Master M2 MVA 2018/2019
Optimisation convexe - HWK 3

Souhaib ATTAIKI

November 10, 2018

Question 1 : Le problème dual de LASSO

Le problème *LASSO* est équivalent à :

$$\begin{cases} min\frac{1}{2} \parallel Z \parallel_2^2 + \lambda \parallel \omega \parallel_1 \\ s.c \\ Z = X\omega - y \end{cases} \tag{1}$$

De plus, les conditions de Slater sont vérifiées, donc on a la propriété de la dualité forte.

Le lagrangien du problème 1 est :

$$\mathcal{L}(Z, \omega, v) = \frac{1}{2} \parallel Z \parallel_2^2 + \lambda \parallel \omega \parallel_1 + v^T(X\omega - y - Z)$$

ce qui est équivalent à :

$$\mathcal{L}(Z, \omega, v) = \big(\frac{1}{2} \parallel Z \parallel_2^2 - v^T Z\big) + \big(\lambda \parallel \omega \parallel_1 + (X^T v)^T \omega\big) - y^T v$$

donc :

$$\mathcal{L}(Z, \omega, v) = \mathcal{L}_1(Z, v) + \mathcal{L}_2(\omega, v) - y^T v$$

Donc on peut obtenir la fonction dual $g(v) = \inf_{Z,\omega} \mathcal{L}$ en séparent le problème de minimisation en deux parties $g(v) = \big(\inf_Z \mathcal{L}_1\big) + \big(\inf_\omega \mathcal{L}_2\big) - y^T v$

Pour la minimisation par rapport à Z, on a la fonction $\mathcal{L}_1$ est une fonction convexe par rapport à Z comme somme positive d'une norme et une fonction linéaire, et on trouve par dérivation que :

$$\inf_Z \mathcal{L}_1 = -\frac{1}{2} \parallel v \parallel_2^2$$

Concernant la minimisation par rapport à ω, on a trouvé au DM2 (exercice 2) que :

$$\inf_{\omega} \mathcal{L}_2 = -\sup_{\omega} \lambda\big(- \parallel \omega \parallel_1 - \frac{1}{\lambda}(X^T v)^T \omega\big) = \begin{cases} 0 & si \parallel X^T v \parallel_\infty \leq \lambda \\ -\infty & sinon \end{cases}$$

Donc le problème dual du *LASSO* est le suivant :

$$\begin{cases} \max -\frac{1}{2} \parallel v \parallel_2^2 - y^T v \\ s.c \\ \parallel X^T v \parallel_\infty \leq \lambda \end{cases}$$

qui est équivalent à :

$$\begin{cases} \min v^T Q v + p^t v \\ s.c \\ Av \preccurlyeq b \end{cases}$$

Avec :

$$Q = \frac{1}{2}\mathbb{I} \in S^+$$
$$p = y$$
$$b = (\lambda, \lambda, ..., \lambda)$$
$$A = [X^T, -X^T]$$

b est un vecteur de taille $2d$, et $A \in \mathbb{R}^{2d \times n}$

Question 2 : Implémentation

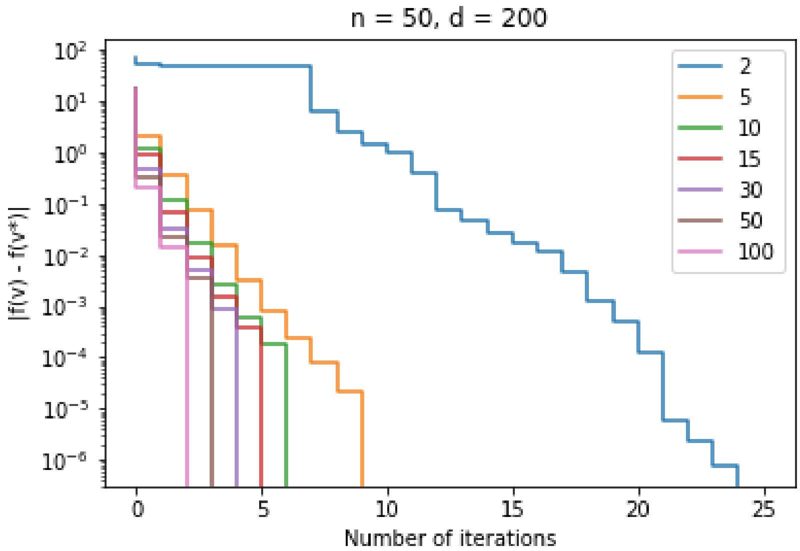
Les deux fonctions sont implémenter dans le fichier *lasso.py*. Pour refaire les simulations de la question3, il suffit d'exécuter le notebook *visualisations.ipynb* ou de lire de *pdf* généré à partir de ce dernier *visualisations.pdf*.

Question 3 : Tests sur des matrices X et observations Y générées de manière aléatoire

Pour cette question, on a pris : n = 50, d = 200, $\lambda = 10$, $\epsilon = 1e^{-6}$. Pour générer les données de tests, on a généré la matrice d'une manière aléatoire X ainsi que le vecteur ω, et on a généré y par $y = X\omega + bruit$.

Pour l'initialisation de l'algorithme, on a prit $v_0 = 0$ qui est un point faisable.

Dans la figure suivante suivante est tracé le gap $f(v) - f^*$ en fonction des itérations en utilisant un échelle logarithmique, et ceci pour $\mu = 2, 5, 10, 15, 30, 50, 100$.



On constate que le nombre d'itérations augmente si μ diminue. On constate aussi que les différentes initialisations donnent pratiquement le même résultat f^*, sauf pour $\mu = 2$, qui donne un résultat meilleur par rapport aux autres.

Pour trouver ω, on utilise la formule $\omega^* = X^+(y + v^*)$ par ce qu'on a prouvé avant que $Z = v$. On constate que les différents μ donnent pratiquement le même ω^*.

Pour les données qu'on a utilisé, on a remarqué qu'un μ plus grand converge rapidement (en terme de temps) vers une solution meilleure. S'il faut choisir, on utilisera $\mu = 100$.