SEMI-STANDARD PARTIAL COVARIANCE VARIABLE SELECTION WHEN IRREPRESENTABLE CONDITIONS FAIL

Fei Xue and Annie Qu

University of Pennsylvania and University of California Irvine

Supplementary Material

In this Supplementary Material, we provide additional conditions, theorems, tables, and corollaries, as well as proofs for Lemma 1, all theorems, propositions and corollaries.

S1 Additional conditions

To simplify statements, we introduce the following notations for limiting behaviors of sequences with certain probability. A function $g_1(n)$ satisfies $g_1(n) = O_p^{(\delta)}(g_2(n))$ if and only if there exists a constant $M > 0$ such that $|g_1(n)| \leq M g_2(n)$ with probability at least $1 - O(n^{-\delta})$, where $\delta > 0$. Similarly, $g_1(n) = o_p^{(\delta)}(g_2(n))$ if and only if for any $\epsilon > 0$ there exists a constant N such that for all $n \geq N$, $|g_1(n)| \leq \epsilon g_2(n)$ with probability at least $1 - O(n^{-\delta})$.

Condition 4. For some constants $0 < \delta < \min\{1 - \kappa_2, 2\kappa_0\}$ and $0 < \kappa_3, \kappa_4 < 1/2$ such that $\kappa_3 + \kappa_4 < 1/2$, $\|\hat{\boldsymbol{C}}_n^{21}\|_\infty = O_p^{(\delta)}(n^{\kappa_3})$, $\lambda_{\min}(\hat{\boldsymbol{C}}_n^{11}) > 0$, $\|(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty = O_p^{(\delta)}(n^{\kappa_4}\sqrt{\log p_n/n})$, $\nu_{n,u} b_n = o(1)$, $\lambda_n = O(n^{1/2+\kappa_3})$, $1/\lambda_n = o(n^{\kappa_0 - 1}/\log n)$, $\max_{1\leq j\leq q_n} d_{jj} = O(\log n)$, and

$\max_{1\le j\le p_n} d_{jj}^{-1} = O(\log n)$, where $\nu_{n,u} = M_{p_n}^2(\log p_n/n)^{\frac{1}{2}} n^{\frac{\delta}{2}}$, $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix, and $b_n = \max\{\log n, n^{\kappa_3+\kappa_4-\kappa_0}(\log n)^2\}$.

Condition 5. Assume that $\boldsymbol{D} \in \mathcal{G}_u(K_{p_n}, M_{p_n})$. There exist positive constants $0 < \kappa_5 < 1/4$ and M_2 such that $E\{\exp(\kappa_6 X_j^2)\} \le M_2$ for all $|\kappa_6| \le \kappa_5$ and all $j = 1, \ldots, p_n$.

Condition 6. For some constants $0 < \delta < \min\{1-\kappa_2, 2\kappa_0\}$ and $0 < \kappa_3, \kappa_4 < 1/2$ such that $\kappa_3 + \kappa_4 < 1/2$, $\|\hat{\boldsymbol{C}}_n^{21}\|_\infty = O_p^{(\delta)}(n^{\kappa_3})$, $\lambda_{\min}(\hat{\boldsymbol{C}}_n^{11}) > (a-1)^{-1}$, $\|(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty = O_p^{(\delta)}(n^{\kappa_4}\sqrt{\log p_n/n})$, $\nu_{n,u} n^{\kappa_3}/\sqrt{\log n} = o(1)$, $p'_{SCAD,\lambda_n^*}(h_{\min}) = o(n^{-\kappa_4})$, $1/\lambda_n = O(n^{\kappa_0}\log^{-2} n)$, and $\max_{1\le j\le p_n} d_{jj}^{-1} = O(\log n)$.

Conditions 4 and 6 are similar. They both require that $\hat{\boldsymbol{C}}_n^{11}$ is invertible, and also control the norms of $(\hat{\boldsymbol{C}}_n^{11})^{-1}$, $\hat{\boldsymbol{C}}_n^{21}$, and $\boldsymbol{d}$, which could be satisfied for exchangeable or AR-1 structures. The lower bounds for eigenvalues of $\hat{\boldsymbol{C}}_n^{11}$ in Conditions 4 and 6 are to ensure strict local minimizers of $L_{SCAD}(\boldsymbol{\gamma}, \hat{\boldsymbol{d}})$ and $L_{Lasso}(\boldsymbol{\gamma}, \hat{\boldsymbol{d}})$ in (3.5) and (3.3), respectively. On the other hand, Condition 4 is for consistency of the SPAC-Lasso, while Condition 6 is for consistency of the SPAC-SCAD. The tuning parameter in $p'_{SCAD,\lambda_n^*}(h_{\min})$ is rescaled to the size of $\boldsymbol{\beta}$, and the condition on $p'_{SCAD,\lambda_n^*}(h_{\min})$ holds when $\lambda_n \ll h_{\min}$. Condition 5 comes from Cai et al. (2011) to ensure consistency of the CLIME.

S2 Consistency for fixed p and q

In this subsection, we assume that p, q, $\boldsymbol{\beta}$, $\boldsymbol{C}$, and $\boldsymbol{\gamma}$ are all constant as $n \to \infty$, so that we can use the inverse of the sample covariance matrix to estimate diagonal elements $\boldsymbol{d}$. For each $1 \le j \le p$, let

$$\hat{d}_{jj} = [(\boldsymbol{X}^T\boldsymbol{X}/n)^{-1}]_{jj} \tag{S2.1}$$

be the j-th diagonal element in $(\boldsymbol{X}^T\boldsymbol{X}/n)^{-1}$. Moreover, we assume the following regularity condition on the existence of the fourth moment of the covariates:

Condition 7. The $E(X_j^4)$ is finite for $j = 1, \ldots, p$.

Condition 7 is automatically satisfied under normality or when the tail distribution is well behaved, for example, the sub-Gaussian distribution.

In the following, we show that the estimator $\hat{\boldsymbol{d}}$ approximates $\boldsymbol{d}$ with a certain rate in Lemma 2, which is useful in establishing the strong sign consistency of the proposed SPAC-Lasso in Theorem 3.

Lemma 2. *Suppose that* $\lambda_n/\sqrt{n \log n} \to \infty$. *With* $\hat{d}_{jj}$ *for* $1 \le j \le p$ *defined in (S2.1), we have* $\max_{1 \le j \le p} |\hat{d}_{jj} - d_{jj}| = O(\lambda_n/n)$ *with probability at least* $1 - O(1/\sqrt{n})$.

Theorem 3. *For* p, q, *and* $\boldsymbol{\gamma}$ *independent of* n, *under the regularity Condition 7, suppose that* $\lambda_n/n \to 0$ *and* $\lambda_n/\sqrt{n \log n} \to \infty$, *and that Condition 1 holds with probability at least* $1 - O(1/\sqrt{n})$ *for some* $0 < \delta < 1/2$. *Then the SPAC-Lasso with* $\hat{d}_{jj}$ $(1 \le j \le p)$ *defined in (S2.1) is strongly sign consistent, that is,*

$$P(\hat{\boldsymbol{\gamma}}_{Lasso}(\lambda_n, \hat{\boldsymbol{d}}) =_s \boldsymbol{\gamma}) \ge 1 - O(1/\sqrt{n}).$$

Let an estimator $\hat{\boldsymbol{\gamma}}$ be **general sign consistent** if

$$\lim_{n \to \infty} P\left\{\text{there exists } \lambda \ge 0 \text{ such that } \hat{\boldsymbol{\gamma}}(\lambda) =_s \boldsymbol{\gamma}\right\} = 1.$$

By definition, the strongly sign consistency implies general sign consistency. We also define the following condition which is slightly weaker than Condition 1.

Condition 8 (Weak irrepresentable condition for SPAC-Lasso)**.** Each element in

$$\left|\boldsymbol{V}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{V}(1)^{-1}\,\mathrm{sign}(\boldsymbol{\beta}(1))\right|$$

is less than 1 for sufficiently large n, where $|\cdot|$ represents taking the absolute value of each entry.

Theorem 4. *For fixed p, q, $\boldsymbol{C}$ and $\boldsymbol{\gamma}$, the SPAC-Lasso with $\hat{d}_{jj}$ $(1 \leq j \leq p)$ defined in (S2.1) is general sign consistent only if Condition 8 holds with probability tending to* 1.

Theorem 3 indicates that the probability of the SPAC-Lasso selecting the true model approaches 1, if Condition 1 (the irrepresentable condition for SPAC-Lasso) and regularity Condition 7 hold. Theorem 4 states that Condition 8 is necessary for general sign consistency. Thus, Conditions 1 and 8 are nearly necessary and sufficient for the sign consistency of the proposed SPAC-Lasso in general.

For strong sign consistency of the proposed SPAC-SCAD, we assume the following Condition 9, which is a typical condition of λ_n for the SCAD penalty, and can be satisfied when $\lambda_n \to 0$.

Condition 9. The SCAD penalty satisfies that $\max_j\{p'_{SCAD,\lambda_n}(|\gamma_j|) : \gamma_j \neq 0\} = O(n^{-1/2+\delta})$ for some $0 < \delta < 1/2$ and $\max_j\{p''_{SCAD,\lambda_n}(|\gamma_j|) : \gamma_j \neq 0\} \to 0$ as $n \to \infty$.

Theorem 5. *For p, q, $\boldsymbol{C}$ and $\boldsymbol{\gamma}$ independent of n, let $\hat{d}_{jj}$ $(1 \leq j \leq p)$ be defined in (S2.1) and suppose that $\sqrt{n}\lambda_n/\sqrt{\log n} \to \infty$. Under the regularity Conditions 7 and 9, we have the following properties for the minimization of $L_{SCAD}(\boldsymbol{\gamma}, \hat{\boldsymbol{d}})$ in (3.5) with probability at least $1 - O(1/\sqrt{n})$:*

(1) Estimation consistency: There exist a positive constant K_0 and a local minimizer $\hat{\boldsymbol{\gamma}}_{SCAD}(\lambda_n, \hat{\boldsymbol{d}})$ such that

$$\|\hat{\boldsymbol{\gamma}}_{SCAD} - \boldsymbol{\gamma}\|_2 \leq K_0 n^{-1/2+\delta}.$$

The corresponding estimator of coefficients $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{\gamma}}_{SCAD}$ *satisfies*

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 \leq K_0 n^{-1/2+\delta},$$

where $\hat{\boldsymbol{V}}^{-1} = diag\{\sqrt{\hat{d}_{11}}, \ldots, \sqrt{\hat{d}_{pp}}\}$ *is a diagonal matrix.*

(2) Strong sign consistency: $\hat{\boldsymbol{\gamma}}_{SCAD} =_s \boldsymbol{\gamma}$.

Theorem 5 demonstrates that the proposed SPAC-SCAD possesses both estimation consistency and strong sign consistency under Conditions 7 and 9. The SPAC-SCAD estimator is almost root-n consistent and selects the true model with probability tending to 1. The above theorem does not require any irrepresentable condition since the dimension of coefficients is fixed and the SCAD penalty becomes a constant when the coefficients are large compared to the tuning parameter.

S3 Additional tables

Table 7: Probes selected by the Lasso, SPAC-Lasso, ALasso, SPAC-Lasso based on all samples.

Method	Selected probes
Lasso	ILMN_2329927, ILMN_1794782, ILMN_1684271, ILMN_1760143, ILMN_1741572, ILMN_1664449, ILMN_1815306, ILMN_2398847, ILMN_1772527, ILMN_3187771, ILMN_2413318, **ILMN_1716728**, ILMN_1793508, ILMN_2139035, ILMN_1682259, ILMN_1783815, ILMN_1762316, ILMN_3247835, **ILMN_3307729**, ILMN_1783337, ILMN_1655930, ILMN_2413251, ILMN_1670134, ILMN_1654851, ILMN_1778240, ILMN_1791222, ILMN_1681802, ILMN_2356574, ILMN_1793201, ILMN_2413537, ILMN_1913676, ILMN_1847402, ILMN_1845157, ILMN_1915345, ILMN_1881192, ILMN_1728676, ILMN_1811507, ILMN_1664294, ILMN_1670693, ILMN_3187362, ILMN_3237534, ILMN_3246424, ILMN_3199438, ILMN_1657022, ILMN_3230260, ILMN_1656111, ILMN_1797332, ILMN_1676289, ILMN_2295879, ILMN_1678032, ILMN_2328378, ILMN_3230157, ILMN_1693259, ILMN_3243705, ILMN_1769751, ILMN_1774836, ILMN_1727127, ILMN_2346573, ILMN_2127328, ILMN_2209766, ILMN_1739423, ILMN_1747192, ILMN_1786972, ILMN_2148679, ILMN_2364535, ILMN_2323302, ILMN_1684802, ILMN_2386269, ILMN_1719039, ILMN_2360291, **ILMN_3248844**, ILMN_1678919, ILMN_1654689, ILMN_2399686,
SPAC-Lasso	ILMN_2263466, ILMN_1760143, ILMN_1741572, ILMN_1815306, ILMN_1697827, ILMN_1772527, ILMN_3187771, ILMN_2413318, **ILMN_1716728**, ILMN_1793508, ILMN_2139035, ILMN_2244484, ILMN_1682259, ILMN_1783815, ILMN_1762316, ILMN_3247835, **ILMN_3307729**, ILMN_1783337, ILMN_1811650, ILMN_1655930, ILMN_2413251, ILMN_1670134, ILMN_1654851, ILMN_1778240, ILMN_1791222, ILMN_2356574, ILMN_1793201, ILMN_2413537, ILMN_1913676, ILMN_1847402, ILMN_1845157, ILMN_1915345, ILMN_1881192, ILMN_1768483, ILMN_1728676, ILMN_1811507, ILMN_1664294, ILMN_1670693, ILMN_3187362, ILMN_3237534, ILMN_3246424, ILMN_3292551, ILMN_3199438, ILMN_1753518, ILMN_1676100, ILMN_1657022, ILMN_3230260, ILMN_1656111, ILMN_1797332, ILMN_1676289, ILMN_2295879, ILMN_1678032, ILMN_2328378, ILMN_3230157, ILMN_1693259, ILMN_3243705, ILMN_1769751, ILMN_1774836, ILMN_1727127, ILMN_2346573, ILMN_2127328, ILMN_2209766, ILMN_1747192, ILMN_1786972, ILMN_2148679, ILMN_2364535, ILMN_2323302, ILMN_1813491, **ILMN_1663035**, ILMN_1684802, ILMN_2386269, ILMN_2360291, **ILMN_3248844**, ILMN_1678919, ILMN_1654689, ILMN_2399686, ILMN_1654357, ILMN_1683854
ALasso	ILMN_1760143, **ILMN_1716728**, ILMN_1793508, ILMN_1682259, **ILMN_3307729**, ILMN_1670134, ILMN_1778240, ILMN_1793201, ILMN_1845157, ILMN_1881192, ILMN_1811507, ILMN_3187362, ILMN_3237534, ILMN_3246424, ILMN_1656111, ILMN_1797332, ILMN_2295879, ILMN_1769751, ILMN_1774836, ILMN_2209766, **ILMN_3248844**
SPAC-ALasso	ILMN_1760143, **ILMN_1716728**, ILMN_1682259, ILMN_3247835, **ILMN_3307729**, ILMN_1655930, ILMN_1670134, ILMN_1778240, ILMN_1793201, ILMN_1881192, ILMN_1811507, ILMN_3237534, ILMN_3246424, ILMN_1656111, ILMN_1797332, ILMN_2295879, ILMN_1693259, ILMN_1769751, ILMN_1774836, ILMN_2209766, ILMN_1813491, **ILMN_1663035**, **ILMN_3248844**, ILMN_1654357

Table 8: Probes selected by the SCAD, SPAC-SCAD, and PC-simple algorithm based on all samples.

Method	Selected probes
SCAD	ILMN_2307883, ILMN_2157709, ILMN_1783023, **ILMN_1716728**, ILMN_1776337, ILMN_1682259, ILMN_2136177, **ILMN_3307729**, ILMN_1670134, ILMN_1805216, ILMN_1793201, ILMN_3270641, ILMN_1836958, ILMN_1811507, ILMN_1715814, ILMN_3187362, ILMN_3294033, ILMN_1653573, ILMN_1734427, ILMN_1803799, ILMN_1715175, ILMN_1793349, ILMN_1656111, ILMN_1675239, ILMN_1672080, ILMN_1802628, ILMN_3225534, ILMN_1656791, ILMN_2148679, ILMN_1763989, ILMN_2086238, ILMN_1659761, ILMN_2194229, ILMN_1672004, ILMN_1664175, **ILMN_3248844**, ILMN_1749809
SPAC-SCAD	ILMN_1719498, ILMN_2306955, ILMN_1760143, ILMN_2124155, ILMN_3230880, ILMN_3240538, **ILMN_1716728**, ILMN_2244484, ILMN_1699610, ILMN_1753468, ILMN_1737195, ILMN_1776337, ILMN_1682259, **ILMN_3307729**, ILMN_1738272, ILMN_1670134, ILMN_1804248, ILMN_2268921, ILMN_1793201, ILMN_1656977, ILMN_1847402, ILMN_1811507, ILMN_1715814, ILMN_3187362, ILMN_1670570, ILMN_1656111, ILMN_1675239, ILMN_1663437, ILMN_3245983, ILMN_2377862, ILMN_2148679, ILMN_1654637, ILMN_1721563, ILMN_2203147, **ILMN_3248844**, ILMN_1751963, ILMN_2399686
Farm-Select	ILMN_2307883, ILMN_2413318, **ILMN_1716728**, ILMN_1670926, ILMN_1682259, **ILMN_3307729**, ILMN_1783337, ILMN_1780601, ILMN_1670134, ILMN_1815668, ILMN_1793201, ILMN_1846499, ILMN_1845157, ILMN_1863939, ILMN_1728676, ILMN_1811507, ILMN_1715814, ILMN_3199438, ILMN_1657022, ILMN_1658015, ILMN_1715175, ILMN_1656111, ILMN_1797332, ILMN_2295879, ILMN_2048822, ILMN_1727127, ILMN_1739583, ILMN_2148679, ILMN_1763989, ILMN_1654637, ILMN_2364535, ILMN_1813491, ILMN_1684802, ILMN_1749403, ILMN_2204726, **ILMN_3248844**, ILMN_1706342
PC-simple	ILMN_1664449, **ILMN_1716728**, **ILMN_3307729**, ILMN_3237534, **ILMN_3248844**

S4 Additional corollary and proofs

Corollary 3. *Let $\hat{\boldsymbol{d}}$ be diagonal elements of the CLIME of $\boldsymbol{D}$. Suppose that $h_{\min} \geq n^{-\kappa_0}$. Under the conditions of Proposition 1 and Condition 6, if there exists a positive constant η such that*

$$|\alpha_2| \leq (1-\eta)\sqrt{\frac{1-\alpha_1}{1-\alpha_3}}\alpha_1/\mathcal{P}_{\lambda_n^*}(h_{\min}), \tag{S4.2}$$

then the SPAC-SCAD possesses estimation consistency and strong sign consistency with sufficiently large q_0 and $p_0 - q_0$.

Proof of Lemma 1

Proof. It can be calculated that $\sigma^{jj} = d_{jj} + \beta_j^2/\sigma_\varepsilon^2$, $\sigma^{yy} = 1/\sigma_\varepsilon^2$, and $\sigma^{jy} = \beta_j/\sigma_\varepsilon^2$ for $j = 1, \ldots, p$. Then,

$$\sigma^{jj}\sigma^{yy} - (\sigma^{jy})^2 = (d_{jj} + \frac{\beta_j^2}{\sigma_\varepsilon^2})\frac{1}{\sigma_\varepsilon^2} - (\frac{\beta_j}{\sigma_\varepsilon^2})^2 = \frac{d_{jj}}{\sigma_\varepsilon^2} = d_{jj}\sigma^{yy}.$$

The jth SPAC is

$$\rho_j s_j = \frac{-\sigma^{jy}}{\sqrt{\sigma^{jj}\sigma^{yy}}}\sqrt{\frac{\sigma^{jj}}{\sigma^{jj}\sigma^{yy} - (\sigma^{jy})^2}} = \frac{-\sigma^{jy}}{\sigma^{yy}\sqrt{d_{jj}}} = \frac{\beta_j}{\sqrt{d_{jj}}} = \gamma_j.$$

The standard deviation of the response conditional on $\boldsymbol{X}_{-j}$ is

$$s_j^2 = \frac{\sigma^{jj}}{\sigma^{jj}\sigma^{yy} - (\sigma^{jy})^2} = \frac{d_{jj} + \beta_j^2\sigma^{yy}}{d_{jj}\sigma^{yy}} = \sigma_\varepsilon^2 + \frac{\beta_j^2}{d_{jj}}.$$

□

Lemma 3. *Suppose that $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n$ are k-dimensional i.i.d. random vectors with mean $\boldsymbol{\theta}_w$ and covariance matrix $\boldsymbol{\Sigma}_w$. Further suppose that $E(\|\boldsymbol{W}_i\|^3)$ is finite for any $1 \le i \le n$.*

(1) We have

$$\sqrt{n}\left(\widetilde{\boldsymbol{W}}_n - \boldsymbol{\theta}_w\right) \xrightarrow{d} N_k\left(\mathbf{0}, \boldsymbol{\Sigma}_w\right), \tag{S4.3}$$

where $\widetilde{\boldsymbol{W}}_n = \sum_{i=1}^n \boldsymbol{W}_i/n$. Moreover,

$$\sup_A \left| P\left(\sqrt{n}\left(\widetilde{\boldsymbol{W}}_n - \boldsymbol{\theta}_w\right) \in A\right) - P\left(\boldsymbol{Z} \in A\right)\right| = O\left(\frac{1}{\sqrt{n}}\right), \tag{S4.4}$$

where $\boldsymbol{Z} \sim N_k(\mathbf{0}, \boldsymbol{\Sigma}_w)$ and the supremum is taken over all measurable convex sets A.

(2) For any continuous function $g : \mathbb{R}^k \to \mathbb{R}$ satisfying the properties that $\nabla g(\boldsymbol{\theta}_w) \neq \mathbf{0}$ and that $\|\nabla^2 g(\boldsymbol{\theta})\|_\infty$ is continuous at $\boldsymbol{\theta}_w$, we have

$$\sqrt{n}\left(g(\widetilde{\boldsymbol{W}}_n) - g(\boldsymbol{\theta}_w)\right) \xrightarrow{d} N\left(0, \{\nabla g(\boldsymbol{\theta}_w)\}^T \boldsymbol{\Sigma}_w \nabla g(\boldsymbol{\theta}_w)\right), \tag{S4.5}$$

and

$$\sup_{x\in\mathbb{R}} \left| P\left(\sqrt{n}\left(g(\widetilde{\boldsymbol{W}}_n) - g(\boldsymbol{\theta}_w)\right) \le x\right) - P\left(Z_g \le x\right)\right| = O\left(\frac{1}{\sqrt{n}}\right), \tag{S4.6}$$

where $Z_g \sim N(0, \{\nabla g(\boldsymbol{\theta}_w)\}^T \boldsymbol{\Sigma}_w \nabla g(\boldsymbol{\theta}_w))$.

(3) Suppose that $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ *are* l*-dimensional identically-distributed random vectors with mean* $\boldsymbol{\theta}_u$ *and covariance matrix* $\boldsymbol{\Sigma}_u$ *such that* $(\boldsymbol{U}_1^T, \boldsymbol{W}_1^T)^T, \ldots, (\boldsymbol{U}_n^T, \boldsymbol{W}_n^T)^T$ *are i.i.d. vectors. Let* $h : \mathbb{R}^l \to \mathbb{R}$ *be a continuous function such that* $\nabla h(\boldsymbol{\theta}_u) \neq \mathbf{0}$ *and that* $\|\nabla^2 h(\boldsymbol{\theta})\|_\infty$ *is continuous at* $\boldsymbol{\theta}_u$*. Then,*

$$h(\widetilde{\boldsymbol{U}}_n) \xrightarrow{p} h(\boldsymbol{\theta}_u), \tag{S4.7}$$

$$\sqrt{n}\ h(\widetilde{\boldsymbol{U}}_n) \left(g(\widetilde{\boldsymbol{W}}_n) - g(\boldsymbol{\theta}_w) \right) \xrightarrow{d} h(\boldsymbol{\theta}_u) N\left(0, \{\nabla g(\boldsymbol{\theta}_w)\}^T \boldsymbol{\Sigma}_w \nabla g(\boldsymbol{\theta}_w)\right), \tag{S4.8}$$

and

$$\sup_{x \in \mathbb{R}} \left| P\left(\sqrt{n}\ h(\widetilde{\boldsymbol{U}}_n) \left(g(\widetilde{\boldsymbol{W}}_n) - g(\boldsymbol{\theta}_w) \right) \leq x \right) - P\left(Z_{hg} \leq x \right) \right| = O\left(\frac{1}{\sqrt{n}} \right), \tag{S4.9}$$

where $\widetilde{\boldsymbol{U}}_n = \sum_{i=1}^n \boldsymbol{U}_i / n$*, and* $Z_{hg} \sim h(\boldsymbol{\theta}_u) N(0, \{\nabla g(\boldsymbol{\theta}_w)\}^T \boldsymbol{\Sigma}_w \nabla g(\boldsymbol{\theta}_w))$.

Proof. (1) The (S4.3) follows from the multivariate central limit theorem. The (S4.4) follows from (Gotze, 1991, Theorem 1.3).

(2) The (S4.5) follows from the delta method. The (S4.6) follows from (Pinelis and Molzon, 2016, Theorem 2.9).

(3) The (S4.7) and (S4.8) follows from the weak law of large numbers and Slutsky's theorem.

Let $\bar{g}(\widetilde{\boldsymbol{U}}_n, \widetilde{\boldsymbol{W}}_n) = h(\widetilde{\boldsymbol{U}}_n) \left(g(\widetilde{\boldsymbol{W}}_n) - g(\boldsymbol{\theta}_w) \right)$. We apply the multivariate central limit theorem to the sequence $\{(\boldsymbol{U}_i^T, \boldsymbol{W}_i^T)^T\}$, and then apply the Delta method to $\bar{g}(\widetilde{\boldsymbol{U}}_n, \widetilde{\boldsymbol{W}}_n)$. Thus, based on (S4.6), we can obtain (S4.9). □

Proof of Lemma 2

Proof. Let $\boldsymbol{v}$ be a vector containing all the elements in upper-triangular part of the true covariance $\boldsymbol{C}$, and $\hat{\boldsymbol{v}}_i$ be a vector containing corresponding sample covariance estimator only using the i-th sample. Since samples are i.i.d, $\hat{\boldsymbol{v}}_i$ for $1 \leq i \leq n$ are i.i.d. Let $\hat{\boldsymbol{v}} = \sum_{i=1}^n \hat{\boldsymbol{v}}_i/n$. Then $\hat{\boldsymbol{v}}$ contains elements in upper-triangular part of $\hat{\boldsymbol{C}}$. By the multivariate central limit theorem,

$$\sqrt{n}(\hat{\boldsymbol{v}} - \boldsymbol{v}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{v}}).$$

where $\boldsymbol{\Sigma}_{\boldsymbol{v}} = \mathrm{Var}(\hat{\boldsymbol{v}}_i)$.

There exists a continuous function $g_j(\boldsymbol{v})$, such that $g_j(\boldsymbol{v}) = d_{jj}$ for $1 \leq j \leq p$. Then $\hat{d}_{jj} = g_j(\hat{\boldsymbol{v}})$. Let $\nabla g_j(\boldsymbol{v})$ be the gradient of g. Since $\boldsymbol{C}$ is positive definite, $\nabla g_j(\boldsymbol{v}) \neq \boldsymbol{0}$. By the delta method,

$$\sqrt{n}(\hat{d}_{jj} - d_{jj}) = \sqrt{n}\{g(\hat{\boldsymbol{v}}) - g(\boldsymbol{v})\} \xrightarrow{d} N(\boldsymbol{0}, \nabla g_j(\boldsymbol{v})^T \boldsymbol{\Sigma}_{\boldsymbol{v}} \nabla g_j(\boldsymbol{v})).$$

Since $\lambda_n/\sqrt{n \log n} \to \infty$, by Lemma 3, we have $P(|\hat{d}_{jj} - d_{jj}| \geq \lambda_n/n) = O(1/\sqrt{n})$ for any $0 < \delta < 1/2$, which implies that $P(\max_{1 \leq j \leq p} |\hat{d}_{jj} - d_{jj}| \geq \lambda_n/n) \leq \sum_{j=1}^p P(|\hat{d}_{jj} - d_{jj}| \geq \lambda_n/n) = O(1/\sqrt{n})$. This completes the proof. □

Proof of Theorem 3

Proof. The proposed estimator with Lasso penalty is

$$\hat{\boldsymbol{\gamma}}_{Lasso} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\boldsymbol{y} - \sum_{j=1}^p \boldsymbol{X}_j \sqrt{\hat{d}_{jj}} \gamma_j \|^2 + \lambda_n \sum_{j=1}^p \hat{d}_{jj} |\gamma_j| \right\}.$$

Let $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1, \ldots, \bar{\beta}_p)$ and $\bar{\boldsymbol{\gamma}} = (\bar{\gamma}_1, \ldots, \bar{\gamma}_p)$ be the true values of the $\boldsymbol{\beta}$ and SPACs, respectively.

Then $\bar{\boldsymbol{\beta}} = \boldsymbol{V}^{-1}\bar{\boldsymbol{\gamma}}$. Denote $\hat{\boldsymbol{u}} = \hat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{\gamma}}_{Lasso} - \boldsymbol{V}^{-1}\bar{\boldsymbol{\gamma}} = \hat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{\gamma}}_{Lasso} - \bar{\boldsymbol{\beta}}$. Then

$$\begin{aligned}
\hat{\boldsymbol{u}} &= \underset{\boldsymbol{u}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\boldsymbol{y} - \sum_{j=1}^{p} \boldsymbol{X}_j(u_j + \bar{\beta}_j)\|^2 + \lambda_n \sum_{j=1}^{p} \sqrt{\hat{d}_{jj}}|u_j + \bar{\beta}_j| \right\} \\
&= \underset{\boldsymbol{u}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\boldsymbol{\varepsilon} - \sum_{j=1}^{p} \boldsymbol{X}_j u_j\|^2 + \lambda_n \sum_{j=1}^{p} \sqrt{\hat{d}_{jj}}|u_j + \bar{\beta}_j| \right\} \\
&= \underset{\boldsymbol{u}}{\operatorname{argmin}} \left\{ \frac{1}{2}\boldsymbol{u}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{u} - \sqrt{n}\boldsymbol{w}^T\boldsymbol{u} + \lambda_n \sum_{j=1}^{p} \sqrt{\hat{d}_{jj}}|u_j + \bar{\beta}_j| \right\},
\end{aligned}$$

where $\boldsymbol{w} = \boldsymbol{X}^T\boldsymbol{\varepsilon}/\sqrt{n}$. Notice that

$$\frac{d[\frac{1}{2}\boldsymbol{u}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{u} - \sqrt{n}\boldsymbol{w}^T\boldsymbol{u}]}{d\boldsymbol{u}} = \sqrt{n}(\sqrt{n}\hat{\boldsymbol{C}}\boldsymbol{u} - \boldsymbol{w}). \tag{S4.10}$$

Let $\hat{\boldsymbol{u}}(1)$, $\boldsymbol{w}(1)$, $\bar{\boldsymbol{\beta}}(1)$, and $\hat{\boldsymbol{u}}(2)$, $\boldsymbol{w}(2)$, $\bar{\boldsymbol{\beta}}(2)$ be the first q and last $p-q$ entries of $\hat{\boldsymbol{u}}$, $\boldsymbol{w}$, and $\bar{\boldsymbol{\beta}}$, respectively. Then, based on the Karush-Kuhn-Tucker (KKT) conditions, if there exists $\hat{\boldsymbol{u}}$ such that

$$\sqrt{n}\hat{\boldsymbol{V}}(1)\hat{\boldsymbol{C}}^{11}\hat{\boldsymbol{u}}(1) - \hat{\boldsymbol{V}}(1)\boldsymbol{w}(1) = -\frac{\lambda_n}{\sqrt{n}}\operatorname{sign}(\bar{\boldsymbol{\beta}}(1)), \tag{S4.11}$$

$$|\hat{\boldsymbol{u}}(1)| < |\bar{\boldsymbol{\beta}}(1)|,$$

$$|\sqrt{n}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}\hat{\boldsymbol{u}}(1) - \hat{\boldsymbol{V}}(2)\boldsymbol{w}(2)| \leq \frac{\lambda_n}{\sqrt{n}}\mathbf{1},$$

then $\operatorname{sign}(\hat{\boldsymbol{\gamma}}_{Lasso}(1)) = \operatorname{sign}(\bar{\boldsymbol{\beta}}(1)) = \operatorname{sign}(\bar{\boldsymbol{\gamma}}(1))$ and $\hat{\boldsymbol{\gamma}}_{Lasso}(2) = \hat{\boldsymbol{u}}(2) = \mathbf{0}$.

Take (S4.11) as a definition of $\hat{\boldsymbol{u}}(1)$. Then the existence of such $\hat{\boldsymbol{u}}$ is implied by the following inequalities:

$$|(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1)| < \sqrt{n}(|\bar{\boldsymbol{\beta}}(1)| - \frac{\lambda_n}{n}|(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\operatorname{sign}(\bar{\boldsymbol{\beta}}(1))|), \tag{S4.12}$$

$$|\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1) - \hat{\boldsymbol{V}}(2)\boldsymbol{w}(2)| < \frac{\lambda_n}{\sqrt{n}}(\mathbf{1} - |\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\operatorname{sign}(\bar{\boldsymbol{\beta}}(1))|). \tag{S4.13}$$

Let A_n denote the event that (S4.12) holds, and B_n be the event that (S4.13) holds. Then

$$P(\hat{\boldsymbol{\gamma}}_{Lasso}(\lambda_n, \hat{\boldsymbol{d}}) =_s \bar{\boldsymbol{\gamma}}) \geq P(A_n \cap B_n).$$

Denote $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q) = (\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1)$, and $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{p-q}) = \hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1) - \hat{\boldsymbol{V}}(2)\boldsymbol{w}(2)$. Let $\boldsymbol{b} = (b_1, \ldots, b_q) = |(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\,\mathrm{sign}(\bar{\boldsymbol{\gamma}}(1))|$ and $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \ldots, \hat{\eta}'_{p-q})^T = \mathbf{1} - |\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\,\mathrm{sign}(\bar{\boldsymbol{\beta}}(1))|$. Then

$$\begin{aligned} 1 - P(A_n \cap B_n) &\leq P(A_n^c) + P(B_n^c) \\ &\leq \sum_{j=1}^{q} P(|\phi_j| \geq \sqrt{n}(|\bar{\boldsymbol{\beta}}(1)| - \frac{\lambda_n}{n} b_j)) + \sum_{j=1}^{p-q} P(|\zeta_j| \geq \frac{\lambda_n}{\sqrt{n}} \eta'_j). \end{aligned}$$

By the multivariate central limit theorem,

$$\boldsymbol{X}^T\boldsymbol{\varepsilon}/\sqrt{n} \xrightarrow{d} N(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{C}).$$

Since $\hat{\boldsymbol{V}}(1)(\hat{\boldsymbol{C}}^{11})^{-1} \xrightarrow{p} \boldsymbol{V}(1)(\boldsymbol{C}^{11})^{-1}$, $\hat{\boldsymbol{V}}(2) \xrightarrow{p} \boldsymbol{V}(2)$, and $\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1} \xrightarrow{p} \boldsymbol{V}(2)\boldsymbol{C}^{21}(\boldsymbol{C}^{11})^{-1}$, by Slutsky's theorem,

$$(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1) \xrightarrow{d} N(\mathbf{0}, \sigma_\varepsilon^2(\boldsymbol{C}^{11})^{-1}),$$

$$\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1) - \hat{\boldsymbol{V}}(2)\boldsymbol{w}(2) \xrightarrow{d} N(\mathbf{0}, \sigma_\varepsilon^2\boldsymbol{V}(2)(\boldsymbol{C}^{22} - \boldsymbol{C}^{21}(\boldsymbol{C}^{11})^{-1}\boldsymbol{C}^{12})\boldsymbol{V}(2)). \quad \text{(S4.14)}$$

Hence, ϕ_j and ζ_j converge in distribution to Gaussian random variables for each j.

By Lemma 2 and Lemma 3, with probability at least $1 - O(1/\sqrt{n})$ for $0 < c < 1/2$, there exists a constant ϵ_c such as $\max_{1\leq j\leq p} \hat{d}_{jj} \leq \epsilon_c$, $\min_{1\leq j\leq p} \hat{d}_{jj} \geq 1/\epsilon_c$,

$$\max_{1\leq j\leq p} |\sqrt{\hat{d}_{jj}} - \sqrt{d_{jj}}| \leq \epsilon_c\lambda_n/n, \quad \text{and} \quad \max_{1\leq j\leq p} |1/\sqrt{\hat{d}_{jj}} - 1/\sqrt{d_{jj}}| \leq \epsilon_c\lambda_n/n. \quad \text{(S4.15)}$$

We have

$$\begin{aligned}\|\boldsymbol{b}\|_\infty &= \|(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\,\mathrm{sign}(\bar{\boldsymbol{\gamma}}(1))\|_\infty \\ &\le \|\{(\hat{\boldsymbol{C}}^{11})^{-1}-(\boldsymbol{C}^{11})^{-1}\}\hat{\boldsymbol{V}}^{-1}(1)\,\mathrm{sign}(\bar{\boldsymbol{\gamma}}(1))\|_\infty \\ &\quad +\|(\boldsymbol{C}^{11})^{-1}\{\hat{\boldsymbol{V}}^{-1}(1)-\boldsymbol{V}^{-1}(1)\}\,\mathrm{sign}(\bar{\boldsymbol{\gamma}}(1))\|_\infty \\ &\quad +\|(\boldsymbol{C}^{11})^{-1}\boldsymbol{V}^{-1}(1)\,\mathrm{sign}(\bar{\boldsymbol{\gamma}}(1))\|_\infty.\end{aligned} \tag{S4.16}$$

Since the elements of sample covariance matrix $\hat{\boldsymbol{C}}$ converge to the corresponding true covariances with probability at least $1-O(1/\sqrt{n})$, $\|(\hat{\boldsymbol{C}}^{11})^{-1}-(\boldsymbol{C}^{11})^{-1}\|_\infty$ converges to 0 with probability at least $1-O(1/\sqrt{n})$. Since $\lambda_n/n \to 0$, the $\lambda_n b_j/n$ converges to 0 with probability at least $1-O(1/\sqrt{n})$. We also have

$$\begin{aligned}\hat{\boldsymbol{\eta}} &= \mathbf{1}-|\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\,\mathrm{sign}(\bar{\boldsymbol{\beta}}(1))| \\ &\ge \mathbf{1}-|\{\hat{\boldsymbol{V}}(2)-\boldsymbol{V}(2)\}\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\,\mathrm{sign}(\bar{\boldsymbol{\beta}}(1))| \\ &\quad -|\boldsymbol{V}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\{\hat{\boldsymbol{V}}^{-1}(1)-\boldsymbol{V}^{-1}(1)\}\,\mathrm{sign}(\bar{\boldsymbol{\beta}}(1))| \\ &\quad -|\boldsymbol{V}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{V}^{-1}(1)\,\mathrm{sign}(\bar{\boldsymbol{\beta}}(1))|\end{aligned} \tag{S4.17}$$

The inequality in (S4.17) is element-wise. By (S4.15) and Condition 1, $\hat{\eta}_j > \eta/2$ holds with probability at least $1-O(1/\sqrt{n})$ for each j. Based on $\lambda_n/\sqrt{n\log n} \to \infty$ and Gaussian distributions of ϕ_j and ζ_j, we have

$$\sum_{j=1}^{q} P(|\phi_j| \ge \frac{\sqrt{n}}{2}(|\bar{\boldsymbol{\beta}}(1)| - \frac{\lambda_n}{n}b_j)) \le \sum_{j=1}^{q}\left[P(|\phi_j| \ge \frac{\sqrt{n}}{4}|\bar{\boldsymbol{\beta}}(1)|) + O(1/\sqrt{n})\right] = O(1/\sqrt{n}),$$

$$\sum_{j=1}^{p-q} P(|\zeta_j| \ge \frac{\lambda_n}{2\sqrt{n}}\hat{\eta}_j) \le \sum_{j=1}^{p-q} P(|\zeta_j| \ge \frac{\lambda_n \eta}{4\sqrt{n}}) + O(1/\sqrt{n}) = O(1/\sqrt{n}).$$

This completes the proof. □

Proof of Theorem 5

Proof. Let $\bar{\boldsymbol{\gamma}} = (\bar{\gamma}_1, \ldots, \bar{\gamma}_p)$ and $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1, \ldots, \bar{\beta}_p)$ be the true values of the SPACs and $\boldsymbol{\beta}$, respectively, $\alpha_n = n^{-1/2+\delta}$, and

$$\tilde{L}(\boldsymbol{\gamma}, \hat{\boldsymbol{d}}) = \frac{1}{2}\|\boldsymbol{y} - \sum_{j=1}^{p} \boldsymbol{X}_j \sqrt{\hat{d}_{jj}}\gamma_j\|^2.$$

By the definition of $\hat{\boldsymbol{d}}$ and Lemma 3, $P(\max_{1\le j\le p} |\hat{d}_{jj} - d_{jj}| \le a_n) = 1 - O(1/\sqrt{n})$.

(1) It suffices to show that there exists a large constant M such that,

$$\min_{\|\boldsymbol{u}\|_2 = M} L_{SCAD}(\bar{\boldsymbol{\gamma}} + \alpha_n \boldsymbol{u}, \hat{\boldsymbol{d}}) > L_{SCAD}(\bar{\boldsymbol{\gamma}}, \hat{\boldsymbol{d}}), \tag{S4.18}$$

with probability at least $1 - O(1/\sqrt{n})$. Let $D_n = L_{SCAD}(\bar{\boldsymbol{\gamma}} + \alpha_n \boldsymbol{u}, \hat{\boldsymbol{d}}) - L_{SCAD}(\bar{\boldsymbol{\gamma}}, \hat{\boldsymbol{d}})$. Then,

$$D_n \ge \tilde{L}(\bar{\boldsymbol{\gamma}} + \alpha_n \boldsymbol{u}, \hat{\boldsymbol{d}}) - \tilde{L}(\bar{\boldsymbol{\gamma}}, \hat{\boldsymbol{d}}) + n\sum_{j=1}^{q} \hat{d}_{jj}\left[p_{SCAD,\lambda_n}(|\bar{\gamma}_j + \alpha_n u_j|) - p_{SCAD,\lambda_n}(|\bar{\gamma}_j|)\right].$$

By Tayler expansion, we have

$$\begin{aligned} D_n &\ge \alpha_n \tilde{L}'(\bar{\boldsymbol{\gamma}}, \hat{\boldsymbol{d}})^T \boldsymbol{u} + \frac{1}{2}\alpha_n^2 \boldsymbol{u}^T \tilde{L}''(\bar{\boldsymbol{\gamma}}, \hat{\boldsymbol{d}})\boldsymbol{u} \\ &+ n\sum_{j=1}^{q} \hat{d}_{jj}\left[\alpha_n p'_{SCAD,\lambda_n}(|\bar{\gamma}_j|)\,\mathrm{sign}(\bar{\gamma}_j) u_j + \alpha_n^2 p''_{SCAD,\lambda_n}(|\bar{\gamma}_j|) u_j^2\right]. \end{aligned} \tag{S4.19}$$

The second term in (S4.19) is bounded below by $b_{0,1} M^2 \alpha_n^2 n$ with probability at least $1 - O(1/\sqrt{n})$ for some constant $b_{0,1} > 0$. By the multivariate central limit theorem, $|\alpha_n \tilde{L}'(\bar{\boldsymbol{\gamma}}, \hat{\boldsymbol{d}})^T \boldsymbol{u}| < b_{0,1}\alpha_n^2 n/3$, dominated by the second term, with probability at least $1 - O(1/\sqrt{n})$. The third term in (S4.19) is bounded above by

$$b_{0,2}\left[qM\alpha_n n \max_j\{p'_{SCAD,\lambda_n}(|\bar{\gamma}_j|) : \bar{\gamma}_j \ne 0\} + M^2\alpha_n^2 n \max_j\{p''_{SCAD,\lambda_n}(|\bar{\gamma}_j|) : \bar{\gamma}_j \ne 0\}\right],$$

with probability at least $1 - O(1/\sqrt{n})$ for some constant $b_{0,2} > 0$. This term is also dominated by the second term via choosing a sufficiently large M. Thus, the (S4.18) holds.

(2) Let $\boldsymbol{w} = \boldsymbol{X}^T\boldsymbol{\varepsilon}/\sqrt{n}$ and $\hat{\boldsymbol{u}} = \hat{\boldsymbol{V}}^{-1}\hat{\gamma}_{SCAD} - \bar{\boldsymbol{\beta}}$. Similarly as in the proof of Theorem 3, if

$$\sqrt{n}\hat{\boldsymbol{V}}(1)\hat{\boldsymbol{C}}^{11}\hat{\boldsymbol{u}}(1) - \hat{\boldsymbol{V}}(1)\boldsymbol{w}(1) = -\sqrt{n}\lambda_n\hat{\mathcal{G}}\,\mathrm{sign}(\bar{\boldsymbol{\beta}}(1)), \tag{S4.20}$$

$$|\hat{\boldsymbol{u}}(1)| < |\bar{\boldsymbol{\beta}}(1)|, \tag{S4.21}$$

$$|\sqrt{n}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}\hat{\boldsymbol{u}}(1) - \hat{\boldsymbol{V}}(2)\boldsymbol{w}(2)| \leq \sqrt{n}\lambda_n\mathbf{1}, \tag{S4.22}$$

then $\mathrm{sign}(\hat{\boldsymbol{\gamma}}_{SCAD}(1)) = \mathrm{sign}(\bar{\boldsymbol{\gamma}}(1))$ and $\hat{\boldsymbol{\gamma}}_{SCAD}(2) = \hat{\boldsymbol{u}}(2) = \mathbf{0}$, where $\hat{\mathcal{G}}$ is a $q \times q$ diagonal matrix with $(\hat{\mathcal{G}})_{jj} = p'_{SCAD,\sqrt{\hat{d}_{jj}}\lambda_n}(|(\hat{\boldsymbol{u}}(1) + \bar{\boldsymbol{\beta}}(1))_j|)$ for $j = 1, \ldots, q$.

Take (S4.20) as a definition of $\hat{\boldsymbol{u}}(1)$. Then the (S4.21) and (S4.22) are implied by the following inequalities:

$$|(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1)| < \sqrt{n}(|\bar{\boldsymbol{\beta}}(1)| - \lambda_n|(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\hat{\mathcal{G}}\,\mathrm{sign}(\bar{\boldsymbol{\beta}}(1))|), \tag{S4.23}$$

$$|\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1) - \hat{\boldsymbol{V}}(2)\boldsymbol{w}(2)| < \sqrt{n}\lambda_n(\mathbf{1} - |\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\hat{\mathcal{G}}\,\mathrm{sign}(\bar{\boldsymbol{\beta}}(1))|). \tag{S4.24}$$

Let A_n denote the event that (S4.23) holds, and B_n be the event that (S4.24) holds. Then

$$P(\hat{\boldsymbol{\gamma}}_{SCAD}(\lambda_n, \hat{\boldsymbol{d}}) =_s \bar{\boldsymbol{\gamma}}) \geq P(A_n \cap B_n).$$

Similarly as in the proof of Theorem 3, let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q) = (\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1)$, $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{p-q}) = \hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1) - \hat{\boldsymbol{V}}(2)\boldsymbol{w}(2)$, $\boldsymbol{b} = (b_1, \ldots, b_q) = |(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\hat{\mathcal{G}}\,\mathrm{sign}(\bar{\boldsymbol{\gamma}}(1))|$ and $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \ldots, \hat{\eta}_{p-q}) = \mathbf{1} - |\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1)\hat{\mathcal{G}}\,\mathrm{sign}(\bar{\boldsymbol{\gamma}}(1))|$.

By Part (1) in this theorem and $\max_j\{p'_{SCAD,\lambda_n}(|\bar{\gamma}_j|) : \bar{\gamma}_j \neq 0\} = O(n^{-1/2+\delta})$, $\lambda_n = O(1)$ and the diagonal elements of $\hat{\mathcal{G}}$ converge to 0 with probability at least $1 - O(1/\sqrt{n})$. Then, $\lambda_n b_j$ converge to 0, and $\hat{\eta}_j > 1/2$, with probability at least $1 - O(1/\sqrt{n})$ for each j. Since $\sqrt{n}\lambda_n/\sqrt{\log n} \to \infty$, by Lemma 3 and the asymptotic Gaussian distributions of ϕ_j and ζ_j,

$$\sum_{j=1}^{q} P(|\phi_j| \geq \frac{\sqrt{n}}{2}(|\bar{\boldsymbol{\gamma}}(1)| - \lambda_n b_j)) \leq \sum_{j=1}^{q}\left[P(|\phi_j| \geq \frac{\sqrt{n}}{4}|\bar{\boldsymbol{\gamma}}(1)|) + O(1/\sqrt{n})\right] = O(1/\sqrt{n}),$$

$$\sum_{j=1}^{p-q} P(|\zeta_j| \geq \frac{\sqrt{n}\lambda_n}{2}\eta'_j) \leq \sum_{j=1}^{p-q} P(|\zeta_j| \geq \frac{\sqrt{n}\lambda_n}{4}) = O(1/\sqrt{n}).$$

This completes the proof. □

We prove Theorem 2 first, and then prove Theorem 1.

Proof of Theorem 2

Proof. Let $\bar{\boldsymbol{\gamma}}$ and $\bar{\boldsymbol{\beta}}$ be the true values of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ respectively, and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_{p_n})^T = \boldsymbol{X}^T\boldsymbol{y} - \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{V}^{-1}\bar{\boldsymbol{\gamma}} = \boldsymbol{X}^T\boldsymbol{\varepsilon}$. Denote the first q_n entries of $\boldsymbol{\omega}$ by $\boldsymbol{\omega}(1)$, and the remaining elements by $\boldsymbol{\omega}(2)$. Define $\bar{\boldsymbol{\gamma}}(1)$, $\bar{\boldsymbol{\beta}}(1)$, and $\bar{\boldsymbol{\gamma}}(2)$, $\bar{\boldsymbol{\beta}}(2)$ in a similar way. Then $\bar{\boldsymbol{\beta}}(1) = \boldsymbol{V}(1)^{-1}\bar{\boldsymbol{\gamma}}(1)$.

For events $A_1 = \{\|\boldsymbol{\omega}(1)\|_\infty \leq \sqrt{2}\sigma_\varepsilon\sqrt{n\log n}\}$ and $A_2 = \{\|\boldsymbol{\omega}(2)\|_\infty \leq \sqrt{2}\sigma_\varepsilon n^{1-\kappa_0}\sqrt{\log n}\}$, since all the covariates are standardized, we have

$$\begin{aligned} P(A_1 \cap A_2) &\geq 1 - \sum_{j\geq 1}^{q_n} P(|w_j| > \sqrt{2}\sigma_\varepsilon\sqrt{n\log n}) - \sum_{j\geq q_n+1}^{p_n} P(|w_j| > \sqrt{2}\sigma_\varepsilon n^{1-\kappa_0}\sqrt{\log n}) \\ &\geq 1 - 2\sigma_\varepsilon\left[q_n n^{-1} + (p_n - q_n)e^{-n^{1-2\kappa_0}\log n}\right] \\ &= 1 - O(n^{-\delta}), \end{aligned}$$

by the multivariate central limit theorem.

Under the event $A_1 \cap A_2$, it suffices to show that there exists a strict local minimizer $\hat{\boldsymbol{\gamma}}_{SCAD}$ of $L_{SCAD}(\boldsymbol{\gamma}, \hat{\boldsymbol{d}})$ such that $\|\hat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{\gamma}}_{SCAD} - \bar{\boldsymbol{\beta}}\|_\infty = O(\alpha_n)$ and $\hat{\boldsymbol{\gamma}}_{SCAD} =_s \bar{\boldsymbol{\gamma}}$, where $\alpha_n = \sqrt{\log p_n/n}$.

We first show that for sufficiently large n, there exists a vector $\hat{\boldsymbol{\gamma}}_{SCAD}(1)$ in

$$\mathcal{H} = \{\boldsymbol{\gamma}(1) \in R^{q_n} : \|\hat{\boldsymbol{V}}(1)^{-1}\boldsymbol{\gamma}(1) - \bar{\boldsymbol{\beta}}(1)\|_\infty \leq \alpha_n\},$$

such that

$$\Psi_1(\hat{\boldsymbol{\gamma}}_{SCAD}(1)) \equiv \boldsymbol{X}_1^T\boldsymbol{y} - \boldsymbol{X}_1^T\boldsymbol{X}_1\hat{\boldsymbol{V}}(1)^{-1}\hat{\boldsymbol{\gamma}}_{SCAD}(1) - n\hat{\boldsymbol{V}}(1)^{-1}p'_{SCAD,\lambda_n}(\hat{\boldsymbol{\gamma}}_{SCAD}(1)) = \mathbf{0}. \tag{S4.25}$$

For any $\boldsymbol{\gamma}(1) = (\gamma_1, \ldots, \gamma_{q_n}) \in \mathcal{H}$, since $h_{\min} \geq \alpha_n$, $\min_{j=1}^{q_n} |\sqrt{\hat{d}_{jj}}\gamma_j| \geq \min_{j=1}^{q_n} |\bar{\beta}_j| - h_{\min} = h_{\min}$, which implies $\boldsymbol{\gamma} =_s \bar{\boldsymbol{\beta}} =_s \bar{\boldsymbol{\gamma}}$. The Ψ_1 can be rewritten as

$$\Psi_1(\boldsymbol{\gamma}(1)) = \boldsymbol{\omega}(1) + \boldsymbol{X}_1^T\boldsymbol{X}_1(\boldsymbol{V}(1)^{-1}\bar{\boldsymbol{\gamma}}(1) - \hat{\boldsymbol{V}}(1)^{-1}\boldsymbol{\gamma}(1)) - n\hat{\boldsymbol{V}}(1)^{-1}p'_{SCAD,\lambda_n}(\boldsymbol{\gamma}(1)).$$

By multiplying $\left[\boldsymbol{X}_1^T\boldsymbol{X}_1\right]^{-1}$ on both sides, we have

$$\begin{aligned} \left[\boldsymbol{X}_1^T\boldsymbol{X}_1\right]^{-1}\Psi_1(\boldsymbol{\gamma}(1)) &= (n\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{\omega}(1) + (\bar{\boldsymbol{\beta}}(1) - \hat{\boldsymbol{V}}(1)^{-1}\boldsymbol{\gamma}(1)) \\ &\quad - (\hat{\boldsymbol{C}}_n^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}p'_{SCAD,\lambda_n}(\boldsymbol{\gamma}(1)). \end{aligned} \tag{S4.26}$$

Since $\|(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty = o_p^{(\delta)}(\alpha_n n^{\kappa_4})$,

$$\|(n\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{\omega}(1)\|_\infty \leq n^{-1}\|(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty\|\boldsymbol{\omega}(1)\|_\infty = o_p^{(\delta)}(\alpha_n).$$

Since $p'_{SCAD,\lambda_n}(\cdot)$ is a decreasing function,

$$p'_{SCAD,\lambda_n}(\gamma_j) = p'_{SCAD,\sqrt{d_{jj}}\lambda_n}\left(\sqrt{d_{jj}}\gamma_j\right) \Big/ \sqrt{d_{jj}} \leq \sqrt{d_{jj}^{-1}}\ p'_{SCAD,\lambda_n^*}(h_{\min}).$$

Since $\hat{d}_{jj}$ is the CLIME of d_{jj}, by Cai et al. (2011, Theorem 5),

$$\max_{1\leq j\leq p_n}|\hat{d}_{jj} - d_{jj}| = O_p^{(\delta)}\left(\nu_{n,u}\right). \tag{S4.27}$$

We have $\lambda_n = O(1)$, $\max_{1\le j\le p_n} d_{jj}^{-1} = O(\log n)$, which implies

$$\max_{1\le j\le p_n} |\sqrt{\hat{d}_{jj}} - \sqrt{d_{jj}}| = O_p^{(\delta)}(\nu_{n,u}\sqrt{\log n}), \qquad \max_{1\le j\le p_n} \hat{d}_{jj}^{-1} = O_p^{(\delta)}(\log n),$$

$$\max_{1\le j\le p_n} |1/\sqrt{\hat{d}_{jj}} - 1/\sqrt{d_{jj}}| = O_p^{(\delta)}(\nu_{n,u}(\log n)^{3/2}).$$

Then by Condition 6,

$$\begin{aligned} & \|(\hat{\boldsymbol{C}}_n^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}p'_{SCAD,\lambda_n}(\boldsymbol{\gamma}(1))\|_\infty \\ \le\ & \|(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty\|\hat{\boldsymbol{V}}(1)^{-1} - \boldsymbol{V}(1)^{-1}\|_\infty\|\boldsymbol{V}(1)\|_\infty p'_{SCAD,\lambda_n^*}(h_{\min}) + \|(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty p'_{SCAD,\lambda_n^*}(h_{\min}) \\ =\ & o_p^{(\delta)}(\alpha_n) \end{aligned}$$

It follows that $\boldsymbol{\gamma}(1) + \hat{\boldsymbol{V}}(1)\left[\boldsymbol{X}_1^T\boldsymbol{X}_1\right]^{-1}\Psi_1(\boldsymbol{\gamma}(1))$ lies in $\mathcal{H}$ for any $\boldsymbol{\gamma}(1) \in \mathcal{H}$. Thus, by the Brouwer fixed-point theorem, $\hat{\boldsymbol{V}}(1)\left[\boldsymbol{X}_1^T\boldsymbol{X}_1\right]^{-1}\Psi_1(\hat{\boldsymbol{\gamma}}_{SCAD}(1)) = \boldsymbol{0}$ for some $\hat{\boldsymbol{\gamma}}_{SCAD}(1) \in \mathcal{H}$, which implies (S4.25).

Let $\hat{\boldsymbol{\gamma}}_{SCAD}(2) = \boldsymbol{0}$, and $\Psi_2 = (n\lambda_n)^{-1}\hat{\boldsymbol{V}}(2)\boldsymbol{X}_2^T(\boldsymbol{y} - \boldsymbol{X}_1\hat{\boldsymbol{V}}(1)^{-1}\hat{\boldsymbol{\gamma}}_{SCAD}(1))$. Next, we will show that $\|\Psi_2\|_\infty < 1$. It can be rewritten as

$$\Psi_2 = (n\lambda_n)^{-1}\hat{\boldsymbol{V}}(2)\boldsymbol{\omega}(2) + \lambda_n^{-1}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(\boldsymbol{V}(1)^{-1}\bar{\boldsymbol{\gamma}}(1) - \hat{\boldsymbol{V}}(1)^{-1}\hat{\boldsymbol{\gamma}}_{SCAD}(1)). \tag{S4.28}$$

The first term on the right-hand side of (S4.28) is $o_p^{(\delta)}(1)$ by Condition 6. Substitute $\hat{\boldsymbol{\gamma}}_{SCAD}(1)$ in the second term on the right-hand side of (S4.28) by the solution to $\Psi_1(\hat{\boldsymbol{\gamma}}_{SCAD}(1)) = \boldsymbol{0}$. Then we have

$$\begin{aligned} & \lambda_n^{-1}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(\boldsymbol{V}(1)^{-1}\bar{\boldsymbol{\gamma}}(1) - \hat{\boldsymbol{V}}(1)^{-1}\hat{\boldsymbol{\gamma}}_{SCAD}(1)) && \text{(S4.29)} \\ =\ & \lambda_n^{-1}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}p'_{SCAD,\lambda_n}(\hat{\boldsymbol{\gamma}}_{SCAD}(1)) - \lambda_n^{-1}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(n\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{\omega}(1). \end{aligned}$$

Since all the covariates are standardized, the second term on the right-hand side of (S4.29) is

$$\|\lambda_n^{-1}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(n\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{\omega}(1)\|_\infty = O_p^{(\delta)}(\frac{n^{\kappa_0}}{\log^2 n}\sqrt{\log n}\ n^{\kappa_3}\frac{n^{\kappa_4}\alpha_n}{n}\sqrt{n\log n}) = o_p^{(\delta)}(1),$$

by Conditions 5 and 6. In the first term on the right-hand side of (S4.29),

$$\lambda_n^{-1}p'_{SCAD,\lambda_n}(\hat{\gamma}_{SCAD,j}) = \mathcal{P}_{\lambda_n}(\hat{\gamma}_{SCAD,j}) = \mathcal{P}_{\sqrt{d_{jj}}\lambda_n}(\sqrt{d_{jj}}\hat{\gamma}_{SCAD,j}) \leq \mathcal{P}_{\lambda_n^*}(h_{\min}),$$

where $\hat{\boldsymbol{\gamma}}_{SCAD,,j}$ is the jth element of $\hat{\boldsymbol{\gamma}}_{SCAD}$. It follows that $\|\lambda_n^{-1}p'_{SCAD,\lambda_n}(\hat{\boldsymbol{\gamma}}_{SCAD}(1))\|_\infty \leq \mathcal{P}_{\lambda_n^*}(h_{\min})$. The first term on the right-hand side of (S4.29) is

$$\begin{aligned}
&\left\|\lambda_n^{-1}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}p'_{SCAD,\lambda_n}(\hat{\boldsymbol{\gamma}}_{SCAD}(1))\right\|_\infty \qquad \text{(S4.30)}\\
= &\left\|\lambda_n^{-1}\left\{\hat{\boldsymbol{V}}(2)-\boldsymbol{V}(2)\right\}\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}p'_{SCAD,\lambda_n}(\hat{\boldsymbol{\gamma}}_{SCAD}(1))\right\|_\infty\\
&+\left\|\lambda_n^{-1}\boldsymbol{V}(2)\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\left\{\hat{\boldsymbol{V}}(1)^{-1}-\boldsymbol{V}(1)^{-1}\right\}\boldsymbol{V}(1)\boldsymbol{V}(1)^{-1}p'_{SCAD,\lambda_n}(\hat{\boldsymbol{\gamma}}_{SCAD}(1))\right\|_\infty\\
&+\left\|\lambda_n^{-1}\boldsymbol{V}(2)\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}p'_{SCAD,\lambda_n}(\hat{\boldsymbol{\gamma}}_{SCAD}(1))\right\|_\infty.
\end{aligned}$$

Then, by Condition 2, the last term on the right-hand side of (S4.30) is less than 1. Other terms on the right-hand side of (S4.30) are $o_p^{(\delta)}(1)$ by Condition 6. Therefore, $\|\Psi_2\|_\infty < 1$ for sufficiently large n.

By Condition 6, $\lambda_{\min}(\hat{\boldsymbol{C}}_n^{11}) > (a-1)^{-1}$ for sufficiently large n. Thus, the $\hat{\boldsymbol{\gamma}}_{SCAD}$ is a strict local minimizer of $L_{SCAD}(\boldsymbol{\gamma}, \hat{\boldsymbol{d}})$ by Fan and Lv (2011, Theorem 1). This completes the proof.

□

Proof of Theorem 1

Proof. We use the same notation as in the proof of Theorem 2.

For events $A_1 = \{\|\boldsymbol{\omega}(1)\|_\infty \leq \sqrt{2}\sigma_\varepsilon\sqrt{n\log n}\}$ and $A_2 = \{\|\boldsymbol{\omega}(2)\|_\infty \leq \sqrt{2}\sigma_\varepsilon n^{1-\kappa_0}\sqrt{\log n}\}$, we still have $P(A_1\cap A_2) \geq 1 - O(n^{-\delta})$ by the multivariate central limit theorem.

Under the event $A_1 \cap A_2$, we first show that for sufficiently large n, there exists a vector $\hat{\boldsymbol{\gamma}}_{Lasso}(1)$ in

$$\mathcal{H} = \{\boldsymbol{\gamma}(1) \in R^{q_n} : \|\hat{\boldsymbol{V}}(1)^{-1}\boldsymbol{\gamma}(1) - \bar{\boldsymbol{\beta}}(1)\|_\infty \leq \alpha_n\},$$

such that

$$\Psi_1(\hat{\boldsymbol{\gamma}}_{Lasso}(1)) \equiv \boldsymbol{X}_1^T\boldsymbol{y} - \boldsymbol{X}_1^T\boldsymbol{X}_1\hat{\boldsymbol{V}}(1)^{-1}\hat{\boldsymbol{\gamma}}_{Lasso}(1) - \lambda_n\hat{\boldsymbol{V}}(1)^{-1}\operatorname{sign}(\hat{\boldsymbol{\gamma}}_{Lasso}(1)) = \mathbf{0}, \quad \text{(S4.31)}$$

where $\alpha_n = \sqrt{\log p_n/n}$. For any $\boldsymbol{\gamma}(1) = (\gamma_1, \ldots, \gamma_{q_n}) \in \mathcal{H}$, since $h_{\min} \geq \alpha_n$, $\min_{j=1}^{q_n} |\sqrt{\hat{d}_{jj}}\gamma_j| \geq \min_{j=1}^{q_n} |\bar{\beta}_j| - h_{\min} = h_{\min}$, which implies $\boldsymbol{\gamma} =_s \bar{\boldsymbol{\beta}} =_s \bar{\boldsymbol{\gamma}}$.

By multiplying $\left[\boldsymbol{X}_1^T\boldsymbol{X}_1\right]^{-1}$ on both sides, the Ψ_1 can be rewritten as

$$\begin{aligned}\left[\boldsymbol{X}_1^T\boldsymbol{X}_1\right]^{-1}\Psi_1(\boldsymbol{\gamma}(1)) &= (n\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{\omega}(1) + (\bar{\boldsymbol{\beta}}(1) - \hat{\boldsymbol{V}}(1)^{-1}\boldsymbol{\gamma}(1)) \\ &\quad -\lambda_n(n\hat{\boldsymbol{C}}_n^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}\operatorname{sign}(\boldsymbol{\gamma}(1)).\end{aligned} \quad \text{(S4.32)}$$

Since $\|(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty = o_p^{(\delta)}(\alpha_n n^{\kappa_4})$ for $\kappa_4 < 1/2$,

$$\|(n\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{\omega}(1)\|_\infty \leq n^{-1}\|(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty\|\boldsymbol{\omega}(1)\|_\infty = O_p^{(\delta)}(n^{-1}\alpha_n n^{\kappa_4}\sqrt{n\log n}) = o_p^{(\delta)}(\alpha_n).$$

Since $\hat{d}_{jj}$ is the CLIME of d_{jj},

$$\max_{1\leq j\leq p_n} |\hat{d}_{jj} - d_{jj}| = O_p^{(\delta)}\left(\nu_{n,u}\right). \quad \text{(S4.33)}$$

We have $\max_{1\leq j\leq q_n} d_{jj} = O(\log n)$, $\max_{1\leq j\leq p_n} d_{jj}^{-1} = O(\log n)$ and $\nu_{n,u}\log n = o(1)$, which implies

$$\max_{1\leq j\leq p_n} |\sqrt{\hat{d}_{jj}} - \sqrt{d_{jj}}| = O_p^{(\delta)}(\nu_{n,u}\sqrt{\log n}), \qquad \max_{1\leq j\leq q_n} \sqrt{\hat{d}_{jj}} = O_p^{(\delta)}(\sqrt{\log n}),$$

$$\max_{1\leq j\leq p_n} |1/\sqrt{\hat{d}_{jj}} - 1/\sqrt{d_{jj}}| = O_p^{(\delta)}(\nu_{n,u}(\log n)^{3/2}), \qquad \max_{1\leq j\leq p_n} 1/\sqrt{\hat{d}_{jj}} = O_p^{(\delta)}(\sqrt{\log n}).$$

Since $\lambda_n = O(n^{1/2+\kappa_3})$, the third term on the right-hand side of (S4.32) is

$$\|\lambda_n(n\hat{\boldsymbol{C}}_n^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}\operatorname{sign}(\boldsymbol{\gamma}(1))\|_\infty \leq \frac{\lambda_n}{n}\|(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty\|\hat{\boldsymbol{V}}(1)^{-1}\|_\infty = o_p^{(\delta)}(\alpha_n).$$

It follows that $\boldsymbol{\gamma}(1) + \hat{\boldsymbol{V}}(1)\left[\boldsymbol{X}_1^T\boldsymbol{X}_1\right]^{-1}\Psi_1(\boldsymbol{\gamma}(1))$ lies in $\mathcal{H}$ for any $\boldsymbol{\gamma}(1) \in \mathcal{H}$. Thus, by the Brouwer fixed-point theorem, $\hat{\boldsymbol{V}}(1)\left[\boldsymbol{X}_1^T\boldsymbol{X}_1\right]^{-1}\Psi_1(\hat{\boldsymbol{\gamma}}_{Lasso}(1)) = \boldsymbol{0}$ for some $\hat{\boldsymbol{\gamma}}_{Lasso}(1) \in \mathcal{H}$, which implies (S4.31).

Let $\hat{\boldsymbol{\gamma}}_{Lasso}(2) = \boldsymbol{0}$ and $\Psi_2 = \lambda_n^{-1}\hat{\boldsymbol{V}}(2)\boldsymbol{X}_2^T(\boldsymbol{y} - \boldsymbol{X}_1\hat{\boldsymbol{V}}(1)^{-1}\hat{\boldsymbol{\gamma}}_{Lasso}(1))$. Next, we will show that $\|\Psi_2\|_\infty < 1$. It can be rewritten as

$$\Psi_2 = \lambda_n^{-1}\hat{\boldsymbol{V}}(2)\boldsymbol{\omega}(2) + n\lambda_n^{-1}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(\boldsymbol{V}(1)^{-1}\bar{\boldsymbol{\gamma}}(1) - \hat{\boldsymbol{V}}(1)^{-1}\hat{\boldsymbol{\gamma}}_{Lasso}(1)). \tag{S4.34}$$

The first term on the right-hand side of (S4.34) is $o_p^{(\delta)}(1)$ since $\lambda_n^{-1} = o(n^{\kappa_0-1}/\log n)$. Substitute $\hat{\boldsymbol{\gamma}}_{Lasso}(1)$ in the second term on the right-hand side of (S4.34) by the solution to $\Psi_1(\hat{\boldsymbol{\gamma}}_{Lasso}(1)) = \boldsymbol{0}$. Then we have

$$\begin{aligned} & n\lambda_n^{-1}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(\boldsymbol{V}(1)^{-1}\bar{\boldsymbol{\gamma}}(1) - \hat{\boldsymbol{V}}(1)^{-1}\hat{\boldsymbol{\gamma}}_{Lasso}(1)) \\ = \;& \hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}\,\mathrm{sign}(\hat{\boldsymbol{\gamma}}_{Lasso}(1)) - n\lambda_n^{-1}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(n\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{\omega}(1). \end{aligned} \tag{S4.35}$$

Since all the covariates are standardized, the second term in (S4.35) is

$$\|n\lambda_n^{-1}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(n\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{\omega}(1)\|_\infty = O_p^{(\delta)}(\frac{n^{\kappa_0}}{\log n}\sqrt{\log n}\; n^{\kappa_3}\frac{\alpha_n n^{\kappa_4}}{n}\sqrt{n\log n}) = o_p^{(\delta)}(1),$$

by Conditions 3 and 4. The first term on the right-hand side of (S4.35) is

$$\begin{aligned} & \left\|\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}\,\mathrm{sign}(\hat{\boldsymbol{\gamma}}_{Lasso}(1))\right\|_\infty \\ = \;& \left\|\left\{\hat{\boldsymbol{V}}(2) - \boldsymbol{V}(2)\right\}\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}\,\mathrm{sign}(\hat{\boldsymbol{\gamma}}_{Lasso}(1))\right\|_\infty \\ & + \left\|\boldsymbol{V}(2)\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\left\{\hat{\boldsymbol{V}}(1)^{-1} - \boldsymbol{V}(1)^{-1}\right\}\mathrm{sign}(\hat{\boldsymbol{\gamma}}_{Lasso}(1))\right\|_\infty \\ & + \left\|\boldsymbol{V}(2)\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\,\mathrm{sign}(\hat{\boldsymbol{\gamma}}_{Lasso}(1))\right\|_\infty. \end{aligned} \tag{S4.36}$$

Since $\mathrm{sign}(\hat{\boldsymbol{\gamma}}_{Lasso}(1)) = \mathrm{sign}(\boldsymbol{\beta}(1))$, the last term of (S4.36) is less than 1 by Condition 1. Other terms on the right-hand side of (S4.36) are $o_p^{(\delta)}(1)$ since $\nu_{n,u}n^{\kappa_3+\kappa_4-\kappa_0}(\log n)^2 = o(1)$.

Therefore, $\|\Psi_2\|_\infty < 1$ for sufficiently large n.

By Condition 4, $\lambda_{\min}(\hat{\boldsymbol{C}}_n^{11}) > 0$ for sufficiently large n. Thus, the $\hat{\boldsymbol{\gamma}}_{Lasso}$ is a strict local minimizer of $L_{Lasso}(\boldsymbol{\gamma}, \hat{\boldsymbol{d}})$ by Fan and Lv (2011, Theorem 1). This completes the proof.

□

Proof of Theorem 4

Proof. Let

$$\begin{aligned} F_{1,n} &= \{\text{there exists } \lambda_n \geq 0 \text{ such that } \hat{\boldsymbol{\gamma}}_{Lasso} =_s \boldsymbol{\gamma}\} \\ &= \{\text{there exists } \lambda_n \geq 0 \text{ such that } \operatorname{sign}(\hat{\boldsymbol{\gamma}}_{Lasso}(1)) = \operatorname{sign}(\boldsymbol{\gamma}(1)) \text{ and } \hat{\boldsymbol{\gamma}}_{Lasso}(2) = \mathbf{0}\}. \end{aligned}$$

Then the SPAC-Lasso is general sign consistent if $\lim_{n\to\infty} P(F_{1,n}) = 1$. Based on the Karush-Kuhn-Tucker (KKT) conditions and (S4.10), $F_{1,n}$ implies that

$$\sqrt{n}\hat{\boldsymbol{V}}(1)\hat{\boldsymbol{C}}^{11}\hat{\boldsymbol{u}}(1) - \hat{\boldsymbol{V}}(1)\boldsymbol{w}(1) = -\frac{\lambda_n}{\sqrt{n}} \operatorname{sign}(\bar{\boldsymbol{\beta}}(1)), \tag{S4.37}$$

$$|\sqrt{n}\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}\hat{\boldsymbol{u}}(1) - \hat{\boldsymbol{V}}(2)\boldsymbol{w}(2)| \leq \frac{\lambda_n}{\sqrt{n}}\mathbf{1}. \tag{S4.38}$$

Solve $\hat{\boldsymbol{u}}(1)$ out of (S4.37) and substitute it into (S4.38). Then we have

$$F_{2,n} := \left\{ \frac{\lambda_n}{\sqrt{n}}\boldsymbol{f}_{1,n} \leq \hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{w}(1) - \hat{\boldsymbol{V}}(2)\boldsymbol{w}(2) \leq \frac{\lambda_n}{\sqrt{n}}\boldsymbol{f}_{2,n} \right\},$$

where

$$\boldsymbol{f}_{1,n} = -\mathbf{1} + \hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1) \operatorname{sign}(\bar{\boldsymbol{\beta}}(1)),$$

$$\boldsymbol{f}_{2,n} = \mathbf{1} + \hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}^{-1}(1) \operatorname{sign}(\bar{\boldsymbol{\beta}}(1)).$$

Let $f_n(\boldsymbol{x}) = P(F_{2,n}|\boldsymbol{X} = \boldsymbol{x})$, and

$$\begin{aligned} H_n &= \{\boldsymbol{X}| \text{ at least one element of } |\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))| \\ &\quad \text{is greater than or equal to } 1\}, \end{aligned}$$

representing the design matrices for which Condition 8 fails. Since $\boldsymbol{\varepsilon}$ follows a Gaussian distribution with mean $\mathbf{0}$, $f_n(\boldsymbol{x}) < 1/2$ when $\boldsymbol{x} \in H_n$. If for any large n there exists $\delta > 0$ such that $P(\boldsymbol{X} \in H_n) > \delta$, then we have

$$P(F_{2,n}) = \int_{\boldsymbol{x}\in H_n} f_n(\boldsymbol{x})dP(\boldsymbol{x}) + P(F_{2,n}, \boldsymbol{X} \in H_n^c) \leq P(\boldsymbol{X} \in H_n)/2 + P(\boldsymbol{X} \in H_n^c) < 1 - \delta/2.$$

Thus, $\limsup P(F_{1,n}) \leq \limsup P(F_{2,n}) < 1$. This contradicts the general sign consistency. Therefore, $|\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))| < \mathbf{1}$.

Since

$$\begin{aligned} & |\boldsymbol{V}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{V}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))| \\ \leq\ & |\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\hat{\boldsymbol{V}}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))| \\ & +|\{\boldsymbol{V}(2) - \hat{\boldsymbol{V}}(2)\}\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{V}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))| \\ & +|\hat{\boldsymbol{V}}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\{\boldsymbol{V}(1)^{-1} - \hat{\boldsymbol{V}}(1)^{-1}\}\operatorname{sign}(\boldsymbol{\beta}(1))|, \end{aligned} \tag{S4.39}$$

where the last two items go to zero as n increases, each element in $|\boldsymbol{V}(2)\hat{\boldsymbol{C}}^{21}(\hat{\boldsymbol{C}}^{11})^{-1}\boldsymbol{V}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))|$ is less than 1 for sufficiently large n. □

Proof of Proposition 1

Proof. By the definition of $\boldsymbol{C}_n$,

$$\|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty = \frac{|\alpha_2 m_0|}{1 - \alpha_1 + \alpha_1 q_0}. \tag{S4.40}$$

Since $\boldsymbol{C}_n$ is positive definite for all large q_0 and p_0, $\alpha_1 \geq 0$ and $\alpha_3 \geq 0$.

We have

$$\begin{aligned} & \|\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty \\ \geq\ & \|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty - \|\boldsymbol{C}_n^{21}\{(\boldsymbol{C}_n^{11})^{-1} - (\hat{\boldsymbol{C}}_n^{11})^{-1}\}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty \\ & -\|(\boldsymbol{C}_n^{21} - \hat{\boldsymbol{C}}_n^{21})(\hat{\boldsymbol{C}}_n^{11})^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty. \end{aligned} \tag{S4.41}$$

By the normality assumption and (Laurent and Massart, 2000, Lemma 1), elements in $\hat{\boldsymbol{C}}_n$ converge to corresponding elements in $\boldsymbol{C}_n$ with probability at least $1 - O(\exp(n^{-c}))$ and $0 < c < 1/2$, which implies that $\|\boldsymbol{C}_n^{21} - \hat{\boldsymbol{C}}_n^{21}\|_\infty = O_p^{(\delta)}(n^{-\tau})$. In addition, similarly as the proof in part (2) of Lemma 3, we can show that $\|(\boldsymbol{C}_n^{11})^{-1} - (\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty = O_p^{(\delta)}(n^{-\tau})$. Thus, the last two terms on the right hand of (S4.41) goes to zero as n increases with probability at least $1 - O(n^{-\delta})$.

On the one hand, if $\alpha_1 = 0$, (4.5) obviously holds for large m_0. If $|\alpha_2| > \alpha_1 L_0 > 0$, then there exists a constant $\epsilon_0 > 0$ such that $|\alpha_2| \geq \alpha_1 q_0/m_0 + \epsilon_0$. Thus, (4.5) follows from (S4.40) and (S4.41) for sufficiently large m_0.

On the other hand, we have

$$\begin{aligned} & \|\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty \\ \leq\ & \|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty + \|\boldsymbol{C}_n^{21}\{(\boldsymbol{C}_n^{11})^{-1} - (\hat{\boldsymbol{C}}_n^{11})^{-1}\}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty \\ & +\|(\boldsymbol{C}_n^{21} - \hat{\boldsymbol{C}}_n^{21})(\hat{\boldsymbol{C}}_n^{11})^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty. \end{aligned} \tag{S4.42}$$

Thus, (4.5) implies that $\|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty \geq 1$, which implies that

$$|\alpha_2| \geq \frac{1-\alpha_1}{m_0} + \alpha_1\frac{q_0}{m_0} > \alpha_1 L_0.$$

Hence $|\alpha_2| > \alpha_1 L_0 \geq \alpha_1$, where the second inequality follows from $m_0 \leq q_0$. By the positive definiteness of $\boldsymbol{C}_n$,

$$\alpha_3 > |\alpha_2|\frac{q_0|\alpha_2|}{1-\alpha_1+\alpha_1 q_0} + \left(\frac{q_0\alpha_2^2}{1-\alpha_1+\alpha_1 q_0} - 1\right)\frac{1}{r_0-1},$$

where $r_0 = p_0 - q_0$. Since the $\boldsymbol{C}_n$ is positive definite for any large r_0, (4.5) implies $\alpha_3 \geq |\alpha_2|$.

We also have

$$\begin{aligned}
& \|\boldsymbol{V}(2)\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty \qquad \text{(S4.43)}\\
\leq\ & \|\boldsymbol{V}(2)\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty \\
& +\|\boldsymbol{V}(2)\boldsymbol{C}_n^{21}\{(\boldsymbol{C}_n^{11})^{-1} - (\hat{\boldsymbol{C}}_n^{11})^{-1}\}\boldsymbol{V}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty \\
& +\|\boldsymbol{V}(2)(\boldsymbol{C}_n^{21} - \hat{\boldsymbol{C}}_n^{21})(\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))\|_\infty.
\end{aligned}$$

Similarly as (S4.41), the last two terms on the right hand side of (S4.43) goes to zero as n increases with probability at least $1 - O(n^{-\delta})$. Based on the definition of $\boldsymbol{C}_n$,

$$\begin{aligned}
& |\boldsymbol{V}(2)\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))| \qquad \text{(S4.44)}\\
=\ & |\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))|\sqrt{\frac{1-\alpha_3}{1-\alpha_1}} \\
& \times\sqrt{\frac{(1+(q_0-2)\alpha_1)(1+(r_0-1)\alpha_3) - (q_0-1)r_0\alpha_2^2}{(1+(q_0-1)\alpha_1)(1+(r_0-2)\alpha_3) - q_0(r_0-1)\alpha_2^2}}.
\end{aligned}$$

Since $\alpha_3 > \alpha_1$ and the last factor on the right hand side of (S4.44) is close to 1 for large q_0 and r_0,

$$|\boldsymbol{V}(2)\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))| < |\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))|. \qquad \text{(S4.45)}$$

Then by (S4.41) and (S4.43), (4.6) holds for sufficiently large m_0 and r_0. □

Proof of Proposition 2

Proof. By the definition of $\boldsymbol{C}_n$,

$$\|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\|_\infty = \frac{|\alpha_2 q_0|}{1-\alpha_1+\alpha_1 q_0}. \qquad \text{(S4.46)}$$

We have

$$\|\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty \leq \|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\|_\infty + \|\boldsymbol{C}_n^{21}\{(\boldsymbol{C}_n^{11})^{-1} - (\hat{\boldsymbol{C}}_n^{11})^{-1}\}\|_\infty + \|(\boldsymbol{C}_n^{21} - \hat{\boldsymbol{C}}_n^{21})(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty.$$

Since the last two terms on the right hand side of (S4.47) goes to zero as n increases with probability at least $1-O(n^{-\delta})$, the inequality (4.7) implies that $\|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\|_\infty \mathcal{P}_\lambda(h_{\min}) \geq 1$, which implies

$$|\alpha_2| \geq \left(\frac{1-\alpha_1}{q_0} + \alpha_1\right)\frac{1}{\mathcal{P}_\lambda(h_{\min})} > \alpha_1.$$

By the positive definiteness of $\boldsymbol{C}_n$,

$$\alpha_3 > |\alpha_2|\frac{q_0|\alpha_2|}{1-\alpha_1+\alpha_1 q_0} + \left(\frac{q_0\alpha_2^2}{1-\alpha_1+\alpha_1 q_0} - 1\right)\frac{1}{r_0-1},$$

where $r_0 = p_0 - q_0$. Since the second term is close to zero for large r_0, (4.7) implies $\alpha_3 \geq |\alpha_2|$.

We also have

$$\begin{aligned}\|\boldsymbol{V}(2)\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\|_\infty \leq\ & \|\boldsymbol{V}(2)\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\|_\infty \\ & + \|\boldsymbol{V}(2)\boldsymbol{C}_n^{21}\{(\boldsymbol{C}_n^{11})^{-1} - (\hat{\boldsymbol{C}}_n^{11})^{-1}\}\boldsymbol{V}(1)^{-1}\|_\infty \\ & + \|\boldsymbol{V}(2)(\boldsymbol{C}_n^{21} - \hat{\boldsymbol{C}}_n^{21})(\hat{\boldsymbol{C}}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\|_\infty,\end{aligned} \tag{S4.47}$$

and

$$\|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\|_\infty \leq \|\hat{\boldsymbol{C}}_n^{21}(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty + \|\boldsymbol{C}_n^{21}\{(\boldsymbol{C}_n^{11})^{-1} - (\hat{\boldsymbol{C}}_n^{11})^{-1}\}\|_\infty + \|(\boldsymbol{C}_n^{21} - \hat{\boldsymbol{C}}_n^{21})(\hat{\boldsymbol{C}}_n^{11})^{-1}\|_\infty. \tag{S4.48}$$

The last two terms on the right hand sides of (S4.47) and (S4.48) goes to zero as n increases with probability at least $1 - O(n^{-\delta})$. Moreover, based on the definition of $\boldsymbol{C}_n$,

$$\begin{aligned}\|\boldsymbol{V}(2)\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\|_\infty &= \|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\|_\infty\sqrt{\frac{1-\alpha_3}{1-\alpha_1}} \\ &\times\sqrt{\frac{(1+(q_0-2)\alpha_1)(1+(r_0-1)\alpha_3)-(q_0-1)r_0\alpha_2^2}{(1+(q_0-1)\alpha_1)(1+(r_0-2)\alpha_3)-q_0(r_0-1)\alpha_2^2}}.\end{aligned}$$

Since $\alpha_3 > \alpha_1$ and the last factor is close to 1 for large q_0 and r_0, (4.8) holds for sufficiently large q_0 and r_0. □

Proof of Corollary 1

Proof. By the definition of $\boldsymbol{C}_n$ and distribution assumption of $\boldsymbol{X}$, Condition 5 is satisfied. By the definition of p_n and q_n, and $h_{\min} \geq n^{-\kappa_0}$, Condition 3 also holds. Since

$$\frac{|\alpha_2 m_0|}{1-\alpha_1+\alpha_1 q_0}\sqrt{\frac{1-\alpha_3}{1-\alpha_1}} \leq 1-\eta \tag{S4.49}$$

is equivalent to

$$|\alpha_2| \leq (1-\eta)\sqrt{\frac{1-\alpha_1}{1-\alpha_3}}(\alpha_1\frac{q_0}{m_0}+\frac{1-\alpha_1}{m_0}),$$

then (S4.49) is implied by (4.9). Then, by (S4.43), (S4.44), and the proof of Proposition 1, Condition 1 holds for sufficiently large q_0 and $p_0 - q_0$ with probability at least $1 - O(n^{-\delta})$. This completes the proof.

□

Proof of Corollary 2

Proof. By the definition of $\boldsymbol{C}_n$ and distribution assumption of $\boldsymbol{X}$, Condition 5 is satisfied. By the definition of p_n and q_n, and $h_{\min} \geq n^{-\kappa_0}$, Condition 3 also holds. The first element of $|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\,\mathrm{sign}(\boldsymbol{\beta}(1))|$ is the largest one, and it is equal to the first element of

$|\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\mathbf{1}|$. Combining with (S4.42), (4.5) implies that

$$\frac{|\alpha_2|(1-\alpha_1\alpha_2)}{(1+\alpha_1)(1-\alpha_2)} \geq 1,$$

which further implies $\alpha_2 > \alpha_1$.

Let $r_0 = p_0 - q_0$, $r_n = p_n - q_n$,

$$T_1 = 1 + \frac{(\alpha_1-\alpha_2)^2(1-\alpha_2^{2q_0-2})}{(1-\alpha_1^2)(1-\alpha_2^2)},$$

$$T_2 = 1 + \frac{(\alpha_3-\alpha_2)^2(1-\alpha_2^{2r_0-2})}{(1-\alpha_3^2)(1-\alpha_2^2)},$$

$T = 1-\alpha_2^2 T_1 T_2$, $\tilde{T} = 1-\alpha_2^4 T_1 T_2$, and

$$S(q_0, r_0, k, \alpha_1, \alpha_2, \alpha_3) = \frac{(\alpha_1-\alpha_2)^2(1-\alpha_1\alpha_2)^2\alpha_2^{2(q_0-k)}}{1-\alpha_1^4}\left[1+\frac{(\alpha_3-\alpha_2)^2(1-\alpha_2^{2r_0-2})}{(1-\alpha_3^2)(1-\alpha_2^2)}\right].$$

Then, d_{jj} can be expressed as

$$d_{jj} = \begin{cases} 1 & \text{if } 1 \leq j \leq q_n - q_0 \text{ and } q_n + r_0 + 1 \leq j \leq p_n \\ \frac{1}{1-\alpha_1^2}\left(1+\frac{S(q_0,r_0,1,\alpha_1,\alpha_2,\alpha_3)}{T}\right) & \text{if } j = q_n - q_0 + 1 \\ \frac{1+\alpha_1^2}{1-\alpha_1^2}\left(1+\frac{S(q_0,r_0,j,\alpha_1,\alpha_2,\alpha_3)}{T}\right) & \text{if } q_n - q_0 + 2 \leq j \leq q_n - 1 \\ \frac{\tilde{T}}{(1-\alpha_1^2)T} & \text{if } j = q_n \\ \frac{\tilde{T}}{(1-\alpha_3^2)T} & \text{if } j = q_n + 1 \\ \frac{1+\alpha_3^2}{1-\alpha_3^2}\left(1+\frac{S(r_0,q_0,p_0-j+1,\alpha_3,\alpha_2,\alpha_1)}{T}\right) & \text{if } q_n + 2 \leq j \leq q_n + r_0 - 1 \\ \frac{1}{1-\alpha_3^2}\left(1+\frac{S(r_0,q_0,1,\alpha_3,\alpha_2,\alpha_1)}{T}\right) & \text{if } j = q_n + r_0. \end{cases}$$

Since $S(q_0, r_0, k, \alpha_1, \alpha_2, \alpha_3) = \alpha_2^2 S(q_0, r_0, k+1, \alpha_1, \alpha_2, \alpha_3)$,

$$\alpha_2 < \sqrt{d_{jj}/d_{j+1j+1}} < 1 \qquad \text{for } q_n - q_0 + 2 \leq j \leq q_n - 2, \tag{S4.50}$$

and

$$\alpha_2 < \sqrt{d_{j+1j+1}/d_{jj}} < 1 \qquad \text{for } q_n + 2 \leq j \leq q_n + r_0 - 2. \tag{S4.51}$$

In addition, since $\alpha_2 > \alpha_1 > 0$,

$$d_{q_n q_n} \geq d_{q_n - 1 q_n - 1}. \tag{S4.52}$$

Let $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{r_n})^T = |\boldsymbol{V}(2)\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\operatorname{sign}(\boldsymbol{\beta}(1))|$. Then by (S4.51), for $2 \leq j \leq r_0 - 1$, $\psi_j/\psi_{j+1} = \sqrt{d_{q_n+j+1 q_n+j+1}/d_{q_n+j q_n+j}}/\alpha_2 > 1$. Since $\boldsymbol{C}_n$ is extended block-AR, $\psi_j = 0$ for $r_0 + 1 \leq j \leq r_n$. Hence, we need to show that there exists a positive constant η_0 such that $\max\{\psi_1, \psi_2\} \leq 1 - \eta_0$. Let $\eta_0 \in (0, \eta)$.

By (S4.50) and (S4.52) for any $c_{0,\eta} \in (0, \eta - \eta_0)$ and sufficiently large p_0 and q_0,

$$\psi_1 \leq \frac{|\alpha_2|(1 - \alpha_1\alpha_2)}{(1 + \alpha_1)(1 - \alpha_2)}\sqrt{\frac{d_{q_n q_n}}{d_{q_n+1 q_n+1}}} + c_{0,\eta} = \sqrt{\frac{1 - \alpha_3^2}{1 - \alpha_1^2}}\frac{|\alpha_2|(1 - \alpha_1\alpha_2)}{(1 + \alpha_1)(1 - \alpha_2)} + c_{0,\eta} \leq 1 - \eta_0.$$

The last inequality follows from (4.10). In addition, $\psi_2 = \psi_1\alpha_2\sqrt{d_{q_n+1 q_n+1}/d_{q_n+2 q_n+2}} \leq \psi_1\alpha_2/|\alpha_2 - \alpha_3|$. Then we have $\|\boldsymbol{\psi}\|_\infty \leq 1 - \eta$ by (4.10). By (S4.43), Condition 1 holds with probability at least $1 - O(n^{-\delta})$. This completes the proof. □

Proof of Corollary 3

Proof. Similarly as the proof in Corollary 1, Conditions 2, 3, and 5 are satisfied, which completes the proof. □

Proof of Proposition 3

Proof. For any $1 \leq i \leq q_n$ and $q_n + 1 \leq j \leq p_n$, it suffices to show that $d_{ii}/d_{jj} \leq g_n^2$. Let $\boldsymbol{v}_{\max,i}$ be an eigenvector corresponding to the largest eigenvalue of $\boldsymbol{C}_{n,i}$, and φ_i^* be the angle between $\boldsymbol{v}_i$ and $\boldsymbol{v}_{\max,i}$. Then,

$$\frac{d_{ii}}{d_{jj}} = \frac{\det(\boldsymbol{C}_{n,i})}{\det(\boldsymbol{C}_{n,j})} = \frac{1 - \boldsymbol{v}_j^T(\boldsymbol{C}_{n,j})^{-1}\boldsymbol{v}_j}{1 - \boldsymbol{v}_i^T(\boldsymbol{C}_{n,i})^{-1}\boldsymbol{v}_i} \leq \frac{1 - \|\boldsymbol{v}_j\|_2^2/\lambda_{\max,j}}{1 - \|\boldsymbol{v}_i\|_2^2\cos^2\varphi_i^*/\lambda_{\max,i} - \|\boldsymbol{v}_i\|_2^2\sin^2\varphi_i^*/\lambda_{\min,i}}.$$

By the Perron–Frobenius theorem, $\boldsymbol{v}_{\max,i}$ can be chosen from the cone spanned by the columns of $\boldsymbol{C}_{n,i}$. Thus, $0 \leq \varphi_i^* \leq \varphi_i \leq \pi/2$. The last inequality follows from $c_{kl} \geq 0$

for any $1 \leq k, l \leq p_n$. Therefore,

$$\frac{d_{ii}}{d_{jj}} \leq \frac{1 - \|\boldsymbol{v}_j\|_2^2/\lambda_{\max,j}}{1 - \|\boldsymbol{v}_i\|_2^2/\lambda_{\max,i} - \|\boldsymbol{v}_i\|_2^2 \sin^2\varphi_i/\lambda_{\min,i}},$$

which implies $\|\boldsymbol{V}(2)\boldsymbol{C}_n^{21}(\boldsymbol{C}_n^{11})^{-1}\boldsymbol{V}(1)^{-1}\|_\infty \leq 1-\eta$. By normality assumption and Conditions 3 and 4 with $\kappa_0 > \max\{\kappa_2+\kappa_3, (\kappa_2+\kappa_4)/2\}$, the last two terms on the right hand sides of (S4.47) goes to zero as n increases with probability at least $1 - O(n^{-\delta})$. This completes the proof.

□

References

Cai, T., W. Liu, and X. Luo (2011). A constrained l_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association 106*(494), 594–607.

Fan, J. and J. Lv (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Transactions on Information Theory 57*(8), 5467–5484.

Gotze, F. (1991). On the rate of convergence in the multivariate CLT. *The Annals of Probability*, 724–739.

Laurent, B. and P. Massart (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics 28*(5), 1302–1338.

Pinelis, I. and R. Molzon (2016). Optimal-order bounds on the rate of convergence to normality in the multivariate delta method. *Electronic Journal of Statistics 10*(1), 1001–1063.