**M1 IM/MPA : TD Optimisation et Éléments Finis**

**Feuille 3 : Algorithmes de descente de gradient (suite).**

**Exercice 1**
**Cas d'une fonctionnelle quadratique dans la méthode de gradient à pas optimal.**
Soit $d \in \mathbb{N}^*$. On considère l'algorithme de gradient à pas optimal appliqué à une fonctionnelle quadratique

$$\mathcal{J} : \mathbb{R}^d \to \mathbb{R}, x \mapsto \frac{1}{2}(Ax, x) - (b, x),$$

avec $A$ une matrice symétrique définie positive et $b$ un vecteur de $\mathbb{R}^d$. On note $x^* \in \mathbb{R}^d$, la solution du système $Ax = b$.

On écrit l'algorithme comme suit :
— *Initialisation :* $x_0 \in \mathbb{R}^d$,
— *Itération* $k \in \mathbb{N}$ *:* $x_{k+1} = x_k + \alpha_k r_k$, où $r_k = b - Ax_k$ et $\alpha_k$ est le pas de descente qui minimise la fonction $\rho \mapsto \mathcal{J}(x_k - \rho \nabla \mathcal{J}(x_k))$.

On admettra le résultat suivant :

**Théorème (Inégalité de Kantorovitch).** *Pour* $x \in \mathbb{R}^d \setminus \{0\}$,

$$1 \leq \frac{(Ax, x)(A^{-1}x, x)}{\|x\|^4} \leq \frac{(\lambda_d + \lambda_1)^2}{4\lambda_d \lambda_1},$$

*où* $\lambda_1$ *la plus petite valeur propre de* $A$ *et* $\lambda_d$ *la plus grande valeur propre de* $A$.

On sait que l'on peut montrer (cf. TD précédents) qu'il existe un unique point de minimum de $\mathcal{J}$ sur $\mathbb{R}^d$.

1(supplément). Montrer que le problème de minimisation se ramène à étudier le minimum de la fonction $\mathcal{E} : \mathbb{R}^d \to \mathbb{R}$, $x \mapsto (A(x - x^*), x - x^*)$ (on pourra développer l'expression de $\mathcal{E}$, ).

1. Donner dans ce cas l'expression du pas optimal $\alpha_k$ à chaque itération $k \in \mathbb{N}$.

On cherche maintenant à montrer que la suite $(x_k)_{k \in \mathbb{N}}$ converge vers $x$ avec un facteur de convergence $\dfrac{\kappa(A) - 1}{\kappa(A) + 1}$ avec $\kappa(A) = \dfrac{\lambda_d}{\lambda_1}$.

2. Montrer que pour tout $k \in \mathbb{N}$,

$$\mathcal{E}(x_{k+1}) = \mathcal{E}(x_k) \left( 1 - \frac{\|r_k\|^4}{(Ar_k, r_k)(A^{-1}r_k, r_k)} \right).$$

3. En déduire que pour tout $k \in \mathbb{N}$, $\exists C_k > 0$, telle que

$$\mathcal{E}(x_k) \leq C_k \mathcal{E}(x_0).$$

On donnera l'expression de $C_k$ en fonction de $\kappa(A)$ et $k$.

4. Conclure.

## Solutions

**Solution exercice 1**

1(supplément). La fonction $\mathcal{J}$ est une fonctionnelle quadratique, elle est donc en particulier $\mathcal{C}^2$ et on a pour tout $x \in \mathbb{R}^d$, $\nabla \mathcal{J}(x) = Ax - b$. De plus, pour tout $x \in \mathbb{R}^d$, on a $Hess(\mathcal{J})(x) = A$. On sait que $A$ est symétrique définie positive, on a donc en particulier $\langle Ax, x\rangle \geq \lambda_1 \|x\|^2$, $\forall x \in \mathbb{R}^d$ et donc $\langle Hess(\mathcal{J})(x)y, y\rangle \geq \lambda_1 \|y\|^2$, $\forall y \in \mathbb{R}^d$. Par caractérisation de l'$\alpha$-convexité, on en déduit que $\mathcal{J}$ est $\lambda_1$-convexe avec $\lambda_1 > 0$. Et donc tout point critique est un point de minimum de $\mathcal{J}$.

1. On a pour tout $x \in \mathbb{R}^d$,

$$\begin{aligned}
\mathcal{E}(x) &= \langle A(x - x^*), x - x^*\rangle, && (1)\\
&= \langle Ax, x\rangle + \langle Ax^*, x^*\rangle - \langle Ax, x^*\rangle - \langle Ax^*, x\rangle, && (2)\\
&= \langle Ax, x\rangle + \langle Ax^*, x^*\rangle - 2\langle Ax^*, x\rangle, && (3)\\
&= \langle Ax, x\rangle + \langle Ax^*, x^*\rangle - 2\langle b, x\rangle, && (4)\\
&= 2\mathcal{J}(x) + \langle Ax^*, x^*\rangle, && (5)\\
& && (6)
\end{aligned}$$

Donc minimiser $\mathcal{E}$ revient à minimiser $\mathcal{J}$ puisqu'elles ne diffèrent que par une constante et que $2 > 0$.

2. On remarque que $\mathcal{E} \geq 0$ et que $\mathcal{E}(x^*) = 0$, $x^*$ est donc un point de minimum de $\mathcal{E}$ et donc $\mathcal{J}$ sur $\mathbb{R}^d$. Mais $\mathcal{J}$ est $\lambda_1$-convexe avec $\lambda_1 > 0$, on sait donc que le problème de minimisation sur $\mathbb{R}^d$ admet un unique point de minimum : c'est $x^*$.

3. Soit $k \in \mathbb{N}$. On écrit qu'à chaque itération, on minimise la fonction

$$\varphi_k : \rho \mapsto \mathcal{J}(x_k - \rho \nabla \mathcal{J}(x_k)),$$

sur tout $\mathbb{R}$. C'est un problème de minimisation sans contraintes et on sait que $\varphi_k$ est $\mathcal{C}^1$, puisque $\mathcal{J}$ est $\mathcal{C}^1$. Si $\alpha_k \in \mathbb{R}$ est un point de minimum de $\varphi$ sur $\mathbb{R}$ (qui est un ouvert), alors $\alpha_k$ est point critique de $\varphi_k$ et vérifie

$$\varphi_k'(\alpha_k) = 0. \tag{7}$$

Cela donne

$$\langle \nabla \mathcal{J}(x_{k+1}), \nabla \mathcal{J}(x_k)\rangle = 0. \tag{8}$$

Autrement dit

$$\langle A(x_k - \alpha_k(Ax_k - b)) - b, Ax_k - b\rangle = 0. \tag{9}$$

Cela donne

$$-\alpha_k \langle A(Ax_k - b), Ax_k - b\rangle = -\langle Ax_k - b, Ax_k - b\rangle. \tag{10}$$

On trouve finalement si $Ax_k - b \neq 0$,

$$\alpha_k = \frac{\langle Ax_k - b, Ax_k - b\rangle}{\langle A(Ax_k - b), Ax_k - b\rangle}. \tag{11}$$

Ou encore

$$\alpha_k = \frac{\langle r_k, r_k\rangle}{\langle Ar_k, r_k\rangle}. \tag{12}$$

4. Montrons que pour tout $k \in \mathbb{N}$,

$$\mathcal{E}(x_{k+1}) = \mathcal{E}(x_k)\left(1 - \frac{\|r_k\|^4}{(Ar_k, r_k)(A^{-1}r_k, r_k)}\right).$$

On a pour $k \in \mathbb{N}$,

$$\begin{aligned}
\mathcal{E}(x_{k+1}) &= \langle Ax_{k+1} - b, x_{k+1} - x^*\rangle, && (13)\\
&= \langle A(x_k - \alpha_k r_k) - b, x_k - \alpha_k r_k - x^*\rangle, && (14)\\
&= \langle Ax_k - b, x_k - \alpha_k r_k - x^*\rangle - \alpha_k\langle Ar_k, x_k - \alpha_k r_k - x^*\rangle, && (15)\\
&= \langle Ax_k - b, x_k - x^*\rangle - \alpha_k\langle Ax_k - b, r_k\rangle - \alpha_k\langle Ar_k, x_k - x^*\rangle + \alpha_k^2\langle Ar_k, r_k\rangle, && (16)\\
&= \mathcal{E}(x_k) - \alpha_k\langle Ax_k - b, r_k\rangle - \alpha_k\langle Ar_k, x_k - x^*\rangle + \alpha_k^2\langle Ar_k, r_k\rangle, && (17)\\
&= \mathcal{E}(x_k) - \frac{\|r_k\|^2}{\langle Ar_k, r_k\rangle}\langle Ax_k - b, r_k\rangle - \frac{\|r_k\|^2}{\langle Ar_k, r_k\rangle}\langle Ar_k, x_k - x^*\rangle + \frac{\|r_k\|^4}{\langle Ar_k, r_k\rangle^2}\langle Ar_k, r_k\rangle, && (18)\\
&= \mathcal{E}(x_k) - \frac{\|r_k\|^4}{\langle Ar_k, r_k\rangle} - \frac{\|r_k\|^2}{\langle Ar_k, r_k\rangle}\langle r_k, A(x_k - x^*)\rangle + \frac{\|r_k\|^4}{\langle Ar_k, r_k\rangle^2}\langle Ar_k, r_k\rangle, && (19)\\
&= \mathcal{E}(x_k) - \frac{\|r_k\|^4}{\langle Ar_k, r_k\rangle} - \frac{\|r_k\|^4}{\langle Ar_k, r_k\rangle} + \frac{\|r_k\|^4}{\langle Ar_k, r_k\rangle}, && (20)\\
&= \mathcal{E}(x_k) - \frac{\|r_k\|^4}{\langle Ar_k, r_k\rangle}, && (21)\\
& && (22)
\end{aligned}$$

Or

$$\begin{aligned}
\mathcal{E}(x_k) &= \langle A(x_k - x^*), x_k - x^*\rangle, && (23)\\
&= \langle r_k, A^{-1}r_k\rangle, && (24)\\
& && (25)
\end{aligned}$$

puisque $r_k = A(x_k - x^*)$. Finalement,

$$\mathcal{E}(x_{k+1}) = \mathcal{E}(x_k)\left(1 - \frac{\|r_k\|^4}{\langle Ar_k, r_k\rangle\langle r_k, A^{-1}r_k\rangle}\right). \tag{26}$$

5. L'inégalité de Kantorovitch donne pour tout $k \in \mathbb{N}$,

$$\mathcal{E}(x_{k+1}) \leq \mathcal{E}(x_k)\left(1 - \frac{4\lambda_d\lambda_1}{(\lambda_d + \lambda_1)^2}\right). \tag{27}$$

Or si $\lambda_1 \neq \lambda d$, comme $0 < (\lambda_1 - \lambda_d)^2 = (\lambda_1 + \lambda_d)^2 - 4\lambda_1\lambda_d$, on en déduit que $C = 1 - \frac{4\lambda_d\lambda_1}{(\lambda_d+\lambda_1)^2} < 1$, et donc

$$\mathcal{E}(x_{k+1}) \leq C\mathcal{E}(x_k), \tag{28}$$

ce qui donne le résultat par une récurrence immédiate avec $C_k = C^k = \left(\frac{1-\kappa(A)}{1+\kappa(A)}\right)^{2k}$.

6. On a donc montré que $\mathcal{E}(x_k) \to 0$, lorsque $k \to +\infty$. Mais on a également

$$\langle A(x_k - x^*), x_k - x^* \rangle \geq \lambda_1 \|x_k - x^*\|^2, \tag{29}$$

et

$$\langle A(x_k - x^*), x_k - x^* \rangle \leq \lambda_d \|x_k - x^*\|^2, \tag{30}$$

On en déduit donc que

$$\|x_k - x^*\|^2 \leq \lambda_1^{-1} C^k \mathcal{E}(x_0) \tag{31}$$

Et donc la suite converge vers $x^*$ avec un facteur $\frac{1-\kappa(A)}{1+\kappa(A)}$.