CC2 optimisation numérique 2023-24 : corrigé

Ex1 (11 pts)

1) (2 pts) La fonction HOptim renvoie le résultat approché (après N itérations de l'algorithme d'Uzawa) du problème :

$$\operatorname{Min} \; x \mapsto J(x)$$

sous la contrainte

$$c_E(x) = 0$$

2) (4 pts) l'exemple étudié ici est :

$$\operatorname{Min} J(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle$$

avec

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$

$$b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

sous la contrainte pour $x \in \mathbb{R}^4$:

$$\sum_{i=1}^{4} x_i = 1$$

Sur cet exemple, la fonction J est coercive et strictement convexe

2

car la matrice A est définie positive.

($^t x A x = \sum_{i=1}^{n+1} (x_i - x_{i-1})^2$, avec la convention $x_{n+1} = x_0 = 0$, et donc $^t x A x > 0$ si $x \neq 0$.) (*)

L'ensemble $\Omega = \{x \in \mathbb{R}^4, \sum x_i = 1\}$ est fermé et convexe. Le problème admet donc une unique solution.

Pour la trouver, on résout KKT :

$$\begin{cases} Ax - b + \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = 0 \\ \sum_{i=1}^{4} x_i = 1 \end{cases}$$

((*) on peut aussi utiliser Sylvester)

soit :

$$\begin{cases} 2x_1 - x_2 + \lambda - 1 = 0 \\ -x_1 + 2x_2 - x_3 + \lambda - 1 = 0 \\ -x_2 + 2x_3 - x_4 + \lambda - 1 = 0 \\ -x_3 + 2x_4 + \lambda - 1 = 0 \\ x_1 + x_2 + x_3 + x_4 = 1 \end{cases}$$

Ceci implique (avec $\lambda' = \lambda - 1$)

$x_1 + x_4 + 4\lambda' = 0$ $(L_1 + L_2 + L_3 + L_4)$

mais aussi :

$\underbrace{x_2 + x_3}_{1 - (x_1 + x_4)} - (x_1 + x_4) + 2\lambda' = 0$ $(L_2 + L_3)$

soit

$$8\lambda' + 1 + 2\lambda' = 0$$

et $\lambda' = -\frac{1}{10}$ puis $x_1 + x_4 = \frac{4}{10}$

et $x_2 + x_3 = 6/10$

$2L_1 + L_2$ donne :

$3x_2 - 2x_3 = 3/10$, soit

$x_2 = x_3 = 3/10$ puis

$x_1 = x_4 = 2/10$ (et $\lambda = \frac{9}{10}$)

Il s'agit de l'unique solution de KKT et donc forcément du problème de minimisation //

3) (2pts) La ligne

`X = X - alpha1*(nablaJ(x) + lamb*nablaCE(x))` consiste à effectuer un pas de gradient (pas constant égal à alpha1) pour minimiser la fonction $x \mapsto \mathcal{L}(x, \lambda)$ (Lagrangien du problème)

La ligne

`lamb = lamb + alpha2*CE(x)` consiste à effectuer un pas de gradient (pas égal à alpha2) pour maximiser la fonction $\lambda \mapsto \mathcal{L}(x, \lambda)$, ceci afin de rechercher un point selle de $\mathcal{L}$.

4) (3 pts) On effectue une itération de l'algorithme :

$X_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $\lambda = 3$

$$X = 0 - 0.1 * \left(A*0 - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + 3 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right)$$

soit

$$X = -\begin{pmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{pmatrix}$$

et $\lambda = 3 + 0.1 * \left(4*(-0.2) - 1 \right)$
$= 3 - 0{,}18 = 2{,}82$

Ex2 (16 pts)

3

1) (2 pts) L'ensemble $S = \{x \in \mathbb{R}^n / Dx = d\}$ est un ensemble convexe et fermé. La fonction J étant continue, coercive et strictement convexe, le problème admet bien une unique solution.

2) (1 pt) Attention : il manque une hypothèse du type :

$$J(y) \geq J(x) + \langle \nabla J(x), y - x \rangle + \underset{>0}{\mu} \|y - x\|^2$$

(J fortement convexe) pour conclure

Dans ce cas, comme

$$\begin{cases} H_x \mathcal{L} = H_x J \text{ et} \\ \nabla_x \mathcal{L} = \nabla_x J + {}^t D \lambda \end{cases}$$

on garde les mêmes hypothèses ($\mathcal{L}(\cdot, \lambda)$ fortement convexe et coercive) et donc les mêmes conclusions (existence et unicité de x^*_λ).

3) (3 pts) La fonction $\lambda \longmapsto \mathcal{L}(x, \lambda)$ est affine et donc concave. Or si f, g sont 2 fonctions concaves, alors $h = \min(f, g)$ est également concave (en effet,

$$f(\alpha u + (1-\alpha)v) \geq \alpha f(u) + (1-\alpha) f(v)$$
$$\geq \alpha h(u) + (1-\alpha) h(v)$$

De même,

$$g(\alpha u + (1-\alpha)v) \geq \alpha h(u) + (1-\alpha) h(v)$$

D'où $h(\alpha u + (1-\alpha)v) \geq \alpha h(u) + (1-\alpha) h(v)$ et h concave). Il en va de même par $\lambda \longmapsto \operatorname{Min}_x (\mathcal{L}(x, \lambda))$

4) (7 pts)

On a dans ce cas

/5

$$\mathcal{L}(x,\lambda) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + \langle {}^tD\lambda, x\rangle$$

et on a alors

$$Ax^*_\lambda = b - {}^tD\lambda$$

soit

$$x^*_\lambda = A^{-1}(b - {}^tD\lambda)$$

2pts

On a alors

$$H(\lambda) = \frac{1}{2}\langle Ax^*_\lambda, x^*_\lambda\rangle - \langle b, x^*_\lambda\rangle + \langle \lambda, Dx^*_\lambda - d\rangle$$

puis

$$H(\lambda) = \frac{1}{2}\langle A^{-1}(b - {}^tD\lambda), b - {}^tD\lambda\rangle - \langle b, A^{-1}(b - {}^tD\lambda)\rangle + \langle \lambda, DA^{-1}(b - {}^tD\lambda) - d\rangle$$

En prenant, le gradient en λ, on trouve :

$$\nabla H(\lambda) = DA^{-1}\,{}^tD\lambda - \frac{1}{2}\cancel{DA^{-1}b} - \frac{1}{2}\,{}^t(A^{-1}\,{}^tD)b + {}^t(A^{-1}\,{}^tD)b + DA^{-1}b - d - 2DA^{-1}\,{}^tD\lambda$$

$DA^{-1}b$

$DA^{-1}b$

$$= -DA^{-1}\,{}^tD\lambda + DA^{-1}b - d$$

Or

$$Dx^*_\lambda - d = DA^{-1}b - DA^{-1}\,{}^tD\lambda - d,$$

soit : $\nabla H(\lambda) = Dx^*_\lambda - d$

3pts

H est de la forme

$$H(\lambda) = \frac{1}{2}\langle \lambda, B\lambda\rangle - \langle f, \lambda\rangle$$

avec $B = -DA^{-1}\,{}^tD$ qui est définie négative. Son unique maximum est donc atteint en λ_0 tq $-DA^{-1}\,{}^tD\lambda_0 = DA^{-1}b - d$

On a alors :

$$\begin{cases} Dx^*_{\lambda_0} = d \text{ (voir précédemment) et} \\ Ax^*_{\lambda_0} = b - {}^tD\lambda_0 \text{ (idem)} \end{cases}$$

Cela correspond exactement au système KKT vérifié par (x^*, λ^*)

(2 pts)

6

5) (3 pts) On a de même

$$H(\lambda) = J(x^*_\lambda) + \langle \lambda, Dx^*_\lambda - d\rangle$$

En supposant $\lambda \mapsto x^*_\lambda$ différentiable la règle de la chaîne donne :

$$\underbrace{\nabla H(\lambda)}_{\in \mathbb{R}^m} = \underbrace{\frac{\partial x^*_\lambda}{\partial \lambda}}_{\in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)} \cdot \nabla J(x^*_\lambda) + (Dx^*_\lambda - d)$$

$$+ \underbrace{\left(\frac{\partial x^*_\lambda}{\partial \lambda}\right)}_{\in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)} \underbrace{{}^tD\lambda}_{\in \mathbb{R}^n}$$

soit $\nabla H(\lambda) = \frac{\partial x^*_\lambda}{\partial \lambda}\left[\nabla J(x^*_\lambda) + {}^tD\lambda\right]$

$$+ Dx^* - d$$

$$= 0 + Dx^* - d$$

car x^*_λ point critique de

$$x \longmapsto J(x) + {}^t\lambda D x$$

et donc $\nabla J(x_\lambda) + {}^t D\lambda = 0$