Supplement to "Adaptive Thresholding for Sparse Covariance Matrix Estimation"

Tony Cai[1] and Weidong Liu[1,2]

Proof of Lemma 2. Without loss of generality, we assume that $\mathsf{E}\mathbf{X} = 0$ and $\mathsf{Var}(X_i) = 1$ for $1 \leq i \leq p$. We first prove (i). Let

$$\tilde{\theta}_{ij} = \frac{1}{n}\sum_{k=1}^{n}\Big[X_{ki}X_{kj} - \widetilde{\sigma}_{ij}\Big]^2 \quad \text{with } \widetilde{\sigma}_{ij} = \tfrac{1}{n}\textstyle\sum_{k=1}^{n} X_{ki}X_{kj}.$$

We shall show that for any $M > 0$, there exists a constant C_1 such that

$$\mathsf{P}\Big(\max_{ij}|\hat{\theta}_{ij} - \tilde{\theta}_{ij}| \geq C_1\sqrt{\log p/n}\Big) = O(p^{-M}). \qquad (1)$$

To prove (1), we write

$$\begin{aligned}\hat{\theta}_{ij} \;=\; & \tilde{\theta}_{ij} + \frac{2}{n}\sum_{k=1}^{n}\Big[X_{ki}X_{kj} - \widetilde{\sigma}_{ij}\Big]\Big[-X_{ki}\bar{X}^j - X_{kj}\bar{X}^i + 2\bar{X}^i\bar{X}^j\Big] \\ & + \frac{1}{n}\sum_{k=1}^{n}\Big[-X_{ki}\bar{X}^j - X_{kj}\bar{X}^i + 2\bar{X}^i\bar{X}^j\Big]^2. \qquad (2)\end{aligned}$$

By the simple inequality $s^2e^s \leq e^{2s}$ for $s > 0$, we have $\mathsf{E}X_{ki}^2 e^{t|X_{ki}|} \leq C_\eta K_1 t^{-2}$ for $t \leq \eta^{1/2}$. It follows from the inequality (24) and (C1) that for any $M > 0$, there exists a constant C_2 such that

$$\mathsf{P}\Big(\max_{i}|\bar{X}^i| \geq C_2\sqrt{\log p/n}\Big) = O(p^{-M}). \qquad (3)$$

[1]Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, tcai@wharton.upenn.edu. The research was supported in part by NSF FRG Grant DMS-0854973.
[2]Department of Mathematics and Institute of Natural Sciences, Shanghai Jiao Tong University, China.

Let

$$Y_{kij} = X_{ki}^2 X_{kj}, \quad \bar{Y}_{kij} = X_{ki}^2 \bar{X}_{kj}, \quad \bar{X}_{kj} = X_{kj} I\{|X_{kj}| \le C_3 \sqrt{\log(p+n)}\},$$

where C_3 satisfies $C_3^2 \eta > M + 1$. Then for any $C_4 > 0$,

$$\begin{aligned} \mathsf{P}\Big(\max_{ij}\Big|\sum_{k=1}^n Y_{kij}\Big| \ge C_4 n\Big) &\le \mathsf{P}\Big(\max_{ij}\Big|\sum_{k=1}^n \bar{Y}_{kij}\Big| \ge C_4 n\Big) + np\max_i \mathsf{P}\Big(|X_i| \ge C_3\sqrt{\log(p+n)}\Big) \\ &= \mathsf{P}\Big(\max_{ij}\Big|\sum_{k=1}^n \bar{Y}_{kij}\Big| \ge C_4 n\Big) + O(p^{-M}). \end{aligned} \quad (4)$$

Let $t = \tau(\log(n+p))^{-1/2}$ and $x = ((M+2)\log p)^{1/2}$ in Lemma 1 with $\tau > 0$ sufficiently small. We have $\bar{B}_n^2 = O(1) n \max_{i,j} (\mathsf{E}Y_{kij}^4)^{1/2} (\mathsf{E}e^{2C_3 \tau X_{ki}^2})^{1/2} = O(n)$. By (C1) and $p \le \exp(n^{1/2})$, we can let C_4 be sufficiently large such that $C_4 n \ge 2C_t \bar{B}_n x$ and $C_4 > 2\max_{ij} \mathsf{E}X_{1i}^2 |X_{1j}|$. It follows from Lemma 1 that

$$\begin{aligned} \mathsf{P}\Big(\max_{ij}\Big|\sum_{k=1}^n \bar{Y}_{kij}\Big| \ge C_4 n\Big) &\le \mathsf{P}\Big(\max_{ij}\Big|\sum_{k=1}^n (\bar{Y}_{kij} - \mathsf{E}\bar{Y}_{kij})\Big| \ge C_4 n/2\Big) \\ &= O(p^{-M}). \end{aligned} \quad (5)$$

Combining (3)-(5), we see that for any $M > 0$, there exists $C_5 > 0$ such that

$$\mathsf{P}\Big(\max_{ij} \frac{1}{n}\Big|\sum_{k=1}^n X_{ki}^2 X_{kj} \bar{X}^j\Big| \ge C_5\sqrt{\frac{\log p}{n}}\Big) = O(p^{-M}). \quad (6)$$

Similar inequalities can be proved for other terms in (2), and hence (1) is proved.

Write

$$\tilde{\theta}_{ij} - \theta_{ij} = \frac{1}{n}\sum_{k=1}^n \Big[(X_{ki}X_{kj})^2 - \mathsf{E}(X_{ki}X_{kj})^2\Big] - \widetilde{\sigma}_{ij}^2 + (\sigma_{ij}^0)^2 - (\widetilde{\sigma}_{ij} - \sigma_{ij}^0)^2.$$

By Lemma 1 and (C1), we see that

$$\mathsf{P}\Big(\max_{ij} |\widetilde{\sigma}_{ij} - \sigma_{ij}^0| \ge C_6\sqrt{\log p/n}\Big) = O(p^{-M}). \quad (7)$$

Take $t = \tau(\log(n+p))^{-1}$ and $x = ((M+2)\log p)^{1/2}$ in Lemma 1. Since $p = \exp(o(n^{1/3}))$, we have $n\varepsilon \ge C_t\sqrt{n}x$ for any $\varepsilon > 0$. Thus by some similar truncation arguments in (4) and

(5), it can be shown that for any $\varepsilon > 0$,

$$\mathsf{P}\Big(\max_{ij}\Big|\frac{1}{n}\sum_{k=1}^{n}\Big[(X_{ki}X_{kj})^2 - \mathsf{E}(X_{ki}X_{kj})^2\Big]\Big| \geq \varepsilon\Big) = O(p^{-M}). \tag{8}$$

Combining (1), (7) and (8) yields that for any $\varepsilon > 0$ and $M > 0$,

$$\mathsf{P}\Big(\max_{ij}\{|\tilde{\theta}_{ij} - \theta_{ij}| + |\hat{\theta}_{ij} - \theta_{ij}|\} \geq \varepsilon\Big) = O(p^{-M}). \tag{9}$$

By (13) and $\mathsf{Var}(X_i) = 1$, we see that $\min_{i,j}\theta_{ij} \geq \tau_0$ which implies

$$\mathsf{P}\Big(\min_{i,j}\tilde{\theta}_{ij} \geq \tau_0/2\Big) \geq 1 - O(p^{-M}). \tag{10}$$

By (1), (3) and (10), it is easy to show that

$$\begin{aligned}
&\mathsf{P}\Big(\max_{ij}|\hat{\sigma}_{ij} - \sigma^0_{ij}|/\hat{\theta}^{1/2}_{ij} \geq \delta\sqrt{\log p/n}\Big) \\
&\leq \mathsf{P}\Big(\max_{ij}\Big\{(n\tilde{\theta}_{ij})^{-1/2}\Big|\sum_{k=1}^{n}(X_{ki}X_{kj} - \sigma^0_{ij})\Big|\Big\} \geq \delta\sqrt{(1 - C_7\sqrt{\log p/n})\log p}\Big) \\
&\quad +O(p^{-M})
\end{aligned} \tag{11}$$

with some $C_7 > 0$ and any $M > 0$. Applying Theorem 2.2 and equation (2.2) in Shao (1999) to the second probability in (11), we have for $\delta \geq 0$,

$$\mathsf{P}\Big(\max_{ij}|\hat{\sigma}_{ij} - \sigma^0_{ij}|/\hat{\theta}^{1/2}_{ij} \geq \delta\sqrt{\log p/n}\Big) = O((\log p)^{-1/2}p^{-\delta+2}).$$

To prove (ii), we only need to show (1), (3) and (8) hold under (C2) with $O(p^{-M})$ being replaced by $O(p^{-M} + n^{-\epsilon/8})$. Let

$$\check{X}_{ki} = X_{ki}I\{|X_{ki}| \leq (n/(\log n)^2)^{1/4}\}.$$

Then we have

$$\begin{aligned}
\mathsf{P}\Big(\max_{i}|\bar{X}^i| \geq C_2\sqrt{\log p/n}\Big) \;\leq\; & \mathsf{P}\Big(\max_{i}\Big|\sum_{k=1}^{n}(\check{X}_{ki} - \mathsf{E}\check{X}_{ki})\Big| \geq 2^{-1}C_2\sqrt{n\log p}\Big) \\
& +np\max_{i}\mathsf{P}\Big(|X_{1i}| \geq (n/(\log n)^2)^{1/4}\Big)
\end{aligned}$$

$$= \quad O(p^{-M}+n^{-\epsilon/8}), \tag{12}$$

where in the last inequality we used Bernstein's inequality (cf. Bennett (1962)) and (C2). Recall Y_{kj} and define $\check{Y}_{kij} = \check{X}_{ki}^2\check{X}_{kj}$. Using Bernstein's inequality again, we have

$$\begin{aligned} \mathsf{P}\Big(\max_{ij}\Big|\sum_{k=1}^{n} Y_{kij}\Big| \geq C_4 n\Big) &\leq \mathsf{P}\Big(\max_{ij}\Big|\sum_{k=1}^{n}(\check{Y}_{kij}-\mathsf{E}\check{Y}_{kij})\Big| \geq 2^{-1}C_4 n\Big) + O(n^{-\epsilon/8}) \\ &= O(p^{-M}+n^{-\epsilon/8}). \end{aligned}$$

Therefore, (6) holds under (C2). Replacing $O(p^{-M})$ with $O(p^{-M}+n^{-\epsilon/8})$, the inequalities (7) and (8) can be similarly proved. Finally, applying Theorem 2.2 and (2.2) in Shao (1999) to the second probability in (11), we complete the proof of (ii). ∎

Proof of Lemma 4. Let $s_1 = Ms_0(p)$ with $M > 0$ being a sufficiently large number. Let

$$\begin{aligned} A^{(i)}_{j_1\cdots j_{s_1}} &= \cap_{k=1}^{s_1}\{|\hat{\sigma}_{ij_k}| \geq \lambda_{nij_k}(\delta)\}, \\ B_i &= \{j : \sigma^0_{ij} = 0; j \neq i\}. \end{aligned}$$

We will show that for any $\delta > \sqrt{2}$,

$$\mathsf{P}\Big(\cup_{i=1}^{p}\cup_{j_1\cdots j_{s_1}\in B_i} A^{(i)}_{j_1\cdots j_{s_1}}\Big) = O(p^{-C_\delta M}) \tag{13}$$

for some $C_\delta > 0$, which implies that with probability $1 - O(p^{-C_\delta M})$, for each i, there are at most s_1 nonzero numbers of $\{|\hat{\sigma}_{ij}|; j \in B_i\}$ and by Lemma 2, they are of order $O(\max_i \sigma^0_{ii}\sqrt{\log p/n})$. This together with (44) proves (36). Let D denote the subset of $\{j_1, \cdots, j_{s_1}\}$ such that the random variables $\{X_i : i \in D\}$ are pairwise uncorrelated. Let $k = \max\{Card(D)\}$ be the largest number of X_j's with $j \in \{j_1, \cdots, j_{s_1}\}$ such that they are uncorrelated. Suppose the lower bound for k is k_0. Then we can write the set

$$\begin{aligned} \{(j_1,\cdots,j_{s_1}) : j_1,\cdots,j_{s_1} \in B_i\} &= \cup_{k=k_0}^{s_1}\{(j_1,\cdots,j_{s_1}) : j_1,\cdots,j_{s_1} \in B_i, \max\{Card(D)\} = k\} \\ &=: \cup_{k=k_0}^{s_1} B_{i,k}. \end{aligned} \tag{14}$$

As in the proof of Theorem 3, we can show that $k_0 \geq M$. The number of elements in $B_{i,k}$ is no more than $(ks)^{s_1} C_p^k$. Define

$$\hat{A}^{(i)}_{j_1 \cdots j_{s_1}} = \bigcap_{k=1}^{s_1} \Big\{ \Big| \sum_{l=1}^{n} Y_{lij_k} \Big| \geq \delta \sqrt{n \log p} \Big\},$$

where $\hat{Y}_{lij_k} = \theta_{ij_k}^{-1/2} X_{li} X_{lj_k}$. To prove (13), we only need to show that for any $\delta > \sqrt{2}$,

$$\mathsf{P}\Big(\cup_{i=1}^{p} \cup_{j_1 \cdots j_{s_1} \in B_i} \hat{A}^{(i)}_{j_1 \cdots j_{s_1}} \Big) = O(p^{-C_\delta M}) \tag{15}$$

for some $C_\delta > 0$. Without loss of generality we assume that $\mathsf{E} X_{j_k} = 0$ and $\mathsf{E} X_{j_k}^2 = 1$. By Lemma 1, we have for any $\varepsilon > 0$,

$$\mathsf{P}\Big(\max_i \Big| \frac{\sum_{k=1}^n Y_{ki}^2}{n} - 1 \Big| \leq \varepsilon \Big) = O(p^{-M})$$

for any $M > 0$. Thus it suffices to prove that for any $\delta > \sqrt{2}$,

$$\sum_{i=1}^{p} \sum_{j_1 \cdots j_{s_1} \in B_i} \mathsf{P}\Big(\bigcap_{k=1}^{s_1} C_k \Big) = O(p^{-C_\delta M}),$$

where

$$C_k = \Big\{ \Big| \frac{\sum_{l=1}^n X_{li} X_{lj_k}}{\sqrt{\sum_{l=1}^n X_{li}^2}} \Big| \geq \delta \sqrt{\log p} \Big\}.$$

Note that X_i and $\{X_{j_1}, \ldots, X_{j_{s_1}}\}$ are independent. So by (14) and conditioning on $\{X_{li}, 1 \leq l \leq n\}$, we can get

$$\sum_{i=1}^{p} \sum_{j_1 \cdots j_{s_1} \in B_i} \mathsf{P}\Big(\bigcap_{k=1}^{s_1} C_k \Big) \leq Cp \sum_{k=k_0}^{s_1} (ks)^{s_1} C_p^k p^{-\delta^2 k/2} = O(p^{-C_\delta M})$$

for some $C_\delta > 0$. This proves (13).

To prove (37), we have for any $M > 0$ in (36),

$$\begin{aligned} \mathsf{E}\|\hat{\Sigma}^\star(\delta) - \Sigma_0\|_2^2 &\leq C_{\gamma,\delta,K,M} s_0^2(p) \frac{\log p}{n} \\ &\quad + \mathsf{E}\|\hat{\Sigma}^\star(\delta) - \Sigma_0\|_2^2 I\{\|\hat{\Sigma}^\star(\delta) - \Sigma_0\|_2 > C_{\gamma,\delta,M} \max_i \sigma_{ii}^0 s_0(p) \Big(\frac{\log p}{n} \Big) \} \end{aligned}$$

$$\begin{aligned}
&\le C_{\gamma,\delta,K,M} s_0^2(p)\frac{\log p}{n} \\
&\quad +2\mathsf{E}\|\Sigma_n-\Sigma_0\|_2^2 I\{\|\hat\Sigma^\star(\delta)-\Sigma_0\|_2 > C_{\gamma,\delta,M}\max_i \sigma_{ii}^0 s_0(p)\Big(\frac{\log p}{n}\Big)\} \\
&\quad +2\mathsf{E}\|\hat\Sigma^\star(\delta)-\Sigma_n-\Sigma_0\|_2^2 I\{\|\hat\Sigma^\star(\delta)-\Sigma_0\|_2 > C_{\gamma,\delta,M}\max_i \sigma_{ii}^0 s_0(p)\Big(\frac{\log p}{n}\Big)\}.
\end{aligned}$$

It is easy to show that $\mathsf{E}\|\Sigma_n-\Sigma_0\|_2^4 \le c_0\max_i(\sigma_{ii}^0)^4 p^5/n^2$, where c_0 is an absolute constant. Note that $\max_{i,j}\mathsf{E}\hat\theta_{ij}^4 \le c_0\max_i(\sigma_{ii}^0)^8$. Then by Lemma 2,

$$\begin{aligned}
\mathsf{E}\|\hat\Sigma^\star(\delta)-\Sigma_n-\Sigma_0\|_2^4 &\le c_0\Big(\max_i(\sigma_{ii}^0)^4p^4+\mathsf{E}\max_i(\sum_{j=1}^p\lambda_{nij}(\delta))^4\Big) \\
&\le c_0\max_i(\sigma_{ii}^0)^4\Big(p^4+p^5\Big(\frac{\log p}{n}\Big)^2\Big) \\
&\quad +c_0p^5\max_{ij}\mathsf{E}\lambda_{nij}^4(\delta)I\{|\hat\theta_{ij}-\theta_{ij}|\ge\max_i(\sigma_{ii}^0)^2\} \\
&\le C\max_i(\sigma_{ii}^0)^4\Big(p^5+p^5\Big(\frac{\log p}{n}\Big)^2+p^5\Big(\frac{\log p}{n}\Big)^2p^{-M}\Big).
\end{aligned}$$

This implies that for $M=5+\xi^{-1}$,

$$\begin{aligned}
\mathsf{E}\|\hat\Sigma^\star(\delta)-\Sigma_0\|_2^2 &\le C\Big(s_0^2(p)\frac{\log p}{n}+p^{5/2-M/2}\log p\Big) \\
&\le Cs_0^2(p)\frac{\log p}{n}. \quad \blacksquare
\end{aligned}$$

Proof of Lemma 5. Take $l=[p^{\tau_2}]$ with $2\epsilon_0+\tau^2/2<\tau_2<1$. Then there exist independent variables $X_{i_0},\dots,X_{i_l}$, where $i_0=i$ and $i_1,\cdots,i_l\in B_i=\{j:\sigma_{ij}^0=0;j\ne i\}$. To prove the result, it suffices to prove (38). In fact, by (38) and the inequality $\|A\|_2\ge\max_i(\sum_{j=1}^p a_{ij}^2)^{1/2}$ for a symmetric matrix $A=(a_{ij})$, we have with probability tending to 1,

$$\|\hat\Sigma^\star(\tau)-\Sigma_0\|_2\ge Cp^{\epsilon_0}\Big(\frac{\log p}{n}\Big)^{1/2}\ge Cp^{\epsilon_0/2}s_0(p)\Big(\frac{\log p}{n}\Big)^{1/2}.$$

Split the set $\{i_1,\cdots,i_l\}$ into $p^{2\epsilon_0}$ subsets $H_1,\cdots,H_{p^{2\epsilon_0}}$ with the same cardinality $[p^{\tau_2-2\epsilon_0}]$. Note that $\tau_2-2\epsilon_0>\tau^2/2$. By Lemma 2, it suffices to show that for some $\epsilon>0$,

$$\mathsf{P}\Big(\min_{i,m}\sum_{j\in H_m}I\Big\{\Big|\sum_{k=1}^nX_{ki}X_{kj}\Big|\ge(\tau+\epsilon)\sqrt{n\log p}\Big\}\ge1\Big)\to1, \tag{16}$$

where we assume that $\mathsf{E}X_j = 0$ and $\mathsf{E}X_j^2 = 1$. As in the proof of Lemma 4, it suffices to show that for some $\epsilon > 0$, $\mathsf{P}\Big(\min_{i,m} \sum_{j\in H_m} I\{C_j\} \geq 1\Big) \to 1$, where

$$C_j = \Big\{\Big|\frac{\sum_{k=1}^n X_{ki}Y_{kj}}{\sqrt{\sum_{k=1}^n X_{ki}^2}}\Big| \geq (\tau+\epsilon)\sqrt{\log p}\Big\}.$$

By conditioning on $\{X_{ki}; 1 \leq k \leq n\}$, we can get

$$\begin{aligned}\mathsf{P}\Big(\bigcup_{j\in H_m} C_j\Big) &\geq 1-(1-p^{-(\tau+2\epsilon)^2/2})^{|H_m|} - O(p^{-M}) \\ &\geq 1-\exp\Big(-|H_m|p^{-(\tau+2\epsilon)^2/2}\Big) - O(p^{-M}),\end{aligned}$$

where $|H_m| = [p^{\tau_2-2\epsilon_0}]$. This implies (16) by letting ϵ satisfy $\tau_2 - 2\epsilon_0 > (\tau+2\epsilon)^2/2$. ∎

References

[1] Bennett, G (1962). Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association* 57: 33-45.

J. Amer. Statist. Assoc. 104: 177-186.

[2] Shao, Q.M. (1999). A Cramér type large deviation result for Student's t-statistic. *Journal of Theoretical Probability* 12: 385-398.