Convex Optimization-Homework 3

Hugo Blanc

November 2021

1 Derive the dual problem of LASSO

Le problème LASSO se présente sous la forme suivante :

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

, tel que $w \in \mathbb{R}^d$ est la variable sur laquelle on minimise, les données sont $X = (x_1^T, x_2^T, ..., x_n^T) \in \mathbb{R}^{n\times d}$,$y = (y_1, ..., y_n) \in \mathbb{R}^n$ et $\lambda > 0$ est un paramètre de régularisation.
Pour déterminer le problème dual de ce problème (LASSO), on commence par introduire une variable $z = Xw - y$ et on réécrit le problème sous la forme suivante :

$$\begin{cases} \text{minimize } \frac{1}{2}||z||_2^2 + \lambda||w||_1 \\ \text{tel que } z = Xw - y \end{cases} \tag{1}$$

Par la suite, on écrit le Lagrangien du problème précédent :

$$\begin{aligned} L(w, z, \nu) &= \frac{1}{2}||z||_2^2 + \lambda||w||_1 + \nu^T(Xw - y - z) \\ &= \frac{1}{2}||z||_2^2 - \nu^T z + \lambda||w||_1 + \nu^T Xw - \nu^T y \end{aligned}$$

Ensuite, la fonction duale de Lagrange est donc :

$$\begin{aligned} g(\nu) &= \inf_{w,z} L(w, z, \nu) \\ &= \inf_z \{\frac{1}{2}||z||_2^2 - \nu^T z\} + \inf_w \{\lambda||w||_1 + \nu^T Xw\} - \nu^T y \\ &= -\sup_z \{\nu^T z - \frac{1}{2}||z||_2^2\} - \sup_w \{-(X^T\nu)^T w - \lambda||w||_1\} - \nu^T y \end{aligned}$$

Aussi, on retrouve des expressions proche de celles étudiés durant le deuxième devoir. Il nous faut en effet calculer les fonctions conjuguées des deux fonctions suivantes :

$$z \mapsto \frac{1}{2}||z||_2^2 \quad w \mapsto \lambda||w||_1$$

Fonction conjuguée de $z \mapsto \frac{1}{2}||z||_2^2$:
Soit $\nu \in \mathbb{R}^n$, on calcule l'expression de la fonction conjuguée évaluée en ν :
On note $f : z \mapsto \nu^T z - \frac{1}{2}||z||_2^2$, cette fonction est de classe $\mathscr{C}^\infty$ et :

$$\nabla f(z) = \nu - z$$
$$\nabla^2 f(z) = -I_n \leq 0$$

Donc, la fonction f est concave, ainsi tout point critique de f maximisera f.
Ainsi, $z = \nu$ maximise la fonction f et donc on en déduit que :

$$\sup_z\{\nu^T z - \frac{1}{2}||z||_2^2\} = \frac{1}{2}||\nu||_2^2$$

Fonction conjuguée de $g : w \mapsto \lambda||w||_1$:
Soit $x \in \mathbb{R}^d$, et soit $w \in \mathbb{R}^d$ avec $w \neq 0$,

$$\begin{aligned}(x, w) - \lambda||w||_1 &= ((x, \frac{w}{||w||_1}) - \lambda)||w||_1 \\ &\leq \left(\sup_{u, ||u||_1 = 1}\{(x, u)\} - \lambda\right)||w||_1\end{aligned}$$

On distingue alors deux cas :

— Si $\sup_{u,||u||_1=1}\{(x, u)\} \leq \lambda$:
  Dans ce cas, on a :

$$(x, w) - \lambda||w||_1 \leq 0$$

  De plus, en $w = 0$ on a $(x, w) - \lambda||w||_1 = 0$.
  Aussi, la fonction conjuguée vaut 0 en x si $\sup_{u,||u||_1=1}\{(x, u)\} \leq \lambda$.

— Si $\sup_{u,||u||_1=1}\{(x, u)\} > \lambda$:
  Dans ce cas, si on prend w_* tel que $(x, \frac{w_*}{||w_*||_1}) > \lambda$ et que l'on pose $w_\alpha = \alpha w_*$, on aura finalement :

$$(x, w_\alpha) - \lambda||w_\alpha||_1 = ((x, \frac{w_\alpha}{||w_\alpha||_1}) - \lambda)||w_\alpha||_1 \underset{\alpha \to +\infty}{\longrightarrow} +\infty$$

  Aussi, on obtient que la fonction conjuguée vaut $+\infty$ en x si $\sup_{u,||u||_1=1}\{(x, u)\} > \lambda$

Par ailleurs, on peut remarquer que $\sup_{u,||u||_1=1}\{(x, u)\} = ||x||_\infty$.
Ainsi, la fonction duale g^* est la fonction indicatrice de la boule de rayon λ pour $||.||_\infty$.

Aussi, le problème dual de (LASSO) est :

$$\begin{cases} \sup_\nu -\frac{1}{2}||\nu||_2^2 - \nu^T y \\ \text{tel que } ||X^T\nu||_\infty \leq \lambda \end{cases} \tag{2}$$

De plus, la contrainte :

$$||X^T\nu||_\infty \leq \lambda \Longleftrightarrow \begin{pmatrix} (X^T\nu)_1 \\ (X^T\nu)_2 \\ \vdots \\ (X^T\nu)_d \\ -(X^T\nu)_1 \\ -(X^T\nu)_2 \\ \vdots \\ -(X^T\nu)_d \end{pmatrix} \leq \begin{pmatrix} \lambda \\ \vdots \\ \lambda \end{pmatrix}$$

Aussi, le dual du problème (LASSO) s'écrit sous la forme :

$$\begin{cases} \text{minimize}_v v^T Q v + p^T v \\ \text{tel que } Av \leq b \end{cases} \tag{3}$$

où on a $Q = \dfrac{1}{2} I_n \geq 0$, $p = y$, $A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix}$ et enfin $b = \begin{pmatrix} \lambda \\ \vdots \\ \lambda \end{pmatrix}$

2 Implémentation

Voir le code pour l'implémentation de l'algorithme de résolution demandé.
Pour résoudre le dual algorithmiquement, on a utilisé la méthode barrière algorithmique associé à la résolution du pas de centrage qui permet de trouver un pas d'amélioration à chaque itération.
Le problème du pas de centrage est résolu en utilisant la méthode de Newton sur le problème intermédiaire.

3 Analyse des résultats

Si on revient brièvement au problème initial, on remarque que le problème primal (où l'on a introduit la contrainte d'égalité) est un problème convexe, et qu'il est réalisable. Ainsi, la contrainte de qualification de Slater est vérifié et il y a forte dualité.
Soit v_* solution du dual et z_*, w_* solution du primal on obtient en utilisant les conditions de Karush-Kuhn-Tucker, sachant que le lagrangien L est différentiable par rapport à z, que le gradient partiel de L par rapport à z s'annulle en (v_*,z_*, w_*), or :

$$\nabla_z L(z) = v - z \tag{4}$$

Aussi, on en déduit que $v_* = z_*$. Par ailleurs, l'optimum est un point réalisable du primal donc :

$$z_* = Xw_* - y \tag{5}$$

Et finalement, on obtient l'expression de w_* dépendant v_* ce qui va nous permettre de répondre à la question :

$$v_* = Xw_* - y \tag{6}$$

Tout d'abord on génère une matrice aléatoire X et on prend $\lambda = 10$, puis on affiche le critère de précision ainsi que l'écart à la solution optimale (obtenue) pour chaque itération de l'algortihme barrière.

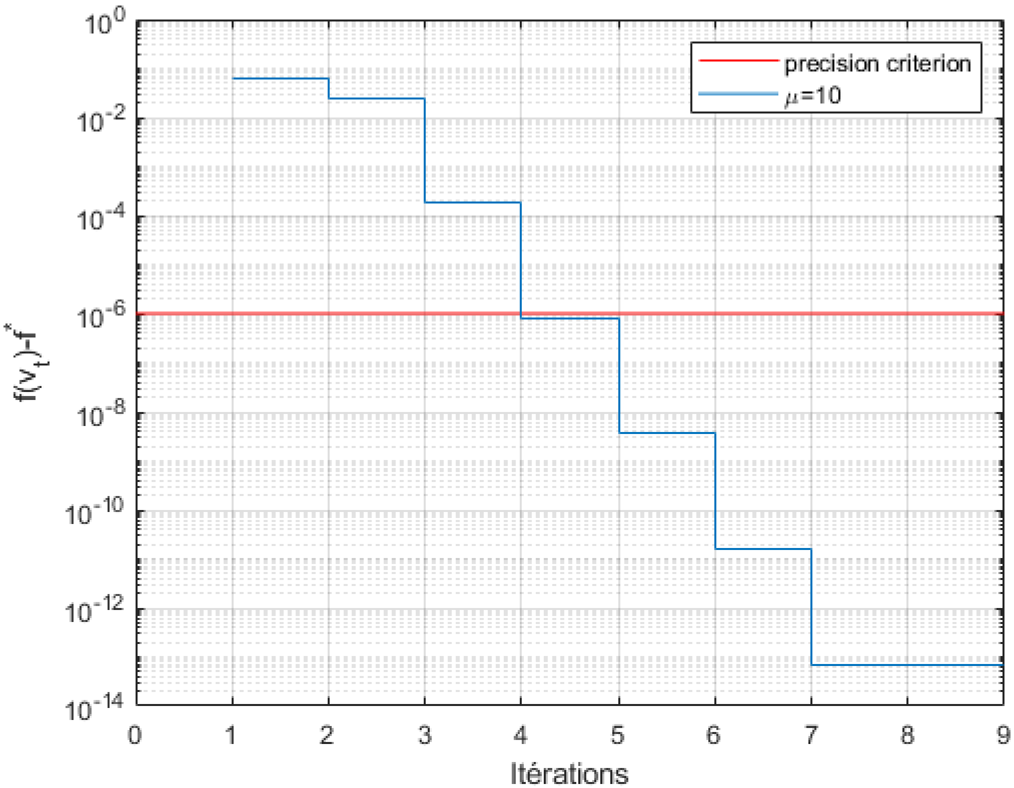

FIGURE 1 – Affichage des résultats de précision pour $\lambda = 10$ (et μ=15)

Aussi, on obtient qu'il y a ici 9 itérations de la méthode barrière logarithmique avant arrêt et convergence. De plus, le critère de précision est bien vérifié.
Désormais, on répète l'expérience pour les valeurs suivantes de $\mu = [2, 7, 15, 50, 100, 200]$ afin d'observer l'influence de ce paramètre.

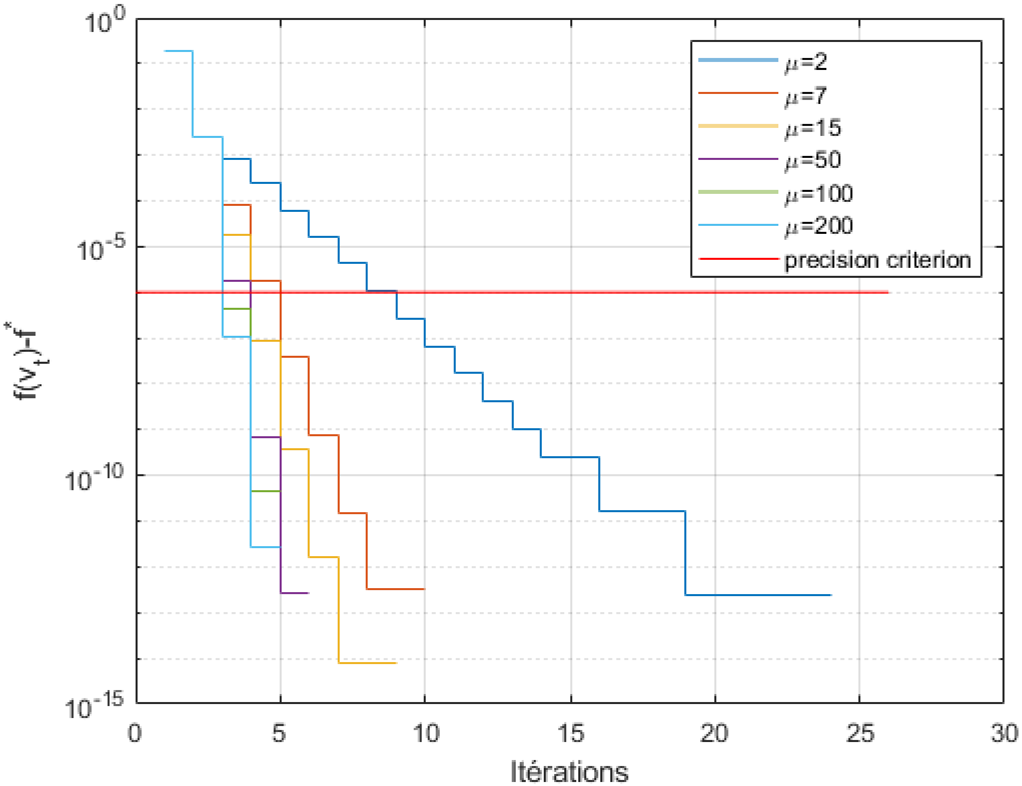

FIGURE 2 – Affichage des résultats de précision pour $\lambda = 10$ (et μ=15)

Globalement, on observe que lorsque μ augmente, il y a de moins en moins d'itérations dans l'algorithme barrière comme l'on pouvait si attendre théoriquement. En effet, le nombre maximal d'itérations est ici de 24 itérations pour $\mu = 2$. Par ailleurs, on sait que lorsque μ augmente le nombre d'itérations pour la résolution du problème intermédiaire par la méthode de Newton augmente. C'est d'ailleurs ce qui a été vérifié (voir ci-après) :

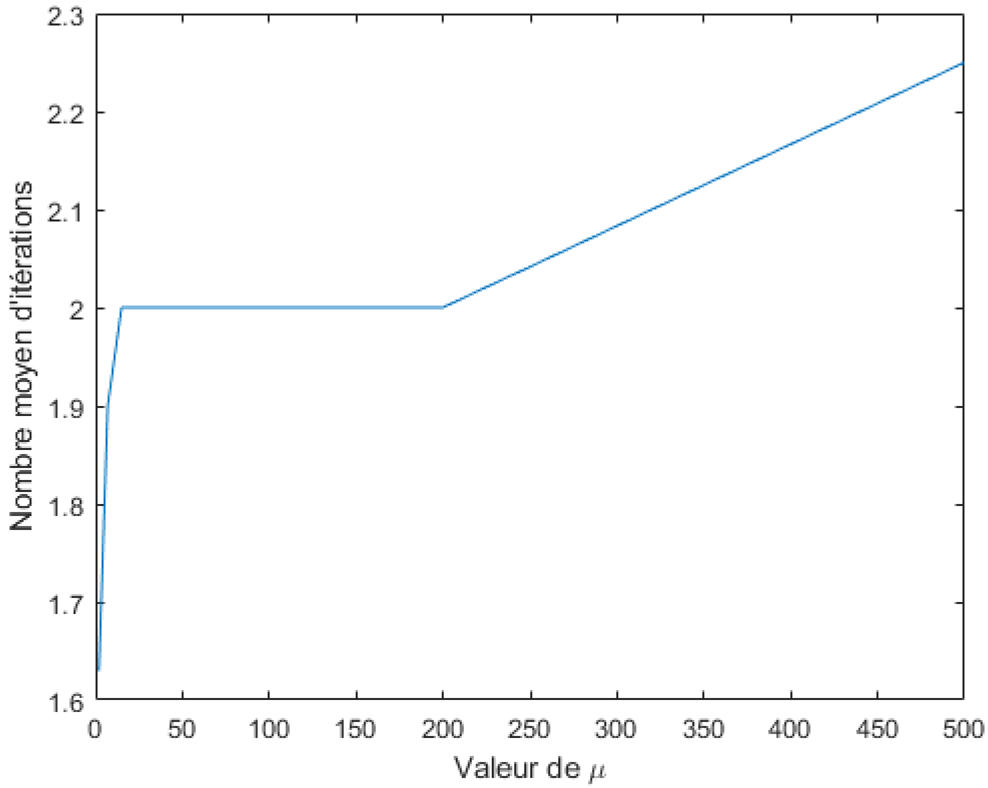

FIGURE 3 – Nombre moyen d'itérations de la méthode de Newton en fonction de μ

Ainsi, il faudra faire un compromis sur la valeur de μ. On peut ici séléctionner $\mu = 15$ ou $\mu = 50$ qui auront ici un nombres d'itérations total faible.
Aussi, il était demandé de vérifier l'impact de μ sur la solution du problème primal w et après vérification μ n'a pas d'impact sur la valeur de w. En effet, on retrouve à chaque fois la même valeur de w.