Sure Screening for Gaussian Graphical Models

Shikai Luo SLUO@NCSU.EDU
Chengchun Shi CSHI4@NCSU.EDU
Rui Song RSONG@NCSU.EDU
Department of Statistics
North Carolina State University
Raleigh, NC 27695, USA

Yuxiang Xie YXXIE@UW.EDU
Daniela Witten DWITTEN@UW.EDU
Departments of Statistics & Biostatistics
University of Washington
Seattle, WA 98195, USA

Editor:

Appendix A. Further Examination of Assumptions 1 and 3

Assumptions 1 and 3 place conditions on the elements of Σ corresponding to elements in $\mathcal{E}$ and $\mathcal{E}^c$. However, it is not immediately clear what types of matrices Σ satisfy these assumptions. In order to clarify this point, we now present three propositions that allow us to recast Assumptions 1 and 3 as conditions on the elements of Σ^{-1}. Define $\alpha = \Lambda_{\max}(\Sigma)/\Lambda_{\min}(\Sigma)$ and $\nu = 2/\left\{\Lambda_{\max}(\Sigma)^{-1} + \Lambda_{\min}(\Sigma)^{-1}\right\}$.

Lemma 1 *Assume $\alpha > 1$. For an integer $M \geq \{\kappa \log n + \log \Lambda_{\max}(\Sigma) - \log(C_1/2)\}/\log\{(\alpha+1)/(\alpha-1)\}$, let $B = \nu \sum_{i=0}^{M-1}(I - \nu\Sigma^{-1})^i$. If $\min_{(j,l)\in\mathcal{E}} |B_{jl}| \geq 2C_1 n^{-\kappa}$, then Assumption 1 holds. If Assumption 1 holds, then $\min_{(j,l)\in\mathcal{E}} |B_{jl}| \geq C_1 n^{-\kappa}/2$.*

Lemma 2 *Assume $\alpha > 1$. For an integer $L \geq \{(1-\xi)\log n + \log \Lambda_{\max}(\Sigma)\}/\log\{(\alpha+1)/(\alpha-1)\}$, let $F = \nu \sum_{i=0}^{L-1}\left(I - \nu\Sigma^{-1}\right)^i$. Then $\max_{(j,l)\in\mathcal{E}^c} |F_{jl}| = o\left\{n^{-(1-\xi)/2}\right\}$ if and only if Assumption 3 holds.*

Appendix B. Technical Proofs

B.1 Proof of Theorem 1

Proof One can show (see, e.g., Lemma 1 of Ravikumar et al. (2011)) that

$$\mathrm{pr}\left(\left|X_j^T X_l/n - \sigma_{jl}\right| \geq C_1 n^{-\kappa}/3\right) \leq c_1 \exp(-c_2 n^{1-2\kappa}). \tag{1}$$

Next, we notice that

$$\mathrm{pr}(\mathcal{E} \not\subseteq \widehat{\mathcal{E}}_{\gamma_n}) = \mathrm{pr}\left\{\bigcup_{(j,l)\in\mathcal{E}} \left(\left|X_j^T X_l/n\right| < 2C_1 n^{-\kappa}/3\right)\right\} \leq \sum_{(j,l)\in\mathcal{E}} \mathrm{pr}\left(\left|X_j^T X_l/n\right| < 2C_1 n^{-\kappa}/3\right).$$

Furthermore, $|\mathcal{E}| < p^2$, and for $(j,l) \in \mathcal{E}$, Assumption 1 implies that

$$\mathrm{pr}(|X_j^T X_l/n| < 2C_1 n^{-\kappa}/3) \leq \mathrm{pr}\left(\left|X_j^T X_l/n - \sigma_{jl}\right| \geq C_1 n^{-\kappa}/3\right).$$

Hence, $\mathrm{pr}(\mathcal{E} \not\subseteq \widehat{\mathcal{E}}_{\gamma_n}) \leq p^2 c_1 \exp(-c_2 n^{1-2\kappa})$. This implies the first part of Theorem 1. Conversely, if $\min_{(j,l)\in\mathcal{E}} |\sigma_{jl}| < C_1 n^{-\kappa}/3$, then there exists $(j',l') \in \mathcal{E}$ with $|\sigma_{j'l'}| < C_1 n^{-\kappa}/3$. This together with (1) implies

$$\mathrm{pr}(\mathcal{E} \subseteq \widehat{\mathcal{E}}_{\gamma_n}) \leq \mathrm{pr}\left\{(j',l') \in \widehat{\mathcal{E}}_{\gamma_n}\right\} \leq \mathrm{pr}\left(\left|X_{j'}^T X_{l'}/n - \sigma_{j'l'}\right| \geq C_1 n^{-\kappa}/3\right) \leq c_1 \exp(-c_2 n^{1-2\kappa}),$$

so that the result holds. ■

B.2 Proof of Lemma 1

Proof It suffices to show that with high probability, graphical sure screening will select no edges between $x_j^{(s)}$ and $x_l^{(t)}$ for all $s \neq t$. This is the case when the event $\left(|X_j^T X_l/n - \sigma_{jl}| \leq C_1 n^{-\kappa}/3\right)$ holds for all $j \neq l$. As was shown in the proof of Theorem 1, $\mathrm{pr}\left\{\bigcap_{j\neq l}\left(|X_j^T X_l/n - \sigma_{jl}| \leq C_1 n^{-\kappa}/3\right)\right\} \geq 1 - C_4 \exp\left(-C_5 n^{1-2\kappa}\right)$. ■

B.3 Preliminaries to Proof of Theorem 2

The following lemma will be used in the proof of Theorem 2.

Lemma 3 *Let $x = (x_1, \ldots, x_p)^T$ be a p-dimensional random vector with mean zero and variance Σ, and let Y be any random variable with $E(Y) = 0$ and $E(Y^2) = 1$. Define $\mathcal{S} = \{j : |E(Yx_j)| > Cn^{-\kappa}\}$ for some constant C. Then $|\mathcal{S}|$, the cardinality of $\mathcal{S}$, satisfies $|\mathcal{S}| \leq C^{-2} n^{2\kappa} \Lambda_{\max}(\Sigma)$.*

Proof Let $\beta = \Sigma^{-1} E(xY)$, and define $\epsilon = Y - x^T\beta$. Then

$$E(x\epsilon) = E\left\{x(Y - x^T\beta)\right\} = E(xY) - E(xx^T)\Sigma^{-1}E(xY) = 0. \tag{2}$$

By the definition of β, we have that $E(Yx_j) = (\Sigma\beta)_j$, the jth element of the vector $\Sigma\beta$. Consequently,

$$\mathcal{S} = \{j : |(\Sigma\beta)_j| > Cn^{-\kappa}\} = \{j : (\Sigma\beta)_j^2 > C^2 n^{-2\kappa}\}. \tag{3}$$

Furthermore,

$$||\Sigma\beta||_2^2 = (\Sigma^{1/2}\beta)^T \Sigma (\Sigma^{1/2}\beta) \leq \Lambda_{\max}(\Sigma) ||\Sigma^{1/2}\beta||_2^2 = \Lambda_{\max}(\Sigma)\beta^T\Sigma\beta.$$

Moreover, recalling from (2) that x and ϵ are uncorrelated, we have that $\beta^T\Sigma\beta = \mathrm{var}(x^T\beta) = \mathrm{var}(Y) - \mathrm{var}(\epsilon) \leq 1$. Thus, we conclude that $||\Sigma\beta||_2^2 \leq \Lambda_{\max}(\Sigma)$. By (3), this implies that $|\mathcal{S}| \leq \Lambda_{\max}(\Sigma)/(C^2 n^{-2\kappa}) = C^{-2} n^{2\kappa} \Lambda_{\max}(\Sigma)$. ■

B.4 Proof of Theorem 2

Proof Let $\mathcal{S}_j = (l : l \neq j, |\sigma_{jl}| \geq C_1 n^{-\kappa}/3)$ and $\mathcal{T}_{j,\gamma_n} = \bigcap_{l:l\neq j} \left(|X_j^T X_l/n - \sigma_{jl}| \leq C_1 n^{-\kappa}/3 \right)$. By definition, $\widehat{\mathcal{E}}_{j,\gamma_n} = \left(l : l \neq j, |X_j^T X_l/n| > 2C_1 n^{-\kappa}/3 \right)$. Then on the set $\mathcal{T}_{j,\gamma_n}$, if l belongs to $\widehat{\mathcal{E}}_{j,\gamma_n}$, then $l \in \mathcal{S}_j$. Thus, we conclude that $\mathrm{pr}(\widehat{\mathcal{E}}_{j,\gamma_n} \subseteq \mathcal{S}_j) \geq \mathrm{pr}(\mathcal{T}_{j,\gamma_n})$. Moreover, an argument similar to that in the proof of Theorem 1 can be used to show that $\mathrm{pr}(\mathcal{T}_{j,\gamma_n}) \geq 1 - C_4 \mathrm{exp}(-C_5 n^{1-2\kappa})$. This implies that $\mathrm{pr}(\widehat{\mathcal{E}}_{j,\gamma_n} \subseteq \mathcal{S}_j) \geq 1 - C_4 \mathrm{exp}(-C_5 n^{1-2\kappa})$. Finally, applying Lemma 3 in conjunction with Assumption 2 yields the desired result. ∎

B.5 Proof of Theorem 3

Proof First, we will show that the assumptions of Theorem 1 are satisfied, so that the sure screening property applies. We must simply show that the new threshold, $\gamma_n = n^{-1/2}\Phi^{-1}\left[1 - m/\left\{p(p-1)\right\}\right]$, is no greater than $2C_1 n^{-\kappa}/3$, the threshold used in Theorem 1. In other words, we must show that

$$\frac{m}{p(p-1)} \geq 1 - \Phi(2C_1 n^{1/2-\kappa}/3). \tag{4}$$

From the fact that $1 - \Phi(x) \leq (2\pi)^{-1/2} x^{-1} \exp(-x^2/2)$, we have that $1 - \Phi(2C_1 n^{1/2-\kappa}/3) \leq C_7 n^{-1/2+\kappa} \exp(-C_8 n^{1-2\kappa})$. Furthermore, since $\log(p) = C_3 n^{\xi}$, we have that $m/\{p(p-1)\} \geq C_9 \exp(-C_{10} n^{\xi})$. Using the fact that $\xi < 1 - 2\kappa$, (4) follows directly.

Next, we show that the threshold value $\gamma_n = n^{-1/2}\Phi^{-1}\left[1 - m/\left\{p(p-1)\right\}\right]$ leads to control of the asymptotic expected false positive rate at $2m/\left\{p(p-1)\right\}$. Recall that the false positive rate is defined as

$$\mathrm{fpr}_n = \frac{1}{|\mathcal{E}^c|} \sum_{(j,l)\in\mathcal{E}^c} 1\left(\left|X_j^T X_l/n\right| > \gamma_n \right).$$

Because $E(X_j^T X_l/n) = \sigma_{jl}$ and $\mathrm{var}(X_j^T X_l/n) = (1+\sigma_{jl}^2)/n$, it follows that $\{(1+\sigma_{jl}^2)/n\}^{-1/2}\left(X_j^T X_l/n - \sigma_{jl}\right)$ converges in distribution to a standard normal. Furthermore, for any $(j,l) \in \mathcal{E}^c$, we have

$$\begin{aligned}
\mathrm{pr}\left(\left|X_j^T X_l/n\right| > \gamma_n\right) &= \mathrm{pr}\left[\left\{(1+\sigma_{jl}^2)/n\right\}^{-1/2}\left(X_j^T X_l/n - \sigma_{jl}\right) > \left\{(1+\sigma_{jl}^2)/n\right\}^{-1/2}(\gamma_n - \sigma_{jl})\right] \\
&+ \mathrm{pr}\left[\left\{(1+\sigma_{jl}^2)/n\right\}^{-1/2}\left(X_j^T X_l/n - \sigma_{jl}\right) < -\left\{(1+\sigma_{jl}^2)/n\right\}^{-1/2}(\gamma_n + \sigma_{jl})\right] \\
&= 1 - \Phi\left[\left\{(1+\sigma_{jl}^2)/n\right\}^{-1/2}(\gamma_n - \sigma_{jl})\right] + 1 - \Phi\left[\left\{(1+\sigma_{jl}^2)/n\right\}^{-1/2}(\gamma_n + \sigma_{jl})\right] \\
&\asymp 2 - 2\Phi(n^{1/2}\gamma_n) = 2m/\{p(p-1)\},
\end{aligned}$$

where the asymptotic equivalence in the previous line follows from the fact that $n^{1/2}\gamma_n = \Phi^{-1}[1 - m/\{p(p-1)\}]$ is of the same order as $n^{\xi/2}$, combined with Assumption 3.

Consequently, the expectation of fpr_n is controlled as desired,

$$E(\mathrm{fpr}_n) = \frac{1}{|\mathcal{E}^c|} \sum_{(j,l)\in\mathcal{E}^c} \mathrm{pr}\left(\left|X_j^T X_l/n\right| > \gamma_n\right) \asymp \frac{\sum_{(j,l)\in\mathcal{E}^c} 2m/\left\{p(p-1)\right\}}{|\mathcal{E}^c|} = 2m/\{p(p-1)\} \leq m/|\mathcal{E}^c|,$$

where the last inequality is due to the fact that $|\mathcal{E}^c| \leq p(p-1)/2$. ∎

B.6 Proof of Lemma 1

Proof

For convenience, we let $\Omega = \Sigma^{-1}$. Note that

$$\|I - \nu\Omega\|_2 = \max\left(\left|\nu\Lambda_{\max}(\Sigma)^{-1} - 1\right|, \left|\nu\Lambda_{\min}(\Sigma)^{-1} - 1\right|\right) = \frac{\alpha - 1}{\alpha + 1} < 1, \tag{5}$$

where $\|\cdot\|_2$ denotes the largest singular value of a matrix. The Neumann series of $I - \nu\Omega$ is of the form

$$\Sigma = \underbrace{\nu \sum_{i=0}^{M-1} (I - \nu\Omega)^i}_{B} + \underbrace{\nu \sum_{i=M}^{\infty} (I - \nu\Omega)^i}_{A}. \tag{6}$$

Then, using (5), we can show that

$$\|A\|_2 \leq |\nu| \sum_{i=M}^{\infty} \left\|(I - \nu\Omega)^i\right\|_2 = \Lambda_{\max}(\Sigma) \left(\frac{\alpha - 1}{\alpha + 1}\right)^M. \tag{7}$$

Since any matrix element is smaller in magnitude than the matrix's largest singular value, and recalling that $M \geq \{\kappa \log n + \log \Lambda_{\max}(\Sigma) - \log(C_1/2)\} / \log\{(\alpha+1)/(\alpha-1)\}$, we see that

$$|A_{jl}| \leq \|A\|_2 \leq \Lambda_{\max}(\Sigma) \left(\frac{\alpha - 1}{\alpha + 1}\right)^M \leq C_1 n^{-\kappa}/2. \tag{8}$$

Now, assume that

$$\min_{(j,l)\in\mathcal{E}} |B_{jl}| \geq 2C_1 n^{-\kappa}. \tag{9}$$

Together, (9), (6), and (8) imply that for any $(j, l) \in \mathcal{E}$,

$$|\sigma_{jl}| \geq |B_{jl}| - |A_{jl}| \geq 2C_1 n^{-\kappa} - C_1 n^{-\kappa}/2 \geq C_1 n^{-\kappa}.$$

Conversely, (6) and (8) imply that under Assumption 1, for any $(j, l) \in \mathcal{E}$,

$$|B_{jl}| \geq |\sigma_{jl}| - |A_{jl}| \geq C_1 n^{-\kappa} - C_1 n^{-\kappa}/2 = C_1 n^{-\kappa}/2.$$

∎

B.7 Proof of Lemma 2

Proof

Using the arguments and notation from the proof of Proposition 1, it can be shown that

$$\Sigma = \underbrace{\nu \sum_{i=0}^{L-1} (I - \nu\Omega)^i}_{F} + \underbrace{\nu \sum_{i=L}^{\infty} (I - \nu\Omega)^i}_{D}, \tag{10}$$

and that for $L \geq \{(1-\xi)\log n + \log \Lambda_{\max}(\Sigma)\} / \log\{(\alpha+1)/(\alpha-1)\}$,

$$|D_{jk}| \leq \|D\|_2 \leq \Lambda_{\max}(\Sigma)\left(\frac{\alpha-1}{\alpha+1}\right)^L = n^{-(1-\xi)}. \tag{11}$$

Assume that

$$\max_{(j,l)\in\mathcal{E}^c} |F_{jl}| = o\left\{n^{-(1-\xi)/2}\right\}. \tag{12}$$

Together, (10), (11), and (12) imply that for $(j,l) \in \mathcal{E}^c$,

$$|\sigma_{jl}| \leq |F_{jl}| + |D_{jl}| = o\left\{n^{-(1-\xi)/2}\right\}.$$

Conversely, (10) and (11) imply that under Assumption 3, for any $(j,l) \in \mathcal{E}^c$,

$$|F_{jl}| \leq |\sigma_{jl}| + |D_{jl}| = o\left\{n^{-(1-\xi)/2}\right\}.$$

■

Appendix C. Sure Screening Property in Finite Samples

Theorem 1 states that under certain conditions, with $\gamma_n \propto n^{-\kappa}$, $\mathrm{pr}\left(\mathcal{E} \subseteq \widehat{\mathcal{E}}_{\gamma_n}\right)$ approaches one. To investigate this in finite samples, we obtained a Monte Carlo estimate of $\mathrm{pr}\left(\mathcal{E} \subseteq \widehat{\mathcal{E}}_{\gamma_n}\right)$ by repeatedly simulating data and computing the fraction of simulated data sets for which $\mathcal{E} \subseteq \widehat{\mathcal{E}}_{\gamma_n}$. We did this under Simulations A–D, for a range of values of n, and for three values of p. Results shown in Fig. 1 indicate that under Simulations A–C, for sufficiently large sample sizes, $\mathrm{pr}\left(\mathcal{E} \subseteq \widehat{\mathcal{E}}_{\gamma_n}\right)$ is arbitrarily close to one. This is to be expected, since in these three settings, elements of $\mathcal{E}$ correspond to non-zero elements in Σ, as required by Assumption 1. In contrast, in Simulation D, $\mathrm{pr}\left(\mathcal{E} \subseteq \widehat{\mathcal{E}}_{\gamma_n}\right)$ is near zero regardless of sample size; this is because many elements of $\mathcal{E}$ are zero in Σ.

References

RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. & YU, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics* **5**, 935–980.

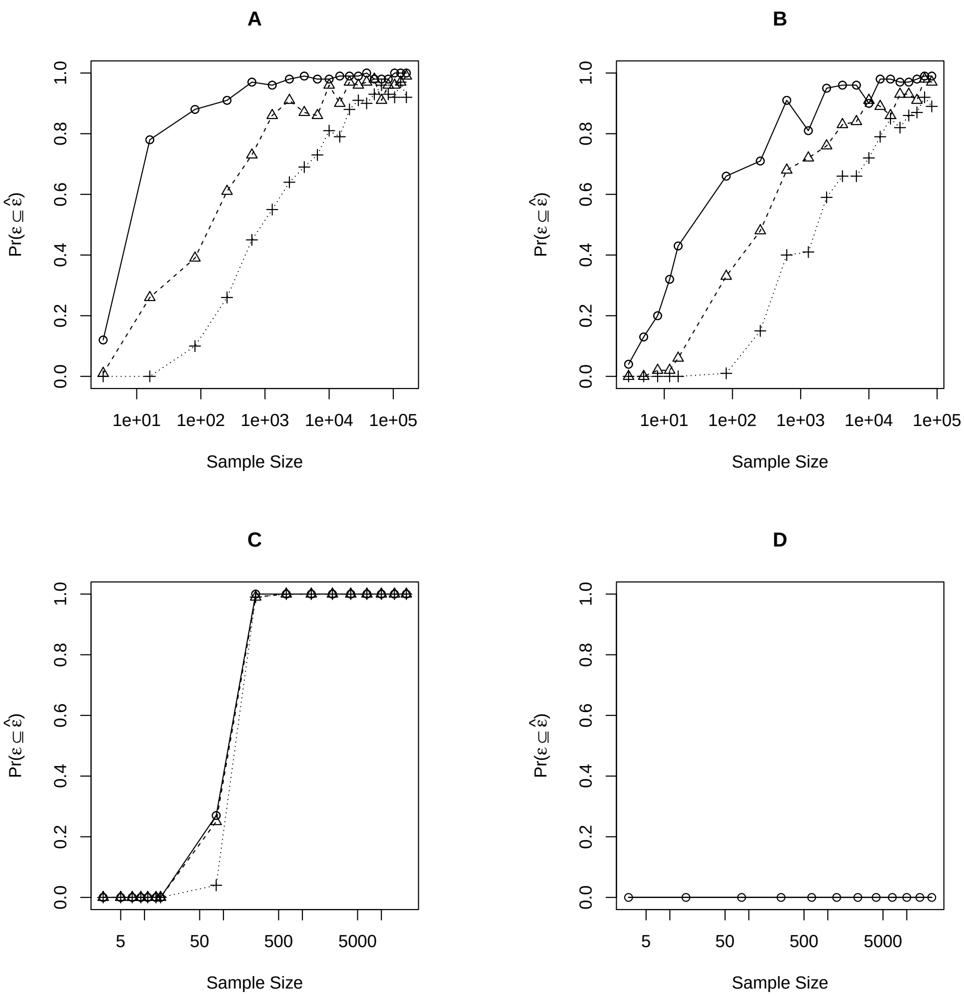

Figure 1: An investigation of $\mathrm{pr}\left(\mathcal{E} \subseteq \widehat{\mathcal{E}}_{\gamma_n}\right)$ in finite samples, in the simulation study described in Section 5. This quantity approaches one in Simulations A, B, and C, but not in Simulation D. Each line corresponds to a fixed number of features p, for a range of values of sample size n.