Ecole Polytechnique, Promotion 2017
Approximation numérique et optimisation (MAP 411)
Premier devoir du lundi 24 septembre 2018
Corrigé proposé par G. Allaire

On considére le schéma aux différences finies, décentré à 3 points

$$u_i^{n+1} = \alpha u_{i-2}^n + \beta u_{i-1}^n + \gamma u_i^n$$

pour l'équation d'advection dans l'intervalle $(0,1)$ avec condition aux limites périodique

$$\begin{cases} \dfrac{\partial u}{\partial t} + a\dfrac{\partial u}{\partial x} = 0 \text{ pour } (x,t) \in (0,1) \times \mathbb{R}_*^+ \\ u(t, x+1) = u(t,x) \text{ pour } (x,t) \in \mathbb{R} \times \mathbb{R}_*^+ \\ u(0,x) = u_0(x) \text{ pour } x \in (0,1). \end{cases} \tag{1}$$

avec une vitesse constante et strictement positive $a > 0$ et $c = a\Delta t/\Delta x$. Dans tout ce qui suit, on suppose que les pas d'espace Δx et de temps Δt tendent vers zéro en gardant leur rapport c constant.

1. Pour étudier la consistance du schéma on doit introduire l'erreur de troncature mais auparavant il faut mettre le schéma sous sa forme standard qui est

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} - \frac{\alpha u_{i-2}^n + \beta u_{i-1}^n + (\gamma - 1)u_i^n}{\Delta t} = 0.$$

L'erreur de troncature est donc

$$E = \frac{u(t_{n+1}, x_i) - u(t_n, x_i)}{\Delta t} - \frac{\alpha u(t_n, x_{i-2}) + \beta u(t_n, x_{i-1}) + (\gamma - 1)u(t_n, x_i)}{\Delta t}$$

$$= \frac{\partial u}{\partial t}(t_n, x_i) - \frac{\alpha + \beta + \gamma - 1}{\Delta t} u(t_n, x_i) + \frac{(2\alpha + \beta)\Delta x}{\Delta t}\frac{\partial u}{\partial x}(t_n, x_i) + \mathcal{O}(\Delta t + \Delta x^2/\Delta t).$$

On rappelle que l'on suppose que le rapport $\Delta x/\Delta t$ est constant. L'erreur de troncature tend alors vers zéro avec Δt et Δx lorsque la fonction $u(t,x)$ est une solution de l'équation d'advection (et seulement dans ce cas) si et seulement si

$$\alpha + \beta + \gamma = 1 \quad \text{ et } 2\alpha + \beta = c.$$

2. On ne peut retrouver le schéma décentré amont qu'en choisissant $\alpha = 0$. Avec les formules de la question précédente on obtient $\beta = c$ et $\gamma = 1 - c$, ce qui est bien formule du schéma décentré amont (qui est donc un cas particulier de la classe de schémas étudiée).

3. Pour trouver un schéma d'ordre 2 il faut poursuivre le développement de Taylor dans l'analyse de l'erreur de troncature. On reprend donc le calcul de la première question

$$E = \frac{\partial u}{\partial t}(t_n, x_i) + \frac{1}{2}\Delta t\frac{\partial^2 u}{\partial t^2}(t_n, x_i) - \frac{\alpha + \beta + \gamma - 1}{\Delta t} u(t_n, x_i)$$

$$+\frac{(2\alpha+\beta)\Delta x}{\Delta t}\frac{\partial u}{\partial x}(t_n,x_i)-\frac{(4\alpha+\beta)\Delta x^2}{2\Delta t}\frac{\partial^2 u}{\partial x^2}(t_n,x_i)+\mathcal{O}(\Delta t^2+\Delta x^3/\Delta t).$$

Bien sûr, on garde les deux relations $\alpha+\beta+\gamma=1$ et $2\alpha+\beta=c$ puisque le schéma doit être consistant. Par ailleurs, si $u(t,x)$ est une solution de l'équation d'advection elle vérifie aussi

$$\frac{\partial^2 u}{\partial t^2}=a^2\frac{\partial^2 u}{\partial x^2}.$$

Dans ce cas, l'erreur de troncature se simplifie en

$$E=\frac{\partial u}{\partial t}(t_n,x_i)+\frac{a^2\Delta t^2-(4\alpha+\beta)\Delta x^2}{2\Delta t}\frac{\partial^2 u}{\partial x^2}(t_n,x_i)+\mathcal{O}(\Delta t^2+\Delta x^2).$$

Le schéma est donc d'ordre 2 si et seulement si

$$4\alpha+\beta=c^2.$$

Un calcul simple donne alors les seules valeurs des paramètres correspondant à un schéma d'ordre 2

$$\alpha=c(c-1)/2,\quad \beta=c(2-c),\quad \gamma=(c-1)(c-2)/2.$$

4. Pour que le schéma vérifie le principe du maximum discret il faut que u_i^{n+1} soit une combinaison convexe des valeurs au pas de temps précédent. On sait déjà que $\alpha+\beta+\gamma=1$. Il faut donc que $\alpha\geq 0$, $\beta\geq 0$, $\gamma\geq 0$. Au vu des formules ci-dessus, sachant que $c\geq 0$ par hypothèse, les deux seules valeurs de c pour lesquelles c'est possible sont $c=1$ et $c=2$.

Lorsque $c=1$ le schéma devient $u_i^{n+1}=u_{i-1}^n$. La solution numérique est donc transportée exactement à la bonne vitesse, d'une maille par pas de temps, sans aucune déformation de son profil initial. En ce sens le schéma est exact et d'ordre infini (il n'y a aucune perte de précision, une fois la donnée initiale discrétisée sur le maillage).

Lorsque $c=2$ le schéma devient $u_i^{n+1}=u_{i-2}^n$. Le même raisonnement que précédemment s'applique. Le schéma est d'ordre infini ou exact.

5. Pour étudier la stabilité L^2 du schéma d'ordre 2 on utilise l'analyse de Fourier. Avec les notations du polycopié, pour $k\in\mathbb{Z}$, le coefficient de Fourier $\hat{u}^n(k)$ de la solution du schéma vérifie

$$\hat{u}^{n+1}(k)=\left(\frac{c(c-1)}{2}e^{-4i\pi k\Delta x}+c(2-c)e^{-2i\pi k\Delta x}+\frac{(c-1)(c-2)}{2}\right)\hat{u}^n(k),$$

ce qui est équivalent à

$$\hat{u}^{n+1}(k)=\left(\frac{c(c-1)}{2}2\cos(2\pi k\Delta x)+c(2-c)-(c-1)e^{2i\pi k\Delta x}\right)e^{-2i\pi k\Delta x}\hat{u}^n(k)$$

En notant $\xi=2\pi k\Delta x$, on obtient après simplification

$$|\hat{u}^{n+1}(k)|^2=|A(k)|^2|\hat{u}^n(k)|^2$$

avec

$$
\begin{aligned}
|A(k)|^2 &= \left((c-1)^2\cos\xi + c(2-c)\right)^2 + (c-1)^2\sin^2\xi \\
&= (c-1)^2 c(c-2)\cos^2\xi + 2c(2-c)(c-1)^2\cos\xi + c^2(2-c)^2 + (c-1)^2
\end{aligned}
$$

On remarque que

$$c^2(2-c)^2 + (c-1)^2 = 1 + (c-1)^2 c(c-2),$$

donc

$$|A(k)|^2 = 1 - (c-1)^2 c(2-c)(1-\cos\xi)^2.$$

Pour $0 \le c \le 2$ le coefficient $(c-1)^2 c(2-c)$ est positif, et on en déduit donc que $|A(k)| \le 1$ pour toute fréquence $k \in \mathbb{Z}$, ce qui prouve que le schéma est stable en norme L^2 sous la condition CFL $0 \le c \le 2$.

L'avantage de ce schéma est qu'il permet de prendre des pas de temps deux fois plus grands que les schémas décentré amont, de Lax-Friedrichs ou de Lax-Wendroff, tout en étant explicite. Par ailleurs, un schéma linéaire consistant et stable est convergent d'après le théorème de Lax. Le schéma est donc convergent sous la condition CFL ci-dessus.

6. Pour obtenir l'équation équivalente du schéma il faut poursuivre le développement de Taylor dans l'erreur de troncature un cran plus loin

$$E = \frac{\partial u}{\partial t}(t_n, x_i) + \frac{1}{2}\Delta t \frac{\partial^2 u}{\partial t^2}(t_n, x_i) + \frac{1}{6}\Delta t^2 \frac{\partial^3 u}{\partial t^3}(t_n, x_i)$$

$$+a\frac{\partial u}{\partial x}(t_n, x_i) - \frac{(4\alpha+\beta)\Delta x^2}{2\Delta t}\frac{\partial^2 u}{\partial x^2}(t_n, x_i) + \frac{(8\alpha+\beta)\Delta x^3}{6\Delta t}\frac{\partial^3 u}{\partial x^3}(t_n, x_i) + \mathcal{O}(\Delta t^3 + \Delta x^3).$$

Avec les valeurs des paramètres α, β, γ et en utilisant les relations

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} \quad \text{et} \quad \frac{\partial^3 u}{\partial t^3} = -a^3 \frac{\partial^3 u}{\partial x^3}$$

pour une solution, on obtient l'équation équivalente

$$\frac{\partial u}{\partial t}(t,x) + a\frac{\partial u}{\partial x}(t,x) - a(c-1)(c-2)\frac{\Delta x^2}{6}\frac{\partial^3 u}{\partial x^3}(t,x) = 0.$$

Remarquons qu'il n'est pas nécessaire de supposer que les relations ci-dessus sur les dérivées secondes et troisièmes soient exactes, mais seulement vérifiées à un terme d'erreur $\mathcal{O}(\Delta t^2 + \Delta x^2)$ près, pour obtenir et justifier l'équation équivalente.

Par définition, le schéma est au moins d'ordre 3 pour cette équation équivalente, alors qu'il n'est que d'ordre 2 (a priori) pour la simple équation d'advection.

Le terme de dérivée d'ordre 3 dans l'équation équivalente est dit "dispersif". Il correspond à une modification de la vitesse d'advection, variable suivant les modes de Fourier de la solution. Cela se traduit par de légères oscillations numériques, notamment pour les solutions peu régulières (par exemple, discontinues en espace, voir le polycopié).