Equations paraboliques

En guise d'introduction aux équations paraboliques, on se propose d'étudier théoriquement et numériquement dans ce chapitre l'équation de la chaleur posée en une dimension d'espace et avec des conditions de bord de Dirichlet homogènes. Ce problème s'écrit

$$(P)\quad \begin{cases} \partial_t u - \partial^2_{xx} u = f & \text{dans }]0,1[\times]0,T] \\ u(x=0,t) = u(x=1,t) = 0 & \forall\, t \in]0,T] \\ u(x,t=0) = u_0(x) & \forall\, x \in]0,1[\end{cases}$$

où $u : [0,1]\times[0,T] \to \mathbb{R}$ est l'inconnue, u_0 la donnée initiale du problème, f un terme source donné et T un réel supposé strictement positif (le T final souhaité).

Commençons par les aspects théoriques et montrons tout d'abord que ce problème ne peut admettre qu'une solution régulière au plus. On montrera ultérieurement qu'une telle solution existe effectivement (on le supposera donc pour le moment).

Proposition

On suppose $u_0 \in L^2(]0,1[)$ et $f \in L^2(]0,T[\times]0,1[)$. Soit u une solution suffisamment régulière de (P). Alors

$$\sup_{t\in[0,T]} \|u(\cdot,t)\|_{L^2(]0,1[)} \leq \|u_0\|_{L^2(]0,1[)} + \|f\|_{L^2(]0,T[\times]0,1[)}$$

Démonstration

Notons v une solution suffisamment régulière du problème (P) avec $f=0$, et w la fonction $u-v$. Par linéarité du problème (P), il est ainsi clair que w est une solution suffisamment régulière du problème (P) avec $u_0 = 0$. En d'autres termes

$$(P') \begin{cases} \partial_t v - \partial_{xx}^2 v = 0 & \text{dans }]0,1[\times]0,T] \\ v(0,t) = v(1,t) = 0 & \forall t \in]0,T] \\ v(x,0) = u_0(x) & \forall x \in]0,1[\end{cases}$$

et

$$(P'') \begin{cases} \partial_t w - \partial_{xx}^2 w = f & \text{dans }]0,1[\times]0,T] \\ w(0,t) = w(1,t) = 0 & \forall t \in]0,T] \\ w(x,0) = 0 \end{cases}$$

La décomposition $u = v + w$ nous permet de séparer l'influence du terme source f et de la condition initiale u_0.

En multipliant (P') par v et en intégrant par partie selon la variable x, nous obtenons successivement

$$\partial_t v - \partial_{xx}^2 v = 0 \Rightarrow v\partial_t v - v\partial_{xx}^2 v = 0 \Rightarrow \int_0^1 \partial_t \frac{v^2}{2}(x,t)\,dx - \int_0^1 v\partial_{xx}^2 v\,(x,t)\,dx = 0$$

$$\Rightarrow \frac{1}{2}\frac{d}{dt}\int_0^1 v^2(x,t)\,dx + \int_0^1 (\partial_x v)^2(x,t)\,dx - \left[v(x,t)\partial_x v(x,t)\right]_0^1 = 0$$

$$\Rightarrow \frac{1}{2}\frac{d}{dt}\int_0^1 v^2(x,t)\,dx + \int_0^1 (\partial_x v)^2(x,t)\,dx = 0 \quad \text{car } v(0,t) = v(1,t) = 0,$$

$$\Rightarrow \frac{1}{2}\frac{d}{dt}\int_0^1 v^2(x,t)\,dx \leq 0, \text{ c'est-à-dire } \frac{d}{dt}\|v(.,t)\|^2_{L^2(]0,1[)} \leq 0.$$

de sorte que $\frac{d}{dt}\int_0^1 v^2(x,t)\,dx \leq 0$, c'est-à-dire $\frac{d}{dt}\|v(.,t)\|^2_{L^2(]0,1[)} \leq 0$.

La fonction $t \to \|v(.,t)\|^2_{L^2(]0,1[)}$ est donc décroissante. Comme $v(.,0) = u_0(.)$, on a donc

$$\|v(.,t)\|^2_{L^2(]0,1[)} \leq \|u_0\|^2_{L^2(]0,1[)}$$

En procédant maintenant de la même manière avec (P''), on arrive facilement à

$$\frac{1}{2}\frac{d}{dt}\int_0^1 w^2(x,t)\,dx + \int_0^1 (\partial_x w)^2(x,t)\,dx = \int_0^1 f w\,(x,t)\,dx.$$

L'inégalité de Cauchy-Schwartz implique alors

$$\frac{1}{2}\frac{d}{dt}\int_0^1 w^2(x,t)\,dx + \|\partial_x w(.,t)\|^2_{L^2(0,1)} \leq \|f(.,t)\|_{L^2(0,1)} \|w(.,t)\|_{L^2(0,1)},$$

mais également, puisque $w(x,t) = \int_0^x \partial_x w(s,t)\,ds$ et donc

$$|w(x,t)| = \int_0^x |\partial_x w(s,t)|\,ds \quad , \quad \text{que}$$

$$|w(x,t)|^2 \leq x \int_0^x |\partial_x w(s,t)|^2 ds \leq x \int_0^1 |\partial_x w(s,t)|^2 ds = x \|\partial_x w(\cdot,t)\|^2_{L^2_{(0,1)}}$$

ce qui implique après intégration en x

$$\|w(\cdot,t)\|^2_{L^2(0,1)} \leq \frac{1}{2} \|\partial_x w(\cdot,t)\|^2_{L^2(0,1)} \leq \|\partial_x w(\cdot,t)\|^2_{L^2(0,1)}.$$

Cette inégalité, qui n'est rien d'autre que l'inégalité de Poincaré, permet alors d'obtenir l'inégalité suivante

$$\frac{1}{2}\frac{d}{dt}\int_0^1 w^2(x,t)\,dx + \|\partial_x w(\cdot,t)\|^2_{L^2(0,1)} \leq \|f(\cdot,t)\|_{L^2(0,1)} \|\partial_x w(\cdot,t)\|_{L^2(0,1)}$$

Ainsi, puisque $ab \leq \frac{(a^2+b^2)}{2}$, on obtient

$$\frac{1}{2}\frac{d}{dt}\int_0^1 w^2(x,t)\,dx + \|\partial_x w(\cdot,t)\|^2_{L^2(0,1)} \leq \frac{1}{2}\left(\|f(\cdot,t)\|^2_{L^2(0,1)} + \|\partial_x w(\cdot,t)\|^2_{L^2(0,1)}\right)$$

ou encore

$$\frac{d}{dt}\int_0^1 w^2(x,t)\,dx \leq \frac{d}{dt}\int_0^1 w^2(x,t)\,dx + \|\partial_x w(\cdot,t)\|^2_{L^2(0,1)} \leq \|f(\cdot,t)\|^2_{L^2(0,1)}$$

En intégrant en temps, on a donc

$$\int_0^1 w^2(x,t)\,dx - \int_0^1 w^2(x,0)\,dx \leq \int_0^t \|f(s,t)\|^2_{L^2(0,1)}\,ds \leq \int_0^T \|f(s,t)\|^2_{L^2(0,1)}\,ds$$

c'est-à-dire, puisque $w(x,0) = 0$,

$$\|w(\cdot,t)\|_{L^2(0,1)} \leq \|f\|_{L^2(]0,1[\times]0,T[)}$$

Finalement, on a donc

$$\|u(\cdot,t)\| \leq \|v(\cdot,t)\|_{L^2(0,1)} + \|w(\cdot,t)\|_{L^2(0,1)}$$

$$\|u(\cdot,t)\| \leq \|u_0\|_{L^2(0,1)} + \|f\|_{L^2(]0,1[\times]0,T[)}$$

ce qui conclut la démonstration de la proposition.

Cette proposition nous permet d'obtenir facilement le résultat d'unicité attendu. En effet, si on suppose que u_1 et u_2 sont deux solutions suffisamment régulières de (P), alors $u_1 - u_2$ vérifie le problème (P) avec $u_0 = f = 0$ de sorte que la proposition permet d'obtenir $\sup_{t\in[0,T]} \|(u_1-u_2)(\cdot,t)\|_{L^2(0,1)} \leq 0$, ce qui implique nécessairement $u_1 = u_2$, d'où l'unicité.

Cette proposition nous permet également d'obtenir une notion de stabilité par rapport aux données u_0 et f de la solution. En effet, si les données u_0 et f sont légèrement perturbées pour être remplacées par u_0^ε et f^ε telles que

$$\|u_0 - u_0^\varepsilon\|_{L^2(0,1)} \leq \varepsilon \quad \text{et} \quad \|f - f^\varepsilon\|_{L^2(]0,1[\times]0,T[)} \leq \varepsilon$$

alors, le problème (P) étant linéaire, on obtient facilement d'après la proposition que

$$\sup_{t\in[0,T]} \|(u-u^\varepsilon)(\cdot,t)\|_{L^2(0,1)} \leq \varepsilon + \varepsilon = 2\varepsilon$$

où u et u^ε représentent respectivement les solutions du problème (P) avec les données (u_0, f) et $(u_0^\varepsilon, f^\varepsilon)$. Ainsi, une perturbation d'ordre ε sur les données induit une perturbation du même ordre sur la solution.

Remarque

La quantité $E(t) = \frac{1}{2}\|u(\cdot,t)\|^2_{L^2(0,1)}$ est appelée énergie du système. On remarque alors que l'énergie du système décroit en l'absence de terme source ($f=0$) et est donc contrôlée par l'énergie initiale dans ce cas.

On s'intéresse maintenant à l'existence d'une solution au problème (P) que l'on va démontrer par une méthode constructive. Les outils mathématiques utilisés étant du type séries de Fourier, nous faisons les brefs rappels suivants.

Rappels sur les séries de Fourier

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$ par morceaux et périodique de période 2π. On appelle série de Fourier trigonométrique de f la série $a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$ avec

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\,dx, \quad a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx,$$

qui peut s'écrire aussi $\sum_{n=-\infty}^{+\infty} c_n e^{inx}$ avec $c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx$.

Si f est une fonction paire (resp. impaire), alors la série de Fourier ne comporte que des cosinus (resp. sinus).

Si f est seulement continue sur $[-\pi, \pi]$, on a alors la convergence quadratique de la série de Fourier de f vers f, ie

$$\lim_{N \to +\infty} \int_{-\pi}^{\pi} \left(f(x) - a_0 - \sum_{n=1}^{N} (a_n \cos nx + b_n \sin nx) \right)^2 dx = 0$$

Il s'agit d'une convergence dans $L^2(-\pi, \pi)$.

Si f est une fonction $\mathcal{C}^1$ par morceaux, on a alors la convergence simple, et même uniforme* de la série de Fourier de f vers f sur $[-\pi, \pi]$. Les hypothèses sur f peuvent être affaiblies.

*sur tout intervalle fermé ne contenant pas de point de discontinuité de f'.

Rappelons enfin le résultat de dérivation de la somme d'une série de fonction, qui nous sera utile par la suite.

Si I est un intervalle de $\mathbb{R}$ et $(f_n)_n$ une suite d'applications définies sur I et telles que

- f_n est de classe $\mathcal{C}^1$ sur I
- la série $\sum_{n\geq 0} f_n$ converge simplement sur I
- la série $\sum_{n\geq 0} f_n'$ converge uniformément sur tout $[a,b]\subset I$

alors la fonction $\sum_{n\geq 0} f_n$ est de classe $\mathcal{C}^1$ sur I et $\left(\sum_{n\geq 0} f_n\right)' = \sum_{n\geq 0} f_n'$

Pour conclure ces rappels, nous rappelons la relation de Parseval que l'on obtient au moins formellement de la façon suivante :

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

$$\Rightarrow f^2(x) = a_0 f(x) + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) f(x)$$

$$\Rightarrow \int_{-\pi}^{\pi} f^2(x)\,dx = a_0 \int_{-\pi}^{\pi} f(x)\,dx + \sum_{n=1}^{\infty} a_n \int_{-\pi}^{\pi} f(x)\cos nx\,dx + \sum_{n=1}^{\infty} b_n \int_{-\pi}^{\pi} f(x)\sin nx\,dx$$

$$\Rightarrow \int_{-\pi}^{\pi} f^2(x)\,dx = 2\pi a_0^2 + \sum_{n=1}^{\infty} \pi(a_n^2 + b_n^2)$$

$$\Rightarrow \|f\|^2_{L^2(-\pi,\pi)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f^2(x)\,dx = a_0^2 + \frac{1}{2}\sum_{n=1}^{\infty} a_n^2 + b_n^2 .$$

Cette égalité n'est rien d'autre que la relation de Parseval.

Passons maintenant à la construction d'une solution régulière du problème (P). On suppose pour cela que la donnée initiale u_0 est telle que $u_0 \in \mathcal{C}^2([0,1])$ et $u_0(0) = u(1) = 0$. Avec un léger abus de notation, on appellera également u_0 le prolongement de u_0 par imparité sur $[-1,0]$, puis par périodicité (de période 2) sur $\mathbb{R}$. La fonction u_0 ainsi obtenue est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et est périodique de période 2. Notons que la fonction $x \to u_0\left(\frac{2}{T}x\right)$

est de période T et de classe $\mathcal{C}^1$ de sorte que les rappels sur les séries de Fourier peuvent lui être appliqués avec $T = 2\pi$.

Des calculs simples montrent alors que la série trigonométrique de u_0 est $a_0 + \sum_{n=1}^{\infty} (a_n \cos n\pi x + b_n \sin n\pi x)$ avec

$$a_0 = \frac{1}{2}\int_{-1}^{1} u_0(x)\,dx\,, \quad a_n = \int_{-1}^{1} u_0(x)\cos(n\pi x)\,dx\,, \quad b_n = \int_{-1}^{1} u_0(x)\sin(n\pi x)\,dx$$

La fonction étant impaire, on a même $a_0 = a_n = 0$ $\forall n$ de sorte que la série de Fourier ne comporte que des sinus :

$$u_0(x) = \sum_{n=1}^{\infty} b_n \sin(n\pi x) \qquad \forall x \in \mathbb{R}\,,$$

avec $$b_n = \int_{-1}^{1} u_0(x)\sin(n\pi x) = 2\int_0^1 u_0(x)\sin(n\pi x)\,dx\,.$$

Remarquons également que par intégration par partie et en utilisant les conditions aux limites $u_0(0) = u_0(1) = 0$, on a

$$b_n = \frac{2}{n\pi}\int_0^1 u_0'(x)\cos(n\pi x)\,dx\,.$$

Une deuxième intégration par partie nous donne alors, puisque $\sin(n\pi) = \sin 0 = 0$

$$b_n = -\frac{2}{(n\pi)^2}\int_0^1 u_0''(x)\sin(n\pi x)\,dx\,.$$

On en déduit donc que pour tout $n \geq 1$

$$|b_n| \leq \frac{2}{(n\pi)^2}\sup_{x\in[0,1]} |u_0''(x)|\,.$$

Supposons que comme u_0, f soit de classe $\mathcal{C}^2$ (ici sur $[0,1]\times[0,T]$) et que $f(t,0) = f(t,1) = 0$, $\forall t \in [0,T]$. Alors, f admet un développement en série de Fourier trigonométrique en espace de la forme

$$f(t,x) = \sum_{n=1}^{\infty} b_n^f(t) \sin(n\pi x) \qquad \forall x \in \mathbb{R},\ \forall t \in [0,1]$$

avec $$b_n^f(t) = \int_{-1}^{1} f(t,x) \sin(n\pi x)\, dx = 2\int_0^1 f(t,x) \sin(n\pi x)\, dx$$

et $$|b_n^f(t)| \leq \frac{2}{(n\pi)^2} \sup_{x\in[0,1]} |\partial_{xx} f(t,x)|, \quad \forall t \in [0,1].$$

Considérons alors la fonction

$$u(x,t) = \sum_{n\geq 1} \left(b_n + \int_0^t b_n^f(s)\, e^{n^2\pi^2 s}\, ds \right) \sin(n\pi x) \exp(-n^2\pi^2 t)$$

Les hypothèses faites sur u_0 et f, et notamment les majorations sur les coefficients $|b_n|$ et $|b_n^f(t)|$ permettent de montrer rigoureusement que cette fonction est $\mathcal{C}^0([0,1]\times[0,T])$, et que $t \to u(t,x) \in \mathcal{C}^1([0,T])$ $\forall x \in [0,1]$, $x \to u(t,x) \in \mathcal{C}^2([0,1])$ $\forall t \in]0,T]$. Par ailleurs, on montre que les fonctions $\partial_t u$ et $\partial_{xx} u$ sont obtenues en dérivant terme à terme la série de u.

On a donc

$$\partial_t u(x,t) = \sum_{n\geq 1} -n^2\pi^2 \left(b_n + \int_0^t b_n^f(s)\, e^{n^2\pi^2 s}\, ds \right) \sin(n\pi x) \exp(-n^2\pi^2 t)$$
$$+ \sum_{n\geq 1} b_n^f(t) \sin(n\pi x)$$

$$\partial_{xx} u(x,t) = -\sum_{n\geq 1} n^2\pi^2 \left(b_n + \int_0^t b_n^f(s)\, e^{n^2\pi^2 s}\, ds \right) \sin(n\pi x) \exp(-n^2\pi^2 t)$$

On a donc

$$\partial_t u(x,t) - \partial_{xx} u(x,t) = f(t,x), \quad \forall t \in]0,T], \ \forall x \in [0,1].$$

Par ailleurs

$$u(x, t=0) = \sum_{n \geq 1} b_n \sin(n\pi x) = u_0(x) \qquad \forall x \in [0,1]$$

et

$$u(x=0,t) = u(x=1,t) = 0 \quad \text{car } \sin 0 = \sin(n\pi) = 0, \ \forall t \in]0,T].$$

La fonction u est donc solution du problème (P).

Approximation numérique par la méthode des différences finies.

Nous poursuivons notre étude du problème (P) par la méthode des différences finies qui permet de calculer de manière approchée la solution de ce problème sur une grille régulière en temps et en espace du domaine $[0,1]\times[0,T]$. Cette grille régulière, ou maillage, sera constituée d'un certain nombre de points $(x_j, t^n) \in [0,1]\times[0,T]$ régulièrement espacés. Plus précisément, on introduit un pas de temps Δt et un pas d'espace Δx tels que

$$\Delta t = \frac{T}{N}, \quad \Delta x = \frac{1}{J+1}$$

où N et $J+1$ représentent le nombre d'intervalles utilisés pour découper $[0,T]$ et $[0,1]$. On note alors

$$x_j = j\,\Delta x \quad \text{et} \quad t^n = n\,\Delta t$$

$\forall j = 0, \dots, J+1$, $\forall n = 0, \dots, N$

Le but de la méthode des différences finies que nous allons considérer est donc de calculer une approximation notée u_j^n de la valeur

exacte $u(x_j, t^n)$ de la solution du problème (P) au point (x_j, t^n), ceci $\forall j = 0, \dots, J+1$ et $\forall n = 0, \dots, N$. Les conditions aux limites du problème nous conduisent naturellement à poser

$$(CL) \quad \begin{cases} u_0^n = 0 & \forall n = 0, \dots, N \\ u_{J+1}^n = 0 & \forall n = 0, \dots, N \\ u_j^0 = u_0(x_j) & \forall j = 0, \dots, J+1 \end{cases}$$

en supposant (pour simplifier) que la donnée initiale u_0 vérifie les conditions de bord $u_0(0) = u_0(1) = 0$. Rappelons que u_0 est une donnée du problème.

Il reste donc à définir les valeurs u_j^n, $\forall j = 1, \dots, J$, $\forall n = 1, \dots, N$, qui approchent les valeurs exactes $u(x_j, t^n)$. Rappelons que ces valeurs exactes sont telles que l'équation du problème (P) est satisfaite, ie

$$(E) \quad \partial_t u(x_j, t^n) - \partial_{xx} u(x_j, t^n) = f(x_j, t^n).$$

Il s'agit donc dans les faits d'approcher les opérateurs ∂_t et ∂_{xx}.

On propose tout d'abord les approximations naturelles suivantes

$$\partial_t u(x_j, t^n) \simeq \frac{u(x_j, t^{n+1}) - u(x_j, t^n)}{\Delta t},$$

$$\partial_{xx} u(x_j, t^n) \simeq \frac{\dfrac{u(x_{j+1}, t^n) - u(x_j, t^n)}{\Delta x} - \dfrac{u(x_j, t^n) - u(x_{j-1}, t^n)}{\Delta x}}{\Delta x} = \frac{u(x_{j+1}, t^n) - 2u(x_j, t^n) + u(x_{j-1}, t^n)}{\Delta x^2},$$

qui conduisent au schéma

$$(S_1) \quad \frac{u_j^{n+1} - u_j^n}{\Delta t} - \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{\Delta x^2} = f(x_j, t^n), \quad \forall j = 1, \dots, J, \ \forall n = 1, \dots, N$$

Notons que ce schéma se réécrit aussi sous la forme

$$(S_1)' \quad u_j^{n+1} = u_j^n + \frac{\Delta t}{\Delta x^2}\left(u_{j+1}^n - 2u_j^n + u_{j-1}^n\right) + \Delta t\, f(x_j, t^n), \quad \forall j = 1, \dots, J, \ \forall n = 1, \dots, N$$

de sorte que la connaissance de $(u_j^0)_{j=0,\dots,J+1}$ suffit pour déterminer

toutes les valeurs $(u_j^n)_{\substack{j=0,\dots,J+1 \\ n=1,\dots,N}}$ par simple récurrence. On dit que le schéma proposé est explicite en temps.

Notons que le schéma (S_1) peut aussi s'écrire vectoriellement sous la forme

$$\frac{U^{n+1} - U^n}{\Delta t} + A U^n = F^n$$

avec $U^n = \begin{pmatrix} u_1^n \\ \vdots \\ u_J^n \end{pmatrix}$, $F^n = \begin{pmatrix} f(x_1, t^n) \\ \vdots \\ f(x_J, t^n) \end{pmatrix}$

et $$A = \frac{1}{\Delta x^2} \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}$$

On se propose maintenant d'étudier plus précisément ce schéma explicite. On commence par cela par définir l'erreur de consistance du schéma (S_1), notée $\varepsilon_j^n(u)$, qui représente l'erreur commise en remplaçant l'équation exacte (E) par la formule approchée. Plus précisément, on a ici

$$\varepsilon_j^n(u) = \frac{u(x_j, t^{n+1}) - u(x_j, t^n)}{\Delta t} - \frac{u(x_{j+1}, t^n) - 2u(x_j, t^n) + u(x_{j-1}, t^n)}{\Delta x^2} - f(x_j, t^n),$$

où u représente une solution régulière du problème (P). On note que la quantité $\partial_t u(x_j, t^n) - \partial_{xx} u(x_j, t^n) - f(x_j, t^n)$ est nulle d'après l'équation exacte satisfaite par u, mais que $\varepsilon_j^n(u)$ n'est pas forcément nulle à cause des erreurs d'approximation commises. On cherche à "mesurer" ces erreurs en étudiant ε_j^n. On s'intéressera plus précisément à la quantité $\max_{n=0,\dots,N-1} \|\varepsilon^n(u)\|_\infty := \max_{n=0,\dots,N-1} \max_{j=1,\dots,J} |\varepsilon_j^n(u)|$.

En utilisant des développements de Taylor, nous avons l'existence d'un temps $\tau \in]t^n, t^{n+1}[$ et de deux abscisses $\xi_1 \in]x_j, x_{j+1}[$ et $\xi_2 \in]x_{j-1}, x_j[$ tels que

$$u(x_j, t^{n+1}) = u(x_j, t^n) + \Delta t\, \partial_t u(x_j, t^n) + \frac{\Delta t^2}{2} \partial_{tt} u(x_j, \tau)$$

$$u(x_{j+1}, t^n) = u(x_j, t^n) + \Delta x\, \partial_x u(x_j, t^n) + \frac{\Delta x^2}{2} \partial_{xx} u(x_j, t^n) + \frac{\Delta x^3}{6} \partial_{xxx} u(x_j, t^n) + \frac{\Delta x^4}{24} \partial_{xxxx} u(\xi_1, t^n)$$

$$u(x_{j-1}, t^n) = u(x_j, t^n) - \Delta x\, \partial_x u(x_j, t^n) + \frac{\Delta x^2}{2} \partial_{xx} u(x_j, t^n) - \frac{\Delta x^3}{6} \partial_{xxx} u(x_j, t^n) + \frac{\Delta x^4}{24} \partial_{xxxx} u(\xi_2, t^n)$$

Des calculs simples montrent alors que

$$\varepsilon_j^n(u) = \partial_t u(x_j, t^n) - \partial_{xx} u(x_j, t^n) - f(x_j, t^n) + \frac{\Delta t}{2} \partial_{tt} u(x_j, \tau) - \frac{\Delta x^2}{24} \left(\partial_{xxxx} u(\xi_1, t^n) + \partial_{xxxx} u(\xi_2, t^n) \right)$$

ou encore, d'après (E),

$$\varepsilon_j^n(u) = \frac{\Delta t}{2} \partial_{tt} u(x_j, \tau) - \frac{\Delta x^2}{24} \left(\partial_{xxxx} u(\xi_1, t^n) + \partial_{xxxx} u(\xi_2, t^n) \right)$$

Par régularité de u, il existe donc une constante $C \in \mathbb{R}$ indépendante de Δt et Δx telle que

$$|\varepsilon_j^n(u)| \leq C(\Delta t + \Delta x^2)$$

et donc $$\max_{n=0,\dots,N-1} \|\varepsilon^n(u)\|_\infty \leq C(\Delta t + \Delta x^2).$$

On note donc que l'erreur commise tend vers 0 lorsque Δt et Δx tendent vers 0, linéairement en Δt et quadratiquement en Δx. On dit que le schéma est d'ordre 1 en temps et d'ordre 2 en espace.

Remarque En reproduisant des calculs semblables, il serait possible d'obtenir un ordre 2 en temps également en remplaçant l'approximation $\partial_t u(x_j;t^n) \simeq \frac{u(x_j,t^{n+1}) - u(x_j;t^n)}{\Delta t}$ par l'approximation

$\partial_t u(x_j,t^n) \simeq \frac{u(x_j,t^{n+1}) - u(x_j,t^{n-1})}{2\Delta t}$. Le schéma à deux niveaux, appelé schéma "saute-mouton" ou "de Richardson", est instable et ne permet donc pas d'obtenir la convergence de la solution approchée vers la solution exacte (au contraire du schema d'ordre 1 en temps proposé, comme nous le verrons ci-dessous).

Remarquons également que l'approximation $\partial_t u(x_j,t^n) \simeq \frac{u(x_j,t^n) - u(x_j,t^{n-1})}{\Delta t}$ conduit au schéma

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} - \frac{u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}}{\Delta x^2} = f(x_j,t^n)$$

qui est d'ordre 1 en temps et 2 en espace comme le précédent, mais qui nécessite la résolution d'un système linéaire pour calculer le vecteur U^{n+1} puisqu'il s'écrit vectoriellement sous la forme $\left(\frac{1}{\Delta t} I + A\right) U^{n+1} = \frac{1}{\Delta t} U^n + F^n$. Il est donc plus difficile à mettre en oeuvre et à analyser (la norme utilisée pour l'étude de ce schéma est la norme $|||V|||_2 := \sqrt{\Delta x \sum_{j=1}^{J} v_j^2}$ pour tout vecteur $V = (v_1, \dots, v_J)$ de $\mathbb{R}^J$). On dit que ce schéma est implicite en temps.

Etudions maintenant la convergence du schema explicite proposé. On definit l'erreur au temps t^n par le vecteur $e^n(u) = (e_1^n(u), \dots, e_J^n(u))$

avec $e_j^n(u) = u_j^n - u(x_j, t^n)$ pour tout $n = 0, \dots, N$ et tout $j = 0, \dots, J+1$. Nous allons montrer que sous la condition CFL sur le pas de temps $\Delta t \leq \frac{\Delta x^2}{2}$, il existe une constante $C \in \mathbb{R}$ telle que $\|e^n(u)\|_\infty \leq C'(\Delta t + \Delta x^2)$ $\forall n = 0, \dots, N$, ce qui garantit la convergence du schéma lorsque Δt et Δx tendent vers 0 (sous la condition $\Delta t \leq \Delta x^2/2$). On a posé ici $\|e^n(u)\|_\infty = \max_{j=1,\dots,J} |e_j^n(u)|$.

Afin de montrer ce résultat, notons tout d'abord que $e_j^0 = 0$ $\forall j$ et $e_0^n = e_{J+1}^n = 0$ $\forall n$. Ensuite, par définition du schéma (S_1) et de l'erreur de consistante $\varepsilon_j^n(u)$, on a

$$\frac{e_j^{n+1} - e_j^n}{\Delta t} - \frac{e_{j+1}^n - 2e_j^n + e_{j-1}^n}{\Delta x^2} = -\varepsilon_j^n(u)$$

ce qui peut s'écrire aussi sous la forme

$$e_j^{n+1} = \left(1 - 2\frac{\Delta t}{\Delta x^2}\right) e_j^n + \frac{\Delta t}{\Delta x^2} e_{j-1}^n + \frac{\Delta t}{\Delta x^2} e_{j+1}^n - \Delta t\, \varepsilon_j^n(u)$$

En supposant que $\Delta t \leq \frac{\Delta x^2}{2}$, on a donc $1 - 2\frac{\Delta t}{\Delta x^2} \geq 0$, de sorte que

$$|e_j^{n+1}| \leq \left(1 - 2\frac{\Delta t}{\Delta x^2}\right)|e_j^n| + \frac{\Delta t}{\Delta x^2}|e_{j-1}^n| + \frac{\Delta t}{\Delta x^2}|e_{j+1}^n| + \Delta t\,|\varepsilon_j^n|$$

$$|e_j^{n+1}| \leq \|e^n(u)\|_\infty + \Delta t\,|\varepsilon_j^n|$$

$$|e_j^{n+1}| \leq \|e^n(u)\|_\infty + \Delta t\,\|\varepsilon^n(u)\|_\infty \quad \forall n = 0, \dots, N-1, \ \forall j = 0, \dots, J+1.$$

Le schéma étant consistant à l'ordre 1 en temps et 2 en espace, on
x donc $|e_j^{n+1}| \leq \|e^n(u)\|_\infty + C\Delta t(\Delta t + \Delta x^2)$

$$\|e^{n+1}(u)\|_\infty \leq \|e^n(u)\|_\infty + C\Delta t(\Delta t + \Delta x^2)$$

e sorte que, puisque $\|e^0(u)\|_\infty = 0$

$$\|e^n(u)\|_\infty \leq C\, n\Delta t\, (\Delta t + \Delta x^2) \qquad \forall n = 0, \dots, N.$$

Comme $n\Delta t \leq T$, on a donc finalement

$$\|e^n(u)\|_\infty \leq \underbrace{CT}_{C'} (\Delta t + \Delta x^2)$$

Le schéma est donc bien convergent sous la condition CFL $\Delta t \leq \frac{\Delta x^2}{2}$

<u>Remarque</u> La condition CFL $\Delta t \leq \frac{\Delta x^2}{2}$ est très contraignante en pratique, à cause notamment de l'exposant 2 en Δx. Il est possible de remédier à ce problème en considérant une classe de schémas implicites, appelés les θ-schémas. Il s'écrivent sous la forme vectorielle

$$\left(\frac{1}{\Delta t} I + \theta A\right) U^{n+1} = \left(\frac{1}{\Delta t} I - (1-\theta) A\right) U^n + F^{n+1/2}$$

avec $\theta \in [0,1]$ et $F^{n+1/2} = \begin{pmatrix} f(x_1, t^n + \Delta t/2) \\ \vdots \\ f(x_J, t^n + \Delta t/2) \end{pmatrix}$

La modification du second membre (par rapport au schéma explicite) va permettre d'obtenir un schéma d'ordre 2 en temps.
On retrouve le schéma explicite lorsque $\theta = 0$ et le schema implicite lorsque $\theta = 1$. Une analyse détaillée de ce schéma permet de montrer - qu'il est d'ordre 1 en temps et 2 en espace, sauf si $\theta = 1/2$ et dans ce cas il est aussi d'ordre 2 en temps. Le $1/2$-schéma est appelé schéma de Crank-Nicholson.

- qu'il est convergent sans condition si $\theta \geq 1/2$ et sous la condition CFL $\Delta t \leq \frac{\Delta x^2}{2}(1-2\theta)$ sinon. On remarque donc que le $1/2$-schéma est convergent et d'ordre 2 sans condition sur Δt !!!

Remarque La convergence du schéma explicite repose sur l'obtention de l'inégalité $|e_j^{n+1}| \leq \|e^{n}(u)\|_\infty + \Delta t\, |\xi_j^n|$ puis sur l'erreur de consistance du schéma. L'inégalité $|e_j^{n+1}| \leq \|e^{n}(u)\|_\infty + \Delta t\, |\xi_j^n|$ peut être associée à une propriété de stabilité du schema, à laquelle on peut donner une définition plus rigoureuse. La stabilité et la consistance d'un schéma entraine sa convergence. C'est le théorème de Lax.