APPENDIX

In this appendix, we first provide some auxiliary lemmas, and then provide the proofs of all the statistical theoretical results in Section IV.

A. Auxiliary Lemmas

Lemma 7. *Let $\mathbf{x}$ be a sub-Gaussian random vector with zero mean and covariance $\boldsymbol{\Sigma}^*$ and $\{\mathbf{x}_i\}_{i=1}^n$ be a collection of i.i.d. samples from $\mathbf{x}$. There exists some constants c_1, c_2, and t_0 such that for all t with $0 < t < t_0$, the SCM $\mathbf{S}$ satisfies the following tail bound*

$$\mathbb{P}\left(|\Sigma^*_{ij} - S_{ij}| > t\right) \leq c_1 \exp(-c_2 n t^2).$$

Lemma 8. *Under the same conditions in Lemma 7, if taking $\lambda = \sqrt{\frac{3\log d}{c_2 n}} \asymp \sqrt{\frac{\log d}{n}} < t_0$, then the following result holds*

$$\mathbb{P}\left(\|\boldsymbol{\Sigma}^* - \mathbf{S}\|_{\max} \leq \lambda\right) \geq 1 - \frac{c_1}{d}.$$

Proof: Applying Lemma 7 and union bound, for any λ such that $0 < \lambda < t_0$, we obtain

$$\begin{aligned}\mathbb{P}\left(\|\boldsymbol{\Sigma}^* - \mathbf{S}\|_{\max} > \lambda\right) &\leq c_1 d^2 \exp(-c_2 n\lambda^2)\\ &= c_1 \exp(-c_2 n \lambda^2 + 2\log d).\end{aligned}$$

For n sufficiently large such that $n > \frac{3\log d}{c_2 t_0^2}$, by taking $\lambda = \sqrt{\frac{3\log d}{c_2 n}} \asymp \sqrt{\frac{\log d}{n}} < t_0$, we obtain

$$\begin{aligned}\mathbb{P}\left(\|\boldsymbol{\Sigma}^* - \mathbf{S}\|_{\max} \leq \lambda\right) &\geq 1 - c_1 \exp(-c_2 n\lambda^2 + 2\log d)\\ &= 1 - \frac{c_1}{d}.\end{aligned}$$

■

Lemma 9. *Under the same conditions in Lemma 7, the following result holds*

$$\left\|(\boldsymbol{\Sigma}^* - \mathbf{S})_{\mathcal{S}^*}\right\|_F = O_p\left(\sqrt{\frac{s^*}{n}}\right).$$

Proof: Applying Lemma 7 and union bound, for any M such that $0 < M\sqrt{\frac{1}{n}} < t_0$, we obtain

$$\begin{aligned}&\mathbb{P}\left(\left\|(\boldsymbol{\Sigma}^* - \mathbf{S})_{\mathcal{S}^*}\right\|_{\max} > M\sqrt{\frac{1}{n}}\right)\\ &\leq c_1 s^* \exp(-c_2 M^2)\\ &= c_1 \exp(-c_2 M^2 + \log s^*).\end{aligned}$$

By taking M such that $\sqrt{\frac{2\log s^*}{c_2}} < M < t_0\sqrt{n}$ and $M \to \infty$ in the above inequality obtains

$$\lim_{M\to\infty} \sup_n \mathbb{P}\left(\left\|(\boldsymbol{\Sigma}^* - \mathbf{S})_{\mathcal{S}^*}\right\|_{\max} > M\sqrt{\frac{1}{n}}\right) = 0.$$

The proof is completed by applying $\left\|(\boldsymbol{\Sigma}^* - \mathbf{S})_{\mathcal{S}^*}\right\|_F \leq \sqrt{s^*}\left\|(\boldsymbol{\Sigma}^* - \mathbf{S})_{\mathcal{S}^*}\right\|_{\max}$. ■

B. Technical Lemmas

Each subproblem in (2) corresponds to a weighted ℓ_1 penalized covariance estimation problem, which generally can be written into the following form:

$$\underset{\beta\mathbf{I}\succeq\boldsymbol{\Sigma}\succeq\alpha\mathbf{I}}{\text{minimize}} \quad f(\boldsymbol{\Sigma}) + \|\boldsymbol{\Lambda}\odot\boldsymbol{\Sigma}\|_{1,\text{off}}, \tag{4}$$

where $\boldsymbol{\Lambda}$ is a $d\times d$ matrix of regularization parameters with $\Lambda_{ij} \in [0,\lambda]$.

Lemma 10. *Suppose that Assumption 2 holds. Consider the general problem in (4). Assume that there exists a set $\mathcal{E}$ such that*

$$\mathcal{S}^* \subseteq \mathcal{E},\ |\mathcal{E}| \leq 2s^*,\ \textit{and}\ \|\boldsymbol{\Lambda}_{\overline{\mathcal{E}}}\|_{\min} \geq \frac{\lambda}{2}.$$

If $\lambda \geq 2\|\nabla f(\boldsymbol{\Sigma}^*)\|_{\max}$, then the optimal solution $\widehat{\boldsymbol{\Sigma}}$ satisfies

$$\begin{aligned}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\right\|_F &\leq \left\|(\nabla f(\boldsymbol{\Sigma}^*))_{\mathcal{E}}\right\|_F + \|\boldsymbol{\Lambda}_{\mathcal{S}^*}\|_F\\ &\leq \frac{2+\sqrt{2}}{2}\lambda\sqrt{s^*}.\end{aligned}$$

Proof: By the mean value theorem, there exists a $\rho \in [0,1]$ such that

$$\langle \nabla f(\boldsymbol{\Sigma}) - \nabla f(\boldsymbol{\Sigma}^*), \boldsymbol{\Delta}\rangle = \text{vec}^\top(\boldsymbol{\Delta})\nabla^2 f(\boldsymbol{\Sigma}^* + \rho\boldsymbol{\Delta})\text{vec}(\boldsymbol{\Delta}),$$

where $\boldsymbol{\Delta} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*$. One has

$$\begin{aligned}&\text{vec}^T(\boldsymbol{\Delta})\nabla^2 f(\boldsymbol{\Sigma}^* + \rho\boldsymbol{\Delta})\text{vec}(\boldsymbol{\Delta})\\ &\geq \lambda_{\min}\left(\nabla^2 f(\boldsymbol{\Sigma}^* + \rho\boldsymbol{\Delta})\right)\|\boldsymbol{\Delta}\|_F^2 = \|\boldsymbol{\Delta}\|_F^2.\end{aligned}$$

Hence, we obtain

$$\|\boldsymbol{\Delta}\|_F^2 \leq \langle \nabla f(\boldsymbol{\Sigma}) - \nabla f(\boldsymbol{\Sigma}^*), \boldsymbol{\Delta}\rangle. \tag{5}$$

Define the Lagrangian function of (4) as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\Sigma}, \mathbf{Z}_1, \mathbf{Z}_2) :=& f(\boldsymbol{\Sigma}) + \|\boldsymbol{\Lambda}\odot\boldsymbol{\Sigma}\|_{1,\text{off}}\\ &- \langle \mathbf{Z}_1, \boldsymbol{\Sigma} - \alpha\mathbf{I}\rangle + \langle \mathbf{Z}_2, \boldsymbol{\Sigma} - \beta\mathbf{I}\rangle\end{aligned}$$

where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are $d\times d$ matrix of dual variables. From convex optimization theory, we know that any optimal solution $\widehat{\boldsymbol{\Sigma}}$ to (4) satisfies the following KKT condition:

$$\begin{aligned}\nabla f(\widehat{\boldsymbol{\Sigma}}) + \boldsymbol{\Lambda}\odot\widehat{\boldsymbol{\Xi}} - \widehat{\mathbf{Z}}_1 + \widehat{\mathbf{Z}}_2 = \mathbf{0},\ &\text{with}\ \widehat{\boldsymbol{\Xi}} \in \partial\|\widehat{\boldsymbol{\Sigma}}\|_{1,\text{off}};\\ \left\langle \widehat{\mathbf{Z}}_1, \widehat{\boldsymbol{\Sigma}} - \alpha\mathbf{I}\right\rangle &= 0;\\ \left\langle \widehat{\mathbf{Z}}_2, \widehat{\boldsymbol{\Sigma}} - \beta\mathbf{I}\right\rangle &= 0;\\ \beta\mathbf{I} \succeq \widehat{\boldsymbol{\Sigma}} &\succeq \alpha\mathbf{I};\\ \widehat{\mathbf{Z}}_1 &\succeq \mathbf{0};\\ \widehat{\mathbf{Z}}_2 &\succeq \mathbf{0},\end{aligned} \tag{6}$$

where $\nabla f(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma} - \mathbf{S} + \tau\mathbf{I}$.

Let $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*$. Applying the inequality (5) with $\boldsymbol{\Sigma} = \widehat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Delta} = \widehat{\boldsymbol{\Delta}}$ yields

$$\begin{aligned}\left\|\widehat{\boldsymbol{\Delta}}\right\|_F^2 \leq& \left\langle \nabla f(\widehat{\boldsymbol{\Sigma}}) - \nabla f(\boldsymbol{\Sigma}^*), \widehat{\boldsymbol{\Delta}}\right\rangle\\ =& \underbrace{\left\langle \nabla f(\widehat{\boldsymbol{\Sigma}}) + \boldsymbol{\Lambda}\odot\widehat{\boldsymbol{\Xi}} - \widehat{\mathbf{Z}}_1 + \widehat{\mathbf{Z}}_2, \widehat{\boldsymbol{\Delta}}\right\rangle}_{\text{I}} - \underbrace{\left\langle \nabla f(\boldsymbol{\Sigma}^*), \widehat{\boldsymbol{\Delta}}\right\rangle}_{\text{II}}\\ &- \underbrace{\left\langle \boldsymbol{\Lambda}\odot\widehat{\boldsymbol{\Xi}}, \widehat{\boldsymbol{\Delta}}\right\rangle}_{\text{III}} + \underbrace{\left\langle \widehat{\mathbf{Z}}_1, \widehat{\boldsymbol{\Delta}}\right\rangle}_{\text{IV}} - \underbrace{\left\langle \widehat{\mathbf{Z}}_2, \widehat{\boldsymbol{\Delta}}\right\rangle}_{\text{V}},\end{aligned} \tag{7}$$

where $\widehat{\boldsymbol{\Xi}} \in \partial\|\widehat{\boldsymbol{\Sigma}}\|_{1,\text{off}}$. It remains to bound terms I, II, III, IV, and V, respectively.

For term I, using the KKT condition in (6), we obtain

$$\text{I} = \left\langle \nabla f(\widehat{\boldsymbol{\Sigma}}) + \boldsymbol{\Lambda}\odot\widehat{\boldsymbol{\Xi}} - \widehat{\mathbf{Z}}_1 + \widehat{\mathbf{Z}}_2, \widehat{\boldsymbol{\Delta}}\right\rangle = 0$$

For term II, separating the support of $\nabla f(\mathbf{\Sigma}^*)$ and $\widehat{\mathbf{\Delta}}$ to $\mathcal{E}$ and $\overline{\mathcal{E}}$, and then using the matrix Hölder's inequality, we obtain

$$\begin{aligned}\mathrm{II} &= \left\langle (\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}}, \widehat{\mathbf{\Delta}}_{\mathcal{E}} \right\rangle + \left\langle (\nabla f(\mathbf{\Sigma}^*))_{\overline{\mathcal{E}}}, \widehat{\mathbf{\Delta}}_{\overline{\mathcal{E}}} \right\rangle \\ &\geq -\left\| (\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}} \right\|_F \left\| \widehat{\mathbf{\Delta}}_{\mathcal{E}} \right\|_F - \left\| (\nabla f(\mathbf{\Sigma}^*))_{\overline{\mathcal{E}}} \right\|_{\max} \left\| \widehat{\mathbf{\Delta}}_{\overline{\mathcal{E}}} \right\|_1 \\ &\geq -\left\| (\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}} \right\|_F \left\| \widehat{\mathbf{\Delta}} \right\|_F - \left\| \nabla f(\mathbf{\Sigma}^*) \right\|_{\max} \left\| \widehat{\mathbf{\Delta}}_{\overline{\mathcal{E}}} \right\|_1 .\end{aligned}$$

For term III, separating the support of $\mathbf{\Lambda} \odot \widehat{\mathbf{\Xi}}$ and $\widehat{\mathbf{\Delta}}$ to $\mathcal{S}^*$ and $\overline{\mathcal{S}^*}$, and then using the matrix Hölder's inequality, we obtain

$$\begin{aligned}\mathrm{III} &= \left\langle \left(\mathbf{\Lambda} \odot \widehat{\mathbf{\Xi}}\right)_{\mathcal{S}^*}, \widehat{\mathbf{\Delta}}_{\mathcal{S}^*} \right\rangle + \left\langle \left(\mathbf{\Lambda} \odot \widehat{\mathbf{\Xi}}\right)_{\overline{\mathcal{S}^*}}, \widehat{\mathbf{\Delta}}_{\overline{\mathcal{S}^*}} \right\rangle \\ &= \left\langle \left(\mathbf{\Lambda} \odot \widehat{\mathbf{\Xi}}\right)_{\mathcal{S}^*}, \widehat{\mathbf{\Delta}}_{\mathcal{S}^*} \right\rangle + \left\langle \mathbf{\Lambda}_{\overline{\mathcal{S}^*}}, \left| \widehat{\mathbf{\Delta}}_{\overline{\mathcal{S}^*}} \right| \right\rangle \\ &\geq -\left\| \mathbf{\Lambda}_{\mathcal{S}^*} \right\|_F \left\| \widehat{\mathbf{\Delta}}_{\mathcal{S}^*} \right\|_F + \left\langle \mathbf{\Lambda}_{\overline{\mathcal{E}}}, \left| \widehat{\mathbf{\Delta}}_{\overline{\mathcal{E}}} \right| \right\rangle \\ &\geq -\left\| \mathbf{\Lambda}_{\mathcal{S}^*} \right\|_F \left\| \widehat{\mathbf{\Delta}} \right\|_F + \left\| \mathbf{\Lambda}_{\overline{\mathcal{E}}} \right\|_{\min} \left\| \widehat{\mathbf{\Delta}}_{\overline{\mathcal{E}}} \right\|_1 ,\end{aligned}$$

where the second equality is due to

$$\left\langle \left(\mathbf{\Lambda} \odot \widehat{\mathbf{\Xi}}\right)_{\overline{\mathcal{S}^*}}, \widehat{\mathbf{\Delta}}_{\overline{\mathcal{S}^*}} \right\rangle = \left\langle \mathbf{\Lambda}_{\overline{\mathcal{S}^*}}, \left| \widehat{\mathbf{\Sigma}}_{\overline{\mathcal{S}^*}} \right| \right\rangle = \left\langle \mathbf{\Lambda}_{\overline{\mathcal{S}^*}}, \left| \widehat{\mathbf{\Delta}}_{\overline{\mathcal{S}^*}} \right| \right\rangle ,$$

and the second inequality is due to

$$\begin{aligned}\left\langle \mathbf{\Lambda}_{\overline{\mathcal{E}}}, \left| \widehat{\mathbf{\Delta}}_{\overline{\mathcal{E}}} \right| \right\rangle &= \sum_{(i,j)\in\overline{\mathcal{E}}} \Lambda_{ij} |\widehat{\Delta}_{ij}| \geq \left\| \mathbf{\Lambda}_{\overline{\mathcal{E}}} \right\|_{\min} \sum_{(i,j)\in\overline{\mathcal{E}}} |\widehat{\Delta}_{ij}| \\ &= \left\| \mathbf{\Lambda}_{\overline{\mathcal{E}}} \right\|_{\min} \left\| \widehat{\mathbf{\Delta}}_{\overline{\mathcal{E}}} \right\|_1 .\end{aligned}$$

For term IV, we obtain

$$\begin{aligned}\mathrm{IV} &= \left\langle \widehat{\mathbf{Z}}_1, \widehat{\mathbf{\Sigma}} \right\rangle - \left\langle \widehat{\mathbf{Z}}_1, \mathbf{\Sigma}^* \right\rangle \\ &= \left\langle \widehat{\mathbf{Z}}_1, \widehat{\mathbf{\Sigma}} - \alpha \mathbf{I} \right\rangle + \left\langle \widehat{\mathbf{Z}}_1, \alpha \mathbf{I} \right\rangle - \left\langle \widehat{\mathbf{Z}}_1, \mathbf{\Sigma}^* \right\rangle \\ &= \left\langle \widehat{\mathbf{Z}}_1, \alpha \mathbf{I} - \mathbf{\Sigma}^* \right\rangle \leq 0.\end{aligned}$$

For term V, we obtain

$$\begin{aligned}\mathrm{V} &= \left\langle \widehat{\mathbf{Z}}_2, \widehat{\mathbf{\Sigma}} \right\rangle - \left\langle \widehat{\mathbf{Z}}_2, \mathbf{\Sigma}^* \right\rangle \\ &= \left\langle \widehat{\mathbf{Z}}_2, \widehat{\mathbf{\Sigma}} - \beta \mathbf{I} \right\rangle + \left\langle \widehat{\mathbf{Z}}_2, \beta \mathbf{I} \right\rangle - \left\langle \widehat{\mathbf{Z}}_2, \mathbf{\Sigma}^* \right\rangle \\ &= \left\langle \widehat{\mathbf{Z}}_2, \beta \mathbf{I} - \mathbf{\Sigma}^* \right\rangle \geq 0.\end{aligned}$$

Substituting the above results into (7) yields

$$\begin{aligned}\left\| \widehat{\mathbf{\Delta}} \right\|_F^2 \leq\ & \left(\left\| (\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}} \right\|_F + \left\| \mathbf{\Lambda}_{\mathcal{S}^*} \right\|_F \right) \left\| \widehat{\mathbf{\Delta}} \right\|_F \\ & + \left(\left\| \nabla f(\mathbf{\Sigma}^*) \right\|_{\max} - \left\| \mathbf{\Lambda}_{\overline{\mathcal{E}}} \right\|_{\min} \right) \left\| \widehat{\mathbf{\Delta}}_{\overline{\mathcal{E}}} \right\|_1 \\ \leq\ & \left(\left\| (\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}} \right\|_F + \left\| \mathbf{\Lambda}_{\mathcal{S}^*} \right\|_F \right) \left\| \widehat{\mathbf{\Delta}} \right\|_F ,\end{aligned} \tag{8}$$

where the second inequality is due to $\left\| \mathbf{\Lambda}_{\overline{\mathcal{E}}} \right\|_{\min} \geq \frac{\lambda}{2} \geq \left\| \nabla f(\mathbf{\Sigma}^*) \right\|_{\max}$. Dividing by $\left\| \widehat{\mathbf{\Delta}} \right\|_F$ on both sides of the inequality (8), we have

$$\begin{aligned}\left\| \widehat{\mathbf{\Delta}} \right\|_F &\leq \left\| (\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}} \right\|_F + \left\| \mathbf{\Lambda}_{\mathcal{S}^*} \right\|_F \\ &\leq \left\| (\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}} \right\|_{\max} \sqrt{|\mathcal{E}|} + \left\| \mathbf{\Lambda}_{\mathcal{S}^*} \right\|_{\max} \sqrt{|\mathcal{S}^*|} \\ &\leq \left\| (\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}} \right\|_{\max} \sqrt{2s^*} + \lambda \sqrt{s^*} \\ &\leq \frac{2+\sqrt{2}}{2} \lambda \sqrt{s^*}.\end{aligned}$$

■

Lemma 11. *Suppose that Assumptions 2 hold. Consider the problem in* (3). *Define the set* $\mathcal{E}^{(k)}$ *by*

$$\mathcal{E}^{(k)} = \mathcal{S}^* \cup \mathcal{S}^{(k)}, \text{ with } \mathcal{S}^{(k)} = \left\{ (i,j) \mid \Lambda_{ij}^{(k-1)} < p'_\lambda(u) \right\},$$

where $u = c\lambda$ *and* $c = \frac{2+\sqrt{2}}{2}$ *is the same to that given in Assumption 2. If* $\lambda \geq 2\left\| \nabla f(\mathbf{\Sigma}^*) \right\|_{\max}$*, then for* $k \geq 1$*, we have* $|\mathcal{E}^{(k)}| \leq 2s^*$*,* $\left\| \mathbf{\Lambda}_{\mathcal{E}^{(k)}}^{(k-1)} \right\|_{\min} \geq \frac{\lambda}{2}$*, and*

$$\begin{aligned}\left\| \widehat{\mathbf{\Sigma}}^{(k)} - \mathbf{\Sigma}^* \right\|_F &\leq \left\| (\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}^{(k)}} \right\|_F + \left\| \mathbf{\Lambda}_{\mathcal{S}^*}^{(k-1)} \right\|_F \\ &\leq \frac{2+\sqrt{2}}{2} \lambda \sqrt{s^*}.\end{aligned}$$

Proof: We first prove $|\mathcal{E}^{(k)}| \leq 2s^*$ holds by induction. For $k=1$, we have $\Lambda_{ij}^{(0)} = \lambda \geq p'_\lambda(u)$ and thus $\mathcal{S}^{(1)} = \emptyset$ and $\mathcal{E}^{(1)} = \mathcal{S}^*$, which implies $|\mathcal{E}^{(1)}| \leq 2s^*$ holds. Assume $|\mathcal{E}^{(k)}| \leq 2s^*$ holds at $k-1$, i.e., $|\mathcal{E}^{(k-1)}| \leq 2s^*$ holds for some $k \geq 2$. Next, we will prove $|\mathcal{E}^{(k)}| \leq 2s^*$ holds at k. For any $(i,j) \in \mathcal{S}^{(k)}$, we obtain $|\widehat{\mathbf{\Sigma}}_{ij}^{(k-1)}| \geq u$ and further have

$$\begin{aligned}\sqrt{|\mathcal{S}^{(k)} \backslash \mathcal{S}^*|} &\leq \sqrt{\sum_{(i,j)\in\mathcal{S}^{(k)}\backslash\mathcal{S}^*} \left(u^{-1} \widehat{\mathbf{\Sigma}}_{ij}^{(k-1)} \right)^2} \\ &= u^{-1} \left\| \widehat{\mathbf{\Sigma}}_{\mathcal{S}^{(k)}\backslash\mathcal{S}^*}^{(k-1)} \right\|_F \\ &= u^{-1} \left\| \left(\widehat{\mathbf{\Sigma}}^{(k-1)} - \mathbf{\Sigma}^* \right)_{\mathcal{S}^{(k)}\backslash\mathcal{S}^*} \right\|_F \\ &\leq u^{-1} \left\| \widehat{\mathbf{\Sigma}}^{(k-1)} - \mathbf{\Sigma}^* \right\|_F .\end{aligned} \tag{9}$$

For any $(i,j) \in \overline{\mathcal{S}^{(k-1)}}$, we have $\Lambda_{ij}^{(k-2)} = p'_\lambda(\widehat{\mathbf{\Sigma}}_{ij}^{(k-2)}) \geq p'_\lambda(u) \geq \frac{\lambda}{2}$, which implies

$$\left\| \mathbf{\Lambda}_{\overline{\mathcal{E}^{(k-1)}}}^{(k-2)} \right\|_{\min} \geq \left\| \mathbf{\Lambda}_{\overline{\mathcal{S}^{(k-1)}}}^{(k-2)} \right\|_{\min} \geq p'_\lambda(u) \geq \frac{\lambda}{2}.$$

One also has $|\mathcal{E}^{(k-1)}| \leq 2s^*$ and $\mathcal{S}^* \subseteq \mathcal{E}^{(k-1)}$. Applying Lemma 10 with $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Sigma}}^{(k-1)}$, $\mathcal{E} = \mathcal{E}^{(k-1)}$, and $\mathbf{\Lambda}_{\mathcal{S}^*} = \mathbf{\Lambda}_{\mathcal{S}^*}^{(k-2)}$ yields

$$\left\| \widehat{\mathbf{\Sigma}}^{(k-1)} - \mathbf{\Sigma}^* \right\|_F \leq \frac{2+\sqrt{2}}{2} \lambda \sqrt{s^*}.$$

Substituting the above result into the inequality (9) yields

$$\sqrt{|\mathcal{S}^{(k)} \backslash \mathcal{S}^*|} \leq \frac{2+\sqrt{2}}{2u} \lambda \sqrt{s^*} = \sqrt{s^*}.$$

Thus, we have

$$|\mathcal{E}^{(k)}| = |\mathcal{S}^* \cup (\mathcal{S}^{(k)} \backslash \mathcal{S}^*)| = |\mathcal{S}^*| + |\mathcal{S}^{(k)} \backslash \mathcal{S}^*| \leq 2s^*,$$

completing the induction.

Then, by the definition of $\mathcal{E}^{(k)}$ and $\mathcal{S}^{(k)}$, we have

$$\left\| \mathbf{\Lambda}_{\overline{\mathcal{E}^{(k)}}}^{(k-1)} \right\|_{\min} \geq \left\| \mathbf{\Lambda}_{\overline{\mathcal{S}^{(k)}}}^{(k-1)} \right\|_{\min} \geq p'_\lambda(u) \geq \frac{\lambda}{2}.$$

Applying Lemma 10 with $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Sigma}}^{(k)}$, $\mathcal{E} = \mathcal{E}^{(k)}$, and $\mathbf{\Lambda}_{\mathcal{S}^*} = \mathbf{\Lambda}_{\mathcal{S}^*}^{(k-1)}$, the optimal solution $\widehat{\mathbf{\Sigma}}^{(k)}$ to (3) satisfies

$$\begin{aligned}\left\| \widehat{\mathbf{\Sigma}}^{(k)} - \mathbf{\Sigma}^* \right\|_F &\leq \left\| (\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}^{(k)}} \right\|_F + \left\| \mathbf{\Lambda}_{\mathcal{S}^*}^{(k-1)} \right\|_F \\ &\leq \frac{2+\sqrt{2}}{2} \lambda \sqrt{s^*}.\end{aligned}$$

■

C. Proof of Theorem 4

Proof: By Lemma 11, we have

$$\left\|\widehat{\mathbf{\Sigma}}^{(k)}-\mathbf{\Sigma}^*\right\|_F \leq \underbrace{\left\|(\nabla f(\mathbf{\Sigma}^*))_{\mathcal{E}^{(k)}}\right\|_F}_{\mathrm{I}}+\underbrace{\left\|\mathbf{\Lambda}_{\mathcal{S}^*}^{(k-1)}\right\|_F}_{\mathrm{II}} \tag{10}$$

Next, we bound the terms I and II, respectively.

For term I, separating the support set into $\mathcal{S}^*$ and $\mathcal{E}^{(k)}\backslash\mathcal{S}^*$, we obtain

$$\begin{aligned}\mathrm{I} &\leq \left\|(\nabla f(\mathbf{\Sigma}^*))_{\mathcal{S}^*}\right\|_F+\|\nabla f(\mathbf{\Sigma}^*)\|_{\max}\sqrt{|\mathcal{E}^{(k)}\backslash\mathcal{S}^*|}\\ &\leq \left\|(\nabla f(\mathbf{\Sigma}^*))_{\mathcal{S}^*}\right\|_F+\frac{1}{2}\lambda u^{-1}\left\|\widehat{\mathbf{\Sigma}}^{(k-1)}-\mathbf{\Sigma}^*\right\|_F,\end{aligned}$$

where the second inequality is due to

$$\sqrt{|\mathcal{E}^{(k)}\backslash\mathcal{S}^*|}=\sqrt{|\mathcal{S}^{(k)}\backslash\mathcal{S}^*|}\leq u^{-1}\left\|\widehat{\mathbf{\Sigma}}^{(k-1)}-\mathbf{\Sigma}^*\right\|_F,$$

which follows from the inequality (9).

By Assumptions 1 and 3, for any $\mathbf{\Sigma}$, if $\left|\Sigma_{ij}-\Sigma_{ij}^*\right|\geq u$, then $p_\lambda'(|\Sigma_{ij}|)\leq\lambda\leq\lambda u^{-1}\left|\Sigma_{ij}-\Sigma_{ij}^*\right|$; otherwise, $p_\lambda'(|\Sigma_{ij}|)\leq p_\lambda'(|\Sigma_{ij}^*|-u)=0$. Therefore, for term V, we have

$$\mathrm{II}\leq\lambda u^{-1}\left\|\widehat{\mathbf{\Sigma}}_{\mathcal{S}^*}^{(k-1)}-\mathbf{\Sigma}_{\mathcal{S}^*}^*\right\|_F\leq\lambda u^{-1}\left\|\widehat{\mathbf{\Sigma}}^{(k-1)}-\mathbf{\Sigma}^*\right\|_F.$$

Substituting the above results into (10) yields

$$\left\|\widehat{\mathbf{\Sigma}}^{(k)}-\mathbf{\Sigma}^*\right\|_F\leq\left\|(\nabla f(\mathbf{\Sigma}^*))_{\mathcal{S}^*}\right\|_F+\delta\left\|\widehat{\mathbf{\Sigma}}^{(k-1)}-\mathbf{\Sigma}^*\right\|_F,$$

where $\delta=\frac{3\lambda}{2u}=\frac{3}{2+\sqrt{2}}\in(0,1)$. ■

D. Proof of Corollary 5

Proof: Since $\nabla f(\mathbf{\Sigma})=\mathbf{\Sigma}-\mathbf{S}+\tau\mathbf{I}$, one has

$$\|\nabla f(\mathbf{\Sigma}^*)\|_{\max}\leq\|\mathbf{\Sigma}^*-\mathbf{S}\|_{\max}+\tau.$$

If λ and τ satisfy

$$\lambda\asymp\sqrt{\frac{\log d}{n}},\ \tau\lesssim\sqrt{\frac{1}{n}},$$

then by Lemma 8, $\lambda\geq 2\|\nabla f(\mathbf{\Sigma}^*)\|_{\max}$ holds with high probability (w.h.p).

Applying Lemma 11 with $k=1$, we obtain

$$\left\|\widehat{\mathbf{\Sigma}}^{(1)}-\mathbf{\Sigma}^*\right\|_F\leq\frac{2+\sqrt{2}}{2}\lambda\sqrt{s^*}.$$

If $\lambda\asymp\sqrt{\frac{\log d}{n}}$, then $\left\|\widehat{\mathbf{\Sigma}}^{(1)}-\mathbf{\Sigma}^*\right\|_F\lesssim\sqrt{\frac{s^*\log d}{n}}$ w.h.p. ■

E. Proof of Corollary 6

Proof: One has

$$\|\nabla f(\mathbf{\Sigma}^*)\|_{\max}\leq\|\mathbf{\Sigma}^*-\mathbf{S}\|_{\max}+\tau.$$

If λ, τ, and ε satisfy

$$\lambda\asymp\sqrt{\frac{\log d}{n}},\ \tau\lesssim\sqrt{\frac{1}{n}},$$

then by Lemma 8, $\lambda\geq 2\|\nabla f(\mathbf{\Sigma}^*)\|_{\max}$ holds w.h.p.

Applying Theorem 4, we obtain

$$\begin{aligned}&\left\|\widehat{\mathbf{\Sigma}}^{(k)}-\mathbf{\Sigma}^*\right\|_F\\ \leq\ &\left\|(\nabla f(\mathbf{\Sigma}^*))_{\mathcal{S}^*}\right\|_F+\delta\left\|\widehat{\mathbf{\Sigma}}^{(k-1)}-\mathbf{\Sigma}^*\right\|_F\\ \leq\ &\frac{1}{1-\delta}\left\|(\nabla f(\mathbf{\Sigma}^*))_{\mathcal{S}^*}\right\|_F+\delta^{k-1}\left\|\widehat{\mathbf{\Sigma}}^{(1)}-\mathbf{\Sigma}^*\right\|_F\\ \leq\ &\frac{1}{1-\delta}\left\|(\nabla f(\mathbf{\Sigma}^*))_{\mathcal{S}^*}\right\|_F+\delta^{k-1}\frac{2+\sqrt{2}}{2}\lambda\sqrt{s^*},\end{aligned}$$

where the last inequality is due to $\left\|\widehat{\mathbf{\Sigma}}^{(1)}-\mathbf{\Sigma}^*\right\|_F\leq\frac{2+\sqrt{2}}{2}\lambda\sqrt{s^*}$, which follows from Lemma 11 with $k=1$.

One has

$$\begin{aligned}\left\|(\nabla f(\mathbf{\Sigma}^*))_{\mathcal{S}^*}\right\|_F &= \left\|(\mathbf{\Sigma}^*-\mathbf{S}+\tau\mathbf{I})_{\mathcal{S}^*}\right\|_F\\ &\leq \left\|(\mathbf{\Sigma}^*-\mathbf{S})_{\mathcal{S}^*}\right\|_F+\tau\left\|(\mathbf{I})_{\mathcal{S}^*}\right\|_F\\ &\leq \left\|(\mathbf{\Sigma}^*-\mathbf{S})_{\mathcal{S}^*}\right\|_F+\tau\sqrt{s^*}.\end{aligned}$$

By Lemma 9, $\left\|(\mathbf{\Sigma}^*-\mathbf{S})_{\mathcal{S}^*}\right\|_F=O_p\left(\sqrt{\frac{s^*}{n}}\right)$. If $\tau\lesssim\sqrt{\frac{1}{n}}$, then $\left\|(\nabla f(\mathbf{\Sigma}^*))_{\mathcal{S}^*}\right\|_F=O_p\left(\sqrt{\frac{s^*}{n}}\right)$.

If $K\geq 1+\frac{\log(\lambda\sqrt{n})}{\log\delta^{-1}}\gtrsim\log(\lambda\sqrt{n})\gtrsim\log\log d$, then we have

$$\delta^{K-1}\lambda\sqrt{s^*}\leq\frac{1}{\lambda\sqrt{n}}\lambda\sqrt{s^*}\leq\sqrt{\frac{s^*}{n}},$$

which yields that $\left\|\widehat{\mathbf{\Sigma}}^{(K)}-\mathbf{\Sigma}^*\right\|_F=O_p\left(\sqrt{\frac{s^*}{n}}\right)$. ■