Simulation Report of Thresholding-based Iterative Selection Procedures for Model Selection and Shrinkage

Huijuan Zhou, Lihao Yin, Tong Wang

Iteration rules：

$$\beta^{(j+1)} = \Theta((\boldsymbol{I} - \frac{1}{k_0^2}\boldsymbol{\Sigma})\beta^{(j)} + \frac{1}{k_0^2}X^T y; \frac{\lambda}{k_0^2})$$

where $\boldsymbol{\Sigma} = \boldsymbol{X}^T\boldsymbol{X}$, $\boldsymbol{X}$ is the standardized design matrix and $k_0 = \mu_{\max}(\boldsymbol{X}) = \|\boldsymbol{X}\|_2$ is max singular value of matrix $\boldsymbol{X}$. It makes the matrix $\boldsymbol{I} - \frac{1}{k_0^2}\boldsymbol{\Sigma}$ to be positive definte.

We demonstrate the empirical performance of TISPs by some simulation data. In addition to the Soft-TISP, i.e., the lasso, we implemented Hard-TISP and SCAD-TISP, the thresholdings of which belong to the hard-thresholding family.

- soft-threshold: $\Theta(x, \lambda) = sign(x)(|x| - \lambda)_+$
- hard-threshold: $\Theta(x, \lambda) = xI_{[|x|>\lambda]}$
- scad-threshold: $\Theta(x, \lambda) = \begin{cases} sign(x)(|x| - \lambda)_+ & |x| < 2\lambda \\ [(a-1)x - sign(x)a\lambda]/(a-2) & 2\lambda \le |x| < a\lambda \\ x & others \end{cases}$

Data Generating and Parameter Setting:

Let $\boldsymbol{\Sigma}$ be the correlation matrix in generating $\boldsymbol{X}$, i.e., each row of $\boldsymbol{X}$ is independently drawn from $N(0, \boldsymbol{\Sigma})$, where $\Sigma_{ij} = \rho^{|i-j|}$ with $\rho = 0.5, 0.85.$
$\beta = (\{3\}^1, \{1.5\}^1, \{0\}^2, \{2\}^1, \{0\}^3)$ $\quad \beta = (\{3\}^1, \{1.5\}^1, \{0\}^2, \{2\}^1, \{0\}^{95}).y = \boldsymbol{X}\beta + \epsilon, \epsilon \sim N(0, \sigma^2)$, we set $\sigma^2 = 4, 16$.Sample size $n = 20$

Penality Parameter Choosing:

We choose penality parameter λ by cross validation.

Performance Criterion:

- Mse: $\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
- sparsit error: $|\{i : sgn(\hat{\beta}_i) \neq sgn(\beta_i)\}|/d$
- proper zero percentages: $|\{i : \beta_i = 0, \hat{\beta}_i = 0\}|/|\{i : \beta_i = 0\}|$
- proper nonzero percentages: $|\{i : \beta_i \neq 0, \hat{\beta}_i \neq 0\}|/|\{i : \beta_i \neq 0\}|$

Results

Case 1 $p < n(p = 8, n = 20, runs = 1000)$

- $1(\rho = 0.5, \sigma = 2)$

	softTISP(Lasso)	hardTISP	scadTISP
Mse	5.774071	5.7740712	6.4273753
Sparse error	0.309750	0.1926250	0.1288750
proper zeros	0.514000	0.7186000	0.8774000
proper nonzeros	0.985000	0.9556667	0.8606667

- $2(\rho = 0.85, \sigma = 2)$

	softTISP(Lasso)	hardTISP	scadTISP
Mse	5.427307	5.4273068	5.601652
Sparse error	0.326250	0.2850000	0.225125
proper zeros	0.536000	0.6102000	0.747400
proper nonzeros	0.912000	0.8953333	0.826000

- $3(\rho = 0.5, \sigma = 8)$

	softTISP(Lasso)	hardTISP	scadTISP
Mse	82.3963662	82.3963662	105.3548877
Sparse error	0.4557500	0.5565000	0.5158750
proper zeros	0.4850000	0.2892000	0.4010000
proper nonzeros	0.6946667	0.8146667	0.7223333

- $4(\rho = 0.85, \sigma = 8)$

	softTISP(Lasso)	hardTISP	scadTISP
Mse	77.4988936	77.4988936	96.914207
Sparse error	0.4293750	0.5313750	0.501125
proper zeros	0.5918000	0.3990000	0.483600
proper nonzeros	0.5806667	0.6863333	0.616000

Case 2 $p > n(p = 100, n = 20, runs = 100)$

- $1(\rho = 0.5, \sigma = 2)$

	softTISP(Lasso)	hardTISP	scadTISP
Mse	9.6522649	9.6522649	12.2808243
Sparse error	0.0681000	0.1806000	0.0453000
proper zeros	0.9329897	0.8214433	0.9592784
proper nonzeros	0.8966667	0.7533333	0.8066667

- $2(\rho = 0.85, \sigma = 2)$

	softTISP(Lasso)	hardTISP	scadTISP
Mse	7.0373040	7.0373040	9.1424540
Sparse error	0.0532000	0.1437000	0.0498000
proper zeros	0.9494845	0.8592784	0.9551546
proper nonzeros	0.8600000	0.7600000	0.7900000

- $3(\rho = 0.5, \sigma = 8)$

	softTISP(Lasso)	hardTISP	scadTISP
Mse	102.4357841	102.4357841	124.2802076
Sparse error	0.1369000	0.4777000	0.2002000
proper zeros	0.8781443	0.5174227	0.8121649
proper nonzeros	0.3766667	0.6900000	0.4000000

- 4($\rho = 0.85, \sigma = 8$)

	softTISP(Lasso)	hardTISP	scadTISP
Mse	99.5372782	99.5372782	153.2313344
Sparse error	0.1184000	0.3640000	0.1948000
proper zeros	0.8981443	0.6379381	0.8184536
proper nonzeros	0.3466667	0.5733333	0.3766667

Results Analysis

From out simulation results, when the noise level is low($\sigma = 2$), the lasso(Soft-TISP) yields a more accurate estimate than the two. And when the noise level is relative high($\sigma = 8$), the Hard-TISP has a better performance. Fix the noise level, the higher correlation of the design matrix, the worse performance of the three types of thresholds. And fix the signal level, the higher noise level, the worse performance of the three types of thresholds.

Files Introduction

Code file contains codes for the three types thresholds and the R code that runs the Rcpp. Figures file contains outcome figures of the eight cases. Outcome file contains performance outcome tables of the eight cases and the estimates of β.