FEUILLE 3 DE TRAVAUX DIRIGÉS

Dans toute la feuille, on considère le modèle linéaire gaussien suivant

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

où les ε_i, $1 \leq i \leq n$, sont i.i.d. $\mathcal{N}(0,1)$ ($\sigma = 1$ connu), $\beta^\star \in \mathbb{R}^p$ est fixé et X est une matrice réelle $n \times p$ avec des colonnes $X_1, \dots, X_p$ satisfaisant $||X_j|| = 1$ pour $1 \leq j \leq p$. Nous considérons également la fonction associée

$$F : \beta \in \mathbb{R}^p \mapsto n^{-1}||Y - X\beta||^2 + \lambda||\beta||_1/n.$$

EXERCICE 1. [*Minimisation coordonnée par coordonnée*]
Montrer que pour tout $j_0 \in \{1, \dots, p\}$ et pour tout $(\beta_j)_{j \neq j_0} \in \mathbb{R}^{p-1}$, la fonction

$$u \mapsto F(\beta_1, \dots, \beta_{j_0-1}, u, \beta_{j_0+1}, \dots, \beta_p)$$

atteint son minimum sur $\mathbb{R}$ en

$$u_{j_0} = R_{j_0}\left(1 - \frac{\lambda}{2|R_{j_0}|}\right)_+, \quad \text{avec } R_{j_0} = X_{j_0}^t\left(Y - \sum_{j \neq j_0} \beta_j X_j\right).$$

EXERCICE 2. [*Lasso dans le cas orthogonal*] On suppose $p \leq n$ et que les colonnes $X_1, \dots, X_p$ sont orthogonales deux à deux avec $||X_j|| = 1$ pour $1 \leq j \leq p$.

Pour $\lambda > 0$ fixé, on s'intéresse à l'estimateur Lasso $\widehat{\beta} = \widehat{\beta}_\lambda$:

$$\widehat{\beta}_\lambda \in \arg\min_{\beta \in \mathbb{R}^p}\{F(\beta)\}.$$

1. Montrer que $\widehat{\beta}$ peut s'écrire sous la forme d'un “seuillage doux" :

$$\widehat{\beta}_j = g_{\lambda/2}(X_j^t Y), \quad 1 \leq j \leq p, \quad \text{pour } g_s(u) = u(1 - s/|u|)_+$$

2. En utilisant l'exercice 3 et l'exercice 4 de la feuille 2, montrer que la procédure pénalisant par la norme $||\cdot||_0$ (au lieu de la norme $||\cdot||_1$) fournit un $\widetilde{\beta}$ qui peut s'écrire sous la forme d'un “seuillage dur" :

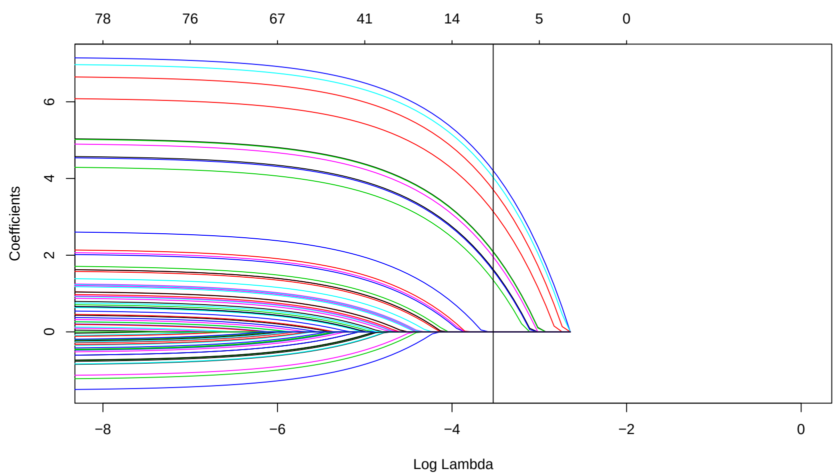
$$\widetilde{\beta}_j = h_{\lambda^{1/2}}(X_j^t Y), \quad \text{pour } h_s(u) = u\mathbf{1}\{|u| \geq s\}$$

3. Représenter graphiquement les fonctions h_s et g_s pour un $s > 0$. Ces fonctions sont-elles continues ?
4. Représenter les chemins de coefficients du Lasso, c'est-à-dire, pour chaque $j \in \{1, \dots, p\}$, la fonction $\lambda \mapsto (\widehat{\beta}_\lambda)_j$. Où peut-on lire les variables sélectionnées par le Lasso pour $\lambda = 1$?

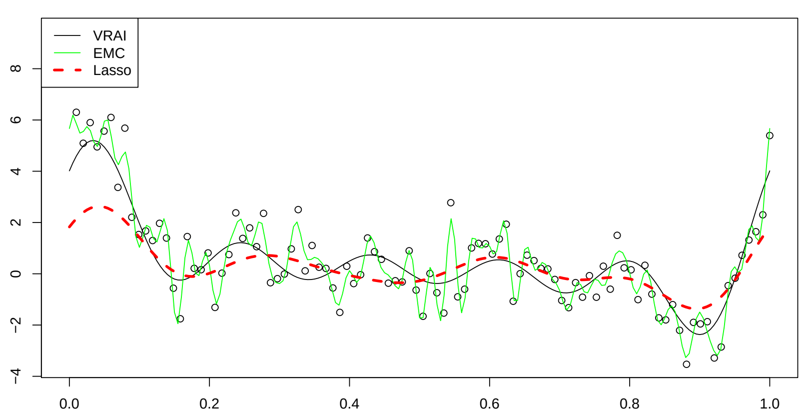
Exercice 3. On reprend l'énoncé du TD 1 exercice 1 (dans lequel $p \leq n$).

1. Écrire le modèle comme un modèle linéaire gaussien avec une matrice X de colonnes X_j, $1 \leq j \leq p$, satisfaisant $||X_j|| = 1$, pour tout j.
2. Calculer l'estimateur Lasso. En déduire un estimateur lasso $\widehat{f}$ de la fonction $f^\star$.
3. On prend $p = 81$ et $n = 101$ avec $\beta^\star_j \neq 0$ pour $j \leq 11$, $\beta^\star_j = 0$ pour $j > 11$. On représente le chemin de paramètres ainsi que l'estimateur $f^\star$. La valeur de λ utilisée est $\lambda = 2\sqrt{2\log p}$ et son log en représenté par le trait vertical dans le chemin de paramètre. Sur la figure (a), où peut-on lire les coordonnées j avec $\widehat{\beta}_j \neq 0$? Où peut-on lire les coordonnées de l'estimateur des moindres carrés ? Sur la figure (b), quel phénomène observe t-on ?

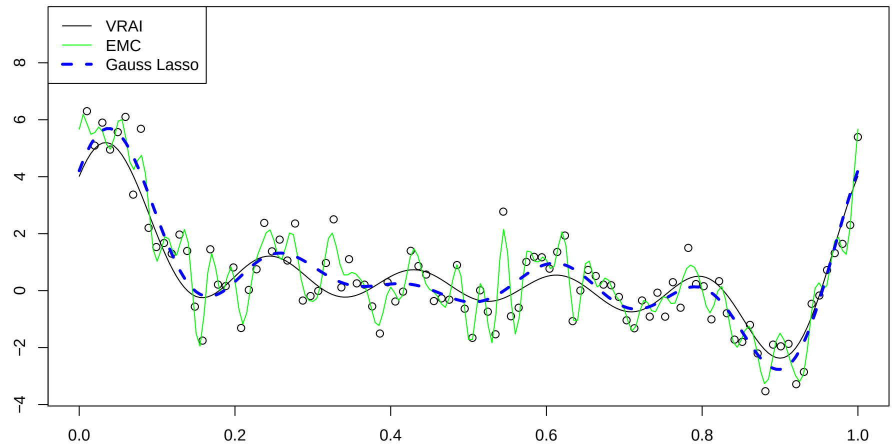
(a)



(b)



4. On cherche à améliorer l'estimateur Lasso. Quelle stratégie proposer ? (Indication : utiliser que le support de $\widehat{\beta}$ est proche de celui de $\beta^\star$). La nouvelle stratégie est appelée Gauss-Lasso. Que peut-on remarquer sur la figure ci-dessous ?

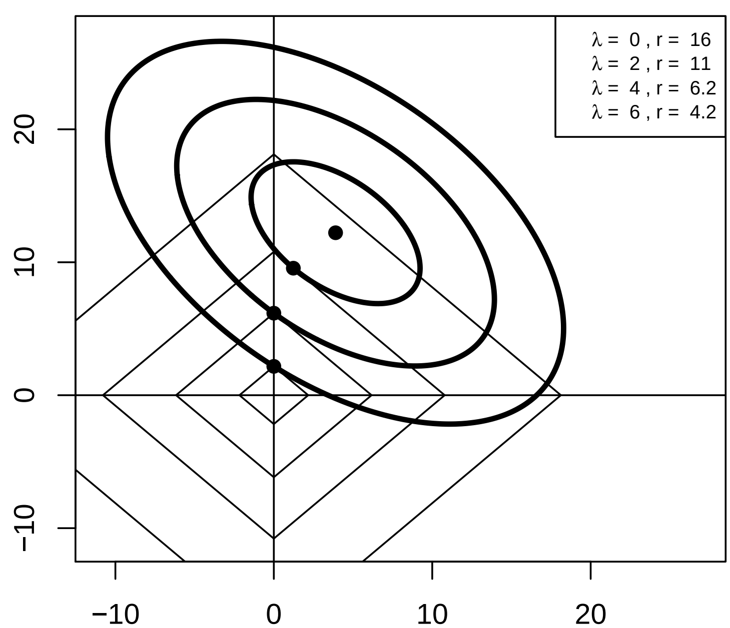


Exercice 4. [*Représentation graphique du Lasso pour $n = p = 2$*]
On considère une matrice X de taille 2×2 de la forme

$$X = \begin{pmatrix} 1 & \cos\theta \\ 0 & \sin\theta \end{pmatrix}, \quad \theta \in]0, \pi/2]$$

et le modèle linéaire gaussien $n = p = 2$ associé.

1. L'estimateur Lasso $\widehat{\beta}$ est-il unique ?
2. Etablir que, pour un $\widehat{r}$ bien choisi,
$$\widehat{\beta} = \arg\min_{\beta \in B_1(\hat{r})} \{||Y - X\beta||^2\}$$
en notant $B_1(r) = \{\beta \in \mathbb{R}^2 \mid ||\beta||_1 \leq r\}$ la boule L^1 de rayon r.
3. Commenter la figure ci-dessous.



Exercice 5. [*Existence de l'estimateur Lasso*]

On rappelle le théorème suivant : *Soit une fonction $f : \mathbb{R}^p \to \mathbb{R}$ convexe et continue. Alors f atteint son minimum sur $\mathbb{R}^p$ si elle n'a pas de direction de récession, c'est-à-dire si pour toute direction $y \in \mathbb{R}^p$ telle que $\forall x \in \mathbb{R}^p$, $u \in \mathbb{R} \mapsto f(x+uy)$ est décroissante, on a $y = 0$.*

Pour $\lambda \geq 0$, on s'intéresse à l'assertion $\arg\min_{\beta \in \mathbb{R}^p}\{n^{-1}||Y - X\beta||^2 + \lambda||\beta||_1/n\} \neq \emptyset$.

1. Montrer que l'assertion est vraie lorsque $\lambda = 0$.
2. En utilisant le résultat plus haut montrer que l'assertion est aussi vraie si $\lambda > 0$.

Exercice 6. [*Propriété des chemins du Lasso*]

On admettra le résultat suivant (lié au calcul de la sous-différentielle de la fonction convexe F) : *Pour tout $\beta \in \mathbb{R}^p$, β minimise F si et seulement si pour tout $j \in \{1, \dots, p\}$,*

$$\begin{cases} |X_j^t Y - X_j^t X\beta| \leq \lambda/2 & \text{si } \beta_j = 0 \\ X_j^t Y - X_j^t X\beta = \lambda \operatorname{sign}(\beta_j)/2 & \text{si } \beta_j \neq 0 \end{cases}$$

On se place sur l'événement où les $X_j^t Y$, $1 \leq j \leq p$, sont tous deux à deux distincts et non nuls. On pose $j_0 = \arg\max_j |X_j^t Y|$.

1. Vérifier que l'estimateur des moindres carrés satisfait la condition pour $\lambda = 0$.
2. Identifier un $M \geq 0$ tel que si $\lambda \geq M$, $\widehat{\beta}_\lambda = 0$ est une solution lasso.
3. On considère la fonction $f(\lambda) = X_{j_0}^t Y - (\lambda/2)\operatorname{sign}(X_{j_0}^t Y)$ pour $\lambda \in [0, M]$. Montrer qu'il existe $\lambda_1 \in]0, M[$ tel que la quantité
$$\max_{j \neq j_0} |X_j^t Y - X_j^t X_{j_0} f(\lambda)| - \lambda/2$$
est nulle en $\lambda = \lambda_1$ et est < 0 pour $\lambda \in]\lambda_1, M[$.
4. Vérifier que $(\widehat{\beta}_\lambda)_{j_0} = f(\lambda)$ et $(\widehat{\beta}_\lambda)_j = 0$ pour $j \neq j_0$ fournit une solution LASSO pour $\lambda \in]\lambda_1, M[$. Que dire du chemin $\lambda \in]\lambda_1, +\infty[\mapsto \widehat{\beta}_\lambda$?

Exercice 7. [*Constante de compatibilité*]

Pour $m \subset \{1, \dots, p\}$ et X une matrice réelle $n \times p$ de colonnes $X_1, \dots, X_p$ de normes égales à 1. On définit la constante de compatibilité :

$$c^2(m, X) = \inf_{z \in \mathcal{C}(m)} \left\{ \frac{|m| \times ||Xz||_2^2}{\left(\sum_{j \in m} |z_j|\right)^2} \right\},$$

pour le cône $\mathcal{C}(m) = \left\{ z \in \mathbb{R}^p \mid \sum_{j \in m^c} |z_j| \leq 5 \sum_{j \in m} |z_j| \right\}$. On note $\rho = \max_{j \neq j'} |X_j^t X_{j'}|$. et on suppose $|m| < (11\rho)^{-1}$. On se propose de montrer que $c^2(m, X) \geq 1 - 11\rho|m|$.

1. En séparant les coordonnées sur m de celles sur m^c, montrer que pour tout $z \in \mathbb{R}^p$, on a $||Xz||_2^2 \geq A - 2B$, avec

$$A = \sum_{j \in m} \sum_{j' \in m} X_j^t X_{j'} z_j z_{j'}$$
$$B = \sum_{j \in m} \sum_{j' \in m^c} |X_j^t X_{j'}| \, |z_j| \, |z_{j'}|$$

2. Montrer que $A \geq \sum_{j \in m} z_j^2 - \rho \left(\sum_{j \in m} |z_j| \right)^2$.

3. Montrer que lorsque $z \in \mathcal{C}(m)$, on a $B \leq 5\rho \left(\sum_{j \in m} |z_j| \right)^2$.

4. En déduire que $||Xz||_2^2 \geq (|m|^{-1} - 11\rho) \left(\sum_{j \in m} |z_j| \right)^2$ et conclure.