Analyse numérique et optimisation

TD n^{o}1 I. Terrasse - Groupes 6&13

Différences Finies en dimension 1

Exercice 1

Problème de Cauchy

Soient T un réel positif, $A : t \in [0,T] \mapsto A(t) \in \mathbb{R}^{m,m}$ une fonction continue de $[0,T]$ à valeurs dans l'ensemble des matrices réelles carrées d'ordre m et u^0 un vecteur de $\mathbb{R}^m$. On s'intéresse au problème de Cauchy :

$$\begin{cases} \text{Chercher} \quad u \in \mathcal{C}^1([0,T],\mathbb{R}^m) \quad \text{tel que} \\ \dfrac{du}{dt}(t) + A(t)u(t) = 0 \quad \forall t \in]0,T[, \\ u(0) = u^0. \end{cases} \tag{1}$$

1. **Cas continu.** Montrer que pour toute donnée initiale $u^0 \in \mathbb{R}^m$, le problème de Cauchy (1) est bien posé au sens d'Hadamard.
   Dans le cas particulier où $A(t)$ est constante égale à A, expliciter la constante $C_{A,T}$ (qui dépend de A et de T) dans l'estimation de stabilité
   $$\max_{t\in[0,T]} \|u(t)\| \leq C_{A,T}\|u^0\|$$
   où $\|.\|$ est une norme quelconque sur $\mathbb{R}^m$. Commenter la dépendance en T.
   Dans le cas où $A(t)$ est à valeurs dans l'ensemble des matrices positives et en considérant pour $\|.\|$ la norme euclidienne, montrer que l'énergie $\dfrac{1}{2}\|u(t)\|^2$ est décroissante (on dit que A est dissipative) et donc que l'on peut prendre $C_{A,T} = 1$.
   Dans la suite de l'exercice, on considérera que pour tout $t \in [0,T]$, $A(t)$ est à valeurs dans l'ensemble des matrices positives

2. **Schémas d'Euler.** On considère un entier $N > 0$ et le pas de temps $\Delta t = \dfrac{T}{N}$. On introduit pour tout $0 \leq n \leq N$, $t^n = n\Delta t$ l'instant discret. On construit 2 schémas dits respectivement schéma d'Euler explicite (ou progressif) et schéma d'Euler implicite (ou rétrograde) pour approcher la solution u du problème de Cauchy (1). Partant de la donnée initiale u^0, on génère des suites d'approximation de $u(t^n)_{0\leq n\leq N}$ notées respectivement $(u_E^n)_{0\leq n\leq N}$ et $(u_I^n)_{0\leq n\leq N}$ construites en observant que pour Δt petit :
   $$\frac{u(t^{n+1}) - u(t^n)}{\Delta t} \simeq \frac{du}{dt}(t^n) \text{ et } \frac{u(t^{n+1}) - u(t^n)}{\Delta t} \simeq \frac{du}{dt}(t^{n+1})$$
   Ecrire (en le justifiant) les 2 schémas sous la forme récurrente $u_E^{n+1} = f_E^n(u_E^n)$, respectivement $u_I^{n+1} = f_I^n(u_I^n)$ pour tout $0 \leq n \leq N-1$.
   Dans la suite de l'exercice, on considérera que pour tout $t \in [0,T]$, $A(t)$ est constante égale à $A \in \mathbb{R}^{m,m}$ positive.
   Les schémas Euler explicite et implicite s'écrivent alors pour tout $0 \leq n \leq N-1$
   $$u_E^{n+1} = B_E\, u_E^n$$
   $$u_I^{n+1} = B_I\, u_I^n$$
   avec B_E et B_I 2 matrices de $\mathbb{R}^{m,m}$ que l'on explicitera.

3. **Propagation des erreurs.** On introduit les suites des erreurs (dites de convergence) $e_E^n = u_E^n - u(t^n)$ et $e_I^n = u_I^n - u(t^n)$. Montrer que pour tout $0 \leq n \leq N-1$

$$\begin{array}{rcl} e_E^{n+1} & = & B_E\, e_E^n - \Delta t \eta_E^n \\ e_I^{n+1} & = & B_I(e_I^n - \Delta t \eta_I^n) \end{array} \tag{2}$$

où η_E^n et η_I^n sont des vecteurs de $\mathbb{R}^m$ que l'on précisera (ces vecteurs représentent **l'erreur de troncature** et dépendent de la solution exacte). Interpréter les expressions (2). En déduire par récurrence que pour tout $0 \leq n \leq N-1$:

$$\begin{array}{rcl} e_E^{n+1} & = & -\Delta t \displaystyle\sum_{k=0}^{n} B_E^{n-k}\, \eta_E^k \\ e_I^{n+1} & = & -\Delta t \displaystyle\sum_{k=0}^{n} B_I^{n-k+1} \eta_I^k \end{array} \tag{3}$$

4. **Consistance (précision) d'ordre 1.** Montrer qu'il existe une constante C_u ne dépendant que de la solution exacte u telle que pour tout choix de Δt (c'est à dire pour tout choix de N),

$$\max_{0 \leq n \leq N-1} \|\eta_E^n\| \leq C_u \Delta t \qquad \max_{0 \leq n \leq N-1} \|\eta_I^n\| \leq C_u \Delta t$$

On justifiera que la solution exacte u a bien la régularité nécessaire.

5. **Stabilité.** On note $< .,. >$ le produit scalaire euclidien sur $\mathbb{R}^m \times \mathbb{R}^m$, $\|.\|$ la norme euclidienne sur $\mathbb{R}^m$ et on note aussi $\|.\|$ la norme matricielle subordonnée sur $\mathbb{R}^{m,m}$. Montrer la stabilité inconditionnelle du schéma d'Euler implicite : pour tout choix de Δt,

$$\|B_I\| \leq 1.$$

On montrera comme dans le cas continu que l'énergie au temps discret est décroissante. Montrer que pour le schéma d'Euler explicite, on a l'identité d'"énergie :

$$\frac{1}{2\Delta t}\left[\|B_E x\|^2 - \|x\|^2\right] = -\left[(Ax, x) - \frac{\Delta t}{2}\|Ax\|^2\right]$$

On suppose qu'il existe $M > 0$ tel que $\forall x \in \mathbb{R}^m$ tel que $Ax \neq 0$,

$$\|Ax\|^2 \leq M(Ax, x)$$

En déduire la stabilité conditionnelle du schéma d'Euler explicite : pour tout $\Delta t \leq \dfrac{2}{M}$,

$$\|B_E\| \leq 1.$$

On suppose pour le reste de cette question que A est une matrice symétrique. On écarte le cas trivial pour lequel la matrice A est la matrice nulle. On rappelle que sous cette hypothèse la matrice A est diagonalisable dans $\mathbb{R}^m$ en base orthonormée, et en ordonnant ses valeurs propres sous la forme $0 \leq \lambda_1 \leq \cdots \leq \lambda_m$, on a $\|A\| = \lambda_m > 0$. En utilisant la base propre, retrouver la stabilité inconditionnelle du schéma d'Euler implicite et la stabilité conditionnelle du schéma d'Euler explicite avec la condition :

$$\Delta t \leq \frac{2}{\lambda_m} = \frac{2}{\|A\|},$$

Comparer avec la condition du cas général.

6. **Convergence.** En déduire sous les hypotèses précédentes sur A, les résultats de convergence à l'ordre 1 suivants :

   Schéma d'Euler implicite : il existe une constante C (dépendant de u et de T) telle que pour tout choix de Δt,
   $$\max_{0 \leq n \leq N} \|e_I^n\| \leq C\Delta t$$
   Schéma d'Euler explicite : il existe une constante C (dépendant de u et de T) telle que pour tout Δt assez petit (à préciser),
   $$\max_{0 \leq n \leq N} \|e_E^n\| \leq C\Delta t$$

Eléments de Correction

Exercice 1

Problème de Cauchy

1. **Cas continu.** La fonction $f : (t,v) \in [0,T] \times \mathbb{R}^m \mapsto f(t,v) := -A(t)\,v \in \mathbb{R}^m$ est globalement lipschitzienne par rapport à la variable v. En effet pour tout $(t, v1, v2) \in [0,T] \times \mathbb{R}^m \times \mathbb{R}^m$,

$$\begin{aligned} \|f(t,v1) - f(t,v2)\| &= \|A(t)(v1 - v2)\| \leq \|A(t)\|\,\|v1 - v2\| \\ &\leq \left(\max_{s\in[0,T]} \|A(s)\|\right) \|v1 - v2\| \end{aligned}$$

où $\|.\|$ désigne une norme quelconque sur $\mathbb{R}^m$ ainsi que la norme subordonnée sur $\mathbb{R}^{m,m}$. En vertu du théorème de Cauchy-Lipschitz, le problème (1) admet une et une seule solution qui dépend continûment de la donnée initiale u^0.
Dans le cas particulier où $A(t)$ est constante égale à A, la solution vérifie explicitement :

$$u(t) = e^{-tA}\,u^0$$

et donc

$$\max_{t\in[0,T]} \|u(t)\| \leq e^{T\|A\|}\|u^0\|$$

La constante $C_{A,T} = e^{T\|A\|}$ explose exponentiellement quand T tend vers l'infini.
On se place dans le cas où $A(t)$ est à valeurs dans l'ensemble des matrices positives. En notant $< .,. >$ le produit scalaire euclidien, il vient pour tout $t \in]0,T[$,

$$\frac{1}{2}\frac{d}{dt}\|u(t)\|^2 = < \frac{du(t)}{dt},\, u(t) > = - < A(t)\,u(t),\, u(t) > \leq 0.$$

donc la fonction $t \in [0,T] \mapsto \|u(t)\|^2 \in \mathbb{R}_+$ est décroissante et

$$\max_{t\in[0,T]} \|u(t)\| \leq \|u^0\|,$$

donc on peut prendre $C_{A,T} = 1$.

2. **Schémas d'Euler.**
Schéma d'Euler explicite :

$$\begin{cases} u_E^0 = u^0 \\ \dfrac{u_E^{n+1} - u_E^n}{\Delta t} + A(t^n)\,u_E^n = 0 \qquad 0 \leq n \leq N-1 \end{cases}$$

Soit

$$\begin{cases} u_E^0 = u^0 \\ u_E^{n+1} = (I - \Delta t A(t^n)\,)\ u_E^n \qquad 0 \leq n \leq N-1 \end{cases}$$

Schéma d'Euler implicite :

$$\left\{\begin{array}{l} u_I^0 = u^0 \\ \\ \dfrac{u_I^{n+1} - u_I^n}{\Delta t} + A(t^{n+1})\, u_I^{n+1} = 0 \qquad 0 \leq n \leq N-1 \end{array}\right.$$

Soit

$$\left\{\begin{array}{l} u_I^0 = u^0 \\ \\ (I + \Delta t A(t^{n+1})\,)\, u_I^{n+1} = u_I^n \qquad 0 \leq n \leq N-1 \end{array}\right.$$

Les matrices $(I + \Delta t A(t^{n+1})\,)$ sont inversibles car leur noyau est réduit à 0 : en effet soit $x \in \mathbb{R}^m$ tel que $(I + \Delta t A(t^{n+1})\,)\, x = 0$,

$$0 =< (I + \Delta t A(t^{n+1})\,)\, x, x >= \|x\|^2 + \Delta t < A(t^{n+1}) x, x > \geq \|x\|^2$$

et donc $x = 0$. Le schéma d'Euler implicite s'écrit alors :

$$\left\{\begin{array}{l} u_I^0 = u^0 \\ \\ u_I^{n+1} = (I + \Delta t A(t^{n+1})\,)^{-1}\, u_I^n \qquad 0 \leq n \leq N-1 \end{array}\right.$$

Dans le cas où $A(t) = A$ matrice constante positive, les schémas d'Euler explicite et implicite s'écrivent alors pour tout $0 \leq n \leq N-1$,

$$\begin{array}{rcccl} u_E^{n+1} & = & (I - \Delta t A)\, u_E^n & = & B_E\, u_E^n \\ \\ u_I^{n+1} & = & (I + \Delta t A)^{-1}\, u_I^n & = & B_I\, u_I^n \end{array}$$

soit

$$B_E = I - \Delta t A \qquad B_I = (I + \Delta t A)^{-1}$$

3. **Propagation des erreurs.** Pour les 2 schémas, on a

$$e^{n+1} = B\, e^n - \Delta t \left[\frac{u(t^{n+1}) - B\, u(t^n)}{\Delta t} \right]$$

Pour le schéma d'Euler explicite,

$$\begin{array}{rcl} \eta_E^n & := & \dfrac{u(t^{n+1}) - B_E\, u(t^n)}{\Delta t} = \dfrac{u(t^{n+1}) - u(t^n)}{\Delta t} + (A\, u)(t^n) \\ & = & \dfrac{u(t^{n+1}) - u(t^n)}{\Delta t} - \dfrac{du}{dt}(t^n) \end{array}$$

représente l'erreur que l'on a commise en approchant l'opérateur différentiel et qui fait que la solution exacte ne vérifie pas le schéma discret dite erreur de troncature.
Pour le schéma d'Euler implicite,

$$\begin{array}{rcl} \eta_I^n & := & B_I^{-1} \dfrac{u(t^{n+1}) - B_I\, u(t^n)}{\Delta t} = \dfrac{u(t^{n+1}) - u(t^n)}{\Delta t} + (A\, u)(t^{n+1}) \\ & = & \dfrac{u(t^{n+1}) - u(t^n)}{\Delta t} - \dfrac{du}{dt}(t^{n+1}) \end{array}$$

représente de même l'erreur que l'on a commise en approchant l'opérateur différentiel et qui fait que la solution exacte ne vérifie pas le schéma discret dite erreur de troncature. On obtient :

$$\begin{array}{rcl} e_E^{n+1} & = & B_E\, e_E^n - \Delta t \eta_E^n \\ e_I^{n+1} & = & B_I (e_I^n - \Delta t \eta_I^n) \end{array}$$

Les expressions précédentes montrent qu'à chaque pas de temps l'erreur est somme de 2 termes, l'un provient de l'erreur au temps précédent qui se propage au temps d'après (ce qui par un effet de boule de neige peut influer sur la convergence, c'est la notion de **stabilité** d'un schéma) et l'autre de l'erreur commise entre les 2 instants t^n et t^{n+1} qui est due au fait que la solution exacte n'est pas solution du schéma discret (mais cette erreur peut être maîtrisée si l'approximation de l'opérateur différentiel est suffisamment précise au moins pour Δt assez petit, c'est la notion de **consistance** d'un schéma). Ces 2 notions stabilité ET consistance sont essentielles pour assurer la convergence d'un schéma d'approximation. Les relations (3) se verifient facilement par réccurence en remarquant que l'erreur au temps initial $e_E^0 = e_I^0 = 0$.

4. **Consistance (précision) d'ordre 1.** La solution exacte u de classe $\mathcal{C}^1([0,T], \mathbb{R}^m)$ vérifie :

$$\frac{du}{dt}(t) = -A(t)u(t) \quad \forall t \in]0,T[, \qquad \text{soit } \frac{du}{dt} \in \mathcal{C}^1([0,T], \mathbb{R}^m)$$

et donc est de classe $\mathcal{C}^2([0,T], \mathbb{R}^m)$ (et même de classe $\mathcal{C}^\infty$). On peut donc faire un développement de Taylor à l'ordre 2 aux temps t^n : $\exists \tau_E^n \in [t^n, t^{n+1}]$, tel que

$$u(t^{n+1}) = u(t^n) + \Delta t \frac{du}{dt}(t^n) + \frac{\Delta t^2}{2} \frac{d^2u}{dt^2}(\tau_E^n)$$

De même, $\exists \tau_I^n \in [t^n, t^{n+1}]$, tel que

$$u(t^n) = u(t^{n+1}) - \Delta t \frac{du}{dt}(t^{n+1}) + \frac{\Delta t^2}{2} \frac{d^2u}{dt^2}(\tau_I^n)$$

Dans les 2 cas explicite et implicite,

$$\max_{0 \le n \le N-1} \|\eta^n\| \le \max_{t \in [0,T]} \|\frac{1}{2} \frac{d^2u(t)}{dt^2}\| \; \Delta t$$

En notant $C_u = \max_{t \in [0,T]} \|\frac{1}{2} \frac{d^2u(t)}{dt^2}\|$, on obtient les estimations demandées qui montrent que l'erreur de troncature tend vers 0 à l'ordre 1 quand Δt tend vers 0. On l'appelle erreur de toncature puisque qu'elle correspond à un développement de Taylor tronqué jusqu'à un certain ordre. Le schéma est dit consistant d'ordre 1. Notons que c'est une estimation a priori qui dépend de la solution exacte que l'on ne connait pas puisque justement on en cherche une approximation mais qui permet d'assurer que l'erreur va bien tendre vers 0 puisque le schéma approche correctement l'équation continue.

5. **Stabilité.** Les relations (3) faisant intervenir des puissances des opérateurs B, il convient d'assurer que la norme de ces opérateurs est inférieure où égale à 1. Dans le cas continu, on avait obtenu la décroissance de l'énergie en effectuant le produit scalaire de l'équation au temps t par la solution $u(t)$. Pour le schéma implicite, en posant $y = B_I x$, on choisit de prendre le produit scalaire du schéma par y :

$$0 = < \frac{y-x}{\Delta t} + A\, y, y > = \frac{1}{2\Delta t}(\|y\|^2 - \|x\|^2 + < y-x, y-x >) + < Ay, y >$$

où on a écrit $y = (y+x)/2 + (y-x)/2$, en utilisant que $y - x = -\Delta t A y$

$$\frac{1}{2\Delta t}\left[\|y\|^2 - \|x\|^2\right] = -\left[< Ay, y > + \frac{\Delta t}{2}\|Ay\|^2\right] \leq 0$$

on en déduit

$$\|y\|^2 \leq \ \|x\|^2$$

et donc pour tout choix de Δt

$$\|B_I\| \leq 1.$$

On suit la même démarche pour le schéma explicite, en posant $y = B_E x$

$$0 =< \frac{y-x}{\Delta t} + A\,x, x >= \frac{1}{2\Delta t}(\|y\|^2 - \|x\|^2 - < y-x, y-x >) + < Ax, x >$$

où on a écrit $x = (y+x)/2 - (y-x)/2$, en utilisant que $y - x = -\Delta t A x$

$$\frac{1}{2\Delta t}\left[\|y\|^2 - \|x\|^2\right] = -\left[< Ax, x > - \frac{\Delta t}{2}\|Ax\|^2\right]$$

Si $Ax = 0$ le terme $(Ax, x) - \frac{\Delta t}{2}\|Ax\|^2$ est nul, sinon en utilisant l'hypothèse de l'énoncé sur l'opérateur A, il est positif sous la condition :

$$\Delta t \leq \frac{2}{M} \leq \frac{2A(x,x)}{\|Ax\|^2}$$

On en déduit que pour tout $\Delta t \leq \frac{2}{M}$,

$$\|B_E\| \leq 1.$$

Soit maintenant A matrice symétrique non identiquement nulle de valeurs propres $0 \leq \lambda_1 \leq \cdots \leq \lambda_m$. Les matrices B_I et B_E diagonalisent dans la même base que A et leurs valeurs propres sont :
pour $B_I = (I + \Delta t A)^{-1}$, $(1 + \Delta t \lambda_k)^{-1}, 1 \leq k \leq m$ qui sont toutes dans l'intervalle $]0, 1]$.
Et par suite

$$\|B_I\| = (1 + \Delta t \lambda_1)^{-1} \leq 1$$

pour $B_E = I - \Delta t A$:

$$1 - \Delta t \lambda_m \leq \cdots \leq 1 - \Delta t \lambda_1 \leq 1$$

Donc sous la condition $-1 \leq 1 - \Delta t \lambda_m$, on a $\|B_E\| \leq 1$. La condition s'écrit encore

$$\Delta t \leq \frac{2}{\lambda_m} = \frac{2}{\|A\|},$$

Si on décompose $x \in \mathbb{R}^m$ sur une base orthonormée de vecteurs propres, on obtient :

$$\|Ax\|^2 = \sum_{j=1}^{m} \lambda_i^2 x_i^2 \leq \lambda_m \sum_{j=1}^{m} \lambda_i x_i^2 = \lambda_m < Ax, x >= \|A\| < Ax, x >$$

et le maximum est atteint pour le m-ème vecteur propre et donc $M = \lambda_m$, on retrouve la condition générale.

6. **Convergence.** On peut donc rassembler les résultats de consistance et stabilité précédents en utilisant les relations (3) :
   Cas implicite : pour tout $1 \leq n \leq N$

$$\begin{aligned} \|e_I^n\| &= \Delta t \left\| \sum_{k=0}^{n-1} B_I^{n-k} \eta_I^k \right\| \leq \Delta t \sum_{k=0}^{n-1} \|B_I^{n-k} \eta_I^k\| \\ &\leq \Delta t \sum_{k=0}^{n-1} \|B_I\|^{n-k} \|\eta_I^k\| \leq \Delta t \sum_{k=0}^{n-1} C_u \Delta t \\ &\leq \Delta t N C_u \Delta t = \Delta t T C_u \end{aligned}$$

   d'où le résultat de convergence à l'ordre 1 avec $C = TC_u$.
   Cas explicite : la démonstration est identique sous l'hypothèse de stabilité $\Delta t \leq \frac{2}{M}$.