. Complément au cours (5).

Erreur de troncature. F. Dubois, mars 2008.

- Un schéma numérique comme par exemple le schéma d'Euler explicite pour la résolution approchée d'une équation différentielle est destiné à apporter une information là où la résolution "analytique" est impossible. Une question naturelle consiste à se demander si on peut a priori évaluer la qualité de cette méthodologie numérique, si on peut évaluer la précision de l'approche, alors même que la connaissance de l'erreur commise par la méthode numérique reste, au moins dans un premier temps, hors de portée. La réponse est oui, grâce à l'emploi de l'erreur de troncature.

- Nous allons définir l'erreur de troncature pour un modèle différentiel simple

(1) $\quad \dfrac{du}{dt} = f(u) \quad , \quad u(t) \in \mathbb{R}, \quad t > 0$

associé à une condition initiale

(2) $\quad u(0) = u_0$.

On suppose (et c'est une hypothèse correcte mathématiquement dès que la fonction $f(\bullet)$ est régulière) que le système dynamique (1)(2) a effectivement une solution unique $t \longmapsto u(t)$ et que cette fonction est assez régulière.

- Pour $\Delta t > 0$ fixé, nous cherchons une approximation u^k de $u(k\Delta t)$ pour les diverses valeurs de l'entier k :

$$(3) \qquad u^k \simeq u(k\Delta t), \quad k \in \mathbb{N}.$$

Nous insistons sur la <u>différence</u> entre u^k et $u(k\Delta t)$. Le nombre u^k est une valeur <u>approchée</u> accessible à l'expérimentation numérique via un algorithme de résolution que nous allons préciser plus bas. Le nombre $u(k\Delta t)$ est la valeur <u>exacte</u> de la solution $u(\bullet)$ du système (1)(2) au temps $t^k = k\Delta t$. Cette valeur exacte existe mais n'est pas connue numériquement a priori. Le but est justement d'en fournir une valeur approchée !

- L'algorithme d'Euler explicite consiste à se simplifier la vie pour la condition initiale (2) en prenant pour valeur "approchée" u^0 la valeur exacte u_0 qui est déjà connue (!) :

$$(4) \qquad u^0 = u_0 .$$

Le schéma est ensuite une relation de récurrence où la valeur u^{k+1} est obtenue à partir de u^k en "mimant" l'équation différentielle (1) :

$$(5) \qquad \frac{1}{\Delta t}\left(u^{k+1} - u^k\right) = f(u^k), \quad k \in \mathbb{N}.$$

Si la fonction $u \mapsto f(u)$ est connue, les relations (4) et ~~(2)~~ (5) permettent le calcul de u^k sans difficulté, pour toutes les valeurs de $k \in \mathbb{N}$.

- La notion d'erreur de troncature n'est pas complètement élémentaire. L'approche est duale. Au lieu d'approcher "naïvement" l'étude de la véritable erreur

$$(6) \quad \varepsilon^k = |u^k - u(k\Delta t)|, \quad k \in \mathbb{N}$$

on observe que la solution exacte $u(k\Delta t)$ n'est <u>pas</u> solution de l'équation approchée (5) ! L'écart obtenu définit l'erreur de troncature :

$$(7) \quad \tau^k = \frac{1}{\Delta t}\Big[u\big((k+1)\Delta t\big) - u(k\Delta t)\Big] - f\big(u(k\Delta t)\big) .$$

- L'intérêt de la définition (7) de l'erreur de troncature est qu'on peut en faire un développement de Taylor puisque $u(t)$ est régulière. Afin d'alléger la notation, nous remplaçons le temps discret courant $t^k = k\Delta t$ par un temps t arbitraire et nous avons alors un paramètre naturel qui est le pas de temps Δt. Nous posons :

$$(8) \quad \tau(\Delta t) \equiv \frac{1}{\Delta t}\Big[u(t+\Delta t) - u(t)\Big] - f\big(u(t)\big) .$$

- On peut alors approcher le second membre de (8) avec un développement de Taylor au premier ordre :

$$(9) \quad u(t+\Delta t) = u(t) + \Delta t \, \frac{du}{dt}(t) + O(\Delta t^2) .$$

Comme la fonction $u(\bullet)$ est solution de l'équation différentielle (1), nous pouvons remplacer au membre de droite de (9) $\frac{du}{dt}$ par sa valeur $f(u(t))$. Nous obtenons donc :

$$(10)\quad u(t+\Delta t) = u(t) + \Delta t\, f(u(t)) + O(\Delta t^2)$$

On déduit alors sans difficulté que l'erreur de troncature définie en (8) est de l'ordre de Δt :

$$(11)\quad \tau(\Delta t) = O(\Delta t).$$

On dit que le schéma d'<u>Euler explicite</u> est <u>d'ordre 1</u>.

• Nous conseillons au lecteur de reprendre les calculs précédents avec les schémas d'Euler implicite et de Crank-Nicolson. On a encore (11) pour Euler implicite mais le schéma de Crank-Nicolson est d'ordre 2 ; si on définit l'erreur de troncature non plus par la relation (8) mais par

$$(12)\quad \tilde{\tau}(\Delta t) \equiv \frac{1}{\Delta t}\left(u(t+\Delta t) - u(t)\right) + \frac{1}{2}\left[f(u(t)) + f(u(t+\Delta t))\right]$$

on peut montrer (nous renouvelons notre conseil au lecteur !) que

$$(13)\quad \tilde{\tau}(\Delta t) = O(\Delta t^2).$$

Le schéma de <u>Crank-Nicolson</u> est <u>d'ordre 2</u>.

- Abordons maintenant l'équation de la chaleur

(14) $\dfrac{\partial u}{\partial t} - \mu \dfrac{\partial^2 u}{\partial x^2} = 0, \quad x \in \mathbb{R}, \; t > 0$

Nous disposons d'un pas d'espace $\Delta x > 0$ et d'un pas de temps $\Delta t > 0$. Nous cherchons une approximation u_j^n de $u(n\Delta t, j\Delta x)$ pour $n \in \mathbb{N}$ et $j \in \mathbb{Z}$:

(15) $u_j^n \simeq u(n\Delta t, j\Delta x), \; n \in \mathbb{N}, j \in \mathbb{Z}.$

Comme $u(\bullet,\bullet)$ est cette fois une fonction de <u>deux</u> variables (l'espace et le temps), nous rappelons l'excellente qualité de l'approximation de la dérivée seconde par un schéma à trois points :

(16) $\dfrac{\partial^2 u}{\partial x^2}(j\Delta x) \simeq \dfrac{1}{\Delta x^2}\left(u_{j+1} - 2u_j + u_{j-1}\right), \; j \in \mathbb{Z}.$

Ainsi, pour $u(\bullet,\bullet)$ fonction régulière (de la variable d'espace essentiellement), nous avons l'erreur de troncature suivante

(17) $\tau^2(\Delta x) \equiv \dfrac{1}{\Delta x^2}\left(u(x+\Delta x) - 2u(x) + u(x-\Delta x)\right) - \dfrac{d^2 u}{dx^2}(x)$

qui mesure l'approximation de la dérivée seconde en espace par le schéma (16). Nous allons montrer à l'aide d'un développement de Taylor que

(18) $\tau^2(\Delta x) = O(\Delta x^2).$

Il suffit de développer $u(x+\Delta x)$ au 3° ordre :

(19) $u(x+\Delta x) = u(x) + \Delta x \dfrac{du}{dx} + \dfrac{\Delta x^2}{2} \dfrac{d^2u}{dx^2} + \dfrac{\Delta x^3}{6} \dfrac{d^3u}{dx^3} + O(\Delta x^4)$

Puis on fait de même pour $u(x-\Delta x)$ (il suffit d'ailleurs de changer Δx en $-\Delta x$ dans la relation (19) !) :

(20) $u(x-\Delta x) = u(x) - \Delta x \dfrac{du}{dx} + \dfrac{\Delta x^2}{2} \dfrac{d^2u}{dx^2} - \dfrac{\Delta x^3}{6} \dfrac{d^3u}{dx^3} + O(\Delta x^4)$

Les termes impairs en Δx disparaissent quand on ajoute (19) et (20). On a donc, après avoir aussi retranché deux fois $u(x)$:

(21) $u(x+\Delta x) - 2u(x) + u(x-\Delta x) = \Delta x^2 \dfrac{d^2u}{dx^2} + O(\Delta x^4)$.

Compte tenu de la définition (17), la relation (18) s'obtient essentiellement par une division de (21) par Δx^2 !

- Le schéma explicite pour la résolution approchée de l'équation de la chaleur (14) s'écrit

(22) $\dfrac{1}{\Delta t}\left(u_j^{n+1} - u_j^n\right) - \dfrac{\mu}{\Delta x^2}\left(u_{j+1}^n - 2u_j^n + u_{j-1}^n\right) = 0,\ n \in \mathbb{N},\ j \in \mathbb{Z}$.

C'est un schéma d'Euler explicite en temps une fois la dérivée seconde en espace approchée à l'aide de la relation (16). Pour $u(\bullet,\bullet)$ solution (régulière !) de (14), nous définissons l'erreur de troncature τ_c par

$$(23)\quad \begin{cases} \tau_c(\Delta t, \Delta x) = \frac{1}{\Delta t}\big(u(t+\Delta t, x) - u(t,x)\big) \\ \qquad\qquad - \frac{\mu}{\Delta x^2}\big[\, u(t, x+\Delta x) - 2u(t,x) + u(t, x-\Delta x)\big]. \end{cases}$$

On a simplement remplacé dans l'expression du schéma u_j^n par $u(n\Delta t, j\Delta x)$, puis $n\Delta t$ par le simple nombre t et $j\Delta x$ par x. Un développement au premier ordre en temps de $u(t+\Delta t, x)$ joint à (17)(18) montre que

$$(24)\quad \tau_c(\Delta t, \Delta x) = \Big(\frac{\partial u}{\partial t} - \mu \frac{\partial^2 u}{\partial x^2}\Big) + O(\Delta t) + O(\Delta x^2).$$

Comme $u(\bullet,\bullet)$ est solution de (14), le premier terme au membre de droite de (24) disparaît. En tenant compte de la condition de stabilité

$$(25)\qquad 0 < \mu \frac{\Delta t}{\Delta x^2} \leqslant \frac{1}{2}$$

on a

$$(26)\qquad \Delta t = O(\Delta x^2)$$

et on a finalement

$$(27)\quad \tau_c(\Delta t, \Delta x) = O(\Delta x^2).$$

On en conclut que le schéma (22) est assez précis (il est du second ordre !) mais il est <u>cher</u> à cause de la condition de stabilité (25) qui impose de prendre de très petits pas de temps !

Dubois, 26 mars 2008.