Ecole Polytechnique, Promotion 2016
Approximation numérique et optimisation (MAP 411)
Devoir obligatoire du mardi 5 décembre 2017
Corrigé proposé par G. Allaire

1 Différences finies

1. Par périodicité il suffit de calculer, à chaque instant $n\Delta t$, les N valeurs $(u_j^n)_{1\leq j\leq N}$. Pour que le schéma

$$\frac{1}{12}\frac{u_{j+1}^{n+1}-u_{j+1}^n}{\Delta t}+\frac{5}{6}\frac{u_j^{n+1}-u_j^n}{\Delta t}+\frac{1}{12}\frac{u_{j-1}^{n+1}-u_{j-1}^n}{\Delta t}-\nu\frac{\sigma_j^{n+1}+\sigma_j^n}{2(\Delta x)^2}=0$$

soit bien défini, il faut donc que la matrice d'ordre N

$$A=\begin{pmatrix} 5/6+2c & 1/12-c & & & 1/12-c \\ 1/12-c & 5/6+2c & 1/12-c & 0 & \\ & \ddots & \ddots & \ddots & \\ & 0 & 1/12-c & 5/6+2c & 1/12-c \\ 1/12-c & & & 1/12-c & 5/6+2c \end{pmatrix} \text{ avec } c=\frac{\nu\Delta t}{2(\Delta x)^2},$$

soit inversible. Pour cela, on vérifie que la forme quadratique associée est définie positive

$$Au\cdot u=c\sum_{j=1}^N(u_{j+1}-u_j)^2+\frac{1}{12}\sum_{j=1}^N(u_{j+1}+u_j)^2+\frac{4}{6}\sum_{j=1}^N u_j^2.$$

Le schéma est clairement consistant car, si $u_j^n=u(t_n,x_j)$, alors

$$\frac{1}{12}\frac{u_{j+1}^{n+1}-u_{j+1}^n}{\Delta t}+\frac{5}{6}\frac{u_j^{n+1}-u_j^n}{\Delta t}+\frac{1}{12}\frac{u_{j-1}^{n+1}-u_{j-1}^n}{\Delta t}=$$

$$\frac{1}{12}\frac{\partial u}{\partial t}(t_n,x_{j+1})+\frac{5}{6}\frac{\partial u}{\partial t}(t_n,x_j)+\frac{1}{12}\frac{\partial u}{\partial t}(t_n,x_{j-1})+\mathcal{O}(\Delta t)=$$

$$\frac{\partial u}{\partial t}(t_n,x_j)+\mathcal{O}(\Delta t+\Delta x),$$

et

$$\frac{\sigma_j^{n+1}+\sigma_j^n}{2(\Delta x)^2}=\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t_{n+1},x_j)+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t_n,x_j)+\mathcal{O}(\Delta x)$$

$$=\frac{\partial^2 u}{\partial x^2}(t_n,x_j)+\mathcal{O}(\Delta x+\Delta t),$$

donc l'erreur de troncature pour une solution régulière u est (au moins) de l'ordre de $\mathcal{O}(\Delta t+\Delta x)$.

2. Pour montrer que le schéma est stable en norme L^2 on utilise l'analyse de Fourier. En reprenant les notations du polycopié, pour $k \in \mathbb{Z}$ on note $\hat{u}^n(k)$ le coefficient de Fourier qui vérifie, par application de la transformée de Fourier au schéma,

$$\frac{\hat{u}^{n+1}(k) - \hat{u}^n(k)}{\Delta t}\left(\frac{1}{12}e^{2i\pi k\Delta x} + \frac{5}{6} + \frac{1}{12}e^{-2i\pi k\Delta x}\right) -$$

$$\frac{\nu}{2(\Delta x)^2}(\hat{u}^{n+1}(k) + \hat{u}^n(k))\left(e^{2i\pi k\Delta x} - 2 + e^{-2i\pi k\Delta x}\right) = 0.$$

En regroupant les termes on obtient, en posant $\phi = 2\pi k\Delta x$,

$$\hat{u}^{n+1}(k) = A(k)\hat{u}^n(k) \quad \text{avec} \quad A(k) = \frac{5 + \cos\phi - \frac{6\nu\Delta t}{(\Delta x)^2}(1-\cos\phi)}{5 + \cos\phi + \frac{6\nu\Delta t}{(\Delta x)^2}(1-\cos\phi)}.$$

Par le théorème de Plancherel, le schéma est stable si et seulement si on a

$$-1 \leq A(k) \leq 1,$$

ce qui est équivalent à

$$-5-\cos\phi-\frac{6\nu\Delta t}{(\Delta x)^2}(1-\cos\phi) \leq 5+\cos\phi-\frac{6\nu\Delta t}{(\Delta x)^2}(1-\cos\phi) \leq 5+\cos\phi+\frac{6\nu\Delta t}{(\Delta x)^2}(1-\cos\phi)$$

qui est toujours vrai puisque $(1-\cos\phi) \geq 0$.

Le schéma est donc stable en norme L^2 et consistant. Par le théorème de Lax 2.2.17 il est automatiquement convergeant.

3. Pour minimiser les calculs on utilise des formules de Taylor centrées sur le point $((n+1/2)\Delta t, j\Delta x)$ (qu'il est "naturel" d'utiliser car il est centré par rapport aux différents points du schéma). Dans toute la suite $\tilde{u}_j^n = u(t_n, x_j)$. On commence par les termes de dérivées en temps :

$$\frac{1}{12}\frac{\tilde{u}_{j+1}^{n+1} - \tilde{u}_{j+1}^n}{\Delta t} = \frac{1}{12}\frac{\partial u}{\partial t}(t_{n+1/2}, x_{j+1}) + \mathcal{O}(\Delta t)^3,$$

$$\frac{1}{12}\frac{\tilde{u}_{j-1}^{n+1} - \tilde{u}_{j-1}^n}{\Delta t} = \frac{1}{12}\frac{\partial u}{\partial t}(t_{n+1/2}, x_{j-1}) + \mathcal{O}(\Delta t)^3,$$

$$\frac{5}{6}\frac{\tilde{u}_j^{n+1} - \tilde{u}_j^n}{\Delta t} = \frac{5}{6}\frac{\partial u}{\partial t}(t_{n+1/2}, x_j) + \mathcal{O}(\Delta t)^3.$$

Par ailleurs, on a

$$\frac{\partial u}{\partial t}(t_{n+1/2}, x_{j+1}) = \frac{\partial u}{\partial t}(t_{n+1/2}, x_j) + \Delta x\frac{\partial^2 u}{\partial t\partial x}(t_{n+1/2}, x_j) +$$

$$\frac{1}{2}(\Delta x)^2\frac{\partial^3 u}{\partial t\partial x^2}(t_{n+1/2}, x_j) + \frac{1}{6}(\Delta x)^3\frac{\partial^4 u}{\partial t\partial x^3}(t_{n+1/2}, x_j) + \mathcal{O}(\Delta x)^4,$$

et, avec un résultat similaire en x_{j-1}, on en déduit

$$\frac{1}{12}\frac{\tilde{u}_{j+1}^{n+1} - \tilde{u}_{j+1}^n}{\Delta t} + \frac{5}{6}\frac{\tilde{u}_j^{n+1} - \tilde{u}_j^n}{\Delta t} + \frac{1}{12}\frac{\tilde{u}_{j-1}^{n+1} - \tilde{u}_{j-1}^n}{\Delta t} = \frac{\partial u}{\partial t}(t_{n+1/2}, x_j) +$$

$$\frac{1}{12}(\Delta x)^2 \frac{\partial^3 u}{\partial t \partial x^2}(t_{n+1/2}, x_j) + \mathcal{O}\left((\Delta t)^3 + (\Delta x)^4\right).$$

On passe maintenant aux termes de dérivées en espace : d'après la formule (1.25) du polycopié on a

$$\frac{\tilde{\sigma}_j^{n+1}}{(\Delta x)^2} = \frac{\partial^2 u}{\partial x^2}(t_{n+1}, x_j) + \frac{(\Delta x)^2}{12}\frac{\partial^4 u}{\partial x^4}(t_{n+1}, x_j) + \mathcal{O}(\Delta x)^4.$$

Or

$$\frac{\partial^2 u}{\partial x^2}(t_{n+1}, x_j) = \frac{\partial^2 u}{\partial x^2}(t_{n+1/2}, x_j) + \frac{\Delta t}{2}\frac{\partial^3 u}{\partial t \partial x^2}(t_{n+1/2}, x_j) + \mathcal{O}(\Delta t)^2,$$

et

$$\frac{\partial^4 u}{\partial x^4}(t_{n+1}, x_j) = \frac{\partial^4 u}{\partial x^4}(t_{n+1/2}, x_j) + \frac{\Delta t}{2}\frac{\partial^5 u}{\partial t \partial x^4}(t_{n+1/2}, x_j) + \mathcal{O}(\Delta t)^2.$$

On a des développements similaires pour $\frac{\tilde{\sigma}_j^n}{(\Delta x)^2}$ mais avec un pas $-\Delta t/2$. Par conséquent, en sommant on obtient

$$\frac{\tilde{\sigma}_j^{n+1} + \tilde{\sigma}_j^n}{2(\Delta x)^2} = \frac{\partial^2 u}{\partial x^2}(t_{n+1/2}, x_j) + \frac{(\Delta x)^2}{12}\frac{\partial^4 u}{\partial x^4}(t_{n+1/2}, x_j) + \mathcal{O}\left((\Delta t)^2 + (\Delta x)^4\right).$$

En dérivant deux fois l'équation, une solution u vérifie

$$\frac{\partial^3 u}{\partial t \partial x^2} - \nu \frac{\partial^4 u}{\partial x^4} = 0,$$

et l'erreur de troncature est donc (en notant $u_j^n = u(t_n, x_j)$)

$$\frac{1}{12}\frac{\tilde{u}_{j+1}^{n+1} - \tilde{u}_{j+1}^n}{\Delta t} + \frac{5}{6}\frac{\tilde{u}_j^{n+1} - \tilde{u}_j^n}{\Delta t} + \frac{1}{12}\frac{\tilde{u}_{j-1}^{n+1} - \tilde{u}_{j-1}^n}{\Delta t} - \nu\frac{\tilde{\sigma}_j^{n+1} + \tilde{\sigma}_j^n}{2(\Delta x)^2} =$$

$$\mathcal{O}\left((\Delta t)^2 + (\Delta x)^4\right).$$

2 Optimisation

On définit

$$J(x) = c \cdot x + \sum_{i=1}^n x_i \log\left(\frac{x_i}{\sum_{j=1}^n x_j}\right)$$

et on considère le problème d'optimisation

$$\min_{x \in K,\ Ax = b} J(x). \tag{1}$$

1. La fonction $J(x)$ est est évidemment continue à l'intérieur de K, et les seules difficultés peuvent venir lorsque $x_i \to 0$. Cependant, on a

$$\lim_{x_i \to 0} x_i \log x_i = 0,$$

et pour $\sum_{j=1}^n x_j \leq 1$

$$0 \leq -x_i \log\left(\sum_{j=1}^n x_j\right) \leq -\sum_{i=1}^n x_i \log\left(\sum_{j=1}^n x_j\right) \to 0 \text{ quand } x \to 0.$$

Donc J est aussi continue sur le bord de K.

2. Par hypothèse, on sait que pour toute colonne $j \in \{1, ..., n\}$ il existe un coefficient a_{ij} de A qui est strictement positif. On a donc pour tout $x \in K$ tel que $Ax = b$

$$0 \leq a_{ij}x_j \leq \sum_{j=1}^n a_{ij}x_j = b_i \Rightarrow 0 \leq x_j \leq \frac{b_i}{a_{ij}}$$

ce qui implique que l'ensemble $\{x \in K,\ Ax = b\}$ est borné. On minimise une fonction continue sur un fermé borné non vide de $\mathbb{R}^n$, donc il existe au moins une solution optimale de (1).

3. On vérifie sans problème que $J(tx) = tJ(x)$ pour tout $x \in K$ et tout $t > 0$. Etudions alors la convexité de J dans $K \cap \{\sum_{i=1}^n x_i = 1\}$:

$$J(x) = c \cdot x + \sum_{i=1}^n x_i \log x_i \quad \text{pour} \quad x \in K \cap \{\sum_{i=1}^n x_i = 1\},$$

qui est une somme de fonctions convexes donc J est convexe dans $K \cap \{\sum_{i=1}^n x_i = 1\}$. Passons au cas général dans K : il nous faut montrer que

$$J(\theta x + (1-\theta)y) \leq \theta J(x) + (1-\theta)J(y)$$

pour tout $x, y \in K$ et $\theta \in [0, 1]$. On définit

$$X = \frac{x}{S} \text{ avec } S = \sum_{i=1}^n x_i, \quad Y = \frac{y}{T} \text{ avec } T = \sum_{i=1}^n y_i, \quad \theta' = \frac{\theta S}{\theta S + (1-\theta)T}.$$

Grâce à l'homogénéité de degré 1 de J on obtient

$$\theta J(x)+(1-\theta)J(y)-J(\theta x+(1-\theta)y) = \theta S J(X)+(1-\theta)TJ(Y)-(\theta S+(1-\theta)T)J(\theta' X+(1-\theta')Y)$$

$$= (\theta S + (1-\theta)T)\Big(\theta' J(X) + (1-\theta')J(Y) - J(\theta' X + (1-\theta')Y)\Big) \leq 0$$

à cause de la convexité de J sur $K \cap \{\sum_{i=1}^n x_i = 1\}$. Donc J est convexe sur K.

4. On trouve que

$$\frac{\partial J}{\partial x_i} = c_i + \log \frac{x_i}{\sum_{j=1}^n x_j},$$

ce qui implique que, si x ne tend pas vers 0,

$$\lim_{x_i \to 0} \frac{\partial J}{\partial x_i} = -\infty.$$

Un développement de Taylor à l'ordre 1 montre donc qu'une solution optimale de (1) ne peut pas se trouver sur le bord de K en dehors de l'origine. Par ailleurs, 0 ne peut pas être solution car b est non nul. Comme l'ensemble $\{x \in K,\ Ax = b\}$ n'est pas inclus dans le bord de K, il existe forcément une solution en dehors du bord de K.

5. Comme toute solution de (1) est à l'intérieur de K les multiplicateurs de Lagrange pour les contraintes $x_i \geq 0$ sont nuls. La condition nécessaire d'optimalité est donc qu'il existe un vecteur $x \in K$ et un multiplicateur de Lagrange $\lambda \in \mathbb{R}^m$ tel que

$$J'(x) + A^t\lambda = 0, \quad Ax = b, \quad x_i > 0.$$

Comme la fonction J ainsi que l'ensemble $\{x \in K,\ Ax = b\}$ sont convexes, cette condition d'optimalité est aussi suffisante en vertu du théorème de Kuhn et Tucker.