# The Geometry of Model Recovery by Penalized and Thresholded Estimators

Patrick Tardivel (joint work with Tomasz Skalski, Piotr Graczyk and Ulrike Schneider)

Institut de mathématiques de Bourgogne, Dijon

# Uniqueness

Consider the optimization problem

$$S_{X,\lambda\mathrm{pen}}(y) := \underset{b\in\mathbb{R}^p}{\mathrm{Argmin}}\ \frac{1}{2}\|y - Xb\|_2^2 + \lambda\mathrm{pen}(b).$$

Where $X \in \mathbb{R}^{n\times p}$, $y \in \mathbb{R}^n$, $\lambda > 0$ and $\mathrm{pen}(b) = \max\{u_0'b, u_1'b, \dots, u_l'b\}$, $u_0 = 0$ and $u_1, \dots, u_l \in \mathbb{R}^p$, is a polyhedral gauge.
Note: $S_{X,\lambda\mathrm{pen}}(y) \neq \emptyset$.

- $\mathrm{pen}(b) = \|b\|_1$–> LASSO.
- $\mathrm{pen}(b) = \|b\|_\infty$.
- $\mathrm{pen}(b) = \sum_{i=1}^p \lambda_i |b|_{\downarrow i}$ where $\lambda_1 > 0$, $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ and $|b|_{\downarrow 1} \geq \dots \geq |b|_{\downarrow p}$ –> SLOPE.
- $\mathrm{pen}(b) = \|Db\|_1$ for some $D \in \mathbb{R}^{m\times p}$ –> Generalized LASSO.

## Theorem

*Let* $X \in \mathbb{R}^{n\times p}$, $\lambda > 0$ *and* $\mathrm{pen}$ *be a polyhedral gauge where* $\mathrm{pen}(b) = \max\{u_0' b, u_1' b, \dots, u_l' b\}$. *There exists* $y \in \mathbb{R}^n$ *for which the minimizer of*

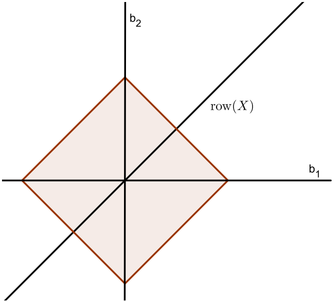
$$\operatorname*{Argmin}_{b\in\mathbb{R}^p} \frac{1}{2}\|y - Xb\|_2^2 + \lambda \mathrm{pen}(b)$$

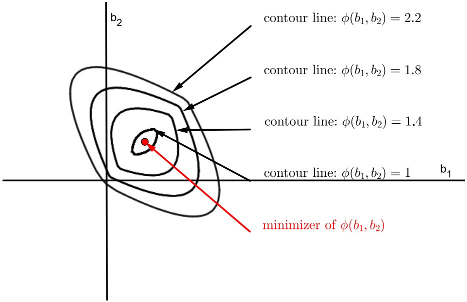
*is not unique if and only if* $\mathrm{row}(X) := \{X'z : z \in \mathbb{R}^n\}$ *intersects a face* $B^* = \mathrm{conv}\{u_0, u_1, \dots, u_l\}$ *whose dimension is* $< \dim(\ker(X))$.

- $\mathrm{pen}(b) = \|b\|_1$ –> $B^* = [-1, 1]^p$.
- $\mathrm{pen}(b) = \|b\|_\infty$ –> $B^* = \{x \in \mathbb{R}^p : \|x\|_1 \le 1\}$.
- $\mathrm{pen}(b) = \|Db\|_1$ for some $D \in \mathbb{R}^{m\times p}$ –> $B^* = D'[-1, 1]^m$.

$$S_{X,\|.\|_\infty}(y) := \underset{b\in\mathbb{R}^2}{\text{Argmin}} \underbrace{\frac{1}{2}\|y - Xb\|_2^2 + \max\{|b_1|, |b_2|\}}_{:=\phi(b)},$$

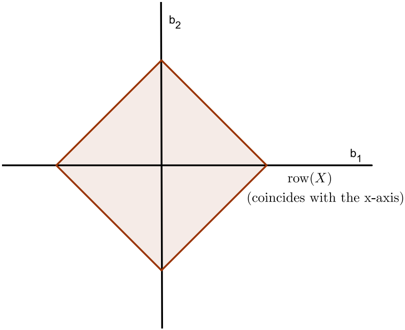
where $X = \begin{pmatrix} 1 & 1 \end{pmatrix}$ and $y = 2$.

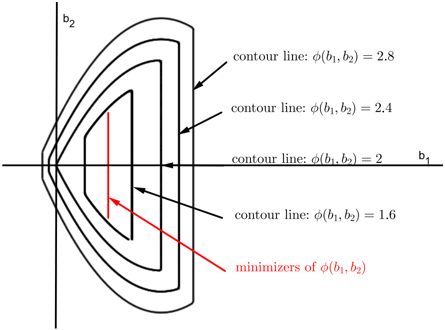


$$S_{X,\|.\|_\infty}(y) := \operatorname*{Argmin}_{b\in\mathbb{R}^2} \underbrace{\frac{1}{2}\|y - Xb\|_2^2 + \max\{|b_1|, |b_2|\}}_{:=\phi(b)},$$

where $X = \begin{pmatrix} 1 & 0 \end{pmatrix}$ and $y = 2$.



# Model pattern recovery

Consider the linear regression model $Y = X\beta + \varepsilon$ where $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ is an unknown parameter and $\varepsilon \in \mathbb{R}^n$ a random noise.
Our goal is to recover $\partial_{\text{pen}}(\beta)$.

Let $f : \mathbb{R}^p \to \mathbb{R}$. $s \in \mathbb{R}^p$ is a subgradient of $f$ at $x \in \mathbb{R}^p$ if

$$f(z) \geq f(x) + s'(z - x) \ \ \forall z \in \mathbb{R}^p.$$

The subdifferential $\partial_f(x)$ at $x$ is the set of all subgradients.

- $\partial_{\|.\|_1}(x) = \partial_{\|.\|_1}(z)$ iif $sign(x) = sign(z)$.
- $$\partial_{\|.\|_\infty}(x) = \partial_{\|.\|_\infty}(z) \text{ iif } \begin{cases} \{i : x_i = \|x\|_\infty\} = \{i : z_i = \|z\|_\infty\} \\ \{i : x_i = -\|x\|_\infty\} = \{i : z_i = -\|z\|_\infty\} \end{cases}$$
- When $D^{\mathrm{tv}}x = (x_2 - x_1, \ldots, x_p - x_{p-1})$ and $\mathrm{pen}(x) = \|D^{\mathrm{tv}}x\|_1$
  $$\partial_{\|D^{\mathrm{tv}}.\|_1}(x) = \partial_{\|D^{\mathrm{tv}}.\|_1}(z) \text{ iif } \begin{cases} \{i : x_{i+1} > x_i\} = \{i : z_{i+1} > z_i\} \\ \{i : x_{i+1} < x_i\} = \{i : z_{i+1} < z_i\} \end{cases}$$
- Let us assume that $\ker(D') = \{0\}$ then $\partial_{\|D.\|_1}(x) = \partial_{\|D.\|_1}(z)$ iif $sign(Dx) = sign(Dz)$.

# Path Condition

$$S_{X,\lambda \mathrm{pen}}(y) = \operatorname*{Argmin}_{b \in \mathbb{R}^p} \frac{1}{2}\|y - Xb\|_2^2 + \lambda \mathrm{pen}(b).$$

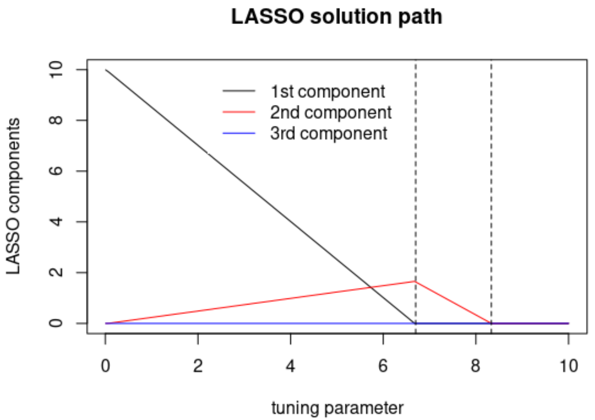
## Definition

$\partial_{\mathrm{pen}}(\beta)$ *satisfies the path condition with respect to* $X$ *and* $\mathrm{pen}$ *when*

$$\exists \lambda > 0 \; \exists \hat{\beta} \in S_{X,\lambda \mathrm{pen}}(X\beta) \text{ such that } \partial_{\mathrm{pen}}(\hat{\beta}) = \partial_{\mathrm{pen}}(\beta)$$

$$X = \begin{pmatrix} 5/6 & 1 & 0 \\ 1/3 & 0 & 1 \end{pmatrix}, \beta = (10, 0, 0).$$

$\mathrm{sign}(\beta)$ does not satisfy the path condition wrt $X$ and $\|.\|_1$.




$$\|X_{\bar{I}}' X_I (X_I' X_I)^{-1} \mathrm{sign}(\beta_I)\|_\infty = 30/29 > 1$$

# Necessary condition for model pattern recovery

**Theorem**

*Let $Y = X\beta + \varepsilon$ where $X \in \mathbb{R}^{n\times p}$, $\beta \in \mathbb{R}^p$ is an unknown parameter and $\varepsilon \in \mathbb{R}^n$ has a symmetric distribution. If the subdifferential of $\beta$ does not satisfies the path condition with respect to $X$ and $\mathrm{pen}$ then*

$$\mathbb{P}(\exists\lambda > 0\ \exists\hat{\beta} \in S_{X,\lambda\mathrm{pen}}(Y) \text{ such that } \partial_{\mathrm{pen}}(\hat{\beta}) = \partial_{\mathrm{pen}}(\beta)) \leq 1/2.$$

Consequently, when $\|X_{\bar{I}}'X_I(X_I'X_I)^{-1}\mathrm{sign}(\beta_I)\|_\infty > 1$ then whatever $\lambda > 0$ we have

$$\mathbb{P}(\mathrm{sign}(\hat{\beta}^{\mathrm{lasso}}(\lambda)) = \mathrm{sign}(\beta)) \leq 1/2.$$

# Accessibility

## Definition (Accessibility condition)

$\partial_{\mathrm{pen}}(\beta)$ *is accessible with respect to* $X$ *and* $\lambda\mathrm{pen}$ *when*

$$\exists y \in \mathbb{R}^p \; \exists \hat{\beta} \in S_{X,\lambda\mathrm{pen}}(y) \text{ such that } \partial_{\mathrm{pen}}(\hat{\beta}) = \partial_{\mathrm{pen}}(\beta)$$

## Proposition

$\partial_{\mathrm{pen}}(\beta)$ *is accessible with respect to* $X$ *and* $\lambda\mathrm{pen}$ *iif for every* $\gamma \in \mathbb{R}^p$ *we have*

$$X\beta = X\gamma \implies \mathrm{pen}(\gamma) \geq \mathrm{pen}(\beta).$$

Note that the accessibility condition does not depend on $\lambda$.

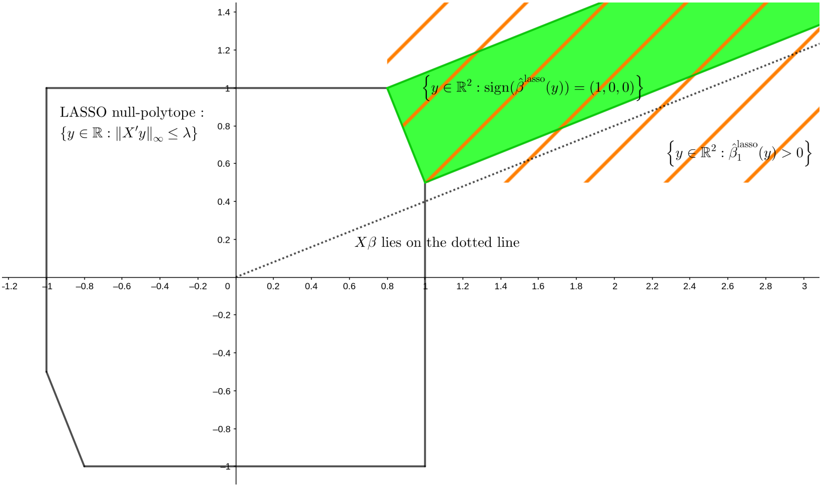
## Proposition

*The path condition implies the accessibility condition.*

$$X = \begin{pmatrix} 5/6 & 1 & 0 \\ 1/3 & 0 & 1 \end{pmatrix}, \beta = (10, 0, 0).$$

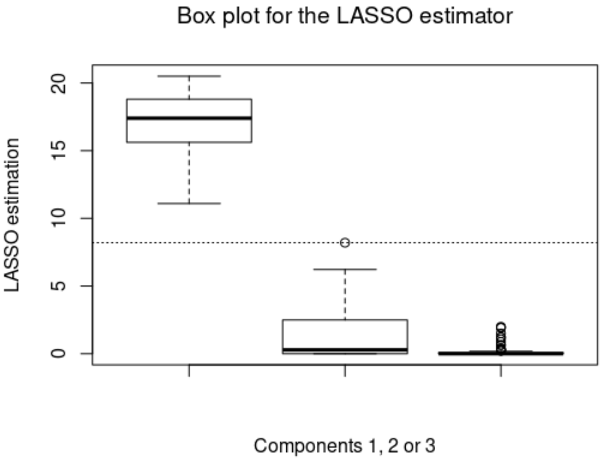
The path condition does not occur but $\partial_{\mathrm{pen}}(\beta)$ is accessible with respect to $X$ and $\|.\|_1$.

$$X = \begin{pmatrix} 5/6 & 1 & 0 \\ 1/3 & 0 & 1 \end{pmatrix}, \beta = (\beta_1, 0, 0) \text{ with } \beta_1 > 0.$$




For this figure, $\lambda = 1$.

$$X = \begin{pmatrix} 5/6 & 1 & 0 \\ 1/3 & 0 & 1 \end{pmatrix}, \beta = (20, 0, 0) \text{ et } \varepsilon \sim N(0, I)$$




For this figure, $\lambda = 1$.

# NSC for sign recovery by generalized LASSO

- One may use the penalty $\mathrm{pen}(b) = \|b\|_1$ in order to recover support: $\mathrm{supp}(\beta) = \{i : \beta_i \neq 0\}$
- One may use the penalty $\mathrm{pen}(b) = \|D^{\mathrm{tv}} b\|_1$ in order to recover the jump set: $\{i : \beta_i \neq \beta_{i+1}\} = \mathrm{supp}(D^{\mathrm{tv}}\beta)$

$$D^{\mathrm{tv}} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix}.$$

- More generally base on the penalty term $\mathrm{pen}(b) = \|Db\|_1$ we aim at recovering $\mathrm{sign}(D\beta)$.

Let $\hat{\beta}$ be a generalized LASSO estimator. The path condition is necessary for recovering $\mathrm{sign}(D\beta)$ via $\mathrm{sign}(D\hat{\beta})$. One may relax the path condition using the estimator $(D\hat{\beta})^{\tau}$.

## Theorem

*Necessary condition) If $\partial_{\|D.\|_1}(\beta)$ is not accessible with respect to $X$ and $\|D.\|_1$ then*

$$\forall y \in \mathbb{R}^n \ \forall \lambda > 0 \ \forall \hat{\beta} \in S_{X,\lambda\|D.\|_1}(y) \ \forall \tau \geq 0 \text{ we have } \\ \text{sign}((D\hat{\beta})^\tau) \neq \text{sign}(D\beta).$$

*Sufficient condition) Given $\varepsilon \in \mathbb{R}^n$, let us set $y^k = X(k\beta) + \varepsilon$. We assume that $S_{X,\lambda\|D.\|_1}(y^k)$ is a singleton and its unique element is $\hat{\beta}$. If $\partial_{\|D.\|_1}(\beta)$ is accessible with respect to $X$ and $\|D.\|_1$ then*

- $\exists k_0 \ \forall k \geq k_0 \ \exists \tau \geq 0 \text{ such that } \text{sign}((D\hat{\beta})^\tau) = \text{sign}(D\beta)$
- $\exists k_0 \ \forall k \geq k_0 \text{ we have } \text{supp}(D\beta) \in \{\emptyset, \{\pi(1)\}, \{\pi(1), \pi(2)\}, \dots, \text{supp}(D\hat{\beta})\}$. *Where $\pi$ is a permutation such that $|(D\hat{\beta})_{\pi(1)}| \geq \dots \geq |(D\hat{\beta})_{\pi(p)}|$.*

**PS:** The article "The Geometry of Model Recovery by Penalized and Thresholded Estimators" provides a similar theorem for penalized least squares estimators (including SLOPE,OSCAR,...).

# Sign recovery by LASSO and thresholded LASSO (Tardivel and Bogdan)

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

- **With the LASSO** one needs the irrepresentability condition (path condition in this presentation)

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

- **With the thresholded LASSO/BP** one needs the identifiability condition (accessibility condition in this presentation)

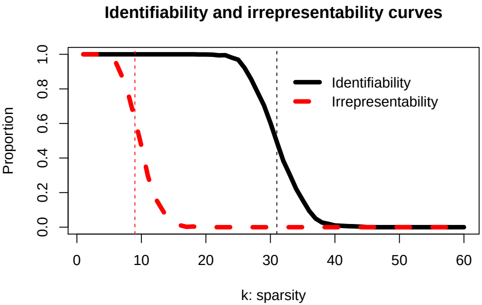
$$X\gamma = X\beta \Rightarrow \|\gamma\|_1 \geq \|\beta\|_1.$$

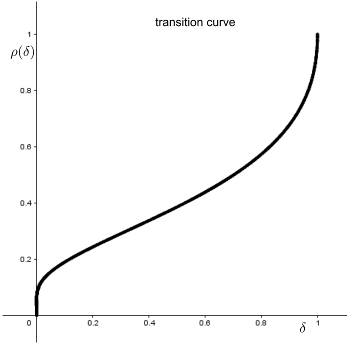
We remind that

$$\underset{\text{Irrepresentability condition}}{\text{(Path condition)}} \Rightarrow \underset{\text{Identifiability condition}}{\text{(accesibility condition)}}$$

# Standard Gaussian design

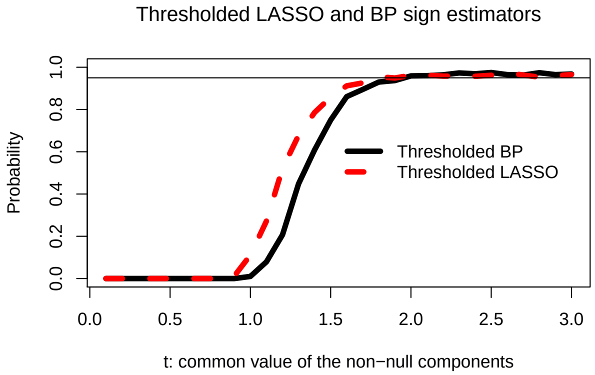
$X \in \mathbb{R}^{100\times 300}$ standard Gaussian matrix




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

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 $\beta$ are all equal to $t > 0$.



Thank you!

- PJC. Tardivel, T. Skalski, P. Graczyk, U. Scheider. The Geometry of Model Recovery by Penalized and Thresholded Estimators.

Related articles


- PJC. Tardivel and M. Bogdan. On the sign recovery by LASSO, thresholded LASSO and thresholded Basis Pursuit Denoising
- U. Schneider, PJC. Tardivel. The Geometry of Uniqueness and Model Selection of Penalized Estimators including SLOPE, LASSO and Basis Pursuit.