SPARSE COVARIANCE THRESHOLDING FOR HIGH-DIMENSIONAL VARIABLE SELECTION

X. Jessie Jeng and Z. John Daye

University of Pennsylvania

Supplementary Material

In this section, we provide detailed proofs of Lemma 7.2 and Theorem 3.3.

S1 Proof of Lemma 7.2

The proof is similar to that of Theorem 1 in Bickel and Levina (2008a) and Theorem 1 in Rothman, Levina and Zhu (2009).

We show (7.31) first. Let $\sigma_{ij}^{\nu} = s_{\nu}(\sigma_{ij})$, then

$$\|\hat{\mathbf{\Sigma}}_{CS}^{\nu} - \mathbf{\Sigma}_{CS}\|_{\infty} = \max_{i\in C}\sum_{j\in S}|\hat{\sigma}_{ij}^{\nu} - \sigma_{ij}| \leq \max_{i\in C}\sum_{j\in S}|\hat{\sigma}_{ij}^{\nu} - \sigma_{ij}^{\nu}| + \max_{i\in C}\sum_{j\in S}|\sigma_{ij}^{\nu} - \sigma_{ij}|.$$

By properties in (2.4), the last term is bounded from above by $\nu \cdot d_{CS}^*$. For the second to the last term, we perform the following decomposition,

$$\begin{aligned}\max_{i\in C}\sum_{j\in S}|\hat{\sigma}_{ij}^{\nu} - \sigma_{ij}^{\nu}| &\leq \max_{i\in C}\sum_{j\in S}\left(|\hat{\sigma}_{ij} - \sigma_{ij}| + |\hat{\sigma}_{ij}^{\nu} - \hat{\sigma}_{ij}| + |\sigma_{ij}^{\nu} - \sigma_{ij}|\right)1(|\hat{\sigma}_{ij}| > \nu, |\sigma_{ij}| > \nu) \\ &+ \max_{i\in C}\sum_{j\in S}|\hat{\sigma}_{ij}^{\nu}|1(|\hat{\sigma}_{ij}| > \nu, |\sigma_{ij}| \leq \nu) + \max_{i\in C}\sum_{j\in S}|\sigma_{ij}^{\nu}|1(|\hat{\sigma}_{ij}| \leq \nu, |\sigma_{ij}| > \nu).\end{aligned}$$

The three terms on the right-hand side of the above inequality are denoted by IV, V, and VI, respectively. By Lemma 7.1, it is easy to see that

$$\max_{i\in C, j\in S}|\hat{\sigma}_{ij} - \sigma_{ij}| \leq O_p\left(\sqrt{\log(s(p-s))}/\sqrt{n}\right). \quad \text{(S1.1)}$$

Therefore,

$$\begin{aligned}\max_{i\in C}\sum_{j\in S}|\hat{\sigma}_{ij} - \sigma_{ij}|1(|\hat{\sigma}_{ij}| > \nu, |\sigma_{ij}| > \nu) &\leq \max_{i\in C, j\in S}|\hat{\sigma}_{ij} - \sigma_{ij}|\max_{i\in C}\sum_{j\in S}1(|\sigma_{ij}| > \nu) \\ &\leq O_p\left(d_{CS}^*\sqrt{\log(s(p-s))}/\sqrt{n}\right)\end{aligned}$$

and

$$\text{IV} \quad \leq \quad O_p\left(d_{CS}^*\sqrt{\log(s(p-s))}/\sqrt{n}\right) + 2\nu \cdot d_{CS}^*.$$

Next, we consider V.

$$\begin{aligned}
\text{V} \quad &\leq \quad \max_{i\in C}\sum_{j\in S}|\hat{\sigma}_{ij}^{\nu} - \sigma_{ij}|1(|\hat{\sigma}_{ij}| > \nu, |\sigma_{ij}| \leq \nu) + \max_{i\in C}\sum_{j\in S}|\sigma_{ij}|1(|\sigma_{ij}| \leq \nu) \\
&\leq \quad \max_{i\in C}\sum_{j\in S}\left(|\hat{\sigma}_{ij}^{\nu} - \hat{\sigma}_{ij}| + |\hat{\sigma}_{ij} - \sigma_{ij}|\right)1(|\hat{\sigma}_{ij}| > \nu, |\sigma_{ij}| \leq \nu) + \nu \cdot d_{CS}^* \\
&\leq \quad \left(\nu + \max_{i\in C, j\in S}|\hat{\sigma}_{ij} - \sigma_{ij}|\right)\max_{i\in C}\left\{\sum_{j\in S}1(|\sigma_{ij}| > \gamma_1\nu) + \sum_{j\in S}1(|\sigma_{ij}| \leq \gamma_1\nu, |\hat{\sigma}_{ij}| > \nu)\right\} + \nu \cdot d_{CS}^* \\
&\leq \quad \left(\nu + \max_{i\in C, j\in S}|\hat{\sigma}_{ij} - \sigma_{ij}|\right)\left(d_{CS}^* + \max_{i\in C}\sum_{j\in S}1(|\hat{\sigma}_{ij} - \sigma_{ij}| > (1-\gamma_1)\nu)\right) + \nu \cdot d_{CS}^*,
\end{aligned}$$

where $\gamma_1 < 1$ and the last inequality above follows from the fact that $1(|\sigma_{ij}| \leq \gamma_1\nu, |\hat{\sigma}_{ij}| > \nu) \leq 1(|\hat{\sigma}_{ij} - \sigma_{ij}| > (1-\gamma_1)\nu)$. Since it is clear that $\max_{i\in C}\sum_{j\in S}1(|\hat{\sigma}_{ij} - \sigma_{ij}| > (1-\gamma_1)\nu)$ is negligible when ν is chosen to be greater than $C\sqrt{\log(s(p-s))}/\sqrt{n}$ for some C large enough, then, by (S1.1),

$$\text{V} \leq O_p\left(d_{CS}^*\sqrt{\log(s(p-s))}/\sqrt{n}\right) + 2\nu \cdot d_{CS}^*.$$

Finally,

$$\begin{aligned}
\text{VI} \quad &\leq \quad \max_{i\in C}\sum_{j\in S}\left(|\sigma_{ij}^{\nu} - \sigma_{ij}| + |\sigma_{ij}|\right)1(|\hat{\sigma}_{ij}| \leq \nu, |\sigma_{ij}| > \nu) \\
&\leq \quad \nu d_{CS}^* + \max_{i\in C, j\in S}|\sigma_{ij}|\max_{i\in C}\left\{\sum_{j\in S}1(|\sigma_{ij}| \leq \gamma_2\nu) + \sum_{j\in S}1(|\hat{\sigma}_{ij}| \leq \nu, |\sigma_{ij}| > \gamma_2\nu)\right\} \\
&\leq \quad (1 + M\gamma_2)\,\nu d_{CS}^* + \max_{i\in C, j\in S}|\sigma_{ij}|\max_{i\in C}\sum_{j\in S}1(|\hat{\sigma}_{ij} - \sigma_{ij}| > (\gamma_2 - 1)\nu) \\
&\leq \quad O_p\left(\nu \cdot d_{CS}^*\right),
\end{aligned}$$

where $\gamma_2 > 1$ and the last inequality follows from the fact that the second term is negligible when ν is chosen to be greater than $C\sqrt{\log(s(p-s))}/\sqrt{n}$ for some C large enough. Summing up the bounds for IV, V, and VI gives (7.31).

The proof for (7.32) is similar as above, thus omitted. □

S2 Proof of Theorem 3.3

Denote A as the event in (3.11), B as the event in (3.12), and

$$C = \left\{\Lambda_{min}\left(\hat{\Sigma}_{SS}^{\nu}\right) > \frac{1}{2}\Lambda_{min}(\mathbf{\Sigma}_{SS})\right\}.$$

Then,
$$P\left(sgn(\hat{\beta}) = sgn(\hat{\beta}^*)\right) \geq P\left(A \cap B \cap C\right).$$
Define events
$$D = \left\{ \left\| \hat{\Sigma}^{\nu}_{CS} (\hat{\Sigma}^{\nu}_{SS})^{-1} \right\|_{\infty} \leq 1 - \frac{\epsilon}{2} \right\},$$
$$E = \left\{ \left\| \frac{1}{n} X_S^T \epsilon \right\|_{\infty} \leq \frac{\epsilon}{4}\lambda_n - s\nu\bar{\rho} \right\}, \qquad F = \left\{ \left\| \frac{1}{n} X_C^T \epsilon \right\|_{\infty} \leq \frac{\epsilon}{4}\lambda_n - s\nu\bar{\rho} \right\}.$$
Then,
$$A \supseteq \left\{ D \cap E \cap F \right\}.$$
Define event
$$G = \left\{ \left\| (\hat{\Sigma}^{\nu}_{SS})^{-1} \right\|_{\infty} \leq \frac{4\underline{\rho}}{5\epsilon\lambda_n} \right\}.$$
Then,
$$B \supseteq \{G \cap E\}.$$
To sum up the above, we have
$$\begin{aligned} P\left(sgn(\hat{\beta}) = sgn(\hat{\beta}^*)\right) &\geq P\left(C \cap D \cap E \cap F \cap G\right) \\ &\geq 1 - P\left(C^c\right) - P\left(D^c\right) - P\left(E^c\right) - P\left(F^c\right) - P\left(G^c\right). \end{aligned}$$
In the following Lemma, we restate the results of Lemma 7.2 in probabilistic terms. The conditions of dimension parameters are reexpressed in the convergence rates of the probabilities. The proof is essentially the same as that of Lemma 7.2.

Lemma S2.1
$$P\left(\left\| \hat{\Sigma}^{\nu}_{CS} - \Sigma_{CS} \right\|_{\infty} > t_B \right) \leq \exp(\log(s(p-s)) - nt_H^2/C')(1+o(1)), \tag{S2.1}$$
where
$$t_H = \frac{t_B}{4d^*_{CS}} - \frac{(5 + M\gamma_2)\nu}{2}, \qquad \gamma_2 > 1,$$
and
$$P\left(\left\| \hat{\Sigma}^{\nu}_{SS} - \Sigma_{SS} \right\|_{\infty} > t_C \right) \leq \exp(2\log s - nt_I^2/C')(1+o(1)), \tag{S2.2}$$
where
$$t_I = \frac{t_C}{4d^*_{SS}} - \frac{(5 + M\gamma_2)\nu}{2}.$$

Using Lemma S2.1, we can calculate $P\left(C^c\right)$, $P\left(G^c\right)$, and $P\left(D^c\right)$, respectively. First of all,
$$\begin{aligned} P\left(C^c\right) &= p\left(\Lambda_{min}\left(\hat{\Sigma}^{\nu}_{SS}\right) \leq \frac{1}{2}\Lambda_{min}(\mathbf{\Sigma}_{SS})\right) \\ &\leq P\left(\left\| \hat{\Sigma}^{\nu}_{SS} - \Sigma_{SS} \right\|_2 > \frac{1}{2}\Lambda_{min}(\mathbf{\Sigma}_{SS}) \right) \\ &\leq P\left(\left\| \hat{\Sigma}^{\nu}_{SS} - \Sigma_{SS} \right\|_{\infty} > \frac{1}{2}\Lambda_{min}(\mathbf{\Sigma}_{SS}) \right) \\ &\leq \exp(2\log s - \alpha_3 n(n^{-c_2} - n^{-c_1})^2)(1+o(1)), \end{aligned}$$

where the last inequality is by (S2.2) and the values of ν and d_S^*. Secondly,

$$\begin{aligned}
P(G^c) &= P\left(\left\|(\hat{\Sigma}_{SS}^{\nu})^{-1}\right\|_\infty > \frac{4\underline{\rho}}{5\epsilon\lambda_n}\right)\\
&\leq P\left(\left\|(\hat{\Sigma}_{SS}^{\nu})^{-1} - (\Sigma_{SS})^{-1}\right\|_\infty > \frac{4\underline{\rho}}{5\epsilon\lambda_n} - \bar{D}\right)\\
&\leq P\left(\left\|\hat{\Sigma}_{SS}^{\nu} - \Sigma_{SS}\right\|_\infty > \alpha_4\right)\\
&\leq \exp(2\log s - \alpha_5 n(n^{-c_2} - n^{-c_1})^2)(1+o(1)),
\end{aligned}$$

where the second inequality is by (7.42). Thirdly,

$$\begin{aligned}
P(D^c) &= P\left(\left\|\hat{\Sigma}_{CS}^{\nu}(\hat{\Sigma}_{SS}^{\nu})^{-1}\right\|_\infty > 1 - \frac{\epsilon}{2}\right)\\
&\leq P\left(\|\mathrm{I}\|_\infty + \|\mathrm{II}\|_\infty > \frac{\epsilon}{2}\right)\\
&\leq P\left(\left\|\hat{\Sigma}_{CS}^{\nu} - \Sigma_{CS}\right\|_\infty > \alpha_6\right) + P\left(d_{CS}^*\left\|\hat{\Sigma}_{SS}^{\nu} - \Sigma_{SS}\right\|_\infty > \alpha_7\right),
\end{aligned}$$

where the last inequality is by (7.44), (7.46), and (7.42). By (S2.1),

$$P\left(\left\|\hat{\Sigma}_{CS}^{\nu} - \Sigma_{CS}\right\|_\infty > \alpha_6\right) \leq \exp(\log(s(p-s)) - \alpha_8 n(n^{-c_2} - n^{-c_1})^2)(1+o(1)),$$

and, by (S2.2),

$$P\left(d_{CS}^*\left\|\hat{\Sigma}_{SS}^{\nu} - \Sigma_{SS}\right\|_\infty > \alpha_7\right) \leq \exp(2\log s - \alpha_9 n(n^{-2c_2} - n^{-c_1})^2)(1+o(1)).$$

To sum up the above, we have

$$P(C^c)+P(G^c)+P(D^c) \leq O\left(\exp(2\log s - \alpha_9 n(n^{-2c_2} - n^{-c_1})^2)\right) \leq O\left(\exp(-\alpha_2 n(n^{-2c_2} - n^{-c_1})^2)\right).$$

Next, we calculate $P(E^c)$ and $P(F^c)$. Since $\epsilon \sim N(0, \sigma^2 I)$, then, given X, we can apply the standard results on the extreme value of multivariate normal as follows:

$$\begin{aligned}
P(E^c) &\leq E\left(\exp(\log s - \alpha_{10} n\lambda_n^2/(4\max_{j\in S}\hat{\sigma}_{jj}))\right)\\
&\leq \exp(-\alpha_1 n^{1-2c}) + \exp(\log s)P\left(\max_{1\leq j\leq p}\hat{\sigma}_{jj} > M/2\right)\\
&\leq \exp(-\alpha_1 n^{1-2c}) + \exp(\log sp - nM/16)\\
&\leq O\left(\exp(-\alpha_1 n^{1-2c})\right).
\end{aligned}$$

Similarly,

$$P(F^c) \leq \exp(-\alpha_1 n^{1-2c}) + \exp(\log(p-s)p - nM/16) \leq O\left(\exp(-\alpha_1 n^{1-2c})\right).$$

Result follows by summing up the above. □

S3 Proof of Lemma S2.1

Similar to the proof of Lemma 7.2, we have

$$\begin{aligned}\left\|\hat{\Sigma}^{\nu}_{CS} - \Sigma_{CS}\right\|_{\infty} &\leq (5 + M\gamma_2)\nu d^*_{CS} + 2d^*_{CS} \max_{i\in C, j\in S} |\hat{\sigma}_{ij} - \sigma_{ij}| \\ &+ (\nu + \max_{i\in C, j\in S} |\hat{\sigma}_{ij} - \sigma_{ij}|) \max_{i\in C} N_i(1-\gamma_1) + M \max_{i\in C} N_i(\gamma_2 - 1).\end{aligned}$$

Then,

$$\begin{aligned}P\left(\left\|\hat{\Sigma}^{\nu}_{CS} - \Sigma_{CS}\right\|_{\infty} > t_B\right) &\leq P\left((5 + M\gamma_2)\nu d^*_{CS} + 2d^*_{CS} \max_{i\in C, j\in S} |\hat{\sigma}_{ij} - \sigma_{ij}| > t_B/2\right) \\ &+ P\left((\nu + \max_{i\in C, j\in S} |\hat{\sigma}_{ij} - \sigma_{ij}|) \max_{i\in C} N_i(1-\gamma_1) > t_B/4\right) \\ &+ P\left(M \max_{i\in C} N_i(\gamma_2 - 1) > t_B/4\right) \\ &= P\left(\max_{i\in C, j\in S} |\hat{\sigma}_{ij} - \sigma_{ij}| > t_H\right)(1 + o(1)),\end{aligned}$$

and (S2.1) follows by Lemma 7.1.

(S2.2) can be derived in a similar way.

□