar{\Sigma}^{(k)}\right)\|_\infty > \delta\right] \leq & \max\left\{2N(N+1)\exp\left\{-\frac{nK\delta^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\}, 2N(N+1)\exp\left\{-\frac{nK}{2}\right\}\right\} \\ = & 2N(N+1)\exp\left(-\frac{nK}{2}\min\left\{\frac{\delta^2}{64\left(1+4\sigma^2\right)^2\gamma^2}, 1\right\}\right)\end{aligned} \tag{50}$$

According to (47) and (50), we have

$$\begin{aligned}\mathbb{P}\left[\|\sum_{k=1}^{K}\frac{T^{(k)}}{T}W^{(k)}\|_\infty \leq \xi\right] \geq & 1-\mathbb{P}\left[\|\sum_{k=1}^{K}\frac{1}{K}\left(\hat{\Sigma}^{(k)}-\bar{\Sigma}^{(k)}\right)\|_\infty > \delta\right] - 2N\exp\left(-\frac{\lambda_{\min}^4 K}{128c_{\max}^2}\left(\frac{\delta^*}{2}-\beta\right)^2\right) \\ \geq & 1-2N(N+1)\exp\left(-\frac{nK}{2}\min\left\{\frac{\delta^2}{64\left(1+4\sigma^2\right)^2\gamma^2}, 1\right\}\right) \\ & -2N\exp\left(-\frac{\lambda_{\min}^4 K}{128c_{\max}^2}\left(\frac{\delta^*}{2}-\beta\right)^2\right)\end{aligned} \tag{51}$$

Namely, with probability at least

$$1-2N(N+1)\exp\left(-\frac{nK}{2}\min\left\{\frac{\delta^2}{64\left(1+4\sigma^2\right)^2\gamma^2}, 1\right\}\right) - 2N\exp\left(-\frac{\lambda_{\min}^4 K}{128c_{\max}^2}\left(\frac{\delta^*}{2}-\beta\right)^2\right)$$

we have $\|\sum_{k=1}^{K}\frac{T^{(k)}}{T}W^{(k)}\|_\infty \leq \xi \leq \delta^*$, supp$(\hat{\Omega}) \subseteq$ supp $\left(\bar{\Omega}\right)$ and according to Lemma 7, we have

$$\|\hat{\Omega}-\bar{\Omega}\|_\infty = \|\Delta\|_\infty \leq 2\kappa_{\bar{\Gamma}}\left(\frac{8}{\alpha}+1\right)\xi = \kappa_{\bar{\Gamma}}\left(\frac{8}{\alpha}+1\right)(2\delta+\delta^*)$$

which completes our proof of Theorem 1.

C. Proof of Theorem 2

We have the following lemma as a sufficient condition for the sign-consistency of (5).

Lemma 10. *For $\xi \in (0, \delta^*]$, if*

$$\|\sum_{k=1}^{K}\frac{T^{(k)}}{T}W^{(k)}\|_\infty \leq xi \tag{52}$$

and

$$\frac{\omega_{\min}}{2} \geq 2\kappa_{\bar{\Gamma}}\left(\frac{8}{\alpha}+1\right)\xi \tag{53}$$

where $\omega_{\min} := \min_{(i,j)\in S}|\bar{\Omega}_{ij}|$, then the estimate $\hat{\Omega}$ of (5) is sign-consistent.

The proof is in Section I.

In the remaining part of the proof, we assume that the condition $\beta \leq \delta^\dagger/2$ stated in Theorem 2 is satisfied. We will consider two cases for different $\omega_{\min} > 0$.

Case (i). If

$$\omega_{\min} \geq \frac{2\alpha}{8+\alpha} \min\left\{\frac{1}{3\kappa_{\bar{\Sigma}} d}, \frac{1}{3\kappa_{\bar{\Sigma}}^3 \kappa_{\bar{\Gamma}} d}\right\} \tag{54}$$

then

$$0 < \delta^\dagger = \delta^*$$

and

$$\frac{\omega_{\min}}{2} \geq 2\kappa_{\bar{\Gamma}} \left(\frac{8}{\alpha} + 1\right) \delta^*$$

Thus for $\xi = \delta^*$, (53) holds. Then according to (51), with probability at least

$$\begin{aligned} & 1 - 2N(N+1)\exp\left(-\frac{nK}{2}\min\left\{\frac{(\delta^*/2)^2}{64\left(1+4\sigma^2\right)^2\gamma^2}, 1\right\}\right) - 2N\exp\left(-\frac{\lambda_{\min}^4 K}{128 c_{\max}^2}\left(\frac{\delta^*}{2} - \beta\right)^2\right) \\ = {} & 1 - 2N(N+1)\exp\left(-\frac{nK}{2}\min\left\{\frac{(\delta^\dagger)^2}{256\left(1+4\sigma^2\right)^2\gamma^2}, 1\right\}\right) - 2N\exp\left(-\frac{\lambda_{\min}^4 K}{128 c_{\max}^2}\left(\frac{\delta^*}{2} - \beta\right)^2\right) \end{aligned}$$

we have $\|\sum_{k=1}^K \frac{T^{(k)}}{T} W^{(k)}\|_\infty \leq \delta^*$ and thus by Lemma 10, we have that (5) is sign-consistent.

Case (ii). If

$$\omega_{\min} < \frac{2\alpha}{8+\alpha} \min\left\{\frac{1}{3\kappa_{\bar{\Sigma}} d}, \frac{1}{3\kappa_{\bar{\Sigma}}^3 \kappa_{\bar{\Gamma}} d}\right\},$$

then

$$\frac{\omega_{\min}}{2} < 2\kappa_{\bar{\Gamma}} \left(\frac{8}{\alpha} + 1\right) \delta^*$$

and

$$0 < \delta^\dagger = \delta' \leq \delta^*$$

Thus

$$\frac{\omega_{\min}}{2} \geq 2\kappa_{\bar{\Gamma}} \left(\frac{8}{\alpha} + 1\right) \delta' \tag{55}$$

Now apply (51) with $\xi = \delta' = \delta^\dagger$, we have

$$\begin{aligned} \mathbb{P}\left[\|\sum_{k=1}^K \frac{T^{(k)}}{T} W^{(k)}\|_\infty \leq \delta'\right] \geq {} & 1 - 2N(N+1)\exp\left(-\frac{nK}{2}\min\left\{\frac{(\delta' - \delta^*/2)^2}{64\left(1+4\sigma^2\right)^2\gamma^2}, 1\right\}\right) \\ & - 2N\exp\left(-\frac{\lambda_{\min}^4 K}{128 c_{\max}^2}\left(\frac{\delta^*}{2} - \beta\right)^2\right) \\ \geq {} & 1 - 2N(N+1)\exp\left(-\frac{nK}{2}\min\left\{\frac{(\delta^\dagger - \delta^\dagger/2)^2}{64\left(1+4\sigma^2\right)^2\gamma^2}, 1\right\}\right) \\ & - 2N\exp\left(-\frac{\lambda_{\min}^4 K}{128 c_{\max}^2}\left(\frac{\delta^*}{2} - \beta\right)^2\right) \\ = {} & 1 - 2N(N+1)\exp\left(-\frac{nK}{2}\min\left\{\frac{(\delta^\dagger)^2}{256\left(1+4\sigma^2\right)^2\gamma^2}, 1\right\}\right) \\ & - 2N\exp\left(-\frac{\lambda_{\min}^4 K}{128 c_{\max}^2}\left(\frac{\delta^*}{2} - \beta\right)^2\right) \end{aligned}$$

Therefore with probability at least

$$1 - 2N(N+1)\exp\left(-\frac{nK}{2}\min\left\{\frac{(\delta^{\dagger})^2}{256\left(1+4\sigma^2\right)^2\gamma^2}, 1\right\}\right) - 2N\exp\left(-\frac{\lambda_{\min}^4 K}{128c_{\max}^2}\left(\frac{\delta^*}{2}-\beta\right)^2\right)$$

we have $\|\sum_{k=1}^{K}\frac{T^{(k)}}{T}W^{(k)}\|_\infty \leq \delta'$ and thus by Lemma 10, sign-consistency is guaranteed.

In conclusion, when $\tau \leq \delta^{\dagger}/2$, with probability at least

$$1 - 2N(N+1)\exp\left(-\frac{nK}{2}\min\left\{\frac{(\delta^{\dagger})^2}{256(1+4\sigma^2)^2\gamma^2}, 1\right\}\right) - 2N\exp\left(-\frac{\lambda_{\min}^4 K}{128c_{\max}^2}\left(\frac{\delta^*}{2}-\beta\right)^2\right)$$

the estimator $\hat{\Omega}$ is sign-consistent and thus $\mathrm{supp}(\hat{\Omega}) = \mathrm{supp}\left(\bar{\Omega}\right)$, which completes our proof of Theorem 2.

D. Proof of Theorem 3

For $\forall Q \in [-1/(2d), 1/(2d)]^{N\times N}$, let $\Omega(E) := I + Q \odot \mathrm{mat}(E)$ for $E \in \mathcal{E}$ where $\mathcal{E}$ is the set of all possible values of E generated according to Theorem 3 and $\mathrm{mat}(E) \in \{0,1\}^{N\times N}$ is defined as follows: $\mathrm{mat}(E)_{ij} = 1$ if $(i,j) \in E$ and $\mathrm{mat}(E)_{ij} = 0$ if $(i,j) \notin E$ for $\forall E \in \mathcal{E}$. Then we know $\Omega(E)$ is real and symmetric. Thus its eigenvalues are real. By Gershgorin circle theorem (Golub & Van Loan, 2012), for any eigenvalue λ of $\Omega(E)$, λ lies in one of the Gershgorin circles, i.e., $|\lambda - \Omega(E)_{jj}| \leq \sum_{l\neq j}|\Omega(E)_{jl}|$ holds for some j. Since $\mathrm{mat}(E)_{jj} = 0$ and $|Q_{jl}| \leq \frac{1}{2d}$ for $1 \leq l \leq N$, we have $\Omega(E)_{jj} = 1$ and $\sum_{l\neq j}|\Omega(E)_{jl}| \leq d\cdot\frac{1}{2d} = \frac{1}{2}$. Thus $\lambda \in \left[\frac{1}{2}, \frac{3}{2}\right]$ and $\Omega(E)$ is positive definite. Thus, we have constructed a multiple Gaussian graphical model. Now consider $\Omega(E)^{-1}$. Because any eigenvalue μ of $[\Omega(E)]^{-1}$ is the reciprocal of an eigenvalue of $\Omega(E)$, we have $|\mu| \in \left[\frac{2}{3}, 2\right]$.

Use $\lambda_1(A)$ to denote the largest eigenvalue of matrix A. for $E, E' \in \mathcal{E}$, according to Theorem H.1.d. in (Marshall et al., 2010), we have

$$\lambda_1(\Omega(E')\Omega(E)^{-1}) \leq \lambda_1(\Omega(E'))\lambda_1(\Omega(E)^{-1}) \leq \frac{3}{2}\cdot 2 = 3$$

which gives us

$$\mathrm{tr}\left(\Omega(E')\Omega(E)^{-1}\right) \leq N\lambda_1(\Omega(E')\Omega(E)^{-1}) \leq 3N \tag{56}$$

For $\mathbf{Q} = \{Q^{(k)}\}_{k=1}^{K}$, we know that there is a bijection between $\mathcal{E}$ and the set of all circular permutations of nodes $V = \{1, ..., N\}$. Thus $|\mathcal{E}|$, i.e., the size of $\mathcal{E}$, is the total number of circular permutations of N elements, which is $C_E := (N-1)!/2$. Since E is uniformly distributed on $\mathcal{E}$, the entropy of E given $\mathbf{Q}$ is $H(E|\mathbf{Q}) = \log C_E$.

Consider a family of N-dimensional random multivariate Gaussian distributions of size K with covariance matrices $\{\bar{\Sigma}^{(k)}\}_{k=1}^{K}$ generated according to Theorem 3. We use $\mathbf{X} := \{X_t^{(k)}\}_{1\leq t\leq n, 1\leq k\leq K}$ to denote the collection of n samples from each of the K distributions. Then for the mutual information $\mathbb{I}(\mathbf{X}; E|\mathbf{Q})$. We have the following bound:

$$\begin{aligned}
\mathbb{I}(\mathbf{X}; E|\mathbf{Q}) \leq &\frac{1}{C_E^2}\sum_{E}\sum_{E'}\mathbb{KL}(P_{\mathbf{X}|E,\mathbf{Q}}\|P_{\mathbf{X}|E',\mathbf{Q}}) \\
= &\frac{1}{C_E^2}\sum_{E}\sum_{E'}\sum_{k=1}^{K}\sum_{t=1}^{n}\mathbb{KL}(P_{X_t^{(k)}|E,Q^{(k)}}\|P_{X_t^{(k)}|E',Q^{(k)}}) \\
= &\frac{n}{C_E^2}\sum_{E}\sum_{E'}\sum_{k=1}^{K}\frac{1}{2}\Big[\mathrm{tr}\left((I+Q^{(k)}\odot\mathrm{mat}(E'))(I+Q^{(k)}\odot\mathrm{mat}(E))^{-1}\right) \\
&-N + \log\frac{\det(I+Q^{(k)}\odot\mathrm{mat}(E))}{\det(I+Q^{(k)}\odot\mathrm{mat}(E'))}\Big]
\end{aligned} \tag{57}$$

Since the summation is taken over all (E, E') pairs, the log term cancels with each other. For the trace term, by (56), we have

$$\mathrm{tr}\left((I+Q^{(k)}\odot\mathrm{mat}(E'))(I+Q^{(k)}\odot\mathrm{mat}(E))^{-1}\right) \leq 3N \tag{58}$$

for $1 \le k \le K$ and $E, E' \in \mathcal{E}$. Putting (58) back to (57) gives

$$\mathbb{I}(\mathbf{X}; E|\mathbf{Q}) \le \frac{n}{C_E^2} \sum_E \sum_{E'} \sum_{k=1}^K \frac{1}{2}(3N - N) = nNK \tag{59}$$

For any estimate $\hat{S}$ of S, define $\hat{E} = \{(i,j) : (i,j) \in \hat{S}, i \neq j\}$. Since $E \subseteq S$, we have $\mathbb{P}\{S \neq \hat{S}\} \ge \mathbb{P}\{E \neq \hat{E}\}$. Then by applying Theorem 1 in (Ghoshal & Honorio, 2017), we get

$$\begin{aligned}\mathbb{P}\{S \neq \hat{S}\} &\ge \mathbb{P}\{E \neq \hat{E}\} \\ &\ge 1 - \frac{\mathbb{I}(\mathbf{X}; S|\mathbf{Q}) + \log 2}{H(S|\mathbf{Q})} \\ &\ge 1 - \frac{nNK + \log 2}{\log[(N-1)!/2]}\end{aligned}$$

For $\log((N-1)!)$, we have:

$$\begin{aligned}\log((N-1)!) &= \sum_{i=1}^{N-1} \log i \\ &\ge \int_1^{N-1} \log x dx \\ &= (N-1)\log(N-1) - N + 2 \\ &= (N-1)\log N + (N-1)\log\frac{N-1}{N} + 2 - N\end{aligned}$$

Since

$$(N-1)\log\frac{N-1}{N} + 2 = 2 - (N-1)\log\left(1 + \frac{1}{N-1}\right) \ge 2 - 1 > 0$$

we have

$$\begin{aligned}\log((N-1)!) &\ge (N-1)\log N - N = N\log N - N - \log N \\ \log((N-1)!/2) &= \log((N-1)!) - \log 2 \ge N\log N - N - \log 2N\end{aligned}$$

For $N \ge 5$, $N\log N - N - \log 2N > 0$, thus we have

$$\mathbb{P}\{S \neq \hat{S}\} \ge 1 - \frac{nNK + \log 2}{\log[(N-1)!/2]} \ge 1 - \frac{nNK + \log 2}{N\log N - N - \log 2N}$$

which completes our proof of Theorem 3.

E. Proof of Theorem 4

By assumption, we have successfully recovered the true support union in the first step, i.e., supp$(\hat{\Omega}) = S$. Since there are constraints that supp$(\Omega) \subseteq$ supp$(\hat{\Omega}) = S$ and diag$(\Omega) =$ diag$(\hat{\Omega})$ in (6), we have

$$\begin{aligned}\ell^{(K+1)}(\Omega) &= \langle \hat{\Sigma}^{(K+1)}, \Omega \rangle - \log\det(\Omega) \\ &= \langle \hat{\Sigma}^{(K+1),S}, \Omega \rangle - \log\det(\Omega)\end{aligned} \tag{60}$$

where $\hat{\Sigma}^{(K+1),S} := \left(\hat{\Sigma}_S^{(K+1)}, 0\right)$. Then the Lagrangian of the problem (6) is

$$\ell^{(K+1)}(\Omega) + \lambda\|\Omega\|_1 + \langle \mu, \Omega \rangle + \langle \nu, \text{diag}(\Omega - \hat{\Omega}) \rangle \tag{61}$$

where $\mu \in \mathbb{R}^{N \times N}, \nu \in \mathbb{R}^N$ are the Lagrange multipliers satisfying $\mu_S = 0$. Here we set $\mu = \bar{\Sigma}^{(K+1),S^c} = (\bar{\Sigma}_{S^c}^{(K+1)}, 0)$ and $\nu = \text{diag}(\bar{\Sigma}^{(K+1)} - \hat{\Sigma}^{(K+1)})$ in (61). Define $W^{(K+1)} := \bar{\Sigma}^{(K+1),S_{\text{off}}} - \hat{\Sigma}^{(K+1),S_{\text{off}}}$. With the primal-dual witness approach, we can get the following lemma similar to Lemma 7.

Lemma 11. *Under Assumption 2, if* $\|W^{(K+1)}\|_\infty \leq \xi$ *with* $\xi \in (0, \delta^{(K+1),*}]$*, then choosing* $\lambda = \frac{8\xi}{\alpha^{(K+1)}}$*, we have supp*$(\hat{\Omega}^{(K+1)}) \subseteq$ *supp* $\left(\bar{\Omega}^{(K+1)}\right)$ *and*

$$\|\hat{\Omega}^{(K+1)} - \bar{\Omega}^{(K+1)}\|_\infty \leq 2\kappa_{\bar{\Gamma}^{(K+1)}} \left(\frac{8}{\alpha^{(K+1)}} + 1\right) \xi \tag{62}$$

The proof is in Section J.

By the definition of $W^{(K+1)}$, we know that $W^{(K+1)}_{S^c_{\text{off}}} = 0$ and $W^{(K+1)}_{S_{\text{off}}} = [\hat{\Sigma}^{(K+1)} - \bar{\Sigma}^{(K+1)}]_{S_{\text{off}}}$. Thus $\|W^{(K+1)}\|_\infty = \|[\hat{\Sigma}^{(K+1)} - \bar{\Sigma}^{(K+1)}]_{S_{\text{off}}}\|_\infty$. Since we have assumed $\|\bar{\Sigma}^{(K+1)}\|_\infty \leq \gamma^{(K+1)}$, according to Lemma 8 and the proof of (50), we have

$$\begin{aligned}\mathbb{P}\left[\|W^{(K+1)}\|_\infty \leq \delta^{(K+1),\dagger}\right] =& \mathbb{P}\left[\|\hat{\Sigma}^{(K+1)} - \bar{\Sigma}^{(K+1)}\|_\infty \leq \delta^{(K+1),\dagger}\right] \\ \leq& 1 - 2|S_{\text{off}}| \exp\left(-\frac{n^{(K+1)}}{2} \min\left\{\frac{(\delta^{(K+1),\dagger})^2}{64(1+4\sigma^2)^2(\gamma^{(K+1)})^2}, 1\right\}\right)\end{aligned} \tag{63}$$

because S_{off} is symmetric.

Similar to Lemma 10, we have the following lemma for the sign-consistency of $\hat{\Omega}^{(K+1)}$ in (6).

Lemma 12. *For* $\xi \in (0, \delta^{(K+1),*}]$*, if*

$$\|W^{(K+1)}\|_\infty \leq \xi \tag{64}$$

and

$$\frac{\omega^{(K+1)}_{\min}}{2} \geq 2\kappa_{\bar{\Gamma}^{(K+1)}} \left(\frac{8}{\alpha^{(K+1)}} + 1\right) \xi \tag{65}$$

where $\omega_{\min} := \min_{(i,j)\in S} |\bar{\Omega}_{ij}|$*, then the estimate* $\hat{\Omega}^{(K+1)}$ *in (6) is sign-consistent.*

The proof is in Section K. Similar to the proof of Theorem 2, we consider two cases of $\omega^{(K+1)}_{\min}$.

Case (i). If

$$\omega^{(K+1)}_{\min} \geq \frac{2\alpha^{(K+1)}}{8 + \alpha^{(K+1)}} \min\left\{\frac{1}{3\kappa_{\bar{\Sigma}^{(K+1)}} d^{(K+1)}}, \frac{1}{3\kappa^3_{\bar{\Sigma}^{(K+1)}} \kappa_{\bar{\Gamma}^{(K+1)}} d^{(K+1)}}\right\} \tag{66}$$

then

$$0 < \delta^{(K+1),\dagger} = \delta^{(K+1),*}$$

and

$$\frac{\omega^{(K+1)}_{\min}}{2} \geq 2\kappa_{\bar{\Gamma}^{(K+1)}} \left(\frac{8}{\alpha^{(K+1)}} + 1\right) \delta^{(K+1),*}$$

Thus for $\xi = \delta^{(K+1),*}$, (65) holds. Then according to (63), with probability at least

$$1 - 2|S_{\text{off}}| \exp\left(-\frac{n^{(K+1)}}{2} \min\left\{\frac{(\delta^{(K+1),\dagger})^2}{64(1+4\sigma^2)^2(\gamma^{(K+1)})^2}, 1\right\}\right)$$

we have $\|W^{(K+1)}\|_\infty \leq \delta = \delta^{(K+1),*}$ and thus by Lemma 12, we have that (6) is sign-consistent.

Case (ii). If

$$\omega^{(K+1)}_{\min} < \frac{2\alpha^{(K+1)}}{8 + \alpha^{(K+1)}} \min\left\{\frac{1}{3\kappa_{\bar{\Sigma}^{(K+1)}} d^{(K+1)}}, \frac{1}{3\kappa^3_{\bar{\Sigma}^{(K+1)}} \kappa_{\bar{\Gamma}^{(K+1)}} d^{(K+1)}}\right\}$$

then

$$\frac{\omega^{(K+1)}_{\min}}{2} < 2\kappa_{\bar{\Gamma}^{(K+1)}} \left(\frac{8}{\alpha^{(K+1)}} + 1\right) \delta^{(K+1),*}$$

and

$$0 < \delta^{(K+1),\dagger} = \delta^{(K+1),\prime} \leq \delta^{(K+1),*}$$

Then

$$\frac{\omega_{\min}^{(K+1)}}{2} \geq 2\kappa_{\bar{\Gamma}^{(K+1)}} \left(\frac{8}{\alpha^{(K+1)}} + 1\right) \delta^{(K+1),\prime} \tag{67}$$

For $\xi = \delta^{(K+1),\prime} = \delta^{(K+1),\dagger}$, (65) holds. Now according to (63), with probability at least

$$1 - 2|S_{\text{off}}| \exp\left(-\frac{n^{(K+1)}}{2} \min\left\{\frac{(\delta^{(K+1),\dagger})^2}{64(1+4\sigma^2)^2(\gamma^{(K+1)})^2}, 1\right\}\right)$$

we have $\|W^{(K+1)}\|_\infty \leq \delta^{(K+1),\prime} = \delta^{(K+1),\dagger}$ and thus by Lemma 12, sign-consistency is guaranteed.

In conclusion, with probability at least

$$1 - 2|S_{\text{off}}| \exp\left(-\frac{n^{(K+1)}}{2} \min\left\{\frac{(\delta^{(K+1),\dagger})^2}{64(1+4\sigma^2)^2(\gamma^{(K+1)})^2}, 1\right\}\right)$$

the estimator $\hat{\Omega}^{(K+1)}$ is sign-consistent and thus $\text{supp}(\hat{\Omega}^{(K+1)}) = \text{supp}\left(\bar{\Omega}^{(K+1)}\right)$, which completes our proof of Theorem 4.

F. Proof of Theorem 5

For $\forall Q \in [-1/(N\log s), 1/(N\log s)]^{N\times N},\ E^{(K+1)} \in \mathcal{E}$, we know $\Omega(E^{(K+1)}) = I + Q \odot \text{mat}(E^{(K+1)})$ is real and symmetric, where $\text{mat}(\cdot) \in \{0,1\}^{N\times N}$ is defined in the proof of Theorem 3. Thus its eigenvalues are real. By Gershgorin circle theorem (Golub & Van Loan, 2012), for any eigenvalue λ of $\Omega(E^{(K+1)})$, λ lies in one of the Gershgorin circles, i.e., $|\lambda - \Omega(E^{(K+1)})_{jj}| \leq \sum_{l\neq j} |\Omega(E^{(K+1)})_{jl}|$ holds for some j. Since $\text{mat}(E^{(K+1)})_{jj} = 0$ and $|Q_{jl}| \leq 1/(N\log s)$ for $1 \leq l \leq N$, we have $\Omega(E^{(K+1)})_{jj} = 1$. Meanwhile, there are at most $s/2$ non-zero elements in any row of $\text{mat}(E^{(K+1)})$ because $|E^{(K+1)}| \leq s$ and $\text{mat}(E^{(K+1)})$ is symmetric. Thus $\sum_{l\neq j} |\Omega(E)_{jl}| \leq \frac{s}{2N\log s}$. Then we have $\lambda \in \left[1 - \frac{s}{2N\log s}, 1 + \frac{s}{2N\log s}\right]$ and $\Omega(E^{(K+1)})$ is positive definite. Thus, we have constructed a Gaussian graphical model. Now consider $\Omega(E^{(K+1)})^{-1}$. Because any eigenvalue μ of $\Omega(E^{(K+1)})^{-1}$ is the reciprocal of an eigenvalue of $\Omega(E^{(K+1)})$, we have $|\mu| \leq 1/(1 - \frac{s}{2N\log s})$.

For any $E^{(K+1)}, \tilde{E}^{(K+1)} \in \mathcal{E}$, according to Theorem H.1.d. in (Marshall et al., 2010), we have

$$\lambda_1(\Omega(\tilde{E}^{(K+1)})\Omega(E^{(K+1)})^{-1}) \leq \lambda_1(\Omega(\tilde{E}^{(K+1)}))\lambda_1(\Omega(E^{(K+1)})^{-1}) \leq \frac{1 + \frac{s}{2N\log s}}{1 - \frac{s}{2N\log s}}$$

which gives us

$$\text{tr}\left(\Omega(\tilde{E}^{(K+1)})\Omega(E^{(K+1)})^{-1}\right) \leq N\lambda_1(\Omega(\tilde{E}^{(K+1)})\Omega(E^{(K+1)})^{-1}) \leq N\frac{1 + \frac{s}{2N\log s}}{1 - \frac{s}{2N\log s}} \tag{68}$$

According to the definition of $\mathcal{E}$, we know that $|\mathcal{E}| = 2^{s/2}$. Since $E^{(K+1)}$ is uniformly distributed on $\mathcal{E}$, the entropy of $E^{(K+1)}$ given Q is

$$H(E^{(K+1)}|Q) = \log|\mathcal{E}| \geq \frac{s}{2}\log 2 \tag{69}$$

Now let $\mathbf{X} := \{X_t\}_{1\leq t\leq n}$ be the samples from a N-dimensional multivariate Gaussian distribution with covariance $\bar{\Sigma}$

generated according to Theorem 5. For the mutual information $\mathbb{I}(\mathbf{X}; E^{(K+1)}|Q)$, we have the following bound:

$$\begin{aligned}\mathbb{I}(\mathbf{X}; E^{(K+1)}|Q) \leq & \frac{1}{|\mathcal{E}|^2} \sum_{E^{(K+1)}} \sum_{\tilde{E}^{(K+1)}} \mathbb{KL}(P_{\mathbf{X}|E^{(K+1)},Q} \| P_{\mathbf{X}|\tilde{E}^{(K+1)},Q}) \\ = & \frac{1}{|\mathcal{E}|^2} \sum_{E^{(K+1)}} \sum_{\tilde{E}^{(K+1)}} \sum_{t=1}^{n} \mathbb{KL}(P_{X_t|E^{(K+1)},Q} \| P_{X_t|\tilde{E}^{(K+1)},Q}) \\ = & \frac{n}{|\mathcal{E}|^2} \sum_{E^{(K+1)}} \sum_{\tilde{E}^{(K+1)}} \frac{1}{2} \Bigg[\mathrm{tr} \left((I + Q \odot \mathrm{mat}(\tilde{E}^{(K+1)}))(I + Q \odot \mathrm{mat}(E^{(K+1)}))^{-1} \right) \\ & - N + \log \frac{\det(I + Q \odot \mathrm{mat}(E^{(K+1)}))}{\det(I + Q \odot \mathrm{mat}(\tilde{E}^{(K+1)}))} \Bigg]\end{aligned} \tag{70}$$

Since the summation is taken over all $(E^{(K+1)}, \tilde{E}^{(K+1)})$ pairs, the $\log$ term cancels with each other. For the trace term, by (68), we have

$$\mathrm{tr} \left((I + Q \odot \mathrm{mat}(\tilde{E}^{(K+1)}))(I + Q \odot \mathrm{mat}(E^{(K+1)}))^{-1} \right) \leq N \frac{1 + \frac{s}{2N \log s}}{1 - \frac{s}{2N \log s}} \tag{71}$$

for $E^{(K+1)}, \tilde{E}^{(K+1)} \in \mathcal{E}$. Putting (71) back to (70) gives

$$\begin{aligned}\mathbb{I}(\mathbf{X}; E^{(K+1)}|Q) \leq & \frac{n}{|\mathcal{E}|^2} \sum_{E^{(K+1)}} \sum_{\tilde{E}^{(K+1)}} \frac{1}{2} \left(N \frac{1 + \frac{s}{2N \log s}}{1 - \frac{s}{2N \log s}} - N \right) \\ = & \frac{ns}{2 \log s} \frac{1}{1 - \frac{s}{2N \log s}} \\ \leq & \frac{2ns}{\log s}\end{aligned} \tag{72}$$

According to our assumption that $4 \leq s \leq N$.

Define $\hat{E}^{(K+1)} := \{(i,j) \in \hat{S}^{(K+1)} : i \neq j\}$. By applying Theorem 1 in (Ghoshal & Honorio, 2017), we get

$$\begin{aligned}\mathbb{P}\{S^{(K+1)} \neq \hat{S}^{(K+1)}\} \geq & \mathbb{P}\{E^{(K+1)} \neq \hat{E}^{(K+1)}\} \\ \geq & 1 - \frac{\mathbb{I}(\mathbf{X}; E^{(K+1)}|Q) + \log 2}{H(E^{(K+1)}|Q)} \\ \geq & 1 - \frac{\frac{2ns}{\log s} + \log 2}{\log |\mathcal{E}|} \\ = & 1 - \frac{\frac{2ns}{\log s} + \log 2}{\frac{s}{2} \log 2} \\ = & 1 - \frac{4n}{(\log 2)(\log s)} - \frac{2}{s}\end{aligned}$$

where the third inequality is by (72).

G. Proof of Lemma 8

We first prove the following lemma showing that (40) and (41) hold for deterministic covariance matrices $\{\Sigma^{(k)}\}_{k=1}^{K}$.

Lemma 13. *For K deterministic matrices $\{\bar{\Sigma}^{(k)}\}_{k=1}^{K}$ and $\gamma \geq \|\bar{\Sigma}^{(k)}\|_\infty$ for $1 \leq k \leq K$, consider the samples $\{X_t^{(k)}\}_{1 \leq t \leq n, 1 \leq k \leq K} \subseteq \mathbb{R}^N$ satisfying the following conditions:*

(i) $\mathbb{E}\left[X_t^{(k)}\right] = 0$, *Cov* $\left(X_t^{(k)}\right) = \bar{\Sigma}^{(k)}$ *for* $1 \leq t \leq n,\ 1 \leq k \leq K$*;*

(ii) $\left\{X_t^{(k)}\right\}_{1 \leq t \leq n,\ 1 \leq k \leq K}$ *are independent;*

(iii) $\frac{X_{t,i}^{(k)}}{\sqrt{\bar{\Sigma}_{ii}^{(k)}}}$ *is sub-Gaussian with parameter* σ *for* $1 \leq i \leq N,\ 1 \leq t \leq n,\ 1 \leq k \leq K$.

Then for the empirical sample covariance matrices $\{\hat{\Sigma}^{(k)}\}_{k=1}^{K}$, (40) *and* (41) *hold for* $1 \leq i, j \leq N$ *and* $0 \leq \nu \leq 8\left(1+4\sigma^2\right)\gamma$.

Proof. First consider the element-wise tail condition. For $1 \leq i, j \leq N$, we need to find an upper bound of the following probability:

$$\mathbb{P}\left[\Big|\frac{1}{nK}\sum_{k=1}^{K}\sum_{t=1}^{n}\left(X_{t,i}^{(k)}X_{t,j}^{(k)} - \bar{\Sigma}_{ij}^{(k)}\right)\Big| > \nu\right] \tag{73}$$

Let $s_i := \max_{1\leq k\leq K}\bar{\Sigma}_{ii}^{(k)}$, $s_j := \max_{1\leq k\leq K}\bar{\Sigma}_{jj}^{(k)}$, $\tilde{X}_{t,i}^{(k)} := \frac{X_{t,i}^{(k)}}{\sqrt{s_i}}$, $\tilde{X}_{t,j}^{(k)} := \frac{X_{t,j}^{(k)}}{\sqrt{s_j}}$, $\tilde{\rho}_{ij}^{(k)} := \frac{\bar{\Sigma}_{ij}^{(k)}}{\sqrt{s_i s_j}}$. We have

$$(73) = \mathbb{P}\left[4\Big|\sum_{k,t}\left(\tilde{X}_{t,i}^{(k)}\tilde{X}_{t,j}^{(k)} - \tilde{\rho}_{ij}^{(k)}\right)\Big| > \frac{4nK\nu}{\sqrt{s_i s_j}}\right]$$

Define $U_{t,ij}^{(k)} := \tilde{X}_{t,i}^{(k)} + \tilde{X}_{t,j}^{(k)}$, $V_{t,ij}^{(k)} := \tilde{X}_{t,i}^{(k)} - \tilde{X}_{t,j}^{(k)}$. Then for any $r \in \mathbb{R}$,

$$4\sum_{k,t}\left(\tilde{X}_{t,i}^{(k)}\tilde{X}_{t,j}^{(k)} - \tilde{\rho}_{ij}^{(k)}\right) = \sum_{k,t}\left\{\left(U_{t,ij}^{(k)}\right)^2 - 2\left(r+\tilde{\rho}_{ij}^{(k)}\right)\right\} - \sum_{k,t}\left\{\left(U_{t,ij}^{(k)}\right)^2 - 2\left(r-\tilde{\rho}_{ij}^{(k)}\right)\right\} \tag{74}$$

Thus

$$\begin{aligned}(73) \leq &\mathbb{P}\left[\Big|\sum_{k,t}\left\{\left(U_{t,ij}^{(k)}\right)^2 - 2\left(r+\tilde{\rho}_{ij}^{(k)}\right)\right\}\Big| > \frac{2nK\nu}{\sqrt{s_i s_j}}\right] \\ &+ \mathbb{P}\left[\Big|\sum_{k,t}\left\{\left(V_{t,ij}^{(k)}\right)^2 - 2\left(r-\tilde{\rho}_{ij}^{(k)}\right)\right\}\Big| > \frac{2nK\nu}{\sqrt{s_i s_j}}\right]\end{aligned} \tag{75}$$

Now define

$$Z_{t,ij}^{(k)} := \left(U_{t,ij}^{(k)}\right)^2 - 2\left(r+\tilde{\rho}_{ij}^{(k)}\right)$$

Applying the inequality $(a+b)^m \leq 2^m(a^m+b^m)$ on $Z_{t,ij}^{(k)}$, we have

$$\mathbb{E}\left[\left|Z_{t,ij}^{(k)}\right|^m\right] \leq 2^m\left\{\mathbb{E}\left[\left|U_{t,ij}^{(k)}\right|^{2m}\right] + \left[2\left(1+\tilde{\rho}_{ij}^{(k)}\right)\right]^m\right\} \tag{76}$$

Let $r_i^{(k)} := \sqrt{\frac{\bar{\Sigma}_{ii}^{(k)}}{s_i}}$, $r_i^{(k)} := \sqrt{\frac{\bar{\Sigma}_{ii}^{(k)}}{s_i}}$, then

$$\tilde{X}_{t,i}^{(k)} = \bar{X}_{t,i}^{(k)}r_i^{(k)},\ \ \tilde{X}_{t,j}^{(k)} = \bar{X}_{t,j}^{(k)}r_j^{(k)}$$

where $\bar{X}_{t,i}^{(k)} := \frac{X_{t,i}^{(k)}}{\sqrt{\bar{\Sigma}_{ii}^{(k)}}}$, $\bar{X}_{t,j}^{(k)} := \frac{X_{t,j}^{(k)}}{\sqrt{\bar{\Sigma}_{jj}^{(k)}}}$.

Assume that $\bar{X}_{t,i}^{(k)}$ is sub-Gaussian with parameter σ for $\leq i \leq N,\ 1 \leq t \leq n, 1 \leq k \leq K$, and then we have

$$\mathbb{E}\left[\exp\left(\lambda\tilde{X}_{t,i}^{(k)}\right)\right] = \mathbb{E}\left[\exp\left(\lambda\bar{X}_{t,i}^{(k)}r_i^{(k)}\right)\right] \leq \exp\left\{\frac{\lambda^2}{2}\sigma^2\left(r_i^{(k)}\right)^2\right\}$$

which shows that $\tilde{X}_{t,i}^{(k)}$ is sub-Gaussian with parameter $\sigma r_i^{(k)}$. Then

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda U_{t,ij}^{(k)}\right)\right] =&\mathbb{E}\left[\exp\left(\lambda\tilde{X}_{t,i}^{(k)}\right)\exp\left(\lambda\tilde{X}_{t,j}^{(k)}\right)\right] \\ \leq&\mathbb{E}\left[\exp\left(2\lambda\tilde{X}_{t,i}^{(k)}\right)\right]^{\frac{1}{2}}\mathbb{E}\left[\exp\left(2\lambda\tilde{X}_{t,j}^{(k)}\right)\right]^{\frac{1}{2}} \\ \leq&\exp\left\{\lambda^2\sigma^2\left[\left(r_i^{(k)}\right)^2 + \left(r_j^{(k)}\right)^2\right]\right\}\end{aligned}$$

Therefore $U_{t,ij}^{(k)}$ is sub-Gaussian with parameter $\sigma_{ij}^{(k)} := \sigma\sqrt{2\left[\left(r_i^{(k)}\right)^2 + \left(r_j^{(k)}\right)^2\right]}$. Similarly, we can prove that $V_{t,ij}^{(k)}$ is sub-Gaussian with parameter $\sigma_{ij}^{(k)}$ as well. Also note that $\sigma_{ij}^{(k)} \leq \sigma\sqrt{2(1+1)} = 2\sigma$.

As it is well-known (see e.g., Lemma 1.4 in (Buldygin & Kozachenko, 2000)), for a sub-Gaussian random variable X with parameter σ, i.e., X that satisfies $\mathbb{E}\left[e^{\lambda X}\right] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$, we have:

$$\mathbb{E}\left[|X|^s\right] \leq 2\left(\frac{s}{e}\right)^{s/2}\sigma^s \tag{77}$$

Apply this lemma on $U_{t,ij}^{(k)}$ with $s = 2m,\ m \geq 2$ and we get

$$\mathbb{E}\left[\left|U_{t,ij}^{(k)}\right|^{2m}\right] \leq 2\left(\frac{2m}{e}\right)^m\left(\sigma_{ij}^{(k)}\right)^{2m}$$

According to the inequality $m! \geq \left(\frac{m}{e}\right)^m$, we have

$$\mathbb{E}\left[\frac{\left|U_{t,ij}^{(k)}\right|^{2m}}{m!}\right] \leq 2^{m+1}\left(\sigma_{ij}^{(k)}\right)^2$$

Plug in (76) and we have

$$\begin{aligned}\left(\frac{\mathbb{E}\left[\left|Z_{t,ij}^{(k)}\right|^m\right]}{m!}\right)^{\frac{1}{m}} &\leq 2^{\frac{1}{m}}\left\{\left[2^{2m+1}\left(\sigma_{ij}^{(k)}\right)^{2m}\right]^{\frac{1}{m}} + \frac{4\left(r+\tilde{\rho}_{ij}^{(k)}\right)}{(m!)^{\frac{1}{m}}}\right\} \\ &\leq \underbrace{2^{\frac{1}{m}}\left\{4\cdot 2^{\frac{1}{m}}\left(\sigma_{ij}^{(k)}\right)^2 + \frac{4\left(r+\tilde{\rho}_{ij}^{(k)}\right)}{(m!)^{\frac{1}{m}}}\right\}}_{h(m)}\end{aligned} \tag{78}$$

Note that $h(m)$ defined above decreases with m and $\left|\tilde{\rho}_{ij}^{(k)}\right| \leq 1$.

Since (74) holds for $\forall r \in \mathbb{R}$, we can choose $r = \frac{\left(r_i^{(k)}\right)^2 + \left(r_j^{(k)}\right)^2}{2}$. Then we have $r < 1$ and

$$Z_{t,ij}^{(k)} := \left(U_{t,ij}^{(k)}\right)^2 - \left(\left(r_i^{(k)}\right)^2 + \left(r_j^{(k)}\right)^2 + 2\tilde{\rho}_{ij}^{(k)}\right)$$

Thus

$$\mathbb{E}\left[Z_{t,ij}^{(k)}\right] = 0$$

and furthermore,

$$\begin{aligned}\sup_{m\geq 2}\left(\frac{\mathbb{E}\left[\left|Z_{t,ij}^{(k)}\right|^m\right]}{m!}\right)^{\frac{1}{m}} &\leq h(2) \\ &= 8\left(\sigma_{ij}^{(k)}\right)^2 + 4\left(r + \left|\tilde{\rho}_{ij}^{(k)}\right|\right) \\ &\leq 8\left(1 + \left(\sigma_{ij}^{(k)}\right)^2\right) \\ &\leq 8\left(1 + 4\sigma^2\right)\end{aligned}$$

Define $B := 8\left(1+4\sigma^2\right)$. If X is a random variable such that $\mathbb{E}[X] = 0$, $\left(\frac{\mathbb{E}[|X|^m]}{m!}\right)^{\frac{1}{m}} \leq B$ for $m \geq 2$, then

$$\mathbb{E}\left[e^{\lambda X}\right] = \mathbb{E}\left[\sum_{k=0}^{\infty}\frac{X^k}{k!}\lambda^k\right] = 1 + \sum_{k=1}^{\infty}\lambda^k\frac{\mathbb{E}\left[X^k\right]}{k!} \leq 1 + \sum_{k=1}^{\infty}(\lambda B)^k \leq 1 + \frac{(\lambda B)^2}{1-|\lambda|B}$$

when $|\lambda| < \frac{1}{B}$. Meanwhile,

$$1 + \frac{(\lambda B)^2}{1-|\lambda|B} \le \exp\left\{\frac{\lambda^2 B^2}{1-|\lambda|B}\right\} \le \exp\left(2\lambda^2 B^2\right)$$

when $|\lambda| \le \frac{1}{2B}$. Therefore for $|\lambda| \le \frac{1}{2B}$,

$$\mathbb{E}\left[e^{\lambda X}\right] \le \exp\left(2\lambda^2 B^2\right) = \exp\left(\frac{\lambda^2 (2B)^2}{2}\right) \tag{79}$$

Then for $X_i,\ 1 \le i \le n$ independent and satisfying $\mathbb{E}[X_i] = 0$, $\left(\frac{\mathbb{E}[|X_i|^m]}{m!}\right)^{\frac{1}{m}} \le B$ when $m \ge 2$, we can claim that for $0 \le \epsilon \le 2B$,

$$\mathbb{P}\left[\Big|\sum_{i=1}^{n} X_i\Big| > n\epsilon\right] \le 2\exp\left(-\frac{n\epsilon^2}{8B^2}\right) \tag{80}$$

In fact, for $0 \le t \le \frac{1}{2B}$,

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^{n} X_i > n\epsilon\right] &\le \mathbb{P}\left[e^{t\sum_{i=1}^{n} X_i} \ge e^{tn\epsilon}\right] \\ &\le e^{-tn\epsilon}\mathbb{E}\left[e^{t\sum_{i=1}^{n} X_i}\right] \\ &= \left(\prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right]\right) e^{-tn\epsilon} \\ &\le \exp\left(2nt^2B^2 - tn\epsilon\right)\end{aligned} \tag{81}$$

Thus choosing $t = \frac{\epsilon}{4B^2} \le \frac{1}{2B}$, we can get

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i > n\epsilon\right] \le \exp\left(-\frac{n\epsilon^2}{8B^2}\right)$$

Similarly, we can also prove that

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i < -n\epsilon\right] \le \exp\left(-\frac{n\epsilon^2}{8B^2}\right)$$

Thus

$$\mathbb{P}\left[\Big|\sum_{i=1}^{n} X_i\Big| > n\epsilon\right] = \mathbb{P}\left[\sum_{i=1}^{n} X_i > n\epsilon\right] + \mathbb{P}\left[\sum_{i=1}^{n} X_i < -n\epsilon\right] \le 2\exp\left(-\frac{n\epsilon^2}{8B^2}\right)$$

Now consider $Z_{t,ij}^{(k)},\ 1 \le t \le n,\ 1 \le k \le K$. These random variables are independent by our assumption and satisfy $\mathbb{E}\left[Z_{t,ij}^{(k)}\right] = 0$, $\sup_{m\ge 2}\left(\frac{\mathbb{E}\left[\left|Z_{t,ij}^{(k)}\right|^m\right]}{m!}\right)^{\frac{1}{m}} \le 8\left(1+4\sigma^2\right) = B$ by our proof. Then according to (80), for $0 \le \frac{2\nu}{\gamma} \le 2B$, i.e., $0 \le \nu \le 8\left(1+4\sigma^2\right)\gamma$, we have:

$$\begin{aligned}\mathbb{P}\left[\Big|\sum_{k,t} Z_{t,ij}^{(k)}\Big| > \frac{2nK\nu}{\gamma}\right] &\le 2\exp\left\{-\frac{4nK\nu^2}{8B^2\gamma^2}\right\} \\ &= 2\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\}\end{aligned} \tag{82}$$

Since $\gamma \ge \max_{1\le k\le K} \|\bar{\Sigma}^{(k)}\|_\infty = \max_{1\le l\le N} s_l \ge \sqrt{s_i s_j}$ for $1 \le i, j \le N$, we have:

$$\mathbb{P}\left[\Big|\sum_{k,t} Z_{t,ij}^{(k)}\Big| > \frac{2nK\nu}{\sqrt{s_i s_j}}\right] \le \mathbb{P}\left[\Big|\sum_{k,t} Z_{t,ij}^{(k)}\Big| > \frac{2nK\nu}{\gamma}\right] \le 2\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\} \tag{83}$$

Plug in the definition of $Z_{t,ij}^{(k)}$, we have

$$\mathbb{P}\left[\Big|\sum_{k,t}\left\{\left(U_{t,ij}^{(k)}\right)^2-2\left(r+\tilde{\rho}_{ij}^{(k)}\right)\right\}\Big|>\frac{2nK}{\sqrt{s_is_j}}\nu\right]\leq 2\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\} \tag{84}$$

Similarly, we can also prove that for $0\leq\nu\leq 8\left(1+4\sigma^2\right)\gamma$,

$$\mathbb{P}\left[\Big|\sum_{k,t}\left\{\left(V_{t,ij}^{(k)}\right)^2-2\left(r-\tilde{\rho}_{ij}^{(k)}\right)\right\}\Big|>\frac{2nK}{\sqrt{s_is_j}}\nu\right]\leq 2\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\} \tag{85}$$

Thus according to (75), we have

$$\begin{aligned}(73)\leq&\mathbb{P}\left[\Big|\sum_{k,t}\left\{\left(U_{t,ij}^{(k)}\right)^2-2\left(r+\tilde{\rho}_{ij}^{(k)}\right)\right\}\Big|>\frac{2nK\nu}{\sqrt{s_is_j}}\right]\\&+\mathbb{P}\left[\Big|\sum_{k,t}\left\{\left(V_{t,ij}^{(k)}\right)^2-2\left(r-\tilde{\rho}_{ij}^{(k)}\right)\right\}\Big|>\frac{2nK\nu}{\sqrt{s_is_j}}\right]\\\leq&4\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\}\end{aligned} \tag{86}$$

i.e.,

$$\begin{aligned}\mathbb{P}\left[\Big|\sum_{k=1}^K\frac{1}{K}\left(\hat{\Sigma}_{ij}^{(k)}-\bar{\Sigma}_{ij}^{(k)}\right)\Big|>\nu\right]&=\mathbb{P}\left[\Big|\frac{1}{nK}\sum_{k=1}^K\sum_{t=1}^n\left(X_{t,i}^{(k)}X_{t,j}^{(k)}-\bar{\Sigma}_{ij}^{(k)}\right)\Big|>\nu\right]\\&\leq 4\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\}\end{aligned} \tag{87}$$

for $0\leq\nu\leq 8\left(1+4\sigma^2\right)\gamma$. Then consider the ℓ_∞-norm of $\hat{\Sigma}^{(k)}-\bar{\Sigma}^{(k)}$. Since $\hat{\Sigma}^{(k)},\ \bar{\Sigma}^{(k)}$ are all symmetric and $N\times N$, we have the following bound:

$$\begin{aligned}\mathbb{P}\left[\|\sum_{k=1}^K\frac{1}{K}\left(\hat{\Sigma}^{(k)}-\bar{\Sigma}^{(k)}\right)\|_\infty>\nu\right]&\leq\frac{N(N+1)}{2}\mathbb{P}\left[\Big|\sum_{k=1}^K\frac{1}{K}\left(\hat{\Sigma}_{ij}^{(k)}-\bar{\Sigma}_{ij}^{(k)}\right)\Big|>\nu\right]\\&\leq 2N(N+1)\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\}\end{aligned} \tag{88}$$

for $0\leq\nu\leq 8\left(1+4\sigma^2\right)\gamma$, which completes our proof of Lemma 13. □

Now consider the setting when $\{\bar{\Sigma}^{(k)}\}_{k=1}^K$ are randomly generated based on Definition 3. According to Lemma 13, we have

$$\mathbb{P}\left[\left|\sum_{k=1}^K\frac{1}{K}\left(\hat{\Sigma}_{ij}^{(k)}-\bar{\Sigma}_{ij}^{(k)}\right)\right|>\nu\Big|\{\bar{\Sigma}^{(k)}\}_{k=1}^K\right]\leq\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\} \tag{89}$$

$$\mathbb{P}\left[\|\sum_{k=1}^K\frac{1}{K}\left(\hat{\Sigma}^{(k)}-\bar{\Sigma}^{(k)}\right)\|_\infty>\nu\Big|\{\bar{\Sigma}^{(k)}\}_{k=1}^K\right]\leq 2N(N+1)\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\} \tag{90}$$

for $\hat{\Sigma}^{(k)}=\frac{1}{n}\sum_{t=1}^n X_t^{(k)}(X_t^{(k)})^{\mathrm{T}}$, $1\leq i,j\leq N$, and $0\leq\nu\leq 8\left(1+4\sigma^2\right)\gamma$ with γ specified in (2) of the corrected condition (ii) in Definition 3.

Then by the law of total expectation (see e.g., (Weiss et al., 2005)), we have

$$\begin{aligned}\mathbb{P}\left[\left|\sum_{k=1}^{K}\frac{1}{K}\left(\hat{\Sigma}_{ij}^{(k)}-\bar{\Sigma}_{ij}^{(k)}\right)\right|>\nu\right]=&\mathbb{E}_{\{\bar{\Sigma}^{(k)}\}_{k=1}^{K}}\left[\mathbb{P}\left[\left|\sum_{k=1}^{K}\frac{1}{K}\left(\hat{\Sigma}_{ij}^{(k)}-\bar{\Sigma}_{ij}^{(k)}\right)\right|>\nu\middle|\{\bar{\Sigma}^{(k)}\}_{k=1}^{K}\right]\right]\\ \leq&\mathbb{E}_{\{\bar{\Sigma}^{(k)}\}_{k=1}^{K}}\left[\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\}\right]\\ =&\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\}\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}\left[\|\sum_{k=1}^{K}\frac{1}{K}\left(\hat{\Sigma}^{(k)}-\bar{\Sigma}^{(k)}\right)\|_\infty>\nu\right]=&\mathbb{E}_{\{\bar{\Sigma}^{(k)}\}_{k=1}^{K}}\left[\mathbb{P}\left[\|\sum_{k=1}^{K}\frac{1}{K}\left(\hat{\Sigma}^{(k)}-\bar{\Sigma}^{(k)}\right)\|_\infty>\nu\middle|\{\bar{\Sigma}^{(k)}\}_{k=1}^{K}\right]\right]\\ \leq&\mathbb{E}_{\{\bar{\Sigma}^{(k)}\}_{k=1}^{K}}\left[2N(N+1)\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\}\right]\\ =&2N(N+1)\exp\left\{-\frac{nK\nu^2}{128\left(1+4\sigma^2\right)^2\gamma^2}\right\}\end{aligned}$$

which completes the proof of Lemma 8. Also notice that the proof above does not rely on any assumption on the distribution of $\{\bar{\Sigma}^{(k)}\}_{k=1}^{K}$. Thus, (40) and (41) hold as long as condition (iii), (iv) and (v) in Definition 3 are satisfied.

H. Proof of Lemma 9

By definition, H is a function that maps K matrices to a symmetric matrix of dimension N, since $\bar{\Omega}^{(k)}=\bar{\Omega}+\Delta^{(k)}\succ 0$ with probability 1 according to condition (ii) in Definition 3. For $\forall k\in\{1,\ldots,K\}$, let $\{\Delta^{(1)},\ldots,\Delta^{(k)},\ldots,\Delta^{(K)},\Delta'^{(k)}\}$ be an i.i.d. family of random matrices following distribution P in Definition 3. Consider $H_1^{(k)}=H(\Delta^{(1)},\ldots,\Delta^{(k)},\ldots,\Delta^{(K)})$ and $H_2^{(k)}=H(\Delta^{(1)},\ldots,\Delta'^{(k)},\ldots,\Delta^{(K)})$. We have

$$\begin{aligned}\left|\!\left|\!\left|H_1^{(k)}-H_2^{(k)}\right|\!\right|\!\right|_2=&\left|\!\left|\!\left|\frac{1}{K}(\bar{\Omega}+\Delta'^{(k)})^{-1}-(\bar{\Omega}+\Delta^{(k)})^{-1}\right|\!\right|\!\right|_2\\ =&\frac{1}{K}\left|\!\left|\!\left|(\bar{\Omega}+\Delta'^{(k)})^{-1}-\bar{\Omega}^{-1}+\bar{\Omega}^{-1}-(\bar{\Omega}+\Delta^{(k)})^{-1}\right|\!\right|\!\right|_2\\ \leq&\frac{1}{K}\left|\!\left|\!\left|(\bar{\Omega}+\Delta'^{(k)})^{-1}-\bar{\Omega}^{-1}\right|\!\right|\!\right|_2+\frac{1}{K}\left|\!\left|\!\left|(\bar{\Omega}+\Delta^{(k)})^{-1}-\bar{\Omega}^{-1}\right|\!\right|\!\right|_2\end{aligned}\tag{91}$$

Since $\mathbb{P}_{\Delta\sim P}[|\!|\!|\Delta|\!|\!|_2\leq c_{\max}\leq\frac{\lambda_{\min}}{2}]=1$ with $\lambda_{\min}=\lambda_{\min}(\bar{\Omega})$ by (2) and since $\bar{\Omega}\succ 0$, we have

$$\left|\!\left|\!\left|(\bar{\Omega}+\Delta^{(k)})^{-1}-\bar{\Omega}^{-1}\right|\!\right|\!\right|_2\leq\frac{c_{\max}}{\lambda_{\min}(\lambda_{\min}-c_{\max})}\leq\frac{2c_{\max}}{\lambda_{\min}^2}$$

and

$$\left|\!\left|\!\left|(\bar{\Omega}+\Delta'^{(k)})^{-1}-\bar{\Omega}^{-1}\right|\!\right|\!\right|_2\leq\frac{c_{\max}}{\lambda_{\min}(\lambda_{\min}-c_{\max})}\leq\frac{2c_{\max}}{\lambda_{\min}^2}$$

according to Equation (7.2) in (El Ghaoui, 2002). Plug the above inequalities in (91) and we can get

$$\left|\!\left|\!\left|H_1^{(k)}-H_2^{(k)}\right|\!\right|\!\right|_2\leq\frac{1}{K}\left|\!\left|\!\left|(\bar{\Omega}+\Delta'^{(k)})^{-1}-\bar{\Omega}^{-1}\right|\!\right|\!\right|_2+\frac{1}{K}\left|\!\left|\!\left|(\bar{\Omega}+\Delta^{(k)})^{-1}-\bar{\Omega}^{-1}\right|\!\right|\!\right|_2\leq\frac{4c_{\max}}{K\lambda_{\min}^2}\tag{92}$$

For $k=1,\ldots,K$, define $A_k=\frac{4c_{\max}}{K\lambda_{\min}^2}I_N$ with $I_N\in\mathbb{R}^{N\times N}$ being an identity matrix. Then by (92), we have

$$(H_1^{(k)}-H_2^{(k)})^2\preceq A_k^2$$

where $X \preceq Y \iff Y - X \succeq 0$.

Define $\sigma_\Delta^2 := \left\|\left\|\sum_{k=1}^K A_k^2\right\|\right\|_2 = \sum_{k=1}^K \left(\frac{4c_{\max}}{K\lambda_{\min}^2}\right)^2 = \frac{16c_{\max}^2}{K\lambda_{\min}^4}$. Then according to Corollary 7.5 in (Tropp, 2011), we have

$$\mathbb{P}\left[\lambda_{\max}(H - \mathbb{E}[H]) > t\right] \le N \exp\left\{-\frac{t^2}{8\sigma_\Delta^2}\right\} \le N \exp\left\{-\frac{\lambda_{\min}^4 K t^2}{128 c_{\max}^2}\right\} \tag{93}$$

Consider $-H(\Delta^{(1)}, \dots, \Delta^{(K)})$. We have

$$((-H_1^{(k)}) - (-H_2^{(k)}))^2 \preceq A_k^2$$

The conditions of Corollary 7.5 in (Tropp, 2011) are also satisfied. Thus, we have

$$\mathbb{P}\left[-\lambda_{\min}(H - \mathbb{E}[H]) > t\right] = \mathbb{P}\left[\lambda_{\max}((-H) - (-\mathbb{E}[H])) > t\right] \le N \exp\left\{-\frac{t^2}{8\sigma_\Delta^2}\right\} \le N \exp\left\{-\frac{\lambda_{\min}^4 K t^2}{128 c_{\max}^2}\right\} \tag{94}$$

By (93) and (94), we have

$$\begin{aligned}\mathbb{P}\left[\|\|H - \mathbb{E}[H]\|\|_2 > t\right] =& \mathbb{P}\left[\lambda_{\max}(H - \mathbb{E}[H]) > t, -\lambda_{\min}(H - \mathbb{E}[H]) > t\right] \\ \le& \mathbb{P}\left[\lambda_{\max}(H - \mathbb{E}[H]) > t\right] + \mathbb{P}\left[-\lambda_{\min}(H - \mathbb{E}[H]) > t\right] \\ \le& 2N \exp\left\{-\frac{\lambda_{\min}^4 K t^2}{128 c_{\max}^2}\right\}\end{aligned} \tag{95}$$

which gives us (43).

I. Proof of Lemma 10

For $\xi \in (0, \delta^*]$, we have proved that if $\|\sum_{k=1}^K \frac{T^{(k)}}{T} W^{(k)}\|_\infty \le \xi$ then $\|\Delta\|_\infty \le 2\kappa_{\bar{\Gamma}}\left(\frac{8}{\alpha} + 1\right)\xi$, $\tilde{\Omega} = \hat{\Omega}$ and $\text{supp}(\hat{\Omega}) \subseteq \text{supp}(\bar{\Omega})$.

Therefore if we further assume that

$$\frac{\omega_{\min}}{2} \ge 2\kappa_{\bar{\Gamma}}\left(\frac{8}{\alpha} + 1\right)\xi$$

we will have

$$\frac{\omega_{\min}}{2} \ge \|\Delta\|_\infty = \|\hat{\Omega} - \bar{\Omega}\|_\infty$$

Then for any $(i,j) \in S^c = \left[\text{supp}(\bar{\Omega})\right]^c$, $\bar{\Omega}_{ij} = 0$, we have $\left[\text{supp}(\bar{\Omega})\right]^c \subseteq \left[\text{supp}(\hat{\Omega})\right]^c$ and thus $(i,j) \in \left[\text{supp}(\hat{\Omega})\right]^c$, $\hat{\Omega}_{ij} = 0 = \bar{\Omega}_{ij}$

For any $(i,j) \in S = \text{supp}(\bar{\Omega})$, we have

$$|\hat{\Omega}_{ij} - \bar{\Omega}_{ij}| \le \|\hat{\Omega} - \bar{\Omega}\|_\infty \le \frac{\omega_{\min}}{2} = \frac{1}{2}\min_{1 \le k,l \le N} \bar{\Omega}_{kl} \le \frac{1}{2}|\bar{\Omega}_{ij}|$$

$$\Rightarrow -\frac{1}{2}|\bar{\Omega}_{ij}| \le \hat{\Omega}_{ij} - \bar{\Omega}_{ij} \le \frac{1}{2}|\bar{\Omega}_{ij}|$$

If $\bar{\Omega}_{ij} > 0$, then

$$-\frac{1}{2}\bar{\Omega}_{ij} \le \hat{\Omega}_{ij} - \bar{\Omega}_{ij}$$

$$\hat{\Omega}_{ij} \ge \frac{1}{2}\bar{\Omega}_{ij} > 0$$

If $\bar{\Omega}_{ij} < 0$, then

$$\hat{\Omega}_{ij} - \bar{\Omega}_{ij} \le -\frac{1}{2}\bar{\Omega}_{ij}$$

$$\hat{\Omega}_{ij} \le \frac{1}{2}\bar{\Omega}_{ij} < 0$$

In conclusion, $\text{sign}(\hat{\Omega}_{ij}) = \text{sign}(\bar{\Omega}_{ij})$ for $\forall\, i,j \in \{1, 2, ..., N\}$. The estimate $\hat{\Omega}$ in (5) is sign-consistent.

J. Proof of Lemma 11

Plug $\mu = \bar{\Sigma}^{(K+1),S^c} = (\bar{\Sigma}^{(K+1)}_{S^c}, 0)$ and $\nu = \text{diag}(\bar{\Sigma}^{(K+1)} - \hat{\Sigma}^{(K+1)})$ in (61). We have the following optimization problem

$$\hat{\Omega}^{(K+1)} = \arg\min_{\Omega \in \mathcal{S}^N_{++}} \ell^{(K+1)}(\Omega) + \lambda\|\Omega\|_1 + \langle \bar{\Sigma}^{(K+1),S^c}, \Omega\rangle + \langle \text{diag}(\bar{\Sigma}^{K+1} - \hat{\Sigma}^{K+1}), \text{diag}(\Omega - \hat{\Omega})\rangle$$

Now we can prove with the five steps in the primal-dual witness approach.

J.1. Step 1

For $(\Omega_{S^{(K+1)}}, 0) \in \mathcal{S}^N_{++}$, we need to verify $[\nabla^2 \ell^{(K+1)}(\Omega)]_{S^{(K+1)}S^{(K+1)}} \succ 0$. In fact,

$$\nabla \ell^{(K+1)}(\Omega) = \hat{\Sigma}^{(K+1),S} - \Omega^{-1} \tag{96}$$

$$\nabla^2 \ell^{(K+1)}(\Omega) = \Gamma(\Omega) = \Omega^{-1} \otimes \Omega^{-1} \tag{97}$$

For $(\Omega_{S^{(K+1)}}, 0) \in \mathcal{S}^N_{++}$, we have $\Gamma((\Omega_{S^{(K+1)}}, 0)) \succ 0$, $\nabla^2 \ell^{(K+1)}(\Omega) \succ 0$. Thus following the same steps in section B.2 , we can prove $[\nabla^2 \ell^{(K+1)}(\Omega)]_{S^{(K+1)}S^{(K+1)}} \succ 0$.

J.2. Step 2

Construct the primal variable $\tilde{\Omega}$ by making $\tilde{\Omega}_{[S^{(K+1)}]^c} = 0$ and solving the restricted problem:

$$\begin{aligned}\tilde{\Omega}_{S^{(K+1)}} = \arg & \min_{(\Omega_{S^{(K+1)}}, 0) \in \mathcal{S}^N_{++}} \ell^{(K+1)}\left((\Omega_{S^{(K+1)}}, 0)\right) + \lambda\|\Omega_{S^{(K+1)}}\|_1 \\ & + \langle \bar{\Sigma}^{(K+1),S^c}, (\Omega_{S^{(K+1)}}, 0)\rangle + \langle \text{diag}(\bar{\Sigma}^{K+1} - \hat{\Sigma}^{K+1}), \text{diag}((\Omega_{S^{(K+1)}}, 0) - \hat{\Omega})\rangle\end{aligned} \tag{98}$$

J.3. Step 3

Choose the dual variable $\tilde{Z}$ in order to fulfill the complementary slackness condition of (61):

$$\begin{cases} \tilde{Z}_{ij} = 1, & \text{if } \tilde{\Omega}_{ij} > 0 \\ \tilde{Z}_{ij} = -1, & \text{if } \tilde{\Omega}_{ij} < 0 \\ \tilde{Z}_{ij} \in [-1, 1], & \text{if } \tilde{\Omega}_{ij} = 0 \end{cases} \tag{99}$$

Therefore we have

$$\|\tilde{Z}\|_\infty \le 1 \tag{100}$$

J.4. Step 4

$\tilde{Z}$ is the subgradient of $\|\tilde{\Omega}\|_1$. Solve for the dual variable $\tilde{Z}_{[S^{(K+1)}]^c}$ in order that $(\tilde{\Omega}, \tilde{Z})$ fulfills the stationarity condition of (61):

$$\left[\nabla \ell^{(K+1)}\left(\left(\tilde{\Omega}_{S^{(K+1)}}, 0\right)\right)\right]_{S^{(K+1)}} + \lambda \tilde{Z}_{S^{(K+1)}} + I_N \text{diag}(\bar{\Sigma}^{(K+1)} - \hat{\Sigma}^{(K+1)}) = 0 \tag{101}$$

$$\left[\nabla \ell^{(K+1)}\left(\left(\tilde{\Omega}_{S^{(K+1)}}, 0\right)\right)\right]_{[S^{(K+1)}]^c} + \lambda \tilde{Z}_{[S^{(K+1)}]^c} + \bar{\Sigma}^{(K+1),S^c}_{[S^{(K+1)}]^c} = 0 \tag{102}$$

where $I_N \in \mathbb{R}^{N \times N}$ is an identity matrix.

J.5. Step 5

Now we need to verify that the dual variable solved by Step 4 satisfied the strict dual feasibility condition:

$$\|\tilde{Z}_{[S^{(K+1)}]^c}\|_\infty < 1 \tag{103}$$

If we can show the strict dual feasibility condition holds, we can claim that the solution in (98) is equal to the solution in (6), i.e., $\tilde{\Omega} = \hat{\Omega}^{(K+1)}$. Thus we will have

$$\text{supp}\left(\hat{\Omega}^{(K+1)}\right) = \text{supp}\left(\tilde{\Omega}\right) \subseteq S^{(K+1)} = \text{supp}\left(\bar{\Omega}^{(K+1)}\right)$$

J.6. Proof of the Strict Dual Feasibility Condition

Plug (96) in the stationarity condition of (6), we have

$$\hat{\Sigma}^{(K+1),S} - \tilde{\Omega}^{-1} + \lambda\tilde{Z} + \bar{\Sigma}^{(K+1),S^c} + I_N \text{diag}(\bar{\Sigma}^{K+1} - \hat{\Sigma}^{K+1}) = 0 \tag{104}$$

Define $\Psi := \tilde{\Omega} - \bar{\Omega}^{(K+1)}$, $R(\Psi) := \tilde{\Omega}^{-1} - \bar{\Sigma}^{(K+1)} + \bar{\Sigma}^{(K+1)}\Psi\bar{\Sigma}^{(K+1)}$. Notice that $W^{(K+1)} = \bar{\Sigma}^{(K+1),S_{\text{off}}} - \hat{\Sigma}^{(K+1),S_{\text{off}}}$. Then we can rewrite (104) as

$$\begin{aligned}
0 =&\hat{\Sigma}^{(K+1),S} - \tilde{\Omega}^{-1} + \lambda\tilde{Z} + \bar{\Sigma}^{(K+1),S^c} + I_N \text{diag}(\bar{\Sigma}^{K+1} - \hat{\Sigma}^{K+1}) \\
=&\hat{\Sigma}^{(K+1),S} - (\tilde{\Omega} - \bar{\Sigma}^{(K+1)} + \bar{\Sigma}^{(K+1)}\Psi\bar{\Sigma}^{(K+1)}) - \bar{\Sigma}^{(K+1)} + \bar{\Sigma}^{(K+1)}\Psi\bar{\Sigma}^{(K+1)} + \bar{\Sigma}^{(K+1),S^c} \\
&+ I_N \text{diag}(\bar{\Sigma}^{K+1} - \hat{\Sigma}^{K+1}) + \lambda\tilde{Z} \\
=&\hat{\Sigma}^{(K+1),S_{\text{off}}} + I_N \text{diag}(\hat{\Sigma}^{(K+1)}) - R(\Psi) - \bar{\Sigma}^{(K+1),S} + I_N \text{diag}(\bar{\Sigma}^{K+1} - \hat{\Sigma}^{K+1}) + \lambda\tilde{Z} \\
=&\hat{\Sigma}^{(K+1),S_{\text{off}}} - \bar{\Sigma}^{(K+1),S_{\text{off}}} + \bar{\Sigma}^{(K+1)}\Psi\bar{\Sigma}^{(K+1)} - R(\Psi) + \lambda\tilde{Z} \\
=&W^{(K+1)} + \bar{\Sigma}^{(K+1)}\Psi\bar{\Sigma}^{(K+1)} - R(\Psi) + \lambda\tilde{Z}
\end{aligned} \tag{105}$$

Now apply Lemma 7 with $K = 1$ and we can get Lemma 11.

K. Proof of Lemma 12

For $\xi \in (0, \delta^{(K+1),*}]$, in Lemma 11, we have proved that if $\|W^{(K+1)}\|_\infty \leq \xi$ then $\|\hat{\Omega}^{(K+1)} - \bar{\Omega}^{(K+1)}\|_\infty \leq 2\kappa_{\bar{\Gamma}^{(K+1)}}\left(\frac{8}{\alpha^{(K+1)}} + 1\right)\xi$ and $\text{supp}(\hat{\Omega}^{(K+1)}) \subseteq \text{supp}(\bar{\Omega}^{(K+1)})$.

Therefore if we further assume that

$$\frac{\omega_{\min}^{(K+1)}}{2} \geq 2\kappa_{\bar{\Gamma}^{(K+1)}}\left(\frac{8}{\alpha^{(K+1)}} + 1\right)\xi$$

we will have

$$\frac{\omega_{\min}^{(K+1)}}{2} \geq \|\hat{\Omega}^{(K+1)} - \bar{\Omega}^{(K+1)}\|_\infty$$

Then for any $(i,j) \in [S^{(K+1)}]^c = \left[\text{supp}(\bar{\Omega}^{(K+1)})\right]^c$, $\bar{\Omega}_{ij}^{(K+1)} = 0$, we have $\left[\text{supp}(\bar{\Omega}^{(K+1)})\right]^c \subseteq \left[\text{supp}(\hat{\Omega}^{(K+1)})\right]^c$ and thus $(i,j) \in \left[\text{supp}(\hat{\Omega}^{(K+1)})\right]^c$, $\hat{\Omega}_{ij}^{(K+1)} = 0 = \bar{\Omega}_{ij}^{(K+1)}$

For any $(i,j) \in S^{(K+1)} = \text{supp}(\bar{\Omega}^{(K+1)})$, we have

$$|\hat{\Omega}_{ij}^{(K+1)} - \bar{\Omega}_{ij}^{(K+1)}| \leq \|\hat{\Omega}^{(K+1)} - \bar{\Omega}^{(K+1)}\|_\infty \leq \frac{\omega_{\min}^{(K+1)}}{2} = \frac{1}{2}\min_{1\leq k,l\leq N} \bar{\Omega}_{kl}^{(K+1)} \leq \frac{1}{2}|\bar{\Omega}_{ij}^{(K+1)}|$$

$$\Rightarrow -\frac{1}{2}|\bar{\Omega}_{ij}^{(K+1)}| \leq \hat{\Omega}_{ij}^{(K+1)} - \bar{\Omega}_{ij}^{(K+1)} \leq \frac{1}{2}|\bar{\Omega}_{ij}^{(K+1)}|$$

If $\bar{\Omega}_{ij}^{(K+1)} > 0$, then

$$-\frac{1}{2}\bar{\Omega}_{ij}^{(K+1)} \leq \hat{\Omega}_{ij}^{(K+1)} - \bar{\Omega}_{ij}^{(K+1)}$$

$$\hat{\Omega}_{ij}^{(K+1)} \geq \frac{1}{2}\bar{\Omega}_{ij}^{(K+1)} > 0$$

If $\bar{\Omega}_{ij}^{(K+1)} < 0$, then

$$\hat{\Omega}_{ij}^{(K+1)} - \bar{\Omega}_{ij}^{(K+1)} \leq -\frac{1}{2}\bar{\Omega}_{ij}^{(K+1)}$$

$$\hat{\Omega}_{ij}^{(K+1)} \leq \frac{1}{2}\bar{\Omega}_{ij}^{(K+1)} < 0$$

In conclusion, $\text{sign}(\hat{\Omega}_{ij}^{(K+1)}) = \text{sign}(\bar{\Omega}_{ij}^{(K+1)})$ for $\forall\, i,j \in \{1, 2, ..., N\}$. The estimate $\hat{\Omega}^{(K+1)}$ in (6) is sign-consistent.