REGULARIZING LASSO: A CONSISTENT VARIABLE SELECTION METHOD

Quefeng Li and Jun Shao

University of Wisconsin, Madison,
East China Normal University and University of Wisconsin, Madison

Supplementary Material

The supplementary material is organized as the follows. In Section S1, we provide additional information regarding the simulation studies in Section 5.1. In Section S2, we give lemmas for establishing asymptotic results when the precision matrix is sparse. The proofs of all lemmas and theorems are given in Section S3.

S1 Additional Tables of Simulation Results

Table 1 provides the average computational time (in minutes) for the eight methods under the simulation settings. SIS clearly requires the least computational effort, whereas RLASSO as well as Scout require much longer computational time. But all methods except RLASSO(CLIME) can be computed under a reasonable amount of time for $p = 5000$ and $n = 100$. RLASSO(CLIME) takes much longer because of inverting a matrix of 5000 dimension. However, 790.8 minutes of computation may still be acceptable. In an unreported simulation with $p = 2000$ and the same other settings, the average computational time for RLASSO(CLIME) is 46.7 minutes.

Finally, because the estimation of $\mathbf{\Sigma}$ is an important step in RLASSO, we provide the average Frobenius norms of estimated $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ in the simulation in Table 2. It is clear that $\boldsymbol{S}$ is not a good estimator of $\mathbf{\Sigma}$ in terms of the Frobenius norm, $\hat{\mathbf{\Sigma}}$ by thresholding (or $\hat{\mathbf{\Omega}}$ by CLIME) is a good estimator when $\mathbf{\Sigma}$ (or $\mathbf{\Omega}$) is sparse but not so good when $\mathbf{\Sigma}$ (or $\mathbf{\Omega}$) is not sparse.

S2 Lemmas

Lemma S1. *Assume conditions (C1)-(C2) and (C3$''$), for any $\lambda_n \to 0$, there exist positive constants C_{14}, C_{15}, C_{16} such that*

$$P\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty > t\right) \le 8\exp\left(-C_{14}n[t/s_h r_q]^{\frac{2}{1-q}}\right) + 4p^2\exp\left(-C_{15}nt^2/s_h^2\right) + 8p\exp(-C_{16}nt^2)$$

for any $0 < t < 8M^{1-h}s_h$.

Table 1: Average computational time of various methods (in minutes)
$p = 5000$ and $n = 100$

	Model 1	Model 2	Model 3	Model 4
RLASSO(AT)	6.6	2.1	4.1	3.6
RLASSO(CLIME)	790.8	700.1	651.8	758.6
RLASSO(GLASSO)	11.0	1.7	7.6	8.2
Scout(1,1)	9.6	1.5	5.8	6.3
LASSO	1.5	0.6	1.2	1.4
LASSO+T	2.8	0.8	1.2	1.9
SLSE+T	4.2	1.5	2.3	2.0
SIS	0.6	0.4	0.6	0.5

Table 2: The average Frobenius norms of estimated $\mathbf{\Sigma}$ and $\mathbf{\Omega}$

		Model 1	Model 2	Model 3	Model 4
$\|\boldsymbol{S}-\mathbf{\Sigma}\|_F$	$\boldsymbol{S}=\boldsymbol{X}'\boldsymbol{X}/n$	618.17	2877.25	2495.43	570.22
$\|\hat{\mathbf{\Sigma}}-\mathbf{\Sigma}\|_F$	RLASSO(AT)	50.88	40.71	385.38	40.17
$\|\hat{\mathbf{\Omega}}-\mathbf{\Omega}\|_F$	RLASSO(CLIME)	41.02	273.01	49.99	42.03

Lemma S2. *Assume conditions (C1′), (C2) and (C3‴), for any $\lambda_n \to 0$, there exist positive constants C_{21}, C_{22}, C_{23} and C_{24} such that*

$$P\left(\|\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_\infty > t\right) \le C_{21}\left[\exp\left(-C_{22}n\,[t/r_q s_h]^{\frac{2}{1-q}}\right) + n^{-\frac{l-1-\tau}{2}}\right] + C_{23}p^2 s_h^{2l} t^{-2l} n^{-l} + C_{24} p t^{-2l} n^{-l},$$

for any $0 < t < 8M^{1-h}s_h$.

S3 Proofs

Proof of Lemma 1. From (1),

$$\begin{aligned}\hat{\beta}_{M_j} - \beta_{M_j} &= \frac{1}{n}\sum_{i=1}^n x_{ij}\left(\mu + \sum_{k=1}^p \beta_k x_{ik} + \sigma_i\epsilon_i\right) - \sum_{k=1}^p \mathrm{E}(x_j x_k)\beta_k \\ &= \sum_{k=1}^p\left[\frac{1}{n}\sum_{i=1}^n x_{ij}x_{ik} - \mathrm{E}(x_j x_k)\right]\beta_k + \frac{1}{n}\sum_{i=1}^n \sigma_i x_{ij}\epsilon_i + \frac{1}{n}\sum_{i=1}^n \mu x_{ij}.\end{aligned}$$

Then,

$$\begin{aligned}P\left(|\hat{\beta}_{M_j} - \beta_{M_j}| > t\right) \le\; & P\left(\sum_{k=1}^p \left|\frac{1}{n}\sum_{i=1}^n x_{ij}x_{ik} - \mathrm{E}(x_j x_k)\right| |\beta_k| > \frac{t}{3}\right) \\ & + P\left(\left|\frac{1}{n}\sum_{i=1}^n \sigma_i x_{ij}\epsilon_i\right| > \frac{t}{3}\right) + P\left(\left|\frac{1}{n}\sum_{i=1}^n \mu x_{ij}\right| > \frac{t}{3}\right)\end{aligned} \tag{S3.1}$$

Under condition (C1), applying Lemma 1 of Cai and Liu (2011) to $\pm(x_{ij}x_{ik} - \mathrm{E}(x_{ij}x_{ik}))$ gives that there exist $D_1 > 0$ and $D_2 > 0$ such that

$$\max_{jk} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_{ij}x_{ik} - \mathrm{E}(x_j x_k)\right| > t\right) \leq 2\exp(-D_1 n t^2),$$

for all $0 < t \leq D_2$. Then, under condition (C2), it follows by Bonferroni inequality that there exist $C_1 > 0$ and $C_3 > 0$ such that, for any $1 \leq j \leq p$,

$$\begin{aligned} & P\left(\sum_{k=1}^{p}\left|\frac{1}{n}\sum_{i=1}^{n} x_{ij}x_{ik} - \mathrm{E}(x_j x_k)\right| |\beta_k| > \frac{t}{3}\right) \\ \leq\; & P\left(M^{1-h}s_h \max_{k}\left|\frac{1}{n}\sum_{i=1}^{n} x_{ij}x_{ik} - \mathrm{E}(x_{ij}x_{ik})\right| > \frac{t}{3}\right) \\ \leq\; & p\cdot\max_{jk} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_{ij}x_{ik} - \mathrm{E}(x_j x_k)\right| > \frac{t}{3M^{1-h}s_h}\right) \\ \leq\; & 2p\exp(-C_1 n t^2/s_h^2), \end{aligned} \tag{S3.2}$$

for all $0 < t \leq C_3 s_h$. Similarly,

$$\max_{1\leq j\leq p} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} \sigma_i x_{ij}\epsilon_i\right| > \frac{t}{3}\right) \leq 2\exp(-C_2 n t^2), \tag{S3.3}$$

$$\max_{1\leq j\leq p} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} \mu x_{ij}\right| > \frac{t}{3}\right) \leq 2\exp(-C_2 n t^2), \tag{S3.4}$$

for some $C_2 > 0$. Therefore,

$$P\left(|\hat{\beta}_{M_j} - \beta_{M_j}| > t\right) \leq 2p\exp(-C_1 n t^2/s_h^2) + 4\exp(-C_2 n t^2). \tag{S3.5}$$

Then, Lemma 1 follows by

$$P\left(\|\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_M\|_\infty > t\right) \leq \sum_{j=1}^{p} P\left(|\hat{\beta}_{M_j} - \beta_{M_j}| > t\right).$$

□

Proof of Lemma 2. By Karush-Kuhn-Tucker conditions, the solution $\tilde{\boldsymbol{\beta}}$ to (8) satisfies that

$$\hat{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_M = -\lambda_n \boldsymbol{Z}, \tag{S3.6}$$

where $\boldsymbol{Z}$ has the form of

$$\boldsymbol{Z} = \begin{cases} 1, & \text{if } \tilde{\beta}_j > 0; \\ -1, & \text{if } \tilde{\beta}_j < 0; \\ \in [-1, 1], & \text{if } \tilde{\beta}_j = 0. \end{cases} \tag{S3.7}$$

Simple algebra from (S3.6) yields

$$\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1}[\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_M - \lambda_n \boldsymbol{Z} - (\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\beta} - (\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})]. \tag{S3.8}$$

Hence,

$$\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty \leq v_p(\|\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_M\|_\infty + \|\lambda_n \boldsymbol{Z}\|_\infty + \|(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\beta}\|_\infty + \|(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_\infty).$$

Equivalently,

$$\frac{1}{v_p}\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty - \|(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_\infty \leq \|\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_M\|_\infty + \|\lambda_n \boldsymbol{Z}\|_\infty + \|(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\beta}\|_\infty$$

Then, by $\|(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_\infty \leq \|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_1 \|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty$, it holds that

$$\begin{aligned}
&P(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty > t) \\
&= P\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty > t, \|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_1 \leq \frac{1}{2v_p}\right) + P\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty > t, \|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_1 > \frac{1}{2v_p}\right) \\
&\leq P\left(\|\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_M\|_\infty + \|\lambda_n \boldsymbol{Z}\|_\infty + \|(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\beta}\|_\infty > \frac{t}{2v_p}, \|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_1 \leq \frac{1}{2v_p}\right) \\
&\quad + P\left(\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_1 > \frac{1}{2v_p}\right) \\
&\leq P\left(\|\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_M\|_\infty > \frac{t}{6v_p}\right) + P\left(\|\lambda_n \boldsymbol{Z}\|_\infty > \frac{t}{6v_p}\right) + P\left(\|(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\beta}\|_\infty > \frac{t}{6v_p}\right) \\
&\quad + P\left(\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_1 > \frac{1}{2v_p}\right) \\
&:= I + II + III + IV
\end{aligned}$$

By Lemma 1, there exist positive constants C_4 and C_5 such that

$$I \leq 2p^2 \exp(-C_4 nt^2/(s_h v_p)^2) + 4p \exp(-C_5 nt^2/v_p^2). \tag{S3.9}$$

By the choice of λ_n, $II = 0$, when n is sufficiently large. For III, under condition (C2), $\|(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\beta}\|_\infty \leq \|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_1 \|\boldsymbol{\beta}\|_\infty \leq M\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_1$. Then, it follows from Theorem 1(i) of Cai and Liu (2011) that,

$$III + IV \leq 2P\left(\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_1 > \frac{t}{6Mv_p}\right) \leq C_6 n^{-1/2} p^{-(\delta-2)} (r_q v_p/t)^{1/(1-q)}, \tag{S3.10}$$

for some $C_6 > 0$. This completes the proof of the lemma. □

Proof of Theorem 1. Note that $\widehat{\mathcal{M}}_{\boldsymbol{\beta}} = \mathcal{M}_{\tilde{\boldsymbol{\beta}},t_n}$ and

$$\begin{aligned}
& P\left(\mathcal{M}_{\boldsymbol{\beta},a_n t_n} \subset \widehat{\mathcal{M}}_{\boldsymbol{\beta}}\right) \\
&= 1 - P\left(\cup_{j:|\beta_j|>a_n t_n}\left\{|\tilde{\beta}_j| \le t_n\right\}\right) \\
&\ge 1 - P\left(\cup_{j:|\beta_j|>a_n t_n}\left\{|\tilde{\beta}_j - \beta_j| > (a_n - 1)t_n\right\}\right) \\
&\ge 1 - P\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty > (a_n - 1)t_n\right) \\
&\ge 1 - O\left[\exp\left(-C_7(\log n)^{-2} n^{1-2\alpha_1-2\alpha_3-2\eta}\right) + (1/p)^{\delta-2}\left((\log n)(1/n)^{\frac{1-q}{2}-\alpha_2-\alpha_3-\eta}\right)^{1/(1-q)}\right]
\end{aligned}$$

by Lemma 2 and the choice of a_n and t_n. Similarly,

$$\begin{aligned}
& P(\widehat{\mathcal{M}}_{\boldsymbol{\beta}} \subset \mathcal{M}_{\boldsymbol{\beta},t_n/a_n}) \\
&= P\left(\cap_{j:|\beta_j|\le t_n/a_n}\left\{|\tilde{\beta}_j| \le t_n\right\}\right) \\
&\ge 1 - P\left(\cup_{j:|\beta_j|\le t_n/a_n}\left\{|\tilde{\beta}_j - \beta_j| > (1 - a_n^{-1})t_n\right\}\right) \\
&\ge 1 - P\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty > (1 - a_n^{-1})t_n\right) \\
&= 1 - O\left[\exp\left(-C_7(\log n)^{-2} n^{1-2\alpha_1-2\alpha_3-2\eta}\right) + (1/p)^{\delta-2}\left((\log n)(1/n)^{\frac{1-q}{2}-\alpha_2-\alpha_3-\eta}\right)^{1/(1-q)}\right].
\end{aligned}$$

This completes the proof of the first part of Theorem 1. In particular, if we choose $h = 0$,

$$\begin{aligned}
& P(\mathcal{M}_{\boldsymbol{\beta}} \subset \widehat{\mathcal{M}}_{\boldsymbol{\beta}}) \\
&= 1 - P\left(\cup_{j\in\mathcal{M}_{\boldsymbol{\beta}}}\left\{|\tilde{\beta}_j| \le t_n\right\}\right) \\
&\ge 1 - P\left(\cup_{j\in\mathcal{M}_{\boldsymbol{\beta}}}\left\{|\beta_j| - |\tilde{\beta}_j - \beta_j| \le t_n\right\}\right) \\
&\ge 1 - P\left(\cup_{j\in\mathcal{M}_{\boldsymbol{\beta}}}\left\{|\tilde{\beta}_j - \beta_j| \ge t_n/2\right\}\right) \\
&= 1 - O\left[\exp\left(-C_8 n^{1-2\alpha_1-2\alpha_3-2\eta}\right) + (1/p)^{\delta-2}(1/n)^{(\frac{1-q}{2}-\alpha_2-\alpha_3-\eta)/(1-q)}\right],
\end{aligned} \tag{S3.11}$$

since under (C5), by the choice of t_n, $\min_{j\in\mathcal{M}_{\boldsymbol{\beta}}}|\beta_j| > \frac{3}{2}t_n$ for large enough n.

On the other hand,

$$\begin{aligned}
& P(\widehat{\mathcal{M}}_{\boldsymbol{\beta}} \subset \mathcal{M}_{\boldsymbol{\beta}}) \\
&= 1 - P\left(\cup_{j\notin\mathcal{M}_{\boldsymbol{\beta}}}\left\{|\tilde{\beta}_j| > t_n\right\}\right) \\
&= 1 - P\left(\cup_{j\notin\mathcal{M}_{\boldsymbol{\beta}}}\left\{|\tilde{\beta}_j - \beta_j| > t_n\right\}\right) \\
&= 1 - O\left[\exp\left(-C_8 n^{1-2\alpha_1-2\alpha_3-2\eta}\right) + (1/p)^{\delta-2}(1/n)^{(\frac{1-q}{2}-\alpha_2-\alpha_3-\eta)/(1-q)}\right].
\end{aligned} \tag{S3.12}$$

(S3.11) and (S3.12) together prove the theorem. □

Proof of Lemma 3. The proof is analogous to that of Lemma 1. Under condition (C1′), it holds that

$$\mathrm{E}\left|x_{ij}x_{ik}-\mathrm{E}(x_jx_k)\right|^{2l}\le 2^{2l-1}\left[\mathrm{E}|x_{ij}x_{ik}|^{2l}+(\mathrm{E}|x_{ij}x_{ik}|)^{2l}\right]=O(1).$$

Then, by Chebyshev Inequality and Theorem 2 in Whittle (1960).

$$\max_{jk}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}x_{ij}x_{ik}-\mathrm{E}(x_jx_k)\right|>t\right)\le t^{-2l}\mathrm{E}\left|\frac{1}{n}\sum_{i=1}^{n}x_{ij}x_{ik}-\mathrm{E}(x_jx_k)\right|^{2l}=O(t^{-2l}n^{-l}).$$

Therefore, by replacing (S3.2) with

$$\begin{aligned}&P\left(\sum_{k=1}^{p}\left|\frac{1}{n}\sum_{i=1}^{n}x_{ij}x_{ik}-\mathrm{E}(x_jx_k)\right||\beta_k|>\frac{t}{3}\right)\\ &\le p\cdot\max_{jk}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}x_{ij}x_{ik}-\mathrm{E}(x_jx_k)\right|>\frac{t}{3M^{1-h}s_h}\right)\\ &\le C_9ps_h^{2l}t^{-2l}n^{-l},\end{aligned}$$

for some $C_9>0$ and replacing (S3.3) and (S3.4) with

$$\max_{1\le j\le p}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_ix_{ij}\epsilon_i\right|>\frac{t}{3}\right)\le\frac{C_{10}}{2}t^{-2l}n^{-l},$$
$$\max_{1\le j\le p}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\mu x_{ij}\right|>\frac{t}{3}\right)\le\frac{C_{10}}{2}t^{-2l}n^{-l},$$

for some $C_{10}>0$, the rest of proof follows from (S3.1). □

Proof of Lemma 4. From Lemma 3, it holds that

$$P\left(\|\hat{\boldsymbol{\beta}}_M-\boldsymbol{\beta}_M\|_\infty>\frac{t}{6v_p}\right)\le C_{11}p^2s_h^{2l}v_p^{2l}t^{-2l}n^{-l}+C_{12}pv_p^{2l}t^{-2l}n^{-l},\tag{S3.13}$$

for some $C_{11}>0$ and $C_{12}>0$.

Under condition (C3′), it follows from Theorem 1(ii) of Cai and Liu (2011) that, there exists $C_{13}>0$ that

$$P\left(\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_1>\frac{t}{6Mv_p}\right)\le C_{13}\left(n^{-1/2}p^{-(\delta-2)}(r_qv_p/t)^{1/(1-q)}+n^{-\frac{l-1-\tau}{2}}\right).\tag{S3.14}$$

Replacing (S3.9) and (S3.10) with (S3.13) and (S3.14) and observing that by the choice of λ_n, $P(\lambda_n\boldsymbol{Z}>\frac{t}{6v_p})=0$, when n is sufficiently large, the rest of the proof resembles the proof of Lemma 2. □

Proof of Theorem 2. The proof is the same as the proof of Theorem 1 by replacing results in Lemma 2 with results in Lemma 4. □

Proof of Lemma S1. From (S3.6),

$$\tilde{\boldsymbol\beta} = \hat{\boldsymbol\Omega}\hat{\boldsymbol\beta}_M - \lambda_n \hat{\boldsymbol\Omega}\boldsymbol Z.$$

Recall that, $\boldsymbol\beta = \boldsymbol\Omega\boldsymbol\beta_M$. Hence,

$$\tilde{\boldsymbol\beta} - \boldsymbol\beta = \hat{\boldsymbol\Omega}\hat{\boldsymbol\beta}_M - \boldsymbol\Omega\boldsymbol\beta_M - \lambda_n \hat{\boldsymbol\Omega}\boldsymbol Z.$$

Then,

$$P(\|\tilde{\boldsymbol\beta} - \boldsymbol\beta\|_\infty > t) \le P(\|\hat{\boldsymbol\Omega}\hat{\boldsymbol\beta}_M - \boldsymbol\Omega\boldsymbol\beta_M\|_\infty > t/2) + P(\|\lambda_n\hat{\boldsymbol\Omega}\boldsymbol Z\|_\infty > t/2). \tag{S3.15}$$

Since

$$\hat{\boldsymbol\Omega}\hat{\boldsymbol\beta}_M - \boldsymbol\Omega\boldsymbol\beta_M = (\hat{\boldsymbol\Omega} - \boldsymbol\Omega)\hat{\boldsymbol\beta}_M + \boldsymbol\Omega(\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M),$$

it holds that

$$P(\|\hat{\boldsymbol\Omega}\hat{\boldsymbol\beta}_M - \boldsymbol\Omega\boldsymbol\beta_M\|_\infty > t/2) \le P(\|(\hat{\boldsymbol\Omega} - \boldsymbol\Omega)\hat{\boldsymbol\beta}_M\|_\infty > t/4) + P(\|\boldsymbol\Omega(\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M)\|_\infty > t/4). \tag{S3.16}$$

The first item in (S3.16) is bounded by

$$\begin{aligned}
&P\left(\|\hat{\boldsymbol\Omega} - \boldsymbol\Omega\|_1\|\hat{\boldsymbol\beta}_M\|_\infty > t/4\right)\\
\le\, &P\left(\|\hat{\boldsymbol\Omega} - \boldsymbol\Omega\|_1\|\hat{\boldsymbol\beta}_M\|_\infty > t/4 \cap \|\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M\|_\infty \le t/8\right) + P(\|\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M\|_\infty > t/8)\\
\le\, &P\left(\|\hat{\boldsymbol\Omega} - \boldsymbol\Omega\|_1\|\boldsymbol\beta_M\|_\infty > t/8 \cap \|\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M\|_\infty \le t/8\right)\\
&+ P\left(\|\hat{\boldsymbol\Omega} - \boldsymbol\Omega\|_1\|\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M\|_\infty > t/8 \cap \|\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M\|_\infty \le t/8\right)\\
&+ P\left(\|\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M\|_\infty > t/8\right)\\
\le\, &P\left(\|\hat{\boldsymbol\Omega} - \boldsymbol\Omega\|_1 > t/[8M^{1-h}s_h]\right) + P\left(\|\hat{\boldsymbol\Omega} - \boldsymbol\Omega\|_1 > 1\right) + P\left(\|\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M\|_\infty > t/8\right).
\end{aligned}$$

For the second item in (S3.16), it follows from the assumption $\|\boldsymbol\Omega\|_1 \le M$ that

$$\begin{aligned}
P(\|\boldsymbol\Omega(\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M)\|_\infty > t/4) &\le P\left(\|\boldsymbol\Omega\|_1\|\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M\|_\infty > t/4\right)\\
&\le P\left(\|\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M\|_\infty > t/[4M]\right).
\end{aligned}$$

Without loss of generality, assume $M \ge 2$. Then, for any $0 < t < 8M^{1-h}s_h$,

$$\begin{aligned}
P\left(\|\tilde{\boldsymbol\beta} - \boldsymbol\beta\|_\infty > t\right) &\le 2P\left(\|\hat{\boldsymbol\Omega} - \boldsymbol\Omega\|_1 > t/[8M^{1-h}s_h]\right)\\
&\quad + 2P\left(\|\hat{\boldsymbol\beta}_M - \boldsymbol\beta_M\|_\infty > t/[4M]\right),
\end{aligned} \tag{S3.17}$$

since $\|\lambda_n\hat{\boldsymbol\Omega}\boldsymbol Z\|_\infty \le |\lambda_n|\|\hat{\boldsymbol\Omega}\|_1\|\boldsymbol Z\|_\infty \le |\lambda_n|\|\hat{\boldsymbol\Omega}\|_1 \le 2|\lambda_n|\|\boldsymbol\Omega\|_1 \le 2M|\lambda_n|$. By the choice of λ_n, when n is sufficiently large, $P(\|\lambda_n\hat{\boldsymbol\Omega}\boldsymbol Z\|_\infty > t/2) = 0$.

Under (C1), it follows by Theorem 1(a) of Cai, Liu, and Luo (2011) that

$$P\left(\|\hat{\boldsymbol\Omega} - \boldsymbol\Omega\|_1 > t/[8M^{1-h}s_h]\right) \le 4\exp\left(-C_{14}n[t/s_hr_q]^{2/(1-q)}\right), \tag{S3.18}$$

for some $C_{14} > 0$.

From Lemma 1, it holds that

$$P(\|\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_M\|_\infty > t/(4M)) \le 2p^2 \exp\left(-C_{15} n t^2 / s_h^2\right) + 4p \exp\left(-C_{16} n t^2\right). \tag{S3.19}$$

(S3.17), (S3.18) together with (S3.19) proves the lemma. □

Proof of Lemma S2. Under conditions (C1′) and (C3‴), by Theorem 1(ii) of Cai, Liu, and Luo (2011).

$$P\left(\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\|_1 > \frac{t}{8M^{1-h} s_h}\right) \le \frac{C_{21}}{2}\left[\exp\left(-C_{22} n \left[t/s_h r_q\right]^{\frac{2}{1-q}}\right) + n^{-\frac{l-1-\tau}{2}}\right].$$

From Lemma 3, it holds that

$$P(\|\hat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_M\|_\infty > t/(4M)) \le \frac{C_{23}}{2} p^2 s_h^{2l} t^{-2l} n^{-l} + \frac{C_{24}}{2} p t^{-2l} n^{-l}.$$

The rest of proof follows by (S3.17). □

Proof of Theorem 3 and Theorem 4. By using results in Lemma S1 and Lemma S2, the proof is the same as that of Theorem 1. □

Proof of (15). Decompose $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{I}_{s_0} & \boldsymbol{\Sigma}_{21}' \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{I}_{p-s_0} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{21} = (\boldsymbol{B}' \ \boldsymbol{0})'$.

$$\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1} = \boldsymbol{S}_{21}(\boldsymbol{S}_{11}^{-1} - \boldsymbol{I}_{s_0}) + (\boldsymbol{S}_{21} - \boldsymbol{\Sigma}_{21})\boldsymbol{I}_{s_0} + \boldsymbol{\Sigma}_{21}. \tag{S3.20}$$

It is well known (e.g. see Bickel and Levina (2008)) that for normally distributed covariates,

$$\max_{1\le i,j\le p} |s_{ij} - \rho_{ij}| = O_P\left(\sqrt{n^{-1}\log p}\right),$$

where s_{ij} is the (i,j)th element of $\boldsymbol{S}$. Then,

$$\|\boldsymbol{S}_{21} - \boldsymbol{\Sigma}_{21}\|_\infty = \max_{s_0 < i \le p} \sum_{j=1}^{s_0} |s_{ij} - \rho_{ij}| = O_P\left(s_0\sqrt{n^{-1}\log p}\right).$$

Hence, $\|(\boldsymbol{S}_{21} - \boldsymbol{\Sigma}_{21})\boldsymbol{I}_{s_0}\|_\infty = O_P\left(s_0\sqrt{n^{-1}\log p}\right)$.

Moreover, $\boldsymbol{S}_{11}^{-1} - \boldsymbol{I}_{s_0} = \boldsymbol{S}_{11}^{-1}(\boldsymbol{I}_{s_0} - \boldsymbol{S}_{11})$. Then,

$$\|\boldsymbol{S}_{11}^{-1} - \boldsymbol{I}_{s_0}\|_\infty \le \|\boldsymbol{S}_{11}^{-1}\|_\infty \|\boldsymbol{I}_{s_0} - \boldsymbol{S}_{11}\|_\infty \le (1 + \|\boldsymbol{S}_{11}^{-1} - \boldsymbol{I}_{s_0}\|_\infty)\|\boldsymbol{I}_{s_0} - \boldsymbol{S}_{11}\|_\infty.$$

Hence, $\|\boldsymbol{S}_{11}^{-1} - \boldsymbol{I}_{s_0}\|_\infty \le \|\boldsymbol{I}_{s_0} - \boldsymbol{S}_{11}\|_\infty / (1 - \|\boldsymbol{I}_{s_0} - \boldsymbol{S}_{11}\|_\infty) = O_P\left(s_0\sqrt{n^{-1}\log p}\right)$. Then,

$$\begin{aligned}
\|\boldsymbol{S}_{21}(\boldsymbol{S}_{11}^{-1} - \boldsymbol{I}_{s_0})\|_\infty &\le \|\boldsymbol{S}_{21}\|_\infty \|\boldsymbol{S}_{11}^{-1} - \boldsymbol{I}_{s_0}\|_\infty \\
&\le (\|\boldsymbol{\Sigma}_{21}\|_\infty + \|\boldsymbol{S}_{21} - \boldsymbol{\Sigma}_{21}\|_\infty)\|\boldsymbol{S}_{11}^{-1} - \boldsymbol{I}_{s_0}\|_\infty \\
&= O_P\left(s_0 + s_0\sqrt{n^{-1}\log p}\right) \cdot O_P\left(s_0\sqrt{n^{-1}\log p}\right) \\
&= O_P\left(s_0^2\sqrt{n^{-1}\log p}\right)
\end{aligned}$$

Since $\|\boldsymbol{\Sigma}_{21}\|_\infty \ge \|\boldsymbol{B}\|_\infty \ge 1 + 2\gamma$, under the assumptions that $n^{-1}\log p \to 0$ and s_0 is fixed, it follows from (S3.20) that $P(\|\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\|_\infty \ge 1 + \gamma) \to 1$. □

References

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics* **36,** 2577–2604.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* **106,** 672–684.

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

Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability and its Applications* **5,** 302.