CONVEX OPTIMIZATION : Homework 3

Guillaume PETIT

25 Novembre 2019

1 Question 1

Le problème primal est

$$\min_{w} \frac{1}{2}||Xw - y||_2^2 + \lambda||w||_1$$

La première étape est de poser $u = Xw - y$ et on obtient

$$\min_{w,u} \frac{1}{2}||u||_2^2 + \lambda||w||_1 \text{ sous contrainte que } u = Xw - y$$

On a bien un problème convexe avec une contrainte d'égalité. On peut donc poser notre Lagrangien et trouver le problème dual.

•**Lagrangien** : Soit $v \in \mathbf{R}^d$,

$$\mathcal{L}(w,u,v) = \frac{1}{2}||u||_2^2 + \lambda||w||_1 + v^T(u - Xw - y)$$
$$\mathcal{L}(w,u,v) = \frac{1}{2}||u||_2^2 + v^T u + \lambda||w||_1 - v^T Xw - v^T y$$

•**fonction duale** :

$$g(v) = \underset{w,u}{Inf}\{\mathcal{L}(w,u,v)\}$$

<u>minimisation par rapport à u</u> :

$$\frac{1}{2}||u||_2^2 + v^T u$$

est une fonction convexe, continue et $\mathcal{C}^1$. Ainsi, son minimum est atteint en :

$$\nabla_u \left(\frac{1}{2}||u||_2^2 + v^T u\right) = 0$$
$$u = -v$$

et le minimum vaut $-\frac{1}{2}||v||_2^2$

<u>minimisation par rapport à w</u> :

$$\underset{w}{\mathrm{Inf}}(\lambda||w||_1 - v^T Xw) = -\underset{w}{\mathrm{Sup}}(v^T Xw - ||w||_1) = \begin{cases} 0 & \text{si } ||X^T v||_\infty \leq \lambda \\ +\infty & \text{sinon.} \end{cases}$$

Finalement, la fonction duale g vaut :

$$g(v) = -\frac{1}{2}v^T v - y^T v \text{ sous contrainte que } ||X^T v||_\infty \leq \lambda$$

On pose

$$Q = \frac{1}{2} I_n,\ A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix},\ p = y,\ b = \lambda \mathbf{1}_{2n}$$

•**fonction duale** : Et on obtient alors le problème dual voulu :

$$\min_v v^T Q v + p^T v \text{ sous contrainte que } Av \preceq b$$

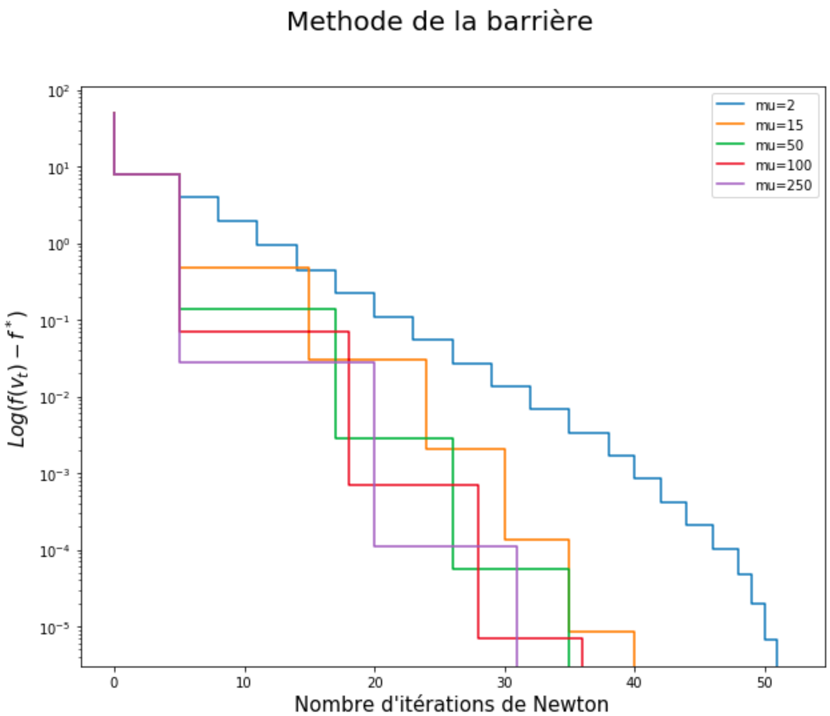
2 Implémentation

On a alors la fonction barrière objective qui est n'est autre que

$$t \left\{ v^T Q v + p^T v \right\} - \sum_{i=1}^{2d} log \left(b_i - a_i^T v \right)$$

Par conséquent, $\nabla \phi(v) = A^T d$ et $\nabla^2 \phi(v) = A^T diag(d)^2 A$ avec d le vecteur tel que $d_i = \dfrac{1}{b_i - a_i^T v}$ et a_i^T sont les lignes de A.

•**Paramètres** : Nous nous plaçons dans le cas $n = 10$ et $d = 50$. Pour l'initialisation, nous avons décidé prendre $t^{(0)} = 1$ pour la méthode de la barrière. On prend $v^{(0)} = 0$, comme cela nous sommes assurés que $Av^{(0)} \preceq \lambda \mathbf{1}_{2n}$ avec $\lambda = 10 > 0$. Pour le backtracking, nous choisissons $\alpha = 0.01$ et $\beta = 0.5$ (valeurs prises dans le livre de Boyd pour un problème similaire).
La contrainte du primal $u = Xw - y$ nous impose une certaine forme pour y à l'initialisation, où nous sélectionnons aléatoirement un w et un u.
On obtient les résultats ci-dessous :



Note : par la faisabilté du primal et l'annulation du gradient, on a $Xw - y + v = 0 \iff w = \left(X^T X\right)^{-1} X^T(y - v)$. Ce qui ressemble de près au w^* des moindres carrés dans une régression linéaire. Dû aux problèmes de dimensions, nous utilisons sur Python la commande "np.linalg.lstsq" pour résoudre cette équation.

La largeur de chaque marche d'escalier (la partie horizontale) est le nombre de pas de Newton requis pour cette outer iteration. La hauteur de chaque contremarche d'escalier (la partie verticale) est exactement égale à (un facteur de) μ, puisque l'écart de dualité est réduit par le facteur μ à la fin de chaque outer iteration. Les graphiques illustrent plusieurs caractéristiques typiques de la méthode de la barrière. Tout d'abord, la méthode fonctionne très bien, avec une convergence à peu près linéaire de l'écart de dualité. Ceci est une conséquence du nombre approximativement constant de pas de Newton nécessaires pour recentrer, pour chaque valeur de μ. Plus μ est petit, plus le nombre total de pas de Newton augmente en raison du grand nombre d'outer iteration nécessaires. Le phénomène opposé se produit quand μ est trop grand.