

**Contrôle continu - Corrigé**

Calculatrice non autorisée – Documents autorisés

Ce sujet comporte 3 exercices sur 2 pages

Durée : 2 heures

**Exercice 1.**

Soit  $P_1 := \{A_1 + sv_1 + tw_1, s, t \in \mathbb{R}\}$  et  $P_2 := \{A_2 + sv_2 + tw_2, s, t \in \mathbb{R}\}$  deux plans de  $\mathbb{R}^n$  définies à l'aide des points  $A_i$  et des vecteurs directeurs non nuls  $v_i$  et  $w_i$ , avec  $w_i$  non colinéaire à  $v_i$ . On cherche les points  $M_1 \in P_1$  et  $M_2 \in P_2$  minimisant la distance euclidienne de  $M_1$  à  $M_2$ .

1. Formaliser le problème ci-dessus comme un problème de minimisation sans contraintes.
2. Sans résolution mathématique, donner une caractéristique intuitive de la solution dans le cas où  $n = 3$ .

**Solution :**

1. On cherche à minimiser  $|M_1 M_2|$  pour  $M_1 \in P_1$  et  $M_2 \in P_2$ . On cherche  $s_1, t_1, s_2$  et  $t_2$  qui minimisent  $f(s_1, t_1, s_2, t_2) = \|A_1 + s_1 v_1 + t_1 w_1 - (A_2 + s_2 v_2 + t_2 w_2)\|_2^2$ .
2. Si les plans sont parallèles, on a une infinité de solutions. On a un  $M_2$  pour chaque  $M_1$  dans  $P_1$ .  
Si les plans ne sont pas parallèles, alors l'ensemble des  $M_1$  acceptables décrit une droite et  $M_2 = M_1$  pour chaque  $M_1$  acceptable.

**Exercice 2.**

Soit  $J$  la fonctionnelle quadratique définie sur  $\mathbb{R}^2$  par  $J(x, y) = ax^2 + by^2$  avec  $a, b > 0$ .

1. (a) Quel est le minimum de  $J$ ? En quel point est-il atteint? Pour  $(\bar{x}, \bar{y})$  donné dans  $\mathbb{R}^2$ , quelle est en  $(\bar{x}, \bar{y})$  la direction de descente **globalement** optimale (c'est à dire la direction permettant de rencontrer directement le point minimisant  $J$ )?  
(b) Dans le cas où la méthode est initialisée à  $(x_0, y_0) \neq (0, 0)$ , établir une condition nécessaire et suffisante sur  $(a, b, x_0, y_0)$  pour que la méthode du gradient à pas optimal converge en une seule itération.
2. (a) Etablir une condition suffisante de la convergence de la méthode du gradient à pas fixe, portant sur le pas de la méthode.  
(b) La méthode du gradient à pas fixe sans contrainte est un cas particulier de la méthode du gradient à pas variable avec contrainte où la contrainte est l'appartenance à l'espace de Hilbert lui-même. En utilisant un résultat intermédiaire d'une démonstration du cours, donner alors une estimation de la vitesse de convergence de la méthode du gradient à pas fixe en fonction du pas  $\rho$  choisi. Pour quelle valeur de  $\rho$  la vitesse estimée est-elle optimale?

**Solution :**

1. (a)  $J$  est minimale en  $(0, 0)$  où elle vaut 0. Depuis  $(\bar{x}, \bar{y})$ , la direction globalement optimale est donc celle qui va nous diriger vers  $(0, 0)$ , à savoir celle de  $-(\bar{x}, \bar{y})$ .

- (b) D'après la question précédente, la convergence a lieu en une seule itération ssi  $-J'(x_0, y_0)$  et  $-(x_0, y_0)$  sont proportionnels. Or  $J'(x_0, y_0) = (2ax_0, 2by_0)$ . Ceci se produit donc ssi  $((a = b) \text{ ou } (x_0 y_0 = 0))$ .
2. (a) La constante d'ellipticité est  $\alpha = 2 \min(a, b)$  et celle de continuité (ou de Lipschitz) est  $M = 2 \max(a, b)$ . Ainsi, la méthode du gradient à pas fixe converge si le pas  $\rho$  vérifie  $0 < \rho < \frac{\min(a, b)}{(\max(a, b))^2}$ .
- (b) Dem du thm 18 : La convergence est géométrique de raison  $r$  telle que  $r^2 = 1 - \alpha\rho(2 - M\rho)$ , polynôme de degré 2 convexe en  $\rho$ . Il suffit de minimiser  $r^2$  en tâchant de garder  $\rho$  dans  $]0, \frac{\min(a, b)}{(\max(a, b))^2}[$  :  $\rho_{optimal} = \min(\frac{1}{2 \max(a, b)}, \frac{\min(a, b)}{(\max(a, b))^2})$ .

### Exercice 3.

Soit  $C \in \mathbb{R}^{m, n}$  une matrice de rang  $m$  avec  $m < n$ . Soient deux vecteurs  $b \in \mathbb{R}^n$  et  $v \in \mathbb{R}^m$ . On se donne enfin une matrice  $A \in \mathbb{R}^{n, n}$ , symétrique définie positive.

On notera  $(x, y) = \sum_i x_i y_i$  le produit scalaire usuel dans  $\mathbb{R}^n$  ou  $\mathbb{R}^m$ , et  $\| \cdot \|$  la norme associée.

On s'intéresse dans un premier temps au problème avec contrainte

$$(P) \quad \begin{cases} \text{Trouver } x \in K \text{ tel que } , \forall y \in K, \\ J(x) \leq J(y). \end{cases}$$

où

$$K = \{y \in \mathbb{R}^n ; Cy = v\} , \\ J(x) = \frac{1}{2}(Ax, x) - (b, x) .$$

N.B. : Par hypothèse, on sait notamment que  $C$  est surjective et que  $\text{Ker } C^T = \{0\}$ .

- (a) Montrer que le problème  $(P)$  admet une solution et une seule, notée  $x$ .
- (b) Montrer que  $\forall y \in \text{Ker } C, (J'(x), y) \geq 0$ . En déduire une relation liant  $b - Ax$  et  $\text{Ker } C$  puis qu'il existe un unique  $\lambda \in \mathbb{R}^m$  tel que  $b - Ax = C^T \lambda$ .
- Considérons aussi le problème sans contrainte

$$(P_C) \quad \begin{cases} \text{Trouver } x_C \in \mathbb{R}^n \text{ tel que } , \forall y \in \mathbb{R}^n, \\ J_C(x_C) \leq J_C(y). \end{cases}$$

où

$$J_C(y) = J(y) + \| Cy - v \|^2 .$$

- Montrer que le problème  $(P_C)$  admet une solution et une seule, notée  $x_C$ .
- En utilisant les conditions d'optimalité, montrer que  $x_C$  est l'unique solution d'un système linéaire que l'on précisera.
- Montrer que l'on a (en désignant par  $J'_C$  le gradient de  $J_C$ )

$$\forall y \in \mathbb{R}^n, (J'_C(x), y) = -(\lambda, Cy).$$

- Notons désormais  $y_C = x_C - x$ .

- Que vaut le hessien  $J''_C$  de  $J_C$  ?
- Montrer que l'on a  $\| Cy_C \| \leq \| \lambda \|$  et  $(Ay_C, y_C) \leq 2 \| \lambda \|^2$ .  
Indication : On pourra calculer (par Taylor)  $J_C(x + y_C) - J_C(x)$ .

- Proposer, sans justification mathématique, une fonctionnelle  $J_\varepsilon$  et un problème  $(P_\varepsilon)$  inspirés de  $J_C$  et  $(P_C)$  tels que la solution  $x_\varepsilon$  de  $(P_\varepsilon)$  soit incitée à tendre vers  $x$  la solution de  $(P)$ . Quel avantage présente  $(P_\varepsilon)$  par rapport à  $(P)$  ?

**Solution :**

1. (a) Théorème 11 du cours. Bien sûr, on se doit de vérifier que toutes les conditions du théorème sont remplies avant son application. Pour l'ellipticité, on utilise le fait que  $A$  est définie positive et donc que  $(Ax, x) \geq \lambda_1 \|x\|^2$  où  $\lambda_1$  est la plus petite valeur propre de  $A$ .
- (b) Condition d'optimalité :  $(J'(x), z - x) \geq 0$  pour tout  $z \in K$ .  
 $y \in \text{Ker } C$  et  $x \in K$  impliquent que  $x - y$  et  $x + y$  sont dans  $K$ , d'où  
 $(J'(x), y) = 0 \ \forall y \in \text{Ker } C$ . Ainsi  $J'(x) \perp \text{Ker } C$ .  
D'autre part  $J'(x) = Ax - b$  et  $(\text{Ker } C)^\perp = \text{Im } C^T$ . D'où l'existence de  $\lambda$  tel que  
 $b - Ax = C^T \lambda$ . L'unicité provient de  $\text{Ker } C^T = \{0\}$ .
2. (a)  $J_C$  hérite des bonnes propriétés de  $J$  (Bien sûr, cela doit être développé).
- (b) Ensemble des contraintes : l'espace vectoriel  $\mathbb{R}^n$  tout entier. La condition d'optimalité devient donc :  $J'_C(x_C) = 0$  (à remonter rapidement en repartant de la condition d'optimalité générale), cad  $(A + 2C^T C)x_C = b + 2C^T v$ .  
Le calcul de  $J'_C(y)$  se fait très simplement en écrivant la forme quadratique :  
 $J_C(y) = \frac{1}{2}((A + 2C^T C)y, y) - (b + 2C^T v, y) + (d, d)$ .
- (c)  $J'_C(x) = (A + 2C^T C)x - b - 2C^T v$ . Or  $Ax - b = -C^T \lambda$  et  $Cx = v$  donc  $J'_C(x) = -C^T \lambda$ .
3. (a)  $J''_C(y) = A + 2C^T C$  qui est constant par rapport à  $y$ .
- (b) Puisque  $J''_C$  est une application constante, la formule de Taylor s'écrit :

$$J_C(x + y_C) - J_C(x) = (J'_C(x), y_C) + \frac{1}{2}(J''_C(x)y_C, y_C)$$

D'autre part, par définition de  $y_C$  et  $x_C$ ,  $J_C(x + y_C) - J_C(x) \leq 0$ .

Il vient  $\frac{1}{2}(Ay_C, y_C) + (C^T C y_C, y_C) \leq (\lambda, C y_C)$ . Or  $(Ay_C, y_C)$  et  $(C^T C y_C, y_C)$  sont tous deux positifs (en effet,  $(C^T C y_C, y_C) = (C y_C, C y_C) = \|C y_C\|^2$ ).

Ainsi, par l'inégalité de Cauchy-Schwarz :

$$\|C y_C\|^2 \leq \|\lambda\| \|C y_C\| \quad \text{et} \quad \frac{1}{2}(Ay_C, y_C) \leq \|\lambda\| \|C y_C\|$$

d'où le résultat.

4.

$$(P_\varepsilon) \quad \begin{cases} \text{Trouver } x_\varepsilon \in \mathbb{R}^n \text{ tel que } , \forall y \in \mathbb{R}^n, \\ J_\varepsilon(x_\varepsilon) \leq J_\varepsilon(y). \end{cases}$$

où

$$J_\varepsilon(y) = J(y) + \frac{1}{\varepsilon} \|C y - v\|^2.$$

Quand  $\varepsilon$  tend vers 0,  $\frac{1}{\varepsilon}$  devient de plus en plus grand. Pour minimiser  $J_\varepsilon$ , intuitivement,  $x_\varepsilon$  devra vérifier que  $\|C x_\varepsilon - v\|$  soit de plus en plus proche de 0. On peut d'autre part, montrer mathématiquement que  $x_\varepsilon$  tend vers  $x$  quand  $\varepsilon$  tend vers 0. Ceci est d'ailleurs l'objet d'un problème dont est issu cet exercice. Le problème  $(P_\varepsilon)$  présente l'avantage d'être sans contrainte.

*Fin du corrigé*