Optimisation Convexe - Devoir 3

Tamim EL AHMAD

Novembre 2019

Question 1

On considère le problème LASSO suivant :

$$\text{minimiser } \frac{1}{2}||Xw - y||_2^2 + \lambda||w||_1 \qquad \text{(LASSO)}$$

Selon la variable $w \in \mathbb{R}^n$ avec $X = (x_1^T, ..., x_n^T) \in \mathbb{R}^{n \times d}$, $y = (y_1, ..., y_n) \in \mathbb{R}^n$ et $\lambda > 0$ une variable de régularisation.
En reformulant les variables, ce problème équivaut au suivant :

$$\begin{aligned} \text{minimiser } & \frac{1}{2}||s||_2^2 + ||t||_1 \qquad \text{(LASSO)} \\ \text{t. q. } & Xw - y - s = 0 \\ & \lambda w - t = 0 \end{aligned}$$

Selon les variables $s \in \mathbb{R}^n$ et $t \in \mathbb{R}^d$.

Soit $\nu_1 \in \mathbb{R}^n$ et $\nu_2 \in \mathbb{R}^d$, on calcule le Lagrangien :

$$\begin{aligned} L(w, s, t, \nu_1, \nu_2) &= \frac{1}{2}||s||_2^2 + ||t||_1 + \nu_1^T(Xw - y - s) + \nu_2^T(\lambda w - t) \\ &= \frac{1}{2}||s||_2^2 - \nu_1^T s + ||t||_1 - \nu_2^T t + (X^T\nu_1 + \lambda\nu_2)^T w - \nu_1^T y \end{aligned}$$

On pose :

$$\begin{aligned} g(\nu_1, \nu_2) &= \inf_{w,s,t} L(w, s, t, \nu_1, \nu_2) \\ &= \begin{cases} \inf_s(\frac{1}{2}||s||_2^2 - \nu_1^T s) - ||\nu_2||_1^* - y^T\nu_1 & \text{si } X^T\nu_1 + \lambda\nu_2 = 0 \\ -\infty & \text{sinon.} \end{cases} \end{aligned}$$

Or, comme prouvé dans le DM2, $||\nu_2||_1^* = 0$ si $|\nu_2| \preceq 1_d$ (avec $|\nu_2| = (|\nu_{21}|, ..., |\nu_{2d}|)$ et $1_d = \underbrace{(1, ..., 1)}_{d \text{ fois}}$)

De plus, soit $f : s \in \mathbb{R}^n \longmapsto \frac{1}{2}||s||_2^2 - \nu_1^T s$:

$$\nabla f(s) = s - \nu_1 \Rightarrow (\nabla f(s) = 0 \iff s = \nu_1)$$

$$\Rightarrow \inf_{s\in\mathbb{R}^n} = f(\nu_1) = \frac{1}{2}||\nu_1||_2^2 - ||\nu_1||_2^2 = -\frac{1}{2}||\nu_1||_2^2$$

Donc :

$$g(\nu_1,\nu_2) = \begin{cases} -y^T\nu_1 - \frac{1}{2}||\nu_1||_2^2 & \text{si } X^T\nu_1 + \lambda\nu_2 = 0 \text{ et } |\nu_2| \preceq 1_d \\ -\infty & \text{sinon.} \end{cases} \tag{1}$$

Par conséquent, le problème dual est le suivant :

$$\begin{aligned} \max_{\nu_1\in\mathbb{R}^n} \ & -y^T\nu_1 - \frac{1}{2}||\nu_1||_2^2 \\ \text{t. q. } & X^T\nu_1 = -\lambda\nu_2 \\ & |\nu_2| \preceq 1_d \end{aligned}$$

Ce qui équivaut au problème suivant :

$$\begin{aligned} \min_{\nu_1\in\mathbb{R}^n} \ & y^T\nu_1 + \frac{1}{2}||\nu_1||_2^2 \\ \text{t. q. } & |X^T\nu_1| \preceq \lambda 1_d \end{aligned}$$

Et donc :

$$\begin{aligned} \min_{\nu_1\in\mathbb{R}^n} \ & y^T\nu_1 + \frac{1}{2}||\nu_1||_2^2 \\ \text{t. q. } & X^T\nu_1 \preceq \lambda 1_d \\ & -X^T\nu_1 \preceq \lambda 1_d \end{aligned}$$

Finalement :

$$\boxed{\begin{aligned} \min_{v\in\mathbb{R}^n} \ & p^Tv + v^TQv \\ \text{t. q. } & Av \preceq b \end{aligned}}$$

Avec :

$$\boxed{\begin{aligned} p &= y \\ Q &= \frac{1}{2}I_n \\ A &= \begin{pmatrix} X^T \\ -X^T \end{pmatrix} \\ b &= \lambda 1_{2d} \end{aligned}}$$

Question 3

On regarde d'abord l'évolution de l'écart entre f^* et $f(v_t)$ en fonction du nombre d'itérations de Newton pendant le déroulement de l'algorithme "barrière". L'instant t correspond à l'itération à laquelle on se trouve pendant l'algorithme "barrière", entre chaque instant t et $t+1$, il peut se dérouler plusieurs itérations de Newton, c'est-à-dire d'itérations pendant chaque appel à la méthode

de Newton. f^* est donné par la valeur de $f(v_{t_f})$, v_{t_f} étant le dernier vecteur v obtenu par la méthode "barrière", correspondant à v^* avec une précision ϵ.
Avec $n = 200$, $d = 20$, $\lambda = 10$, $t = 1$ pour la méthode "barrière", $\alpha = 0.1$ et $\beta = 0.6$ pour le "backtracking", et la matrice X générée aléatoirement avec chaque élément choisi selon une loi uniforme entre 0 et 1, le vecteur w_0 de référence choisi de même et finalement $Y = Xw_0 + \xi$ avec ξ vecteur gaussien dont chaque élément suit une loi normale centrée réduite.
Finalement, on choisit le vecteur de départ $v_0 = O_n$ et $\epsilon = 10^{-6}$.

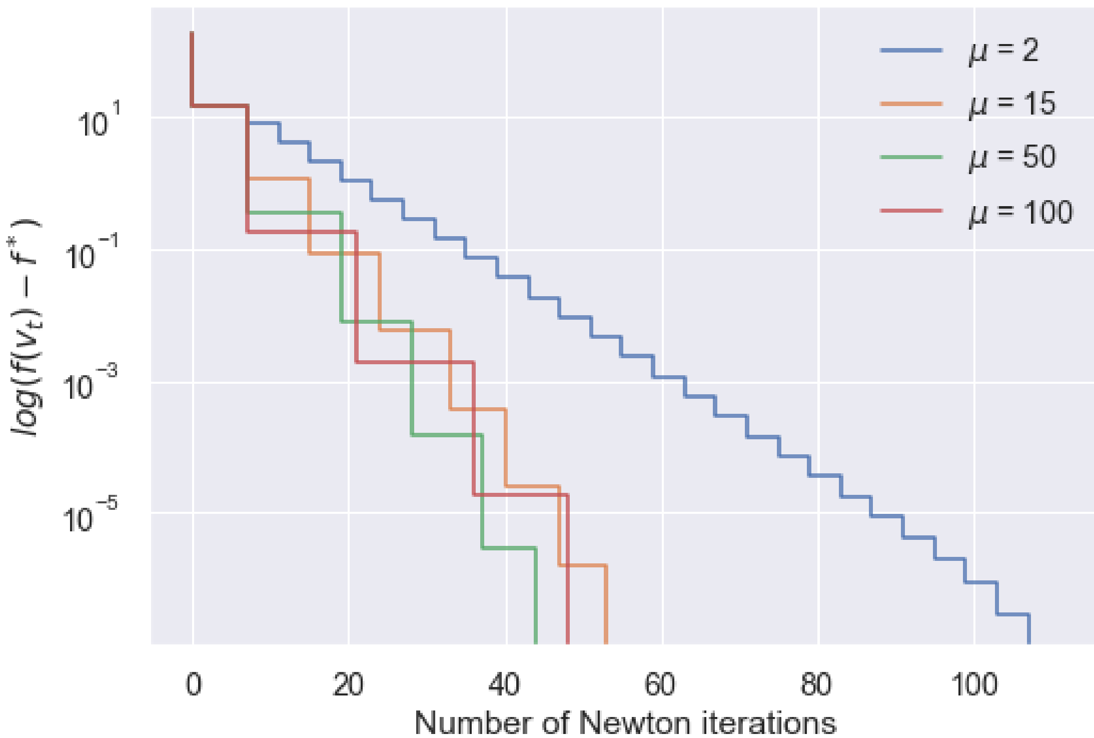

FIGURE 1 – Evolution de l'écart entre f^* et $f(v_t)$ en fonction du nombre d'itérations de Newton pendant le déroulement de l'algorithme "barrière"

On s'intéresse maintenant à retrouver l'optimal w^* en fonction de la solution v^* du problème dual.
Puisque le problème satisfait les conditions Slater, les solutions s^*, t^* et w^* et du dual v^* et u^* satisfont les conditions KKT :

$$\begin{aligned}\frac{\partial}{\partial s} L(w^*, s^*, t^*, v^*, u^*) = 0 &\iff s^* = v^* \\ &\iff Xw^* - y = v^* \\ &\iff w^* = (X^T X)^{-1} X^T (v^* + y)\end{aligned}$$

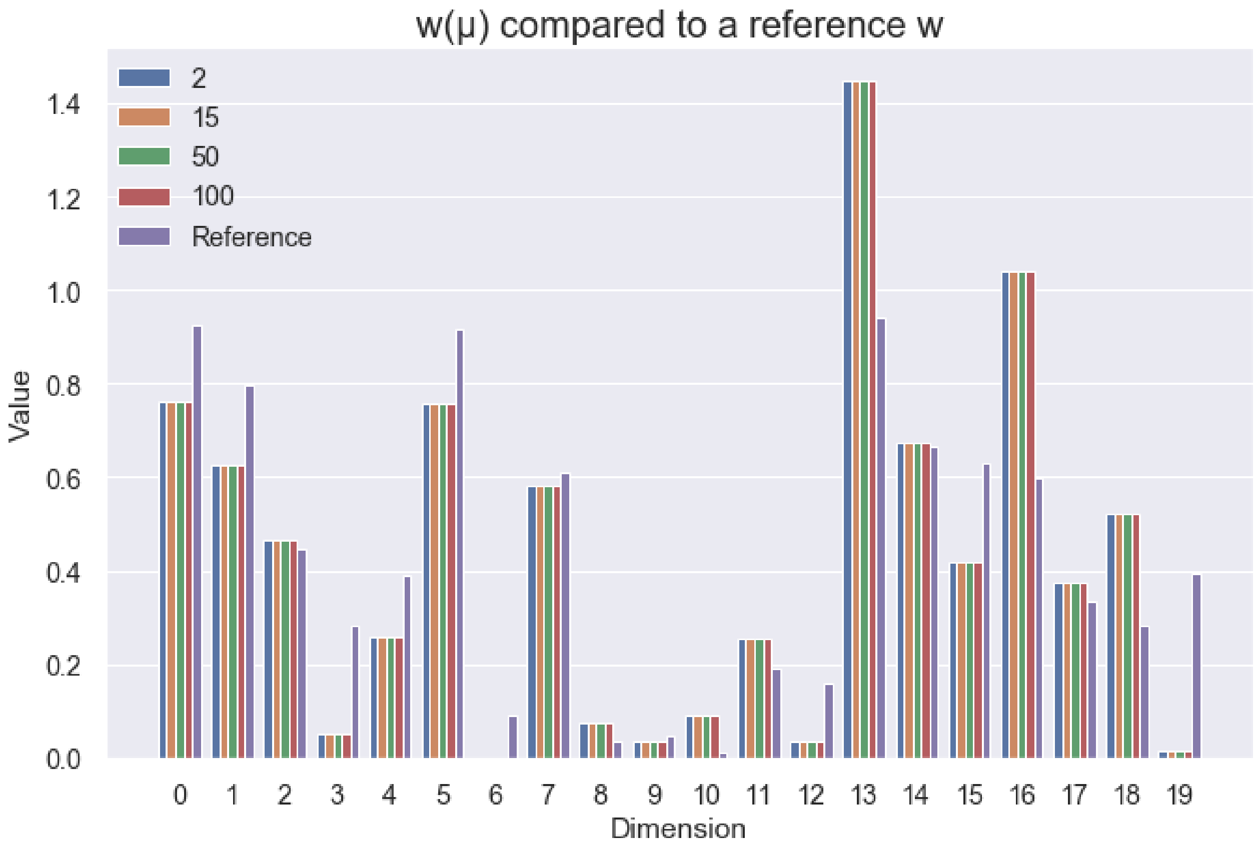

FIGURE 2 – Valeurs de chaque coefficient de w^*, pour chaque $\mu \in \{5, 15, 50, 100\}$, et du w_0 de référence. On voit que μ n'a pas d'influence sur le w^* trouvé et que la méthode "barrière" permet de trouver une bonne approximation de la solution optimale.