High-Dimensional Inference Via Hybrid Orthogonalization

Yang Li[1], Jia Zhou[*2], Zemin Zheng[1], Ziwei Zhu[3]

[1] *University of Science and Technology of China*

[2] *Hefei University of Technology,* [3] *Radix Trading*

Supplementary Material

This supplementary material consists of three parts. Section S1 considers the sub-Gaussian design for HOT. Section S2 lists the key lemmas and presents the proofs of the main results. Additional technical proofs for the lemmas are provided in Section S3.

S1 Sub-Gaussian design

S1.1 Theoretical results for sub-Gaussian design

In this section, we extend the theoretical results in Theorem 4 to allow for sub-Gaussian features. The formal definition of a sub-Gaussian random vector is presented in Definition 2 below.

Definition 2. A mean-zero random vector $\mathbf{x} \in \mathbb{R}^d$ is said to be K-sub-

[*] Corresponding author

[†] Yang Li and Jia Zhou contributed equally to this work.

Gaussian if its moment generating function satisfies

$$\mathbb{E}[\exp(s\mathbf{u}^\top \mathbf{x})] \leq \exp\left(\frac{K^2 s^2}{2}\right),$$

for any $s \in \mathbb{R}$ and any unit vector $\mathbf{u} \in \mathbb{R}^d$. Equivalently, we also write $\mathbf{x} \sim \mathrm{subG}(K^2)$.

As before, consider a random design matrix $\mathbf{X}$ with i.i.d. rows of mean zero and population covariance $\mathbf{\Sigma}$. Moreover, the strong/weak signal assumptions are identical to those in Section 3. Here we relax the Gaussian assumption in Condition 1 to the following sub-Gaussian assumption.

Condition 4. *Suppose that the random design matrix $\mathbf{X}$ has i.i.d. K-sub-Gaussian rows for some constant K, and the eigenvalues of $\mathbf{\Phi} = \mathbf{\Sigma}^{-1} = (\phi_{ij})_{p\times p}$ are bounded within the interval $[1/L, L]$ for some constant $L \geq 1$.*

Condition 4 assumes the design matrix has i.i.d. sub-Gaussian rows, with bounded precision matrix eigenvalues. A similar sub-Gaussian assumption has been considered in van de Geer et al. (2014). Now we are ready to present the main theoretical results.

Theorem 5. *For $1 \leq j \leq p$, let $\lambda_j = (1+\epsilon)\sqrt{2\delta\log(p)/(nc_{K,L})}$ for some constants $\epsilon > 0$, $\delta > 1$, and $c_{K,L} > 0$ depending on K and L. Assume that Condition 4 holds and that $\widehat{S}$ satisfies Definition 1. Then for any $1 \leq j \leq p$,* *if* $s_1(\log p)/n = o(1)$ *and* $s_\Phi(\log p)/n = o(1)$*, we have*

$$\widehat{\beta}_j - \beta_j = W_j + \Delta_j,$$

where $W_j = \mathbf{z}_j^\top \boldsymbol{\varepsilon} / \mathbf{z}_j^\top \mathbf{x}_j$ *and* $|\Delta_j| = o_{\mathbb{P}}(n^{-1/2})$*. Moreover, it holds with probability at least* $1 - o(p^{1-\delta}) - \theta_{n,p}$ *that*

$$\tau_j := \|\mathbf{z}_j\|_2 / |\mathbf{z}_j^\top \mathbf{x}_j| \asymp n^{-1/2}.$$

Conditional on $\mathbf{X}$*, suppose that the components of* $\boldsymbol{\varepsilon}$ *are independent and identically distributed with mean 0 and variance* σ^2*. For any given* $\mathbf{X}$*, if there is some* $\varsigma > 0$ *such that the inequality*

$$\lim_{n\to\infty} \frac{1}{(\sigma\|\mathbf{z}_j\|_2)^{2+\varsigma}} \sum_{k=1}^n \mathbb{E}|z_{jk}\varepsilon_k|^{2+\varsigma} = 0$$

holds, we further have

$$\lim_{n\to\infty} \mathbb{P}(W_j \leq \tau_j \sigma x) = \Phi(x), \quad \forall x \in \mathbb{R},$$

where $\Phi(\cdot)$ *is the distribution function of the standard normal distribution.*

As in Theorem 4, Theorem 5 establishes the asymptotic normality of the noise term W_j under the classical Lyapunov condition, which relaxes the Gaussian noise assumption in Zhang and Zhang (2014). In other words, Theorem 5 accommodates non-Gaussian designs and non-Gaussian errors, exhibiting a similar spirit to Theorem 2.4 in van de Geer et al. (2014). Note that in their work, the design $\mathbf{X}$ is assumed to have either i.i.d. sub-Gaussian rows or i.i.d. bounded rows, while the errors are assumed to be sub-exponential.

Table 5: Four performance measures by different methods over 100 replications in Section S1.2.

	H-SIS(I)	H-SIS	H-HOLP(I)	H-HOLP	LDPE	adLDPE	STPS
				Setting 1			
CP_all	0.956	0.947	0.957	0.951	0.949	0.976	0.969
CP_max	0.958	0.955	0.962	0.940	0.856	0.796	0.936
Length	0.708	0.719	0.719	0.708	0.613	0.877	1.409
Time (s)	22.32	23.12	23.34	23.77	42.35	41.57	122.77
				Setting 2			
CP_all	0.961	0.956	0.965	0.957	0.949	0.965	0.960
CP_max	0.964	0.956	0.950	0.966	0.848	0.806	0.962
Length	0.702	0.686	0.703	0.695	0.615	0.814	1.432
Time (s)	23.12	22.78	22.93	23.37	41.79	41.48	120.12
				Setting 3			
CP_all	0.965	0.957	0.962	0.957	0.951	0.959	0.961
CP_β_1	0.97	0.95	0.94	0.95	0.92	0.94	0.96
Length	0.694	0.704	0.699	0.704	0.629	0.801	1.420
Time (s)	22.76	23.08	23.27	23.88	41.65	41.79	119.07

S1.2 Simulation studies

In this section, all simulation settings are identical to those in Section 4, with the only difference being that the rows of the design matrix follow a sub-Gaussian distribution instead of a Gaussian one. Specifically, similar to Janková and van de Geer (2017), we set $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}^{1/2}$, where $\mathbf{U}$ is an

$n \times p$ matrix whose entries are i.i.d. draws from the uniform distribution $U[-\sqrt{3}, \sqrt{3}]$, and $\mathbf{\Sigma} = (\rho^{|j-k|})_{p\times p}$ with $\rho = 0.9$. All simulations were performed on a Lenovo Legion Y9000P (Intel Core i9-13900HX, 32 GB RAM). The reported time is the CPU time (in seconds) for a single replication.

The simulation results in Table 5 are consistent with those in Section 4. Namely, the proposed HOT method achieves an attractive balance among accurate coverage, short interval length, and low computational cost across all settings, demonstrating a clear advantage over other competing methods.

S2 Proofs of main results

S2.1 Lemmas

The following lemmas will be used in the proofs of the main results.

Lemma 1. *Suppose that* $\mathbf{X} \sim N(0, \mathbf{I}_n \otimes \mathbf{\Sigma})$ *and the eigenvalues of* $\mathbf{\Phi} = \mathbf{\Sigma}^{-1} = (\phi_{ij})_{p\times p}$ *are bounded within the interval* $[1/L, L]$ *for some constant* $L \geq 1$. *For a given set* $\widehat{S}$ *and any* $j \in \widehat{S}^c$, *we have the following conditional distribution*

$$\mathbf{x}_j = \mathbf{X}_{\widehat{S}}\boldsymbol{\gamma}_j^{(j)} + \boldsymbol{\rho}_j^{(j)},$$

where the residual vector $\boldsymbol{\rho}_j^{(j)} = \mathbf{x}_j - \mathbb{E}(\mathbf{x}_j|\mathbf{X}_{\widehat{S}})$ *is independent of* $\mathbf{X}_{\widehat{S}}$ *and* $\boldsymbol{\gamma}_j^{(j)} = \mathbf{\Sigma}_{\widehat{S}\widehat{S}}^{-1}\mathbf{\Sigma}_{\widehat{S}j}$ *is the corresponding regression coefficient vector. Moreover,* *the residual matrix* $\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\backslash\{j\}}$ *consisting of the residual vectors* $\boldsymbol{\rho}^{(j)}_k$ *for* $k \in \widehat{S}^c\backslash\{j\}$ *satisfies*

$$\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\backslash\{j\}} \sim N\left(\mathbf{0}, \mathbf{I}_n \otimes \left[(\boldsymbol{\Phi}_{\widehat{S}^c\widehat{S}^c})^{-1}\right]_{\widehat{S}^c\backslash\{j\},\widehat{S}^c\backslash\{j\}}\right).$$

Lemma 2. *Suppose that the eigenvalues of a symmetric matrix* $\boldsymbol{\Phi} = (\phi_{ij})_{p\times p} = \boldsymbol{\Sigma}^{-1} = (\sigma_{ij})^{-1}_{p\times p}$ *are bounded within the interval* $[1/L, L]$ *for some constant* $L \geq 1$*. Then we have*

1. *the eigenvalues of* $\boldsymbol{\Sigma}$ *are also bounded within* $[1/L, L]$*;*
2. *the diagonal components of* $\boldsymbol{\Phi}$ *and* $\boldsymbol{\Sigma}$ *are bounded within* $[1/L, L]$*;*
3. *the eigenvalues of any principal submatrix of* $\boldsymbol{\Phi}$ *and* $\boldsymbol{\Sigma}$ *are bounded within* $[1/L, L]$*;*
4. *the* L_2*-norms of the rows of* $\boldsymbol{\Phi}$ *and* $\boldsymbol{\Sigma}$ *are bounded within* $[1/L, L]$*;*
5. $\sigma_{jj}\phi_{jj} \geq 1$ *for any* $j \in \{1, \ldots, p\}$*.*

Lemma 3. *Consider the following linear regression model*

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n),$$

where $\mathbf{y} \in \mathbb{R}^n$ *is a response vector,* $\mathbf{X} \in \mathbb{R}^{n\times d}$ *is a random design matrix,* $\boldsymbol{\beta} \in \mathbb{R}^d$ *is a vector of unknown regression coefficients, and* $\boldsymbol{\varepsilon}$ *is a random error vector that is independent of* $\mathbf{X}$*. Suppose that* $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$ *and the eigenvalues of* $\boldsymbol{\Sigma} = (\sigma_{ij})_{d\times d}$ *are bounded within the interval* $[1/L, L]$ *for* *some constant $L \geq 1$. If $n \geq d$, then the ordinary least squares estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$ exists with probability one. Moreover, for any $p > 1$ and any constant $\delta > 1$, with probability at least $1 - 2p^{-2\delta}$, it satisfies*

$$\max\{d - 2\sqrt{2\delta d \log p}, 0\}\sigma^2 \leq \|\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_2^2 \leq \{d + 2\sqrt{2\delta d \log p} + 4\delta \log p\}\sigma^2.$$

Lemma 4. *Suppose that the $n \times p$ random matrix $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$, where the eigenvalues of $\boldsymbol{\Sigma}$ are bounded within the interval $[1/L, L]$ for some constant $L \geq 1$. For a positive constant ξ and a set $S \subset \{1, \ldots, p\}$, the cone invertibility factor and the sign-restricted cone invertibility factor are defined as*

$$F_2^*(\xi, S) = \inf\left\{s^{1/2}\|\mathbf{X}^\top \mathbf{X}\mathbf{u}/n\|_\infty / \|\mathbf{u}\|_2 : \mathbf{u} \in \mathcal{G}(\xi, S)\right\} \text{ and}$$
$$F_2(\xi, S) = \inf\left\{s^{1/2}\|\mathbf{X}^\top \mathbf{X}\mathbf{u}/n\|_\infty / \|\mathbf{u}\|_2 : \mathbf{u} \in \mathcal{G}_-(\xi, S)\right\},$$

respectively, where $s = |S|$, the cone $\mathcal{G}(\xi, S) = \{\mathbf{u} \in \mathbb{R}^{p-1} : \|\mathbf{u}_{S^c}\|_1 \leq \xi\|\mathbf{u}_S\|_1 \neq 0\}$, and the sign-restricted cone $\mathcal{G}_-(\xi, S) = \{\mathbf{u} \in \mathbb{R}^{p-1} : \|\mathbf{u}_{S^c}\|_1 \leq \xi\|\mathbf{u}_S\|_1 \neq 0, u_j \mathbf{x}_j^\top \mathbf{X}\mathbf{u}/n \leq 0, \forall j \notin S\}$. Then for $s = o(n/\log(p))$, there exist positive constants c_1, c_2, c_3 such that with probability at least $1 - c_2 \exp(-c_3 n)$,

$$F_2(\xi, S) \geq F_2^*(\xi, S) \geq c_1.$$

Lemma 5. *Suppose that $\mathbf{X}$ has i.i.d. K-sub-Gaussian rows of mean zero and population covariance $\boldsymbol{\Sigma}$, and the eigenvalues of $\boldsymbol{\Phi} = \boldsymbol{\Sigma}^{-1} = (\phi_{ij})_{p\times p}$ are bounded within the interval $[1/L, L]$ for some constant $L \geq 1$. For a* *given set $\widehat{S}$ and any $j \in \widehat{S}^c$, we have*

$$\mathbf{x}_j = \mathbf{X}_{\widehat{S}} \boldsymbol{\gamma}_j^{(j)} + \boldsymbol{\rho}_j^{(j)},$$

where $\boldsymbol{\gamma}_j^{(j)} = \arg\min_{\boldsymbol{\gamma}} \mathbb{E} \left\| \mathbf{x}_j - \mathbf{X}_{\widehat{S}} \boldsymbol{\gamma} \right\|_2^2 = \boldsymbol{\Sigma}_{\widehat{S}\widehat{S}}^{-1} \boldsymbol{\Sigma}_{\widehat{S}j}$ is the population least squares regression coefficient and the residual vector $\boldsymbol{\rho}_j^{(j)}$ satisfies $\mathbb{E}(\mathbf{X}_{\widehat{S}}^\top \boldsymbol{\rho}_j^{(j)}) = \mathbf{0}$. Moreover, the rows of the residual matrix $\boldsymbol{\rho}_{\widehat{S}^c \setminus \{j\}}^{(j)}$ are i.i.d. K'-sub-Gaussian random vectors with zero mean and common covariance matrix $\left[(\boldsymbol{\Phi}_{\widehat{S}^c \widehat{S}^c})^{-1} \right]_{\widehat{S}^c \setminus \{j\}, \widehat{S}^c \setminus \{j\}}$, where K' is a constant depending on K and L.

Lemma 6. *Consider the following linear regression model*

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y} \in \mathbb{R}^n$ is the response vector, $\mathbf{X} \in \mathbb{R}^{n \times d}$ is the random design matrix, $\boldsymbol{\beta} \in \mathbb{R}^d$ is the vector of regression coefficients, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ is the error vector. Suppose that $(\mathbf{X}_{i\cdot}^\top, \varepsilon_i)$, $1 \leq i \leq n$, are i.i.d. copies of $(\mathbf{x}, \varepsilon)$, where $\mathbf{x}$ is a mean-zero K_2-sub-Gaussian random vector with covariance matrix $\boldsymbol{\Sigma}$ and ε is a mean-zero K_1-sub-Gaussian random variable with variance $\sigma^2 > 0$. Assume further that $\mathbb{E}(\mathbf{x}\varepsilon) = \mathbf{0}$ and that the eigenvalues of $\boldsymbol{\Sigma} = (\sigma_{ij})_{d \times d}$ are bounded within the interval $[1/L, L]$ for some constant $L \geq 1$. For any $p > d$ satisfying $d \log(p)/n = o(1)$ and any constant $\delta > 1$, with probability at least $1 - o(p^{1-\delta})$, the ordinary least squares estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ exists and satisfies $\|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\|_2 \asymp \sqrt{n}$.

Lemma 7. *Suppose that $\mathbf{X}$ has i.i.d. K-sub-Gaussian rows of mean zero* *and population covariance* $\mathbf{\Sigma}$*, and the eigenvalues of* $\mathbf{\Phi} = \mathbf{\Sigma}^{-1} = (\phi_{ij})_{p\times p}$ *are bounded within the interval* $[1/L, L]$ *for some constant* $L \geq 1$*. For a positive constant* ξ *and a set* $S \subset \{1, \ldots, p\}$*, the cone invertibility factor and the sign-restricted cone invertibility factor are defined as*

$$F_2^*(\xi, S) = \inf\left\{s^{1/2}\|\mathbf{X}^\top\mathbf{X}\mathbf{u}/n\|_\infty/\|\mathbf{u}\|_2 : \mathbf{u} \in \mathcal{G}(\xi, S)\right\} \text{ and}$$

$$F_2(\xi, S) = \inf\left\{s^{1/2}\|\mathbf{X}^\top\mathbf{X}\mathbf{u}/n\|_\infty/\|\mathbf{u}\|_2 : \mathbf{u} \in \mathcal{G}_-(\xi, S)\right\},$$

respectively, where $s = |S|$*, the cone* $\mathcal{G}(\xi, S) = \{\mathbf{u} \in \mathbb{R}^p : \|\mathbf{u}_{S^c}\|_1 \leq \xi\|\mathbf{u}_S\|_1 \neq 0\}$*, and the sign-restricted cone* $\mathcal{G}_-(\xi, S) = \{\mathbf{u} \in \mathbb{R}^p : \|\mathbf{u}_{S^c}\|_1 \leq \xi\|\mathbf{u}_S\|_1 \neq 0, u_j\mathbf{x}_j^\top\mathbf{X}\mathbf{u}/n \leq 0, \forall j \notin S\}$*. Then for* $s = o(n/\log(p))$*, there exist positive constants* c_1*,* c_2*,* c_3 *such that with probability at least* $1 - c_2\exp(-c_3 n)$*,*

$$F_2(\xi, S) \geq F_2^*(\xi, S) \geq c_1.$$

S2.2 Proof of Proposition 1

We first prove the equivalence of the optimization problems. For any j, $1 \leq j \leq p$, by definition we have

$$\widehat{\boldsymbol{\pi}}_{-j}^{(j)} = \underset{\mathbf{b}_{-j}}{\arg\min}\Big\{(2n)^{-1}\big\|\mathbf{x}_j - \mathbf{X}_{-j}\mathbf{b}_{-j}\big\|_2^2 + \lambda_j \sum_{k\in\widehat{S}^c\setminus\{j\}} v_{jk}\cdot|b_k|\Big\}. \tag{S2.1}$$

Since the coefficients in the set $\widehat{S}\setminus\{j\}$ are not penalized, the Karush-Kuhn-Tucker (KKT) condition for $\widehat{\boldsymbol{\pi}}_{\widehat{S}\setminus\{j\}}^{(j)}$ gives

$$-\mathbf{X}_{\widehat{S}\setminus\{j\}}^\top(\mathbf{x}_j - \mathbf{X}_{\widehat{S}\setminus\{j\}}\widehat{\boldsymbol{\pi}}_{\widehat{S}\setminus\{j\}}^{(j)} - \mathbf{X}_{\widehat{S}^c\setminus\{j\}}\widehat{\boldsymbol{\pi}}_{\widehat{S}^c\setminus\{j\}}^{(j)}) = \mathbf{0},$$

which entails

$$\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}\backslash\{j\}} = (\mathbf{X}^\top_{\widehat{S}\backslash\{j\}}\mathbf{X}_{\widehat{S}\backslash\{j\}})^{-1}\mathbf{X}^\top_{\widehat{S}\backslash\{j\}}(\mathbf{x}_j - \mathbf{X}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}). \tag{S2.2}$$

When $\mathbf{b}^{(j)}_{\widehat{S}\backslash\{j\}} = (\mathbf{X}^\top_{\widehat{S}\backslash\{j\}}\mathbf{X}_{\widehat{S}\backslash\{j\}})^{-1}\mathbf{X}^\top_{\widehat{S}\backslash\{j\}}(\mathbf{x}_j - \mathbf{X}_{\widehat{S}^c\backslash\{j\}}\mathbf{b}^{(j)}_{\widehat{S}^c\backslash\{j\}})$, we have

$$\mathbf{x}_j - \mathbf{X}_{-j}\mathbf{b}^{(j)}_{-j} = (\mathbf{I}_n - \mathbf{P}_{\widehat{S}\backslash\{j\}})(\mathbf{x}_j - \mathbf{X}_{\widehat{S}^c\backslash\{j\}}\mathbf{b}^{(j)}_{\widehat{S}^c\backslash\{j\}}) = \boldsymbol{\psi}^{(j)}_j - \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\mathbf{b}^{(j)}_{\widehat{S}^c\backslash\{j\}}. \tag{S2.3}$$

Thus, it follows from (S2.2) and (S2.3) that the optimization problem (S2.1) is equivalent to

$$\begin{cases} \widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}} = \underset{\mathbf{b}_{\widehat{S}^c\backslash\{j\}}}{\arg\min}\Big\{(2n)^{-1}\big\|\boldsymbol{\psi}^{(j)}_j - \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\mathbf{b}_{\widehat{S}^c\backslash\{j\}}\big\|_2^2 + \lambda_j \sum_{k\in\widehat{S}^c\backslash\{j\}} v_{jk}\cdot|b_k|\Big\}, \\ \widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}\backslash\{j\}} = (\mathbf{X}^\top_{\widehat{S}\backslash\{j\}}\mathbf{X}_{\widehat{S}\backslash\{j\}})^{-1}\mathbf{X}^\top_{\widehat{S}\backslash\{j\}}(\mathbf{x}_j - \mathbf{X}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}). \end{cases}$$

Therefore, $\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}$ and $\widehat{\boldsymbol{\omega}}_j$ in (8) are the optimal solutions of the same optimization problem. This implies that $\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}$ and $\widehat{\boldsymbol{\omega}}_j$ are equivalent. Moreover, in view of (S2.2), we also have

$$\mathbf{x}_j - \mathbf{X}_{-j}\widehat{\boldsymbol{\pi}}^{(j)}_{-j} = (\mathbf{I}_n - \mathbf{P}_{\widehat{S}\backslash\{j\}})(\mathbf{x}_j - \mathbf{X}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}) = \boldsymbol{\psi}^{(j)}_j - \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}},$$

which entails that $\mathbf{x}_j - \mathbf{X}_{-j}\widehat{\boldsymbol{\pi}}^{(j)}_{-j}$ is equivalent to the hybrid orthogonalization vector $\mathbf{z}_j = \boldsymbol{\psi}^{(j)}_j - \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\omega}}_j$ defined in (8) due to the equivalence of $\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}$ and $\widehat{\boldsymbol{\omega}}_j$.

Second, although the penalized regression problem of $\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}$ is not strictly convex and may not have a unique minimizer, we continue to show that different minimizers would give the same value of $\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}$. It implies that the value of the hybrid orthogonalization vector $\mathbf{z}_j$ is unique.

In what follows, we will prove this by contradiction.

Suppose that there are two solutions $\widehat{\boldsymbol{\pi}}^{(1j)}_{\widehat{S}^c\backslash\{j\}}$ and $\widehat{\boldsymbol{\pi}}^{(2j)}_{\widehat{S}^c\backslash\{j\}}$ such that $\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(1j)}_{\widehat{S}^c\backslash\{j\}} \neq \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(2j)}_{\widehat{S}^c\backslash\{j\}}$. Denote by t^* the minimum value of the optimization problem attained at either $\widehat{\boldsymbol{\pi}}^{(1j)}_{\widehat{S}^c\backslash\{j\}}$ or $\widehat{\boldsymbol{\pi}}^{(2j)}_{\widehat{S}^c\backslash\{j\}}$. Then for any $0 < \alpha < 1$, the optimization problem attains a value of t at $\alpha\widehat{\boldsymbol{\pi}}^{(1j)}_{\widehat{S}^c\backslash\{j\}} + (1-\alpha)\widehat{\boldsymbol{\pi}}^{(2j)}_{\widehat{S}^c\backslash\{j\}}$ as follows:

$$\begin{aligned} t = &\frac{\big\|\boldsymbol{\psi}^{(j)}_j - \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}(\alpha\widehat{\boldsymbol{\pi}}^{(1j)}_{\widehat{S}^c\backslash\{j\}} + (1-\alpha)\widehat{\boldsymbol{\pi}}^{(2j)}_{\widehat{S}^c\backslash\{j\}})\big\|_2^2}{2n} \\ &+ \lambda_j \sum_{k\in\widehat{S}^c\backslash\{j\}} v_{jk}|\alpha\widehat{\pi}^{(1j)}_k + (1-\alpha)\widehat{\pi}^{(2j)}_k|, \end{aligned}$$

where $\widehat{\pi}^{(1j)}_k$ and $\widehat{\pi}^{(2j)}_k$ denote the kth elements of $\widehat{\boldsymbol{\pi}}^{(1j)}$ and $\widehat{\boldsymbol{\pi}}^{(2j)}$, respectively.

On the one hand, due to the strict convexity of the function $f(\mathbf{x}) = \|\mathbf{y}-\mathbf{x}\|_2^2$, we get

$$\begin{aligned} &\big\|\boldsymbol{\psi}^{(j)}_j - \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}(\alpha\widehat{\boldsymbol{\pi}}^{(1j)}_{\widehat{S}^c\backslash\{j\}} + (1-\alpha)\widehat{\boldsymbol{\pi}}^{(2j)}_{\widehat{S}^c\backslash\{j\}})\big\|_2^2 \\ &< \alpha\big\|\boldsymbol{\psi}^{(j)}_j - \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(1j)}_{\widehat{S}^c\backslash\{j\}}\big\|_2^2 + (1-\alpha)\big\|\boldsymbol{\psi}^{(j)}_j - \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(2j)}_{\widehat{S}^c\backslash\{j\}}\big\|_2^2. \end{aligned}$$

On the other hand, it follows from the convexity of the function $f(x) = |x|$ that

$$|\alpha\widehat{\pi}^{(1j)}_k + (1-\alpha)\widehat{\pi}^{(2j)}_k| \leq \alpha|\widehat{\pi}^{(1j)}_k| + (1-\alpha)|\widehat{\pi}^{(2j)}_k|.$$

Thus, combining these two inequalities gives

$$t < \alpha t^* + (1-\alpha)t^* = t^*,$$

which means that $\alpha\widehat{\boldsymbol{\pi}}^{(1j)}_{\widehat{S}^c\backslash\{j\}} + (1-\alpha)\widehat{\boldsymbol{\pi}}^{(2j)}_{\widehat{S}^c\backslash\{j\}}$ attains a smaller value than t^*.

This contradicts the definition of t^*, which completes the proof.

S2.3 Proof of Theorem 1

We first prove the first part of this theorem. With some simple algebra, we have

$$|\Delta_j| = \Big| \sum_{k\in\widehat{S}^c\setminus\{j\}} \frac{\mathbf{z}_j^\top \mathbf{x}_k \beta_k}{\mathbf{z}_j^\top \mathbf{x}_j} \Big| \le \left(\max_{k\in\widehat{S}^c\setminus\{j\}} \Big| \frac{\mathbf{z}_j^\top \mathbf{x}_k}{\mathbf{z}_j^\top \mathbf{x}_j} \Big| \right) \sum_{k\in\widehat{S}^c} |\beta_k|.$$

In addition, by the KKT condition and the equality (9), for any $k \in \widehat{S}^c\setminus\{j\}$, we can get

$$|\mathbf{z}_j^\top \mathbf{x}_k| = |\mathbf{z}_j^\top \boldsymbol{\psi}_k^{(j)}| = |(\boldsymbol{\psi}_j^{(j)} - \boldsymbol{\psi}_{\widehat{S}^c\setminus\{j\}} \widehat{\boldsymbol{\omega}}_j)^\top \boldsymbol{\psi}_k^{(j)}| \le \sqrt{n}\lambda_j \|\boldsymbol{\psi}_k^{(j)}\|_2,$$

which along with $\widehat{S}^c \subset S_1^c$ yields

$$|\Delta_j| \le \left(\max_{k\in\widehat{S}^c\setminus\{j\}} \Big| \frac{\sqrt{n}\lambda_j \|\boldsymbol{\psi}_k^{(j)}\|_2}{\mathbf{z}_j^\top \mathbf{x}_j} \Big| \right) \sum_{k\in S_1^c} |\beta_k|.$$

We proceed to prove the second part of this theorem. Under the Lyapunov condition (13), the Lyapunov central limit theorem gives

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{\mathbf{z}_j^\top \boldsymbol{\varepsilon}}{\sigma \|\mathbf{z}_j\|_2} \le x\Big) = \Phi(x),$$

where x is any real number. In view of $W_j = \frac{\mathbf{z}_j^\top \boldsymbol{\varepsilon}}{\mathbf{z}_j^\top \mathbf{x}_j}$, the above equality can be rewritten as

$$\lim_{n\to\infty} \mathbb{P}(W_j \le \tau_j \sigma x) = \Phi(x),$$

which completes the proof of this theorem.

S2.4 Proof of Corollary 1

The proof of this corollary is the same as that of Theorem 1, and, therefore, has been omitted.

S2.5 Proof of Proposition 2

Since the design matrix $\mathbf{X}$ here is random, rather than fixed as in Zhang and Zhang (2014), the statistics related to $\mathbf{X}$ are also random variables. We will first analyze the properties of some key statistics before proving Proposition 2.

Part 1: Deviation bounds of $\|\mathbf{x}_k\|_2$. Under Condition 1, $\|\mathbf{x}_k\|_2^2/\sigma_{kk} \sim \chi^2_{(n)}$ for any $1 \leq k \leq p$, where σ_{kk} denotes the (k,k)th entry of $\boldsymbol{\Sigma}$. We use the following tail bound for the chi-squared distribution with n degrees of freedom (Ren et al., 2015, inequality (93)):

$$\mathbb{P}\Big\{\Big|\frac{\chi^2_{(n)}}{n} - 1\Big| \geq t\Big\} \leq 2\exp\big(-nt(t \wedge 1)/8\big). \tag{S2.4}$$

Applying (S2.4) with $t = 4\sqrt{\delta\log(p)/n}$ shows that

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

holds with probability at least $1 - 2p^{-2\delta}$. Because Condition 1 states that the eigenvalues of $\boldsymbol{\Phi}$ are within the interval $[1/L, L]$, Lemma 2 gives $1/L \leq \sigma_{kk} \leq L$ for any $1 \leq k \leq p$. In view of $\log p = o(n)$ by Condition 2, it follows that for sufficiently large n, with probability at least $1-2p^{-2\delta}$,

$$\begin{aligned}\widetilde{M}_1^* &\leq \sqrt{[1-4\sqrt{\delta\log(p)/n}]/L} \leq \|\mathbf{x}_k\|_2/\sqrt{n} \\ &\leq \sqrt{[1+4\sqrt{\delta\log(p)/n}]L} \leq \widetilde{M}_1,\end{aligned} \tag{S2.5}$$

where $\widetilde{M}_1$ and $\widetilde{M}_1^*$ are positive constants satisfying $\widetilde{M}_1 \geq \widetilde{M}_1^*$.

Therefore, we have

$$\mathbb{P}(\max_{k\neq j}\|\mathbf{x}_k\|_2/\sqrt{n} > \widetilde{M}_1) \leq \sum_{k\neq j}\mathbb{P}(\|\mathbf{x}_k\|_2/\sqrt{n} > \widetilde{M}_1) \leq p\cdot 2p^{-2\delta} = 2p^{1-2\delta},$$

as well as

$$\mathbb{P}(\min_{k\neq j}\|\mathbf{x}_k\|_2/\sqrt{n} < \widetilde{M}_1^*) \leq \sum_{k\neq j}\mathbb{P}(\|\mathbf{x}_k\|_2/\sqrt{n} < \widetilde{M}_1^*) \leq p\cdot 2p^{-2\delta} = 2p^{1-2\delta}.$$

Combining the above two inequalities further yields

$$\mathbb{P}\left(\bigcap_{k\neq j}\{\widetilde{M}_1^* \leq \|\mathbf{x}_k\|_2/\sqrt{n} \leq \widetilde{M}_1\}\right) \geq 1-2p^{1-2\delta}-2p^{1-2\delta} = 1-o(p^{1-\delta}). \tag{S2.6}$$

Part 2: Deviation bounds of $\|\boldsymbol{\psi}_k^{(j)}\|_2$ *for* $k\in\widehat{S}^c$. Suppose that k is the k'th element of the set $\widehat{S}^{(k)} = \widehat{S}\cup\{k\}$. On the one hand, in view of $\mathbf{P}_{\widehat{S}} = \mathbf{X}_{\widehat{S}}(\mathbf{X}_{\widehat{S}}^\top\mathbf{X}_{\widehat{S}})^{-1}\mathbf{X}_{\widehat{S}}^\top$, we can rewrite $\boldsymbol{\psi}_k^{(j)}$ as

$$\boldsymbol{\psi}_k^{(j)} = (\mathbf{I}_n-\mathbf{P}_{\widehat{S}})\mathbf{x}_k = \mathbf{x}_k - \mathbf{X}_{\widehat{S}}(\mathbf{X}_{\widehat{S}}^\top\mathbf{X}_{\widehat{S}})^{-1}\mathbf{X}_{\widehat{S}}^\top\mathbf{x}_k \triangleq \mathbf{x}_k - \mathbf{X}_{\widehat{S}}\widehat{\boldsymbol{\gamma}}_k^{(j)}, \tag{S2.7}$$

where $\widehat{\boldsymbol{\gamma}}_k^{(j)} = (\mathbf{X}_{\widehat{S}}^\top\mathbf{X}_{\widehat{S}})^{-1}\mathbf{X}_{\widehat{S}}^\top\mathbf{x}_k$ is the least squares estimate of the coefficient vector obtained by regressing $\mathbf{x}_k$ on $\mathbf{X}_{\widehat{S}}$.

On the other hand, since the rows of $\mathbf{X}_{\widehat{S}^{(k)}}$ are independent and follow the distribution $N(0,(\boldsymbol{\Theta}^{(k)})^{-1})$ for $\boldsymbol{\Theta}^{(k)} = (\boldsymbol{\Sigma}_{\widehat{S}^{(k)},\widehat{S}^{(k)}})^{-1} = (\theta_{ij}^{(k)})_{(d+1)\times(d+1)}$, the conditional distribution of $\mathbf{x}_k$ given $\mathbf{X}_{\widehat{S}}$ is Gaussian:

$$\mathbf{x}_k|\mathbf{X}_{\widehat{S}} \sim N(\mathbf{X}_{\widehat{S}}\boldsymbol{\gamma}_k^{(j)}, (1/\theta_{k'k'}^{(k)})\mathbf{I}_n),$$

where $\boldsymbol{\gamma}_k^{(j)} = -(\boldsymbol{\Theta}_{k',-k'}^{(k)})^\top/\theta_{k'k'}^{(k)}$. Here the d-dimensional coefficient vector $\boldsymbol{\Theta}_{k',-k'}^{(k)}$ is the k'th row of $\boldsymbol{\Theta}^{(k)}$ with the k'th component removed. Based on this conditional distribution, we know that the random error vector

$$\boldsymbol{\rho}_k^{(j)} = \mathbf{x}_k - \mathbf{X}_{\widehat{S}}\boldsymbol{\gamma}_k^{(j)} \tag{S2.8}$$

is independent of $\mathbf{X}_{\widehat{S}}$ and follows the distribution $N(0, (1/\theta_{k'k'}^{(k)})\mathbf{I}_n)$. It is worth mentioning that $\theta_{k'k'}^{(k)} \in [1/L, L]$ can be easily deduced from Condition 1 and Lemma 2.

We now derive a bound for the deviation of $\|\boldsymbol{\psi}_k^{(j)}\|_2$ from its population counterpart $\|\boldsymbol{\rho}_k^{(j)}\|_2$. For notational simplicity, let

$$a_{d,p,\delta} = \max\{d - 2\sqrt{2\delta d \log p}, 0\} \quad \text{and} \quad b_{d,p,\delta} = d + 2\sqrt{2\delta d\log p} + 4\delta \log p.$$

Combining (S2.7) and (S2.8) and applying Lemma 3 yields, with probability at least $1 - 2p^{-2\delta}$,

$$\left(\frac{a_{d,p,\delta}}{\theta_{k'k'}^{(k)}}\right)^{1/2} \leq \|\mathbf{X}_{\widehat{S}}(\widehat{\boldsymbol{\gamma}}_k^{(j)} - \boldsymbol{\gamma}_k^{(j)})\|_2 \leq \left(\frac{b_{d,p,\delta}}{\theta_{k'k'}^{(k)}}\right)^{1/2}.$$

Moreover, in view of (S2.7) and (S2.8), we also have

$$\boldsymbol{\rho}_k^{(j)} - \boldsymbol{\psi}_k^{(j)} = \mathbf{X}_{\widehat{S}}(\widehat{\boldsymbol{\gamma}}_k^{(j)} - \boldsymbol{\gamma}_k^{(j)}),$$

which together with the above inequality gives, with probability at least

$1-2p^{-2\delta}$,

$$\left(\frac{a_{d,p,\delta}}{\theta^{(k)}_{k'k'}}\right)^{1/2} \leq \|\boldsymbol{\rho}_k^{(j)} - \boldsymbol{\psi}_k^{(j)}\|_2 \leq \left(\frac{b_{d,p,\delta}}{\theta^{(k)}_{k'k'}}\right)^{1/2}. \tag{S2.9}$$

Since $\left|\|\boldsymbol{\psi}_k^{(j)}\|_2 - \|\boldsymbol{\rho}_k^{(j)}\|_2\right| \leq \|\boldsymbol{\rho}_k^{(j)} - \boldsymbol{\psi}_k^{(j)}\|_2$, the above inequality entails

$$\mathbb{P}\left\{\left|\|\boldsymbol{\psi}_k^{(j)}\|_2 - \|\boldsymbol{\rho}_k^{(j)}\|_2\right| \geq \left(\frac{b_{d,p,\delta}}{\theta^{(k)}_{k'k'}}\right)^{1/2}\right\} \leq 2p^{-2\delta}. \tag{S2.10}$$

In order to proceed, we will then construct the deviation bounds of $\|\boldsymbol{\rho}_k^{(j)}\|_2$. Note that $\|\boldsymbol{\rho}_k^{(j)}\|_2^2/(1/\theta^{(k)}_{k'k'}) \sim \chi^2_{(n)}$. Applying (S2.4) with $t = 4\sqrt{\delta\log(p)/n}$ shows that

$$\left(\frac{n(1-4\sqrt{\delta\log(p)/n})}{\theta^{(k)}_{k'k'}}\right)^{1/2} \leq \|\boldsymbol{\rho}_k^{(j)}\|_2 \leq \left(\frac{n(1+4\sqrt{\delta\log(p)/n})}{\theta^{(k)}_{k'k'}}\right)^{1/2}$$

holds with probability at least $1-2p^{-2\delta}$. This inequality together with (S2.10) entails that with probability at least $1-4p^{-2\delta}$,

$$\frac{(n(1-4\sqrt{\delta\log(p)/n}))^{1/2} - b_{d,p,\delta}^{1/2}}{(\theta^{(k)}_{k'k'})^{1/2}} \leq \|\boldsymbol{\psi}_k^{(j)}\|_2 \leq \frac{(n(1+4\sqrt{\delta\log(p)/n}))^{1/2} + b_{d,p,\delta}^{1/2}}{(\theta^{(k)}_{k'k'})^{1/2}}.$$

Since $d = o(n/s_\Phi)$ and $\log p = o(n)$, we have $b_{d,p,\delta} = O(d+\log p) = o(n)$. It follows that for sufficiently large n, with probability at least $1-4p^{-2\delta}$,

$$\widetilde{M}_2^* \leq \|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n} \leq \widetilde{M}_2, \tag{S2.11}$$

where $\widetilde{M}_2$ and $\widetilde{M}_2^*$ are positive constants satisfying $\widetilde{M}_2 \geq \widetilde{M}_2^*$. Thus, we have

$$\mathbb{P}(\max_{k\in\widehat{S}^c\setminus\{j\}} \|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n} > \widetilde{M}_2) \leq \sum_{k\neq j}\mathbb{P}(\|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n} > \widetilde{M}_2) \leq p\cdot 4p^{-2\delta} = 4p^{1-2\delta}$$

and

$$\mathbb{P}(\min_{k\in\widehat{S}^c\setminus\{j\}} \|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n} < \widetilde{M}_2^*) \leq \sum_{k\neq j}\mathbb{P}(\|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n} < \widetilde{M}_2^*) \leq p\cdot 4p^{-2\delta} = 4p^{1-2\delta},$$

respectively. Combining the above two inequalities further yields

$$\mathbb{P}\left(\bigcap_{k\in\widehat{S}^c\setminus\{j\}}\{\widetilde{M}_2^*\leq\|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n}\leq\widetilde{M}_2\}\right)\geq 1-4p^{1-2\delta}-4p^{1-2\delta}=1-o(p^{1-\delta}).\tag{S2.12}$$

We proceed to prove Proposition 2. Since the following argument applies to any j, $1\leq j\leq p$, we drop the superscript (j) in $F_2^{(j)}(\xi,S_{j*})$ and $\mathcal{G}_-^{(j)}(\xi,S_{j*})$ for notational simplicity. Recall that the sign-restricted cone invertibility factor here is defined as

$$F_2(\xi,S_{j*})=\inf\left\{\frac{|S_{j*}|^{1/2}\|(\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}})^\top\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}\mathbf{u}/n\|_\infty}{\|\mathbf{u}\|_2}:\mathbf{u}\in\mathcal{G}_-(\xi,S_{j*})\right\}$$

with the sign-restricted cone

$$\mathcal{G}_-(\xi,S_{j*})=\left\{\mathbf{u}:\|\mathbf{u}_{S_{j*}^c}\|_1\leq\xi\|\mathbf{u}_{S_{j*}}\|_1, u_k(\boldsymbol{\psi}_k^{(j)})^\top\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}\mathbf{u}/n\leq 0,\quad\forall\ k\in S_{j*}^c\right\}.$$

In what follows, we mainly present the proof when $j\in\widehat{S}^c$ and the same argument applies to $j\in\widehat{S}$.

First, since $b_{d,p,\delta}=O(d+\log p)$ and $\theta^{(k)}_{k'k'}\in[1/L,L]$, inequality (S2.9) implies that there is a positive constant C such that, with probability at least $1-2p^{-2\delta}$,

$$\|\boldsymbol{\rho}_k^{(j)}-\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{d+\log p}\leq C.$$

Thus, we further have

$$\begin{aligned}&\mathbb{P}(\max_{k\in\widehat{S}^c\setminus\{j\}}\|\boldsymbol{\rho}_k^{(j)}-\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{d+\log p}>C)\\ \leq&\sum_{k\neq j}\mathbb{P}(\|\boldsymbol{\rho}_k^{(j)}-\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{d+\log p}>C)\leq p\cdot 2p^{-2\delta}=2p^{1-2\delta}.\end{aligned}\tag{S2.13}$$

We then prove the conclusions of this proposition. For any $\mathbf{u} \in \mathbb{R}^{p-d-1}$, we will first give the deviation of $\|\boldsymbol{\psi}_k^{(j)}\mathbf{u}\|_2$ from its population counterpart $\|\boldsymbol{\rho}_k^{(j)}\mathbf{u}\|_2$. For any matrix $\mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_p) \in \mathbb{R}^{n\times p}$ and any vector $\mathbf{t} = (t_1, \ldots, t_p)^\top \in \mathbb{R}^p$, it is clear that

$$\|\mathbf{H}\mathbf{t}\|_2 = \|t_1\mathbf{h}_1 + \ldots + t_p\mathbf{h}_p\|_2 \le |t_1|\|\mathbf{h}_1\|_2 + \ldots + |t_p|\|\mathbf{h}_p\|_2 \le (\max_{1\le i\le p}\|\mathbf{h}_i\|_2)\|\mathbf{t}\|_1.$$

Based on this inequality, we have

$$\|\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\backslash\{j\}}\mathbf{u}\|_2 - \|\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\mathbf{u}\|_2 \le \|(\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\backslash\{j\}} - \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}})\mathbf{u}\|_2 \le \max_{k\in\widehat{S}^c\backslash\{j\}}\|\boldsymbol{\rho}_k^{(j)} - \boldsymbol{\psi}_k^{(j)}\|_2\|\mathbf{u}\|_1.$$

Therefore, for any $\mathbf{u} \in \mathcal{G}(\xi, S_{j*})$ with $\mathcal{G}(\xi, S_{j*}) = \{\mathbf{u} \in \mathbb{R}^{p-d-1} : \|\mathbf{u}_{S_{j*}^c}\|_1 \le \xi\|\mathbf{u}_{S_{j*}}\|_1\}$, in view of (S2.13), $|S_{j*}| \le s_\Phi$, and $\|\mathbf{u}_{S_{j*}}\|_1 \le s_\Phi^{1/2}\|\mathbf{u}_{S_{j*}}\|_2$, with probability at least $1 - o(p^{1-\delta})$,

$$\begin{aligned}
&\|\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\mathbf{u}\|_2 \ge \|\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\backslash\{j\}}\mathbf{u}\|_2 - C(d+\log p)^{1/2}\|\mathbf{u}\|_1\\
&\ge \|\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\backslash\{j\}}\mathbf{u}\|_2 - C(d+\log p)^{1/2}(1+\xi)\|\mathbf{u}_{S_{j*}}\|_1\\
&\ge \|\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\backslash\{j\}}\mathbf{u}\|_2 - C(1+\xi)\{s_\Phi(d+\log p)\}^{1/2}\|\mathbf{u}_{S_{j*}}\|_2. \qquad \text{(S2.14)}
\end{aligned}$$

Now we turn our attention to $\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\backslash\{j\}}$. For notational simplicity, let $B_j = \widehat{S}^c\backslash\{j\}$. By Lemma 1, we have

$$\boldsymbol{\rho}^{(j)}_{B_j} \sim N(0, \mathbf{I}_n \otimes \mathbf{V}_j),$$

where

$$\mathbf{V}_j = \boldsymbol{\Sigma}_{B_jB_j} - \boldsymbol{\Sigma}_{B_j\widehat{S}}\boldsymbol{\Sigma}^{-1}_{\widehat{S}\widehat{S}}\boldsymbol{\Sigma}_{\widehat{S}B_j} = \left[(\boldsymbol{\Phi}_{\widehat{S}^c\widehat{S}^c})^{-1}\right]_{B_jB_j}.$$

Since the eigenvalues of $\boldsymbol{\Phi}$ are bounded within the interval $[1/L, L]$ for some constant $L \ge 1$, the eigenvalues of $(\boldsymbol{\Phi}_{\widehat{S}^c\widehat{S}^c})^{-1}$ are also bounded within this interval by Lemma 2. As $\mathbf{V}_j$ is a principal submatrix of $(\mathbf{\Phi}_{\widehat{S}^c\widehat{S}^c})^{-1}$, its eigenvalues are bounded within $[1/L, L]$ as well. Thus, for any $\mathbf{u} \in \mathcal{G}(\xi, S_{j*})$, applying the same argument as in (S3.68) in Lemma 4 shows that

$$\frac{\|\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\setminus\{j\}}\mathbf{u}\|_2}{n^{1/2}} \geq \left\{\frac{1}{4L^{1/2}} - 9(1+\xi)L^{1/2}\sqrt{\frac{s_\Phi \log(p)}{n}}\right\}\|\mathbf{u}\|_2$$

holds with probability at least $1 - c_2^* \exp(-c_3^* n)$, where c_2^* and c_3^* are some positive constants.

Combining this result with (S2.14), we find that for any $\mathbf{u} \in \mathcal{G}(\xi, S_{j*})$,

$$\begin{aligned}
&\frac{\|\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}\mathbf{u}\|_2}{n^{1/2}} \geq \frac{\|\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\setminus\{j\}}\mathbf{u}\|_2}{n^{1/2}} - C(1+\xi)\left\{\frac{s_\Phi(d+\log p)}{n}\right\}^{1/2}\|\mathbf{u}_{S_{j*}}\|_2 \\
&\geq \left[\frac{1}{4L^{1/2}} - 9(1+\xi)L^{1/2}\sqrt{\frac{s_\Phi \log(p)}{n}} - C(1+\xi)\left\{\frac{s_\Phi(d+\log p)}{n}\right\}^{1/2}\right]\|\mathbf{u}\|_2
\end{aligned} \tag{S2.15}$$

This bound holds with probability at least $1-c_2^* \exp(-c_3^* n)-o(p^{1-\delta})$. Moreover, since $\log(p)/n = o(1)$, we have

$$\frac{\exp(-c_3^* n)}{p^{1-\delta}} = \frac{p^{\delta-1}}{\exp(c_3^* n)} = \exp\{(\delta-1)\log p - c_3^* n\} = o(1),$$

which implies that $c_2^* \exp(-c_3^* n) = o(p^{1-\delta})$.

Therefore, since $s_\Phi(\log p)/n = o(1)$ and $d = o(n/s_\Phi)$ together imply $s_\Phi(d + \log p)/n = o(1)$, we conclude from (S2.15) that, for some constant $c_0^* > 0$,

$$RE_2(\xi, S_{j*}) \geq (c_0^*)^{1/2} \tag{S2.16}$$

holds with probability at least $1 - o(p^{1-\delta})$, where $RE_2(\xi, S_{j*})$ is defined as

$$RE_2(\xi, S_{j*}) = \inf\left\{\frac{\|\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}\mathbf{u}\|_2}{n^{1/2}\|\mathbf{u}\|_2} : \mathbf{u} \in \mathcal{G}(\xi, S_{j*})\right\}.$$

Note that the cone $\mathcal{G}(\xi, S_{j*}) = \{\mathbf{u} \in \mathbb{R}^{p-d-1} : \|\mathbf{u}_{S^c_{j*}}\|_1 \leq \xi\|\mathbf{u}_{S_{j*}}\|_1\}$.

Finally, by (Ye and Zhang, 2010, Proposition 5) and an argument similar to the proof of Lemma 4, we obtain

$$F_2(\xi, S_{j*}) \geq F_2^*(\xi, S_{j*}) \geq \frac{RE_2^2(\xi, S_{j*})}{1+\xi} \geq \frac{c_0^*}{1+\xi} =: c^*$$

with probability at least $1 - o(p^{1-\delta})$, where the cone invertibility factor $F_2^*(\xi, S_{j*})$ is defined as

$$F_2^*(\xi, S_{j*}) = \inf\left\{\frac{|S_{j*}|^{1/2}\|(\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}})^\top\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}\mathbf{u}/n\|_\infty}{\|\mathbf{u}\|_2} : \mathbf{u} \in \mathcal{G}(\xi, S_{j*})\right\}.$$

This completes the proof of Proposition 2.

S2.6 Proof of Theorem 2

Under the conditions of Theorem 2, it is clear that the results in *Part 1* and *Part 2* of the proof of Proposition 2 are still valid. In what follows, we mainly present the proof when $j \in \widehat{S}^c$ and the same argument applies to $j \in \widehat{S}$ for a given set $\widehat{S}$. Throughout the proof, $\lambda_j = (1+\epsilon)\sqrt{(2\delta/(n\phi_{jj}))\log(p)}$, where $\epsilon = \widetilde{M}(\xi'+1)/\{\widetilde{M}^*(\xi'-1)\} - 1 > 0$, and $\widetilde{M}$ and $\widetilde{M}^*$ are positive constants satisfying $\widetilde{M} \geq \widetilde{M}^*$. Here ϕ_{jj} is defined in Condition 1. In addition, C is some positive constant which can vary from expression to expression.

Part 1: Deviation bounds of $\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_2$. Let $\widetilde{M} = \max(\widetilde{M}_1, \widetilde{M}_2)$ and $\widetilde{M}^* = \min(\widetilde{M}_1^*, \widetilde{M}_2^*)$, respectively. Recall that we have the conditional distribution

$$\mathbf{x}_j = \mathbf{X}_{\widehat{S}}\boldsymbol{\pi}_{\widehat{S}}^{(j)} + \mathbf{X}_{\widehat{S}^c\backslash\{j\}}\boldsymbol{\pi}_{\widehat{S}^c\backslash\{j\}}^{(j)} + \mathbf{e}_j,$$

where $\mathbf{e}_j = \mathbf{x}_j - \mathbb{E}(\mathbf{x}_j|\mathbf{X}_{-j})$ is independent of $\mathbf{X}_{-j}$ and follows the distribution $N(\mathbf{0}, (1/\phi_{jj})\mathbf{I}_n)$. Thus, it follows that

$$(\mathbf{I}_n - \mathbf{P}_{\widehat{S}})\mathbf{x}_j = (\mathbf{I}_n - \mathbf{P}_{\widehat{S}})(\mathbf{X}_{\widehat{S}}\boldsymbol{\pi}_{\widehat{S}}^{(j)} + \mathbf{X}_{\widehat{S}^c\backslash\{j\}}\boldsymbol{\pi}_{\widehat{S}^c\backslash\{j\}}^{(j)} + \mathbf{e}_j),$$

which can be simplified to

$$\boldsymbol{\psi}_j^{(j)} = \boldsymbol{\psi}_{\widehat{S}^c\backslash\{j\}}^{(j)}\boldsymbol{\omega}_j + \mathbf{e}_j^* \tag{S2.17}$$

according to the definition of $\boldsymbol{\psi}_k^{(j)}$. Here $\boldsymbol{\omega}_j = \boldsymbol{\pi}_{\widehat{S}^c\backslash\{j\}}^{(j)} = -\boldsymbol{\Phi}_{j,\widehat{S}^c\backslash\{j\}}^\top/\phi_{jj}$ and $\mathbf{e}_j^* = (\mathbf{I}_n - \mathbf{P}_{\widehat{S}})\mathbf{e}_j$.

Denote by $S_{j*} = \{1 \leq k \leq p : k \in \widehat{S}^c\backslash\{j\}, \phi_{jk} \neq 0\}$, $S_{j*}^c = \widehat{S}^c\backslash(\{j\} \cup S_{j*})$, and $z_\infty^{j*} = \|(\boldsymbol{\psi}_{\widehat{S}^c\backslash\{j\}}^{(j)})^\top(\boldsymbol{\psi}_j^{(j)} - \boldsymbol{\psi}_{\widehat{S}^c\backslash\{j\}}^{(j)}\boldsymbol{\omega}_j)/n\|_\infty$. In what follows, we first use arguments similar to those in the proof of Theorem 3 to construct bounds for $\|\boldsymbol{\psi}_{\widehat{S}^c\backslash\{j\}}^{(j)}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)\|_2$; the main difficulty here is that the predictor matrix $\boldsymbol{\psi}_{\widehat{S}^c\backslash\{j\}}^{(j)}$ is much more complex than a usual design matrix with i.i.d. rows. For the remainder of the proof, let $\mathcal{E}'' = \bigcap_{k\in\widehat{S}^c\backslash\{j\}}\{\widetilde{M}^* \leq v_{jk} \leq \widetilde{M}\}$, where $v_{jk} = \|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n}$. First of all, the same argument as in (S2.30) gives

$$\mathbf{v}^\top(\boldsymbol{\psi}_{\widehat{S}^c\backslash\{j\}}^{(j)})^\top\boldsymbol{\psi}_{\widehat{S}^c\backslash\{j\}}^{(j)}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)/n \leq (z_\infty^{j*} + \widetilde{M}\lambda_j)\|\mathbf{v}_{S_{j*}}\|_1 + (z_\infty^{j*} - \widetilde{M}^*\lambda_j)\|\mathbf{v}_{S_{j*}^c}\|_1$$

for all vectors $\mathbf{v} \in \mathbb{R}^{p-d-1}$ satisfying $\operatorname{sign}(\mathbf{v}_{S_{j*}^c}) = \operatorname{sign}((\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)_{S_{j*}^c}) =$

$\mathrm{sign}((\widehat{\boldsymbol{\omega}}_j)_{S_{j*}^c})$. Based on the above inequality, if $z_\infty^{j*} < \widetilde{M}^*\lambda_j$, it follows from $\mathbf{v} = \widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j$ that

$$(z_\infty^{j*} + \widetilde{M}\lambda_j)\|(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)_{S_{j*}}\|_1 + (z_\infty^{j*} - \widetilde{M}^*\lambda_j)\|(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)_{S_{j*}^c}\|_1 \geq 0,$$

which entails that

$$\|(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)_{S_{j*}^c}\|_1 \leq \xi\|(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)_{S_{j*}}\|_1 \tag{S2.18}$$

for $\xi \geq (\widetilde{M}\lambda_j + z_\infty^{j*})/(\widetilde{M}^*\lambda_j - z_\infty^{j*})$.

In addition, if $z_\infty^{j*} \leq \widetilde{M}^*\lambda_j$, the same argument as in (S2.33) gives

$$(\widehat{\omega}_{j,k} - \omega_{j,k})(\boldsymbol{\psi}_k^{(j)})^\top \boldsymbol{\psi}_{\widehat{S}^c\setminus\{j\}}^{(j)}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)/n \leq 0 \tag{S2.19}$$

for any $k \in S_{j*}^c$, where $\widehat{\omega}_{j,k}$ is the kth component of $\widehat{\boldsymbol{\omega}}_j$. In view of (S2.18) and (S2.19), if $z_\infty^{j*} < \widetilde{M}^*\lambda_j$, we can get

$$\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j \in \mathcal{G}_-(\xi, S_{j*}) \tag{S2.20}$$

for $\xi \geq (\widetilde{M}\lambda_j + z_\infty^{j*})/(\widetilde{M}^*\lambda_j - z_\infty^{j*})$. Here the sign-restricted cone $\mathcal{G}_-(\xi, S_{j*})$ is defined as

$$\mathcal{G}_-(\xi, S_{j*}) = \left\{\mathbf{u} \in \mathbb{R}^{p-d-1} : \|\mathbf{u}_{S_{j*}^c}\|_1 \leq \xi\|\mathbf{u}_{S_{j*}}\|_1, u_k(\boldsymbol{\psi}_k^{(j)})^\top \boldsymbol{\psi}_{\widehat{S}^c\setminus\{j\}}^{(j)}\mathbf{u}/n \leq 0, \forall k \in S_{j*}^c\right\}.$$

In order to proceed, our analysis will then be conditional on the event $\mathcal{E}'''^* = \{z_\infty^{j*} \leq \lambda_j\widetilde{M}^*(\xi' - 1)/(\xi' + 1)\}$ for some constant $\xi' > 1$. Note that $z_\infty^{j*} \leq \lambda_j\widetilde{M}^*(\xi' - 1)/(\xi' + 1) < \widetilde{M}^*\lambda_j$ holds on this event, which together with (S2.20) implies that

$$\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j \in \mathcal{G}_-(\xi, S_{j*})$$

for some constant $\xi \geq (\widetilde{M}\lambda_j + z_\infty^{j*})/(\widetilde{M}^*\lambda_j - z_\infty^{j*})$. Similarly, we first clarify that the magnitude of $(\widetilde{M}\lambda_j + z_\infty^{j*})/(\widetilde{M}^*\lambda_j - z_\infty^{j*})$ is around a constant level. The same argument as in (S2.35) gives

$$1 \leq \frac{\widetilde{M}\lambda_j + z_\infty^{j*}}{\widetilde{M}^*\lambda_j - z_\infty^{j*}} \leq 1 + \frac{\xi' - 1}{2} + \frac{(\xi'+1)\widetilde{M} - 2\widetilde{M}^*}{2\widetilde{M}^*}.$$

Moreover, since $z_\infty^{j*} \leq \lambda_j \widetilde{M}^*(\xi'-1)/(\xi'+1) \leq \lambda_j \widetilde{M}(\xi'-1)/(\xi'+1)$, it follows that

$$\|(\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}})^\top \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)/n\|_\infty \leq \widetilde{M}\lambda_j + z_\infty^{j*} \leq 2\widetilde{M}\lambda_j\xi'/(\xi'+1).$$

Thus, in view of $|S_{j*}| \leq s_\Phi$, combining the above results gives

$$\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_2 \leq \frac{|S_{j*}|^{1/2}(\widetilde{M}\lambda_j + z_\infty^{j*})}{F_2(\xi, S_{j*})} \leq \frac{\{2\widetilde{M}\xi'/(\xi'+1)\}s_\Phi^{1/2}\lambda_j}{F_2(\xi, S_{j*})}, \tag{S2.21}$$

where the sign-restricted cone invertibility factor $F_2(\xi, S_{j*})$ is defined as

$$F_2(\xi, S_{j*}) = \inf\left\{\frac{|S_{j*}|^{1/2}\|(\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}})^\top \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}\mathbf{u}/n\|_\infty}{\|\mathbf{u}\|_2} : \mathbf{u} \in \mathcal{G}_-(\xi, S_{j*})\right\}.$$

We then derive the probability of the event $\mathcal{E}'''^{*}$. Since $\mathbf{e}_j$ is independent of $\mathbf{X}_{\widehat{S}}$ and $\mathbf{X}_{\widehat{S}^c\backslash\{j\}}$, it is also independent of $\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}$. In addition, it follows from the definition of $\mathbf{e}_j^* = (\mathbf{I}_n - \mathbf{P}_{\widehat{S}})\mathbf{e}_j$ that

$$(\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}})^\top \mathbf{e}_j^* = (\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}})^\top \mathbf{e}_j.$$

On the event $\mathcal{E}''$, $\|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n} \leq \widetilde{M}$ holds for any $k \in \widehat{S}^c\backslash\{j\}$. Since $\mathbf{e}_j$ follows the distribution $N(0, (1/\phi_{jj})\mathbf{I}_n)$, for

$$\lambda_j = \widetilde{M}\sqrt{(2\delta/(n\phi_{jj}))\log(p)}(\xi'+1)/\{\widetilde{M}^*(\xi'-1)\},$$

the same argument as in (S2.38) gives

$$1 - \mathbb{P}\left(\mathcal{E}'''^*\right) \leq o(p^{1-\delta}). \tag{S2.22}$$

We now show the magnitude of $F_2(\xi, S_{j*})$. By Proposition 2, under the assumption that $d = o(n/s_\Phi)$, there is some positive constant c^* such that

$$F_2(\xi, S_{j*}) \geq c^*$$

holds with probability at least $1-o(p^{1-\delta})$. Moreover, with the aid of (S2.12) and (S2.22), we get

$$\begin{aligned}\mathbb{P}(\mathcal{E}'' \cap \mathcal{E}'''^*) &= 1 - \mathbb{P}\left((\mathcal{E}'')^c \cup (\mathcal{E}'''^*)^c\right)\\ &\geq 1 - \mathbb{P}((\mathcal{E}'')^c) - \mathbb{P}((\mathcal{E}'''^*)^c) = 1 - o(p^{1-\delta}).\end{aligned}$$

Thus, combining these results with (S2.21) gives

$$\mathbb{P}\left\{\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_2 \leq C s_\Phi^{1/2} \lambda_j\right\} \geq 1 - o(p^{1-\delta}). \tag{S2.23}$$

Part 2: Deviation bounds of $n^{-1/2}\|\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)\|_2$ *and* $\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_q$ *for* $q \in [1, 2)$. We first construct bounds for $\|\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)\|_2$. Some simple algebra shows that

$$\|\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)\|_2^2 \leq \|(\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}})^\top \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)\|_\infty \cdot \|(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)\|_1. \tag{S2.24}$$

Conditional on the event $\mathcal{E}'''^* = \{z_\infty^{j*} \leq \lambda_j \widetilde{M}^*(\xi' - 1)/(\xi' + 1)\}$, it follows that

$$\|(\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}})^\top \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)/n\|_\infty \leq \widetilde{M}\lambda_j + z_\infty^{j*} \leq \widetilde{M}\lambda_j + \lambda_j \widetilde{M}^*(\xi' - 1)/(\xi' + 1),$$

which further yields

$$\|(\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}})^\top \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)\|_\infty \leq n\{\widetilde{M} + \widetilde{M}^*(\xi' - 1)/(\xi' + 1)\}\lambda_j. \tag{S2.25}$$

Moreover, since $\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j \in \mathcal{G}_-(\xi, S_{j*})$, we have

$$\begin{aligned}\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_1 &= \|(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)_{S^c_{j*}}\|_1 + \|(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)_{S_{j*}}\|_1 \leq (1+\xi)\|(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)_{S_{j*}}\|_1 \\ &\leq (1+\xi)s_\Phi^{1/2}\|(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)_{S_{j*}}\|_2 \leq (1+\xi)s_\Phi^{1/2}\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_2.\end{aligned}$$

This, together with (S2.23), entails that

$$\mathbb{P}\left\{\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_1 \leq C s_\Phi \lambda_j\right\} \geq 1 - o(p^{1-\delta}). \tag{S2.26}$$

Then by inequalities (S2.24), (S2.25), and (S2.26), we finally have

$$\mathbb{P}\left\{\|\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)\|_2 \leq C(s_\Phi n)^{1/2}\lambda_j\right\} \geq 1 - o(p^{1-\delta}), \tag{S2.27}$$

which entails that

$$\mathbb{P}\left\{n^{-1/2}\|\boldsymbol{\psi}^{(j)}_{\widehat{S}^c\backslash\{j\}}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j)\|_2 \leq C s_\Phi^{1/2}\lambda_j\right\} \geq 1 - o(p^{1-\delta}).$$

We then derive the deviation bounds of $\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_q$ for $q \in (1,2)$. Note that the two inequalities $\sum_{j=1}^p a_j \leq b_1$ and $\sum_{j=1}^p a_j^2 \leq b_2$ for $a_j \geq 0$ imply

$$\sum_{j=1}^p a_j^q = \sum_{j=1}^p a_j^{2-q}a_j^{2q-2} \leq \left(\sum_{j=1}^p a_j\right)^{2-q}\left(\sum_{j=1}^p a_j^2\right)^{q-1} \leq b_1^{2-q}b_2^{q-1},$$

for any $q \in (1,2)$. Thus, we have

$$\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_q^q \leq \|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_1^{2-q} \cdot (\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_2^2)^{q-1},$$

which together with (S2.23) and (S2.26) gives

$$\mathbb{P}\left\{\|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_q \leq C s_\Phi^{1/q}\lambda_j\right\} \geq 1 - o(p^{1-\delta})$$

for any $q \in (1,2)$. This concludes the proof of Theorem 2.

S2.7 Proof of Theorem 3

Under the conditions of Theorem 3, it is clear that the results in *Part 1* and *Part 2* of the proof of Proposition 2 are still valid. In what follows, we mainly present the proof when $j \in \widehat{S}^c$ and the same argument applies to $j \in \widehat{S}$ for a given set $\widehat{S}$. Throughout the proof, the regularization parameters are $\lambda_j = (1+\epsilon)\sqrt{(2\delta/(n\phi_{jj}))\log(p)}$, where $\epsilon = \widetilde{M}(\xi'+1)/\{\widetilde{M}^*(\xi'-1)\}-1 > 0$; these are the same as those in the proof of Theorem 2. In addition, C is some positive constant which can vary from expression to expression.

Part 1: Deviation bounds of $\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_2$. In this part, our analysis will be conditional on the event $\mathcal{E}' \cap \mathcal{E}''$, where $\mathcal{E}' = \bigcap_{k\neq j}\{\widetilde{M}^* \leq \|\mathbf{x}_k\|_2/\sqrt{n} \leq \widetilde{M}\}$ and $\mathcal{E}'' = \bigcap_{k\in\widehat{S}^c\setminus\{j\}}\{\widetilde{M}^* \leq \|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n} \leq \widetilde{M}\}$, respectively. For clarity, we start with some additional analyses. Consider the conditional distribution

$$\mathbf{x}_j = \mathbf{X}_{-j}\boldsymbol{\pi}_{-j}^{(j)} + \mathbf{e}_j, \tag{S2.28}$$

where $\mathbf{e}_j = \mathbf{x}_j - \mathbb{E}(\mathbf{x}_j|\mathbf{X}_{-j})$ is independent of $\mathbf{X}_{-j}$ and follows the distribution $N(0,(1/\phi_{jj})\mathbf{I}_n)$, and the regression coefficient vector $\boldsymbol{\pi}_{-j}^{(j)} = -\boldsymbol{\Phi}_{j,-j}^\top/\phi_{jj}$. Let $S^j = \{1 \leq k \leq p : k \neq j, \phi_{jk} \neq 0\}$, $z_\infty^j = \|\mathbf{X}_{-j}^\top(\mathbf{x}_j - \mathbf{X}_{-j}\boldsymbol{\pi}_{-j}^{(j)})/n\|_\infty$, and $J_j = \widehat{S} \cup S^j$.

Since $j \in \widehat{S}^c$, we can rewrite (S2.28) as

$$\mathbf{x}_j = \mathbf{X}_{\widehat{S}}\boldsymbol{\pi}_{\widehat{S}}^{(j)} + \mathbf{X}_{\widehat{S}^c\setminus\{j\}}\boldsymbol{\pi}_{\widehat{S}^c\setminus\{j\}}^{(j)} + \mathbf{e}_j.$$

In view of Proposition 1, we have

$$(\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}}, \widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}) = \underset{\mathbf{b}_{\widehat{S}},\mathbf{b}_{\widehat{S}^c\backslash\{j\}}}{\arg\min} \Big\{(2n)^{-1}\big\|\mathbf{x}_j - \mathbf{X}_{\widehat{S}}\mathbf{b}_{\widehat{S}} - \mathbf{X}_{\widehat{S}^c\backslash\{j\}}\mathbf{b}_{\widehat{S}^c\backslash\{j\}}\big\|_2^2 + \lambda_j \sum_{k\in\widehat{S}^c\backslash\{j\}} v_{jk}|b_k|\Big\},$$

whose KKT conditions are

$$\begin{cases} -\mathbf{X}_{\widehat{S}}^\top(\mathbf{x}_j - \mathbf{X}_{\widehat{S}}\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}} - \mathbf{X}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}) = \mathbf{0}, \\ -\mathbf{X}_{\widehat{S}^c\backslash\{j\}}^\top(\mathbf{x}_j - \mathbf{X}_{\widehat{S}}\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}} - \mathbf{X}_{\widehat{S}^c\backslash\{j\}}\widehat{\boldsymbol{\pi}}^{(j)}_{\widehat{S}^c\backslash\{j\}}) = -n\lambda_j\widehat{\boldsymbol{\kappa}}^{(j)}_{\widehat{S}^c\backslash\{j\}}. \end{cases}$$

Here the kth component of $\widehat{\boldsymbol{\kappa}}^{(j)}$ satisfies

$$\widehat{\kappa}^{(j)}_k \begin{cases} = v_{jk}\cdot \mathrm{sign}(\widehat{\pi}^{(j)}_k) & \text{if } \widehat{\pi}^{(j)}_k \neq 0, k\in \widehat{S}^c\backslash\{j\}, \\ \in [-v_{jk}, v_{jk}] & \text{if } \widehat{\pi}^{(j)}_k = 0, k\in \widehat{S}^c\backslash\{j\}, \end{cases}$$

where $v_{jk} = \|\boldsymbol{\psi}^{(j)}_k\|_2/\sqrt{n}$ and $\widehat{\pi}^{(j)}_k$ denotes the kth component of $\widehat{\boldsymbol{\pi}}^{(j)}$. Without loss of generality, we assume that $\widehat{S} = \{1,\ldots,d\}$. Thus, based on the above KKT condition, it follows that

$$\mathbf{X}_{-j}^\top\mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}^{(j)}_{-j} - \boldsymbol{\pi}^{(j)}_{-j})/n = \mathbf{X}_{-j}^\top(\mathbf{x}_j - \mathbf{X}_{-j}\boldsymbol{\pi}^{(j)}_{-j})/n - \lambda_j\widehat{\boldsymbol{\kappa}}^{(j*)}, \tag{S2.29}$$

where $\widehat{\boldsymbol{\kappa}}^{(j*)} = (\mathbf{0}^\top, (\widehat{\boldsymbol{\kappa}}^{(j)}_{\widehat{S}^c\backslash\{j\}})^\top)^\top$.

In what follows, we use arguments similar to those in Ye and Zhang (2010) to construct bounds for $\|\widehat{\boldsymbol{\pi}}^{(j)}_{-j} - \boldsymbol{\pi}^{(j)}_{-j}\|_2$; the main difference here is that the partially penalized regression problem can be more complex. To this end, we first analyze the properties of $\widehat{\boldsymbol{\pi}}^{(j)}_{-j} - \boldsymbol{\pi}^{(j)}_{-j}$. Denote by $J_j^c = -j\backslash J_j = \{1,\ldots,p\}\backslash(J_j\cup\{j\})$. Thus, it follows from (S2.29) and

$\mathbf{v}^\top \mathbf{X}_{-j}^\top(\mathbf{x}_j - \mathbf{X}_{-j}\boldsymbol{\pi}_{-j}^{(j)})/n \le z_\infty^j \|\mathbf{v}\|_1$ that

$$\begin{aligned}\mathbf{v}^\top \mathbf{X}_{-j}^\top \mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})/n &= \mathbf{v}^\top \mathbf{X}_{-j}^\top(\mathbf{x}_j - \mathbf{X}_{-j}\boldsymbol{\pi}_{-j}^{(j)})/n - \lambda_j \mathbf{v}^\top \widehat{\boldsymbol{\kappa}}^{(j*)} \\ &\le z_\infty^j \|\mathbf{v}\|_1 - \lambda_j \mathbf{v}^\top \widehat{\boldsymbol{\kappa}}^{(j*)}\end{aligned}$$

for all vectors $\mathbf{v} \in \mathbb{R}^{p-1}$. If $\mathbf{v}$ satisfies $\mathrm{sign}(\mathbf{v}_{J_j^c}) = \mathrm{sign}((\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})_{J_j^c}) = \mathrm{sign}((\widehat{\boldsymbol{\pi}}_{-j}^{(j)})_{J_j^c})$, with the aid of $v_{jk} \ge \widetilde{M}^*$ for any $k \in \widehat{S}^c \backslash \{j\}$, it follows that

$$-\mathbf{v}^\top \widehat{\boldsymbol{\kappa}}^{(j*)} = -(\mathbf{v}_{J_j}^\top \widehat{\boldsymbol{\kappa}}_{J_j}^{(j*)} + \mathbf{v}_{J_j^c}^\top \widehat{\boldsymbol{\kappa}}_{J_j^c}^{(j*)}) \le -\mathbf{v}_{J_j}^\top \widehat{\boldsymbol{\kappa}}_{J_j}^{(j*)} - \widetilde{M}^* \|\mathbf{v}_{J_j^c}\|_1,$$

which together with $\widehat{S} \subset J_j$ and $v_{jk} \le \widetilde{M}$ for any $k \in \widehat{S}^c \backslash \{j\}$ further gives

$$-\mathbf{v}^\top \widehat{\boldsymbol{\kappa}}^{(j*)} \le \widetilde{M} \|\mathbf{v}_{J_j}\|_1 - \widetilde{M}^* \|\mathbf{v}_{J_j^c}\|_1.$$

Combining these results gives

$$\mathbf{v}^\top \mathbf{X}_{-j}^\top \mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})/n \le (z_\infty^j + \widetilde{M}\lambda_j)\|\mathbf{v}_{J_j}\|_1 + (z_\infty^j - \widetilde{M}^*\lambda_j)\|\mathbf{v}_{J_j^c}\|_1 \tag{S2.30}$$

for all vectors $\mathbf{v} \in \mathbb{R}^{p-1}$ satisfying $\mathrm{sign}(\mathbf{v}_{J_j^c}) = \mathrm{sign}((\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})_{J_j^c}) = \mathrm{sign}((\widehat{\boldsymbol{\pi}}_{-j}^{(j)})_{J_j^c})$. Based on the above inequality, if $z_\infty^j < \widetilde{M}^*\lambda_j$, it follows from $\mathbf{v} = \widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}$ that

$$(z_\infty^j + \widetilde{M}\lambda_j)\|(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})_{J_j}\|_1 + (z_\infty^j - \widetilde{M}^*\lambda_j)\|(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})_{J_j^c}\|_1 \ge 0,$$

which entails that

$$\|(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})_{J_j^c}\|_1 \le \xi \|(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})_{J_j}\|_1 \tag{S2.31}$$

for $\xi \ge (\widetilde{M}\lambda_j + z_\infty^j)/(\widetilde{M}^*\lambda_j - z_\infty^j)$.

Moreover, in view of (S2.29), with $\mathbf{e}^k = (0, \ldots, 0, \underbrace{1}_{k\text{th}}, 0, \ldots, 0) \in \mathbb{R}^{p-1}$, it follows from $\mathbf{v} = (\widehat{\pi}_k^{(j)} - \pi_k^{(j)})\mathbf{e}^k$ that

$$(\widehat{\pi}_k^{(j)} - \pi_k^{(j)})\mathbf{x}_k^\top \mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})/n \leq z_\infty^j |\widehat{\pi}_k^{(j)} - \pi_k^{(j)}| - \widetilde{M}^* \lambda_j |\widehat{\pi}_k^{(j)} - \pi_k^{(j)}| \tag{S2.32}$$

for any $k \in J_j^c$, where the inequality can be derived by applying an argument similar to that in (S2.30). With the aid of (S2.32), if $z_\infty^j \leq \widetilde{M}^* \lambda_j$, it follows that

$$(\widehat{\pi}_k^{(j)} - \pi_k^{(j)})\mathbf{x}_k^\top \mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})/n \leq 0 \tag{S2.33}$$

for any $k \in J_j^c$. In view of (S2.31) and (S2.33), if $z_\infty^j < \widetilde{M}^* \lambda_j$, we finally get

$$\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)} \in \mathcal{G}_-(\xi, J_j) \tag{S2.34}$$

for $\xi \geq (\widetilde{M}\lambda_j + z_\infty^j)/(\widetilde{M}^* \lambda_j - z_\infty^j)$, where the sign-restricted cone $\mathcal{G}_-(\xi, J_j)$ introduced in Ye and Zhang (2010) is defined as

$$\mathcal{G}_-(\xi, J_j) = \left\{ \mathbf{u} \in \mathbb{R}^{p-1} : \|\mathbf{u}_{J_j^c}\|_1 \leq \xi \|\mathbf{u}_{J_j}\|_1, u_k \mathbf{x}_k^\top \mathbf{X}_{-j}\mathbf{u}/n \leq 0, \quad \forall k \in J_j^c \right\}.$$

In order to proceed, our analysis will be conditional on the event $\mathcal{E}''' = \{z_\infty^j \leq \lambda_j \widetilde{M}^*(\xi' - 1)/(\xi' + 1)\}$ for some constant $\xi' > 1$. Note that $z_\infty^j \leq \lambda_j \widetilde{M}^*(\xi' - 1)/(\xi' + 1) < \widetilde{M}^* \lambda_j$ holds on this event, which together with (S2.34) implies that

$$\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)} \in \mathcal{G}_-(\xi, J_j)$$

for $\xi \geq (\widetilde{M}\lambda_j + z_\infty^j)/(\widetilde{M}^* \lambda_j - z_\infty^j)$. We next show that the ratio $(\widetilde{M}\lambda_j + z_\infty^j)/(\widetilde{M}^* \lambda_j - z_\infty^j)$ is bounded by a constant, so ξ can be chosen as a constant.

In view of $z_\infty^j \le \lambda_j \widetilde{M}^*(\xi' - 1)/(\xi' + 1)$, it follows that

$$\begin{aligned} &\frac{\widetilde{M}\lambda_j + z_\infty^j}{\widetilde{M}^*\lambda_j - z_\infty^j} + 1 = \frac{\widetilde{M}\lambda_j + \widetilde{M}^*\lambda_j}{\widetilde{M}^*\lambda_j - z_\infty^j} \\ \le &\frac{\widetilde{M}\lambda_j + \widetilde{M}^*\lambda_j}{\widetilde{M}^*\lambda_j - \lambda_j \widetilde{M}^*(\xi' - 1)/(\xi' + 1)} = \frac{(\xi' + 1)(\widetilde{M} + \widetilde{M}^*)}{2\widetilde{M}^*}, \end{aligned}$$

which entails that

$$\frac{\widetilde{M}\lambda_j + z_\infty^j}{\widetilde{M}^*\lambda_j - z_\infty^j} \le \frac{(\xi' + 1)(\widetilde{M} + \widetilde{M}^*)}{2\widetilde{M}^*} - 1 = 1 + \frac{\xi' - 1}{2} + \frac{(\xi' + 1)\widetilde{M} - 2\widetilde{M}^*}{2\widetilde{M}^*},$$

which is a constant larger than one in view of $\xi' > 1$. Combining this inequality with $(\widetilde{M}\lambda_j + z_\infty^j)/(\widetilde{M}^*\lambda_j - z_\infty^j) \ge (\widetilde{M}\lambda_j)/(\widetilde{M}^*\lambda_j) \ge 1$ further gives

$$1 \le \frac{\widetilde{M}\lambda_j + z_\infty^j}{\widetilde{M}^*\lambda_j - z_\infty^j} \le 1 + \frac{\xi' - 1}{2} + \frac{(\xi' + 1)\widetilde{M} - 2\widetilde{M}^*}{2\widetilde{M}^*}. \tag{S2.35}$$

Moreover, since $z_\infty^j \le \lambda_j \widetilde{M}^*(\xi' - 1)/(\xi' + 1) \le \lambda_j \widetilde{M}(\xi' - 1)/(\xi' + 1)$, it follows from (S2.29) that

$$\begin{aligned} &\|\mathbf{X}_{-j}^\top \mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})/n\|_\infty \le \widetilde{M}\lambda_j + z_\infty^j \\ &\qquad\qquad \le \widetilde{M}\lambda_j + \lambda_j \widetilde{M}(\xi' - 1)/(\xi' + 1) = 2\widetilde{M}\lambda_j \xi'/(\xi' + 1). \end{aligned}$$

Thus, in view of $|J_j| \le s_\Phi + d$, combining the above results gives

$$\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_2 \le \frac{|J_j|^{1/2}(\widetilde{M}\lambda_j + z_\infty^j)}{F_2(\xi, J_j)} \le \frac{\{2\widetilde{M}\xi'/(\xi' + 1)\}(s_\Phi + d)^{1/2}\lambda_j}{F_2(\xi, J_j)}, \tag{S2.36}$$

where the sign-restricted cone invertibility factor $F_2(\xi, J_j)$ introduced in Ye and Zhang (2010) is defined as

$$F_2(\xi, J_j) = \inf\left\{|J_j|^{1/2}\|\mathbf{X}_{-j}^\top \mathbf{X}_{-j}\mathbf{u}/n\|_\infty / \|\mathbf{u}\|_2 : \mathbf{u} \in \mathcal{G}_-(\xi, J_j)\right\}.$$

To derive a probability bound for the event $\mathcal{E}'''$, we use the following inequality, valid for $x > 0$:

$$\int_x^\infty e^{-t^2/2} dt \le e^{-x^2/2}/x. \tag{S2.37}$$

Since $\mathbf{e}_j$ is independent of $\mathbf{X}_{-j}$ and follows the distribution $N(0, (1/\phi_{jj})\mathbf{I}_n)$, we have

$$\mathbf{e}_j^\top \mathbf{x}_k \mid \mathbf{x}_k \sim N(0, \|\mathbf{x}_k\|_2^2/\phi_{jj})$$

for any $k \neq j$, which together with (S2.37) gives

$$\mathbb{P}\Big(\frac{\mathbf{e}_j^\top \mathbf{x}_k}{\sqrt{\|\mathbf{x}_k\|_2^2/\phi_{jj}}} > \sqrt{2\delta \log(p)} \,\Big|\, \mathbf{x}_k\Big) \le \frac{1}{\sqrt{2\pi}} \cdot \frac{p^{-\delta}}{\sqrt{2\delta\log(p)}}.$$

Since the right-hand side of the above inequality is independent of $\mathbf{x}_k$, integrating the conditional bound over $\mathbf{x}_k$ yields

$$\mathbb{P}\Big(\frac{\mathbf{e}_j^\top \mathbf{x}_k}{\sqrt{\|\mathbf{x}_k\|_2^2/\phi_{jj}}} > \sqrt{2\delta \log(p)}\Big) \le \frac{1}{\sqrt{2\pi}} \cdot \frac{p^{-\delta}}{\sqrt{2\delta\log(p)}}.$$

Combining this inequality with the probability bound for the event $\mathcal{E}'$ and setting $\lambda_j = [\widetilde{M}(\xi'+1)/\{\widetilde{M}^*(\xi'-1)\}]\sqrt{(2\delta/(n\phi_{jj}))\log(p)}$, we obtain

$$\begin{aligned}
1 - \mathbb{P}\left(z_\infty^j \le \frac{\lambda_j \widetilde{M}^*(\xi'-1)}{\xi'+1}\right) &= \mathbb{P}\Big(\max_{k\neq j, 1\le k\le p} \frac{|\mathbf{e}_j^\top \mathbf{x}_k|}{n} > \widetilde{M}\sqrt{\frac{2\delta\log(p)}{n\phi_{jj}}}\Big) \\
\le \mathbb{P}\Big(\max_{k\neq j, 1\le k\le p} \frac{|\mathbf{e}_j^\top \mathbf{x}_k|}{n} > \widetilde{M}\sqrt{\frac{2\delta\log(p)}{n\phi_{jj}}}\Big|\mathcal{E}'\Big) &+ \mathbb{P}((\mathcal{E}')^c) \\
\le 2\sum_{k\neq j, 1\le k\le p} \mathbb{P}\Big(\frac{\mathbf{e}_j^\top \mathbf{x}_k}{\sqrt{\|\mathbf{x}_k\|_2^2/\phi_{jj}}} > \frac{\widetilde{M}\sqrt{n}}{\|\mathbf{x}_k\|_2}\sqrt{2\delta\log(p)}\Big|\mathcal{E}'\Big) &+ o(p^{1-\delta}) \\
\le 2\sum_{k\neq j, 1\le k\le p} \mathbb{P}\Big(\frac{\mathbf{e}_j^\top \mathbf{x}_k}{\sqrt{\|\mathbf{x}_k\|_2^2/\phi_{jj}}} > \sqrt{2\delta\log(p)}\Big) &+ o(p^{1-\delta}) \le o(p^{1-\delta}).
\end{aligned} \tag{S2.38}$$

We still need to know the magnitude of $F_2(\xi, J_j)$. Since the eigenvalues of $\mathbf{\Sigma}$ are bounded within the interval $[1/L, L]$ for some constant $L \ge 1$, we can deduce that the eigenvalues of $\mathbf{\Sigma}_{-j,-j}$ are also bounded within the interval $[1/L, L]$ by Lemma 2. Since $s_\Phi + d = o(n/\log(p))$ follows from Condition 2 and $d = o(n/\log(p))$, in view of Lemma 4, there are positive constants c_1, c_2, c_3 such that

$$F_2(\xi, J_j) \geq c_1$$

with probability at least $1 - c_2 \exp(-c_3 n)$. Since $\log(p)/n = o(1)$, it follows that

$$\frac{\exp(-c_3 n)}{p^{1-\delta}} = \frac{p^{\delta-1}}{\exp(c_3 n)} = \exp\{(\delta - 1)\log p - c_3 n\} = o(1),$$

which entails $c_2 \exp(-c_3 n) = o(p^{1-\delta})$.

Moreover, with the aid of (S2.6), (S2.12), and (S2.38), we know that

$$\begin{aligned}
&\mathbb{P}(\mathcal{E}' \cap \mathcal{E}'' \cap \mathcal{E}''') = 1 - \mathbb{P}\left((\mathcal{E}')^c \cup (\mathcal{E}'')^c \cup (\mathcal{E}''')^c\right) \\
\geq &1 - \mathbb{P}((\mathcal{E}')^c) - \mathbb{P}((\mathcal{E}'')^c) - \mathbb{P}((\mathcal{E}''')^c) = 1 - o(p^{1-\delta}).
\end{aligned}$$

Thus, combining these results with (S2.36) gives

$$\mathbb{P}\left\{\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_2 \leq C(s_\Phi + d)^{1/2}\lambda_j\right\} \geq 1 - o(p^{1-\delta}). \tag{S2.39}$$

Part 2: Deviation bounds of $n^{-1/2}\|\mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})\|_2$ *and* $\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_q$ *with* $q \in [1, 2)$. We first construct bounds for $\|\mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})\|_2$. Some simple algebra shows that

$$\|\mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})\|_2^2 \leq \|\mathbf{X}_{-j}^\top \mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})\|_\infty \|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_1. \tag{S2.40}$$

In the event $\mathcal{E}''' = \{z_\infty^j \le \lambda_j \widetilde{M}^*(\xi'-1)/(\xi'+1)\}$, it follows that

$$\|\mathbf{X}_{-j}^\top \mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})/n\|_\infty \le \widetilde{M}\lambda_j + z_\infty^j \le \widetilde{M}\lambda_j + \lambda_j \widetilde{M}^*(\xi'-1)/(\xi'+1),$$

which further yields

$$\|\mathbf{X}_{-j}^\top \mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})\|_\infty \le n\{\widetilde{M} + \widetilde{M}^*(\xi'-1)/(\xi'+1)\}\lambda_j. \tag{S2.41}$$

Moreover, since $\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)} \in \mathcal{G}_-(\xi, J_j)$, we have

$$\begin{aligned}
\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_1 &= \|(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})_{J_j^c}\|_1 + \|(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})_{J_j}\|_1 \\
&\le (1+\xi)\|(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})_{J_j}\|_1 \\
&\le (1+\xi)(s_\Phi + d)^{1/2}\|(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})_{J_j}\|_2 \\
&\le (1+\xi)(s_\Phi + d)^{1/2}\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_2,
\end{aligned}$$

which together with (S2.39) entails that

$$\mathbb{P}\left\{\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_1 \le C(s_\Phi + d)\lambda_j\right\} \ge 1 - o(p^{1-\delta}). \tag{S2.42}$$

Then with the aid of (S2.40), (S2.41), and (S2.42), we finally get

$$\mathbb{P}\left\{\|\mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})\|_2 \le C\{(s_\Phi + d)n\}^{1/2}\lambda_j\right\} \ge 1 - o(p^{1-\delta}), \tag{S2.43}$$

which further entails that

$$\mathbb{P}\left\{n^{-1/2}\|\mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)})\|_2 \le C(s_\Phi + d)^{1/2}\lambda_j\right\} \ge 1 - o(p^{1-\delta}).$$

We can also derive the deviation bounds of $\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_q$ with $q \in (1,2)$. Note that the two inequalities $\sum_{j=1}^p a_j \le b_1$ and $\sum_{j=1}^p a_j^2 \le b_2$ for $a_j \ge 0$ imply

$$\sum_{j=1}^p a_j^q = \sum_{j=1}^p a_j^{2-q} a_j^{2q-2} \le \left(\sum_{j=1}^p a_j\right)^{2-q} \left(\sum_{j=1}^p a_j^2\right)^{q-1} \le b_1^{2-q} b_2^{q-1},$$

for any $q \in (1, 2)$. Thus, we have

$$\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_q^q \le \|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_1^{2-q} \cdot (\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_2^2)^{q-1},$$

which together with (S2.39) and (S2.42) gives

$$\mathbb{P}\left\{\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_q \le C(s_\Phi + d)^{1/q}\lambda_j\right\} \ge 1 - o(p^{1-\delta})$$

for any $q \in (1, 2)$. This completes the proof of Theorem 3.

S2.8 Proof of Theorem 4

Under the conditions of Theorem 4, it is clear that the results in *Part 1* and *Part 2* of the proof of Proposition 2 are still valid. Denote by $\mathcal{E} = \{S_1 \subset \widehat{S}\}$ the event in Definition 1. By Definition 1, $\widehat{S}$ is independent of the data $(\mathbf{X}, \mathbf{y})$, so we treat $\widehat{S}$ as given. Conditional on the event $\mathcal{E}$, we will first analyze the properties of some key statistics before deriving the confidence intervals of β_j. In what follows, we mainly present the proof when $j \in \widehat{S}^c$ and the same argument applies to $j \in \widehat{S}$. Throughout the proof, C is some positive constant which can vary from expression to expression.

Part 1: Deviation bounds of $\|\mathbf{z}_j\|_2$. We derive the deviation bounds of $\|\mathbf{z}_j\|_2$ in two different cases that correspond to the asymptotic results in Theorems 2 and 3, respectively. In each case, the assumption on s_1 is different, but we obtain asymptotically equivalent deviation bounds of $\|\mathbf{z}_j\|_2$. In both cases, the regularization parameters are $\lambda_j = (1 + \epsilon)\sqrt{(2\delta/(n\phi_{jj}))\log(p)}$, where

$\epsilon = \widetilde{M}(\xi' + 1)/\{\widetilde{M}^*(\xi' - 1)\} - 1 > 0$; these are the same as those in the proof of Theorem 3.

Case 1: $s_1 = o(n/s_\Phi)$. Since $s_\Phi = o(n/\log(p))$ and $d = O(s_1) = o(n/s_\Phi)$, the convergence rates established in Theorem 2 are asymptotically vanishing. Then in this case, we mainly utilize the results in Theorem 2. Since $\mathbf{e}_j$ is independent of $\mathbf{X}_{\widehat{S}}$, it follows from $\mathbf{e}_j^* = (\mathbf{I}_n - \mathbf{P}_{\widehat{S}})\mathbf{e}_j$ defined in (S2.17) that

$$\phi_{jj}\|\mathbf{e}_j^*\|_2^2|\mathbf{X}_{\widehat{S}} \sim \chi^2_{(n-d)}.$$

In view of $d = o(n)$, applying the same argument as in (S2.5) gives

$$\mathbb{P}\left\{\widetilde{M}_4^* \leq (\|\mathbf{e}_j^*\|_2/\sqrt{n}) \leq \widetilde{M}_4 \mid \mathbf{X}_{\widehat{S}}\right\} \geq 1 - o(p^{1-\delta}),$$

where $\widetilde{M}_4$ and $\widetilde{M}_4^*$ are positive constants satisfying $\widetilde{M}_4 \geq \widetilde{M}_4^*$. Since the right-hand side of the above inequality is independent of $\mathbf{X}_{\widehat{S}}$, we further have

$$\mathbb{P}\left\{\widetilde{M}_4^* \leq \|\mathbf{e}_j^*\|_2/\sqrt{n} \leq \widetilde{M}_4\right\} \geq 1 - o(p^{1-\delta}). \tag{S2.44}$$

In addition, it follows from $\boldsymbol{\psi}_j^{(j)} = \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}\boldsymbol{\omega}_j + \mathbf{e}_j^*$ and $\mathbf{z}_j = \boldsymbol{\psi}_j^{(j)} - \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}\widehat{\boldsymbol{\omega}}_j$ that

$$\mathbf{e}_j^* - \mathbf{z}_j = \boldsymbol{\psi}^{(j)}_{\widehat{S}^c\setminus\{j\}}(\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j),$$

which together with inequality (S2.27) gives

$$\mathbb{P}\left\{\|\mathbf{e}_j^*\|_2 - C(s_\Phi n)^{1/2}\lambda_j \leq \|\mathbf{z}_j\|_2 \leq \|\mathbf{e}_j^*\|_2 + C(s_\Phi n)^{1/2}\lambda_j\right\} \geq 1 - o(p^{1-\delta}). \tag{S2.45}$$

In view of $s_\Phi(\log p)/n = o(1)$, the above inequality together with the inequality (S2.44) yields that

$$\mathbb{P}\left\{\|\mathbf{z}_j\|_2 \asymp n^{1/2}\right\} \geq 1 - o(p^{1-\delta}). \tag{S2.46}$$

Case 2: $s_1 = o(n/\log(p))$. In view of $d = O(s_1)$ by Definition 1, it follows from $s_1 \log(p)/n = o(1)$ that $d \log(p)/n = o(1)$, so that the convergence rates in Theorem 3 are asymptotically vanishing. Then in this case, we mainly utilize the asymptotic results in Theorem 3. For $\mathbf{e}_j = \mathbf{x}_j - \mathbf{X}_{-j}\boldsymbol{\pi}_{-j}^{(j)}$ defined in (S2.28), applying the same argument as in (S2.5) gives

$$\mathbb{P}\left\{\widetilde{M}_3^* \leq \|\mathbf{e}_j\|_2/\sqrt{n} \leq \widetilde{M}_3\right\} \geq 1 - o(p^{1-\delta}), \tag{S2.47}$$

where $\widetilde{M}_3$ and $\widetilde{M}_3^*$ are positive constants satisfying $\widetilde{M}_3 \geq \widetilde{M}_3^*$. By Proposition 1, we have $\mathbf{z}_j = \mathbf{x}_j - \mathbf{X}_{-j}\widehat{\boldsymbol{\pi}}_{-j}^{(j)}$. This implies that

$$\mathbf{e}_j - \mathbf{z}_j = \mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}),$$

which together with the inequality (S2.43) gives, with probability at least $1 - o(p^{1-\delta})$,

$$\begin{aligned} \|\mathbf{e}_j\|_2 - C\{(s_\Phi + d)\log p\}^{1/2} &\leq \|\mathbf{z}_j\|_2 \\ &\leq \|\mathbf{e}_j\|_2 + C\{(s_\Phi + d)\log p\}^{1/2}. \end{aligned} \tag{S2.48}$$

In view of $s_\Phi(\log p)/n = o(1)$ and $d(\log p)/n = o(1)$, the above inequality together with inequality (S2.47) shows

$$\mathbb{P}\left\{\|\mathbf{z}_j\|_2 \asymp n^{1/2}\right\} \geq 1 - o(p^{1-\delta}). \tag{S2.49}$$

Therefore, in both cases, we have $\|\mathbf{z}_j\|_2 \asymp n^{1/2}$ with probability at least

$1 - o(p^{1-\delta})$.

Part 2: Deviation bounds of τ_j. First of all, it follows from the equality (2.9) that

$$\begin{aligned}\mathbf{z}_j^\top \mathbf{x}_j = \mathbf{z}_j^\top \boldsymbol{\psi}_j^{(j)} &= \|\mathbf{z}_j\|_2^2 + (\boldsymbol{\psi}_{\widehat{S}^c\backslash\{j\}}^{(j)} \widehat{\boldsymbol{\omega}}_j)^\top \mathbf{z}_j \\ &= \|\mathbf{z}_j\|_2^2 + \sqrt{n}\lambda_j \sum_{k\in\widehat{S}^c\backslash\{j\}} (\|\boldsymbol{\psi}_k^{(j)}\|_2 \cdot |\widehat{\omega}_{j,k}|) \\ &\geq \|\mathbf{z}_j\|_2^2,\end{aligned} \tag{S2.50}$$

where the third equality above follows from the Karush-Kuhn-Tucker (KKT) condition of the lasso estimator, which gives

$$\begin{aligned}\mathbf{z}_j^\top \boldsymbol{\psi}_k^{(j)} &= (\boldsymbol{\psi}_j^{(j)} - \boldsymbol{\psi}_{\widehat{S}^c\backslash\{j\}} \widehat{\boldsymbol{\omega}}_j)^\top \boldsymbol{\psi}_k^{(j)} \\ &= \sqrt{n}\lambda_j \|\boldsymbol{\psi}_k^{(j)}\|_2 \cdot \mathrm{sgn}(\widehat{\omega}_{j,k})\end{aligned} \tag{S2.51}$$

for any $k \in A = \{k \in \widehat{S}^c\backslash\{j\} : \mathrm{sgn}(\widehat{\omega}_{j,k}) \neq 0\}$, where $\widehat{\omega}_{j,k}$ is the kth component of $\widehat{\boldsymbol{\omega}}_j$. By (S2.50), we immediately have

$$\tau_j = \|\mathbf{z}_j\|_2 / |\mathbf{z}_j^\top \mathbf{x}_j| \leq 1/\|\mathbf{z}_j\|_2. \tag{S2.52}$$

It remains to find the lower bound of τ_j. With the aid of (S2.12), it follows that with probability at least $1 - o(p^{1-\delta})$,

$$\begin{aligned}\mathbf{z}_j^\top \mathbf{x}_j = \mathbf{z}_j^\top \boldsymbol{\psi}_j^{(j)} &= \|\mathbf{z}_j\|_2^2 + \sqrt{n}\lambda_j \sum_{k\in\widehat{S}^c\backslash\{j\}} (\|\boldsymbol{\psi}_k^{(j)}\|_2 \cdot |\widehat{\omega}_{j,k}|) \\ &\leq \|\mathbf{z}_j\|_2^2 + n\widetilde{M}_2 \lambda_j \|\widehat{\boldsymbol{\omega}}_j\|_1.\end{aligned} \tag{S2.53}$$

Based on (S2.52) and (S2.53), it follows that

$$\mathbb{P}\left\{ \frac{\|\mathbf{z}_j\|_2}{\|\mathbf{z}_j\|_2^2 + \widetilde{M}_2 n\lambda_j \|\widehat{\boldsymbol{\omega}}_j\|_1} \leq \tau_j \leq \frac{1}{\|\mathbf{z}_j\|_2} \right\} \geq 1 - o(p^{1-\delta}). \tag{S2.54}$$

We then derive the bounds of $\|\widehat{\boldsymbol{\omega}}_j\|_1$. Note that by the triangle inequality,

$$\|\widehat{\boldsymbol{\omega}}_j\|_1 \leq \|\boldsymbol{\omega}_j\|_1 + \|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_1.$$

According to the equality (S2.17), $\boldsymbol{\omega}_j = \boldsymbol{\pi}^{(j)}_{\widehat{S}^c\setminus\{j\}} = -\boldsymbol{\Phi}^\top_{j,\widehat{S}^c\setminus\{j\}}/\phi_{jj}$. Thus, we have $\|\boldsymbol{\omega}_j\|_1 \leq \|\boldsymbol{\pi}^{(j)}_{-j}\|_1 \leq s_\Phi^{1/2}\|\boldsymbol{\pi}^{(j)}_{-j}\|_2$, where $\boldsymbol{\pi}^{(j)}_{-j} = -\boldsymbol{\Phi}^\top_{j,-j}/\phi_{jj}$. Since the eigenvalues of $\boldsymbol{\Phi}$ are bounded within the interval $[1/L, L]$ for some constant $L \geq 1$, it follows from Lemma 2 that

$$\|\boldsymbol{\Phi}_{j,-j}\|_2 \leq L, \quad 1/L \leq \sigma_{jj} \leq L, \quad \text{and} \quad \sigma_{jj}\phi_{jj} \geq 1.$$

These facts imply that

$$\|\boldsymbol{\pi}^{(j)}_{-j}\|_2 \leq \|\boldsymbol{\Phi}_{j,-j}\|_2/\phi_{jj} \leq L\sigma_{jj} \leq L^2 = O(1),$$

which further gives

$$\|\boldsymbol{\omega}_j\|_1 \leq \|\boldsymbol{\pi}^{(j)}_{-j}\|_1 \leq s_\Phi^{1/2}\|\boldsymbol{\pi}^{(j)}_{-j}\|_2 \leq O(s_\Phi^{1/2}).$$

In *Case 1*, since $s_\Phi(\log p)/n = o(1)$, combining these results with (S2.26) gives, with probability at least $1 - o(p^{1-\delta})$,

$$\|\widehat{\boldsymbol{\omega}}_j\|_1 \leq \|\boldsymbol{\omega}_j\|_1 + \|\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j\|_1 = O(s_\Phi^{1/2}) + O(s_\Phi\lambda_j) = O(s_\Phi^{1/2}).$$

Thus, $\lambda_j\|\widehat{\boldsymbol{\omega}}_j\|_1 = O\left[\{s_\Phi \log(p)/n\}^{1/2}\right] = o(1)$ with the same probability.

In *Case 2*, Proposition 1 and (S2.42) give, with probability at least $1 - o(p^{1-\delta})$,

$$\|\widehat{\boldsymbol{\omega}}_j\|_1 \leq \|\boldsymbol{\pi}^{(j)}_{-j}\|_1 + \|\widehat{\boldsymbol{\pi}}^{(j)}_{-j} - \boldsymbol{\pi}^{(j)}_{-j}\|_1 = O(s_\Phi^{1/2}) + O\{(s_\Phi + d)\lambda_j\}.$$

Since $\lambda_j^2 = O\{\log(p)/n\}$, Condition 2 and $d\log(p)/n = o(1)$ yield

$$\lambda_j\|\widehat{\boldsymbol{\omega}}_j\|_1 = O\left\{\left(\frac{s_\Phi \log p}{n}\right)^{1/2}\right\} + O\left\{\frac{(s_\Phi + d)\log p}{n}\right\} = o(1)$$

with the same probability. Therefore, in both cases, $n\lambda_j\|\widehat{\boldsymbol{\omega}}_j\|_1 = o(n)$. Combining this result with (S2.46) in *Case 1*, (S2.49) in *Case 2*, and (S2.54) gives

$$\mathbb{P}\left\{\tau_j \asymp n^{-1/2}\right\} \geq 1 - o(p^{1-\delta}).$$

Moreover, we show that

$$\frac{n^{1/2}\tau_j}{\phi_{jj}^{1/2}} \xrightarrow{\mathbb{P}} 1.$$

The previously established inequalities (S2.50) and (S2.53), together with $n\lambda_j\|\widehat{\boldsymbol{\omega}}_j\|_1 = o(n)$ and $\|\mathbf{z}_j\|_2^2 \asymp n$, imply that

$$\frac{|\mathbf{z}_j^\top \mathbf{x}_j|}{\|\mathbf{z}_j\|_2^2} \xrightarrow{\mathbb{P}} 1.$$

It remains to derive the limit of $\|\mathbf{z}_j\|_2/\sqrt{n}$. In *Case 1*, since

$$\phi_{jj}\|\mathbf{e}_j^*\|_2^2 \mid \mathbf{X}_{\widehat{S}} \sim \chi^2_{(n-d)}$$

and $d = o(n)$, we have $\|\mathbf{e}_j^*\|_2/\sqrt{n} \xrightarrow{\mathbb{P}} \phi_{jj}^{-1/2}$. This result and (S2.45), together with $s_\Phi(\log p)/n = o(1)$, give

$$\frac{\|\mathbf{z}_j\|_2}{\sqrt{n}} \xrightarrow{\mathbb{P}} \phi_{jj}^{-1/2}.$$

In *Case 2*, (S2.48) and the assumptions $s_\Phi(\log p)/n = o(1)$ and $d(\log p)/n = o(1)$ give

$$\frac{\big|\|\mathbf{z}_j\|_2 - \|\mathbf{e}_j\|_2\big|}{\sqrt{n}} = o_{\mathbb{P}}(1).$$

Since $\mathbf{e}_j = \mathbf{x}_j - \mathbb{E}(\mathbf{x}_j \mid \mathbf{X}_{-j})$ follows the distribution $N(\mathbf{0}, (1/\phi_{jj})\mathbf{I}_n)$, we again have

$$\frac{\|\mathbf{z}_j\|_2}{\sqrt{n}} \xrightarrow{\mathbb{P}} \phi_{jj}^{-1/2}.$$

Therefore, in both cases,

$$\frac{n^{1/2}\tau_j}{\phi_{jj}^{1/2}} = \frac{n^{1/2}}{\phi_{jj}^{1/2}\|\mathbf{z}_j\|_2} \frac{\|\mathbf{z}_j\|_2^2}{|\mathbf{z}_j^\top \mathbf{x}_j|} \xrightarrow{\mathbb{P}} 1.$$

Part 3: Deviation bounds of the inference bias Δ_j. Note that the inference bias Δ_j here is defined as

$$\Delta_j = \sum_{k \in \widehat{S}^c \setminus \{j\}} \frac{\mathbf{z}_j^\top \mathbf{x}_k \beta_k}{\mathbf{z}_j^\top \mathbf{x}_j}.$$

Then we have

$$|\Delta_j| = \Big| \sum_{k \in \widehat{S}^c \setminus \{j\}} \frac{\mathbf{z}_j^\top \mathbf{x}_k \beta_k}{\mathbf{z}_j^\top \mathbf{x}_j} \Big| \le \left(\max_{k \in \widehat{S}^c \setminus \{j\}} \Big| \frac{\mathbf{z}_j^\top \mathbf{x}_k}{\mathbf{z}_j^\top \mathbf{x}_j} \Big| \right) \sum_{k \in \widehat{S}^c} |\beta_k|.$$

As in (S2.50) and (S2.51), the KKT condition and equality (9) give, for any $k \in \widehat{S}^c \setminus \{j\}$,

$$|\mathbf{z}_j^\top \mathbf{x}_k| = |\mathbf{z}_j^\top \boldsymbol{\psi}_k^{(j)}| = |(\boldsymbol{\psi}_j^{(j)} - \boldsymbol{\psi}_{\widehat{S}^c \setminus \{j\}} \widehat{\boldsymbol{\omega}}_j)^\top \boldsymbol{\psi}_k^{(j)}| \le \sqrt{n}\lambda_j \|\boldsymbol{\psi}_k^{(j)}\|_2.$$

Recall that the inequality (S2.11) gives

$$\begin{aligned} \mathbb{P}\big(\max_{k \in \widehat{S}^c \setminus \{j\}} \frac{\|\boldsymbol{\psi}_k^{(j)}\|_2}{\sqrt{n}} > \widetilde{M}_2\big) &\le \sum_{k \in \widehat{S}^c \setminus \{j\}} \mathbb{P}\big(\frac{\|\boldsymbol{\psi}_k^{(j)}\|_2}{\sqrt{n}} > \widetilde{M}_2\big) \\ &\le p \cdot 4p^{-2\delta} = 4p^{1-2\delta} = o(p^{1-\delta}). \end{aligned}$$

Thus, by (S2.46) in *Case 1*, (S2.49) in *Case 2*, and (S2.50), the following bound holds with probability at least $1 - o(p^{1-\delta})$:

$$\max_{k \in \widehat{S}^c \setminus \{j\}} \Big| \frac{\mathbf{z}_j^\top \mathbf{x}_k}{\mathbf{z}_j^\top \mathbf{x}_j} \Big| \le \max_{k \in \widehat{S}^c \setminus \{j\}} \frac{\sqrt{n}\lambda_j \|\boldsymbol{\psi}_k^{(j)}\|_2}{\|\mathbf{z}_j\|_2^2} \le C\sqrt{\log(p)/n}.$$

Since $S_1 \subset \widehat{S}$ on the event $\mathcal{E}$, the above inequality together with $\|\boldsymbol{\beta}_{S_1^c}\|_1 = o(1/\sqrt{\log p})$ in Condition 3 entails that

$$\mathbb{P}\left\{|\Delta_j| \le C\sqrt{\log(p)/n} \cdot o(1/\sqrt{\log p}) = o(n^{-1/2})\right\} \ge 1 - o(p^{1-\delta}).$$

Part 4: Confidence intervals of the coefficients β_j.

The proof of this part is the same as that of Theorem 1, and, therefore, has been omitted.

S2.9 Proof of Theorem 5

Denote by $\mathcal{E} = \{S_1 \subset \widehat{S}\}$ the event in Definition 1. By Definition 1, $\widehat{S}$ is independent of the data $(\mathbf{X}, \mathbf{y})$, so we treat $\widehat{S}$ as given. Conditional on the event $\mathcal{E}$, we first prove the deviation bounds in Parts 1–5 and then establish the confidence intervals for β_j in Part 6. In what follows, we mainly present the proof when $j \in \widehat{S}^c$ and the same argument applies to $j \in \widehat{S}$. Throughout the proof, the regularization parameters are $\lambda_j = (1+\epsilon)\sqrt{(2\delta/(nc_{K,L}))\log(p)}$, where $\epsilon = \widetilde{M}(\xi'+1)/\{\widetilde{M}^*(\xi'-1)\} - 1 > 0$. Throughout, C denotes a positive constant that may vary from line to line.

Part 1: Deviation bounds of $\|\mathbf{x}_k\|_2$ and $\|\boldsymbol{\psi}_k^{(j)}\|_2$. Let x_{ki} denote the ith component of $\mathbf{x}_k$. For each fixed k, the random variables x_{ki}, $i = 1, \ldots, n$, are i.i.d. sub-Gaussian, and their squares x_{ki}^2 are i.i.d. sub-exponential.

Bernstein's inequality for i.i.d. sub-exponential random variables then gives

$$\mathbb{P}\left(\bigcap_{k\neq j}\{\widetilde{M}^* \leq \|\mathbf{x}_k\|_2/\sqrt{n} \leq \widetilde{M}\}\right) \geq 1 - o(p^{1-\delta}).$$

Moreover, for each $k \in \widehat{S}^c \backslash \{j\}$, Lemma 5 shows that the population regression of $\mathbf{x}_k$ on $\mathbf{X}_{\widehat{S}}$ satisfies the assumptions of Lemma 6. Applying Lemma 6 with δ replaced by $\delta + 1$ and using a union bound, we obtain

$$\mathbb{P}\left(\bigcap_{k\in\widehat{S}^c\backslash\{j\}}\{\widetilde{M}^* \leq \|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n} \leq \widetilde{M}\}\right) \geq 1 - o(p^{1-\delta}).$$

Part 2: Deviation bounds of $\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_2$, $\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_1$, *and the prediction error.* Let $S^j = \{1 \leq k \leq p : k \neq j, \phi_{jk} \neq 0\}$, $z_\infty^j = \|\mathbf{X}_{-j}^\top(\mathbf{x}_j - \mathbf{X}_{-j}\boldsymbol{\pi}_{-j}^{(j)})/n\|_\infty$, and $J_j = \widehat{S} \cup S^j$. In this part, our analysis will be conditional on the event $\mathcal{E}' \cap \mathcal{E}'' \cap \mathcal{E}'''$, where $\mathcal{E}' = \bigcap_{k\neq j}\{\widetilde{M}^* \leq \|\mathbf{x}_k\|_2/\sqrt{n} \leq \widetilde{M}\}$, $\mathcal{E}'' = \bigcap_{k\in\widehat{S}^c\backslash\{j\}}\{\widetilde{M}^* \leq \|\boldsymbol{\psi}_k^{(j)}\|_2/\sqrt{n} \leq \widetilde{M}\}$, and $\mathcal{E}''' = \{z_\infty^j \leq \lambda_j \widetilde{M}^*(\xi'-1)/(\xi'+1)\}$ for some constant $\xi' > 1$.

Consider the linear regression

$$\mathbf{x}_j = \mathbf{X}_{-j}\boldsymbol{\pi}_{-j}^{(j)} + \mathbf{e}_j,$$

where $\boldsymbol{\pi}_{-j}^{(j)} = \arg\min_{\boldsymbol{\pi}} \mathbb{E}\,\|\mathbf{x}_j - \mathbf{X}_{-j}\boldsymbol{\pi}\|_2^2 = -\boldsymbol{\Phi}_{j,-j}^\top/\phi_{jj}$ is the population least squares regression coefficient. Moreover, the elements of the residual vector $\mathbf{e}_j$ are i.i.d. K'-sub-Gaussian random variables with zero mean and common variance $1/\phi_{jj}$, where K' is a constant depending on K and L. Let e_{ij} be the ith element of $\mathbf{e}_j$. By the definition of the population least squares coefficient, we have

$$\mathbb{E}(X_{i,-j}e_{ij}) = \mathbf{0}.$$

For every $k \neq j$, the random variables $X_{ik}e_{ij}$, $1 \leq i \leq n$, are i.i.d. and have mean zero. Since X_{ik} and e_{ij} are sub-Gaussian, these products are sub-exponential. Bernstein's inequality then gives positive constants c_0 and c_1, depending only on K and L, such that

$$\mathbb{P}\left(\left|\mathbf{x}_k^\top \mathbf{e}_j/n\right| > t\right) \leq 2\exp(-c_0 n t^2), \qquad 0 \leq t \leq c_1. \tag{S2.55}$$

Recall that $1+\epsilon = \widetilde{M}(\xi'+1)/\{\widetilde{M}^*(\xi'-1)\}$. Choose the constant $c_{K,L}$ in the definition of λ_j such that $c_{K,L} \leq c_0\widetilde{M}^2$. Hence,

$$t_n := \frac{\lambda_j \widetilde{M}^*(\xi'-1)}{\xi'+1} = \widetilde{M}\sqrt{\frac{2\delta\log(p)}{nc_{K,L}}}.$$

Since $\log(p)/n = o(1)$, we have $t_n \leq c_1$ for sufficiently large n. Applying (S2.55) and a union bound gives

$$\mathbb{P}\{(\mathcal{E}''')^c\} \leq 2(p-1)\exp(-c_0 n t_n^2) \leq 2p^{1-2\delta} = o(p^{1-\delta}). \tag{S2.56}$$

Combining this result with the probability bounds for $\mathcal{E}'$ and $\mathcal{E}''$ in Part 1 yields

$$\mathbb{P}(\mathcal{E}' \cap \mathcal{E}'' \cap \mathcal{E}''') \geq 1 - o(p^{1-\delta}). \tag{S2.57}$$

Let $\mathbf{h}_j = \widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}$. On $\mathcal{E}'\cap\mathcal{E}''\cap\mathcal{E}'''$, repeating the KKT argument used to derive (S2.29)–(S2.34) shows that, for some constant $\xi > 1$ independent of n and p,

$$\mathbf{h}_j \in \mathcal{G}_-(\xi, J_j), \qquad \|\mathbf{X}_{-j}^\top \mathbf{X}_{-j}\mathbf{h}_j/n\|_\infty \leq \widetilde{M}\lambda_j + z_\infty^j \leq C\lambda_j. \tag{S2.58}$$

On $\mathcal{E}'''$, (S2.35) shows that $(\widetilde{M}\lambda_j + z_\infty^j)/(\widetilde{M}^*\lambda_j - z_\infty^j)$ is bounded above by a constant; hence ξ can be chosen independently of n and p.

By Definition 1, $d = O(s_1)$. Hence, the assumptions $s_1 \log(p)/n = o(1)$ and $s_\Phi \log(p)/n = o(1)$ imply that

$$|J_j| \log(p)/n \leq (s_\Phi + d)\log(p)/n = o(1).$$

Since $\widehat{S}$ is treated as fixed, Lemma 7 applied to $\mathbf{X}_{-j}$ gives positive constants c_F, c_2, and c_3 such that the event $\mathcal{E}_F = \{F_2(\xi, J_j) \geq c_F\}$ satisfies

$$\mathbb{P}(\mathcal{E}_F) \geq 1 - c_2\exp(-c_3 n) = 1 - o(p^{1-\delta}).$$

On $\mathcal{E}' \cap \mathcal{E}'' \cap \mathcal{E}''' \cap \mathcal{E}_F$, (S2.58) and $F_2(\xi, J_j) \geq c_F$ give

$$\|\mathbf{h}_j\|_2 \leq \frac{|J_j|^{1/2}\|\mathbf{X}_{-j}^\top \mathbf{X}_{-j}\mathbf{h}_j/n\|_\infty}{F_2(\xi, J_j)} \leq C(s_\Phi + d)^{1/2}\lambda_j.$$

Together with (S2.57) and the probability bound for $\mathcal{E}_F$, this gives

$$\mathbb{P}\left\{\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_2 \leq C(s_\Phi + d)^{1/2}\lambda_j\right\} \geq 1 - o(p^{1-\delta}), \tag{S2.59}$$

Moreover, the cone condition in (S2.58) gives

$$\|\mathbf{h}_j\|_1 \leq (1+\xi)\|(\mathbf{h}_j)_{J_j}\|_1 \leq (1+\xi)|J_j|^{1/2}\|\mathbf{h}_j\|_2.$$

Combining this inequality with (S2.59) yields

$$\mathbb{P}\left\{\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)} - \boldsymbol{\pi}_{-j}^{(j)}\|_1 \leq C(s_\Phi + d)\lambda_j\right\} \geq 1 - o(p^{1-\delta}). \tag{S2.60}$$

Finally, by Hölder's inequality and (S2.58),

$$n^{-1}\|\mathbf{X}_{-j}\mathbf{h}_j\|_2^2 = \mathbf{h}_j^\top \mathbf{X}_{-j}^\top \mathbf{X}_{-j}\mathbf{h}_j/n \leq \|\mathbf{h}_j\|_1\|\mathbf{X}_{-j}^\top \mathbf{X}_{-j}\mathbf{h}_j/n\|_\infty \leq C(s_\Phi + d)\lambda_j^2.$$

Therefore,

$$\mathbb{P}\left\{n^{-1/2}\|\mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)}-\boldsymbol{\pi}_{-j}^{(j)})\|_2 \le C(s_\Phi+d)^{1/2}\lambda_j\right\} \ge 1-o(p^{1-\delta}). \quad \text{(S2.61)}$$

Part 3: Deviation bounds of $\|\mathbf{z}_j\|_2$. Bernstein's inequality for i.i.d. sub-exponential random variables gives

$$\mathbb{P}\left\{\widetilde{M}_3^* \le \|\mathbf{e}_j\|_2/\sqrt{n} \le \widetilde{M}_3\right\} \ge 1-o(p^{1-\delta}), \quad \text{(S2.62)}$$

where $\widetilde{M}_3$ and $\widetilde{M}_3^*$ are positive constants satisfying $\widetilde{M}_3 \ge \widetilde{M}_3^*$. By Proposition 1, we have $\mathbf{z}_j = \mathbf{x}_j - \mathbf{X}_{-j}\widehat{\boldsymbol{\pi}}_{-j}^{(j)}$. This implies that

$$\mathbf{e}_j - \mathbf{z}_j = \mathbf{X}_{-j}(\widehat{\boldsymbol{\pi}}_{-j}^{(j)}-\boldsymbol{\pi}_{-j}^{(j)}),$$

which together with the inequality (S2.61) gives, with probability at least $1-o(p^{1-\delta})$,

$$\begin{aligned}\|\mathbf{e}_j\|_2 - \widetilde{M}\{(s_\Phi+d)\log p\}^{1/2} &\le \|\mathbf{z}_j\|_2 \\ &\le \|\mathbf{e}_j\|_2 + \widetilde{M}\{(s_\Phi+d)\log p\}^{1/2}.\end{aligned}$$

In view of $s_\Phi(\log p)/n = o(1)$ and $d(\log p)/n = o(1)$, the above inequality together with inequality (S2.62) shows

$$\mathbb{P}\left\{\|\mathbf{z}_j\|_2 \asymp n^{1/2}\right\} \ge 1-o(p^{1-\delta}). \quad \text{(S2.63)}$$

Part 4: Deviation bounds of τ_j. The identity $\mathbf{z}_j = \mathbf{x}_j - \mathbf{X}_{-j}\widehat{\boldsymbol{\pi}}_{-j}^{(j)}$ gives

$$\mathbf{z}_j^\top\mathbf{x}_j = \|\mathbf{z}_j\|_2^2 + (\mathbf{X}_{-j}\widehat{\boldsymbol{\pi}}_{-j}^{(j)})^\top\mathbf{z}_j \ge \|\mathbf{z}_j\|_2^2.$$

This immediately gives

$$\tau_j = \|\mathbf{z}_j\|_2/|\mathbf{z}_j^\top\mathbf{x}_j| \le 1/\|\mathbf{z}_j\|_2.$$

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

$$\mathbf{z}_j^\top \mathbf{x}_j \leq \|\mathbf{z}_j\|_2^2 + n\widetilde{M}\lambda_j \|\widehat{\boldsymbol{\pi}}_{-j}^{(j)}\|_1.$$

Therefore,

$$\mathbb{P}\left\{ \frac{\|\mathbf{z}_j\|_2}{\|\mathbf{z}_j\|_2^2 + \widetilde{M}n\lambda_j \|\widehat{\boldsymbol{\pi}}_{-j}^{(j)}\|_1} \leq \tau_j \leq \frac{1}{\|\mathbf{z}_j\|_2} \right\} \geq 1 - o(p^{1-\delta}). \tag{S2.64}$$

We then derive bounds for $\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)}\|_1$. Note that $\boldsymbol{\pi}_{-j}^{(j)} = -\boldsymbol{\Phi}_{j,-j}^\top / \phi_{jj}$. Since the eigenvalues of $\boldsymbol{\Phi}$ are bounded within the interval $[1/L, L]$ for some constant $L \geq 1$, it follows from Lemma 2 that

$$\|\boldsymbol{\Phi}_{j,-j}\|_2 \leq L, \quad 1/L \leq \sigma_{jj} \leq L, \quad \text{and} \quad \sigma_{jj}\phi_{jj} \geq 1.$$

Thus,

$$\|\boldsymbol{\pi}_{-j}^{(j)}\|_2 \leq \|\boldsymbol{\Phi}_{j,-j}\|_2 / \phi_{jj} \leq L\sigma_{jj} \leq L^2 = O(1),$$

which further gives

$$\|\boldsymbol{\pi}_{-j}^{(j)}\|_1 \leq s_\Phi^{1/2} \|\boldsymbol{\pi}_{-j}^{(j)}\|_2 \leq O(s_\Phi^{1/2}).$$

Combining these results with (S2.60) gives, with probability at least $1 - o(p^{1-\delta})$,

$$\|\widehat{\boldsymbol{\pi}}_{-j}^{(j)}\|_1 \leq C\{s_\Phi^{1/2} + (s_\Phi + d)\lambda_j\}.$$

Consequently,

$$\lambda_j \|\widehat{\boldsymbol{\pi}}_{-j}^{(j)}\|_1 \leq C\{s_\Phi^{1/2}\lambda_j + (s_\Phi + d)\lambda_j^2\} = o(1),$$

where the last equality follows from $\lambda_j \asymp \sqrt{\log(p)/n}$, $s_\Phi \log(p)/n = o(1)$, and $d \log(p)/n = o(1)$. Hence, $n\lambda_j \|\widehat{\boldsymbol{\pi}}_{-j}^{(j)}\|_1 = o(n)$. Combining this result with (S2.63) and (S2.64) yields

$$\mathbb{P}\left\{\tau_j \asymp n^{-1/2}\right\} \geq 1 - o(p^{1-\delta}).$$

Part 5: Deviation bounds of the inference bias Δ_j*.* The proof of this part is the same as that of Theorem 4, and, therefore, has been omitted.

Part 6: Confidence intervals of the coefficients β_j*.*

The proof of this part is the same as that of Theorem 1, and, therefore, has been omitted.

S3 Additional technical details

This section presents the proofs of the lemmas used to establish the main theoretical results.

S3.1 Proof of Lemma 1

We first consider the conditional distribution

$$\mathbf{x}_j = \mathbf{X}_{-j}\boldsymbol{\pi}_{-j}^{(j)} + \mathbf{e}_j,$$

where $\mathbf{e}_j = \mathbf{x}_j - \mathbb{E}(\mathbf{x}_j|\mathbf{X}_{-j}) \sim N(\mathbf{0}, (1/\phi_{jj})\mathbf{I}_n)$ and $\boldsymbol{\pi}_{-j}^{(j)} = -\boldsymbol{\Phi}_{j,-j}^{\top}/\phi_{jj}$. Without loss of generality, we write $\boldsymbol{\Sigma}$ as

$$\begin{pmatrix} \boldsymbol{\Sigma}_{\widehat{S}\widehat{S}} & \boldsymbol{\Sigma}_{\widehat{S}\widehat{S}^c} \\ \boldsymbol{\Sigma}_{\widehat{S}^c\widehat{S}} & \boldsymbol{\Sigma}_{\widehat{S}^c\widehat{S}^c} \end{pmatrix}.$$

By some simple algebra, we can get

$$\mathbf{\Phi} = \mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{\Sigma}_{\widehat{S}\widehat{S}} & \mathbf{\Sigma}_{\widehat{S}\widehat{S}^c} \\ \mathbf{\Sigma}_{\widehat{S}^c\widehat{S}} & \mathbf{\Sigma}_{\widehat{S}^c\widehat{S}^c} \end{pmatrix}^{-1} = \begin{pmatrix} * & * \\ * & \mathbf{\Sigma}^{-1}_{\widehat{S}^c\widehat{S}^c.1} \end{pmatrix}, \tag{S3.65}$$

where $\mathbf{\Sigma}_{\widehat{S}^c\widehat{S}^c.1} = \mathbf{\Sigma}_{\widehat{S}^c\widehat{S}^c} - \mathbf{\Sigma}_{\widehat{S}^c\widehat{S}}\mathbf{\Sigma}^{-1}_{\widehat{S}\widehat{S}}\mathbf{\Sigma}_{\widehat{S}\widehat{S}^c}$.

On the other hand, the conditional distribution of $\mathbf{X}_{\widehat{S}^c}$ given $\mathbf{X}_{\widehat{S}}$ yields

$$\boldsymbol{\rho}^{(j)}_{\widehat{S}^c} \sim N(\mathbf{0}, \mathbf{I}_n \otimes (\mathbf{\Sigma}_{\widehat{S}^c\widehat{S}^c} - \mathbf{\Sigma}_{\widehat{S}^c\widehat{S}}\mathbf{\Sigma}^{-1}_{\widehat{S}\widehat{S}}\mathbf{\Sigma}_{\widehat{S}\widehat{S}^c})).$$

Then it follows from (S3.65) that

$$(\mathbf{\Sigma}_{\widehat{S}^c\widehat{S}^c} - \mathbf{\Sigma}_{\widehat{S}^c\widehat{S}}\mathbf{\Sigma}^{-1}_{\widehat{S}\widehat{S}}\mathbf{\Sigma}_{\widehat{S}\widehat{S}^c})^{-1} = \mathbf{\Sigma}^{-1}_{\widehat{S}^c\widehat{S}^c.1} = \mathbf{\Phi}_{\widehat{S}^c\widehat{S}^c}.$$

Therefore, we immediately have

$$\boldsymbol{\rho}^{(j)}_{\widehat{S}^c} \sim N(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Phi}^{-1}_{\widehat{S}^c\widehat{S}^c}),$$

which further yields

$$\boldsymbol{\rho}^{(j)}_{\widehat{S}^c\setminus\{j\}} \sim N\left(\mathbf{0}, \mathbf{I}_n \otimes \left[(\mathbf{\Phi}_{\widehat{S}^c\widehat{S}^c})^{-1}\right]_{\widehat{S}^c\setminus\{j\},\widehat{S}^c\setminus\{j\}}\right).$$

This completes the proof of Lemma 1.

S3.2 Proof of Lemma 2

We prove the conclusions of Lemma 2 in turn.

1. In view of the fact that the eigenvalues of $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ are the inverses of each other, the first conclusion is clear.

2. Suppose that the eigenvalues of $\mathbf{\Phi}$ are $1/L \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p \leq$

L. Then there exists an orthonormal matrix $\mathbf{Q} = (q_{ij})_{p\times p}$ such that

$$\boldsymbol{\Phi} = \mathbf{Q} \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{pmatrix} \mathbf{Q}^\top. \tag{S3.66}$$

Thus, for any $1 \le k \le p$, we have

$$1/L \le \lambda_1 \le \phi_{kk} = \lambda_1 q_{k1}^2 + \ldots + \lambda_p q_{kp}^2 \le \lambda_p \le L,$$

which completes the proof of this conclusion.

3. By Cauchy's interlace theorem for eigenvalues in (Hwang, 2004, Theorem 1), this conclusion is also clear.

4. Based on (S3.66), we have

$$\boldsymbol{\Phi}\boldsymbol{\Phi}^\top = \mathbf{Q} \begin{pmatrix} \lambda_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p^2 \end{pmatrix} \mathbf{Q}^\top. \tag{S3.67}$$

Denote by $\|\boldsymbol{\phi}_j\|_2$ the L_2-norm of the jth row of $\boldsymbol{\Phi}$. Since $\boldsymbol{\Phi}$ is symmetric, $\|\boldsymbol{\phi}_j\|_2^2$ is equal to the jth diagonal component of $\boldsymbol{\Phi}\boldsymbol{\Phi}^\top$ for any $j \in \{1, \ldots, p\}$.

In view of (S3.67), we know that the eigenvalues of the symmetric matrix $\boldsymbol{\Phi}\boldsymbol{\Phi}^\top$ are bounded within the interval $[L^{-2}, L^2]$. Thus, the second conclusion of this lemma gives that $\|\boldsymbol{\phi}_j\|_2^2$ is bounded within the interval $[L^{-2}, L^2]$, which yields that $\|\boldsymbol{\phi}_j\|_2$ is bounded within the interval $[1/L, L]$. This completes the proof of this conclusion.

5. Based on (S3.66), we have

$$\boldsymbol{\Sigma} = \mathbf{Q} \begin{pmatrix} 1/\lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\lambda_p \end{pmatrix} \mathbf{Q}^\top.$$

Thus, for any $1 \leq j \leq p$, we obtain

$$\sigma_{jj} = \frac{1}{\lambda_1} q_{j1}^2 + \ldots + \frac{1}{\lambda_p} q_{jp}^2.$$

By the Cauchy–Schwarz inequality, we get

$$\begin{aligned} \sigma_{jj}\phi_{jj} &= \left(\frac{1}{\lambda_1} q_{j1}^2 + \ldots + \frac{1}{\lambda_p} q_{jp}^2\right)(\lambda_1 q_{j1}^2 + \ldots + \lambda_p q_{jp}^2) \\ &= \left(\sum_{k=1}^p \left(\frac{q_{jk}}{\sqrt{\lambda_k}}\right)^2\right)\left(\sum_{k=1}^p (q_{jk}\sqrt{\lambda_k})^2\right) \geq \left(\sum_{k=1}^p q_{jk}^2\right)^2 = 1. \end{aligned}$$

This concludes the proof of Lemma 2.

S3.3 Proof of Lemma 3

Since $\mathbf{X}\boldsymbol{\Sigma}^{-1/2}$ has independent standard Gaussian entries and $n \geq d$, it has full column rank with probability one. Moreover, $\boldsymbol{\Sigma}$ is positive definite. Hence, $\mathbf{X}$ has full column rank with probability one, so that $\mathbf{X}^\top\mathbf{X}$ is positive definite and the ordinary least squares estimator exists with probability one.

We next establish a bound for the prediction error of $\widehat{\boldsymbol{\beta}}$. Since $\mathbf{X}$ is independent of $\boldsymbol{\varepsilon}$, standard properties of the ordinary least squares estimator imply that, conditional on $\mathbf{X}$,

$$\frac{\|\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_2^2}{\sigma^2} \big| \mathbf{X} \sim \chi^2_{(d)}.$$

For a random variable $Z \sim \chi^2_{(d)}$ and any $x > 0$, the additive chi-squared tail bounds give

$$\mathbb{P}\{Z - d \geq 2\sqrt{dx} + 2x\} \leq \exp(-x) \quad \text{and} \quad \mathbb{P}\{d - Z \geq 2\sqrt{dx}\} \leq \exp(-x).$$

Taking $x = 2\delta \log p$ and applying the union bound, conditional on $\mathbf{X}$ we obtain

$$\mathbb{P}\Big[\max\{d - 2\sqrt{2\delta d \log p}, 0\}\sigma^2 \leq \|\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_2^2 \\ \leq \{d + 2\sqrt{2\delta d \log p} + 4\delta \log p\}\sigma^2 \,\Big|\, \mathbf{X}\Big] \geq 1 - 2p^{-2\delta}.$$

Since the right-hand side is independent of $\mathbf{X}$, the same bound holds unconditionally, which completes the proof of Lemma 3.

S3.4 Proof of Lemma 4

Since $\mathcal{G}_-(\xi, S) \subset \mathcal{G}(\xi, S)$, it is immediate that $F_2(\xi, S) \geq F_2^*(\xi, S)$. To show the constant lower bound, we first establish a connection between $F_2(\xi, S)$ and the restricted eigenvalues (RE), which are defined as

$$RE_m(\xi, S) = \inf \left\{ \frac{s^{1/m}\|\mathbf{X}\mathbf{u}\|_2}{(ns)^{1/2}\|\mathbf{u}\|_m} : \mathbf{u} \in \mathcal{G}(\xi, S) \right\}$$

for a positive integer m and the cone $\mathcal{G}(\xi, S)$ defined as in Lemma 4. By (Ye and Zhang, 2010, Proposition 5), we have

$$F_2^*(\xi, S) \geq RE_1(\xi, S) RE_2(\xi, S).$$

For every $\mathbf{u} \in \mathcal{G}(\xi, S)$, the cone constraint gives $\|\mathbf{u}\|_1 \leq (1+\xi)s^{1/2}\|\mathbf{u}\|_2$. It follows from the definitions of $RE_1(\xi, S)$ and $RE_2(\xi, S)$ that

$$RE_1(\xi, S) \geq \frac{RE_2(\xi, S)}{1+\xi},$$

and consequently

$$F_2^*(\xi, S) \geq \frac{RE_2^2(\xi, S)}{1+\xi}.$$

We then prove a constant lower bound for $RE_2(\xi, S)$ with high probability. First, it follows from (Raskutti et al., 2010, Theorem 1) that for any $\mathbf{u} \in \mathbb{R}^p$, there exist positive constants c_2 and c_3 so that

$$\frac{\|\mathbf{X}\mathbf{u}\|_2}{n^{1/2}} \geq \frac{1}{4}\|\mathbf{\Sigma}^{1/2}\mathbf{u}\|_2 - 9\rho(\mathbf{\Sigma})\sqrt{\frac{\log p}{n}}\|\mathbf{u}\|_1$$

holds with probability at least $1 - c_2\exp(-c_3 n)$, where

$$\rho(\mathbf{\Sigma}) = \max_{j\in\{1,\ldots,p\}} \mathbf{\Sigma}_{jj}^{1/2}.$$

On the other hand, for $\mathbf{u} \in \mathcal{G}(\xi, S)$, it follows from the definition of $\mathcal{G}(\xi, S)$ that

$$\|\mathbf{u}\|_1 = \|\mathbf{u}_{S^c}\|_1 + \|\mathbf{u}_S\|_1 \leq (1+\xi)\|\mathbf{u}_S\|_1 \leq (1+\xi)s^{1/2}\|\mathbf{u}_S\|_2.$$

In addition, we have $\|\mathbf{\Sigma}^{1/2}\mathbf{u}\|_2 \geq \lambda_{\min}(\mathbf{\Sigma}^{1/2})\|\mathbf{u}\|_2$, where $\lambda_{\min}(\mathbf{\Sigma}^{1/2})$ is the smallest eigenvalue of $\mathbf{\Sigma}^{1/2}$. Thus, it follows that

$$\frac{\|\mathbf{X}\mathbf{u}\|_2}{n^{1/2}} \geq \left\{\frac{1}{4}\lambda_{\min}(\mathbf{\Sigma}^{1/2}) - 9(1+\xi)\rho(\mathbf{\Sigma})\sqrt{\frac{s\log(p)}{n}}\right\}\|\mathbf{u}\|_2 \tag{S3.68}$$

holds with probability at least $1 - c_2\exp(-c_3 n)$. Since the eigenvalues of $\mathbf{\Sigma}$ are bounded within the interval $[1/L, L]$, both $\lambda_{\min}(\mathbf{\Sigma}^{1/2})$ and $\rho(\mathbf{\Sigma})$ are bounded from above and below by some positive constants according to

Lemma 2.

Therefore, when $s = o(n/\log(p))$, it follows from the definition of $RE_2(\xi, S)$ that there exists some positive constant c_0 such that for sufficiently large n,

$$RE_2(\xi, S) \geq c_0^{1/2}$$

holds with probability at least $1 - c_2 \exp(-c_3 n)$. Consequently,

$$F_2(\xi, S) \geq F_2^*(\xi, S) \geq \frac{c_0}{1+\xi} =: c_1$$

on the same event, which completes the proof of Lemma 4.

S3.5 Proof of Lemma 5

Let $C = \widehat{S}^c$ and $B = C \backslash \{j\}$. Expanding the population least squares objective and differentiating with respect to $\boldsymbol{\gamma}$ give the normal equation

$$\boldsymbol{\Sigma}_{\widehat{S}\widehat{S}} \boldsymbol{\gamma} = \boldsymbol{\Sigma}_{\widehat{S}j}.$$

Since $\boldsymbol{\Sigma}_{\widehat{S}\widehat{S}}$ is positive definite, the unique solution is

$$\boldsymbol{\gamma}_j^{(j)} = \boldsymbol{\Sigma}_{\widehat{S}\widehat{S}}^{-1} \boldsymbol{\Sigma}_{\widehat{S}j}.$$

Consequently, $\mathbb{E}(\boldsymbol{\rho}_j^{(j)}) = \mathbf{0}$, and

$$\mathbb{E}(\mathbf{X}_{\widehat{S}}^\top \boldsymbol{\rho}_j^{(j)}) = n \left(\boldsymbol{\Sigma}_{\widehat{S}j} - \boldsymbol{\Sigma}_{\widehat{S}\widehat{S}} \boldsymbol{\gamma}_j^{(j)} \right) = \mathbf{0}.$$

For the ith row, the residual vector corresponding to the variables in C is

$$\boldsymbol{\rho}_{i,C}^\top = \mathbf{X}_{i,C}^\top - \boldsymbol{\Sigma}_{C\widehat{S}} \boldsymbol{\Sigma}_{\widehat{S}\widehat{S}}^{-1} \mathbf{X}_{i,\widehat{S}}^\top.$$

The coefficient matrix satisfies

$$\left\|\mathbf{\Sigma}_{C\widehat{S}}\mathbf{\Sigma}_{\widehat{S}\widehat{S}}^{-1}\right\|_{\mathrm{op}} \le \|\mathbf{\Sigma}\|_{\mathrm{op}}\|\mathbf{\Sigma}_{\widehat{S}\widehat{S}}^{-1}\|_{\mathrm{op}} \le L^2.$$

By Definition 2, $\boldsymbol{\rho}_{i,C}^{\top}$ is a mean-zero K'-sub-Gaussian random vector, where K' depends only on K and L. Since the rows of $\mathbf{X}$ are i.i.d., the residual rows are also i.i.d.

Finally, their common covariance matrix is the Schur complement

$$\mathrm{Cov}(\boldsymbol{\rho}_{i,C}^{\top}) = \mathbf{\Sigma}_{CC} - \mathbf{\Sigma}_{C\widehat{S}}\mathbf{\Sigma}_{\widehat{S}\widehat{S}}^{-1}\mathbf{\Sigma}_{\widehat{S}C} = (\mathbf{\Phi}_{CC})^{-1},$$

where the last equality follows from the block inverse formula. Taking the principal submatrix indexed by $B = C\backslash\{j\}$ gives

$$\mathrm{Cov}(\boldsymbol{\rho}_{i,B}^{\top}) = \left[(\mathbf{\Phi}_{CC})^{-1}\right]_{B,B}.$$

This is the covariance matrix stated in the lemma, which completes the proof.

S3.6 Proof of Lemma 6

Since $d\log(p)/n = o(1)$, we have $d = o(n)$. Moreover, $\mathbf{X}\mathbf{\Sigma}^{-1/2}$ has independent isotropic sub-Gaussian rows. It follows from (Eldar and Kutyniok, 2012, Theorem 5.39) that, for some constant $c_1 > 0$, the event

$$\mathcal{E}_1 = \{\lambda_{\min}(\mathbf{X}^{\top}\mathbf{X}) \ge n/(2L)\}$$

satisfies

$$\mathbb{P}(\mathcal{E}_1) \ge 1 - 2\exp(-c_1 n). \tag{S3.69}$$

Thus, $\mathbf{X}^\top\mathbf{X}$ is positive definite and the ordinary least squares estimator exists on $\mathcal{E}_1$. Denote by

$$\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$$

the projection matrix onto the column space of $\mathbf{X}$.

We then derive upper and lower bounds for the residual sum of squares. Since the ε_i are i.i.d. sub-Gaussian random variables with variance σ^2, the random variables $\varepsilon_i^2 - \sigma^2$ are i.i.d. sub-exponential. Thus, Bernstein's inequality gives a constant $c_2 > 0$ such that the event

$$\mathcal{E}_2 = \left\{\frac{\sigma^2}{2}n \leq \|\boldsymbol{\varepsilon}\|_2^2 \leq 2\sigma^2 n\right\}$$

satisfies

$$\mathbb{P}(\mathcal{E}_2) \geq 1 - 2\exp(-c_2 n). \tag{S3.70}$$

For any $1 \leq j \leq d$, the random variables $X_{ij}\varepsilon_i$, $1 \leq i \leq n$, are i.i.d. with zero mean. Since both X_{ij} and ε_i are sub-Gaussian, their products are sub-exponential. Thus, by Bernstein's inequality and a union bound, there exists a constant $C_0 > 0$ such that the event

$$\mathcal{E}_3 = \left\{\|\mathbf{X}^\top\boldsymbol{\varepsilon}\|_\infty \leq C_0\{n\log(p)\}^{1/2}\right\}$$

satisfies

$$\mathbb{P}(\mathcal{E}_3^c) \leq 2dp^{-(\delta+1)} \leq 2p^{-\delta}. \tag{S3.71}$$

On $\mathcal{E}_1 \cap \mathcal{E}_3$, it follows that

$$\boldsymbol{\varepsilon}^\top\mathbf{P}_{\mathbf{X}}\boldsymbol{\varepsilon} \leq \lambda_{\max}\{(\mathbf{X}^\top\mathbf{X})^{-1}\}\|\mathbf{X}^\top\boldsymbol{\varepsilon}\|_2^2 \leq \frac{2L}{n}d\|\mathbf{X}^\top\boldsymbol{\varepsilon}\|_\infty^2 \leq 2LC_0^2 d\log(p) = o(n).$$

Consequently, on $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ and for sufficiently large n,

$$\|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\|_2^2 = \boldsymbol{\varepsilon}^\top(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})\boldsymbol{\varepsilon} = \|\boldsymbol{\varepsilon}\|_2^2 - \boldsymbol{\varepsilon}^\top \mathbf{P}_{\mathbf{X}}\boldsymbol{\varepsilon},$$

which gives

$$\frac{\sigma^2}{4}n \leq \|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\|_2^2 \leq 2\sigma^2 n.$$

Finally, $\log(p)/n = o(1)$ gives $\exp(-c_k n) = o(p^{1-\delta})$ for $k = 1, 2$, while $p^{-\delta} = o(p^{1-\delta})$. Combining these results with (S3.69)–(S3.71), we obtain

$$\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3) \geq 1 - o(p^{1-\delta}),$$

and hence $\|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\|_2 \asymp \sqrt{n}$ on this event, which completes the proof of Lemma 6.

S3.7 Proof of Lemma 7

Since $\mathcal{G}_-(\xi, S) \subset \mathcal{G}(\xi, S)$, it follows immediately that $F_2(\xi, S) \geq F_2^*(\xi, S)$. Recall the restricted eigenvalues $RE_m(\xi, S)$ defined in the proof of Lemma 4. By (Ye and Zhang, 2010, Proposition 5),

$$F_2^*(\xi, S) \geq RE_1(\xi, S) RE_2(\xi, S).$$

Moreover, the cone constraint gives $\|\mathbf{u}\|_1 \leq (1+\xi)s^{1/2}\|\mathbf{u}\|_2$ for every $\mathbf{u} \in \mathcal{G}(\xi, S)$. Thus,

$$F_2^*(\xi, S) \geq \frac{RE_2^2(\xi, S)}{1+\xi}. \tag{S3.72}$$

We then establish a constant lower bound for $RE_2(\xi, S)$. Let $A > 1$ be a constant to be specified later, let $r = \lceil As \rceil$, and set $q = s + r$. For a fixed set $T \subset \{1, \ldots, p\}$ with $|T| \le q$, the matrix $\mathbf{X}_T \boldsymbol{\Sigma}_{TT}^{-1/2}$ has independent isotropic sub-Gaussian rows whose sub-Gaussian norms are bounded by a constant depending only on K and L. Theorem 5.39 of Eldar and Kutyniok (2012) and the eigenvalue bounds for $\boldsymbol{\Sigma}_{TT}$ yield uniform lower and upper singular-value bounds for $\mathbf{X}_T$, except on an event of probability at most $2\exp(-cn)$ for some $c > 0$. The number of such sets is at most qp^q. Since $q\log(p)/n = o(1)$, a union bound gives positive constants a_0, b_0, c_2, and c_3 such that, with probability at least $1 - c_2\exp(-c_3 n)$,

$$a_0\|\mathbf{v}\|_2 \le n^{-1/2}\|\mathbf{X}\mathbf{v}\|_2 \le b_0\|\mathbf{v}\|_2 \tag{S3.73}$$

simultaneously for all vectors $\mathbf{v}$ supported on at most q coordinates. The corresponding failure probability is bounded by

$$2q\exp\{q\log(p) - cn\} \le c_2\exp(-c_3 n)$$

for sufficiently large n.

On the event in (S3.73), fix $\mathbf{u} \in \mathcal{G}(\xi, S)$. Let T_1 contain the indices of the r largest components of $\mathbf{u}_{S^c}$ in absolute value, and partition the remaining indices into disjoint sets $T_2, T_3, \ldots$ of size at most r in decreasing order of the magnitudes of the corresponding components. By the ordering of the coordinates in $T_1, T_2, \ldots$,

$$\sum_{\ell \ge 2}\|\mathbf{u}_{T_\ell}\|_2 \le \frac{\|\mathbf{u}_{S^c}\|_1}{\sqrt{r}} \le \xi\sqrt{\frac{s}{r}}\|\mathbf{u}_S\|_2 \le \frac{\xi}{\sqrt{A}}\|\mathbf{u}_{S\cup T_1}\|_2. \tag{S3.74}$$

Since $|S \cup T_1| \le q$ and $|T_\ell| \le q$, inequalities (S3.73) and (S3.74) yield

$$\begin{aligned} n^{-1/2}\|\mathbf{X}\mathbf{u}\|_2 &\ge n^{-1/2}\|\mathbf{X}\mathbf{u}_{S\cup T_1}\|_2 - \sum_{\ell\ge 2} n^{-1/2}\|\mathbf{X}\mathbf{u}_{T_\ell}\|_2 \\ &\ge \left(a_0 - \frac{b_0\xi}{\sqrt{A}}\right)\|\mathbf{u}_{S\cup T_1}\|_2. \end{aligned}$$

Choose A sufficiently large so that $b_0\xi/\sqrt{A} \le a_0/2$. Moreover, (S3.74) gives

$$\|\mathbf{u}\|_2 \le \left(1 + \frac{\xi}{\sqrt{A}}\right)\|\mathbf{u}_{S\cup T_1}\|_2.$$

Consequently, there exists a positive constant c_0 depending only on K, L, and ξ such that

$$RE_2(\xi, S) \ge c_0$$

with probability at least $1 - c_2\exp(-c_3 n)$. Combining this result with (S3.72) gives

$$F_2(\xi, S) \ge F_2^*(\xi, S) \ge \frac{c_0^2}{1+\xi} =: c_1,$$

which completes the proof.

References

Eldar, Y. C. and Kutyniok, G. (2012). *Compressed Sensing: Theory and Applications*. Cambridge University Press, Cambridge.

Hwang, S.-G. (2004). Cauchy's interlace theorem for eigenvalues of Hermitian matrices. *Amer. Math. Monthly 111*(2), 157–159.

Janková, J. and van de Geer, S. (2017). Honest confidence regions and optimality in high-dimensional precision matrix estimation. *TEST 26*(1), 143–162.

Raskutti, G., Wainwright, M. J. and Yu, B. (2010). Restricted eigenvalue properties for correlated Gaussian designs. *J. Mach. Learn. Res. 11*(78), 2241–2259.

Ren, Z., Sun, T., Zhang, C.-H. and Zhou, H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist. 43*(3), 991–1026.

van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist. 42*(3), 1166–1202.

Ye, F. and Zhang, C.-H. (2010). Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *J. Mach. Learn. Res. 11*(114), 3519–3540.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol. 76*(1), 217–242.