CONVEX OPTIMIZATION

Homework 3

Thomase FAURÉ

22 Novembre 2019

1 Question 1

Nous avons : le primal

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

On réécrit le problème de manière équivalente en posant : $u = Xw - y$.
On obtient ainsi :

$$\min_{w,u} \frac{1}{2}||u||_2^2 + \lambda||w||_1 \text{ s.t. } u = Xw - y$$

Notre problème est convexe avec une contrainte d'égalité affine. Pour trouver notre problème dual on va chercher à calculer :

$$g(\mu) = \underset{w,u}{Inf}(\frac{1}{2}||u||_2^2 + \mu^T u + \lambda||w||_1 - \mu^T Xw - \mu^T y)$$

<u>minimisation par rapport à u</u> :

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

est une fonction convexe et différentiable (car quadratique). Ainsi, son minimum est atteint en :

$$u = -\mu$$

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

<u>minimisation par rapport à w</u> :

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

Finalement, nous avons :

$$g(\mu) = -\frac{1}{2}\mu^T \mu - y^T \mu \text{ s.t. } ||X^T w||_\infty \leq \lambda$$

On obtient ainsi le problème dual :

$$\min_{\mu} \mu^T Q \mu + p^T \mu \ \text{ s.t. } \ A\mu \preceq b$$

En écrivant :

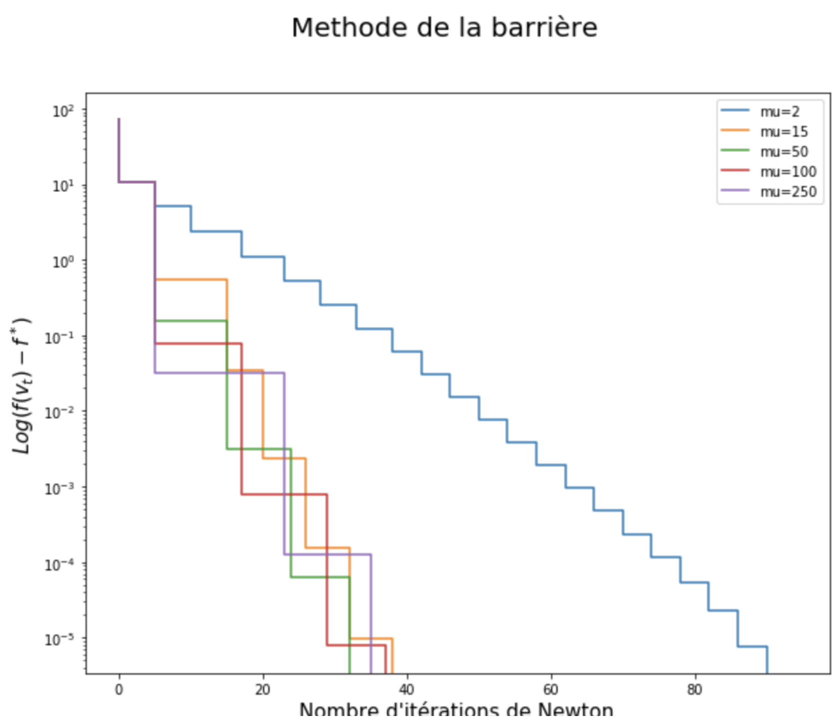
$$Q = \frac{1}{2} I_n$$
$$p = y$$

$$A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix}$$

$$b = \lambda \mathbf{1}_{2n}$$

2 Résultats

Nous obtenons les résultats suivant :



Tout d'abord on peut voir que plus μ est petit plus la partie horizontale de "l'escalier" est petite. Cette partie correspond au nombre d'itérations de l'algorithme de Newton. Ceci s'explique car plus μ est petit moins t évolue et donc notre ancienne itération (le point optimal de Newton à l'itération précédente) est un bon point de départ pour notre nouvelle itération de Newton qui va, par conséquent, converger très vite. Mais à contrario t lui ne va pas beaucoup évoluer (on met à jour t avec $\mu * t$) et la condition $m/t < eps$ va mettre un certain nombre d'itérations avant d'être vérifier et la convergence va prendre un certain temps. Si μ est grand au contraire Newton va mettre du temps à converger mais la seconde condition va être rempli facilement (la partie verticale de notre escalier est proportionnelle a μ. On remarque donc un trade-off sur la valeur de notre paramètre, ici la valeur qui semble la plus adéquate est $\mu = 50$.

3 Annexe

Notre fonction objective de la méthode de la barrière est :

$$t \left\{ v^T Q v + p^T v \right\} - \sum_{i=1}^{2d} log \left(b_i - a_i^T v \right) \quad \text{où } a_i^T \text{ les lignes de A.}$$

Nous trouvons après calcul :

$$\nabla \phi(v) = A^T d$$

$$\nabla^2 \phi(v) = A^T diag(d)^2 A.$$

avec d le vecteur tel que $d_i = \dfrac{1}{b_i - a_i^T v}$

Nous prenons $t^0 = 1$ pour la méthode de la barrière, et $v^{(0)} = 0$, pour assurer $Av^{(0)} \preceq \lambda \mathbf{1}_{2n}$ (en effet $\lambda = 10$). Pour le backtracking line search on prend $\alpha = 0.01$ et $\beta = 0.7$.