Supplemental Material to "Penalized Sparse Covariance Regression with High Dimensional Covariates"

A Appendix

A.1 Proof of Theorem 1

Proof. We follow the proof idea of Theorem 7.13 (a) in Wainwright (2019). Recall that $\mathbf{y}\mathbf{y}^\top = \sum_{k=0}^{K} \beta_k^{(0)} \mathbf{W}_k + \mathcal{E}$. Define $\widehat{\boldsymbol{\delta}} \overset{\text{def}}{=} \widehat{\boldsymbol{\beta}}^{\text{lasso}} - \boldsymbol{\beta}^{(0)}$. We first show that, if $\lambda_0 \ge (2/p) \max_{0\le k\le K} |\text{tr}(\mathbf{W}_k\mathcal{E})|$ holds, then $\widehat{\boldsymbol{\delta}} \in \mathbb{C}_3(\mathcal{S}) \overset{\text{def}}{=} \{\boldsymbol{\delta} \in \mathbb{R}^{K+1} : \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1 \le 3\|\boldsymbol{\delta}_{\mathcal{S}}\|_1\}$. Subsequently, we show that $\big\{\lambda_0 \ge (2/p) \max_{k\in\mathcal{S}} |\text{tr}(\mathbf{W}_k\mathcal{E})|\big\}$ holds with high probability.

Step 1. Since $\widehat{\boldsymbol{\beta}}^{\text{lasso}}$ is the solution to the problem (2.4), we have

$$Q(\widehat{\boldsymbol{\beta}}^{\text{lasso}}) + \lambda_0\|\widehat{\boldsymbol{\beta}}^{\text{lasso}}\|_1 = \frac{1}{2p}\left\|\mathcal{E} - \sum_{k=0}^{K}\widehat{\delta}_k\mathbf{W}_k\right\|_F^2 + \lambda_0\|\widehat{\boldsymbol{\beta}}^{\text{lasso}}\|_1 \le \frac{1}{2p}\|\mathcal{E}\|_F^2 + \lambda_0\|\boldsymbol{\beta}^{(0)}\|_1.$$

Rearranging the above inequality, we obtain that

$$0 \le \frac{1}{2p}\left\|\sum_{k=0}^{K}\widehat{\delta}_k\mathbf{W}_k\right\|_F^2 \le \frac{1}{p}\text{tr}\left(\mathcal{E}\sum_{k=0}^{K}\widehat{\delta}_k\mathbf{W}_k\right) + \lambda_0\Big\{\|\boldsymbol{\beta}^{(0)}\|_1 - \|\widehat{\boldsymbol{\beta}}^{\text{lasso}}\|_1\Big\} \tag{A.1}$$

Note that

$$\text{tr}\left(\mathcal{E}\sum_{k=0}^{K}\widehat{\delta}_k\mathbf{W}_k\right) \le \sum_{k=0}^{K}|\widehat{\delta}_k|\cdot|\text{tr}\left(\mathbf{W}_k\mathcal{E}\right)| \le \|\widehat{\boldsymbol{\delta}}\|_1 \max_{0\le k\le K}|\text{tr}(\mathbf{W}_k\mathcal{E})|. \tag{A.2}$$

Since $\boldsymbol{\beta}^{(0)}$ is supported on $\mathcal{S}$, we can write $\|\boldsymbol{\beta}^{(0)}\|_1 - \|\widehat{\boldsymbol{\beta}}^{\text{lasso}}\|_1 = \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_1 - \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)} + \widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 -$

$\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1$. Substituting it into the inequality (A.1) and using the inequality (A.2) yields

$$
\begin{aligned}
0 \leq & \frac{1}{p}\left\|\sum_{k=0}^{K}\widehat{\delta}_k\mathbf{W}_k\right\|_F^2 \leq \frac{2}{p}\max_{0\leq k\leq K}|\mathrm{tr}(\mathbf{W}_k\mathcal{E})|\cdot\|\widehat{\boldsymbol{\delta}}\|_1 + 2\lambda_0\Big\{\|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_1 - \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)} + \widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1\Big\} \\
\leq & \lambda_0\|\widehat{\boldsymbol{\delta}}\|_1 + 2\lambda_0\Big\{\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1\Big\} \leq \lambda_0\Big\{3\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1\Big\}, \qquad \text{(A.3)}
\end{aligned}
$$

where we have used the condition $\lambda_0 \geq (2/p)\max_{0\leq k\leq K}|\mathrm{tr}(\mathbf{W}_k\mathcal{E})|$ in the third inequality. Thus, we conclude that $\widehat{\boldsymbol{\delta}} \in \mathbb{C}_3(\mathcal{S})$. Then, by the RE Condition (C5) and the inequality (A.3), we can obtain that

$$
\kappa\|\widehat{\boldsymbol{\delta}}\|^2 \leq \frac{1}{p}\left\|\sum_{k=0}^{K}\widehat{\delta}_k\mathbf{W}_k\right\|_F^2 \leq \lambda_0\Big\{3\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1\Big\} \leq 3\lambda_0\sqrt{s+1}\|\widehat{\boldsymbol{\delta}}\|,
$$

where the last inequality follows from (A.17) in Lemma 1 with $\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 \leq \sqrt{s+1}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\| \leq \sqrt{s+1}\|\widehat{\boldsymbol{\delta}}\|$. This implies the conclusion $\|\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}} - \boldsymbol{\beta}^{(0)}\| = \|\widehat{\boldsymbol{\delta}}\| \leq (3/\kappa)\sqrt{s+1}\lambda_0$.

Step 2. It remains to show that the event $\big\{\lambda_0 \geq (2/p)\max_{0\leq k\leq K}|\mathrm{tr}(\mathbf{W}_k\mathcal{E})|\big\}$ holds with high probability. Recall that $\mathrm{tr}(\mathbf{W}_k\mathcal{E}) = \mathbf{y}^\top\mathbf{W}_k\mathbf{y} - \mathrm{tr}(\mathbf{W}_k\boldsymbol{\Sigma}_0)$. Further note that Condition (C4) and norm inequality (A.20) in Lemma 1 imply that $\sup_{p,k}\|\mathbf{W}_k\| \leq \sup_{p,k}\|\mathbf{W}_k\|_1 \leq w$ and $\|\boldsymbol{\Sigma}_0\| \leq \|\boldsymbol{\Sigma}_0^{1/2}\|^2 \leq \|\boldsymbol{\Sigma}_0^{1/2}\|_1^2 \leq \sigma_{\max}$. Then by union bound and Lemma 2, we have

$$
\begin{aligned}
P\left\{\frac{2}{p}\max_{0\leq k\leq K}|\mathrm{tr}(\mathbf{W}_k\mathcal{E})| \geq \lambda_0\right\} \leq & \sum_{k=0}^{K}P\left(\left|\mathbf{y}^\top\mathbf{W}_k\mathbf{y} - \mathrm{tr}(\mathbf{W}_k\boldsymbol{\Sigma}_0)\right| \geq \frac{p\lambda_0}{2}\right) \\
\leq & 2(K+1)\exp\left\{-\min\left(\frac{C_1p\lambda_0^2}{w^2\sigma_{\max}^2}, \frac{C_2p\lambda_0}{w\sigma_{\max}}\right)\right\}.
\end{aligned}
$$

Thus, we should have the event $\big\{\lambda_0 \geq (2/p)\max_{0\leq k\leq K}|\mathrm{tr}(\mathbf{W}_k\mathcal{E})|\big\}$ holds with the probability at least $1 - 2(K+1)\exp\Big\{-\min\Big(\frac{C_1p\lambda_0^2}{w^2\sigma_{\max}^2}, \frac{C_2p\lambda_0}{w\sigma_{\max}}\Big)\Big\}$. This completes the proof of the theorem. $\square$

Remark. In Theorem 1, we establish the ℓ_2-bound for the lasso estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}}$. In the subsequent analysis for the LLA algorithm, this ℓ_2-bound is used to obtain the ℓ_∞-bound $\|\widehat{\boldsymbol{\beta}}^{\text{lasso}} - \boldsymbol{\beta}^{(0)}\|_\infty$ by applying the norm inequality (A.18) in Lemma 1. This will lead to an extra factor $\sqrt{s}$ between the two tuning parameters λ_0 and λ. In fact, we may get rid of the factor $\sqrt{s}$ by directly establishing the ℓ_∞-bound of the Lasso estimator. Then we can relax the the requirement of λ in Theorem 2 to be $\lambda \geq c\lambda_0$ for some constant $c > 0$. This can be done by replacing the restricted eigenvalue (RE) Condition (C5) with a restricted invertibility factor (RIF) type condition (Zhang and Zhang, 2012):

(C5') (Restricted Invertibility Factor) Define the set $\mathbb{C}_3(\mathcal{S}) \stackrel{\text{def}}{=} \{\boldsymbol{\delta} \in \mathbb{R}^{K+1} : \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1 \leq 3\|\boldsymbol{\delta}_{\mathcal{S}}\|_1\}$. Assume $\{\mathbf{W}_k\}_{0\leq k\leq K}$ satisfies the restricted invertibility factor (RIF) condition, that is,

$$\frac{1}{p}\|\boldsymbol{\Sigma}_W\boldsymbol{\delta}\|_\infty \geq \kappa'\|\boldsymbol{\delta}\|_\infty, \quad \text{for all } \boldsymbol{\delta} \in \mathbb{C}_3(\mathcal{S})$$

for some constant $\kappa' > 0$, where $\boldsymbol{\Sigma}_W = \{\mathrm{tr}(\mathbf{W}_k\mathbf{W}_l) : 0 \leq k, l \leq K\} \in \mathbb{R}^{(K+1)\times(K+1)}$.

We next use Condition (C5') to establish the ℓ_∞-bound. By (A.3) in the proof of Theorem 1, we know that $\widehat{\boldsymbol{\delta}} = \widehat{\boldsymbol{\beta}}^{\text{lasso}} - \boldsymbol{\beta}^{(0)} \in \mathbb{C}_3(\mathcal{S})$. Thus, RIF condition implies that $\|\widehat{\boldsymbol{\delta}}\|_\infty \leq \|\boldsymbol{\Sigma}_W\widehat{\boldsymbol{\delta}}\|_\infty/(p\kappa')$. Note that

$$\boldsymbol{\Sigma}_W\widehat{\boldsymbol{\delta}} = \boldsymbol{\Sigma}_W(\widehat{\boldsymbol{\beta}}^{\text{lasso}} - \boldsymbol{\beta}^{(0)}) = \mathrm{tr}\left\{\mathbf{W}_k\left(\sum_{l=0}^{K}\widehat{\beta}_l^{\text{lasso}}\mathbf{W}_l - \mathbf{y}\mathbf{y}^\top\right)\right\}_{0\leq k\leq K} + \mathrm{tr}(\mathbf{W}_k\mathcal{E})_{0\leq k\leq K}.$$

Since $p^{-1}\max_{0\leq k\leq K}|\mathrm{tr}(\mathbf{W}_k\mathcal{E})| \leq \lambda_0/2$ by the assumption, we are left with bounding the first term. The optimality of $\widehat{\boldsymbol{\beta}}^{\text{lasso}}$ implies that

$$\frac{1}{2p}\left\|\mathbf{y}\mathbf{y}^\top - \sum_{l=0}^{K}\widehat{\beta}_l^{\text{lasso}}\mathbf{W}_l\right\|_F^2 + \lambda_0\|\widehat{\boldsymbol{\beta}}^{\text{lasso}}\|_1 \leq \frac{1}{2p}\left\|\mathbf{y}\mathbf{y}^\top - \sum_{l=0}^{K}\widehat{\beta}_l^{\text{lasso}}\mathbf{W}_l - t\mathbf{W}_k\right\|_F^2 + \lambda_0\|\widehat{\boldsymbol{\beta}}^{\text{lasso}}\|_1 + \lambda_0|t|,$$

for any $t \in \mathbb{R}$ and $0 \leq k \leq K$. Then we have

$$\frac{t}{p}\mathrm{tr}\left\{\mathbf{W}_k\left(\mathbf{y}\mathbf{y}^\top - \sum_{l=0}^{K}\widehat{\beta}_l^{\mathrm{lasso}}\mathbf{W}_l\right)\right\} \leq \frac{t^2}{2p}\|\mathbf{W}_k\|_F^2 + \lambda_0|t| \leq \frac{w^2t^2}{2} + \lambda_0|t|,$$

where we have used Condition (C4) and $\|\mathbf{W}_k\|_F^2 \leq p\|\mathbf{W}_k\|_1^2 \leq pw^2$ in the last inequality. Since t is arbitrary, we conclude that $\left|\mathrm{tr}\left\{\mathbf{W}_k\left(\mathbf{y}\mathbf{y}^\top - \sum_{l=0}^{K}\widehat{\beta}_l^{\mathrm{lasso}}\mathbf{W}_l\right)\right\}\right| \leq \lambda_0$ for each $0 \leq k \leq K$. Arranging these results, we conclude that

$$\|\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}} - \boldsymbol{\beta}^{(0)}\|_\infty = \|\widehat{\boldsymbol{\delta}}\|_\infty \leq \frac{1}{p\kappa'}\|\boldsymbol{\Sigma}_W\widehat{\boldsymbol{\delta}}\|_\infty \leq \frac{1}{\kappa'}(\frac{\lambda_0}{2} + \lambda_0) = \frac{3}{2\kappa'}\lambda_0.$$

This gives the desired ℓ_∞-bound for the Lasso estimator. We can see that the error bound $\|\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}} - \boldsymbol{\beta}^{(0)}\|_\infty = O(\lambda_0)$ is free of the factor $\sqrt{s}$.

A.2 Proof of Theorem 2

Following the idea of Fan et al. (2014), we prove the results in two steps. In the first step, we prove that the LLA algorithm converges under the given event. In the second step, we give the upper bounds for the three probabilities. In the last step, we show that the LLA algorithm converges to the oracle estimator with probability tending to one under the assumed conditions.

Step 1. Recall that $a_0 = \min\{1, a_2\}$. We first define three events as

$$\begin{aligned}
E_0 &= \left\{\|\widehat{\boldsymbol{\beta}}^{\mathrm{initial}} - \boldsymbol{\beta}^{(0)}\|_\infty \leq a_0\lambda\right\},\\
E_1 &= \left\{\|\nabla_{\mathcal{S}^c}Q(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{oracle}})\|_\infty < a_1\lambda\right\},\\
E_2 &= \left\{\|\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{oracle}}\|_{\min} \geq \gamma\lambda\right\}.
\end{aligned}$$

In the following, we prove that the LLA algorithm converges under the event $E_1 \cap E_2 \cap E_3$ in two further steps. We first show that the LLA algorithm initialized by

$\widehat{\boldsymbol{\beta}}^{\text{initial}}$ finds $\widehat{\boldsymbol{\beta}}^{\text{oracle}}$ after one iteration, under the event $E_0 \cap E_1$. We next show that if $\widehat{\boldsymbol{\beta}}^{\text{oracle}}$ is obtained, then the LLA algorithm will find $\widehat{\boldsymbol{\beta}}^{\text{oracle}}$ again in the next iteration, under the event $E_1 \cap E_2$. Then, we can immediately obtain that the LLA algorithm initialized by $\widehat{\boldsymbol{\beta}}^{\text{initial}}$ should converge to $\widehat{\boldsymbol{\beta}}^{\text{oracle}}$ after two iterations with probability at least $P(E_0 \cap E_1 \cap E_2) \geq 1 - P(E_0^c) - P(E_1^c) - P(E_2^c) = 1 - \delta_0 - \delta_1 - \delta_2$.

Step 1.1. Recall that $\widehat{\boldsymbol{\beta}}^{(0)} = \widehat{\boldsymbol{\beta}}^{\text{initial}}$. Under the event E_0, due to Assumption 1, we have $\widehat{\beta}_k^{(0)} \leq \|\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}^{(0)}\|_\infty \leq a_0\lambda \leq a_2\lambda$ for $k \in \mathcal{S}^c$, and $\widehat{\beta}_k^{(0)} \geq \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \|\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}^{(0)}\|_\infty > \gamma\lambda$ for $k \in \mathcal{S}$. By property (iv) of $p_\lambda(\cdot)$, we have $p_\lambda'(|\widehat{\beta}_k^{(0)}|) = 0$ for $k \in \mathcal{S}$. Thus, according to step (2.a) of the Algorithm 1, $\widehat{\boldsymbol{\beta}}^{(1)}$ should be the solution to the problem

$$\widehat{\boldsymbol{\beta}}^{(1)} = \operatorname{argmin}_{\boldsymbol{\beta}} Q(\boldsymbol{\beta}) + \sum_{k \in \mathcal{S}^c} p_\lambda'(|\widehat{\boldsymbol{\beta}}_k^{(0)}|)|\beta_k|. \tag{A.4}$$

By properties (ii) and (iii), $p_\lambda'(|\widehat{\boldsymbol{\beta}}_k^{(0)}|) \geq a_1\lambda$ holds for $k \in \mathcal{S}^c$. We next show that $\widehat{\boldsymbol{\beta}}^{\text{oracle}}$ is the unique global solution to (A.4) under the event E_1. By Condition (C2), we can verify that $\widehat{\beta}^{\text{oracle}}$ is the unique solution to $\operatorname{argmin}_{\boldsymbol{\beta}:\boldsymbol{\beta}_{\mathcal{S}^c = \mathbf{0}}} Q(\boldsymbol{\beta})$ and

$$\nabla_{\mathcal{S}} Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}) \stackrel{\text{def}}{=} \Big(\nabla_k Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}), k \in \mathcal{S}\Big) = \mathbf{0}. \tag{A.5}$$

Thus, for any $\boldsymbol{\beta}$ we have

$$\begin{aligned} Q(\boldsymbol{\beta}) \geq & Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}) + \sum_{k=0}^{K} \nabla_k Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}})(\beta_k - \widehat{\boldsymbol{\beta}}_k^{\text{oracle}}) \\ = & Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}) + \sum_{k \in \mathcal{S}^c} \nabla_k Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}})(\beta_k - \widehat{\boldsymbol{\beta}}_k^{\text{oracle}}). \end{aligned} \tag{A.6}$$

By (A.6), $\widehat{\boldsymbol{\beta}}^{\text{oracle}}_{\mathcal{S}^c} = \mathbf{0}$ and under the event E_1, for any $\boldsymbol{\beta}$ we have

$$
\begin{aligned}
&\left\{ Q(\boldsymbol{\beta}) + \sum_{k \in \mathcal{S}^c} p'_\lambda(|\widehat{\beta}^{(0)}_k|)|\beta_k| \right\} - \left\{ Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}) + \sum_{k \in \mathcal{S}^c} p'_\lambda(|\widehat{\beta}^{(0)}_k|)|\widehat{\beta}^{\text{oracle}}_k| \right\} \\
&\geq \sum_{k \in \mathcal{S}^c} \left\{ p'_\lambda(|\widehat{\beta}^{(0)}_k|) + \nabla_k Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}) \text{sign}(\beta_k) \right\} |\beta_k| \\
&\geq \sum_{k \in \mathcal{S}^c} \left\{ a_1 \lambda + \nabla_k Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}) \text{sign}(\beta_k) \right\} |\beta_k| \geq 0.
\end{aligned}
$$

The strict inequality holds unless $\beta_k = 0$ for all $k \in \mathcal{S}^c$. By uniqueness of the oracle estimator, we should have $\widehat{\boldsymbol{\beta}}^{\text{oracle}}$ is the unique solution to (A.4). This proves $\widehat{\boldsymbol{\beta}}^{(1)} = \widehat{\boldsymbol{\beta}}^{\text{oracle}}$.

Step 1.2. Given the LLA algorithm finds the oracle estimator, we denote $\widehat{\boldsymbol{\beta}}$ as the solution to the optimization problem in the next iteration of the LLA algorithm. By using $\widehat{\boldsymbol{\beta}}^{\text{oracle}}_{\mathcal{S}^c} = \mathbf{0}$ and $\nabla_k Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}) = 0, \forall k \in \mathcal{S}$, then under the event $E_2 = \left\{ \|\widehat{\boldsymbol{\beta}}^{\text{oracle}}_{\mathcal{S}}\|_{\min} \geq \gamma\lambda \right\}$, we have

$$
\widehat{\boldsymbol{\beta}} = \text{argmin}_{\boldsymbol{\beta}} Q(\boldsymbol{\beta}) + \sum_{k \in \mathcal{S}^c} p'_\lambda(0)|\beta_k|. \tag{A.7}
$$

Recall that $p'_\lambda(0) \geq a_1\lambda$. Then by similar procedures in Step 1, we can show that $\widehat{\boldsymbol{\beta}}^{\text{oracle}}$ is the unique solution to (A.7), under the event $E_1 = \left\{ \|\nabla_{\mathcal{S}^c} Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}_{\mathcal{S}})\|_\infty < a_1\lambda \right\}$. Hence, the LLA algorithm converges, under the event $E_1 \cap E_2$. This completes the proof of Step 1.

Step 2. We next give the upper bounds for $\delta_0 = P(E_0^c)$, $\delta_1 = P(E_1^c)$ and $\delta_2 = P(E_2^c)$ under the additional conditions. The three bounds are derived in the three further steps.

Step 2.1. Note that we use $\widehat{\boldsymbol{\beta}}^{\text{lasso}}$ as the initial estimator. Then by Theorem 1 and the condition $\lambda \geq (3\sqrt{s+1}\lambda_0)/(a_0\kappa)$, we have

$$
\|\widehat{\boldsymbol{\beta}}^{\text{initial}} - \boldsymbol{\beta}^{(0)}\|_\infty \leq \|\widehat{\boldsymbol{\beta}}^{\text{lasso}} - \boldsymbol{\beta}^{(0)}\| \leq \frac{3}{\kappa}\sqrt{s+1}\lambda_0 \leq a_0\lambda
$$

holds with probability at least $1-\delta_0'$ with

$$\delta_0' = 2(K+1)\exp\left\{-\min\left(\frac{C_1 p\lambda_0^2}{w^2\sigma_{\max}^2}, \frac{C_2 p\lambda_0}{w\sigma_{\max}}\right)\right\}.$$

Consequently, we should have $\delta_0 = P(E_0^c) = P(\|\widehat{\boldsymbol{\beta}}^{\text{initial}} - \boldsymbol{\beta}^{(0)}\|_\infty > a_0\lambda) \le \delta_0'$. This completes the proof of Step 2.1.

Step 2.2. We next bound the probability $\delta_1 = P(E_1^c) = P\big(\|\nabla_{\mathcal{S}^c} Q(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}})\|_\infty \ge a_1\lambda\big)$. Let $\mathbf{Y} = \text{vec}(\mathbf{y}\mathbf{y}^\top) \in \mathbb{R}^{p^2}$, $\mathbf{E} = \text{vec}(\mathcal{E}) \in \mathbb{R}^{p^2}$, and $\mathbf{V}_k = \text{vec}(\mathbf{W}_k) \in \mathbb{R}^{p^2}$. Further define $\mathbb{V} = (\mathbf{V}_k : 1 \le k \le K) \in \mathbb{R}^{p^2\times K}$, $\mathbb{V}_{\mathcal{S}} = (\mathbf{V}_k : k \in \mathcal{S}) \in \mathbb{R}^{p^2\times(s+1)}$, and $\mathbb{V}_{\mathcal{S}^c} = (\mathbf{V}_k : k \in \mathcal{S}^c) \in \mathbb{R}^{p^2\times(K-s)}$. Then we should have $\mathbf{Y} = \mathbb{V}_{\mathcal{S}}\boldsymbol{\beta}_{\mathcal{S}}^{(0)} + \mathbf{E}$, and $Q(\boldsymbol{\beta}) = (2p)^{-1}\|\mathbf{Y} - \mathbb{V}\boldsymbol{\beta}\|^2$. Let $\mathbb{H}_{\mathcal{S}} \overset{\text{def}}{=} \mathbb{V}_{\mathcal{S}}(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\mathbb{V}_{\mathcal{S}}^\top \in \mathbb{R}^{p^2\times p^2}$. Then we can compute that $\nabla_{\mathcal{S}^c} Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}) = \big\{\nabla_k Q(\widehat{\boldsymbol{\beta}}^{\text{oracle}}), k \in \mathcal{S}^c\big\} = -p^{-1}\mathbb{V}_{\mathcal{S}^c}^\top(\mathbf{I}_{p^2} - \mathbb{H}_{\mathcal{S}})\mathbf{E}$. By union bound, we have

$$\begin{aligned}
\delta_1 =& P\big(\|\nabla_{\mathcal{S}^c} Q(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}})\|_\infty \ge a_1\lambda\big) \le \sum_{k\in\mathcal{S}^c} P\Big(|\mathbf{V}_k^\top(\mathbf{I}_{p^2} - \mathbb{H}_{\mathcal{S}})\mathbf{E}| \ge pa_1\lambda\Big) \\
&\le \sum_{k\in\mathcal{S}^c}\Big\{P\Big(|\mathbf{V}_k^\top\mathbf{E}| \ge pa_1\lambda/2\Big) + P\Big(|\mathbf{V}_k^\top\mathbb{H}_{\mathcal{S}}\mathbf{E}| \ge pa_1\lambda/2\Big)\Big\}. \qquad (A.8)
\end{aligned}$$

Note that $\mathbf{V}_k^\top\mathbf{E} = \text{tr}(\mathbf{W}_k\mathcal{E}) = \text{tr}\{\mathbf{W}_k(\mathbf{y}\mathbf{y}^\top - \boldsymbol{\Sigma}_0)\} = \mathbf{y}^\top\mathbf{W}_k\mathbf{y} - \text{tr}(\mathbf{W}_k\boldsymbol{\Sigma}_0)$. Then by Lemma 2 and Conditions (C3) and (C4), we have $P\Big(|\mathbf{V}_k^\top\mathbf{E}| \ge pa_1\lambda/2\Big) =$

$$P\Big(\big|\mathbf{y}^\top\mathbf{W}_k\mathbf{y} - \text{tr}(\mathbf{W}_k\boldsymbol{\Sigma}_0)\big| > pa_1\lambda/2\Big) \le 2\exp\left\{-\min\left(\frac{C_3 a_1^2 p\lambda^2}{w^2\sigma_{\max}^2}, \frac{C_4 a_1 p\lambda}{w\sigma_{\max}}\right)\right\}.$$

By Condition (C4) and inequality (A.20) in Lemma 1, we have $\|\mathbf{W}_k\| \le \|\mathbf{W}_k\|_1 \le w$

for each $1 \le k \le K$. Then we can derive that

$$
\begin{aligned}
|\mathbf{V}_k^\top \mathbb{H}_{\mathcal{S}} \mathbf{E}| \le & \|(\mathbb{V}_{\mathcal{S}}^\top \mathbb{V}_{\mathcal{S}})^{-1} \mathbb{V}_{\mathcal{S}}^\top \mathbf{V}_k\| \|\mathbb{V}_{\mathcal{S}}^\top \mathbf{E}\| \le \|(\mathbb{V}_{\mathcal{S}}^\top \mathbb{V}_{\mathcal{S}})^{-1}\| \|\mathbb{V}_{\mathcal{S}}^\top \mathbf{V}_k\| \|\mathbb{V}_{\mathcal{S}}^\top \mathbf{E}\| \\
\le & \|\mathbf{\Sigma}_{W,\mathcal{S}}^{-1}\| \Big\{ \sqrt{s+1} \max_{l \in \mathcal{S}} |\mathrm{tr}(\mathbf{W}_l \mathbf{W}_k)| \Big\} \Big\{ \sqrt{s+1} \max_{l \in \mathcal{S}} |\mathrm{tr}(\mathbf{W}_l \mathcal{E})| \Big\} \\
\le & \Big\{ (p\tau_{\min})^{-1} \Big\} \Big\{ \sqrt{s+1} (pw^2) \Big\} \Big\{ \sqrt{s+1} \max_{l \in \mathcal{S}} |\mathrm{tr}(\mathbf{W}_l \mathcal{E})| \Big\} \\
= & \tau_{\min}^{-1} w^2 (s+1) \max_{l \in \mathcal{S}} \big| \mathbf{y}^\top \mathbf{W}_l \mathbf{y} - \mathrm{tr}(\mathbf{W}_l \mathbf{\Sigma}_0) \big|,
\end{aligned}
$$

where the third inequality is due to inequality (A.18) in Lemma 1, and the last inequality is due to the following two facts: (i) by Condition (C4) and inequality (A.20) in Lemma 1, we have $|\mathrm{tr}(\mathbf{W}_l \mathbf{W}_k)| \le p \|\mathbf{W}_l\| \|\mathbf{W}_k\| \le pw^2$; (ii) by Condition (C2), we have $\big\|\mathbf{\Sigma}_{W,\mathcal{S}}^{-1}\big\| = \lambda_{\min}^{-1}(\mathbf{\Sigma}_{W,\mathcal{S}}) \le (p\tau_{\min})^{-1}$. Then by Lemma 2 and Conditions (C3) and (C4), we have $P\Big(|\mathbf{V}_k^\top \mathbb{H}_{\mathcal{S}} \mathbf{E}| \ge p^2 a_1 \lambda / 2 \Big) \le$

$$
\begin{aligned}
& \sum_{l \in \mathcal{S}} P \left\{ \big| \mathbf{y}^\top \mathbf{W}_l \mathbf{y} - \mathrm{tr}(\mathbf{W}_l \mathbf{\Sigma}_0) \big| > \frac{a_1 \tau_{\min} p \lambda}{2(s+1) w^2} \right\} \\
\le & 2(s+1) \exp \left[-\min \left\{ \frac{C_5 a_1^2 \tau_{\min}^2 p \lambda^2}{w^6 \sigma_{\max}^2 (s+1)^2}, \frac{C_6 a_1 \tau_{\min} p \lambda}{w^3 \sigma_{\max} (s+1)} \right\} \right]
\end{aligned}
$$

Together with (A.8), we have

$$
\begin{aligned}
\delta_1 \le & 2(K-s) \exp \left\{ -\min \left(\frac{C_3 a_1^2 p \lambda^2}{w^2 \sigma_{\max}^2}, \frac{C_4 a_1 p \lambda}{w \sigma_{\max}} \right) \right\} \\
& + 2(K-s)(s+1) \exp \left[-\min \left\{ \frac{C_5 a_1^2 \tau_{\min}^2 p \lambda^2}{w^6 \sigma_{\max}^2 (s+1)^2}, \frac{C_6 a_1 \tau_{\min} p \lambda}{w^3 \sigma_{\max} (s+1)} \right\} \right].
\end{aligned}
$$

Step 2.3. We next bound $\delta_2 = P(E_2^c) = P(\|\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}}\|_{\min} < \gamma\lambda)$. Note that $\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}} = \boldsymbol{\beta}_{\mathcal{S}}^{(0)} + (\mathbb{V}_{\mathcal{S}}^\top \mathbb{V}_{\mathcal{S}})^{-1} \mathbb{V}_{\mathcal{S}}^\top \mathbf{E}$, and thus $\|\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}}\|_{\min} \ge \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \|(\mathbb{V}_{\mathcal{S}}^\top \mathbb{V}_{\mathcal{S}})^{-1} \mathbb{V}_{\mathcal{S}}^\top \mathbf{E}\|_\infty$. Then we have

$$
\delta_2 \le P\Big(\|(\mathbb{V}_{\mathcal{S}}^\top \mathbb{V}_{\mathcal{S}})^{-1} \mathbb{V}_{\mathcal{S}}^\top \mathbf{E}\|_\infty \ge \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \gamma\lambda \Big). \tag{A.9}
$$

Note that

$$
\begin{aligned}
&\|(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\mathbb{V}_{\mathcal{S}}^\top\mathbf{E}\|_\infty \le \|(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\mathbb{V}_{\mathcal{S}}^\top\mathbf{E}\| \le \|(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\|\|\mathbb{V}_{\mathcal{S}}^\top\mathbf{E}\| \\
\le&(p\tau_{\min})^{-1}\sqrt{s+1}\|\mathbb{V}_{\mathcal{S}}^\top\mathbf{E}\|_\infty = \sqrt{s+1}(p\tau_{\min})^{-1}\max_{k\in\mathcal{S}}|\mathbf{y}^\top\mathbf{W}_k\mathbf{y} - \mathrm{tr}(\mathbf{W}_k\mathbf{\Sigma}_0)|,
\end{aligned}
$$

where the first inequality is due to inequality (A.18) in Lemma 1, and the third inequality is due to Condition (C2) and (A.18) in Lemma 1. Together with (A.9) and using Lemma 2, we have

$$
\begin{aligned}
\delta_2 \le& \sum_{k\in\mathcal{S}} P\left\{\left|\mathbf{y}^\top\mathbf{W}_k\mathbf{y} - \mathrm{tr}(\mathbf{W}_k\mathbf{\Sigma}_0)\right| \ge \frac{\tau_{\min}p}{(s+1)^{1/2}}(\|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \gamma\lambda)\right\} \\
\le& 2(s+1)\exp\left[-\min\left\{\frac{C_5\tau_{\min}^2 p(\|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \gamma\lambda)^2}{w^2\sigma_{\max}^2(s+1)}, \frac{C_6\tau_{\min}p(\|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \gamma\lambda)}{w\sigma_{\max}(s+1)^{1/2}}\right\}\right].
\end{aligned}
$$

This competes the proof of Step 2.

Step 3. To obtain the desired result, it suffices to prove that δ_1, δ_2, and δ_0' tend to 0 as $p\to\infty$ under the assumed conditions. By Condition (C1), we know that $\|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \gamma\lambda > \lambda$. Then, by inspecting the forms of upper bounds of δ_0, δ_1, δ_2, it remains to prove that

$$
\min\left\{\frac{p\lambda^2}{s^2}, \frac{p\lambda}{s}, \frac{p\lambda^2}{s}, \frac{p\lambda}{\sqrt{s}}, p\lambda_0^2, p\lambda_0,\right\}\Big/\log(K) \to 0 \qquad \text{(A.10)}
$$

as $p\to\infty$. Further note $\lambda \ge (3\sqrt{s+1}\lambda_0)/(a_0\kappa)$. Then we can easily verify that, (A.10) holds as long as $p\lambda_0^2/\{s\log(K)\}\to\infty$ as $p\to\infty$. This completes the proof of Step 3 and completes the proof of the theorem.

A.3 Proof of Theorem 3

Recall that the oracle estimator is computed with the knowledge of the true support set of $\boldsymbol{\beta}^{(0)}$. That is, $\widehat{\boldsymbol{\beta}}^{\text{oracle}} = \text{argmin}_{\boldsymbol{\beta}:\boldsymbol{\beta}_{\mathcal{S}^c=\mathbf{0}}} Q(\boldsymbol{\beta})$, where $Q(\boldsymbol{\beta})$ is defined in (2.2). Equivalently, we should have

$$\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}} - \boldsymbol{\beta}_{\mathcal{S}}^{(0)} = \boldsymbol{\Sigma}_{W,\mathcal{S}}^{-1}\boldsymbol{\Sigma}_{WY,\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}}^{(0)} = \boldsymbol{\Sigma}_{W,\mathcal{S}}^{-1} S_p,$$

where $\boldsymbol{\Sigma}_{W,\mathcal{S}} = \{\text{tr}(\mathbf{W}_k\mathbf{W}_l) : k, l \in \mathcal{S}\} \in \mathbb{R}^{(s+1)\times(s+1)}$, $\boldsymbol{\Sigma}_{WY,\mathcal{S}} = \{\mathbf{y}^\top\mathbf{W}_k\mathbf{y} : k \in \mathcal{S}\}^\top \in \mathbb{R}^{s+1}$, and

$$S_p = \begin{pmatrix} \text{vec}^\top(\mathbf{W}_0) \\ \vdots \\ \text{vec}^\top(\mathbf{W}_s) \end{pmatrix} \text{vec}(\mathbf{y}\mathbf{y}^\top - \boldsymbol{\Sigma}_0) = \begin{pmatrix} \text{vec}^\top(\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_0\boldsymbol{\Sigma}_0^{1/2}) \\ \vdots \\ \text{vec}^\top(\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_s\boldsymbol{\Sigma}_0^{1/2}) \end{pmatrix} \text{vec}(\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}_p).$$

Here we have used the facts that $\mathbf{y} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z}$, and $\text{vec}(\mathbf{M}_1\mathbf{M}_2\mathbf{M}_3) = (\mathbf{M}_3^\top \otimes \mathbf{M}_1)\text{vec}(\mathbf{M}_2)$ for three arbitrary matrices $\mathbf{M}_1$, $\mathbf{M}_2$, $\mathbf{M}_3$ of shapes $p_1 \times p_2$, $p_2 \times p_3$, and $p_3 \times p_4$ (see, e.g., (1.3.6) in Golub and Van Loan, 2013, p. 28). Re-express $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_L)^\top$, where $\mathbf{a}_l = (a_{l0}, \ldots, a_{ls})^\top \in \mathbb{R}^{s+1}$. Let $\widetilde{S}_p = (s+1)^{-1/2}\mathbf{A}\boldsymbol{\Sigma}_{W,\mathcal{S}}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}} - \boldsymbol{\beta}_{\mathcal{S}}^{(0)}) = (s+1)^{-1/2}\mathbf{A}S_p$. Then we should have

$$\widetilde{S}_p = \begin{pmatrix} \text{vec}^\top(\boldsymbol{\Delta}_1) \\ \vdots \\ \text{vec}^\top(\boldsymbol{\Delta}_L) \end{pmatrix} \text{vec}(\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}_p) \in \mathbb{R}^L,$$

where $\boldsymbol{\Delta}_l = (s+1)^{-1/2}\sum_{k=0}^{s} a_{lk}(\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_k\boldsymbol{\Sigma}_0^{1/2})$ for $1 \le l \le L$. Further note that

$$\frac{1}{\sqrt{s+1}}\max_{1\le l\le L}\sum_{k=0}^{s}|a_{lk}| = \frac{1}{\sqrt{s+1}}\|\mathbf{A}\|_\infty \le \|\mathbf{A}\| < \infty,$$

where the first inequality follows from (A.20) in Lemma 1. By Condition (C4), we have $\sup_{p,k}\|\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_k\boldsymbol{\Sigma}_0^{1/2}\|_1 < \infty$. Then it follows that

$$\begin{aligned}\sup_p\|\boldsymbol{\Delta}_l\|_1 \le& \sup_p\frac{1}{\sqrt{s+1}}\sum_{k=0}^{s}|a_{lk}|\cdot\|\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_k\boldsymbol{\Sigma}_0^{1/2}\|_1\\ \le&\left\{\frac{1}{\sqrt{s+1}}\max_{1\le l\le L}\sum_{k=0}^{s}|a_{lk}|\right\}\left\{\sup_{p,k}\|\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_k\boldsymbol{\Sigma}_0^{1/2}\|_1\right\} < \infty,\end{aligned}$$

for each $1 \le l \le L$. By using Lemma 3, we know that

$$\mathrm{cov}(\widetilde{S}_p) = 2\{\mathrm{tr}(\boldsymbol{\Delta}_k\boldsymbol{\Delta}_l) : 1 \le l \le L\} + (\mu_4 - 3)\{\mathrm{tr}(\boldsymbol{\Delta}_k \circ \boldsymbol{\Delta}_l) : 1 \le k, l \le L\}.$$

By assumed conditions in the theorem, we can verify that $p^{-1}\mathrm{cov}(\widetilde{S}_p) \to \mathbf{C}$. Then by Lemma 3, we should have

$$\sqrt{p/(s+1)}\mathbf{A}(p^{-1}\boldsymbol{\Sigma}_{W,\mathcal{S}})(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{oracle}} - \boldsymbol{\beta}_{\mathcal{S}}^{(0)}) = p^{-1/2}\widetilde{S}_p \to_d \mathcal{N}(\mathbf{0}, \mathbf{C}).$$

By Condition (C6), we know that $p^{-1}\boldsymbol{\Sigma}_{W,\mathcal{S}} \to \mathbf{G}_0$ in the Frobenius norm. With the help of Slutsky's theorem, we obtain that $\sqrt{p/(s+1)}\mathbf{A}\mathbf{G}_0\left(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{oracle}} - \boldsymbol{\beta}_{\mathcal{S}}^{(0)}\right) \to_d \mathcal{N}(\mathbf{0}, \mathbf{C})$ as $p \to \infty$. This completes the proof of the theorem.

A.4 Proofs of Theorems 4 and 5

Proof of Theorem 4. The proof is very similar to the proof of Theorem 1 in Appendix A.1. Note that $\mathbf{y}_i\mathbf{y}_i^\top = \sum_{k=0}^{K}\beta_k^{(0)}\mathbf{W}_k + \mathcal{E}_i$ for $1 \le i \le n$. Define $\widehat{\boldsymbol{\delta}} \overset{\mathrm{def}}{=} \widehat{\boldsymbol{\beta}}_n^{\mathrm{lasso}} - \boldsymbol{\beta}^{(0)}$. We first show that, if $\lambda_0 \ge (2/p)\max_{0\le k\le K}|n^{-1}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{W}_k\mathcal{E}_i)|$ holds, then

$\widehat{\boldsymbol{\delta}} \in \mathbb{C}_3(\mathcal{S}) \overset{\text{def}}{=} \{\boldsymbol{\delta} \in \mathbb{R}^{K+1} : \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1 \leq 3\|\boldsymbol{\delta}_{\mathcal{S}}\|_1\}$. Subsequently, we show that $\big\{\lambda_0 \geq (2/p)\max_{0\leq k\leq K}|n^{-1}\sum_{i=1}^n \mathrm{tr}(\mathbf{W}_k\mathcal{E}_i)|\big\}$ holds with high probability.

Step 1. Since $\widehat{\boldsymbol{\beta}}_n^{\text{lasso}}$ is the solution to $\mathrm{argmin}_{\boldsymbol{\beta}} Q_n(\boldsymbol{\beta}) + \lambda_0\|\boldsymbol{\beta}\|_1$, we have

$$\begin{aligned} Q_n(\widehat{\boldsymbol{\beta}}_n^{\text{lasso}}) + \lambda_0\|\widehat{\boldsymbol{\beta}}_n^{\text{lasso}}\|_1 =& \frac{1}{2np}\sum_{i=1}^n \left\|\mathcal{E}_i - \sum_{k=0}^K \widehat{\delta}_k \mathbf{W}_k\right\|_F^2 + \lambda_0\|\widehat{\boldsymbol{\beta}}_n^{\text{lasso}}\|_1 \\ \leq& \frac{1}{2np}\sum_{i=1}^n \|\mathcal{E}_i\|_F^2 + \lambda_0\|\boldsymbol{\beta}^{(0)}\|_1. \end{aligned}$$

Rearranging the above inequality, we obtain that

$$0 \leq \frac{1}{2p}\left\|\sum_{k=0}^K \widehat{\delta}_k \mathbf{W}_k\right\|_F^2 \leq \frac{1}{np}\sum_{i=1}^n \mathrm{tr}\left(\mathcal{E}_i \sum_{k=0}^K \widehat{\delta}_k \mathbf{W}_k\right) + \lambda_0\Big\{\|\boldsymbol{\beta}^{(0)}\|_1 - \|\widehat{\boldsymbol{\beta}}^{\text{lasso}}\|_1\Big\} \quad \text{(A.11)}$$

Note that

$$\frac{1}{n}\sum_{i=1}^n \mathrm{tr}\left(\mathcal{E}_i \sum_{k=0}^K \widehat{\delta}_k \mathbf{W}_k\right) \leq \sum_{k=0}^K |\widehat{\delta}_k| \cdot \left|n^{-1}\sum_{i=1}^n \mathrm{tr}\left(\mathbf{W}_k\mathcal{E}_i\right)\right| \leq \|\widehat{\boldsymbol{\delta}}\|_1 \max_{0\leq k\leq K}\left|n^{-1}\sum_{i=1}^n \mathrm{tr}\left(\mathbf{W}_k\mathcal{E}_i\right)\right|. \quad \text{(A.12)}$$

Since $\boldsymbol{\beta}^{(0)}$ is supported on $\mathcal{S}$, we can write $\|\boldsymbol{\beta}^{(0)}\|_1 - \|\widehat{\boldsymbol{\beta}}^{\text{lasso}}\|_1 = \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_1 - \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)} + \widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1$. Substituting it into the inequality (A.11) and using the inequality (A.12) yields

$$\begin{aligned} 0 \leq& \frac{1}{p}\left\|\sum_{k=0}^K \widehat{\delta}_k \mathbf{W}_k\right\|_F^2 \leq \frac{2}{p}\max_{0\leq k\leq K}\left|n^{-1}\sum_{i=1}^n \mathrm{tr}(\mathbf{W}_k\mathcal{E}_i)\right| \cdot \|\widehat{\boldsymbol{\delta}}\|_1 + 2\lambda_0\Big\{\|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_1 - \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)} + \widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1\Big\} \\ \leq& \lambda_0\|\widehat{\boldsymbol{\delta}}\|_1 + 2\lambda_0\Big\{\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1\Big\} \leq \lambda_0\Big\{3\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1\Big\}, \end{aligned} \quad \text{(A.13)}$$

where we have used the condition $\lambda_0 \geq (2/p)\max_{0\leq k\leq K}|n^{-1}\sum_{i=1}^n \mathrm{tr}(\mathbf{W}_k\mathcal{E}_i)|$ in the third inequality. Thus, we conclude that $\widehat{\boldsymbol{\delta}} \in \mathbb{C}_3(\mathcal{S})$. Then, by the RE Condition (C5) and the inequality (A.13), we can obtain that

$$\kappa\|\widehat{\boldsymbol{\delta}}\|^2 \leq \frac{1}{p}\left\|\sum_{k=0}^K \widehat{\delta}_k \mathbf{W}_k\right\|_F^2 \leq \lambda_0\Big\{3\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1\Big\} \leq 3\lambda_0\sqrt{s+1}\|\widehat{\boldsymbol{\delta}}\|,$$

where the last inequality follows from (A.17) in Lemma 1 with $\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 \le \sqrt{s+1}\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\| \le \sqrt{s+1}\|\widehat{\boldsymbol{\delta}}\|$. This implies the conclusion $\|\widehat{\boldsymbol{\beta}}_n^{\text{lasso}} - \boldsymbol{\beta}^{(0)}\| = \|\widehat{\boldsymbol{\delta}}\| \le (3/\kappa)\sqrt{s+1}\lambda_0$.

Step 2. It remains to show that the event $\left\{\lambda_0 \ge (2/p)\max_{0\le k\le K}|n^{-1}\sum_{i=1}^n \operatorname{tr}(\mathbf{W}_k\mathcal{E}_i)|\right\}$ holds with high probability. Recall that $n^{-1}\sum_{i=1}^n \operatorname{tr}(\mathbf{W}_k\mathcal{E}_i) = n^{-1}\sum_{i=1}^n \mathbf{y}_i^\top \mathbf{W}_k \mathbf{y}_i - \operatorname{tr}(\mathbf{W}_k\boldsymbol{\Sigma}_0)$. Further note that Condition (C4) and norm inequality (A.20) in Lemma 1 imply that $\sup_{p,k}\|\mathbf{W}_k\| \le \sup_{p,k}\|\mathbf{W}_k\|_1 \le w$ and $\|\boldsymbol{\Sigma}_0\| \le \|\boldsymbol{\Sigma}_0^{1/2}\|^2 \le \|\boldsymbol{\Sigma}_0^{1/2}\|_1^2 \le \sigma_{\max}$. Then by union bound and Lemma 2, we have

$$
\begin{aligned}
P\left\{\frac{2}{p}\max_{0\le k\le K}|n^{-1}\sum_{i=1}^n \operatorname{tr}(\mathbf{W}_k\mathcal{E}_i)| \ge \lambda_0\right\} \le & \sum_{k=0}^K P\left(\left|n^{-1}\sum_{i=1}^n \mathbf{y}_i^\top\mathbf{W}_k\mathbf{y}_i - \operatorname{tr}(\mathbf{W}_k\boldsymbol{\Sigma}_0)\right| \ge \frac{p\lambda_0}{2}\right) \\
\le & 2(K+1)\exp\left\{-\min\left(\frac{C_1 np\lambda_0^2}{w^2\sigma_{\max}^2}, \frac{C_2 np\lambda_0}{w\sigma_{\max}}\right)\right\}.
\end{aligned}
$$

Thus, we should have the event $\left\{\lambda_0 \ge (2/p)\max_{0\le k\le K}|n^{-1}\sum_{i=1}^n \operatorname{tr}(\mathbf{W}_k\mathcal{E}_i)|\right\}$ holds with the probability at least $1-2(K+1)\exp\left\{-\min\left(\frac{C_1 np\lambda_0^2}{w^2\sigma_{\max}^2}, \frac{C_2 np\lambda_0}{w\sigma_{\max}}\right)\right\}$. This completes the proof of the theorem.

Proof of Theorem 5. The proof is very similar to the proof of Theorem 2 in Appendix A.2. There are three steps. In the first step, we need to prove that the LLA algorithm converges under the event $E_1 \cap E_2 \cap E_3$, where

$$
\begin{aligned}
E_0 &= \left\{\|\widehat{\boldsymbol{\beta}}_n^{\text{lasso}} - \boldsymbol{\beta}^{(0)}\|_\infty \le a_0\lambda\right\}, \\
E_1 &= \left\{\|\nabla_{\mathcal{S}^c} Q(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}})\|_\infty < a_1\lambda\right\}, \\
E_2 &= \left\{\|\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}}\|_{\min} \ge \gamma\lambda\right\}.
\end{aligned}
$$

In the second step, we derive the upper bounds for $P(E_0^c)$, $P(E_1^c)$ and $P(E_2^c)$. In the last step, we show that the LLA algorithm converges to the oracle estimator with probability tending to one under the assumed conditions. Since the first step is almost the same as that in Appendix A.2, we omit the details.

Step 2. In this step, we give the upper bounds for $\delta_0 = P(E_0^c)$, $\delta_1 = P(E_1^c)$ and $\delta_2 = P(E_2^c)$ under the assumed conditions. The three bounds are derived in the three further steps.

Step 2.1. Note that we use $\widehat{\boldsymbol{\beta}}_n^{\text{lasso}}$ as the initial estimator. Then by Theorem 4 and the condition $\lambda \geq (3\sqrt{s+1}\lambda_0)/(a_0\kappa)$, we have

$$\|\widehat{\boldsymbol{\beta}}_n^{\text{lasso}} - \boldsymbol{\beta}^{(0)}\|_\infty \leq \|\widehat{\boldsymbol{\beta}}_n^{\text{lasso}} - \boldsymbol{\beta}^{(0)}\| \leq \frac{3}{\kappa}\sqrt{s+1}\lambda_0 \leq a_0\lambda$$

holds with probability at least $1 - \delta_0'$ with

$$\delta_0' = 2(K+1)\exp\left\{-\min\left(\frac{C_1 np\lambda_0^2}{w^2\sigma_{\max}^2}, \frac{C_2 np\lambda_0}{w\sigma_{\max}}\right)\right\}.$$

Consequently, we should have $\delta_0 = P(E_0^c) = P(\|\widehat{\boldsymbol{\beta}}_n^{\text{lasso}} - \boldsymbol{\beta}^{(0)}\|_\infty > a_0\lambda) \leq \delta_0'$. This completes the proof of Step 2.1.

Step 2.2. We next bound the probability $\delta_1 = P(E_1^c) = P\big(\|\nabla_{n,\mathcal{S}^c}Q(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}})\|_\infty \geq a_1\lambda\big)$. Let $\mathbf{Y}_i = \text{vec}(\mathbf{y}_i\mathbf{y}_i^\top) \in \mathbb{R}^{p^2}$, $\mathbf{E}_i = \text{vec}(\mathcal{E}_i) \in \mathbb{R}^{p^2}$, and $\mathbf{V}_k = \text{vec}(\mathbf{W}_k) \in \mathbb{R}^{p^2}$. Further define $\mathbb{V} = (\mathbf{V}_k : 1 \leq k \leq K) \in \mathbb{R}^{p^2\times K}$, $\mathbb{V}_{\mathcal{S}} = (\mathbf{V}_k : k \in \mathcal{S}) \in \mathbb{R}^{p^2\times(s+1)}$, and $\mathbb{V}_{\mathcal{S}^c} = (\mathbf{V}_k : k \in \mathcal{S}^c) \in \mathbb{R}^{p^2\times(K-s)}$. Then we should have $\mathbf{Y}_i = \mathbb{V}_{\mathcal{S}}\boldsymbol{\beta}_{\mathcal{S}}^{(0)} + \mathbf{E}_i$, and $Q_n(\boldsymbol{\beta}) = (2np)^{-1}\sum_{i=1}^n \|\mathbf{Y}_i - \mathbb{V}\boldsymbol{\beta}\|^2$. Let $\mathbb{H}_{\mathcal{S}} \stackrel{\text{def}}{=} \mathbb{V}_{\mathcal{S}}(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\mathbb{V}_{\mathcal{S}}^\top \in \mathbb{R}^{p^2\times p^2}$, and $\overline{\mathbf{E}} = n^{-1}\sum_{i=1}^n \mathbf{E}_i$. Then we can compute that $\nabla_{\mathcal{S}^c}Q(\widehat{\boldsymbol{\beta}}_n^{\text{oracle}}) = \big\{\nabla_k Q(\widehat{\boldsymbol{\beta}}_n^{\text{oracle}}), k \in \mathcal{S}^c\big\} = -p^{-1}\mathbb{V}_{\mathcal{S}^c}^\top(\mathbf{I}_{p^2} - \mathbb{H}_{\mathcal{S}})\overline{\mathbf{E}}$. By union bound, we have

$$\begin{aligned}\delta_1 =& P\big(\|\nabla_{\mathcal{S}^c}Q(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\text{oracle}})\|_\infty \geq a_1\lambda\big) \leq \sum_{k\in\mathcal{S}^c} P\Big(|\mathbf{V}_k^\top(\mathbf{I}_{p^2} - \mathbb{H}_{\mathcal{S}})\overline{\mathbf{E}}| \geq pa_1\lambda\Big) \\ &\leq \sum_{k\in\mathcal{S}^c}\Big\{P\Big(|\mathbf{V}_k^\top\overline{\mathbf{E}}| \geq pa_1\lambda/2\Big) + P\Big(|\mathbf{V}_k^\top\mathbb{H}_{\mathcal{S}}\overline{\mathbf{E}}| \geq pa_1\lambda/2\Big)\Big\}.\end{aligned} \tag{A.14}$$

Note that $\mathbf{V}_k^\top\overline{\mathbf{E}} = \text{tr}(n^{-1}\sum_{i=1}^n \mathbf{W}_k\mathcal{E}_i) = \text{tr}\{n^{-1}\sum_{i=1}^n \mathbf{W}_k(\mathbf{y}_i\mathbf{y}_i^\top - \boldsymbol{\Sigma}_0)\} = n^{-1}\sum_{i=1}^n \mathbf{y}_i^\top\mathbf{W}_k\mathbf{y}_i - \text{tr}(\mathbf{W}_k\boldsymbol{\Sigma}_0)$. Then by Lemma 2 and Conditions (C3) and (C4), we have $P\Big(|\mathbf{V}_k^\top\overline{\mathbf{E}}| \geq$

$pa_1\lambda/2\Big) =$

$$P\Big(n^{-1}\sum_{i=1}^{n}\big|\mathbf{y}_i^\top\mathbf{W}_k\mathbf{y}_i - \mathrm{tr}(\mathbf{W}_k\boldsymbol{\Sigma}_0)\big| > pa_1\lambda/2\Big) \le 2\exp\left\{-\min\left(\frac{C_3a_1^2np\lambda^2}{w^2\sigma_{\max}^2}, \frac{C_4a_1np\lambda}{w\sigma_{\max}}\right)\right\}.$$

By Condition (C4) and inequality (A.20) in Lemma 1, we have $\|\mathbf{W}_k\| \le \|\mathbf{W}_k\|_1 \le w$ for each $1 \le k \le K$. Then we can derive that

$$\begin{aligned}
|\mathbf{V}_k^\top\mathbb{H}_{\mathcal{S}}\overline{\mathbf{E}}| \le& \|(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\mathbb{V}_{\mathcal{S}}^\top\mathbf{V}_k\|\|\mathbb{V}_{\mathcal{S}}^\top\overline{\mathbf{E}}\| \le \|(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\|\|\mathbb{V}_{\mathcal{S}}^\top\mathbf{V}_k\|\|\mathbb{V}_{\mathcal{S}}^\top\overline{\mathbf{E}}\| \\
\le& \|\boldsymbol{\Sigma}_{W,\mathcal{S}}^{-1}\|\left\{\sqrt{s+1}\max_{l\in\mathcal{S}}|\mathrm{tr}(\mathbf{W}_l\mathbf{W}_k)|\right\}\left\{\sqrt{s+1}\max_{l\in\mathcal{S}}|\mathrm{tr}(n^{-1}\sum_{i=1}^{n}\mathbf{W}_l\mathcal{E}_i)|\right\} \\
\le& \left\{(p\tau_{\min})^{-1}\right\}\left\{\sqrt{s+1}(pw^2)\right\}\left\{\sqrt{s+1}\max_{l\in\mathcal{S}}|\mathrm{tr}(n^{-1}\sum_{i=1}^{n}\mathbf{W}_l\mathcal{E}_i)|\right\} \\
=& \tau_{\min}^{-1}w^2(s+1)\max_{l\in\mathcal{S}}\big|n^{-1}\sum_{i=1}^{n}\mathbf{y}_i^\top\mathbf{W}_l\mathbf{y}_i - \mathrm{tr}(\mathbf{W}_l\boldsymbol{\Sigma}_0)\big|,
\end{aligned}$$

where the third inequality is due to inequality (A.18) in Lemma 1, and the last inequality is due to the following two facts: (i) by Condition (C4) and inequality (A.20) in Lemma 1, we have $|\mathrm{tr}(\mathbf{W}_l\mathbf{W}_k)| \le p\|\mathbf{W}_l\|\|\mathbf{W}_k\| \le pw^2$; (ii) by Condition (C2), we have $\big\|\boldsymbol{\Sigma}_{W,\mathcal{S}}^{-1}\big\| = \lambda_{\min}^{-1}(\boldsymbol{\Sigma}_{W,\mathcal{S}}) \le (p\tau_{\min})^{-1}$. Then by Lemma 2 and Conditions (C3) and (C4), we have $P\Big(|\mathbf{V}_k^\top\mathbb{H}_{\mathcal{S}}\overline{\mathbf{E}}| \ge p^2a_1\lambda/2\Big) \le$

$$\begin{aligned}
&\sum_{l\in\mathcal{S}}P\left\{\big|n^{-1}\sum_{i=1}^{n}\mathbf{y}_i^\top\mathbf{W}_l\mathbf{y}_i - \mathrm{tr}(\mathbf{W}_l\boldsymbol{\Sigma}_0)\big| > \frac{a_1\tau_{\min}p\lambda}{2(s+1)w^2}\right\} \\
\le& 2(s+1)\exp\left[-\min\left\{\frac{C_5a_1^2\tau_{\min}^2np\lambda^2}{w^6\sigma_{\max}^2(s+1)^2}, \frac{C_6a_1\tau_{\min}np\lambda}{w^3\sigma_{\max}(s+1)}\right\}\right]
\end{aligned}$$

Together with (A.14), we have

$$\begin{aligned}
\delta_1 \le& 2(K-s)\exp\left\{-\min\left(\frac{C_3a_1^2p\lambda^2}{w^2\sigma_{\max}^2}, \frac{C_4a_1p\lambda}{w\sigma_{\max}}\right)\right\} \\
&+ 2(K-s)(s+1)\exp\left[-\min\left\{\frac{C_5a_1^2\tau_{\min}^2p\lambda^2}{w^6\sigma_{\max}^2(s+1)^2}, \frac{C_6a_1\tau_{\min}p\lambda}{w^3\sigma_{\max}(s+1)}\right\}\right].
\end{aligned}$$

Step 2.3. We next bound $\delta_2 = P(E_2^c) = P(\|\widehat{\boldsymbol{\beta}}_{n,\mathcal{S}}^{\text{oracle}}\|_{\min} < \gamma\lambda)$. Note that $\widehat{\boldsymbol{\beta}}_{n,\mathcal{S}}^{\text{oracle}} = \boldsymbol{\beta}_{\mathcal{S}}^{(0)} + (\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\mathbb{V}_{\mathcal{S}}^\top\overline{\mathbf{E}}$, and thus $\|\widehat{\boldsymbol{\beta}}_{n,\mathcal{S}}^{\text{oracle}}\|_{\min} \geq \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \|(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\mathbb{V}_{\mathcal{S}}^\top\overline{\mathbf{E}}\|_\infty$. Then we have

$$\delta_2 \leq P\Big(\|(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\mathbb{V}_{\mathcal{S}}^\top\overline{\mathbf{E}}\|_\infty \geq \|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \gamma\lambda\Big). \tag{A.15}$$

Note that

$$\begin{aligned}
&\|(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\mathbb{V}_{\mathcal{S}}^\top\overline{\mathbf{E}}\|_\infty \leq \|(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\mathbb{V}_{\mathcal{S}}^\top\overline{\mathbf{E}}\| \leq \|(\mathbb{V}_{\mathcal{S}}^\top\mathbb{V}_{\mathcal{S}})^{-1}\|\|\mathbb{V}_{\mathcal{S}}^\top\overline{\mathbf{E}}\| \\
\leq&(p\tau_{\min})^{-1}\sqrt{s+1}\|\mathbb{V}_{\mathcal{S}}^\top\overline{\mathbf{E}}\|_\infty = \sqrt{s+1}(p\tau_{\min})^{-1}\max_{k\in\mathcal{S}}|n^{-1}\sum_{i=1}^n \mathbf{y}_i^\top\mathbf{W}_k\mathbf{y}_i - \operatorname{tr}(\mathbf{W}_k\boldsymbol{\Sigma}_0)|,
\end{aligned}$$

where the first inequality is due to inequality (A.18) in Lemma 1, and the third inequality is due to Condition (C2) and (A.18) in Lemma 1. Together with (A.15) and using Lemma 2, we have

$$\begin{aligned}
\delta_2 \leq&\sum_{k\in\mathcal{S}} P\left\{\Big|n^{-1}\sum_{i=1}^n \mathbf{y}_i^\top\mathbf{W}_k\mathbf{y}_i - \operatorname{tr}(\mathbf{W}_k\boldsymbol{\Sigma}_0)\Big| \geq \frac{\tau_{\min}p}{(s+1)^{1/2}}(\|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \gamma\lambda)\right\} \\
\leq&2(s+1)\exp\left[-\min\left\{\frac{C_5\tau_{\min}^2 np(\|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \gamma\lambda)^2}{w^2\sigma_{\max}^2(s+1)}, \frac{C_6\tau_{\min}np(\|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \gamma\lambda)}{w\sigma_{\max}(s+1)^{1/2}}\right\}\right].
\end{aligned}$$

This competes the proof of Step 2.

Step 3. To obtain the desired result, it suffices to prove that δ_1, δ_2, and δ_0' tend to 0 as $p \to \infty$ under the assumed conditions. By Condition (C1), we know that $\|\boldsymbol{\beta}_{\mathcal{S}}^{(0)}\|_{\min} - \gamma\lambda > \lambda$. Then, by inspecting the forms of upper bounds of $\delta_0, \delta_1, \delta_2$, it remains to prove that

$$\min\left\{\frac{np\lambda^2}{s^2}, \frac{np\lambda}{s}, \frac{np\lambda^2}{s}, \frac{np\lambda}{\sqrt{s}}, np\lambda_0^2, np\lambda_0,\right\}\Big/\log(K) \to 0 \tag{A.16}$$

as $p \to \infty$. Further note $\lambda \geq (3\sqrt{s+1}\lambda_0)/(a_0\kappa)$. Then we can easily verify that, (A.16) holds as long as $np\lambda_0^2/\{s\log(K)\} \to \infty$ as $np \to \infty$. This completes the proof of Step 3 and completes the proof of the theorem.

A.5 Useful Lemmas

Lemma 1. (NORM INEQUALITIES) *Let $\mathbf{v} \in \mathbb{R}^p$ be an arbitrary vector, and $\mathbf{\Delta} \in \mathbb{R}^{p\times p}$ be an arbitrary symmetric matrix. Then we should have*

$$\|\mathbf{v}\| \le \|\mathbf{v}\|_1 \le \sqrt{p}\|\mathbf{v}\|, \tag{A.17}$$

$$\|\mathbf{v}\|_\infty \le \|\mathbf{v}\| \le \sqrt{p}\|\mathbf{v}\|_\infty, \tag{A.18}$$

$$\|\mathbf{\Delta}\| \le \|\mathbf{\Delta}\|_F \le \sqrt{p}\|\mathbf{\Delta}\|, \tag{A.19}$$

$$\|\mathbf{\Delta}\| \le \|\mathbf{\Delta}\|_1 = \|\mathbf{\Delta}\|_\infty \le \sqrt{p}\|\mathbf{\Delta}\|. \tag{A.20}$$

Proof. The inequalities (A.17), (A.18), and (A.19) are directly from (2.2.5), (2.2.6), and (2.3.7) in (Golub and Van Loan, 2013, p. 69, 72), respectively. Since $\mathbf{\Delta}$ is symmetric, we immediately obtain that $\|\mathbf{\Delta}\|_1 = \|\mathbf{\Delta}\|_\infty$ by definitions of the two norms; see for example (2.3.9) and (2.3.10) in (Golub and Van Loan, 2013, p. 72). Then by Corollary 2.3.2 in (Golub and Van Loan, 2013, p. 73), we have

$$\|\mathbf{\Delta}\| \le \sqrt{\|\mathbf{\Delta}\|_1\|\mathbf{\Delta}\|_\infty} = \|\mathbf{\Delta}\|_1 = \mathbf{\Delta}\|_\infty.$$

The rightmost inequality $\|\mathbf{\Delta}\|_\infty \le \sqrt{p}\|\mathbf{\Delta}\|$ follows from (2.3.11) in (Golub and Van Loan, 2013, p. 72). This completes the proof. ☐

Lemma 2. (HANSON-WRIGHT INEQUALITY) *Let $\mathbf{y} = \mathbf{\Sigma}^{1/2}\mathbf{Z}$, where $\mathbf{Z} = (Z_1, \dots, Z_p)^\top \in \mathbb{R}^p$ is a random vector with independent and identically distributed sub-Gaussian coordinates. Assume that $E(Z_j) = 0$, $var(Z_j) = 1$ for each $1 \le j \le p$, and $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ is a positive definite matrix. Let $\mathbf{\Delta} \in \mathbb{R}^{p\times p}$ be a symmetric matrix. Then, for every $t \ge 0$,* *we have*

$$P\Big\{\big|\mathbf{y}^\top\mathbf{\Delta}\mathbf{y}-tr(\mathbf{\Delta}\mathbf{\Sigma})\big|\geq t\Big\}\leq 2\exp\left\{-\min\left(\frac{C_1t^2}{p\|\mathbf{\Delta}\|^2\|\mathbf{\Sigma}\|^2},\frac{C_2t}{\|\mathbf{\Delta}\|\|\mathbf{\Sigma}\|}\right)\right\},$$

where C_1 and C_2 are two positive constants. Furthermore, suppose that $\mathbf{y}_i$ $(1\leq i\leq n)$ are n independent copies of $\mathbf{y}$, then we have

$$P\Big\{\Big|n^{-1}\sum_{i=1}^n\mathbf{y}_i^\top\mathbf{\Delta}\mathbf{y}_i-tr(\mathbf{\Delta}\mathbf{\Sigma})\Big|\geq t\Big\}\leq 2\exp\left\{-\min\left(\frac{C_1nt^2}{p\|\mathbf{\Delta}\|^2\|\mathbf{\Sigma}\|^2},\frac{C_2nt}{\|\mathbf{\Delta}\|\|\mathbf{\Sigma}\|}\right)\right\}.$$

Proof. By using ordinary Hanson-Wright inequality (e.g., Theorem 6.2.1 in Vershynin, 2018), we have $P\big\{\big|\mathbf{y}^\top\mathbf{\Delta}\mathbf{y}-\mathrm{tr}(\mathbf{\Delta}\mathbf{\Sigma})\big|\geq t\big\}=$

$$P\Big\{\big|\mathbf{Z}^\top(\mathbf{\Sigma}^{1/2}\mathbf{\Delta}\mathbf{\Sigma}^{1/2})\mathbf{Z}-\mathrm{tr}(\mathbf{\Delta}\mathbf{\Sigma})\big|\geq t\Big\}\leq 2\exp\left\{-\min\left(\frac{C_1t^2}{\|\mathbf{\Sigma}^{1/2}\mathbf{\Delta}\mathbf{\Sigma}^{1/2}\|_F^2},\frac{C_2t}{\|\mathbf{\Sigma}^{1/2}\mathbf{\Delta}\mathbf{\Sigma}^{1/2}\|}\right)\right\}.$$

By norm inequality (A.19) in Lemma 1, we have $\|\mathbf{\Sigma}^{1/2}\mathbf{\Delta}\mathbf{\Sigma}^{1/2}\|_F^2\leq p\|\mathbf{\Sigma}^{1/2}\mathbf{\Delta}\mathbf{\Sigma}^{1/2}\|^2$. Further note that $\|\mathbf{\Sigma}^{1/2}\mathbf{\Delta}\mathbf{\Sigma}^{1/2}\|\leq\|\mathbf{\Sigma}^{1/2}\|^2\|\mathbf{\Delta}\|=\|\mathbf{\Delta}\|\|\mathbf{\Sigma}\|$. Then we can immediately obtain the first inequality of the lemma.

We next prove the second inequality of the lemma. Note that $\mathbf{y}_i=\mathbf{\Sigma}^{1/2}\mathbf{Z}_i$, where $\mathbf{Z}_i$ $(1\leq i\leq n)$ are n independent and identically distributed random vectors, and $\mathbb{Z}=(\mathbf{Z}_1^\top,\ldots,\mathbf{Z}_n^\top)^\top\in\mathbb{R}^{np}$ independent and identically distributed sub-Gaussian coordinates. Denote $\mathbb{A}=\mathbf{I}_n\otimes(\mathbf{\Sigma}^{1/2}\mathbf{\Delta}\mathbf{\Sigma}^{1/2})\in\mathbb{R}^{(np)\times(np)}$. Then, by using ordinary Hanson-Wright inequality, we have

$$\begin{aligned}&P\big\{\big|n^{-1}\sum_{i=1}^n\mathbf{y}_i^\top\mathbf{\Delta}\mathbf{y}_i-\mathrm{tr}(\mathbf{\Delta}\mathbf{\Sigma})\big|\geq t\big\}=P\Big\{\Big|\sum_{i=1}^n\mathbf{Z}_i^\top(\mathbf{\Sigma}^{1/2}\mathbf{\Delta}\mathbf{\Sigma}^{1/2})\mathbf{Z}_i-n\mathrm{tr}(\mathbf{\Delta}\mathbf{\Sigma})\Big|>nt\Big\}\\ =&P\Big\{\Big|\mathbb{Z}^\top\mathbb{A}\mathbb{Z}-\mathrm{tr}(\mathbb{A})\Big|>nt\Big\}\leq 2\exp\left\{-\min\left(\frac{C_1n^2t^2}{\|\mathbb{A}\|_F^2},\frac{C_2nt}{\|\mathbb{A}\|}\right)\right\}.\end{aligned}$$

By using the relationship between matrix norm and Kronecker product (e.g., results on Page 709 of Golub and Van Loan, 2013), we have $\|\mathbb{A}\|_F^2=\|\mathbf{I}_n\|_F^2\|\mathbf{\Sigma}^{1/2}\mathbf{\Delta}\mathbf{\Sigma}^{1/2}\|_F^2\leq$

$np\|\mathbf{\Delta}\|^2\|\mathbf{\Sigma}\|^2$, and $\|\mathbb{A}\| = \|\mathbf{I}_n\|\|\mathbf{\Sigma}^{1/2}\mathbf{\Delta}\mathbf{\Sigma}^{1/2}\| \leq \|\mathbf{\Delta}\|\|\mathbf{\Sigma}\|$. Then we can immediately obtain the second inequality of the lemma. This completes the proof of the lemma.

□

Lemma 3. *Let* $\mathbf{Z} = (Z_1, \ldots, Z_p)^\top \in \mathbb{R}^p$, *where* $Z_1, \ldots, Z_p$ *are independent and identically distributed with mean* 0 *and variance* 1. *Define*

$$S_p = \begin{pmatrix} vec^\top(\mathbf{\Delta}_1) \\ \vdots \\ vec^\top(\mathbf{\Delta}_L) \end{pmatrix} vec(\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}_p),$$

where $\mathbf{\Delta}_l \in \mathbb{R}^{p\times p}$ *is a symmetric matrix for* $1 \leq l \leq L$ *with* $L < \infty$. *Suppose that* $\sup_p \|\mathbf{\Delta}_l\|_1 < \infty$ *for* $1 \leq l \leq L$, *and* $E|Z_j|^{4+\eta} < \infty$ *for some* $\eta > 0$. *Then we have* $E(S_p) = 0$, *and*

$$cov(S_p) = 2\{tr(\mathbf{\Delta}_k\mathbf{\Delta}_l) : 1 \leq l \leq L\} + (\mu_4 - 3)\{tr(\mathbf{\Delta}_k \circ \mathbf{\Delta}_l) : 1 \leq k, l \leq L\},$$

where $\mu_4 = E(Z_j^4)$. *Moreover,* $p^{-1/2-\varepsilon}S_p \to^{L_2} 0$ *for ant* $\varepsilon > 0$. *In addition, assume that there is a positive definite matrix* $\mathbf{V} \in \mathbb{R}^{L\times L}$ *such that* $p^{-1}cov(S_p) \to \mathbf{V}$, *then we have* $p^{-1/2}S_p \to_d \mathcal{N}(\mathbf{0}, \mathbf{V})$ *as* $p \to \infty$.

Proof. This is directly modified from Lemma 4 in the supplementary material of Zou et al. (2021). □

A.6 Verification of Conditions (C2), (C5), and (C6)

We consider a specific example to verify Conditions (C2), (C5), and (C6). Specifically, we assume that $\mathbf{W}_k$ $(1 \leq k \leq K)$ are K similarity matrices independently generated as follows. More specifically, assume that $\mathbf{W}_k = (w_{k,j_1j_2}) \in \mathbb{R}^{p\times p}$ is a symmetric matrix, whose diagonal elements are set to be zeros, and off-diagonal elements are independently and identically generated from Bernoulli distributions with probability $\theta/(p-1) \in (0,1)$ for some constant $\theta \geq 1$. We then have the following lemma, which is useful for the subsequent verification of the conditions.

Lemma 4. *Let $\widehat{\omega}_{k_1k_2} = p^{-1}tr(\mathbf{W}_{k_1}\mathbf{W}_{k_2})$ for each $1 \leq k_1, k_2 \leq K$. Then for any $t \geq 0$, we have*

$$P\Big(|\widehat{\omega}_{kk} - \theta)| \geq t\Big) \leq 2\exp\left\{-\frac{pt^2}{4\theta + 4t/3}\right\}, \tag{A.21}$$

for any $1 \leq k \leq K$. In addition, for any $t \geq 2\theta^2/p$, we have

$$P\Big(|\widehat{\omega}_{k_1k_2}| \geq t\Big) \leq 2\exp\left\{-\frac{p\big(t - 2\theta^2/p\big)^2}{4\theta^2 + 4t/3}\right\}, \tag{A.22}$$

for any $k_1 \neq k_2$.

Proof. We first prove (A.21). In fact, we can compute that $\widehat{\omega}_{kk} = p^{-1}\mathrm{tr}(\mathbf{W}_k^2) = 2p^{-1}\sum_{j_1>j_2} w_{k,j_1j_2}^2 = 2p^{-1}\sum_{j_1>j_2} w_{k,j_1j_2}$, since w_{k,j_1j_2}s are Bernoulli random variables. Note that $E(w_{k,j_1j_2}) = \theta/(p-1)$ and $\mathrm{var}(w_{k,j_1j_2}) = \{\theta/(p-1)\}\{1-\theta/(p-1)\} \leq \theta/(p-1)$. Then by Bernstein's inequality for sum of independent bounded random variables (e.g., Theorem 2.8.4 in Vershynin, 2018), we have

$$P\left(\Big|\sum_{j_1>j_2}\Big(w_{k,j_1j_2} - \frac{\theta}{p-1}\Big)\Big| \geq t\right) \leq 2\exp\left\{-\frac{t^2/2}{p\theta/2 + t/3}\right\},$$

for any $t \geq 0$. By Replacing t with $pt/2$, we can directly obtain (A.21).

We next prove (A.22). Note that $\widehat{\omega}_{k_1k_2} = p^{-1}\mathrm{tr}(\mathbf{W}_{k_1}\mathbf{W}_{k_2}) = 2p^{-1}\sum_{j_1>j_2} w_{k_1,j_1j_2}w_{k_2,j_1j_2}$. Then it is easy to compute that $E(w_{k_1,j_1j_2}w_{k_2,j_1j_2}) = \theta^2/(p-1)^2$ and $\mathrm{var}(w_{k_1,j_1j_2}w_{k_2,j_1j_2}) \leq$

$\theta^2/(p-1)^2$. Similarly, by using Bernstein's inequality we have

$$P\left(\left|\sum_{j_1>j_2}\left(w_{k_1,j_1j_2}w_{k_2,j_1j_2}-\frac{\theta^2}{(p-1)^2}\right)\right|\geq t\right)\leq 2\exp\left\{-\frac{t^2/2}{\theta^2+t/3}\right\},$$

for any $t\geq 0$. By Replacing t with $pt/2$, we can obtain that

$$P\Big(\Big|\widehat{\omega}_{k_1k_2}-\theta^2/(p-1)\Big|\geq t\Big)\leq 2\exp\left\{-\frac{pt^2}{8\theta^2/p+4t/3}\right\}.$$

Then by using $(p-1)^{-1}\leq 2/p$ for $p\geq 2$, we can derive that for any $t\geq 2\theta^2/p$,

$$P\Big(|\widehat{\omega}_{k_1k_2}|\geq t\Big)\leq P\Big(|\widehat{\omega}_{k_1k_2}-\theta^2/(p-1)|\geq t-\theta^2/(p-1)\Big)\leq 2\exp\left\{-\frac{p\big(t-2\theta^2/p\big)^2}{4\theta^2+4t/3}\right\}.$$

This proves (A.22) and completes the proof of the lemma. □

Verification of Condition (C2). Define $\widehat{\mathbf{\Omega}}_{\mathcal{S}}=p^{-1}\mathbf{\Sigma}_{W,\mathcal{S}}=(\widehat{\omega}_{k_1k_2})\in\mathbb{R}^{(s+1)\times(s+1)}$ with $\widehat{\omega}_{k_1k_2}=p^{-1}\mathrm{tr}(\mathbf{W}_{k_1}\mathbf{W}_{k_2})$ for $k_1,k_2\in\mathcal{S}$. Recall that $\mathbf{W}_0=\mathbf{I}_p$. Then one can easily verify that $\widehat{\omega}_{k0}=\widehat{\omega}_{0k}=1$ if $k=1$ and $\widehat{\omega}_{k0}=\widehat{\omega}_{0k}=0$ otherwise. Further define $\mathbf{\Omega}_{\mathcal{S}}=\mathrm{diag}\{1,\theta,\ldots,\theta\}\in\mathbb{R}^{(s+1)\times(s+1)}$. Then by Lemma 4, we know that

$$P\left\{\|\widehat{\mathbf{\Omega}}_{\mathcal{S}}-\mathbf{\Omega}_{\mathcal{S}}\|_{\max}\geq t\right\}\leq 2s^2\exp\left\{-\frac{p\big(t-2\theta^2/p\big)^2}{4\theta^2+4t/3}\right\},$$

for any $t\geq 2\theta^2/p$. Here, $\|\mathbf{M}\|_{\max}=\max_{i,j}|m_{ij}|$ denotes the element-wise max-norm for an arbitrary matrix $\mathbf{M}=(m_{ij})$. This implies that $\mathbf{\Omega}_{\mathcal{S}}$ should be the probabilistic limit of $\widehat{\mathbf{\Omega}}_{\mathcal{S}}$. By matrix norm inequality, we know that $\|\widehat{\mathbf{\Omega}}_{\mathcal{S}}-\mathbf{\Omega}_{\mathcal{S}}\|\leq(s+1)\|\widehat{\mathbf{\Omega}}_{\mathcal{S}}-\mathbf{\Omega}_{\mathcal{S}}\|_{\max}$. Since $2s\geq s+1$, we can deduce that

$$P\left\{\|\widehat{\mathbf{\Omega}}_{\mathcal{S}}-\mathbf{\Omega}_{\mathcal{S}}\|\geq t\right\}\leq P\left\{\|\widehat{\mathbf{\Omega}}_{\mathcal{S}}-\mathbf{\Omega}_{\mathcal{S}}\|_{\max}\geq t/(s+1)\right\}\leq 2s^2\exp\left\{-\frac{p\big\{t/(2s)-2\theta^2/p\big\}^2}{4\theta^2+4t/3}\right\},$$

for any $t\geq 4\theta^2 s/p$. This implies that $\lambda_{\min}(\widehat{\mathbf{\Omega}}_{\mathcal{S}})\geq\lambda_{\min}(\mathbf{\Omega}_{\mathcal{S}})-\|\widehat{\mathbf{\Omega}}_{\mathcal{S}}-\mathbf{\Omega}_{\mathcal{S}}\|\to_p 1$ as

$p \to \infty$, provided $p/\{s^2 \log(s)\} \to \infty$ as $p \to \infty$. Consequently, we should expect that Condition (C2) holds with high probability.

Verification of Condition (C5). Similarly, define $\widehat{\mathbf{\Omega}} = p^{-1}\mathbf{\Sigma}_W = (\widehat{\omega}_{k_1k_2}) \in \mathbb{R}^{(K+1)\times(K+1)}$ with $\widehat{\omega}_{k_1k_2} = p^{-1}\mathrm{tr}(\mathbf{W}_{k_1}\mathbf{W}_{k_2})$ for $0 \le k_1, k_2 \le K$. Recall that $\boldsymbol{\delta} \in \mathbb{C}_3(\mathcal{S}) \stackrel{\text{def}}{=} \{\boldsymbol{\delta} \in \mathbb{R}^{K+1} : \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1 \le 3\|\boldsymbol{\delta}_{\mathcal{S}}\|_1\}$. Let $\mathcal{T} \subset \mathcal{S}^c$ collect the indexes of the $s+1$ largest $|\delta_k|$ in $\mathcal{S}^c$. Further define $\overline{\mathcal{S}} = \mathcal{S} \cup \mathcal{T}$. Then we should have

$$\begin{aligned}\frac{1}{p}\left\|\sum_{k=0}^{K}\delta_k\mathbf{W}_k\right\|_F^2 =&\frac{1}{p}\left\|\sum_{k\in\overline{\mathcal{S}}}\delta_k\mathbf{W}_k\right\|_F^2 + 2\sum_{k_1\in\overline{\mathcal{S}}}\sum_{k_2\in\overline{\mathcal{S}}^c}\delta_{k_1}\delta_{k_2}\widehat{\omega}_{k_1k_2} + \frac{1}{p}\left\|\sum_{k\in\overline{\mathcal{S}}^c}\delta_k\mathbf{W}_k\right\|_F^2\\ \ge&\frac{1}{p}\left\|\sum_{k\in\overline{\mathcal{S}}}\delta_k\mathbf{W}_k\right\|_F^2 + 2\sum_{k_1\in\overline{\mathcal{S}}}\sum_{k_2\in\overline{\mathcal{S}}^c}\delta_{k_1}\delta_{k_2}\widehat{\omega}_{k_1k_2} = Q_1 + Q_2.\end{aligned}$$

We next investigate Q_1 and Q_2, respectively.

Let $\widehat{\mathbf{\Omega}}_{\overline{\mathcal{S}}} = \big(\widehat{\omega}_{k_1k_2} : k_1, k_2 \in \overline{\mathcal{S}}\big) \in \mathbb{R}^{(2s+2)\times(2s+2)}$ be the sub-matrix of $\widehat{\mathbf{\Omega}}$. Similarly, let $\mathbf{\Omega}_{\overline{\mathcal{S}}} = \mathrm{diag}\{1, \theta, \dots, \theta\} \in \mathbb{R}^{(2s+2)\times(2s+2)}$. Then by similar procedures in the verification of Condition (C2), we can derive that $\|\widehat{\mathbf{\Omega}}_{\overline{\mathcal{S}}} - \mathbf{\Omega}_{\overline{\mathcal{S}}}\| \to_p 0$ as long as $p/\{s^2\log(s)\} \to \infty$ as $p \to \infty$. Then it follows that

$$Q_1 = \frac{1}{p}\left\|\sum_{k\in\overline{\mathcal{S}}}\delta_k\mathbf{W}_k\right\|_F^2 = \boldsymbol{\delta}_{\overline{\mathcal{S}}}^\top\widehat{\mathbf{\Omega}}_{\overline{\mathcal{S}}}\boldsymbol{\delta}_{\overline{\mathcal{S}}} \ge \lambda_{\min}(\mathbf{\Omega}_{\overline{\mathcal{S}}})\|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\|^2 + \boldsymbol{\delta}_{\overline{\mathcal{S}}}^\top(\widehat{\mathbf{\Omega}}_{\overline{\mathcal{S}}} - \mathbf{\Omega}_{\overline{\mathcal{S}}})\boldsymbol{\delta}_{\overline{\mathcal{S}}} = \|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\|^2\{1 + o_p(1)\},$$

as long as $p/\{s^2\log(s)\} \to \infty$ as $p \to \infty$.

For the term Q_2, we can derive that

$$\begin{aligned}|Q_2| = &\left|2\sum_{k_1\in\overline{\mathcal{S}}}\sum_{k_2\in\overline{\mathcal{S}}^c}\delta_{k_1}\delta_{k_2}\widehat{\omega}_{k_1k_2}\right| \le 4(s+1)\max_{k_1\in\overline{\mathcal{S}}}|\delta_{k_1}| \cdot \max_{k_1\in\overline{\mathcal{S}}, k_2\in\overline{\mathcal{S}}^c}|\widehat{\omega}_{k_1k_2}| \cdot \sum_{k_2\in\overline{\mathcal{S}}^c}|\delta_{k_2}|\\ \le&4(s+1)\|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\| \cdot \max_{k_1\in\overline{\mathcal{S}}, k_2\in\overline{\mathcal{S}}^c}|\widehat{\omega}_{k_1k_2}| \cdot \|\boldsymbol{\delta}_{\overline{\mathcal{S}}^c}\|_1 \le 12(s+1)^{3/2}\|\boldsymbol{\delta}\|^2 \cdot \max_{k_1\in\overline{\mathcal{S}}, k_2\in\overline{\mathcal{S}}^c}|\widehat{\omega}_{k_1k_2}|,\end{aligned}$$

where we have used the facts that $\|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\| \le \|\boldsymbol{\delta}\|$ and $\|\boldsymbol{\delta}_{\overline{\mathcal{S}}^c}\|_1 \le \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1 \le 3\|\boldsymbol{\delta}_{\mathcal{S}}\|_1 \le$

$3(s+1)^{1/2}\|\boldsymbol{\delta}_{\mathcal{S}}\| \leq 3(s+1)^{1/2}\|\boldsymbol{\delta}\|$. By (A.22) in Lemma 4, we know that

$$P\Big(\max_{k_1\in\overline{\mathcal{S}},k_2\in\overline{\mathcal{S}}^c} |\widehat{\omega}_{k_1k_2}| \geq t\Big) \leq 4(s+1)(K-2s-1)\exp\left\{-\frac{p\big(t-2\theta^2/p\big)^2}{4\theta^2+4t/3}\right\},$$

for any $t \geq 2\theta^2/p$. Hence, we should have $\max_{k_1\in\overline{\mathcal{S}},k_2\in\overline{\mathcal{S}}^c} |\widehat{\omega}_{k_1k_2}| = O_p(\sqrt{\log(Ks)/p})$. This indicates that $|Q_2| = o_p(\|\boldsymbol{\delta}\|^2)$ as long as $p/\{s^3\log(Ks)\} \to \infty$ as $p \to \infty$.

By far, we have shown that $p^{-1}\left\|\sum_{k=0}^K \delta_k \mathbf{W}_k\right\|_F^2 \geq \|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\|^2\{1+o_p(1)\} + o_p(\|\boldsymbol{\delta}\|^2) = \|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\|^2 + o_p(\|\boldsymbol{\delta}\|^2)$. Thus, if we can show that $\|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\|^2 \geq \kappa\|\boldsymbol{\delta}\|^2$ for some $\kappa > 0$ and $\boldsymbol{\delta} \in \mathbb{C}_3(\mathcal{S})$, then Condition (C5) should hold with high probability. In fact, by Lemma 2.2 of van de Geer and Bühlmann (2009), we have $\|\boldsymbol{\delta}_{\overline{\mathcal{S}}^c}\| \leq (s+1)^{-1/2}\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1$. Since $\boldsymbol{\delta} \in \mathbb{C}_3(\mathcal{S})$, it follows that $\|\boldsymbol{\delta}_{\overline{\mathcal{S}}^c}\| \leq 3(s+1)^{-1/2}\|\boldsymbol{\delta}_{\mathcal{S}}\|_1 \leq 3\|\boldsymbol{\delta}_{\mathcal{S}}\| \leq 3\|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\|$, where we have used $\|\boldsymbol{\delta}_{\mathcal{S}}\|_1 \leq (s+1)^{1/2}\|\boldsymbol{\delta}_{\mathcal{S}}\|$ in the second inequality. Then we should have $\|\boldsymbol{\delta}\|^2 = \|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\|^2 + \|\boldsymbol{\delta}_{\overline{\mathcal{S}}^c}\|^2 \leq 10\|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\|^2$, or equivalently, $\|\boldsymbol{\delta}_{\overline{\mathcal{S}}}\|^2 \geq 0.1\|\boldsymbol{\delta}\|^2$. Combine above results, we can obtain that $p^{-1}\left\|\sum_{k=0}^K \delta_k \mathbf{W}_k\right\|_F^2 \geq 0.1\|\boldsymbol{\delta}\|^2 + o_p(\|\boldsymbol{\delta}\|^2)$, as long as $p/\{s^3\log(Ks)\} \to \infty$ as $p \to \infty$. Thus, we should expect that RE Condition (C5) holds with high probability.

Verification of Condition (C6). We consider a special case that $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\beta}^{(0)}) = \beta_0^{(0)}\mathbf{I}_p + \beta_1^{(0)}\mathbf{W}_1$ with $\beta_0^{(0)}, \beta_1^{(0)} > 0$. By our above results, we can show that $\mathbf{G}_{0,p} = p^{-1}\boldsymbol{\Sigma}_{W,\mathcal{S}} \to_p \mathbf{G}_0 \stackrel{\text{def}}{=} \text{diag}\{1,\theta\}$, which is positive definite. In addition, we have

$$\mathbf{G}_{1,p} = p^{-1}\begin{bmatrix} \text{tr}(\boldsymbol{\Sigma}_0^2) & \text{tr}(\boldsymbol{\Sigma}_0^2\mathbf{W}_1) \\ \text{tr}(\boldsymbol{\Sigma}_0^2\mathbf{W}_1) & \text{tr}\{(\boldsymbol{\Sigma}_0\mathbf{W}_1)^2\} \end{bmatrix}.$$

We next examine each entry of $\mathbf{G}_{1,p}$. First, we can compute that $p^{-1}\text{tr}(\boldsymbol{\Sigma}_0^2) = (\beta_0^{(0)})^2 + p^{-1}\text{tr}(\mathbf{W}_1^2)(\beta_1^{(0)})^2 \to_p (\beta_0^{(0)})^2 + \theta(\beta_1^{(0)})^2$. For the off-diagonal entries, we shoud have $p^{-1}\text{tr}(\boldsymbol{\Sigma}_0^2\mathbf{W}_1) = 2p^{-1}\text{tr}(\mathbf{W}_1^2)\beta_0^{(0)}\beta_1^{(0)} + p^{-1}\text{tr}(\mathbf{W}_1^3)(\beta_1^{(0)})^2$. By Corollary 2.1.2 of Aguilar (2021), we can show that $p^{-1}\text{tr}(\mathbf{W}_1^3) \to_p 0$. Then we should have $p^{-1}\text{tr}(\boldsymbol{\Sigma}_0^2\mathbf{W}_1) \to_p$

$2\theta\beta_0^{(0)}\beta_1^{(0)}$. Last, note that $p^{-1}\mathrm{tr}\{(\boldsymbol{\Sigma}_0\mathbf{W}_1)^2\} = p^{-1}\mathrm{tr}(\mathbf{W}_1^2)(\beta_0^{(0)})^2 + 2p^{-1}\mathrm{tr}(\mathbf{W}_1^3)\beta_0^{(0)}\beta_1^{(0)} + p^{-1}\mathrm{tr}(\mathbf{W}_1^4)(\beta_1^{(0)})^2$. By Corollary 2.1.2 of Aguilar (2021), we can show that $p^{-1}\mathrm{tr}(\mathbf{W}^4) \to_p 2\theta^2 + \theta$. Then we should have $p^{-1}\mathrm{tr}\{(\boldsymbol{\Sigma}_0\mathbf{W}_1)^2\} \to_p \theta(\beta_0^{(0)})^2 + (2\theta^2+\theta)(\beta_1^{(0)})^2$. Thus, we obtain that $\mathbf{G}_{1,p} \to_p \mathbf{G}_1$ with

$$\mathbf{G}_1 = \begin{bmatrix} (\beta_0^{(0)})^2 + \theta(\beta_1^{(1)})^2 & 2\theta\beta_0^{(0)}\beta_1^{(0)} \\ 2\theta\beta_0^{(0)}\beta_1^{(0)} & \theta(\beta_0^{(0)})^2 + (2\theta^2+\theta)(\beta_1^{(1)})^2 \end{bmatrix}.$$

It can be verified that the determinant $|\mathbf{G}_1| > 0$, which implies $\mathbf{G}_1$ is also positive definite. This indicates that Condition (C6) (i) can hold with high probability.

We next verify Condition (C6) (ii). Suppose the eigen-decomposition of $\mathbf{W}_1$ is $\mathbf{W}_1 = \mathbf{V}\mathbf{D}\mathbf{V}^\top$, where $\mathbf{V}$ is an orthogonal matrix, and $\mathbf{D}$ is a diagonal matrix collecting the eigenvalues of $\mathbf{W}_1$. Then we can derive that,

$$\begin{aligned}
&\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_1\boldsymbol{\Sigma}_0^{1/2} = (\beta_0^{(0)}\mathbf{I}_p + \beta_1^{(0)}\mathbf{W}_1)^{1/2}\mathbf{W}_1(\beta_0^{(0)}\mathbf{I}_p + \beta_1^{(0)}\mathbf{W}_1)^{1/2} \\
=&\beta_0^{(0)}\mathbf{V}\Big\{\mathbf{I}_p + (\beta_1^{(0)}/\beta_0^{(0)})\mathbf{D}\Big\}^{1/2}\mathbf{V}^\top\Big(\mathbf{V}\mathbf{D}\mathbf{V}^\top\Big)\mathbf{V}\Big\{\mathbf{I}_p + (\beta_1^{(0)}/\beta_0^{(0)})\mathbf{D}\Big\}^{1/2}\mathbf{V}^\top \\
=&\beta_0^{(0)}\mathbf{V}\Big\{\mathbf{I}_p + (\beta_1^{(0)}/\beta_0^{(0)})\mathbf{D}\Big\}^{1/2}\mathbf{D}\Big\{\mathbf{I}_p + (\beta_1^{(0)}/\beta_0^{(0)})\mathbf{D}\Big\}^{1/2}\mathbf{V}^\top \\
=&\beta_0^{(0)}\mathbf{V}\Big\{\mathbf{D} + (\beta_1^{(0)}/\beta_0^{(0)})\mathbf{D}^2\Big\}\mathbf{V}^\top = \beta_0^{(0)}\mathbf{W}_1 + \beta_1^{(0)}\mathbf{W}_1^2.
\end{aligned}$$

Consequently, it follows that

$$\begin{aligned}
\mathbf{H}_p =& p^{-1}\begin{bmatrix} \mathrm{tr}(\boldsymbol{\Sigma}_0 \circ \boldsymbol{\Sigma}_0) & \mathrm{tr}\{(\boldsymbol{\Sigma}_0 \circ (\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_1\boldsymbol{\Sigma}_0^{1/2})\} \\ \mathrm{tr}\{(\boldsymbol{\Sigma}_0 \circ (\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_1\boldsymbol{\Sigma}_0^{1/2})\} & \mathrm{tr}\{(\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_1\boldsymbol{\Sigma}_0^{1/2}) \circ (\boldsymbol{\Sigma}_0^{1/2}\mathbf{W}_1\boldsymbol{\Sigma}_0^{1/2})\} \end{bmatrix} \\
=& \begin{bmatrix} (\beta_0^{(0)})^2 & p^{-1}\mathrm{tr}(\mathbf{W}_1^2)\beta_0^{(0)}\beta_1^{(0)} \\ p^{-1}\mathrm{tr}(\mathbf{W}_1^2)\beta_0^{(0)}\beta_1^{(0)} & p^{-1}\mathrm{tr}(\mathbf{W}_1^2 \circ \mathbf{W}_1^2)(\beta_1^{(0)})^2 \end{bmatrix}.
\end{aligned}$$

Recall that $p^{-1}\mathrm{tr}(\mathbf{W}_1^2) \to_p \theta$. We can also derive that $p^{-1}\mathrm{tr}(\mathbf{W}_1^2 \circ \mathbf{W}_1^2) \to_p \theta^2 + \theta$. Then we should have $\mathbf{H}_p \to_p \mathbf{H}$ with

$$\mathbf{H} = \begin{bmatrix} (\beta_0^{(0)})^2 & \theta\beta_0^{(0)}\beta_1^{(0)} \\ \theta\beta_0^{(0)}\beta_1^{(0)} & (\theta^2+\theta)(\beta_1^{(0)})^2 \end{bmatrix}.$$

One can easily verify that the determinant $|\mathbf{H}| > 0$, which implies $\mathbf{H}$ is also positive definite. This indicates that Condition (C6) (ii) can also hold with high probability.

A.7 Additional Simulation Results

In this subsection, we conduct three additional experiments to better evaluate our method. For the first two experiments, we try two different data generation processes of the components of $\mathbf{Z}$, while holding other simulation settings in Section 5.1 unchanged. Specifically, the components of $\mathbf{Z}$ are assumed to be independently and identically generated from a mixture normal distribution $\xi \cdot \mathcal{N}(0, 5/9) + (1-\xi) \cdot \mathcal{N}(0, 5)$ with $P(\xi = 1) = 0.9$ and $P(\xi = 0) = 0.1$, or a standardized exponential distribution $\mathrm{Exp}(1) - 1$. The simulation results are presented in Tables A.1–A.2, respectively. For the third experiment, we construct $\mathbf{W}_k$s with moderate correlation , while generating $\mathbf{Z}$ from the standard normal distribution and holding other simulation settings in Section 5.1 unchanged. Specifically, we independently generate each $\mathbf{x}_j = (X_{j1}, \ldots, X_{jK})^\top \in \mathbb{R}^K$ $(1 \le j \le p)$ from the multivariate normal distribution $\mathcal{N}_K(\mathbf{0}, \mathbf{\Sigma}_x)$, where $\mathbf{\Sigma}_x = (0.5^{|k_1-k_2|})_{1\le k_1,k_2\le K} \in \mathbb{R}^{K\times K}$. Then we should have X_{jk}s with the same j but different k are linearly correlated with $\mathrm{corr}(X_{j,k_1}, X_{j,k_2}) = 0.5^{|k_1-k_2|}$. We then construct $\mathbf{W}_k = (w_{k,j_1j_2})_{1\le j_1,j_2\le p} \in \mathbb{R}^{p\times p}$ with $w_{k,j_1,j_2} = X_{j_1,k}X_{j_2,k} \times \exp\{-p(X_{j_1,k} - X_{j_2,k})^2\}$ for each $1 \le k \le K$. The simulation results are presented in Table A.3. By the three tables, we can see that all the results are qualitatively similar to those in Table 1 of the main text. This further demonstrates the robustness and broad applicability of our proposed method.

Table A.1: Simulation results for $\mathbf{Z}$ generated from the mixture normal distribution.

(p, K)	Penalty	TPR	FPR	CS	RMSE	Bias	SD	$\|\cdot\|_2$	$\|\cdot\|_F$
(200,10)	SCAD	0.787	0.061	0.290	0.602	0.052	0.596	8.053	2.883
	MCP	0.790	0.060	0.290	0.602	0.052	0.596	8.037	2.875
	OLS	–	–	–	0.616	0.049	0.612	8.090	3.057
	ORACLE	1.000	0.000	1.000	0.535	0.026	0.531	5.403	2.058
(500,100)	SCAD	0.927	0.060	0.580	0.125	0.004	0.124	6.093	1.883
	MCP	0.927	0.060	0.580	0.125	0.004	0.125	6.130	1.885
	OLS	–	–	–	0.250	0.018	0.249	19.142	5.305
	ORACLE	1.000	0.000	1.000	0.105	0.001	0.105	3.973	1.356
(1000,1000)	SCAD	0.993	0.047	0.800	0.025	0.000	0.025	3.466	1.113
	MCP	0.993	0.047	0.800	0.025	0.000	0.025	3.460	1.112
	OLS	–	–	–	0.161	0.013	0.160	31.005	11.299
	ORACLE	1.000	0.000	1.000	0.022	0.000	0.022	2.482	0.878

A.8 Selection of Tuning Parameters

To implement the LLA algorithm, we need first compute the Lasso estimator (2.4) as an initial estimator. This requires selecting two tuning parameters: λ_0 for the Lasso estimator, and λ in the folded concave penalized loss function (2.5). We can separately select the two tuning parameters λ_0 and λ. However, this approach can be very time-consuming because we need to consider all possible pairs (λ_0, λ). In addition, we can expect that $\lambda \asymp \lambda_0$ as remarked at the end of Appendix A.1 Therefore, another approach is to select a single value for both λ_0 and λ by setting $\lambda_0 = \lambda$. We conducted a preliminary experiment to assess the performance of the two approaches. Specifically, we adopt the same simulation setting as in Section 5.1 with $(p, K) = (200, 10)$ and $\mathbf{Z}$ generated from a normal distribution. For both approaches, we use the

Table A.2: Simulation results for $\mathbf{Z}$ generated from the standardized exponential distribution.

(p, K)	Penalty	TPR	FPR	CS	RMSE	Bias	SD	$\|\cdot\|_2$	$\|\cdot\|_F$
(200,10)	SCAD	0.823	0.074	0.260	0.635	0.058	0.630	7.938	2.886
	MCP	0.820	0.070	0.280	0.635	0.059	0.630	7.922	2.870
	OLS	–	–	–	0.644	0.045	0.642	8.958	3.038
	ORACLE	1.000	0.000	1.000	0.573	0.023	0.571	5.564	2.098
(500,100)	SCAD	0.940	0.076	0.510	0.124	0.005	0.123	5.146	1.782
	MCP	0.938	0.074	0.510	0.124	0.005	0.123	5.183	1.788
	OLS	–	–	–	0.247	0.019	0.246	15.220	5.166
	ORACLE	1.000	0.000	1.000	0.104	0.001	0.104	3.240	1.198
(1000,1000)	SCAD	0.995	0.034	0.830	0.027	0.000	0.027	3.339	1.132
	MCP	0.995	0.034	0.830	0.027	0.000	0.027	3.339	1.132
	OLS	–	–	–	0.162	0.013	0.161	29.949	11.331
	ORACLE	1.000	0.000	1.000	0.025	0.000	0.025	2.757	0.973

Table A.3: Simulation results for $\mathbf{Z}$ generated from the standard normal distribution and $\mathbf{W}_k$s constructed with moderate correlation.

(p, K)	Penalty	TPR	FPR	CS	RMSE	Bias	SD	$\|\cdot\|_2$	$\|\cdot\|_F$
(200,10)	SCAD	0.588	0.103	0.060	0.793	0.164	0.748	18.883	4.172
	MCP	0.575	0.115	0.050	0.830	0.182	0.776	18.925	4.222
	OLS	–	–	–	0.833	0.062	0.826	18.902	4.398
	ORACLE	1.000	0.000	1.000	0.619	0.043	0.610	15.865	3.277
(500,100)	SCAD	0.745	0.054	0.160	0.210	0.021	0.155	18.136	3.615
	MCP	0.733	0.051	0.150	0.218	0.023	0.150	18.234	3.679
	OLS	–	–	–	0.453	0.022	0.451	26.706	7.355
	ORACLE	1.000	0.000	1.000	0.118	0.004	0.115	12.488	2.322
(1000,1000)	SCAD	0.845	0.093	0.280	0.066	0.002	0.039	17.189	3.281
	MCP	0.848	0.087	0.320	0.068	0.003	0.038	17.068	3.311
	OLS	–	–	–	0.264	0.013	0.263	56.135	15.673
	ORACLE	1.000	0.000	1.000	0.024	0.000	0.024	10.051	1.751

Table A.4: Simulation results for two different tuning parameter selection approaches. Approach (I) is to separately select λ_0 and λ, and Approach (II) is to select a single value for both λ_0 and λ.

Approach	Penalty	TPR	FPR	CS	RMSE	Bias	SD	$\|\cdot\|_2$	$\|\cdot\|_F$
(I)	SCAD	0.796	0.069	0.235	0.464	0.051	0.458	7.667	2.642
(II)	SCAD	0.792	0.070	0.230	0.465	0.053	0.459	7.732	2.656
(I)	MCP	0.796	0.070	0.230	0.464	0.051	0.458	7.690	2.645
(II)	MCP	0.794	0.071	0.220	0.465	0.053	0.459	7.730	2.656

BIC-type criterion (5.1). We replicate the experiment 200 times and compute the same measurements as those in Table 1. The results are given in Table A.4. From Table A.4, we observe that the results of Approach (I) are slightly better than Approach (II). This is expected because Approach (I) explores all possible pairs (λ_0, λ), while Approach (II) only considers pairs with $\lambda_0 = \lambda$. Nevertheless, the two approaches perform very similarly for both the SCAD and MCP estimators. In addition, Approach (II) requires less computational time. Consequently, we adopt Approach (II) in the subsequent simulation experiments and real data analysis.

References

Aguilar, C. O. (2021), "An Introduction to Algebraic Graph Theory," *New York: Geneseo*, 41–57.

Fan, J., Xue, L., and Zou, H. (2014), "Strong oracle optimality of folded concave penalized estimation," *Annals of Statistics*, 42, 819–849.

Golub, G. and Van Loan, C. (2013), *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, 4th ed.

van de Geer, S. A. and Bühlmann, P. (2009), "On the conditions used to prove oracle results for the Lasso," *Electronic Journal of Statistics*, 3, 1360–1392.

Vershynin, R. (2018), *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Wainwright, M. (2019), *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Zhang, C.-H. and Zhang, T. (2012), "A general theory of concave regularization for high-dimensional sparse estimation problems," *Statistical Science*, 27, 576–593.

Zou, T., Luo, R., Lan, W., and Tsai, C.-L. (2021), "Network influence analysis," *Statistica Sinica*, 31, 1727–1748.