Leçon 233 : Analyse numérique matricielle : résolution approchée de systèmes linéaires, recherche de vecteurs propres, exemples

[L3] p11 <u>déf 1</u> : Soit $A \in \mathcal{M}_n(\mathbb{K})$, $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. On appelle rayon spectral de A et on note $\rho(A)$ le maximum des modules des valeurs propres de A.

<u>déf 2</u> : La norme $\|\ \|$ sur $\mathcal{M}_n(\mathbb{K})$ induite par la norme $\mathbb{K}$ sur $\mathbb{K}^n$ est définie par : $\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$

[L3] p14 <u>Rq</u> : Pour toute norme $\|\ \|$, on a $\rho(A) \leq \|A\|$

I Conditionnement d'un système linéaire [L3] p19

<u>déf 3</u> : Soit $A \in GL_n(\mathbb{K})$. Le conditionnement de A par rapport à la norme $\|\ \|$ est le nombre
$$\text{cond}(A) = \|A\|\,\|A^{-1}\|.$$

<u>prop 4</u> : Soit $A \in GL_n(\mathbb{K})$. Soient x et $x + \delta x$ les solutions de $Ax = b$ et $A(x+\delta x) = b + \delta b$ où $b \in \mathbb{K}^n$ alors
$$\frac{\|\delta x\|}{\|x\|} \leq \text{cond}(A) \frac{\|\delta b\|}{\|b\|}$$

<u>prop 5</u> : Soient $A \in GL_n(\mathbb{K})$ et x et $x + \delta x$ les solutions de $Ax = b$ et $(A+\delta A)(x+\delta x) = b$, où $b \in \mathbb{K}^n$ alors
$$\frac{\|\delta x\|}{\|x+\delta x\|} \leq \text{cond}(A) \frac{\|\delta A\|}{\|A\|}$$

<u>Rq</u> : Plus cond(A) est petit, mieux le système est conditionné.

II Méthodes directes

[L3] p26-27 Lorsque l'on cherche à approcher la solution de l'équation de la chaleur, on est conduit à résoudre $\frac{\partial}{\partial x}\left(K \frac{\partial T}{\partial x}\right) = f(x)$, K constant, le système
$$-\begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & & -1 \\ 0 & & -1 & 2 \end{pmatrix} \begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix} = \frac{1}{K} \begin{pmatrix} \Delta x^2 f_1 - \alpha \\ \Delta x^2 f_2 \\ \vdots \\ \Delta x^2 f_n - \beta \end{pmatrix}$$

On va donc chercher à résoudre des systèmes de la forme $Ax = b$ avec $A \in \mathcal{M}_n(\mathbb{K})$ et $b = \mathbb{K}^n$

Avec les formules de Cramer, on a un coût en $O(n^2(n+1)!)$

1. Méthode du pivot de Gauss

[L3] p28 On va chercher à obtenir un système équivalent avec une matrice A triangulaire qui est facile à résoudre.

<u>algorithme</u> :

$A = (a_{ij}^{(1)})$ $\quad b = (b_i^{(1)})$

Pour k allant de 1 à n. On suppose $a_{kk}^{(k)} \neq 0$. [L3] p30

- pour $i = k+1, \dots, n$
  - $m_i^{(k)} = \frac{a_{i1}^{(k)}}{a_{kk}^{(k)}}$ $\quad$ • $b_i^{(k+1)} = b_i^{(k)} - m_i^{(k)} b_k^{(k)}$
  - $\forall j = k+1, \dots, n \quad a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_i^{(k)} a_{kj}^{(k)}$

sortie :
$$A^{(n)} = \begin{pmatrix} a_{11}^{(1)} & * & \cdots & * \\ & \ddots & & \vdots \\ & & \ddots & * \\ 0 & & & a_{nn}^{(n)} \end{pmatrix} \qquad b^{(n)} = \left(b_{ij}^{(n)}\right)$$

<u>prop 7</u> : Le coût du pivot de Gauss est en $O(n^3)$ [L3] p32

<u>prop 8</u> : Si A est symétrique alors à l'étape k, on a $\forall i,j \in [\![k+1, n]\!]$ $a_{ij}^{(k+1)} = a_{ji}^{(k+1)}$. On a donc deux fois moins de calculs. [L3] p34

<u>ex</u> : $\begin{pmatrix} 2 & 7 & 8 \\ 3 & 3 & 4 \\ 1 & 6 & 5 \end{pmatrix} X = \begin{pmatrix} 0 \\ 3 \\ 4 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 2 & 7 & 8 \\ 0 & -15/2 & -8 \\ 0 & 0 & -5/3 \end{pmatrix} X = \begin{pmatrix} 0 \\ 3 \\ 5 \end{pmatrix}$

<u>appli</u> : calcul du déterminant d'une matrice. [L3] p32

<u>déf 9</u> : Soit $A = (a_{ij})$. Le plus petit entier K, tel que si $|i-j| > K$ alors $a_{ij} = 0$, est appelé demi-largeur de bande. [L3] p34

<u>prop 10</u> : Si $A \in \mathcal{M}_n(\mathbb{K})$ est de demi-largeur de bande K et si les $a_{kk}^{(k)}$ sont tous non nuls alors les matrices $A^{(k)}$ sont toutes des matrices de demi-largeur de bande $\leq K$.

2. Décomposition LU

<u>prop 11</u> : Soit A une matrice dont les mineurs principaux sont tous non nuls. Alors : [L3] p36

i) les pivots $a_{kk}^{(k)} \neq 0$

ii) $\exists$ L triangulaire inférieure et U triangulaire supérieure tq $A = LU$

[L3] p36-37 Avec les mêmes notations que précédemment, on a

$$L = \begin{pmatrix} 1 & & & 0 \\ m_2^{(1)} & \ddots & & \\ \vdots & & \ddots & \\ m_n^{(1)} & \cdots & m_n^{(n-1)} & 1 \end{pmatrix} \qquad U = A^{(n)}$$

ex: $\begin{pmatrix} 2 & 7 & 8 \\ 3 & 3 & 4 \\ 1 & 6 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 3/2 & 1 & 0 \\ 1/2 & -1/3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 7 & 8 \\ 0 & -15/2 & -8 \\ 0 & 0 & -5/3 \end{pmatrix}$

[L3] p37 Rq: Si l'on impose que $\forall i$, $l_{ii} = 1$ alors la décomposition est unique.

[L3] p38 Rq: Pour résoudre le système $Ax = b$, on résout le système $Ly = b$ puis le système $Ux = y$.

3. Décomposition de Cholesky

[L3] p38 thm 12: Une matrice $A \in S_n^{++}(\mathbb{K}) \Leftrightarrow \exists L$ triang inf tq $A = L\,{}^tL$

Rq: Cette décomposition est unique si tous les coefficients de L sont positifs.

[L3] p40 algorithme:

$L(1:n; 1:n) = 0$

$L(1,1) = \sqrt{a(1,1)}$

pour $i = 2 \dots n$ faire

$L(i,1) = \dfrac{a(i,1)}{L(1,1)}$

pour $j = 2 \dots n$ faire

$L(j,j) = \sqrt{a(j,j) - \sum_{k=1}^{j-1} L(j,k)^2}$

pour $j = i \dots n$ faire

$L(i,j) = \dfrac{1}{L(j,j)}\left(a(i,j) - \sum_{k=1}^{j-1} L(i,k)L(j,k)\right)$

ex: $A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} = L \times {}^tL$

avec $L = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ -1/\sqrt{2} & \sqrt{3}/\sqrt{2} & 0 \\ 0 & -\sqrt{\frac{2}{3}} & \frac{2}{\sqrt{3}} \end{pmatrix}$

[L3] p40 prop 13: Si $A \in S_n(\mathbb{K})$ est de demi-largeur de bande K alors L est de demi-largeur de bande K.

4. Décomposition QR

[L3] p41 thm 14: Soit $A \in \mathcal{M}_{n,m}(\mathbb{R})$ de rang $n \leq m$. Alors $\exists Q \in GL_n(\mathbb{R})$ et $R \in \mathcal{M}_{n,m}(\mathbb{R})$ triang inférieure telle que $A = QR$

ex: · moindres carrés

· $\begin{pmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{pmatrix} X = b \Leftrightarrow \begin{pmatrix} 6/7 & 69/175 & -58/175 \\ 3/7 & -158/175 & 6/175 \\ -2/7 & -6/35 & -33/35 \end{pmatrix} \begin{pmatrix} 14 & 21 & -14 \\ 0 & -175 & 70 \\ 0 & 0 & 35 \end{pmatrix} X = b$

5. FFT

Soit $N \in \mathbb{N}$ et $\omega = e^{\frac{2i\pi}{N}}$ [LAS] p464

déf 15 : On appelle transformée de Fourier discrète de longueur N, l'application qui à toute suite de nombres complexes $(x_j)_{j=1\dots N}$ fait correspondre la suite $(y_i)_{i=1}^N$ définie par $y_i = \sum_{j=1}^N \omega^{-ij} x_j$

prop 16 : Réciproquement, $x_k = \frac{1}{N}\sum_{i=1}^N \omega^{ki} y_i$

III. Méthodes itératives

But: On ne cherche pas à trouver une solution à $Ax = b$ mais à construire une suite de vecteurs (x_k) qui converge vers la solution. [L3] p44-4

Dans le cas de l'équation de la chaleur, avec K non constant ne dépendant que de T, on est ramené à un système $AT = B$ où A dépend de T. D'où une résolution avec une méthode itérative.

1. Méthode de décomposition

[L3] p $Ax = b$. On décompose $A = M - N$ avec M inversible

La solution vérifie $x = M^{-1}Nx + M^{-1}b$.

$\begin{cases} Mx^{(k+1)} = Nx^{(k)} + b \\ x^{(0)} \in \mathbb{R}^n \end{cases}$ Si $x^{(k)} \xrightarrow[k\to\infty]{} x$ alors x est solution.

prop 17 : La méthode converge ssi $\rho(M^{-1}N) < 1$

prop 18 : Soit $A \in S_n^{++}(\mathbb{R})$. Si $A = M - N$ avec M inversible alors [L3] p47

i) ${}^tM + N \in S_n(\mathbb{R})$

ii) Si ${}^tM + N \in S_n^{++}(\mathbb{R})$ alors $\rho(M^{-1}N) < 1$

Méthode de Jacobi

On effectue la décomposition $A = D - M$ où D est la matrice formée des éléments diagonaux de A. [L3] p47

[L3] p48

déf 19: On dit que $A=(a_{ij})\in\mathcal{M}_n(\mathbb{R})$ est à diagonale dominante stricte si $\forall i=1,\dots,n \quad |a_{ii}| > \sum_{j\neq i} |a_{ij}|$

prop 20: Si A est diagonale dominante stricte, alors $\rho(D^{-1}M)<1$ et la méthode converge.

Méthode de Gauss-Seidel

[L3] p51

On effectue la décomposition $A=M-N$ avec $M=\begin{pmatrix} a_{11} & & 0 \\ \vdots & \ddots & \\ a_{n1} & \cdots & a_{nn}\end{pmatrix}$

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left[-\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right] + \frac{b_i}{a_{ii}}$$

[L3] p52

prop 21 : Soit $A\in S_n^{++}(\mathbb{R})$, $A={}^tL+L+D$. Alors la méthode de Gauss-Seidel converge.

[L3] p53

prop 22 : Si $A\in\mathcal{M}_n(\mathbb{K})$ est à diagonale dominante stricte alors la méthode de Gauss-Seidel converge.

2. Méthode du gradient

déf 23 : Pour A matrice symétrique définie positive l'algorithme

DVT

[MEN]

$$\begin{cases} r_0 = b - Ax_0\ ,\ p_0 = r_0\ ,\quad x_0 \text{ donné dans } \mathbb{R}^n \\ x_{k+1} = x_k + \alpha_k p_k\ ,\ r_{k+1} = r_k - \alpha_k A p_k\ ,\ p_{k+1} = r_{k+1} + \beta_k p_k \\ \alpha_k = \dfrac{\|r_k\|^2}{\langle Ap_k, p_k\rangle}\ ,\ \beta_k = \dfrac{\|r_{k+1}\|^2}{\|r_k\|^2} \\ k\in\mathbb{N},\ \text{tant que } r_k\neq 0 \end{cases}$$

est appelé méthode du gradient conjugué.

thm 24 : Si $A\in S_n^{++}(\mathbb{R})$, alors la méthode du gradient conjugué converge en au plus n itérations.

IV Recherche de vecteurs propres et valeurs propres

[L3] p68

Lorsque l'on étudie un système de trois masses reliées par des ressorts et que l'on cherche des solutions sous la forme $X(t)=\sin(\omega t)\hat{X}$, il faut que ω et $\hat{X}$ vérifient $(K-\omega^2M)\hat{X}=0$. Cela revient à une recherche des valeurs propres de KM^{-1}.

1. Conditionnement, localisation et sensibilité

prop 25 : Soit A diagonalisable, P tq $P^{-1}AP = D = \operatorname{diag}(\lambda_i)$ et $\|\cdot\|$ une norme tq $\|\operatorname{diag}(d_i)\| = \max|d_i|$. Alors $\forall \delta A$, $Sp(A+\delta A)\subset \bigcup D_i$ où $D_i = \{z\in\mathbb{C};\ |z-\lambda_i|\leq \operatorname{cond}(P)\times\|\delta A\|\}$ [L3] p69

thm 26 : (Gershgorin) Soit $A=(a_{ij})\in\mathcal{M}_n(\mathbb{K})$ diagonale dominante stricte. Pour $k\in\{1,\dots,n\}$, on définit les disques de Gershgorin $D_k\subset\mathbb{C}$ associés à A : [L3] p48

$$z\in D_k \Leftrightarrow |z-a_{kk}|\leq \sum_{j\neq k}|a_{k,j}|$$

On a $Sp(A)\subset\bigcup_{k=1}^n D_k$

prop 27 : Soit $A\in\mathcal{M}_n(\mathbb{C})$ hermitienne. Supposons que λ soit une valeur propre simple de A et que x soit un vecteur propre unitaire associé. Alors : [AMO] p207

$\exists$ un voisinage ouvert V_A de A et une unique application C^∞ (et même analytique réelle) $(X,\Lambda): V_A \longrightarrow \mathbb{C}^n\times\mathbb{R}$ tq :

i) $X(A)=x$ et $\Lambda(A)=\lambda$

ii) $BX(B)=\Lambda(B)X(B)$ et $\|X(B)\|_2 = 1 \quad \forall B\in V_A$

2. Calcul de valeurs propres

prop 28 : Soit $A\in\mathcal{M}_n(\mathbb{K})$ diagonalisable dont la valeur propre de plus grand module λ_1 est unique. Soit $q_0\in\mathbb{R}^n$ qui n'est pas orthogonal au sous-espace propre associé à λ_1. Alors la suite définie par $x^{(k)} = Aq^{(k-1)}$ et $q^{(k)} = \frac{x^{(k)}}{\|x^{(k)}\|}$ vérifie : [L3] p71

i) $q=\lim_{k\to\infty}\left(\frac{\lambda_1}{|\lambda_1|}\right)^k q^{(k)}$ est un vecteur propre de norme 1 associé à λ_1 DVT

ii) $\lim_{k\to\infty}\|Aq^{(k)}\| = |\lambda_1|$

iii) Si $q_j^{(k)}$ est la j-ème coordonnée de $q^{(k)}$
$\lim_{k\to\infty}\frac{x_j^{(k+1)}}{q_j^{(k)}} = \lambda_1$

ex : trouver les deux plus grandes valeurs propres de : $\begin{pmatrix} 2 & 1 & 2 & 3 & 4 \\ 1 & 2 & 1 & 2 & 3 \\ 2 & 1 & 2 & 1 & 2 \\ 3 & 2 & 1 & 2 & 1 \\ 4 & 3 & 2 & 1 & 2\end{pmatrix}$ [L3] p72

3. Calcul de vecteurs propres

prop 29 : Soient A une matrice diagonalisable et λ une valeur propre quelconque. On suppose $\tilde{\lambda}$ tq $\tilde{\lambda}\neq\lambda$ et $|\tilde{\lambda}-\lambda| < |\tilde{\lambda}-\mu| \quad \forall\mu\in Sp(A)\setminus\{\lambda\}$ alors : [L3] p72

$\lim_{k\to\infty}\frac{(\lambda-\tilde{\lambda})^k}{|\lambda-\tilde{\lambda}|^k}\frac{u^{(k)}}{\|u^{(k)}\|} = q$, q est un vecteur propre associé à λ.

<u>Références</u> :

[L3] : L3 Mathématiques appliquées, YGER et WEIL

[LAS] : Analyse numérique matricielle appliquée à l'art de l'ingénieur. tome 2   Lascaux, théodor

[MEN] : Analyse numérique des équations aux dérivées partielles, Di Menza

[AMO] : Analyse numérique matricielle, Amodei

Développement n°1

—. Méthode du gradient conjugué .—

Soit $A \in M_n(\mathbb{R})$ symétrique définie positive (sdp), $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.

<u>METHODES DU GRADIENT</u>. Minimiser sur $\mathbb{R}^n$ la fonction f définie par $f(y) = \frac{1}{2}\langle Ay, y\rangle - \langle b, y\rangle + c$, $y \in \mathbb{R}^n$.

<u>Rque</u>. • $\{x, h\} \subset \mathbb{R}^n$

$$\begin{aligned} f(x+h) &= \tfrac{1}{2}\langle A(x+h), x+h\rangle - \langle b, x+h\rangle + c \\ &= f(x) + \tfrac{1}{2}\big[\langle Ax, h\rangle + \underbrace{\langle Ah, x\rangle}_{=\langle Ax, h\rangle}\big] + \tfrac{1}{2}\langle Ah, h\rangle - \langle b, h\rangle \\ &= f(x) + \underbrace{\langle Ax - b, h\rangle}_{=Df(x).h} + \underbrace{\tfrac{1}{2}\langle Ah, h\rangle}_{=o(h)} \qquad (1) \end{aligned}$$

$\Rightarrow \nabla f(x) = Ax - b$.

• $\{x, h, k\} \subset \mathbb{R}^n$

Df affine $\Rightarrow D^2 f(x).(h,k) = \langle Ak, h\rangle$.

A sdp $\Rightarrow f$ convexe $\Rightarrow$ tout point critique est min. global

• Bilan : A inversible $\Rightarrow \exists!$ point critique ($\nabla f(x) = 0$).

+ A sdp $\Rightarrow \exists!$ min. global.

<u>Thm</u>. Si A est sdp, alors résoudre $Ax = b$, $x \in \mathbb{R}^n$ est équivalent à ~~résoudre~~ minimiser f sur $\mathbb{R}^n$.

<u>Preuve</u>. Si x est solution de $Ax = b$ (sur $\mathbb{R}^n$), alors pour $h \in \mathbb{R}^n \setminus \{0\}$, (1) devient $f(x+h) = f(x) + \frac{1}{2}\langle Ah, h\rangle$ et on a $f(x+h) > f(x)$. Si x minimise f (sur $\mathbb{R}^n$) alors $\nabla f(x) = 0$, i.e. $Ax = b$.

Idée. Faire décroître f en suivant la direction du gradient car « gradient = direction de plus grande pente » (Considérer (1) ; appliquer l'inégalité de Cauchy-Schwarz à $|\langle \nabla f(x), h \rangle|$ et utiliser le cas d'égalité.)

MÉTHODE DU GRADIENT À PAS CONSTANT.

$x_{k+1} := x_k - \alpha \nabla f(x_k)$, $\alpha > 0$ fixé.

MÉTHODE DU GRADIENT À PAS OPTIMAL.

$x_{k+1} := x_k + \alpha_k \nabla f(x_k)$,

avec α_k qui minimise $g_k : \alpha \mapsto f(x_k + \alpha \nabla f(x))$.

MÉTHODE DU GRADIENT CONJUGUÉ.

$\boxed{x_{k+1} := x_k + \alpha_k p_k}$,

avec α_k qui minimise $g_k : \alpha \mapsto f(x_k + \alpha p_k)$

et $\langle p_{k+1}, A p_k \rangle = 0$ (: directions conjuguées)

Recherche des ~~expressions~~ relations de récurrence.

- $r_k := b - Ax_k$, le résidu de x_k pour le système $Ax_k = b$ (écart avec la solution, $= -\nabla f(x_k)$).
- $p_k := r_k + \gamma_k (x_k - x_{k-1})$, modification de la direction du gradient proportionnellement au déplacement précédent $x_k - x_{k+1}$.

$$\boxed{p_k = r_k + \underbrace{\gamma_k \alpha_{k-1}}_{=: \beta_{k-1}} p_{k-1}}$$

- Expression de r_{k+1} en fonction de r_k, α_k et p_k :

$r_{k+1} = b - Ax_{k+1} = b - Ax_k - A(\alpha_k p_k)$

$$\boxed{r_{k+1} = r_k - \alpha_k A p_k}$$

- Expression de α_k en fonction de r_k et p_k (r_k et p_k étant connus à l'étape k) :

$$g_k'(\alpha) = Df(x_k + \alpha p_k) \cdot \underbrace{p_k}_{=(\alpha \mapsto x_k + \alpha p_k)'(1)} = \langle A(x_k + \alpha p_k) - b, p_k \rangle =$$

$$= \langle -r_k + A\alpha p_k, p_k \rangle$$

$$g_k'(\alpha) = 0 \iff \alpha \langle A p_k, p_k \rangle = \langle p_k, r_k \rangle$$

$$\boxed{\alpha_k := \frac{\langle p_k, r_k \rangle}{\langle A p_k, p_k \rangle}}$$ tant que $p_k \neq 0$ ($\langle A p_k, p_k \rangle$ est alors non nul)

- Expression de $f(x_{k+1})$ en fonction de $f(x_k)$, r_k et p_k :

$$\underset{\text{cf (1)}}{f(x_{k+1}) =} f(x_k) + \alpha_k \langle \underbrace{A x_k - b}_{= -r_k}, p_k \rangle + \frac{1}{2} \alpha_k^2 \langle A p_k, p_k \rangle$$

$$= f(x_k) - \frac{1}{2} \frac{\langle p_k, r_k \rangle^2}{\langle A p_k, p_k \rangle} .$$

Or $\langle p_k, r_{k+1} \rangle = \langle p_k, r_k - \alpha_k A p_k \rangle = \langle p_k, r_k \rangle - \alpha_k \langle p_k, A p_k \rangle = 0$ et donc

$$\langle p_k, r_k \rangle = \langle r_k + \beta_{k-1} p_{k-1}, r_k \rangle = \| r_k \|^2 . \qquad (2)$$

D'où

$$f(x_{k+1}) = f(x_k) - \frac{\| r_k \|^2}{2 \langle A p_k, r_k \rangle}$$

- r_k étant donné, on veut p_k qui minimise $\langle A p_k, p_k \rangle$ (pour minimiser $f(x_{k+1})$). Or

$$\langle A p_k, p_k \rangle = \langle A(r_k + \beta_{k-1} p_{k-1}), r_k + \beta_{k-1} p_{k-1} \rangle$$

$$= \langle A r_k, r_k \rangle + 2 \beta_{k-1} \langle A p_{k-1}, r_k \rangle + \beta_{k-1}^2 \langle A p_{k-1}, p_{k-1} \rangle .$$

$$=: P_k(\beta_{k-1}) .$$

On veut $\overset{\text{donc}}{}$ β_{k-1} qui minimise $P_k(\beta)$ (polynôme de degré 2 à coefficient dominant strictement positif (tant que $p_{k-1} \neq 0$).

$$\underline{\beta_{k-1} = -\frac{\langle Ap_{k-1}, r_k \rangle}{\langle Ap_{k-1}, p_{k-1} \rangle}} \qquad (\Leftrightarrow P_k'(\beta_{k-1}) = 0).$$

- Les directions sont conjuguées et les résidus consécutifs sont orthogonaux :

$$\langle p_{k+1}, Ap_k \rangle = \langle r_{k+1} + \beta_k p_k, Ap_k \rangle = \qquad (3)$$
$$= \langle r_{k+1}, Ap_k \rangle + \beta_k \langle p_k, Ap_k \rangle = 0$$

$$\langle r_{k+1}, r_k \rangle = \langle r_k - \alpha_k Ap_k, r_k \rangle =$$
$$= \|r_k\|^2 - \frac{\langle p_k, r_k \rangle}{\langle Ap_k, p_k \rangle} \langle Ap_k, r_k \rangle$$
$$\overset{\text{d'ap.}(2)}{=} \|r_k\|^2 \left[1 - \frac{\langle Ap_k, p_k \rangle}{\langle Ap_k, p_k \rangle}\right]$$
$$= 0 \qquad (4)$$

car $\langle Ap_k, p_k \rangle = \langle Ap_k, r_k + \beta_{k-1} p_{k-1} \rangle \underset{(3)}{=} \langle Ap_k, r_k \rangle$

- Expression de β_k en fonction de r_{k+1} et r_k :

$$\beta_k = -\frac{\langle Ap_k, r_{k+1} \rangle}{\langle Ap_k, p_k \rangle} = -\frac{1}{\alpha_k} \frac{\langle r_{k+1} - r_k, r_{k+1} \rangle}{\langle Ap_k, p_k \rangle}$$

$Ap_k = (r_{k+1} - r_k)/\alpha_k$.

$$= \frac{\langle Ap_k, p_k \rangle}{\langle p_k, r_k \rangle} \frac{\|r_{k+1}\|^2 - \langle r_k, r_{k+1} \rangle}{\langle Ap_k, p_k \rangle}.$$

$$\boxed{\beta_k \underset{(2)\,\&(4)}{=} \frac{\|r_{k+1}\|^2}{\|r_k\|^2}}$$

<u>Définition</u>. Pour A sdp, l'algorithme itératif

$$\begin{cases} r_0 = b - Ax_0 \ , \ p_0 = r_0 \ , \ (x_0 \text{ donné dans } \mathbb{R}^n) \\ x_{k+1} = x_k + \alpha_k p_k \ , \ r_{k+1} = r_k - \alpha_k A p_k \ , \ p_{k+1} = r_{k+1} + \beta_k p_k , \\ \alpha_k = \dfrac{\|r_k\|^2}{\langle Ap_k , p_k \rangle} \ , \ \beta_k = \dfrac{\|r_{k+1}\|^2}{\|r_k\|^2} \\ k \in \mathbb{N} , \text{ tant que } r_k \neq 0 . \end{cases}$$

est appelé <u>méthode du gradient conjugué</u>.

<u>Thm</u>. Si $A \in M_n(\mathbb{R})$ est sdp, alors la méthode du gradient converge en au plus n itérations.

<u>Lemme</u>. Pour chaque j et k tels que $j \neq k$, on a

$$\langle r_j , r_k \rangle = 0$$

<u>Preuve</u> (du lemme) On montre par récurrence sur k que $\langle r_j , r_k \rangle = 0 = \langle Ap_j , p_k \rangle \quad \forall j \in [0, \dots, k-1]$.

Initialisation : $\langle r_0 , r_1 \rangle = 0 = \langle Ap_0 , p_1 \rangle$ d'après (3) et (4)

Hérédité : On suppose la proposition vraie au rang k.

Si $j = k$, on a bien $\langle r_k , r_{k+1} \rangle = 0 = \langle Ap_k , p_{k+1} \rangle$ par (3) et (4). Si $j < k$,

$$\begin{aligned} \langle r_j , r_{k+1} \rangle &= \langle r_j , r_k \rangle - \alpha_k \langle r_j , Ap_k \rangle \\ &= \begin{cases} -\alpha_k \langle p_j - \beta_{j-1} p_{j-1} , Ap_k \rangle & \text{si } j > 0 \\ -\alpha_k \langle p_0 , Ap_k \rangle & \text{si } j = 0 \end{cases} \\ &\overset{HR}{=} 0 \end{aligned} \qquad (5)$$

$$\begin{aligned} \langle Ap_j , p_{k+1} \rangle &= \langle Ap_j , r_{k+1} \rangle + \beta_k \underbrace{\langle Ap_j , p_k \rangle}_{\overset{HR}{=} 0} \\ &= \frac{1}{\alpha_j} \langle r_j - r_{j+1} , r_{k+1} \rangle \\ &\underset{(5)}{=} 0 \end{aligned}$$

<u>Preuve</u> (du thm). S'il existe $k_0 < n$ tel que $r_{k_0} = 0$, alors $r_k = 0$ pour $k \geq 0$.

Sinon, $r_k \neq 0$ pour $k = 0, \ldots, n-1$. Or d'après le lemme r_n est orthogonal à $\mathrm{Vect}(r_0, \ldots, r_{n-1})$, qui est égal à E car $r_0, \ldots, r_{n-1}$ sont tous non nul et orthogonaux et forment donc une famille libre*. Donc $r_n = 0$ et $r_k = 0$ si $k \geq n$.

Ainsi, la suite (r_k) stationne en 0 à partir d'un rang $k_0 \leq n$. ~~et on a~~

<u>Conclusion</u> : On obtient x_{k_0} tel que $Ax_{k_0} = b$ (i.e. $r_{k_0} = 0$) en au plus n itérations.

<u>RAPPEL</u>. [Rouvière, PGCD, 3e éd]

Une fonction est convexe (: en dessous des cordes)
- ssi $f(y) - f(x) \geq Df(x).(y-x)$ (: au dessus des tg)
  (Considérer $g : t \mapsto f(ty + (1-t)x)$.) [E42]
- ssi $f' \nearrow$ [E42].
- ssi $D^2 f$ est une f.q.d.p en tt pt. [E108].
  (E42 + Taylor-Young).
- alors : pt critique $\Rightarrow$ min. global [E119]
  (corollaire de E42).

<u>RÉFÉRENCE</u> : Analyse numérique des équations aux dérivées partielles, Laurent Di Menza, CASSINI. [7.3. Méth. de grad.]
(Voir [Analyse numérique ; Héron, Issard-Roch, Picard] pour une preuve plus courte, mais non constructive.)

---

* Si $\sum \lambda_i\, i = 0$ alors $0 = \langle 0, j \rangle = \langle \sum \lambda_i\, i, j \rangle = \lambda_j \| j \|^2$

Développement n° 2

—. Formule du rayon spectral. —

__Lemme__. Soit A une matrice $n \times n$. Si v_j, resp u_i, est un vecteur propre à gauche, resp. à droite, de A associé à la valeur propre λ_j (${}^t v_j A = \lambda_j {}^t v_j$), resp. λ_i, avec $\lambda_i \neq \lambda_j$, alors ${}^t v_j u_i = 0$.

__Preuve__. On a

$$\lambda_j {}^t v_j u_i = {}^t v_j A u_i = {}^t v_j \lambda_i u_i = \lambda_i {}^t v_j u_i ,$$

soit $(\lambda_j - \lambda_i) {}^t v_j u_i = 0$ et donc ${}^t v_j u_i = 0$ puisque $\lambda_j \neq \lambda_i$.

__Proposition__ (Formule du rayon spectral). Soit A une matrice $n \times n$, diagonalisable dont la valeur propre de plus grand module λ_1 est unique. Soit $q^{(0)} \in \mathbb{R}^n$ unitaire et non orthogonal au sous-espace propre à gauche associé à λ_1. Alors la suite définie par

$$x^{(k)} = A q^{(k-1)}, \quad k \geqslant 1 \quad \text{et} \quad q^{(k)} = \frac{x^{(k)}}{\|x^{(k)}\|}$$

vérifie

1. $q = \lim_{k \to +\infty} \left(\frac{\lambda_1}{|\lambda_1|} \right)^k q^{(k)}$ est un vecteur propre de norme 1 associé à λ_1;

2. $\lim_{k \to +\infty} \| A q^{(k)} \| = |\lambda_1|$;

3. ~~Si $q_j^{(k)}$ $(1 \leqslant j \leqslant n)$ est~~
$$\lim_{k \to +\infty} \frac{x_j^{(k+1)}}{q_j^{(k)}} = \lambda_1 .$$

<u>Preuve</u>. A étant diagonalisable, on note $u_1, \dots, u_n$ une base de vecteurs propre de A associés resp. aux valeurs propre $\lambda_1, \dots, \lambda_n$. On note aussi $q^{(0)} = \sum_{i=1}^{n} \alpha_i u_i$ et $v_1, \dots, v_n$ les vecteurs propres à gauche de A vérifiant ${}^t v_i u_i = 1$.*

Montrons par récurrence que $\boxed{q^{(k)} = A^k q^{(0)} / \|A^k q^{(0)}\|}$ (1)

— Initialisation : $q^{(0)} = \dfrac{q^{(0)}}{\|q^{(0)}\|}$ (car unitaire)

<u>ou</u> $q^{(1)} = \dfrac{A q^{(0)}}{\|A q^{(0)}\|}$ $\left(= \dfrac{x^{(1)}}{\|x^{(1)}\|} \right)$.

— Hérédité : On suppose la formule vraie au rang k.

$$q^{(k+1)} = \frac{A q^{(k)}}{\|A q^{(k)}\|} = A \frac{A^k q^{(0)}}{\|A^k q^{(0)}\|} \left(\frac{\|A^{k+1} q^{(0)}\|}{\|A^k q^{(0)}\|} \right)^{-1} =$$

$$= \frac{A^{k+1} q^{(0)}}{\|A^{k+1} q^{(0)}\|}.$$

<u>Cas où λ_1 est une valeur propre simple</u>.

$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$.

• Grâce au lemme, $\alpha_1 = {}^t v_1 q^{(0)}$ et donc <u>α_1 est non nul</u> par hypothèse sur $q^{(0)}$.

• ~~Calcul~~ Réécriture de $A^k q^{(0)}$: $A^k q^{(0)} = \sum_{i=1}^{n} \alpha_i \lambda_i^k u_i$ et comme $\alpha_1 \neq 0$,

$$A^k q^{(0)} = \alpha_1 \lambda_1^k \Big(u_1 + \underbrace{\sum_{i=2}^{n} \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1} \right)^k u_i}_{=: e^{(k)}} \Big) \qquad \boxed{A^k q^{(0)} = \alpha_1 \lambda_1^k (u_1 + e^{(k)})}$$

$e^{(k)} = O\left(\left(\frac{\lambda_2}{\lambda_1} \right)^k \right)$ car $|\lambda_i| < |\lambda_1|$ pour $i \geq 2$.

$\xrightarrow[k \to +\infty]{} 0$

• Réécriture de $\|A^k q^{(0)}\|$: Par inégalité triangulaire

$$\big| \|u_1 + e^{(k)}\| - \|u_1\| \big| \leq \|e^{(k)}\|$$

---

* $\det({}^tA - \lambda I_n) = \det({}^t(A - \lambda I_n)) = \det(A - \lambda I_n)$ donc $Sp(A) = Sp({}^tA)$; or ${}^tvA = \lambda {}^tv \iff {}^tAv = \lambda v$, ainsi les val. propres à gauche et à droites sont les mêmes. (Il suffit ensuite de quotienter)

Donc

$$e^{(k)} = \mathcal{O}\left(\left(\frac{\lambda_2}{\lambda_1}\right)^k\right) \Longrightarrow \|u_1 + e^{(k)}\| = \|u_1\| + \varepsilon_k$$

$$\text{où } \varepsilon_k = \mathcal{O}\left(\left(\frac{\lambda_2}{\lambda_1}\right)^k\right).$$

Il vient

$$\boxed{\|A^k q^{(0)}\| = |\alpha_1|\,|\lambda_1|^k\,\|u_1 + e^{(k)}\| = |\alpha_1|\,|\lambda_1|^k\left(\|u_1\| + \varepsilon_k\right)}.$$

- Point 1 :

$$\left(\frac{\lambda_1}{|\lambda_1|}\right)^k q^{(k)} = \frac{\alpha_1 \cancel{\lambda_1^{2k}}\,(u_1 + e^{(k)})}{|\alpha_1|\,\cancel{|\lambda_1|^{2k}}\,(\|u_1\| + \varepsilon_k)} \xrightarrow[k\to+\infty]{} \frac{\alpha_1}{|\alpha_1|}\,\frac{u_1}{\|u_1\|}.$$

- Point 2 :

$$\|A\,q^{(k)}\| = \frac{\|A^{k+1} q^{(0)}\|}{\|A^k q^{(0)}\|} = \frac{|\alpha_1|\,|\lambda_1|^{k+1}}{|\alpha_1|\,|\lambda_1|} \times \underbrace{\frac{\|u_1\| + \varepsilon_{k+1}}{\|u_1\| + \varepsilon_k}}_{\xrightarrow[k\to+\infty]{} 1} \xrightarrow[k\to+\infty]{} |\lambda_1|$$

- Point 3 : Si $(u_1)_j \neq 0$, $(u_1)_j + e_j^{(k+1)} \neq 0$ pour k assez grand et dans ce cas

$$\frac{(A q^{(k)})_j}{q_j^{(k)}} = \frac{(A^{k+1} q^{(0)})_j}{(A^k q^{(0)})_j} = \frac{\alpha_1 \lambda_1^{k+1}}{\alpha_1 \lambda_1^k} \cdot \frac{(u_1)_j + e_j^{(k+1)}}{(u_1)_j + e_j^{(k)}} \xrightarrow[k\to+\infty]{} \lambda_1.$$

Cas où λ_1 est une valeur propre multiple.

$$\lambda_1 = \lambda_2 = \dots = \lambda_p \text{ et } |\lambda_1| > |\lambda_{p+1}| \geq \dots \geq |\lambda_n|.$$

Même schéma en écrivant

$$A^k q^{(0)} = \lambda_1^k \Bigg(\underbrace{\sum_{i=1}^{p} \alpha_i u_i}_{=:\, u} + \underbrace{\sum_{i=p+1}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k u_i}_{=:\, e^{(k)}} \Bigg).$$

Par hypothèse sur $q^{(0)}$, ~~au moins un α_i,~~ il existe $i_0 \in [1, \dots, p]$ tel que $\alpha_{i_0} \neq 0$, et donc u est un vecteur propre de A associé à λ_1. De plus $e^{(k)} = \mathcal{O}\left(\left(\frac{\lambda_{p+1}}{\lambda_1}\right)^k\right) \xrightarrow[k\to+\infty]{} 0$

Exemple • $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ admet 1 comme valeur propre double. Le vecteur propre unitaire associé est $u = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.
Soit $q^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. On a $\|q^{(0)}\|_\infty = 1$ et $\underbrace{A^k}_{\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}} q^{(0)} = \begin{pmatrix} k+1 \\ 1 \end{pmatrix}$.

D'où, d'après (1)

$$q_k = \frac{A^k q^{(0)}}{\|A^k q^{(0)}\|} = \begin{pmatrix} 1 \\ 1/k+1 \end{pmatrix}.$$

Alors $$A q_k = \begin{pmatrix} 1 + \frac{1}{1+k} \\ \frac{1}{1+k} \end{pmatrix} \quad \text{et} \quad \|A q_k\|_\infty = 1 + \frac{1}{1+k}.$$

Pour $\varepsilon > 0$, il faut $k \geqslant \frac{1}{\varepsilon} - 1$ itérations pour avoir

$$\left| \|A q^{(k)}\|_\infty - \underbrace{1}_{\|u\|_\infty} \right| < \varepsilon.$$

• Soit $P(X) = x^n + a_{n-1} x^{n-1} + \cdots + a_0$, la racine de plus grand module est unique. Alors la méthode de la puissance itérée permet de la détermin. en l'appliquant à la matrice compagnon

$$C_P = \begin{pmatrix} 0 & \cdots & 0 & -a_0 \\ +1 & \ddots & \vdots & -a_1 \\ 0 & \ddots & 0 & \vdots \\ 0 & \cdots 0 & +1 & -a_{n-1} \end{pmatrix},$$

car $\det(A - XI_n) = P(X)$.
(lien avec la méthode de Bernouilli).

RÉFÉRENCE : [L3] Mathématiques appliquées L3 ; PEARSON.
(I.1.VI.3. Calcul de la plus gde v.p.)
[LT] Analyse numérique matricielle appliquée à l'art de l'ingénieur ; Tome 2 : Méthodes itératives ; P. Lascaux, R. Théodor ; 2e éd ; MASSON.
{10.1. La méthode de la puissance} (Pour les exemples et des cas plus gal).