MÉTHODES NUMÉRIQUES POUR LES EDP

UNIVERSITÉ DE CERGY

Contents

1. Approximation des dérivées par différence finie

Un problème qu'on rencontre souvent en analyse numérique est l'approximation de la dérivée d'une fonction $f : [a, b] \to \mathbb{R}$ sur un intervalle donné $[a, b]$.

1.1. **Méthode générale.** Une approche naturelle consiste à introduire $n+1$ nœuds $x_k \in [a, b]$ uniformément répartis, c'est-à-dire tels que

$$x_0 = a, \qquad x_n = b, \qquad x_{k+1} = x_k + h, \quad \forall k \in \{0, \dots, n-1\},$$

où

$$h := \frac{b-a}{n}.$$

On approche alors $f'(x_i)$ en utilisant les valeurs nodales $f(x_k)$, dont on considère avoir l'accès. On note u'_i l'approximation de $f'(x_i)$, donc

$$u'_i \simeq f'(x_i).$$

De manière générale, on définit les u'_i via l'équation

$$h \sum_{k=-m}^{m} \alpha_k \, u'_{i-k} = \sum_{k=-m'}^{m'} \beta_k \, f(x_{i-k}), \tag{1}$$

où $\{\alpha_k\}, \{\beta_k\} \in \mathbb{R}$ sont $m + m' + 1$ coefficients à déterminer, et où on peut utiliser la convention $u'_j = 0$ et $f(x_j) = 0$ pour tout $j \notin \{0, \dots, n\}$. Cette équation déterminant une approximation est appelée schéma.

Le coût du calcul est un critère important dans le choix du schéma, il faut par exemple noter que si $m \neq 0$, la détermination des quantités u'_i requiert la résolution d'un système linéaire.

Definition 1.1 (Stencil)**.** *L'ensemble des nœuds impliqués dans la construction de la dérivée de y en un nœud donné est appelé* stencil.

1.2. **Méthodes des différences finies classiques.**

1.2.1. *Méthode "forward".* Le moyen le plus simple pour construire une formule du type (1) consiste à revenir à la définition de la dérivée. Si $f'(x_i)$ existe, alors

$$f'(x_i) = \lim_{h\to 0^+} \frac{f(x_i+h) - f(x_i)}{h}. \tag{2}$$

Definition 1.2 (Différence finie progressive)**.** *En remplaçant la limite par le taux d'accroissement, avec h fini, on obtient l'approximation*

$$u'_{i,FD} = \frac{f(x_{i+1}) - f(x_i)}{h}, \qquad \forall i \in \{0, \ldots, n-1\}. \tag{3}$$

Cette relation est un cas particulier de (1) où $m = 0$, $\alpha_0 = 1$, $m' = 1$, $\beta_{-1} = 1$, $\beta_0 = -1$, $\beta_1 = 0$. Le second membre de (3) est appelé différence finie progressive, ou "en avant".

L'approximation que l'on fait revient à remplacer $f'(x_i)$ par la pente de la droite passant par les points $(x_i, f(x_i))$ et $(x_{i+1}, f(x_{i+1}))$.

Pour estimer l'erreur commise, il suffit d'écrire le développement de Taylor de f (qui sera toujours supposée assez régulière). En effet, par le théorème de Taylor-Lagrange, il existe $\xi_i \in]x_i, x_{i+1}[$ tel que

$$f(x_{i+1}) = f(x_i) + hf'(x_i) + \frac{h^2}{2} f''(\xi_i).$$

Ainsi,

$$f'(x_i) - u'_{i,FD} = -\frac{h}{2} f''(\xi_i).$$

1.2.2. *Méthode centrée.* Au lieu de (3), on aurait pu utiliser un taux d'accroissement centré, obtenant alors l'approximation suivante.

Definition 1.3 (Différence finie centrée)**.**

$$u'_{i,CD} = \frac{f(x_{i+1}) - f(x_{i-1})}{2h}, \qquad \forall i \in \{1, \ldots, n-1\}. \tag{4}$$

Le schéma (4) est un cas particulier de (1) où $m = 0$, $\alpha_0 = 1$, $m' = 1$, $\beta_{-1} = \frac{1}{2}$, $\beta_0 = 0$, $\beta_1 = -\frac{1}{2}$. Le second membre de (4) est appelé différence finie centrée. Géométriquement, l'approximation revient à remplacer $f'(x_i)$ par la pente de la droite passant par les points $(x_{i-1}, f(x_{i-1}))$ et $(x_{i+1}, f(x_{i+1}))$.

Lemma 1.4. *Il existe $\xi_i \in [x_{i-1}, x_{i+1}]$ tel que*

$$f'(x_i) - u'_{i,CD} = -\frac{h^2}{6} f^{(3)}(\xi_i).$$

Démonstration. On utilise le développement de Taylor autour de x_i aux points $x_{i+1} = x_i + h$ et $x_{i-1} = x_i - h$ et le théorème de Taylor-Lagrange, on obtient

$$f(x_i + h) = f(x_i) + hf'(x_i) + \frac{h^2}{2} f''(x_i) + \frac{h^3}{6} f^{(3)}(\xi_1),$$
$$f(x_i - h) = f(x_i) - hf'(x_i) + \frac{h^2}{2} f''(x_i) - \frac{h^3}{6} f^{(3)}(\xi_2),$$

où $\xi_1 \in]x_i, x_i + h[$ et $\xi_2 \in]x_i - h, x_i[$. Ainsi,

$$f'(x_i) - u'_{i,CD} = -\frac{h^2}{12}\big(f^{(3)}(\xi_1) + f^{(3)}(\xi_2)\big).$$

Puisque $f^{(3)}$ est continue sur $]x_i - h, x_i + h[$, la moyenne

$$\frac{f^{(3)}(\xi_1) + f^{(3)}(\xi_2)}{2}$$

est une valeur intermédiaire de $f^{(3)}$ sur cet intervalle. Par théorème des valeurs intermédiaires, il existe $\xi_i \in]x_i - h, x_i + h[$ tel que

$$f^{(3)}(\xi_i) = \frac{f^{(3)}(\xi_1) + f^{(3)}(\xi_2)}{2}.$$

□

La formule (4) fournit donc une approximation de $f'(x_i)$ qui est du second ordre par rapport à h.

1.2.3. *Méthode "backward".* Enfin, on peut définir de manière analogue un troisième schéma.

Definition 1.5 (Différence finie rétrograde)**.**

$$u'_{i,BD} = \frac{f(x_i) - f(x_{i-1})}{h}, \qquad \forall i \in \{1, \ldots, n\}. \tag{5}$$

L'erreur suivante lui correspond

$$f'(x_i) - u'_{i,BD} = \frac{h}{2} f''(\xi_i),$$

pour un certain $\xi_i \in]x_{i-1}, x_i[$. Les valeurs des paramètres dans (5) sont $m = 0$, $\alpha_0 = 1$, $m' = 1$ et $\beta_{-1} = 0$, $\beta_0 = 1$, $\beta_1 = -1$.

1.2.4. *Approximation de dérivées d'ordres supérieurs.* Des schémas d'ordre élevé, ou encore des approximations par différences finies de dérivées de f d'ordre supérieur, peuvent être construits en augmentant l'ordre des développements de Taylor. Voici un exemple concernant l'approximation de f''. Si $f \in C^4([a, b])$, on obtient

$$f''(x_i) = \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1})}{h^2} - \frac{h^2}{24}\Big(f^{(4)}(x_i + \theta_i h) + f^{(4)}(x_i - \omega_i h)\Big),$$

où $0 < \theta_i, \omega_i < 1$, d'où on déduit le schéma aux différences finies centrées

$$u''_i = \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1})}{h^2}, \qquad \forall i \in \{1, \ldots, n-1\}. \tag{6}$$

L'erreur correspondante est

$$f''(x_i) - u''_i = -\frac{h^2}{24}\Big(f^{(4)}(x_i + \theta_i h) + f^{(4)}(x_i - \omega_i h)\Big).$$

La formule (6) fournit donc une approximation de $f''(x_i)$ du second ordre par rapport à h.

1.2.5. *Différences finies compactes.* Pour abréger on note $f_i^{(k)} = f^{(k)}(x_i)$ et $f_i := f(x_i)$. Des approximations plus précises de f' sont données par les formules suivantes

Definition 1.6 (Différences finies compactes)**.** *On définit u'_i via les équations*

$$\alpha u'_{i-1} + u'_i + \alpha u'_{i+1} = \frac{\beta}{2h}\big(f_{i+1} - f_{i-1}\big) + \frac{\gamma}{4h}\big(f_{i+2} - f_{i-2}\big), \tag{7}$$

où $i \in \{2, \dots, n-2\}$.

Les coefficients α, β et γ doivent être déterminés de manière à ce que les relations (7) conduisent à des valeurs de u_i qui approchent $f'(x_i)$ à l'ordre le plus élevé par rapport à h. Pour cela, on choisit des coefficients qui minimisent l'erreur de consistance

$$\sigma_i := \alpha f'_{i-1} + f'_i + \alpha f'_{i+1} - \left[\frac{\beta}{2h}(f_{i+1} - f_{i-1}) + \frac{\gamma}{4h}(f_{i+2} - f_{i-2})\right]. \tag{8}$$

Nous pouvons donner une définition non rigoureuse mais générale des erreurs de consistance.

Definition 1.7 (Erreur de consistance)**.** *L'erreur de consistance d'un schéma consiste à considérer le schéma, à y remplacer la grandeur approximée par la grandeur exacte, et à regarder l'erreur qui y est faite.*

Definition 1.8 (Erreur de convergence)**.** *L'erreur de convergence est l'erreur entre une quantité exacte et son approximation.*

On considère une norme $\|\cdot\|$ sur $\mathbb{R}^{N+1}$ quelconque.

Lemma 1.9 (Consistance implique convergence)**.** *Considérons un schéma de différences finies compactes pour approcher f', écrit sous forme matricielle*

$$A\, u' = BF,$$

où

$$u' := (u'_i)_{i=0}^N, \qquad F' := \big(f'(x_i)\big)_{i=0}^N, \qquad F := \big(f(x_i)\big)_{i=0}^N,$$

et où on a $u'_i \simeq f'(x_i)$. B peut dépendre de h mais pas A, et A est inversible. Supposons qu'il existe $C > 0$ et $n \in \mathbb{N}$ tels que pour tout $h > 0$ dans un voisinage de 0,

$$\|AF' - BF\| \leqslant Ch^n,$$

qui est l'erreur de consistance. Alors l'erreur de convergence est

$$\|u' - F'\| \leqslant C\,\|A^{-1}\|\,h^n.$$

Démonstration. On a $Au' = BF$ et on définit l'erreur de consistance $\sigma = (\sigma_i)_{i=0}^N$ par $\sigma := AF' - BF$. En soustrayant ces deux relations, on obtient $A(u' - F') = -\sigma$ et donc $u' - F' = -A^{-1}\sigma$. $\square$

Autrement dit, l'ordre de convergence global est égal à l'ordre de consistance n. Dans (7) on a $N = 3$, et A est une matrice ayant 1 sur sa diagonale et α en-dessous et au-dessus de sa diagonale. Plus explicitement,

$$A = \begin{pmatrix} 1 & \alpha & 0 & \cdots & 0 \\ \alpha & 1 & \alpha & \ddots & \vdots \\ 0 & \alpha & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \alpha \\ 0 & \cdots & 0 & \alpha & 1 \end{pmatrix} \in \mathbb{R}^{(N+1)\times(N+1)}$$

$$B = \frac{1}{h} \begin{pmatrix} 0 & \frac{\beta}{2} & 0 & -\frac{\gamma}{4} & \cdots & 0 \\ -\frac{\beta}{2} & 0 & \frac{\beta}{2} & 0 & -\frac{\gamma}{4} & \vdots \\ 0 & -\frac{\beta}{2} & 0 & \frac{\beta}{2} & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & -\frac{\beta}{2} & 0 & \frac{\beta}{2} \\ 0 & \cdots & \frac{\gamma}{4} & 0 & -\frac{\beta}{2} & 0 \end{pmatrix} \in \mathbb{R}^{(N+1)\times(N+1)}.$$

Lemma 1.10 (Ordre 6 des différences finies compactes)**.** *Dans le cas de* (7)*, il existe un unique schéma d'ordre* 6 *et il correspond aux paramètres*

$$\alpha = \frac{1}{3}, \qquad \beta = \frac{14}{9}, \qquad \gamma = \frac{1}{9}. \tag{9}$$

Démonstration. En supposant que $f \in C^5([a,b])$ et en écrivant le développement de Taylor en x_i, on trouve

$$f_{i\pm1} = f_i \pm hf_i' + \frac{h^2}{2}f_i^{(2)} \pm \frac{h^3}{6}f_i^{(3)} + \frac{h^4}{24}f_i^{(4)} \pm \frac{h^5}{120}f_i^{(5)} + \frac{h^6}{6!}f_i^{(6)} + O(h^7),$$
$$f_{i\pm2} = f_i \pm 2hf_i' + 2h^2 f_i^{(2)} \pm \frac{4}{3}h^3 f_i^{(3)} + \frac{2}{3}h^4 f_i^{(4)} \pm \frac{4}{15}h^5 f_i^{(5)} + \frac{2^6h^6}{6!}f_i^{(6)} + O(h^7),$$
$$f_{i\pm1}' = f_i' \pm hf_i^{(2)} + \frac{h^2}{2}f_i^{(3)} \pm \frac{h^3}{6}f_i^{(4)} + \frac{h^4}{24}f_i^{(5)} \pm \frac{h^5}{120}f_i^{(5)} + O(h^6).$$

Ainsi

$$\alpha f_{i-1}' + f_i' + \alpha f_{i+1}' = (2\alpha+1)f_i' + \alpha h^2 f_i^{(3)} + \alpha\frac{h^4}{12}f_i^{(5)} + O(h^6).$$

On calcule ensuite

$$f_{i+1} - f_{i-1} = 2hf_i' + \frac{h^3}{3}f_i^{(3)} + \frac{h^5}{60}f_i^{(5)} + O(h^7),$$

et

$$f_{i+2} - f_{i-2} = 4hf_i' + \frac{8}{3}h^3 f_i^{(3)} + \frac{8}{15}h^5 f_i^{(5)} + O(h^7).$$

Par conséquent, le second membre vaut

$$\begin{aligned} &\frac{\beta}{2h}\big(f_{i+1} - f_{i-1}\big) + \frac{\gamma}{4h}\big(f_{i+2} - f_{i-2}\big) \\ &\qquad = (\beta+\gamma)f_i' + \Big(\frac{\beta}{6} + \frac{2\gamma}{3}\Big)h^2 f_i^{(3)} + \Big(\frac{\beta}{120} + \frac{2\gamma}{15}\Big)h^4 f_i^{(5)} + O(h^6). \end{aligned}$$

Par substitution dans (8), on obtient

$$\sigma_i = (2\alpha+1)f_i' + \alpha\frac{h^2}{2}f_i^{(3)} + \alpha\frac{h^4}{12}f_i^{(5)} - (\beta+\gamma)f_i'$$
$$- \frac{h^2}{2}\left(\frac{\beta}{6}+\frac{2\gamma}{3}\right)f_i^{(3)} - \frac{h^4}{60}\left(\frac{\beta}{2}+8\gamma\right)f_i^{(5)} + O(h^6).$$

On construit des schémas du second ordre en annulant le coefficient de f_i', c'est-à-dire en imposant

$$2\alpha+1 = \beta+\gamma,$$

des schémas d'ordre 4 en annulant aussi le coefficient de $f_i^{(3)}$,

$$6\alpha = \beta+4\gamma,$$

et des schémas d'ordre 6 en annulant aussi le coefficient de $f_i^{(5)}$,

$$10\alpha = \beta+16\gamma.$$

Le système linéaire formé par ces trois dernières relations est non singulier et a une unique solution (9).

Par le Lemme 1.9, l'erreur de convergence est la même que l'erreur de consistance. □

Il y a une seule méthode d'ordre 6 mais il existe en revanche une infinité de méthodes du second et du quatrième ordre. Parmi celles-ci, citons un schéma très utilisé qui correspond aux coefficients

$$\alpha = \frac{1}{4}, \qquad \beta = \frac{3}{2}, \qquad \gamma = 0.$$

Des schémas d'ordre plus élevé peuvent être construits au prix d'un accroissement supplémentaire du stencil.

1.2.6. *Conditions de bord.* Les schémas aux différences finies traditionnels correspondent au choix $\alpha = 0$ et permettent de calculer de manière explicite l'approximation de la dérivée première de f en un nœud, contrairement aux schémas compacts qui nécessitent dans tous les cas la résolution d'un système linéaire de la forme $Au = BF$.

Pour pouvoir résoudre le système, il est nécessaire de se donner les valeurs des variables u_i pour $i < 0$ et $i > n$. On est dans une situation simple quand f est une fonction périodique de période $b-a$, auquel cas

$$u_{i+n} = u_i \qquad \forall i \in \mathbb{Z}.$$

Dans le cas non périodique, le système (7) doit être complété par des relations aux nœuds voisins des extrémités de l'intervalle d'approximation. Par exemple, la dérivée première en x_0 peut être calculée en utilisant la relation

$$u_0' + \alpha u_1' = \frac{1}{h}\big(Af_1 + Bf_2 + Cf_3 + Df_4\big),$$

et en imposant

$$A = \frac{-3+\alpha+2D}{2}, \qquad B = 2+3D, \qquad C = \frac{-1-\alpha+6D}{2},$$

afin que le schéma soit au moins précis à l'ordre deux. Dans ce document, nous essaierons le plus possible d'éviter les problématiques liées aux conditions de bord.

2. Résolution numérique des équations différentielles ordinaires

2.1. **Le problème de Cauchy.** Soit $d \in \mathbb{N}$, I désigne un intervalle de $\mathbb{R}$, $t_0 \in I$, le problème de Cauchy associé à une EDO du premier ordre s'écrit de la manière suivante. Il faut trouver une fonction réelle $y \in C^1(I, \mathbb{R}^d)$ telle que

$$\begin{cases} y'(t) = f(t, y(t)) & \text{si } t \in I \\ y(t_0) = y_0 \end{cases} \tag{10}$$

où $f : I \times \mathbb{R}^d \to \mathbb{R}^d$ est continue par rapport aux deux variables. Si f ne dépend pas explicitement de t, l'équation différentielle est dite autonome. Le cas scalaire correspond à $d = 1$.

2.1.1. *Forme intégrale.* En intégrant (10) entre t_0 et t, on obtient

$$y(t) - y_0 = \int_{t_0}^{t} f(\tau, y(\tau))\, d\tau. \tag{11}$$

La solution de (10) est donc nécessairement de classe C^1 sur I et satisfait l'équation intégrale (11). Inversement, si y est définie par (11), alors elle est continue sur I et $y(t_0) = y_0$. De plus, en tant que primitive de la fonction continue $f(\cdot, y(\cdot))$, on a $y \in C^1(I)$ et elle satisfait l'équation différentielle :

$$y'(t) = f(t, y(t)).$$

Ainsi, si f est continue, le problème de Cauchy (10) est équivalent à l'équation intégrale (11). Nous verrons plus loin comment tirer parti de cette équivalence pour les méthodes numériques.

2.1.2. *Existence locale et unicité.* Rappelons maintenant deux résultats d'existence et d'unicité pour (10). On supposera $f : I \times \mathbb{R}^d \to \mathbb{R}^d$ localement lipschitzienne en (t_0, y_0) par rapport à y, ce qui signifie qu'il existe une boule ouverte $J \subseteq I$ centrée en t_0 de rayon r_J, une boule ouverte Σ centrée en y_0 de rayon r_Σ et une constante $L > 0$ telles que :

$$|f(t, y_1) - f(t, y_2)| \leqslant L|y_1 - y_2| \qquad \forall t \in J,\ \forall y_1, y_2 \in \Sigma.$$

Cette condition est automatiquement vérifiée si la dérivée de f par rapport à y est continue. En effet, dans ce cas, il suffit de prendre

$$L = \max_{(t,y) \in \overline{J \times \Sigma}} |\partial_y f(t, y)|\,.$$

Lemma 2.1 (Rappel sur l'existence de la solution locale)**.** *Soit $f : I \times \mathbb{R}^d \to \mathbb{R}^d$ localement lipschitzienne en (t_0, y_0) par rapport à y. Alors le problème de Cauchy* (10) *admet une unique solution dans une boule ouverte de centre t_0 et de rayon $r_0 > 0$.*

Cette solution est appelée solution locale.

2.1.3. *Existence globale et unicité.*

Lemma 2.2 (Rappel sur l'existence d'une solution globale)**.** *Le problème de Cauchy admet une solution globale unique si f est uniformément lipschitzienne par rapport à y, c'est-à-dire si on peut prendre $J = I$, $\Sigma = \mathbb{R}$.*

2.1.4. *Stabilité sous perturbation.* En vue de l'analyse de stabilité du problème de Cauchy, on considère le problème suivant :

$$\begin{cases} \dot{z}(t) = f(t, z(t)) + \delta(t), & t \in I, \\ z(t_0) = y_0 + \delta_0, \end{cases} \tag{12}$$

où $\delta_0 \in \mathbb{R}$ et où δ est une fonction continue sur I. Le problème (12) est déduit de (10) en perturbant la donnée initiale y_0 par δ_0 et la fonction f par δ. Caractérisons à présent la sensibilité de la solution z par rapport à ces perturbations. Intuitivement, la stabilité correspond au fait que si l'EDO est parturbée, alors la solution change d'une manière "continue".

Definition 2.3 (Problème de Cauchy stable)**.** *Soit I un ensemble borné. Le problème de Cauchy* (10) *est dit stable sur I si, pour toute perturbation $(\delta_0, \delta(t))$ satisfaisant*

$$|\delta_0| \leqslant \varepsilon, \qquad |\delta(t)| \leqslant \varepsilon \quad \forall t \in I,$$

avec $\varepsilon > 0$ assez petit pour garantir l'existence de la solution du problème perturbé (12)*, alors*

$$\exists C > 0 \quad \text{tel que} \quad |y(t) - z(t)| \leqslant C\varepsilon \quad \forall t \in I. \tag{13}$$

La constante C dépend en général de t_0, y et f, mais pas de ε.

Quand I n'est pas borné supérieurement, on dit que (10) *est asymptotiquement stable si, en plus de* (13)*, on a*

$$|y(t) - z(t)| \to 0 \quad \text{quand } t \to +\infty,$$

si $|\delta(t)| \to 0$ quand $t \to +\infty$.

2.1.5. *Grönwall.* Rappelons le lemme de Grönwall pour le problème de Cauchy.

Lemma 2.4 (Grönwall)**.** *Soit p une fonction positive intégrable sur l'intervalle $]t_0, t_0 + T[$, et soient g et φ deux fonctions continues sur $[t_0, t_0 + T]$, avec g croissante. Si φ satisfait*

$$\varphi(t) \leqslant g(t) + \int_{t_0}^{t} p(\tau)\, \varphi(\tau)\, d\tau \qquad \forall t \in [t_0, t_0 + T],$$

alors

$$\varphi(t) \leqslant g(t) \exp\left(\int_{t_0}^{t} p(\tau)\, d\tau \right) \qquad \forall t \in [t_0, t_0 + T].$$

2.1.6. *Utilité du numérique.* On ne sait intégrer qu'un très petit nombre d'EDO non linéaires. De plus, même quand c'est possible, il n'est pas toujours facile d'exprimer explicitement la solution ; considérer par exemple l'équation très simple :

$$y' = \frac{y - t}{y + t},$$

dont la solution n'est définie que de manière implicite par la relation :

$$\frac{1}{2}\log(t^2+y^2)+\arctan\left(\frac{y}{t}\right)=C,$$

où C est une constante dépendant de la condition initiale.

Pour cette raison, nous sommes conduits à considérer des méthodes numériques. Celles-ci peuvent en effet être appliquées à n'importe quelle EDO, sous la seule condition qu'elle admette une unique solution.

2.2. **Méthodes numériques à un pas.** Abordons à présent l'approximation numérique du problème de Cauchy (10). On fixe $0 < T < +\infty$ et on note $I =]t_0, t_0 + T[$ l'intervalle d'intégration. Pour $h > 0$, soit

$$t_n = t_0 + nh, \qquad n = 0, 1, 2, \ldots, N_h,$$

une suite de nœuds de I induisant une discrétisation de I en sous-intervalles $I_n := [\,t_n, t_{n+1}\,]$.

La longueur h de ces sous-intervalles est appelée pas de discrétisation. Le nombre N_h est le plus grand entier tel que

$$t_{N_h} \leqslant t_0 + T.$$

On a donc $hN_h \simeq T$.

Soit u_j l'approximation au nœud t_j de la solution exacte $y(t_j) =: y_j$,

$$u_j \simeq y_j.$$

De même, $f_j := f(t_j, u_j)$. On pose naturellement

$$u_0 = y_0.$$

Definition 2.5 (Méthode à un pas, méthode multipas)**.** *Une méthode numérique pour l'approximation du problème* (10) *est dite à un pas si* $\forall n \geqslant 0$*, le schéma définissant* u_{n+1} *ne dépend que de* u_n*. Autrement, on dit que le schéma est une méthode multi-pas (ou à pas multiples).*

Une méthode multipas est par exemple quand u_{n+1} dépend de u_n et u_{n-1}. Pour l'instant, nous concentrons notre attention sur les méthodes à un pas. En voici quelques-unes.

Definition 2.6 (Méthode d'Euler explicite)**.**

$$u_{n+1} = u_n + hf(t_n, u_n).$$

Definition 2.7 (Méthode d'Euler implicite)**.**

$$u_{n+1} = u_n + hf(t_{n+1}, u_{n+1}).$$

Dans les deux cas, y' est approchée par un schéma aux différences finies (resp. progressif puis rétrograde). Puisque ces deux schémas sont des approximations au premier ordre par rapport à h de la dérivée première de y, on s'attend à obtenir une approximation d'autant plus précise que le pas du maillage h est petit.

Definition 2.8 (Méthode du trapèze, ou de Crank–Nicolson)**.**

$$u_{n+1} = u_n + \frac{h}{2}(f(t_n, u_n) + f(t_{n+1}, u_{n+1})).$$

Cette méthode provient de l'approximation de l'intégrale (11) par la formule de quadrature du trapèze.

Definition 2.9 (Méthode de Heun)**.**

$$u_{n+1} = u_n + \frac{h}{2} \left(f(t_n, u_n) + f(t_{n+1}, u_n + hf_n) \right).$$

Definition 2.10 (Méthode explicite, implicite)**.** *Une méthode est dite explicite si la valeur u_{n+1} peut être calculée directement à l'aide des valeurs précédentes $(u_k)_{k \leqslant n}$ (ou d'une partie d'entre elles). Une méthode est dite implicite si u_{n+1} n'est défini que par une relation implicite faisant intervenir la fonction f.*

Ainsi, la substitution opérée dans la méthode de Heun a pour effet de transformer la méthode implicite du trapèze en une méthode explicite. La méthode d'Euler explicite est explicite, tandis que celle d'Euler implicite est implicite. Noter que les méthodes implicites nécessitent à chaque pas de temps la résolution d'un problème non linéaire (si f dépend non linéairement de la seconde variable).

Pour les méthodes implicites, il faut à chaque itération résoudre un problème consistant à trouver le zéro d'une fonction. Pour Euler implicite, afin de déterminer u_{n+1} à partir de u_n et t_{n+1} (auxquels on a accès), il faut résoudre l'équation

$$F(x) = 0,$$

où $F(x) := x - u_n - hf(t_{n+1}, x)$. On trouve donc le nombre $x = u_{n+1}$, comme solution.

2.3. **Analyse des méthodes à un pas.**

2.3.1. *Convergence.* Comme en Définition 1.7, la consistance mesure à quel point le schéma numérique reproduit l'équation originale quand le pas tend vers 0. Par ailleurs, la convergence dit quelque chose au niveau de la solution.

On rappelle que le max est une norme

$$\|(u_n)_{0 \leqslant n \leqslant j}\|_{\ell^\infty} := \max_{n \in \{0,\dots,j\}} |u_n|.$$

Definition 2.11 (Méthode convergente et ordre de convergence)**.** *Une méthode est dite convergente si*

$$\max_{0 \leqslant n \leqslant N} |u_n - y_n| \leqslant C(h)$$

où $C(h) \underset{h \to 0}{\longrightarrow} 0$. On dit que l'ordre de convergence est $p > 0$ s'il existe $c > 0$ tel que $C(h) = ch^p$.

2.3.2. *Grönwall discret.*

Lemma 2.12 (Grönwall discret)**.** *Soit $(k_n)_{n \in \mathbb{N}}$ et $(A_n)_{n \in \mathbb{N}}$ des suites de réels positifs et $(\phi_n)_{n \in \mathbb{N}}$ une suite telle que pour tout $n \in \mathbb{N}$,*

$$\phi_n \leqslant A_n + \sum_{s=0}^{n-1} k_s \phi_s,$$

Si (A_n) est croissante pour tout $n \geqslant 0$, alors pour tout $n \in \mathbb{N}$,

$$\phi_n \leqslant A_n \exp\left(\sum_{s=0}^{n-1} k_s\right).$$

Démonstration. L'idée de la preuve est d'éliminer les termes récurrents de type ϕ_s dans la somme, en les remplaçant par leur majorant inductif. Nous allons montrer par récurrence sur n que

$$\phi_n \leqslant A_n \exp\left(\sum_{s=0}^{n-1} k_s\right).$$

- Initialisation. On a $\phi_0 \leqslant A_0$, et comme $\sum_{s=0}^{-1} k_s = 0$, alors

$$\phi_0 \leqslant A_0 = A_0 e^0 = A_0 \exp\left(\sum_{s=0}^{-1} k_s\right).$$

- Hérédité. Supposons le résultat vrai pour tout $s < n$, c'est-à-dire

$$\phi_s \leqslant A_s \exp\left(\sum_{i=0}^{s-1} k_i\right), \qquad \forall s < n.$$

En partant de l'inégalité fondamentale,

$$\phi_n \leqslant A_n + \sum_{s=0}^{n-1} k_s \phi_s,$$

nous remplaçons chaque ϕ_s dans la somme par sa borne inductive :

$$\phi_n \leqslant A_n + \sum_{s=0}^{n-1} k_s\, A_s \exp\left(\sum_{i=0}^{s-1} k_i\right) \underset{A_s \text{ croissante}}{\leqslant} A_n \left(1 + \sum_{s=0}^{n-1} k_s \exp\left(\sum_{i=0}^{s-1} k_i\right)\right).$$

On reconnaît maintenant la forme discrète de l'intégrale exponentielle.

Lemma 2.13. *Nous voulons montrer que, si $k_s \geqslant 0$ pour tout s, alors*

$$1 + \sum_{s=0}^{n-1} k_s \exp\left(\sum_{i=0}^{s-1} k_i\right) \leqslant \exp\left(\sum_{s=0}^{n-1} k_s\right). \tag{14}$$

Démonstration. Pour cela, on introduit les notations $S_0 := 0$,

$$S_s := \sum_{i=0}^{s-1} k_i, \qquad s \geqslant 0, \qquad B_n := 1 + \sum_{s=0}^{n-1} k_s e^{S_s}, \qquad n \geqslant 0.$$

L'inégalité (14) s'écrit donc simplement $B_n \leqslant e^{S_n}$. Nous allons le prouver par récurrence sur n.

Pour $n = 0$, on a $B_0 = 1$, $S_0 = 0$, donc $B_0 = 1 = e^{S_0}$.

Supposons que, pour un certain $n \geqslant 0$, on ait $B_n \leqslant e^{S_n}$. Nous allons montrer que cela implique $B_{n+1} \leqslant e^{S_{n+1}}$. Par définition de B_{n+1}, on a

$$B_{n+1} = B_n + k_n e^{S_n} \leqslant e^{S_n} + k_n e^{S_n} = (1 + k_n)\, e^{S_n} \underset{1+x \leqslant e^x}{\leqslant} e^{k_n + S_n} = e^{S_{n+1}}.$$

$\square$

Nous obtenons donc finalement :
$$\phi_n \leqslant A_n \exp\left(\sum_{s=0}^{n-1} k_s\right),$$
ce qui conclut l'hérédité et prouve le Lemme 2.12. □

Corollary 2.14. *Soit (a_n) une suite positive. Si pour tout $n \in \{0, \ldots, N_h\}$,*
$$a_{n+1} \leqslant (1 + ch)a_n + Ch^{p+1},$$
alors $a_n \leqslant (a_0 + CTh^p)\, e^{cT}$.

Le lemme de Grönwall n'est pas nécessaire dans ce cas mais on va l'utiliser.

Démonstration. On a
$$a_{n+1} - a_n \leqslant ch\, a_n + Ch^{p+1}.$$
En sommant cette inégalité de $n = 0$ à $n = m - 1$ (avec $m \geqslant 1$ arbitraire), on obtient
$$a_m - a_0 = \sum_{n=0}^{m-1} \big(a_{n+1} - a_n\big) \leqslant \sum_{s=0}^{m-1} \big(ch\, a_s + Ch^{p+1}\big) = ch \sum_{s=0}^{m-1} a_s + Ch^{p+1}\, m.$$
On passe a_0 à droite et en utilisant le lemme de Grönwall discret, on obtient
$$a_m \leqslant \big(a_0 + Ch^{p+1} m\big)\, e^{cmh} \leqslant (a_0 + CTh^p)\, e^{cT}.$$
□

2.3.3. *Consistance implique convergence.* Considérons un schéma du type
$$u_{n+1} = \Phi\big(t_n, u_n, h\big). \tag{15}$$
On constate que ces schémas sont explicites. Par exemple, Euler progressif et la méthode de Heun se mettent sous cette forme, on a

- $\Phi\big(t, y, h\big) = y + hf(t, y)$ pour Euler progressif
- pour la méthode de Heun,
$$\Phi(t, y, h) = y + \frac{h}{2}\Big(f(t, y) + f\big(t + h,\ y + hf(t, y)\big)\Big).$$

On définit l'erreur de troncature locale
$$\tau_{n+1} := y_{n+1} - \Phi\left(t_n, y_n, h\right).$$

Definition 2.15 (Consistance d'un schéma)**.** *Une méthode est dite consistante si*
$$\max_{0 \leqslant n \leqslant N_h - 1} |\tau_n| \underset{h \to 0}{\longrightarrow} 0.$$

Proposition 2.16. *Prenons un schéma du type* (15)*. Supposons que*
$$|\Phi(t, y, h) - \Phi(t, z, h)| \leqslant (1 + Ch)|y - z|.$$
Si $|\tau_n| \leqslant Ch^{p+1}$ pour un $C > 0$ indépendant de h et de n (i.e. si la méthode est consistante d'ordre $p + 1$), alors la méthode est convergente d'ordre p, c'est-à-dire
$$\|u_n - y_n\| \leqslant ch^p$$
pour un $c > 0$ indépendant de h et de n.

Démonstration. On considère l'erreur $e_n := u_n - y_n$. On a

$$e_{n+1} = e_n + \Phi(t_n, u_n, h) - \Phi(t_n, y_n, h) - \tau_{n+1},$$

donc

$$|e_{n+1}| \leqslant (1 + Ch)|e_n| + Ch^{p+1}.$$

On termine en appliquant le Corollaire 2.14. □

2.3.4. *Méthode d'Euler explicite.*

Theorem 2.17 (Ordre de la méthode d'Euler explicite)**.** *Considérons que f est Lipschitzienne en sa seconde variable. La méthode d'Euler explicite est convergente d'ordre 1, c'est-à-dire que*

$$\max_{0 \leqslant n \leqslant N_h} |u_n - y_n| \leqslant Ch.$$

Démonstration. On utilise la formule de Taylor sur $y(t)$ autour de t_n, via Taylor-Lagrange

$$y_{n+1} = y(t_n + h) = y_n + hy'(t_n) + \frac{h^2}{2} y''(\xi_n),$$

pour un certain $\xi_n \in]t_n, t_{n+1}[$. Mais comme $y'(t) = f(t, y(t))$, cela donne

$$y_{n+1} = y_n + hf(t_n, y_n) + \frac{h^2}{2} y''(\xi_n).$$

L'erreur de consistance est

$$\sigma_{n+1} := y_{n+1} - \Phi\left(t_n, y_n, h\right) = \frac{h^2}{2} y''(\xi_n).$$

Sous l'hypothèse que y'' est bornée sur $[0, T]$, il existe $M > 0$ tel que $|y''(t)| \leqslant M$, et ainsi, pour tout n,

$$|\sigma_{n+1}| \leqslant \frac{M}{2} h^2 =: C_0 h^2,$$

et on voit que la méthode est consistante.

De plus, f est L-Lipschitzienne par rapport à sa seconde variable, donc

$$|\Phi(t, y, h) - \Phi(t, z, h)| \leqslant (1 + Lh)|y - z|.$$

On termine en appliquant la Proposition 2.16. □

2.3.5. *Méthode de Heun.*

Theorem 2.18 (Ordre de la méthode de Heun)**.** *La méthode de Heun est convergente d'ordre 2, c'est-à-dire que*

$$\max_{0 \leqslant n \leqslant N_h} |u_n - y_n| \leqslant Ch^2.$$

Démonstration. On a

$$y'(t_n) = f(t_n, y_n), \qquad y''(t_n) = \frac{\partial f}{\partial t}(t_n, y_n) + \frac{\partial f}{\partial y}(t_n, y_n)\, f(t_n, y_n).$$

Développons

$$\begin{aligned} y_{n+1} = y(t_n+h) &= y_n + h\,y'(t_n) + \frac{h^2}{2}y''(t_n) + O(h^3) \\ &= y_n + hf(t_n,y_n) + \frac{h^2}{2}\left(\frac{\partial f}{\partial t}(t_n,y_n) + \frac{\partial f}{\partial y}(t_n,y_n)f(t_n,y_n)\right) + O(h^3). \end{aligned}$$

Développons maintenant

$$\begin{aligned} &f\big(t_n+h,\, y_n+hf(t_n,y_n)\big) \\ &\quad = f(t_n,y_n) + h\,\frac{\partial f}{\partial t}(t_n,y_n) + h\,\frac{\partial f}{\partial y}(t_n,y_n)\,f(t_n,y_n) + O(h^2). \end{aligned}$$

On en déduit

$$\begin{aligned} \Phi(t_n,y_n,h) &= y_n + \frac{h}{2}\Big(f(t_n,y_n) + f\big(t_n+h,\, y_n+hf(t_n,y_n)\big)\Big) \\ &= y_n + hf(t_n,y_n) + \frac{h^2}{2}\left(\frac{\partial f}{\partial t}(t_n,y_n) + \frac{\partial f}{\partial y}(t_n,y_n)f(t_n,y_n)\right) + O(h^3). \end{aligned}$$

On a donc

$$\sigma_{n+1} = y_{n+1} - \Phi(t_n,y_n,h) = O(h^3).$$

Par ailleurs,

$$\begin{aligned} \Phi(t,y,h) - \Phi(t,z,h) = y - z &+ \frac{h}{2}\left(f(t,y) - f(t,z)\right) \\ &+ \frac{h}{2}\left(f\left(t+h, y+hf(t,y)\right) - f\left(t+h, z+hf(t,z)\right)\right) \end{aligned}$$

et

$$\begin{aligned} &\left|f\left(t+h, y+hf(t,y)\right) - f\left(t+h, z+hf(t,z)\right)\right| \\ &\quad \leqslant L\left|y+hf(t,y) - (z+hf(t,z))\right| \leqslant L\left(|y-z| + h\left|f(t,y) - f(t,z)\right|\right) \\ &\quad \leqslant L(1+Lh)\left|y-z\right|. \end{aligned}$$

Enfin,

$$|\Phi(t,y,h) - \Phi(t,z,h)| \leqslant (1 + Lh\,(1+Lh/2))\,|y-z|.$$

On termine en appliquant la Proposition 2.16. □

2.4. **Méthode d'Euler implicite.**

Theorem 2.19 (Ordre de la méthode d'Euler implicite)**.** *On suppose que f est L-Lipschitzienne en sa seconde variable. Pour $h < 1/L$, la méthode d'Euler implicite est convergente d'ordre 1.*

Démonstration. On définit d'abord l'erreur de consistance

$$\sigma_{n+1} := y_{n+1} - \big(y_n + hf(t_{n+1},y_{n+1})\big).$$

Un développement de Taylor de y autour de t_{n+1} donne

$$y_n = y_{n+1} - hy'(t_{n+1}) + \frac{h^2}{2}y''(\xi_{n+1}),$$

d'où

$$\sigma_{n+1} = -\frac{h^2}{2} y''(\xi_{n+1}), \qquad |\sigma_{n+1}| \leqslant Ch^2.$$

On introduit l'erreur $e_n := u_n - y_n$. En soustrayant l'identité vérifiée par y et le schéma numérique, on obtient :

$$e_{n+1} = e_n + h\big(f(t_{n+1}, u_{n+1}) - f(t_{n+1}, y_{n+1})\big) - \sigma_{n+1}.$$

Par hypothèse que f est L-Lipschitz,

$$|e_{n+1}| \leqslant |e_n| + hL\,|e_{n+1}| + |\sigma_{n+1}|,$$

donc

$$(1 - hL)|e_{n+1}| \leqslant |e_n| + |\sigma_{n+1}|.$$

Pour h assez petit tel que $1 - hL > 0$, on obtient

$$|e_{n+1}| \leqslant \frac{1}{1 - hL}\big(|e_n| + |\sigma_{n+1}|\big) \leqslant (1 + C_1 h)|e_n| + C_2 h^2,$$

avec des constantes C_1, C_2 indépendantes de h. On termine en appliquant le Corollaire 2.14. □

Proposition 2.20 (Ordre de la méthode de Crank–Nicolson)**.** *La méthode de Crank–Nicolson est convergente d'ordre* 2.

Démonstration. • On commence par prouver la consistance. On définit l'erreur de consistance

$$\sigma_{n+1} := y_{n+1} - \Big(y_n + \frac{h}{2}\big(f(t_n, y_n) + f(t_{n+1}, y_{n+1})\big)\Big),$$

et on veut connaître son comportement quand h est petit.

On part de la formulation intégrale de l'EDO,

$$y_{n+1} = y_n + \int_{t_n}^{t_{n+1}} f\big(s, y(s)\big)\, ds.$$

On applique la formule du trapèze à l'intégrale, en gardant $y(s)$ exact :

$$\int_{t_n}^{t_{n+1}} f\big(s, y(s)\big)\, ds = \frac{h}{2}\Big(f\big(t_n, y_n\big) + f\big(t_{n+1}, y_{n+1}\big)\Big) + \sigma_{n+1},$$

et on veut connaître l'ordre de σ_{n+1} en h.

On pose

$$g(s) := f\big(s, y(s)\big).$$

On suppose $g \in C^2$ sur l'intervalle considéré. On note $t_{n+1} = t_n + h$. On veut estimer le reste σ_{n+1} défini par

$$\int_{t_n}^{t_{n+1}} g(s)\, ds = \frac{h}{2}\Big(g(t_n) + g(t_{n+1})\Big) + \sigma_{n+1}.$$

Pour tout $s \in [t_n, t_{n+1}]$, il existe $\xi_s \in [t_n, t_{n+1}]$ tel que

$$g(s) = g(t_n) + (s - t_n)g'(t_n) + \frac{(s - t_n)^2}{2} g''(\xi_s).$$

On intègre cette identité de $s = t_n$ à $s = t_{n+1} = t_n + h$. En posant $\sigma = s - t_n$, on obtient

$$\begin{aligned}\int_{t_n}^{t_{n+1}} g(s)\,ds &= \int_0^h \left[g(t_n) + \sigma g'(t_n) + \frac{\sigma^2}{2} g''(\xi_{t_n+\sigma})\right] d\sigma \\ &= h\,g(t_n) + \frac{h^2}{2} g'(t_n) + \frac{1}{2}\int_0^h \sigma^2 g''(\xi_{t_n+\sigma})\,d\sigma.\end{aligned}$$

On applique Taylor en t_n au point $t_{n+1} = t_n + h$, il existe $\eta_n \in [t_n, t_{n+1}]$ tel que

$$g(t_{n+1}) = g(t_n) + hg'(t_n) + \frac{h^2}{2} g''(\eta_n).$$

La formule du trapèze vaut donc

$$\begin{aligned}\frac{h}{2}\big(g(t_n) + g(t_{n+1})\big) &= \frac{h}{2}\left(g(t_n) + g(t_n) + hg'(t_n) + \frac{h^2}{2} g''(\eta_n)\right), \\ &= h\,g(t_n) + \frac{h^2}{2} g'(t_n) + \frac{h^3}{4} g''(\eta_n).\end{aligned}$$

Par définition,

$$\sigma_{n+1} = \int_{t_n}^{t_{n+1}} g(s)\,ds - \frac{h}{2}\big(g(t_n) + g(t_{n+1})\big).$$

En utilisant les deux expressions précédentes, on obtient

$$\sigma_{n+1} = \frac{1}{2}\int_0^h \tau^2 g''(\xi_{t_n+\tau})\,d\tau - \frac{h^3}{4} g''(\eta_n).$$

On suppose que g'' est bornée sur l'intervalle $[t_n, t_{n+1}]$, par exemple

$$\big|g''(s)\big| \leqslant M \quad \text{pour tout } s \in [t_n, t_{n+1}].$$

Alors

$$\left|\frac{1}{2}\int_0^h \tau^2 g''(\xi_{t_n+\tau})\,d\tau\right| \leqslant \frac{1}{2} M \int_0^h \tau^2\,d\tau = \frac{Mh^3}{6},$$

et

$$\left|\frac{h^3}{4} g''(\eta_n)\right| \leqslant \frac{Mh^3}{4}.$$

On en déduit qu'il existe une constante $C > 0$, indépendante de h, telle que

$$|\tau_{n+1}| \leqslant Ch^3.$$

• Prouvons maintenant la convergence. On introduit l'erreur $e_n := u_n - y_n$. En soustrayant l'identité vérifiée par y et le schéma numérique, on obtient

$$e_{n+1} = e_n + \frac{h}{2}\big(f(t_n, u_n) - f(t_n, y_n) + f(t_{n+1}, u_{n+1}) - f(t_{n+1}, y_{n+1})\big) - \tau_{n+1}.$$

Par hypothèse que f est L-Lipschitz,

$$|e_{n+1}| \leqslant |e_n|\left(1 + \frac{h}{2}L\right) + \frac{h}{2}L|e_{n+1}| + |\tau_{n+1}|.$$

et

$$\left(1 - \frac{h}{2}L\right)|e_{n+1}| \leqslant |e_n|\left(1 + \frac{h}{2}L\right) + |\tau_{n+1}|.$$

Pour $h < 2/L$, on termine de la même manière que pour la méthode d'Euler implicite. $\square$

References