A Proofs of Theorems

A.1 Proof of Theorem 1

Proof. Suppose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are *i.i.d.* samples drawn from p-dimensional L-GMRF with the parameters $(\mathbf{0}, \boldsymbol{\Omega})$. Then the sample covariance matrix is computed by $\boldsymbol{S} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^\top$. Recall that the adjoint operator $\mathcal{L}^*$ of $\mathcal{L}$ is defined by $[\mathcal{L}^* \boldsymbol{Y}]_k = Y_{i,i} - Y_{i,j} - Y_{j,i} + Y_{j,j}$, where $[\mathcal{L}^* \boldsymbol{Y}]_k$ denotes the k-th element of the vector $\mathcal{L}^* \boldsymbol{Y}$, and $i, j \in [p]$ satisfying $k = i - j + \frac{j-1}{2}(2p - j)$ and $i > j$. By simple calculation, one obtains, for any $k \in [p(p-1)/2]$,

$$[\mathcal{L}^* \boldsymbol{S}]_k = \frac{1}{n}\sum_{t=1}^{n} \left[\mathcal{L}^*\left(\boldsymbol{x}_t \boldsymbol{x}_t^\top\right)\right]_k = \frac{1}{n}\sum_{t=1}^{n} \left([\boldsymbol{x}_t]_i - [\boldsymbol{x}_t]_j\right)^2, \tag{21}$$

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

Define a set $\mathcal{A}_{ij} := \left\{\boldsymbol{x} \in V^{p-1} | x_i = x_j\right\}$, where $V^{p-1} = \{\boldsymbol{x} \in \mathbb{R}^p |\, \mathbf{1}^\top \boldsymbol{x} = 0\}$. Obviously, $\mathcal{A}_{ij}$ has measure zero in V^{p-1} for any $i \neq j$. Therefore, for $n \geq 1$, one has

$$\mathbb{P}\left[\min_{k \in [p(p-1)/2]} [\mathcal{L}^* \boldsymbol{S}]_k > 0\right] = 1. \tag{22}$$

By the usage of the linear operator $\mathcal{L}$, the optimization (3) can be rewritten as

$$\min_{\boldsymbol{w}} -\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) + \operatorname{tr}\left(\mathcal{L}\boldsymbol{w}\boldsymbol{S}\right), \quad \text{subject to} \quad \boldsymbol{w} \geq \mathbf{0}. \tag{23}$$

Following from the the fact that $\langle \mathcal{L}\boldsymbol{x}, \boldsymbol{X}\rangle = \langle \boldsymbol{x}, \mathcal{L}^* \boldsymbol{X}\rangle$ holds for any $\boldsymbol{x} \in \mathbb{R}^{p(p-1)/2}$ and $\boldsymbol{X} \in \mathbb{R}^{p \times p}$, we further rewrite (23) as

$$\min_{\boldsymbol{w}} -\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) + \langle \mathcal{L}^* \boldsymbol{S}, \boldsymbol{w}\rangle, \quad \text{subject to} \quad \boldsymbol{w} \geq \mathbf{0}. \tag{24}$$

By the reformulation in (24), it is equivalent to prove that the global minimizer of (24) exists and is unique almost surely if $n \geq 1$. The existence is established by the coercivity of the objective function in (24). A function $g : \Omega \to \mathbb{R} \cup \{+\infty\}$ is called *coercive* over Ω, if every sequence $\boldsymbol{x}_k \in \Omega$ with $\|\boldsymbol{x}_k\| \to +\infty$ obeys $\lim_{k\to\infty} g(\boldsymbol{x}_k) = +\infty$, where $\Omega \subset \mathbb{R}^q$. Define $F(\boldsymbol{w}) = -\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) + \langle \mathcal{L}^* \boldsymbol{S}, \boldsymbol{w}\rangle$. We can bound $F(\boldsymbol{w})$ from the below as follows

$$\begin{aligned}
F(\boldsymbol{w}) &= -\log\left(\prod_{i=2}^{p} \lambda_i(\mathcal{L}\boldsymbol{w})\right) + \langle \mathcal{L}^* \boldsymbol{S}, \boldsymbol{w}\rangle \\
&\geq -(p-1)\log\left(\sum_{i=1}^{p} \lambda_i(\mathcal{L}\boldsymbol{w})\right) + \langle \mathcal{L}^* \boldsymbol{S}, \boldsymbol{w}\rangle + (p-1)\log(p-1) \\
&= -(p-1)\log\left(\sum_{i=1}^{p} [\mathcal{L}\boldsymbol{w}]_{ii}\right) + \langle \mathcal{L}^* \boldsymbol{S}, \boldsymbol{w}\rangle + (p-1)\log(p-1) \\
&= -(p-1)\log\left(2\sum_{t=1}^{p(p-1)/2} w_t\right) + \langle \mathcal{L}^* \boldsymbol{S}, \boldsymbol{w}\rangle + (p-1)\log(p-1) \\
&\geq -(p-1)\log\left(\sum_{t=1}^{p(p-1)/2} w_t\right) + \min_k [\mathcal{L}^* \boldsymbol{S}]_k \sum_{t=1}^{p(p-1)/2} w_t + (p-1)\log\frac{p-1}{2},
\end{aligned} \tag{25}$$

where the first equality follows from the fact that $\mathcal{L}\boldsymbol{w} + \boldsymbol{J}$ admits the eigenvalue decomposition that

$$\mathcal{L}\boldsymbol{w} + \boldsymbol{J} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{u} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{U} & \boldsymbol{u} \end{bmatrix}^\top, \tag{26}$$

where $\mathcal{L}\boldsymbol{w} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ and $\boldsymbol{J} = \boldsymbol{u}\boldsymbol{u}^\top$. The first inequality in (25) is established due to the inequality $\frac{a_1 + a_2 + \ldots + a_n}{n} \geq \sqrt[n]{a_1 \cdot a_2 \cdots a_n}$ for any non-negative real numbers of $a_1, \ldots, a_n$, and the fact that the smallest eigenvalue of $\mathcal{L}\boldsymbol{w}$ is zero. The last equality directly follows from the definition of the linear operator $\mathcal{L}$.

Let $\alpha = \min_k [\mathcal{L}^* \boldsymbol{S}]_k$. According to (22), we obtain that $\alpha > 0$ holds almost surely. Let $z = \sum_{t=1}^{p(p-1)/2} w_t$. Then the objective function can be lower bounded by

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

A simple calculation yields $\lim_{z\to+\infty} h(z) = +\infty$ if $\alpha > 0$. Note that the feasible set of the optimization (24) is

$$\mathcal{S}_{\boldsymbol{w}} = \left\{\boldsymbol{w} \,|\, \boldsymbol{w} \geq \boldsymbol{0}, \boldsymbol{w} \in \operatorname{dom}(F)\right\}. \tag{27}$$

For any sequence $\boldsymbol{w}_k$ which is in the closure of the feasible set $\operatorname{cl}(\mathcal{S}_{\boldsymbol{w}})$ and satisfies $\|\boldsymbol{w}_k\| \to +\infty$, one has $\sum_{t=1}^{p(p-1)/2} [\boldsymbol{w}_k]_t \to +\infty$. Then one can establish that

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

holds if $\alpha > 0$. Therefore, $F(\boldsymbol{w})$ is coercive over $\operatorname{cl}(\mathcal{S}_{\boldsymbol{w}})$ almost surely. By the Extreme Value Theorem (Drábek and Milota, 2007), if $\Omega \subset \mathbb{R}^q$ is non-empty and closed, and $g : \Omega \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous and coercive, then the optimization $\min_{\boldsymbol{x}\in\Omega} g(\boldsymbol{x})$ has at least one global minimizer. Therefore, by the coercivity of $F(\boldsymbol{w})$, the optimization (24) has at least one global minimizer in $\operatorname{cl}(\mathcal{S}_{\boldsymbol{w}})$ almost surely.

We can rewrite $\mathcal{S}_{\boldsymbol{w}} = \Omega_A \cap \Omega_B$, where $\Omega_A := \{\boldsymbol{w} \in \mathbb{R}^{p(p-1)/2} \,|\, \boldsymbol{w} \geq \boldsymbol{0}\}$ and $\Omega_B := \{\boldsymbol{w} \in \mathbb{R}^{p(p-1)/2} \,|\, (\mathcal{L}\boldsymbol{w}+\boldsymbol{J}) \in \mathcal{S}_{++}^p\}$. Notice that Ω_A is a closed set, while Ω_B is an open set. Consider the set $V := \operatorname{cl}(\mathcal{S}_{\boldsymbol{w}}) \setminus \mathcal{S}_{\boldsymbol{w}}$, we have

$$V \subseteq \{\operatorname{cl}(\Omega_A) \cap \operatorname{cl}(\Omega_B)\} \setminus \{\Omega_A \cap \Omega_B\} = \Omega_A \cap \partial\Omega_B, \tag{28}$$

where $\partial\Omega_B$ is the boundary of Ω_B. Note that every matrix on the boundary of the set of positive definite matrices is positive semi-definite and has zero determinant. Hence, one has $\partial\Omega_B = \{\boldsymbol{w} \in \mathbb{R}^{p(p-1)/2} \,|\, (\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) \in \mathcal{S}_+^p,\ \det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) = 0\}$. As a result, for any $\boldsymbol{w}_k \in \operatorname{cl}(\mathcal{S}_{\boldsymbol{w}}) \setminus \mathcal{S}_{\boldsymbol{w}}$, $F(\boldsymbol{w}_k) = +\infty$. Therefore, (24) has at least one global minimizer in the set $\mathcal{S}_{\boldsymbol{w}}$ almost surely.

The uniqueness of the minimizer is established by proving that the optimization (24) is strictly convex. For any $\boldsymbol{w} \in \mathcal{S}_{\boldsymbol{w}}$, the minimum eigenvalue of $\nabla^2 F(\boldsymbol{w})$ can be bounded from the below as follows

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

where the second equality is obtained by calculating the Hessian $\nabla^2 F(\boldsymbol{w})$; the second inequality is based on the property of Kronecker product that the eigenvalues of $\boldsymbol{A} \otimes \boldsymbol{B}$ are $\lambda_i \mu_j$ for $i, j \in [p]$, in which λ_i and μ_j are the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively; the last inequality is establised by the facts that $\lambda_{\min}(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) > 0$ for any $\boldsymbol{w} \in \mathcal{S}_{\boldsymbol{w}}$, and $\|\mathcal{L}\boldsymbol{x}\|_{\mathrm{F}}^2 > 0$ for any $\boldsymbol{x} \neq \boldsymbol{0}$. Therefore, the optimization (24) is strictly convex, and thus (24) has at most one global minimizer.

Combining the existence and uniqueness, we conclude that the minimizer of the optimization (24) exists and is unique almost surely as long as $n \geq 1$. □

A.2 Proof of Theorem 2

Proof. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be *i.i.d.* samples drawn from a p-dimensional L-GMRF with the parameters $(\boldsymbol{0}, \boldsymbol{\Omega})$, and the sample size n is lower bounded by

$$n \geq \max\left(8c_0^{-1} c_d \left(\|\boldsymbol{A}_{\mathcal{S}}\|_{\max} + \|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}\right)^2 M^2 |\mathcal{S}| \log p, 8c_d \|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}^2 \log p\right), \tag{29}$$

where $c_0 = 1/\left(8\left\|\mathcal{L}^*(\boldsymbol{\Omega}+\boldsymbol{J})^{-1}\right\|_{\max}^2\right)$, and $c_d \geq (d+2)\left\|\boldsymbol{A}_{\mathcal{S}^c}\right\|_{\min}^{-2}$ with a constant $d>0$.

Recall that the parameter space of the Laplacian constrained precision matrices we consider is $\mathcal{F}(d_{n,p}, M)$ defined in (3.1.3). For any $\boldsymbol{\Omega} \in \mathcal{F}(d_{n,p}, M)$, there exists an unique $\boldsymbol{w}$ such that $\boldsymbol{\Omega} = \mathcal{L}\boldsymbol{w}$. Therefore, we can obtain an equivalent parameter space of $\mathcal{F}(d_{n,p}, M)$ with the form of $\boldsymbol{w}$,

$$\mathcal{F}'(d_{n,p}, M) = \left\{ \begin{array}{c} \boldsymbol{w} \in \mathbb{R}^{p(p-1)/2} | \max_i \sum_{j=1}^{p} I\left\{[\mathcal{L}\boldsymbol{w}]_{ij} \neq 0\right\} \leq d_{n,p}, \\ \frac{1}{M} \leq \lambda_2(\mathcal{L}\boldsymbol{w}) \leq \lambda_{\max}(\mathcal{L}\boldsymbol{w}) \leq M \end{array} \right\},$$

For any given $\boldsymbol{w}^\star \in \mathcal{F}'(d_{n,p}, M)$, we define a local region around $\boldsymbol{w}^\star$ by

$$\mathcal{B}\left(\boldsymbol{w}^\star; \lambda_{\max}\left(\mathcal{L}\boldsymbol{w}^\star\right)\right) = \left\{\boldsymbol{w}\,|\boldsymbol{w} \in \mathbb{B}\left(\boldsymbol{w}^\star; \lambda_{\max}\left(\mathcal{L}\boldsymbol{w}^\star\right)\right) \cap \mathcal{S}_{\boldsymbol{w}}\right\},$$

where $\mathbb{B}(\boldsymbol{w}^\star; r) = \{\boldsymbol{w} \in \mathbb{R}^{p(p-1)/2} \,|\, \|\mathcal{L}\boldsymbol{w} - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}} \leq r\}$ and $\mathcal{S}_{\boldsymbol{w}}$ is defined in (27). It is easy to verify that $\boldsymbol{w}^\star \in \mathcal{B}(\boldsymbol{w}^\star; \lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star))$.

Next, we prove that $\hat{\boldsymbol{w}} \in \mathcal{B}(\boldsymbol{w}^\star; \lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star))$, where $\hat{\boldsymbol{w}}$ is the proposed estimator defined in (8). We can see the optimization problems (8) and (24) have the same feasible set. Therefore, $\mathcal{S}_{\boldsymbol{w}}$ is also the feasible set of (24) and thus $\hat{\boldsymbol{w}} \in \mathcal{S}_{\boldsymbol{w}}$ must hold. We construct an intermediate estimator,

$$\boldsymbol{w}_t = \boldsymbol{w}^\star + t\left(\hat{\boldsymbol{w}} - \boldsymbol{w}^\star\right), \tag{30}$$

where t is taken such that $\|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}} = \lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star)$ if $\|\mathcal{L}\hat{\boldsymbol{w}} - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}} \geq \lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star)$, and $t=1$ otherwise. Hence $\|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}} \leq \lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star)$ always holds and $t \in [0,1]$. One further has $\boldsymbol{w}_t \in \mathcal{S}_{\boldsymbol{w}}$ because both $\boldsymbol{w}^\star, \hat{\boldsymbol{w}} \in \mathcal{S}_{\boldsymbol{w}}$ and $\mathcal{S}_{\boldsymbol{w}}$ is a convex set as shown in the following. For any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S}_{\boldsymbol{w}}$, define $\boldsymbol{x}_t = t\boldsymbol{x}_1 + (1-t)\boldsymbol{x}_2$, $t \in [0,1]$. It is clear that $\boldsymbol{x}_t \geq \mathbf{0}$. Since $\mathcal{S}_{++}^p$ is a convex cone, one has

$$\mathcal{L}\boldsymbol{x}_t + \boldsymbol{J} = t(\mathcal{L}\boldsymbol{x}_1 + \boldsymbol{J}) + (1-t)(\mathcal{L}\boldsymbol{x}_2 + \boldsymbol{J}) \in \mathcal{S}_{++}^p, \tag{31}$$

indicating that $\boldsymbol{x}_t \in \mathcal{S}_{\boldsymbol{w}}$ and thus the set $\mathcal{S}_{\boldsymbol{w}}$ is convex. Hence $\mathcal{B}(\boldsymbol{w}^\star; r)$ is a convex set. Therefore, we conclude that $\boldsymbol{w}_t \in \mathcal{B}(\boldsymbol{w}^\star; \lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star))$.

The following lemma characterizes the local region around $\boldsymbol{w}^\star$.

Lemma 3. *(Ying et al., 2020a) Let $f(\boldsymbol{w}) = -\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J})$. Then for any $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathcal{B}(\boldsymbol{w}^\star; r)$ defined in* (30)*, we have*

$$\langle \nabla f(\boldsymbol{w}_1) - \nabla f(\boldsymbol{w}_2), \boldsymbol{w}_1 - \boldsymbol{w}_2 \rangle \geq \left(\|\mathcal{L}\boldsymbol{w}^\star\|_2 + r\right)^{-2} \|\mathcal{L}\boldsymbol{w}_1 - \mathcal{L}\boldsymbol{w}_2\|_{\mathrm{F}}^2.$$

Applying Lemma 3 with $\boldsymbol{w}_1 = \boldsymbol{w}_t$, $\boldsymbol{w}_2 = \boldsymbol{w}^*$ and $r = \lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star)$ yields

$$t\langle -\mathcal{L}^*\left(\mathcal{L}\boldsymbol{w}_t + \boldsymbol{J}\right)^{-1} + \mathcal{L}^*\left(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J}\right)^{-1},\ \hat{\boldsymbol{w}} - \boldsymbol{w}^* \rangle \geq \left(2\lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star)\right)^{-2} \|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^*\|_{\mathrm{F}}^2. \tag{32}$$

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

$$q'(a) = \langle -\mathcal{L}^*\left(\mathcal{L}\boldsymbol{w}_a + \boldsymbol{J}\right)^{-1} + \mathcal{L}^*\left(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J}\right)^{-1},\ \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \rangle, \tag{33}$$

and

$$q''(a) = \left\langle \mathcal{L}^*\left(\left(\mathcal{L}\boldsymbol{w}_a + \boldsymbol{J}\right)^{-1}\left(\mathcal{L}\hat{\boldsymbol{w}} - \mathcal{L}\boldsymbol{w}^\star\right)\left(\mathcal{L}\boldsymbol{w}_a + \boldsymbol{J}\right)^{-1}\right),\ \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \right\rangle = \mathrm{tr}\left(\boldsymbol{ABAB}\right),$$

where $\boldsymbol{w}_a = \boldsymbol{w}^\star + a(\hat{\boldsymbol{w}} - \boldsymbol{w}^\star)$, $\boldsymbol{A} = (\mathcal{L}\boldsymbol{w}_a + \boldsymbol{J})^{-1}$ and $\boldsymbol{B} = (\mathcal{L}\hat{\boldsymbol{w}} - \mathcal{L}\boldsymbol{w}^\star)$. Note that $\boldsymbol{A}$ is symmetric and positive definite because $\boldsymbol{w}_t \in \mathcal{S}_{\boldsymbol{w}}$ and $\boldsymbol{B}$ is symmetric. Let $\boldsymbol{C} = \boldsymbol{AB}$. By Theorem 1 in Drazin and Haynsworth (1962), we know that all the eigenvalues of a matrix $\boldsymbol{X} \in \mathbb{R}^{p\times p}$ are real if there exists a symmetric and positive definite matrix $\boldsymbol{Y} \in \mathbb{R}^{p\times p}$ such that $\boldsymbol{XY}$ are symmetric. We can see the matrix $\boldsymbol{CA}$ is symmetric with $\boldsymbol{A}$ symmetric and positive definite, and thus all the eigenvalues of $\boldsymbol{C}$ are real. Suppose $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\boldsymbol{C}$. Then the eigenvalues of $\boldsymbol{CC}$ are $\lambda_1^2, \ldots, \lambda_p^2$. 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

$$t\langle \mathcal{L}^*\left(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J}\right)^{-1} - \mathcal{L}^*\left(\mathcal{L}\hat{\boldsymbol{w}} + \boldsymbol{J}\right)^{-1}, \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \rangle = tq'(1) \geq tq'(t) \geq \left(2\lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star)\right)^{-2} \|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^*\|_{\mathrm{F}}^2. \tag{34}$$

where the first inequality holds because $q'(a)$ is non-decreasing and $t \leq 1$, and the second inequality follows from (32).

The Lagrangian of the optimization (8) is

$$L(\boldsymbol{w}, \boldsymbol{\nu}) = -\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) + \operatorname{tr}(\mathcal{L}\boldsymbol{w}\boldsymbol{S}) + \lambda \boldsymbol{a}^\top \boldsymbol{w} - \boldsymbol{v}^\top \boldsymbol{w},$$

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

$$\begin{cases} -\mathcal{L}^*\left((\mathcal{L}\hat{\boldsymbol{w}} + \boldsymbol{J})^{-1}\right) + \mathcal{L}^*\boldsymbol{S} + \lambda\boldsymbol{a} - \hat{\boldsymbol{v}} = \boldsymbol{0}; \\ \hat{\boldsymbol{v}}^\top \hat{\boldsymbol{w}} = \boldsymbol{0}, \ \hat{\boldsymbol{w}} \geq \boldsymbol{0}, \ \hat{\boldsymbol{v}} \geq \boldsymbol{0}; \end{cases} \tag{35}$$

According to the first condition in (35), one has

$$\left\langle -\mathcal{L}^*\left(\mathcal{L}\hat{\boldsymbol{w}} + \boldsymbol{J}\right)^{-1} + \mathcal{L}^*\boldsymbol{S},\, \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \right\rangle = \langle \hat{\boldsymbol{v}} - \lambda\boldsymbol{a}, \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \rangle. \tag{36}$$

Substituting (36) into (34) yields

$$\|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}}^2 = 4t\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)\left(\underbrace{\langle \hat{\boldsymbol{v}}, \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \rangle}_{I_1} - \underbrace{\langle \lambda\boldsymbol{a}, \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \rangle}_{I_2} + \underbrace{\left\langle \mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right), \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \right\rangle}_{I_3}\right). \tag{37}$$

Next we bound term I_1, I_2 and I_3, respectively. The term I_1 can be directly bounded by

$$\langle \hat{\boldsymbol{v}}, \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \rangle = -\langle \hat{\boldsymbol{v}}, \boldsymbol{w}^\star \rangle \leq 0, \tag{38}$$

which follows from the second condition in (35) and the fact that $\boldsymbol{w}^\star \geq \boldsymbol{0}$.

For term I_2, we separate the support of $\boldsymbol{a}$ into two parts, $\mathcal{S}$ and its complementary set $\mathcal{S}^c$, where $\mathcal{S}$ is the support of $\boldsymbol{w}^\star$. Let $|\mathcal{S}| = s$. A simple algebra yields

$$\langle \lambda\boldsymbol{a}, \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \rangle = \langle \lambda\boldsymbol{a}_{\mathcal{S}}, (\hat{\boldsymbol{w}} - \boldsymbol{w}^\star)_{\mathcal{S}} \rangle + \langle \lambda\boldsymbol{a}_{\mathcal{S}^c}, (\hat{\boldsymbol{w}} - \boldsymbol{w}^\star)_{\mathcal{S}^c} \rangle \geq -\lambda\|\boldsymbol{a}_{\mathcal{S}}\|\,\|(\hat{\boldsymbol{w}} - \boldsymbol{w}^\star)_{\mathcal{S}}\| + \lambda\langle \boldsymbol{a}_{\mathcal{S}^c}, \hat{\boldsymbol{w}}_{\mathcal{S}^c} \rangle, \tag{39}$$

where the inequality follows from the Cauchy–Schwarz inequality and $\boldsymbol{w}^\star_{\mathcal{S}^c} = \boldsymbol{0}$.

For term I_3, we separate the support of $\mathcal{L}^*\big((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\big)$ into parts, $\mathcal{S}$ and $\mathcal{S}^c$. Then one has

$$\begin{aligned} \left\langle \mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right), \hat{\boldsymbol{w}} - \boldsymbol{w}^\star \right\rangle \leq &\left\|\left(\mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right)\right)_{\mathcal{S}}\right\| \|(\hat{\boldsymbol{w}} - \boldsymbol{w}^\star)_{\mathcal{S}}\| \\ &+ \left\langle \left(\mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right)\right)_{\mathcal{S}^c}, \hat{\boldsymbol{w}}_{\mathcal{S}^c} \right\rangle. \end{aligned} \tag{40}$$

Substituting (38), (39) and (40) into (37) yields

$$\begin{aligned} \|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}}^2 \leq &4t\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)\left(\left\|\left(\mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right)\right)_{\mathcal{S}}\right\| \|(\hat{\boldsymbol{w}} - \boldsymbol{w}^\star)_{\mathcal{S}}\|\right. \\ &\left. + \left\langle \left(\mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right)\right)_{\mathcal{S}^c} - \lambda\boldsymbol{a}_{\mathcal{S}^c}, \hat{\boldsymbol{w}}_{\mathcal{S}^c} \right\rangle + \lambda\|\boldsymbol{a}_{\mathcal{S}^\star}\|\,\|(\hat{\boldsymbol{w}} - \boldsymbol{w}^\star)_{\mathcal{S}^\star}\|\right). \end{aligned} \tag{41}$$

Define an event

$$\mathcal{J} = \left\{\left\|\mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right)\right\|_{\max} \leq \lambda\|\boldsymbol{a}_{\mathcal{S}^c}\|_{\min}\right\}.$$

Under the event $\mathcal{J}$, one obtains

$$\left\langle \left(\mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right)\right)_{\mathcal{S}^c} - \lambda\boldsymbol{a}_{\mathcal{S}^c}, \hat{\boldsymbol{w}}_{\mathcal{S}^c} \right\rangle \leq 0. \tag{42}$$

Combining (41) and (42) together yields

$$\|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}}^2 \leq 4t\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)\left(\lambda\|\boldsymbol{a}_{\mathcal{S}}\| + \left\|\left(\mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right)\right)_{\mathcal{S}}\right\|\right)\|\hat{\boldsymbol{w}} - \boldsymbol{w}^\star\|, \tag{43}$$

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

On the other hand, one has

$$\|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}} = t\,\|\mathcal{L}\hat{\boldsymbol{w}} - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}} \geq t\left(\sum_{i\neq j}\left[\mathcal{L}\hat{\boldsymbol{w}} - \mathcal{L}\boldsymbol{w}^\star\right]_{ij}^2\right)^{\frac{1}{2}} = \sqrt{2}t\,\|\hat{\boldsymbol{w}} - \boldsymbol{w}^\star\|\,. \tag{44}$$

Combining (43) and (44) together yields

$$\|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}} \leq 2\sqrt{2}\lambda_{\max}^2\left(\mathcal{L}\boldsymbol{w}^\star\right)\left(\lambda\,\|\boldsymbol{a}_{\mathcal{S}}\| + \left\|\left(\mathcal{L}^*\left(\left(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J}\right)^{-1} - \boldsymbol{S}\right)\right)_{\mathcal{S}}\right\|\right). \tag{45}$$

One also has

$$\left\|\left(\mathcal{L}^*\left(\left(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J}\right)^{-1} - \boldsymbol{S}\right)\right)_{\mathcal{S}}\right\| \leq \left(|\mathcal{S}|\ \left\|\mathcal{L}^*\left(\left(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J}\right)^{-1} - \boldsymbol{S}\right)\right\|_{\max}^2\right)^{\frac{1}{2}} \leq \sqrt{s}\lambda\,\|\boldsymbol{a}_{\mathcal{S}^c}\|_{\min}. \tag{46}$$

Plugging (46) and the inequality $\|\boldsymbol{a}_{\mathcal{S}}\| \leq \sqrt{s}\,\|\boldsymbol{a}_{\mathcal{S}}\|_{\max}$ into (45), one has

$$\|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}} \leq 2\sqrt{2}\lambda_{\max}^2\left(\mathcal{L}\boldsymbol{w}^\star\right)\left(\|\boldsymbol{a}_{\mathcal{S}}\|_{\max} + \|\boldsymbol{a}_{\mathcal{S}^c}\|_{\min}\right)\sqrt{s}\lambda. \tag{47}$$

Note that $\|\boldsymbol{A}_{\mathcal{S}}\|_{\max} = \|\boldsymbol{a}_{\mathcal{S}}\|_{\max}$ and $\|\boldsymbol{A}_{\mathcal{S}}\|_{\min} = \|\boldsymbol{a}_{\mathcal{S}}\|_{\min}$. By taking $\lambda = \sqrt{c_0^{-1}c_d \log p/n}$ and $n \geq 8c_0^{-1}c_1c_d\left(\|\boldsymbol{A}_{\mathcal{S}}\|_{\max} + \|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}\right)^2 s\log p$, one has

$$2\sqrt{2}\lambda_{\max}\left(\mathcal{L}\boldsymbol{w}^\star\right)\left(\|\boldsymbol{A}_{\mathcal{S}}\|_{\max} + \|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}\right)\sqrt{s}\lambda < 1. \tag{48}$$

Combining (47) and (48), one has

$$\|\mathcal{L}\boldsymbol{w}_t - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}} < \lambda_{\max}\left(\mathcal{L}\boldsymbol{w}^\star\right). \tag{49}$$

Thus $t = 1$ in (30), and $\boldsymbol{w}_t = \hat{\boldsymbol{w}}$. Therefore, under the event $\mathcal{J}$, together with the fact that $s \leq d_{n,p}p$, we conclude that

$$\frac{1}{p}\,\|\mathcal{L}\hat{\boldsymbol{w}} - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}}^2 \lesssim \frac{d_{n,p}\log p}{n}. \tag{50}$$

Finally, we compute the probability that the event $\mathcal{J}$ holds using the following lemma.

Lemma 4. *(Ying et al., 2020a) Consider a zero-mean random vector $\boldsymbol{x} = [x_1, \ldots, x_p]^\top \in \mathbb{R}^p$ is a L-GMRF with precision matrix $\mathcal{L}\boldsymbol{w}^\star \in \mathcal{S}_L$. Given n i.i.d samples $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}$, the associated sample covariance matrix $\boldsymbol{S} = \frac{1}{n}\sum_{k=1}^n \boldsymbol{x}^{(k)}\left(\boldsymbol{x}^{(k)}\right)^\top$ satisfies, for $t \in [0, t_0]$,*

$$\mathbb{P}\left[\,\left|[\mathcal{L}^*\boldsymbol{S}]_i - \left(\mathcal{L}^*(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1}\right)_i\right| \geq t\right] \leq 2\exp(-c_0nt^2), \quad \textit{for } i \in [p(p-1)/2],$$

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

By applying Lemma 4 with $t = \lambda\,\|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}$ and union sum bound, then get

$$\mathbb{P}\left[\left\|\mathcal{L}^*\left(\left(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J}\right)^{-1} - \boldsymbol{S}\right)\right\|_{\max} \geq \lambda\,\|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}\right] \leq p^2\exp\left(-c_0n\lambda^2\,\|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}^2\right),$$

for any $\lambda\,\|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min} \leq t_0$, where $t_0 = \left\|\mathcal{L}^*(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1}\right\|_{\max}$ and $c_0 = 1/\left(8\left\|\mathcal{L}^*(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1}\right\|_{\max}^2\right)$. Take $\lambda = \sqrt{c_0^{-1}c_d\log p/n}$, where $c_d \geq (d+2)\,\|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}^{-2}$ with a constant $d > 0$. To ensure $\lambda\,\|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min} \leq t_0$, one takes $n \geq 8c_d\log p\,\|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}^2$. By calculation, we establish

$$\mathbb{P}\left[\left\|\mathcal{L}^*\left(\left(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J}\right)^{-1} - \boldsymbol{S}\right)\right\|_{\max} \leq \lambda\,\|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}\right] \geq 1 - p^2\exp\left(-c_0n\lambda^2\,\|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}^2\right) \geq 1 - p^{-d}.$$

Note that the inequality (50) holds for any $\boldsymbol{w}^\star \in \mathcal{F}'(d_{n,p}, M)$. By taking $\boldsymbol{\Omega} = \mathcal{L}\boldsymbol{w}^\star$ and $\widehat{\boldsymbol{\Omega}} = \mathcal{L}\hat{\boldsymbol{w}}$, we obtain

$$\inf_{\boldsymbol{\Omega}\in\mathcal{F}(d_{n,p},M)}\mathbb{P}\left\{\frac{1}{p}\,\|\mathcal{L}\hat{\boldsymbol{w}} - \mathcal{L}\boldsymbol{w}^\star\|_{\mathrm{F}}^2 \lesssim \frac{d_{n,p}\log p}{n}\right\} \geq 1 - p^{-d},$$

for some constant $C > 0$.

□