Improved bounds for Square-root Lasso and Square-root Slope

Alexis Derumigny, CREST - ENSAE, Palaiseau (France)

The framework

The sparse linear regression model

$$Y = \mathbb{X}\beta^* + \varepsilon,$$

with

- $Y \in \mathbb{R}^n$: vector of observations
- $\mathbb{X} \in \mathbb{R}^{n \times p}$: design matrix
- $\beta^* \in \mathbb{R}^p$: unknown true parameter with at most s non null components
- $\varepsilon \in \mathbb{R}^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$: random noise.

Definition of the estimators

Square-Root Lasso :

$$\hat{\beta}^{SQL} := \arg\min_{\beta} \Big(||Y - \mathbb{X}\beta||_n + \lambda|\beta|_1 \Big)$$

Square-Root Slope :

$$\hat{\beta}^{SQS} := \arg\min_{\beta} \Big(||Y - \mathbb{X}\beta||_n + |\beta|_* \Big),$$

where

- $|\cdot|_q$ is the l_q norm and $||\cdot||_n^2 := |\cdot|_2^2/n$ is the prediction norm
- $|u|_* := \sum_{i=1}^p \lambda_j |u|_{(j)}$ is the sorted l_1 norm
- with tuning parameters $\lambda_1 \geq \cdots \geq \lambda_p > 0$, and $\lambda > 0$.

$SRE(s, c_0)$ condition :

The design matrix $\mathbb{X}$ satisfies $\max_{j=1,\dots,p} ||\mathbb{X}e_j||_n \leq 1$ and

$$\kappa(s) := \min_{\delta \in C_{SRE}(s,c_0):\delta \neq 0} \frac{||\mathbb{X}\delta||_n}{|\delta|_2} > 0,$$

where $C_{SRE}(s, c_0) := \{\delta \in \mathbb{R}^p : |\delta|_1 \leq (1 + c_0)\sqrt{s}|\delta|_2\}$.

$WRE(s, c_0)$ condition :

The design matrix $\mathbb{X}$ satisfies $\max_{j=1,\dots,p} ||\mathbb{X}e_j||_n \leq 1$ and

$$\kappa' := \min_{\delta \in C_{WRE}(s,c_0):\delta \neq 0} \frac{||\mathbb{X}\delta||_n}{|\delta|_2} > 0,$$

where $C_{WRE}(s, c_0) := \Big\{\delta \in \mathbb{R}^p : |\delta|_* \leq (1 + c_0)|\delta|_2 \sqrt{\sum_{j=1}^s \lambda_j^2}\Big\}$.

Choice of the tuning parameters

Square-Root Lasso:

$$\lambda = \gamma\sqrt{\frac{1}{n}\log\left(\frac{2p}{s}\right)}, \quad \text{with } \gamma \geq 16 + 4\sqrt{2},$$

Square-Root Slope:

$$\lambda_j = \gamma'\sqrt{\frac{1}{n}\log\left(\frac{2p}{j}\right)}, \text{ for } j = 1, \dots, p, \text{ with } \gamma' \geq 16 + 4\sqrt{2},$$

Minimax optimal rates for Square-root Lasso

If $(s/n)\log(2p/s) < 9\kappa^2/256\gamma^2$ and under $SRE(s, 5/3)$, then with probability greater than $1 - (p/s)^{-s} - (1 + e^2)e^{-n/24}$, we have

$$||\mathbb{X}(\hat{\beta}^{SQL} - \beta^*)||_n \leq \frac{C_1}{\kappa^2}\sigma\sqrt{\frac{s}{n}\log\left(\frac{p}{s}\right)},$$

$$|\hat{\beta}^{SQL} - \beta^*|_q \leq \frac{C_2}{\kappa^2}\sigma s^{1/q}\sqrt{\frac{1}{n}\log\left(\frac{2p}{s}\right)},$$

where $1 \leq q \leq 2$ and C_1, C_2 are constants.

Algorithm 1: Algorithm for making $\hat{\beta}^{SQL}$ adaptive to s.

Input: a distance $d(\cdot, \cdot)$ on $\mathbb{R}^p$
Input: a function $w(\cdot) : [1, s_*] \to \mathbb{R}_+$
Input: a family of estimators $\big(\hat{\beta}_{(s)}\big)_{s=1,\dots,s_*}$
$M \leftarrow \lfloor \log_2(s_*) \rfloor$;
for $m \leftarrow 1$ ***to*** $M + 1$ **do**
  compute the estimator $\hat{\beta}_{(2^m)}$;
end
compute $\hat{\sigma} \leftarrow ||Y - \mathbb{X}\hat{\beta}_{(2^{M+1})}||_n$;
compute the set $S_1 \leftarrow \Big\{ m \in \{1, \dots, M\}$:
$d\big(\hat{\beta}_{(2^{k-1})}, \hat{\beta}_{(2^k)}\big) \leq 4\hat{\sigma}C_0 w(2^k)$, for all $k \geq m \Big\}$;
if $S_1 \neq \emptyset$ **then** $\tilde{m} \leftarrow \min S_1$ **else** $\tilde{m} \leftarrow M$;
Output: $\tilde{s} \leftarrow 2^{\tilde{m}}$
Output: $\tilde{\beta} \leftarrow \hat{\beta}_{(\tilde{s})}$

Computation of the Square-root Slope

As in the case of the Square-root Lasso, we still have for any β,

$$||Y - \mathbb{X}\beta||_n = \min_{\sigma > 0}\left(\sigma + \frac{||Y - \mathbb{X}\beta||_n^2}{\sigma}\right),$$

where the minimum is attained for $\hat{\sigma} = ||Y - \mathbb{X}\beta||_n$.
As a consequence,

$$\hat{\beta}^{SQS} \in \arg\min_{\beta \in \mathbb{R}^p}\Big(||Y - \mathbb{X}\beta||_n + |\beta|_*\Big)$$

is equivalent to take the estimator $\hat{\beta}$ in the joint minimization program

$$(\hat{\beta}, \hat{\sigma}) \in \arg\min_{\beta \in \mathbb{R}^p,\, \sigma > 0}\left(\sigma + \frac{||Y - \mathbb{X}\beta||_n^2}{\sigma} + |\beta|_*\right).$$

Algorithm 2: Scaled Slope algorithm

Input: explained variable Y, design matrix $\mathbb{X}$;
Input: tuning parameters $\lambda_1 \leq \cdots \leq \lambda_p$;
choose some initialization value for $\hat{\sigma}$, for example the standard deviation of Y ;
repeat
  estimate $\hat{\beta}$ by the Slope algorithm with the parameters $\hat{\sigma} \cdot \lambda_1, \dots, \hat{\sigma} \cdot \lambda_p$;
  estimate $\hat{\sigma}$ by $||Y - \mathbb{X}\hat{\beta}||_n$;
until *convergence*;
Output: a joint estimator $(\hat{\beta}, \hat{\sigma})$;

Minimax optimal rates for Square-root Slope

If $(s/n)\log(2ep/s) < \kappa'^2/256\gamma'^2$ and under $WRE(s, 20)$, then with probability greater than $1 - (p/s)^{-s} - (1 + e^2)e^{-n/24}$, we have

$$||\mathbb{X}(\hat{\beta}^{SQS} - \beta^*)||_n \leq \frac{C_1'}{\kappa'}\sigma\sqrt{\frac{s}{n}\log\left(\frac{p}{s}\right)},$$

$$|\hat{\beta}^{SQS} - \beta^*|_* \leq \frac{C_1'}{\kappa'^2}\sigma\frac{s}{n}\log\left(\frac{p}{s}\right),$$

$$|\hat{\beta}^{SQS} - \beta^*|_2 \leq \frac{C_1'}{\kappa'^2}\sigma\sqrt{\frac{s}{n}\log\left(\frac{p}{s}\right)},$$

where C_1' is a constant,
denoting by $|\cdot|_q$ the l_q norm (for estimation), and $||\cdot||_n^2 := |\cdot|_2^2/n$ (for prediction).

Reference

- Derumigny, A. (2018). Improved bounds for Square-Root Lasso and Square-Root Slope. *Electronic Journal of Statistics*, 12(1), 741-766.