Ecole Polytechnique **Promotion 2016**

MAP 411 - Approximation numérique et optimisation
CORRECTION de l'examen classant.

Exercice 1

Question 1. On introduit la solution exacte dans le schéma, en remplaçant u_j^n par $u(n\Delta t, j\Delta x)$ et on fait des développements de Taylor. On trouve l'erreur de troncature :

$$F_{\Delta t,\Delta x} = \frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} + \mathcal{O}(\Delta t + \Delta x).$$

Le schéma est donc consistant et d'ordre au moins 1 en espace et en temps.

Question 2. En écrivant le schéma sous la forme :

$$u_j^{n+1} = \left(1 - \frac{2a\Delta t}{\Delta x}\right) u_j^n + \frac{3a\Delta t}{2\Delta x} u_{j-1}^n + \frac{a\Delta t}{2\Delta x} u_{j+1}^n$$

on voit que si $\frac{2a\Delta t}{\Delta x} \leq 1$, alors u_j^{n+1} est une combinaison convexe de u_j^n, u_{j-1}^n, u_{j+1}^n, et donc le schéma satisfait le principe du maximum discret. Il est donc stable en norme L^∞. Ainsi, le schéma est linéaire, stable et consistant. Il est donc convergent en norme L^∞ d'après le théorème de Lax.

Question 3. Pour étudier la stabilité en norme L^2, on introduit un mode de Fourier $\rho_k(n)e^{2i\pi kx}$ dans le schéma. On trouve, pour $k \in \mathbb{Z}$,

$$\rho_k(n+1) = \left[1 - 2\nu + 2\nu\cos(2\pi k\Delta x) + i\nu\sin(2\pi k\Delta x)\right]\rho_k(n) = A_k\rho_k(n)$$

On trouve :

$$|A_k|^2 = P(\xi) = 3\nu^2\xi^2 + 4\nu(1-2\nu)\xi + \nu^2 + (1-2\nu)^2.$$

où $\xi = \cos(2\pi k\Delta x)$. Comme ξ varie dans $[-1,1]$ et $3\nu^2 > 0$, P atteint son maximum en -1 ou +1 :

$$\max_{\xi\in[-1,1]} P(\xi) = \max(P(-1), P(1)) = \max((1-4\nu)^2, 1).$$

En conclusion, si $\nu \leq 1/2$, i.e. si $\frac{2a\Delta t}{\Delta x} \leq 1$, alors le schéma est stable en norme L^2. Comme il est de plus consistant, il est convergent norme L^2.

Question 4. En poursuivant le développement de Taylor de la question 1,

$$F_{\Delta t,\Delta x} = \frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} + \frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2} - a\Delta x\frac{\partial^2 u}{\partial x^2} + \mathcal{O}(\Delta t^2 + \Delta x^2).$$

Si u est solution de l'équation, on a $\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0$ et $\frac{\partial^2 u}{\partial t^2} = a^2\frac{\partial^2 u}{\partial x^2}$, donc :

$$F_{\Delta t,\Delta x} = a\Delta x\left(\frac{a\Delta t}{2\Delta x} - 1\right)\frac{\partial^2 u}{\partial x^2} + \mathcal{O}(\Delta t^2 + \Delta x^2).$$

Sous la condition de stabilité $\frac{2a\Delta t}{\Delta x} \leq 1$ le terme en Δx ne peut pas s'annuler. Le schéma n'est donc pas d'ordre 2.

Exercice 2

Question 1. Le lagrangien associé au problème (3) est donné par :

$$\mathcal{L}(v,\mu) = J(v) + (h(v)-\kappa)\mu, \quad \forall (v,\mu) \in \mathbb{R}^n \times \mathbb{R}_+.$$

Si (u,λ) est un point selle de $\mathcal{L}$. Par définition :

$$J(u) + (h(u)-\kappa)\mu \leq J(u) + (h(u)-\kappa)\lambda \leq J(v) + (h(v)-\kappa)\lambda, \quad \forall v \in \mathbb{R}^n, \forall \mu \in \mathbb{R}_+.$$

Question 2. On est dans les conditions d'application du théorème 4.3.3 (Kuhn et Tucker). Il existe donc $\lambda \in \mathbb{R}_+$ tel que (u,λ) est un point selle de $\mathcal{L}$. En particulier :

$$J(u) + \lambda h(u) \leq J(v) + \lambda h(v), \quad \forall v \in \mathbb{R}^n,$$

i.e. u est solution du problème (4) avec $\alpha = \lambda$.

Question 3. Soit u solution de (4). On a :

$$J(u) + \alpha h(u) \leq J(v) + \alpha h(v), \quad \forall v \in \mathbb{R}^n.$$

Donc pour tout κ,

$$J(u) + \alpha(h(u)-\kappa) \leq J(v) + \alpha(h(v)-\kappa), \quad \forall v \in \mathbb{R}^n.$$

En choisissant $\kappa = h(u)$, on a donc :

$$J(u) + \mu(h(u)-\kappa) \leq J(u) + \alpha(h(u)-\kappa) \leq J(v) + \alpha(h(v)-\kappa), \quad \forall v \in \mathbb{R}^n, \forall \mu \in \mathbb{R}_+$$

Donc (u,α) est un point selle de $\mathcal{L}$. D'après le théorème 4.3.3, u est donc solution de (3) avec $\kappa = h(u)$.

<u>Alternative</u> : on peut également raisonner sur les conditions d'optimalité du premier ordre (Kuhn et Tucker).

Question 4. La difficulté vient de la non dérivabilité de $v \mapsto \|v\|_1$.

Question 5. Le lagrangien associé à (6) est donné par :

$$\mathcal{L}(x,y,\lambda,\mu) = g(x,y) - \lambda\cdot x - \mu\cdot y = J(x-y) + \alpha\mathbb{1}\cdot(x+y) - \lambda\cdot x - \mu\cdot y,$$

où $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^n_+$ et $\mu \in \mathbb{R}^n_+$. La fonction g est convexe et dérivable (composition d'une fonction convexe dérivable et d'une fonction affine), les fonctions définissant les contraintes sont évidemment dérivables et convexes. Les contraintes sont qualifiées en tout point (x,y) car elles sont affines. Le théorème de Kuhn et Tucker s'applique donc. Si (x,y) est solution de (6), il existe $\lambda \in \mathbb{R}^n_+$ et $\mu \in \mathbb{R}^n_+$ tels que (x,y,λ,μ) est un point selle du lagrangien.

On a :

$$\frac{\partial g}{\partial x_i}(x,y) = \frac{\partial J}{\partial v_i}(x-y) + \alpha, \quad \frac{\partial g}{\partial y_i}(x,y) = -\frac{\partial J}{\partial v_i}(x-y) + \alpha$$

Les conditions d'optimalité s'écrivent donc, pour $i = 1, \dots, n$:

$$\begin{cases} \dfrac{\partial J}{\partial v_i}(x-y) + \alpha - \lambda_i = 0, \\ -\dfrac{\partial J}{\partial v_i}(x-y) + \alpha - \mu_i = 0, \\ x_i \geq 0, \lambda_i \geq 0, x_i\lambda_i = 0, \\ y_i \geq 0, \mu_i \geq 0, y_i\mu_i = 0. \end{cases}$$

Question 6. Soit (x, y) solution de (6). D'après les conditions d'optimalité, on a :

$$\lambda_i + \mu_i = 2\alpha, \quad i = 1, \dots, n.$$

Comme $\alpha > 0$, λ_i et μ_i ne peuvent pas être tous les deux nuls. Or, si $x_i \neq 0$ alors $\lambda_i = 0$, donc $\mu_i \neq 0$. Mais comme $y_i\mu_i = 0$, cela implique que $y_i = 0$. On montre de même que si $y_i \neq 0$ alors $x_i = 0$. On en déduit que $x_iy_i = 0$, $\forall i = 1, \dots, n$.

Question 7. Soit (z, t) solution de (6). On pose $u = z - t$. Comme $z_i \geq 0$, $t_i \geq 0$, $z_it_i = 0$, $\forall i = 1, \dots, n$, on a $\mathbb{1} \cdot (z + t) = \|u\|_1$. Donc :

$$g(z, t) = J(u) + \alpha\|u\|_1.$$

Soit $v \in \mathbb{R}^n$, on pose $x = v_+$ et $y = v_-$, les parties positive et négative de v. Ainsi $x - y = v$ et $\mathbb{1} \cdot (x + y) = \|v\|_1$. Par définition de (z, t), on a $g(z, t) \leq g(x, y)$. Or $g(x, y) = J(v) + \alpha\|v\|_1$. On a donc bien $J(u) \leq J(v)$, $\forall v \in \mathbb{R}^n$.

Question 8. Le problème (6) a un inconvénient évident : sa taille est le double de celle de (5). En revanche, il a l'avantage de ne faire intervenir que des fonctions dérivables. On peut donc par exemple utiliser une méthode de gradient avec projection (les contraintes étant très faciles à prendre en compte), ou un algorithme d'Uzawa (maximisation de l'énergie duale par une méthode de gradient).

Exercice 3

Question 1. Soit u solution de (7). On multiplie (7) par $v \in V$ et on intègre sur $(0, L)$:

$$-\eta \int_0^L u''(x)v(x)\,dx + w\int_0^L u'(x)v(x)\,dx = \int_0^L f(x)v(x)\,dx.$$

En intégrant par parties le premier terme, et en utilisant le fait que les fonctions s'annulent en 0 et L, on trouve le résultat voulu.

Question 2. On a par définition de u_h :

$$a(u_h, v_h) = L(v_h), \forall v_h \in V_h.$$

Comme de plus $V_h \subset V$ et u est solution de (8) :

$$a(u, v_h) = L(v_h), \forall v_h \in V_h.$$

Par linéarité, on en déduit que $a(\epsilon_h, v_h) = 0, \forall v_h \in V_h$. Donc, comme $e_h \in V_h$, $a(\epsilon_h, e_h) = 0$. D'où le résultat puisque $\epsilon_h = e_h + \pi_h$.

Question 3. Soit $v_h \in V_h$, on a :

$$a(v_h, v_h) = \eta \int_0^L (v_h'(x))^2 \, dx + \int_0^L w v_h(x) v_h'(x) \, dx.$$

Or

$$\int_0^L w v_h(x) v_h'(x) \, dx = \frac{w}{2} \int_0^L (v_h^2(x))' \, dx = \frac{w}{2} \left[v_h^2(x) \right]_0^L = 0,$$

car les fonctions de V_h s'annulent en 0 et L. D'où le résultat.

Question 4. On commence par une intégration par parties :

$$a(\pi_h, e_h) = \int_0^L (\eta \pi_h' e_h' + w \pi_h' e_h) = \int_0^L (\eta \pi_h' e_h' - w \pi_h e_h'),$$

puis on applique les inégalités de Cauchy-Schwarz et Young avec $\alpha = 2$ puis avec $\alpha = 2/\eta$:

$$|a(\pi_h, e_h)| \leq \eta \int_0^L (\pi_h')^2 + \frac{\eta}{4} \int_0^L (e_h')^2 + \frac{w^2}{\eta} \int_0^L (\pi_h)^2 + \frac{\eta}{4} \int_0^L (e_h')^2,$$

d'où l'inégalité voulue.

Question 5. D'après les trois questions précédentes :

$$\eta \int_0^L (e_h')^2 = |a(\pi_h, e_h)| \leq \frac{\eta}{2} \int_0^L (e_h')^2 + \eta \int_0^L (\pi_h')^2 + \frac{w^2}{\eta} \int_0^L (\pi_h)^2.$$

Or

$$\int_0^L (\pi_h')^2 = \sum_{K \in \mathcal{T}_h} \int_K (\pi_h')^2 \leq C_0^2 \sum_{K \in \mathcal{T}_h} h_K^{2k} \int_K (u^{(k+1)})^2 \leq C_0^2 h^{2k} \|u^{(k+1)}\|_{L^2}^2,$$

et

$$\int_0^L \pi_h^2 = \sum_{K \in \mathcal{T}_h} \int_K \pi_h^2 \leq C_0^2 \sum_{K \in \mathcal{T}_h} h_K^{2k+2} \int_K (u^{(k+1)})^2 \leq C_0^2 h^{2k+2} \|u^{(k+1)}\|_{L^2}^2.$$

Donc :

$$\frac{\eta}{2} \int_0^L (e_h')^2 \leq \eta C_0^2 h^{2k} \|u^{(k+1)}\|_{L^2}^2 + \frac{w^2}{\eta} C_0^2 h^{2k+2} \|u^{(k+1)}\|_{L^2}^2,$$

d'où le résultat voulu avec $C_1 = \sqrt{2} C_0$:

$$\|e_h'\|_{L^2} \leq \sqrt{2} C_0 \sqrt{1 + \frac{w^2 h^2}{\eta^2}} \|u^{(k+1)}\|_{L^2} h^k.$$

Question 6. La majoration de $\|e_h'\|_{L^2}$ est optimale, dans le sens où elle se comporte en h^k, c'est-à-dire avec la même puissance de k que l'erreur d'interpolation π_h mesurée dans la même norme. Cependant, la constante qui multiplie h^k peut être grande quand $\eta \ll |w| h$, c'est-à-dire quand le transport domine la diffusion à l'échelle du pas de discrétisation h.

Question 7. Soit $u \in V$ solution de (8). Comme $V_h \subset V$, on a

$$a(u, v_h) = L(v_h), \forall v_h \in V_h.$$

Mais comme une solution de (8) est aussi solution de (7),

$$\sum_{K \in \mathcal{T}_h} \int_K w v_h'(x)(-\eta u_h''(x) + w u_h'(x))\, dx = \sum_{K \in \mathcal{T}_h} \tau_K \int_K w v_h'(x) f(x)\, dx$$

et donc u est également solution du problème (12) :

$$a_h(u, v_h) = L_h(v_h), \forall v_h \in V_h.$$

Comme par ailleurs, on a $a_h(u_h, v_h) = L_h(v_h)$, $\forall v_h \in V_h$, on en déduit à nouveau par linéarité :

$$a_h(\epsilon_h, v_h) = 0, \forall v_h \in V_h, \quad \text{et} \quad a_h(e_h, e_h) = -a_h(\pi_h, e_h).$$

Question 8. On a $\tau_K = \frac{\lambda h_K}{|w|}$ et $\frac{1}{|w|} < \frac{h_K}{\eta}$, donc

$$\tau_K \leq \lambda \frac{h_K^2}{\eta}.$$

On a :

$$a_h(v_h, v_h) = a(v_h, v_h) + \sum_{K \in \mathcal{T}_h} \tau_K \int_K w v_h'(x)(-\eta v_h''(x) + w v_h'(x))\, dx.$$

On a déjà vu que $a(v_h, v_h) = \eta \int_0^L (v_h'(x))^2\, dx$. Par ailleurs :

$$\begin{aligned}
\left| \sum_{K \in \mathcal{T}_h} \tau_K \int_K w v_h'(x)(-\eta v_h''(x)) \right| &\leq \frac{1}{2} \int_0^L \tau(x) w^2 (v_h'(x))^2\, dx + \sum_{K \in \mathcal{T}_h} \frac{\tau_K}{2} \int_K \eta^2 (v_h''(x))^2\, dx \\
&\leq \frac{1}{2} \int_0^L \tau(x) w^2 (v_h'(x))^2\, dx + \sum_{K \in \mathcal{T}_h} \frac{\lambda h_K^2}{2\eta} \int_K \eta^2 (v_h''(x))^2\, dx \\
&\leq \frac{1}{2} \int_0^L \tau(x) w^2 (v_h'(x))^2\, dx + \sum_{K \in \mathcal{T}_h} \frac{\lambda d_0 \eta}{2} \int_K (v_h'(x))^2\, dx.
\end{aligned}$$

On en déduit :

$$\begin{aligned}
a(v_h, v_h) &= \eta \int_0^L (v_h'(x))^2\, dx + \int_0^L \tau(x) w^2 (v_h'(x))^2\, dx + \sum_{K \in \mathcal{T}_h} \tau_K \int_K w v_h'(x)(-\eta v_h''(x))\, dx \\
&\geq \eta \int_0^L (v_h'(x))^2\, dx + \int_0^L \tau(x) w^2 (v_h'(x))^2\, dx - \left| \sum_{K \in \mathcal{T}_h} \tau_K \int_K w v_h'(x)(-\eta v_h''(x))\, dx \right| \\
&\geq \eta \left(1 - \frac{\lambda d_0}{2}\right) \int_0^L (v_h'(x))^2\, dx + \frac{1}{2} \int_0^L \tau(x) w^2 (v_h'(x))^2\, dx.
\end{aligned}$$

Question 9. D'après la question précédente, en choisissant $\lambda \leq 1/d_0$, on a la coercivité de a_h pour la norme $\| \cdot \|_h$ avec une constante $1/2$.

Question 10. On a vu que $a_h(e_h, e_h) = -a_h(\pi_h, e_h)$. Avec la coercivité de a_h on a donc :

$$\frac{1}{2}\|e_h\|_h^2 \leq -\eta\int_0^L \pi_h'(x)e_h'(x)\,dx - \int_0^L w\pi_h'(x)e_h(x)\,dx - \sum_{K\in\mathcal{T}_h}\int_K \tau_K\big(-\eta\pi_h''(x) + w\pi_h'(x)\big)we_h'(x)\,dx.$$

En intégrant par parties le second terme du second membre et en utilisant l'inégalité triangulaire on a donc :

$$\|e_h\|_h^2 \leq 2\left|\eta\int_0^L \pi_h'(x)e_h'(x)\,dx\right| + 2\left|\int_0^L w\pi_h(x)e_h'(x)\,dx\right| + 2\left|\sum_{K\in\mathcal{T}_h}\int_K \tau_K\eta w\pi_h''(x)e_h'(x)\,dx\right| + 2\left|\int_0^L \tau(x)w^2\pi_h'(x)e_h'(x)\,dx\right|.$$

Question 11. On étudie chacun des termes de l'inégalité précédente avec les inégalités de Cauchy-Schwarz et de Young :

$$2\left|\eta\int_0^L \pi_h'(x)e_h'(x)\,dx\right| \leq \frac{\eta}{\alpha_1}\int_0^L (e_h'(x))^2\,dx + \alpha_1\eta\int_0^L (\pi_h'(x))^2\,dx,$$

et

$$\begin{aligned}2\left|\int_0^L w\pi_h(x)e_h'(x)\,dx\right| &\leq \sum_{K\in\mathcal{T}_h}\left(\frac{\tau_K}{\alpha_2}\int_K (we_h'(x))^2\,dx + \frac{\alpha_2}{\tau_K}\int_K (\pi_h(x))^2\,dx\right)\\ &\leq \frac{1}{\alpha_2}\int_0^L \tau(x)w^2(e_h'(x))^2\,dx + \sum_{K\in\mathcal{T}_h}\frac{\alpha_2}{\tau_K}\int_K (\pi_h(x))^2\,dx.\end{aligned}$$

En utilisant de plus (14) et (15) :

$$\begin{aligned}2\left|\sum_{K\in\mathcal{T}_h}\int_K \tau_K\eta w\pi_h''(x)e_h'(x)\,dx\right| &\leq \sum_{K\in\mathcal{T}_h}\left(\frac{\tau_K}{\alpha_3}\int_K (we_h'(x))^2\,dx + \alpha_3\tau_K\eta^2\int_K (\pi_h''(x))^2\,dx\right)\\ &\leq \frac{1}{\alpha_3}\int_0^L \tau(x)w^2(e_h'(x))^2\,dx + \sum_{K\in\mathcal{T}_h}\alpha_3\lambda\eta h_K^2\int_K (\pi_h''(x))^2\,dx\\ &\leq \frac{1}{\alpha_3}\int_0^L \tau(x)w^2(e_h'(x))^2\,dx + \alpha_3\lambda\eta d_0\int_0^L (\pi_h'(x))^2\,dx.\end{aligned}$$

Enfin :

$$2\left|\int_0^L \tau(x)w^2\pi_h'(x)e_h'(x)\,dx\right| \leq \frac{1}{\alpha_4}\int_0^L \tau(x)w^2(e_h'(x))^2\,dx + \alpha_4\sum_{K\in\mathcal{T}_h}\tau_K\int_K (w\pi_h'(x))^2\,dx.$$

On choisit maintenant $\alpha_1 = 2$, $\alpha_2 = \alpha_3 = \alpha_4 = 6$ pour absorber les termes $\int_0^L (e_h'(x))^2\,dx$ et $\int_0^L \tau(x)w^2(e_h'(x))^2\,dx$ dans le membre de gauche de (16). On en déduit :

$$\|e_h\|_h^2 \leq c_1\eta\int_0^L (\pi_h'(x))^2\,dx + c_2\sum_{K\in\mathcal{T}_h}\tau_K\int_K (w\pi_h'(x))^2\,dx + c_3\sum_{K\in\mathcal{T}_h}\frac{1}{\tau_K}\int_K (\pi_h(x))^2\,dx$$

avec $c_1 = 2\alpha_1 + 2\alpha_3\lambda d_0$, $c_2 = 2\alpha_4$, $c_3 = 2\alpha_2$

Question 12. On déduit de la question précédente et des majorations des erreurs d'interpolation données dans l'énoncé :

$$\eta\|e_h'\|_{L^2}^2 \leq \left(c_1\eta C_0^2 h^{2k} + c_2\lambda h|w|C_0^2 h^{2k} + c_3\frac{|w|}{\lambda\gamma h}C_0^2 h^{2k+2}\right)\|u^{(k+1)}\|_{L^2}^2$$

d'où

$$\|e_h'\|_{L^2}^2 \leq \left(c_1 C_0^2 + c_2\lambda C_0^2\frac{|w|h}{\eta} + \frac{c_3 C_0^2}{\lambda\gamma}\frac{|w|h}{\eta}\right)h^{2k}\|u^{(k+1)}\|_{L^2}^2$$

et donc

$$\|e_h'\|_{L^2}^2 \leq C_2^2\left(1 + \frac{|w|h}{\eta}\right)h^{2k}\|u^{(k+1)}\|_{L^2}^2$$

avec $C_2^2 = \max(2c_1C_0^2, 2c_2\lambda C_0^2 + \frac{2c_3C_0^2}{\lambda\gamma})$, d'où le résultat voulu. La constante dans la majoration de l'erreur est maintenant $\frac{|w|h}{\eta}$ au lieu de $\frac{w^2h^2}{\eta^2}$ pour la méthode de Galerkin classique. Dans des régimes où l'advection domine beaucoup la diffusion à l'échelle h, on s'attend donc à de meilleurs résultats avec cette nouvelle formulation, ce qu'on observe effectivement en pratique.