DM3 - Optimisation Convexe

Coërchon Colin

2 Décembre 2024

1 Question 1

Afin de transformer le problème en un problème contraint, nous ajoutons une variable auxiliaire $z \in \mathbb{R}^n$:

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

Le Lagrangien de ce problème est alors :

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

Pour un ν fixe, nous avons :

$$\begin{aligned}
g(\nu) = \inf_{w,z} \mathcal{L}(w, z, \nu) &= \inf_{w,z} \left[\frac{1}{2}\|z\|_2^2 + \lambda\|w\|_1 + \nu^T(z - Xw + y)\right] \\
&= \inf_{z} \left[\frac{1}{2}\|z\|_2^2 + \nu^T z\right] + \inf_{w} \big(\lambda\|w\|_1 - \nu^T Xw\big) + \nu^T y \\
&= \inf_{z} \left[\frac{1}{2}\|z\|_2^2 + \nu^T z - \lambda \sup_{w} \left(\frac{1}{\lambda}(X^T\nu)^T w - \|w\|_1\right)\right] + \nu^T y \\
&= \inf_{z} \left[\frac{1}{2}\|z\|_2^2 + \nu^T z - \lambda f^* \left(\frac{1}{\lambda}X^T\nu\right)\right] + \nu^T y
\end{aligned}$$

où f^* est le conjugé de la fonction $\|\cdot\|_1$, que nous avons déjà calculé dans le devoir maison numéro 2 ("HW2").

Ainsi, nous obtenons :

$$g(\nu) = \begin{cases} \inf_{z} \left[\frac{1}{2}\|z\|_2^2 + \nu^T z\right] + \nu^T y & \text{si} \quad \|X^T\nu\|_\infty \leq \lambda \\ -\infty & \text{sinon} \end{cases}$$

La fonction $z \mapsto \frac{1}{2}\|z\|_2^2 + \nu^T z$ est convexe et différentiable, et donc elle atteint un minimum en un point critique. Nous trouvons que $z = -\nu$ est un point critique, ce qui donne :

$$\inf_{z} \left[\frac{1}{2}\|z\|_2^2 + \nu^T z\right] = -\frac{1}{2}\|\nu\|_2^2$$

et enfin :

$$\boxed{g(\nu) = \begin{cases} -\frac{1}{2}\|\nu\|_2^2 + \nu^T y & \text{si} \quad \|X^T\nu\|_\infty \leq \lambda \\ -\infty & \text{sinon} \end{cases}}$$

Le problème dual de LASSO est donc :

$$\max_{\nu} -\frac{1}{2}\|\nu\|_2^2 + \nu^T y \quad \text{s.t.} \quad \|X^T\nu\|_\infty \leq \lambda$$

Puisque $\|X^T\nu\|_\infty \leq \lambda$ équivaut à $\begin{pmatrix} X^T \\ -X^T \end{pmatrix} \nu \leq \lambda \mathbf{1}_{2d}$, le problème se réécrit sous la forme :

$$\max_{\nu} -\frac{1}{2}\nu^T I_n \nu + y^T\nu \quad \text{s.t.} \quad \begin{pmatrix} X^T \\ -X^T \end{pmatrix} \nu \preceq \lambda \mathbf{1}_{2d}$$

Il suffit alors de résoudre :

$$\min_{\nu} \frac{1}{2}\nu^T I_n \nu - y^T\nu \quad \text{s.t.} \quad \begin{pmatrix} X^T \\ -X^T \end{pmatrix} \nu \preceq \lambda \mathbf{1}_{2d}$$

qui devient le problème quadratique :

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

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

$$\boxed{\begin{array}{ll} \min\limits_{\nu} & \nu^T Q\nu + p^T\nu \\ \text{s.c.} & A\nu \preceq \lambda b \end{array}}$$

2 Question 2

Dans tout le code, nous définissons $f_t : \nu \mapsto t(\nu^T Q\nu + p^T) - \sum_{i=1}^d \log(b_i - L_i\nu)$, où L_i est la i-ème ligne de A. Pour l'étape de Newton, il est nécessaire de calculer le gradient et le Hessien de f_t, qui sont donnés par :

$$\nabla f_t(\nu) = t\left((Q + Q^T)\nu + p\right) + \sum_{i=1}^{2d} \frac{1}{b_i - L_i\nu} L_i^T,$$

$$H(f_t)(\nu) = 2tQ + \sum_{i=1}^{2d} \frac{1}{(b_i - L_i\nu)^2} L_i^T L_i.$$

La taille du pas dans la méthode de Newton est obtenue par une recherche de pas avec retour en arrière.
Les paramètres de la méthode peuvent être décomposés comme suit :

- dans la recherche de pas : $\alpha \in \left(0, \frac{1}{2}\right)$ et $\beta \in (0, 1)$,
- dans la méthode de Newton : une tolérance $\varepsilon_{\text{Newton}}$,
- dans la méthode du barrier : un $t^{(0)} > 0$, une tolérance $\varepsilon_{\text{barrier}}$, et un $\mu > 1$.

$\hookrightarrow$ La suite dans le fichier `HW3_Colin_Coerchon.ipynb`