A Constrained ℓ_1 Minimization Approach to Sparse Precision Matrix Estimation

Presented by Xingguo Li

CSCI 8980, UMN

Authors : Tony Cai, Weidong Liu, and Xi Luo

April 14, 2014

Outline

- CLIME
- Statistical Properties
  - Convergence Rates under Norms, *i.e.* $\|\cdot\|_2, \|\cdot\|_F^2$ and $|\cdot|_\infty$
  - Convergence Rates of Expectation & Ordered Variables
  - Model Selection Consistency
  - Comparison with ℓ_1-MLE
- Numerical Experiments

Motivation

- ℓ_1 regularized log-determinant (Banerjee *et al.*, 2008):

$$\widehat{\mathbf{\Omega}}_{\text{Glasso}} := \operatorname*{argmin}_{\mathbf{\Omega} \succ 0} \{\langle \mathbf{\Omega}, \mathbf{\Sigma}_n \rangle - \log\det(\mathbf{\Omega}) + \lambda_n \|\mathbf{\Omega}\|_1\} \quad (1)$$

  Optimality condition:

$$\widehat{\mathbf{\Omega}}^{-1}_{\text{Glasso}} - \mathbf{\Sigma}_n = \lambda_n \widehat{\mathbf{Z}}, \; \widehat{\mathbf{Z}} \in \partial \|\widehat{\mathbf{\Omega}}_{\text{Glasso}}\|_1$$

  Dantzig type problem:

$$\min \|\mathbf{\Omega}\|_1 \; s.t. \; |\mathbf{\Omega}^{-1} - \mathbf{\Sigma}_n|_\infty \leq \lambda_n, \mathbf{\Omega} \in \mathbb{R}^{p \times p}$$

- CLIME

$$\min \|\mathbf{\Omega}\|_1 \; s.t. \; |\mathbf{I} - \mathbf{\Sigma}_n \mathbf{\Omega}|_\infty \leq \lambda_n, \mathbf{\Omega} \in \mathbb{R}^{p \times p}, \quad (2)$$

Properties

Compared with ℓ_1-MLE (1),

- No requirement of positive definiteness of $\mathbf{\Omega}$
- Columnwise decomposibility: For all $i = 1, \ldots, p$,

$$\min \|\boldsymbol{\beta}\|_1 \ s.t. \ \ |\mathbf{e}_i - \mathbf{\Sigma}_n \boldsymbol{\beta}|_\infty \leq \lambda_n, \boldsymbol{\beta} \in \mathbb{R}^p. \tag{3}$$

Lemma 1

Let $\{\widehat{\mathbf{\Omega}}_1\}$ be the solution set of (2), and let $\{\widehat{\mathbf{B}}\} := \{(\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_p)\}$, where $\widehat{\boldsymbol{\beta}}_i$ are solutions to (3) for $i = 1, \ldots, p$. Then $\{\widehat{\mathbf{\Omega}}_1\} = \{\widehat{\mathbf{B}}\}$.

- Improved convergence rate (polynomial-type tails)
- Improved model selection consistency (polynomial-type tails)

Proof of Lemma 1

Remind:
$\widehat{\mathbf{\Omega}}_1 = (\widehat{\omega}_1^1, \ldots, \widehat{\omega}_p^1) = \operatorname{argmin}_{\mathbf{\Omega} \in \mathbb{R}^{p \times p}} \|\mathbf{\Omega}\|_1 \ s.t. \ \ |\mathbf{I} - \mathbf{\Sigma}_n \mathbf{\Omega}|_\infty \leq \lambda_n$
$\widehat{\mathbf{B}} = (\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_p), \ \widehat{\boldsymbol{\beta}}_i = \operatorname{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \ s.t. \ |\mathbf{e}_i - \mathbf{\Sigma}_n \boldsymbol{\beta}|_\infty \leq \lambda_n, \ \forall i$

(1) We have

$$|\widehat{\omega}_i^1|_1 \geq |\widehat{\boldsymbol{\beta}}_i|_1, \ \forall 1 \leq i \leq p \tag{4}$$
$$\|\widehat{\mathbf{\Omega}}_1\|_1 \leq \|\widehat{\mathbf{B}}\|_1, \tag{5}$$
$$\Rightarrow \widehat{\mathbf{B}} \in \{\widehat{\mathbf{\Omega}}_1\}$$

(2) If $\widehat{\mathbf{\Omega}}_1 \notin \{\widehat{\mathbf{B}}\}$, $\exists i$ s.t. $|\widehat{\omega}_i^1|_1 > |\widehat{\boldsymbol{\beta}}_i|_1$, then by (4) $\|\widehat{\mathbf{\Omega}}_1\|_1 > \|\widehat{\mathbf{B}}\|_1$
$\Rightarrow\Leftarrow$ (5)

Therefore, $\{\widehat{\mathbf{B}}\} = \{\widehat{\mathbf{\Omega}}_1\}$. $\square$

Properties

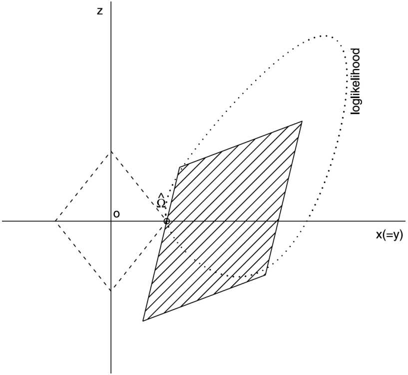

Figure 1 : Plot of the elementwise ℓ_∞ constrained feasible set (shaded polygon) and the elementwise ℓ_1 norm objective (dashed diamond near the origin) from CLIME. The log-likelihood function as in Glasso is represented by the dotted line. (Cai *et al.*, 2011)

Parameter Class

$\mathbf{\Omega}_0 \in \mathcal{U}$: Uniformity class of matrices,

$$\mathcal{U} := \mathcal{U}(q, x_0(p))$$
$$= \left\{ \mathbf{\Omega} : \mathbf{\Omega} \succ 0, \|\mathbf{\Omega}\|_{L_1} \leq M, \max_{1 \leq i \leq p} \sum_{j=1}^{p} |w_{ij}|^q \leq s_0(p), 0 \leq q < 1 \right\}$$

- $q = 0$, $\mathcal{U}(0, s_0(p))$ is a class of $s_0(p)$-sparse matrices
- Wider class of precision matrix than truly sparse matrices, *i.e.* $s_0(p)$ is small when many entries are small.

Tail Class

Two types of tails:

(C1) Exponential-type tails: $\exists$ some constant $0 < \eta < 1/4$ such that $\log p/n \leq \eta$ and for bounded constant K

$$\mathbb{E}e^{t(X_i-\mu_i)^2} \leq K < \infty \quad \text{for all } |t| \leq \eta, \quad \text{for all } i$$

(C2) Polynomial-type tails: For some $\gamma, c_1, \delta > 0$ and $p \leq c_1 n^\gamma$,

$$\mathbb{E}|X_i - \mu_i|^{4\gamma+4+\delta} \leq K \quad \text{for all } i$$

- Bounded $\theta := \max_{ij} \theta_{ij} = \max_{i,j} \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j) - \sigma_{ij}^0]^2$

Convergence Rate under Norms

- Symmetrizing operation: $\widehat{\mathbf{\Omega}} = (\widehat{\omega}_{ij})$, where $\widehat{\mathbf{\Omega}}_1 = (\widehat{\omega}_{ij}^1)$ and $\widehat{\omega}_{ij} = \widehat{\omega}_{ji} = \widehat{\omega}_{ij}^1 I\{|\widehat{\omega}_{ij}^1| \le |\widehat{\omega}_{ji}^1|\} + \widehat{\omega}_{ji}^1 I\{|\widehat{\omega}_{ij}^1| > |\widehat{\omega}_{ji}^1|\}$

Theorem 1

Suppose $\mathbf{\Omega}_0 \in \mathcal{U}(q, s_0(p))$. Assume (C1) or (C2) holds. Let $\lambda_n = C_{0i} M \sqrt{\log p/n}$ and $\tau > 0$, then w.h.p.

$$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2 \le C_{1i} M^{2-2q} s_0(p) \left(\log p/n\right)^{(1-q)/2} \tag{6}$$

$i = 1, 2$ for (C1) and (C2) respectively.

- ℓ_1-MLE estimator for polynomial-type of tails when $q = 0$:

$$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2 = \mathcal{O}(s_0(p)\sqrt{\frac{p^{\tau/(\gamma+1+\delta/4)}}{n}})$$

- Other norms: $\frac{1}{p}\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_F^2 = \mathcal{O}(s_0(p)(\frac{\log p}{n})^{1-q/2})$

$$|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0|_\infty = \mathcal{O}(\sqrt{\frac{\log p}{n}})$$

Convergence Rate of $\sup_{\boldsymbol{\Omega}_0 \in \mathcal{U}} \mathbb{E}\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_2^2$

- Replace $\boldsymbol{\Sigma}_n$ with $\boldsymbol{\Sigma}_{n,\rho} = \boldsymbol{\Sigma}_n + \rho \mathbf{I}$ to (1) ensure the existence of $\mathbb{E}\|\widehat{\boldsymbol{\Omega}}_\rho - \boldsymbol{\Omega}_0\|_2^2$ and (2) get a feasible initial value of $\widehat{\boldsymbol{\Omega}}$.

Theorem 2

Suppose $\boldsymbol{\Omega}_0 \in \mathcal{U}(q, s_0(p))$ and (C1) holds. Let $\rho = \sqrt{\log p/n}$, $\lambda_n = C_0 M \sqrt{\log p/n}$ and τ sufficiently large. If $p = n^\xi$ for some $\xi > 0$, then

$$\sup_{\boldsymbol{\Omega}_0 \in \mathcal{U}} \mathbb{E}\|\widehat{\boldsymbol{\Omega}}_\rho - \boldsymbol{\Omega}_0\|_2^2 = \mathcal{O}\left(M^{4-4q} s_0(p)^2 \left(\frac{\log p}{n} \right)^{1-q} \right). \tag{7}$$

- Hold for $\min\left(\sqrt{\frac{\log p}{n}}, p^{-\alpha} \right) \leq \rho \leq \sqrt{\frac{\log p}{n}}$ with any $\alpha > 0$.
- Same order of rate for $|\cdot|_\infty^2$ and $\|\cdot\|_F^2$ with the rates under norms.

Convergence Rate of Ordered Variables

- $\mathcal{U}_o(\alpha, B) = \{\mathbf{\Omega} : \max_j \sum_i \{|\omega_{ij}| : |i-j| \geq k\} \leq B(k+1)^{-\alpha},$
  $\mathbf{\Omega} \succ 0, \forall k \geq 0\}$ for some $\alpha > 0$
- Better rates can be obtained

Theorem 3

Let $\mathbf{\Omega}_0 \in \mathcal{U}_o(\alpha, B)$ and $\lambda_n = CB\sqrt{\log p/n}$ with sufficiently large C.

(a) *If (C1) or (C2) holds, then w.h.p.,*

$$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2 = \mathcal{O}\left(B^2(\log p/n)^{\alpha/(2\alpha+2)}\right) \tag{8}$$

(b) *Suppose $p \geq n^\xi, \xi > 0$. If (C1) holds and $\rho = \sqrt{\log p/n}$, then*

$$\sup_{\mathbf{\Omega}_0 \in \mathcal{U}_o(\alpha, B)} \mathbb{E}\|\widehat{\mathbf{\Omega}}_\rho - \mathbf{\Omega}_0\|_2^2 = \mathcal{O}\left(B^4(\log p/n)^{\alpha/(\alpha+1)}\right) \tag{9}$$

- $s_0(p)$ term disappears from the bounds.

A General Result

Theorem 6

Let $\boldsymbol{\Omega}_0 \in \mathcal{U}(q, s_0(p))$ *and* $\rho > 0$. *If* $\lambda_n \geq \|\boldsymbol{\Omega}_0\|_{L_1}(\max_{ij} |\widehat{\sigma}_{ij} - \sigma^0_{ij}| + \rho)$, *then*

$$|\widehat{\boldsymbol{\Omega}}_\rho - \boldsymbol{\Omega}_0|_\infty = \mathcal{O}(\|\boldsymbol{\Omega}_0\|_{L_1}\lambda_n), \tag{10}$$

$$\|\widehat{\boldsymbol{\Omega}}_\rho - \boldsymbol{\Omega}_0\|_2 = \mathcal{O}(\|\boldsymbol{\Omega}_0\|_{L_1}^{1-q} s_0(p)\lambda_n^{1-q}), \tag{11}$$

$$\frac{1}{p}\|\widehat{\boldsymbol{\Omega}}_\rho - \boldsymbol{\Omega}_0\|_F^2 = \mathcal{O}(\|\boldsymbol{\Omega}_0\|_{L_1}^{2-q} s_0(p)\lambda_n^{2-q}). \tag{12}$$

- Need to show $\max_{ij} |\widehat{\sigma}_{ij} - \sigma^0_{ij}| = \mathcal{O}(\sqrt{\log p/n})$ w.h.p. with corresponding constant for each result.

Graphical Model Selection Consistency

- Threshold $\widetilde{\mathbf{\Omega}} = (\widetilde{\omega}_{ij})$ with $\widetilde{\omega}_{ij} = \widehat{\omega}_{ij} I\{|\widehat{\omega}_{ij}| \geq \tau_n\}$, for $\tau_n \geq 4M\lambda_n$
- Define: $\mathcal{M}(\mathbf{\Omega}) = \{\mathrm{sign}(\omega_{ij}), 1 \leq i, j \leq p\}$,
  $S(\mathbf{\Omega}) = \{(i,j) : \omega_{ij} \neq 0\}$, $\theta_{\min} = \min_{(i,j)\in S(\mathbf{\Omega}_0)} |\omega^0_{ij}|$

Theorem 7

Suppose (C1) or (C2) holds and $\mathbf{\Omega}_0 \in \mathcal{U}(0, s_0(p))$. If $\theta_{\min} > 2\tau_n$, then $\mathcal{M}(\widetilde{\mathbf{\Omega}}) = \mathcal{M}(\mathbf{\Omega}_0)$ w.h.p.

- Sign consistency: Recover both sparsity pattern and signs of nonzero elements
- $\theta_{\min} > 2\tau_n$: Ensure nonzero elements are correctly retained
- If $M \perp\!\!\!\perp n, p$, then $\tau_n = \mathcal{O}(\sqrt{\log p/n})$

Comparison CLIME with ℓ_1-MLE (Ravikumar *et al.*, 2008)

- CLIME: $\min \|\boldsymbol{\Omega}\|_1 \ s.t. \ \ |\mathbf{I} - \boldsymbol{\Sigma}_n \boldsymbol{\Omega}|_\infty \leq \lambda_n, \boldsymbol{\Omega} \in \mathbb{R}^{p \times p}$
- ℓ_1-MLE: $\min_{\Theta \succ 0} \{\langle \boldsymbol{\Omega}, \boldsymbol{\Sigma}_n \rangle - \log\det(\boldsymbol{\Omega}) + \lambda_n \|\boldsymbol{\Omega}\|_{1,\text{off}}\}$

	CLIME	ℓ_1-MLE
Irrepresentability[1]	No	Yes
(n, p) Scale	$\log p = o(n)$	$n > C s_0^2(p) \log p$
Sparsity	Allow small values	Only truly sparse
Conv. Rate (poly)	$\mathcal{O}\left(s_0(p)\sqrt{\log p / n}\right)$	$\mathcal{O}\left(s_0(p)\sqrt{p^{\tau/(\gamma+1+\delta/4)}/n}\right)$
Model Selection	$\theta_{\min} \geq C\sqrt{\log p/n}$	$\theta_{\min} \geq C\sqrt{p^{\tau/\gamma+1+\delta/4}/n}$

[1] $\|\boldsymbol{\Gamma}_{S^cS}(\boldsymbol{\Gamma}_{SS})^{-1}\|_{L_1} \leq 1-\alpha$, $\alpha \in (0,1]$, $\boldsymbol{\Gamma} = \boldsymbol{\Omega}_0^{-1} \otimes \boldsymbol{\Omega}_0^{-1}$

Numerical Experiments

- CLIME: $\min \|\boldsymbol{\beta}\|_1 \ s.t. \ |\mathbf{e}_i - \boldsymbol{\Sigma}_n \boldsymbol{\beta}|_\infty \le \lambda_n, \boldsymbol{\beta} \in \mathbb{R}^p, \ i = 1, \dots, p$
- LP: $\min \sum_{j=1}^p u_j \ s.t. \ \forall 1 \le j \le p, \ \forall 1 \le k \le p$
  $$\begin{aligned} -\beta_j \le u_j, \quad & -\widehat{\sigma}_k^T + I\{k = i\} \le \lambda_n \\ +\beta_j \le u_j, \quad & +\widehat{\sigma}_k^T - I\{k = i\} \le \lambda_n \end{aligned}$$
- Refit (correct bias): Let $\widehat{S} = S(\widetilde{\boldsymbol{\Omega}})$, $\widehat{S}^c = \{\omega_{ij}, (i,j) \in \widehat{S}^c\}$,
  $$\check{\boldsymbol{\Omega}} = \operatorname{argmin}_{\boldsymbol{\Omega}_{\widehat{S}^c}=0} \ \langle \boldsymbol{\Omega}, \boldsymbol{\Sigma} \rangle - \log\det(\boldsymbol{\Omega})$$
- Compare with Glasso and SCAD (Fan *et al.*, 2001)
  $$\widehat{\boldsymbol{\Omega}}_{\text{Glasso}} := \operatorname{argmin}_{\Theta \succ 0} \langle \boldsymbol{\Omega}, \boldsymbol{\Sigma}_n \rangle - \log\det(\boldsymbol{\Omega}) + \lambda_n \|\boldsymbol{\Omega}\|_1$$
  $$\widehat{\boldsymbol{\Omega}}_{\text{SCAD}} := \operatorname{argmin}_{\Theta \succ 0} \langle \boldsymbol{\Omega}, \boldsymbol{\Sigma}_n \rangle - \log\det(\boldsymbol{\Omega}) + \sum_{i=1}^p \sum_{j=1}^p P_{\lambda_n,a}^{\text{SCAD}}(\omega_{ij})$$
  where $$P_{\lambda,a}^{\text{SCAD}}(x) = \begin{cases} \lambda|x| & \text{if } |x| \le \lambda; \\ -\left(\frac{|x|^2 - 2a\lambda|x| + \lambda^2}{2(a-1)}\right) & \text{if } \lambda \le |x| \le a\lambda; \\ \frac{(a+1)\lambda^2}{2} & \text{if } |x| \ge a\lambda \end{cases}$$

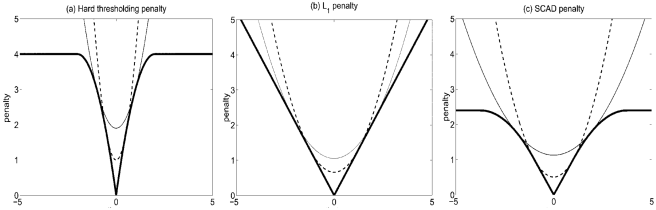

Figure 2 : Penalty functions and their quadratic approximations. (Fan *et al.*, 2001)

- Model 1. $\omega_{ij}^0 = 0.6^{|i-j|}$

(a) Truth (b) CLIME (c) Glasso (d) SCAD

Figure 3 : Heatmaps of the frequency of the zeros identified for each entry of the precision matrix (when $p = 60$) out of 100 replications. (Cai *et al.*, 2011)

Breast Cancer Dataset (Hass *et al.*, 2006)

Classification performance criterion:

- Specificity: $\frac{\text{TN}}{\text{TN+FP}}$
- Sensitivity: $\frac{\text{TP}}{\text{TP+FN}}$
- MCC: $\frac{\text{TP}\times\text{TN}-\text{FP}\times\text{FN}}{\sqrt{(\text{TP+FP})(\text{TP+FN})(\text{TN+FP})(\text{TN+FN})}}$

Method	Specificity	Sensitivity	MCC	Nonzero entries in $\hat{\boldsymbol{\Omega}}$
Glasso	0.768 (0.009)	0.630 (0.021)	0.366 (0.018)	3923 (2)
Adaptive lasso	0.787 (0.009)	0.622 (0.022)	0.381 (0.018)	1233 (1)
SCAD	0.794 (0.009)	0.634 (0.022)	0.402 (0.020)	674 (1)
CLIME	0.749 (0.005)	0.806 (0.017)	0.506 (0.020)	492 (7)

Figure 4 : Comparison of classification performance. Glasso, Adaptive lasso, and SCAD results are taken from Fan *et al.*, 2009.

Proof of Theorem 6

Let $\rho = 0$. Same proof for $\rho > 0$.

Assumption: $\lambda_n \geq \|\boldsymbol{\Omega}_0\|_{L_1}(\max_{ij}|\hat{\sigma}_{ij} - \sigma_{ij}^0|) \Leftrightarrow |\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_n|_\infty \leq \lambda_n / \|\boldsymbol{\Omega}_0\|_{L_1}$

(1) $\underline{|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0|_\infty} \leq 4\|\boldsymbol{\Omega}_0\|_{L_1}\lambda_n$

$\Uparrow$

$$|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0|_\infty \leq |\widehat{\boldsymbol{\Omega}}_1 - \boldsymbol{\Omega}_0|_\infty \leq \|\boldsymbol{\Omega}_0\|_{L_1}\underbrace{|\boldsymbol{\Sigma}_0(\widehat{\boldsymbol{\Omega}}_1 - \boldsymbol{\Omega}_0)|_\infty}_{\text{(i)}} \leq 4\|\boldsymbol{\Omega}_0\|_{L_1}\lambda_n$$

$(*)$ $|\mathbf{AB}|_\infty \leq \|\mathbf{A}\|_{L_1}|\mathbf{B}|_\infty$

$$\text{(i)} \leq \underbrace{|\boldsymbol{\Sigma}_n(\widehat{\boldsymbol{\Omega}}_1 - \boldsymbol{\Omega}_0)|_\infty}_{\text{(ii)}} + \underbrace{|(\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}_0)(\widehat{\boldsymbol{\Omega}}_1 - \boldsymbol{\Omega}_0)|_\infty}_{\text{(iii)}} \leq 4\lambda_n$$

$$\text{(ii)} \leq |\boldsymbol{\Sigma}_n\widehat{\boldsymbol{\Omega}}_1 - \mathbf{I}|_\infty + |\mathbf{I} - \boldsymbol{\Sigma}_n\boldsymbol{\Omega}_0|_\infty \leq \lambda_n + \|\boldsymbol{\Omega}_0\|_{L_1}|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_n|_\infty \leq 2\lambda_n$$

$$\begin{aligned}\text{(iii)} &\leq \|\widehat{\boldsymbol{\Omega}}_1 - \boldsymbol{\Omega}_0\|_{L_1}|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}_0|_\infty \\ &\leq \|\widehat{\boldsymbol{\Omega}}_1\|_{L_1}|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}_0|_\infty + \|\boldsymbol{\Omega}_0\|_{L_1}|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}_0|_\infty \leq 2\lambda_n\end{aligned}$$

Proof of Theorem 6

(2) $\underline{\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_2} \le C_4 s_0(p)(4\|\boldsymbol{\Omega}_0\|_{L_1}\lambda_n)^{1-q}, \ C_4 \le 2(1 + 2^{1-q} + 3^{1-q})$

$\Uparrow$

$\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_2 \le \|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_{L_1} = \max_j |\widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j^0|_1$

$(*)\ \|\mathbf{A}\|_2 \le \sqrt{\|\mathbf{A}\|_{L_1}\|\mathbf{A}\|_\infty}, \ \mathbf{A} = \mathbf{A}^T \Rightarrow \|\mathbf{A}\|_2 \le \|\mathbf{A}\|_{L_1}$

Let $\mathbf{h}_j = \widehat{\boldsymbol{\omega}}_j - \boldsymbol{\omega}_j^0$, $\mathbf{h}_j^1 = (\widehat{\omega}_{ij}\mathbf{I}\{|\widehat{\omega}_{ij}| \ge 2t_n\}; 1 \le i \le p)^T - \boldsymbol{\omega}_j^0$,

$\mathbf{h}_j^2 = \mathbf{h}_j - \mathbf{h}_j^1, \ t_n = |\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0|_\infty \le 4\|\boldsymbol{\Omega}_0\|_{L_1}\lambda_n$

$\Rightarrow |\boldsymbol{\omega}_j^0|_1 - |\mathbf{h}_j^1|_1 + |\mathbf{h}_j^2|_1 \le |\boldsymbol{\omega}_j^0 + \mathbf{h}_j^1|_1 + |\mathbf{h}_j^2|_1 = |\widehat{\boldsymbol{\omega}}_j|_1 \le |\widehat{\boldsymbol{\omega}}_j^1|_1 \le |\boldsymbol{\omega}_j^0|_1$

$\Rightarrow |\mathbf{h}_j^2|_1 \le |\mathbf{h}_j^1|_1 \Rightarrow |\mathbf{h}_j|_1 \le |\mathbf{h}_j^1|_1 + |\mathbf{h}_j^2|_1 \le 2|\mathbf{h}_j^1|_1$

$\cdots \Rightarrow |\mathbf{h}_j|_1 \le 2|\mathbf{h}_j^1|_1 \le 2(1 + 2^{1-q} + 3^{1-q})t_n^{1-q}s_0(p)$

(3) $\frac{1}{p}\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_F^2 \le C_5 s_0(p)(4\|\boldsymbol{\Omega}_0\|_{L_1}\lambda_n)^{2-q}, \ C_5 \le C_4$

$(*)\ \|\mathbf{A}\|_F^2 \le p\|\mathbf{A}\|_{L_1}|\mathbf{A}|_\infty$

Proof of Other Theorems

Based on Theorem 6, bound $\max_{ij} |\widehat{\sigma}_{ij} - \sigma^0_{ij}|$ w.h.p.

- Theorem 1 (a) and 4 (a), *i.e.* exponential-type tails,

$$\max_{ij} |\widehat{\sigma}_{ij} - \sigma^0_{ij}| \leq 2\eta^{-2}(2 + \tau + \eta^{-1}e^2K^2)^2\sqrt{\log p/n},$$

  w.p. $\geq 1 - 4p^{-\tau}$.

- Theorem 1 (b) and 4 (b), *i.e.* polynomial-type tails,

$$\max_{ij} |\widehat{\sigma}_{ij} - \sigma^0_{ij}| \leq \sqrt{(\theta + 1)(5 + \tau)\log p/n},$$

  w.p. $\geq 1 - \mathcal{O}(n^{-\delta/8} + p^{-\tau/2})$.

- Theorem 2,3,5 are direct results of Theorem 6,1,4.

Thank you!