Chapitre 3

Equation de la Chaleur

3.1 Modélisation

Pour modéliser la diffusion de la chaleur par un dispositif quelconque enfermé dans un volume V, le critère choisi est celui de l'évolution de la température $T(x,t)$ répartie dans le volume à l'instant $t>0$. En admettant qu'en l'absence de forces extérieures, le seul phénomène à prendre en compte est le flux de chaleur au travers de la surface ∂V du volume généré par le gradient de température, on obtient l'équation de convservation :

$$\frac{d}{dt}\int_V T(x,t)d\Omega_x = \int_{\partial V} \vec{\nabla} T \cdot \overrightarrow{dS}$$

où le vecteur $\overrightarrow{dS}$ est orienté dans le sens de la normale extérieure au volume V, en accord avec l'observation que la température du volume V augmente à mesure qu'il diffuse de la chaleur autour de lui, i.e. la température cumulée $\int_V T(x,t)d\Omega_x$ augmente dès que $\vec{\nabla} T \cdot \overrightarrow{dS} > 0$ dans V. D'après la formule de Stokes :

$$\int_{\partial V} \underbrace{\vec{\nabla} T \cdot \overrightarrow{dS}}_{:=\omega} = \int_V d\omega = \int_V \operatorname{div}(\vec{\nabla} V) d\Omega_x.$$

On suppose de plus que l'application $(x,t) \mapsto T(x,t)$ est suffisamment régulière pour écrire :

$$\frac{d}{dt}\int_V T(x,t)d\Omega_x = \int_V \frac{\partial}{\partial t} T(x,t)d\Omega_x.$$

On en déduit, le volume V étant arbitraire :

$$\frac{\partial}{\partial t} T(x,t) = \underbrace{\operatorname{div}(\vec{\nabla} T)}_{=\Delta T} \quad \text{dans} \quad \mathbb{R}^n$$

i.e., par définition de l'opérateur Δ :

$$\frac{\partial}{\partial t}T(x,t) = \Delta T \quad \text{dans} \quad \mathbb{R}^n$$

3.2 Existence et unicité

Soit $\Omega \subset \mathbb{R}^n$ un ouvert de frontière $\partial\Omega$ assez régulière, par exemple $\mathcal{C}^1$ par morceaux. Soit $f : \Omega \times]0,+\infty[\to \mathbb{R}$, $u_0 : \Omega \to \mathbb{R}$. On considère le problème : trouver $u : \Omega \times]0,+\infty[\to \mathbb{R}$ solution de :

$$\begin{cases} \dfrac{\partial u}{\partial t} - \Delta u = f & \text{dans} \quad \Omega \times]0,+\infty[, \\ u = 0 & \text{sur} \quad \partial\Omega \times]0,+\infty[, \\ u(x,0) = u_0(x) & \text{dans} \quad \Omega \end{cases} \tag{3.1}$$

Proposition 3.2.1. *Si $u_0 \in L^2(\Omega)$ et si $f \in L^2(\Omega \times]0,+\infty[)$, le problème (3.1) admet une unique solution $u \in L^2(]0,+\infty[, H_0^1(\Omega)) \cap \mathcal{C}([0,+\infty[, L^2(\Omega))$ telle que, de façon équivalente à (3.1) :*

$$\forall v \in H_0^1(\Omega), \quad \frac{d}{dt}\int_\Omega u(t)v dx + \int_\Omega \nabla u \nabla v dx = \int_\Omega f(t) v dx.$$

Démonstration. Soit $\varphi \in \mathcal{C}_c^1(\Omega)$. Par intégration par parties sur Ω, on obtient :

$$\frac{d}{dt}\int_\Omega u(t)\varphi dx + \int_\Omega \nabla u(t) \nabla \varphi dx = \int_\Omega f(t)\varphi dx. \tag{3.2}$$

Par densité de $\mathcal{C}_c^1(\Omega)$ dans $L^2(\Omega)$ et dans $H_0^1(\Omega)$, la formulation variationnelle (3.2) se généralise à $\varphi = v \in H_0^1(\Omega)$. On retrouve (3.1) à partir de (3.2) immédiatement par intégration par parties de (3.2). Le choix $v = u(t)$ dans (3.2) conduit à, pour tout $t > 0$:

$$\int_\Omega |u(t)|^2 dx + \int_0^t \int_\Omega |\nabla u(s)|^2 dx ds = \int_0^t \int_\Omega f(s)u(s) dx ds + \int_\Omega |u_0|^2 dx$$

$$\Rightarrow \sup_{t\in[0,+\infty[} \int_\Omega |u(t)|^2 dx + \int_0^{+\infty}\int_\Omega |\nabla u|^2 dx dt \leq \|f\|_2 \|u\|_2 + \int_\Omega |u_0|^2 dx$$

i.e. :

$$\sup_{t\in[0,+\infty[} \|u(t)\|_{L^2(\Omega)} + \|\nabla u\|_{L^2(\Omega\times]0,+\infty[)} < +\infty.$$

$\square$

Proposition 3.2.2. *Si $u_0 \in L^2(\Omega)$ et si $f = 0$, alors la solution u de (3.1) vérifie :*

$$u \in \mathcal{C}^1(]0, +\infty[, L^2(\Omega))$$

et

$$u \in \mathcal{C}^\infty([\varepsilon, +\infty[\times\overline{\Omega}), \quad \forall\varepsilon > 0.$$

Démonstration. Brézis Théorème X.1. □

Proposition 3.2.3 (Principe du maximum)**.** *Si $u_0 \in L^2(\Omega)$ et si $f = 0$, alors la solution u de (3.1) vérifie :*

$$\min(0, \inf_\Omega u_0) \leq u \leq \max(0, \sup_\Omega u_0).$$

1. *Si $u_0 \geq 0$ p.p. dans Ω, alors $u \geq 0$ dans $\Omega\times]0, +\infty[$.*
2. *Si $u_0 \in L^\infty(\Omega)$, alors $u \in L^\infty(\Omega\times]0, +\infty[)$ et*

$$\|u\|_\infty \leq \|u_0\|_\infty.$$

Démonstration. Dans le cas où $\Omega = \mathbb{R}^n$, c'est une conséquence directe de l'expression de u obtenue explicitement. En supposant que u admet une transformée de Fourier à tout instant $t > 0$, soit

$$\hat{u}(\xi, t) = \int_\mathbb{R} e^{-i\xi x} u(x, t)dx, \quad \xi \in \mathbb{R}^n, \quad t > 0$$

ainsi que ses dérivées, on obtient l'équation avec condition initiale :

$$\frac{\partial}{\partial t}\hat{u}(\xi, t) + |\xi|^2\hat{u}(\xi, t) = 0, \quad \hat{u}(\xi, 0) = \hat{u}_0(\xi), \quad \xi \in \mathbb{R}^n, \quad t > 0.$$

de solution :

$$\hat{u}(\xi, t) = e^{-|\xi|^2 t}\hat{u}_0(\xi), \quad \xi \in \mathbb{R}^n, \quad t > 0.$$

On en déduit, par transformation de Fourier inverse :

$$u(x, t) = \frac{1}{4\pi t}\int_\mathbb{R} e^{-\frac{|x-y|^2}{4t}} u_0(y)dy, \quad x \in \mathbb{R}^n, \quad t > 0. \tag{3.3}$$

□

Définition 3.2.1. Pour tout $t > 0$, on appelle noyau de la chaleur l'application :

$$K_t : x \mapsto \frac{1}{4\pi t}e^{-\frac{x^2}{4t}}$$

Dans la suite, on considère le problème lorsque $n = 1$ car c'est la discrétisation en temps qui nous intéresse ici. Plus précisément, soit $T > 0$. On considère le problème avec conditions aux limites :

$$\begin{cases} \dfrac{\partial u}{\partial t} - \dfrac{\partial^2 u}{\partial x^2} = 0 & \text{dans} \quad]0, L[\times]0, +\infty[=: Q, \\ u(x, 0) = u_0(x) & \text{dans} \quad]0, L[\\ u(0, t) = u(L, t) = 0 & \text{dans} \quad]0, +\infty[\end{cases} \tag{3.4}$$

Théorème 3.2.4 (Existence et unicité)**.** *Soit $u_0 \in \mathcal{C}(]0, L[, \mathbb{R})$. Alors, il existe une unique fonction $u \in \mathcal{C}^2(]0, L[\times]0, +\infty[, \mathbb{R}) \cap \mathcal{C}([0, L] \times [0, +\infty[, \mathbb{R})$ solution de (3.4).*

Remarque 12 (Effet régularisant de l'équation de la chaleur). Si u est solution de (3.4) et si $u_0 \in \mathcal{C}(]0, L[, \mathbb{R})$, alors $u \in \mathcal{C}^\infty(]0, L[\times]0, T[)$.

Proposition 3.2.5 (Principe du maximum)**.** *Sous les hypothèses du Théorème 3.2.4, la solution u du problème (3.4) vérifie*

1. *Si $u_0(x) \geq 0$, $\forall x \in [0, L]$, alors $u(x, t) \geq 0$, $\forall t \geq 0$, $\forall x \in]0, L[$.*
2. *$\|u\|_{L^\infty(]0,L[\times]0,+\infty[} \leq \|u_0\|_{L^\infty(]0,L[)}$.*

Résolution analytique par les séries de Fourier

Proposition 3.2.6. *On suppose que $u_0 \in \mathcal{C}^1([0, L], \mathbb{R})$ et que $u_0(0) = u_0(L) = 0$. Alors la solution du problème (3.4) se développe en série de Fourier sous la forme :*

$$u(x, t) = \sum_{n \in \mathbb{Z}} c_n(u_0) e^{-(n\pi)^2 t} \sin\left(\frac{n\pi x}{L}\right) \tag{3.5}$$

où $(c_n(u_0))_{n \in \mathbb{Z}}$ est la suite des coefficients de Fourier de u_0 définis par :

$$c_n(0) = \frac{2}{L} \int_0^L u_0(x) \sin\left(\frac{n\pi x}{L}\right) dx, \quad \forall n \geq 0.$$

Démonstration. On commence par chercher u sous la forme $u(x, t) = X(x)T(t)$ où X et T sont de classe $\mathcal{C}^2$, en accord avec le Thérème 3.2.4. Après report dans (3.4), on obtient :

$$\frac{X''}{X} = \frac{T'}{T} = \lambda \in \mathbb{R}.$$

Si $\lambda = \omega^2 > 0$ avec $\omega > 0$, alors

$$X(x) = ae^{\omega x} + be^{\omega x}, \quad a, b \in \mathbb{R}$$

avec : $X(0) = X(L) = 0$, ce qui entraîne. que (a, b) est slution du système :

$$\begin{cases} a + b & = \quad 0, \\ ae^{\omega L} + be^{-\omega L} = 0 \end{cases}$$

dont l'unique solution est $a = b = 0$, en contradiction avec $u \not\equiv 0$. Donc $\lambda = -\omega^2 < 0$ et on a :

$$X(x) = a\cos(\omega x) + b\sin(\omega x)$$

avec : $X(0) = 0 \Rightarrow a = 0$. Il reste : $X(L) = b\sin(\omega L) = 0$ et donc $\omega \in \frac{\pi}{L}\mathbb{N}$. On obtient dons la suite de solutions $(u_n)_{n\in\mathbb{N}}$ définies par :

$$u_n(x,t) = e^{-(\frac{n\pi}{L})^2 t}\sin\left(\frac{n\pi x}{L}\right), \quad \forall (x,t) \in Q.$$

Il reste à vérifier la condition au bord en $t = 0$, ce qu'on cherche a priori pour $u = \sum_{n\in\mathbb{N}} b_n u_n$. Formellement, on trouve que :

$$u(x,0) = \sum_{n\in\mathbb{N}} b_n u_n(x,0) = \sum_{n\in\mathbb{N}} b_n \sin\left(\frac{n\pi x}{L}\right) = u_0(x).$$

Le problème admet une solution ssi u_0 coïncide avec sa série de Fourier et si cette dernière est impaire. Comme $u_0(0) = 0$, on prolonge u_0 en une fonction impaire de classe $\mathcal{C}^1$ sur $[-L, L]$. La condition $u_0(L) = 0$ permet de prolonger u_0 par périodicité à $\mathbb{R}$ en une fonction périodique de période $2L > 0$, continue et $\mathcal{C}^1$ par morceaux. On en déduit que la série de Fourier de u_0 est absolument convergente sur $\mathbb{R}$ vers u_0 et que la série des coefficients est dans ℓ^1. Il reste à vérifier qu'on peut dériver sous le signe somme dans $u = \sum_{n\in\mathbb{N}} b_n u_n$ lorsque $(b_n)_{n\in\mathbb{N}} \in \ell^1$. Soit $T > 0$. On a :

$$|b_n u_n(x,t)| \le |b_n| e^{-(\frac{n\pi}{L})^2 t} \le |b_n| e^{-(\frac{n\pi}{L})^2 T} \le C e^{-(\frac{n\pi}{L})^2 T}, \quad \forall (x,t) \in \mathbb{R} \times [0,T]$$

et la série majorante est convergente. On en déduit que la série de fonctions continues $\sum b_n u_n$ est uniformément convergente sur tout compact de $[0, L] \times \mathbb{R}^+$ donc de somme continue sur $[0, L] \times \mathbb{R}^+$. On a aussi : $\forall k, \ell > 0$,

$$\left|\frac{\partial^{k+\ell}}{\partial t^k \partial x^\ell} u_n(x,t)\right| \le C\left(\frac{n\pi}{L}\right)^{2k+\ell} e^{-(\frac{n\pi}{L})^2 T}, \quad \forall (x,t) \in \mathbb{R} \times [0,T]$$

où la série majorante est convergente. On en déduit que la série des dérivées partielles $\sum \frac{\partial^{k+\ell}}{\partial t^k \partial x^\ell} u_n$ est uniformément convergente sur tout compact de $[0, L] \times \mathbb{R}^+$, donc que la somme de la série $\sum b_n u_n$ est de classe $\mathcal{C}^\infty$ sur $[0, L] \times \mathbb{R}^+$. On conclut par unicité de la série de Fourier de u_0. $\square$

Proposition 3.2.7. *Si $u_0 \in \mathcal{C}^1([0,L],\mathbb{R})$, alors (3.5) est l'unique solution de (3.4).*

Démonstration. C'est une conséquence du principe du maximum. □

3.3 Approximation numérique

Le Schéma d'Euler explicite

Soit $L, T > 0$. On considère les subdivisions :

$$x_0 = 0 < x_1 < \cdots < x_I < x_{I+1} = L, \quad 0 = t_0 < t_1 < \cdots < t_N < t_{N+1} = T.$$

Pour simplifier les notations, on suppose que les subdivisions sont régulières de pas $\Delta x > 0$, $\Delta > 0$. Soit $(u_i^n)_{0\leq i\leq I+1, 0\leq n\leq N+1}$ la suite solution du schéma :

$$\begin{cases} \dfrac{1}{\Delta t}(u_i^{n+1} - u_i^n) + \dfrac{1}{\Delta x^2}(-u_{i-1}^n + 2u_i^n - u_{i-1}^n) = f(x_i), & i = 1, \cdots, I, \\ & n = 1, \cdots N, \\ \\ u_i^0 = u_0(x_i) & i = 1, \cdots I, \\ u_0^n = u_{I+1}^n = 0 & n = 1, \cdots, N \end{cases} \tag{3.6}$$

Définition 3.3.1 (Consistance)**.** On définit l'erreur de consistance au point (x_i, t_n), par :

$$\varepsilon_i^n = \frac{1}{\Delta t}(u(x_i,t_{n+1}) - u(x_i,t_n)) + \frac{1}{\Delta x^2}(-u(x_{i-1},t_n) + 2u(x_i,t_n) - u(x_{i+1},t_n)) - f(x_i).$$

Le schéma est dit consistant si

$$\lim_{(\Delta t,\Delta x)\to(0,0)} \max_{1\leq i\leq I, 1\leq n\leq N} |\varepsilon_i^n| = 0.$$

Proposition 3.3.1. *On pose :*

$$\|\varepsilon^n\|_\infty = \max_{1\leq i\leq I} |\varepsilon_i^n|, \quad \forall n \geq 0.$$

Si $u \in \mathcal{C}^4(]0,L[\times]0,T[)$, alors

$$\sup_{1\leq n\geq N} \|\varepsilon^n\|_\infty \leq C\|u^{(4)}\|_\infty(\Delta t + \Delta x^2)$$

Démonstration. On pose :

$$\varepsilon(x,t) = \frac{1}{\Delta t}(u(x,t+\Delta t)-u(x,t)+\frac{1}{\Delta x^2}(-u(x-\Delta x,t)+2u(x,t)-u(x+\Delta x,t))-f(x).$$

La formule de Taylor donne :

$$\varepsilon(x,t) = \frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2}u(x,t) - \frac{(\Delta x)^2}{12}\frac{\partial^4 u}{\partial x^4}u(x,t) + o(\Delta t) + o((\Delta x)^2) \qquad (3.7)$$

□

Définition 3.3.2 (Convergence)**.** On définit l'erreur de convergence par :

$$e_i^n = u(x_i,t_n) - u_i^n, \quad i = 1,\cdots,I, \quad n = 1,\cdots,N.$$

et on pose :

$$\|e^n\|_\infty = \max_{i=1,\cdots,I} \|e_i^n|, \quad n = 1,\cdots,N.$$

Le schéma est dit convergent si

$$\lim_{(\Delta t,\Delta x)\to(0,0)} \|e^n\|_\infty = 0.$$

Proposition 3.3.2. *On suppose que* $u \in \mathcal{C}^4([0,L]\times[0,T])$ *et que*

$$0 < \frac{\Delta t}{(\Delta x)^2} < \frac{1}{2}. \qquad (3.8)$$

Alors :

$$\max_{1\le n\le N} \|e^n\|_\infty \le CT\|u^{(4)}\|_\infty(\Delta t + (\Delta x)^2).$$

Démonstration. Soit (μ_i^n) une suite de réels et soit (z_i^n) la suite définie par le schéma :

$$\begin{cases} \frac{1}{\Delta t}(z_i^{n+1} - z_i^n) + \frac{1}{\Delta x^2}(-z_{i-1}^n + 2z_i^n - z_{i-1}^n) = \mu_i^n, & i = 1,\cdots,I, \quad n = 1,\cdots N, \\ z_i^0 \in \mathbb{R} & i = 1,\cdots I, \\ z_0^n = z_{I+1}^n = 0 & n = 1,\cdots,N \end{cases} \qquad (3.9)$$

Alors :

$$z_i^{n+1} = \left(1 - 2\frac{\Delta t}{(\Delta x)^2}\right) z_i^n + \frac{\Delta t}{(\Delta x)^2}\left(z_{i-1}^n + z_{i+1}^n\right) + \Delta t \mu_i^n, \quad n = 1,\cdots,N, \quad i = 1,\cdots,I.$$

On en déduit, compte tenu de (3.8) :

$$\|z^{n+1}\|_\infty \leq \|z^n\|_\infty + \Delta t\|\mu^n\|_\infty \leq \|z^0\|_\infty + \Delta t\sum_{k=0}^{n}\|\mu^k\|_\infty$$

Si $\mu_i^n = f(x_i)$, $\forall n \in [[1, N]]$ et si $z_i^0 = u_0(x_i)$, $i = 1, \cdots, I$, alors $z_i^n = u_i^n$ et on en déduit :

$$\|u^{n+1}\|_\infty \leq \|u^n\|_\infty + \Delta t\|f\|_\infty$$

i.e., le schéma est stable, donc convergent puisque consistant. Si $z_i^n = e_i^n$, alors $\mu_i^n = \varepsilon_i^n$ est l'erreur de consistance et $z_i^0 = 0 \Rightarrow$

$$\|e^n\|_\infty \leq C\|u^{(4)}\|_\infty N\Delta t(\Delta t + (\Delta x)^2) \leq C\|u^{(4)}\|_\infty T(\Delta t + (\Delta x)^2).$$

□

Définition 3.3.3 (Erreur d'arrondis)**.** On appelle erreur sur les arrondis au point (x_i, t_n) le terme z_i^n dans la suite définie par le schéma (3.9) lorsque $\mu_i^n = 0$.

Proposition 3.3.3. *Le schéma (3.6) est stable au sens des erreurs d'arrondi.*

Démonstration. Soit (z_i^n) la suite résultant du schéma (3.9). On pose :

$$Z^{(n)} = \begin{pmatrix} z_1^n \\ \vdots \\ z_N^n \end{pmatrix} \in \mathbb{R}^N.$$

Le schéma (3.9) se réécrit sous forme matricielle :

$$Z^{n+1} = \underbrace{\left(I - \frac{\Delta t}{(\Delta x)^2}A_N\right)}_{=:B_N} Z^n + \Delta t\,\mu_i^n, \quad n \geq 0,$$

où $A_N \in \mathbb{R}^{N\times N}$ est la matrice (2.13). On en déduit :

$$Z^{(n)} = B_N^n Z^{(0)} + \sum_{k=0}^{n-1} B_N^k \mu^{n-k}, \quad \forall n \geq 1.$$

La matrice B_N étant symétrique réelle, on a :

$$\|B^n\|_2 = \rho(B_N^n) = \rho(B_N)^n.$$

D'après (2.17), les valeurs propres $\lambda_i^{(N)}$, $i = 1, \cdots, N$ de A_N vérifient :

$$0 < \lambda_1^{(N)} = 4\sin\left(\frac{\pi}{(N+1)}\right)^2 < \cdots < \lambda_N^{(N)} = 4\cos\left(\frac{\pi}{(N+1)}\right)^2 < 4$$

On en déduit :

$$-1 \underset{(3.8)}{<} 1 - \frac{\Delta t}{(\Delta x)^2}\lambda_N^{(N)} < \cdots < 1 - \frac{\Delta t}{(\Delta x)^2}\lambda_1^{(N)} < 1$$

i.e. $\rho(B_N) < 1$ et donc le schéma est stable. $\square$

Proposition 3.3.4 (Stabilité au sens de Von Neumann)**.** *On suppose que* $\sum_{k\in\mathbb{Z}} |c_k| < +\infty$ *et que (3.8) est vérifié. Alors, le schéma (3.6) converge au sens de Von Neumann, i.e. :*

$$\lim_{(\Delta t, \Delta x)\to(0,0)} \|e^{N+1}\|_\infty = 0.$$

Démonstration. Pour simplifier les notations, on pose $L = \pi$. On suppose que la donnée initiale u_0 dans (3.4) coïncide avec son développement en série de Fourier, soit :

$$u_0(x) = \sum_{k\in\mathbb{Z}} c_k e^{ikx}, \quad \forall x \in [0, \pi].$$

Alors, le schéma (3.6) est défini avec :

$$u_j^0 = \sum_{k\in\mathbb{Z}} c_k e^{ijk\Delta x}, \quad j = 0, \cdots, N+1.$$

On en déduit : $\forall j \in [[1, N]]$,

$$u_j^1 = \sum_{k\in\mathbb{Z}} c_k \underbrace{\left(1 - \frac{4\Delta t}{(\Delta x)^2}\left(\sin\left(\frac{k}{2}\Delta x\right)\right)^2\right)}_{=:\gamma_k} e^{ijk\Delta x}$$

puis, par récurrence sur $n \geq 0$,

$$u_j^n = \sum_{k\in\mathbb{Z}} c_k \gamma_k^n e^{ijk\Delta x}, \quad n \geq 0,$$

avec :

$$1 \geq \gamma_k \underset{(3.8)}{>} 1 - 2\left(\sin\left(\frac{k}{2}\Delta x\right)\right)^2 \geq -1$$

Soit $j \in [[0, I]]$. On pose $L = \pi$ dans (3.5). D'après (3.5), on a alors :

$$u(x_j, T) - u_j^{N+1} = \sum_{k \in \mathbb{Z}} c_k (e^{-k^2 T} - \gamma_k^{N+1}) e^{ikx_j}$$

avec :

$$|c_k (e^{-k^2 T} - \gamma_k^{N+1}) e^{ikx_j}| = |c_k| |e^{-k^2 T} - \gamma_k^{N+1}| \leq 2|c_k|$$

et la série majorante est convergente par hypothèse. Soit $k \in \mathbb{Z}$. On a :

$$\gamma_k - 1 = -\frac{4\Delta t}{(\Delta x)^2} \left(\sin\left(\frac{k}{2}\Delta x\right)\right)^2 \underset{(\Delta t, \Delta x) \to (0,0)}{\sim} -\frac{4\Delta t}{(\Delta x)^2} \left(\frac{k}{2}\Delta x\right)^2 = -k^2 \Delta t$$

$$\begin{aligned} \gamma_k - 1 = -\frac{4\Delta t}{(\Delta x)^2} \left(\sin\left(\frac{k}{2}\Delta x\right)\right)^2 &= -\frac{4\Delta t}{(\Delta x)^2} \left(\frac{k}{2}\Delta x + o(k\Delta x)^2\right)^2 \\ &= -k^2 \Delta t (1 + o(k\Delta x)) \end{aligned}$$

donc :

$$\begin{aligned} \ln(\gamma_k^{N+1}) = (N+1) \ln \gamma_k &= (N+1) \ln(1 - k^2 \Delta t (1 + o(k\Delta x))) \\ &\underset{(\Delta t, \Delta x) \to (0,0)}{\sim} -(N+1) k^2 \Delta t = -k^2 T \end{aligned}$$

i.e. :

$$\lim_{(\Delta t, \Delta x) \to (0,0)} \ln(\gamma_k^{N+1}) = -k^2 T.$$

On en déduit, par continuité de l'exponentielle :

$$\lim_{(\Delta t, \Delta x) \to (0,0)} \gamma_k^{N+1} = e^{-k^2 T}$$

i.e :

$$\lim_{(\Delta t, \Delta x) \to (0,0)} (\gamma_k^{N+1} - e^{-k^2 T}) = 0.$$

Du Théorème de convergence dominée il résulte que

$$\lim_{(\Delta t, \Delta x) \to (0,0)} \sum_{k \in \mathbb{Z}} |c_k| |\gamma_k^{N+1} - e^{-k^2 T}| = 0.$$

On remarque que :

$$|e_j^{(N+1)}| \leq \sum_{k \in \mathbb{Z}} |c_k| |\gamma_k^{N+1} - e^{-k^2 T}|, \quad j = 0, \cdots I + 1.$$

et donc :

$$\|e^{(N+1)}\|_\infty \leq \sum_{k \in \mathbb{Z}} |c_k| |\gamma_k^{N+1} - e^{-k^2 T}| \underset{(\Delta t, \Delta x) \to (0,0)}{\to} 0.$$

$\square$

Schéma implicite et schéma de Crank-Nickolson

Soit $\theta \in [0,1]$. On considère le schéma :

$$\left\{\begin{array}{l} \dfrac{u_i^{n+1}-u_i^n}{\Delta x} + \theta\dfrac{(-u_{i-1}^{n+1}+2u_i^{n+1}-u_{i+1}^{n+1})}{(\Delta x)^2} + (1-\theta)\dfrac{(-u_{i-1}^{n}+2u_i^{n}-u_{i+1}^{n})}{(\Delta x)^2} = 0, \\ \\ u_i^0 = u_0(x_i), \quad u_0^n = u_{N+1}^n = 0, \quad i = 1,\cdots,N, \quad n \geq 0. \end{array}\right. \tag{3.10}$$

Proposition 3.3.5. *Le schéma (3.10) est consistant d'ordre 2 en espace. Il est consistant d'ordre 2 en temps si $\theta = \frac{1}{2}$, d'ordre 1 en temps sinon.*

Démonstration. Soit $(x,t) \in [0,1] \times \mathbb{R}^+$. On a :

$$\varepsilon_1(x,t) := \frac{u(x,t+\Delta t)-u(x,t)}{\Delta t} + \frac{-u(x-\Delta x,t+\Delta t)+2u(x,t+\Delta t)-u(x+\Delta x,t+\Delta t)}{\Delta x^2} =$$

$$= \frac{\partial u}{\partial t}(x,t) + \frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2} - \underbrace{\frac{\partial^2 u}{\partial x^2}(x,t+\Delta t)}_{=\frac{\partial u}{\partial t}(x,t+\Delta t)} + O(\Delta t^2) + O(\Delta x^2)$$

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

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

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

$$\varepsilon_2(x,t) := \frac{u(x,t+\Delta t)-u(x,t)}{\Delta t} + \frac{-u(x-\Delta x,t)+2u(x,t)-u(x+\Delta x,t)}{\Delta x^2} =$$

$$\underset{(3.7)}{=} \frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2}(x,t) + O(\Delta t^2) + O(\Delta x^2)$$

On en déduit :

$$\theta\varepsilon_1(x,t) + (1-\theta)\varepsilon_2(x,t) = \left(\frac{1}{2}-\theta\right)\frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2}(x,t) + O(\Delta t^2) + O(\Delta x^2)$$

$$= \left\{\begin{array}{ll} O(\Delta t^2) + O(\Delta x^2) & \text{si} \quad \theta = \frac{1}{2}, \\ \\ O(\Delta t) + O(\Delta x^2) & \text{sinon} \end{array}\right.$$

□

Proposition 3.3.6 (Convergence du θ-Schéma)**.** *Sous les hypothèses de la Proposition 3.2.6, le schéma (3.10) est convergent.*

Démonstration. Soit $\theta \in [0,1]$ et soit $(\mu_i^n)_{0\leq i\leq N, n\geq 0}$ une suite de réels. On considère le schéma :

$$\begin{cases} \dfrac{z_i^{n+1}-z_i^n}{\Delta x} + \theta\dfrac{(-z_{i-1}^{n+1}+2z_i^{n+1}-z_{i+1}^{n+1})}{(\Delta x)^2} + (1-\theta)\dfrac{(-z_{i-1}^{n}+2z_i^{n}-z_{i+1}^{n})}{(\Delta x)^2} = \mu_i^n, \\ \\ \mu_i^0 \in \mathbb{R}, \quad \mu_0^n = \mu_{N+1}^n = 0, \quad i = 1,\cdots,N, \quad n \geq 0. \end{cases}$$

On note $(Z^n)_{n\geq 0}$ la suite des vecteurs de composantes z_i^n, $i = 1,\cdots,N$. Alors :

$$\left(1+\theta\frac{\Delta t}{(\Delta x)^2}A_N\right) Z^{n+1} = \left(1-(1-\theta)\frac{\Delta t}{(\Delta x)^2}A_N\right) Z^{n+1} + \Delta t \mu^n$$

d'où on déduit que :

$$\begin{aligned} \|Z^{n+1}\|_2 &\leq \left\| \left(1+\theta\frac{\Delta t}{(\Delta x)^2}A_N\right)^{-1} \right\|_2 \left(\left\| 1-(1-\theta)\frac{\Delta t}{(\Delta x)^2}A_N \right\|_2 \|Z^n\|_2 + \Delta t\|\mu^n\|_2 \right) \\ &\leq \frac{\left(1-(1-\theta)\frac{\Delta t}{(\Delta x)^2}\lambda_1(A_N)\right)}{\left(1+\theta\frac{\Delta t}{(\Delta x)^2}\lambda_N(A_N)\right)}\|Z^n\|_2 + \frac{\Delta t\|\mu^n\|_2}{\left(1+\theta\frac{\Delta t}{(\Delta x)^2}\lambda_N(A_N)\right)} \\ &\leq \underbrace{\left(1-\frac{\frac{\Delta t}{(\Delta x)^2}}{1+\theta\frac{\Delta t}{(\Delta x)^2}\lambda_N(A_N)}\right)}_{=:\tau_N}\|Z^n\|_2 + \Delta t\|\mu^n\|_2. \\ &\leq \tau_N^n\|Z^0\|_2 + \Delta t\sum_{k=0}^{n-1}\tau_N^k\|\mu^{n-k}\|_2. \end{aligned}$$

Si $\mu_i^n = \varepsilon_i^n$, alors $z_i^n = e_i^n$, $\forall i \in [[1,N]]$. Par construction : $Z^0 = 0$, donc

$$\begin{aligned} \|e^n\|_2 \leq C\Delta t(\Delta t + (\Delta x)^2)\frac{\tau_N^n - 1}{\tau_N - 1} &\underset{0<\tau_N<1}{=} C\Delta t(\Delta t + (\Delta x)^2)\frac{1-\tau_N^n}{1-\tau_N} \\ &\leq C(\Delta t + (\Delta x)^2)\frac{\Delta t}{1-\tau_N} \leq \\ &\leq C\left(1+\theta\frac{\Delta t}{(\Delta x)^2}\lambda_N(A_N)\right)(\Delta x)^2(\Delta t + (\Delta x)^2) \end{aligned}$$

□

Proposition 3.3.7 (Principe du maximum). *On suppose (3.8) réalisé. Alors, la solution du schéma (3.10) vérifie :*

$$\min_{0\leq i\leq N+1} u_i^0 \leq \min_{0\leq i\leq N+1} u_i^n \leq \max_{0\leq i\leq N+1} u_i^n \leq \max_{0\leq i\leq N+1} u_i^0$$

Démonstration. Soit $\theta \in [0,1]$ et soit $u_{i_0}^{n+1} = \min u_i^{n+1}$. On a :

$$\left(1+2\theta\frac{\Delta t}{(\Delta x)^2}\right)u_{i_0}^{n+1} - \theta\frac{\Delta t}{(\Delta x)^2}(u_{i_0-1}^{n+1}+u_{i_0+1}^{n+1}) =$$

$$= \left(1-2(1-\theta)\frac{\Delta t}{(\Delta x)^2}\right)u_{i_0}^n + (1-\theta)\frac{\Delta t}{(\Delta x)^2}(u_{i_0-1}^n+u_{i_0+1}^n)$$

On en déduit, par définition de i_0 :

$$\left(1+2\theta\frac{\Delta t}{(\Delta x)^2}\right)u_{i_0}^{n+1} \geq 2\theta\frac{\Delta t}{(\Delta x)^2}u_{i_0}^{n+1}+$$

$$+\left(1-2(1-\theta)\frac{\Delta t}{(\Delta x)^2}\right)u_{i_0}^n + (1-\theta)\frac{\Delta t}{(\Delta x)^2}(u_{i_0-1}^n+u_{i_0+1}^n)$$

i.e. :

$$u_{i_0}^{n+1} \geq \left(1-2(1-\theta)\frac{\Delta t}{(\Delta x)^2}\right)u_{i_0}^n + (1-\theta)\frac{\Delta t}{(\Delta x)^2}(u_{i_0-1}^n+u_{i_0+1}^n)$$

$$\underset{(3.8)}{\geq} \left(1-2(1-\theta)\frac{\Delta t}{(\Delta x)^2}\right)\min_i u_i^n + 2(1-\theta)\frac{\Delta t}{(\Delta x)^2}\min_i u_i^n = \min_i u_i^n$$

i.e. : $\min\limits_i u_i^{n+1} \geq \min\limits_i u_i^n \geq \min\limits_i u_i^0$.
De même, si $u_{i_0}^{n+1} = \max u_i^{n+1}$. On a, par définition de i_0 :

$$\left(1+2\theta\frac{\Delta t}{(\Delta x)^2}\right)u_{i_0}^{n+1} \leq 2\theta\frac{\Delta t}{(\Delta x)^2}u_{i_0}^{n+1}+$$

$$+\left(1-2(1-\theta)\frac{\Delta t}{(\Delta x)^2}\right)u_{i_0}^n + (1-\theta)\frac{\Delta t}{(\Delta x)^2}(u_{i_0-1}^n+u_{i_0+1}^n)$$

i.e. :

$$u_{i_0}^{n+1} \leq \left(1-2(1-\theta)\frac{\Delta t}{(\Delta x)^2}\right)u_{i_0}^n + (1-\theta)\frac{\Delta t}{(\Delta x)^2}(u_{i_0-1}^n+u_{i_0+1}^n)$$

$$\underset{(3.8)}{\leq} \left(1-2(1-\theta)\frac{\Delta t}{(\Delta x)^2}\right)\max_i u_i^n + 2(1-\theta)\frac{\Delta t}{(\Delta x)^2}\max_i u_i^n = \max_i u_i^n$$

i.e. : $\max\limits_i u_i^{n+1} \geq \max\limits_i u_i^n \geq \max\limits_i u_i^0$.

□

Proposition 3.3.8 (Convergence au sens de Von Neumann)**.** *On suppose que* $\sum_{k\in\mathbb{Z}}|c_k| < +\infty$*. Alors, le schéma (3.10) converge au sens de Von Neumann, i.e. :*

$$\lim_{(\Delta t,\Delta x)\to(0,0)} \|e^{N+1}\|_\infty = 0.$$

Démonstration. Par hypothèse :

$$u_j^0 = \sum_{j\in\mathbb{Z}} c_k e^{ikj\Delta x}, \quad i = 1,\cdots,N.$$

Au temps $n = 0$, le schéma (3.10) se réécrit sous forme matricielle :

$$\left(I + \theta\frac{\Delta t}{(\Delta x)^2}A_N\right)u^1 = \left(I - (1-\theta)\frac{\Delta t}{(\Delta x)^2}A_N\right)u^0.$$

On cherche u^1 sous la forme :

$$u_j^1 = \sum_{k\in\mathbb{Z}} c_k^1 e^{ikj\Delta x}, \quad j = 1,\cdots,N.$$

Le calcul donne directement

$$c_k^1 = \underbrace{\left(1 - \frac{4\frac{\Delta t}{(\Delta x)^2}\left(\sin\left(\frac{k}{2}\Delta x\right)\right)^2}{1 + 4\theta\frac{\Delta t}{(\Delta x)^2}\left(\sin\left(\frac{k}{2}\Delta x\right)\right)^2}\right)}_{=:\tau_k} c_k, \quad \forall k\in\mathbb{Z}. \tag{3.11}$$

Par récurrence sur $n \geq 0$, on trouve que :

$$u_j^n = \sum_{k\in\mathbb{Z}} \tau_k^n c_k e^{ikj\Delta x}, \quad j = 1,\cdots,N.$$

avec

$$\lim_{\Delta x\to 0} \frac{4}{(\Delta x)^2}\left(\sin\left(\frac{k}{2}\Delta x\right)\right)^2 = k^2, \quad \forall k\in\mathbb{Z}.$$

donc

$$\lim_{\Delta x\to 0} \tau_k = 1.$$

Il en résulte :

$$\ln(\tau_k) \underset{(\Delta t,\Delta x)\to(0,0)}{\sim} -\frac{4\frac{\Delta t}{(\Delta x)^2}\left(\sin\left(\frac{k}{2}\Delta x\right)\right)^2}{1 + 4\theta\frac{\Delta t}{(\Delta x)^2}\left(\sin\left(\frac{k}{2}\Delta x\right)\right)^2} \underset{(\Delta t,\Delta x)\to(0,0)}{\sim} -k^2\Delta t$$

$$\Rightarrow n\ln(\tau_k) \underset{(\Delta t,\Delta x)\to(0,0)}{\sim} -k^2 t_n,$$

i.e. :

$$\lim_{(\Delta t,\Delta x)\to(0,0)} \tau_k^n = e^{-k^2 t_n}, \quad \forall k\in\mathbb{Z}, \quad \forall n\geq 0.$$

On conclut comme pour la Proposition 3.3.4. □

Définition 3.3.4. Le schéma (3.10) est dit stable au sens de Von Neumann si dans (3.11) :

$$|\tau_k| < 1, \quad \forall k \in \mathbb{Z}.$$

Proposition 3.3.9 (Stabilité au sens de Von Neumann)**.** *1. Si $\theta \geq \frac{1}{2}$, le schéma (3.10) est inconditionnellement stable.*

2. Si $\theta > \frac{1}{2}$, le schéma (3.10) est stable si en outre :

$$\frac{\Delta t}{(\Delta x)^2} \leq \frac{1}{2(1-\theta)}$$

Démonstration. Soit $k \in \mathbb{Z}$ et soit $\theta \in [0,1]$. On pose :

$$s_k = \frac{4\Delta t}{(\Delta x)^2}\left(\sin\left(\frac{k}{2}\Delta x\right)\right)^2.$$

On a :

$$|\tau_k| < 1 \iff 0 < \underbrace{\frac{s_k}{1+\theta s_k}}_{=:f_\theta(s_k)} < 2 \tag{3.12}$$

1. Si $\theta \geq \frac{1}{2}$, l'étude des variations de f_θ montre que $f_\theta(\mathbb{R}^{+*}) \subset]0,2[$, i.e. que (3.12) est réalisé pour tout $k \in \mathbb{Z}$.
2. Si $\theta < \frac{1}{2}$, l'étude des variations de f_θ montre que $]0,2[= f_\theta\left(]0, \frac{2}{1-2\theta}[\right)$. On conclut en remarquant que :

$$0 \leq s_k \leq \frac{4\Delta t}{(\Delta x)^2}$$

avec :

$$\frac{4\Delta t}{(\Delta x)^2} < \frac{2}{1-2\theta} \iff \frac{\Delta t}{(\Delta x)^2} < \frac{1}{2(1-2\theta)}$$

□

Corollaire 3.3.10. *1. Les schémas d'Euler implicite ($\theta = 1$ dans (3.10)) et de Crank-Nicolson ($\theta = \frac{1}{2}$ dans (3.10)) sont inconditionnellement stables.*

2. Le schéma d'Euler explicite ($\theta = 0$ dans (3.10)) n'est stable que si

$$\frac{\Delta t}{(\Delta x)^2} \leq \frac{1}{2}.$$

Convection-diffusion

Dans l'approximation de l'équation de transport (1.10) par le schéma convegent (1.16), l'ereur de consistance est donnée par

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

de sorte qu'un schéma convergeant d'ordre 1 en temps et d'ordre 2 en espace est donné par :

$$\begin{cases} \dfrac{u_i^{n+1} - u_i^n}{\Delta t} + \dfrac{u_i^n - u_{i-1}^n}{\Delta x} + \dfrac{1}{2\Delta x}(-u_{i-1}^n + 2u_i^n - u_{i+1}^n) = 0, \\ \\ u_i^0 = u_0(x_i), \quad i \in \mathbb{Z}, \quad n \geq 0. \end{cases} \tag{3.13}$$

Proposition 3.3.11. *Sous la condition :*

$$0 < \frac{\Delta t}{\Delta x} < \frac{1}{2}$$

le schéma (3.13) est convergent d'ordre 1 *en temps et d'ordre* 2 *en espace.*

Remarque 13. Le schéma (3.13) approche aussi l'équation de convection-diffusion :

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} - \varepsilon\frac{\partial^2 u}{\partial x^2} = 0, \quad x \in \mathbb{R}, \quad t \geq 0$$

lorsque $\varepsilon > 0$ est un petit paramètre.

Bibliographie

[1] Haïm Brezis. Analyse Fonctionnelle. Théorie et applications.Masson, Paris, 1983. *Chapitres X.1 et X.2.*

[2] Thierry Gallouët, Raphaèle Herbin. Analyse numérique des équations aux dérivées partielles. Master. Marseille, France. 2011. (https ://cel.hal.science/cel-00637008v2) *Chapitre 2.*

[3] P.A. Raviart et J.M. Thomas Introduction à l'analyse numérique des équations aux dérivés partielles, Masson, Paris 1996. *Chapitre 7.*

[4] Lionel Sainsaulieu. Calcul Scientifique. Cours et exercices corrigés pour le second cycle et les écoles d'ingénieurs, Masson, Paris 1996. *Chapitre 2.4.*

[5] Brigitte Lucquin, Olivier Pironneau. Introduction au calcul scientifique. Masson, Paris, 1996. *Chapitres VII.3 et VII.5*

[6] Alfio Quarteroni, Riccardo Sacco, Fausto Saleri Numerical Mathematics. Springer, Berlin, 2007. *Chapitres 13.1, 13.2, 13.8*

[7] Alfio Quarteroni, Riccardo Sacco, Paola Gervasio Calcul scientifique. Springer, Milan, 2010. *Chapitre 8.2.6*