Chapitre 3 : Problème d'optimisation de LASSO

Le problème LASSO (Least Absolute Shrinkage & Selection Operator) a été introduit par Chen & Donoho (1994) et par Tibshirani (1996).

Définition : (LASSO) → méthode de MdC pénalisée

Soit $X \in \mathbb{R}^{n \times p}$ et $Y \in \mathbb{R}^n$ et $\lambda > 0$. L'ensemble des solutions du problème d'optimisation LASSO est,

$$S_{X,\lambda}(x) = \underset{b \in \mathbb{R}^p}{\operatorname{Argmin}} \left\{ \frac{1}{2} \| y - Xb \|_2^2 + \lambda \| b \|_1 \right\}.$$

↳ souvent utile en stats. ↳ variable réponse

Proposition : Soit $X \in \mathbb{R}^{n \times p}$, $y \in \mathbb{R}^n$ et $\lambda > 0$, alors,

i) L'ensemble $S_{X,\lambda}(x)$ est non-vide.

ii) Il existe $\hat{\beta} \in S_{X,\lambda}(x)$ tel que $\|\hat{\beta}\|_0 \leq rg(X)$

↳ Lorsque $p \gg n$, $\hat{\beta}$ a au moins $p-n$ 0

Proposition : (Valeurs ajustées LASSO)

Soit $X \in \mathbb{R}^{n \times p}$, $y \in \mathbb{R}^n$, $\lambda > 0$ et $S_{X,\lambda}(y)$ l'ens. des solutions du problème LASSO.

Si $\hat{\beta} \in S_{X,\lambda}(y)$ et $\tilde{\beta} \in S_{X,\lambda}(y)$ alors $X\hat{\beta} = X\tilde{\beta}$ et $\|\hat{\beta}\|_1 = \|\tilde{\beta}\|_1$.

↳ Preuve : Montrons que $X\hat{\beta} = X\tilde{\beta}$

Supposons que $X\hat{\beta} \neq X\tilde{\beta}$. Comme la fonction $u \in \mathbb{R}^n \mapsto \|u\|^2$ est strictement convexe alors en posant $\beta^\circ = \frac{\hat{\beta} + \tilde{\beta}}{2}$

on a, $\|y - X\beta^\circ\|_2^2 = \|\frac{1}{2}(y - X\hat{\beta}) + \frac{1}{2}(y - X\tilde{\beta})\|_2^2$

Par conséquent, puisque $\|\beta^\circ\|_1 \leq \frac{1}{2}\|\hat{\beta}\|_1 + \frac{1}{2}\|\tilde{\beta}\|_1$

On en déduit,

ineg triang.

$$\frac{1}{2}\|y - X\beta^\circ\|_2^2 + \lambda\|\beta^\circ\|_1$$

$$< \frac{1}{2}\left(\underbrace{\frac{1}{2}\|y - X\hat{\beta}\|_2^2 + \lambda\|\hat{\beta}\|_1}_{(1)} + \underbrace{\frac{1}{2}\|y - X\tilde{\beta}\|_2^2 + \lambda\|\tilde{\beta}\|_1}_{(2)}\right)$$

Ce qui implique que β° est soit strictement inférieur au premier terme (1) soit au deuxième (2), ce qui signifierait $\hat{\beta} \notin S_{X,\lambda}(y)$ ou $\tilde{\beta} \notin S_{X,\lambda}(y)$.

Ainsi par contradiction on a, $X\hat{\beta} = X\tilde{\beta}$. Par conséquent, comme $\frac{1}{2}\|y - X\hat{\beta}\|_2^2 + \lambda\|\hat{\beta}\|_1 = \frac{1}{2}\|y - X\tilde{\beta}\| + \lambda\|\tilde{\beta}\|_1$.

car minimiseurs LASSO

On en déduit $\|\hat{\beta}\|_1 = \|\tilde{\beta}\|_1$. $\square$

colonnes 2 à 2 ⊥

I / Expression explicite du LASSO où X orthogonal

Lemme : (opérateur proximal)

Soit $y \in \mathbb{R}^p$. La fonction $f : b \in \mathbb{R} \mapsto \frac{1}{2}\|y - b\|_2^2 + \lambda\|b\|_1$ atteint son minimum en un unique point :

$$\hat{\beta} = \begin{pmatrix} \mathrm{sign}(y_1)(|y_1| - \lambda)_+ \\ \vdots \\ \mathrm{sign}(y_p)(|y_p| - \lambda)_+ \end{pmatrix}$$

↳ Preuve :

On a,

$$\frac{1}{2}\|y-b\|_2^2 + \lambda\|b\|_1 = \sum_{i=1}^{p}\left(\frac{1}{2}(y_i-b_i)^2 + \lambda|b_i|\right)$$

On peut se restreindre à $p=1$, pour $b \neq 0$ on a,

$$f'(b) = \begin{cases} b - y + \lambda & b > 0 \\ b - y - \lambda & b < 0 \end{cases}.$$

• Distinction de cas : $\quad f'(b) = 0 \Leftrightarrow \begin{cases} b = y - \lambda \\ b = y + \lambda \end{cases}$

* Si $y < -\lambda$: (1^{er} cas)

b	$-\infty$		$y+\lambda$		0		$+\infty$
$f'(b)$		$-$	0	$+$	$\|$	$+$	
$f(b)$		↘		↗		↗	

Donc $\hat{\beta} = y + \lambda = -(|y| - \lambda) = \operatorname{sign}(y)(|y|-\lambda)_+$

* Si $-\lambda \leq y \leq \lambda$: (2^{e} cas)

b	$-\infty$		0		$+\infty$
$f'(b)$		$-$	$\|$	$+$	
$f(b)$		↘		↗	

Alors, $\hat{\beta} = 0 = \operatorname{sign}(y)(|y|-\lambda)_+$

* Si $y > \lambda$: (3^{e} cas)

b	$-\infty$		0		$y-\lambda$		$+\infty$
$f'(b)$		$-$	$\|$	$-$	0	$+$	
$f(b)$		↘		↘		↗	

Ainsi on obtient, $\hat{\beta} = y - \lambda = \text{sign}(y)(|y| - \lambda)_+$. $\square$

Corollaire : Soit $X = (X_1 | \dots | X_p) \in \mathbb{R}^{n \times p}$ telle que ${}^tXX = Id_p$, $y \in \mathbb{R}^n$ et $\lambda > 0$.

L'unique élément de $S_{X,\lambda}(y)$ est,

$$\boxed{\hat{\beta} = \begin{pmatrix} \text{sign}({}^tX_1 y)\,(|{}^tX_1 y| - \lambda)_+ \\ | \\ \text{sign}({}^tX_p y)\,(|{}^tX_p y| - \lambda)_+ \end{pmatrix}}$$

↳ Preuve : On pose $\theta = {}^tXy = ({}^tX_1 y | \dots | {}^tX_p y)$.

On a,
$$\|y - Xb\|_2^2 = \|y - X\theta + X\theta - Xb\|_2^2$$
$$= \|y - X\theta\|_2^2 + \|X\theta - Xb\|_2^2 + 2\,{}^t(y - X\theta)(X\theta - Xb)$$

Comme, ${}^t(y - X\theta)(X\theta - Xb) = {}^t(\underbrace{{}^tXy}_{=\theta} - \underbrace{{}^tXX\theta}_{=I\theta})(\theta - b)$
$$= 0.$$

On en déduit que,
$$\|y - Xb\|_2^2 = \|y - X\theta\|_2^2 + \underbrace{\|X\theta - Xb\|_2^2}_{=\|\theta - b\|_2^2}$$

Finalement, la fonction $b \in \mathbb{R}^p \mapsto \frac{1}{2}\|y - Xb\|_2^2 + \lambda\|b\|_1$ atteint son minimum au même point que la fonction précédente $b \in \mathbb{R}^p \longmapsto \frac{1}{2}\|\theta - b\|_2^2 + \lambda\|b\|_1$ d'où, d'après le lemme précédent.

$$\hat\beta = \left(\text{sign}(v_1)(|v_1|-\lambda)_+ , \dots , \text{sign}(v_p)(|v_p|-\lambda)_+\right) \quad \square$$

II / Caractérisation des solutions du LASSO

Proposition : Soit $X = (X_1 | \dots | X_p) \in \mathbb{R}^{n\times p}$, $y \in \mathbb{R}^n$ et $\lambda > 0$.

Les assertions suivantes sont équivalentes :

i) $\hat\beta \in S_{X,\lambda}(y)$

ii) $\begin{cases} \| {}^tX(y - X\hat\beta)\|_\infty \leq \lambda \\ {}^tX_i\,(y - X\hat\beta) = \lambda\, \text{sign}(\hat\beta_i) \quad \forall i \in \text{supp}(\hat\beta) \end{cases}$

iii) $\begin{cases} \| {}^tX(y - X\hat\beta)\|_\infty \leq \lambda \\ {}^t\hat\beta\, {}^tX(y - X\hat\beta) = \lambda \|\hat\beta\|_1 . \end{cases}$

Corollaire : Soit $X \in \mathbb{R}^{n\times p}$ et $\lambda > 0$,

Alors $S_{X,\lambda}(y) = \{0\}$ si et seulement si $\|{}^tXy\|_\infty \leq \lambda$

↳ Preuve :

• Existence :

$\Rightarrow$) Supposons que $\hat\beta = 0 \in S_{X,\lambda}(y)$, alors,

$$\| {}^tX(y - \underbrace{X\beta}_{0})\|_\infty \leq \lambda \iff \|{}^tXy\|_\infty \leq \lambda$$

$\Leftarrow$ Inversement, si $\|{}^tXy\| \leq \lambda$ alors $\begin{cases} \| {}^tX(y - X0)\|_\infty \leq \lambda \\ {}^t0\,{}^tX(y - X0) = \lambda \|0\|_1 \end{cases}$

Donc $0 \in S_{X,\lambda}(y)$

- Unicité : Si $\hat\beta \in S_{X,\lambda}(y)$ alors $\|\hat\beta\|_1 = 0$ donc $\hat\beta = 0$ et ainsi $S_{X,\lambda}(y) = \{0\}$ $\square$

Exercice : $X = \begin{pmatrix} 1 & 0 & 2/3 \\ 0 & 1 & 2/3 \end{pmatrix}$, $y = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ et $\lambda = 1$

Montrons que $\hat\beta = {}^t\left(0, 0, \frac{15}{8}\right) \in S_{X,\lambda}(y)$.

On rappelle que $\hat\beta$ doit vérifier,

$$\begin{cases} \|{}^tX(y - X\hat\beta)\|_\infty \leq 1 \\ {}^tX_i\,(y - X\hat\beta) = \lambda \operatorname{sign}(\hat\beta_i) \quad \forall i \in \operatorname{supp}(\hat\beta) \end{cases}$$

On a,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \\ \frac{2}{3} & \frac{2}{3} \end{pmatrix} \left(\underbrace{\begin{pmatrix} 2 \\ 2 \end{pmatrix} - \underbrace{\begin{pmatrix} 1 & 0 & 2/3 \\ 0 & 1 & 2/3 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \frac{15}{8} \end{pmatrix}}_{\begin{pmatrix} 5/4 \\ 5/4 \end{pmatrix}}}_{\begin{pmatrix} 3/4 \\ 3/4 \end{pmatrix}} \right) = \frac{1}{4} \begin{pmatrix} 3 \\ 3 \\ 4 \end{pmatrix}$$

Donc, $\left\| \frac{1}{4} \begin{pmatrix} 3 \\ 3 \\ 4 \end{pmatrix} \right\|_\infty = 1 \leq 1$

et $${}^t\hat\beta \left(\frac{1}{4} \begin{pmatrix} 3 \\ 3 \\ 4 \end{pmatrix} \right) = \frac{15}{8} = \underbrace{1}_{\lambda} \times \underbrace{\left\| \begin{pmatrix} 0 \\ 0 \\ \frac{15}{8} \end{pmatrix} \right\|_1}_{\|\hat\beta\|_1}$$

Donc $\hat\beta \in S_{X,\lambda}(y)$.

III / Condition nécessaire et suffisante pour l'unicité de LASSO.

Proposition : (condition suffisante pour l'unicité)

Soit $X \in \mathbb{R}^{n \times p}$ telle que $\mathrm{Ker}(X) = \{0\}$ alors pour tout $y \in \mathbb{R}^n$ et tout $\lambda > 0$, $S_{X,\lambda}(y) = \{0\}$.

Preuve : La fonction $b \in \mathbb{R}^p \longmapsto \frac{1}{2}\| y - Xb \|_2^2$ est strictement convexe.

En effet, $\mathrm{Hess} f(b) = {}^tXX$ est définie positive.

Ainsi la fonction $\phi : b \in \mathbb{R}^{n \times p} \longmapsto \frac{1}{2}\| y - Xb\|_2^2 + \lambda \|b\|_1$ est strictement convexe donc $S_{X,\lambda}(y)$ est $\mathcal{L}$ un singleton.

$\square$.

Proposition : (condition necessaire et suffisante pour l'unicité)

Soit $X \in \mathbb{R}^{n \times p}$, il existe $y \in \mathbb{R}^n$, il existe $\lambda > 0$ tel que $S_{X,\lambda}(y)$ n'est pas un singleton si et seulement si $\mathrm{im}({}^tX)$ coupe une face de $[-1;1]^p$ dont la dim est strictement inférieure à $\dim(\mathrm{Ker}(X))$.

Remarque : Lorsque $\mathrm{Ker}(X) = \{0\}$ alors $\mathrm{Im}({}^tX)$ ne coupe pas une face de $[-1;1]^p$ dont la dimension est strictement inférieure à $\dim(\mathrm{Ker}(X)) = 0$.

Donc pour tout $y \in \mathbb{R}^n$, $\forall \lambda > 0$, $S_{X,\lambda}(y)$ est un singleton.

IV - Chemin des solutions de LASSO.

Proposition : (Chemin des solutions de LASSO)

Soit $X \in \mathbb{R}^{n \times p}$ et $y \in \mathbb{R}^n$,

i) Le chemin des valeurs ajustées du LASSO, $\lambda \longmapsto X\hat{\beta}(\lambda)$, où $\hat{\beta} \in S_{X,\lambda}(y)$ est une fonction affine et continue par morceaux.

ii) Pour tout $\lambda > 0$, on suppose que $S_{X,\lambda}(y)$ a un unique éléments $\hat{\beta}(\lambda)$ alors le chemin des solutions du LASSO $\lambda \longmapsto \hat{\beta}(\lambda)$ est continue et affine par morceaux.

V / Relation entre LASSO et poursuite de base

Proposition : (relation entre LASSO et poursuite de base)

Soit $X \in \mathbb{R}^{n \times p}$,

i) Soit $y \in \mathbb{R}^n$, $\lambda > 0$ un élément $\hat{\beta} \in S_{X,\lambda}(y)$ est identifiable relativement à X et à la norme ℓ_1.

ii) Soit $y \in \mathrm{Im}(X)$. Pour tout $\lambda > 0$ et pour tout $\hat{\beta} \in S_{X,\lambda}(y)$, $\|\beta\|_1 \leq N_X(y)$.

iii) Supposons que $\mathrm{Im}({}^tX)$ ne coupe pas une face de $[-1;1]^p$ (dont la dim est $<$ dim(Ker(X^t)))

Si $y \in \mathrm{Im}(X)$ alors la fonction $\lambda > 0 \longmapsto \hat{\beta}(\lambda)$, où $\hat{\beta}$ est l'unique élément de $S_{X,\lambda}(y)$, converge lorsque $\lambda \to 0$ vers l'unique solution du pb PB $Xb = y$.

VI / Problème d'optimisation lié au LASSO.

Proposition : (Tibshirani et al 2012)

Soit $X \in \mathbb{R}^{n\times p}$, $\lambda > 0$ et $\hat\beta \in S_{X,\lambda}(y)$.

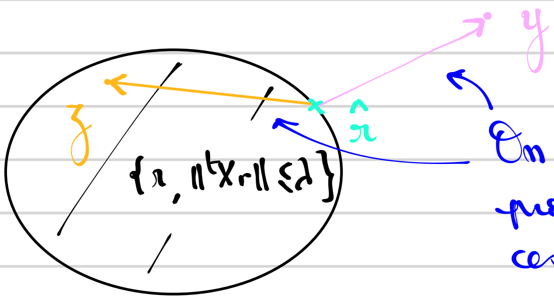
On pose $\hat r = y - X\hat\beta$ (appelés résidus de LASSO), alors

$$\hat r = \operatorname{argmin} \|y - r\|_2 \text{ s.c. } \|{}^tXr\|_\infty \leq \lambda$$

Preuve: Comme $\hat\beta \in S_{X,\lambda}(y)$ alors

$\lambda \geq \|{}^tX(y - X\beta)\|_\infty = \|{}^tX\hat r\|_\infty$ (contrainte satisf.)

Montrons que la projection de y sur l'ensemble convexe fermé $\{r \in \mathbb{R}^n : \|{}^tXr\|_\infty \leq \lambda\}$ vaut $\hat r$.



On veut montrer que le produit scalaire entre ces deux vecteurs est ≤ 0.

Soit $z \in \mathbb{R}^n : \|{}^tXz\| \leq \lambda$ alors,

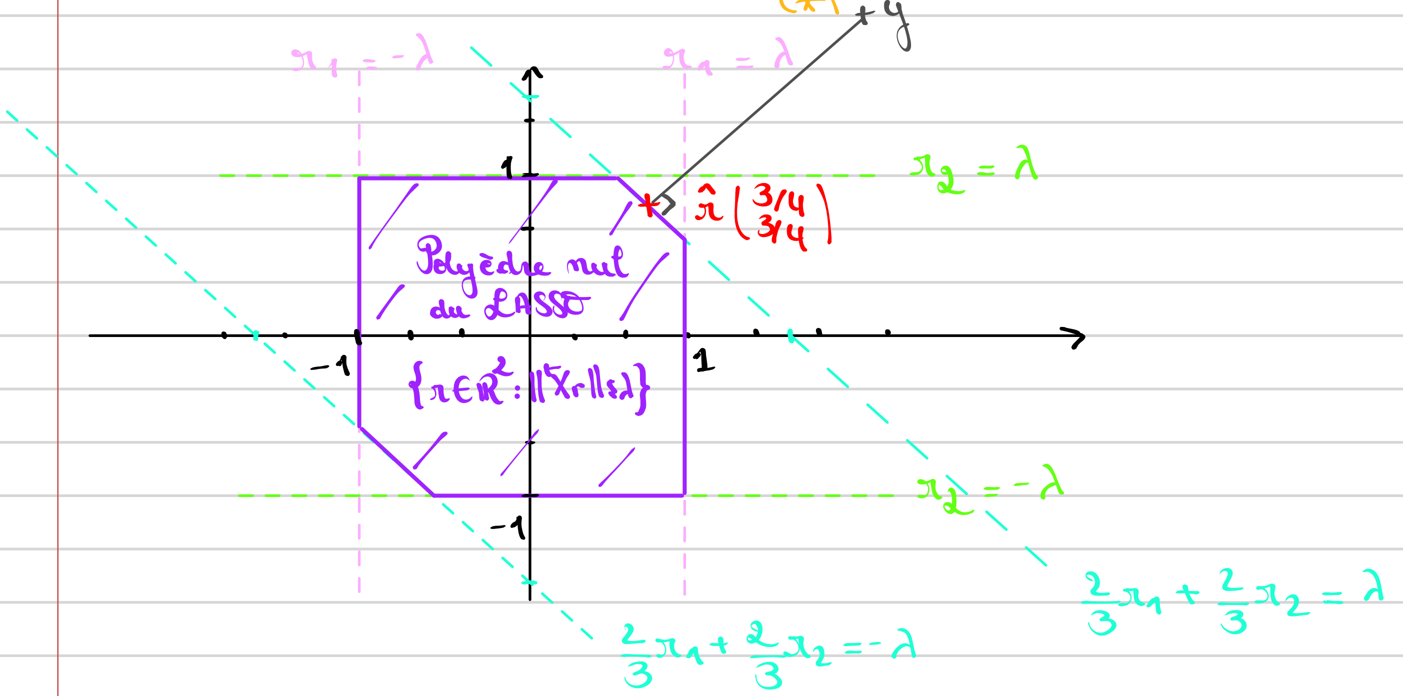
$$\begin{aligned}
{}^t(y - \hat r)(z - \hat r) &= {}^t\hat\beta\, {}^tX(z - \hat r)\\
&= {}^t\hat\beta\, {}^tXz - {}^t\hat\beta\, {}^tX\hat r\\
&= {}^t\hat\beta\, {}^tXz - \underbrace{{}^t\hat\beta\, {}^tX(y - X\hat\beta)}\\
&= \underbrace{{}^t\hat\beta\, {}^tXz} - \lambda\|\hat\beta\|_1\\
&\leq \|\hat\beta\|_1 \|{}^tXz\|_\infty - \lambda\|\hat\beta\|_1\\
&\leq \lambda\|\hat\beta\|_1 - \lambda\|\hat\beta\|_1 = 0
\end{aligned}$$

□

Exemple : $X = \begin{pmatrix} 1 & 0 & 2/3 \\ 0 & 1 & 2/3 \end{pmatrix}$, $y = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ et $\lambda = 1$

On rappelle que $\hat{\beta} = (0, 0, 15/8) \in S_{X,\lambda}(y)$.
La projection de y sur $\{ r \in \mathbb{R}^2 : \|{}^tXr\|_\infty \le \lambda \}$ vaut $\hat{r} = y - X\hat{\beta} = \begin{pmatrix} 3/4 \\ 3/4 \end{pmatrix}$

$$\|{}^tXr\|_\infty \le \lambda \iff \begin{cases} |r_1| \le \lambda \\ |r_2| \le \lambda \\ |\underbrace{\frac{2}{3}r_1 + \frac{2}{3}r_2}_{(*)}| \le \lambda \end{cases} \iff \begin{cases} -\lambda \le r_1 \le \lambda \\ -\lambda \le r_2 \le \lambda \\ -\lambda \le (*) \le \lambda \end{cases}$$



Exercice : $X = (X_1 | X_2 | X_3) = \begin{pmatrix} 1 & 1 & 2 \\ 1 & -1 & 0 \end{pmatrix}$

$y = \begin{pmatrix} 0.6 \\ 1.2 \end{pmatrix}$ et $\lambda = 1$

On commence par chercher $\|{}^tXr\|_\infty \le \lambda$,

$$\begin{cases} |r_1 + r_2| \le \lambda \\ |-r_1 + r_2| \le \lambda \\ |2r_1| \le \lambda \end{cases} \iff \begin{cases} -1 \le r_1 + r_2 \le 1 \\ -1 \le -r_1 + r_2 \le 1 \\ -\frac{1}{2} \le r_1 \le \frac{1}{2} \end{cases}$$

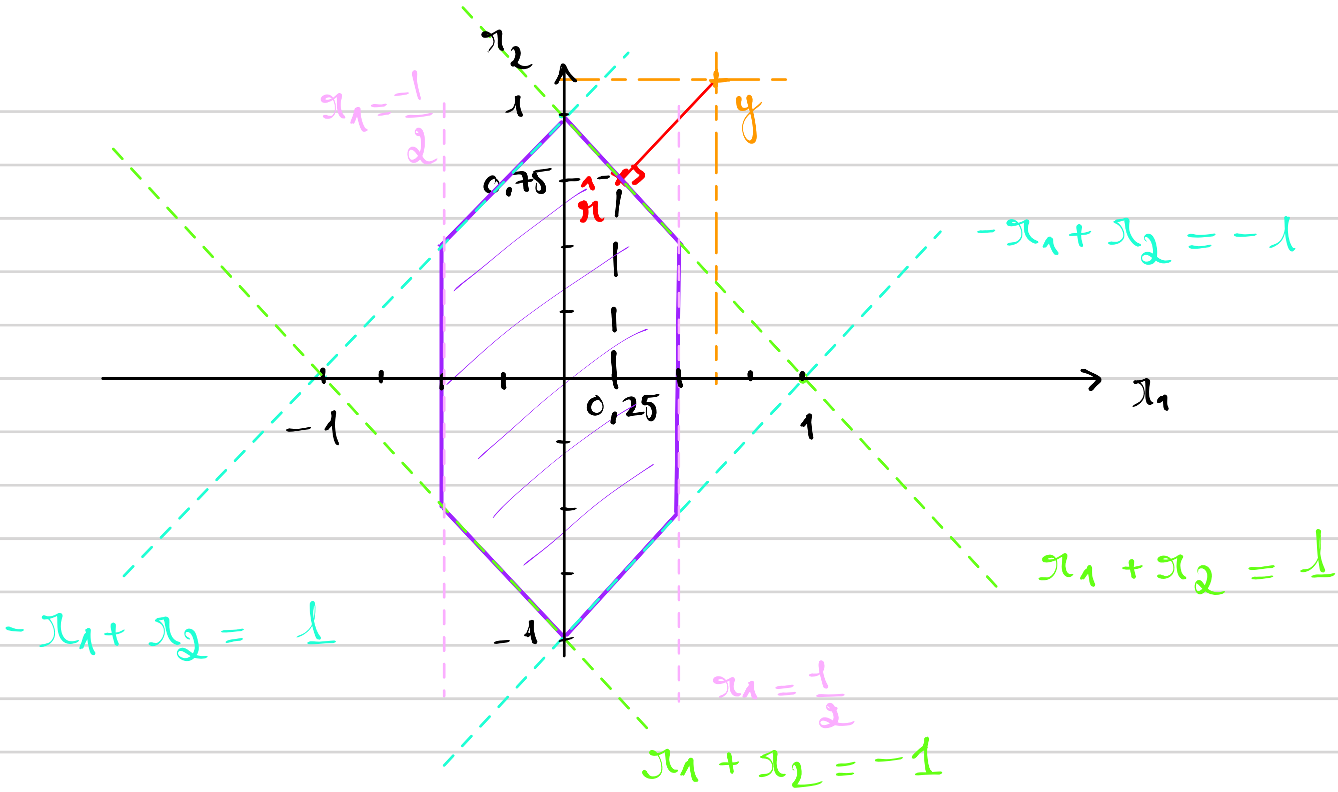


Figure : Polyèdre nul du LASSO

Graphiquement on observe que,

* $\hat{x}_1 + \hat{x}_2 = {}^tX \hat{x}_1 = \lambda$ donc $\hat{\beta}_1 \geq 0$
  ↳ $\hat{x}$ sur le bord $x_1 + x_2 = \lambda$

* $|\hat{x}_1 - \hat{x}_2| = |{}^tX_2 \hat{x}| = |{}^tX_2(y - X\hat{\beta})| < \lambda$
  ↳ $\hat{x}$ n'est pas sur le bord : $x_1 - x_2 = \pm\lambda$

donc $\hat{\beta}_2 = 0$.

* $|2x_1| = |{}^tX_3 \hat{x}| = |{}^tX_3(y - X\hat{\beta})| < \lambda$

donc $\hat{\beta}_3 = 0$.

Déterminons $\hat{\beta}_1$:

$$1 = \lambda = {}^tX_1(y - X\hat{\beta}) = {}^tX_1(y - \hat{\beta}_1 X_1)$$
$$= {}^tX_1 y - \hat{\beta}_1 {}^tX_1 X_1 = 1{,}8 - 2\hat{\beta}_1$$

d'où $\hat{\beta}_1 = 0{,}4$

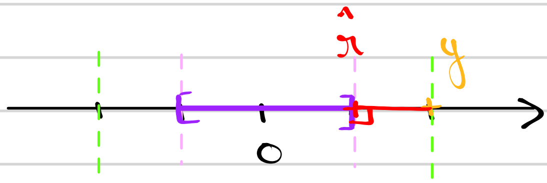
Ainsi, $\hat{\beta} = \begin{pmatrix} 0.4 \\ 0 \\ 0 \end{pmatrix}$

Exercice : $X = (2 \ \ 1)$ $y = 1$ et $\lambda = 1$

Déterminer $S_{X,\lambda}(y)$:

On cherche r tq $\|{}^tXr\|_\infty \leq \lambda = 1$,

$$\begin{cases} -1 \leq 2r \leq 1 \\ -1 \leq r \leq 1 \end{cases}$$



On observe que $|2\hat{r}| = |{}^tX_1\hat{r}| = |{}^tX_1(y - X\hat{\beta}_1)|$
$= \lambda$

donc $\hat{\beta}_1 \geq 0$

et $|\hat{r}| = |{}^tX_2\hat{r}| = |{}^tX_2(y - X\hat{\beta}_2)| < \lambda$

donc $\hat{\beta}_2 = 0$.

On détermine $\hat{\beta}_1$:

$$1 = \lambda = {}^tX_1(y - X\hat{\beta}) = {}^tX_1 y - \hat{\beta}_1 {}^tX_1 X_1$$
$$\Leftrightarrow 1 = 2 - \hat{\beta}_1 4 \Leftrightarrow \hat{\beta}_1 = \frac{1}{4}$$

On obtient $\tilde{\beta} = \begin{pmatrix} 1/4 \\ 0 \end{pmatrix}$.