© *J.-B. A. K. <jean-baptiste.apoung@math.u-psud.fr>*

Exercices d'entraînement

IMPORTANT-

Les examens finaux (première et seconde session) cette année seront exclusivement théoriques. C'est-à-dire, ils n'auront pas de volet mise en oeuvre sur machine. Aussi les exercices qui suivent sont fournis afin de vous y préparer. **Des indications** (priorité, pertinence, solution, etc) **sur ces exercices seront fournies en séances de cours.** Ces exercices sont en **majorité** des exercices d'apprentissage du cours, c'est-à-dire qu'ils peuvent être résolus entièrement avec les notes de cours sous la main. Le **Thème 4** porte sur des exercices type examen : ils ne sont pas fournis ici puisqu'ils sont laissés sous leur format original et sont accessibles sous Dokeos. Le **Thème 5** n'est pas obligatoire n'hésitez pas à m'écrire en cas de soucis à l'adresse indiquée ci-dessus.

Thème - 1 *Sur le chapitre I du cours*

Exercice-1 **:** **Résolution de l'équation de la chaleur sur un domaine borné** $[0, L]$

On s'intéresse au problème aux limites suivant, dans lequel u_0, g_0, g_L, f sont des fonctions de classe C^2 et ou $\alpha > 0$:

$$(\mathcal{P}) \begin{cases} \dfrac{\partial u}{\partial t}(t,x) = \alpha \dfrac{\partial^2 u}{\partial x^2}(t,x) + f(t,x), & t > 0, \quad x \in]0, L[, & (1) \\ u(t,0) = g_0(t), & t > 0, & (2) \\ u(t,L) = g_L(t), & t > 0, & (3) \\ u(0,x) = u_0(x), & x \in]0, L[. & (4) \end{cases}$$

avec les relations $u_0(0) = g_0(0), u_0(L) = g_L(0)$ qui assurent la compatibilité de la condition initiale et des conditions aux limites.

Q-1 **:** **Relèvement des conditions aux limites**

a. Montrer que l'on peut trouver une fonction $\bar{u}(t,x)$ affine en x telle que

$$\bar{u}(t,0) = g_0(t) \text{ et } \bar{u}(t,L) = g_L(t) \quad \forall t > 0$$

b. Montrer que la fonction $\bar{u}$ est de classe C^2 sur $\mathbb{R} \times [0, L]$.

Q-2 **:** **Problème équivalent avec conditions aux limites homogènes**

En effectuant le changement d'inconnue $v(t,x) = u(t,x) - \bar{u}(t,x)$, montrer que le problème $(\mathcal{P})$ est équivalent au problème suivant

$$(\mathcal{H}) \begin{cases} \dfrac{\partial v}{\partial t}(t,x) = \alpha \dfrac{\partial^2 v}{\partial x^2}(t,x) + \bar{f}(t,x), & t > 0, \quad x \in]0, L[, & (1) \\ v(t,0) = v(t,L) = 0, & t > 0, & (2) \\ v(0,x) = v_0(x), & x \in]0, L[. & (3) \end{cases}$$

où les fonctions de classe C^2, v_0 et $\bar{f}$ sont à préciser.

Q-3 **:** **Problème homogène sans terme source : c'est-à-dire** $\bar{f} \equiv 0$.

On considère le problème $(\mathcal{H})$ avec $\bar{f} = 0$

a. En recherchant la solution de (1) sous la forme $v(t,x) = \varphi(x)\psi(t)$, montrer que

$$v_k(t,x) = exp\left(-k^2\pi^2\alpha t/L^2\right)\sin\left(k\pi x/L\right)$$

est solution de (1), (2), pour tout $k \in \mathbb{N}$.

b. On pose formellement

$$v(t,x) = \sum_{k=0}^{+\infty} B_k\, v_k(t,x)$$

pour $x \in]0,L[$, $t \geq 0$. Ecrire la relation vérifiée par les coefficients B_k pour que v vérifie, (1), (2) et (3). En déduire que les coefficients B_k sont les coefficients de Fourier de la fonction notée V_0, prolongement impair $2L-$périodique de v_0 sur $\mathbb{R}$.

c. On définit la fonction E par $E(t) = \|v(t,\cdot)\|^2_{L^2(]0,L[)}$, pour $t \geq 0$. Montrer que si v est solution de $(\mathcal{P})$, alors $E(t) \leq E(0)$, $\forall t \geq 0$. En déduire que le problème $(\mathcal{H})$ admet une solution unique.

d. Montrer qu'il existe $K > 0$ et $\beta > 0$ telles que la solution du problème $(\mathcal{H})$ vérifie l'estimation

$$(\mathcal{J}) \qquad \max_{x\in]0,L[} |v(t,x)| \leq K\, e^{-\beta t},\ t > 0.$$

Q-4 : **Prise en compte du terme source :** $\bar{f} \neq 0$

On pose formellement

$$v(t,x) = \sum_{k=0}^{+\infty} B_k(t)\, v_k(t,x)$$

pour $x \in]0,L[$, $t \geq 0$. Où les $v_k(t,x)$ sont définies ci-dessus.

(*Attirons l'attention sur le fait que ce choix est moins général que celui couramment adopté :* $v(t,x) = \sum_{k=0}^{+\infty} B_k(t)\, \sin\left(k\pi x/L\right)$.

Poursuivons néanmoins avec notre choix.)

a. Montrer que pour tout $k \in \mathbb{N}$, $B_k(t)$ est solution d'une équation différentielle ordinaire du premier ordre, dont on précisera les conditions initiales et le terme source en fonction des coefficients de Fourrier des prolongements impair $2L-$périodique de v_0 et $\bar{f}(t,\cdot)$ sur $\mathbb{R}$.

b. Déterminer les $B_k(t)$ et donner l'expression de la solution $v(t,x)$ du problème $(\mathcal{H})$.

c. Montrer que la solution du problème $(\mathcal{H})$ est unique et donner un résultat analogue à la formule $(\mathcal{J})$ de 3.d.

Q-5 : **Solution du problème** $(\mathcal{P})$

a. Montrer que si le problème $(\mathcal{P})$ admet une solution, celle-ci est unique.

b. A l'aide des questions précédentes, donner l'expression de la solution du problème $(\mathcal{P})$.

c. Donner un résultat analogue à la formule $(\mathcal{J})$ de 3.d.

Exercice-2 : **Analyse d'une classe de schémas explicites pour l'équation de la chaleur**

Les notations que nous utiliserons et qui ne seront pas explicitement renseignées devront être comprises dans le sens donné en cours.

On considère un schéma numérique linéaire à un pas

$$\left\{\begin{array}{rcl} (P_{k,h}v)_j^n &=& 0, \quad n \in \mathbb{N}, j \in \mathbb{Z}, \\ v_j^0 &=& u_0(jh), \quad j \in \mathbb{Z}, \end{array}\right. \tag{1}$$

introduit pour la résolution numérique de l'équation

$$\left\{\begin{array}{rcl} (Pu)(t,x) &=& 0, \quad t > 0, x \in \mathbb{R} \\ u(0,x) &=& u_0(x) \qquad x \in \mathbb{R}, \end{array}\right. \tag{2}$$

sur une grille de pas h en espace et k en temps. Ici , $v_j^n, j \in \mathbb{Z}, n \in \mathbb{N}$ est une approximation de la valeur $u(t^n, x_j)$ de u au point $x_j = jh$, à l'instant $t^n = nk$.

Dans toute la suite de l'exercice, on prendra $Pu(t,x) = \dfrac{\partial u}{\partial t}(t,x) - D\dfrac{\partial^2 u}{\partial x^2}(t,x)$. avec $D > 0$.

Un exemple de schéma numérique est le schéma d'Euler explicite en temps associé au schéma à trois points en espace. Dans ce cas on a :

$$(P_{k,h}v)_j^n = \frac{1}{k}(v_j^{n+1} - v_j^n) + \frac{D}{h^2}(-v_{j-1}^n + 2v_j^n - v_{j+1}^n) = 0 \tag{3}$$

Nous nous plaçons cependant dans un cadre plus générale et considérons le schéma écrit sous la forme

$$\left\{ \begin{array}{rcl} v_j^{n+1} & = & \alpha v_{j-1}^n + \beta\ v_j^n + \gamma v_{j+1}^n, \quad n \in \mathbb{N}, j \in \mathbb{Z}, \\ v_j^0 & & \text{donné} \quad , \ j \in \mathbb{Z}, \end{array} \right. \tag{4}$$

où $\alpha, \beta\ , \gamma$ sont des réels ne dépendant que de k, h et D. On rappelle dans ce cas que

$$(P_{k,h}v)_j^n = \frac{1}{k}\left(v_j^{n+1} - \alpha v_{j-1}^n - \beta\ v_j^n - \gamma v_{j+1}^n\right) \tag{5}$$

Remarque 1.

Bien qu'on se soit limité ici à l'équation de la chaleur, la démarche qui suit s'adapte aussi à une équation de type convection diffusion réaction où l'on a

$$Pu(t,x) = \frac{\partial u}{\partial t}(t,x) - D\,\frac{\partial^2 u}{\partial x^2}(t,x) + V\,\frac{\partial u}{\partial x}(t,x) + R\,u(t,x),$$

avec $R\ \in \mathbb{R}^+$ et $V\ \in \mathbb{R}$.

Autour de la consistance

La consistance de ce schéma a été vue en cours lorsque le problème est posé sur un domaine borné. Nous n'y reviendrons plus. Voir par contre le **Thème -5**, *pour une utilisation plus élaborée de la notion de consistance dans le cadre de l'équation de transport.*

Autour de la stabilité

Le fait que le problème soit posé sur tout $\mathbb{R}$ nous incite à adopter une autre approche plus avantageuse dans l'étude de la stabilité du schéma. C'est l'analyse de la stabilité (l^2) par transformée de Fourier encore appelée stabilité au sens de Von-Neumann. L'objectif de cette partie est de se familiariser avec cette notion à travers un exemple simple.

Définition 2.

On dit que le schéma aux différences finies $P_{k,h}v = 0$ est stable au sens V si, $\forall T > 0$,

i) $v^n \in V$ pour tout n tel que $nk < T$.

ii) $\exists C_T > 0$ tel que $\sup_{nk<T} \|v^n\|_V \leq C_T \|v^0\|_V$.

Remarque 3.

Dans la pratique on s'intéresse aux stabilités l^2 et l^∞. Dans ces cas l'espace V sera :

- Cas l^∞

$$l^\infty(\mathbb{Z}) = \{v = (v_j)_{j\in\mathbb{Z}} : \sup_{j\in\mathbb{Z}} |v_j| < \infty\} \quad \text{muni de la norme} \quad \|v\|_{l^\infty(\mathbb{Z})} = \sup_{j\in\mathbb{Z}} |v_j|$$

- Cas l^2

$$l^2(h\mathbb{Z}) = \{v = (v_j)_{j\in\mathbb{Z}} : \sum_{j\in\mathbb{Z}} h|v_j|^2 < \infty\} \quad \text{muni de la norme} \quad \|v\|_{l^2(h\mathbb{Z})} = \sqrt{\sum_{j\in\mathbb{Z}} h|v_j|^2}.$$

Q-1 : Etude abstraite dans un espace V

On suppose que le schéma $P_{k,h}v = 0$ est tel que :

$$\exists\, \mathbf{C} \geq \mathbf{0} \quad \text{tel que} \quad \|v^{n+1}\|_V \leq (1 + Ck)\, \|v^n\|_V \quad \forall v \in \mathbb{N}. \tag{6}$$

Q-1-1 : Montrer que le schéma est stable au sens V. *On remarquera que :* $1 + x \leq e^x \quad \forall x \geq 0$.

Q-1-2 : Transcrire la relation (6) dans les cas de stabilité l^2 et l^∞.

Q-2 : Stabilité dans le cas l^∞

On considère le schéma (4). On suppose que :

- $\alpha + \beta \ + \gamma = 1$, on dit que le schéma **préserve les constantes**.
- $\alpha \geq 0, \beta \ \geq 0, \gamma \geq 0$, le schéma est dit **monotone** : $(v_j^n > w_j^n \quad \forall j \in \mathbb{Z}) \Longrightarrow (v_j^{n+1} > w_j^{n+1}\ \forall j \in \mathbb{Z})$.

Q-2-1 : Montrer que $\forall n \in \mathbb{N}, \forall j \in \mathbb{Z}$

$$\begin{aligned} -\alpha\|v^n\|_{l^\infty(\mathbb{Z})} \leq \alpha v_{j-1}^n \leq \alpha\|v^n\|_{l^\infty(\mathbb{Z})} \\ -\beta \ \|v^n\|_{l^\infty(\mathbb{Z})} \leq \beta \ v_j^n \leq \beta \ \|v^n\|_{l^\infty(\mathbb{Z})} \\ -\gamma\, \|v^n\|_{l^\infty(\mathbb{Z})} \leq \gamma\, v_{j+1}^n \leq \gamma\, \|v^n\|_{l^\infty(\mathbb{Z})} \end{aligned}$$

Q-2-2 : En déduire que le schéma est stable l^∞.

Q-2-3 : Déterminer alors les conditions suffisantes sur k, h, D, pour que le schéma (3) soit stable l^∞.

Q-3 : Stabilité dans le cas l^2. Commençons par quelques définitions nécessaires.

Définition 4 (Transformée de Fourier discrète)**.**

On désigne par transformation de Fourier discrète l'application $\widehat{\cdot}$ définie par

$$\begin{array}{rccc} \widehat{\cdot} \ : & l^2(h\mathbb{Z}) & \to & L^2\left(\left]-\dfrac{\pi}{h}, \dfrac{\pi}{h}\right[\right) \\ & v & \mapsto & \widehat{v}, \end{array}$$

avec

$$\widehat{v}(\xi) = \sum_{j\in\mathbb{Z}} e^{-ijh\xi} v_j h, \quad \xi \in \left[-\frac{\pi}{h}, \frac{\pi}{h}\right].$$

Remarque 5 (Formule d'inversion de la transformée de Fourier discrète)**.**

Il est également possible de reconstruire la suite v si $\widehat{v}$ est connue sur $[-\pi/h, \pi/h]$: si $j \in \mathbb{Z}$ est fixé, on multiplie en effet l'équation par $e^{ijh\xi}$ et on intègre sur l'intervalle $[-\pi/h, \pi/h]$. On obtient

$$\int_{-\pi/h}^{\pi/h} \widehat{v}(\xi)\ d\xi = h \int_{-\pi/h}^{\pi/h} v_j\ d\xi + h \sum_{k \neq j} \int_{-\pi/h}^{\pi/h} e^{-i(j-k)h\xi} v_k\ d\xi.$$

Par périodicité, on obtient

$$v_j = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} e^{-ijh\xi}\ \widehat{v}(\xi)\ d\xi.$$

Proposition 6 (Formule de Parseval)**.**

Pour $v \in l^2(h\mathbb{Z})$, on a

$$\frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} |\widehat{v}(\xi)|^2\ d\xi = \|v\|_h^2.$$

Q-3-1 : Pour tout $l \in \mathbb{Z}$ et $v \in l^2(h\mathbb{Z})$ on définit la suite translatée $\tau_l v$ par $(\tau_l v)_j := v_{j+l}, \quad j \in \mathbb{Z}$. Montrer que

$$\widehat{\tau_l v}(\xi) = e^{ilh\xi} \widehat{v}(\xi), \xi \in \left[-\frac{\pi}{h}, \frac{\pi}{h}\right]. \tag{7}$$

Q-3-2 : En appliquant la transformation de Fourier discrète au schéma (4) montrer que

$$\widehat{v}^{n+1}(\xi) = g(h\xi)\, \widehat{v}^n(\xi), \forall \xi \in \left[-\frac{\pi}{h}, \frac{\pi}{h}\right], \forall n \in \mathbb{N}. \tag{8}$$

où la fonction g à valeur dans $\mathbb{C}$ appelée **facteur d'amplification**, est donnée dans ce cas par :

$$g(\xi h) = \alpha\, e^{-i\xi h} + \beta\ + \gamma\, e^{i\xi h} \quad , \forall \xi \in \left[-\frac{\pi}{h}, \frac{\pi}{h}\right]. \tag{9}$$

Q-3-3 : En déduire une condition nécessaire et suffisante sur g assurant la stabilité l^2 du schéma (4).

Q-3-4 : Montrer que si $\alpha + \beta\ + \gamma = 1$, on a :

$$|\alpha e^{-i\theta} + \beta\ + \gamma e^{+i\theta}|^2 = 1 - 4\Big(\beta\ (\alpha + \gamma) + 4\alpha\gamma\Big) \sin^2(\frac{\theta}{2}) + m\alpha\gamma \sin^4(\frac{\theta}{2}), \quad \forall \theta \in [-\pi, \pi]$$

où m est un entier positif à déterminer.

Q-3-5 : Etudier alors la stabilité l^2 du schémas (3).

Thème - 2 *Sur le chapitre II du cours*

Exercice-1 **:** **Stockage de matrices creuses**

Pour chacune des matrices suivantes, donner leur stockage **COO**, **CSR**, **CSC**, et **bande** si elle est sous forme bande.

$$A = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 3 & 4 & 0 & 0 \\ 0 & 5 & 6 & 0 \\ 0 & 0 & 7 & 8 \end{bmatrix} \qquad A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \qquad A = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}$$

Exercice-2 : **Renumérotation de matrices creuses : apprentissage du cours**

Q-1 : Relation entrer les renumérotations *Cuthill-Mackey* (**CM**) et celle inversée *Reverse Cuthill-Mackey* (**RCM**)

Soit A une matrice carrée d'ordre n, et B_{cm} respectivement B_{rcm} les matrices obtenues par renumérotation CM respectivement RCM de A.

Soit p_{cm} respectivement p_{rcm} les vecteurs de permutation associés à la renumérotation CM respectivement RCM d'une matrice A. C'est-à-dire

$$\begin{array}{l} B_{cm}(i,j) = A(p_{cm}(i), p_{cm}(j)) \\ B_{rcm}(i,j) = A(p_{rcm}(i), p_{rcm}(j)) \end{array} \qquad 1 \leq i,j \leq n$$

On pose ip_{cm} respectivement ip_{rcm} les *vecteurs de permutation inverse* respectifs de p_{cm} et p_{rcm}. C'est-à-dire $ip_{cm}(p_{cm}(k)) = k, \quad 1 \leq k \leq n$.

Q-1-1 : En regardant le cours, donner la relation entre p_{rcm} et p_{cm}.
Q-1-2 : A-t-on une relation analogue à la précédente entre ip_{rcm} et ip_{cm} ?

Q-2 : Illustrations

Soit la matrice suivante :

$$A = \begin{bmatrix} 1 & 0 & 0 & 2 & 0 & 3 \\ 0 & 4 & 0 & 0 & 5 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 7 & 0 & 0 & 8 & 0 & 9 \\ 0 & 10 & 0 & 0 & 11 & 0 \\ 12 & 0 & 0 & 13 & 0 & 15 \end{bmatrix}$$

Q-2-1 : Pour la matrice A ci-dessus, donner sa matrice B obtenue par renumérotation **CM** et **RCM**.

Q-2-2 : Préciser pour chacune des numérotation ci-dessus la matrice de permutation P telle que $B = P\,A\,P^T$

Q-2-3 : Pour chacune des matrices de permutation P obtenue ci-dessus, proposer sont stockage optimisé au moyen d'un vecteur. On justifiera le choix du stockage (directe ou inverse voir cours).

Q-2-4 : Fournir un stockage bande de la matrice B. Ce stockage bande est-il avantageux ici ?
On regardera le nombre de zéros supplémentaires à stocker pour satisfaire le stockage bande.

Q-2-5 : Vérifier que l'inverse de la matrice B peut-être obtenu par action sur les blocs diagonaux et ceci sans aucun remplissage. En déduire un stockage possible de cette matrice.

Q-2-6 : On pose $p = [6, 3, 5, 1, 2, 4]$ et on définit la matrice C par : $C(p(i), p(j)) = A(i,j), 1 \leq i,j \leq 6$.

1. Calculer C. Est-il bénéfique de stocker C sous forme bande ?
2. Appliquer l'algorithme de décomposition LU à la matrice C et vérifier que le *phénomène de remplissage* (voir cours) n'est pas produit.

Exercice-3 : **Prise en compte des complexités**

Soit $A \in \mathcal{M}_{n,n}(\mathbb{R})$ et $B \in \mathcal{M}_{n,n}(\mathbb{R})$ deux matrices symétriques définies positives telles que $\varrho(A^{-1}B) < 1$. Soit $F \in \mathbb{R}^n$ un vecteur donné et $m \in \mathbb{N}^*$. On se propose de calculer le vecteur U^m le $m+1$-ième terme de la suite $AU^{k+1} = BU^k + F, \quad U^0 = U_0$.
Pour cela on considère les deux algorithmes suivants :

Algorithme 1 (avec résolution répétée)

Initialisation :
U^0 = U_0

Iterations :
Pour k = 0 à m-1
Calculer: V = B U^k + F
Resoudre: A U^{k+1} = V
fin Pour

Algorithme 2 (avec inversion de la matrice)

Initialisation :
C = A^{-1} B
E = A^{-1} F
U^0 = U_0

Iterations :
Pour k = 0 à m-1
Calculer: U^{k+1} = C U^k + E
fin Pour

Q-1 : On considère l'algorithme 1

Q-1-1 : Quels formats de stockage sont mieux adaptés pour les matrices A et B ?

Q-1-2 : Peut-on avoir des soucis d'espace mémoire pour cet algorithme ?

Q-1-3 : Quelle est la complexité de cet algorithme lorsque les matrices A, B sont pleines ?

Q-2 : On considère l'algorithme 2

Q-2-1 : En s'inspirant de l'exercice 1, justifier la non nécessité de calculer explicitement l'inverse A^{-1} de la matrice A pour déterminer C et E.

Q-2-2 : Quelle est la complexité de cet algorithme lorsque les matrices A, B sont pleines ?

Q-2-3 : Sur un exemple simple de matrice de taille 3, montrer que $A^{-1}B$ peut-être pleine alors que A et B sont creuses.

Q-2-4 : Peut-on être confronté à un problème d'espace mémoire lié au stockage dans cet algorithme ?

Q-3 : Conseillez sur le choix de l'un de ces algorithmes, en vous appuyant sur le potentiel problème d'espace mémoire et sur le fait que pour m fixé, on peut se trouver dans les cas $m \gg n$ ou $n \gg m$ (avec $\gg$ *signifiant très grand).*

Thème - 3 *Sur le chapitre III du cours*

Note 1.

Pour chacun des exercices qui suivent, on indiquera à la fin s'il s'agit d'une **méthode de projection ou pas**. Si oui on précisera les **espaces de Krylov** associés et on indiquera si c'est une version avec **redémarrage**.

Exercice-1 : Méthode du gradient à pas fixe (ou de Richardson)

On considère la fonction f définie de $\mathbb{R}^n$ dans $\mathbb{R}$ par $f(x) = \frac{1}{2}(Ax, x) - (b, x)$.

Q-1 : Montrer que x_0 est solution de $Ax = b$ si et seulement si x_0 minimise la fonction f.

Q-2 : On considère la méthode itérative : $\|x_0$ étant un vecteur initial donné, $x_{k+1} = x_k - \alpha \nabla f(x_k),\ k = 1, \ldots$.

Q-2-1 : Donner la matrice d'itération B de cette méthode et conclure que cette méthode converge si et seulement si $0 < \alpha < \frac{2}{\lambda_n}$ où $0 < \lambda_1 \leq \ldots \leq \lambda_n$ sont les valeurs propres de la matrice symétrique définie positive A.

Q-2-2 : Montrer que le meilleur choix de α est : $\alpha_{opt} = \frac{2}{\lambda_n + \lambda_1}$ et qu'alors $\varrho(B) = \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \equiv \frac{cond(A)-1}{cond(A)+1}$.

Exercice-2 : Méthode du gradient à pas variable

Q-1 : On suppose construite une suite $(p_0, p_1, \ldots p_k, \ldots)$ de vecteurs linéairement indépendants. Et on considère la méthode itérative suivante :

x_0 vecteur initial donné,
$x_{k+1} = x_k + \alpha_k p_k$.
Où α_k est choisi tel qu'il réalise le minimum de $f(x_k + \alpha p_k)$.

Q-1-1 : Montrer que $\alpha_k = \frac{(r_k,p_k)}{(Ap_k,p_k)}$, où $r_k = b - Ax_k$, et conclure que $(r_{k+1}, p_k) = 0, \forall k \geq 0$.

Q-1-2 : On pose $E(x_k) = (A(x_k - x), (x_k - x))$ où x est la solution de $Ax = b$.
Montrer que pour la valeur de α_k ci-dessus, on a $E(x_{k+1}) = E(x_k) - \frac{(r_k,p_k)^2}{(Ap_k,p_k)}$.

Q-1-3 : En remarquant que $E(x_k) = (A^{-1}r_k, r_k)$, montrer que $E(x_{k+1}) = E(x_k)\left(1 - \frac{(r_k,p_k)^2}{(A^{-1}r_k,r_k)(Ap_k,p_k)}\right)$.

Q-1-4 : Déduire de la question précédente que $E(x_{k+1}) \leq E(x_k)\left[1 - \frac{1}{cond(A)}\left(\frac{r_k}{\|r_k\|}, \frac{p_k}{\|p_k\|}\right)^2\right]$.
Où $cond(A) = \frac{\lambda_n}{\lambda_1}$ désigne le conditionnement de la matrice A.
on utilisera le fait que $(Ay, y) \leq \lambda_n(y, y), \quad (A^{-1}y, y) \leq \frac{1}{\lambda_1}(y, y) \forall y$.
Conclure qu'une condition suffisante de convergence est de choisir les p_k tels que
$\forall k \geq 0 \quad \left(\frac{r_k}{\|r_k\|}, \frac{p_k}{\|p_k\|}\right) \geq \mu > 0$, où μ est une constante indépendante de k.

Q-1-5 : Déduire de la question précédente que $p_k = r_k, \forall k \geq 0$ est un choix possible assurant la convergence.

Q-2 : **On prend dans cette question** $p_k = r_k \ \ \forall k \geq 0$.

Q-2-1 : Écrire l'algorithme ainsi obtenu.

Q-2-2 : Que devient $E(x_{k+1})$ de la question 1-c ci-dessus ?

Q-2-3 : *On admet l'inégalité de Kantorovich suivante :* $\frac{(Ay,y)(A^{-1}y,y)}{(y,y)^2} \leq \frac{(\lambda_n+\lambda_1)^2}{4\lambda_n\lambda_1} \ \forall y \neq 0$..
Montrer qu'on a alors $E(x_{k+1}) = E(x_k)\left(\frac{cond(A)-1}{cond(A)+1}\right)^2$.

Q-2-4 : Conclure que $\|x_k - x\| \leq \sqrt{\frac{E(x_0)}{\lambda_1}}\left(\frac{cond(A)-1}{cond(A)+1}\right)^k$.

Exercice-3 : **Méthode du gradient conjugué**

Une amélioration de la méthode du gradient à pas variable consiste à choisir les directions p_k telles que x_{k+1} réalise le minimum de f sur $x_0 + [p_0, p_1, \ldots p_k]$. La particularité de la méthode réside dans le fait que ce problème de minimisation globale, doit se réduire au problème de minimisation locale : $\min_{x=x_0+\bar{x}+y, y\in[p_k]} f(x)$, dans lequel $\bar{x} \in [p_0, p_1, \ldots p_{k-1}]$ est connu et issu d'une précédente minimisation. Les deux premières questions ci-dessous expliquent pourquoi on doit choisir les p_k A-conjugués.

Q-1 : Monter que $f(x_0 + \bar{x} + \alpha p_k) = f(x_0 + \bar{x}) + \alpha(p_k, A\bar{x}) + \frac{\alpha^2}{2}(p_k, Ap_k) - \alpha(p_k, r_0)$.

Q-2 : Conclure qu'un moyen de découpler le problème de minimisation globale en une succession de problèmes de minimisation locale consiste à prendre $(p_k, A\bar{x}) = 0$, ce qui équivaut à $(p_k, Ap_i) = 0, \ \forall 0 \leq i \leq k-1$. **On dit dans ce cas que les vecteurs** $p_i, i = 1 \ldots k$ **sont A-conjugués**.

La méthode du gradient conjugué consiste alors en deux points :

- choisir les directions p_k A-conjuguées ; c'est-à-dire $(Ap_k, p_j) = 0 \ \forall 0 \leq j \leq k-1$,
- initialiser $p_0 = r_0$, et prendre p_{k+1} dans le plan contenant r_{k+1} et p_k c'est-à-dire $p_{k+1} = r_{k+1} + \beta_{k+1}p_k$.

Q-3 : Montrer que les vecteurs p_{k+1} et p_k sont A-conjugués si et seulement si $\beta_{k+1} = -\frac{(r_{k+1},Ap_k)}{(p_k,Ap_k)}$.

Q-4 : En remarquant que $(r_k, p_{k-1}) = 0 \; \forall k$, montrer que $(r_k, p_k) = (r_k, r_k) \; \forall k$.
Simplifier alors l'expression de α_k.

Q-5 : En écrivant $r_k = p_k - \beta_k p_{k-1}$, montrer que $(r_{k+1}, r_k) = 0$. (*On utilisera l'expression de* β_k).

Q-6 : En écrivant $Ap_{k-1} = \frac{1}{\alpha_{k-1}}(r_{k-1} - r_k)$, montrer que

$$(Ap_{k-1}, r_k) = -\frac{1}{\alpha_{k-1}}(r_k, r_k) \quad \text{et} \quad (Ap_{k-1}, p_{k-1}) = \frac{1}{\alpha_{k-1}}(r_{k-1}, r_{k-1}).$$

Simplifier alors l'expression de β_{k+1}.

Q-7 : Montrer que $(r_k, r_j) = 0 \; \forall \; 0 \leq j \leq k-1$.

En déduire qu'en arithmétique exacte, l'algorithme du gradient conjugué converge en au plus n itérations, où n est la taille du système.

(*On remarquera que si* $r_k \neq 0, \forall \; 0 \leq k \leq n-1$, *alors* $[r_0, \ldots, r_{n-1}]$ *est une base de* $\mathbb{R}^n$).

Thème - 4 *Exercices tirés des sujets d'examen ou des tests*

Note 2.

Récupérer les sujets d'examen des années antérieures dans le répertoire dédié sur Dokeos.

Thème - 5 Approfondissement : consistance dans l'équation de transport *(ne faire que si on a assez de temps)*

Bien que nous ne soyons concernés que par les problèmes paraboliques dans notre cours, le schéma donné sous la forme (5) intègre aussi une grande classe de schémas pour les problèmes de type transport pure c'est-à-dire

$Pu(t,x) = \partial_t u(t,x) + c\partial_x(t,x)$, avec $c \neq 0$.

En guise d'exemple citons

Remarque 7 (Quelques schémas pour équation de transport écrits sous la forme (1)).

- Décentré à gauche
$$\frac{1}{k}(v_j^{n+1} - v_j^n) + \frac{c}{h}(v_j^n - v_{j-1}^n) = 0.$$
- Décentré à droite
$$\frac{1}{k}(v_j^{n+1} - v_j^n) + \frac{c}{h}(v_{j+1}^n - v_j^n) = 0$$
- Centré
$$\frac{1}{k}(v_j^{n+1} - v_j^n) + \frac{c}{2h}(v_{j+1}^n - v_{j-1}^n) = 0$$
- Lax-Friedrichs
$$\frac{1}{k}\left(v_j^{n+1} - \frac{1}{2}(v_{j+1}^n + v_{j-1}^n)\right) + \frac{c}{2h}(v_{j+1}^n - v_{j-1}^n) = 0.$$
- Lax-Wendroff
$$\frac{1}{k}\left(v_j^{n+1} - v_j^n\right) + \frac{c}{2h}(v_{j+1}^n - v_{j-1}^n) - \frac{c^2 k}{2h^2}(v_{j+1}^n - 2v_j^n + v_{j-1}^n) = 0$$

Ici l'analyse de la stabilité rentre dans le cadre décrit ci-dessus. Mais pour l'analyse de la consistance, on peut aller un peu plus loin. C'est le but de cette partie

Autour de la consistance

Définition 8.

Le schéma aux différences finies $P_{k,h}v = 0$ est consistant d'ordre p en temps et q en espace avec le problème $Pu = 0$ si pour toute fonction régulière $\phi := \phi(t,x)$ telle que $P\phi = 0$, on a

$$P_{k,h}\phi = \mathcal{O}(k^p) + \mathcal{O}(h^q). \tag{10}$$

Le schéma est dit consistant s'il est consistant d'ordre au moins 1 en temps et au moins 1 en espace.

Remarque 9.

Dans la définition précédente $P_{k,h}v$ désigne la suite $(P_{k,h}v)_j^n, j \in \mathbb{Z}, n \in \mathbb{N}$. La relation (10) doit donc être comprise au sens d'une certaine norme. Mais dans la pratique on se contente de la vérifier ponctuellement (en chaque point (t_n, x_j) de la grille).

Q-8 : Montrer que pour le schéma considéré (4), on a

$$(P_{k,h}v)_j^n = \frac{1}{k}\Big(v_j^{n+1} - \alpha v_{j-1}^n - \beta\ v_j^n - \gamma v_{j+1}^n\Big), \quad n \in \mathbb{N}, j \in \mathbb{Z} \tag{11}$$

Q-9 : Développement limité de $P_{k,h}\phi$.
Soit ϕ une fonction de classe C^m sur $\mathbb{R}_+ \times \mathbb{R}$. Montrer que

$$k\,(P_{k,h}\phi)_j^n = (1-\alpha-\beta\ -\gamma)\phi_j^n + \sum_{s=1}^{m-1}\Big(\frac{k^s}{s!}(\partial_{t^s}\phi)_j^n - (\gamma+(-1)^s\alpha)\frac{h^s}{s!}(\partial_{x^s}\phi)_j^n\Big) + \mathcal{O}(k^m) + \mathcal{O}(h^m), \quad n \in \mathbb{N}, j \in \mathbb{Z} \tag{12}$$

Où on a posé $\partial_{x^s}\phi = \frac{\partial^s\phi}{\partial_{x^s}}, \quad \partial_{t^s}\phi = \frac{\partial^s\phi}{\partial_{t^s}}, \quad \forall s \in \mathbb{N}$.

Q-10 : On suppose que ϕ est au moins de classe C^4, et vérifie $\boxed{\partial_t\phi + c\,\partial_x\phi = e}$. On pose $\lambda = \dfrac{k}{h}$.

Q-10-1 : Montrer que

$$\begin{aligned}
&\partial_t\phi = -c\partial_x\phi + e,\\
&\partial_{tt}\phi = c^2\partial_{xx}\phi - c\partial_x e + \partial_t e,\\
&\partial_{ttt}\phi = -c^3\partial_{xxx}\phi + c^3\partial_{xx}e - c\,\partial_{xt}e + \partial_{tt}e.
\end{aligned}$$

Q-10-2 : En déduire que

$$\begin{aligned}
(P_{k,h}\phi)_j^n = \frac{1-\alpha-\beta\ -\gamma}{k}\phi_j^n \quad &- \quad \tfrac{1}{\lambda}\Big(\gamma-\alpha+(c\lambda)\Big)(\partial_x\phi)_j^n\\
&- \quad \tfrac{h}{2\lambda}\Big(\gamma+\alpha-(c\lambda)^2\Big)(\partial_{xx}\phi)_j^n\\
&- \quad \tfrac{h^2}{6\lambda}\Big(\gamma-\alpha+(c\lambda)^3\Big)(\partial_{xxx}\phi)_j^n\\
&+ \quad e_j^n + \tfrac{k}{2}\Big(\partial_t e - c\partial_x e\Big)_j^n + \tfrac{k^2}{6}\Big(\partial_{tt}e - c\partial_{xt}e + c^2\partial_{xx}e\Big)_j^n\\
&+ \quad \mathcal{O}(k^3) + \mathcal{O}(h^3) \quad n \in \mathbb{N}, j \in \mathbb{Z}
\end{aligned} \tag{13}$$

Q-11 : Evaluation de l'ordre de consistance.

Les notations sont celles de la question précédente. $\boxed{\text{On suppose ici que } e = 0}$

Q-11-1 : Que représente ϕ dans ce cas ?

Q-11-2 : Donner l'ordre de consistance en temps et en espace des schémas fournis à la Remarque 7.

Q-11-3 : Déterminer les équations vérifiées par $\alpha, \beta\ , \gamma$ pour que le schéma (4) soit consistant d'ordre maximum. Déterminer ce schéma et déduire qu'un schéma de type (4) ne peut excéder l'ordre 2 en temps et en espace.

Q-12 : Equations équivalentes et applications.

Il peut être intéressant de comparer des schémas qui ont les mêmes ordres de consistance. Pour cette fin on utilise l'équation équivalente, qui est l'équation aux dérivées partielles dont le schéma numérique (4) approche la solution à un ordre encore meilleur comparé à l'équation de transport.

Les notations sont celles de la question précédente. Soient p et q les ordres de consistance en temps et en espace du schéma (4). On suppose

- $e \neq 0$ à déterminer.
- $v_j^n = \phi(t_n, x_j) + \mathcal{O}(k^{p+1}) + \mathcal{O}(h^{q+1}) \quad n \in \mathbb{N}, j \in \mathbb{Z}$.
  C'est-à-dire que la solution v_j^n fournie par le schéma (4) approche encore mieux la solution de $\partial_t\phi + \partial_x\phi = e$ que celle de $\partial_t\phi + \partial_x\phi = 0$.

Q-12-1 : Montrer que dans ce cas $(P_{k,h}\phi)_j^n = \mathcal{O}(k^p) + \mathcal{O}(h^q), \quad n \in \mathbb{N}, j \in \mathbb{Z}$.

Q-12-2 : A l'aide de (13) déduire que $e = \mathcal{O}(k^p) + \mathcal{O}(h^q)$.

Q-12-3 : En déduire que

$$
\begin{aligned}
e = -\frac{1-\alpha-\beta\ -\gamma}{k}\phi_j^n \quad &+ \quad \tfrac{1}{\lambda}\Big(\gamma-\alpha+(c\lambda)\Big)(\partial_x\phi)_j^n \\
&+ \quad \tfrac{h}{2\lambda}\Big(\gamma+\alpha-(c\lambda)^2\Big)(\partial_{xx}\phi)_j^n \\
&+ \quad \tfrac{h^2}{6\lambda}\Big(\gamma-\alpha+(c\lambda)^3\Big)(\partial_{xxx}\phi)_j^n \\
&+ \quad \mathcal{O}(k^{p+1})+\mathcal{O}(kh^q) \\
&+ \quad \mathcal{O}(k^3)+\mathcal{O}(h^3) \quad n\in\mathbb{N}, j\in\mathbb{Z}.
\end{aligned}
\tag{14}
$$

Q-12-4 : Préciser e (*son terme principal*) pour chacun des schémas de la Remarque 7.

Q-12-5 : Déduire les équations équivalentes associées a chacun des schémas de la Remarque 7.

Q-12-6 : A la lueur de l' **Exercice-2 du Tème-1**

- Préciser pour chaque schéma de la Remarque 7, les conditions sur λ (*en fonction du signe de* c) pour que le schéma soit *dissipatif*. C'est-dire qu'il soit à même de faire décroître la norme L^2 de la solution approchée et par conséquent atténuer les oscillations numériques au cours du temps.
- Pour les schémas de la Remarque 7 ayant le même ordre de consistance, dire lesquels seront moins dissipatifs. *Il faut noter que la dissipativité est une propriété intéressante, mais on aimerait en pratique que son effet soit le moindre possible.*

Remarque 10 (Exploitation supplémentaire du facteur d'amplification)**.**

On a vu précédemment que le facteur d'amplification renseignait sur la stabilité l^2 du schéma.
Ce facteur d'amplification peut aussi permettre d'obtenir les informations supplémentaires suivantes sur le schéma numérique pour le transport :

- l'erreur de dissipation et l'ordre de dissipation du schéma.
- l'erreur de phase et les propriétés dispersives du schéma.

Essayez de vous renseigner sur ces éléments.

Remarque 11 (Exemples supplémentaires de schémas sous la forme (4))**.**

On pose $\lambda = \frac{k}{h}$.

- Lax-Friedrichs (*version évoluée*)

$$v_j^{n+1} = \frac{(1+c\lambda)^2}{4}v_{j-1}^n + \frac{1-(c\lambda)^2}{2}v_j^n + \frac{(1-c\lambda)^2}{4}v_{j+1}^n, \quad n\in\mathbb{N}, j\in\mathbb{Z},$$

- Rusanov : Soit κ un réel tel que $\kappa \geq |c|$

$$v_j^{n+1} = \lambda\frac{\kappa+c}{2}v_{j-1}^n + (1-\kappa\,\lambda)v_j^n - \lambda\frac{\kappa-c}{2}v_{j+1}^n, \quad n\in\mathbb{N}, j\in\mathbb{Z},$$