Does the l1 -norm Learn a Sparse Graph under Laplacian Constrained Graphical Models?

Jiaxi Ying
Jos´e Vin´ıcius de M. Cardoso
Daniel P. Palomar

Model

connected, connected graphs

$$\mathcal{S}_L = \{\boldsymbol{\Theta} \in \mathcal{S}_+^p \,|\, \Theta_{ij} = \Theta_{ji} \leq 0,\, \forall\, i \neq j,\, \boldsymbol{\Theta} \cdot \mathbf{1} = \mathbf{0},\ \operatorname{rank}(\boldsymbol{\Theta}) = p - 1\},$$

Definition 2.1. A zero-mean random vector $\boldsymbol{x} = [x_1, \dots, x_p]^\top \in V^{p-1}$ is called a Laplacian constrained Gaussian Markov Random Fields (L-GMRF) with parameters $(\mathbf{0}, \boldsymbol{\Theta})$ with $\boldsymbol{\Theta} \in \mathcal{S}_L$, if and only if its density function $q_L : V^{p-1} \to \mathbb{R}$ follows

$$q_L(\boldsymbol{x}) = (2\pi)^{-\frac{p-1}{2}} \det{}^\star(\boldsymbol{\Theta})^{\frac{1}{2}} \exp\Big(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Theta} \boldsymbol{x}\Big), \tag{3}$$

$$V^{p-1} := \left\{\boldsymbol{x} \in \mathbb{R}^p \mid \mathbf{1}^\top \boldsymbol{x} = 0\right\}$$

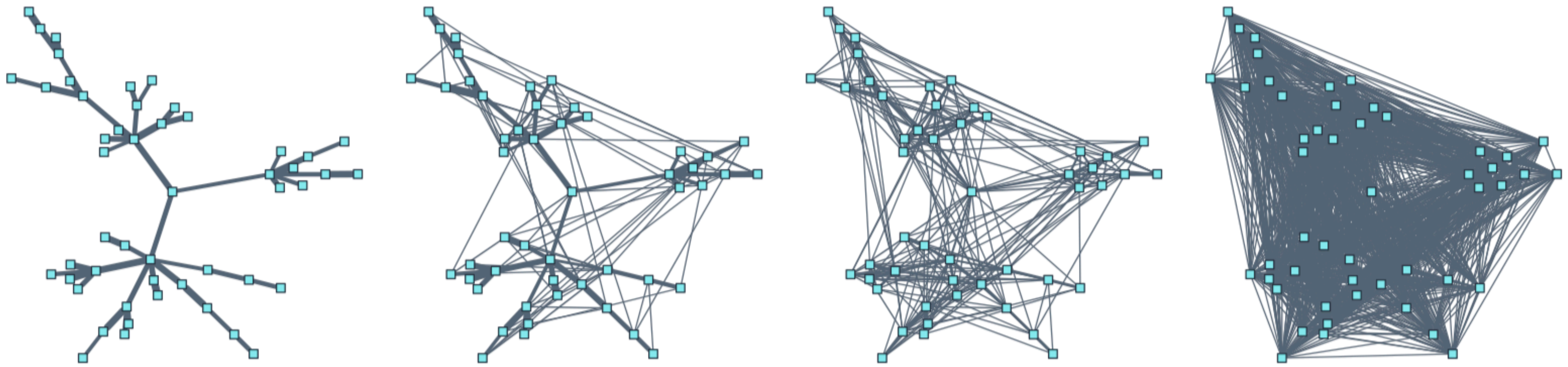
$$\min_{\boldsymbol{\Theta} \in \mathcal{S}_L} -\log\det(\boldsymbol{\Theta} + \boldsymbol{J}) + \operatorname{tr}(\boldsymbol{\Theta}\boldsymbol{S}) + \sum_{i>j} h_\lambda(\Theta_{ij}),$$

Motivations

$$\min_{\boldsymbol{\Theta} \in \mathcal{S}_L} -\log\det(\boldsymbol{\Theta} + \boldsymbol{J}) + \operatorname{tr}(\boldsymbol{\Theta}\boldsymbol{S}) + \lambda \sum |\Theta_{ij}|,$$

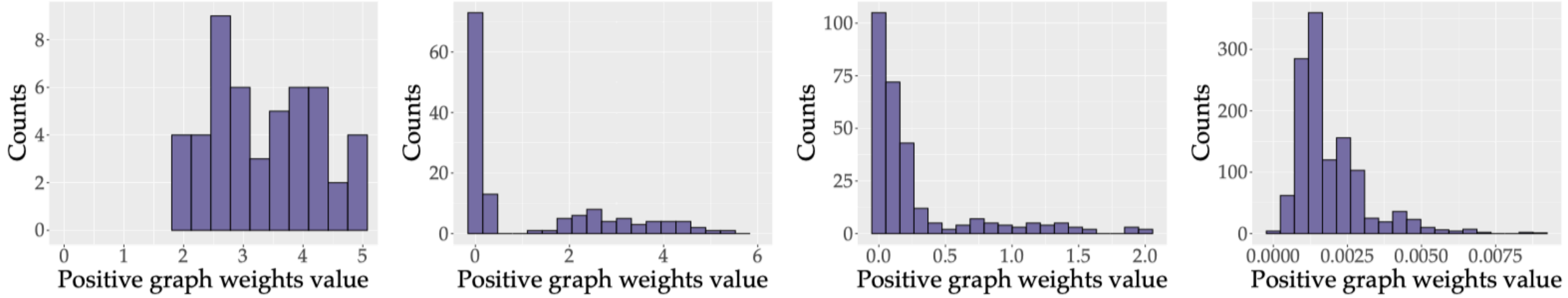
Theorem 3.1. *Let $\hat{\boldsymbol{\Theta}} \in \mathbb{R}^{p \times p}$ be the global minimum of* $\boxed{5}$ *with $p > 3$. Define $s_1 = \max_k S_{kk}$ and $s_2 = \min_{ij} S_{ij}$. If the regularization parameter λ in* $\boxed{5}$ *satisfies $\lambda \in [(2 + 2\sqrt{2})(p+1)(s_1 - s_2), +\infty)$, then the estimated graph weight $\hat{W}_{ij} = -\hat{\Theta}_{ij}$ obeys*

$$\hat{W}_{ij} \geq \frac{1}{(s_1 - (p+1)s_2 + \lambda)p} > 0, \quad \forall\, i \neq j.$$



(a) Ground-truth (b) $\lambda = 0$ (c) $\lambda = 0.1$ (d) $\lambda = 10$

Figure 1: Graph learning using ℓ_1-norm regularization with different regularization parameters. The number of positive edges in (a), (b), (c) and (d) are 49, 135, 286 and 1225, respectively. The graph in (d) is fully connected. The relative errors of the learned graphs in (b), (c) and (d) are 0.14, 0.64 and 0.99, respectively.



(a) Ground-truth (b) $\lambda = 0$ (c) $\lambda = 0.1$ (d) $\lambda = 10$

Figure 2: Histograms of nonzero graph weights learned by the ℓ_1-norm regularization method with different regularization parameters, corresponding to the graphs learned in Figure $\boxed{1}$. The histograms count the number of nonzero graph weights falling into each interval.

Simplify the Laplacian structural constraints

Definition 3.2. The linear operator $\mathcal{L} : \mathbb{R}^{p(p-1)/2} \to \mathbb{R}^{p\times p}$, $\boldsymbol{x} \mapsto \mathcal{L}\boldsymbol{x}$, is defined by

$$[\mathcal{L}\boldsymbol{x}]_{ij} = \begin{cases} -x_k & i > j, \\ [\mathcal{L}\boldsymbol{x}]_{ji} & i < j, \\ -\sum_{j\neq i}[\mathcal{L}\boldsymbol{x}]_{ij} & i = j, \end{cases} \tag{7}$$

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

A simple example is given below which illustrates the definition of the operator $\mathcal{L}$. Let $\boldsymbol{x} \in \mathbb{R}^3$. Then we have

$$\mathcal{L}\boldsymbol{x} = \begin{bmatrix} \sum_{i=1,2} x_i & -x_1 & -x_2 \\ -x_1 & \sum_{i=1,3} x_i & -x_3 \\ -x_2 & -x_3 & \sum_{i=2,3} x_i \end{bmatrix}.$$

The adjoint operator $\mathcal{L}^*$ of $\mathcal{L}$ is defined so as to satisfy $\langle \mathcal{L}\boldsymbol{x}, \boldsymbol{Y}\rangle = \langle \boldsymbol{x}, \mathcal{L}^*\boldsymbol{Y}\rangle$, $\forall \boldsymbol{x} \in \mathbb{R}^{p(p-1)/2}$ and $\boldsymbol{Y} \in \mathbb{R}^{p\times p}$.

Simplify the Laplacian structural constraints

Definition 3.3. The adjoint operator $\mathcal{L}^* : \mathbb{R}^{p\times p} \to \mathbb{R}^{p(p-1)/2}$, $\boldsymbol{Y} \mapsto \mathcal{L}^*\boldsymbol{Y}$, is defined by

$$[\mathcal{L}^*\boldsymbol{Y}]_k = Y_{i,i} - Y_{i,j} - Y_{j,i} + Y_{j,j}, \tag{8}$$

where $i, j \in [p]$ obeying $k = i - j + \frac{j-1}{2}(2p - j)$ and $i > j$.

By introducing the linear operator $\mathcal{L}$, we can simplify the definition of $\mathcal{S}_L$ in ($\boxed{2}$) as below.

Theorem 3.4. *The Laplacian set $\mathcal{S}_L$ defined in ($\boxed{2}$) can be written as*

$$\mathcal{S}_L = \left\{\mathcal{L}\boldsymbol{x} \middle| \; \boldsymbol{x} \geq \boldsymbol{0}, \; (\mathcal{L}\boldsymbol{x} + \boldsymbol{J}) \in \mathcal{S}_{++}^p \right\}. \tag{9}$$

where $\boldsymbol{J} = \frac{1}{p}\boldsymbol{1}_{p\times p}$ and $\boldsymbol{x} \geq \boldsymbol{0}$ means every entry of $\boldsymbol{x}$ is non-negative.

$$\mathcal{S}_L = \{\boldsymbol{\Theta} \in \mathcal{S}_+^p \,|\, \Theta_{ij} = \Theta_{ji} \leq 0, \forall\, i \neq j, \; \boldsymbol{\Theta} \cdot \boldsymbol{1} = \boldsymbol{0}, \; \operatorname{rank}(\boldsymbol{\Theta}) = p - 1\},$$

Original:

$$\min_{\boldsymbol{\Theta} \in \mathcal{S}_L} -\log\det(\boldsymbol{\Theta} + \boldsymbol{J}) + \operatorname{tr}(\boldsymbol{\Theta}\boldsymbol{S}) + \sum_{i>j} h_\lambda(\Theta_{ij}),$$

Modified:

$$\min_{\boldsymbol{w} \geq \boldsymbol{0}} -\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) + \operatorname{tr}(\boldsymbol{S}\mathcal{L}\boldsymbol{w}) + \sum_{i} h_\lambda(w_i).$$

MM Framework

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

linearizing ∑ ih λ(w i)

$$f_k(\boldsymbol{w}) = -\log\det(\mathcal{L}\boldsymbol{w}+\boldsymbol{J})+\operatorname{tr}(\boldsymbol{S}\mathcal{L}\boldsymbol{w})+\sum_i h'_\lambda(\hat{w}_i{}^{(k-1)})w_i, \tag{14}$$

By minimizing $f_k(\boldsymbol{w})$, we establish a sequence $\{\hat{\boldsymbol{w}}^{(k)}\}_{k\geq 1}$ by

$$\hat{\boldsymbol{w}}^{(k)} = \arg\min_{\boldsymbol{w}\geq \boldsymbol{0}} -\log\det(\mathcal{L}\boldsymbol{w}+\boldsymbol{J})+\operatorname{tr}(\boldsymbol{S}\mathcal{L}\boldsymbol{w})+\sum_i z_i^{(k-1)}w_i, \tag{15}$$

Algorithm

Algorithm 1 Nonconvex Graph Learning (NGL)

Input: Sample covariance $\boldsymbol{S}$, λ, $\hat{\boldsymbol{w}}^{(0)}$;

$k \leftarrow 1$;

1: **while** Stopping criteria not met **do**

2: Update $z_i^{(k-1)} = h'_\lambda(\hat{w}_i^{(k-1)})$, for $i = 1, \ldots, p(p-1)/2$;

3: Update $\hat{\boldsymbol{w}}^{(k)} = \arg\min_{\boldsymbol{w} \geq \boldsymbol{0}} - \log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) + \operatorname{tr}(\boldsymbol{S}\mathcal{L}\boldsymbol{w}) + \sum_i z_i^{(k-1)} w_i$;

4: $k \leftarrow k + 1$;

5: **end while**

Output: $\hat{\boldsymbol{w}}$.

projected gradient descent (k) with backtracking line search

Algorithm 2 Update $\hat{\boldsymbol{w}}^{(k)}$

Input: Sample covariance $\boldsymbol{S}$, λ, $\hat{\boldsymbol{w}}^{(k-1)}$, β;

$\boldsymbol{w}_0^{(k)} = \hat{\boldsymbol{w}}^{(k-1)}$;

$t \leftarrow 1$;

1: **while** Stopping criteria not met **do**

2: Update $\boldsymbol{w}_t^{(k)} = \mathcal{P}_+\big(\boldsymbol{w}_{t-1}^{(k)} - \eta \nabla f_k(\boldsymbol{w}_{t-1}^{(k)})\big)$;

3: **if** $f_k(\boldsymbol{w}_t^{(k)}) > f_k(\boldsymbol{w}_{t-1}^{(k)}) + \langle \nabla f_k(\boldsymbol{w}_{t-1}^{(k)}), \boldsymbol{w}_t^{(k)} - \boldsymbol{w}_{t-1}^{(k)} \rangle + \frac{1}{2\eta} \left\| \boldsymbol{w}_t^{(k)} - \boldsymbol{w}_{t-1}^{(k)} \right\|^2$ **then**

4: $\eta \leftarrow \beta\eta$;

5: Back to Step 2;

6: **end if**

7: $t \leftarrow t + 1$;

8: **end while**

Output: $\hat{\boldsymbol{w}}^{(k)}$.

$$\text{where } \mathcal{P}_+(a) = \max(a, 0) \text{ and } \nabla f_k(\boldsymbol{w}_{t-1}^{(k)}) = -\mathcal{L}^*(\mathcal{L}\boldsymbol{w}_{t-1}^{(k)} + \boldsymbol{J})^{-1} + \mathcal{L}^*\boldsymbol{S} + \boldsymbol{z}^{(k-1)}.$$

Theoretical Results

the true graph weights $\boldsymbol{w}^\star \in \mathbb{R}^{p(p-1)/2}$ $\boldsymbol{w}^\star \geq \mathbf{0}$ $\mathscr{L}\boldsymbol{w}^\star \in \mathcal{S}_L$

$\mathcal{S}^\star = \left\{ i \in [p(p-1)/2] \mid w_i^\star > 0 \right\}$ the support set of w* and s be the number of the nonzero weights

Assumption 3.5. The function $h_\lambda : \mathbb{R} \to \mathbb{R}$ satisfies the following conditions:

1. $h_\lambda(0) = 0$, and $h'_\lambda(x)$ is monotone and Lipschitz continuous for $x \in [0, +\infty)$;
2. There exists a $\gamma > 0$ such that $h'_\lambda(x) = 0$ for $x \geq \gamma\lambda$;
3. $h'_\lambda(x) = \lambda$ for $x \leq 0$ and $h'_\lambda(c\lambda) \geq \lambda/2$, where $c = (2+\sqrt{2})\lambda^2_{\max}(\mathcal{L}\boldsymbol{w}^\star)$ is a constant.

Assumption 3.6. The minimal nonzero graph weight satisfies $\min_{i \in \mathcal{S}^\star} w_i^\star \geq (c+\gamma)\lambda \gtrsim \lambda$, where c and γ are defined in Assumption $\boxed{3.5}$. There exists $\tau \geq 1$ such that

$$1/\tau \leq \lambda_2(\mathcal{L}\boldsymbol{w}^\star) \leq \lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star) \leq \tau, \tag{17}$$

where $\lambda_2(\mathcal{L}\boldsymbol{w}^\star)$ and $\lambda_{\max}(\mathcal{L}\boldsymbol{w}^\star)$ are the second smallest eigenvalue and maximum eigenvalue of $\mathcal{L}\boldsymbol{w}^\star$, respectively. Note that the smallest eigenvalue of $\mathcal{L}\boldsymbol{w}^\star$ is 0.

Theoretical Results

Theorem 3.8. *Under Assumptions 3.5 and 3.6, take the regularization parameter* $\lambda = \sqrt{4\alpha c_0^{-1} \log p / n}$ *for some* $\alpha > 2$. *If the sample size* n *is lower bounded by*

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

then with probability at least $1 - 1/p^{\alpha-2}$, *the sequence* $\hat{\boldsymbol{w}}^{(k)}$ *returned by Algorithm 1 satisfies*

$$\left\|\hat{\boldsymbol{w}}^{(k)} - \boldsymbol{w}^\star\right\| \leq \underbrace{2(3\sqrt{2}+4)\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)\sqrt{\alpha c_0^{-1} s \log p / n}}_{\textit{Statistical error}} + \underbrace{\left(\frac{3}{2+\sqrt{2}}\right)^k \left\|\hat{\boldsymbol{w}}^{(0)} - \boldsymbol{w}^\star\right\|}_{\textit{Optimization error}},$$

where $c_0 = 1/\left(8\left\|\mathcal{L}^*(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1}\right\|_{\max}^2\right)$ *is a constant.*

Corollary 3.9. *Under the same assumptions and conditions as stated in Theorem 3.8, the sequence* $\hat{\boldsymbol{w}}^{(k)}$ *returned by Algorithm 1 satisfies*

$$\left\|\mathcal{L}\hat{\boldsymbol{w}}^{(k)} - \mathcal{L}\boldsymbol{w}^\star\right\|_{\mathrm{F}} \leq \underbrace{4(2\sqrt{2}+3)\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)\sqrt{\alpha c_0^{-1} s \log p / n}}_{\textit{Statistical error}} + \underbrace{\left(\frac{3}{2+\sqrt{2}}\right)^k \left\|\mathcal{L}\hat{\boldsymbol{w}}^{(0)} - \mathcal{L}\boldsymbol{w}^\star\right\|_{\mathrm{F}}}_{\textit{Optimization error}},$$

with probability at least $1 - 1/p^{\alpha-2}$. *If* $k \geq \lceil 4\log(4\alpha c_0^{-1}) \rceil$, *then the estimation error is dominated by the statistical error and we further obtain*

$$\left\|\mathcal{L}\hat{\boldsymbol{w}}^{(k)} - \mathcal{L}\boldsymbol{w}^\star\right\|_{\mathrm{F}} \lesssim \sqrt{s \log p / n},$$

where $c_0 = 1/\left(8\left\|\mathcal{L}^*(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1}\right\|_{\max}^2\right)$ *is a constant.*

Analysis of Algorithm Convergence

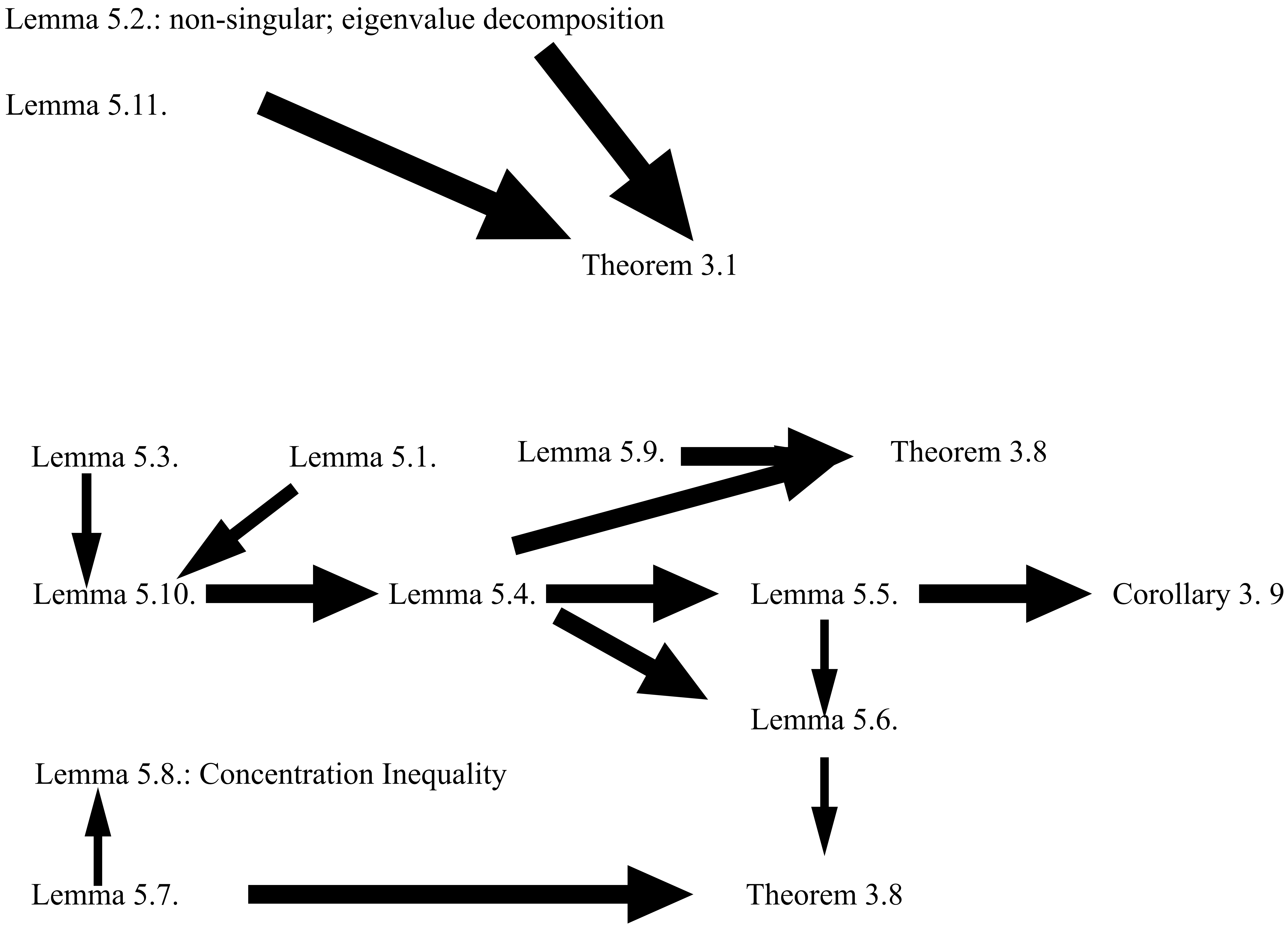
Theorem 3.10. *Under Assumptions* $\boxed{3.5}$ *and* $\boxed{3.6}$*, take the regularization parameter* $\lambda = \sqrt{4\alpha c_0^{-1} \log p / n}$ *for some* $\alpha > 2$*. If the sample size* n *satisfies*

$$n \geq c_0^{-1} \max\left(n \geq \max\left(840\alpha c_0^{-1} \frac{(\delta\tau^2+1)^4}{\delta^2\tau^2} sp \log p, 8\alpha \log p\right)\right),$$

then with probability at least $1 - 1/p^{\alpha-2}$*, the sequence* $\{\boldsymbol{w}_t^{(k)}\}_{t \geq 1}$ *returned by Algorithm* $\boxed{2}$ *obeys*

$$\left\|\boldsymbol{w}_t^{(k)} - \hat{\boldsymbol{w}}^{(k)}\right\|^2 \leq \rho^t \left\|\boldsymbol{w}_0^{(k)} - \hat{\boldsymbol{w}}^{(k)}\right\|^2, \quad \forall\, k \geq 2,$$

where $\rho = 1 - \frac{\beta(1-\delta^{-1})^2}{p\tau^4(1+\delta^{-1})^2} < 1$ *with* $\delta > 1$ *and* $\beta \in (0,1)$*, and* $c_0 = 1/\left(8\left\|\mathcal{L}^*(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1}\right\|_{\max}^2\right)$ *is a constant.*


Lemma 5.2.: non-singular; eigenvalue decomposition
Lemma 5.11.
Theorem 3.1
Lemma 5.3.
Lemma 5.1.
Lemma 5.9.
Theorem 3.8
Lemma 5.10.
Lemma 5.4.
Lemma 5.5.
Corollary 3. 9
Lemma 5.6.
Lemma 5.8.: Concentration Inequality
Lemma 5.7.
Theorem 3.8

Technical Lemmas

5.1 Technical Lemmas

Lemma 5.1. *Let* $f(\boldsymbol{w}) = -\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J})$, *with* $\boldsymbol{w} \in \mathbb{R}^{p(p-1)/2}$. *Then for any* $\boldsymbol{x} \in \mathbb{R}^{p(p-1)/2}$, *we have*

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

Lemma 5.2. *For any given* $\boldsymbol{w} \in \mathbb{R}^{p(p-1)/2}$ *satisfying* $(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) \in \mathcal{S}_{++}^p$, *there must exist an unique* $\boldsymbol{x} \in \mathbb{R}^{p(p-1)/2}$ *such that*

$$\mathcal{L}\boldsymbol{x} + \frac{1}{b}\boldsymbol{J} = (\mathcal{L}\boldsymbol{w} + b\boldsymbol{J})^{-1} \tag{19}$$

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

Lemma 5.3. *Let* $\mathcal{G} = \mathcal{L}^*\mathcal{L} : \mathbb{R}^{p(p-1)/2} \to \mathbb{R}^{p(p-1)/2}$, $\boldsymbol{x} \mapsto \mathcal{L}^*\mathcal{L}\boldsymbol{x}$. *For any* $\boldsymbol{x} \in \mathbb{R}^{p(p-1)/2}$, $\mathcal{G}\boldsymbol{x} = \boldsymbol{M}\boldsymbol{x}$ *with* $\boldsymbol{M} \in \mathbb{R}^{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}$ *satisfying*

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

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*

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

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

Technical Lemmas

Lemma 5.4. *Take* $\lambda = \sqrt{4\alpha c_0^{-1} \log p / n}$ *and suppose* $n \geq 94\alpha c_0^{-1} \lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star) s \log p$ *for some* $\alpha > 2$, *where* c_0 *is a constant defined in Lemma* $\boxed{5.8}$. *Let*

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

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

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

where $\mathcal{S}^\star$ *is the support of* $\boldsymbol{w}^\star$.

Lemma 5.5. *Take* $\lambda = \sqrt{4\alpha c_0^{-1} \log p / n}$ *and suppose* $n \geq 94\alpha c_0^{-1} \lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star) s \log p$ *for some* $\alpha > 2$, *where* c_0 *is a constant defined in Lemma* $\boxed{5.8}$. *Define the set* $\mathcal{E}^{(k)}$ *by*

$$\mathcal{E}^{(k)} = \{\mathcal{S}^\star \cup \mathcal{S}^{(k)}\}, \quad \textit{with} \quad \mathcal{S}^{(k)} = \{i \in [p(p-1)/2] \,|\, \hat{w}_i^{(k-1)} \geq b\}, \tag{20}$$

where $\hat{\boldsymbol{w}}^{(k)}$ *for* $k \geq 1$ *is defined in* ($\boxed{15}$), $\mathcal{S}^\star$ *is the support of* $\boldsymbol{w}^\star$ *with* $|\mathcal{S}^\star| \leq s$ *and* $b = (2+\sqrt{2})\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)\lambda$ *is a constant. Under Assumption* $\boxed{3.5}$, *if* $\left\| \mathcal{L}^*\big((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S} \big) \right\|_{\max} \leq \lambda/2$ *holds and* $\hat{\boldsymbol{w}}^{(0)}$ *satisfies* $|supp^+(\hat{\boldsymbol{w}}^{(0)})| \leq s$, *then* $\mathcal{E}^{(k)}$ *obeys* $|\mathcal{E}^{(k)}| \leq 2s$, *for any* $k \geq 1$.

Lemma 5.6. *Take* $\lambda = \sqrt{4\alpha c_0^{-1} \log p / n}$ *and suppose* $n \geq 94\alpha c_0^{-1} \lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star) s \log p$ *for some* $\alpha > 2$, *where* c_0 *is a constant defined in Lemma* $\boxed{5.8}$. *Under Assumptions* $\boxed{3.5}$ *and* $\boxed{3.6}$, *if* $\left\| \mathcal{L}^*\big((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S} \big) \right\|_{\max} \leq \lambda/2$ *holds and* $\hat{\boldsymbol{w}}^{(0)}$ *satisfies* $|supp^+(\hat{\boldsymbol{w}}^{(0)})| \leq s$, *then for any* $k \geq 1$, $\hat{\boldsymbol{w}}^{(k)}$ *defined in* ($\boxed{15}$) *obeys*

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

and

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

Technical Lemmas

Lemma 5.7. *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 5.8. Then one has*

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

Lemma 5.8. *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* $t \in [0, t_0]$,

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

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.*

Lemma 5.9. *Let* $f(\boldsymbol{w}) = -\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J})$. *Define a local region of* $\boldsymbol{w}^\star$ *by*

$$\mathcal{B}(\boldsymbol{w}^\star; r) = \{\boldsymbol{w} | \boldsymbol{w} \in \mathbb{B}(\boldsymbol{w}^\star; r) \cap \mathcal{S}_{\boldsymbol{w}}\}.$$

where $\mathbb{B}(\boldsymbol{w}^\star; r) = \{\boldsymbol{w} \in \mathbb{R}^{p(p-1)/2} \,|\, \|\boldsymbol{w} - \boldsymbol{w}^\star\| \leq r\}$, *and* $\mathcal{S}_{\boldsymbol{w}} = \{\boldsymbol{w}\,|\,\boldsymbol{w} \geq \boldsymbol{0}, (\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) \in \mathcal{S}_{++}^p\}$. *Then, under Assumption 3.6,* $g(\boldsymbol{w})$ *is* $\frac{2}{(1+\delta^{-1})^2\tau^2}$*-strongly convex and* $\frac{2p\tau^2}{(1-\delta^{-1})^2}$*-smooth in the region* $\mathcal{B}(\boldsymbol{w}^\star; \frac{1}{\sqrt{2}p\delta\tau})$ *where* τ *is defined in (17) and* $\delta > 1$. *In other words, for any* $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathcal{B}(\boldsymbol{w}^\star; \frac{1}{\sqrt{2}p\delta\tau})$, *we have*

$$\frac{1}{(1+\delta^{-1})^2\tau^2}\|\boldsymbol{w}_2 - \boldsymbol{w}_1\|^2 \leq f(\boldsymbol{w}_2) - f(\boldsymbol{w}_1) - \langle\nabla f(\boldsymbol{w}_1), \boldsymbol{w}_2 - \boldsymbol{w}_1\rangle \leq \frac{p\tau^2}{(1-\delta^{-1})^2}\|\boldsymbol{w}_2 - \boldsymbol{w}_1\|^2.$$

Technical Lemmas

Lemma 5.10. *Let $f(\boldsymbol{w}) = -\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J})$. Define a local region of $\boldsymbol{w}^\star$ by*

$$\mathcal{B}_{\boldsymbol{M}}(\boldsymbol{w}^\star; r) = \{\boldsymbol{w} | \boldsymbol{w} \in \mathbb{B}_{\boldsymbol{M}}(\boldsymbol{w}^\star; r) \cap \mathcal{S}_{\boldsymbol{w}}\}.$$

where $\mathbb{B}_{\boldsymbol{M}}(\boldsymbol{w}^\star; r) = \{\boldsymbol{w} \in \mathbb{R}^{p(p-1)/2} \mid \|\boldsymbol{w} - \boldsymbol{w}^\star\|_{\boldsymbol{M}} \leq r\}$, in which $\|\boldsymbol{x}\|_{\boldsymbol{M}} = \langle \boldsymbol{x}, \boldsymbol{M}\boldsymbol{x}\rangle^{\frac{1}{2}}$ for any $\boldsymbol{x} \in \mathbb{R}^{p(p-1)/2}$ with $\boldsymbol{M}$ defined in Lemma 5.3, and $\mathcal{S}_{\boldsymbol{w}} = \{\boldsymbol{w} \,|\, \boldsymbol{w} \geq \boldsymbol{0}, (\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) \in \mathcal{S}_{++}^p\}$. Then for any $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathcal{B}_{\boldsymbol{M}}(\boldsymbol{w}^\star; r)$, we have

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

Lemma 5.11. *[66] Suppose a positive matrix $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ is diagonally scaled such that $A_{ii} = 1$, $i = 1, \ldots, p$, 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 $\boldsymbol{A}$ exists and $\boldsymbol{A}$ is an inverse M-matrix if $s^{-1} \geq p - 2$ with $p > 3$.

Lemma 5.12. *[63] (Sub-exponential tail bound) Suppose X is sub-exponential with parameters (υ, α). Then*

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

Proof

Theorem 3.1. *Let $\hat{\boldsymbol{\Theta}} \in \mathbb{R}^{p\times p}$ be the global minimum of (5) with $p > 3$. Define $s_1 = \max_k S_{kk}$ and $s_2 = \min_{ij} S_{ij}$. If the regularization parameter λ in (5) satisfies $\lambda \in [(2+2\sqrt{2})(p+1)(s_1 - s_2), +\infty)$, then the estimated graph weight $\hat{W}_{ij} = -\hat{\Theta}_{ij}$ obeys*

$$\hat{W}_{ij} \geq \frac{1}{(s_1 - (p+1)s_2 + \lambda)p} > 0, \quad \forall\, i \neq j.$$

Lemma 5.2. *For any given $\boldsymbol{w} \in \mathbb{R}^{p(p-1)/2}$ satisfying $(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) \in \mathcal{S}^p_{++}$, there must exist an unique $\boldsymbol{x} \in \mathbb{R}^{p(p-1)/2}$ such that*

$$\mathcal{L}\boldsymbol{x} + \frac{1}{b}\boldsymbol{J} = (\mathcal{L}\boldsymbol{w} + b\boldsymbol{J})^{-1} \tag{19}$$

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

Lemma 5.11. *[66] Suppose a positive matrix $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ is diagonally scaled such that $A_{ii} = 1$, $i = 1, \ldots, p$, 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 $\boldsymbol{A}$ exists and $\boldsymbol{A}$ is an inverse M-matrix if $s^{-1} \geq p - 2$ with $p > 3$.

Proof

Theorem 3.8. *Under Assumptions* $\boxed{3.5}$ *and* $\boxed{3.6}$*, take the regularization parameter* $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ *for some* $\alpha > 2$*. If the sample size* n *is lower bounded by*

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

then with probability at least $1 - 1/p^{\alpha-2}$*, the sequence* $\hat{\boldsymbol{w}}^{(k)}$ *returned by Algorithm* $\boxed{1}$ *satisfies*

$$\left\|\hat{\boldsymbol{w}}^{(k)} - \boldsymbol{w}^\star\right\| \leq \underbrace{2(3\sqrt{2}+4)\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)\sqrt{\alpha c_0^{-1} s \log p/n}}_{\text{Statistical error}} + \underbrace{\left(\frac{3}{2+\sqrt{2}}\right)^k \left\|\hat{\boldsymbol{w}}^{(0)} - \boldsymbol{w}^\star\right\|}_{\text{Optimization error}},$$

where $c_0 = 1/\left(8\left\|\mathcal{L}^*(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1}\right\|_{\max}^2\right)$ *is a constant.*

Lemma 5.7. *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* $\boxed{5.8}$*. Then one has*

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

Lemma 5.6. *Take* $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ *and suppose* $n \geq 94\alpha c_0^{-1}\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star) s \log p$ *for some* $\alpha > 2$*, where* c_0 *is a constant defined in Lemma* $\boxed{5.8}$*. Under Assumptions* $\boxed{3.5}$ *and* $\boxed{3.6}$*, if* $\left\|\mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right)\right\|_{\max} \leq \lambda/2$ *holds and* $\hat{\boldsymbol{w}}^{(0)}$ *satisfies* $|\mathrm{supp}^+(\hat{\boldsymbol{w}}^{(0)})| \leq s$*, then for any* $k \geq 1$*,* $\hat{\boldsymbol{w}}^{(k)}$ *defined in (*$\boxed{15}$*) obeys*

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

and

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

Proof

Corollary 3.9. *Under the same assumptions and conditions as stated in Theorem 3.8, the sequence* $\hat{\boldsymbol{w}}^{(k)}$ *returned by Algorithm 1 satisfies*

$$\left\|\mathcal{L}\hat{\boldsymbol{w}}^{(k)} - \mathcal{L}\boldsymbol{w}^\star\right\|_{\mathrm{F}} \leq \underbrace{4(2\sqrt{2}+3)\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)\sqrt{\alpha c_0^{-1} s \log p/n}}_{\textit{Statistical error}} + \underbrace{\left(\frac{3}{2+\sqrt{2}}\right)^k \left\|\mathcal{L}\hat{\boldsymbol{w}}^{(0)} - \mathcal{L}\boldsymbol{w}^\star\right\|_{\mathrm{F}}}_{\textit{Optimization error}},$$

with probability at least $1 - 1/p^{\alpha-2}$. *If* $k \geq \lceil 4\log(4\alpha c_0^{-1})\rceil$, *then the estimation error is dominated by the statistical error and we further obtain*

$$\left\|\mathcal{L}\hat{\boldsymbol{w}}^{(k)} - \mathcal{L}\boldsymbol{w}^\star\right\|_{\mathrm{F}} \lesssim \sqrt{s\log p/n},$$

where $c_0 = 1/\left(8\left\|\mathcal{L}^*(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1}\right\|_{\max}^2\right)$ *is a constant.*

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

Lemma 5.4. *Take* $\lambda = \sqrt{4\alpha c_0^{-1}\log p/n}$ *and suppose* $n \geq 94\alpha c_0^{-1}\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)s\log p$ *for some* $\alpha > 2$, *where* c_0 *is a constant defined in Lemma 5.8. Let*

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

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

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

where $\mathcal{S}^\star$ *is the support of* $\boldsymbol{w}^\star$.

Lemma 5.5. *Take* $\lambda = \sqrt{4\alpha c_0^{-1}\log p/n}$ *and suppose* $n \geq 94\alpha c_0^{-1}\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)s\log p$ *for some* $\alpha > 2$, *where* c_0 *is a constant defined in Lemma 5.8. Define the set* $\mathcal{E}^{(k)}$ *by*

$$\mathcal{E}^{(k)} = \{\mathcal{S}^\star \cup \mathcal{S}^{(k)}\}, \quad \textit{with} \quad \mathcal{S}^{(k)} = \{i \in [p(p-1)/2] \,|\, \hat{w}_i^{(k-1)} \geq b\}, \tag{20}$$

where $\hat{\boldsymbol{w}}^{(k)}$ *for* $k \geq 1$ *is defined in (15),* $\mathcal{S}^\star$ *is the support of* $\boldsymbol{w}^\star$ *with* $|\mathcal{S}^\star| \leq s$ *and* $b = (2+\sqrt{2})\lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star)\lambda$ *is a constant. Under Assumption 3.5, if* $\left\|\mathcal{L}^*\left((\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1} - \boldsymbol{S}\right)\right\|_{\max} \leq \lambda/2$ *holds and* $\hat{\boldsymbol{w}}^{(0)}$ *satisfies* $|supp^+(\hat{\boldsymbol{w}}^{(0)})| \leq s$, *then* $\mathcal{E}^{(k)}$ *obeys* $|\mathcal{E}^{(k)}| \leq 2s$, *for any* $k \geq 1$.

Proof

Theorem 3.10. *Under Assumptions 3.5 and 3.6, take the regularization parameter* $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ *for some* $\alpha > 2$. *If the sample size* n *satisfies*

$$n \geq c_0^{-1} \max\left(n \geq \max\left(840\alpha c_0^{-1} \frac{(\delta\tau^2+1)^4}{\delta^2\tau^2} sp \log p, 8\alpha \log p\right)\right),$$

then with probability at least $1 - 1/p^{\alpha-2}$, *the sequence* $\{\boldsymbol{w}_t^{(k)}\}_{t\geq 1}$ *returned by Algorithm 2 obeys*

$$\left\|\boldsymbol{w}_t^{(k)} - \hat{\boldsymbol{w}}^{(k)}\right\|^2 \leq \rho^t \left\|\boldsymbol{w}_0^{(k)} - \hat{\boldsymbol{w}}^{(k)}\right\|^2, \quad \forall\ k \geq 2,$$

where $\rho = 1 - \frac{\beta(1-\delta^{-1})^2}{p\tau^4(1+\delta^{-1})^2} < 1$ *with* $\delta > 1$ *and* $\beta \in (0,1)$, *and* $c_0 = 1/\left(8\left\|\mathcal{L}^*(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1}\right\|_{\max}^2\right)$ *is a constant.*

Lemma 5.4. *Take* $\lambda = \sqrt{4\alpha c_0^{-1} \log p/n}$ *and suppose* $n \geq 94\alpha c_0^{-1} \lambda_{\max}^2(\mathcal{L}\boldsymbol{w}^\star) s \log p$ *for some* $\alpha > 2$, *where* c_0 *is a constant defined in Lemma 5.8. Let*

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

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

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

where $\mathcal{S}^\star$ *is the support of* $\boldsymbol{w}^\star$.

Lemma 5.9. *Let* $f(\boldsymbol{w}) = -\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J})$. *Define a local region of* $\boldsymbol{w}^\star$ *by*

$$\mathcal{B}(\boldsymbol{w}^\star; r) = \{\boldsymbol{w} | \boldsymbol{w} \in \mathbb{B}(\boldsymbol{w}^\star; r) \cap \mathcal{S}_{\boldsymbol{w}}\}.$$

where $\mathbb{B}(\boldsymbol{w}^\star; r) = \{\boldsymbol{w} \in \mathbb{R}^{p(p-1)/2} \,|\, \|\boldsymbol{w} - \boldsymbol{w}^\star\| \leq r\}$, *and* $\mathcal{S}_{\boldsymbol{w}} = \{\boldsymbol{w}\,|\,\boldsymbol{w} \geq \boldsymbol{0}, (\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) \in \mathcal{S}_{++}^p\}$. *Then, under Assumption 3.6,* $g(\boldsymbol{w})$ *is* $\frac{2}{(1+\delta^{-1})^2\tau^2}$*-strongly convex and* $\frac{2p\tau^2}{(1-\delta^{-1})^2}$*-smooth in the region* $\mathcal{B}(\boldsymbol{w}^\star; \frac{1}{\sqrt{2p}\delta\tau})$ *where* τ *is defined in (17) and* $\delta > 1$. *In other words, for any* $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathcal{B}(\boldsymbol{w}^\star; \frac{1}{\sqrt{2p}\delta\tau})$, *we have*

$$\frac{1}{(1+\delta^{-1})^2\tau^2}\|\boldsymbol{w}_2 - \boldsymbol{w}_1\|^2 \leq f(\boldsymbol{w}_2) - f(\boldsymbol{w}_1) - \langle \nabla f(\boldsymbol{w}_1), \boldsymbol{w}_2 - \boldsymbol{w}_1\rangle \leq \frac{p\tau^2}{(1-\delta^{-1})^2}\|\boldsymbol{w}_2 - \boldsymbol{w}_1\|^2.$$