On the sign recovery by LASSO, thresholded LASSO and thresholded Basis Pursuit Denoising

Patrick Tardivel and Małgorzata Bogdan

Instytut Mathematyczny
Wrocław

July 3rd 2020

Introduction

Let us consider the high-dimensional linear regression model

$$Y = X\beta + \varepsilon,$$

where $X \in \mathbb{R}^{n \times p}$ with $p > n$, $\mathrm{rank}(X) = n$.

We aim at recovering $\mathrm{sign}(\beta) := (\mathrm{sign}(\beta_i))_{1 \leq i \leq p}$, where $\beta \in \mathbb{R}^p$ is an unknown sparse parameter.

LASSO and BP estimators

The LASSO estimator $\hat{\beta}(\lambda)$ is defined by

$$\hat{\beta}(\lambda) := \underset{b \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{2}\|Y - Xb\|_2^2 + \lambda\|b\|_1, \lambda > 0$$

The Basis Pursuit (BP) estimator $\hat{\beta}^{bp}$ is defined by

$$\hat{\beta}^{bp} := \underset{b \in \mathbb{R}^p}{\operatorname{argmin}} \|b\|_1 \text{ subject to } Y = Xb.$$

The noiseless case ($Y = X\beta$)

The BP (estimator) $\hat{\beta}^{bp} := \underset{b \in \mathbb{R}^p}{\text{argmin}} \|b\|_1$ subject to $Y = Xb$ recovers $\text{sign}(\beta)$ iff β is identifiable (with respect to X and the L_1 norm)

$$X\gamma = X\beta \text{ and } \gamma \neq \beta \Rightarrow \|\gamma\|_1 > \|\beta\|_1.$$

Note that β is identifiable iff $\text{sign}(\beta)$ is identifiable.

The LASSO (estimator) $\hat{\beta}(\lambda) := \underset{b \in \mathbb{R}^p}{\text{argmin}} \frac{1}{2}\|Y - Xb\|_2^2 + \lambda\|b\|_1$ recovers $\text{sign}(\beta)$ iff the irrepresentability condition

$$\|X_{\bar{I}}' X_I (X_I' X_I)^{-1} sign(\beta_I)\|_\infty < 1, \text{ where } \begin{cases} I := \{i : \beta_i \neq 0\}, \\ \bar{I} := \{i : \beta_i = 0\} \end{cases}$$

and non-null components of β are sufficiently large (Buhlmann and van de Geer(2011)).

Irrepresentability condition

Theorem (Wainwright, 2009)

Let $Y = X\beta + \varepsilon$ *where* ε *has a symmetric distribution. If the following inequality holds*

$$\|X_{\bar{I}}' X_I (X_I' X_I)^{-1} \mathrm{sign}(\beta_I)\|_\infty > 1$$

then whatever $\lambda > 0$*, we have* $\mathbb{P}(\mathrm{sign}(\hat{\beta}(\lambda)) = \mathrm{sign}(\beta)) \leq 1/2$*.*

Proposition (Tardivel and Bogdan)

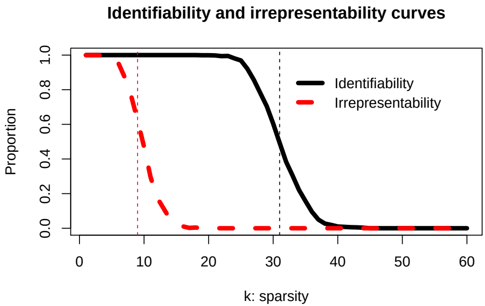
If the following inequality holds

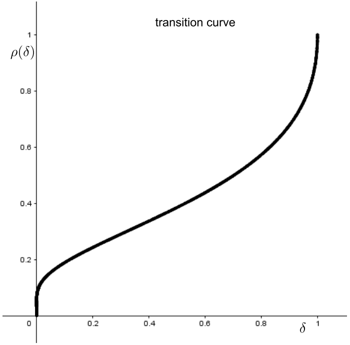
$$\|X_{\bar{I}}' X_I (X_I' X_I)^{-1} sign(\beta_I)\|_\infty < 1,$$

then the parameter β *is identifiable with respect to* X *and the* L_1 *norm.*

Standard Gaussian design

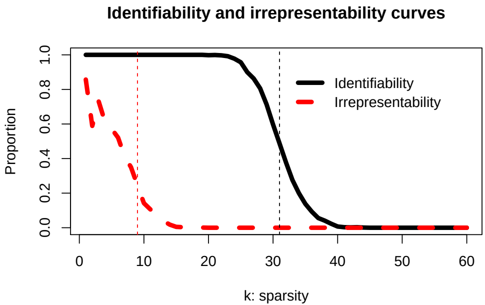
$X \in \mathbb{R}^{100\times 300}$ standard Gaussian matrix





- black line $k = \rho(100/300) \times 100 = 31$
- red line $k = 100/(2\log(300)) = 9$

$X \in \mathbb{R}^{100\times300}$ where columns are extremely correlated (columns of X are almost all equal)



Theorem (Tardivel and Bogdan)

$Y = X\beta + \varepsilon$, $X \in \mathbb{R}^{n\times p}$ and $\beta \in \mathbb{R}^p$ an unknown parameter.

$$\hat{\beta} \text{ represents } \begin{cases} \hat{\beta}(\lambda) := \underset{b\in\mathbb{R}^p}{\operatorname{argmin}} \frac{1}{2}\|Y - Xb\|_2^2 + \lambda\|b\|_1, \text{ or} \\ \hat{\beta}^{bp} := \underset{b\in\mathbb{R}^p}{\operatorname{argmin}} \|b\|_1 \text{ subject to } Y = Xb. \end{cases}$$

If β is not identifiable then: $\exists i$ such that $\beta_i \neq 0$ and $\hat{\beta}_i\beta_i \leq 0$.
Thus, thresholded LASSO/BP cannot recover $\text{sign}(\beta)$.

If β is identifiable and non-null components of β are large then

$$\max_{i:\beta_i<0}\left\{\hat{\beta}_i\right\} < \min_{i:\beta_i=0}\left\{\hat{\beta}_i\right\} \leq \max_{i:\beta_i=0}\left\{\hat{\beta}_i\right\} < \min_{i:\beta_i>0}\left\{\hat{\beta}_i\right\}.$$

Thus, thresholded LASSO/BP can recover $\text{sign}(\beta)$.

To summarize

To recover $\mathrm{sign}(\beta)$ one needs the following conditions

- **With the LASSO** one needs the irrepresentability condition

$$\|X_{\bar{I}}'X_I(X_I'X_I)^{-1}\mathrm{sign}(\beta_I)\|_\infty < 1$$

- **With the thresholded LASSO/BP** one needs the identifiability condition

$$X\gamma = X\beta \text{ and } \gamma \neq \beta \text{ then } \|\gamma\|_1 > \|\beta\|_1.$$

We remind that

Irrepresentability condition $\Rightarrow$ Identifiability condition

Method to compute the threshold when the design is standard Gaussian:

Input: design matrix X, response Y and $\lambda > 0$ (for LASSO)

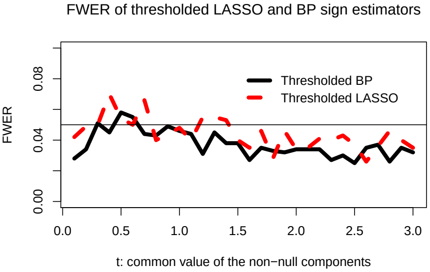
1. Generate $Z_1, \dots, Z_l$ be i.i.d random vectors having $\mathcal{N}(0, I_n)$ distribution and solve the following optimization problems:

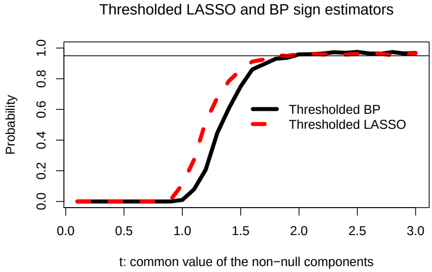
$$(\hat{b}^{(i)}, \hat{c}^{(i)}) = \underset{b\in\mathbb{R}^p, c\in\mathbb{R}}{\operatorname{argmin}} \frac{1}{2}\|Y - Xb - Z_i c\|_2^2 + \lambda(\|b\|_1 + |c|),$$

$$(\hat{b}^{(i)}, \hat{c}^{(i)}) = \underset{b\in\mathbb{R}^p, c\in\mathbb{R}}{\operatorname{argmin}} \|b\|_1 + |c| \text{ subject to } Xb + Z_i c = Y.$$

2. Compute the threshold as the empirical $(1-\alpha)^{1/p}$ quantile of $\hat{c}^{(1)}, \dots, \hat{c}^{(l)}$.

Let $Y = X\beta + \varepsilon$ where $X \in \mathbb{R}^{100\times 300}$ is a standard Gaussian matrix, $\varepsilon \sim \mathcal{N}(0, I_n)$, $\|\beta\|_0 = 20$, non null components of β are all equal to $t > 0$.





Thank you!

- Tardivel and Bogdan. On the sign recovery by LASSO, thresholded LASSO and thresholded Basis Pursuit Denoising

Related article

- U. Schneider, PJC. Tardivel. The Geometry of Uniqueness and Model Selection of Penalized Estimators including SLOPE, LASSO and Basis Pursuit.