Optimisation & Analyse convexe
Séance 3 : Algorithmes pour l'optimisation sans contrainte

Exercice 1 (Algorithme du gradient conjugué).
Dans tout cet exercice, A désigne une matrice symétrique définie positive $N \times N$, de valeurs propres $0 < \lambda_1 \leq \ldots \leq \lambda_N$, et $b \in \mathbb{R}^N$. On note $\|v\|_A^2 = (Av, v)$ pour $v \in \mathbb{R}^N$. Considérons la fonction :

$$J(v) = \frac{1}{2}(Av, v) - (b, v) \quad v \in \mathbb{R}^N,$$

et notons u le point de minimum de J sur $\mathbb{R}^N$. On considère l'algorithme de descente à pas optimal :

$$u^{n+1} = u^n + \rho^n d^n \tag{1}$$

$$\text{avec} \quad d^0 = -\nabla J(u^0), \quad d^n = -\nabla J(u^n) + \frac{\|\nabla J(u^n)\|^2}{\|\nabla J(u^{n-1})\|^2} d^{n-1}, \quad \rho^n = \frac{(-\nabla J(u^n), d^n)}{(Ad^n, d^n)}.$$

Rappelons que le principe de cette méthode est le suivant :

partant de $u^0 \in \mathbb{R}^n$, une suite (u^n) est construite telle que :

$$u^{n+1} \in u^n + G^n \quad \text{et} \quad J(u^{n+1}) = \inf_{v \in u^n + G^n} J(v) \tag{2}$$

avec $G^n = \text{vect}(\nabla J(u^0), \ldots, \nabla J(u^n))$.

1. Vérifier que $J(v) - J(u) = \dfrac{1}{2}\|v - u\|_A^2 \quad \forall v \in \mathbb{R}^N$.
2. Montrer que l'espace G^n est engendré par $\{\nabla J(u^0), A\nabla J(u^0), \ldots, A^n \nabla J(u^0)\}$.
3. Pour $k \geq 0$, on pose $e^k = u^k - u$.
   (a) Soit $v \in u^n + G^n$. Montrer qu'il existe un polynôme Q de degré n tel que : $v = u^0 + Q(A)\nabla J(u^0)$.
   (b) En utilisant (2) et (**??**), montrer que :
   $$\|e^{n+1}\|_A = \min\{\|P(A)e^0\|_A, \; P \in \mathcal{P}_{n+1}\} \tag{3}$$
   où et $\mathcal{P}_n$ désigne l'ensemble des polynômes P de degré n tels que $P(0) = 1$.
4. Montrer que
   $$\|P(A)e^0\|_A^2 \leq \|e^0\|_A^2 \max_i P^2(\lambda_i) \qquad \forall P \in \mathcal{P}_k.$$
   En déduire que
   $$\|e^n\|_A \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n \|e^0\|_A \qquad \text{avec} \;\; \kappa = \frac{\lambda_N}{\lambda_1}.$$

(Ind. $\min\{\|p\|_{L^\infty([a,b])} : p \in \mathcal{P}_k\} = 1/T_k\left(\frac{b+a}{b-a}\right)$, *où* T_k *désigne le polynôme de Tchebychev de degré k. De plus on a :*

$$T_k\left(\frac{b+a}{b-a}\right) = T_k\left(\frac{\frac{b}{a}+1}{\frac{b}{a}-1}\right) > \frac{1}{2}\left(\frac{\sqrt{\frac{b}{a}}+1}{\sqrt{\frac{b}{a}}-1}\right)^k, \qquad \textit{pour} \;\; b > a.$$

Corrigé 1 .

1. L'égalité $J(v) - J(u) = \dfrac{1}{2}\|v-u\|_A^2 \quad \forall v \in \mathbb{R}^N$ s'obtient par des calculs directs.
2. Nous allons raisonner par récurrence. Pour $n = 0$, l'assertion est évidente. Supposons que l'hypothèse

   $$\text{(HR)} \qquad G^{n-1} \text{ est engendré par } \{\nabla J(u^0), A\nabla J(u^0), \dots, A^{n-1}\nabla J(u^0)\}$$

   soit vraie à l'ordre $n-1$ et montrons que (HR) reste encore vraie à l'ordre n. Notons d'abord que, d'après (2), l'itéré $u^n \in u^{n-1} + G^{n-1}$ et s'écrit donc sous la forme $u^n = u^0 + \displaystyle\sum_{i=0}^{n-1} \gamma_i \nabla J(u^i)$, avec $\gamma_i \in \mathbb{R}$. Comme on a supposé (HR) vraie à l'ordre $n-1$, pour tout $i \leq n-1$ $\nabla J(u^i)$ s'écrit sous la forme d'une combinaison linéaire de $\{\nabla J(u^0), A\nabla J(u^0), \dots, A^{n-1}\nabla J(u^0)\}$. Donc il existe $\alpha_1, \cdots, \alpha_{n-1}$ tels que

   $$\begin{aligned} u^n &= u^0 + \sum_{i=0}^{n-1} \alpha_i A^i \nabla J(u^0), \\ \text{et} \quad \nabla J(u^n) &= Au^n - b = Au^0 - b + \sum_{i=0}^{n-1} \alpha_i A^{i+1}\nabla J(u^0) \\ &= \nabla J(u^0) + \sum_{i=1}^{n} \alpha_{i-1} A^i \nabla J(u^0). \end{aligned}$$

   On en déduit que G^n est bien engendré par $\{\nabla J(u^0), A\nabla J(u^0), \dots, A^n\nabla J(u^0)\}$.
3. (a) Soit $v \in u^n + G^n$. De la question précédente, on sait qu'ils existent $(\alpha_i)_{i=0}^n \in \mathbb{R}$ et $(\beta_i)_{i=0}^n \in \mathbb{R}$ tels que :

   $$u^n = u^0 + \sum_{i=0}^{n-1} \alpha_i A^i \nabla J(u^0) \qquad \text{et} \qquad v = u^0 + \left(\sum_{i=0}^{n-1} \alpha_i A^i\right)\nabla J(u^0) + \left(\sum_{i=0}^{n} \beta_i A^i\right)\nabla J(u^0).$$

   On en déduit alors qu'il existe un polynôme de degré n : $Q(x) = \displaystyle\sum_{i=0}^{n-1}(\alpha_i + \beta_i)x^i + \beta_n x^n$ tel que

   $$v = u^0 + Q(A)\nabla J(u^0).$$

(b) Remarquons d'abord que $\nabla J(u^0) = Au^0 - b = Au^0 - Au = Ae^0$. D'autre part, tenant compte de (2), (**??**) et de la question (a), il vient (on notera $\mathbb{P}_n$ l'ensemble des polynômes de degré n) :

$$\begin{aligned}
\frac{1}{2}\|e^{n+1}\|_A^2 = J(u^{n+1}) - J(u) &= \min_{v \in u^n + G^n} \{J(v) - J(u)\} \\
&= \frac{1}{2} \min_{v \in u^n + G^n} \|v - u\|_A^2 \\
&= \frac{1}{2} \min_{Q \in \mathbb{P}_n} \|u^0 - u + Q(A)\nabla J(u^0)\|_A^2 \\
&= \frac{1}{2} \min_{Q \in \mathbb{P}_n} \|(I + Q(A)A)e^0\|_A^2. \\
&= \frac{1}{2} \min_{P \in \mathcal{P}_{n+1}} \|P(A)e^0\|_A^2.
\end{aligned}$$

4. On va traduire (3) dans la base orthonormée $(v_l)_{l=1}^n$ des vecteurs propres de A. Cette base est A-orthogonale puisque $(Av_l, v_k) = \lambda_l (v_l, v_k)$, avec en particulier $\|v_l\|_A^2 = \lambda_l$. On pose $e^0 = \sum_{i=1}^N \alpha_i v_i$, on a :

$$P(A)e^0 = \sum_{i=1}^N \alpha_i P(\lambda_i) v_i, \quad \|e^0\|_A^2 = \sum_{i=1}^N \alpha_i^2 \|v_i\|_A^2 = \sum_{i=1}^N \lambda_i \alpha_i^2$$

et

$$\|P(A)e^0\|_A^2 = \sum_{i=1}^N \alpha_i^2 P^2(\lambda_i) \|v_i\|_A^2 = \sum_{i=1}^N \lambda_i \alpha_i^2 P^2(\lambda_i).$$

D'où

$$\begin{aligned}
\|P(A)e^0\|_A^2 &\leq \left(\sum_{i=1}^N \lambda_i \alpha_i^2\right) \max_i P^2(\lambda_i) \\
&= \|e^0\|_A^2 \max_i P^2(\lambda_i).
\end{aligned}$$

Et d'après (3)

$$\|e^{n+1}\|_A \leq \min\left\{ \max_{1 \leq i \leq N} |P(\lambda_i)| : P \in \mathcal{P}_{n+1} \right\} \|e^0\|_A.$$

Il en découle que pour tout $n \geq 0$:

$$\|e^n\|_A \leq \min\left\{ \|P\|_{L^\infty([\lambda_1, \lambda_N])} : P \in \mathcal{P}_n \right\} \|e^0\|_A,$$

où on a noté $\|P\|_{L^\infty([\lambda_1, \lambda_N])} = \max_{x \in [\lambda_1, \lambda_N]} |P(x)|$.

En utilisant le résultat donné en indication, le min ci dessus est égal à $1/T_n \left(\frac{\lambda_N + \lambda_1}{\lambda_N - \lambda_1}\right)$, et

$$\begin{aligned}
\|e^n\|_A &\leq 1/T_n \left(\frac{\lambda_N + \lambda_1}{\lambda_N - \lambda_1}\right) \|e^0\|_A, \\
&\leq 1/T_n \left(\frac{\kappa + 1}{\kappa - 1}\right) \|e^0\|_A, \\
&\leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^n \|e^0\|_A.
\end{aligned}$$

Exercice 2 Méthodes de Jacobi et Gauss-Seidel. Corrigé 2 . On calcule d'abord le rayon spectrale $\rho(\mathcal{J})$ de la matrice de Jacobi. Pour assurer la convergence de la méthode il faut et il suffit que ce rayon vérifie $\rho(\mathcal{J}) < 1$. Il est à noté que la matrice est symètrique, et $\rho(\mathcal{G}) = \rho(\mathcal{J})^2$.

Exercice 3 (Comparaison avec les méthodes itératives) .

Corrigé 3 .

1. Considérons le problème de corde vibrante fixée à ces deux extrémités. La statique de ce problème est fournie par la minimisation de l'énergie potentielle :

$$\begin{cases} \text{Trouver } u \in C^1[0,1] \text{ tel que } u(0) = 0, \; u(1) = 0, \text{ et} \\ E_p(u) = \min_{v \in C^1[0,1]} E_p(v), \text{ avec } E_p(v) = \displaystyle\int_0^1 \frac{1}{2}\left(\frac{dv}{dx}(x)\right)^2 - f(x)v(x)dx \end{cases} \tag{4}$$

Le problème de minimisation quadratique :

$$\begin{cases} \text{Trouver } u \in \mathbb{R}^N \text{ tel que :} \\ \frac{1}{2}(Au;\ u) - (b;\ u) = \min_{v \in \mathbb{R}^N} \frac{1}{2}(Av;\ v) - (b;\ v), \end{cases} \tag{5}$$

correspond à la discrétisation du problème (4), en effet si nous introduisons les deux réels $v_0 = 0$ et $v_{N+1} = 0$ alors :

$$\begin{aligned} (Av; v) &= \frac{1}{h^2}\sum_{i=1}^{N}(-v_{i-1} + 2\, v_i - v_{i+1})\, v_i, \\ &= \frac{1}{h^2}\sum_{i=1}^{N} v_i^2 - \sum_{i=1}^{N} v_{i-1}v_i - \sum_{i=1}^{N} v_{i+1}v_i + \sum_{i=1}^{N} v_i^2, \\ &= \frac{1}{h^2}\sum_{i=1}^{N} v_i^2 - \sum_{i=0}^{N-1} v_i v_{i+1} - \sum_{i=1}^{N} v_{i+1}v_i + \sum_{i=0}^{N-1} v_{i+1}^2, \\ &= \frac{1}{h^2}\sum_{i=0}^{N} v_i^2 - 2\, v_i v_{i+1} + v_{i+1}^2 = \frac{1}{h^2}\sum_{i=0}^{N}(v_{i+1} - v_i)^2. \end{aligned}$$

De là nous tirons

$$\begin{aligned} h\ J(v) &= h\ \sum_{i=0}^{N} \frac{1}{2}\left(\frac{v_{i+1} - v_i}{h}\right)^2 - h\ \sum_{i=1}^{n} v_i\ f(x_i) \\ &\simeq \int_0^1 \frac{1}{2}\left(\frac{dv}{dx}(x)\right)^2 dx - \int_0^1 f(x)v(x)dx. \end{aligned}$$

2. Pour répondre à cette question, définissions v^k comme l'image de φ^k par A et calculons sa $l^{\text{ème}}$ coordonnée v_l^k :

$$v_l^k = -\frac{1}{h^2}\sin((l-1)k\pi h) + \frac{2}{h^2}\sin(lk\pi h) - \frac{1}{h^2}\sin((l+1)k\pi h), \quad \forall l \in [\![2, N-1]\!]$$

Il est encore licite d'écrire cette relation pour $l = 1$ et $l = N$; en effet comme $(N+1)h = 1$, nous avons :

$$\sin(0k\pi h) \;=\; 0 \text{ et } \sin\left((N+1)k\pi h\right) \;=\; 0.$$

Ainsi, des formules trigonométriques, nous tirons :

$$v_l^k \;=\; \frac{2}{h^2}(1-\cos(k\pi h))\sin(lk\pi h) \;=\; \frac{4}{h^2}\sin^2(\frac{k\pi h}{2})\sin(lk\pi h).$$

Nous venons de montrer que les φ^k sont les vecteurs propres de A associés aux valeurs propres λ^k.

$$A\,\varphi^k \;=\; \lambda^k\,\varphi^k.$$

Le cardinal de cet ensemble étant égal à N, nous venons d'exhiber l'ensemble des valeurs propres de A.

3. Commençons par expliciter la plus petite et la plus grande valeur propre de A, elles correspondent à $k = 1$ et $k = N$

$$\begin{cases} \lambda_1 \;=\; \dfrac{4}{h^2}\sin^2(\dfrac{\pi h}{2}) \;=\; \pi^2 + O(h^2) \;\simeq \pi^2, \\ \lambda_N \;=\; \dfrac{4}{h^2}\sin^2(\dfrac{N\pi h}{2}) \;=\; \dfrac{4}{h^2}\cos^2(\dfrac{\pi h}{2}) \;=\; \dfrac{4}{h^2} - \pi^2 + O(h^2) \;\simeq \dfrac{4}{h^2}. \end{cases}$$

Le conditionnement de la matrice A est donc :

$$\kappa := \frac{\lambda_N}{\lambda_1} \sim \frac{4}{\pi^2 h^2}.$$

Ainsi, les vitesses de convergence pour les méthodes de gradient à pas fixe et optimal sont données par :

$$-\ln(\frac{\kappa-1}{\kappa+1}) = -\ln(\frac{1-\frac{1}{\kappa}}{1+\frac{1}{\kappa}}) \simeq \frac{2}{\kappa} \simeq \frac{\pi^2 h^2}{2}. \tag{6}$$

Et la vitesse de convergence du gradient conjugué est donnée par :

$$-\ln\left(\frac{1-1/\sqrt{\kappa}}{1+1/\sqrt{\kappa}}\right) \simeq \frac{2}{\sqrt{\kappa}} \simeq \pi h. \tag{7}$$

4. Pour la méthode de Jacobi, nous avons :

$$\mathcal{J} = D^{-1}(E+F) \;=\; \frac{h^2}{2}\,(\frac{2}{h^2}I - A)$$

Ainsi l'ensemble des valeurs propres $\lambda_k(\mathcal{J})$ de $\mathcal{J}$ est donné par :

$$\lambda_k(\mathcal{J}) = \frac{h^2}{2}\left(\frac{2}{h^2} - \lambda_k\right) = \cos(k\pi h) \quad k \in \{1,\cdots,N\},$$

dont les plus grandes en valeur absolue sont données pour $k = 1$ ou $k = N$:

$$\rho(\mathcal{J}) \;=\; \cos(\pi h) \;=\; 1 - \frac{\pi^2 h^2}{2} + O(h^4).$$

Rappelons l'algorithme de Jacobi :

$$D u^{n+1} \;=\; (E+F) u^n \;+\; b \tag{8}$$

Ce qui nous donne pour l'erreur $e^n \;=\; u^n - u$, comme $Du \;=\; (E+F)u + b$:

$$D e^{n+1} \;=\; (E+F) e^n \;\Longrightarrow\; u^{n+1} \;=\; \mathcal{J}\, u^n \tag{9}$$

Et comme $\mathcal{J}$ est une matrice symétrique :

$$\|e^{n+1}\|_2 \;\leqslant\; \rho(\mathcal{J}) \|e^n\|_2 \tag{10}$$

La vitesse de convergence est donc de

$$-\ln \rho(\mathcal{J}) \;=\; -\ln(1 - \frac{\pi^2 h^2}{2} + O(h^4)) \;\simeq\; \frac{\pi^2 h^2}{2}. \tag{11}$$

La vitesse de convergence de la méthode de Jacobi est comparable à celle des méthodes de gradient à pas fixe et à pas optimal.

5. Le rayons spectral $\rho(\mathcal{G})$ de la matrice de Gauss-Seidel nous est fournie par la formule :

$$\rho(G) \;=\; [\rho(\mathcal{J})]^2 = [1 - \frac{\pi^2 h^2}{2} + O(h^4)] \;=\; 1 - \pi^2 h^2 + O(h^4). \tag{12}$$

Ainsi la vitesse de convergence de la méthode de Gauss-Seidel et donnée par :

$$-\ln \rho(G) \;=\; \ln(1 - \pi^2 h^2 + O(h^4)) \;\simeq\; \pi^2 h^2. \tag{13}$$

Ce qui nous montre que sa vitesse de convergence est deux fois plus rapide que celle de la méthode des gradient à pas fixe et optimal mais beaucoup moins rapide que la méthode du gradient conjugué.

6. Considérons la méthode du gradient à pas fixe. Nous avons la majoration d'erreur :

$$\|e^n\|_2 \;\leqslant\; \left(\frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}\right)^n \|e^0\|_2. \tag{14}$$

Pour assurer $\|e^n\|_2 \;\leqslant\; \varepsilon$, il suffit d'imposer

$$\left(\frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}\right)^n \|e^0\|_2 \;\leqslant\; \varepsilon \tag{15}$$

C'est à dire :

$$n \geqslant \frac{\ln \|e^0\|_2 - \ln \varepsilon}{-\ln \frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}} \;\simeq\; \frac{-\ln \varepsilon}{\frac{\pi^2 h^2}{2}} \;=\; \frac{-2 \ln \varepsilon}{\pi^2 h^2}. \tag{16}$$

Application numérique :

$$n \;\geqslant\; 1,4 \;\times\; 10^4. \tag{17}$$