

Supplementary Material: Nonconvex Sparse Graph Learning under Laplacian Constrained Graphical Model

Jiaxi Ying ^{*} José Vinícius de M. Cardoso [†] Daniel P. Palomar [‡]

Abstract

In this supplementary material, Section 1 provides the proofs of Theorems 3.1, 3.4, 3.8, and Corollary 3.9. The proofs of technical lemmas are collected in Section 2. Section 3 presents the additional experimental results.

1 Proofs of Theorems

This section contains the proofs of Theorems 3.1, 3.4, 3.8, and Corollary 3.9. Before starting to prove the theorems, we first present some definitions and technical lemmas.

Recall that the optimization problem we aim to solve is

$$\min_{\mathbf{w} \geq \mathbf{0}} -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J}) + \text{tr}(\mathbf{S}\mathcal{L}\mathbf{w}) + \sum_i h_\lambda(w_i). \quad (1)$$

Let $F(\mathbf{w}) = -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J}) + \text{tr}(\mathbf{S}\mathcal{L}\mathbf{w}) + \sum_i h_\lambda(w_i)$. The feasible set of (1) is $\mathcal{S}_w = \{\mathbf{w} \mid \mathbf{w} \geq \mathbf{0}, \mathbf{w} \in \text{dom}(F)\}$, where $\text{dom}(F)$ denotes the domain of the function F . One can verify that

$$\text{dom}(F) = \{\mathbf{w} \in \mathbb{R}^{p(p-1)/2} \mid \det(\mathcal{L}\mathbf{w} + \mathbf{J}) > 0\}. \quad (2)$$

The set \mathcal{S}_w can be equivalently written as

$$\mathcal{S}_w = \{\mathbf{w} \in \mathbb{R}^{p(p-1)/2} \mid \mathbf{w} \geq \mathbf{0}, (\mathcal{L}\mathbf{w} + \mathbf{J}) \in \mathcal{S}_{++}^p\}, \quad (3)$$

which is due to the reason that $\mathcal{L}\mathbf{w} + \mathbf{J}$ must be positive semi-definite, because $\mathcal{L}\mathbf{w}$ is positive semi-definite for any $\mathbf{w} \geq \mathbf{0}$ following from (27), while the matrix \mathbf{J} is rank one, and the nonzero eigenvalue of \mathbf{J} is 1 whose eigenvector is orthogonal to the row and column spaces of $\mathcal{L}\mathbf{w}$. The condition $\det(\mathcal{L}\mathbf{w} + \mathbf{J}) > 0$ in (2) implies that $\mathcal{L}\mathbf{w} + \mathbf{J}$ is non-singular. The non-singularity and positive semi-definiteness of $\mathcal{L}\mathbf{w} + \mathbf{J}$ together lead to $(\mathcal{L}\mathbf{w} + \mathbf{J}) \in \mathcal{S}_{++}^p$.

Next, we will show that \mathcal{S}_w is a convex set. For any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{S}_w$, define $\mathbf{x}_t = t\mathbf{x}_1 + (1-t)\mathbf{x}_2$, $t \in [0, 1]$. It is clear that $\mathbf{x}_t \geq \mathbf{0}$. Since \mathcal{S}_{++}^p is a convex cone, one has

$$\mathcal{L}\mathbf{x}_t + \mathbf{J} = t(\mathcal{L}\mathbf{x}_1 + \mathbf{J}) + (1-t)(\mathcal{L}\mathbf{x}_2 + \mathbf{J}) \in \mathcal{S}_{++}^p, \quad (4)$$

indicating that $\mathbf{x}_t \in \mathcal{S}_w$ and thus \mathcal{S}_w is convex.

^{*}Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR; Email: jx.ying@connect.ust.hk.

[†]Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR; Email: jvdmc@connect.ust.hk.

[‡]Department of Electronic and Computer Engineering, Department of Industrial Engineering and Data Analytics, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR; Email: palomar@ust.hk.

Recall that the sequence $\{\hat{\mathbf{w}}^{(k)}\}_{k \geq 1}$ is established by solving

$$\hat{\mathbf{w}}^{(k)} = \arg \min_{\mathbf{w} \geq \mathbf{0}} -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J}) + \text{tr}(\mathcal{S}\mathcal{L}\mathbf{w}) + \sum_i z_i^{(k-1)} w_i, \quad (5)$$

where $z_i^{(k-1)} = h'_\lambda(\hat{w}_i^{(k-1)})$, $i \in [p(p-1)/2]$, and h_λ satisfies the conditions in Assumption 3.5. Let $\|\mathbf{x}\|_{\max} = \max_i |x_i|$ and $\|\mathbf{x}\|_{\min} = \min_i |x_i|$. We present the technical lemmas in Section 1.1 for establishing the theorems.

1.1 Technical Lemmas

Lemma 1.1. *Let $f(\mathbf{w}) = -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J})$, with $\mathbf{w} \in \mathbb{R}^{p(p-1)/2}$. Then for any $\mathbf{x} \in \mathbb{R}^{p(p-1)/2}$, we have*

$$\mathbf{x}^\top \nabla^2 f(\mathbf{w}) \mathbf{x} = \text{vec}(\mathcal{L}\mathbf{x})^\top \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \otimes (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right) \text{vec}(\mathcal{L}\mathbf{x}).$$

Lemma 1.2. *For any given $\mathbf{w} \in \mathbb{R}^{p(p-1)/2}$ satisfying $(\mathcal{L}\mathbf{w} + \mathbf{J}) \in \mathcal{S}_{++}^p$, there must exist a unique $\mathbf{x} \in \mathbb{R}^{p(p-1)/2}$ such that*

$$\mathcal{L}\mathbf{x} + \frac{1}{b}\mathbf{J} = (\mathcal{L}\mathbf{w} + b\mathbf{J})^{-1} \quad (6)$$

holds for any $b \neq 0$, where $\mathbf{J} = \frac{1}{p}\mathbf{1}_{p \times p}$, in which $\mathbf{1}_{p \times p} \in \mathbb{R}^{p \times p}$ with each element equals 1.

Lemma 1.3. *Let $\mathcal{G} = \mathcal{L}^* \mathcal{L} : \mathbb{R}^{p(p-1)/2} \rightarrow \mathbb{R}^{p(p-1)/2}$, $\mathbf{x} \mapsto \mathcal{L}^* \mathcal{L}\mathbf{x}$. For any $\mathbf{x} \in \mathbb{R}^{p(p-1)/2}$, $\mathcal{G}\mathbf{x} = \mathbf{M}\mathbf{x}$ with $\mathbf{M} \in \mathbb{R}^{\frac{p(p-1)}{2} \times \frac{p(p-1)}{2}}$ satisfying*

$$M_{kl} = \begin{cases} 4 & l = k, \\ 1 & l \in (\Omega_i \cup \Omega_j) \setminus k, \\ 0 & \text{Otherwise,} \end{cases}$$

where $i, j \in [p]$ satisfying $k = i - j + \frac{j-1}{2}(2p-j)$ and $i > j$, and Ω_t is an index set defined by

$$\Omega_t := \left\{ l \in [p(p-1)/2] \mid [L\mathbf{x}]_{tt} = \sum_l x_l \right\}, \quad t \in [p].$$

Furthermore, we have $\lambda_{\min}(\mathbf{M}) = 2$ and $\lambda_{\max}(\mathbf{M}) = 2p$.

Lemma 1.4. *Take $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ and suppose $n \geq 94\alpha c_0^{-1} \lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) s \log p$ for some $\alpha > 2$, where c_0 is a constant defined in Lemma 1.8. Let*

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \geq \mathbf{0}} -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J}) + \text{tr}(\mathcal{L}\mathbf{w}\mathcal{S}) + \mathbf{z}^\top \mathbf{w},$$

where \mathbf{z} obeys $0 \leq z_i \leq \lambda$ for $i \in [p(p-1)/2]$. If $\|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathcal{S})\|_{\max} \leq \lambda/2 \leq \|\mathbf{z}_{\mathcal{E}^c}\|_{\min}$ holds with the set \mathcal{E} satisfying $\mathcal{S}^* \subseteq \mathcal{E}$ and $|\mathcal{E}| \leq 2s$, then $\hat{\mathbf{w}}$ obeys

$$\|\mathcal{L}\hat{\mathbf{w}} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq 2\sqrt{2}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left(\|\mathbf{z}_{\mathcal{S}^*}\| + \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathcal{S}))_{\mathcal{E}}\| \right) \leq 2(1 + \sqrt{2})\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\sqrt{s}\lambda,$$

where \mathcal{S}^* is the support of \mathbf{w}^* .

Lemma 1.5. Under Assumption 3.5, take $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ and suppose $n \geq 94\alpha c_0^{-1} \lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)s \log p$ for some $\alpha > 2$, where c_0 is a constant defined in Lemma 1.8. Define the set $\mathcal{E}^{(k)}$ by

$$\mathcal{E}^{(k)} = \{\mathcal{S}^* \cup \mathcal{S}^{(k)}\}, \quad \text{with } \mathcal{S}^{(k)} = \{i \in [p(p-1)/2] \mid \hat{w}_i^{(k-1)} \geq b\}, \quad (7)$$

where $\hat{w}^{(k)}$ for $k \geq 1$ is defined in (5), \mathcal{S}^* is the support of \mathbf{w}^* with $|\mathcal{S}^*| \leq s$ and $b = (2 + \sqrt{2})\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\lambda$ is a constant. If $\|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})\|_{\max} \leq \lambda/2$ holds and $\hat{\mathbf{w}}^{(0)}$ satisfies $|\text{supp}^+(\hat{\mathbf{w}}^{(0)})| \leq s$, then $\mathcal{E}^{(k)}$ obeys $|\mathcal{E}^{(k)}| \leq 2s$, for any $k \geq 1$.

Lemma 1.6. Under Assumptions 3.5 and 3.6, take $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ and suppose $n \geq 94\alpha c_0^{-1} \lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)s \log p$ for some $\alpha > 2$, where c_0 is a constant defined in Lemma 1.8. If $\|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})\|_{\max} \leq \lambda/2$ holds and $\hat{\mathbf{w}}^{(0)}$ satisfies $|\text{supp}^+(\hat{\mathbf{w}}^{(0)})| \leq s$, then for any $k \geq 1$, $\hat{\mathbf{w}}^{(k)}$ defined in (5) obeys

$$\|\hat{\mathbf{w}}^{(k)} - \mathbf{w}^*\| \leq 2\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left\| (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*} \right\| + \frac{3}{2 + \sqrt{2}} \|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\|,$$

and

$$\|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq 2\sqrt{2}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left\| (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*} \right\| + \frac{3}{2 + \sqrt{2}} \|\mathcal{L}\hat{\mathbf{w}}^{(k-1)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}}.$$

Lemma 1.7. Take $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ and suppose $n \geq 8\alpha \log p$ for some $\alpha > 2$, where c_0 is a constant defined in Lemma 1.8. Then one has

$$\mathbb{P}[\|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})\|_{\max} \leq \lambda/2] \geq 1 - 1/p^{\alpha-2}.$$

Lemma 1.8. Consider a zero-mean random vector $\mathbf{X} = [X_1, \dots, X_p]^\top \in \mathbb{R}^p$ is a LGMRF with precision matrix $\mathcal{L}\mathbf{w}^* \in \mathcal{S}_L$. Given n i.i.d samples $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$, the associated sample covariance matrix $\mathbf{S} = \frac{1}{n} \sum_{k=1}^n \mathbf{X}^{(k)} \mathbf{X}^{(k)T}$ satisfies, for $t \in [0, t_0]$,

$$\mathbb{P} [|[\mathcal{L}^*\mathbf{S}]_i - (\mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1})_i| \geq t] \leq 2 \exp(-c_0 n t^2), \quad \text{for } i = 1, \dots, p(p-1)/2,$$

where $t_0 = \|\mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1}\|_{\max}$ and $c_0 = 1/(8\|\mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1}\|_{\max}^2)$ are two constants.

Lemma 1.9. Let $f(\mathbf{w}) = -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J})$. Define a local region of \mathbf{w}^* by

$$\mathcal{B}_{\mathbf{M}}(\mathbf{w}^*; r) = \{\mathbf{w} \mid \mathbf{w} \in \mathbb{B}_{\mathbf{M}}(\mathbf{w}^*; r) \cap \mathcal{S}_{\mathbf{w}}\}.$$

where $\mathbb{B}_{\mathbf{M}}(\mathbf{w}^*; r) = \{\mathbf{w} \in \mathbb{R}^{p(p-1)/2} \mid \|\mathbf{w} - \mathbf{w}^*\|_{\mathbf{M}} \leq r\}$, and $\mathcal{S}_{\mathbf{w}} = \{\mathbf{w} \mid \mathbf{w} \geq \mathbf{0}, (\mathcal{L}\mathbf{w} + \mathbf{J}) \in \mathcal{S}_{++}^p\}$. Then for any $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{B}_{\mathbf{M}}(\mathbf{w}^*; r)$, we have

$$\langle \nabla f(\mathbf{w}_1) - \nabla f(\mathbf{w}_2), \mathbf{w}_1 - \mathbf{w}_2 \rangle \geq (\|\mathcal{L}\mathbf{w}^*\|_2 + r)^{-2} \|\mathcal{L}\mathbf{w}_1 - \mathcal{L}\mathbf{w}_2\|_{\mathbb{F}}^2.$$

Lemma 1.10. [7] Suppose a positive matrix $\mathbf{A} \in \mathbb{R}^{p \times p}$ is diagonally scaled such that $A_{ii} = 1$, $i = 1, \dots, p$, and $0 < A_{ij} < 1$, $i \neq j$. Let y and x be the lower and upper bounds satisfying

$$0 < y \leq A_{ij} \leq x < 1, \quad \forall i \neq j,$$

and define s by

$$x^2 = sy + (1-s)y^2.$$

Then the inverse matrix of \mathbf{A} exists and \mathbf{A} is an inverse M -matrix if $s^{-1} \geq p-2$ with $p > 3$.

Lemma 1.11. [6] (Sub-exponential tail bound) Suppose X is sub-exponential with parameters (v, α) . Then

$$\mathbb{P}[X - \mu \geq t] \leq \begin{cases} e^{-\frac{t^2}{2v^2}} & t \in [0, \frac{v^2}{\alpha}], \\ e^{-\frac{t}{2\alpha}} & t \in (\frac{v^2}{\alpha}, +\infty). \end{cases}$$

1.2 Proof of Theorem 3.1

Proof. The ℓ_1 -norm regularized maximum likelihood estimation under the Laplacian constrained Gaussian graphical model can be formulated as

$$\min_{\Theta \in \mathcal{S}_L} -\log \det(\Theta + \mathbf{J}) + \text{tr}(\Theta \mathbf{S}) + \lambda \sum_{i>j} |\Theta_{ij}|, \quad (8)$$

where \mathbf{S} is the sample covariance matrix with the samples generated from LGMRF. As a result of Theorem 3.4, the optimization (8) can be equivalently written as

$$\min_{\mathbf{w} \geq \mathbf{0}} -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J}) + \text{tr}(\mathcal{L}\mathbf{w}\mathbf{S}) + \lambda \|\mathbf{w}\|_1. \quad (9)$$

Due to the non-negativity constraint $\mathbf{w} \geq \mathbf{0}$, (9) can be further rewritten as

$$\min_{\mathbf{w} \geq \mathbf{0}} -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J}) + \langle \mathcal{L}^* \mathbf{S} + \lambda \mathbf{1}, \mathbf{w} \rangle, \quad (10)$$

where $\mathbf{1} = [1, \dots, 1]^\top$.

We first prove that the optimization (10) has one global minimizer if $\lambda > 0$. Let $f(\mathbf{w}) = -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J}) + \langle \mathcal{L}^* \mathbf{S} + \lambda \mathbf{1}, \mathbf{w} \rangle$. The feasible set of (10) is $\mathcal{S}_{\mathbf{w}} = \{\mathbf{w} \in \mathbb{R}^{p(p-1)/2} \mid \mathbf{w} \geq \mathbf{0}, (\mathcal{L}\mathbf{w} + \mathbf{J}) \in \mathcal{S}_{++}^p\}$, which is the same with the feasible set of (1). For any $\mathbf{w} \in \mathcal{S}_{\mathbf{w}}$, the minimum eigenvalue of $\nabla^2 f(\mathbf{w})$ can be lower bounded by

$$\begin{aligned} \lambda_{\min}(\nabla^2 f(\mathbf{w})) &= \inf_{\|\mathbf{x}\|=1} \mathbf{x}^\top \nabla^2 f(\mathbf{w}) \mathbf{x} \\ &= \inf_{\|\mathbf{x}\|=1} (\text{vec}(\mathcal{L}\mathbf{x}))^\top \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \otimes (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right) \text{vec}(\mathcal{L}\mathbf{x}) \\ &\geq \inf_{\|\mathbf{x}\|=1} \frac{(\text{vec}(\mathcal{L}\mathbf{x}))^\top \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \otimes (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right) \text{vec}(\mathcal{L}\mathbf{x})}{(\text{vec}(\mathcal{L}\mathbf{x}))^\top \text{vec}(\mathcal{L}\mathbf{x})} \cdot \inf_{\|\mathbf{x}\|=1} \|\mathcal{L}\mathbf{x}\|_{\text{F}}^2 \\ &\geq \lambda_{\min} \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \otimes (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right) \cdot \inf_{\|\mathbf{x}\|=1} \|\mathcal{L}\mathbf{x}\|_{\text{F}}^2 \\ &= \lambda_{\min} \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \otimes (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right) \cdot \inf_{\|\mathbf{x}\|=1} \mathbf{x}^\top \mathbf{M} \mathbf{x} \\ &= 2\lambda_{\min} \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right)^2 \\ &> 0, \end{aligned}$$

where the second equality is due to Lemma 1.1; the third equality follows from Lemma 1.3; the last equality follows from the property of Kronecker product that the eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ are $\lambda_i \mu_j$ for $i, j \in [p]$, where λ_i and μ_j are the eigenvalues of $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\mathbf{B} \in \mathbb{R}^{p \times p}$, respectively, and $\lambda_{\min}(\mathbf{M}) = 2$ following from Lemma 1.3; the last inequality follows from (3). Therefore, the optimization (10) is strictly convex, and thus (10) has at most one global minimizer.

The existence of minimizers of (10) can be guaranteed by the coercivity of $f(\mathbf{w})$. The function $f(\mathbf{w})$ can be lower bounded by

$$\begin{aligned} f(\mathbf{w}) &= -\log \left(\prod_{i=1}^p \lambda_i(\mathcal{L}\mathbf{w} + \mathbf{J}) \right) + \langle \mathcal{L}^* \mathbf{S} + \lambda \mathbf{1}, \mathbf{w} \rangle \\ &= -\log \left(\prod_{i=2}^p \lambda_i(\mathcal{L}\mathbf{w}) \right) + \langle \mathcal{L}^* \mathbf{S} + \lambda \mathbf{1}, \mathbf{w} \rangle \\ &\geq -(p-1) \log \left(\sum_{i=1}^p \lambda_i(\mathcal{L}\mathbf{w}) \right) + \langle \mathcal{L}^* \mathbf{S} + \lambda \mathbf{1}, \mathbf{w} \rangle + (p-1) \log(p-1) \end{aligned}$$

$$\begin{aligned}
&= -(p-1) \log \left(\sum_{i=1}^p [\mathcal{L}\mathbf{w}]_{ii} \right) + \langle \mathcal{L}^* \mathbf{S} + \lambda \mathbf{1}, \mathbf{w} \rangle + (p-1) \log(p-1) \\
&= -(p-1) \log \left(2 \sum_{t=1}^{p(p-1)/2} w_t \right) + \langle \mathcal{L}^* \mathbf{S} + \lambda \mathbf{1}, \mathbf{w} \rangle + (p-1) \log(p-1) \\
&\geq -(p-1) \log \left(\sum_{t=1}^{p(p-1)/2} w_t \right) + \lambda \sum_{t=1}^{p(p-1)/2} w_t + (p-1) \log \frac{p-1}{2}, \tag{11}
\end{aligned}$$

where the second equality follows from (34) with $b = 1$; the forth equality follows from $\mathcal{L}\mathbf{w} \cdot \mathbf{1} = \mathbf{0}$; the last inequality holds because $\mathbf{w} \geq \mathbf{0}$, and $\mathcal{L}^* \mathbf{S} \geq \mathbf{0}$, which follows from (70); the first inequality holds because the smallest eigenvalue $\lambda_1(\mathcal{L}\mathbf{w}) = 0$ and

$$\frac{a_1 + a_2 + \dots + a_n}{n} \geq \sqrt[p]{a_1 \cdot a_2 \cdots a_n}$$

holds for any non-negative real numbers of a_1, \dots, a_n . A function $g : \Omega \rightarrow \mathbb{R} \cup \{+\infty\}$ is called coercive over Ω , if every sequence $\mathbf{x}_k \in \Omega$ with $\|\mathbf{x}_k\| \rightarrow +\infty$ obeys $\lim_{k \rightarrow \infty} g(\mathbf{x}_k) = +\infty$, where $\Omega \subset \mathbb{R}^n$. Let

$$h(z) = -(p-1) \log z + \lambda z + (p-1) \log \frac{p-1}{2}.$$

A simple calculation yields $\lim_{z \rightarrow +\infty} h(z) = +\infty$ if $\lambda > 0$. For any sequence $\mathbf{w}_k \in \text{cl}(\mathcal{S}_{\mathbf{w}})$ with $\|\mathbf{w}_k\| \rightarrow +\infty$, where $\text{cl}(\mathcal{S}_{\mathbf{w}})$ is the closure of $\mathcal{S}_{\mathbf{w}}$, one has $\sum_{t=1}^{p(p-1)/2} [\mathbf{w}_k]_t \rightarrow +\infty$, because $\sum_{t=1}^{p(p-1)/2} [\mathbf{w}_k]_t \geq \|\mathbf{w}_k\|$. Then one obtains

$$\lim_{k \rightarrow \infty} f(\mathbf{w}_k) \geq \lim_{k \rightarrow \infty} h \left(\sum_{t=1}^{p(p-1)/2} [\mathbf{w}_k]_t \right) = \lim_{z \rightarrow +\infty} h(z) = +\infty,$$

where the first inequality follows from (11). Hence, $f(\mathbf{w})$ is coercive over $\text{cl}(\mathcal{S}_{\mathbf{w}})$. Following from the Extreme Value Theorem in [1], if $\Omega \subset \mathbb{R}^n$ is non-empty and closed, and $g : \Omega \rightarrow \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous and coercive, then the optimization $\min_{\mathbf{x} \in \Omega} g(\mathbf{x})$ has at least one global minimizer. Therefore, by the coercivity of $f(\mathbf{w})$, (10) has at least one global minimizer in $\text{cl}(\mathcal{S}_{\mathbf{w}})$.

Let $\Omega_A = \{\mathbf{w} \in \mathbb{R}^{p(p-1)/2} \mid \mathbf{w} \geq \mathbf{0}\}$ and $\Omega_B = \{\mathbf{w} \in \mathbb{R}^{p(p-1)/2} \mid (\mathcal{L}\mathbf{w} + \mathbf{J}) \in \mathcal{S}_{++}^p\}$. Ω_A is a closed set and Ω_B is an open set. Then $\mathcal{S}_{\mathbf{w}}$ can be rewritten as $\mathcal{S}_{\mathbf{w}} = \Omega_A \cap \Omega_B$. Consider the set $V := \text{cl}(\mathcal{S}_{\mathbf{w}}) \setminus \mathcal{S}_{\mathbf{w}}$, we have

$$V \subseteq \left\{ \text{cl}(\Omega_A) \cap \text{cl}(\Omega_B) \right\} \setminus \left\{ \Omega_A \cap \Omega_B \right\} = \Omega_A \cap \partial\Omega_B, \tag{12}$$

where $\partial\Omega_B$ is the boundary of Ω_B . Notice that every matrix on the boundary of the set of positive definite matrices is positive semi-definite and has zero determinant. Hence, one has $\partial\Omega_B = \{\mathbf{w} \in \mathbb{R}^{p(p-1)/2} \mid (\mathcal{L}\mathbf{w} + \mathbf{J}) \in \mathcal{S}_+^p, \det(\mathcal{L}\mathbf{w} + \mathbf{J}) = 0\}$. As a result, for any $\mathbf{w}_b \in \text{cl}(\mathcal{S}_{\mathbf{w}}) \setminus \mathcal{S}_{\mathbf{w}}$, $f(\mathbf{w}_b) = +\infty$. Therefore, (10) has at least one global minimizer and the minimizer must belong to the set $\mathcal{S}_{\mathbf{w}}$. On the other hand, by the strict convexity of $f(\mathbf{w})$, (10) has at most one global minimizer in $\mathcal{S}_{\mathbf{w}}$. Totally, we conclude that (10) has an unique global minimizer in $\mathcal{S}_{\mathbf{w}}$ if $\lambda > 0$.

We prove the theorem through the KKT conditions. The Lagrangian of the optimization (10) is

$$L(\mathbf{w}, \mathbf{v}) = -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J}) + \langle \mathcal{L}^* \mathbf{S} + \lambda \mathbf{1}, \mathbf{w} \rangle - \mathbf{v}^\top \mathbf{w},$$

where \mathbf{v} is a KKT multiplier. Let $(\hat{\mathbf{w}}, \hat{\mathbf{v}})$ be any pair of points that satisfies the KKT conditions. Then we have

$$-\mathcal{L}^* \left((\mathcal{L}\hat{\mathbf{w}} + \mathbf{J})^{-1} \right) + \mathcal{L}^* \mathbf{S} + \lambda \mathbf{1} - \hat{\mathbf{v}} = \mathbf{0}; \tag{13}$$

$$\hat{w}_i \hat{v}_i = 0, \text{ for } i = 1, \dots, p(p-1)/2; \tag{14}$$

$$\hat{\mathbf{w}} \geq \mathbf{0}, \hat{\mathbf{v}} \geq \mathbf{0}; \tag{15}$$

As we know, for any convex optimization with differentiable objective and constraint functions, any point that satisfies the KKT conditions (under Slater's constraint qualification) must be primal and dual optimal. Therefore, $\hat{\boldsymbol{w}}$ must obey $\hat{\boldsymbol{w}} = \arg \min_{\boldsymbol{w} \geq \mathbf{0}} f(\boldsymbol{w})$. Note that the pair of points $(\hat{\boldsymbol{w}}, \hat{\boldsymbol{v}})$ that satisfies the KKT conditions is unique. To prove the optimal solution $\hat{\boldsymbol{w}} > \mathbf{0}$ holds for (10), we can equivalently prove that the KKT conditions (13)-(15) hold for $(\hat{\boldsymbol{w}} > \mathbf{0}, \hat{\boldsymbol{v}} = \mathbf{0})$. It is further equivalent to prove that

$$\mathcal{L}^* ((\mathcal{L}\hat{\boldsymbol{w}} + \boldsymbol{J})^{-1}) = \mathcal{L}^* \boldsymbol{S} + \lambda \mathbf{1}. \quad (16)$$

holds for $\hat{\boldsymbol{w}} > \mathbf{0}$. Following from Lemma 1.2 with the fact that $\hat{\boldsymbol{w}} \in \mathcal{S}_w$, there must exist a unique \boldsymbol{x} such that

$$\mathcal{L}\boldsymbol{x} + \frac{1}{b}\boldsymbol{J} = (\mathcal{L}\hat{\boldsymbol{w}} + b\boldsymbol{J})^{-1} \quad (17)$$

holds for any $b \neq 0$. Thus one has

$$\mathcal{L}^* ((\mathcal{L}\hat{\boldsymbol{w}} + \boldsymbol{J})^{-1}) = \mathcal{L}^* (\mathcal{L}\boldsymbol{x} + \boldsymbol{J}) = \mathcal{L}^* \mathcal{L}\boldsymbol{x}, \quad (18)$$

where the first equality follows from (17) with $b = 1$; the second equality holds because $\boldsymbol{J} \in \mathcal{N}(\mathcal{L}^*)$ where $\mathcal{N}(\mathcal{L}^*)$ is the null space of \mathcal{L}^* defined by $\mathcal{N}(\mathcal{L}^*) := \{\boldsymbol{X} \in \mathbb{R}^{p \times p} \mid \mathcal{L}^* \boldsymbol{X} = \mathbf{0}\}$. Combining (16) and (18) yields

$$\boldsymbol{x} = (\mathcal{L}^* \mathcal{L})^{-1} (\mathcal{L}^* \boldsymbol{S} + \lambda \mathbf{1}), \quad (19)$$

where $\mathcal{L}^* \mathcal{L}$ is invertible according to Lemma 1.3. Recall that \boldsymbol{S} is the sample covariance matrix defined by

$$\boldsymbol{S} = \sum_{k=1}^n \boldsymbol{X}^{(k)} (\boldsymbol{X}^{(k)})^\top,$$

where $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ are the samples independently drawn from the LGMRF in Definition 2.1. According to the density function of LGMRF, we get $\mathbf{1}^\top \boldsymbol{X}^{(k)} = 0$ for $k = 1, \dots, n$. Therefore, \boldsymbol{S} is symmetric and obeys $\boldsymbol{S} \cdot \mathbf{1} = \mathbf{0}$. It is easy to verify that $\boldsymbol{S} \in \mathcal{R}(\mathcal{L})$, where $\mathcal{R}(\mathcal{L})$ is the range space of \mathcal{L} defined by $\mathcal{R}(\mathcal{L}) := \{\mathcal{L}\boldsymbol{y} \mid \boldsymbol{y} \in \mathbb{R}^{p(p-1)/2}\}$. Hence, there must exist a $\boldsymbol{y} \in \mathbb{R}^{p(p-1)/2}$ such that $\boldsymbol{S} = \mathcal{L}\boldsymbol{y}$. Thus $\mathcal{L}^* \boldsymbol{S} = \mathcal{L}^* \mathcal{L}\boldsymbol{y}$. One further obtains

$$\boldsymbol{y} = (\mathcal{L}^* \mathcal{L})^{-1} \mathcal{L}^* \boldsymbol{S}.$$

Then a simple calculation yields

$$\mathcal{L}(\mathcal{L}^* \mathcal{L})^{-1} \mathcal{L}^* \boldsymbol{S} = \mathcal{L}\boldsymbol{y} = \boldsymbol{S}. \quad (20)$$

Next, we construct a matrix $\boldsymbol{X} = \mathcal{L}\boldsymbol{x} + a\boldsymbol{J}$ with $a > 0$ and have

$$\begin{aligned} \mathcal{L}\boldsymbol{x} + a\boldsymbol{J} &= \mathcal{L}(\mathcal{L}^* \mathcal{L})^{-1} (\mathcal{L}^* \boldsymbol{S} + \lambda \mathbf{1}) + a\boldsymbol{J} \\ &= \boldsymbol{S} + \lambda \mathcal{L}(\mathcal{L}^* \mathcal{L})^{-1} \mathbf{1} + a\boldsymbol{J} \\ &= \boldsymbol{S} + \frac{\lambda}{2p} \mathcal{L} \mathbf{1} + a\boldsymbol{J}, \end{aligned} \quad (21)$$

where the first equality follows from (19); the second equality follows from (20); the third equality holds because $\mathcal{L}^* \mathcal{L} \mathbf{1} = 2p \mathbf{1}$ and then one has $(\mathcal{L}^* \mathcal{L})^{-1} \mathbf{1} = \frac{1}{2p} \mathbf{1}$.

Let $\hat{\boldsymbol{X}} = \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{X} \boldsymbol{D}^{-\frac{1}{2}}$ be the normalized matrix of \boldsymbol{X} , where \boldsymbol{D} is a diagonal matrix containing the diagonals of \boldsymbol{X} . Notice that each diagonal element of $\hat{\boldsymbol{X}}$ is 1. Next, we will prove that, under some conditions, $\hat{\boldsymbol{X}}$ is an inverse M -matrix, that is, $(\hat{\boldsymbol{X}})^{-1}$ is a M -matrix. We say $\boldsymbol{A} \in \mathbb{R}^{p \times p}$ is an M -matrix if

$$\boldsymbol{A} = s\boldsymbol{I} - \boldsymbol{B}, \quad (22)$$

where $\mathbf{B} \in \mathbb{R}^{p \times p}$ is an element-wise non-negative matrix and $s > \rho(\mathbf{B})$, the spectral radius of \mathbf{B} . According to (21), one has

$$X_{ij} = S_{ij} - \frac{\lambda}{2p} + \frac{a}{p}, \quad \text{for } i \neq j,$$

and

$$X_{ii} = S_{ii} + \frac{p-1}{2p}\lambda + \frac{a}{p}, \quad i = 1, \dots, p.$$

Define $\tilde{S}_{ij} = \max_{i \neq j} S_{ij}$, $\bar{S}_{ij} = \min_{i \neq j} S_{ij}$, $\tilde{S}_{kk} = \max_k S_{kk}$ and $\bar{S}_{kk} = \min_k S_{kk}$. By the definition of $\hat{\mathbf{X}}$, the lower bound y and upper bound x of the elements off the diagonal of $\hat{\mathbf{X}}$ can be obtained as below,

$$\begin{aligned} \hat{X}_{ij} &= \frac{X_{ij}}{\sqrt{X_{ii}X_{jj}}} = \frac{S_{ij} - \frac{\lambda}{2p} + \frac{a}{p}}{\sqrt{S_{ii} + \frac{p-1}{2p}\lambda + \frac{a}{p}} \cdot \sqrt{S_{jj} + \frac{p-1}{2p}\lambda + \frac{a}{p}}} \\ &\geq \frac{\tilde{S}_{ij} - \frac{\lambda}{2p} + \frac{a}{p}}{\tilde{S}_{kk} + \frac{p-1}{2p}\lambda + \frac{a}{p}} =: y, \quad \forall i \neq j. \end{aligned} \quad (23)$$

and

$$\hat{X}_{ij} \leq \frac{\tilde{S}_{ij} - \frac{\lambda}{2p} + \frac{a}{p}}{\bar{S}_{kk} + \frac{p-1}{2p}\lambda + \frac{a}{p}} =: x, \quad \forall i \neq j. \quad (24)$$

Define s by $x^2 = sy + (1-s)y^2$. According to Lemma 1.10, if $0 < y \leq x < 1$ and $s^{-1} \geq p-2$ with $p > 3$, then $\hat{\mathbf{X}}$ is an inverse M -matrix. We can see, provided that $y = \frac{1}{p+1}$ and the inequalities

$$0 < y \leq x \leq \sqrt{2}y < 1, \quad (25)$$

hold, then one has

$$s^{-1} = \frac{y - y^2}{x^2 - y^2} \geq \frac{1 - y}{y} > p - 2,$$

where the first inequality follows from $x \leq \sqrt{2}y$. Therefore, if $y = \frac{1}{p+1}$ and (25) holds, then $\hat{\mathbf{X}}$ is an inverse M -matrix.

Next, we prove that if $a = \tilde{S}_{kk} - (p+1)\bar{S}_{ij} + \lambda$ with $\lambda \geq 2(\sqrt{2}+1)(p+1)(\tilde{S}_{kk} - \bar{S}_{ij})$, then $y = \frac{1}{p+1}$ and (25) holds. Substituting $a = \tilde{S}_{kk} - (p+1)\bar{S}_{ij} + \lambda$ into (23) yields $y = \frac{1}{p+1}$. Then it is clear that $y > 0$ and $\sqrt{2}y < 1$. Comparing x and y defined in (23) and (24), respectively, one has $y \leq x$. A simple algebra yields

$$\begin{aligned} x &= \frac{\tilde{S}_{ij} - \frac{\lambda}{2p} + \frac{a}{p}}{\bar{S}_{kk} + \frac{p-1}{2p}\lambda + \frac{a}{p}} \leq \frac{p\tilde{S}_{kk} - \frac{\lambda}{2} + a}{p\bar{S}_{ij} + \frac{p-1}{2}\lambda + a} \\ &= \frac{(p+1)(\tilde{S}_{kk} - \bar{S}_{ij}) + \frac{\lambda}{2}}{\tilde{S}_{kk} - \bar{S}_{ij} + \frac{(p+1)\lambda}{2}} \leq \frac{\sqrt{2}}{p+1} = \sqrt{2}y, \end{aligned}$$

where the first inequality follows from $\tilde{S}_{ij} \leq \tilde{S}_{kk}$ and $\bar{S}_{kk} \geq \bar{S}_{ij}$ because $\mathbf{S} \in \mathcal{S}_+^p$ and $\mathbf{S} \cdot \mathbf{1} = \mathbf{0}$. It is easy to verify that $\lambda \geq 2(\sqrt{2}+1)(p+1)(\tilde{S}_{kk} - \bar{S}_{ij})$ is large enough to establish the second inequality. Therefore, all the inequalities in (25) hold.

Consequently, by Lemma 1.10, we conclude that $\hat{\mathbf{X}}$ is an inverse M -matrix when $a = \tilde{S}_{kk} - (p+1)\bar{S}_{ij} + \lambda$ with $\lambda \geq 2(\sqrt{2}+1)(p+1)(\tilde{S}_{kk} - \bar{S}_{ij})$. Therefore, $\hat{\mathbf{X}}^{-1} = \mathbf{D}^{\frac{1}{2}}\mathbf{X}^{-1}\mathbf{D}^{\frac{1}{2}}$ is an M -matrix. Notice that the

elements off the diagonal of an M -matrix are non-positive according to (22). As a result, the elements off the diagonal of $\mathbf{D}^{\frac{1}{2}}\mathbf{X}^{-1}\mathbf{D}^{\frac{1}{2}}$ are non-positive, implying that the elements off the diagonal of \mathbf{X}^{-1} are also non-positive, because $\mathbf{X} = \mathcal{L}\mathbf{x} + a\mathbf{J}$ is positive definite and thus the diagonal elements of \mathbf{D} are positive. The application of (17) with $b = \frac{1}{a}$ yields

$$\mathbf{X}^{-1} = [\mathcal{L}\mathbf{x} + a\mathbf{J}]^{-1} = \mathcal{L}\hat{\mathbf{w}} + \frac{1}{a}\mathbf{J}.$$

One further obtains

$$[\mathcal{L}\hat{\mathbf{w}} + \frac{1}{a}\mathbf{J}]_{ij} = -\hat{w}_k + \frac{1}{ap} \leq 0, \quad \forall i \neq j,$$

where $k = i - j + \frac{i-1}{2}(2p - j)$. Therefore, we establish

$$\hat{w}_k \geq \frac{1}{ap} = \frac{1}{(\tilde{S}_{kk} - (p+1)\tilde{S}_{ij} + \lambda)p} > 0, \quad \forall k,$$

concluding that (16) holds for $\hat{\mathbf{w}} > \mathbf{0}$. Note that $a = \tilde{S}_{kk} - (p+1)\tilde{S}_{ij} + \lambda > 0$ because $\tilde{S}_{kk} > 0, \lambda > 0$, and $\tilde{S}_{ij} \leq 0$, where $\tilde{S}_{ij} \leq 0$ following from $\mathbf{S} \cdot \mathbf{1} = \mathbf{0}$ and each diagonal element $S_{ii} \geq 0$ since \mathbf{S} is positive semi-definite. \square

1.3 Proofs of Theorem 3.4

Proof. Let $\mathbf{x} \in \mathbb{R}^{p(p-1)/2}$. According to the definition of \mathcal{L} , $\mathcal{L}\mathbf{x}$ must obey $[\mathcal{L}\mathbf{x}]_{ij} = [\mathcal{L}\mathbf{x}]_{ji}$, for any $i \neq j$ and $(\mathcal{L}\mathbf{x}) \cdot \mathbf{1} = \mathbf{0}$.

Next, we will show that $\mathcal{L}\mathbf{x}$ is positive semi-definite for any $\mathbf{x} \geq \mathbf{0}$ by the Gershgorin circle theorem [5]. Given a matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ with entries X_{ij} . Let $R_i(\mathbf{X}) = \sum_{j \neq i} |X_{ij}|$ be the sum of the absolute values of the non-diagonal entries in the i -th row. Then a Gershgorin disc is the disc $D(X_{ii}, R_i(\mathbf{X}))$ centered at X_{ii} on the complex plane with radius $R_i(\mathbf{X})$. Gershgorin circle theorem [3] shows that each eigenvalue of \mathbf{X} lies within at least one of the Gershgorin discs. For any $\mathbf{x} \geq \mathbf{0}$, $R_i(\mathcal{L}\mathbf{x}) = [\mathcal{L}\mathbf{x}]_{ii}$ holds for each i because $(\mathcal{L}\mathbf{x}) \cdot \mathbf{1} = \mathbf{0}$ and $[\mathcal{L}\mathbf{x}]_{ij} \leq 0$ for any $i \neq j$. For any given eigenvalue λ of $\mathcal{L}\mathbf{x}$, by Gershgorin circle theorem, there must exist one Gershgorin disc $D([\mathcal{L}\mathbf{x}]_{ii}, R_i(\mathcal{L}\mathbf{x}))$ such that

$$|\lambda - [\mathcal{L}\mathbf{x}]_{ii}| \leq R_i(\mathcal{L}\mathbf{x}) = [\mathcal{L}\mathbf{x}]_{ii}, \quad (26)$$

indicating that $\lambda \geq 0$. Note that the eigenvalues of $\mathcal{L}\mathbf{x}$ are real since $\mathcal{L}\mathbf{x}$ is symmetric. Therefore, one has

$$\mathcal{L}\mathbf{x} \in \mathcal{S}_+^p, \quad \forall \mathbf{x} \geq \mathbf{0}. \quad (27)$$

Finally, we will prove that $\text{rank}(\mathcal{L}\mathbf{x}) = p - 1 \Leftrightarrow (\mathcal{L}\mathbf{x} + \mathbf{J}) \in \mathcal{S}_{++}^p$, for any $\mathbf{x} \geq \mathbf{0}$. On one hand, if $\text{rank}(\mathcal{L}\mathbf{x}) = p - 1$, then $\mathcal{L}\mathbf{x} + \mathbf{J}$ admits the eigenvalue decomposition $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where $\mathbf{U} = [\mathbf{U}_s \ \frac{1}{\sqrt{p}}\mathbf{1}]$ and $\mathbf{\Lambda}$ is a diagonal matrix with the diagonal elements $[\lambda_2, \dots, \lambda_p, 1]$. Here $\lambda_{i=2}^p$ are the nonzero eigenvalues of $\mathcal{L}\mathbf{x}$ and \mathbf{U}_s is a $p \times (p - 1)$ matrix whose columns are the corresponding eigenvectors of $\mathcal{L}\mathbf{x}$. Note that the nonzero eigenvalue $\lambda_{i=2}^p > 0$ because $\mathcal{L}\mathbf{x} \in \mathcal{S}_+^p$. Therefore, one has $(\mathcal{L}\mathbf{x} + \mathbf{J}) \in \mathcal{S}_{++}^p$. On the other hand, if $(\mathcal{L}\mathbf{x} + \mathbf{J}) \in \mathcal{S}_{++}^p$, then $\text{rank}(\mathcal{L}\mathbf{x}) \geq \text{rank}(\mathcal{L}\mathbf{x} + \mathbf{J}) - \text{rank}(\mathbf{J}) = p - 1$ because $\mathcal{L}\mathbf{x} + \mathbf{J}$ is full rank and \mathbf{J} is rank one. Furthermore, $\text{rank}(\mathcal{L}\mathbf{x}) \leq p - 1$ because $(\mathcal{L}\mathbf{x}) \cdot \mathbf{1} = \mathbf{0}$. Therefore, we conclude that $\text{rank}(\mathcal{L}\mathbf{x}) = p - 1$, completing the proof. \square

1.4 Proofs of Theorem 3.8 and Corollary 3.9

Proof. We first prove Theorem 3.8. Take the regularization parameter $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ for some $\alpha > 2$, and the sample size

$$n \geq \max(94\alpha c_0^{-1} \lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)s \log p, 8\alpha \log p), \quad (28)$$

where c_0 is a constant defined in Lemma 1.8. Notice that the sample size n in (28) satisfies the conditions on the number of samples in Lemmas 1.4, 1.5, 1.6 and 1.7. We choose the initial point $\hat{\mathbf{w}}^{(0)}$ of Algorithm 1 satisfying $|\text{supp}^+(\hat{\mathbf{w}}^{(0)})| \leq s$.

Define an event $\mathcal{J} = \{\|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})\|_{\max} \leq \lambda/2\}$. According to Lemma 1.7, $\|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})\|_{\max} \leq \lambda/2$ holds with probability at least $1 - 1/p^{\alpha-2}$. Under the event \mathcal{J} , one applies Lemma 1.6 and obtains, for any $k \geq 1$,

$$\|\hat{\mathbf{w}}^{(k)} - \mathbf{w}^*\| \leq 2\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| + \frac{3}{2 + \sqrt{2}}\|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\|,$$

By induction, if $d_k \leq a_0 + \rho d_{k-1}$ for any $k \geq 1$ with $\rho \in [0, 1)$, then

$$d_k \leq \frac{1 - \rho^k}{1 - \rho}a_0 + \rho^k d_0. \quad (29)$$

Taking $a_0 = 2\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\|$, $\rho = \frac{3}{2 + \sqrt{2}}$ and $d_k = \|\hat{\mathbf{w}}^{(k)} - \mathbf{w}^*\|$, one obtains

$$\|\hat{\mathbf{w}}^{(k)} - \mathbf{w}^*\| \leq 2\sqrt{2}(\sqrt{2} + 1)^2\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| + \left(\frac{3}{2 + \sqrt{2}}\right)^k \|\hat{\mathbf{w}}^{(0)} - \mathbf{w}^*\|.$$

Under the event \mathcal{J} , $\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\|$ can be bounded by

$$\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| \leq \sqrt{s}\lambda/2 \leq \sqrt{\alpha c_0^{-1} s \log p/n}. \quad (30)$$

Therefore, under the event \mathcal{J} , which holds with probability at least $1 - 1/p^{\alpha-2}$, one has

$$\|\hat{\mathbf{w}}^{(k)} - \mathbf{w}^*\| \leq 2(3\sqrt{2} + 4)\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\sqrt{\alpha c_0^{-1} s \log p/n} + \left(\frac{3}{2 + \sqrt{2}}\right)^k \|\hat{\mathbf{w}}^{(0)} - \mathbf{w}^*\|.$$

Next, we prove Corollary 3.9. Under the event \mathcal{J} , one applies Lemma 1.6 and obtains,

$$\|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq 2\sqrt{2}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| + \frac{3}{2 + \sqrt{2}}\|\mathcal{L}\hat{\mathbf{w}}^{(k-1)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \quad (31)$$

holds for any $k \geq 1$. Taking $a_0 = 2\sqrt{2}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\|$, $\rho = \frac{3}{2 + \sqrt{2}}$ and $d_k = \|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}}$, by (29) one has

$$\|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq 4(\sqrt{2} + 1)^2\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| + \left(\frac{3}{2 + \sqrt{2}}\right)^k \|\mathcal{L}\hat{\mathbf{w}}^{(0)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}}.$$

Similarly, according to (30), one obtains

$$\|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq 4(3 + 2\sqrt{2})\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\sqrt{\alpha c_0^{-1} s \log p/n} + \left(\frac{3}{2 + \sqrt{2}}\right)^k \|\mathcal{L}\hat{\mathbf{w}}^{(0)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}}$$

holds at least $1 - 1/p^{\alpha-2}$.

Alternative to (29), one obtains

$$d_k \leq \frac{1 - \rho^{k-1}}{1 - \rho}a_0 + \rho^{k-1}d_1,$$

and correspondingly establishes

$$\|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq 4(3 + 2\sqrt{2})\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\sqrt{\alpha c_0^{-1} s \log s/n} + \left(\frac{3}{2 + \sqrt{2}}\right)^{k-1} \|\mathcal{L}\hat{\mathbf{w}}^{(1)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}}. \quad (32)$$

To apply Lemma 1.4, we first check the necessary conditions of the lemma. Let $\mathbf{z}^{(0)}$ satisfy $z_i^{(0)} = h'_\lambda(\hat{w}_i^{(0)})$, $i \in [p(p-1)/2]$. Notice that $z_i^{(0)} \in [0, \lambda]$ for $i \in [p(p-1)/2]$ by Assumption 3.5. According to (7), $\mathcal{E}^{(1)} = \{\mathcal{S}^* \cup \mathcal{S}^{(1)}\}$, where $\mathcal{S}^{(1)} = \{i \in [p(p-1)/2] \mid w_i^{(0)} \geq b\}$. For any $i \in \{\mathcal{S}^{(1)}\}^c$, one has

$$z_i^{(0)} = h'_\lambda(\hat{w}_i^{(0)}) \geq h'_\lambda(b) \geq \frac{\lambda}{2},$$

where the first inequality holds because $\hat{w}_i^{(0)} < b$ for any $i \in \{\mathcal{S}^{(1)}\}^c$, and h'_λ is non-increasing according to Assumption 3.5; the second inequality directly follows from Assumption 3.5. Hence one has

$$\left\| \mathbf{z}_{\{\mathcal{E}^{(1)}\}^c}^{(0)} \right\|_{\min} \geq \left\| \mathbf{z}_{\{\mathcal{S}^{(1)}\}^c}^{(0)} \right\|_{\min} \geq \lambda/2.$$

One also obtains $|\mathcal{E}^{(1)}| < 2s$ by Lemma 1.5, and $\mathcal{S}^* \subseteq \mathcal{E}^{(1)}$ by the definition of $\mathcal{E}^{(1)}$. Therefore, one can apply Lemma 1.4 with $\mathcal{E} = \mathcal{E}^{(1)}$ and $\mathbf{z} = \mathbf{z}^{(0)}$ and obtains

$$\|\mathcal{L}\hat{\mathbf{w}}^{(1)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq 2(1 + \sqrt{2})\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\sqrt{s}\lambda.$$

If $t \geq \log_{\frac{2+\sqrt{2}}{3}}(\lambda\sqrt{n/\log p}) = \log\left(\sqrt{4\alpha c_0^{-1}}\right)/\log\frac{2+\sqrt{2}}{3}$, a simple algebra yields

$$\left(\frac{3}{2+\sqrt{2}}\right)^{t-1} \|\mathcal{L}\hat{\mathbf{w}}^{(1)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq \frac{2(3\sqrt{2}+4)}{3}(\lambda\sqrt{n/\log p})^{-1}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\sqrt{s}\lambda \lesssim \sqrt{s \log p/n}. \quad (33)$$

Taking $k \geq \lceil 4 \log(4\alpha c_0^{-1}) \rceil \geq \log\left(\sqrt{4\alpha c_0^{-1}}\right)/\log\frac{2+\sqrt{2}}{3}$, and combining (32) and (33) together, we can conclude that

$$\|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \lesssim \sqrt{s \log p/n},$$

completing the proof. \square

2 Proof of Technical Lemmas

This section contains the proofs of technical lemmas used in Section 1.

2.1 Proof of Lemma 1.1

Proof. The gradient of $f(\mathbf{w})$ is $\nabla f(\mathbf{w}) = -\mathcal{L}^*(\mathcal{L}\mathbf{w} + \mathbf{J})^{-1}$ and its Hessian matrix is $\nabla^2 f(\mathbf{w})$ with the k -th column being

$$\begin{aligned} [\nabla^2 f(\mathbf{w})]_{:,k} &= \frac{\partial(\nabla f(\mathbf{w}))}{\partial w_k} = -\mathcal{L}^* \left(\frac{\partial(\mathcal{L}\mathbf{w} + \mathbf{J})^{-1}}{\partial w_k} \right) \\ &= \mathcal{L}^* \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \frac{\partial(\mathcal{L}\mathbf{w} + \mathbf{J})}{\partial w_k} (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right) \\ &= \mathcal{L}^* \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \mathbf{A}_k (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right), \end{aligned}$$

where $\mathbf{A}_k \in \mathbb{R}^{p \times p}$ is a matrix with $[\mathbf{A}_k]_{ii} = [\mathbf{A}_k]_{jj} = 1$, $[\mathbf{A}_k]_{ij} = [\mathbf{A}_k]_{ji} = -1$ and zeros for the other elements, in which $i, j \in \mathbb{Z}^+$ obeying $k = i - j + \frac{i-1}{2}(2p-j)$ and $i > j$. Therefore, $\nabla^2 f(\mathbf{w})$ can be written as

$$\nabla^2 f(\mathbf{w}) = [\mathcal{L}^* \mathbf{B}_1, \mathcal{L}^* \mathbf{B}_2, \dots, \mathcal{L}^* \mathbf{B}_{p(p-1)/2}],$$

where $\mathbf{B}_k = (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1}\mathbf{A}_k(\mathcal{L}\mathbf{w} + \mathbf{J})^{-1}$, for $k = 1, 2, \dots, p(p-1)/2$. Then one has

$$\begin{aligned}
\mathbf{x}^\top \nabla^2 f(\mathbf{w}) \mathbf{x} &= \mathbf{x}^\top [\mathcal{L}^* \mathbf{B}_1, \mathcal{L}^* \mathbf{B}_2, \dots, \mathcal{L}^* \mathbf{B}_{p(p-1)/2}] \mathbf{x} \\
&= \mathbf{x}^\top \mathcal{L}^* \left(\sum_{k=1}^{p(p-1)/2} x_k \mathbf{B}_k \right) \\
&= \mathbf{x}^\top \mathcal{L}^* \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \left(\sum_{k=1}^{p(p-1)/2} x_k \mathbf{A}_k \right) (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right) \\
&= \mathbf{x}^\top \mathcal{L}^* \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \mathcal{L} \mathbf{x} (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right) \\
&= \langle \mathcal{L} \mathbf{x}, (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \mathcal{L} \mathbf{x} (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \rangle \\
&= \text{vec}(\mathcal{L} \mathbf{x})^\top \text{vec} \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \mathcal{L} \mathbf{x} (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right) \\
&= \text{vec}(\mathcal{L} \mathbf{x})^\top \left((\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \otimes (\mathcal{L}\mathbf{w} + \mathbf{J})^{-1} \right) \text{vec}(\mathcal{L} \mathbf{x}),
\end{aligned}$$

where the fourth equality follows from the definition of \mathcal{L} , and the last equality follows from the property of Kronecker product that $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{B})$. \square

2.2 Proof of Lemma 1.2

Proof. Let $\mathbf{X} = \mathcal{L}\mathbf{w} + b\mathbf{J}$ with any $b \neq 0$ and any given $\mathbf{w} \in \mathbb{R}^{p(p-1)/2}$ obeying $(\mathcal{L}\mathbf{w} + \mathbf{J}) \in \mathcal{S}_{++}^p$. It is easy to verify that the column spaces as well as row spaces of $\mathcal{L}\mathbf{w}$ and \mathbf{J} are orthogonal with each other. Hence \mathbf{X} admits the eigenvalue decomposition

$$\mathbf{X} = \mathcal{L}\mathbf{w} + b\mathbf{J} = \begin{bmatrix} \mathbf{U} & \mathbf{u} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & b \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{u} \end{bmatrix}^\top, \quad (34)$$

where $\mathcal{L}\mathbf{w} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ and $b\mathbf{J} = b\mathbf{u}\mathbf{u}^\top$ with $\mathbf{u} = \frac{1}{\sqrt{p}}\mathbf{1}_p$, in which $\mathbf{1}_p \in \mathbb{R}^p$ with each element equal to 1. Notice that $\mathbf{\Lambda}$ is non-singular. \mathbf{X}^{-1} admits the eigenvalue decomposition

$$\mathbf{X}^{-1} = \begin{bmatrix} \mathbf{U} & \mathbf{u} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{b} \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{u} \end{bmatrix}^\top = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^\top + \frac{1}{b}\mathbf{J},$$

It is easy to check that $\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^\top$ is symmetric and $\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^\top \cdot \mathbf{1}_p = \mathbf{0}$. Therefore, there must exist a \mathbf{x} such that $\mathcal{L}\mathbf{x} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^\top$. One further obtains $x_k = [\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^\top]_{ij}$, for $k = 1, \dots, p(p-1)/2$, where $i, j \in \mathbb{Z}^+$ satisfying $k = i - j + \frac{i-1}{2}(2p-j)$ and $i > j$. Hence such \mathbf{x} is fixed and unique for a given \mathbf{w} . Note that \mathbf{x} is independent of b , and thus $\mathcal{L}\mathbf{x} + \frac{1}{b}\mathbf{J} = (\mathcal{L}\mathbf{w} + b\mathbf{J})^{-1}$ holds for any $b \neq 0$, completing the proof. \square

2.3 Proof of Lemma 1.3

Proof. Define an index set Ω_t by

$$\Omega_t := \left\{ l \in [p(p-1)/2] \mid [\mathcal{L}\mathbf{x}]_{tl} = \sum_l x_l \right\}, \quad t \in [p]. \quad (35)$$

According to the definition of \mathcal{L} , for any $\mathbf{x} \in \mathbb{R}^{p(p-1)/2}$, one obtains that $\mathcal{L}\mathbf{x} \in \mathbb{R}^{p \times p}$ obeys

$$[\mathcal{L}\mathbf{x}]_{ij} = \begin{cases} -x_k & i > j, \\ [\mathcal{L}\mathbf{x}]_{ji} & i < j, \\ \sum_{l \in \Omega_i} x_l & i = j, \end{cases} \quad (36)$$

where $k = i - j + \frac{i-1}{2}(2p - j)$. By the definition of \mathcal{L}^* , one further obtains that $\mathcal{L}^* \mathcal{L} \mathbf{x} \in \mathbb{R}^{p(p-1)/2}$ satisfies

$$[\mathcal{L}^* \mathcal{L} \mathbf{x}]_k = [\mathcal{L} \mathbf{x}]_{ii} + [\mathcal{L} \mathbf{x}]_{jj} + 2x_k = \sum_{l \in \Omega_i \cup \Omega_j} x_l + 2x_k = \sum_{l \in \Omega_i \cup \Omega_j \setminus k} x_l + 4x_k, \quad (37)$$

where $i, j \in [p]$ satisfying $k = i - j + \frac{i-1}{2}(2p - j)$ and $i > j$. The last equality holds because

$$\sum_{l \in \Omega_i \cup \Omega_j} x_l = [\mathcal{L} \mathbf{x}]_{ii} + [\mathcal{L} \mathbf{x}]_{jj} = - \sum_{m \neq i} [\mathcal{L} \mathbf{x}]_{im} - \sum_{m \neq j} [\mathcal{L} \mathbf{x}]_{jm} = 2x_k - \sum_{m \neq i, j} [\mathcal{L} \mathbf{x}]_{im} - \sum_{m \neq i, j} [\mathcal{L} \mathbf{x}]_{jm},$$

where the last equality follows from $[\mathcal{L} \mathbf{x}]_{ij} = -x_k$ according to (36).

According to (37), we conclude that there exists a matrix $\mathbf{M} \in \mathbb{R}^{\frac{p(p-1)}{2} \times \frac{p(p-1)}{2}}$ such that $\mathcal{L}^* \mathcal{L} \mathbf{x} = \mathbf{M} \mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^{p(p-1)/2}$, and \mathbf{M} obeys

$$M_{kl} = \begin{cases} 4 & l = k, \\ 1 & l \in (\Omega_i \cup \Omega_j) \setminus k, \\ 0 & \text{Otherwise,} \end{cases} \quad (38)$$

where $i, j \in [p]$ satisfying $k = i - j + \frac{i-1}{2}(2p - j)$ and $i > j$. Note that we use the fact that $\{\Omega_i \setminus k\} \cap \{\Omega_j \setminus k\} = \emptyset$ with $k = i - j + \frac{i-1}{2}(2p - j)$.

Finally, we will compute the minimum and maximum eigenvalues of \mathbf{M} . To compute the minimum eigenvalue of \mathbf{M} , one has

$$\lambda_{\min}(\mathbf{M}) = \inf_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^\top \mathbf{M} \mathbf{x}}{\|\mathbf{x}\|^2} = \inf_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathcal{L} \mathbf{x}\|_F^2}{\|\mathbf{x}\|^2} = \inf_{\mathbf{x} \neq \mathbf{0}} \frac{2 \sum_{k=1}^{p(p-1)/2} x_k^2 + \sum_{i=1}^p ([\mathcal{L} \mathbf{x}]_{ii})^2}{\|\mathbf{x}\|^2} \geq 2.$$

with equality when $[\mathcal{L} \mathbf{x}]_{11} = \dots = [\mathcal{L} \mathbf{x}]_{pp} = 0$, which can be written as $\mathbf{Q} \mathbf{x} = \mathbf{0}$ with $\mathbf{Q} \in \mathbb{R}^{p \times \frac{p(p-1)}{2}}$. Obviously, there must exist a nonzero solution to $\mathbf{Q} \mathbf{x} = \mathbf{0}$, and thus $\lambda_{\min}(\mathbf{M}) = 2$. To compute the maximum eigenvalue of \mathbf{M} , one has

$$\begin{aligned} \lambda_{\max}(\mathbf{M}) &= \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^\top \mathbf{M} \mathbf{x}}{\|\mathbf{x}\|^2} = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{2 \sum_{k=1}^{p(p-1)/2} x_k^2 + \sum_{t=1}^p ([\mathcal{L} \mathbf{x}]_{tt})^2}{\|\mathbf{x}\|^2} \\ &= \sup_{\mathbf{x} \neq \mathbf{0}} \frac{4 \sum_{k=1}^{p(p-1)/2} x_k^2 + \sum_{t=1}^p \sum_{i, j \in \Omega_t, i \neq j} x_i x_j}{\|\mathbf{x}\|^2} \\ &\leq \sup_{\mathbf{x} \neq \mathbf{0}} \frac{4 \sum_{k=1}^{p(p-1)/2} x_k^2 + \frac{1}{2} \sum_{t=1}^p \sum_{i, j \in \Omega_t, i \neq j} (x_i^2 + x_j^2)}{\|\mathbf{x}\|^2} \\ &= \sup_{\mathbf{x} \neq \mathbf{0}} \frac{4 \sum_{k=1}^{p(p-1)/2} x_k^2 + \sum_{t=1}^p (|\Omega_t| - 1) \sum_{i \in \Omega_t} x_i^2}{\|\mathbf{x}\|^2} \\ &= \sup_{\mathbf{x} \neq \mathbf{0}} \frac{4 \sum_{k=1}^{p(p-1)/2} x_k^2 + (p-2) \sum_{t=1}^p \sum_{i \in \Omega_t} x_i^2}{\|\mathbf{x}\|^2} \\ &= 2p, \end{aligned}$$

with equality when each element of \mathbf{x} is equal with each other, and thus $\lambda_{\max}(\mathbf{M}) = 2p$. The last second equality is obtained by plugging $|\Omega_t| = p - 1$ with $t \in [p]$, which is easy to verify according to the definition of Ω_t in (35); the last equality follows from $\sum_{t=1}^p \sum_{i \in \Omega_t} x_i^2 = 2 \sum_{k=1}^{p(p-1)/2} x_k^2$, because for any $k \in [p(p-1)/2]$, $k \in \Omega_t$ only holds with $t = \{i, j\}$, where $i, j \in [p]$ obeying $k = i - j + \frac{i-1}{2}(2p - j)$ and $i > j$. \square

2.4 Proof of Lemma 1.4

Proof. Take $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ and $n \geq 94\alpha c_0^{-1} \lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) s \log p$. Define a local region

$$\mathcal{B}_M(\mathbf{w}^*; \lambda_{\max}(\mathcal{L}\mathbf{w}^*)) = \{\mathbf{w} \mid \mathbf{w} \in \mathbb{B}_M(\mathbf{w}^*; \lambda_{\max}(\mathcal{L}\mathbf{w}^*)) \cap \mathcal{S}_w\},$$

where $\mathbb{B}_M(\mathbf{w}^*; r) = \{\mathbf{w} \in \mathbb{R}^{p(p-1)/2} \mid \|\mathbf{w} - \mathbf{w}^*\|_M \leq r\}$, in which $\|\mathbf{x}\|_M^2 = \langle \mathbf{x}, \mathbf{M}\mathbf{x} \rangle = \|\mathcal{L}\mathbf{x}\|_F^2$ with $\mathbf{M} \succ \mathbf{0}$ defined in Lemma 1.3, and $\mathcal{S}_w = \{\mathbf{w} \in \mathbb{R}^{p(p-1)/2} \mid \mathbf{w} \geq \mathbf{0}, (\mathcal{L}\mathbf{w} + \mathbf{J}) \in \mathcal{S}_{++}^p\}$. It is easy to check that $\mathbf{w}^* \in \mathcal{B}_M(\mathbf{w}^*; \lambda_{\max}(\mathcal{L}\mathbf{w}^*))$.

Recall that $\hat{\mathbf{w}}$ minimizes the optimization

$$\min_{\mathbf{w} \geq \mathbf{0}} -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J}) + \text{tr}(\mathcal{L}\mathbf{w}\mathbf{S}) + \mathbf{z}^\top \mathbf{w}, \quad (39)$$

where $0 \leq z_i \leq \lambda$ for $i \in [p(p-1)/2]$. We can see the optimization problems (1) and (39) have the same feasible set. Therefore, \mathcal{S}_w is also the feasible set of (39) and thus $\hat{\mathbf{w}} \in \mathcal{S}_w$ must hold.

Next, we will prove that $\hat{\mathbf{w}} \in \mathcal{B}_M(\mathbf{w}^*; \lambda_{\max}(\mathcal{L}\mathbf{w}^*))$. We first construct an intermediate estimator,

$$\mathbf{w}_t = \mathbf{w}^* + t(\hat{\mathbf{w}} - \mathbf{w}^*), \quad (40)$$

where t is taken such that $\|\mathbf{w}_t - \mathbf{w}^*\|_M = \lambda_{\max}(\mathcal{L}\mathbf{w}^*)$ if $\|\hat{\mathbf{w}} - \mathbf{w}^*\|_M > \lambda_{\max}(\mathcal{L}\mathbf{w}^*)$, and $t = 1$ otherwise. Hence $\|\mathbf{w}_t - \mathbf{w}^*\|_M \leq \lambda_{\max}(\mathcal{L}\mathbf{w}^*)$ always holds and $t \in [0, 1]$. One further has $\mathbf{w}_t \in \mathcal{S}_w$ because both $\mathbf{w}^*, \hat{\mathbf{w}} \in \mathcal{S}_w$ and \mathcal{S}_w is a convex set as shown in (4). Therefore, we conclude that $\mathbf{w}_t \in \mathcal{B}_M(\mathbf{w}^*; \lambda_{\max}(\mathcal{L}\mathbf{w}^*))$. Applying Lemma 1.9 with $\mathbf{w}_1 = \mathbf{w}_t$, $\mathbf{w}_2 = \mathbf{w}^*$ and $r = \lambda_{\max}(\mathcal{L}\mathbf{w}^*)$ yields

$$t \langle -\mathcal{L}^*(\mathcal{L}\mathbf{w}_t + \mathbf{J})^{-1} + \mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle \geq (2\lambda_{\max}(\mathcal{L}\mathbf{w}^*))^{-2} \|\mathcal{L}\mathbf{w}_t - \mathcal{L}\mathbf{w}^*\|_F^2. \quad (41)$$

Let $q(a) = -\log \det(\mathcal{L}(\mathbf{w}^* + a(\hat{\mathbf{w}} - \mathbf{w}^*)) + \mathbf{J}) + a \langle \mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle$ and $a \in [0, 1]$. One has

$$q'(a) = \langle -\mathcal{L}^*(\mathcal{L}\mathbf{w}_a + \mathbf{J})^{-1} + \mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle, \quad (42)$$

and

$$q''(a) = \left\langle \mathcal{L}^* \left((\mathcal{L}\mathbf{w}_a + \mathbf{J})^{-1} (\mathcal{L}\hat{\mathbf{w}} - \mathcal{L}\mathbf{w}^*) (\mathcal{L}\mathbf{w}_a + \mathbf{J})^{-1} \right), \hat{\mathbf{w}} - \mathbf{w}^* \right\rangle = \text{tr}(\mathbf{A}\mathbf{B}\mathbf{A}\mathbf{B}),$$

where $\mathbf{w}_a = \mathbf{w}^* + a(\hat{\mathbf{w}} - \mathbf{w}^*)$, $\mathbf{A} = (\mathcal{L}\mathbf{w}_a + \mathbf{J})^{-1}$ and $\mathbf{B} = (\mathcal{L}\hat{\mathbf{w}} - \mathcal{L}\mathbf{w}^*)$. Note that \mathbf{A} is symmetric and positive definite because $\mathbf{w}_t \in \mathcal{S}_w$ and \mathbf{B} is symmetric. Let $\mathbf{C} = \mathbf{A}\mathbf{B}$. According to Theorem 1 in [2], all the eigenvalues of a matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ are real if there exists a symmetric and positive definite matrix $\mathbf{Y} \in \mathbb{R}^{p \times p}$ such that $\mathbf{X}\mathbf{Y}$ are symmetric. It is easy to check that the matrix $\mathbf{C}\mathbf{A}$ is symmetric with \mathbf{A} symmetric and positive definite, and thus all the eigenvalues of \mathbf{C} are real. Suppose $\lambda_1, \dots, \lambda_p$ are the eigenvalues of \mathbf{C} . Then the eigenvalues of $\mathbf{C}\mathbf{C}$ are $\lambda_1^2, \dots, \lambda_p^2$. Therefore, $q''(a) = \sum_{i=1}^p \lambda_i^2 \geq 0$, implying that $q'(a)$ is non-decreasing with the increase of a . Then one obtains

$$t \langle \mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathcal{L}^*(\mathcal{L}\hat{\mathbf{w}} + \mathbf{J})^{-1}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle = tq'(1) \geq tq'(t) \geq (2\lambda_{\max}(\mathcal{L}\mathbf{w}^*))^{-2} \|\mathcal{L}\mathbf{w}_t - \mathcal{L}\mathbf{w}^*\|_F^2. \quad (43)$$

where the first inequality holds because $q'(a)$ is non-decreasing and $t \leq 1$, and the second inequality follows from (41).

The Lagrangian of the optimization (39) is

$$L(\mathbf{w}, \boldsymbol{\nu}) = -\log \det(\mathcal{L}\mathbf{w} + \mathbf{J}) + \text{tr}(\mathcal{L}\mathbf{w}\mathbf{S}) + \mathbf{z}^\top \mathbf{w} - \mathbf{v}^\top \mathbf{w},$$

where \mathbf{v} is a KKT multiplier. Let $(\hat{\mathbf{w}}, \hat{\mathbf{v}})$ be the primal and dual optimal point. Then $(\hat{\mathbf{w}}, \hat{\mathbf{v}})$ must satisfy the KKT conditions as below

$$-\mathcal{L}^* \left((\mathcal{L}\hat{\mathbf{w}} + \mathbf{J})^{-1} \right) + \mathcal{L}^*\mathbf{S} + \mathbf{z} - \hat{\mathbf{v}} = \mathbf{0}; \quad (44)$$

$$\hat{w}_i \hat{v}_i = 0, \text{ for } i = 1, \dots, p(p-1)/2; \quad (45)$$

$$\hat{\mathbf{w}} \geq \mathbf{0}, \hat{\mathbf{v}} \geq \mathbf{0}; \quad (46)$$

According to (44), one has

$$\langle -\mathcal{L}^*(\mathcal{L}\hat{\mathbf{w}} + \mathbf{J})^{-1} + \mathcal{L}^*\mathbf{S}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle = \langle \hat{\mathbf{v}} - \mathbf{z}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle. \quad (47)$$

Substituting (47) into (43) yields

$$\begin{aligned} \|\mathcal{L}\mathbf{w}_t - \mathcal{L}\mathbf{w}^*\|_{\text{F}}^2 &\leq 4t\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left(\langle \hat{\mathbf{v}} - \mathbf{z}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle + \langle \mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}), \hat{\mathbf{w}} - \mathbf{w}^* \rangle \right) \\ &= 4t\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left(\underbrace{\langle \hat{\mathbf{v}}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle}_{\text{term I}} - \underbrace{\langle \mathbf{z}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle}_{\text{term II}} + \underbrace{\langle \mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}), \hat{\mathbf{w}} - \mathbf{w}^* \rangle}_{\text{term III}} \right). \end{aligned} \quad (48)$$

Next we will bound term I, II and III, respectively. The term I can be directly bounded by

$$\langle \hat{\mathbf{v}}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle = -\langle \hat{\mathbf{v}}, \mathbf{w}^* \rangle \leq 0. \quad (49)$$

where the equality follows from (45) and the inequality follows from $\hat{\mathbf{v}} \geq \mathbf{0}$ in (46) and $\mathbf{w}^* \geq \mathbf{0}$.

For term II, we separate the support of \mathbf{z} into two parts, \mathcal{S}^* and its complementary set $\{\mathcal{S}^*\}^c$, where \mathcal{S}^* is the support of \mathbf{w}^* with $|\mathcal{S}^*| \leq s$. Take a set \mathcal{E} satisfying $\mathcal{S}^* \subseteq \mathcal{E}$ and $|\mathcal{E}| \leq 2s$. A simple algebra yields

$$\begin{aligned} \langle \mathbf{z}, \hat{\mathbf{w}} - \mathbf{w}^* \rangle &= \langle \mathbf{z}_{\mathcal{S}^*}, (\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{S}^*} \rangle + \langle \mathbf{z}_{\{\mathcal{S}^*\}^c}, (\hat{\mathbf{w}} - \mathbf{w}^*)_{\{\mathcal{S}^*\}^c} \rangle \\ &= \langle \mathbf{z}_{\mathcal{S}^*}, (\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{S}^*} \rangle + \langle \mathbf{z}_{\{\mathcal{S}^*\}^c}, \hat{\mathbf{w}}_{\{\mathcal{S}^*\}^c} \rangle \\ &\geq -\|\mathbf{z}_{\mathcal{S}^*}\| \|(\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{S}^*}\| + \langle \mathbf{z}_{\{\mathcal{S}^*\}^c}, \hat{\mathbf{w}}_{\{\mathcal{S}^*\}^c} \rangle \\ &\geq -\|\mathbf{z}_{\mathcal{S}^*}\| \|(\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{S}^*}\| + \langle \mathbf{z}_{\mathcal{E}^c}, \hat{\mathbf{w}}_{\mathcal{E}^c} \rangle, \end{aligned} \quad (50)$$

where the first inequality follows from Cauchy–Schwarz inequality and the second inequality follows from $\mathbf{z} \geq \mathbf{0}$, $\hat{\mathbf{w}} \geq \mathbf{0}$ and $\mathcal{E}^c \subseteq \{\mathcal{S}^*\}^c$.

For term III, we separate the support of $\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})$ into parts, \mathcal{E} and \mathcal{E}^c . Then one has

$$\begin{aligned} \langle \mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}), \hat{\mathbf{w}} - \mathbf{w}^* \rangle &= \langle (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}}, (\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{E}} \rangle \\ &\quad + \langle (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}^c}, (\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{E}^c} \rangle \\ &\leq \left\| (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}} \right\| \|(\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{E}}\| \\ &\quad + \langle (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}^c}, \hat{\mathbf{w}}_{\mathcal{E}^c} \rangle. \end{aligned} \quad (51)$$

Substituting (49), (50) and (51) into (48) yields

$$\begin{aligned} \|\mathcal{L}\mathbf{w}_t - \mathcal{L}\mathbf{w}^*\|_{\text{F}}^2 &\leq 4t\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left(\left\| (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}} \right\| \|(\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{E}}\| \right. \\ &\quad \left. + \langle (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}^c} - \mathbf{z}_{\mathcal{E}^c}, \hat{\mathbf{w}}_{\mathcal{E}^c} \rangle + \|\mathbf{z}_{\mathcal{S}^*}\| \|(\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{S}^*}\| \right). \end{aligned} \quad (52)$$

Notice that the inequality

$$\langle (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}^c} - \mathbf{z}_{\mathcal{E}^c}, \hat{\mathbf{w}}_{\mathcal{E}^c} \rangle \leq 0, \quad (53)$$

holds because $\hat{\mathbf{w}} \geq \mathbf{0}$, $\|\mathbf{z}_{\mathcal{E}^c}\|_{\min} \geq \lambda/2$ and

$$\left\| (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}^c} \right\|_{\max} \leq \left\| \mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}) \right\|_{\max} \leq \frac{\lambda}{2},$$

where the last inequality follows from the conditions in Lemma 1.4. Combining (52) and (53) together yields

$$\begin{aligned} \|\mathcal{L}\mathbf{w}_t - \mathcal{L}\mathbf{w}^*\|_{\text{F}}^2 &\leq 4t\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left(\|\mathbf{z}_{\mathcal{S}^*}\| \|(\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{S}^*}\| + \left\| (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}} \right\| \|(\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{E}}\| \right) \\ &\leq 4t\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left(\|\mathbf{z}_{\mathcal{S}^*}\| + \left\| (\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}} \right\| \right) \|\hat{\mathbf{w}} - \mathbf{w}^*\|, \end{aligned} \quad (54)$$

where the last inequality follows from $\|\hat{\mathbf{w}} - \mathbf{w}^*\| \geq \|(\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{E}}\| \geq \|(\hat{\mathbf{w}} - \mathbf{w}^*)_{\mathcal{S}^*}\|$.

On the other hand, one has

$$\|\mathcal{L}\mathbf{w}_t - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} = t \|\mathcal{L}\hat{\mathbf{w}} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \geq t \left(\sum_{i \neq j} ([\mathcal{L}\hat{\mathbf{w}} - \mathcal{L}\mathbf{w}^*]_{ij})^2 \right)^{\frac{1}{2}} = \sqrt{2}t \|\hat{\mathbf{w}} - \mathbf{w}^*\|. \quad (55)$$

Combining (54) and (55) together yields

$$\|\mathcal{L}\mathbf{w}_t - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq 2\sqrt{2}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left(\|\mathbf{z}_{\mathcal{S}^*}\| + \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}}\| \right). \quad (56)$$

Recall that $\|\mathbf{z}\|_{\max} \leq \lambda$ and $|\mathcal{S}^*| \leq s$. Thus one has

$$\|\mathbf{z}_{\mathcal{S}^*}\| \leq \sqrt{s}\lambda. \quad (57)$$

One also has

$$\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}}\| \leq \left(|\mathcal{E}| \|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})\|_{\max}^2 \right)^{\frac{1}{2}} \leq \frac{\sqrt{2}}{2} \sqrt{s}\lambda. \quad (58)$$

Substituting (57) and (58) into (56) yields

$$\|\mathcal{L}\mathbf{w}_t - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq 2(\sqrt{2} + 1)\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\sqrt{s}\lambda < \lambda_{\max}(\mathcal{L}\mathbf{w}^*), \quad (59)$$

which implies that $t = 1$ in (40), i.e., $\mathbf{w}_t = \hat{\mathbf{w}}$. The last inequality is established by plugging $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ with $n \geq 94\alpha c_0^{-1} \lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)s \log p$. Therefore, we conclude that

$$\|\mathcal{L}\hat{\mathbf{w}} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \leq 2\sqrt{2}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left(\|\mathbf{z}_{\mathcal{S}^*}\| + \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}}\| \right) \leq 2(1 + \sqrt{2})\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\sqrt{s}\lambda,$$

where the first inequality is established by (56) with $t = 1$, and the second inequality is established by plugging (57) and (58). \square

2.5 Proof of Lemma 1.5

Proof. Recall that $\mathcal{E}^{(k)} = \{\mathcal{S}^* \cup \mathcal{S}^{(k)}\}$ and $\mathcal{S}^{(k)} = \{i \in [p(p-1)/2] \mid \hat{w}_i^{(k-1)} \geq b\}$ with $b = (2 + \sqrt{2})\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\lambda$.

We prove $|\mathcal{E}^{(k)}| \leq 2s$ holds by induction. For $k = 1$, $\forall i \notin \text{supp}^+(\hat{\mathbf{w}}^{(0)})$, i.e., $w_i^{(0)} \leq 0$, one has $w_i^{(0)} < b$, implying that $i \notin \mathcal{S}^{(1)}$. In other words, $\mathcal{S}^{(1)} \subseteq \text{supp}^+(\hat{\mathbf{w}}^{(0)})$. Then one has

$$|\mathcal{E}^{(1)}| = |\mathcal{S}^* \cup \mathcal{S}^{(1)}| \leq |\mathcal{S}^* \cup \text{supp}^+(\hat{\mathbf{w}}^{(0)})| \leq s + s = 2s.$$

Therefore, $|\mathcal{E}^{(k)}| \leq 2s$ holds for $k = 1$.

Assume $|\mathcal{E}^{(k-1)}| \leq 2s$ holds for some $k \geq 2$. We separate the set $\mathcal{E}^{(k)}$ into two parts, \mathcal{S}^* and $\mathcal{S}^{(k)} \setminus \mathcal{S}^*$. For any $i \in \mathcal{S}^{(k)} \setminus \mathcal{S}^*$, one has $\hat{w}_i^{(k-1)} \geq b$, and further obtains

$$\begin{aligned} \sqrt{|\mathcal{S}^{(k)} \setminus \mathcal{S}^*|} &\leq \sqrt{\sum_{i \in \mathcal{S}^{(k)} \setminus \mathcal{S}^*} \left(\frac{\hat{w}_i^{(k-1)}}{b} \right)^2} = \frac{\|\hat{\mathbf{w}}_{\mathcal{S}^{(k)} \setminus \mathcal{S}^*}^{(k-1)}\|}{b} = \frac{\|(\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*)_{\mathcal{S}^{(k)} \setminus \mathcal{S}^*}\|}{b} \\ &\leq \frac{\|\hat{\mathbf{w}}_{\mathcal{E}^{(k)}}^{(k-1)} - \mathbf{w}^*\|}{b} \leq \frac{\|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\|}{b}. \end{aligned} \quad (60)$$

Let $\mathbf{z}^{(k-2)}$ satisfy $z_i^{(k-2)} = h'_\lambda(\hat{w}_i^{(k-2)})$, $i \in [p(p-1)/2]$. By Assumption 3.5, one has $z_i^{(k-2)} \in [0, \lambda]$ for $i \in [p(p-1)/2]$. For any $i \in \{\mathcal{S}^{(k-1)}\}^c$, one further has

$$z_i^{(k-2)} = h'_\lambda(\hat{w}_i^{(k-2)}) \geq h'_\lambda(b) \geq \frac{\lambda}{2}, \quad (61)$$

where the first inequality holds because $\hat{w}_i^{(k-2)} < b$ for any $i \in \{\mathcal{S}^{(k-1)}\}^c$ by the definition of $\mathcal{S}^{(k-1)}$, and h'_λ is non-increasing by Assumption 3.5; the second inequality follows from Assumption 3.5. Therefore, one obtains

$$\left\| \mathbf{z}_{\{\mathcal{E}^{(k-1)}\}^c}^{(k-2)} \right\|_{\min} \geq \left\| \mathbf{z}_{\{\mathcal{S}^{(k-1)}\}^c}^{(k-2)} \right\|_{\min} \geq \lambda/2,$$

where the first inequality follows from $\{\mathcal{E}^{(k-1)}\}^c \subseteq \{\mathcal{S}^{(k-1)}\}^c$ and the second inequality follows from (61). One also has $|\mathcal{E}^{(k-1)}| \leq 2s$ and $\mathcal{S}^* \subseteq \mathcal{E}^{(k-1)}$. Hence we can apply Lemma 1.4 with $\mathcal{E} = \mathcal{E}^{(k-1)}$ and $\mathbf{z} = \mathbf{z}^{(k-2)}$ and obtain

$$\left\| \hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^* \right\| \leq \frac{\sqrt{2}}{2} \left\| \mathcal{L} \hat{\mathbf{w}}^{(k-1)} - \mathcal{L} \mathbf{w}^* \right\|_{\text{F}} \leq (2 + \sqrt{2}) \lambda_{\max}^2(\mathcal{L} \mathbf{w}^*) \sqrt{s} \lambda, \quad (62)$$

where the first inequality holds with the proof similar to (69). Combining (60) and (62) together yields

$$\sqrt{|\mathcal{S}^{(k)} \setminus \mathcal{S}^*|} \leq \frac{(2 + \sqrt{2}) \lambda_{\max}^2(\mathcal{L} \mathbf{w}^*) \sqrt{s} \lambda}{b} = \sqrt{s},$$

where the last equality follows from $b = (2 + \sqrt{2}) \lambda_{\max}^2(\mathcal{L} \mathbf{w}^*) \lambda$. Therefore, one gets

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

completing the induction. \square

2.6 Proof of Lemma 1.6

Proof. For any $k \geq 1$, one has $|\mathcal{E}^{(k)}| \leq 2s$ by Lemma 1.5. According to the definition of $\mathcal{E}^{(k)}$ in (7), one has $\mathcal{S}^* \subseteq \mathcal{E}^{(k)}$. Let $z_i^{(k-1)} = h'_\lambda(\hat{w}_i^{(k-1)})$, $i \in [p(p-1)/2]$. By Assumption 3.5, one has $z_i^{(k-1)} \in [0, \lambda]$ for $i \in [p(p-1)/2]$. For any $i \in \{\mathcal{S}^{(k)}\}^c$, one has

$$z_i^{(k-1)} = h'_\lambda(\hat{w}_i^{(k-1)}) \geq h'_\lambda(b) \geq \frac{\lambda}{2},$$

where the first inequality holds because $\hat{w}_i^{(k-1)} < b$ for any $i \in \{\mathcal{S}^{(k)}\}^c$ by the definition of $\mathcal{S}^{(k)}$ in (7), and h'_λ is non-increasing by Assumption 3.5; the second inequality follows from Assumption 3.5. Therefore, $\left\| \mathbf{z}_{\{\mathcal{E}^{(k)}\}^c}^{(k-1)} \right\|_{\min} \geq \left\| \mathbf{z}_{\{\mathcal{S}^{(k)}\}^c}^{(k-1)} \right\|_{\min} \geq \lambda/2$. Applying Lemma 1.4 with $\mathcal{E} = \mathcal{E}^{(k)}$ and $\mathbf{z} = \mathbf{z}^{(k-1)}$ yields

$$\left\| \mathcal{L} \hat{\mathbf{w}}^{(k)} - \mathcal{L} \mathbf{w}^* \right\|_{\text{F}} \leq 2\sqrt{2} \lambda_{\max}^2(\mathcal{L} \mathbf{w}^*) \left(\left\| \mathbf{z}_{\mathcal{S}^*}^{(k-1)} \right\| + \left\| (\mathcal{L}^* ((\mathcal{L} \mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}^{(k)}} \right\| \right). \quad (63)$$

We will show that the term $\left\| \mathbf{z}_{\mathcal{S}^*}^{(k-1)} \right\|$ in (63) can be bounded in terms of $\left\| \hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^* \right\|$. For any given $\mathbf{w} \in \mathbb{R}^{p(p-1)/2}$, if $|w_i^* - w_i| \geq b$, then one has

$$0 \leq h'_\lambda(w_i) \leq \lambda \leq \lambda b^{-1} |w_i^* - w_i|,$$

where $b = (2 + \sqrt{2}) \lambda_{\max}^2(\mathcal{L} \mathbf{w}^*) \lambda$, and the first two inequalities follows from Assumption 3.5. Otherwise, one has $w_i^* - w_i \leq |w_i^* - w_i| \leq b$, then $0 \leq h'_\lambda(w_i) \leq h'_\lambda(w_i^* - b)$ because h'_λ is non-increasing. Totally, one has

$$h'_\lambda(w_i) \leq \lambda b^{-1} |w_i^* - w_i| + h'_\lambda(w_i^* - b), \quad \forall i \in [p(p-1)/2]. \quad (64)$$

Collecting the indices $i \in \mathcal{S}^*$ together and applying (64) with $\mathbf{w} = \hat{\mathbf{w}}^{(k-1)}$ yields

$$\left\| \mathbf{z}_{\mathcal{S}^*}^{(k-1)} \right\| = \left\| h'_\lambda(\hat{\mathbf{w}}_{\mathcal{S}^*}^{(k-1)}) \right\| \leq \frac{\lambda}{b} \left\| \hat{\mathbf{w}}_{\mathcal{S}^*}^{(k-1)} - \mathbf{w}_{\mathcal{S}^*}^* \right\| + \left\| h'_\lambda(\mathbf{w}_{\mathcal{S}^*}^* - \mathbf{b}) \right\| \leq \frac{\lambda}{b} \left\| \hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^* \right\| + \left\| h'_\lambda(\mathbf{w}_{\mathcal{S}^*}^* - \mathbf{b}) \right\|, \quad (65)$$

where $h'_\lambda(\hat{\mathbf{w}}_{\mathcal{S}^*}^{(k-1)}) = (h'_\lambda(\hat{w}_i^{(k-1)}))_{i \in \mathcal{S}^*}$, and $\mathbf{b} = [b, \dots, b]^\top$ is a constant vector. Combining (63) and (65) together yields

$$\begin{aligned} \|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} &\leq 2\sqrt{2}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left(\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}^{(k)}}\| + \|h'_\lambda(\mathbf{w}_{\mathcal{S}^*}^* - \mathbf{b})\| \right) \\ &\quad + 2\sqrt{2}\frac{\lambda}{b}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\|. \end{aligned} \quad (66)$$

By separating the set $\mathcal{E}^{(k)}$ into two parts, \mathcal{S}^* and $\mathcal{S}^{(k)} \setminus \mathcal{S}^*$, one has

$$\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}^{(k)}}\| \leq \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| + \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^{(k)} \setminus \mathcal{S}^*}\|.$$

We will show that the term $\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^{(k)} \setminus \mathcal{S}^*}\|$ can be bounded in terms of $\|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\|$.

$$\begin{aligned} \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^{(k)} \setminus \mathcal{S}^*}\| &\leq \sqrt{|\mathcal{E}^{(k)} \setminus \mathcal{S}^*|} \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^{(k)} \setminus \mathcal{S}^*}\|_{\max} \\ &\leq \sqrt{|\mathcal{E}^{(k)} \setminus \mathcal{S}^*|} \|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})\|_{\max} \\ &\leq \frac{1}{b} \|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\| \|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})\|_{\max} \\ &\leq \frac{\lambda}{2b} \|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\|, \end{aligned}$$

where the last second equality follows from (60). Thus one has

$$\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{E}^{(k)}}\| \leq \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| + \frac{\lambda}{2b} \|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\|. \quad (67)$$

Substituting (67) into (66) yields

$$\begin{aligned} \|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} &\leq 2\sqrt{2}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \left(\|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| + \|h'_\lambda(\mathbf{w}_{\mathcal{S}^*}^* - \mathbf{b})\| \right) \\ &\quad + 3\sqrt{2}\frac{\lambda}{b}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\| \\ &= 2\sqrt{2}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| + \frac{3\sqrt{2}}{2 + \sqrt{2}} \|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\| \\ &\leq 2\sqrt{2}\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| + \frac{3}{2 + \sqrt{2}} \|\mathcal{L}\hat{\mathbf{w}}^{(k-1)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}}, \end{aligned} \quad (68)$$

where the equality is established by plugging $b = (2 + \sqrt{2})\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*)\lambda$ and following from $\|h'_\lambda(\mathbf{w}_{\mathcal{S}^*}^* - \mathbf{b})\| = 0$ because $\|\mathbf{w}_{\mathcal{S}^*}^*\|_{\min} - b \geq \gamma\lambda$ and $h'_\lambda(x) = 0$ for any $x \geq \gamma\lambda$ following from Assumption 3.6; the last inequality follows from

$$\|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} = \left(2 \sum_{i=1}^{p(p-1)/2} (\hat{w}_i^{(k)} - w_i^*)^2 + \sum_{j=1}^p ([\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*]_{jj})^2 \right)^{\frac{1}{2}} \geq \sqrt{2} \|\hat{\mathbf{w}}^{(k)} - \mathbf{w}^*\|. \quad (69)$$

Similarly, one also obtains

$$\begin{aligned} \|\hat{\mathbf{w}}^{(k)} - \mathbf{w}^*\| &\leq \frac{\sqrt{2}}{2} \|\mathcal{L}\hat{\mathbf{w}}^{(k)} - \mathcal{L}\mathbf{w}^*\|_{\mathbb{F}} \\ &\leq 2\lambda_{\max}^2(\mathcal{L}\mathbf{w}^*) \|(\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S}))_{\mathcal{S}^*}\| + \frac{3}{2 + \sqrt{2}} \|\hat{\mathbf{w}}^{(k-1)} - \mathbf{w}^*\|. \end{aligned}$$

□

2.7 Proof of Lemma 1.7

Proof. We apply Lemma 1.8 with $t = \lambda/2$ and union sum bound, then get

$$\mathbb{P}[\|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})\|_{\max} \geq \lambda/2] \leq p(p-1) \exp(-\frac{1}{4}c_0 n \lambda^2) \leq p^2 \exp(-\frac{1}{4}c_0 n \lambda^2),$$

for any $\lambda \leq 2t_0$, where $t_0 = \|\mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1}\|_{\max}$ and $c_0 = 1/(8\|\mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1}\|_{\max}^2)$. Take $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ for some $\alpha > 2$. To guarantee $\lambda \leq 2t_0$, one takes $n \geq 8\alpha \log p$. By calculation, we establish

$$\mathbb{P}[\|\mathcal{L}^*((\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1} - \mathbf{S})\|_{\max} \leq \lambda/2] \geq 1 - p^2 \exp(-\frac{1}{4}c_0 n \lambda^2) \geq 1 - 1/p^{\alpha-2},$$

completing the proof. \square

2.8 Proof of Lemma 1.8

Proof. The LGMRF is a constrained GMRF model with $\mathbf{1}^\top \mathbf{X} = 0$ and we can follow the method called *conditioning by Kriging* [4] to sample LGMRF. More specifically, to sample \mathbf{X} for LGMRF with precision matrix $\mathcal{L}\mathbf{w}^*$, we could first sample from a unconstrained GMRF $\tilde{\mathbf{X}} \sim N(\mathbf{0}, (\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1})$ and then correct for the constraint $\mathbf{1}^\top \mathbf{X} = 0$ by

$$\mathbf{X}^{(k)} = \tilde{\mathbf{X}}^{(k)} - \frac{1}{p} \mathbf{1} \mathbf{1}^\top \tilde{\mathbf{X}}^{(k)}, \quad \text{for } k = 1, \dots, n.$$

For any $i \in [p(p-1)/2]$, one has

$$\begin{aligned} [\mathcal{L}^* \mathbf{S}]_i &= [\mathcal{L}^* (\frac{1}{n} \sum_{k=1}^n \mathbf{X}^{(k)} \mathbf{X}^{(k)T})]_i = \frac{1}{n} \sum_{k=1}^n [\mathcal{L}^* (\mathbf{X}^{(k)} \mathbf{X}^{(k)T})]_i = \frac{1}{n} \sum_{k=1}^n (X_a^{(k)} - X_b^{(k)})^2 \\ &= \frac{1}{n} \sum_{k=1}^n \left(\left(\tilde{X}_a^{(k)} - \left(\frac{1}{p} \mathbf{1} \mathbf{1}^\top \tilde{\mathbf{X}}^{(k)} \right)_a \right) - \left(\tilde{X}_b^{(k)} - \left(\frac{1}{p} \mathbf{1} \mathbf{1}^\top \tilde{\mathbf{X}}^{(k)} \right)_b \right) \right)^2 \\ &= \frac{1}{n} \sum_{k=1}^n (\tilde{X}_a^{(k)} - \tilde{X}_b^{(k)})^2, \end{aligned} \quad (70)$$

where the indices a and b obey $i = a - b + (b-1)(2p-b)/2$ and $a > b$.

Let $\tilde{\Sigma} = (\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1}$ and thus $\tilde{\mathbf{X}} \sim N(\mathbf{0}, \tilde{\Sigma})$. We first introduce two auxiliary random variables $Y_{k,i} := \tilde{X}_a^{(k)} - \tilde{X}_b^{(k)}$ and $Z_{k,i} := Y_{k,i}^2$. Together with (70), one has

$$\frac{1}{n} \sum_{k=1}^n Z_{k,i} = [\mathcal{L}^* \mathbf{S}]_i. \quad (71)$$

We can see $Y_{k,i} \sim N(0, \sigma_i^2)$ because of the fact that any linear combination of p components in $\tilde{\mathbf{X}}$ has a univariate normal distribution. The variance of $Y_{k,i}$ is

$$\begin{aligned} \sigma_i^2 &= \mathbb{E}[(Y_{k,i} - \mathbb{E}(Y_{k,i}))^2] = \mathbb{E}[(\tilde{X}_a^{(k)} - \tilde{X}_b^{(k)})^2] \\ &= \tilde{\Sigma}_{aa} + \tilde{\Sigma}_{bb} - \tilde{\Sigma}_{ab} - \tilde{\Sigma}_{ba} = (\mathcal{L}^* \tilde{\Sigma})_i. \end{aligned} \quad (72)$$

Therefore $Z_{k,i}/\sigma_i^2 \sim \chi^2(1)$ and $\mathbb{E}[Z_{k,i}/\sigma_i^2] = 1$. We say a random variable X is *sub-exponential* if there are non-negative parameters (ν, α) such that

$$\mathbb{E}[\exp(\lambda(X - \mathbb{E}[X]))] \leq \exp\left(\frac{\nu^2 \lambda^2}{2}\right), \quad \text{for all } |\lambda| < \frac{1}{\alpha}. \quad (73)$$

By checking the condition in (73), one can conclude that $Z_{k,i}/\sigma_i^2$ is sub-exponential with parameters $(2, 4)$. Furthermore, if random variables $\{Y_k\}_{k=1}^n$ are independent and sub-exponential with parameters (v_k, α_k) , then $\sum_{k=1}^n Y_k$ is still sub-exponential with parameters (v_*, α_*) where

$$v_* := \sqrt{\sum_{k=1}^n v_k^2} \quad \text{and} \quad \alpha_* := \max_{k=1, \dots, n} \alpha_k.$$

Thus, $\sum_{k=1}^n Z_{k,i}/\sigma_i^2$ is sub-exponential with parameters $(2\sqrt{n}, 4)$. The application of the sub-exponential tail bound in Lemma 1.11 yields

$$\mathbb{P}\left[\left|\sum_{k=1}^n Z_{k,i}/\sigma_i^2 - n\right| \geq t_0\right] \leq 2 \exp\left(-\frac{t_0^2}{8n}\right), \quad \text{for } t_0 \in [0, n].$$

By taking $t_0 = nt/\max_i \sigma_i^2$, one has

$$\begin{aligned} \mathbb{P}\left[\left|\frac{1}{n} \sum_{k=1}^n Z_{k,i} - \sigma_i^2\right| \geq t\right] &= \mathbb{P}\left[\left|\sum_{k=1}^n Z_{k,i}/\sigma_i^2 - n\right| \geq \frac{nt}{\sigma_i^2}\right] \\ &\leq \mathbb{P}\left[\left|\sum_{k=1}^n Z_{k,i}/\sigma_i^2 - n\right| \geq \frac{nt}{\max_i \sigma_i^2}\right] \\ &\leq 2 \exp\left(-\frac{nt^2}{8(\max_i \sigma_i^2)^2}\right) \end{aligned} \quad (74)$$

holds for $t \in [0, \max_i \sigma_i^2]$. Notice that $\sigma_i^2 = (\mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1})_i$ according to (72). Substituting (71) into (74) yields

$$\mathbb{P}\left[\left|[\mathcal{L}^* \mathbf{S}]_i - (\mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1})_i\right| \geq t\right] \leq 2 \exp(-c_0 nt^2), \quad \text{for } t \in [0, t_0]. \quad (75)$$

where $t_0 = \|\mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1}\|_{\max}$ and $c_0 = 1/(8\|\mathcal{L}^*(\mathcal{L}\mathbf{w}^* + \mathbf{J})^{-1}\|_{\max}^2)$, completing the proof. \square

2.9 Proof of Lemma 1.9

Proof. Recall that $\mathbb{B}_M(\mathbf{w}^*; r) = \{\mathbf{w} \in \mathbb{R}^{p(p-1)/2} \mid \|\mathbf{w} - \mathbf{w}^*\|_M \leq r\}$, where $\|\mathbf{x}\|_M^2 = \langle \mathbf{x}, \mathbf{M}\mathbf{x} \rangle = \|\mathcal{L}\mathbf{x}\|_{\mathbb{F}}^2$ with $\mathbf{M} \succ \mathbf{0}$ defined in Lemma 1.3, and $\mathcal{S}_{\mathbf{w}} = \{\mathbf{w} \mid \mathbf{w} \geq \mathbf{0}, (\mathcal{L}\mathbf{w} + \mathbf{J}) \in \mathcal{S}_{++}^p\}$. We can see $\mathcal{B}_M(\mathbf{w}^*; r)$ is a convex set because both $\mathbb{B}_M(\mathbf{w}^*; r)$ and $\mathcal{S}_{\mathbf{w}}$ are convex. It is easy to check that $\mathcal{S}_{\mathbf{w}}$ is convex (See (4) for more details). For any $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{B}_M(\mathbf{w}^*; r)$, by Mean Value Theorem, one obtains

$$f(\mathbf{w}_2) = f(\mathbf{w}_1) + \langle \nabla f(\mathbf{w}_1), \mathbf{w}_2 - \mathbf{w}_1 \rangle + \frac{1}{2} \langle \mathbf{w}_2 - \mathbf{w}_1, \nabla^2 f(\mathbf{w}_t)(\mathbf{w}_2 - \mathbf{w}_1) \rangle, \quad (76)$$

where $\mathbf{w}_t = t\mathbf{w}_2 + (1-t)\mathbf{w}_1$ with $t \in [0, 1]$. For any nonzero $\mathbf{x} \in \mathbb{R}^{p(p-1)/2}$, one has

$$\begin{aligned} \mathbf{x}^\top \nabla^2 f(\mathbf{w}_t) \mathbf{x} &= (\text{vec}(\mathcal{L}\mathbf{x}))^\top \left((\mathcal{L}\mathbf{w}_t + \mathbf{J})^{-1} \otimes (\mathcal{L}\mathbf{w}_t + \mathbf{J})^{-1} \right) \text{vec}(\mathcal{L}\mathbf{x}) \\ &= \frac{(\text{vec}(\mathcal{L}\mathbf{x}))^\top \left((\mathcal{L}\mathbf{w}_t + \mathbf{J})^{-1} \otimes (\mathcal{L}\mathbf{w}_t + \mathbf{J})^{-1} \right) \text{vec}(\mathcal{L}\mathbf{x})}{(\text{vec}(\mathcal{L}\mathbf{x}))^\top \text{vec}(\mathcal{L}\mathbf{x})} \cdot \|\mathcal{L}\mathbf{x}\|_{\mathbb{F}}^2 \\ &\geq \inf_{\mathbf{y}} \frac{(\text{vec}(\mathcal{L}\mathbf{y}))^\top \left(((\mathcal{L}\mathbf{w}_t)^\dagger + \mathbf{J}) \otimes ((\mathcal{L}\mathbf{w}_t)^\dagger + \mathbf{J}) \right) \text{vec}(\mathcal{L}\mathbf{y})}{(\text{vec}(\mathcal{L}\mathbf{y}))^\top \text{vec}(\mathcal{L}\mathbf{y})} \cdot \|\mathcal{L}\mathbf{x}\|_{\mathbb{F}}^2 \\ &= \inf_{\mathbf{y}} \frac{(\text{vec}(\mathcal{L}\mathbf{y}))^\top \left((\mathcal{L}\mathbf{w}_t)^\dagger \otimes (\mathcal{L}\mathbf{w}_t)^\dagger \right) \text{vec}(\mathcal{L}\mathbf{y})}{(\text{vec}(\mathcal{L}\mathbf{y}))^\top \text{vec}(\mathcal{L}\mathbf{y})} \cdot \|\mathcal{L}\mathbf{x}\|_{\mathbb{F}}^2 \\ &\geq \lambda_2^2((\mathcal{L}\mathbf{w}_t)^\dagger) \cdot \|\mathcal{L}\mathbf{x}\|_{\mathbb{F}}^2, \end{aligned} \quad (77)$$

where $\lambda_2((\mathcal{L}\mathbf{w}_t)^\dagger)$ denotes the second smallest eigenvalue of $(\mathcal{L}\mathbf{w}_t)^\dagger$ and $(\mathcal{L}\mathbf{w}_t)^\dagger$ is the pseudo inverse of $\mathcal{L}\mathbf{w}_t$. The first equality follows from Lemma 1.1; the last equality follows from $(\text{vec}(\mathcal{L}\mathbf{y}))^\top ((\mathcal{L}\mathbf{w}_t)^\dagger \otimes \mathbf{J}) \text{vec}(\mathcal{L}\mathbf{y}) = 0$, $(\text{vec}(\mathcal{L}\mathbf{y}))^\top (\mathbf{J} \otimes (\mathcal{L}\mathbf{w}_t)^\dagger) \text{vec}(\mathcal{L}\mathbf{y}) = 0$, and $(\text{vec}(\mathcal{L}\mathbf{y}))^\top (\mathbf{J} \otimes \mathbf{J}) \text{vec}(\mathcal{L}\mathbf{y}) = 0$ which are easy to verify; the last inequality holds because of the property of Kronecker product that the eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ are $\lambda_i \mu_j$ with the corresponding eigenvector $\mathbf{a}_i \otimes \mathbf{b}_j$, where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of $\mathbf{A} \in \mathbb{R}^{p \times p}$ with the corresponding eigenvectors $\mathbf{a}_1, \dots, \mathbf{a}_p$, and μ_1, \dots, μ_p are the eigenvalues of $\mathbf{B} \in \mathbb{R}^{p \times p}$ with the corresponding eigenvectors $\mathbf{b}_1, \dots, \mathbf{b}_p$. Notice that there is one and only one zero eigenvalue for $\mathcal{L}\mathbf{w}_t$ because $\mathbf{w}_t \in \mathcal{S}_w$. Assume $\lambda_1, \dots, \lambda_p$ are the eigenvalues of $(\mathcal{L}\mathbf{w}_t)^\dagger$ with the corresponding eigenvectors $\mathbf{a}_1, \dots, \mathbf{a}_p$. Without loss of generality, let $\lambda_1 = 0$ and then $\mathbf{a}_1 = \frac{1}{\sqrt{p}}\mathbf{1}$. By calculation, one obtains

$$(\text{vec}(\mathcal{L}\mathbf{y}))^\top \text{vec}(\mathbf{a}_i \otimes \mathbf{a}_1) = 0, \quad \text{and} \quad (\text{vec}(\mathcal{L}\mathbf{y}))^\top \text{vec}(\mathbf{a}_1 \otimes \mathbf{a}_i) = 0,$$

for any $i = 1, \dots, p$ and any $\mathbf{y} \in \mathbb{R}^{p(p-1)/2}$, indicating that $\text{vec}(\mathcal{L}\mathbf{y})$ is orthogonal to all the eigenvectors of $(\mathcal{L}\mathbf{w}_t)^\dagger \otimes (\mathcal{L}\mathbf{w}_t)^\dagger$ corresponding to zero eigenvalues. The smallest nonzero eigenvalue of $(\mathcal{L}\mathbf{w}_t)^\dagger \otimes (\mathcal{L}\mathbf{w}_t)^\dagger$ is $\lambda_2^2((\mathcal{L}\mathbf{w}_t)^\dagger)$, establishing (77). Notice that (77) also holds with $\mathbf{x} = \mathbf{0}$. One further obtains

$$\lambda_2^2((\mathcal{L}\mathbf{w}_t)^\dagger) \geq (\|\mathcal{L}\mathbf{w}^*\|_2 + (1-t)\|\mathcal{L}\mathbf{w}_1 - \mathcal{L}\mathbf{w}^*\|_2 + t\|\mathcal{L}\mathbf{w}_2 - \mathcal{L}\mathbf{w}^*\|_2)^{-2} \geq (\|\mathcal{L}\mathbf{w}^*\|_2 + r)^{-2}, \quad (78)$$

where the second inequality is established by $\|\mathcal{L}\mathbf{x}\|_2 \leq \|\mathcal{L}\mathbf{x}\|_F$ and the fact that both $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{B}_M(\mathbf{w}^*; r)$. Substituting (78) into (77) yields

$$\mathbf{x}^\top \nabla^2 f(\mathbf{w}_t) \mathbf{x} \geq (\|\mathcal{L}\mathbf{w}^*\|_2 + r)^{-2} \cdot \|\mathcal{L}\mathbf{x}\|_F^2. \quad (79)$$

Combining (76) and (79) yields

$$f(\mathbf{w}_2) \geq f(\mathbf{w}_1) + \langle \nabla f(\mathbf{w}_1), \mathbf{w}_2 - \mathbf{w}_1 \rangle + \frac{1}{2} (\|\mathcal{L}\mathbf{w}^*\|_2 + r)^{-2} \|\mathcal{L}\mathbf{w}_1 - \mathcal{L}\mathbf{w}_2\|_F^2, \quad (80)$$

and

$$f(\mathbf{w}_1) \geq f(\mathbf{w}_2) + \langle \nabla f(\mathbf{w}_2), \mathbf{w}_1 - \mathbf{w}_2 \rangle + \frac{1}{2} (\|\mathcal{L}\mathbf{w}^*\|_2 + r)^{-2} \|\mathcal{L}\mathbf{w}_1 - \mathcal{L}\mathbf{w}_2\|_F^2, \quad (81)$$

Combining (80) and (81), we establish

$$\langle \nabla f(\mathbf{w}_1) - \nabla f(\mathbf{w}_2), \mathbf{w}_1 - \mathbf{w}_2 \rangle \geq (\|\mathcal{L}\mathbf{w}^*\|_2 + r)^{-2} \|\mathcal{L}\mathbf{w}_1 - \mathcal{L}\mathbf{w}_2\|_F^2, \quad (82)$$

completing the proof. \square

3 Additional Experimental Results

In this section, we present additional numerical simulation results on synthetic data and real-world data.

3.1 Synthetic Data

We present additional simulation results on random Barabasi-Albert graphs. Figure 1 illustrates the histograms of the nonzero graph weights, which are learned by the ℓ_1 -norm regularization method with different regularization parameters as shown in Figure 1 of the paper. Figure 1 (d) depicts the histogram for the graph in Figure 1 (d) of the paper, which is a fully connected graph, i.e., every graph weight is strictly positive. It is observed in Figure 1 (d) that all the graph weights are very small. Therefore, a large regularization parameter will lead to learn a graph with every weight strictly positive and small. We can see that the histogram in Figure 1 (d) is significantly different from the true histogram in Figure 1 (a), implying that the estimated model will fail to identify the true relationships among the data variables.

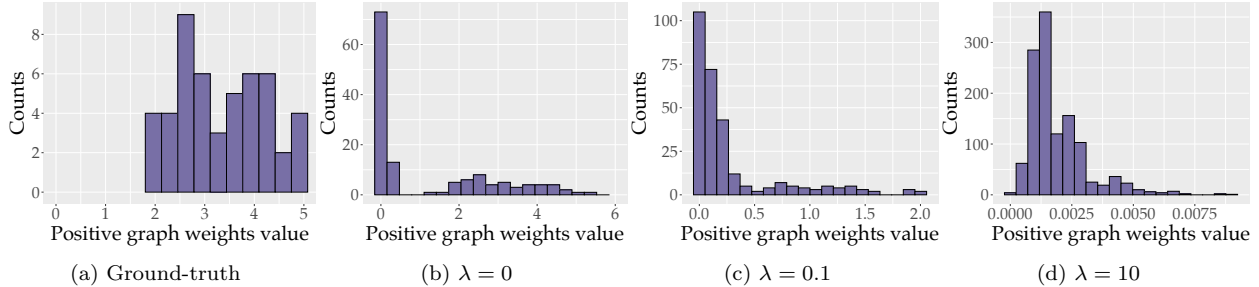


Figure 1: Histograms of nonzero graph weights learned by the ℓ_1 -norm regularization method with different regularization parameters, corresponding to the graphs in Figure 1 of the paper. The histograms count the number of nonzero graph weights falling into each interval.

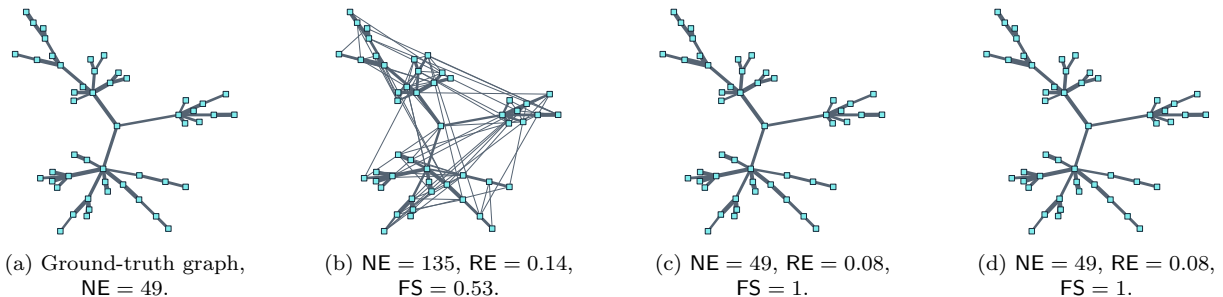


Figure 2: A sample result of learning a Barabasi-Albert graph of degree one by (b) GLE-ADMM [8], (c) NGL-SCAD (proposed) and (d) NGL-MCP (proposed). The sample size ratio is $n/p = 6$. NE denotes the number of positive edges in the graph, i.e., the number of nonzero graph weights. The regularization parameters for each method are set as $\lambda_{\text{ADMM}} = 0$, $\lambda_{\text{SCAD}} = \lambda_{\text{MCP}} = 0.5$.

Figure 2 shows a sample result of learning a random Barabasi-Albert graph via GLE-ADMM [8], NGL-SCAD and NGL-MCP. It is observed that the graphs learned by NGL-SCAD and NGL-MCP present the connection between any two nodes correctly, while there are many incorrect connections in the graph learned by GLE-ADMM. In addition, performance measures including sparsity, relative error, and F-score also indicate a better performance of the proposed method.

Table 1 shows the running time for different numbers of nodes p , where the graph weight \mathbf{w} has the dimension $p(p-1)/2$. It is observed that the computational time of the proposed NGL-SCAD is much less than that of GLE-ADMM.

3.2 Real-world Data

In this section, we compare the proposed method with the benchmark GLE-ADMM [8] on two real-world data sets.

We conduct experiments on the COVID-19 data set¹ provided by the Israelite Hospital Albert Einstein in Brazil. The data set contains anonymized data from patients who had samples collected to perform the test

¹The data set is freely available at: <https://www.kaggle.com/einsteindata4u/covid19>

Table 1: Comparison of computational time (seconds)

p	50	100	500	1000
GLE-ADMM	0.117	0.676	57.032	485.465
NGL-SCAD	0.023	0.097	9.000	67.018

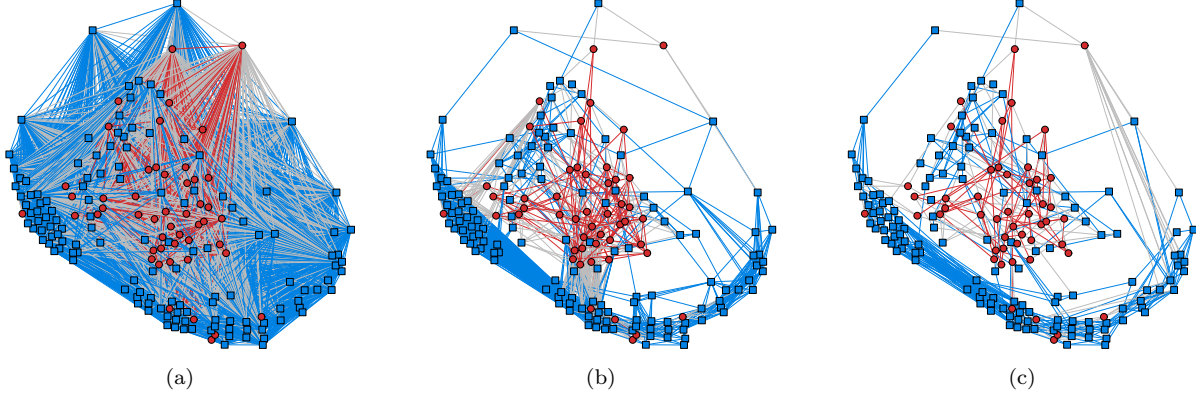


Figure 3: Learned graphs using the Brazilian COVID-19 data set via (a) GLE-ADMM, (b) NGL-SCAD (proposed), and (c) NGL-MCP (proposed). The computational time for GLE-ADMM, NGL-SCAD, and NGL-MCP are 9.9, 55.8 and 57.9 seconds, respectively. The regularization parameters are set as $\lambda_{\text{ADMM}} = 0$, $\lambda_{\text{SCAD}} = 0.1$, and $\lambda_{\text{MCP}} = 0.5$.

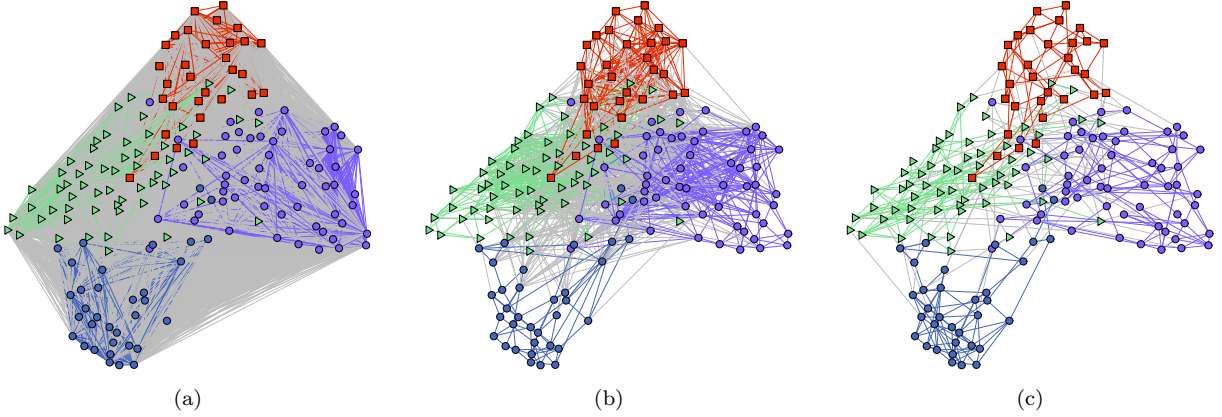


Figure 4: Stock graphs learned via (a) GLE-ADMM, (b) NGL-SCAD (proposed), and (c) NGL-MCP (proposed). The computational time for GLE-ADMM, NGL-SCAD, and NGL-MCP are 20.5, 1.3 and 90.3 seconds, respectively. The regularization parameters are set as $\lambda_{\text{ADMM}} = 0$, $\lambda_{\text{SCAD}} = 0.25$, and $\lambda_{\text{MCP}} = 0.5$.

for SARS-CoV-2. The features in the data set are mainly clinical coming from blood, urine, and saliva exams, e.g., hemoglobin level, platelets, red blood cells, etc. The original data set contains 108 features from 558 patients. We do not consider features that were measured for at most 10 patients due to the high number of missing values. In addition, a large number of patients had no record of any features. Finally, we end up with a data matrix of 182 patients with 57 features, i.e., $p = 182$ and $n = 57$. The remaining missing values were filled in with zeros. We then compare the proposed method with the GLE-ADMM method on this data set. It is observed in Figure 3 that the proposed NGL-SCAD and NGL-MCP output a more interpretable representation of the network, where the blue and red nodes denote patients who tested negative and positive for SARS-CoV-2, respectively.

We also conduct experiments on the data set from stocks composing the S&P 500 index. We select log-returns from 181 stocks from 4 sectors, namely: "Industrials", "Consumer Staples", "Energy", "Information Technology", during a period of 4 years from January 1st 2016 to May 20th 2020, with a total of 1101 observations. Then the data matrix \mathbf{X} has a size of 181×1101 . The graphs are learned on the basis of the sample correlation matrix. Figure 4 shows that the graphs learned by NGL-SCAD and NGL-MCP are able to vividly display the sectors modularity, whereas the graph learned via GLE-ADMM fails to do so.

Finally, we present additional experimental results of GLE-ADMM on the 2019-nCoV data set with

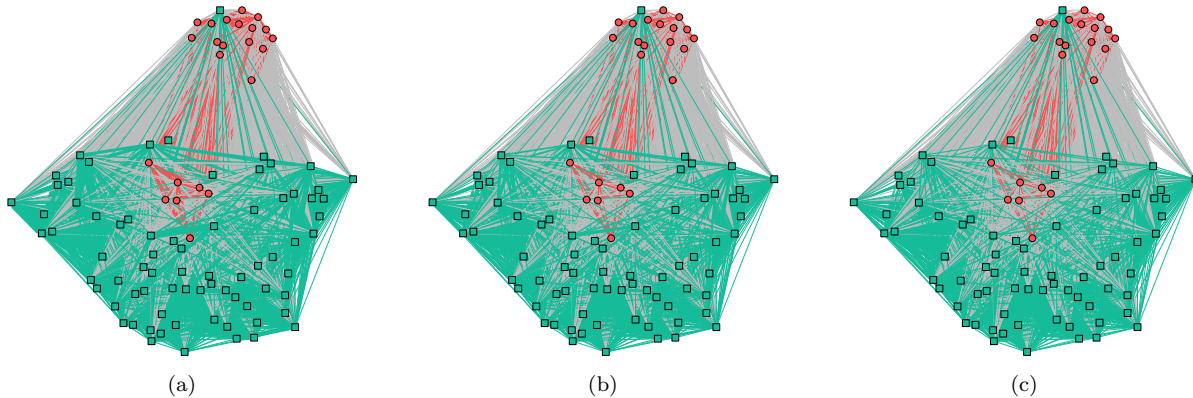


Figure 5: The learned graphs by GLE-ADMM with (a) $\lambda = 0$, (b) $\lambda = 0.02$ and (c) $\lambda = 0.1$ on the 2019-nCoV data set.

different regularization parameters. It is observed in Figure 5 that the learned graphs learned by GLE-ADMM are always dense along the increase of λ , implying that the ℓ_1 -norm is not able to impose a sparse solution here.

References

- [1] P. Drábek and J. Milota. *Methods of nonlinear analysis: applications to differential equations*. Springer Science & Business Media, 2007.
- [2] M. P. Drazin and E. V. Haynsworth. Criteria for the reality of matrix eigenvalues. *Mathematische Zeitschrift*, 78(1):449–452, 1962.
- [3] S. A. Gershgorin. Über die abgrenzung der eigenwerte einer matrix. *Bulletin of the Russian Academy of Sciences*, (6):749–754, 1931.
- [4] H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, 2005.
- [5] R. S. Varga. *Gershgorin and his circles*. Berlin, Germany: Springer-Verlag, 2004.
- [6] M. J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.
- [7] R. A. Willoughby. The inverse M-matrix problem. *Linear Algebra and its Applications*, 18(1):75–94, 1977.
- [8] L. Zhao, Y. Wang, S. Kumar, and D. P. Palomar. Optimization algorithms for graph Laplacian estimation via ADMM and MM. *IEEE Transactions on Signal Processing*, 67(16):4231–4244, 2019.