Option Calcul Scientifique: Equations aux Dérivées partielles

Préparation Agrégation de Mathématiques

Université de Rennes 1
Isabelle Gruais

16 février 2023

Chapitre 1

Equation de transport

1.1 Modélisation

On considère un matériau en mouvement à la vitesse dans l'espace caratérisé par sa densité $\rho(x,t)$ en un point x à la date $t > 0$. On suppose que le mouvement de ce matériau est entièrement décrit par le champ de vitesses $\vec{v}$. On note V un élément de volume arbitraire fixe dans l'espace. La quantité de matière présente dans ce volume à la date t est alors

$$\rho_V(t) := \int_V \rho(x,t) d\Omega(x).$$

En l'absence de forces extérieures, la variation de ρ_V coïncide avec la quantité de matière qui traverse le volume V et est modélisée par le flux au travers de la frontière de V, soit :

$$\underbrace{d\rho_V}_{=\rho'_V dt} = -\int_{\partial V} \rho(x,t) \underbrace{\overrightarrow{dM}}_{=\vec{v}dt} \cdot d\vec{S}$$

i.e. :

$$\rho'_V(t)dt = -\int_{\partial V} \rho(x,t)\vec{v}(x,t) \cdot d\vec{S}dt \tag{1.1}$$

où le vecteur surfacique $d\vec{S} = \|d\vec{S}\|\vec{n}$ est orienté dans la direction de la normale extérieure à V. Avec cette convention, $\vec{v} \cdot d\vec{S} > 0$ dans le cas d'un flux sortant et alors ρ_V décroît ($\rho'_V < 0$), $\vec{v} \cdot d\vec{S} < 0$ dans le cas d'un flux entrant et alors ρ_V croît ($\rho'_V > 0$). On suppose que ρ est assez régulière, de classe $\mathcal{C}^1$, pour écrire :

$$\rho'_V(t) = \int_V \frac{\partial \rho}{\partial t}(x,t) d\Omega(x).$$

et alors (1.1) devient :

$$\int_V \frac{\partial \rho}{\partial t}(x,t)d\Omega = -\int_{\partial V} \underbrace{\rho(x,t)\vec{v}(x,t)\cdot d\vec{S}}_{=:\omega(x)},$$

i.e., en appliquant successivement la formule de Stokes et la définition de la divergence :

$$\int_V \frac{\partial \rho}{\partial t}(x,t)d\Omega = -\int_{\partial V} \omega \underset{Stokes}{=} -\int_V d\omega = -\int_V \mathrm{div}(\rho\vec{v})d\Omega$$

ce qui équivaut à :

$$\int_V \left(\frac{\partial \rho}{\partial t} + \mathrm{div}(\rho\vec{v})\right) d\Omega = 0.$$

Le volume V étant arbitraire dans l'espace, on en déduit :

$$\frac{\partial \rho}{\partial t} + \mathrm{div}(\rho\vec{v}) = 0, \quad \text{dans} \quad \Omega \times \mathbb{R}^+.$$

Exemple

On suppose connu l'emplacement d'une nappe de pétrole due au dégazement intempestif d'un supertanker au large des côtes et on cherche à anticiper son déplacement le long des côtes dans les heures à venir, par exemple pour la mise en oeuvre efficace de barrages. On suppose connu $v : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}^2$, $(x,t) \mapsto v(x,t)$, le champ des vecteurs vitesses des courants marins, donné par exemple par la table des marées. A $t = 0$, on connaît la densité initiale d'hydrocarbure $\rho_0(x)$ et on cherche à calculer la densité d'hydrocarbure $\rho(x,t)$ à l'instant t au point $x \in \mathbb{R}^2$. L'équation de conservation de la masse s'écrit :

$$\frac{\partial \rho}{\partial t} + \mathrm{div}(\rho v) = 0, \quad \rho(x,0) = \rho_0(x)$$

avec

$$\rho_0(x) = \begin{cases} 1 & \text{si} \quad x \in A, \\ 0 & \text{si} \quad x \in A^c, \end{cases} \tag{1.2}$$

où A représente le lieu initial de la nappe. Dans le cas d'un déplacement maritime, le vecteur $v : \mathbb{R}^2 \times \mathbb{R}^+$ n'est évidemment pas constant (la marée n'est pas la même d'une ville à l'autre). De plus, le déplacement de la nappe dépend également du vent qui affecte donc le vecteur v. On supposera pourtant ici, pour slmplifier l'exposé, que le vecteur $v$vest constant en espace et en temps. Alors, le problème (1.2) admet pour solution :

$$\rho(x,t) = \rho_0(x - tv) \tag{1.3}$$

qui exprime que la nappe perçue au point x à la date t a été transportée dans la direction du vecteur v à la distance tv du point initial. En fait, il est clair que (1.3) n'est pas une solution classique de (1.2) :comme ρ_0 n'est pas continue, lq fonction ρ définie par (1.3) ne l'est pas non plus et ses dérivées partielles ne sont donc pas définies au sens classique.

Dans la suite on verra comment on peut donner une formulation correcte des solutions de (1.2). Plus généralement, les équations de transport sont très importantes en mécanique des fluides : par exemple les équations d'Euler sont utilisées pour modéliser l'écoulement de l'air autour d'une aile d'avion.

Dans ce cours on se limitera aux équations scalaires en une dimension d'espace d'abord dans le cas relativement simple d'une équation linéaire (paragraphes 1.2 et 1.3) pusi dans le cas nettement plus difficile d'une équation non linéaire (paragraphes 1.4 et 1.5)

1.2 Solutions classiques et solutions faibles. Le cas linéaire

Soit à résoudre : chercher $u : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ solution du problème dit de Cauchy

$$\begin{aligned} &\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t > 0 \\ &u(x,0) = u_0(x), \quad x \in \mathbb{R} \end{aligned} \tag{1.4}$$

où la vitesse $c \in \mathbb{R}$ et la condition initiale $u_0 : \mathbb{R} \to \mathbb{R}$ sont données.

Définition 1.2.1 (Solution classique)**.** On dit que $u : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ est une solution classique de (1.4) si $u \in \mathcal{C}^1(\mathbb{R} \times \mathbb{R}^+, \mathbb{R})$ et si u vérifie (1.4).

Une CN pour que u soit une solution classique de (1.4) est que $u_0 \in \mathcal{C}^1(\mathbb{R})$.

Proposition 1.2.1. *Si $u_0 \in \mathcal{C}^1(R)$, alors il existe une unique solution classique de (1.4) et elle s'écrit :*

$$u(x,t) = u_0(x - ct), \quad \forall (x,t) \in \mathbb{R} \times \mathbb{R}^+. \tag{1.5}$$

Démonstration. Pour montrer l'existence, il suffit de remarquer que u définie par (1.5) est de classe $\mathcal{C}^1$ et vérifie en tout point :

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = -cu_0'(x - ct) + cu_0'(x - ct) = 0.$$

Pour montrer l'unicité, on introduit la notion de caractéristique. Plus précisément, on cherche a priori des solutions sous la forme de courbes $s \mapsto u(x(s), t(s))$ en partant de la relation :

$$\frac{d}{ds}u(x(s),t(s)) = \dot{x}\frac{\partial u}{\partial x} + \dot{t}\frac{\partial u}{\partial t}.$$

Cette équation est identique à (1.4) ssi :

$$\begin{pmatrix} \dot{x} \\ \dot{t} \end{pmatrix} \quad // \quad d\begin{pmatrix} c \\ 1 \end{pmatrix}$$

i.e. ssi

$$\frac{\dot{x}}{c} = \frac{\dot{t}}{1} \iff \dot{x} - c\dot{t} = 0 \iff \frac{d}{ds}(x - ct) = 0 \iff x - ct = Cste.$$

Les droites $x - ct$ =Cste sont appelés les caractéristiques du problème (cf. figure 1.1). Soit $x_0 \in \mathbb{R}$ et soit $x - ct = x_0$ la caratéristqiue issue du point $(x_0, 0)$ du plan. Par construction, si u est une solution classique, alors, le long de cette droite :

$$u(x,t) = u(x_0,0) = u_0(x_0) = u_0(x - ct).$$

i.e. (1.5). Il reste à montrer que toute solution classique de (1.4) est constante le long des droites $\mathcal{D}_{x_0} : x - ct = x_0$, $x_0 \in \mathbb{R}$. Soit $x_0 \in \mathbb{R}$ et soit $\varphi_{x_0} : \mathbb{R}^+ \to \mathbb{R}$ définie par $\varphi_{x_0}(t) = u(x_0 + ct, t)$, $\forall t \geq 0$. On a :

$$\varphi'_{x_0}(t) = c\frac{\partial u}{\partial x}(x_0 + ct, t) + \frac{\partial u}{\partial t}(x_0 + ct, t) \underset{(1.4)}{=} 0.$$

On en déduit : $\varphi_{x_0}(t) = \varphi_{x_0}(0) = u(x_0, 0) = u_0(x_0)$, $\forall t \in \mathbb{R}^+$, i.e. u est constanbte le long de $\mathcal{D}_{x_0}$, $\forall x_0 \in \mathbb{R}$. □

Remarque 1 (Terme source). Le problème physique peut conduire à une équation avec terme source $f \in \mathcal{C}(\mathbb{R} \times \mathbb{R}^+)$ au second membre :

$$\begin{aligned} \frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} &= f(x,t) \quad x \in \mathbb{R}, \quad t \in \mathbb{R}^+ \\ u(x,0) &= u_0(x), \quad x \in \mathbb{R}. \end{aligned} \tag{1.6}$$

où $u_0 \in \mathcal{C}^1(\mathbb{R})$. La méthode des caractéristiques conduit à chercher les courbes $s \mapsto (x(s), t(s))$ solutions de

$$\frac{d}{ds}u(x(s),t(s)) = \dot{t}\frac{\partial u}{\partial t} + \dot{x}\frac{\partial u}{\partial x} = f(x(s),t(s))$$

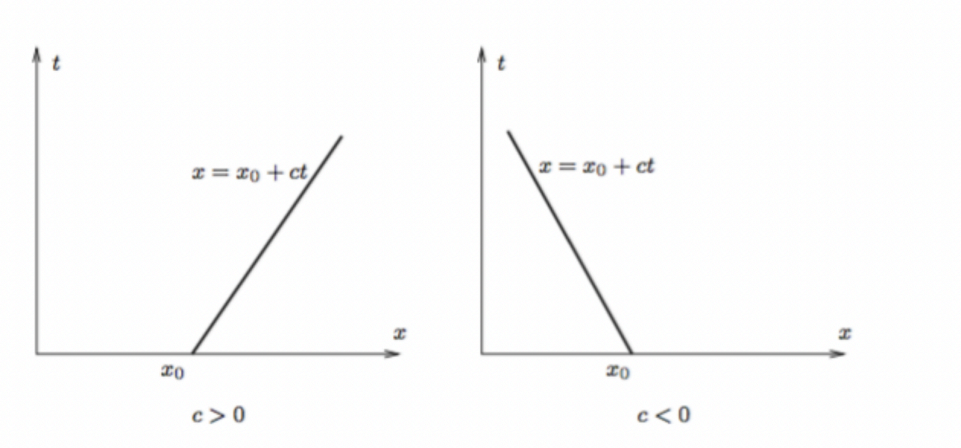

FIGURE 1.1 – Droites caractéristiques. Le cas linéaire.

avec

$$\dot{x} - c\dot{t} = 0 \iff x - ct = Cste.$$

Le choix

$$\dot{t} = 1, \quad \dot{c} = c, \quad t(0) = 0, \quad x(0) = x_0$$

conduit à : $t = s$, $x(t) = ct + x_0$, i.e. aux droites caractéristiques $x - ct = x_0$, $x_0 \in \mathbb{R}$, le long desquelles :

$$u(x,t) = u(x_0 + ct, t) = u_0(x_0) + \int_0^t f(x - cs, s)ds =$$

$$= u_0(x - ct) + \int_0^t f(x - c(t-s), s)ds.$$

On en déduit l'unicité de la solution. Inversement, on vérifie directement que u ainsi définie est solution de (1.6) et que cette solution est classique car $u_0 \in \mathcal{C}^1(\mathbb{R})$.

Définition 1.2.2 (Solution faible)**.** On appelle solution faible de (1.4) toute solution u du problème : $u \in L^\infty(\mathbb{R} \times \mathbb{R}^+)$ et $\forall \varphi \in \mathcal{C}_c^1(\mathbb{R} \times \mathbb{R}, \mathbb{R})$,

$$\int_{\mathbb{R}} \int_{\mathbb{R}^+} u \left(\frac{\partial \varphi}{\partial t} + c \frac{\partial \varphi}{\partial x} \right) dxdt + \int_{\mathbb{R}} u_0(x)\varphi(x,0)dx = 0. \tag{1.7}$$

Proposition 1.2.2. *Si u est solution classique de (1.4), alors u est solution faible. Réciproquement, si $u \in \mathcal{C}^1(\mathbb{R}\times]0, +\infty[) \cap \mathcal{C}(\mathbb{R} \times \mathbb{R}^+)$ est une solution faible de (1.4), alors u est une solution classique.*

Théorème 1.2.3 (Existence et unicité de la solution faible)**.** *Si* $u_0 \in L^\infty_{\text{loc}}(\mathbb{R})$*, alors il existe une unique solution faible de (1.4).*

Démonstration. On va montrer que $u(x,t) = u_0(x-ct)$ est solution faible. Par hypothèse sur u_0, $u \in L^\infty(\mathbb{R}\times\mathbb{R}^+)$. Soit $\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}^+,\mathbb{R})$. On a :

$$\int_{\mathbb{R}}\int_{\mathbb{R}^+} u_0(x-ct)\left(\frac{\partial\varphi}{\partial t}(x,t)+c\frac{\partial\varphi}{\partial x}(x,t)\right)dxdt =$$

$$\underset{y=x-ct}{=}\int_{\mathbb{R}}\int_{\mathbb{R}^+} u_0(y)\left(\frac{\partial\varphi}{\partial t}(y+ct,t)+c\frac{\partial\varphi}{\partial x}(y+ct,t)\right)dydt =$$

$$\underset{Fubini}{=}\int_{\mathbb{R}}\int_{\mathbb{R}^+} u_0(y)\left(\frac{\partial\varphi}{\partial t}(y+ct,t)+c\frac{\partial\varphi}{\partial x}(y+ct,t)\right)dtdy =$$

$$= \int_{\mathbb{R}} u_0(y)\left(\int_0^{+\infty}\frac{d}{dt}\varphi(y+ct,t)dt\right)dy = -\int_{\mathbb{R}} u_0(y)\varphi(y,0)dy$$

i.e. $u(x,t) = u_0(x-ct)$ est solution faible de (1.4). Il reste à montrer qu'elle est unique. Soit v une solution faible de (1.4) et soit $w = u - v$. On a :

$$\int_{\mathbb{R}}\int_{\mathbb{R}^+} w(x,t)\left(\frac{\partial\varphi}{\partial t}(x,t)+c\frac{\partial\varphi}{\partial x}(x,t)\right)dxdt = 0, \quad \forall\varphi\in\mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}^+). \quad (1.8)$$

D'après le Lemme 1.2.4 ci-dessous, pour tout $f \in \mathcal{C}_c(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R})$, il existe $\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R})$ solution de

$$\frac{\partial\varphi}{\partial t}+c\frac{\partial\varphi}{\partial x} = f,$$

et on déduit de (1.8) que :

$$\int_{\mathbb{R}}\int_{\mathbb{R}^+} w(x,t)f(x,t)dxdt = 0, \quad \forall f\in\mathcal{C}_c(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R}).$$

i.e. $w = 0$ par densité de $\mathcal{C}_c(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R})$ dans $L^\infty(\mathbb{R}\times\mathbb{R}^+,\mathbb{R})$. □

Lemme 1.2.4 (Résultat d'existence)**.** *Pour tout* $f \in \mathcal{C}_c(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R})$ *il existe* $\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R})$ *t.q.*

$$\frac{\partial\varphi}{\partial t}+c\frac{\partial\varphi}{\partial x} = f.$$

Démonstration. Soit $f \in \mathcal{C}_c(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R})$ et soit $T > 0$ t.q. $f(x,t) = 0$, si $\max(|x|,t) \geq T$ En particulier : $f(x,t) = 0$, $\forall x \in \mathbb{R}$, $\forall t \geq T$. On considère le problème :

$$\frac{\partial\varphi}{\partial t}+c\frac{\partial\varphi}{\partial x} = f, \quad \varphi|_{t=T} = 0. \quad (1.9)$$

On vérifie directement que φ définie par

$$\varphi(x,t) = -\int_t^T f(x + c(s-t), s)ds, \quad \forall (x,t) \in \mathbb{R} \times \mathbb{R}^{+*}$$

est solution classique de (1.9). De plus, si $t > T$, alors

$$\forall s \in [T,t], \quad s \geq T \Rightarrow f(x - c(t-s), s) = 0, \quad \forall x \in \mathbb{R},$$

donc $\varphi(x,t) = 0$, $\forall x \in \mathbb{R}$. Si $0 < t \leq T$, alors : $\forall s \in [t,T]$,

$$|x + c(s-t)| \geq ||x| - c|s-t|| \geq |x| - c|s-t| = |x| - c(s-t) \geq$$

$$\geq |x| - c(T-t) \geq |x| - cT.$$

Donc si $|x| \geq (c+1)T$ alors $f(x + c(s-t), s) = 0$, $\forall s \in [t,T]$. On en déduit que φ a son support dans le compact $K = [-(c+1)T, (c+1)T] \times [0,T]$. □

Remarque 2. La solution faible de (1.4) a les propriétés suivantes :

1. si $u_0 \geq 0$ p.p., alors $u \geq 0$ p.p..
2. $\|u(\cdot,t)\|_{L^p(\mathbb{R})} = \|u_0\|_{L^p(\mathbb{R})}, \quad \forall p \in [1, +\infty], \quad$ p.p.t. $t > 0$

Dans la suite, on s'attachera à vérifier que ces propriétés sont conservées par les schémas numériques étudiés.

1.3 Schémas numériques dans le cas linéaire

On considère le problème (1.4) avec $c = 1$:

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0, \quad u(\cdot, 0) = u_0 \in L^\infty(\mathbb{R}). \tag{1.10}$$

Que sait que ce problème admet une solution unique et que cette solution s'écrit : $u(x,t) = u_0(x-t)$. On rappelle que cette solution est classique, de classe $\mathcal{C}^1$, si $u_0 \in \mathcal{C}^1(\mathbb{R})$, et faible de régularité L^∞ si $u_0 \in L^\infty(\mathbb{R})$. On va chercher à approcher cete solution par un schéma numérique. Un tel schéma est évidemment inutile dans le cas linéaire puisque la solution exacte est alors connue, mais on commencera par ce cas pour faciliter l'exposé.

Schéma explicite différences finies centrées

On se donne les subdivisions régulières de pas $\Delta x > 0$, $\Delta t > 0$, en espace et en temps resp. :

$$x_i = i\Delta x, \quad t_n = n\Delta t, \quad i \in \mathbb{Z}, \quad n \in \mathbb{N},$$

et on considère la suite $(u_i^n)_{n\geq 0, i\in\mathbb{Z}}$ solution du schéma numérique

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} = 0, \quad u_i^0 = u_0(x_i), \quad i \in \mathbb{Z}, \quad n \geq 0 \tag{1.11}$$

Pour tout $n \geq 0$, on note $u^n = (u^n)_{i\in\mathbb{Z}}$. On a la formule de récurrence :

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{2\Delta x}(u_{i+1}^n - u_{i-1}^n), \quad i \in \mathbb{Z}, \quad n \geq 0$$

Proposition 1.3.1. *Le schéma (1.11) ne respecte pas la positivité.*

Démonstration. On suppose que $u^0 \in \mathcal{C}^1(\mathbb{R})$ est telle que : $u^0 \geq 0$ et

$$u_i^0 = \begin{cases} 1 & \text{si} \quad i > 0, \\ 0 & \text{si} \quad i \leq 0. \end{cases} \tag{1.12}$$

Alors :

$$u_0^1 = -\frac{\Delta t}{2\Delta x} < 0.$$

□

Définition 1.3.1 (Stabilité)**.** 1. Le schéma numérique (1.11) est L^∞-stable, resp. L^2-stable, s'il existe une constante $C > 0$ indépendante des pas $\Delta x > 0$, $\Delta t > 0$ t.q. :

$$\|u^n\|_\infty := \sup_{i\in\mathbb{Z}} |u_i^n| \leq C, \quad \forall n \geq 0$$

resp.

$$\|u^n\|_2 := (\sum_{i\in\mathbb{Z}} |u_i^n|^2)^{\frac{1}{2}} \leq C, \quad \forall n \geq 0$$

2. Le schéma numérique (1.11) est stable au sens de Von Neumann si la suite $(u^n)_{n\geq 0}$ construite à partir de la donnée initiale $u_0(x) = e^{ipx}$, $p \in \mathbb{Z}$, conserve les propriétés d'amortissemnt de la solution exacte de (1.10).

Proposition 1.3.2. *Le schéma (1.11) est inconditionnellement instable.*

Démonstration. Si u^0 est définie par (1.12), alors

$$u_i^1 = \begin{cases} 0 & \text{si} \quad i < 0, \\ -\frac{\Delta t}{2\Delta x} & \text{si} \quad i = 0, \\ 1 - \frac{\Delta t}{2\Delta x} & \text{si} \quad i = 1, \\ 1 & \text{si} \quad i > 1, \end{cases}$$

On en déduit que $\|u^1\|_\infty = \max\left(\frac{\Delta t}{2\Delta x}, \left|1 - \frac{\Delta t}{2\Delta x}\right|, 1\right)$ dépend des pas de discrétisation Δt et Δx, contrairement à $\|u^0\|_\infty = 1$, i.e. que le schéma (1.11) n'est pas L^∞-stable. Si u^0 est définie par

$$u_0^0 = \begin{cases} 1 & \text{si} \quad i = 0, \\ 0 & \text{sinon} \end{cases}$$

alors $\|u^0\|_2 = 1$ et

$$u_i^1 = \begin{cases} -\frac{\Delta t}{2\Delta x} & \text{si} \quad i = -1, \\ 1 & \text{si} \quad i = 0, \\ \frac{\Delta t}{2\Delta x} & \text{si} \quad i = 1, \\ 0 & \text{si} \quad |i| > 1, \end{cases}$$

et donc $\|u^1\|_2 = \sqrt{1 + \frac{1}{2}\left(\frac{\Delta t}{\Delta x}\right)^2}$, d'où on déduit que le schéma (1.11) n'est pas L^2-stable.

La solution exacte de (1.10) avec la donnée initiale $u^0(x) = e^{ipx}$, où $p \in \mathbb{Z}$ est fixé quelconque, est définie par $u(x,t) = u_0(x-t) = e^{ip(x-t)} = e^{-ipt}u_0(x)$, ret donc $|u(x,t)| = |u_0(x)|$, $\forall x \in \mathbb{R}$. Par ailleurs, le calcul donne :

$$u_j^1 = e^{ijp\Delta x}\left(1 - i\frac{\Delta t}{\Delta x}\sin(p\Delta x)\right) = \underbrace{\left(1 - i\frac{\Delta t}{\Delta x}\sin(p\Delta x)\right)}_{=:\mathcal{J}(\Delta t, \Delta x)} u_j^0, \quad \forall j \in \mathbb{Z}.$$

où

$$|\mathcal{J}(\Delta t, \Delta x)| = \sqrt{1 + \left(\frac{\Delta t}{\Delta x}\sin(p\Delta x)\right)^2} > 1 \quad \text{si} \quad \sin(p\Delta x) \neq 0,$$

i.e. le schéma (1.11) n'est pas stable au sens de Von Neumann.

Plus généralement, on suppose que u_0 est 2π-périodique. Alors u_0 se décompose en série de Fourier :

$$u_0(x) = \sum_{k=-\infty}^{+\infty} a_k e^{ikx} \quad \text{avec} \quad a_k = \frac{1}{2\pi}\int_0^{2\pi} u_0(y)e^{-iky}dy, \quad k \in \mathbb{Z}. \quad (1.13)$$

On en déduit :

$$u_j^0 = \sum_{k=-\infty}^{+\infty} a_k e^{ikj\Delta x}, \quad \forall j \in \mathbb{Z}.$$

Après application de (1.11), on obtient :

$$u_j^1 = \sum_{k=-\infty}^{+\infty} a_k \underbrace{\left(1 - i\frac{\Delta t}{\Delta x}\sin(k\Delta x)\right)}_{=:\gamma_k} e^{ikj\Delta x}, \quad \forall j \in \mathbb{Z}$$

Par récurrence sur $n \geq 0$, on en déduit :

$$u_j^n = \sum_{k=-\infty}^{+\infty} a_k \gamma_k^n e^{ikj\Delta x}, \quad \forall j \in \mathbb{Z}.$$

où γ_k, $k \in \mathbb{Z}$, est appelé le coefficient d'amplification de la $k$ème harmonique. On a :

$$|\gamma_k| = \sqrt{1 + \left(\frac{\Delta t}{\Delta x}\sin(k\Delta x)\right)^2} > 1 \quad \text{si} \quad \sin(k\Delta x) \neq 0.$$

On remarque que si $0 < \Delta x < \pi$, alors il existe $k_0 \in \mathbb{N}$ t.q. $\sin(k\Delta x) \neq 0$. On en déduit : $\forall j \in \mathbb{Z}$

$$\|u_j^n\|_2^2 \geq |a_{k_0}\gamma_{k_0}^n|^2 = |a_{k_0}|^2 \left(1 + \left(\frac{\Delta t}{\Delta x}\sin(k_0\Delta x)\right)^2\right)^n \underset{n\to+\infty}{\to} +\infty$$

et donc

$$\lim_{n\to+\infty} \|u^n\|_2^2 = +\infty.$$

□

Définition 1.3.2 (Consistance)**.** On appelle erreur de consistance locale du schéma (1.11) au point (x_i, t_n) la quantité :

$$\varepsilon_i^n = \frac{u(x_i, t_{n+1}) - u(x_i, t_n)}{\Delta t} + \frac{u(x_{i+1}, t_n) - u(x_{i-1}, t_n)}{2\Delta x}, \quad i \in \mathbb{Z}, \quad n \geq 0.$$

L'erreur de consistance du schéma (1.11) est alors définie par :

$$\mathcal{E}(\Delta x, \Delta t) = \max_{i\in\mathbb{Z}, n\geq 0} |\varepsilon_i^n|.$$

Le schéma (1.11) est dit consistant si $\lim_{(\Delta x,\Delta t)\to(0,0)} \mathcal{E}(\Delta x, \Delta t) = 0$.

Définition 1.3.3 (Convergence)**.** On introduit l'erreur de discrétisation :

$$e_i^n = u(x_i, t_n) - u_i^n, \quad \forall i \in \mathbb{Z}, \quad \forall n \geq 0$$

Le schéma (1.11) est dit convergent si :

$$\lim_{n \to +\infty} \|e^n\|_\infty = 0$$

où on a posé :

$$\|e^n\|_\infty := \max_{i \in \mathbb{Z}} |e_i^n|.$$

Proposition 1.3.3. *Le schéma (1.11) n'est pas convergent.*

Démonstration. Soit $u_0 \in \mathcal{C}^1(\mathbb{R})$ t.q.

$$u_0(x) = \begin{cases} 1 & \text{si} \quad x = 0, \\ 0 & \text{si} \quad |x| \geq \frac{1}{2}. \end{cases}$$

Soit $\Delta x = \Delta t \in]\frac{1}{2}, 1[$. Alors :

$$u_i^0 = \begin{cases} 1 & \text{si} \quad i = 0, \\ 0 & \text{si} \quad i \neq 0. \end{cases}$$

De plus :

$$u_1^1 = u_1^0 - \frac{\Delta t}{2\Delta x}(u_2^0 - u_0^0) = \frac{\Delta t}{2\Delta x}$$

$$u_{-1}^1 = u_{-1}^0 - \frac{\Delta t}{2\Delta x}(u_0^0 - u_{-2}^0) = -\frac{\Delta t}{2\Delta x}$$

$$u_i^1 = u_i^0 - \frac{\Delta t}{2\Delta x}(u_{i+1}^0 - u_{i-1}^0) = 0 \quad \text{si} \quad i > 1$$

$$u_i^1 = u_i^0 - \frac{\Delta t}{2\Delta x}(u_{i+1}^0 - u_{i-1}^0) = 0 \quad \text{si} \quad i < -1$$

On suppose que

$$u_n^n = \left(\frac{\Delta t}{2\Delta x}\right)^n, \tag{1.14}$$

$$u_i^n = 0 \quad \text{si} \quad i > n \tag{1.15}$$

Alors

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

i.e. (1.14)–(1.15) est vrai, $\forall n \geq 0$. De plus :

$$u(x_n, t_n) = u_0(n(\Delta x - \Delta t)) = u_0(0) = 1, \quad \forall n \geq 1,$$

On en déduit :

$$\|e^n\|_\infty \geq |u_n^n - u(x_n, t_n)| = \left|1 - \frac{1}{2^n}\right| \underset{n\to+\infty}{\to} 1 \neq 0$$

i.e. le schéma (1.11) ne converge pas pour ce choix de u_0, Δx, Δt.

□

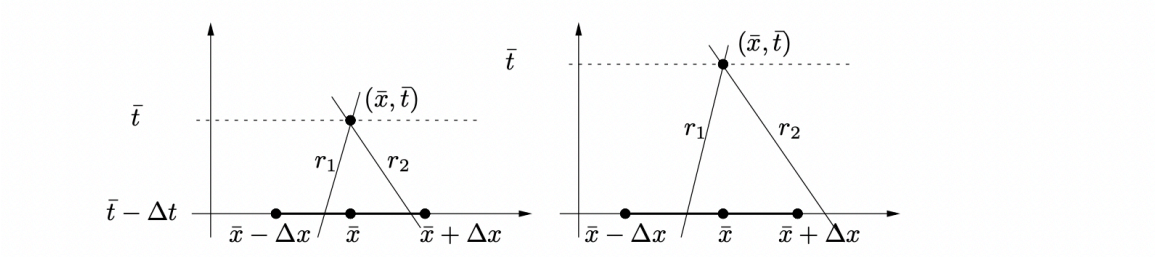

FIGURE 1.2 – Condition CFL : vérifiée à gauche, non vérifiée à droite

Schéma différences finies décentré amont

On considère le schéma explicite décentré amont :

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{u_i^n - u_{i-1}^n}{\Delta x} = 0, \quad u_i^0 = u_0(x_i), \quad i \in \mathbb{Z}, \quad n \geq 0 \tag{1.16}$$

Proposition 1.3.4. *Le schéma (1.16) est stable sous la condition dite de Courant-Friedrichs-Levy (CFL) :*

$$\frac{\Delta t}{\Delta x} \leq 1. \tag{1.17}$$

i.e. : $\forall n \geq 0$, $\forall A, B > 0$,

$$A \leq u^n \leq B \Rightarrow A \leq u^{n+1} \leq B.$$

Démonstration. On suppose que (1.17) est vérifié. On a :

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x}(u_i^n - u_{i-1}^n) = \left(1 - \frac{\Delta t}{\Delta x}\right) u_i^n + \frac{\Delta t}{\Delta x} u_{i-1}^n, \quad i \in \mathbb{Z}, \quad n \geq 0.$$

On suppose qu'il existe $A > 0$, $B > 0$ t.q. $A \leq u^n \leq B$. Alors : $\forall i \in \mathbb{Z}$,

$$u_i^{n+1} \underset{(1.17)}{\leq} \left(1 - \frac{\Delta t}{\Delta x}\right) B + \frac{\Delta t}{\Delta x} B = B.$$

et de même :

$$u_i^{n+1} \underset{(1.17)}{\geq} \left(1 - \frac{\Delta t}{\Delta x}\right) A + \frac{\Delta t}{\Delta x} A = A,$$

i.e. le schéma (1.16) est stable. □

Théorème 1.3.5 (Convergence du schéma décentré amont)**.** *On suppose que $u_0 \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$ et que u_0, u_0', u_0'' sont bornées.*

1. *On pose :*
$$A = \inf_{x \in \mathbb{R}} u_0(x), \quad B = \sup_{x \in \mathbb{R}} u_0(x).$$
Alors :
$$A \leq u_i^n \leq B, \quad \forall i \in \mathbb{Z}, \quad \forall n \geq 0.$$

2. *Soit $T > 0$. Alors, il existe une constante $C > 0$ ne dépendant que de u_0 t.q. :*
$$\sup_{i \in \mathbb{Z}, n\Delta t \leq T} |u(x_i, t_n) - u_i^n| \leq CT(\Delta t + \Delta x).$$

Démonstration. 1. C'est une conséquence immédiate de la Proposition 1.3.4.

2. Soit $(\mu_i^n)_{i \in \mathbb{Z}, n \geq 0}$ une suite de réels et soit $(z_i^n)_{i \in \mathbb{Z}, n \geq 0}$ la suite solution du schéma numérique :
$$\frac{z_i^{n+1} - z_i^n}{\Delta t} + \frac{z_i^n - z_{i-1}^n}{\Delta x} = \mu_i^n, \quad i \in \mathbb{Z}, \quad n \geq 0.$$
On a : $\forall i \in \mathbb{Z}$, $\forall n \geq 0$,
$$\|z^{n+1}\|_\infty \leq \|z^n\|_\infty + \Delta t \|\mu^n\|_\infty \leq \|z^0\|_\infty + \Delta t \sum_{k=0}^{n} \|\mu^k\|_\infty, \quad \forall i \in \mathbb{Z},$$
SI $\mu_i^n = 0$, $\forall i \in \mathbb{Z}$, alors $z_i^n = u_i^n$, $\forall n \geq 0$, $\forall i \in \mathbb{Z}$. On en déduit :
$$\|z^{n+1}\|_\infty \leq \|z^n\|_\infty, \quad \forall n \geq 0,$$
i.e. le schéma est stable.
Si $z_i^n = u(x_i, t_n) - u_i^n$, $\forall i \in \mathbb{Z}$, $\forall n \geq 0$, alors $z_i^0 = 0$ et μ_i^n coïncide avec l'erreur de consistance ε_i^n définie par :
$$\varepsilon_i^n = \frac{1}{\Delta t}(u(x_i, t_{n+1}) - u(x_i, t_n)) + \frac{1}{\Delta x}(u(x_i, t_n) - u(x_{i-1}, t_n))$$

Le calcul direct montre que :

$$\varepsilon_i^n = \frac{1}{2}(c^2\Delta t - \Delta x)u_0''(x_i - t_n) + o(\Delta t + \Delta x), \quad \forall i \in \mathbb{Z}, \quad \forall n \geq 0.$$

d'où on déduit que :

$$\|\varepsilon^n\|_\infty \leq C(\Delta t + \Delta x)(1 + \|u_0''\|_\infty), \quad \forall n \geq 0;$$

Il en résulte :

$$\|u(\cdot, t_n) - u^n\|_\infty \leq \Delta t \sum_{k=0}^{n} \|\varepsilon^n\|_\infty \leq Cn\Delta t(\Delta t + \Delta x)(1 + \|u_0''\|_\infty)$$

et donc

$$\sup_{n\Delta t \leq T} \|u(\cdot, t_n) - u^n\|_\infty \leq CT(1 + \|u_0''\|_\infty)(\Delta t + \Delta x).$$

□

Remarque 3 (Décentrement). Soit le schéma explicite décentré en aval :

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{u_{i+1}^n - u_i^n}{\Delta x} = 0, \quad u_i^0 = u_0(x_i), \quad i \in \mathbb{Z}, \quad n \geq 0.$$

On a :

$$u_i^{n+1} = \left(1 + \frac{\Delta t}{\Delta x}\right) u_i^n - \frac{\Delta t}{\Delta x} u_{i+1}^n, \quad \forall i \in \mathbb{Z}, \quad \forall n \geq 0.$$

On suppose que : $u_0(x) = 0$, $\forall x \geq 0$. Alors : $u_i^0 = 0$, $\forall i \in \mathbb{N}$. On vérifie directement que : $u_i^n = 0$, $\forall i \in \mathbb{N}$, $\forall n \geq 0$. On en déduit : $\forall \geq 0$,

$$u_{-1}^{n+1} = \left(1 + \frac{\Delta t}{\Delta x}\right) u_{-1}^n = \left(1 + \frac{\Delta t}{\Delta x}\right)^{n+1} u_{-1}^0$$

et donc $\lim_{n\to+\infty} |u_{-1}^n| = +\infty$.

Remarque 4. Si $u_0 \notin \mathcal{C}(\mathbb{R})$ la donnée initiale dans (1.16) n'est plus définie. On considère alors plutôt le schéma :

$$\begin{cases} \dfrac{u_i^{n+1} - u_i^n}{\Delta t} + \dfrac{u_i^n - u_{i-1}^n}{\Delta t} = 0, & i \in \mathbb{Z}, \quad n \geq 0, \\ u_i^0 = \dfrac{1}{\Delta x} \displaystyle\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u_0(y)dy, & i \in \mathbb{Z}. \end{cases} \tag{1.18}$$

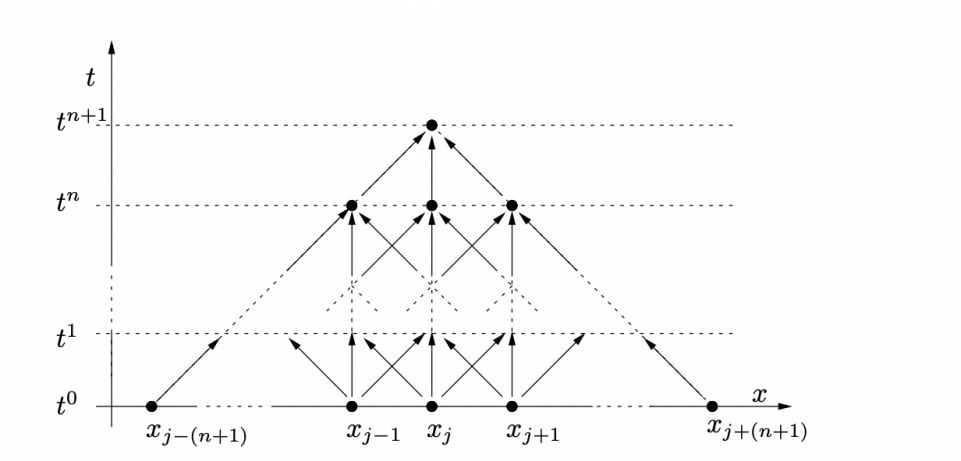

FIGURE 1.3 – Domaine de dépendance numérique au point (x_j, t^{n+1})

1.4 Le cas non linéaire

Soit $f \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$ et soit $u_0 \in \mathcal{C}^1(\mathbb{R})$. On considère le problème : trouver $u : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$, $((x,t) \mapsto u(x,t)$ solution de

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(f(u)) = 0, \quad u(x,0) = u_0(x), \quad (x,t) \in \mathbb{R} \times \mathbb{R}^+. \tag{1.19}$$

Définition 1.4.1 (Solution classique)**.** Soit $f \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$ et soit $u_0 \in \mathcal{C}^1(\mathbb{R})$. On appelle solution classique de (1.19) toute solution u du problème : trouver $u \in \mathcal{C}^1(\mathbb{R} \times \mathbb{R}^+)$ t.q.

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(f(u)) = 0, \quad u(x,0) = u_0(x), \quad (x,t) \in \mathbb{R} \times \mathbb{R}^+.$$

On commence par le résultat préliminaire sur les edos.

Proposition 1.4.1. *Soit $x_0 \in \mathbb{R}$ et soit $f \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$. Si $x : \mathbb{R}^+ \to \mathbb{R}$ est solution de l'équation différentielle :*

$$x'(t) = f(x(t)), \quad x(0) = x_0, \quad t \in \mathbb{R}^+,$$

et si $T_{\max} > 0$ désigne le temps d'existence de x, alors

$$T_{\max} < +\infty \Rightarrow \lim_{t \to T_{\max}} \|x(t)\| = +\infty.$$

Définition 1.4.2 (Courbe caractéristique)**.** On appelle courbe caractéristique du problème (1.19) issue de $x_0 \in \mathbb{R}$ la courbe définie par le problème de Cauchy :

$$x' = f'(u(x(t),t)), \quad x(0) = x_0, \quad t \in \mathbb{R}^+. \tag{1.20}$$

Théorème 1.4.2 (Non existence)**.** *Soit $f \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$. On suppose que f' n'est pas constante. Alors, il existe $u_0 \in \mathcal{C}_c^\infty(\mathbb{R})$ t.q. (1.19) n'admette pas de solution classique.*

Démonstration. La méthode des caractéristiques conduit à chercher les courbes $s \mapsto u(x(s),t(s))$ t.q.

$$\frac{d}{ds}u(x(s),t(s)) = \dot{t}\frac{\partial u}{\partial t} + \dot{x}\frac{\partial u}{\partial x} = 0.$$

Par comparaison avec (1.19), on en déduit qu'il faut avoir :

$$\det \begin{vmatrix} \dot{x} & f'(u) \\ \dot{t} & 1 \end{vmatrix} = 0$$

i.e. :

$$\dot{x} - \dot{t}f'(u) = 0. \tag{1.21}$$

Soit $s \mapsto (x(s),t(s))$ solution de $\dot{x} = \dot{t}f'(u)$, $\dot{t} \neq 0$. Le choix $\dot{t} = 1$ revient à choisir $s = t$ pour paramètre et alors $t \mapsto (x(t),t)$ est une caractéristique de (1.19) au sens de la Définition 1.4.2.

Comme $f \in \mathcal{C}^2(\mathbb{R},\mathbb{R})$, on a $f' \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$ et donc le théorème de Cauchy-Lispschitz s'applique à (1.20). Il existe donc une solution maximale x de (1.20) définie sur $[0,T_{\max}[$ et $\lim_{t\to T_{\max}} |x(t)| = +\infty$ si $T_{\max} < +\infty$. Par construction :

$$\frac{d}{dt}u(x(t),t) = \dot{x}\frac{\partial u}{\partial x} + \frac{\partial u}{\partial t} = f'(u)\frac{\partial u}{\partial x} + \frac{\partial u}{\partial t} = \frac{\partial}{\partial x}f(u(x,t)) + \frac{\partial u}{\partial t} = 0$$

i.e.

$$u(x(t),t) = Cste = u(x(0),0) = u_0(x_0), \quad \forall t \in [0,T_{\max}[.$$

On en déduit, par définition de (1.20) :

$$\dot{x}(t) = f'(u(x(t),t) = f'(u_0(x_0)), \quad \forall t \in [0,T_{\max}[.$$

et donc

$$x(t) = f'(u_0(x_0))t + x_0, \quad \forall t \in [0,T_{\max}[.$$

i.e. le système (1.20) décrit une droite issue de x_0, $\forall x_0 \in \mathbb{R}$, et on en déduit : $T_{\max} = +\infty$.

Comme f' est non constante, il existe $v_0, v_1 \in \mathbb{R}$ t.q. $f'(v_0) > f'(v_1)$. On peut construire $u_0 \in \mathcal{C}_c^\infty(\mathbb{R})$ t.q. $u_0(x_0) = v_0$, $u_0(x_1) = v_1$ et $x_0 < x_1$ (voir Figure 1.4.) On suppose que u est une solution classique avec cette donnée initiale. Alors :

$$u(x_0 + f'(u_0(x_0))t, t) = u_0(x_0) = v_0, \quad u(x_1 + f'(u_0(x_1))t, t) = u_0(x_1) = v_1.$$

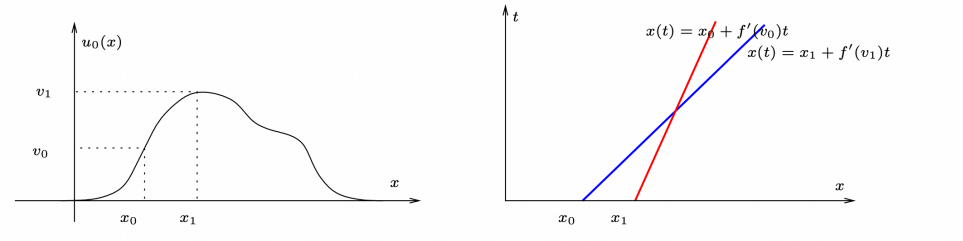

FIGURE 1.4 – Droites caractéristiques. Le cas non linéaire.

Soit $T > 0$ t.q. $x_0 + f'(u_0(x_0))T = x_1 + f'(u_0(x_1))T =: \overline{x}$, i.e.

$$T = \frac{x_1 - x_0}{f'(v_0) - f'(v_1)}.$$

On a :

$$u(\overline{x}, T) = u(x_0 + f'(u_0(x_0))T, T) = u_0(x_0) = v_0$$
$$u(\overline{x}, T) = u(x_1 + f'(u_0(x_1))T, T) = u_0(x_1) = v_1$$

ce qui contredit $v_0 \neq v_1$. □

Définition 1.4.3 (Solution faible)**.** Soit $u_0 \in L^\infty(\mathbb{R})$ et soit $f \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$. On appelle solution faible de (1.19) toute solution du problème : trouver $u \in L^\infty(\mathbb{R} \times \mathbb{R}^+)$ t.q. : $\forall \varphi \in \mathcal{C}^1_c(\mathbb{R} \times \mathbb{R}, \mathbb{R})$,

$$\int_{\mathbb{R}} \int_{\mathbb{R}^+} \left(u(x,t) \frac{\partial \varphi}{\partial t} + f(u(x,t)) \frac{\partial \varphi}{\partial x} \right) dxdt + \int_{\mathbb{R}} u_0(x) \varphi(x, 0) dx = 0. \quad (1.22)$$

Le Théorème 1.4.2 montre qu'une solution faible de (1.19) perd sa régularité le long des droites caractéristiques, par exemple à l'intersection de deux telles droites. La Proposition 1.4.3 ci-dessous précise le rapport entre solution faible et solution classique. En particulier, elle montre qu'une solution de (1.19) qui serait classique en-dehors de ses droites caractéristiques est globalement une solution faible.

Proposition 1.4.3. *Soit $f \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$ et soit $u_0 \in \mathcal{C}(\mathbb{R}, \mathbb{R})$.*

1. *Toute solution classique de (1.19) est une solution faible de (1.19).*
2. *Si $u \in \mathcal{C}^1(\mathbb{R} \times \mathbb{R}^{+*}, \mathbb{R}) \cap \mathcal{C}(\mathbb{R} \times \mathbb{R}^+, \mathbb{R})$ est une solution faible de (1.19), alors u est une solution classique de (1.19).*
3. *Soit $\sigma \in \mathbb{R}$. On pose :*

$$D_1 = \{(x,t) \in \mathbb{R} \times \mathbb{R}^+ \mid x < \sigma t\}, \quad D_2 = \{(x,t) \in \mathbb{R} \times \mathbb{R}^+ \mid x > \sigma t\}$$

Soit $u \in \mathcal{C}(\mathbb{R}\times\mathbb{R}^+,\mathbb{R})$ vérifiant : $u|_{D_i} \in \mathcal{C}^1(D_i,\mathbb{R})$, $i \in \{1,2\}$, et u vérifie (1.19) en tout $(x,t) \in D_i$, $i \in \{1,2\}$. Alors u est une solution faible de (1.19).

Démonstration. 1. Soit u une solution lassique de (1.19) et soit $\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R},\mathbb{R})$. On a

$$\int_{\mathbb{R}}\int_{\mathbb{R}^+} \frac{\partial u}{\partial t}(x,t)\varphi(x,t)dxdt + \int_{\mathbb{R}}\int_{\mathbb{R}^+} \frac{\partial}{\partial x}(f(u(x,t)))\varphi(x,t)dxdt = 0.$$

On applique le théorème de Fubini et on intègre par parties :

$$0 = \int_{\mathbb{R}}\int_{\mathbb{R}^+} \frac{\partial u}{\partial t}(x,t)\varphi(x,t)dtdx + \int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{\partial}{\partial x}(f(u(x,t)))\varphi(x,t)dxdt$$

$$= -\int_{\mathbb{R}}\int_{\mathbb{R}^+} u(x,t)\frac{\partial\varphi}{\partial t}(x,t)dtdx - \int_{\mathbb{R}^+}\int_{\mathbb{R}} f(u(x,t))\frac{\partial\varphi}{\partial x}(x,t)dxdt + \int_{\mathbb{R}} u_0(x)\varphi(x,0)dx,$$

car $\mathrm{supp}(\varphi) \subset \mathbb{R}\times\mathbb{R}$ est compact.

2. Soit $u \in \mathcal{C}^1(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R}) \cap \mathcal{C}(\mathbb{R}\times\mathbb{R}^+,\mathbb{R})$ une solution faible de (1.19). On a suffisamment de régularité pour intègrer par parties dans (1.22), dans un premier temps avec $\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R})$, ce qui donne :

$$\int_{\mathbb{R}}\int_{\mathbb{R}^+} \underbrace{\left(\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(f(u(x,t)))\right)}_{\in\mathcal{C}(\mathbb{R}\times\mathbb{R}^+,\mathbb{R})}\varphi(x,t)dxdt = 0, \quad \forall\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R}).$$

Par densité de $\mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R})$ dans $\mathcal{C}(\mathbb{R}\times\mathbb{R}^{+*},\mathbb{R}$, on en déduit que

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(f(u(x,t))) = 0 \quad \text{dans} \quad \mathbb{R}\times\mathbb{R}^{+*}.$$

On intègre à nouveau par parties dans (1.22) avec $\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R},\mathbb{R})$, ce qui donne :

$$\int_{\mathbb{R}}(u_0(x) - u(x,0)\varphi(x,0)dx - \int_{\mathbb{R}}\int_{\mathbb{R}^+} \underbrace{\left(\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(f(u(x,t)))\right)}_{=0}\varphi(x,t)dxdt = 0.$$

On en déduit :

$$\int_{\mathbb{R}} \underbrace{(u_0(x) - u(x,0))}_{\in\mathcal{C}(\mathbb{R},\mathbb{R})}\varphi(x,0)dx = 0, \quad \forall\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R},\mathbb{R})$$

Par densité de $\mathcal{C}^1_c(\mathbb{R},\mathbb{R})$ dans $\mathcal{C}(\mathbb{R},\mathbb{R})$, on en déduit que $u(x,0) = u_0(x)$,$\forall x \in \mathbb{R}$. Donc u est une solution classique de (1.19).

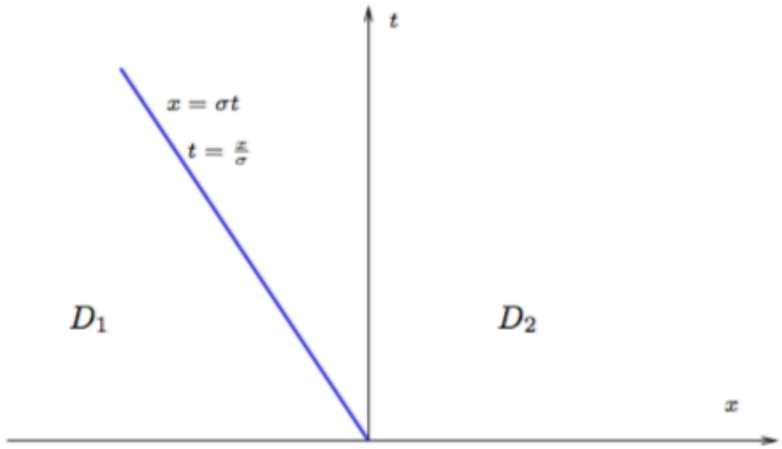

FIGURE 1.5 – Domaines D_1 et D_2

3. Soit $u \in \mathcal{C}(\mathbb{R} \times \mathbb{R}^+, \mathbb{R})$ vérifiant (1.19) en tout $(x,t) \in D_i$, $i \in \{1,2\}$. Soit $\varphi \in \mathcal{C}_c^1(\mathbb{R} \times \mathbb{R}, \mathbb{R})$. Soit $\sigma \in \mathbb{R}$. On suppose $\sigma < 0$ (voir figure 1.5) pour fixer les idées. Le cas $\sigma > 0$ se traite de la même façon. On a :

$$\int_{\mathbb{R}} \int_{\mathbb{R}^+} \left(u \frac{\partial \varphi}{\partial t} + f(u) \frac{\partial \varphi}{\partial x} \right) dxdt + \int_{\mathbb{R}} u_0(x) \varphi(x,0) dx =$$

$$= \sum_{i=1,2} \int_{D_i} \left(u \frac{\partial \varphi}{\partial t} + f(u) \frac{\partial \varphi}{\partial x} \right) dxdt + \int_{\mathbb{R}} u_0(x) \varphi(x,0) dx =$$

$$= - \sum_{i=1,2} \int_{D_i} \underbrace{\left(\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) \right)}_{=0} \varphi dxdt + \int_{\mathbb{R}} u_0(x) \varphi(x,0) dx +$$

$$+ \sum_{i=1,2} \int_{\partial D_i} (u \varphi \nu_t^i + f(u) \varphi \nu_x^i) d\sigma_i$$

où ν^i, $i \in \{1,2\}$, désigne le vecteur normal unitaire le long de ∂D_i orienté vers l'extérieur de D_i. On pose :

$$\nu = \begin{pmatrix} 1 \\ -\sigma \end{pmatrix}.$$

Alors : $\nu^1 = -\nu^2 = \nu$ le long de la droite $x = \sigma t$ et la continuité de $u\varphi$, resp. $f(u)\varphi$, entraîne : $u\varphi|_{\partial D_1 \cap \{x=\sigma t\}} = u\varphi|_{\partial D_2 \cap \{x=\sigma t\}} = u|_{x=\sigma t}$, resp. $f(u)\varphi|_{\partial D_1 \cap \{x=\sigma t\}} = f(u)\varphi|_{\partial D_2 \cap \{x=\sigma t\}} = u|_{x=\sigma t}$. Il en résulte :

$$\sum_{i=1,2} \int_{\partial D_i \cap \{x=\sigma t\}} (u\varphi\nu_t^i + f(u)\varphi\nu_x^i) d\sigma_i = 0.$$

Il reste :

$$\sum_{i=1,2} \int_{\partial D_i \setminus \{x=\sigma t\}} (u\varphi\nu_t^i + f(u)\varphi\nu_x^i) d\sigma_i = \sum_{i=1,2} \int_{\partial D_i \cap \{t=0\}} (u\varphi\nu_t^i + f(u)\varphi \underbrace{\nu_x^i}_{=0}) d\sigma_i =$$

$$= -\int_{\mathbb{R}} u(x,0)\varphi(x,0)dx = -\int_{\mathbb{R}} u_0(x)\varphi(x,0)dx.$$

Finalement :

$$\int_{\mathbb{R}} \int_{\mathbb{R}^+} \left(u\frac{\partial \varphi}{\partial t} + f(u)\frac{\partial \varphi}{\partial x} \right) dxdt + \int_{\mathbb{R}} u_0(x)\varphi(x,0)dx =$$

$$= -\int_{\mathbb{R}} u_0(x)\varphi(x,0)dx + \int_{\mathbb{R}} u_0(x)\varphi(x,0)dx = 0.$$

□

La démonstration du Théorème 1.4.2 a montré qu'une solution de (1.19) perd sa régularité à l'intersection de deux droites carcatéristques. En de tels points, le théorème d'existence et d'unicuté de Cauchy-Lipschitz ne s'applique plus au système des caractéristiques (1.20) et donc en particulier, on peut s'attendre à l'existence de plusieurs solutions faibles. Un exemple de non unicité de la solution faible est donné par l'équation de Burgers qui s'écrit :

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(u^2) = 0. \tag{1.23}$$

On complète (1.23) par une donnée initiale ad hoc permettant de déterminer analytiquement les solutions du problème de Cauchy associé aux caractéristiques qui sont les droites :

$$x - 2u_0(x_0)t = x_0, \quad x_0 \in \mathbb{R}$$

où on choisit :

$$u_0(x) = \begin{cases} u_d & \text{si} \quad x > 0, \\ u_g & \text{si} \quad x < 0, \end{cases} \tag{1.24}$$

avec $u_d, u_g \in \mathbb{R}$. On suppose $u_g < 0 < u_d$. Alors la méthode des caractéristiques donne :

$$u(x,t) = \begin{cases} u_d & \text{si} \quad x - 2u_d t < 0, \\ u_g & \text{si} \quad x - 2u_g t > 0, \end{cases}$$

A priori, $u(x,t)$ n'existe pas si $x - 2u_g t < 0$ et $x - 2u_d t > 0$. D'après la Proposition 1.4.3, on obtient une solution faible en posant :

$$u(x,t) = \begin{cases} u_d & \text{si} \quad x - 2u_d t < 0, \\ \dfrac{x}{2t} & \text{si} \quad x - 2u_d t > 0 \quad \text{et} \quad x - 2u_g t < 0, \\ u_g & \text{si} \quad x - 2u_g t > 0, \end{cases}$$

On peut aussi chercher une solution faible sous la forme

$$u(x,t) = \begin{cases} u_g & \text{si} \quad x < \sigma t \\ u_d & \text{si} \quad x > \sigma t \end{cases}$$

où le paramètre $\sigma \in \mathbb{R}$ reste à choisir. mais alors le critère de la Proposition 1.4.3 ne s'applique plus puisque la solution u est discontinue sur la droite $x - \sigma t = 0$. Après report dans (1.22), on obtient, $\forall \varphi \in \mathcal{C}_c^1(\mathbb{R} \times \mathbb{R}, \mathbb{R})$,

$$\int_{\{x<\sigma t\}} u_g \frac{\partial \varphi}{\partial t} dxdt + \int_{\{x>\sigma t\}} u_d \frac{\partial \varphi}{\partial t} dxdt + \int_{\{x<\sigma t\}} f(u_g) \frac{\partial \varphi}{\partial x} dxdt +$$

$$+ \int_{\{x>\sigma t\}} f(u_d) \frac{\partial \varphi}{\partial x} dxdt + \int_{-\infty}^{0} u_g \varphi(x,0) dx + \int_{0}^{+\infty} u_d \varphi(x,0) dx = 0.$$

On remarque que la normale à la droite $x - \sigma t = 0$ admet pour vecteur directeur (non unitaire, masi la relation est linéaire) $\nu \in \mathbb{R}^2$ de composantes $\nu_x = 1$, $\nu_t = \sigma$. Soit $\varphi \in \mathcal{C}1_c(\mathbb{R} \times \mathbb{R}^{+*}, \mathbb{R})$. Alors $\varphi(x,0) = 0$, $\forall x \in \mathbb{R}$ et on a :

$$\int_{\{x<\sigma t\}} u_g \frac{\partial \varphi}{\partial t} dxdt + \int_{\{x>\sigma t\}} u_d \frac{\partial \varphi}{\partial t} dxdt = \int_{0}^{+\infty} [u] \varphi(\sigma t, t) \nu_t dt =$$

$$= -(u_d - u_g)\sigma \int_{0}^{+\infty} \varphi(\sigma t, t) dt$$

$$\int_{\{x<\sigma t\}} f(u_g) \frac{\partial \varphi}{\partial x} dxdt + \int_{\{x>\sigma t\}} f(u_d) \frac{\partial \varphi}{\partial x} dxdt = \int_{0}^{+\infty} [f(u)] \varphi(\sigma t, t) \nu_x dt =$$

$$= (f(u_d) - f(u_g)) \int_0^{+\infty} \varphi(\sigma t, t) dt.$$

Il en résulte : $\forall \varphi \in \mathcal{C}_c^1(\mathbb{R} \times \mathbb{R}^{+*}, \mathbb{R})$,

$$0 = -(u_d - u_g)\sigma \int_0^{+\infty} \varphi(\sigma t, t) dt + (f(u_d) - f(u_g)) \int_0^{+\infty} \varphi(\sigma t, t) dt,$$

i.e.

$$\sigma(u_d - u_g) = f(u_d) - f(u_g). \tag{1.25}$$

La relation (1.25), appelée condition de Rankine et Hugoniot, donne pour $f(u) = u^2$:

$$\sigma = u_d + u_g,$$

et fournit la solution faible :

$$u(x,t) = \begin{cases} u_g & \text{si} \quad x < (u_g + u_d)t, \\ u_d & \text{si} \quad x > (u_g + u_d)t. \end{cases} \tag{1.26}$$

Néanmoins, parmi les solutions faibles, on peut distinguer une unique solution dite entropique.

Définition 1.4.4 (Solution entropique)**.** Soit $u_0 \in L^\infty(\mathbb{R})$ et soit $f \in \mathcal{C}(\mathbb{R}, \mathbb{R})$. On dit que $u \in L^\infty(\mathbb{R} \times \mathbb{R}^+)$ est une solution entropique de (1.19) si pour toute fonction $\eta \in \mathcal{C}^1(\mathbb{R})$ convexe appelée entropie et pour toute fonction $\phi \in \mathcal{C}^1(\mathbb{R})$ t.q. $f'\eta' = \phi'$ appelée flux d'entropie on a :

$$\int_{\mathbb{R}} \int_{\mathbb{R}^+} \left(\eta(u) \frac{\partial \varphi}{\partial t} + \phi(u) \frac{\partial \varphi}{\partial x} \right) + \int_{\mathbb{R}} \eta(u_0) \varphi(x, 0) dx \geq 0, \quad \forall \varphi \in \mathcal{C}_c^1(\mathbb{R} \times \mathbb{R}, \mathbb{R}^+). \tag{1.27}$$

Théorème 1.4.4 (Kruskov)**.** *Soit $u_0 \in L^\infty(\mathbb{R})$ et soit $f \in \mathcal{C}^1(\mathbb{R})$. Il existe une unique solution entropique de (1.19) au sens de la Définition 1.4.4.*

Proposition 1.4.5. *Toute solution classique de (1.19) est une solution entropique.*

Démonstration. Soit u une solution classique de (1.19). Soit $\eta \in \mathcal{C}^1(\mathbb{R})$ convexe une entropie et soit $\phi \in \mathcal{C}^1(\mathbb{R})$ t.q. $\phi' = f'\eta'$. Soit $\varphi \in \mathcal{C}_c^1(\mathbb{R} \times \mathbb{R}, \mathbb{R})$. On a :

$$\int_{\mathbb{R}} \int_{\mathbb{R}^+} \left(\eta(u) \frac{\partial \varphi}{\partial t} + \phi(u) \frac{\partial \varphi}{\partial x} \right) dx dt + \int_{\mathbb{R}} \eta(u_0) \varphi(x, 0) dx =$$

$$- \int_{\mathbb{R}} \eta(u_0) \varphi(x, 0) dx - \int_{\mathbb{R}} \int_{\mathbb{R}^+} \left(\eta'(u) \frac{\partial u}{\partial t} + \phi'(u) \frac{\partial u}{\partial x} \right) \varphi dx dt +$$

$$+\int_{\mathbb{R}}\eta(u_0)\varphi(x,0)dx \underset{\phi'=f'\eta'}{=} -\int_{\mathbb{R}}\int_{\mathbb{R}^+}\eta'(u)\underbrace{\left(\frac{\partial u}{\partial t}+f'(u)\frac{\partial u}{\partial x}\right)}_{=0}\varphi dxdt.$$

□

Proposition 1.4.6. *Toute solution faible entropique de (1.19) est une solution faible de (1.19).*

Démonstration. On remarque que $\eta \in \mathcal{C}^1(\mathbb{R})$ définie par : $\eta(x) = x$, $\forall x \in \mathbb{R}$ est convexe, donc une entropie. Alors $\phi = f \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$ est une flux d'entropie associé. □

On déduit de la Proposition 1.4.5 et du Théorème 1.4.4 de Kruskov que si (1.19) admet plusieurs solutions faibles et si l'une d'entre elles est régulière, alors cette dernière est nécessairement la solution entropique. Enfin, la caractŕisation suivante, que l'on admettra, est souvent utilisée en pratique.

Proposition 1.4.7 (Entropies de Kruskov)**.** *Si $u_0 \in L^\infty(\mathbb{R})$ et si $f \in \mathcal{C}^1(\mathbb{R})$, alors $u \in L^\infty(\mathbb{R}\times\mathbb{R}^+)$ est une solution entropique de (1.19) au sens de la Définition 1.4.4 ssi pour tout $\kappa \in \mathbb{R}$, la caractérisation (1.27) des solutions entropiques est vérifiée avec l'entropie η_κ définie par $\eta_\kappa(s) = |s-\kappa|$ et le flux denbtropie associé ϕ_κ défini par :*

$$\phi_\kappa(u) = \max(f(u),\kappa) - \min(f(u),\kappa), \quad \forall u \in \mathbb{R}.$$

Remarque 5. Dans la Proposition 1.4.7, l'entropie η_κ n'est pas de classe $\mathcal{C}^1$.

Proposition 1.4.8 (Estimation L^∞)**.** *Soit $u_0 \in L^\infty(\mathbb{R})$ et soit $A, B \in \mathbb{R}$ t.q. $A \le u_0 \le B$ p.p. Soit $f \in \mathcal{C}^1(\mathbb{R})$. Alors la solution entropique $u \in L^\infty(\mathbb{R}\times\mathbb{R}^+)$ de (1.19) vérifie : $A \le u \le B$ p.p.*

Cette propriété est fondamentale dans les problèmes de transport et il est donc important qu'elle soit conservée par les schémas numériques.

On termine cette Section par un résultat sur les solutions du problème de Riemann dont on s'est servi d'ailleurs pour montrer la non unicité des solutions faibles de (1.24).

Définition 1.4.5 (Problème de Riemann)**.** Soit $f \in \mathcal{C}^1(\mathbb{R})$. On appelle problème de Riemann avec données $u_g, u_d \in \mathbb{R}$ le problème : trouver u solution de :

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial}{\partial x}(f(u)) = 0, \quad x \in \mathbb{R}, \quad t > 0, \\ \\ u(x,0) = \begin{cases} u_g & \text{si} \quad x < 0, \\ u_d & \text{si} \quad x > 0, \end{cases} \end{cases} \tag{1.28}$$

Lorsque f est convexe ou concave, alors les solutions de (1.28) se calculent facilement. En effet, on peut montrer :

Proposition 1.4.9. *Soit $f \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$ strictement convexe et soit $u_g, u_d \in \mathbb{R}$.*

1. *Si $u_d < u_g$, on pose :*

$$\sigma = \frac{[f(u)]}{[u]} \quad avec \quad [u] = u_d - u_g, \quad [f(u)] = f(u_d) - f(u_g).$$

Alors, la fonction u définie par :

$$u(x,t) = \begin{cases} u_g & si \quad x < \sigma t, \\ u_d & si \quad x > \sigma t, \end{cases} \tag{1.29}$$

est l'unique solution entropique de (1.28). Une solution de la forme (1.29) est appelée une onde de choc.

2. *Si $u_g < u_d$, alors la fonction u définie par :*

$$u(x,t) = \begin{cases} u_g & si \quad x < f'(u_g)t, \\ u_d & si \quad x > f'(u_d)t, \\ \xi & si \quad x = f'(\xi)t \quad avec \quad u_g < \xi < u_d \end{cases} \tag{1.30}$$

est l'unique solution entropique de (1.28). Une solution de la forme (1.30) est appelée une onde de détente.

Démonstration. 1. On cherche une solution u sous la forme (1.29). Le même raisonnement que pour (1.25) montre que nécessairement :

$$\sigma = \frac{[f(u)]}{[u]}.$$

la méthode des caractéristiques fournit la solution :

$$u(x,t) = u_0(x - f'(u_0(x_0))t), \quad \forall (x,t) \in \mathbb{R} \times \mathbb{R}^+$$

i.e : $u(x,t) = u_0(x - f'(u_d)t)$ resp. $u(x,t) = u_0(x - f'(u_g)t)$ le long des caractéristiques issues de $x_0 > 0$, resp. $x_0 < 0$. Comme f est strictement convexe, f' est strictement croissante, et donc les deux caractéristiques issues de $x_0 = 0$ ont pour pentes respectives $f'(u_d) < f'(u_g)$. On en déduit que u vérifie :

$$u(x,t) = \begin{cases} u_d & \text{si} \quad x > f'(u_g)t, \\ u_g & \text{si} \quad x < f'(u_d)t. \end{cases}$$

Il reste à définir $u(x,t)$ pour $f'(u_d)t < x < f'(u_g)t$, $t > 0$. De l'hypothèse $u_d < u_g$ on déduit que $u(x,t)$ peut prendre les deux valeurs u_d, u_g si $f'(u_d)t < x < f'(u_g)t$, $t > 0$, i.e. qu'il existe une infinité de possibilités pour u. On discrimine en retenant la valeur σ ci-dessus compatible avec la caractérisation des solutions faibles. La stricte convexité de f entraîne :

$$f(u_g) > f(u_d) + (u_g - u_d)f'(u_d) \underset{u_g > u_d}{\Rightarrow} \sigma = \frac{f(u_g) - f(u_d)}{u_g - u_d} > f'(u_d)$$

$$f(u_d) > f(u_g) + (u_d - u_g)f'(u_g) \underset{u_g > u_d}{\Rightarrow} -\sigma = \frac{f(u_d) - f(u_g)}{u_g - u_d} > -f'(u_g)$$

$$\iff \sigma < f'(u_g)$$

Finalement :

$$f'(u_d) < \sigma < f'(u_g)$$

et on peut prolonger u par continuité le long des caractéritiques $x - f'(u_d)t = 0$ et $x - f'(u_g)t = 0$ en posant :

$$u(x,t) = \begin{cases} u_g & \text{si} \quad x < \sigma t, \\ u_d & \text{si} \quad x > \sigma t. \end{cases}$$

Il reste à vérifier que u est la solution entropique. Soit $\eta \in \mathcal{C}^1(\mathbb{R})$ convexe et soit $\phi \in \mathcal{C}^1(\mathbb{R})$ t.q. $\phi' = f'\eta'$. Soit $\varphi \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}, \mathbb{R}^+)$. Après intégration par parties on obtient :

$$\int_{\mathbb{R}} \int_{\mathbb{R}^+} \left(\eta(u) \frac{\partial \varphi}{\partial t} + \phi(u) \frac{\partial \varphi}{\partial x} \right) dxdt + \int_{\mathbb{R}} \eta(u_0(x)) \varphi(x,0) dx =$$

$$= \underbrace{(-\sigma[\eta(u)] + [\phi(u)])}_{\underset{Lemme\ 1.4.10}{\geq 0}} \underbrace{\int_0^{+\infty} \varphi(\sigma t, t) dt}_{>0} \geq 0.$$

2. On remarque que par convexité stricte de $f \in \mathcal{C}^1(\mathbb{R})$, f' est continue et strictement croissante donc inversible. En particulier : $u_g < u_d \Rightarrow f'(u_g) < f'(u_d)$ Par la méthode des caractéristiques, on obtient la solution :

$$u(x,t) = \begin{cases} u_d & \text{si} \quad x > f'(u_d)t, \\ u_g & \text{si} \quad x < f'(u_g)t. \end{cases}$$

Il reste à définir $u(x,t)$ pour $f'(u_g)t < x < f'(u_d)t$, $t > 0$. On proilonge u par continuité le long des caractéristiques $x = f'(u_d)t$ et $x = f'(u_g)t$ en posant :

$$u(x,t) = f'^{-1}\left(\frac{x}{t}\right) \quad \text{si} \quad f'(u_g)t < x < f'(u_d)t, \quad t > 0. \qquad (1.31)$$

Il reste à vérifier que u est la solution entropique. Soit $\eta \in \mathcal{C}^1(\mathbb{R})$ convexe et soit $\phi \in \mathcal{C}^1(\mathbb{R})$ t.q. $\phi' = f'\eta'$. Soit $\varphi \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}, \mathbb{R}^+)$. Après intégration par parties on obtient :

$$\int_{\mathbb{R}} \int_{\mathbb{R}^+} \left(\eta(u) \frac{\partial \varphi}{\partial t} + \phi(u) \frac{\partial \varphi}{\partial x} \right) dxdt + \int_{\mathbb{R}} \eta(u_0(x)) \varphi(x,0) dx =$$

$$= - \iint_{\{f'(u_g)t < x < f'(u_d)t\}} \eta'(u) \left(\frac{\partial u}{\partial t} + f'(u) \frac{\partial u}{\partial x} \right) \varphi(x,t) dxdt =$$

$$= - \iint_{\{f'(u_g)t < x < f'(u_d)t\}} \eta' \left(\frac{x}{t} \right) (f'^{-1})' \left(\frac{x}{t} \right) \left(-\frac{x}{t^2} + f'(u) \frac{1}{t} \right) \varphi(x,t) dxdt =$$

$$= - \iint_{\{f'(u_g)t < x < f'(u_d)t\}} \eta' \left(\frac{x}{t} \right) \underbrace{\left(-\frac{x}{t} + f'(u) \right)}_{\substack{=0 \\ (1.31)}} \varphi(x,t) \frac{dx}{t} dt = 0$$

□

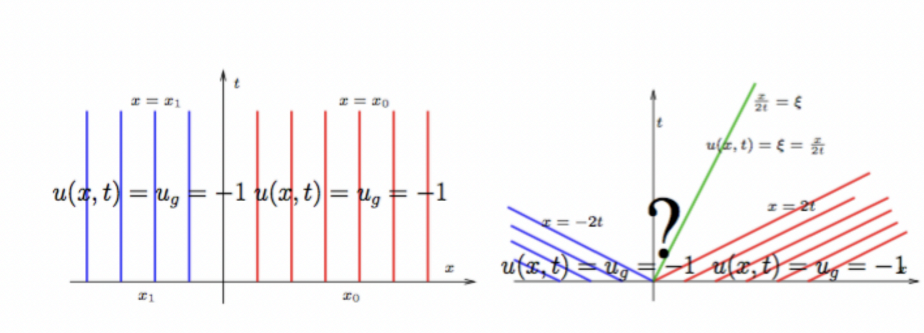

FIGURE 1.6 – Problème de Riemann pour l'équation de Burgers.

Lemme 1.4.10. *Soit $a < b$, soit $f \in \mathcal{C}^1(\mathbb{R})$ et $\eta \in \mathcal{C}^1(\mathbb{R})$ des fonctions convexes et soit $\phi \in \mathcal{C}^1(\mathbb{R})$ t.q. $\phi' = f'\eta'$. Alors :*

$$(b-a)(\phi(b) - \phi(a)) \geq (f(b) - f(a))(\eta(b) - \eta(a)).$$

Démonstration. On a :

$$\phi(b) - \phi(a) = \int_a^b \phi'(x) dx = \int_a^b f'(x) \eta'(x) dx$$

donc :

$$\phi(b)-\phi(a) = \int_a^b f'(x)(\eta'(x)-\eta'(y))dx+\eta'(y)\int_a^b f'(x)\eta'(x)dx, \quad \forall y \in [a,b].$$

On intègre par rapport à y :

$$(b-a)(\phi(b)-\phi(a)) = \int_a^b\int_a^b f'(x)(\eta'(x)-\eta'(y))dxdy+\int_a^b \eta'(y)dy\int_a^b f'(x)dx =$$

$$= \int_a^b\int_a^b f'(x)(\eta'(x)-\eta'(y))dxdy + (\eta(b)-\eta(a))(f(b)-f(a)).$$

avec, par convexité de η et f :

$$\int_a^b\int_a^b f'(x)(\eta'(x)-\eta'(y))dxdy = \iint_{\{x<y\}} f'(x)(\eta'(x)-\eta'(y))dxdy+$$

$$+\iint_{\{x>y\}} f'(x)\underbrace{(\eta'(x)-\eta'(y))}_{>0}dxdy$$

$$\geq \iint_{\{x<y\}} f'(x)(\eta'(x)-\eta'(y))dxdy + \iint_{\{x>y\}} f'(y)(\eta'(x)-\eta'(y))dxdy = 0$$

donc

$$(b-a)(\phi(b)-\phi(a)) \geq (\eta(b)-\eta(a))(f(b)-f(a)).$$

□

1.5 Le cas non linéaire : schémas numériques

Soit $u_0 \in L^\infty(\mathbb{R})$ et soit $f \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$. On cherche une approximation numérique de la solution entropique de (1.19). On utilise les mêmes notations que pour le schéma (1.18). On note $f^n_{i+\frac{1}{2}} \sim f(u(x_{i+\frac{1}{2}},t_n))$ l'approximation du flux numérique. On considère le schéma :

$$\begin{cases} \dfrac{u_i^{n+1}-u_i^n}{\Delta t} + \dfrac{f^n_{i+\frac{1}{2}}-f^n_{i-\frac{1}{2}}}{\Delta x} = 0, \quad i \in \mathbb{Z}, \quad n \geq 0, \\ u_i^0 = \dfrac{1}{\Delta x}\displaystyle\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u_0(x)dx, \qquad i \in \mathbb{Z}. \end{cases} \tag{1.32}$$

dans lequel $f^n_{i+\frac{1}{2}}$ reste à définir. Un premier choix possible est le schélma centré :

$$f^n_{i+\frac{1}{2}} = \frac{f(u^n_{i+1}) + f(u^n_i)}{2}$$

dont on a vu qu'il est à proscrire puisque, dans le cas linéaire, il est instable. On va s'intéresser aux schémas les plus simples à trois points, i.e. dans lesquels l'équation associée à l'inconnue u^n_i fait intervenir les trois inconnues discrètes $u^n_{i\pm1}$ et u^n_i, $i \in \mathbb{Z}$. Le flux numérique s'écrit sous la forme $f^n_{i+\frac{1}{2}} = g(u^n_i, u^n_{i+1})$. Un bon schéma est obtenu en choisissant un flux monotone au sens suivant.

Définition 1.5.1. Une fonction $g : \mathbb{R}^2 \to \mathbb{R}$ est un flux monotone pour la discrétisation de (1.19) si :

1. g est consistante par rapport à f, i.e. $g(u, u) = f(u)$;
2. $(x_1, x_2) \mapsto g(x_1, x_2)$ est croissante par rapport à la variable x_1 et décroissante par rapport à la variable x_2 ;
3. g est lipschitzienne sur $[A, B]$ où $A = \inf_{\mathbb{R}} u_0$ et $B = \sup_{\mathbb{R}} u_0$.

Remarque 6 (Flux monotones et schémas monotones). Si le shéma (1.32) est à flux monotone et s'il vérifie la condition de CFL, alors on peut montrer qu'il est monotone, i.e. qu'il peut s'écrire sous la forme $u^{n+1}_i = H(u^n_{i-1}, u^n_i, u^n_{i+1})$ où H est une fonction croissante de chacun de ses trois arguments.

Cas où f est monotone

Pour illustrer le choix de g, on suppose par exemple que f est croissante. Un choix très simple consiste alors à prendre $g(u^n_i, u^n_{i+1}) = f(u^n_i)$. Alors : $g(u, u) = f(u)$, $\forall u \in \mathbb{R}$, i.e. g est consistante par rapport à f. Par construction :

$$g(x_1, x_2) = f(x_1), \quad \forall (x_1, x_2) \in \mathbb{R}^2$$

i.e. g est croissante par rapport à la variable x_1 par hypothèses sur f, et mnanifestement décroissante par rapport à la variable x_2. Soit $x, y \in [A, B]^2$. On a :

$$|g(x) - g(y)| = |f(x_1) - f(y_1)| = \left| \int_{x_1}^{y_1} f'(t)dt \right| \leq \left| \int_{x_1}^{y_1} |f'(t)|dt \right| \leq$$

$$\leq |x_1 - y_1| \sup_{[A,B]} |f'| \leq C\|x - y\| \sup_{[A,B]} |f'| < +\infty$$

où $C > 0$ est une constante ne dépendant que du choix de la norme $\|\cdot\|$ de $\mathbb{R}^2$, i.e. g est lipschitzienne sur $[A, B]$. Le schéma résultant est dit décentré amont et s'écrit :

$$\begin{cases} \dfrac{u_i^{n+1} - u_i^n}{\Delta t} + \dfrac{f(u_i^n) - f(u_{i-1}^n)}{\Delta x} = 0, & i \in \mathbb{Z}, \quad n \geq 0, \\ u_i^0 = \dfrac{1}{\Delta x} \displaystyle\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u_0(x)dx, & i \in \mathbb{Z}. \end{cases}$$

On vérifie directement que dans le cas linéaire on retrouve le schéma décentré amont (1.18).

Schéma à décomposition de flux

Le schéma à décomposition de flux (flux splitting) est caractérisé par le choix $f = f_1 + f_2$ où f_1, resp. f_2, est croissante, resp. décroissante. Alors on pose : $g(u_1, u_2) = f_1(u_1) + f_2(u_2)$, $\forall u = (u_1, u_2) \in \mathbb{R}^2$, et alors : $g(u_i^n, u_{i+1}^n) = f_1(u_i^n) + f_2(u_{i+1}^n)$, $\forall i \in \mathbb{Z}$, $\forall n \geq 0$.

Schéma de Lax-Friedrich

Le schéma de Lax-Friedrich consiste à modifier le schéma centré de façon à le rendre stable en posant :

$$f_{i+\frac{1}{2}}^n = g(u_i^n, u_{i+1}^n) = \frac{1}{2}\left(f(u_i^n) + f(u_{i+1}^n)\right) + D(u_i^n - u_{i+1}^n)$$

où $D \geq 0$ est suffisamment grand pour que $(u_1, u_2) \mapsto (u_1, u_2)$ soit croissante, resp. décroissante, par rapport à u_1, resp. u_2.

Schéma de Godunov

Le schéma de Godunov est un des plus connus et il a inspiré de nombreux autres schémas. Le fux numérique du schéma de Godunov s'écrit :

$$g(u, v) = w_R(0, u, v) := \begin{cases} f(u) & \text{si} \quad f'(u) > 0, \\ f \circ f'^{-1}(0) & \text{si} \quad f'(u) > 0 \quad \text{et} \quad f'(v) > 0, \\ f(v) & \text{si} \quad f'(v) < 0. \end{cases} \tag{1.33}$$

On montre que le flux de Godunov (1.33) vérifie les conditions de la Définition 1.5.1. Pour le voir on montre que le flux de Godunov s'écrit (*Exercice*) :

$$g(u_i^n, u_{i+1}^n) = \begin{cases} \min\limits_{\xi\in[u_i^n, u_{i+1}^n]} f(\xi) & \text{si} \quad u_i^n \leq u_{i+1}^n, \\ \\ \max\limits_{\xi\in[u_{i+1}^n, u_i^n]} f(\xi) & \text{si} \quad u_{i+1}^n \leq u_i^n, \end{cases}$$

Théorème 1.5.1 (Stabilité et convergence)**.** *Soit $(u_i^n)_{i\in\mathbb{Z},n\geq 0}$ la suite définie par le schéma numérique :*

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

On suppose que g est un flux monotone au sens de la Définition 1.5.1 et lipschitzienne de constante M sur $[A, B]$ avec

$$A = \inf_{\mathbb{R}} u_0, \quad B = \sup_{\mathbb{R}} u_0.$$

On suppose de plus que

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

Alors le schéma est stable en temps fini et converge vers la solution du problème (1.19). De plus :

$$A \leq u_i^n \leq B, \quad \forall i \in \mathbb{Z}, \quad \forall n \geq 0.$$

Démonstration. On pose : $\forall (x,t) \in \mathbb{R}\times\mathbb{R}^+$,

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

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

$$g(u(x,t), u(x+\Delta x, t)) = f(u(x,t)) + \Delta x \frac{\partial g}{\partial u_2} \circ (u,u) \frac{\partial u}{\partial x} + O((\Delta x)^2)$$

$$g(u(x-\Delta x, t), u(x,t)) = f(u(x,t)) - \Delta x \frac{\partial g}{\partial u_1} \circ (u,u) \frac{\partial u}{\partial x} + O((\Delta x)^2)$$

d'où :

$$\varepsilon(x,t) = \frac{\partial u}{\partial t} + \underbrace{\left(\frac{\partial g}{\partial u_2} + \frac{\partial g}{\partial u_1}\right) \circ (u,u) \frac{\partial u}{\partial x}}_{= \frac{\partial}{\partial x} g(u,u) = f'(u)} + O(\Delta t) + O(\Delta x)$$

$$= \underbrace{\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(f(u(x,t)))}_{\substack{=0 \\ (1.19)}} + O(\Delta t) + O(\Delta x) = O(\Delta t) + O(\Delta x).$$

Il en résulte :

$$\varepsilon_i^n := \varepsilon(x_i, t_n) = O(\Delta t) + O(\Delta x), \quad \forall i \in \mathbb{Z}, \quad \forall n \geq 0.$$

On en déduit l'erreur de consistance :

$$\mathcal{E}(\Delta x, \Delta t) := \sup_{i \in \mathbb{Z}, n \geq 0} |\varepsilon_i^n| = O(\Delta t) + O(\Delta x).$$

Soit $(\mu_i^n)_{i \in \mathbb{Z}, n \geq 0}$ une suite de réels > 0 et soit $(z_i^n)_{i \in \mathbb{Z}, n \geq 0}$ la suite définie par le schéma :

$$\begin{cases} \dfrac{z_i^{n+1} - z_i^n}{\Delta t} + \dfrac{1}{\Delta x}\left(g(z_i^n, z_{i+1}^n) - g(z_{i-1}^n, z_i^n)\right) = \mu_i^n, \quad i \in \mathbb{Z}, \quad n \geq 0, \\ \\ z_i^0 \in \mathbb{R}, \quad i \in \mathbb{Z}. \end{cases} \tag{1.34}$$

On pose :

$$H(u, v, w) := v - \frac{\Delta t}{\Delta x}(g(v, w) - g(u, v)), \quad \forall u, v, w \in \mathbb{R},$$

de sorte que le schéma (1.34) se réécrit :

$$z_i^{n+1} = H(z_{i-1}^n, z_i^n, z_{i+1}^{n+1}) + \Delta \mu_i^n, \quad i \in \mathbb{Z}, \quad n \geq 0.$$

La fonction H est croissante par rapport à chacune de ses variables. En effet : $\forall u = (u_0, u_1, u_2) \in \mathbb{R}^3$, $\forall v = (v_0, v_1, v_2) \in \mathbb{R}^3$,

$$\frac{H(u_0, v_1, u_2) - H(u_0, u_1, u_2)}{v_1 - u_1} = 1 + \frac{\Delta t}{\Delta x}\left(\underbrace{-\frac{g(v_1, u_2) - g(u_1, u_2)}{v_1 - u_1}}_{>0} + \underbrace{\frac{g(u_0, v_1) - g(u_0, u_1)}{v_1 - u_1}}_{<0}\right)$$

$$\geq 1 + \frac{\Delta t}{\Delta x}(-M - M) = 1 - 2M\frac{\Delta t}{\Delta x} > 0,$$

$$\frac{H(v_0, u_1, u_2) - H(u_0, u_1, u_2)}{v_0 - u_0} = \frac{\Delta t}{\Delta x}\left(0 + \underbrace{\frac{g(v_0, u_1) - g(u_0, u_1)}{v_0 - u_0}}_{>0}\right) > 0,$$

$$\frac{H(u_0,u_1,v_2)-H(u_0,u_1,u_2)}{v_2-u_2}=\frac{\Delta t}{\Delta x}\left(-\underbrace{\frac{g(u_1,v_2)-g(u_1,u_2)}{v_2-u_2}}_{<0}+0\right)>0$$

On en déduit : $\forall i\in\mathbb{Z}$,

$$z_i^1=H(z_{i-1}^0,z_i^0,z_{i+1}^0)+\Delta t\mu_i^0\leq H(\|z^0\|_\infty,\|z^0\|_\infty,\|z^0\|_\infty)+\Delta t\|\mu^0\|_\infty=\|z^0\|_\infty+\Delta t\|\mu^0\|_\infty$$

et

$$z_i^1=H(z_{i-1}^0,z_i^0,z_{i+1}^0)+\Delta t\mu_i^0\geq H(-\|z^0\|_\infty,-\|z^0\|_\infty,-\|z^0\|_\infty)-\Delta t\|\mu^0\|_\infty$$

$$=-\|z^0\|_\infty-\Delta t\|\mu^0\|_\infty$$

i.e. : $\|z^1\|_\infty\leq\|z^0\|_\infty+\Delta t\|\mu^0\|_\infty$. On suppose que :

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

Alors :

$$z_i^{n+1}=H(z_{i-1}^n,z_i^n,z_{i+1}^n)+\Delta t\mu_i^n\leq H(C_n,C_n,C_n)+\Delta t\|\mu^n\|_\infty=$$

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

et

$$z_i^{n+1}=H(z_{i-1}^n,z_i^n,z_{i+1}^n)+\Delta t\mu_i^n\geq H(-C_n,-C_n,-C_n)-\Delta t\|\mu^n\|_\infty=$$

$$=-C_n-\Delta t\|\mu^n\|_\infty=-\|z^0\|_\infty-\Delta t\sum_{k=0}^{n}\|\mu^k\|_\infty$$

i.e. :

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

Si $\mu_i^n=0$, alors $z_i^n=u_i^n$, $\forall i\in\mathbb{Z}$, $\forall n\geq 0$, et alors :

$$\|u^{n+1}\|_\infty\leq\|u^0\|_\infty,\quad\forall n\geq 0,$$

i.e., le schéma est stable.

Si $z_i^n = u(x_i, t_n) - u_i^n =: e_i^n$, alors $\mu^i = \varepsilon_i^n = O(\Delta t + \Delta x)$, $\forall i \in \mathbb{Z}$, $\forall n \geq 0$, et alors :

$$\|e^n\|_\infty \leq C\Delta t \sum_{k=0}^{n-1} (\Delta t + \Delta x) = n\Delta t(\Delta t + \Delta x)$$

i.e. :

$$\sup_{n\Delta t \leq T} \leq CT(\Delta t + \Delta x), \quad \forall T > 0.$$

□

Remarque 7. Sans l'hypothèses de monotonie sur g, on peut seulement montrer que le schéma est convergent quand $\Delta \to 0$. Plus précisément : $\forall i \in \mathbb{Z}$, $\forall n \geq 0$,

$$|z_i^{n+1} - u_i^{n+1}| \leq |z_i^n - u_i^n| + \frac{M\Delta t}{\Delta x}(|z_{i+1}^n - u_{i+1}^n| + 2|z_i^n - u_i^n| + |z_{i-1}^n - u_{i-1}^n|) + \Delta t|\mu_i^n|$$

$$\leq \left(1 + \frac{4M\Delta t}{\Delta x}\right) \|z^n - u^n\|_\infty + \Delta t\|\mu^n\|_\infty$$

$$\Rightarrow \|z^{n+1} - u^{n+1}\|_\infty \leq \left(1 + \frac{4M\Delta t}{\Delta x}\right) \|z^n - u^n\|_\infty + \Delta t\|\mu^n\|_\infty$$

$$\leq \left(1 + \frac{4M\Delta t}{\Delta x}\right)^{n+1} \underbrace{\|z^0 - u^0\|_\infty}_{\substack{=0 \\ (1.34)}} + \Delta t \sum_{k=0}^{n} \left(1 + \frac{4M\Delta t}{\Delta x}\right)^k \|\mu^{n-k}\|_\infty$$

Si $z_i^n = u(x_i, t_n)$, alors $\mu_i^n = \varepsilon_i^n$ et on note $e_i^n = z_i^n - u_i^n$ l'erreur de convergence en (x_i, t_n), $\forall i \in \mathbb{Z}$, $\forall n \geq 0$. On en déduit :

$$\|e^n\|_\infty \leq C\Delta t \left(\frac{\left(1 + \frac{4M\Delta t}{\Delta x}\right)^n - 1}{\frac{4M\Delta t}{\Delta x}}\right) (\Delta t + \Delta x)$$

$$\Rightarrow \sup_{i\in\mathbb{Z}, n\Delta t \leq T} \|e^n\|_\infty \leq \frac{C}{4M}(\Delta x)^2 \left(1 + \frac{\Delta t}{\Delta x}\right) e^{\frac{4MT}{\Delta x}}, \quad \forall T > 0.$$

Bibliographie

[1] Thierry Gallouët, Raphaèle Herbin. Analyse numérique des équations aux dérivées partielles. Master. Marseille, France. 2011. (https ://cel.hal.science/cel-00637008v2) *Chapitre 5.*

[2] Lionel Sainsaulieu. Calcul Scientifique. Cours et exercices corrigés pour le second cycle et les écoles d'ingénieurs, Masson, Paris 1996. *Chapitres 1.4 et 2.3.*

[3] Brigitte Lucquin, Olivier Pironneau. Introduction au calcul scientifique. Masson, Paris, 1996. *Chapitres I.6 et VII.4*

[4] Alfio Quarteroni, Riccardo Sacco, Fausto Saleri Numerical Mathematics. Springer, Berlin, 2007. *Chapitres 13.5 à 13.8*

[5] Alfio Quarteroni, Riccardo Sacco, Paola Gervasio Calcul scientifique. Springer, Milan, 2010. *Chapitres 8.3.1 et 8.3.2*

Chapitre 2

Equation de Laplace

2.1 Modélisation

Dans $\mathbb{R}^n$ on considère la variation d'une quantité telle que la température T sous l'effet de forces volumiques de densité f en l'absence de tout champ de vitesses (par exemple, la chaleur n'est pas mue par un champ de vitesses). Autrement dit, l'énergie du matériau considérée se réduit à son énergie potentielle de la forme $\frac{1}{2}\|\nabla T\|^2$ dont la force associée dérive d'un potentiel, soit $\vec{\nabla} T$ (loi de Fourier). Si V est un volume quelconque fixe, la force volumique $f_V := \int_V f(x) d\Omega$ compense exactement le flux à la frontière de V d'origine potentielle :

$$f_V = -\int_{\partial V} \vec{\nabla} T \cdot \overrightarrow{dS}$$

où $\overrightarrow{dS}$ est la normale extérieure à V le long de ∂V. On en déduit :

$$\int_V f(x) d\Omega = -\int_{\partial V} \underbrace{\vec{\nabla} T \cdot \overrightarrow{dS}}_{=:\omega} = -\int_{\partial V} \omega \underset{Stokes}{=} -\int_V d\omega =$$

$$= -\int_V \underbrace{\mathrm{div}(\vec{\nabla} T)}_{=\Delta T} d\Omega = -\int_V \Delta T d\Omega.$$

Ceci est vrai pour tout volume $V \subset \mathbb{R}^n$ suffisamment régulier donc

$$f = -\Delta T \quad \text{dans} \quad \mathbb{R}^n.$$

Définition 2.1.1 (Laplacien)**.** Dans $\mathbb{R}^n$ on appelle Laplacien l'opérateur :

$$\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}.$$

dit elliptique car sa transformée de Fourier

$$\mathcal{F}(\Delta)(\xi) = \xi_1^2 + \xi_2^2$$

est le terme principal de l'équation d'une sphère et plus généralement d'une ellipse. Du fait de sa symétriel le laplacien estr le modèle des pŕatuers elliptiques :

$$x \mapsto \sum_{i=1}^{n} a_i(x) \frac{\partial^2}{\partial x_i^2}.$$

2.2 Le Laplacien comme opérateur non borné

Soit $\Omega \subset \mathbb{R}^n$ un ouvert connexe, borné ou non. L'opérateur Δ est bien défini au sens des distributions sur $L^2(\Omega)$:

$$\langle \Delta u, \varphi \rangle := \int_\Omega u \Delta \varphi dx, \quad \forall \varphi \in \mathcal{D}(\Omega)$$

mais il n'opère pas dans $L^2(\Omega)$. Pour remédier à cela, on utilise la notion d'opérateur non borné.

Définition 2.2.1 (Opérateur borné)**.** Un opérateur $A : \mathcal{H} \to \mathcal{H}$ défini sur un espace de Hilbert $\mathcal{H}$ est dit borné s'il est continu, i.e. s'il existe une constante $C > 0$ t.q.

$$\|Ax\| \leq C\|x\|, \quad \forall x \in \mathcal{H}.$$

Un opérateur non borné est défini, en général, sur un sous-espace $D(A)$ dense de $\mathcal{H}$. Un opérateur non borné est donc une application linéaire $A : D(A) \to \mathcal{H}$. Si A est un opérateur à domaine dense $D(A)$, on définit $D(A^*)$ comme l'ensemble des vecteurs $\varphi \in \mathcal{H}$ pour lesquels il existe un vecteur $\varphi' \in \mathcal{H}$ vérifiant :

$$\forall \psi \in D(A), \quad \langle A\psi, \varphi \rangle = \langle \psi, \varphi' \rangle.$$

Pour tout $\varphi \in D(A^*)$, domaine de l'adjoint, on note $A^*\varphi = \varphi'$. Par densité de $D(A)$, φ' est défini de façon unique. Un opérateur A sur $\mathcal{H}$ est dit-auto-adjoint si $D(A^*) = D(A)$ et $A = A^*$ sur $D(A)$. Malheureusement, dans beaucoup de cas intéressants, $D(A^*)$ est beaucoup plus petit que $D(A)$, voire est réduit à $\{0\}$. Avec ces notations, on pose :

$$D(-\Delta) = \{u \in L^2(\Omega) \mid -\Delta u \in L^2(\Omega)\}.$$

On remarque que pour tout $u \in D(-\Delta)$, on peut définir la trace $u|_{\partial\Omega}$ et la dérivée normale $\dfrac{\partial u}{\partial n}$ le long de $\partial\Omega$ en posant : $\forall \varphi \in \mathcal{C}^\infty(\Omega)$,

$$\int_{\partial\Omega} u\frac{\partial\varphi}{\partial n}d\sigma - \int_{\partial\Omega}\frac{\partial u}{\partial n}\varphi d\sigma = \int_\Omega (-\Delta u)\varphi dx + \int_\Omega u\Delta\varphi dx, \quad \text{où} \quad \frac{\partial\varphi}{\partial n} := \vec{n}\cdot\vec{\nabla}\varphi$$

L'adjoint de $-\Delta$ est bien défini quand il est associé à une condition sur le bord de type Dirichlet, resp. Neumann,

Proposition 2.2.1. *Les opérateurs $-\Delta_d$ et $-\Delta_n$, définis par :*

$$D(-\Delta_d) = \{u \in L^2(\Omega) \mid \Delta u \in L^2(\Omega) \quad et \quad u|_{\partial\Omega} = 0\}$$

et : $-\Delta_d u = -\Delta u$, $\forall u \in D(-\Delta_d)$, resp. :

$$D(-\Delta_n) = \{u \in L^2(\Omega) \mid \Delta u \in L^2(\Omega) \quad et \quad \left.\frac{\partial u}{\partial n}\right|_{\partial\Omega} = 0\}$$

et : $-\Delta_n u = -\Delta u$, $\forall u \in D(-\Delta_n)$, sont auto-adjoints et positifs.

Démonstration. On a les égalités :

$$\int_\Omega (-\Delta u)\varphi dx = \int_\Omega u(-\Delta\varphi)dx, \quad \forall u, \varphi \in D(-\Delta_d)$$

i.e. $(-\Delta_d)^* = -\Delta_d$, resp.

$$\int_\Omega (-\Delta u)\varphi dx = \int_\Omega u(-\Delta\varphi)dx, \quad \forall u, \varphi \in D(-\Delta_n)$$

i.e. $(-\Delta_n)^* = -\Delta_n$.

De plus, si $u \in D(-\Delta)$, on prolonge $u \in L^2(\Omega)$ et $-\Delta u \in L^2(\Omega)$ par 0 dans $\mathbb{R}^n \setminus \Omega$. On en déduit alors :

$$\int_\Omega (-\Delta u)u dx = \int_{\mathbb{R}^n} \overline{\mathcal{F}(-\Delta u)}\mathcal{F}(u)dx = \int_{\mathbb{R}^n} |\xi|^2 |\mathcal{F}(u)(\xi)|^2 d\xi \geq 0$$

i.e. $-\Delta$ est un opérateur positif. □

Corollaire 2.2.2. *Le spectre de l'opérateur $-\Delta_d$, resp. $-\Delta_n$, est formé d'une suite de valeurs propres $(\lambda_k^{(d)})_{k\geq 0} \in (\mathbb{R}^+)^{\mathbb{N}}$, resp.$(\lambda_k^{(n)})_{k\geq 0} \in (\mathbb{R}^+)^{\mathbb{N}}$, sans point d'accumulation à distance finie, t.q. $\lim_{k\to+\infty}\lambda_k^{(d)} = +\infty$, resp. $\lim_{k\to+\infty}\lambda_k^{(n)} = +\infty$, et les vecteurs propres correspondants forment une base hilbertienne de $L^2(\Omega)$.*

Démonstration. C'est une conséquence (admise) de la théorie des opérateurs auto-adjoints positifs. □

L'exemple le plus simple de la situation décrite dans le Corollaire 2.2.2 s'obtient en dimension $n = 1$ d'espace. On pose $\Omega =]0, L[$ avec $L > 0$ et on considère l'équation :

$$-u'' - \lambda u = 0, \quad u(0) = u(L) = 0.$$

Par le calcul, on trouve directement que les valeurs propres sont les réels :

$$\lambda_k = \left(\frac{k\pi}{L}\right)^2 \geq 0, \quad k \geq 0$$

de vecteurs propres associés :

$$t \mapsto w_{\lambda_k}(t) = \sin\left(\frac{k\pi}{L}\right), \quad k \geq 0.$$

On vérifie que la suite $(w_k)_{k\geq 0}$ est une base hilbertienne de $L^2(0, T)$.

Remarque 8. L'analyse ci-dessus se généralise au cas du problème de Dirichlet (resp. de Neumann, et d'autres) dans un ouvert Ω borné. Pour évaluer le comporteme,nt asymptotique des valeurs propres on introduit la fonction de comptage :

$$N(\lambda) := \sharp\{\lambda_k \leq \lambda\}, \quad \forall \lambda \in \mathbb{R}.$$

Ces valeurs propres apparaisent de manière fondamentale dans les équations de propagation des ondes et de la chaleur sur un ouvert Ω.

Elles apparaissent aussi en théorie des nombres. Si Ω est un carré de côté 1, alors les fonctions propres et les valeurs propres du Laplacien sont donnés par :

$$w_{\lambda_{n,m}}(x, y) = \sin(n\pi x)\sin(m\pi y), \quad \lambda_{n,m} = (n^2 + m^2)\pi^2, \quad n, m \geq 0.$$

L'évaluation de $N(\lambda)$ dans ce cas est bien un problème de théorie des nombres : compter le nombre de points à coordonnées entières contenus dans un cercle de rayon $\sqrt{\lambda}$, $\lambda > 0$. On démontre que :

$$N(\lambda) = \frac{\sqrt{\lambda}}{4\pi} + o(\sqrt{\lambda}), \quad \forall \lambda > 0.$$

2.3 Formulation variationnelle

Introduction et formalisme

Soit $\Omega \subset \mathbb{R}^n$ un ouvert borné de $\mathbb{R}^n$ et soit $f \in L^2(\Omega)$. On considère le problème aux limites : trouver u solution de (2.1)

$$u - \Delta u = f \quad \text{dans} \quad \Omega, \quad u = 0 \quad \text{sur} \quad \partial\Omega. \tag{2.1}$$

ce qui, avec les notations de la Section 2.2, se réécrit : trouver $u \in D(-\Delta_d)$ t.q.

$$(I - \Delta_d)u = f.$$

D'après le Corollaire 2.2.2, l'opérateur $I - \Delta_d$ est inversible et le problème (2.1) admet $u := (I - \Delta_d)^{-1} f$ pour unique solution dans $D(-\Delta_d)$.

Par multiplication de (2.1) par $v \in \mathcal{D}(\Omega)$, on obtient : (2.2)

$$\underbrace{\int_\Omega (u\overline{v} + \nabla u \overline{\nabla v})dx}_{=:((u,v))} = \int_\Omega f\overline{v}dx, \quad \forall v \in \mathcal{D}(\Omega). \tag{2.2}$$

ce qui peut s'interpréter comme la résolution d'un problème de représentation d'une forme sesquilinéaire par le théorème de représentation de Riesz. Pour cela, on introduit :

$$H^1(\Omega) = \{v \in L^2(\Omega) \mid \vec{\nabla} v \in L^2(\Omega)^n\}$$

L'espace $H^1(\Omega)$ muni du produit scalaire $((\cdot,\cdot))$ est un espace de Hilbert. On remarque que la relation :

$$\int_\Omega \vec{\varphi} \cdot \vec{\nabla} u dx = -\int_\Omega \mathrm{div}(\vec{\varphi}) u dx + \int_{\partial\Omega} \underbrace{\vec{n} \cdot \vec{\varphi}}_{=:\varphi_n} u d\sigma, \quad \forall \vec{\varphi} \in \mathcal{C}^\infty(\Omega)^n$$

permet de définir les valeurs de $u \in H^1(\Omega)$ sur $\partial\Omega$, ce qui permet de définir le sous-espace :

$$H_0^1(\Omega) = \{v \in H^1(\Omega) \mid v|_{\partial\Omega} = 0\}.$$

Proposition 2.3.1. *Le sous-espace $H_0^1(\Omega)$ est fermé dans $H^1(\Omega)$.*

Démonstration. Soit $(u_k)_{k\geq 0} \in H_0^1(\Omega)^{\mathbb{N}}$ t.q. $u_k \underset{k\to+\infty}{\to} u$ dans $H^1(\Omega)$, i.e. :

$$u_k \underset{k\to+\infty}{\to} u \quad \text{et} \quad \vec{\nabla} u_k \underset{k\to+\infty}{\to} \vec{\nabla} u \quad \text{dans} \quad L^2(\Omega).$$

Alors : $\forall \vec{\varphi} \in \mathcal{C}^\infty(\Omega)^n$,

$$\int_\Omega \vec{\varphi} \cdot \vec{\nabla} u_k dx \underset{k\to+\infty}{\to} \int_\Omega \vec{\varphi} \cdot \vec{\nabla} u dx,$$

d'une part, et

$$\int_\Omega \vec{\varphi} \cdot \vec{\nabla} u_k dx = -\int_\Omega \mathrm{div}(\vec{\varphi}) u_k dx \underset{k\to+\infty}{\to} -\int_\Omega \mathrm{div}(\vec{\varphi}) u dx$$

d'autre part, d'où on déduit que :

$$\int_{\partial\Omega} \varphi_n u d\sigma = 0, \quad \forall \vec{\varphi} \in \mathcal{C}^\infty(\Omega)^n$$

i.e. $u = 0$ sur $\partial\Omega$ et $u \in H_1^0(\Omega)$. □

On admettra le résultat de densité :

Proposition 2.3.2. *Le sous-espace $H_0^1(\Omega)$ de $H^1(\Omega)$ coïncide avec l'adhérence de $\mathcal{D}(\Omega)$ pour la norme $\|\cdot\|$ de $H^1(\Omega)$:*

$$\overline{\mathcal{D}(\Omega)}^{H^1} = H_0^1(\Omega) \underset{\neq}{\subset} H^1(\Omega)$$

Proposition 2.3.3. *Le problème (2.1) admet une unique solution $u \in H_0^1(\Omega)$*

Démonstration. Compte tenu de (2.2) et de la Proposition 2.3.2, on est ramené à résoudre : trouver $u \in H_0^1(\Omega)$ t.q. :

$$\forall v \in H_0^1(\Omega), \quad ((u,v)) = \int_\Omega f\overline{v} dx. \tag{2.3}$$

On remarque que $\ell : v \mapsto \int_\Omega f\overline{v} dx$ est une forme antilinéaire continue sur $H^1(\Omega)$. D'après le Théorème de représentation de Riesz, il existe un unique $w \in H^1(\Omega)$ t.q. : $\ell(v) = ((w,v))$, $\forall v \in H^1(\Omega)$. Soit P la projection orthogonale : $H^1(\Omega) \to H_0^1(\Omega)$, bien définie d'après la Proposition 2.3.1. Alors, (2.3) se réécrit :

$$\forall v \in H^1(\Omega), \quad ((u, Pv)) = ((w, Pv)).$$

Nécessairement : $u \in H_0^1(\Omega) \Rightarrow u = Pw$. □

On généralise le résultat de la Proposition 2.3.3 dans le cadre fonctionnel abstrait de deux espaces de Hilbert H, V ($H = L^2(\Omega)$ et $V = H_0^1(\Omega)$ dans l'exemple précédent) munis de deux produits scalaires différents. On suppose que $V \subset H$ et que cette injection est continue et dense ($\overline{V}^H = H$). Alors, l'identification de H avec son dual (antidual si on considère des espaces de fonctions complexes) réalise une injection de cet espace dans V^*, ce qui donne le triplet :

$$V \subset H \simeq H^* \subset V^*.$$

Dans ce cadre mes formes sesquilinéaires continues $(u, v) \mapsto a(u, v)$ s'identifient à des opérateurs linéaires continus $A_a : V \to V^*$ suivant la formule :

$$\forall (u, v) \in V \times V, \quad A_a(u)(v) = a(u, v).$$

et par restriction de leur image à H définissent encore des opérateurs non bornés dans H dont les domaines et actions sont définis par :

$$D(A_a^H) = \{u \in V \mid A_a(u) \in H\} \quad \text{et} \quad A_a^H(u) = A_a(u), \quad \forall u \in D(A_a^H).$$

On en déduit, en particulier :

$$\forall u \in D(A_a^H), \quad \forall v \in V, \quad (A_a^H(u), v)_H = A_a(u)(v) = {}_{V^*}\langle A_a(u), v\rangle_V.$$

Le Théorème de Lax-Milgram

Dans le formalisme ci-dessus, le Théorème de Lax-Milgram généralise le Théorème de représentation de Riesz.

Théorème 2.3.4. *Soit V un espace de Hilbert et soit $(u, v) \mapsto a(u, v)$ une forme sesquilinéaire continue sur $V \times V$. On suppose qu'il existe une constante $\alpha > 0$ t.q.*

$$|a(u, u)| \geq \alpha \|u\|_V^2, \quad \forall u \in V.$$

Alors, l'opérateur $A_a : V \to V^*$, $u \mapsto A_a(u)(\cdot) = a(u, \cdot)$, réalise un isomorphisme de $V \to V^*$ et on a :

$$\|A_a^{-1}\|_{\mathcal{L}_c(V^*, V)} \leq \frac{1}{\alpha}$$

Démonstration. On commence par remarquer que A_a est linéaire (immédiat) injective. En effet, soit $u \in V$ t.q. $A_a(u) = 0$. Alors :

$$0 = |A_a(u)(u)| = |a(u, u)| \geq \alpha \|u\|_V^2 \Rightarrow \|u\| = 0$$

i.e. $u = 0$. Il reste à vérifier que $A_a(V) = V^*$. Pour cela on va montrer successivement que $A_a(V)$ est fermé et dense dans V^*. Comme V est complet, il en est de même de V^* et on est ramené à montrer que $A_a(V)$ est complet dans V^*. Soit $(u_k)_{k\geq 0} \in V^{\mathbb{N}}$. On suppose que la suite $(V_a(u_k))_{k\geq 0}$ est de Cauchy. On a : $\forall k, p \geq 0$,

$$\alpha\|u_k - u_p\|_V^2 \leq |a(u_k - u_p, u_k - u_p)| = |A_a(u_k - u_p)(u_k - u_p)| \leq$$
$$\leq \|A_a(u_k - u_p)\|\|u_k - u_p\|_V$$

donc

$$\|u_k - u_p\|_V \leq \frac{1}{\alpha}\|A_a(u_k - u_p)\|.$$

On en déduit que la suite $(u_k)_{k\geq 0}$ est de Cauchy dans V, donc convergente dans V qui est complet. Soit $u \in V$ sa limite dans V. On remarque que A_a est continue sur V par continuité de a. En effet : $\forall w, v \in V$,

$$|A_a(w)(v)| = |a(w, v)| \leq \|a\|\|w\|_W\|v\|_V$$
$$\Rightarrow \|A_a(w)\|_{V^*} = \sup_{v\in V\setminus\{0\}} \frac{|A_a(w)(v)|}{\|v\|_V} \leq \|a\|\|w\|_V$$

On en déduit que $A_a(u_k) \underset{k\to+\infty}{\to} A_a(u)$ dans V^*, i.e. que $A_a(V)$ est complet. Pour montrer que $A_a(V)$ est dense dans V^*, on remarque que :

$$A_a(V) = \overline{A_a(V)} = \overline{\mathrm{Im}(A_a)} = \mathrm{Ker}(A_a^*)^\perp \Rightarrow \overline{A_a(V)} = V^*$$
$$\iff \mathrm{Ker}(A_a^*)^\perp = V^* \iff \mathrm{Ker}(A_a^*)^{\perp\perp} = V^{*\perp} \iff \underset{=\mathrm{Ker}(A_a^*)}{\overline{\mathrm{Ker}(A_a^*)}} = \{0\}$$

Par définition de l'adjoint : $\forall u, v \in V$,

$$A_a(u)(v) = A_a^*(v)(u) = a(u, v).$$

Soit $A_a^*(v) = 0$ avec $v \in V$. Alors :

$$0 = A_a^*(v)(v) = a(v, v) \geq \alpha\|v\|_V^2 \Rightarrow \|v\|_V = 0$$

i.e. : $v = 0$, ce qui achève de montrer que $A_a(V) = V^*$, et finalement que A_a est un isomorphisme de $V \to V^*$. Soit $f \in V^*$. On a :

$$\alpha\|A_a^{-1}(f)\|_V^2 =\leq a(A_a^{-1}(f), A_a^{-1}(f)) = f(A_a^{-1}(f)) \leq \|f\|_{V^*}\|A_a^{-1}(f)\|_V$$
$$\Rightarrow \alpha\|A_a^{-1}(f)\|_V \leq \|f\|_{V^*}, \quad \forall f \in V^*.$$

Il en résulte :

$$\|A_a^{-1}\|_{\mathcal{L}_c(V^*,V)} = \sup_{f\in V^*\setminus\{0\}} \frac{\|A_a^{-1}(f)\|_V}{\|f\|_{V^*}} \leq \frac{1}{\alpha}.$$

□

On va donner différents types d'applications de ce formalisme. Il convient de garder à l'esprit que ces extensions peuvent être combinées entre elles. Pour simplifier l'exposé et parce que c'est aussi le cadre de ces applications, on se limite à des fonctiosn à valeurs réelles.

Conditions aux limites et inégalité de Pioincaré

Le formalisme ci-dessous dit formulation variationnelle a l'avantage de prendre en compte les conditions aux limites attachées à l'edp étudiée. En choisissant adéquatement V et la forme linéaire ℓ dans le problème : trouver u solution de

$$u \in V \quad \text{et} \quad ((u,v))_V = \ell(v), \quad \forall v \in V$$

on peut traiter différentes conditions aux limites. Soit $\Gamma \subset \partial\Omega$ une partie de mesure non nulle du bord $\partial\Omega$. On pose :

$$V_\Gamma := \{v \in H^1(\Omega) \mid v|_\Gamma = 0\}. \tag{2.4}$$

Proposition 2.3.5. *L'espace V_Γ défini par (2.4) est un sous-espace fermé de $H^1(\Omega)$.*

Démonstration. Par définition, V_Γ est un sous-espace de $H^1(\Omega)$.

Soit $(u_k)_{k\geq 0} \in V_\Gamma^{\mathbb{N}}$ t.q. $u_k \underset{k\to+\infty}{\to} u$ dans $H^1(\Omega)$. Soit $\forall \vec{\varphi} \in \mathcal{C}^\infty(\Omega)^n$. On a :

$$\int_\Omega \vec{\varphi}\cdot\vec{\nabla}u_k dx = -\int_\Omega \operatorname{div}(\vec{\varphi})u_k dx + \int_{\partial\Omega} \underbrace{\vec{n}\cdot\vec{\varphi}}_{=:\varphi_n} u_k d\sigma$$

$$\underset{u_k\in V_\Gamma}{=} -\int_\Omega \operatorname{div}(\vec{\varphi})u_k dx + \int_{\partial\Omega\setminus\Gamma} \varphi_n\, u_k d\sigma, \quad \forall k \geq 0,$$

d'où on déduit :

$$\int_{\partial\Omega} \varphi_n\, u d\sigma = \int_\Omega \vec{\varphi}\cdot\vec{\nabla}u dx + \int_\Omega \operatorname{div}(\vec{\varphi})u dx = \lim_{k\to+\infty}\left(\int_\Omega \vec{\varphi}\cdot\vec{\nabla}u dx + \int_\Omega \operatorname{div}(\vec{\varphi})u dx\right)$$

$$= \lim_{k\to+\infty}\int_{\partial\Omega\setminus\Gamma} \varphi_n\, u_k d\sigma.$$

En particulier, si $\operatorname{supp}(\vec{\varphi}) \subset \Gamma$ alors :

$$\int_{\partial\Omega} \varphi_n\, u d\sigma = \int_\Gamma \varphi_n\, u d\sigma = 0.$$

On en déduit que $u|_\Gamma = 0$, i.e. $u \in V_\Gamma$. □

Soit $f \in L^2(\Omega)$ et soit $g \in L^2(\partial\Omega)$. On introduit la forme linéaire continue sur V_Γ :

$$\ell : v \mapsto \int_\Omega f v dx + \int_{\partial\Omega} g v d\sigma$$

Soit $\lambda > 0$. D'après le théorème de Lax-Milgram (ou le Théorème de représentation de Riesz appliqué à V_Γ), il existe un unique $u \in V_\Gamma$ t.q. :

$$\forall v \in V_\Gamma, \quad \lambda \int_\Omega uv dx + \int_\Omega \vec{\nabla} u \cdot \vec{\nabla} v dx = \ell(v). \tag{2.5}$$

En prenant $v = \varphi \in \mathcal{D}(\Omega) \subset V_\Gamma$, et en intégrant par parties, on obtient :

$$\lambda u - \Delta u = f \quad \text{dans} \quad \mathcal{D}'(\Omega), \quad u = 0 \quad \text{sur} \quad \Gamma. \tag{2.6}$$

On suppose que u est assez régulier pour appliquer la formule de Green. Après multiplication de (2.6) par $v \in V_\Gamma$:

$$\int_\Omega f v dx = \int_\Omega (\lambda u - \Delta u) v dx = \underbrace{\lambda \int_\Omega uv dx + \int_\Omega \vec{\nabla} u \vec{\nabla} v dx}_{\underset{(2.5)}{=} \ell(v)} - \int_{\partial\Omega\setminus\Gamma} \frac{\partial u}{\partial n} v d\sigma.$$

Par comparaison avec (2.5), on en déduit que :

$$\frac{\partial u}{\partial n} = g \quad \text{sur} \quad \partial\Omega \setminus \Gamma.$$

et u est finalement solution du problème aux limites :

$$\lambda u - \Delta u = f \quad \text{dans} \quad \mathcal{D}'(\Omega), \quad u = 0 \quad \text{sur} \quad \Gamma, \quad \frac{\partial u}{\partial n} = g \quad \text{sur} \quad \partial\Omega \setminus \Gamma;$$

Résolution analytique par les séries de Fourier

On considère le problème avec conditions aux limites :

$$\begin{cases} \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} = 0 & \text{dans} \quad]0, L[\times]0, M[=: Q, \\ u(x, M) = 0 & \text{dans} \quad]0, L[\\ u(x, 0) = u_0(x) & \text{dans} \quad]0, L[\\ u(0, y) = u(L, y) = 0 & \text{dans} \quad]0, M[\end{cases} \tag{2.7}$$

Théorème 2.3.6 (Existence et unicité). *Soit* $u_0 \in \mathcal{C}^2(]0,L[,\mathbb{R}) \cap \mathcal{C}([0,L],\mathbb{R})$. *Alors, il existe une unique fonction* $u \in \mathcal{C}^2(]0,L[\times]0,M[,\mathbb{R}) \cap \mathcal{C}([0,L] \times [0,M],\mathbb{R})$ *solution de (2.7).*

Proposition 2.3.7. *On suppose que* $u_0 \in \mathcal{C}^2([0,L],\mathbb{R})$ *et que* $u_0(0) = u_0(L) = 0$. *Alors la solution du problème (2.7) se développe en série de Fourier sous la forme :*

$$u(x,t) = \sum_{n\in\mathbb{N}} c_n(u_0) \frac{\sinh(\frac{n\pi}{L}(M-y))}{\sinh(\frac{n\pi M}{L})} \sin\Big(\frac{n\pi x}{L}\Big) \tag{2.8}$$

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

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

Démonstration. On commence par chercher u sous la forme $u(x,t) = X(x)Y(y)$ où X et Y sont de classe $\mathcal{C}^2$, en accord avec le Thérème 2.3.6. Après report dans (2.7), on obtient :

$$\frac{X''}{X} = -\frac{Y''}{Y} = \lambda \in \mathbb{R}.$$

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 solution du système :

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

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 donc la suite de solutions $(u_n)_{n\in\mathbb{N}}$ définies par :

$$u_n(x,y) = \frac{\sinh(\frac{n\pi}{L}(M-y))}{\sinh(\frac{n\pi M}{L})} \sin\Big(\frac{n\pi x}{L}\Big), \quad \forall (x,t) \in Q.$$

Il reste à vérifier la condition au bord en $y = 0$, ce qu'on cherche a priori pour $u = \sum_{n\in\mathbb{N}} \alpha_n u_n$. Formellement, on trouve que :

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

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 $u = \sum_{n\in\mathbb{N}} \alpha_n u_n$ lorsque $(\alpha_n)_{n\in\mathbb{N}} \in \ell^1$. Soit $T > 0$. On a :

$$|u_n(x, y)| \leq |\alpha_n|, \quad \forall (x, t) \in \mathbb{R} \times [0, T]$$

et la série majorante est convergente par hypothèse sur u_0. On en déduit que la série de fonctions continues $\sum \alpha_n u_n$ est uniformément convergente sur tout compact de $[0, L] \times \mathbb{R}^+$ donc de somme continue sur $[0, L] \times \mathbb{R}^+$. Soit $\delta \in]0, M[$. On a : $\forall (x, y) \in [0, L] \times [\delta, M]$,

$$|u_n(x, y)| \leq C \frac{\sinh(\frac{n\pi}{L}(M - y))}{\sinh(\frac{n\pi M}{L})} \leq C \frac{\sinh(\frac{n\pi}{L}(M - \delta))}{\sinh(\frac{n\pi M}{L})} \underset{n\to+\infty}{\sim} C e^{-\frac{n\pi}{L}\delta}$$

ainsi que : $\forall k, \ell > 0$,

$$\left| \frac{\partial^{k+\ell}}{\partial x^k \partial y^\ell} u_n(x, y) \right| \leq C \left(\frac{n\pi}{L} \right)^{k+\ell} e^{-\frac{n\pi}{L}\delta}, \quad \forall (x, y) \in [0, L] \times [\delta, M]$$

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 $[0, L]\times]0, M]$, donc que la somme de la série $\sum \alpha_n u_n$ est de classe $\mathcal{C}^\infty$ sur $[0, L]\times]0, M]$. On conclut par unicité de la série de Fourier de u_0. □

Proposition 2.3.8. *Si $u_0 \in \mathcal{C}^2([0, L], \mathbb{R})$, alors (2.8) est l'unique solution de (2.7).*

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

Le principe du maximum

Définition 2.3.1 (Principe du maximum)**.** On appelle principe du maximum continu le fait que si $f \geq 0$ alors le minimum de la solution u du problème (2.24) est atteint sur le bord du domaine de définition de u.

Proposition 2.3.9 (Le principe du maximum)**.** *Soit $\Omega \subset \mathbb{R}^n$ un ouvert borné et soit $f \in \mathcal{C}^2(\Omega)$.*

a) Si $\Delta f > 0$ dans Ω alors :

$$\forall x \in \Omega, \quad f(x) < \max_{y \in \partial\Omega} f(y).$$

b) Si $\Delta f = 0$ dans Ω (f est dite harmonique), alors

$$\forall x \in \Omega, \quad \min_{y \in \partial\Omega} f(y) \leq f(x) \leq \max_{y \in \partial\Omega} f(y).$$

Démonstration. On remarque que $\partial\Omega = \overline{\Omega} \cap \Omega^c$ est fermé comme intersection de fermés, et borné par hypothèse sur Ω, donc c'est un compact de $\mathbb{R}^n$ et f y atteint ses bornes.

a) Soit $\Delta f > 0$ dans Ω. On suppose qu'il existe $x_0 \in \Omega$ t.q. :

$$f(x_0) \geq \max_{y \in \partial\Omega} f(y).$$

Comme $\overline{\Omega}$ est un compact de $\mathbb{R}^n$ (en dimension finie), on peut supposer que

$$f(x_0) = \max_{y \in \overline{\Omega}} f(x) \geq \max_{y \in \partial\Omega} f(x)$$

Comme Ω est ouvert, on a $\nabla f(x_0) = 0$. Soit $r > 0$ t.q., si $B_r(x_0)$ est la boule ouverte de centre x_0 et de rayon $r > 0$, on ait :$B_r(x_0) \subset \Omega$, et soit $x \in B_r(x_0) \setminus \{x_0\}$. On pose :

$$\varphi(t) = f(x_0 + t(x - x_0)), \quad \forall t \in [0, 1].$$

Alors $f \in \mathcal{C}^2(\Omega, \mathbb{R}) \Rightarrow \varphi \in \mathcal{C}^2([0, 1], \mathbb{R})$ et on a

$$\varphi(1) = f(x) = f(x_0) + \int_0^1 (1-t) f''(x_0 + t(x - x_0)) \cdot (x - x_0)^2 dt \leq f(x_0)$$

$$\Rightarrow \int_0^1 (1 - t) f''(x_0 + t(x - x_0)) \cdot (x - x_0)^2 dt \leq 0.$$

Soit $u = \frac{1}{\|x - x_0\|}(x - x_0)$. On a, par bilinéarité de la différentielle :

$$\int_0^1 (1 - t) f''(x_0 + t(x - x_0)) \cdot u^2 dt \leq 0.$$

On en déduit, par continuité de f'' :

$$\lim_{x \to x_0} \int_0^1 (1-t) f''(x_0 + t(x - x_0)) \cdot u^2 dt = \int_0^1 (1-t) dt f''(x_0) \cdot u^2 = \frac{1}{2} f''(x_0) \cdot u^2 \leq 0.$$

Ceci étant vrai pour tout $x \in B_r(x_0)$, on en déduit, par définition des dérivées partielles d'ordre 2 :

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \leq 0, \quad i,j = 1, \cdots, n.$$

en contradiction avec $\Delta f > 0$ dans Ω et $x_0 \in \Omega$.

b) Soit $\Delta f = 0$ dans Ω. Il suffit de montrer que $f \leq \max_{y\in\partial\Omega} f$ dans Ω, l'autre inégalité étant alors obtenue en considérant $-f$. Soit $\varepsilon > 0$. On pose :

$$g_\varepsilon(x) = f(x) + \varepsilon\|x\|^2 \quad \forall x \in \Omega.$$

Alors :

$$\Delta g_\varepsilon = \Delta f + 2n\varepsilon = 2n\varepsilon > 0 \quad \text{dans} \quad \Omega.$$

On remarque que $\partial\Omega$ étant borné, il existe $M > 0$ t.q. $\|x\| \leq M$, $\forall x \in \partial\Omega$. De a), on déduit que : $\forall x \in \Omega$,

$$g_\varepsilon(x) < \max_{y\in\partial\Omega} g \leq \max_{y\in\partial\Omega} f + \varepsilon M$$

et donc : $\forall x \in \Omega$,

$$f(x) \leq g_\varepsilon(x) < \max_{y\in\partial\Omega} f + \varepsilon M.$$

On en déduit :

$$\sup_{x\in\Omega} f(x) \leq \max_{y\in\partial\Omega} f + \varepsilon M.$$

Ceci étant vrai pour tout $\varepsilon > 0$, il en résulte que :

$$\sup_{x\in\Omega} f(x) \leq \max_{y\in\partial\Omega} f.$$

□

Proposition 2.3.10 (Principe du maximum faible)**.** *Soit $\Omega \subset \mathbb{R}^N$ un ouvert borné de $\mathbb{R}^N$ et soit $u \in H^1(\Omega)$ solution de :*

$$\begin{cases} \Delta u \geq 0 & \text{p.p. dans } \Omega, \\ u^+ \in H_0^1(\Omega). \end{cases}$$

Alors : $u \leq 0$ p.p. dans Ω.

Lemme 2.3.11 (Lemme préliminaire)**.** *Soit $f \in H^{-1}(\Omega) = (H_0^1(\Omega))'$ t.q. $f \geq 0$ p.p. dans $\mathcal{D}'(\Omega)$, i.e. :*

$$_{H^{-1}}\langle f, \varphi\rangle_{H_0^1} \geq 0, \quad \forall \varphi \in \mathcal{D}(\Omega, \mathbb{R}^+).$$

Alors

$$\forall v \in H_0^1(\Omega), \quad _{H^{-1}}\langle f, v^+\rangle_{H_0^1} \geq 0.$$

Démonstration. Soit $v \in H_0^1(\Omega)$ et soit $(\varphi_n)_{n\geq 0} \in \mathcal{D}(\Omega)^{\mathbb{N}}$ t.q. $\varphi_n \underset{n\to+\infty}{\to} v$ dans $H_0^1(\Omega)$. On a :

$$\|\nabla v^+ - \nabla \varphi_n^+\|_2 \leq \|\nabla v - \nabla \varphi_n\|_2 \underset{n\to+\infty}{\to} 0$$

i.e. $\varphi_n^+ \underset{n\to+\infty}{\to} v^+$ dans $H_0^1(\Omega)$. Soit $n \geq 0$. On remarque que φ_n^+ est à support compact car fermé dans le support de φ_n qui est fermé. Donc il existe une suite régularisante $(\rho_\delta)_{\delta>0}$ t.q. $\rho_\delta \star \varphi_n^+ \in \mathcal{D}(\Omega)$ dès que $\delta \in]0, \delta_n'[$ est assez petit, i.e. inférieur à la distance du support de φ_n^+ au bord $\partial\Omega$, et $\rho_\delta \star \varphi_n^+ \underset{\delta\to 0}{\to} \varphi_n^+$ dans $H_0^1(\Omega)$. Par définition de la convolution : $\rho_\delta \star \varphi_n^+ \geq 0$ dans Ω, $\forall \delta \in]0, \delta_n'[$, $\forall n \geq 0$. Par hypothèse sur f :

$$_{H^{-1}}\langle f, \rho_\delta \star \varphi_n^+ \rangle_{H_0^1} \geq 0, \quad \forall \delta \in]0, \delta_n'[.$$

On conclut après extraction d'une suite diagonale $(\rho_{\delta_n})_{n\geq 0}$ t.q. $\rho_{\delta_n} \star \varphi_n^+ \underset{n\to+\infty}{\to} v^+$ dans $H_0^1(\Omega)$:

$$_{H^{-1}}\langle f, v^+ \rangle_{H_0^1} = \lim_{n\to+\infty} \underbrace{_{H^{-1}}\langle f, \rho_{\delta_n} \star \varphi_n^+ \rangle_{H_0^1}}_{\geq 0} \geq 0.$$

□

Démonstration de la Proposition 2.3.10. Soit $f = -\Delta u$. Alors $f \in H^{-1}(\Omega)$ et $f \leq 0$ p.p. dans Ω. Par hypothèse sur u^+ :

$$_{H^{-1}}\langle f, u^+ \rangle_{H_0^1} = \int_\Omega \nabla u \nabla u^+ dx = \int_\Omega |\nabla u^+|^2 dx \geq 0 \Rightarrow \nabla u^+ = 0 \quad \text{p.p. dans } \Omega$$

i.e. $u^+ = 0$ p.p. dans Ω.

□

Remarque 9. On a utilisé le fait que $H_0^1(\Omega)$ est un espace de Hilbert pour le produit scalaire $(u, v) \mapsto \int_\Omega \nabla u \nabla v dx$.

Corollaire 2.3.12. *Soit $\Omega \subset \mathbb{R}^N$ un ouvert borné et soit $g \in H^1(\Omega)$. Si $u \in H^1(\Omega)$ est solution de :*

$$\begin{cases} \Delta u = 0 \quad \textit{dans } \Omega, \\ u - g \in H_0^1(\Omega), \end{cases}$$

alors $|u| \leq \|g\|_\infty$ p.p. dans Ω.

Démonstration. Si $g \notin L^\infty(\Omega)$, il n'y a rien à montrer. On suppose que $g \in L^\infty(\Omega)$. On remarque que $(u - \|g\|_\infty)^+ \in H_0^1(\Omega)$. D'après la Proposition 2.3.10 appliquée à $u - \|g\|_\infty$ avec $f = 0$, on a : $u \leq \|g\|_\infty$ p.p. dans Ω. De même, $(u + \|g\|_\infty)^- \in H_0^1(\Omega)$. D'après la Proposition 2.3.10 appliquée à $u + \|g\|_\infty$ avec $f = 0$, on a : $u \geq -\|g\|_\infty$ p.p. dans Ω. □

Proposition 2.3.13. *Soit $\Omega \subset \mathbb{R}^N$ un ouvert borné de $\mathbb{R}^N$ et soit $f, c \in L^\infty(\Omega)$. On suppose que $c \geq \eta > 0$ p.p. dans Ω pour une constante $\eta > 0$. Soit $u \in H_0^1(\Omega)$ solution de :*

$$-\Delta u + cu = f \quad dans \quad \mathcal{D}'(\Omega)$$

Alors

$$\|u\|_\infty \leq \frac{\|f\|_\infty}{\eta}.$$

Démonstration. Soit $k \in \mathbb{R}^+$. On remarque que $(u-k)^+ \in H_0^1(\Omega$. On en déduit :

$$\int_\Omega \nabla u \nabla (u-k)^+ dx + \int_\Omega cu(u-k)^+ dx = \int_\Omega f(u-k)^+ dx.$$

avec

$$\int_\Omega \nabla u \nabla (u-k)^+ dx = \int_\Omega \nabla (u-k) \nabla (u-k)^+ dx = \int_\Omega \nabla (u-k)^+ \nabla (u-k)^+ dx =$$

$$= \int_\Omega |\nabla (u-k)^+|^2 dx \geq 0,$$

$$\int_\Omega cu(u-k)^+ dx = \int_\Omega c(u-k)(u-k)^+ dx + k \int_\Omega c(u-k)^+ dx =$$

$$= \int_\Omega c|(u-k)^+|^2 dx + k \int_\Omega c(u-k)^+ dx$$

$$\geq \eta \int_\Omega |(u-k)^+|^2 dx + k \int_\Omega c(u-k)^+ dx.$$

On en déduit :

$$0 \leq \int_\Omega |\nabla (u-k)^+|^2 dx + \eta \int_\Omega |(u-k)^+|^2 dx \leq \int_\Omega (f - ck)(u-k)^+ dx.$$

On pose :

$$k = \frac{\|f\|_\infty}{\eta}.$$

Alors :

$$ck \geq \|f\|_\infty \Rightarrow f - ck \leq 0 \quad \text{p.p. dans } \Omega \Rightarrow \int_\Omega (f - ck)(u-k)^+ dx \leq 0$$

Il en résulte :

$$\eta \int_\Omega |(u-k)^+|^2 dx \leq 0$$

i.e. : $(u-k)^+ = 0$ p.p. Le même raisonnement ave $(u+k)^-$ cpnduit à $u \geq -\frac{\|f\|_\infty}{\eta}$ p.p. dans Ω. □

2.4 Calcul approché par les différences finies en dimension 1

Soit $f \in \mathcal{C}([0,1])$. On cherche $u : [0,1] \to \mathbb{R}$ solution de :

$$-u''(x) = f(x), \quad x \in]0,1[, \quad u(0) = u(1) = 0. \tag{2.9}$$

Cette équation modélise par exemple la diffusion de la chaleur dans un barreau conducteur chauffé (terme source f) dont les deux extrémités snt plongées dans de la glace. On se donne une subdivision de $[0,1]$:

$$x_0 = 0 < x_1 < \cdots < x_N < x_{N+1} = 1$$

supposée régulière de pas $h = \dfrac{1}{N+1} > 0$ pour simplifier, et on veut calculer des approximations $u_i \sim u(x_i)$, $i = 0, \cdots, N+1$, en chaque point x_i de la subdivision. Pour cela, on considère le problème : trouver $u^{(N)} \in \mathbb{R}^N$, de composantes $u_1, \cdots, u_N$, solution du système :

$$\begin{cases} \dfrac{1}{h^2}(-u_{i-1} + 2u_i - u_{i+1}) = f_i, & i = 1, \cdots, N, \\ \\ u_0 = u_{N+1} = 0 \end{cases} \tag{2.10}$$

où $f_i = f(x_i)$, $i = 1, \cdots, N$, sont donnés au second membre.

Définition 2.4.1 (Erreur de consistance)**.** On pose :

$$\varepsilon_i^{(N)} = \frac{1}{h^2}(-u(x_{i-1}) + 2u(x_i) - u(x_{i+1})) - f(x_i), \quad i = 1, \cdots, N$$

On appelle erreur de consistance du schéma (2.10) la quantité

$$\|\varepsilon^{(N)}\|_\infty = \max_{1 \leq i \leq N} |\varepsilon_i^{(N)}|, \quad N > 0.$$

Le schéma (2.10) est dit consistant si $\lim_{N \to +\infty} \|\varepsilon^{(N)}\|_\infty = 0$.

Par application de la formule de Taylor, on trouve :

$$\varepsilon^{(N)} = O(h) = O\left(\frac{1}{N+1}\right) \underset{N\to+\infty}{\to} 0,$$

i.e. le schéma (2.10) est consistant.

Remarque 10. Si $u \in \mathcal{C}^4(]0,1[)$ alors il existe une constante $C > 0$ indépendante de u t.q. :

$$\|\varepsilon^{(N)}\|_\infty \leq C\|u^{(4)}\|h^2.$$

Définition 2.4.2. Le schéma (2.10) est dit convergent si

$$\lim_{N\to+\infty} \max_{1\leq i\leq N} |u_i^{(N)} - u(x_i)| = 0.$$

Proposition 2.4.1. *Le schéma (2.10) est convergent.*

Démonstration. Soit $(\mu^{(N)})_{1\leq i\leq N} \in \mathbb{R}^N$ et soit $z^{(N)} \in \mathbb{R}^N$ solution du schéma :

$$\begin{cases} \dfrac{1}{h^2}(-z_{i-1} + 2z_i - z_{i+1}) = \mu_i^{(N)}, \qquad i = 1, \cdots, N, \\ \\ z_0 = z_{N+1} = 0 \end{cases} \tag{2.11}$$

qui se réécrit sous forme matricielle :

$$\frac{1}{h^2} A_N z^{(N)} = \mu^{(N)} \tag{2.12}$$

où $A_N \in \mathbb{R}^{N\times N}$ est la matrice carrée d'ordre N de coefficients :

$$a_{ij}^{(N)} = \begin{cases} 2 & \text{si} \quad i = j, \\ -1 & \text{si} \quad |i-j| = 1, \\ 0 & \text{si} \quad |i-j| > 1. \end{cases} \tag{2.13}$$

On remarque que A_N est symétrique réelle, donc diagonalisable dans une bon, de valeurs propres dans $\mathbb{R}$. Soit $x \in \mathbb{R}^N$ un vecteur propre de A_N de valeur propre $\lambda \in \mathbb{R}$. L'équation $A_N x = \lambda x$ s'écrit composante par composante :

$$-x_{i-1} + (2-\lambda)x_i - x_{i+1} = 0, \quad i = 1, \cdots, N \tag{2.14}$$

$$x_0 = x_{N+1} = 0. \tag{2.15}$$

L'équation caractéristique associée s'écrit :

$$r^2 + (\lambda - 2)r + 1 = 0. \tag{2.16}$$

On vérifie que (2.14)–(2.15) admet des solutions non nulles ssi le discriminant de (2.16) est ≤ 0, i.e. ssi $|\lambda - 2| < 2$. On pose : $\lambda = 2 + 2\cos\theta$, $\theta \in]0, \pi[$. Alors (2.16) admet les racines

$$r^{\pm} = -e^{\pm i\theta}.$$

On en déduit

$$x_k = a\cos(k\theta) + b\sin(k\theta), \quad 0 \leq k \leq N+1$$

avec $x_0 = a = 0$. On en déduit :

$$x_{N+1} = b\sin((N+1)\theta) = 0 \iff \theta \in \frac{\pi}{N+1}\mathbb{N}.$$

Plus précisément :

$$0 < \theta < \pi \Rightarrow \theta = \frac{p\pi}{(N+1)} = \theta_p, \quad 1 \leq p \leq N$$

On en déduit que A_N admet N valeurs propres distinctes > 0 :

$$\begin{aligned} 4 > \lambda_1^{(N)} = 4\left(\cos\left(\frac{\pi}{N+1}\right)\right)^2 > \cdots > \lambda_p^{(N)} = 4\left(\cos\left(\frac{p\pi}{N+1}\right)\right)^2 > \\ > \cdots > \lambda_N^{(N)} = 4\left(\cos\left(\frac{N\pi}{N+1}\right)\right)^2 = 4\left(\sin\left(\frac{\pi}{N+1}\right)\right)^2 > 0. \end{aligned} \tag{2.17}$$

En particulier A_N est inversible et donc

$$z^{(N)} = h^2 A_N^{-1}\mu^{(N)} \Rightarrow \|z^{(N)}\|_2 \leq h^2 \underbrace{\|A_N^{-1}\|_2}_{=\rho(A_N^{-1})} \|\mu^{(N)}\|_2$$

où $\|A_N^{-1}\|_2 = \rho(A_N^{-1})$ par symétrie de A_N^{-1}. Les valeurs propres de A_N^{-1} sont :

$$0 < \frac{1}{\lambda_1^{(N)}} < \cdots \frac{1}{\lambda_N^{(N)}} \Rightarrow \rho(A_N^{-1}) = \frac{1}{\lambda_N^{(N)}} = \frac{1}{4\left(\sin\left(\frac{\pi}{N+1}\right)\right)^2} = \frac{1}{4(\sin(\pi h))^2}$$

et donc :

$$\|z^{(N)}\|_2 \leq \left(\frac{h}{2\sin(\pi h)}\right)^2 \|\mu^{(N)}\|_2. \tag{2.18}$$

Soit $X^{(N)} \in \mathbb{R}^N$ le vecteur de composantes $X_i^{(N)} = x_i$, $1 \leq i \leq N$. Si $z_i^{(N)} = u(x_i) - u_i$, $1 \leq i \leq N$, alors $z^{(N)} = u(X^{(N)}) - u^{(N)}$ et $\mu^{(N)} = \varepsilon^{(N)}$ avec

$$\|\varepsilon^{(N)}\|_2 \leq \sqrt{N}\|\varepsilon^{(N)}\|_\infty \leq C\sqrt{N}h \leq C\sqrt{h}$$

donc

$$\|u(X^{(N)}) - u^{(N)}\|_\infty \leq \|u(X^{(N)}) - u^{(N)}\|_2 \leq C\sqrt{h}\left(\frac{h}{2\sin(\pi h)}\right)^2$$

$$\leq C\sqrt{h} \leq \frac{C}{\sqrt{N}} \underset{N\to+\infty}{\to} 0$$

□

Autres conditions aux limites

Conditions de Dirichlet non homogènes

On considère les conditions sur le bord non homogènes :

$$u(0) = a, \quad u(1) = b.$$

Le schéma (2.10) reste inchangé à l'exception du second membre qui devient :

$$f_i = \begin{cases} f(x_1) + a & \text{si} \quad i = 1, \\ f(x_N) + b & \text{si} \quad i = N, \\ f(x_i) & \text{si} \quad 1 < i < N, \end{cases} \tag{2.19}$$

Conditions de Neumann et de Fourier

On considère les conditions sur le bord :

(i) de type Neumann en $x = 0$: $u'(0) = a$,

(ii) de type Fourier en $x = 1$: $u'(1) + \alpha u(1) = b$ avec $\alpha > 0$.

Les premiers termes des développements en série de Taylor de $u'(0)$ et $u'(1)$ suggèrent de choisir :

$$u_0 = u_1 - ah, \quad u_{N+1} = \frac{u_N + bh}{1 + \alpha h}.$$

Alors le schéma (2.10) devient : trouver $u_0, u_1, \cdots, u_N, u_{N+1}$ t.q.

$$\begin{cases} \dfrac{1}{h}(u_0 - u_1) = -a, \\ \dfrac{1}{h^2}(-u_{i-1} + 2u_i - u_{i+1}) = f_i, \qquad i = 1, \cdots, N, \\ \dfrac{1}{h}(-u_N + (1+\alpha h)u_{N+1}) = b. \end{cases} \tag{2.20}$$

Schéma numérique bien posé

Définition 2.4.3. Le schéma numérique (2.10) est dit bien posé s'il admet une unique solution.

Le schéma (2.10) se réécrit sous forme matricielle : trouver $U_N \in \mathbb{R}^N$ solution de :

$$A_N U_N = f \tag{2.21}$$

où $A_N \in \mathbb{R}^{N\times N}$ est la matrice carrée introduite dans (2.12) et définie par (2.13).

Proposition 2.4.2. *La matrice (2.13) du système (2.21) est définie positive.*

Démonstration. Soit $x \in \mathbb{R}^N$. On a :

$$Ax \cdot x = 2\sum_{i=1}^{N} x_i^2 - 2\sum_{i=1}^{N-1} x_i x_{i+1} = \sum_{i=1}^{N-1}(x_i - x_{i+1})^2 + x_1^2 + x_N^2 \geq 0.$$

Si $Ax \cdot x = 0$, alors

$$x_1 = x_N = 0 \quad \text{et} \quad x_i = x_{i+1}, \quad i = 1 \cdots N \Rightarrow x = 0.$$

□

Corollaire 2.4.3. *La matrice (2.13) du système (2.21) est inversible.*

Démonstration. C'est une conséquence directe de la Proposition 2.4.2. □

On en déduit que le schéma numérique (2.10) est bien posé.

Exemple 1. Le problème (2.20) se réécrit sous forme matricielle : trouver $U_h \in \mathbb{R}^{N+2}$. solution de

$$\frac{1}{h^2} A_h U_h = f_h$$

où $A_h = (a_{ij}^{(h)})_{0 \leq i,j \leq N+1}$ est la matrice carrée d'ordre $N+2$ définie par :

$$a_{ij}^{(h)} = \begin{cases} 1 & \text{si} \quad i = j = 0, \\ -1 & \text{si} \quad (i,j) \in \{(1,0),(N+1,N)\} \\ 2 & \text{si} \quad 1 \leq i = j \leq N \\ -1 & \text{si} \quad |i-j| = 1,\ 1 \leq i,j \leq N \\ 1+\alpha h & \text{si} \quad i = j = N+1, \\ 0 & \text{si} \quad |i-j| > 1, \end{cases}$$

et où le second membre $f_h = (f_i^{(h)})_{0 \leq i \leq N+1} \in \mathbb{R}^{N+2}$ est défini par :

$$f_i^{(h)} = \begin{cases} -\dfrac{a}{h} & \text{si} \quad i = 0, \\ f(x_i) & \text{si} \quad 1 \leq i \leq N, \\ \dfrac{b}{h} & \text{si} \quad i = N+1. \end{cases}$$

Soit $x \in \mathbb{R}^{N+2}$. On a :

$$A_h x \cdot x = \sum_{i=0}^{N} (x_i - x_{i+1})^2 + \alpha h x_{N+1}^2 \geq 0$$

car $\alpha h > 0$. Si $Ax \cdot x = 0$, alors :

$$x_0 = x_1 = \cdots = x_{N+1} = 0$$

i.e. $x = 0$ et A_h est définie positive, donc inversible.

Exemple 2. On considère le problème aux limites avec conditions de Neumann :

$$-u'' = f \quad \text{dans} \quad]0,1[, \quad u'(0) = u'(1) = 0. \tag{2.22}$$

Les approximations

$$u'(0) \sim \frac{1}{h}(u_1 - u_0), \quad u'(1) \sim \frac{1}{h}(u_{N+1} - u_N)$$

conduisent au schéma numérique :

$$\begin{cases} \dfrac{1}{h^2}(-u_{i-1}+2u_i-u_{i+1})=f_i, \qquad i=1,\cdots,N, \\ \\ u_0=u_1, \\ u_{N+1}=u_N. \end{cases} \tag{2.23}$$

Sous forme matricielle, le schéma (2.23) se réécrit : trouver $U_h \in \mathbb{R}^{N+2}$ solution de

$$\frac{1}{h^2}A_hU_h=f_h$$

où $A_h=(a_{ij}^{(h)})_{0\le i,j\le N+1}$ est la matrice carrée d'ordre $N+2$ définie par :

$$a_{ij}^{(h)}=\begin{cases} 1 & \text{si} \quad i=j=0, \\ 2 & \text{si} \quad 1\le i=j\le N, \\ -1 & \text{si} \quad |i-j|=1, \\ 1 & \text{si} \quad i=j=N+1, \\ 0 & \text{si} \quad |i-j|>1 \end{cases}$$

et où le second membre $f_h=(f_i^{(h)})_{0\le i\le N+1}\in\mathbb{R}^{N+2}$ est le vecteur de composantes :

$$f_i^{(h)}=\begin{cases} 0 & \text{si} \quad i=0, \\ f(x_i) & \text{si} \quad 1\le i\le N, \\ 0 & \text{si} \quad i=N+1. \end{cases}$$

Soit $x\in\mathbb{R}^{N+2}$. On a :

$$A_hx\cdot x=\sum_{i=0}^{N}(x_i-x_{i+1})^2\ge 0$$

et

$$Ax\cdot x=0\Rightarrow x_0=\cdots=x_{N+1}.$$

On vérifie directement que

$$A_h\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}=0$$

i.e. A_h est semi-définitive positive et non inversible. De fait, le problème (2.22) admet les constantes pour solutions.

Définition 2.4.4 (Matrice monotone)**.** Une matrice réelle est dite monotone si elle est inversible, d'inverse à coefficients ≥ 0.

Proposition 2.4.4 (Caractérisation des matrices monotones)**.** *Une matrice réelle $A \in \mathbb{R}^{N\times N}$ est monotone ssi : $\forall v \in \mathbb{R}^N$,*

$$Av \geq 0 \Rightarrow v \geq 0.$$

(Les inégalités s'entendent composante par composante)

Démonstration. Soit $A \in \mathbb{R}^{N\times}$. $\Rightarrow$ Soit $A \in \mathbb{R}^{N\times N}$ monotone et soit $v \in \mathbb{R}^N$. On suppose que $Av \geq 0$. Alors :

$$v_i = \underbrace{A^{-1}_{ik}}_{\geq 0}\underbrace{(Av)_k}_{\geq 0} \geq 0, \quad \forall i \in [[1, N]].$$

$\Leftarrow$ Inversement, on suppose que : $\forall v \in \mathbb{R}^N$,

$$Av \geq 0 \Rightarrow v \geq 0.$$

Soit $v \in \mathbb{R}^N$ t.q. $Av = 0$. Alors :

$$Av = 0 \Rightarrow v \geq 0 \quad \text{et} \quad -Av = 0 \Rightarrow -v \geq 0$$

donc $v = 0$, i.e. A est inversible. Soit $i \in [[1, N]]$. On note e_i le $i$ème vecteur de la base canonique de $\mathbb{R}^N$. Par hypothèse sur A :

$$e_i = A(A^{-1}e_i) \geq 0 \Rightarrow A^{-1}e_i \geq 0.$$

On en déduit : $(A^{-1}e_i)_j = A^{-1}_{ij} \geq 0$, $\forall j \in [[1, N]]$. Ceci étant vrai pour tout $i \in [[1, N]]$, il en résulte : $A^{-1} \geq 0$, i.e. A est monotone. □

Principe du maximum discret

On considère le problème : trouver u solution de

$$\begin{cases} -u''(x) + c(x)u(x) = f(x) \quad \text{si} \quad 0 < x < 1, \\ u(0) = 0, \\ u(1) = 0, \end{cases} \tag{2.24}$$

où $c \in \mathcal{C}([0, 1], \mathbb{R}^+)$ et $f \in \mathcal{C}([0, 1], \mathbb{R})$. L'analogue du problème discrétisé (2.10) s'écrit :

$$\begin{cases} \dfrac{1}{h^2}(-u_{i-1} + 2u_i - u_{i+1}) + c_i u_i = f_i, \qquad i = 1, \cdots, N, \\ \\ u_0 = u_{N+1} = 0 \end{cases} \tag{2.25}$$

où $c_i = c(x_i)$, $f_i = f(x_i)$, $i = 1, \cdots, N$, sont donnés. Sous forme matricielle, le problème (2.25) se réécrit :

$$\frac{1}{h^2} A_h U_h = f_h$$

où $A_h = (a_{ij}^{(h)})_{1 \leq i,j \leq N} \in \mathbb{R}^{N \times N}$ est la matrice carrée d'ordre N définie par :

$$a_{ij}^{(h)} = \begin{cases} 2 + c_i h^2 & \text{si} \quad i = j, \\ -1 & \text{si} \quad |i - j| = 1, \\ 0 & \text{si} \quad |i - j| > 1, \end{cases} \tag{2.26}$$

Proposition 2.4.5. *Soit $c = (c_1, \cdots, c_N)^T \in \mathbb{R}^N$ t.q. $c_i \geq 0$, $\forall i \in [[1, N]]$. Alors la matrice A_h définie par (2.26) est symétrique, définie positive et donc inversible.*

Démonstration. La matrice A_h est symétrique par définition. Soit $x \in \mathbb{R}^N$. On a :

$$A_h x \cdot x = \sum_{i=1}^{N-1} (x_i - x_{i+1})^2 + x_1^2 + x_N^2 + h^2 \sum_{i=1}^{N} c_i x_i^2 \underset{c \geq 0}{\geq} 0$$

On suppose que $A_h x \cdot x = 0$. Alors :

$$x_1 = x_N = 0 \quad \text{et} \quad x_i = x_{i+1}, \quad i = 1, \cdots, N \Rightarrow x_1 = \cdots = x_N = 0.$$

□

Remarque 11. Si la solution u de (2.24) vérifie le principe du maximum, on souhaite qu'il en soit de même pour la solution approchée.

Lemme 2.4.6. *Soit $c = (c_1, \cdots, c_N)^T \in \mathbb{R}^N$ t.q. $c_i \geq 0$, $\forall i \in [[1, N]]$. Alors la matrice A_h définie par (2.26) est monotone.*

Démonstration. On commence par remarquer que la matrice A_h est inversible d'après la Proposition 2.4.5. Soit $v \in \mathbb{R}^N$ t.q. $A_h v \geq 0$. Soit $i_0 \in [[1, N]]$ t.q. $v_{i_0} = \min_{1 \leq j \leq N} v_j$. Si $i_0 = 1$, alors :

$$(2 + h^2 c_1) v_1 \geq v_2 \geq v_1 \Rightarrow \underbrace{(1 + h^2 c_1)}_{>0} v_1 \geq 0 \Rightarrow v_1 \geq 0.$$

Si $i_0 = N$, alors :

$$(2 + h^2 c_N) v_N \geq v_{N-1} \geq v_N \Rightarrow \underbrace{(1 + h^2 c_N)}_{>0} v_N \geq 0 \Rightarrow v_N \geq 0.$$

On suppose que $1 < i_0 < N$. On a :

$$(2 + h^2 c_{i_0}) v_{i_0} \geq v_{i_0-1} + v_{i_0+1} \geq 2v_{i_0} \Rightarrow h^2 c_{i_0} v_{i_0} \geq 0.$$

Si $c_{i_0} > 0$, alors $v_{i_0} \geq 0$. Sinon, si $v_{i_0} = \min_{1 \leq j \leq N} v_j \Rightarrow c_{i_0} = 0$, alors

$$v_{i_0} - v_{i_0-1} \geq v_{i_0+1} - v_{i_0} \geq 0 \Rightarrow v_{i_0} = v_{i_0-1} = \min_j v_j.$$

Par récurrence sur $j \in [[1, i_0]]$, on en déduit :

$$v_{i_0} = v_{i_0-1} = \cdots = v_1 = \min_j v_j \geq 0.$$

□

Définition 2.4.5 (Ordre du schéma)**.** On dit que le schéma (2.25), resp. (2.10), est d'ordre p s'il existe $C > 0$ t.q.

$$\|\varepsilon^{(N)}\|_\infty \leq C h^p$$

où $\varepsilon^{(N)} = (\varepsilon_i^{(N)})_{1 \leq i \leq N}$ est l'erreur de consistance du schéma (2.25) resp. (2.10), définie par : $\forall i \in [[1, N]]$,

$$\varepsilon_i^{(N)} = \frac{1}{h^2}(-u(x_{i-1} + 2u(x_i) - u(x_{i+1})) + c(x_i)u(x_i) - f(x_i). \tag{2.27}$$

resp. par la Définition 1.3.2.

Proposition 2.4.7. *Si la solution u de (2.24), resp. de (2.9), est de classe $\mathcal{C}^4$, alors :*

$$\|\varepsilon^{(N)}\|_\infty \leq \frac{h^2}{12} \sup_{0 \leq x \leq 1} |u^{(4)}|. \tag{2.28}$$

Démonstration. D'après la formule de Taylor avec reste intégral : $\forall x \in [0, 1]$,

$$u(x+h)+u(x-h)-2u(x) = hu'(x)+h^2u''(x)+\int_x^{x+h} \frac{(x+h-t)^3}{6} u^{(4)}(t)dt+$$

$$+\int_x^{x-h} \frac{(x-h-t)^3}{6} u^{(4)}(t)dt$$

$$\Rightarrow \left| \frac{1}{h^2}(-u(x-h) + 2u(x) - u(x-h)) + c(x)u(x) - f(x) \right| =$$

$$= \frac{1}{h^2} \left| \int_x^{x+h} \frac{(x+h-t)^3}{6} u^{(4)}(t)dt + \int_x^{x-h} \frac{(x-h-t)^3}{6} u^{(4)}(t)dt \right|$$

$$\leq \frac{2}{h^2} \left(\frac{h^4}{24} \|u^{(4)}\|_\infty \right) = \frac{h^2}{12} \|u^{(4)}\|_\infty.$$

□

Définition 2.4.6 (Stabilité)**.** Le schéma (2.25), resp. (2.10), est stable si, pour tout $\mu \in \mathbb{R}^N$, la solution $z^{(h)}$ du schéma :

$$\frac{1}{h^2} A_h z^{(h)} = \mu \tag{2.29}$$

où $A_h \in \mathbb{R}^{N\times N}$ est la matrice carrée définie par (2.26), resp. par (2.13), vérifie :

$$\|z^{(h)}\|_\infty \leq C\|\mu\|_\infty,$$

pour une constante $C > 0$ indépendante de $h > 0$.

Proposition 2.4.8. *Le schéma (2.25), resp. (2.10), est stable.*

Démonstration. La stabilité du schéma (2.10) résulte de (2.18) et de l'équivalence des normes sur $\mathbb{R}^N$. On note A_{0h} la matrice définie par (2.13). Alors $A_h = A_{0h} + h^2 C_h$ où C_h est la matrice diagonale définie par :

$$(C_h)_{ii} = c_i, \quad i = 1, \cdots, N.$$

On a :

$$A_{0h}^{-1} - A_h^{-1} = A_{0h}^{-1} A_h A_h^{-1} - A_{0h}^{-1} A_{0h} A_h^{-1} = A_{0h}^{-1} \underbrace{(A_h - A_{0h})}_{=h^2 C_h \geq 0} A_h^{-1}$$

Soit $v \geq 0$. Alors :

$$A_h - A_{0h} \geq 0 \Rightarrow (A_h - A_{0h})v \geq 0$$

D'après le Lemme 2.4.6, les matrices A_h et A_{0h} sont monotones, donc

$$(A_h - A_{0h})v \geq 0 \Rightarrow A_h^{-1}(A_h - A_{0h})v = v - A_h^{-1} A_{0h} v \geq 0.$$

Ceci est également vrai pour $A_{0h}^{-1} v \geq 0$, donc $A_{0h}^{-1} v \geq A_h^{-1} v$. Finalement :

$$A_{0h}^{-1} \geq A_h^{-1} \geq 0.$$

On remarque que si B est une matrice positive, soit $B \geq 0$, alors

$$\|B\|_\infty = \max_i \sum_j |B_{ij}| = \max_i \sum_j B_{ij}.$$

On en déduit :

$$\|A_h^{-1}\|_\infty = \max_i \sum_j (A_h^{-1})_{ij} \leq \max_i \sum_j (A_{0h}^{-1})_{ij} = \|A_{0h}^{-1}\|_\infty.$$

Il reste donc à estimer $\|A_{0h}^{-1}\|_\infty$. On a aussi :

$$\|A_{0h}^{-1}\|_\infty = \|A_{0h}^{-1}e\|_\infty \quad \text{où} \quad e = (1, \cdots, 1)^T \in \mathbb{R}^N$$

Soit $d_h = h^2 A_{0h}^{-1} e \in \mathbb{R}^N$. De façon équivalente, d_h est solution du système $h^{-2}A_{0h}d_h = e$ résultant de la discrétisation de

$$-u'' = 1 \quad \text{dans} \quad]0,1, \quad u(0) = u(1) = 0$$

dont la solution exacte est

$$u_0(x) = \frac{1}{2}x(1-x), \quad \forall x \in [0,1].$$

Soit $x^{(h)} \in \mathbb{R}^N$ le vecteur de composantes $x_1, \cdots, x_N$, et soit $u_0(x^{(h)})$ le vecteur de composnates $u_0(x_1), \cdots, u_0(x_N)$. Comme u_0 est polynomiale de degré 2, on a :

$$\frac{1}{h^2}A_{0h}u_0(x^{(h)}) = e.$$

Il en résulte :

$$h^2\|A_{0h}^{-1}e\|_\infty = \|u_0(x^{(h)})\|_\infty \leq \sup_{[0,1]}|u_0| = \frac{1}{8}$$

i.e. :

$$\|A_h^{-1}\|_\infty \leq \|A_{0h}^{-1}\|_\infty = \|A_{0h}^{-1}e\|_\infty \leq \frac{1}{8h^2}.$$

Soit $\mu \in \mathbb{R}^N$ et soit $z^{(h)} \in \mathbb{R}^N$ solution de (2.29). On a :

$$\|z^{(h)}\|_\infty = h^2\|A_n^{-1}\mu\|_\infty \leq h^2\|A_n^{-1}\|_\infty h^2\|\mu\|_\infty \leq \frac{1}{8}\|\mu\|_\infty, \tag{2.30}$$

i.e. le schéma (2.25) est stable. □

Définition 2.4.7 (Erreur de discrétisation)**.** On appelle erreur de discrétisation au point x_i la quantité :

$$e_i^{(h)} = u(x_i) - u_i, \quad i = 1, \cdots, N.$$

Théorème 2.4.9. *Soit u la solution du problème (2.24). On suppose que $u \in \mathcal{C}^4([0,1])$. Alors, l'erreur de discrétisation $e^{(h)}$ du schéma (2.25) vérifie :*

$$\|e^{(h)}\|_\infty \leq \frac{h^2}{96}\|u^{(4)}\|_\infty$$

Le schéma (2.25) est donc convergent d'ordre 2.

Démonstration. Le choix $z^{(h)} = e^{(h)}$ dans (2.29) conduit à $\mu = \varepsilon^{(N)}$ défini par (2.27). De (2.30) et (2.28) on déduit :

$$\|e^{(h)}\|_\infty \underset{(2.30)}{\leq} \frac{1}{8}\|\varepsilon^{(N)}\|_\infty \underset{(2.28)}{\leq} \frac{h^2}{96} \sup_{x\in[0,1]} |u^{(4)}(x)|$$

□

Proposition 2.4.10 (Le principe du maximum discret)**.** *La solution du problème (2.24) avec $f = 0$ et la condition de Dirichlet non homogène $u_0 = a$, $u_{N+1} = b$, $a, b \in \mathbb{R}$ donnés, vérifie :*

$$\min(a, b) \leq u_i \leq \max(a, b), \quad i = 1, \cdots, N.$$

Démonstration. Soit $U^{(h)} \in \mathbb{R}^N$ le vecteur de composantes $U_i^{(h)} = u_i$, $i = 1, \cdots, N$. Par définition, et compte tenu de (2.19) :

$$\frac{1}{h^2} A_h U^{(h)} = \begin{pmatrix} a \\ 0 \\ \vdots \\ 0 \\ b \end{pmatrix}.$$

Soit $\lambda \in \mathbb{R}$. Le calcul direct donne :

$$\frac{1}{h^2} A_h (U^{(h)} + \lambda e) = \begin{pmatrix} a+\lambda \\ 0 \\ \vdots \\ 0 \\ b+\lambda \end{pmatrix}, \quad e := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Si $\lambda = -\max(a, b)$ alors $A_h(U^{(h)} + \lambda e) \leq 0 \Rightarrow U^{(h)} + \lambda e \leq 0$, i.e. $u_i \leq \max(a, b)$, $i = 1, \cdots, N$. Si $\lambda = -\min(a, b)$ alors $A_h(U^{(h)} + \lambda e) \geq 0 \Rightarrow U^{(h)} + \lambda e \geq 0$, i.e. $u_i \geq \min(a, b)$, $i = 1, \cdots, N$. □

2.5 Diffusion bidimensionnelle

On considère le problème de diffusion dans un ouvert $\Omega \subset \mathbb{R}^2$ de $\mathbb{R}^2$:

$$\left\{ \begin{array}{rcl} -\Delta u & = & f \quad \text{dans} \quad \Omega, \\ u & = & 0 \quad \text{sur} \quad \partial\Omega. \end{array} \right. \tag{2.31}$$

Le problème est bien posé au sens où si $f \in \mathcal{C}^1(\overline{\Omega})$, alors il existe une unique solution $u \in \mathcal{C}(\overline{\Omega}) \cap \mathcal{C}^2(\Omega)$ de (2.31). Si $f \in L^2(\Omega)$ et si Ω est convexe (ou à bord régulier), alors il existe une unique solution faible $u \in H^2(\Omega)$ de (2.31) ; i.e. vérifiant :

$$\begin{cases} u \in H_0^1(\Omega), \\ \\ \displaystyle\int_\Omega \nabla u \nabla v = \int_\Omega f v dx, \quad \forall v \in H_0^1(\Omega). \end{cases}$$

On peut montrer que si $u \in \mathcal{C}^2(\overline{\Omega})$, alors u est solution de (2.31) ssi u est solution faible de (2.31). Pour discréttiser le problème on se donne un nombre fini de points alignés dans les directions des axes $0x$ et $0y$ comme représentés dans la Figure 2.1 (on choisit un maillage régulier de pas Δx et Δy dans les directions de $0x$ et $0y$ resp.) Certains points sont à l'intérieur de Ω, d'autres sur le bord. Cimme dans le cas de la dimension 1, les inconnues discrètes u_{ij} sont associées aux points $P(x_i, y_j)$ du maillage de sorte que $u_{ij} \sim u(P(x_i, y_j))$. On note $\{P_i,\ i \in I\}$ l'ensemble des points de discrétisation. Dans le cas de points vraiment intérieurs, tels que le point P_1 sur la Figure 2.1, i.e. pour lesquels les points voisins dans le schéma sont aussi intérieurs, on a :

$$-\Delta u(P_1) = \frac{-u(P_2) + 2u(P_1) - u(P_3)}{(\Delta x)^2} + \frac{-u(P_4) + 2u(P_1) - u(P_5)}{(\Delta y)^2} + \\ + O((\Delta x)^2 + (\Delta y)^2)$$

Pour des points proches du bord, i.e. pour lesquels l'un des points voisins est hors de Ω, on doit prendre en compte les conditions sur le bord, ce qui dégrade l'approximation qui devient de l'ordre de $O((\Delta x) + (\Delta y))$. Le schéma numérique peut se réécrire sous forme matricielle $A_{\Delta x, \Delta y} U_{\Delta x, \Delta y} = F_{\Delta x, \Delta y}$ où la matrice $A_{\Delta x, \Delta y}$ est tridiagonake par bandes et dont la largeur de bande dépend du système de numérotation des noeuds choisi. On peut montrer qu ela matrice $A_{\Delta x, \Delta y}$ est inversible et monotone, et que le schéma est stable. De la stabilité et de la consistance, on déduit comme en dimension 1 la convergence du schéma.

Pour simplifier la suite de l'exposé, on suppose que $\Omega =]0, a[\times]0, b[$ est un rectangle. Soit $M > 0$ et soit $N > 0$. On définit les pas du maillage :

$$h_x := \frac{a}{N+1}, \quad h_y := \frac{b}{M+1}$$

et on considère les subdivisions

$$x_i = i h_x, \quad y_j = j h_y, \quad i = 0, \cdots, N+1, \quad j = 0, \cdots, M+1.$$

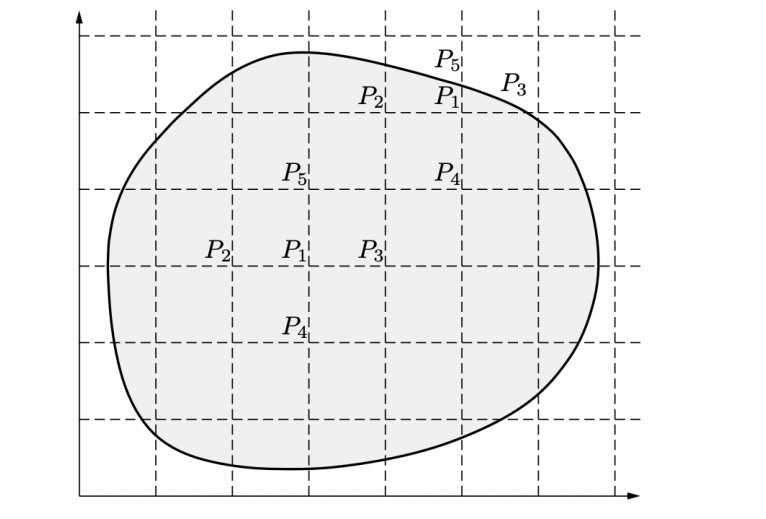

FIGURE 2.1 – Différences finies : discrétisation bidimensionnelle

de $[0, a]$ et $[0, b]$ resp.

L'analogue des problèmes discrétisés (2.10) et (2.25) s'écrit :

$$\begin{cases} \dfrac{1}{h_x^2}(-u_{i-1,j} + 2u_{i,j} - u_{i+1,j}) + \dfrac{1}{h_y^2}(-u_{i,j-1} + 2u_{i,j} - u_{i,j+1}) = f_{ij}, \\ \qquad\qquad i = 1, \cdots, N, \quad j = 1, \cdots, M \\ u_{0,j} = u_{N+1,j} = 0, \qquad\qquad j = 0, \cdots, M+1, \\ u_{i,0} = u_{i,N+1} = 0, \qquad\qquad i = 0, \cdots, N+1 \end{cases} \tag{2.32}$$

où $f_{i,j} = f(x_i, y_j)$, $i = 1, \cdots, N$, $j = 1, \cdots, M$, sont donnés. Sous forme matricielle, le problème (2.25) se réécrit :

$$A_h U_h = f_h \tag{2.33}$$

où $U_h \in \mathbb{R}^{NM}$ est le vecteur

$$U_h = \begin{pmatrix} U_1^{(h)} \\ \vdots \\ U_N^{(h)} \end{pmatrix}, \quad \text{avec} \quad U_j^{(h)} = \begin{pmatrix} u_{1j} \\ \vdots \\ u_{Nj} \end{pmatrix}, \quad j = 1, \cdots M,$$

où $A_h = (A_{ij}^{(h)})_{1 \le i,j \le M} \in \mathbb{R}^{NM \times NM}$ est la matrice carrée d'ordre NM formée de M^2 blocs de taille $N \times N$:

$$A_{ij}^{(h)} = \begin{cases} B_h & \text{si} \quad i = j, \\ C_h & \text{si} \quad |i-j| = 1, \\ 0_h & \text{si} \quad |i-j| > 1, \end{cases}$$

où $B_h = (b_{ij}^{(h)})_{1\leq i,j\leq N} \in \mathbb{R}^{N\times N}$, est la matrice carrée d'ordre N définie par ses coefficients :

$$b_{ij}^{(h)} = \begin{cases} \dfrac{2}{h_x^2} + \dfrac{2}{h_y^2} & \text{si} \quad i = j, \\ -\dfrac{1}{h_x^2} & \text{si} \quad |i-j| = 1, \\ 0 & \text{si} \quad |i-j| > 1, \end{cases} \tag{2.34}$$

$C_h = (c_{ij}^{(h)})_{1\leq i,j\leq N} \in \mathbb{R}^{N\times N}$, est la matrice carrée d'ordre N définie par ses coefficients :

$$c_{ij}^{(h)} = \begin{cases} -\dfrac{1}{h_y^2} & \text{si} \quad i = j, \\ 0 & \text{si} \quad i \neq j \end{cases} \tag{2.35}$$

$0_h \in \mathbb{R}^{N\times N}$, est la matrice carrée nulle d'ordre N.

Proposition 2.5.1. *Si la solution u de (2.31) est de classe $\mathcal{C}^4$, alors : l'erreur de consistance $\varepsilon^{(N,M)}$ vérifie :*

$$\|\varepsilon^{(N,M)}\|_\infty \leq C(h_x^2 + h_y^2)\|u^{(4)}\|_\infty. \tag{2.36}$$

où la constante $C > 0$ est indépendante de u et des pas h_x, h_y.

Proposition 2.5.2. *Le schéma (2.32) est stable.*

Théorème 2.5.3. *Soit u la solution du problème (2.31). On suppose que $u \in \mathcal{C}^4(\overline{\Omega})$. Alors, l'erreur de discrétisation $e^{(h)}$ du schéma (2.32) vérifie :*

$$\|e^{(h)}\|_\infty \leq C(h_x^2 + h_y^2)\|u^{(4)}\|_\infty$$

Le schéma (2.32) est donc convergent d'ordre 2.

La méthode de Gauss-Seidel

On se propose de résoudre le système (2.33) par la méthode de Gauss-Seidel (version Jacobi), ce qui donne le schéma point par point :

$$\begin{cases} \dfrac{1}{h_x^2}(-u_{i-1,j}^n + 2u_{i,j}^{n+1} - u_{i+1,j}^n) \quad + \quad \dfrac{1}{h_y^2}(-u_{i,j-1}^n + 2u_{i,j}^{n+1} - u_{i,j+1}^n) = f_{ij}, \\ u_{ij}^0 = 0, \\ u_{0,j}^n = u_{N+1,j}^n = 0, \\ u_{i,0}^n = u_{i,N+1}^n = 0, \\ i = 1, \cdots, N, \quad j = 1, \cdots, M \qquad n \geq 0. \end{cases} \tag{2.37}$$

Proposition 2.5.4. *Soit $T > 0$ et soit $h = \max(h_x, h_y)$. Si la solution u de (2.31) est de classe $\mathcal{C}^4$, alors, le schéma (2.37) est convergent d'ordre 2 :*

$$\max_{nh \leq T} \|e^n\|_\infty \leq C_T h^2 \|u^{(4)}\|_\infty.$$

où $C_T > 0$ ne dépend que de $T > 0$.

Démonstration. On commence par remarquer que l'erreur de consistance est donnée par :

$$\varepsilon_{ij}^n = \varepsilon_{i,j}^{(N,M)}, \quad \forall (i,j) \in [[1,N]] \times [[1,M]].$$

et vérifie donc

$$\|\varepsilon^n\|_\infty \leq C(h_x^2 + h_y^2)\|u^{(4)}\|_\infty. \tag{2.38}$$

où la constante $C > 0$ est indépendante de u et des pas h_x, h_y. Soit $(\mu_{ij}^n)_{n \geq 0}$ une suite de réels et soit (z_{ij}^n) la suite définie par le schéma :

$$z_{ij}^n = \frac{h_y^2}{2(h_x^2 + h_y^2)}\left(z_{i-1,j}^n + z_{i+1,j}^n\right) + \frac{h_x^2}{2(h_x^2 + h_y^2)}\left(z_{i,j-1}^n + z_{i,j+1}^n\right) +$$

$$+ \frac{h_x^2 h_y^2}{2(h_x^2 + h_y^2)}\mu_{ij}^n, \quad i = 1, \cdots, N, \quad j = 1, \cdots, M$$

$$z_{0j}^n = z_{N+1,j}^n = z_{i,0}^n = z_{i,M+1}^n = 0, \quad i = 1, \cdots, N, \quad j = 1, \cdots, M.$$

Soit $n \geq 0$. On a :

$$|z_{ij}^{n+1}| \leq \frac{1}{2(h_x^2 + h_y^2)}(2h_x^2\|z^n\|_\infty + 2h_y^2\|z^n\|_\infty) + \frac{h_x^2 h_y^2}{2(h_x^2 + h_y^2)}\|\mu^n\|_\infty$$

$$\leq \|z^n\|_\infty + \frac{h_x^2 h_y^2}{2(h_x^2+h_y^2)}\|\mu^n\|_\infty.$$

On pose $\mu_{ij}^n = \varepsilon_{ij}$. alors $z_{ij}^n = u(x_i, y_j) - u_{ij}^n = e_{ij}^n$ coïncide avec l'erreur de convergence et on a, tant que $nh \leq T$:

$$\|e^n\|_\infty \leq \|e^0\|_\infty + Cnh^4\|u^{(4)}\|_\infty \leq Ch^2\|u^{(4)}\|_\infty + CTh^3\|u^{(4)}\|_\infty$$

$$\leq C(1+hT)h^2\|u^{(4)}\|_\infty$$

□

Bibliographie

[1] Thierry Gallouët, Raphaèle Herbin. Analyse numérique des équations aux dérivées partielles. Master. Marseille, France. 2011. (https ://cel.hal.science/cel-00637008v2) *Chapitres 1.2 et 1.4.*

[2] Xavier Gourdon. Les Maths en Tête. Analyse. Elllpses Marketing, Paris, 2020, *p. 338–339*

[3] Lionel Sainsaulieu. Calcul Scientifique. Cours et exercices corrigés pour le second cycle et les écoles d'ingénieurs, Masson, Paris 1996. *Chapitres 1.2, 2.1, 2.2, 3.5.*

[4] Brigitte Lucquin, Olivier Pironneau. Introduction au calcul scientifique. Masson, Paris, 1996. *Chapitre VII.1*

[5] Alfio Quarteroni, Riccardo Sacco, Fausto Saleri Numerical Mathematics. Springer, Berlin, 2007. *Chapitre 12.1*

[6] Alfio Quarteroni, Riccardo Sacco, Paola Gervasio Calcul scientifique. Springer, Milan, 2010. *Chapitres 8.2.1 à 8.2.5*

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 :

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

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 $\overrightarrow{\nabla} T \cdot \overrightarrow{dS} > 0$ dans V. D'après la formule de Stokes :

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

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 :

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

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 $t > 0$:

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

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

i.e. :

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

□

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 :

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

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 :*

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

où $(c_n(u_0))_{n\in\mathbb{N}}$ 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 Théorème 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 :

$$\left\{ \begin{array}{lll} a + b & = & 0, \\ ae^{\omega L} + be^{-\omega L} = 0 & & \end{array} \right.$$

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 :

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

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 : $\forall k, \ell > 0$,

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

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. □

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 $(u_i^n)_{0\le i\le I+1, 0\le n\le N+1}$ 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

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

Proposition 3.3.1. *On pose :*

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

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

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

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 :

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

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. □

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. :*

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

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 $j \in [[0, N+1]]$ et soit $n \geq 0$. On pose $L = \pi$ dans (3.5). D'après (3.5), on a alors :

$$u(x_j, t_n) - u_j^n = \sum_{k \in \mathbb{Z}} c_k (e^{-k^2 t_n} - \gamma_k^n) e^{ikx_j}$$

avec :

$$|c_k(e^{-k^2 t_n} - \gamma_k^n)e^{ikx_j}| = |c_k||e^{-k^2 t_n} - \gamma_k^n| \leq 2|c_k|$$

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 :

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

i.e. :

$$\lim_{(\Delta t, \Delta x)\to(0,0)} \ln(\gamma_k^n) = -k^2 t_n.$$

On en déduit, par continuité de l'exponentielle :

$$\lim_{(\Delta t, \Delta x)\to(0,0)} \gamma_k^n = e^{-k^2 t_n}$$

i.e :

$$\lim_{(\Delta t, \Delta x)\to(0,0)} (\gamma_k^n - e^{-k^2 t_n}) = 0.$$

Du Théorème de convergence dominée il résulte que

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

On remarque que :

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

et donc :

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

□

Schéma implicite et schéma de Crank-Nickolson

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

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

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 :

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

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

□

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 :

$$\|Z^{n+1}\|_2 \leq \left\|\left(1+\theta\frac{\Delta t}{(\Delta x)^2}A_N\right)^{-1}\right\|_2 \left(\left\|1-(1-\theta)\frac{\Delta t}{(\Delta x)^2}A_N\right\|_2 \|Z^n\|_2 + \Delta t\|\mu^n\|_2\right)$$

$$\leq \frac{\left(1-(1-\theta)\frac{\Delta t}{(\Delta x)^2}\lambda_N(A_N)\right)}{\left(1+\theta\frac{\Delta t}{(\Delta x)^2}\lambda_N(A_N)\right)}\|Z^n\|_2 + \frac{\Delta t\|\mu^n\|_2}{\left(1+\theta\frac{\Delta t}{(\Delta x)^2}\lambda_N(A_N)\right)}$$

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

$$\leq \tau_N^n\|Z^0\|_2 + \Delta t\sum_{k=0}^{n-1}\tau_N^k\|\mu^{n-k}\|_2.$$

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

$$\|e^n\|_2 \underset{0<\tau_N<1}{\leq} C\Delta t(\Delta t + (\Delta x)^2)\frac{1-\tau_N^n}{1-\tau_N}$$

$$\leq C(\Delta t + (\Delta x)^2)\frac{\Delta t}{1-\tau_N} \leq$$

$$\leq C\left(1+\theta\frac{\Delta t}{(\Delta x)^2}\lambda_N(A_N)\right)(\Delta x)^2(\Delta t + (\Delta x)^2)$$

□

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. : $\min_i u_i^{n+1} \geq \min_i u_i^n \geq \min_i u_i^0$.

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. : $\max_i u_i^{n+1} \geq \max_i u_i^n \geq \max_i u_i^0$.

□

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 :

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

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 $]0,2[= f_\theta\left(\left]0, \frac{2}{1-2\theta}\right[\right)$. 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*

Chapitre 4

Equation des Ondes

4.1 Introduction et généralités

Une onde naît quand une perturbation locale d'une grandeur physique Ψ mesurée par la dérivée temporelle $\frac{\partial \Psi}{\partial t}$ induit la variation spatiale d'une autre grandeur physique Φ, soit :

$$\frac{\partial \Psi}{\partial t} = a \frac{\partial \Phi}{\partial x}$$

et inversement :

$$\frac{\partial \Phi}{\partial t} = b \frac{\partial \Psi}{\partial x}$$

pour des constantes $a, b > 0$. On en déduit que Ψ et Φ sont solutions de la même équation :

$$\frac{\partial^2 \Psi}{\partial t^2} = ab \frac{\partial^2 \Psi}{\partial x^2} \quad \text{et} \quad \frac{\partial^2 \Phi}{\partial t^2} = ab \frac{\partial^2 \Phi}{\partial x^2}$$

Ondes acoustiques

Dans un $\Omega \subset \mathbb{R}^3$ un ouvert rempli de fluide, la propagation d'ondes acoustiques dépend de la densité $\rho(x)$ du fluide au point $x \in \Omega$, de la vitesse de propagation locale $c(x)$ des ondes. La pression $p(x,t)$ et la vitesse du fluide sont liées par les équations :

$$\rho(x) \frac{\partial \vec{v}}{\partial t} = \vec{\nabla} p, \quad \frac{1}{\rho(x) c^2(x)} \frac{\partial p}{\partial t} = \operatorname{div}(\vec{v})$$

d'où on déduit que p est solution de :

$$\frac{1}{\rho(x) c^2(x)} \frac{\partial^2 p}{\partial t^2} - \operatorname{div}\left(\frac{\vec{\nabla} p}{\rho(x)} \right) = 0, \quad \Omega, \quad t > 0.$$

En milieu homogène, cette équation devient :

$$\frac{\partial^2 p}{\partial t^2} - c^2 \Delta p = 0, \quad \Omega, \quad t > 0.$$

4.2 Propriétés de l'équation des ondes $1D$

Le problème modèle

Soit le problème : trouver $u : \mathbb{R}^N \times \mathbb{R}^+ \to \mathbb{R}$ solution de (4.1)

$$\begin{cases} \rho \dfrac{\partial^2 u}{\partial t^2} - \operatorname{div}(\mu \nabla u) = f, \quad \mathbb{R}^N \times \mathbb{R}^+, \\ \\ u(x,0) = u_0(x), \quad \mathbb{R}^N \\ \\ \dfrac{\partial u}{\partial t}(x,0) = u_1(x), \quad \mathbb{R}^N \end{cases} \tag{4.1}$$

Les coefficients $\rho(x)$ et $\mu(x)$ qui caractérisent le milieu de propagation sont supposés mesurables et vérifier :

$$0 < \rho_- \leq \rho(x) \leq \rho_+ < +\infty \quad \text{p.p. dans} \quad \mathbb{R}^N$$
$$0 < \mu_- \leq \mu(x) \leq \mu_+ < +\infty \quad \text{p.p. dans} \quad \mathbb{R}^N$$

Les données sont donc, outre ρ et μ, les donées initiales $u_0(x)$, $u_1(x)$ et le second membre $f(x,t)$.

La formule de d'Alembert

On suppose que $N = 1$ et que l'onde se propage avec la vitesse constante $c > 0$. L'équation (4.1) devient :

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} - c^2 \dfrac{\partial^2 u}{\partial x^2} = f, \\ \\ u(x,0) = u_0(x), \\ \\ \dfrac{\partial u}{\partial t}(x,0) = u_1(x), \end{cases} \tag{4.2}$$

Théorème 4.2.1. *La solution du problème (4.2) est donnée par la formule de d'Alembert :*

$$u(x,t) = \frac{1}{2}(u_0(x+ct) + u_0(x-ct)) + \frac{1}{2c}\int_{x-ct}^{x+ct} u_1(s)ds + \frac{1}{2c}\int_0^t \int_{|y-x| \leq c(t-s)} f(y,s)dsdy \tag{4.3}$$

Démonstration. On remarque que l'opérateur des ondes se décompose sous la forme :

$$\frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x^2} = \left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right) \circ \left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right)$$

Alors :

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = f \iff \begin{cases} \dfrac{\partial v}{\partial t} - c\dfrac{\partial v}{\partial x} = f, \\ \dfrac{\partial u}{\partial t} + c\dfrac{\partial u}{\partial x} = v, \\ u(x,0) = u_0(x), \quad v_0(x) = u_1(x) + cu_0'(x) \end{cases}$$

Par la méthode des caractéristiques, on trouve successivement :

$$v(x,t) = v_0(x+ct) + \int_0^t f(x + c(t-s), s)ds$$

$$u(x,t) = u_0(x-ct) + \int_0^t v(x - c(t-s), s)ds$$

avec : $\forall s \in [0,t]$,

$$v(x - c(t-s), s) = v_0(x - c(t-2s)) + \int_0^s f(x - c(t - 2s + \tau), \tau)d\tau.$$

Il en résulte :

$$u(x,t) = u_0(x-ct) + \int_0^t v_0(x - c(t-2s))ds + \int_0^t \int_0^s f(x - c(t-2s+\tau), \tau)d\tau ds.$$

Par définition de v_0 :

$$\int_0^t v_0(x - c(t-2s))ds \underset{y = x - c(t-2s)}{=} \int_{x-ct}^{x+ct} v_0(y)\frac{dy}{2c} =$$

$$= \frac{1}{2}\int_{x-ct}^{x+ct} u_1(y)dy + \frac{1}{2}(u_0(x+ct) + u_0(x-ct)).$$

De plus :

$$\int_0^t \int_0^s f(x - c(t-2s+\tau), \tau)d\tau ds = \frac{1}{2c}\int\int_{\substack{|y-x| \le t-\tau \\ 0 \le \tau \le t}} f(y,\tau)dyd\tau$$

□

Remarque 14. Si $f = 0$ alors la solution se décompose sous la forme :

$$u(x,t) = \underbrace{\frac{1}{2}u_0(x+ct) + \frac{1}{2c}\int_0^{x+ct} u_1(s)ds}_{=:u^+(x+ct)} + \underbrace{\frac{1}{2}u_0(x-ct) - \frac{1}{2c}\int_0^{x-ct} u_1(s)ds}_{=:u^-(x-ct)}$$

où u^+, resp. u^-, correspond à une onde se propageant dans la direction des $x > 0$, resp. des $x < 0$.

Cône de dépendance et propagation à vitesse finie

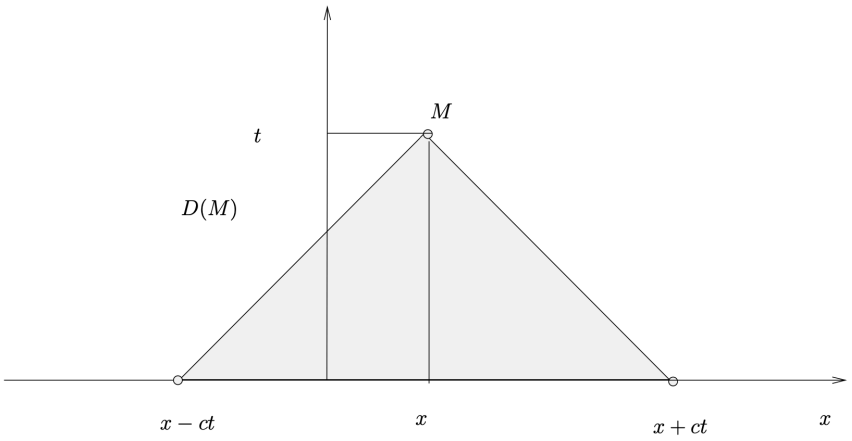

FIGURE 4.1 – Cône de dépendance.

La formule de d'Alembert (4.3) montre que la solution $u(x,t)$ au point $M(x,t)$ est entièrement déterminée par les valeurs des données initiales u_0, u_1 et du second membre f aux points du cône de dépendance $D(M)$ (figure 4.1). Autrement dit, les ondes se propagent à vitesse finie. En effet, si u_0, u_1 et f sont à support compact dans $K = [a,b]$, alors, à l'instant $t > 0$, $u(\cdot,t)$ est à support dans $K_t := [a-ct, b+ct]$.

Régularité de la solution de (4.2)

Si $f = 0$, la formule de d'Alembert (4.3) montre que si $u_0 \in \mathcal{C}^{k+1}(\mathbb{R})$ et si $u_1 \in \mathcal{C}^k(\mathbb{R})$, alors $u \in \mathcal{C}^{k+1}(\mathbb{R} \times \mathbb{R}^{+*})$, i.e. l'équation des ondes conserve la régularité.

Dans le cas général où $f \neq 0$, alors, contrairement au cas du Laplacien, on ne gagne pas deux crans de régularité lorsqu'on passe de f à u. On na voir qu'on ne gagne qu'un cran de régularité.

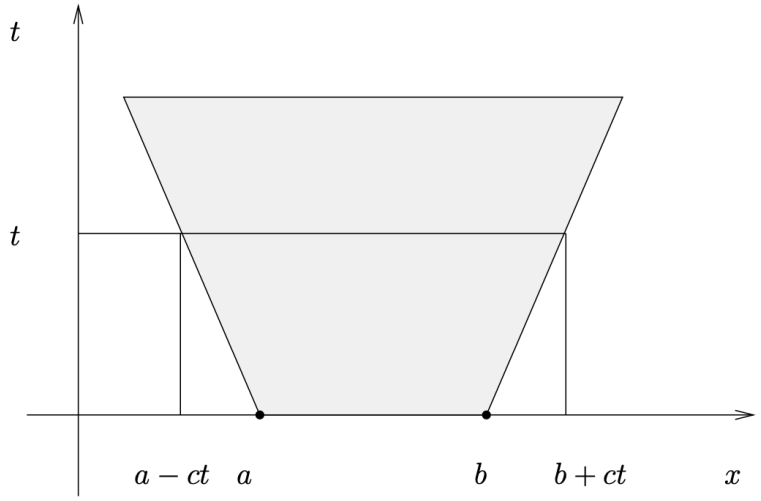

FIGURE 4.2 – Cône de de dépendance.

Pour cela, on suppose que $u_0 = u_1 = 0$ par linéarité de l'équation des ondes, puisque l'influence des données initiales a été étudiée. Alors, on vérifie directement que la transformée de Fourier de u définie par :

$$\hat{u}(\xi, t) = \int_0^{+\infty} u(x,t)e^{-i\xi x}dx, \quad \forall \xi \in \mathbb{R}, \quad \forall t > 0.$$

est solution de :

$$\begin{cases} \dfrac{\partial^2 \hat{u}}{\partial t^2} + (c\xi)^2 \hat{u} = \hat{f}, \\ \\ \hat{u}(0) = \dfrac{\partial \hat{u}}{\partial t}(0) = 0. \end{cases}$$

d'où on déduit que

$$\hat{u}(\xi, t) = a(t)\cos(c\xi t) + b(t)\sin(c\xi t)$$

avec, d'après la méthode de variation des constantes :

$$a'(t)\cos(c\xi t) + b'(t)\sin(c\xi t) = 0.$$

Il en résulte :

$$\frac{\partial \hat{u}}{\partial t} = -c\xi a \sin(c\xi t) + c\xi b \cos(c\xi t)$$

$$\frac{\partial^2 \hat{u}}{\partial t^2} = -c\xi a' \sin(c\xi t) + c\xi b' \cos(c\xi t) - (c\xi)^2 \underbrace{(a\cos(c\xi t) + b\sin(c\xi t))}_{=\hat{u}(\xi,t)}$$

$$\Rightarrow \frac{\partial^2 \hat{u}}{\partial t^2} + (c\xi)^2 \hat{u} = \hat{f} = -c\xi a' \sin(c\xi t) + c\xi b' \cos(c\xi t).$$

On en déduit que a' et b' sont solutions du système :

$$\begin{cases} a' \cos(c\xi t) + b' \sin(c\xi t) = 0, \\ -a' \sin(c\xi t) + b' \cos(c\xi t) = \dfrac{\hat{f}}{c\xi} \end{cases}$$

i.e. :

$$a'(t) = -\frac{\hat{f}(\xi,t)}{c\xi} \sin(c\xi t), \quad b'(t) = \frac{\hat{f}(\xi,t)}{c\xi} \cos(c\xi t)$$

Finalement :

$$\hat{u}(\xi,t) = a_0 \cos(c\xi t) + b_0 \sin(c\xi t) +$$

$$+ \int_0^t \frac{\hat{f}(\xi,s)}{c\xi} \left(-\cos(\xi t)\sin(\xi s) + \sin(c\xi t)\cos(c\xi s)\right) ds =$$

$$= \int_0^t \frac{\sin(c\xi(t-s))}{c\xi} \hat{f}(\xi,s) ds, \quad \forall \xi \in \mathbb{R}, \quad \forall t > 0.$$

On en déduit :

$$c\widehat{\frac{\partial u}{\partial x}}(\xi,t) = ic\xi \hat{u}(\xi,t) = i \int_0^t \sin(c\xi(t-s)) \hat{f}(\xi,s) ds, \quad \forall \xi \in \mathbb{R}, \quad \forall t > 0.$$

$$\frac{\partial \hat{u}}{\partial t}(\xi,t) = \int_0^t \cos(c\xi(t-s)) \hat{f}(\xi,s) ds, \quad \forall \xi \in \mathbb{R}, \quad \forall t > 0.$$

d'où, par Cauchy-Schwartz :

$$\left| c\widehat{\frac{\partial u}{\partial x}}(\xi,t) \right|^2 \leq t \int_0^t \|\hat{f}(\xi,s)\|^2 ds$$

$$\left| \frac{\partial \hat{u}}{\partial t}(\xi,t) \right|^2 \leq t \int_0^t \|\hat{f}(\xi,s)\|^2 ds$$

ce qui donne, par l'égalité de Plancherel :

$$\left\| \frac{\partial u}{\partial x}(t) \right\|^2_{L^2(\mathbb{R})} \leq t \int_0^t \int_{\mathbb{R}} \|f(x,s)\|^2 dx ds$$

$$\left\| \frac{\partial u}{\partial t}(t) \right\|^2_{L^2(\mathbb{R})} \leq t \int_0^t \int_{\mathbb{R}} \|f(x,s)\|^2 dx ds$$

ce qui montre que

$$f \in L^2_{\text{loc}}(\mathbb{R}^+, L^2(\mathbb{R})) \Rightarrow u \in L^\infty(\mathbb{R}^+, H^1(\mathbb{R})) \quad \text{et} \quad \frac{\partial u}{\partial t} \in L^\infty(\mathbb{R}^+, L^2(\mathbb{R}))$$

Développement en série de Fourier

Soit à résoudre :

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} - \dfrac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1, \quad t > 0, \\ \\ u(0,t) = u(1,t) = 0, \quad t > 0, \\ \\ u(x,0) = u_0(x), \quad \dfrac{\partial u}{\partial t}(x,0) = u_1(x), \quad 0 < x < 1. \end{cases} \tag{4.4}$$

On cherche une solution u sous la forme :

$$u(x,t) = X(x)T(t), \quad x \in]0,1[, \quad t > 0$$

Après report dans (4.4), on obtient :

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

avec la condition sur le bord :

$$u(0,t) = u(1,t) = 0 \underset{T(t)\neq 0}{\Rightarrow} X(0) = X(1) = 0. \tag{4.5}$$

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

$$X(x) = ae^{\omega x} + be^{-\omega x} \underset{(4.5)}{\equiv} 0$$

ce qui contredit $u \not\equiv 0$. Donc nécessairement : $\lambda = \omega^2 < 0$ et

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

avec

$$X(0) = X(1) = 0 \Rightarrow a = 0 \quad \text{et} \quad \omega \in \pi\mathbb{N}.$$

On en déduit :

$$u(x,t) = \sum_{n\geq 1} T_n(t) \underbrace{\sin(n\pi x)}_{=:X_n(x)} =: \sum_{n\geq 1} u_n(x,t) \tag{4.6}$$

On suppose que

$$u_0 \in H_0^1(0,1) \quad \text{et} \quad u_1 \in L^2(0,1) \tag{4.7}$$

sont développables en séries de Fourier sous la forme :

$$u_0(x) = \sum_{n\geq 0} b_n^0 \sin(n\pi x), \quad u_1(x) = \sum_{n\geq 0} b_n^1 \sin(n\pi x),$$

avec, compte tenu de (4.7) :

$$\sum_{n\geq 0} (n\pi)^2 |b_n^0|^2 < +\infty \quad \text{et} \quad \sum_{n\geq 0} |b_n^1|^2 < +\infty$$

Alors les fonctions T_n, $n \geq 0$, dans (4.6) sont, au moins formellement, solutions de

$$\begin{cases} T_n'' + (n\pi)^2 T_n = 0, \quad t > 0, \\ \\ T_n(0) = b_n^0, \quad T'(0) = b_n^1, \quad n \geq 0. \end{cases} \tag{4.8}$$

On vérifie immédiatement que (4.8) admet pour unique solution :

$$T_n(t) = b_n^0 \cos(n\pi t) + \frac{b_n^1}{n\pi} \sin(n\pi t), \quad t > 0, \quad n \geq 1. \tag{4.9}$$

Proposition 4.2.2. *Soit $u_0 \in H_0^1(0,1)$ et soit $u_1 \in L^2(0,1)$. On suppose que u_0 et u_1 sont développables en séries de Fourier sous la forme :*

$$u_0(x) = \sum_{n\geq 0} b_n^0 \sin(n\pi x), \quad u_1(x) = \sum_{n\geq 0} b_n^1 \sin(n\pi x),$$

Alors, pour tout $T > 0$, la série (4.6), (4.9) converge uniformément dans $\mathcal{C}^0([0,T])$ vers $u \in \mathcal{C}^0([0,T], H_0^1(0,1)) \cap \mathcal{C}^1([0,T], L^2(0,1))$ solution de :

$$u \in \mathcal{C}^0([0,T], H_0^1(0,1)),$$

$$\frac{\partial u}{\partial t} \in \mathcal{C}^0([0,T], L^2(0,1)),$$

$$\frac{d}{dt}\int_0^1 \frac{\partial u(t)}{\partial t} v dx + \int_0^1 \frac{\partial u(t)}{\partial x} v' dx = 0, \quad \forall v \in H_0^1(0,1), \quad \textit{dans} \quad \mathcal{D}'(0,T)$$

$$u(0) = u_0, \quad \frac{\partial u}{\partial t}(0) = u_1.$$

Démonstration. Soit $T > 0$. On a : $\forall n \geq 1$, $\forall x \in]0,1[$, $\forall t \in [0,T]$,

$$|u_n(x,t)| \leq |T_n(t)| \leq |b_n^0| + \frac{|b_n^1|}{n\pi}$$

d'où : $\forall n \geq 1$, $\forall t > 0$,

$$\int_0^1 |u(t)|^2 dx = \sum_{n\geq 1} |T_n(t)|^2 \leq 2 \sum_{n\geq 1} \left(|b_n^0|^2 + \frac{|b_n^1|^2}{(n\pi)^2} \right) < +\infty$$

La série majorante étant convergente, on en déduit que $t \mapsto u(t)$ est somme d'une série de fonctions continues uniformément convergente sur $[0,T]$, donc que $u \in \mathcal{C}^0([0,T], L^2(0,1))$. De même :

$$\frac{\partial u_n}{\partial x}(x,t) = n\pi T_n(t) \cos(n\pi x), \quad \forall x \in [0,1], \quad \forall t \in [0,T], \quad \forall n \geq 1$$

donc : $\forall n \geq 1$, $\forall t > 0$,

$$\int_0^1 \left| \sum_{n\geq 1} \frac{\partial u_n}{\partial x}(t) \right|^2 dx = \sum_{n\geq 1} |T_n(t)|^2 \leq 2 \sum_{n\geq 1} \left((n\pi)^2 |b_n^0|^2 + |b_n^1|^2 \right) < +\infty$$

où la série majorante est convergente. On en déduit que la série des dérivées $\sum \frac{\partial u_n}{\partial x}$ est uniformément convergente sur $[0,T]$ et que $u \in \mathcal{C}^0([0,T], H_0^1(0,1))$ avec

$$\frac{\partial u}{\partial x} = \sum_{n\geq 1} n\pi T_n(t) \cos(n\pi x), \quad \forall t \in [0,T], \quad \text{p.p. en } x \in]0,1[.$$

D'autre part : $\forall n \geq 1$, $\forall x \in]0,1[$, $\forall t \in [0,T]$,

$$\frac{\partial u_n}{\partial t} = T_n'(t) \sin(n\pi x)$$

avec

$$|T_n'(t)|^2 \leq 2((n\pi)^2 |b_n^0|^2 + |b_n^1|^2)$$

$$\Rightarrow \int_0^1 \left| \frac{\partial u}{\partial t} \right|^2 dx = \sum_{n\geq 1} |T_n'(t)|^2 \leq 2 \sum_{n\geq 1} ((n\pi)^2 |b_n^0|^2 + |b_n^1|^2) < +\infty$$

La série majorante étant convergente, on en déduit que la série des dérivées $\frac{\partial u_n}{\partial t}$ est uniformément convergente sur $[0,T]$ de somme :

$$\frac{\partial u}{\partial t} = \sum_{n\geq 1} \frac{\partial u_n}{\partial t} = \sum_{n\geq 1} T_n'(t) \sin(n\pi x) \quad \forall t \in [0,T], \quad \text{p.p. en } x \in]0,1[$$

et donc $\dfrac{\partial u}{\partial t} \in \mathcal{C}^0([0,T], L^2(0,1))$.

Soit $\varphi \in \mathcal{D}(0,1)$ et soit $N \geq 1$. Par construction :

$$\sum_{n=1}^{N} \int_0^1 \frac{\partial^2 u_n(t)}{\partial t^2} \varphi dx + \sum_{n=1}^{N} \int_0^1 \frac{\partial u_n(t)}{\partial x} \varphi' dx = 0, \quad \forall t > 0.$$

Soit $\phi \in \mathcal{D}(0,T)$. Il en résulte :

$$-\sum_{n=1}^{N} \int_0^T \phi'(t) \int_0^1 \frac{\partial u_n(t)}{\partial t} \varphi dx dt + \sum_{n=1}^{N} \int_0^T \phi(t) \int_0^1 \frac{\partial u_n(t)}{\partial x} \varphi' dx dt = 0.$$

avec, par convergence uniforme de la série $\sum u_n$ vers u dans $\mathcal{C}^0([0,T], H_0^1(\Omega))$:

$$\lim_{N \to +\infty} \sum_{n=1}^{N} \int_0^T \phi(t) \int_0^1 \frac{\partial u_n(t)}{\partial x} \varphi' dx dt = \int_0^T \phi(t) \int_0^1 \frac{\partial u(t)}{\partial x} \varphi' dx dt,$$

et par convergence uniforme de la série $\sum \frac{\partial u_n}{\partial t}$ vers $\frac{\partial u}{\partial t}$ dans $\mathcal{C}^0([0,T], L^2(\Omega))$:

$$\lim_{N \to +\infty} \sum_{n=1}^{N} \int_0^T \phi'(t) \int_0^1 \frac{\partial u_n(t)}{\partial t} \varphi dx dt = \int_0^T \phi'(t) \int_0^1 \frac{\partial u(t)}{\partial t} \varphi dx dt,$$

i.e. :

$$-\int_0^T \phi'(t) \int_0^1 \frac{\partial u(t)}{\partial t} \varphi dx dt + \int_0^T \phi(t) \int_0^1 \frac{\partial u(t)}{\partial x} \varphi' dx dt = 0.$$

i.e. encore, par densité de $\mathcal{D}(0,1)$ dans $H_0^1(0,1)$: $\forall v \in H_0^1(0,1)$, $\forall \phi \in \mathcal{D}(0,T)$,

$$-\int_0^T \phi'(t) \int_0^1 \frac{\partial u(t)}{\partial t} v dx dt + \int_0^T \phi(t) \int_0^1 \frac{\partial u(t)}{\partial x} v' dx dt = 0.$$

Finalement :

$$\frac{d}{dt} \int_0^1 \frac{\partial u(t)}{\partial t} v dx + \int_0^1 \frac{\partial u(t)}{\partial x} v' dx = 0, \quad \forall v \in H_0^1(0,1), \quad \text{dans} \quad \mathcal{D}'(0,T).$$

□

Proposition 4.2.3. *Soit $u_0 \in H_0^1(0,1)$ t.q. $u_0' \in H^1(0,1)$ et soit $u_1 \in H^1(0,1)$. On suppose que u_0 et u_1 sont développables en séries de Fourier sous la forme :*

$$u_0(x) = \sum_{n \geq 0} b_n^0 \sin(n\pi x), \quad u_1(x) = \sum_{n \geq 0} b_n^1 \sin(n\pi x),$$

Alors, pour tout $T > 0$, la série (4.6), (4.9) converge uniformément dans $\mathcal{C}^0([0,T])$ vers $u \in \mathcal{C}^0([0,T], H_0^1(0,1)) \cap \mathcal{C}^2([0,T], L^2(0,1))$ solution de :

$$u \in \mathcal{C}^0([0,T], H_0^1(0,1)),$$
$$\frac{\partial u}{\partial t} \in \mathcal{C}^0([0,T], H^1(0,1)),$$
$$\frac{\partial^2 u}{\partial t^2} \in \mathcal{C}^0([0,T], L^2(0,1)),$$
$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0 \quad dans \quad L^2((0,1)\times(0,T))$$
$$u(0) = u_0, \quad \frac{\partial u}{\partial t}(0) = u_1.$$

Démonstration. Par régularité de u_0 et u_1 :

$$\sum_{n\geq 1} \left((n\pi)^4 |b_n^0|^2 + (n\pi)^2 |b_n^1|^2\right) < +\infty.$$

Par le même raisonnement que précédemment, on montre que la série $\sum \frac{\partial^2 u_n}{\partial t^2}$ converge uniformément dans $\mathcal{C}^0([0,T], L^2(0,1))$ vers $\frac{\partial^2 u}{\partial t^2}$. □

Proposition 4.2.4. *Sous les hypothèses de la Proposition 4.2.2, l'énergie se conserve au sens suivant :*

$$\frac{1}{2}\left|\frac{\partial u(t)}{\partial t}\right|^2 + \frac{1}{2}\int_0^1 \left|\frac{\partial u(t)}{\partial x}\right|^2 dx = \frac{1}{2}\int_0^1 |u_0'|^2 dx + \frac{1}{2}\int_0^1 |u_1|^2 dx, \quad \forall t \in [0,T].$$

Démonstration. Soit $t >$ et soit $n \geq 1$. par définition de T_n :

$$0 = \int_0^t (T_n''(s)T_n'(s) + (n\pi)^2 T_n(s)T_n'(s))ds =$$
$$= \frac{1}{2}\int_0^t (\frac{d}{ds}\left((T_n'(s))^2\right) + (n\pi)^2(|T_n(s)|^2)')ds$$
$$= \frac{1}{2}(|T_n'(t)|^2 + (n\pi)^2|T_n(t)|^2) - \frac{1}{2}((b_n^1)^2 + (n\pi)^2|b_n^0|^2).$$

On en déduit :

$$0 = \frac{1}{2}\sum_{n\geq 1}(|T_n'(t)|^2 + (n\pi)^2|T_n(t)|^2) - \frac{1}{2}\sum_{n\geq 1}((b_n^1)^2 + (n\pi)^2|b_n^0|^2)$$

$$= \frac{1}{2}\int_0^1 \left|\frac{\partial u(t)}{\partial t}\right|^2 dx + \frac{1}{2}\int_0^1 \left|\frac{\partial u(t)}{\partial x}\right|^2 dx - \frac{1}{2}\int_0^1 |u_1|^2 dx - \frac{1}{2}\int_0^1 |u_0'|^2 dx$$

□

4.3 Approximation numérique

On considère les subdivisions :

$$x_0 = 0 < x_1 < \cdots < x_N < x_{N+1} = 1$$

$$t_0 = 0 < t_1 < \cdots < t_n < \cdots$$

supposées régulières de pas $\Delta x > 0$ et $\Delta t > 0$ resp.

Soit $(u_i^n)_{0\leq i\leq N+1, n\geq 0}$ la suite solution du schéma :

$$\begin{cases} u_i^0 = u_0(x_i), \quad i = 0, \cdots, N+1, \\ \\ v_i^0 = u_1(x_i) + cu_0'(x_i), \quad i = 1, \cdots, N, \\ \\ u_i^1 = \left(1 - c\dfrac{\Delta t}{\Delta x}\right) u_i^0 + c\dfrac{\Delta t}{\Delta x} u_{i-1}^0 + \Delta t v_i^0, \quad i = 1, \cdots, N. \end{cases} \tag{4.10}$$

$$\begin{cases} \dfrac{u_i^{n-1} - 2u_i^n + u_i^{n+1}}{\Delta t^2} - \dfrac{c^2}{(\Delta x)^2}(u_{i-1}^n - 2u_i^n + u_{i+1}^n) = f(x_i), \quad i = 1 \cdots, N, \quad n \geq 1 \\ \\ u_0^n = u_{N+1}^n = 0, \quad n \geq 1, \end{cases} \tag{4.11}$$

Remarque 15. L'initialisation (4.10) est le premier pas du schéma :

$$\begin{cases} \dfrac{v_i^{n+1} - v_i^n}{\Delta t} - c\dfrac{v_{i+1}^n - v_i^n}{\Delta x} = f(x_i), \\ \\ \dfrac{u_i^{n+1} - u_i^n}{\Delta t} + c\dfrac{u_i^n - u_{i-1}^n}{\Delta x} = v_i^n, \quad i = 1, \cdots, N, \quad n \geq 0, \\ \\ v_i^0 = u_1(x_i) + cu_0'(x_i), \quad u_i^0 = u_0(x_i), \quad i = 1, \cdots, N \end{cases}$$

qui découle imédialement de la réécriture de (4.2)sous la forme :

$$\begin{cases} \dfrac{\partial v}{\partial t} - c\dfrac{\partial v}{\partial x} = f, \\ \\ \frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = v, \quad 0 < x < 1 \quad t > 0. \end{cases}$$

Remarque 16. On veut avoir : $u_i^n \sim u(x_i, t_n)$, $i = 1, \cdots, N$, $n \geq 0$.

Proposition 4.3.1. *Le schéma (4.10)–(4.11) est consistant d'ordre* 2 *en temps et en espace.*

Démonstration. Soit $n \geq 1$. On définit l'erreur de consistance ε_i^n au point (x_i, t_n) en posant : $\varepsilon_i^n = \varepsilon(x_i, t_n)$ avec : $\forall x \in [0,1]$, $\forall t > 0$.

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

$$-\frac{c^2}{(\Delta x)^2}\left(u(x-\Delta x,t) - 2u(x,t) + u(x+\Delta x,t)\right) - f(x).$$

La formule de Taylor donne directement : $\forall i \in [[1,N]]$, $\forall n \geq 0$,

$$|\varepsilon_i^n| \leq C\left(\left\|\frac{\partial^4 u}{\partial t^4}\right\|_\infty (\Delta t)^2 + \left\|\frac{\partial^4 u}{\partial x^4}\right\|_\infty (\Delta x)^2\right), \quad \forall i \in [[1,N]], \quad \forall n \geq 0. \quad (4.12)$$

i.e. :

$$\varepsilon_i^n = O((\Delta t)^2 + (\Delta x)^2), \quad i = 1, \ldots, N, \quad n \geq 0.$$

□

Proposition 4.3.2. *On suppose que*

$$c\frac{\Delta t}{\Delta x} < 1.$$

Alors, le schéma (4.10)–(4.11) est convergent d'ordre 2 *en temps et en espace.*

Démonstration. Soit $(\mu_i^n)_{n\geq 0}$ une suite de réels et soit z_i^n, $n \geq 0$, $i = 1, \cdots, N$, définie par :

$$\begin{cases} \dfrac{z_i^{n-1} - 2z_i^n + z_i^{n+1}}{\Delta t^2} - \left(\dfrac{c}{\Delta x}\right)^2 (z_{i-1}^n - 2z_i^n + z_{i+1}^n) = \mu_i^n, \quad i = 1 \cdots, N, \quad n \geq 1 \\ z_0^n = z_{N+1}^n = 0, \quad n \geq 0, \\ z^0 \in \mathbb{R}^N, \quad z^1 \in \mathbb{R}^N. \end{cases}$$

ce qu'on réécrit sous la forme :

$$\begin{cases} z^{n+1} - z^n = z^n - z^{n-1} - \left(\dfrac{c\Delta t}{\Delta x}\right)^2 A_N z^n + (\Delta t)^2 \mu^n, \quad n \geq 0 \\ z_0^n = z_{N+1}^n = 0, \quad n \geq 0, \\ z^0 \in \mathbb{R}^N, \quad z^1 \in \mathbb{R}^N. \end{cases}$$

Soit $n \geq 0$. On en déduit :

$$(z^{n+1} - z^n) - (z^n - z^{n-1}) + \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N z^n = (\Delta t)^2 \mu^n. \tag{4.13}$$

On pose :

$$v^k = z^{k+1} - z^k, \quad \forall k \geq 0.$$

En multipliant les deux membres de (4.13) par $z^{n+1} - z^{n-1} = v^n + v^{n-1}$, on obtient :

$$\|v^n\|_2^2 - \|v^{n-1}\|_2^2 + \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N z^n \cdot (z^{n+1} - z^{n-1}) = (\Delta t)^2 \mu^n \cdot (z^{n+1} - z^{n-1})$$

i.e., compte tenu de la symétrie de A_N :

$$E^n := \|v^n\|_2^2 + \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N z^n \cdot z^{n+1} = \|v^{n-1}\|_2^2 + \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N z^{n-1} \cdot z^n + (\Delta t)^2 \mu^n \cdot (v^n + v^{n-1})$$

$$= \underbrace{\|v^0\|_2^2 + \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N z^0 \cdot z^1}_{=E^0} + (\Delta t)^2 \sum_{k=1}^{n} \mu^k \cdot (v^k + v^{k-1})$$

$$= E^0 + (\Delta t)^2 \mu^1 \cdot v^0 + (\Delta t)^2 \sum_{k=1}^{n} (\mu^{k+1} + \mu^k) \cdot v^k + (\Delta t)^2 \mu^{n+1} \cdot v^n.$$

On remarque que :

$$A_N z^n \cdot z^{n+1} = A_N z^n \cdot v^n + A_N z^n \cdot z^n$$

avec : $\forall \alpha > 0$:

$$|A_N z^n \cdot v^n| \leq \frac{\alpha}{2} \|A_N z^n\|_2^2 + \frac{1}{2\alpha} \|v^n\|_2^2.$$

Soit $\alpha > 0$. Il en résulte :

$$E^n \geq \left(1 - \frac{1}{2\alpha}\left(\frac{c\Delta t}{\Delta x}\right)^2\right) \|v^n\|_2^2 + \left(\frac{c\Delta t}{\Delta x}\right)^2 (1 - 2\alpha) A_N z^n \cdot z^n,$$

et on choisit $\alpha \in]0, \frac{1}{2}[$. Alors :

$$1 > \frac{1}{2\alpha}\left(\frac{c\Delta t}{\Delta x}\right)^2 > \left(\frac{c\Delta t}{\Delta x}\right)^2 \Rightarrow 0 < \left(\frac{c\Delta t}{\Delta x}\right)^2 < 1.$$

On remarque que

$$\min_{0<\alpha<\frac{1}{2}} \left(1 - \frac{1}{2\alpha}\left(\frac{c\Delta t}{\Delta x}\right)^2, 1-2\alpha\right)$$

atteint son maximum en $\alpha > 0$ t.q.

$$1 - \frac{1}{2\alpha}\left(\frac{c\Delta t}{\Delta x}\right)^2 = 1 - 2\alpha$$

i.e. en $\alpha = \alpha_0$ solution. de :

$$\frac{1}{2\alpha}\left(\frac{c\Delta t}{\Delta x}\right)^2 = 2\alpha \iff \alpha_0 = \frac{1}{2}\left(\frac{c\Delta t}{\Delta x}\right).$$

On en déduit :

$$E^n \geq \left(1 - \frac{c\Delta t}{\Delta x}\right)\left(\|v^n\|_2^2 + A_N z^n \cdot z^n\right) \tag{4.14}$$

En particulier, si $E^0 = 0$, alors :

$$\|v^n\|_2^2 \leq \frac{\sqrt{2}(\Delta t)^2}{\left(1 - \frac{c\Delta t}{\Delta x}\right)}(\sum_{k=0}^{n+1} \|\mu^k\|_2^2)^{\frac{1}{2}}(\sum_{k=0}^{n} \|v^k\|_2^2)^{\frac{1}{2}}, \quad \forall n \geq 0. \tag{4.15}$$

$$\Rightarrow (\sum_{k=0}^{n} \|v^k\|_2^2)^{\frac{1}{2}} \leq \frac{\sqrt{2}t_n\Delta t}{\left(1 - \frac{c\Delta t}{\Delta x}\right)}(\sum_{k=0}^{n+1} \|\mu^k\|_2^2)^{\frac{1}{2}}$$

On définit l'erreur de convergence au point (x_i, t_n), $1 \leq i \leq N$, $n \geq 0$, par :

$$e_i^n = u(x_i, t_n) - u_i^n, \quad i = 1, \cdots, N, \quad n \geq 0.$$

Par construction : $e_i^0 = u(x_i, 0) - u_0(x_i) = 0$, $i = 0, \cdots, N+1$. Pour $n = 1$, l'erreur de convergence au point (x_i, t_0) est associée à l'étape d'initialisation (4.10) par la relation : $e_i^1 = e_0(x_i)$ où : $\forall x \in [0, 1]$,

$$e_0(x) = u(x, \Delta t) - \left(1 - c\frac{\Delta t}{\Delta x}\right)u_0(x) - c\frac{\Delta t}{\Delta x}u_0(x - \Delta x) - \Delta t(u_1(x) + cu_0'(x)). \tag{4.16}$$

On a :

$$\left(1 - c\frac{\Delta t}{\Delta x}\right)u_0(x) + c\frac{\Delta t}{\Delta x}u_0(x - \Delta x) = u_0(x) + c\frac{\Delta t}{\Delta x}\left(-\Delta x u_0'(x) + O((\Delta x)^2)\right)$$

$$= u_0(x) - c\Delta t u_0'(x) + \underbrace{\Delta t O(\Delta x)}_{=O((\Delta t)^2+(\Delta x)^2)} ,$$

donc

$$e_0(x) = u(x,\Delta t) - u_0(x) + c\Delta t u_0'(x) - c\Delta t u_0'(x) - \Delta t u_1(x) + O((\Delta t)^2+(\Delta x)^2)$$

$$= u(x,\Delta t) - u(x,0) - \Delta t \frac{\partial u}{\partial t}(x,0) + O((\Delta t)^2 + (\Delta x)^2) = O((\Delta t)^2 + (\Delta x)^2).$$

i.e. :

$$e_i^1 = O((\Delta t)^2 + (\Delta x)^2).$$

Alors, pour le choix $\mu_i^n = \varepsilon_i^n$, $i = 1, \cdots, N$,$n \geq 1$, $z^0 = z^1 = 0$, on obtient $z^n = e^n$, $n \geq 2$. et (4.15) s'applique : $\forall n \geq 2$,

$$\|z^{n+1}\|_2 \leq \|\underbrace{z^1}_{=0}\|_2 + \sum_{k=0}^{n} \|\underbrace{z^{k+1} - z^k}_{=v^k}\|_2 \leq \sqrt{n} \left(\sum_{k=0}^{n} \|v^k\|_2^2 \right)^{\frac{1}{2}}$$

$$\underset{(4.12)}{\leq} \frac{\sqrt{2} t_n^{\frac{3}{2}} \sqrt{\Delta t}}{\left(1 - \dfrac{c\Delta t}{\Delta x}\right)} \left(\sum_{k=1}^{n+1} \|\mu^k\|_2^2\right)^{\frac{1}{2}}$$

En particulier : $\forall n \geq 2$,

$$\sup_{n\Delta t \leq T} \|e^n\|_2 \leq C \left(1 + \frac{\sqrt{2} T^2}{\left(1 - \dfrac{c\Delta t}{\Delta x}\right)} \right) ((\Delta t)^2 + (\Delta x)^2)$$

□

Remarque 17. Le choix $z^1 := 0$ est justifié par l'étape d'initialisation (4.10) qui conduit à la formule :

$$u_i^1 = \left(1 - \frac{c\Delta t}{\Delta x}\right) u_0(x_i) + \frac{c\Delta t}{\Delta x} u_0(x_{i-1}) + \Delta t (u_1(x_i -) c u_0'(x_i))$$

dont le membre de droite est combinaison linéaire de termes exacts.

Proposition 4.3.3. *Si $f = 0$ dans le schéma (4.11), alors l'énergie est conservée. Plus précisément, si on définit l'énergie discrète au temps t_n, $n \geq 0$, par :*

$$E^n := \|u^n - u^{n-1}\|_2^2 + \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N u^n \cdot u^{n-1},$$

alors : $E^{n+1} = E^n$, $\forall n \geq 0$.

Démonstration. Soit $n \geq 0$. Si $f = 0$ alors :

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

Après multiplication des deux membres de cette égalité par $u^{n+1} - u^{n-1} = (u^{n+1} - u^n) + (u^n - u^{n-1})$ on obtient :

$$\|u^{n+1} - u^n\|^2 - \|u^n - u^{n-1}\|^2 + \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N u^n \cdot (u^{n+1} - u^{n-1}) = 0.$$

On conclut grâce à la symétrie de A_N. □

Remarque 18. Avec les notations de la Proposition 4.3.3, si on pose $v^n = u^{n+1} - u^n$, alors

$$E^n = \|v^n\|_2^2 + \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N v^n \cdot u^{n-1} + \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N u^{n-1} \cdot u^{n-1}$$

et le même raisonnement que pour (4.14) conduit à :

$$E^n \geq \left(1 - \frac{c\Delta t}{\Delta x}\right)\left(\|v^n\|_2^2 + A_N u^n \cdot u^n\right).$$

Proposition 4.3.4. *Soit $N > 0$ et soit $\Delta x = \frac{2\pi}{(N+1)}$. Le schéma (4.10)–(4.11) est stable au sens de Von Neumann ssi $\frac{c\Delta t}{\Delta x} < 1$. Si de plus*

$$u_j^0 = \sum_{k\in\mathbb{Z}} c_k^0 e^{ikj\Delta x} \quad et \quad u_j^1 = \sum_{k\in\mathbb{Z}} c_k^1 e^{ijk\Delta x}$$

alors :

$$u_j^n = \sum_{k\in\mathbb{Z}} c_k^n e^{ikj\Delta x} \tag{4.17}$$

avec

$$c_k^n = a_k e^{in\theta_k} + b_k e^{-in\theta_k}, \quad \forall k \in \mathbb{Z}, \quad n \geq 0 \tag{4.18}$$

où les constantes a_k, b_k, θ_k ne dépendent que de Δt, Δx, et $k \in \mathbb{Z}$, et où

$$\theta_k \underset{(\Delta t,\Delta x)\to(0,0)}{\sim} ckt_n, \quad \forall n \geq 0.$$

Démonstration. On vérifie directement que

$$c_k^{n+1} - (2 - \alpha(k)^2)c_k^n + c_k^{n-1} = 0 \tag{4.19}$$

où :

$$\alpha(k) = 2\frac{c\Delta t}{\Delta x}\sin\left(\frac{k\Delta x}{2}\right) \underset{(\Delta t,\Delta x)\to(0,0)}{\sim} ck\Delta t.$$

L'équation caractéristique associée à (4.19) s'écrit :

$$r^2 + (\alpha(k)^2 - 2)r + 1 = 0$$

et admet pour discriminant

$$\Delta = (\alpha(k)^2 - 2)^2 - 4 < 0.$$

On pose :

$$\alpha(k)^2 - 2 = 2\cos\theta_k$$

et alors on trouve que les valeurs propres de $S(k)$ sont :

$$\pm e^{\pm i\theta_k}.$$

On remarque aussi que :

$$\alpha(k)^2 = 4\left(\cos\left(\frac{\theta_k}{2}\right)\right)^2$$

ce qui est licite ssi $c\frac{\Delta t}{\Delta x} < 1$. Les valeurs propres de $S(k)$ s'écrivent : $e^{\pm i\theta_k}$ avec

$$e^{\pm i\theta_k} = 1 - \frac{\alpha(k)^2}{2} \pm \sqrt{1 - \left(1 - \frac{\alpha(k)^2}{2}\right)^2}.$$

On en déduit :

$$\tan\theta_k = \frac{\sqrt{1 - \left(1 - \frac{\alpha(k)^2}{2}\right)^2}}{1 - \frac{\alpha(k)^2}{2}} \underset{(\Delta t,\Delta x)\to(0,0)}{\sim} \alpha(k) \underset{(\Delta t,\Delta x)\to(0,0)}{\to} 0$$

donc

$$\theta_k \underset{(\Delta t,\Delta x)\to(0,0)}{\sim} \alpha(k).$$

Finalement, on en déduit (4.18) avec :

$$n\alpha(k) \underset{(\Delta t,\Delta x)\to(0,0)}{\sim} ckn\Delta t = ckt_n.$$

i.e. : $\forall n \geq 1$, $\forall k \in \mathbb{Z}$,

$$|c_k^n e^{ijk\Delta x} - a_k e^{ik(j\Delta x + ct_n)} - b_k e^{ik(j\Delta x - ct_n)}| \leq |a_k||e^{in\theta_k} - e^{ikct_n}| + |b_k||e^{-in\theta_k} - e^{-ikct_n}|$$

$$= 2(|a_k| + |b_k|)\left|\sin\left(\frac{n\theta_k - ckt_n}{2}\right)\right| \underset{(\Delta t,\Delta x)\to(0,0)}{\to} 0.$$

□

Corollaire 4.3.5. *Sous les hypothèses de la Proposition 4.3.4, si en outre :*

$$\sum_{k\in\mathbb{Z}} |c_k^0|^2 < +\infty \quad et \quad \sum_{k\in\mathbb{Z}} |c_k^1|^2 < +\infty$$

alors la série de fonctions (4.17)–(4.18) converge dans $\mathcal{C}^0([0,T], L^2(0,1))$ *vers la somme de la série :*

$$\sum_{k\in\mathbb{Z}} \frac{1}{2i\sin(\theta_k)} \left((c_k^1 - e^{-i\theta_k}c_k^0)e^{ik(j\Delta x+ct_n)} + (e^{i\theta_k}c_k^0 - c_k^1)e^{ik(j\Delta x-ct_n)}\right)$$

solution de (4.2).

Démonstration. Des relatons :

$$\begin{cases} c_k^0 = a_k + b_k, \\ c_k^1 = a_k e^{ik\theta_k} + b_k e^{-ik\theta_k}, \end{cases}$$

on déduit directement que :

$$a_k = \frac{1}{2i\sin(\theta_k)}(c_k^1 - e^{-i\theta_k}c_k^0), \quad b_k = \frac{1}{2i\sin(\theta_k)}(e^{i\theta_k}c_k^0 - c_k^1)$$

□

Schéma centré

Soit $\theta \in [0,1]$ et soit $(u_i^n)_{0\le i\le N+1, n\ge 0}$ la suite solution du schéma centré :

$$\begin{cases} \dfrac{u^{n-1} - 2u^n + u^{n+1}}{\Delta t^2} + \dfrac{c^2}{(\Delta x)^2} A_N \left(\dfrac{\theta}{2}(u^{n+1} + u^{n-1}) + (1-\theta)u^n\right) = f^N, \\ \\ u_0^n = u_{N+1}^n = 0, \quad i = 1\cdots, N, \quad n \ge 1 \end{cases} \tag{4.20}$$

initialisé par (4.10), où $f^N \in \mathbb{R}^N$ est le vecteur de composantes $f_i^N = f(x_i)$, $i = 1, \cdots, N$,

Proposition 4.3.6. *Le schéma centré (4.10), (4.20) est consistant d'ordre* 2 *en temps et en espace, pour tout* $\theta \in [0,1]$.

Démonstration. On définit l'erreur de consistance au point (x_i, t_n) par : $\varepsilon_i^n = \varepsilon_0(x_i)$ si $n = 0$ avec ε_0 définie par (4.16), et par $\varepsilon_i^n = \varepsilon(x_i, t_n)$ si $n \ge 1$ avec : $\forall x \in [0,1]$, $\forall t > 0$,

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

$$+\frac{c^2}{(\Delta x)^2}A_N\left(\frac{\theta}{2}(u(x,t+\Delta t)+u(x,t-\Delta t))+(1-\theta)u(x,t)\right)-f(x)$$

La formule de Taylor donne :

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

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

On conclut comme dans la Proposition 4.3.1. □

Proposition 4.3.7. *Soit $\theta\in[0,1[$. On suppose que*

$$\frac{c\Delta t}{\Delta x}<\frac{1}{\sqrt{1-\theta}}. \tag{4.21}$$

Alors, le schéma centré (4.10), (4.20) est convergent d'ordre 2 en temps et en espace, pour tout $\theta\in[0,1]$.

Démonstration. Soit $(\mu_i^n)_{n\geq 0}$ une suite de réels et soit z_i^n, $n\geq 0$, $i=1,\cdots,N$, définie par :

$$\begin{cases}\dfrac{z^{n-1}-2z^n+z^{n+1}}{\Delta t^2}+\dfrac{c^2}{(\Delta x)^2}A_N\left(\dfrac{\theta}{2}(z^{n+1}+z^{n-1})+(1-\theta)z^n\right)=\mu^n,\\ \\ z_0^n=z_{N+1}^n=0,\quad n\geq 0,\\ \\ z^0\in\mathbb{R}^N,\quad z^1\in\mathbb{R}^N.\end{cases}$$

Soit $n\geq 1$. On pose :

$$v^n=z^{n+1}-z^n.$$

On a :

$$E^n:=\|v^n\|_2^2+\frac{\theta}{2}\left(\frac{c\Delta t}{\Delta x}\right)^2A_Nv^n\cdot v^n+(1-\theta)\left(\frac{c\Delta t}{\Delta x}\right)^2A_Nz^n\cdot z^{n+1}=E^{n-1}+\mu^n\cdot(v^n+v^{n-1})$$

et

$$E^n\geq\left(1-\sqrt{1-\theta}\frac{c\Delta t}{\Delta x}\right)\left(\|v^n\|_2^2+A_Nz^n\cdot z^n\right)$$

On conclut comme dans la Proposition 4.3.2 avec $\frac{c\Delta t}{\Delta x}$ par $\sqrt{1-\theta}\frac{c\Delta t}{\Delta x}$ □

Schéma implicite non centré

Soit $(u_i^n)_{0\leq i\leq N+1, n\geq 0}$ la suite solution du schéma implicite non centré :

$$\begin{cases} \dfrac{u_i^{n-1}-2u_i^n+u_i^{n+1}}{\Delta t^2}+\dfrac{c^2}{(\Delta x)^2}(A_N u^{n+1})_i = f(x_i), \\ \\ u_0^n = u_{N+1}^n = 0, \quad i = 1\cdots, N, \quad n\geq 1 \end{cases} \tag{4.22}$$

initialisé par (4.10).

Proposition 4.3.8. *Alors Le schéma implicite non centré (4.10), (4.22) est consistant d'ordre* 1 *en temps et d'ordre* 2 *en espace.*

Démonstration. On définit l'erreur de consistance au point $(x_i t_n)$ par :

$$\varepsilon_i^n = \begin{cases} \varepsilon(x_i,t_n), \quad i=1,\cdots,N & \text{si} \quad n\geq 1, \\ \varepsilon^0(x_i), \quad i=1,\cdots,N & \text{si} \quad n=0, \end{cases}$$

où ε^0, est définie par (4.16), et où : $\forall x\in[0,1]$, $\forall t>0$,

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

$$+\left(\frac{c}{\Delta x}\right)^2(-u(x-\Delta x,t+\Delta t)+2u(x,t+\Delta t)-u(x+\Delta x,t+\Delta t))-f(x).$$

La formule de Taylor donne :

$$\begin{aligned}\varepsilon(x,t) &= \frac{\partial^2 u}{\partial t^2}(x,t)+O((\Delta t)^2)-c^2\frac{\partial^2 u}{\partial x^2}(x,t+\Delta t)+O((\Delta x)^2)-f(x) \\ &= O(\Delta t)+O(\Delta x)^2.\end{aligned}$$

□

Proposition 4.3.9. *Le schéma implicite non centré (4.10), (4.22) est convergeant d'ordre* 1 *en temps et d'ordre* 2 *en espace.*

Démonstration. Soit $(\mu_i^n)_{n\geq 0}$ une suite de réels et soit z_i^n, $n\geq 0$, $i=1,\cdots,N$, définie par :

$$\begin{cases} \dfrac{1}{(\Delta x)^2}(z^{n-1}-2z^n+z^{n+1})+\dfrac{c^2}{(\Delta x)^2}A_N z^{n+1} = \mu^n, \\ \\ z_0^n = z_{N+1}^n = 0, \quad n\geq 0, \\ \\ z^0\in\mathbb{R}^N, \quad z^1\in\mathbb{R}^N. \end{cases}$$

$$\begin{cases} \dfrac{z_i^{n-1} - 2z_i^n + z_i^{n+1}}{\Delta t^2} + \dfrac{c^2}{(\Delta x)^2}(A_N z^{n+1})_i = \mu_i^n, \\ \\ z_0^n = z_{N+1}^n = 0, \quad n \geq 0, \\ \\ z^0 \in \mathbb{R}^N, \quad z^1 \in \mathbb{R}^N. \end{cases}$$

On pose

$$v^n = z^{n+1} - z^n,$$

Alors :

$$E^n := \|v^n\|_2^2 + \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N z^{n+1} \cdot z^{n+1} = E^{n-1} - \|v^n - v^{n-1}\|_2^2 - \left(\frac{c\Delta t}{\Delta x}\right)^2 A_N v^n \cdot v^n + 2\mu^n \cdot v^n$$

$$\leq E^{n-1} + 2\mu^n \cdot v^n \leq E^0 + 2\sum_{k=0}^{n} \mu^k \cdot v^k.$$

On en déduit : si $E^0 = 0$,

$$\|v^n\|_2^2 \leq E^0 + 2\left(\sum_{k=0}^{n} \|\mu^k\|_2^2\right)^{\frac{1}{2}} \left(\sum_{k=0}^{n} \|v^k\|_2^2\right)^{\frac{1}{2}}$$

$$\Rightarrow \left(\sum_{k=0}^{n} \|v^k\|_2^2\right)^{\frac{1}{2}} \leq 2\left(\sum_{k=0}^{n} \|\mu^k\|_2^2\right)^{\frac{1}{2}}$$

On conclut comme dans la Proposition 4.3.1 compte tenu de la Proposition 4.3.8. □

Equation des ondes à deux dimensions d'espace

On considère le problème : trouver $u :]0,1[\times]0,1[\times]0,+\infty[\to \mathbb{R}$ solution de :

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} - c^2\left(\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2}\right) = f(x), \quad (x,y) \in]0,1[^2, \quad t > 0, \\ \\ u(x,y,0) = u_0(x,y), \quad \dfrac{\partial u}{\partial t}(x,y,0) = u_1(x,y), \quad (x,y) \in]0,1[^2, \\ \\ u(x,y,t) = 0 \quad \text{si} \quad x, y \in \{0,1\} \end{cases}$$

L'approximation numérique es basée sur les subdivisions en espace,

$$0 = x_0 < x_1 < \cdots < x_I < x_{I+1} = 1, \quad 0 = y_0 < y_1 < \cdots < y_J < y_{J+1} = 1$$

resp. en temps :

$$0 = t_0 < \cdots < t_n < \cdots,$$

supposées régulières i.e. définies par :

$$x_i = i\Delta x, \quad y_j = j\Delta y, \quad t_n = n\Delta t, \quad i = 0, \cdots, I{+}1, \quad j = 0, \cdots, J{+}1, \quad n \geq 0.$$

Soit $\theta \in [0,1]$. Avec ces notations on définit la suite $(u^n_{i,j})_{n\geq 0}$ par :

$$\begin{cases} \dfrac{u^n - 2u^n + u^{n-1}}{(\Delta t)^2} + c^2 \left(\dfrac{1}{(\Delta x)^2} A_I + \dfrac{1}{(\Delta y)^2} A_J \right) \left(\dfrac{\theta}{2}(u^{n+1} + u^{n-1}) + (1-\theta)u^n \right) = f^{I,J}, \\ u^n_{0,j} = u^n_{I+1,j} = u^n_{i,0} = u^n_{i,J+1} = 0, \\ u^0_{ij} = u_0(x_i, y_j), \quad u^1 \in \mathbb{R}^{(I+1)(J+1)} \end{cases} \tag{4.23}$$

où $f^{I,J} \in \mathbb{R}^{IJ}$ est définie par :

$$f^{I,J}_{i,j} = f^{I,J}(x_i, y_j), \quad i = 1, \cdots, I, \quad J = 1, \cdots, J,$$

et où les opérateurs A_I, A_J sont définis par :

$$(A_I U)_{i,j} = -U_{i+1,j} + 2U_{i,j} - U_{i-1,j}, \quad i = 1, \cdots, I, \quad j =, \cdots, J.$$

$$(A_J U)_{i,j} = -U_{i,j+1} + 2U_{i,j} - U_{i,j-1}, \quad i = 1, \cdots, I, \quad j =, \cdots, J.$$

Proposition 4.3.10. *On suppose que* $0 \leq \theta < 1$ *et que*

$$(c\Delta t)^2 \left(\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} \right) < \frac{1}{1-\theta}.$$

Si l'étape d'initialisation donnant u^1 *est consistante d'ordre* 2 *en temps et en espace, alors le schéma (4.23) est convergent d'ordre* 2 *en temps et en espace.*

Bibliographie

[1] Robert Dautray, Jacques-Louis Lions. Analyse mathématique et calcul numérique pour les sciences et les techniques, Tome 3, Masson, Paris, 1985, et Volumes 7 à 9, Masson, Paris, 1988. *Chapitres XIV.3, XV.4 et XX.3.*

Table des matières

Table des figures

Liste des tableaux

Bibliographie

[1] Thierry Gallouët, Raphaèle Herbin. Analyse numérique des équations aux dérivées partielles. Master. Marseille, France. 2011. (https ://cel.hal.science/cel-00637008v2) *Chapitre 5.*

[2] Lionel Sainsaulieu. Calcul Scientifique. Cours et exercices corrigés pour le second cycle et les écoles d'ingénieurs, Masson, Paris 1996. *Chapitres 1.4 et 2.3.*

[3] Brigitte Lucquin, Olivier Pironneau. Introduction au calcul scientifique. Masson, Paris, 1996. *Chapitres I.6 et VII.4*

[4] Alfio Quarteroni, Riccardo Sacco, Fausto Saleri Numerical Mathematics. Springer, Berlin, 2007. *Chapitres 13.5 à 13.8*

[5] Alfio Quarteroni, Riccardo Sacco, Paola Gervasio Calcul scientifique. Springer, Milan, 2010. *Chapitres 8.3.1 et 8.3.2*