Optimisation & Analyse convexe
Séance 4 : Algorithmes pour l'optimisation sans contrainte

Dans tous les exercices suivants, A désigne une matrice symétrique définie positive $N \times N$, de valeurs propres $0 < \lambda_1 \leq \ldots \leq \lambda_N$, et $b \in \mathbb{R}^N$. Pour $v \in \mathbb{R}^N$, on pose

$$J(v) = \frac{1}{2}(Av, v) - (b, v).$$

Notons u le point de minimum de J sur $\mathbb{R}^N$.

Exercice 1 (Algorithme de gradient à pas fixe).

1. Montrer que l'algorithme de gradient à pas fixe

$$u^{n+1} = u^n - \rho \nabla J(u^n)$$

converge pour $0 < \rho < 2/\lambda_N$.

2. Montrer que la vitesse maximale de convergence est obtenue pour $\rho = 2/(\lambda_1 + \lambda_N)$ et préciser sa valeur.

Corrigé 1 .

1. Supposons que $0 < \rho < 2/\lambda_N$ et notons $e^n = u^n - u$. Pour établir la convergence de la méthode GPF, nous allons montrer que la norme de l'erreur e^n tend vers 0 lorsque $n \to +\infty$. Observons d'abord qu'on a : $b = Au$ et

$$e^{n+1} := u^{n+1} - u = u^n - \rho(Au^n - b) - u = \Big(I_N - \rho A\Big)\, e_n\,.$$

On sait qu'il existe une base orthonormale de vecteurs propres de A, notons-la $(p_i)_{1 \leq i \leq N}$. Soit $v \in \mathbb{R}^n$, on pose $v = \sum_{i=1}^N v_i\, p_i$. On a alors

$$\Big(I_N - \rho A\Big)\, v = \sum_{i=1}^N \Big(I_N - \rho A\Big)\, v_i\, p_i = \sum_{i=1}^N (1 - \rho\, \lambda_i)\, v_i\, p_i\,;$$

ainsi,

$$\begin{aligned} \|(I_N - \rho A)\, v\|_2^2 &= \langle \sum_{i=1}^N (1 - \rho\, \lambda_i)\, v_i\, p_i, \sum_{j=1}^N (1 - \rho\, \lambda_j)\, v_j\, p_j \rangle = \sum_{i=1}^N (1 - \rho\, \lambda_i)^2\, v_i^2 \\ &\leq \max_i (1 - \rho\, \lambda_i)^2 \sum_{j=1}^N v_j^2 = \Big(\max_i |1 - \rho\, \lambda_i|\Big)^2\, \|v\|_2^2\,, \end{aligned}$$

pour tout $v \in \mathbb{R}^n$. On en déduit donc que

$$\|e^{n+1}\|_2 \leq \max_i |1 - \rho\, \lambda_i|\, \|e^n\|_2\,.$$

Notons $\gamma_\rho = \max_i |1 - \rho\, \lambda_i|$. Avec un simple raisonnement par récurrence, on obtient

$$\|e^n\|_2 \leq \gamma_\rho^n \|e^0\|_2\,.$$

Une condition nécessaire et suffisante pour la convergence de l'algorithme est d'avoir γ_ρ strictement plus petit que 1. Vérifions donc que $\gamma_\rho < 1$.
Pour tout $1 \leq i \leq N$, on a : $\lambda_1 \leq \lambda_i \leq \lambda_N$ et donc $1 - \rho\,\lambda_N \leq 1 - \rho\,\lambda_i \leq 1 - \rho\,\lambda_1$. Il en résulte :

$$\gamma_\rho = \max_i |1 - \rho\lambda_i| = \max\left(|1 - \rho\lambda_1|, |1 - \rho\lambda_N|\right).$$

Cette inégalité avec l'hypothèse $0 < \rho < \frac{2}{\lambda_N}$, implique qu'on a bien $\gamma_\rho < 1$.

2. Rappelons que γ_ρ est appelé taux de convergence (voir définition 5.2.2, chapitre 5 du poly) et la vitesse de convergence est définie par $R_{GPF}(\rho) := -\ln(\gamma_\rho)$. Cette vitesse est maximale lorsque γ_ρ est minimal. Or, on a :

$$\gamma_\rho = \max\left(1 - \rho\lambda_1, \rho\lambda_N - 1\right).$$

Donc γ_ρ est minimal lorsque $1 - \rho\lambda_1 = \rho\lambda_N - 1$ ce qui correspond à $\rho \equiv \rho_o := \dfrac{2}{\lambda_1 + \lambda_N}$, et dans ce cas, $\gamma_{\rho_o} = \dfrac{\lambda_N - \lambda_1}{\lambda_1 + \lambda_N}$. La vitesse maximale est alors :

$$R_{GPF} = -\ln\left(1 - \frac{2\lambda_1}{\lambda_1 + \lambda_N}\right) = -\ln\left(1 - \frac{2}{1+\kappa}\right), \qquad \text{où} \;\; \kappa = \frac{\lambda_n}{\lambda_1}.$$

Commentaire. Lorsque κ est grand, la vitesse de convergence maximale est de l'ordre de $R_{GPF} \sim \dfrac{2}{\kappa}$. Plus κ est grand, plus la vitesse est lente ...

Exercice 2 (Algorithme de gradient à pas optimal).
On considère l'algorithme du gradient à pas optimal :

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

où $d^n = -\nabla J(u^n)$ et ρ^n réalise le minimum de $\inf_{\rho \in \mathbb{R}} J(u^n + \rho d^n)$.

1. Montrer que pour tout n, $(d^{n+1}, d^n) = 0$. Calculer ρ^n.
2. On pose $e^n = u^n - u$. Montrer que

$$\|e^{n+1}\|_A^2 = \left(1 - \frac{(d^n, d^n)^2}{(Ad^n, d^n)\;(A^{-1}d^n, d^n)}\right)\|e^n\|_A^2,$$

où $\|v\|_A^2 = (Av, v)$. En déduire que :

$$\|e^{n+1}\|_A \leq \left(\frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}\right)\|e^n\|_A. \tag{1}$$

Ind. On pourra utiliser le résultat suivant (admis) :

$$(\textit{Inégalité de Kantorovich}) \qquad \min_{x \in \mathbb{R}^N} \frac{(x,x)^2}{(Ax,x)(A^{-1}x,x)} = \frac{4\lambda_1\lambda_N}{(\lambda_1 + \lambda_N)^2}.$$

Corrigé 2 .

1. Le pas optimal ρ^n minimise la fonction $f : \rho \longmapsto J(u^n + \rho d^n)$. Donc, en se rappelant que $d^{n+1} = -\nabla J(u^n + \rho^n d^n)$, on obtient :
$$0 = f'(\rho^n) = (\nabla J(u^n + \rho^n d^n), d^n) = -(d^{n+1}, d^n).$$
Rappelons aussi que $d^{n+1} = b - Au^n - \rho^n Ad^n = d^n - \rho^n Ad^n$. Donc
$$\begin{aligned}(d^{n+1}, d^n) = 0 \quad &\Longrightarrow \quad (d^n, d^n) - \rho^n (Ad^n, d^n) = 0 \\ &\Longrightarrow \quad \rho^n = \frac{(d^n, d^n)}{(Ad^n, d^n)}.\end{aligned}$$
Remarque : A chaque itération n, le pas ρ^n est bien définie puisque $(Ad^n, d^n) \neq 0$. En effet, si $(Ad^n, d^n) = 0$ alors $d^n = b - Au^n = 0$ (et $u^n = u$), auquel cas l'algorithme s'arrêterait !
2. Notons d'abord que $e^{n+1} = u^{n+1} - u = u^n + \rho^n d^n - u = e^n + \rho^n d^n$. Donc
$$\begin{aligned}\|e^{n+1}\|_A^2 &= (A(e^n + \rho^n d^n), e^n + \rho^n d^n) \\ &= (Ae^n, e^n) + (\rho^n)^2 (Ad^n, d^n) + 2\rho^n (Ae^n, d^n) \\ &= (Ae^n, e^n) + \frac{(d^n, d^n)^2}{(Ad^n, d^n)} + 2\frac{(d^n, d^n)(Ae^n, d^n)}{(Ad^n, d^n)}.\end{aligned}$$
De plus, on a : $Ae^n = Au^n - Au = Au^n - b = -d^n$, et $e^n = -A^{-1}d^n$. Par conséquent :
$$\begin{aligned}\|e^{n+1}\|_A^2 &= (Ae^n, e^n) + \frac{(d^n, d^n)^2}{(Ad^n, d^n)} - 2\frac{(d^n, d^n)^2}{(Ad^n, d^n)} \\ &= (Ae^n, e^n)\left(1 - \frac{(d^n, d^n)^2}{(Ad^n, d^n)(Ae^n, e^n)}\right) \\ &= \|e^n\|_A^2 \left(1 - \frac{(d^n, d^n)^2}{(Ad^n, d^n)(d^n, A^{-1}d^n)}\right).\end{aligned}$$
En utilisant l'inégalité de Kantorovich, on obtient :
$$\|e^{n+1}\|_A^2 \leq \|e^n\|_A^2 \left(1 - \frac{4\lambda_1\lambda_N}{(\lambda_1 + \lambda_N)^2}\right), \qquad \text{ou encore} \quad \|e^{n+1}\|_A \leq \|e^n\|_A \left(\frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}\right).$$

Commentaire. La méthode GPO a le même taux de convergence que la méthode du gradient à pas fixe avec $\rho = 2/(\lambda_1 + \lambda_N)$.

Exercice 3 (Algorithme du gradient conjugué).
Il s'agit de l'algorithme de descente à pas optimal :
$$u^{n+1} = u^n + \rho^n d^n \tag{2}$$
avec
$$d^0 = -\nabla J(u^0), \quad d^n = -\nabla J(u^n) + \frac{\|\nabla J(u^n)\|^2}{\|\nabla J(u^{n-1})\|^2} d^{n-1}, \quad \rho^n = \frac{(-\nabla J(u^n), d^n)}{(Ad^n, d^n)}.$$
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 :

$$u^{n+1} \in u^n + G^n \quad \text{et} \quad J(u^{n+1}) = \inf_{v \in u^n + G^n} J(v) \tag{3}$$

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

Dans tout l'exercice, on note $\|v\|_A^2 = (Av, v)$ pour $v \in \mathbb{R}^N$.

1. Vérifier que
$$J(v) - J(u) = \frac{1}{2}\|v-u\|_A^2 \quad \forall v \in \mathbb{R}^N. \tag{4}$$
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 (3) et (4), montrer que :
   $$\|e^{n+1}\|_A = \min\{\|P(A)e^0\|_A,\ P \in \mathcal{P}_{n+1}\} \tag{5}$$
   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 :*

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

Corrigé 3 .

1. L'égalité (4) s'obtient par des calculs directs.
2. Nous allons raisonner par récurrence. Pour $n = 0$, l'assertion est évidente. Supposons que l'hypothèse
$$\text{(HR)} \qquad G^{n-1} \text{ est engendré par } \{\nabla J(u^0), A\nabla J(u^0), \ldots, A^{n-1}\nabla J(u^0)\}$$
soit vraie à l'ordre $n-1$ et montrons que (HR) reste encore vraie à l'ordre n. Notons d'abord que, d'après (3), l'itéré $u^n \in u^{n-1} + G^{n-1}$ et s'écrit donc sous la forme $u^n = u^0 + \sum_{i=0}^{n-1} \gamma_i \nabla J(u^i)$, 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 $\{\nabla J(u^0), A\nabla J(u^0), \ldots, A^{n-1}\nabla J(u^0)\}$. 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 $\{\nabla J(u^0), A\nabla J(u^0), \ldots, A^n \nabla J(u^0)\}$.

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 : $Q(x) = \sum_{i=0}^{n-1} (\alpha_i + \beta_i) x^i + \beta_n x^n$ 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 (3), (4) 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 (5) 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 (5)

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

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 4 (Comparaison avec les méthodes itératives) . On suppose maintenant que la matrice A et le vecteur b sont donnés par :

$$A = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \dots & \ddots & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{pmatrix}, \qquad b = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}.$$

où $h = 1/(N+1)$, $x_i = ih$ et f est une fonction continue sur $\mathbb{R}$. Le but de cet exercice est de comparer les méthodes de descente gradient et gradient conjugué avec les méthodes itératives Jacobi et Gauss-Seidel.

1. Donner une interprétation physique au problème de minimisation.

2. Montrer que les valeurs propres de la matrice A sont les :

$$\lambda^k = \frac{4}{h^2}\sin^2(\frac{k\pi h}{2}) \qquad 1 \leq k \leq N,$$

associées aux vecteurs propres φ^k

$$\varphi^k = \begin{bmatrix} \sin(k\pi h) & \sin(2k\pi h) & \ldots & \sin((N-1)k\pi h) & \sin(Nk\pi h) \end{bmatrix}^T.$$

3. Evaluer les vitesses de convergence du gradient et gradient conjugué pour h petit.
4. Calculer le rayon spectral $\rho(\mathcal{J})$ de la matrice de Jacobi utilisée pour la résolution du système d'optimalité. Evaluer $\rho(\mathcal{J})$ et la vitesse de convergence de la méthode pour h petit. Que peut-on en déduire ?
5. Calculer le rayon spectral $\rho(\mathcal{G})$ de la matrice de Gauss-Seidel. Evaluer la vitesse de convergence pour h petit. Que peut-on en déduire ?
6. Estimer le nombre d'itérations nécessaires pour obtenir la convergence à une précision ε donnée pour l'une quelconque de ces méthodes.
   Application : $n = 99,\ \varepsilon = 10^{-3}$.

Corrigé 4 .

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 :

$$\begin{cases} \text{Trouver } u \in C^1[0,1] \text{ tel que } u(0) \ = \ 0,\ u(1) \ = \ 0, \text{ et} \\ E_p(u) \ = \ \displaystyle\min_{v \in C^1[0,1]} E_p(v), \text{ avec } E_p(v) \ = \ \int_0^1 \frac{1}{2}\left(\frac{dv}{dx}(x)\right)^2 - f(x)v(x)dx \end{cases} \tag{6}$$

Le problème de minimisation quadratique :

$$\begin{cases} \text{Trouver } u \in \mathbb{R}^N \text{ tel que :} \\ \dfrac{1}{2}(Au;\ u) - (b;\ u) \ = \ \displaystyle\min_{v \in \mathbb{R}^N} \frac{1}{2}(Av;\ v) - (b;\ v), \end{cases} \tag{7}$$

correspond à la discrétisation du problème (6), 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 :

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

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

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

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 valeurs propres de A associées 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 :

$$-\ln(\frac{\kappa-1}{\kappa+1}) = -\ln(\frac{1-\frac{1}{\kappa}}{1+\frac{1}{\kappa}}) \simeq \frac{2}{\kappa} \simeq \frac{\pi^2 h^2}{2}. \tag{8}$$

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

$$-\ln\left(\frac{1-1/\sqrt{\kappa}}{1+1/\sqrt{\kappa}}\right) \simeq \frac{2}{\sqrt{\kappa}} \simeq \pi h. \tag{9}$$

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 :

$$Du^{n+1} \;=\; (E+F)u^n \;+\; b \tag{10}$$

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

$$De^{n+1} \;=\; (E+F)e^n \quad\Longrightarrow\quad u^{n+1} \;=\; \mathcal{J}\,u^n \tag{11}$$

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

$$\|e^{n+1}\|_2 \;\leqslant\; \rho(\mathcal{J})\|e^n\|_2 \tag{12}$$

La vitesse de convergence est donc de

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

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 :

$$\rho(G) \;=\; [\rho(\mathcal{J})]^2 = [1 - \frac{\pi^2 h^2}{2} + O(h^4)] \;=\; 1 - \pi^2 h^2 + O(h^4). \tag{14}$$

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

$$-\ln\rho(G) \;=\; \ln(1 - \pi^2 h^2 + O(h^4)) \;\simeq\; \pi^2 h^2. \tag{15}$$

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 :

$$\|e^n\|_2 \;\leqslant\; \left(\frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}\right)^n \|e^0\|_2. \tag{16}$$

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

$$\left(\frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}\right)^n \|e^0\|_2 \leqslant \varepsilon \tag{17}$$

C'est à dire :

$$n \geqslant \frac{\ln \|e^0\|_2 - \ln \varepsilon}{-\ln \frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}} \simeq \frac{-\ln \varepsilon}{\frac{\pi^2 h^2}{2}} = \frac{-2\ln \varepsilon}{\pi^2 h^2}. \tag{18}$$

Application numérique :

$$n \geqslant 1,4 \times 10^4. \tag{19}$$