Partitioned Approach for High-dimensional Confidence Intervals with Large Split Sizes

University of Science and Technology of China

Supplementary Material

S1 Proof of Proposition 1

We would like to apply a similar argument as that in the proof of Zhang and Zhang (2014, Theorem 1) to derive the confidence intervals of β_j. The fundamental difference is that the design matrix $\mathbf{X}$ is now random instead of fixed. Thus, the statistics related to $\mathbf{X}$ such as $\mathbf{z}_j$, η_j and τ_j are also random variables (vectors). We will derive the properties of these statistics before deriving the confidence intervals of β_j.

Part 1: Deviation bounds of $\|\mathbf{z}_j\|_2^2$. Recall that $\eta_j = \max_{k\neq j} |\mathbf{z}_j^T\mathbf{x}_k|/\|\mathbf{z}_j\|_2$, $\tau_j = \|\mathbf{z}_j\|_2/|\mathbf{z}_j^T\mathbf{x}_j|$, defined in (3.5), and $\mathbf{z}_j$ is the relaxed residual vector of regressing $\mathbf{x}_j$ on $\mathbf{X}_{-j}$ in (3.4) such that

$$\mathbf{z}_j = \mathbf{x}_j - \mathbf{X}_{-j}\widehat{\boldsymbol{\gamma}}_j,$$

$$\{\widehat{\boldsymbol{\gamma}}_j, \hat{\sigma}_j\} = \underset{\mathbf{b}\in\mathbb{R}^{p-1},\sigma_j\in\mathbb{R}^+}{\arg\min}\left\{\frac{\|\mathbf{x}_j - \mathbf{X}_{-j}\mathbf{b}\|_2^2}{2n\sigma_j} + \frac{\sigma_j}{2} + \lambda_0\sum_{k\neq j}\frac{\|\mathbf{x}_k\|_2}{\sqrt{n}}|b_k|\right\},$$

with components of $\widehat{\boldsymbol{\gamma}}_j = \{\hat{\gamma}_{j,k}; k = 1,\cdots,p, k\neq j\}$, where the regularization parameter $\lambda_0 = (1+\varepsilon)\sqrt{2\delta\log(p)/n}$ for some $\delta \geq 1$ and $\varepsilon > 0$.

We first derive the deviation bound for $\|\mathbf{z}_j\|_2^2$. Note that $\mathbf{X} = (x_{ij})_{n\times p} = (\mathbf{x}_1,\cdots,\mathbf{x}_p)$, where the rows of $\mathbf{X}$ are i.i.d. from $N(\mathbf{0}, \mathbf{\Sigma})$. Let $\mathbf{\Sigma} = (\sigma_{ij})_{p\times p}$ and $\mathbf{x}_{i,-j}$ be the ith row of $\mathbf{X}$ after taking the jth component off. Similarly, the notation $\mathbf{\Sigma}^{-1}_{j,-j}$ denotes a subvector of the jth row of $\mathbf{\Sigma}^{-1}$ without the jth component. Let $\sigma_j = 1/\mathbf{\Sigma}^{-1}_{j,j}$. By the conditional distribution of multivariate normal vector, we have

$$x_{ij}|\mathbf{x}_{i,-j} = N(-\sigma_j \mathbf{x}_{i,-j}(\mathbf{\Sigma}^{-1}_{j,-j})^T, \sigma_j),$$

independent over i. It follows that $x_{ij} = -\sigma_j \mathbf{x}_{i,-j}(\mathbf{\Sigma}^{-1}_{j,-j})^T + \rho_{ij}$, where $\rho_{ij} \sim N(0, \sigma_j)$ are i.i.d. over i. Denote by $\boldsymbol{\gamma}_j = -\sigma_j(\mathbf{\Sigma}^{-1}_{j,-j})^T$ and $\boldsymbol{\rho}_j = (\rho_{1j}, \cdots, \rho_{nj})^T$. In matrix notation, we have

$$\mathbf{x}_j = \mathbf{X}_{-j}\boldsymbol{\gamma}_j + \boldsymbol{\rho}_j,$$

with components of $\gamma_j = \{\gamma_{j,k}; k = 1, \cdots, p, k \neq j\}$, where $\mathbf{X}_{-j}$ is the submatrix of $\mathbf{X}$ by taking the jth column off.

Note that $\mathbf{z}_j$ is the residual of the scaled Lasso estimator in the regression model of $\mathbf{x}_j$ against $\mathbf{X}_{-j}$ with $\boldsymbol{\gamma}_j = -\sigma_j(\mathbf{\Sigma}^{-1}_{j,-j})^T$, and we can get the sparsity of $\boldsymbol{\gamma}_j$ through the assumption that the rows of $\mathbf{\Sigma}^{-1}$ satisfy the L_0 sparsity condition. Thus, by applying the estimation error bound of the residual vector of the scaled Lasso in Ren et al. (2015, Inequality (18)), we can get

$$\max_{1\le j\le p} P\Big(\frac{1}{n}\big|\|\mathbf{z}_j\|_2^2 - \|\boldsymbol{\rho}_j\|_2^2\big| > Cs\frac{\log p}{n}\Big) \le o(p^{-\delta+1}), \tag{S1.1}$$

which gives the deviation of $\|\mathbf{z}_j\|_2^2$ from its population counterpart $\|\boldsymbol{\rho}_j\|_2^2$.

With $\|\boldsymbol{\rho}_j\|_2^2/\sigma_j \sim \chi^2_{(n)}$ for any $1 \le j \le p$, applying the following tail probability bound with $t = 2\sqrt{2\delta\log(p)/n}$ for the chi-squared distribution with n degrees of freedom Ren et al. (2015, Inequality (93)):

$$P\{|\frac{\chi^2_{(n)}}{n} - 1| \geq t\} \leq 2\exp\big(-nt(t\wedge 1)/8\big) \tag{S1.2}$$

gives

$$1 - 2\sqrt{2\delta\log(p)/n} \leq \|\boldsymbol{\rho}_j\|_2^2/(n\sigma_j) \leq 1 + 2\sqrt{2\delta\log(p)/n},$$

holding with probability at least $1-2p^{-\delta}$. This inequality together with (S1.1) entails that with probability at least $1-o(p^{-\delta+1})$,

$$[1-2\sqrt{2\delta\log(p)/n}]\sigma_j - Cs\log(p)/n \leq \|\mathbf{z}_j\|_2^2/n \leq [1+2\sqrt{2\delta\log(p)/n}]\sigma_j + Cs\log(p)/n,$$

for any $1 \leq j \leq p$. In view of $s = o(n/\log p)$, we have $s \leq c_0 n/\log p$ with some sufficiently small constant c_0. Combining these results leads to

$$[1-2\sqrt{2\delta\log(p)/n}]\sigma_j - Cc_0 \leq \|\mathbf{z}_j\|_2^2/n \leq [1+2\sqrt{2\delta\log(p)/n}]\sigma_j + Cc_0, \tag{S1.3}$$

with probability at least $1-o(p^{-\delta+1})$, which completes the proof of **Part 1**.

Part 2: Deviation bounds of $\max_{k\neq j}\|\mathbf{x}_k\|_2$ **and** $\min_{k\neq j}\|\mathbf{x}_k\|_2$. In order to proceed, we need to construct an upper bound for $\max_{k\neq j}\|\mathbf{x}_k\|_2$ and a lower bound for $\min_{k\neq j}\|\mathbf{x}_k\|_2$, respectively. Since $\|\mathbf{x}_k\|_2^2/\sigma_{kk} \sim \chi^2_{(n)}$ for any $1 \leq k \leq p$, by applying (S1.2) with $t = 4\sqrt{\delta\log(p)/n}$ for the chi-squared distribution with n degrees of freedom, we have

$$[1-4\sqrt{\delta\log(p)/n}]\sigma_{kk} \leq \|\mathbf{x}_k\|_2^2/n \leq [1+4\sqrt{\delta\log(p)/n}]\sigma_{kk}$$

holding with probability at least $1-2p^{-2\delta}$. By the condition that the eigenvalues of $\boldsymbol{\Sigma}$ are within $[M_*, M^*]$, we have $M_* \leq \sigma_{kk} \leq M^*$ for any $1 \leq k \leq p$. It follows that for sufficiently large n, with probability at least $1-2p^{-2\delta}$,

$$\widetilde{M}^* \leq \sqrt{[1+4\sqrt{\delta\log(p)/n}]M_*} \leq \|\mathbf{x}_k\|_2/\sqrt{n} \leq \sqrt{[1+4\sqrt{\delta\log(p)/n}]M^*} \leq \widetilde{M},$$

where $\widetilde{M}^*$ and $\widetilde{M}$ are some positive constants. Thus we have

$$P(\max_{k\neq j} \|\mathbf{x}_k\|_2/\sqrt{n} > \widetilde{M}) \leq \sum_{k\neq j} P(\|\mathbf{x}_k\|_2/\sqrt{n} > \widetilde{M}) \leq p \cdot 2p^{-2\delta} = 2p^{1-2\delta} = o(p^{1-\delta}), \tag{S1.4}$$

and

$$P(\min_{k\neq j} \|\mathbf{x}_k\|_2/\sqrt{n} < \widetilde{M}^*) \leq \sum_{k\neq j} P(\|\mathbf{x}_k\|_2/\sqrt{n} < \widetilde{M}^*) \leq p \cdot 2p^{-2\delta} = 2p^{1-2\delta} = o(p^{1-\delta}), \tag{S1.5}$$

respectively, which entail that $\max_{k\neq j} \|\mathbf{x}_k\|_2/\sqrt{n} \leq \widetilde{M}$ and $\min_{k\neq j} \|\mathbf{x}_k\|_2/\sqrt{n} \geq \widetilde{M}^*$ hold with probability at least $1 - o(p^{-\delta+1})$. It completes the proof of **Part 2**.

Part 3: Deviation bounds of τ_j. Then we turn to the deviation bound of τ_j. In order to proceed, it is worthwhile to notice a basic inequality that

$$\mathbf{z}_j^T\mathbf{x}_j = \|\mathbf{z}_j\|_2^2 + (\mathbf{X}_{-j}\widehat{\boldsymbol{\gamma}}_j)^T\mathbf{z}_j = \|\mathbf{z}_j\|_2^2 + \sqrt{n}\hat{\sigma}_j\lambda_0 \sum_{k\neq j}(\|\mathbf{x}_k\|_2 \cdot |\widehat{\boldsymbol{\gamma}}_{j,k}|) \geq \|\mathbf{z}_j\|_2^2, \tag{S1.6}$$

where the second equality above follows from the Karush-Kuhn-Tucker (KKT) condition for the scaled Lasso estimator, which gives $\mathbf{x}_k^T\mathbf{z}_j = \mathbf{x}_k^T(\mathbf{x}_j - \mathbf{X}_{-j}\widehat{\boldsymbol{\gamma}}_j) = \sqrt{n}\hat{\sigma}_j\lambda_0\|\mathbf{x}_k\|_2 \cdot \operatorname{sgn}(\widehat{\boldsymbol{\gamma}}_{j,k})$ with $\widehat{\boldsymbol{\gamma}}_{j,k}$ being the kth component of $\widehat{\boldsymbol{\gamma}}_j$, for any $k \in A = \{k \neq j : \operatorname{sgn}(\widehat{\boldsymbol{\gamma}}_{j,k}) \neq 0\}$.

With the aid of (S1.6), we will first establish the upper bound of τ_j. It follows easily $\mathbf{z}_j^T\mathbf{x}_j \geq \|\mathbf{z}_j\|_2^2$ in (S1.6) that $\tau_j = \|\mathbf{z}_j\|_2/|\mathbf{z}_j^T\mathbf{x}_j| \leq 1/\|\mathbf{z}_j\|_2$. Since $\sqrt{\log(p)/n} \to 0$ as $n \to \infty$ and c_0 is sufficiently small, in view of (S1.3) and $\tau_j \leq 1/\|\mathbf{z}_j\|_2$, we know that when n is large enough, there exists some constant c_j depending on j such that

$$\tau_j \leq \frac{1}{\|\mathbf{z}_j\|_2} = \frac{1}{\sqrt{n}}\frac{1}{\sqrt{\|\mathbf{z}_j\|_2^2/n}} \leq \frac{1}{\sqrt{n}}\frac{1}{([1-2\sqrt{2\delta\log(p)/n}]\sigma_j - Cc_0)^{1/2}} \leq \frac{c_j}{\sqrt{n}}, \tag{S1.7}$$

holding with probability at least $1-o(p^{-\delta+1})$.

It remains to find the lower bound of τ_j. In view of (S1.4) and the basic inequality (S1.6), it follows that with probability at least $1-o(p^{-\delta+1})$,

$$\mathbf{z}_j^T\mathbf{x}_j = \|\mathbf{z}_j\|_2^2 + \sqrt{n}\hat{\sigma}_j\lambda_0\sum_{k\neq j}(\|\mathbf{x}_k\|_2\cdot|\widehat{\boldsymbol{\gamma}}_{j,k}|) \leq \|\mathbf{z}_j\|_2^2 + n\widetilde{M}\hat{\sigma}_j\lambda_0\|\widehat{\boldsymbol{\gamma}}_j\|_1,$$

which yields that $\tau_j = \|\mathbf{z}_j\|_2/|\mathbf{z}_j^T\mathbf{x}_j| \geq 1/(\|\mathbf{z}_j\|_2 + \frac{n\widetilde{M}\hat{\sigma}_j\lambda_0\|\widehat{\boldsymbol{\gamma}}_j\|_1}{\|\mathbf{Z}_j\|_2})$. Now we need to construct an upper bound for $\|\widehat{\boldsymbol{\gamma}}_j\|_1$.

Since $\widehat{\boldsymbol{\gamma}}_j$ is the scaled lasso estimator with $\lambda_0 = (1+\varepsilon)\sqrt{2\delta\log(p)/n}$ for some $\delta \geq 1$ and $\varepsilon > 0$, combining the estimator error bound of the scaled lasso estimator Ren et al. (2015, Inequality (17)) and inequality (S1.5) yields

$$P\Big\{\|\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j\|_1 \leq \frac{C_1^* s_j^*\sqrt{\delta\log p}}{\sqrt{n}}\Big\} \geq 1-o(p^{-\delta+1}), \tag{S1.8}$$

where C_1^* is a constant and $s_j^* = \|\gamma_j\|_0$. Thus, it follows that with probability at least $1-o(p^{-\delta+1})$,

$$\|\widehat{\boldsymbol{\gamma}}_j\|_1 \leq \|\boldsymbol{\gamma}_j\|_1 + \frac{C^* s_j^*\sqrt{\delta\log p}}{\sqrt{n}}.$$

Returning to derive the lower bound of τ_j. In view of $\lambda_0 = (1+\varepsilon)\sqrt{2\delta\log(p)/n}$, $\sqrt{\log p/n} \to 0$ as $n\to\infty$ and $\boldsymbol{\gamma}_j = -\sigma_j(\boldsymbol{\Sigma}^{-1}_{j,-j})^T$, as well as the assumption that the rows of $\boldsymbol{\Sigma}^{-1}$ is L_0 sparse, we have

$$\widetilde{M}\lambda_0\|\widehat{\boldsymbol{\gamma}}_j\|_1 \leq \widetilde{M}(1+\varepsilon)\sqrt{2\delta\log(p)/n}(\|\boldsymbol{\gamma}_j\|_1 + \frac{C^* s_j^*\sqrt{\delta\log p}}{\sqrt{n}}) \leq c_j'\widetilde{M}\sqrt{s\log(p)/n},$$

where c_j' is a constant. Combining this inequality and $\|\mathbf{z}_j\|_2 \geq \sqrt{n}/c_j$ from (S1.7) along with $\sqrt{s\log(p)/n} = o(1)$ gives that there exist some constant c_j'' such that

$$\frac{1}{\|\mathbf{z}_j\|_2 + \frac{n\widetilde{M}\hat{\sigma}_j\lambda_0\|\widehat{\boldsymbol{\gamma}}_j\|_1}{\|\mathbf{Z}_j\|_2}} \geq \frac{1}{\|\mathbf{z}_j\|_2 + c_jc_j'\widetilde{M}\hat{\sigma}_j\sqrt{s\log p}} \geq \frac{c_j''}{\|\mathbf{z}_j\|_2}.$$

In view of this inequality and (S1.3), we may come to the conclusion that with probability at least $1 - o(p^{-\delta+1})$, there exists a constant $\widetilde{c}_j$ such that

$$\tau_j \geq \frac{c_j''}{\|\mathbf{z}_j\|_2} \geq \frac{c_j''}{\sqrt{n}} \frac{1}{[1 + 2\sqrt{2\delta \log(p)/n}]\sigma_j + Cc_0} \geq \frac{\widetilde{c}_j}{\sqrt{n}},$$

which together with (S1.7) entails that $\tau_j \asymp n^{-1/2}$ with probability at least $1 - o(p^{-\delta+1})$. Moreover, conditional on this event, it is not difficult to see from the previous proof that

$$\lim_{n\to\infty} \tau_j n^{1/2} = \lim_{n\to\infty} n^{1/2}\|\mathbf{z}_j\|_2/|\mathbf{z}_j^T\mathbf{x}_j| = \lim_{n\to\infty} n^{1/2}/\|\mathbf{z}_j\|_2 = \lim_{n\to\infty} n^{1/2}/\|\boldsymbol{\rho}_j\|_2 = \boldsymbol{\Sigma}_{j,j}^{-1/2}.$$

It completes the proof of **Part 3**.

Part 4: Deviation bounds of η_j. In this part, we continue to find the deviation bound for $\eta_j = \max_{k\neq j} |\mathbf{z}_j^T\mathbf{x}_k|/\|\mathbf{z}_j\|_2$. By the KKT condition, we have for any $k \neq j$, $1 \leq k \leq p$,

$$|\mathbf{x}_k^T\mathbf{z}_j| = |\mathbf{x}_k^T(\mathbf{x}_j - \mathbf{X}_{-j}\widehat{\boldsymbol{\gamma}}_j)| \leq \sqrt{n}\hat{\sigma}_j\lambda_0\|\mathbf{x}_k\|_2.$$

Combining this inequality and (S1.7) along with the upper bound of $\max_{k\neq j} \|\mathbf{x}_k\|_2/\sqrt{n}$ in **Part 2** yields

$$\eta_j \leq \sqrt{n}\hat{\sigma}_j\lambda_0 \max_{k\neq j} \|\mathbf{x}_k\|_2/\|\mathbf{z}_j\|_2 \leq c_j\widetilde{M}\sqrt{n}\hat{\sigma}_j\lambda_0. \tag{S1.9}$$

On the other hand, in view of Ren et al. (2015, Inequality(18)), we have

$$P\{|\widehat{\sigma}_j/\sigma_j^* - 1| \geq 1/2\} \leq o(p^{-\delta+1}),$$

where $\sigma_j^* = \|\boldsymbol{\rho}_j\|_2/\sqrt{n}$ is the oracle estimator of σ_j. With the aid of $\frac{n(\sigma_j^*)^2}{\mathbb{E}(\sigma_j^*)^2} \sim \chi^2_{(n)}$, (S1.2) justifies the replacement of σ_j^* by $\sqrt{\mathbb{E}(\sigma_j^*)}$ or a constant C^* in the above inequality, which entails that $\widehat{\sigma}_j \leq \frac{3}{2}C^*$ can hold with probability at least $1 - o(p^{-\delta+1})$.

In view of the above fact and (S1.9), as well as $\lambda_0 = (1+\varepsilon)\sqrt{2\delta\log(p)/n}$, for sufficiently large n, we get

$$\eta_j \le c_j \widetilde{M}\sqrt{n}\hat{\sigma}_j\lambda_0 \le \frac{3}{2}c_j\widetilde{M}C^*(1+\varepsilon)\sqrt{2\delta\log(p)} = C_j\sqrt{\log(p)},$$

holding with probability at least $1-o(p^{-\delta+1})$, where $C_j = \frac{3}{2}c_j\widetilde{M}C^*(1+\varepsilon)\sqrt{2\delta}$. It completes the proof of **Part 4**.

Part 5: Confidence intervals of β_j. By the definition of the LDPE estimator given in (3.3), replacing $\mathbf{y}$ with $\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ along with some simplification gives for any j, $1 \le j \le p$,

$$\widehat{\beta}_j - \beta_j = \frac{\mathbf{z}_j^T\boldsymbol{\varepsilon}}{\mathbf{z}_j^T\mathbf{x}_j} + \frac{\sum_{k\neq j}\mathbf{z}_j^T\mathbf{x}_k(\beta_k - \widehat{\beta}_k^{(\text{init})})}{\mathbf{z}_j^T\mathbf{x}_j}. \tag{S1.10}$$

Moving the term $\mathbf{z}_j^T\boldsymbol{\varepsilon}/\mathbf{z}_j^T\mathbf{x}_j$ to the left hand side and then dividing both sides by τ_j gives

$$|\tau_j^{-1}(\widehat{\beta}_j - \beta_j) - \mathbf{z}_j^T\boldsymbol{\varepsilon}/\|\mathbf{z}_j\|_2| \le (\max_{k\neq j}|\mathbf{z}_j^T\mathbf{x}_k|/\|\mathbf{z}_j\|_2)\|\widehat{\boldsymbol{\beta}}^{(\text{init})} - \boldsymbol{\beta}\|_1 = \eta_j\|\widehat{\boldsymbol{\beta}}^{(\text{init})} - \boldsymbol{\beta}\|_1. \tag{S1.11}$$

For simplicity, denote by $\mathcal{E}$ the probability event in **Parts 1-4** such that the deviation bounds of τ_j and η_j still hold. Then $P(\mathcal{E}) \ge 1-o(p^{-\delta+1})$. Define two new events $\mathcal{E}_1$ and $\mathcal{E}_2$ as

$$\mathcal{E}_1 = \{|\tau_j^{-1}(\widehat{\beta}_j - \beta_j) - \mathbf{z}_j^T\boldsymbol{\varepsilon}/\|\mathbf{z}_j\|_2| \le \sigma^*\epsilon_n'\},$$

$$\mathcal{E}_2 = \{|\widehat{\sigma}/\sigma^* - 1| \le \epsilon_n''\}.$$

We first derive two probability inequalities, which will be used in the next proof. First, in view of $C_2 s(2/n)\log(p/\epsilon) \le \epsilon_n''$, it follows from the Condition 1 that $P(\mathcal{E}_2^c) \le \epsilon$.

Second, combining inequality (S1.11) with the assumptions in Proposition 1 gives

$$\begin{aligned} &P(\mathcal{E}_1^c \cap \mathcal{E}) \leq P(\mathcal{E}_1^c|\mathcal{E}) \leq P(\eta_j\|\widehat{\boldsymbol{\beta}}^{(\text{init})} - \boldsymbol{\beta}\|_1 > \sigma^*\epsilon_n'|\mathcal{E}) \\ \leq& P\big\{C_j\sqrt{\log(p)}\|\widehat{\boldsymbol{\beta}}^{(\text{init})} - \boldsymbol{\beta}\|_1 > \sigma^* C_1 C_j s\sqrt{(2/n)} \cdot \sqrt{\log(p)\log(p/\epsilon)}\big\} \\ \leq& P\big\{\|\widehat{\boldsymbol{\beta}}^{(\text{init})} - \boldsymbol{\beta}\|_1 > \sigma^* C_1 s\sqrt{(2/n)\log(p/\epsilon)}\big\} \leq \epsilon. \end{aligned} \tag{S1.12}$$

Returning to the confidence intervals of β_j. Conditional on the event $\mathcal{E} \cap \mathcal{E}_1 \cap \mathcal{E}_2$, we know that $\tau_j^{-1}|\widehat{\beta}_j - \beta_j| \geq \widehat{\sigma}t$ implies $|\mathbf{z}_j^T\boldsymbol{\varepsilon}|/\|\mathbf{z}_j\|_2 \geq \widehat{\sigma}t - \sigma^*\epsilon_n' \geq \sigma^*\{(1-\epsilon_n'')t - \epsilon_n'\}$ for any $t > (1+\epsilon_n')/(1-\epsilon_n'')$. Let $x = (1-\epsilon_n'')t - \epsilon_n'$. Since $\mathbf{z}_j$ only depends on $\mathbf{X}$, along with the fact that $\mathbf{X}$ and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$ are independent, conditional on each realization of $\mathbf{z}_j$, we have $\mathbf{z}_j^T\boldsymbol{\varepsilon}/(\|\mathbf{z}_j\|_2\sigma^*) \sim \sqrt{n}\varepsilon_1/\|\boldsymbol{\varepsilon}\|_2$ with $\sigma^* = \|\boldsymbol{\varepsilon}\|_2/\sqrt{n}$. It follows that

$$P\Big(\frac{|\mathbf{z}_j^T\boldsymbol{\varepsilon}|}{\|\mathbf{z}_j\|_2} \geq \sigma^* x\big|\mathbf{z}_j\Big) = P\{(n-x^2)\varepsilon_1^2 \geq x^2(\varepsilon_2^2 + \cdots + \varepsilon_n^2)\} \leq 2\Phi_{n-1}(-x\sqrt{1-n^{-1}}), \tag{S1.13}$$

where $\Phi_{n-1}(t)$ is the Student t-distribution function with $n-1$ degrees of freedom.

Since the right hand side of inequality (S1.13) is independent of the realization of $\mathbf{z}_j$, along with the fact that $\mathbf{z}_j$ and $\boldsymbol{\varepsilon}$ are independent, we have $P(|\mathbf{z}_j^T\boldsymbol{\varepsilon}|/\|\mathbf{z}_j\|_2 \geq \sigma^* x) \leq 2\Phi_{n-1}(-x\sqrt{1-n^{-1}})$. With the aid of the analysis in previous paragraph and taking the probabilities of the events $\mathcal{E}^c$, $\mathcal{E}_1^c \cap \mathcal{E}$ and $\mathcal{E}_2^c$ into consideration, we conclude that for sufficiently large n,

$$\begin{aligned} &P(|\widehat{\beta}_j - \beta_j| \geq \tau_j\widehat{\sigma}t) \leq P(|\widehat{\beta}_j - \beta_j| \geq \tau_j\widehat{\sigma}t|\mathcal{E} \cap \mathcal{E}_1 \cap \mathcal{E}_2) + P(\mathcal{E}^c \cup \mathcal{E}_1^c \cup \mathcal{E}_2^c) \\ \leq& P(\tau_j^{-1}|\widehat{\beta}_j - \beta_j| \geq \widehat{\sigma}t|\mathcal{E} \cap \mathcal{E}_1 \cap \mathcal{E}_2) + P(\mathcal{E}^c) + P(\mathcal{E}_1^c \cap \mathcal{E}) + P(\mathcal{E}_2^c) \\ \leq& 2\Phi_{n-1}(-x\sqrt{1-n^{-1}}) + 2\epsilon + o(p^{-\delta+1}). \end{aligned}$$

Since $\max(\epsilon_n', \epsilon_n'') \to 0$ and when $n \to \infty$, the t-distribution will converge to the normal distribution, by letting $n \to \infty$ and $t = \Phi^{-1}(1-\alpha/2)$, we further have

$$\lim_{n\to\infty} P\big\{|\widehat{\beta}_j - \beta_j| \le \tau_j \widehat{\sigma}\Phi^{-1}(1-\alpha/2)\big\} = 1-\alpha,$$

which completes the proof of Proposition 1.

S2 Proof of Theorem 1

The proof of Theorem 1 is to conduct delicate analysis on some events with significant probability and we will break the communication barriers between different subsamples by considering certain overall statistics. Similar to (3.5), the bias factor $\eta_j^{(l)}$ and noise factor $\tau_j^{(l)}$ of the lth subsample are defined as

$$\eta_j^{(l)} = \max_{k\neq j} |(\mathbf{z}_j^{(l)})^T \mathbf{x}_k^{(l)}|/\|\mathbf{z}_j^{(l)}\|_2,\ \tau_j^{(l)} = \|\mathbf{z}_j^{(l)}\|_2/|(\mathbf{z}_j^{(l)})^T \mathbf{x}_j^{(l)}|.$$

The overall bias and noise factors $\widetilde{\eta}_j$ and $\widetilde{\tau}_j$ are

$$\widetilde{\eta}_j = \max_{1\le l\le K} \eta_j^{(l)} \quad \text{and} \quad \widetilde{\tau}_j = \max_{1\le l\le K} \tau_j^{(l)}.$$

We will first derive the deviation bounds for $\widetilde{\tau}_j$ and $\widetilde{\eta}_j$. Since similar conditions are imposed for each subsample as those in Proposition 1, by (S1.7), we know that for sufficiently large $\widetilde{n}$,

$$\tau_j^{(l)} \le 1/\|\mathbf{z}_j^{(l)}\|_2 \le c_j/\sqrt{\widetilde{n}}$$

holds with probability at least $1 - o(p^{-\delta+1})$. It follows that

$$P(\widetilde{\tau}_j > c_j/\sqrt{\widetilde{n}}) \le \sum_{l=1}^{K} P(\tau_j^{(l)} > c_j/\sqrt{\widetilde{n}}) = o(Kp^{-\delta+1}). \tag{S2.1}$$

Thus, we get $\widetilde{\tau}_j \le c_j/\sqrt{\widetilde{n}}$ with probability at least $1-o(Kp^{-\delta+1})$. By the same argument, $\widetilde{\tau}_j \ge \widetilde{c}_j/\sqrt{\widetilde{n}}$ with probability at least $1-o(Kp^{-\delta+1})$ such that $\widetilde{\tau}_j \asymp \widetilde{n}^{-1/2}$. Similarly, we have $\widetilde{\eta}_j \le C_j\sqrt{\log(p)}$ with probability at least $1-o(Kp^{-\delta+1})$. Define event $\widetilde{\mathcal{E}}$ such that the deviation bounds for both $\widetilde{\tau}_j$ and $\widetilde{\eta}_j$ hold. It follows that $P(\widetilde{\mathcal{E}}) \ge 1-o(Kp^{-\delta+1})$.

Then we would like to apply an argument similar to the proof of Proposition 1 after taking the communication barriers into consideration, and derive the confidence intervals for components of the bagging estimator $\widehat{\boldsymbol{\beta}}^{(\mathrm{mean})}$. For the LDPE estimator $\widehat{\beta}_j^{(l)}$ of the lth subsample, $1 \le l \le K$, similar to (S1.10), by definition we have for any coordinate j, $1 \le j \le p$,

$$\widehat{\beta}_j^{(l)} - \beta_j = \frac{(\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}}{(\mathbf{z}_j^{(l)})^T\mathbf{x}_j^{(l)}} + \frac{\sum_{k\neq j}(\mathbf{z}_j^{(l)})^T\mathbf{x}_k^{(l)}(\beta_k - \widehat{\beta}_k^{(\mathrm{init})})}{(\mathbf{z}_j^{(l)})^T\mathbf{x}_j^{(l)}}.$$

Therefore, the bagging estimator $\widehat{\boldsymbol{\beta}}^{(\mathrm{mean})} = K^{-1}\sum_{l=1}^K \widehat{\boldsymbol{\beta}}^{(l)}$ satisfies that

$$\widehat{\beta}_j^{(\mathrm{mean})} - \beta_j = K^{-1}\sum_{l=1}^K \frac{(\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}}{(\mathbf{z}_j^{(l)})^T\mathbf{x}_j^{(l)}} + K^{-1}\sum_{l=1}^K \frac{\sum_{k\neq j}(\mathbf{z}_j^{(l)})^T\mathbf{x}_k^{(l)}(\beta_k - \widehat{\beta}_k^{(\mathrm{init})})}{(\mathbf{z}_j^{(l)})^T\mathbf{x}_j^{(l)}}.$$

So we have

$$\begin{aligned}&\Big|\widetilde{\tau}_j^{-1}(\widehat{\beta}_j^{(\mathrm{mean})} - \beta_j) - K^{-1}\sum_{l=1}^K(\widetilde{\tau}_j^{-1}\tau_j^{(l)})\frac{(\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}}{\|\mathbf{z}_j^{(l)}\|_2}\Big| \\ &\le K^{-1}\sum_{l=1}^K(\widetilde{\tau}_j^{-1}\tau_j^{(l)})\eta_j^{(l)}\|\widehat{\boldsymbol{\beta}}^{(\mathrm{init})} - \boldsymbol{\beta}\|_1 \le \widetilde{\eta}_j\|\widehat{\boldsymbol{\beta}}^{(\mathrm{init})} - \boldsymbol{\beta}\|_1. \qquad \text{(S2.2)}\end{aligned}$$

Modifying the event $\mathcal{E}_1$ a bit and keep $\mathcal{E}_2$ the same as that defined in the proof of Proposition 1, we denote by

$$\widetilde{\mathcal{E}}_1 = \Big\{\Big|\widetilde{\tau}_j^{-1}(\widehat{\beta}_j^{(\mathrm{mean})} - \beta_j) - K^{-1}\sum_{l=1}^K(\widetilde{\tau}_j^{-1}\tau_j^{(l)})\cdot(\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}/\|\mathbf{z}_j^{(l)}\|_2\Big| \le \sigma^*\epsilon_n'\Big\} \text{ and}$$

$$\mathcal{E}_2 = \big\{|\widehat{\sigma}/\sigma^* - 1| \le \epsilon_n''\big\}.$$

By inequality (S2.2), the definition of $\widetilde{\mathcal{E}}$ and conditions in Theorem 1, similar to (S1.12), we get $P(\mathcal{E}_2^c) \le \epsilon$ and $P(\widetilde{\mathcal{E}}_1^c|\widetilde{\mathcal{E}}) \le \epsilon$.

Conditioning on the event $\widetilde{\mathcal{E}}_1 \cap \mathcal{E}_2 \cap \widetilde{\mathcal{E}}$, we know that $\sqrt{K}\widetilde{\tau}_j^{-1}|\widehat{\beta}_j^{(\mathrm{mean})} - \beta_j| \ge \widehat{\sigma}t$ implies

$$\big|K^{-1/2}\sum_{l=1}^{K}(\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}/\|\mathbf{z}_j^{(l)}\|_2\big| \ge \widehat{\sigma}t - \sqrt{K}\sigma^*\epsilon_n' \ge \sigma^*\{(1-\epsilon_n'')t - \sqrt{K}\epsilon_n'\}, \qquad \text{(S2.3)}$$

for any $t > \sqrt{K}\epsilon_n'/(1-\epsilon_n'')$. Since $K^{-1/2}\sum_{l=1}^{K}(\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}/\|\mathbf{z}_j^{(l)}\|_2 \sim \varepsilon_1$, it follows that

$$P\big(\frac{1}{\sqrt{K}}\sum_{l=1}^{K}\frac{(\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}}{\|\mathbf{z}_j^{(l)}\|_2} \ge \sigma^*\{(1-\epsilon_n'')t - \sqrt{K}\epsilon_n'\}\big) \le \ P(\sqrt{n}\frac{\varepsilon_1}{\|\boldsymbol{\varepsilon}\|_2} \ge (1-\epsilon_n'')t - \sqrt{K}\epsilon_n')$$

$$\le 2\Phi_{n-1}(-(1-\epsilon_n'')t + \sqrt{K}\epsilon_n'). \qquad \text{(S2.4)}$$

Therefore, we get

$$P(\sqrt{K}\widetilde{\tau}_j^{-1}|\widehat{\beta}_j^{(\mathrm{mean})} - \beta_j| \ge \widehat{\sigma}t) \le 2\Phi_{n-1}(-(1-\epsilon_n'')t + \sqrt{K}\epsilon_n') + 2\epsilon + o(Kp^{-\delta+1}).$$

By the same argument as that in the proof of Proposition 1, if $\sqrt{K}\epsilon_n' \to 0$, we get

$$\lim_{n\to\infty} P\big\{|\widehat{\beta}_j^{(\mathrm{mean})} - \beta_j| \le K^{-1/2}\widetilde{\tau}_j\widehat{\sigma}\Phi^{-1}(1-\alpha/2)\big\} = 1-\alpha.$$

It completes the proof of Part (**A**).

For Part (**B**), we first derive the bounds on the key quantity K_j. On one hand, in view of $K_j = K^2/\sum_{l=1}^{K}(\omega_j^{(l)})^2$ and $\omega_j^{(l)} = \widetilde{\tau}_j^{-1}\tau_j^{(l)} \le 1$, it is clear that $K_j \ge K$. On the other hand, by Proposition 1 and an argument similar to (S2.1), we know that with probability at least $1 - o(Kp^{-\delta+1})$, $\tau_j^{(l)} \ge \widetilde{c}_j\widetilde{n}^{-1/2}$ for any $1 \le l \le K$. Thus, together with (S2.1), there exists positive constant $c_j^* \ge 1$ such that $\min_{l=1}^{K}\omega_j^{(l)} \ge \sqrt{c_j^*}$ and $K_j \le c_j^* K$ hold with probability at least $1 - o(Kp^{-\delta+1})$.

We now proceed to derive confidence intervals for the refined inference. Similar to the proof of Part (**A**), conditioning on the event $\widetilde{\mathcal{E}}_1 \cap \mathcal{E}_2 \cap \widetilde{\mathcal{E}}$, we know that $\sqrt{K_j}\widetilde{\tau}_j^{-1}|\widehat{\beta}_j^{(\text{mean})} - \beta_j| \geq \widehat{\sigma} t$ implies

$$\Big|\frac{1}{\sqrt{\sum_{l=1}^K (\omega_j^{(l)})^2}} \sum_{l=1}^K \omega_j^{(l)} (\mathbf{z}_j^{(l)})^T \boldsymbol{\varepsilon}^{(l)} / \|\mathbf{z}_j^{(l)}\|_2\Big| \geq \widehat{\sigma} t - \sqrt{K_j}\sigma^* \epsilon_n' \geq \sigma^*\{(1-\epsilon_n'')t - \sqrt{K_j}\epsilon_n'\},$$

for any $t > (1+\sqrt{K_j})\epsilon_n'/(1-\epsilon_n'')$. Since $\frac{1}{\sqrt{\sum_{l=1}^K (\omega_j^{(l)})^2}} \sum_{l=1}^K \omega_j^{(l)} (\mathbf{z}_j^{(l)})^T \boldsymbol{\varepsilon}^{(l)} / \|\mathbf{z}_j^{(l)}\|_2 \sim \varepsilon_1$, similar to (S2.4), it follows that

$$P\Big(\frac{\sum_{l=1}^K \omega_j^{(l)} (\mathbf{z}_j^{(l)})^T \boldsymbol{\varepsilon}^{(l)}}{\|\mathbf{z}_j^{(l)}\|_2 \sqrt{\sum_{l=1}^K (\omega_j^{(l)})^2}} \geq \sigma^*\{(1-\epsilon_n'')t - \sqrt{K_j}\epsilon_n'\}\Big) \leq\ 2\Phi_{n-1}(-(1-\epsilon_n'')t + \sqrt{K_j}\epsilon_n').$$

Therefore, we get

$$P(\sqrt{K_j}\widetilde{\tau}_j^{-1}|\widehat{\beta}_j^{(\text{mean})} - \beta_j| \geq \widehat{\sigma} t) \leq 2\Phi_{n-1}(-(1-\epsilon_n'')t + \sqrt{K_j}\epsilon_n') + 2\epsilon + o(Kp^{-\delta+1}).$$

If $\sqrt{K_j}\epsilon_n' \to 0$, similarly we have

$$\lim_{n\to\infty} P\{|\widehat{\beta}_j^{(\text{mean})} - \beta_j| \leq K_j^{-1/2}\widetilde{\tau}_j \widehat{\sigma}\Phi^{-1}(1-\alpha/2)\} = 1-\alpha.$$

It concludes the proof of Theorem 1.

S3 Proof of Theorem 2

The proof of Theorem 2 can be finished by applying the union bound to some key inequalities in the proof of Theorem 1, which is detailed as follows. In view of (S2.2), we have

$$\max_{j\in S}\Big|\widetilde{\tau}_j^{-1}(\widehat{\beta}_j^{(\text{mean})} - \beta_j) - K^{-1}\sum_{l=1}^K (\widetilde{\tau}_j^{-1}\tau_j^{(l)})\frac{(\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}}{\|\mathbf{z}_j^{(l)}\|_2}\Big| \leq \max_{j\in S}\widetilde{\eta}_j\|\widehat{\boldsymbol{\beta}}^{(\text{init})} - \boldsymbol{\beta}\|_1 \cdot K^{-1}\sum_{l=1}^K (\widetilde{\tau}_j^{-1}\tau_j^{(l)}).$$

Since the event $\widetilde{\mathcal{E}}$ holds with probability at least $1 - o(Kp^{-\delta+1})$ and S is a set with finite number of elements, it is clear that $\max_{j\in S}\widetilde{\eta}_j \leq \max_{j\in S} C_j\sqrt{\log p}$ also holds with probability at least $1 - o(Kp^{-\delta+1})$. Conditioning on this event (denoted by $\mathcal{E}_3$), under the assumptions of Theorem 2, similar to (S1.12), we get

$$P\{\max_{j\in S} |\widetilde{\tau}_j^{-1}(\widehat{\beta}_j^{(\text{mean})} - \beta_j) - K^{-1}\sum_{l=1}^{K}(\widetilde{\tau}_j^{-1}\tau_j^{(l)}) \cdot (\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}/\|\mathbf{z}_j^{(l)}\|_2| \geq \sigma^*\epsilon_n'|\mathcal{E}_3\} \leq \epsilon.$$

Then by arguments similar to (S2.3) and (S2.4) together with the union bound, we know that for any $t > \sqrt{K}\epsilon_n'/(1-\epsilon_n'')$, $\max_{j\in S}\sqrt{K}\widetilde{\tau}_j^{-1}|\widehat{\beta}_j^{(\text{mean})} - \beta_j| \geq \widehat{\sigma}t$ implies

$$\min_{j\in S}\big|K^{-1/2}\sum_{l=1}^{K}(\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}/\|\mathbf{z}_j^{(l)}\|_2\big| \geq \widehat{\sigma}t - \sqrt{K}\sigma^*\epsilon_n' \geq \sigma^*\{(1-\epsilon_n'')t - \sqrt{K}\epsilon_n'\},$$

and that

$$P\Big(\min_{j\in S}\frac{1}{\sqrt{K}}\sum_{l=1}^{K}\frac{(\mathbf{z}_j^{(l)})^T\boldsymbol{\varepsilon}^{(l)}}{\|\mathbf{z}_j^{(l)}\|_2} \geq \sigma^*\{(1-\epsilon_n'')t - \sqrt{K}\epsilon_n'\}\Big) \leq |S|\cdot 2\Phi_{n-1}(-(1-\epsilon_n'')t + \sqrt{K}\epsilon_n').$$

Therefore, we have

$$P(\max_{j\in S}\sqrt{K}|\widehat{\beta}_j^{(\text{mean})} - \beta_j|/\widetilde{\tau}_j \geq \widehat{\sigma}t) \leq |S|\cdot 2\Phi_{n-1}[-(1-\epsilon_n'')t + \sqrt{K}\epsilon_n'] + 2\epsilon + o(Kp^{-\delta+1}).$$

Under the extra assumption in Part(**B**), together with $\min_{l=1}^{K}\omega_j^{(l)} \geq \sqrt{c_j^*}$ with probability at least $1 - o(Kp^{-\delta+1})$ (shown in the proof of Part(**B**) of Theorem 1), similarly we have

$$P(\max_{j\in S}\sqrt{K_j}|\widehat{\beta}_j^{(\text{mean})} - \beta_j|/\widetilde{\tau}_j \geq \widehat{\sigma}t) \leq \sum_{j\in S}2\Phi_{n-1}[-(1-\epsilon_n'')t + \sqrt{K_j}\epsilon_n'] + 2\epsilon + o(Kp^{-\delta+1}).$$

It completes the proof of Theorem 2.

S4 Proof of Theorem 3

We first present some definitions and three lemmas that will be used in the rest proofs. Define $\iota_j^2 = \mathbb{E}\|\boldsymbol{\rho}_j\|_2^2/n = \sigma_j$ and $(\widehat{\iota}_j^{(l)})^2 = \frac{(\mathbf{Z}_j^{(l)})^T\mathbf{X}_j^{(l)}}{\widetilde{n}}$ for $1 \leq j \leq p$ and $1 \leq l \leq K$. Denote by

$$\mathbf{C} = \begin{pmatrix} 1 & -\gamma_{1,2} & \cdots & -\gamma_{1,p} \\ -\gamma_{2,1} & 1 & \cdots & -\gamma_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma_{p,1} & -\gamma_{p,2} & \cdots & 1 \end{pmatrix},$$

and

$$\widehat{\mathbf{C}}^{(l)} = \begin{pmatrix} 1 & -\widehat{\gamma}_{1,2}^{(l)} & \cdots & -\widehat{\gamma}_{1,p}^{(l)} \\ -\widehat{\gamma}_{2,1}^{(l)} & 1 & \cdots & -\widehat{\gamma}_{2,p}^{(l)} \\ \vdots & \vdots & \ddots & \vdots \\ -\widehat{\gamma}_{p,1}^{(l)} & -\widehat{\gamma}_{p,2}^{(l)} & \cdots & 1 \end{pmatrix}.$$

Write $\mathbf{T}^2 = \mathrm{diag}\{\iota_1^2, \cdots, \iota_p^2\}$ and $(\widehat{\mathbf{T}}^{(l)})^2 = \mathrm{diag}\{(\widehat{\iota}_1^{(l)})^2, \cdots, (\widehat{\iota}_p^{(l)})^2\}$ as diagonal matrixes for $1 \leq j \leq p$ and $1 \leq l \leq K$. Let $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1} = \mathbf{T}^{-2}\mathbf{C}$. Then, the nodewise Lasso estimator for $\boldsymbol{\Theta}$ can be constructed as $\widehat{\boldsymbol{\Theta}}^{(l)} = (\widehat{\mathbf{T}}^{(l)})^{-2}\widehat{\mathbf{C}}^{(l)}$. Denote the jth row of $\mathbf{X}^{(l)}$ and $\widehat{\boldsymbol{\Theta}}^{(l)}$ by $\widetilde{\mathbf{x}}_j^{(l)} = (x_{j1}^{(l)}, \cdots, x_{jp}^{(l)})^T$ and $\widehat{\boldsymbol{\Theta}}_j^{(l)}$, where $\mathbf{X}^{(l)}$ is the lth subsample for $1 \leq l \leq K$. With the above definitions, we have $\mathbf{Z}^{(l)} = \mathbf{X}^{(l)}(\widehat{\boldsymbol{\Theta}}^{(l)})^T$, where $\mathbf{Z}^{(l)} = (\mathbf{z}_1^{(l)}, \cdots, \mathbf{z}_p^{(l)})$. Thus the multiplier bootstrap statistic can be rewritten as

$$W_G = \max_{j\in G} \frac{1}{\sqrt{n}K} \sum_{l=1}^{K} \sum_{i=1}^{\widetilde{n}} (\widehat{\boldsymbol{\Theta}}_j^{(l)})^T \widetilde{\mathbf{x}}_i^{(l)} \widehat{\sigma} e_i^{(l)}.$$

Lemma 1. *Assume that* $(\log(pn))^7/n \leq C_3 n^{-c_3}$ *for some constants* $C_3, c_3 > 0$. *Define* $\xi_{ij} = \frac{1}{K}\boldsymbol{\Theta}_j^T \widetilde{\mathbf{x}}_i^{(l)} \varepsilon_i^{(l)}$. *Then under the assumptions of Theorem 1, we have for any* $G \subseteq \{1, 2, \ldots, p\}$,

$$\sup_{x\in\mathbb{R}} \left| P\left(\max_{j\in G} \sum_{l=1}^{K} \sum_{i=1}^{\widetilde{n}} \xi_{ij}/\sqrt{n} \leq x \right) - P\left(\max_{j\in G} \sum_{i=1}^{n} u_{ij}/\sqrt{n} \leq x \right) \right| \lesssim n^{-c'},$$

where $c' > 0$ *and* $\left\{\mathbf{u}_i = (u_{i1}, \ldots, u_{ip})^T\right\}$ *is a sequence of mean zero independent Gaussian vector with* $\mathbb{E}\mathbf{u}_i\mathbf{u}_i^T = \frac{1}{K}\mathbf{\Theta}_j^T\mathbf{\Sigma}\mathbf{\Theta}_j\sigma^2$

Since this lemma is a direct corollary to Zhang and Cheng (2017, Lemma 1.1), we omit the proof.

Lemma 2. *Assume that* $\max_j s(\log(p\widetilde{n}))^3(\log(\widetilde{n}))^2/\widetilde{n} = o(1)$. *Define* $\widehat{\xi}_{ij} = \frac{1}{K}(\widehat{\mathbf{\Theta}}_j^{(l)})^T\widetilde{\mathbf{x}}_i^{(l)}\varepsilon_i^{(l)}$. *Then under the assumptions of Theorem 1, there exist* $\zeta_1, \zeta_2 > 0$ *such that*

$$P\left(\max_{1\le j\le p}\left|\sum_{l=1}^{K}\sum_{i=1}^{\widetilde{n}}\widehat{\xi}_{ij}/\sqrt{n} - \sum_{l=1}^{K}\sum_{i=1}^{\widetilde{n}}\xi_{ij}/\sqrt{n}\right| \ge \zeta_1\right) < \zeta_2,$$

where $\zeta_1\sqrt{1\vee\log(p/\zeta_1)} = o(1)$ *and* $\zeta_2 = o(1)$.

Lemma 3. *Define*

$$\Gamma = \max_{1\le j,\ k\le p}\left|\frac{\widehat{\sigma}^2}{K^2}\sum_{l=1}^{K}(\widehat{\mathbf{\Theta}}_j^{(l)})^T\widehat{\mathbf{\Sigma}}^{(l)}\widehat{\mathbf{\Theta}}_k^{(l)} - \frac{\sigma^2}{K}\mathbf{\Theta}_j^T\mathbf{\Sigma}\mathbf{\Theta}_k\right|, \quad \widehat{\mathbf{\Sigma}}^{(l)} = (\mathbf{X}^{(l)})^T\mathbf{X}^{(l)}/\widetilde{n}.$$

Then we have $\Gamma = O_P(\frac{\left|\widehat{\sigma}^2-\sigma^2\right|}{K^2} + K\sqrt{\frac{s\log p}{\widetilde{n}}})$.

We proceed to prove the Theorem 3. Without loss of generality, we set $G = \{1, 2, \cdots, p\}$. Define

$$T_G = \max_{j\in G}\sqrt{n}\left(\hat{\beta}_j^{(\text{mean})} - \beta_j\right), \quad T_{0,G} = \max_{j\in G}\sum_{l=1}^{K}\sum_{i=1}^{\widetilde{n}}\xi_{ij}.$$

Notice that

$$|T_G - T_{0,G}| \le \max_{1\le j\le p}\left|\sum_{l=1}^{K}\sum_{i=1}^{\widetilde{n}}\widehat{\xi}_{ij}/\sqrt{n} - \sum_{l=1}^{K}\sum_{i=1}^{\widetilde{n}}\xi_{ij}/\sqrt{n}\right| + \|\Delta\|_\infty,$$

where

$$\begin{aligned}\|\Delta\|_\infty &= \max_j\left(\frac{\sqrt{n}}{K}\sum_{l=1}^{K}\frac{\sum_{k\ne j}(\mathbf{z}_j^{(l)})^T\mathbf{x}_k^{(l)}(\beta_k - \widehat{\beta}_k^{(\text{init})})}{(\mathbf{z}_j^{(l)})^T\mathbf{x}_j^{(l)}}\right)\\ &\le \frac{\sqrt{n}}{K}\|\widehat{\boldsymbol{\beta}}^{(\text{init})} - \boldsymbol{\beta}\|_1\max_j\sum_{l=1}^{K}\tau_j^{(l)}\eta_j^{(l)} = O_P\left(K^{1/2}s\log(p)/\sqrt{n}\right),\end{aligned}$$

Thus by Lemma 2 and the assumption that $s^2(\log(p))^3/\widetilde{n} = o(1)$, we have

$$P\left(|T_G - T_{0,G}| > \zeta_1\right) < \zeta_2 \tag{S4.1}$$

for $\zeta_1\sqrt{1 \vee \log\left(p/\zeta_1\right)} = o(1)$ and $\zeta_2 = o(1)$.

Finally, with Lemmas 1-3 and (S4.1), applying the same arguments as in Zhang and Cheng (2017, Theorem 2.2) gives

$$\sup_{\alpha\in(0,1)}\left|P\left(\max_{j\in G}\sqrt{n}\left(\widehat{\beta}_j^{(\text{mean})} - \beta_j\right) > c_G^*(\alpha)\right) - \alpha\right| = o(1),$$

where $c_G^*(\alpha) = \inf\{t \in \mathbb{R} : P\left(W_G^* \le t|(\mathbf{y}, \mathbf{X})\right) \ge 1 - \alpha\}$ with

$$W_G^* = \max_{j\in G}\frac{\sqrt{n}}{K}\sum_{l=1}^{K}\sum_{i=1}^{\widetilde{n}}\frac{z_{i,j}^{(l)}\widehat{\sigma}e_i^{(l)}}{(\mathbf{z}_j^{(l)})^T\mathbf{x}_j^{(l)}}.$$

Since $\max_{j\in G}\sqrt{n}|\widehat{\beta}_j^{(\text{mean})} - \beta_j| = \sqrt{n}\max_{j\in G}\max\{\widehat{\beta}_j^{(\text{mean})} - \beta_j, \beta_j - \widehat{\beta}_j^{(\text{mean})}\}$, similar arguments yields

$$\sup_{\alpha\in(0,1)}\left|P\left(\max_{j\in G}\sqrt{n}\left|\widehat{\beta}_j^{(\text{mean})} - \beta_j\right| > c_G(\alpha)\right) - \alpha\right| = o(1),$$

which completes the proof of Theorem 3.

S5 Proof of Theorem 4

The proof of Theorem 4 is similar to the proof of Theorem 3 in Zhang and Zhang (2014). Following their arguments, we immediately have the equivalence of the following two statements:

$$(\widehat{\sigma}/\sigma) \vee (\sigma/\widehat{\sigma}) - 1 + \epsilon_n'\sigma^*/(\widehat{\sigma} \wedge \sigma) \le \{1 - (\widehat{\sigma}/\sigma - 1)_+\}\, c_n;$$

$$\widetilde{t}_j + \epsilon_n'\left(\sigma^*/\sigma\right)\widetilde{t}_j \le \widehat{t}_j = \left(1 + c_n\right)(\widehat{\sigma}/\sigma)\widetilde{t}_j,\ \widehat{t}_j - \widetilde{t}_j + \epsilon_n'\left(\sigma^*/\sigma\right)\widetilde{t}_j \le 2c_n\widetilde{t}_j. \tag{S5.1}$$

We proceed to prove the first part of Theorem 4. For any given $\mathbf{X}$, let $\widetilde{\varepsilon}_j = K^{-1}\sum_{l=1}^K \tau_j^{(l)} \frac{(\mathbf{Z}_j^{(l)})^T \boldsymbol{\varepsilon}^{(l)}}{\|\mathbf{Z}_j^{(l)}\|_2} \sim N\left(0, K^{-2}\sum_{l=1}^K (\tau_j^{(l)})^2\sigma^2\right)$, $\widetilde{\beta}_j = \beta_j + \widetilde{\varepsilon}_j$ and

$$\Omega_n = \left\{\left|\widetilde{\beta}_j - \widehat{\beta}_j^{(\text{mean})}\right| \le \epsilon_n' \left(\sigma^*/\sigma\right)\widetilde{t}_j, \quad \text{(S5.1) holds}, \forall j \le p\right\}.$$

As in the proof of Theorem 1, $|\widetilde{\beta}_j - \widehat{\beta}_j^{(\text{mean})}| \le K^{-1}\sum_{l=1}^K \tau_j^{(l)}\eta_j^{(l)}\|\widehat{\boldsymbol{\beta}}^{(\text{init})} - \boldsymbol{\beta}\|_1$. By the assumption that $\max_{j\le p}\eta_j^{(l)} C_1 s/\sqrt{n} \le \epsilon_n'$, we have $|\widetilde{\beta}_j - \widehat{\beta}_j^{(\text{mean})}| \le \epsilon_n' \left(\sigma^*/\sigma\right)\widetilde{t}_j$ when $\|\widehat{\boldsymbol{\beta}}^{(\text{init})} - \boldsymbol{\beta}\|_1 \le C_1 s\sigma^* L_0/\sqrt{n}$, which yields $P\left\{\Omega_n\right\} \ge 1 - 3\epsilon$. On the event Ω_n, (S5.1) gives

$$\widehat{t}_j \ge \widetilde{t}_j + |\widehat{\beta}_j^{(\text{mean})} - \widetilde{\beta}_j|, \quad |\widehat{\beta}_j^{(\text{mean})} - \widetilde{\beta}_j| + |\widehat{t}_j - \widetilde{t}_j| \le 2c_n\widetilde{t}_j.$$

Then by choosing $\Delta = 2c_n\widetilde{t}_j$ in the Lemma 1 of Zhang and Zhang (2014), we can directly come to the conclusion that

$$\begin{aligned} E\|\widehat{\boldsymbol{\beta}}^{(\text{t})} - \boldsymbol{\beta}\|_2^2 I_{\Omega_n} \le & \sum_{j=1}^p \min\left\{\beta_j^2, K^{-2}\sum_{l=1}^K (\tau_j^{(l)})^2\sigma^2\left(L_0^2\left(1+2c_n\right)^2 + 1\right)\right\} \\ & + K^{-1}\left(\epsilon L_n/p\right)\sigma^2\sum_{j=1}^p \widetilde{\tau}_j^2, \end{aligned}$$

where $L_n = 4/L_0^3 + 4c_n/L_0 + 12c_n^2 L_0$.

It remains to prove the second part of Theorem 4. Following the argument of Zhang and Zhang (2014), in view of $\widehat{t}_j \ge \widetilde{t}_j + \left|\widehat{\beta}_j - \widetilde{\beta}_j\right|$, thus $\left|\widehat{\beta}_j\right| > \widehat{t}_j$ implies $|\widetilde{\varepsilon}_j| > \widetilde{t}_j$ for $\beta_j = 0$; in view of $\left|\widehat{\beta}_j - \widetilde{\beta}_j\right| + \left|\widehat{t}_j - \widetilde{t}_j\right| \le 2c_n\widetilde{t}_j$, thus $\left|\widehat{\beta}_j\right| \le \widehat{t}_j$ implies $|\widetilde{\varepsilon}_j| > \widetilde{t}_j$ for $|\beta_j| > \left(2+2c_n\right)\widetilde{t}_j$. Combining the above results gives

$$P\left(\left\{j : |\beta_j| > \left(2+2c_n\right)\widetilde{t}_j\right\} \subseteq \widehat{S}^{(\text{t})} \subseteq \{j : \beta_j \ne 0\}\right) \ge P\left\{\Omega_n^c\right\} + pP\left\{|\widetilde{\varepsilon}_j| > \widetilde{t}_j\right\}.$$

Clearly, we have

$$P\left\{|\widetilde{\varepsilon}_j| > \widetilde{t}_j \middle| \mathbf{X}\right\} \le P\left\{|\widetilde{\varepsilon}_j| > K^{-1}(\sum_{l=1}^K (\tau_j^{(l)})^2)^{1/2}\sigma L_0 \middle| \mathbf{X}\right\} = 2\Phi(-L_0) \le \alpha/p.$$

Thus combining the above two inequalities completes the proof of the second part of Theorem 4.

S6 Proofs of Lemmas

S6.1 Proof of Lemma 2

With some simple algebra, we obtain

$$\begin{aligned}\left|\sum_{l=1}^{K}\sum_{i=1}^{\widetilde{n}}\widehat{\xi}_{ij}/\sqrt{n}-\sum_{l=1}^{K}\sum_{i=1}^{\widetilde{n}}\xi_{ij}/\sqrt{n}\right| &= \left|\frac{1}{K}\sum_{l=1}^{K}\left((\widehat{\mathbf{\Theta}}_j^{(l)})^T-\mathbf{\Theta}_j^T\right)\sum_{i=1}^{\widetilde{n}}\widetilde{\mathbf{x}}_i^{(l)}\varepsilon_i^{(l)}/\sqrt{n}\right| \\ &\le \frac{1}{K}\sum_{l=1}^{K}\left\|\widehat{\mathbf{\Theta}}_j^{(l)}-\mathbf{\Theta}_j\right\|_1\left\|\sum_{i=1}^{\widetilde{n}}\widetilde{\mathbf{x}}_i^{(l)}\varepsilon_i^{(l)}/\sqrt{n}\right\|_\infty\end{aligned} \tag{S6.1}$$

Since the same argument in the proof of Lemma 1.2 in Zhang and Cheng (2017) gives

$$\mathbb{E}\left\{\max_{1\le j\le p}\left|\sum_{i=1}^{\widetilde{n}}x_{ij}^{(l)}\varepsilon_i/\widetilde{n}\right|\right\}\lesssim\sqrt{\log(p)/\widetilde{n}}+\log(\widetilde{n}p)\log\widetilde{n}\log(p)/\widetilde{n},$$

for any $1\le l\le K$, we proceed to derive the bounds of $\left\|\widehat{\mathbf{\Theta}}_j^{(l)}-\mathbf{\Theta}_j\right\|_1$.

By the definitions of $\widehat{\mathbf{\Theta}}_j^{(l)}$ and $\mathbf{\Theta}_j$, it follows that

$$\begin{aligned}\left\|\widehat{\mathbf{\Theta}}_j^{(l)}-\mathbf{\Theta}_j\right\|_1 &= \left\|\widehat{\mathbf{C}}_j^{(l)}/(\widehat{\iota}_j^{(l)})^2-\widetilde{\mathbf{C}}_j/\iota_j^2\right\|_1 \\ &\le \underbrace{\left\|\widehat{\boldsymbol{\gamma}}_j^{(l)}-\boldsymbol{\gamma}_j\right\|_1/(\widehat{\iota}_j^{(l)})^2}_{i}+\underbrace{\left\|\boldsymbol{\gamma}_j\right\|_1\left(1/(\widehat{\iota}_j^{(l)})^2-1/\iota_j^2\right)}_{ii},\end{aligned} \tag{S6.2}$$

where $\widehat{\mathbf{C}}_j^{(l)}$ and $\widetilde{\mathbf{C}}_j$ are the jth rows of $\widehat{\mathbf{C}}^{(l)}$ and $\mathbf{C}$, respectively. Moreover, we have

$$\begin{aligned}\left|(\widehat{\iota}_j^{(l)})^2-\iota_j^2\right| &= \underbrace{\left|(\boldsymbol{\rho}_j^{(l)})^T\boldsymbol{\rho}_j^{(l)}/\widetilde{n}-\iota_j^2\right|}_{I}+\underbrace{\left|(\boldsymbol{\rho}_j^{(l)})^T\mathbf{X}_{-j}^{(l)}\left(\widehat{\boldsymbol{\gamma}}_j^{(l)}-\boldsymbol{\gamma}_j\right)/\widetilde{n}\right|}_{II} \\ &\quad+\underbrace{\left|(\boldsymbol{\rho}_j^{(l)})^T\mathbf{X}_{-j}^{(l)}\boldsymbol{\gamma}_j/\widetilde{n}\right|}_{III}+\underbrace{\left|\left(\boldsymbol{\gamma}_j\right)^T(\mathbf{X}_{-j}^{(l)})^T\mathbf{X}_{-j}^{(l)}\left(\widehat{\boldsymbol{\gamma}}_j^{l}-\boldsymbol{\gamma}_j\right)/\widetilde{n}\right|}_{IV},\end{aligned}$$

where $\boldsymbol{\rho}_j^{(l)} = \mathbf{x}_j^{(l)} - \mathbf{X}_{-j}^{(l)}\boldsymbol{\gamma}_j$.

As for i in (S6.2), by the same argument as in (S1), we have

$$(\widehat{\iota}_j^{(l)})^2 = \frac{(\mathbf{z}_j^{(l)})^T\mathbf{x}_j^{(l)}}{\widetilde{n}} = O(1), \quad \|\widehat{\boldsymbol{\gamma}}_j^{(l)} - \boldsymbol{\gamma}_j\|_1 = O(\frac{s_j^*\sqrt{\log p}}{\sqrt{\widetilde{n}}})$$

with probability at least $1 - o(p^{-\delta+1})$ for some $\delta > 1$, where $s_j^* = \|\gamma_j\|_0$. As for ii in (S6.2), since $\|\boldsymbol{\rho}_j^{(l)}\|_2^2/\sigma_j \sim \chi^2_{(\widetilde{n})}$ for any $1 \leq j \leq p$, applying the same argument as in (S1) gives

$$I = O(\sqrt{\log(p)/\widetilde{n}}),$$

holding with probability at least $1 - 2p^{-\delta}$. Second, under the Gaussian assumption of $\boldsymbol{\rho}_j^{(l)}$, it follows that

$$\left\|(\boldsymbol{\rho}_j^{(l)})^T\mathbf{X}_{-j}^{(l)}\right\|_\infty/\widetilde{n} = O(\sqrt{\log(p)/\widetilde{n}}),$$

holding with probability at least $1 - o(p^{-\delta+1})$, which entails

$$II \leq \left\|(\boldsymbol{\rho}_j^{(l)})^T\mathbf{X}_{-j}^{(l)}\right\|_\infty \left\|\widehat{\boldsymbol{\gamma}}_j^{(l)} - \boldsymbol{\gamma}_j\right\|_1/\widetilde{n} = O(\frac{s_j^*\log p}{\widetilde{n}}),$$

holding with probability at least $1 - o(p^{-\delta+1})$. Similarly, since $\|\boldsymbol{\gamma}_j\|_1 \leq \sqrt{s_j^*}\|\boldsymbol{\gamma}_j\|_2 \leq \sqrt{s_j^*\sigma_{jj}}/\lambda_{\min}(\boldsymbol{\Sigma}) = O(\sqrt{s_j^*})$ with $\lambda_{\min}(\boldsymbol{\Sigma})$ indicating the minimum eigenvalue of $\boldsymbol{\Sigma}$, we have

$$III \leq \left\|(\boldsymbol{\rho}_j^{(l)})^T\mathbf{X}_{-j}^{(l)}\right\|_\infty \left\|\boldsymbol{\gamma}_j\right\|_1/\widetilde{n} = O(\sqrt{\frac{s_j^*\log p}{\widetilde{n}}}),$$

with probability at least $1 - o(p^{-\delta+1})$.

As for IV, the KKT condition yields

$$\|(\mathbf{X}_{-j}^{(l)})^T(\mathbf{x}_j^{(l)} - \mathbf{X}_{-j}(\widehat{\boldsymbol{\gamma}}_j^{(l)}))\|_\infty/\widetilde{n} \leq \frac{\max_{k\neq j}\|\mathbf{x}_k^{(l)}\|_2}{\sqrt{\widetilde{n}}}\widehat{\sigma}_j\lambda_0.$$

Combining the facts $\left\|(\boldsymbol{\rho}_j^{(l)})^T\mathbf{X}_{-j}^{(l)}\right\|_\infty/\widetilde{n} = O_P(\sqrt{\log p/\widetilde{n}})$ and $\frac{\|\mathbf{X}_k^{(l)}\|_2}{\sqrt{\widetilde{n}}} = O_P(1)$ gives

$$IV = O(\sqrt{s_j^* \log(p)/\widetilde{n}})$$

holding with probability at least $1-o(p^{-\delta+1})$. Thus with probability at least $1-o(p^{-\delta+1})$, we have

$$1/(\widehat{\iota}_j^{(l)})^2 - 1/\iota_j^2 = O(\sqrt{s_j^* \log(p)/\widetilde{n}}).$$

We can come to the conclusion that

$$i = O_P(s_j^* \sqrt{\log(p)/\widetilde{n}}), \quad ii = O_P(s_j^* \sqrt{\log(p)/\widetilde{n}}),$$

which entails that

$$\left\|\widehat{\boldsymbol{\Theta}}_j^{(l)} - \boldsymbol{\Theta}_j\right\|_1 = O_P(s_j^* \sqrt{\log(p)/\widetilde{n}}).$$

Returning to the equality (S6.1), with assumption that $o(Kp^{-\delta+1}) = o(1)$, we now have

$$\begin{aligned}
&\left|\sum_{l=1}^{K}\sum_{i=1}^{\widetilde{n}} \widehat{\xi}_{ij}/\sqrt{n} - \sum_{l=1}^{K}\sum_{i=1}^{\widetilde{n}} \xi_{ij}/\sqrt{n}\right| \le \frac{1}{K}\sum_{l=1}^{K}\left\|\widehat{\boldsymbol{\Theta}}_j^{(l)} - \boldsymbol{\Theta}_j\right\|_1 \left\|\sum_{i=1}^{\widetilde{n}} \widetilde{\mathbf{x}}_i^{(l)}\varepsilon_i^{(l)}/\sqrt{n}\right\|_\infty \\
&= O_P(\frac{\sqrt{\log(p)\widetilde{n}} + \log(\widetilde{n}p)\log\widetilde{n}\log(p)}{\sqrt{n}K}\sum_{l=1}^{K}\left\|\widehat{\boldsymbol{\Theta}}_j^{(l)} - \boldsymbol{\Theta}_j\right\|_1) \\
&= O_P(\frac{s_j^*\log p}{\sqrt{n}} + \frac{s_j^*(\log p)^{3/2}\log(\widetilde{n}p)\log\widetilde{n}}{n}) \le O_P(\max_j \frac{\sqrt{s}\log p}{\sqrt{n}}).
\end{aligned}$$

Choosing ζ_1 such that $\max_j \sqrt{s}\log(p)/(\sqrt{n}\zeta_1) = o(1)$ and $\zeta_1\sqrt{1 \vee \log(p/\zeta_1)} = o(1)$, then we can get the conclusion of Lemma 2 and finish the proof.

S6.2 Proof of Lemma 3

We need to derive the bounds of $\|(\widehat{\mathbf{\Theta}}^{(l)})^T\widehat{\mathbf{\Sigma}}^{(l)}\widehat{\mathbf{\Theta}}^{(l)} - \mathbf{\Theta}\|_\infty$. With some simple algebra, we have

$$\begin{aligned}\|(\widehat{\mathbf{\Theta}}^{(l)})^T\widehat{\mathbf{\Sigma}}^{(l)}\widehat{\mathbf{\Theta}}^{(l)} - \mathbf{\Theta}\|_\infty &= \|((\widehat{\mathbf{\Theta}}^{(l)})^T\widehat{\mathbf{\Sigma}}^{(l)} - \mathbf{I})\widehat{\mathbf{\Theta}}^{(l)} + \widehat{\mathbf{\Theta}}^{(l)} - \mathbf{\Theta}\|_\infty \\ &\leq \|((\widehat{\mathbf{\Theta}}^{(l)})^T\widehat{\mathbf{\Sigma}}^{(l)} - \mathbf{I})\widehat{\mathbf{\Theta}}^{(l)}\|_\infty + \|\widehat{\mathbf{\Theta}}^{(l)} - \mathbf{\Theta}\|_\infty.\end{aligned}$$

On the one hand, applying the same argument as in (S1) gives

$$\|\widehat{\boldsymbol{\gamma}}_j^{(l)}\|_1 \leq \|\boldsymbol{\gamma}_j^{(l)}\|_1 + \|\widehat{\boldsymbol{\gamma}}_j^{(l)} - \boldsymbol{\gamma}_j^{(l)}\|_1 = O(\sqrt{s_j^*}) + O(s_j^*\sqrt{\frac{\log p}{\widetilde{n}}}) = O(\sqrt{s_j^*}),$$

holding with probability at least $1 - o(p^{-\delta+1})$, which entails that $\|\widehat{\mathbf{\Theta}}_j^{(l)}\|_1 = O_P(\sqrt{s_j^*})$. On the other hand, since $\|\widehat{\boldsymbol{\gamma}}_j^{(l)} - \boldsymbol{\gamma}_j^{(l)}\|_2 \leq \|\widehat{\boldsymbol{\gamma}}_j^{(l)} - \boldsymbol{\gamma}_j^{(l)}\|_1 = O_P(s_j^*\sqrt{\frac{\log p}{\widetilde{n}}})$, we have $\|\widehat{\mathbf{\Theta}}_j^{(l)} - \mathbf{\Theta}_j\|_2 = O(s_j^*\sqrt{\frac{\log p}{\widetilde{n}}})$ holding with probability at least $1 - o(p^{-\delta+1})$. Combining these results gives

$$\begin{aligned}\|(\widehat{\mathbf{\Theta}}^{(l)})^T\widehat{\mathbf{\Sigma}}^{(l)}\widehat{\mathbf{\Theta}}^{(l)} - \mathbf{\Theta}\|_\infty &\leq \max_j \frac{\max_{k\neq j}\|\mathbf{x}_k^{(l)}\|_2}{\sqrt{\widetilde{n}}}\widehat{\sigma}_j\lambda_0\|\widehat{\mathbf{\Theta}}_j^{(l)}\|_1 + \max_j\|\widehat{\mathbf{\Theta}}_j^{(l)} - \mathbf{\Theta}_j\|_2 \\ &= O(\max_j s_j^*\sqrt{\frac{\log p}{\widetilde{n}}}),\end{aligned}$$

holding with probability at least $1 - o(p^{-\delta+1})$, which yileds that

$$\max_{1\leq j,k\leq p}\left|(\widehat{\mathbf{\Theta}}_j^{(l)})^T\widehat{\mathbf{\Sigma}}^{(l)}\widehat{\mathbf{\Theta}}_k^{(l)} - \mathbf{\Theta}_j^T\mathbf{\Sigma}\mathbf{\Theta}_k\right| = O_P(\max_j s_j^*\sqrt{\frac{\log p}{\widetilde{n}}}). \tag{S6.3}$$

Moreover, by the same arguments as in the proof of Zhang and Cheng (2017, Theorem 2.2), we have

$$|\mathbf{\Theta}_j^T\mathbf{\Sigma}\mathbf{\Theta}_k| \leq 1/(\iota_j\iota_k) = O(1),$$

uniformly for $1 \leq j, k \leq p$. Thus, with assumption that $o(Kp^{-\delta+1}) = o(1)$, combining this result and inequality (S6.3) gives

$$\Gamma = O_P(\frac{|\widehat{\sigma}^2 - \sigma^2|}{K^2} + K\sqrt{\frac{s \log p}{\widetilde{n}}}),$$

which completes the proof of Lemma 3.

References

Ren, Z., Sun, T., Zhang, C.-H. and Zhou, H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical model. *Ann. Statist.* **43**, 991–1026.

Zhang, X. and Cheng, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* **112**, 757–768.

Zhang, S. and Zhang, C.-H. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Statist. Soc. Ser.* B **76**, 217–242.