Sparsit??, Estimation et S??lection de Variables

Evgenii Chzhen & Henry VONG

10 november 2015

Contents

Introduction

We observe $y_1, ..., y_M$ which satisfy the following model of gaussian sequence :

$$y_j = a\eta_j + \xi_j; \;\; j = 1, ..., M;$$

with $a \in \mathbb{R}$ and $\eta_j \in \{0,1\}$ which are parametres chosen such that :

$$\sum_{j=1}^{M} = \left[M^{1-\beta}\right], \text{ for one fixed } \beta \in (0,1),$$

where the random variables $\xi_j \overset{\text{i.i.d}}{\sim} \mathbb{N}(0,1)$. We consider $M = 50$, $\beta = 0.3$ and $a \in [1,4]$ and let $\tau = \sqrt{2logM}$. We want to apply the following estimator for given problem and compute $R(a)$ for $a \in [1,4]$.

1. Estimation by hard threshold

We define an estimator by hard threshold as it follows

$$\hat{\theta}_j^H = y_j \mathbb{I}(|y_j| > \tau), \;\; j = 1, ..., M,$$

and $R(a) = \left\|\hat{\theta}^H - a\eta\right\|_2$.

2. Estimation by soft threshold

We define an estimator by soft threshold as it follows

$$\hat{\theta}_j^S = y_j \left(1 - \frac{\tau}{|y_j|}\right)_+, \;\; j = 1, ..., M,$$

and $R(a) = \left\|\hat{\theta}^S - a\eta\right\|_2$.

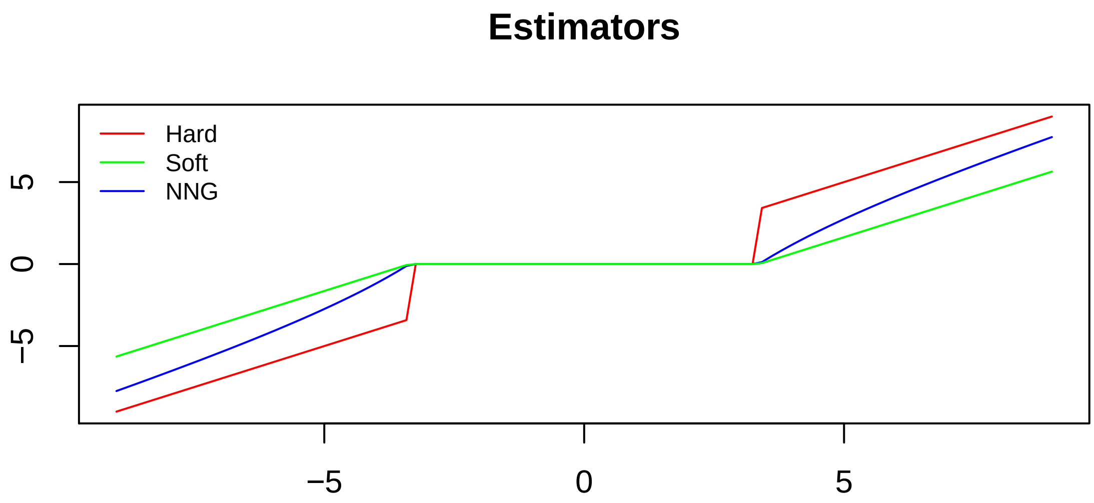
3. Non-negative garrot

We define an estimator by soft threshold as it follows

$$\hat{\theta}_j^{NG} = y_j \left(1 - \frac{\tau^2}{y_j^2}\right)_+, \ \ j = 1, ..., M,$$

and $R(a) = \left\|\hat{\theta}^{NG} - a\eta\right\|_2$.

To understand the difference between the estimators we plot each of them for given M.



4. Seclection of non zero coefficients

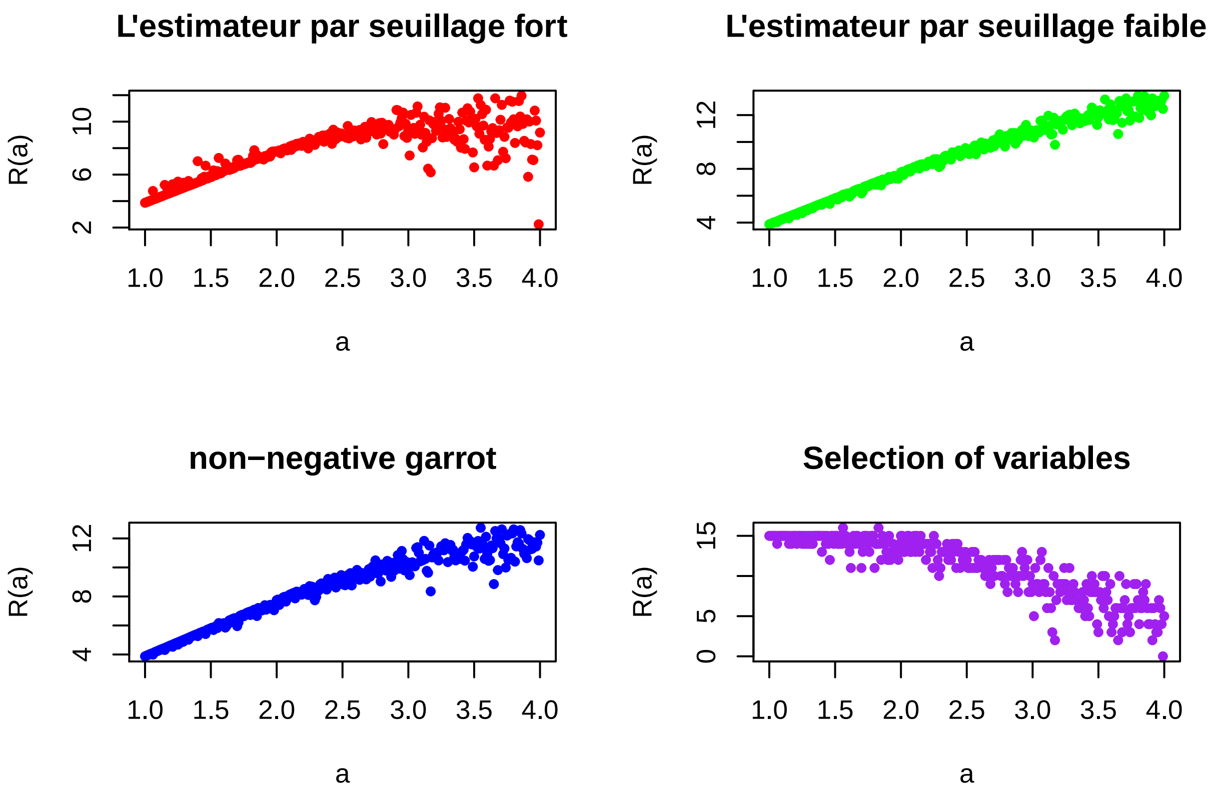
We consider a selection of non zero coefficients of $(a\eta_j)_{j=1,...,M}$ by hard threshold :

$$\hat{\eta}_j = \mathbb{I}(|y_j| \geq \sqrt{2logM}).$$

And $R(a) = \sum\limits_{j=1}^{M} |\eta_j - \hat{\eta}_j|$.

5. Plotting

We plot $R(a)$ for each estimator.



L'estimateur par seuillage fort

L'estimateur par seuillage faible

non−negative garrot

Selection of variables

Corollary: One can see that $R(a)$ for the first three estimators is increasing, whenever $R(a)$ in case os selection a non-zero variables is decreasing. It can be explained in this way: with a growth of a the difference between $a\eta_j + \xi_j$ and ξ_j is increasing so it's easier to distinguish just a noise from a non-zero value with a noise therefore the risk is decreasing.

6. Application for previous problem

Consider a following model from TP2 :

$$Y = \mathbb{X} \cdot \beta + \xi,$$

we want to apply given estimators to this model. First step is to transform model from TP2 to model of gaussien sequence for this purpose we assume that $\mathbf{X}^T\mathbf{X}/n = I_N$ it allows us to rewrite the model as it follows, see for instance (Tsybakov 2008, 68–69) :

$$\frac{1}{n}\mathbb{X}^T Y = \beta + \frac{1}{n}\mathbb{X}^T \xi,$$

introducing new notation one can obtain :

$$z = \beta + \psi,$$

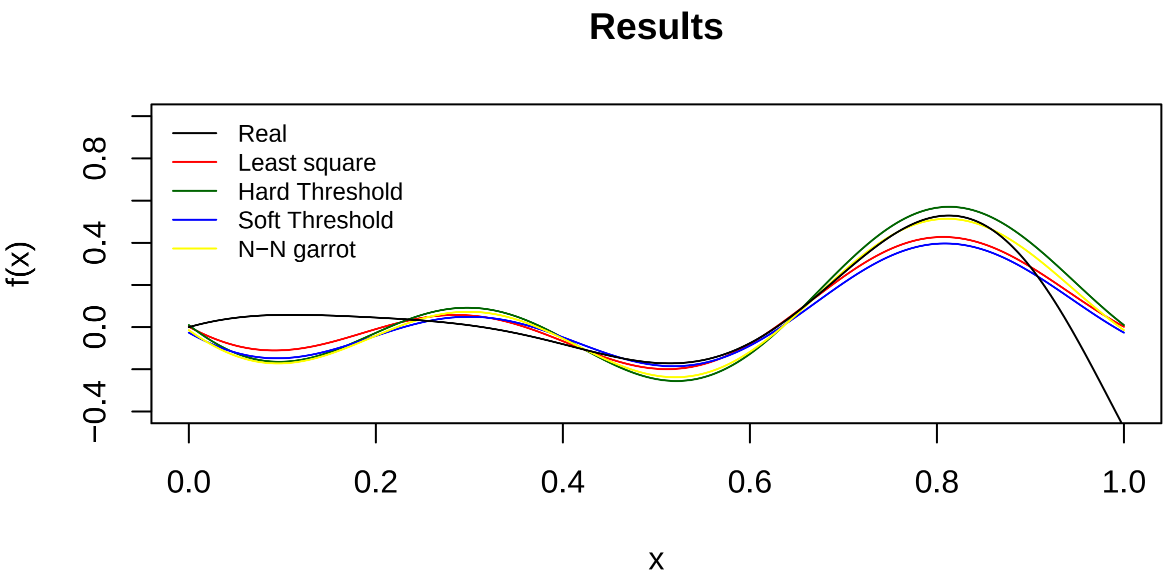
where $z = \frac{1}{n}^T \mathbb{X}^T Y$, $\psi \overset{\text{i.i.d}}{\sim} \mathbb{N}\left(0, \frac{1}{n^2}\mathbb{X}^T\mathbb{X}\right) = \mathbb{N}\left(0, \frac{1}{n^2} I_N\right)$ and finally we apply hard threshold, soft threshold and non-negative garrot estimators to modified model with number of projections $N = 5$.

Here is the results that we obtained

Coefficients	Linear regression	seuillage fort	seuillage faible	non-negative garrot
β_1	0.0546969	0.0742491	0.0383667	0.0569082

Coefficients	Linear regression	seuillage fort	seuillage faible	non-negative garrot
β_2	0.0703495	0.0905122	0.0546298	0.0762871
β_3	-0.1150686	-0.1467884	-0.110906	-0.138017
β_4	-0.1076622	-0.1360941	-0.1002116	-0.1266333
β_5	-0.0758448	-0.1192068	-0.0833244	-0.1084059

We plot all the results to compare estimators visually



Corollary: One can see that the difference between all estimators visually is not critical, however least square estimator is more computationaly complicated therefore for the given problem we would recommend to use threshold estimators.

References

Tsybakov, Alexandre B. 2008. *Introduction to Nonparametric Estimation*. Springer.