MA650

Laurent Dumas

CHAPITRE 1

PARTIE A : ANALYSE NUMERIQUE MATRICIELLE

cours réalisé en présentiel

CHAPITRE 2

PARTIE B : ETUDE NUMERIQUE DE FONCTIONS

2.1 Interpolation de Lagrange

cours réalisé en présentiel

2.2 Approximation de fonctions. Polynômes orthogonaux

cours réalisé en présentiel

2.3 Méthodes de quadrature

voir poly

CHAPITRE 3

PARTIE C : RESOLUTION NUMERIQUE DES EDO

3.1 Résolution numérique des EDO : définitions et notations

On part du problème de Cauchy suivant : trouver $y \in C^1([t_0, t_0+T], \mathbf{R}^m)$ tel que

$$\begin{cases} y'(t) = f(t, y(t)), \quad t \in [t_0, t_0+T], \\ y(t_0) \in \mathbf{R}^m \text{ fixé.} \end{cases} \tag{1}$$

Pour que ce problème possède une unique solution grâce au théorème de Cauchy-Lipschitz, nous supposerons par la suite (sauf mention contraire) que la fonction f est continue de $[t_0, t_0+T] \times \mathbf{R}^m$ dans $\mathbf{R}^m$ et est globalemement Lipschitzienne par rapport à sa seconde variable avec un coefficient de Lipschitz noté L pour la norme choisie sur $\mathbf{R}^m$:

$$\forall t \in [t_0, t_0+T], \quad \forall (y_1, y_2) \in (\mathbf{R}^m)^2, \quad ||f(t, y_2) - f(t, y_1)|| \leq L||y_2 - y_1||.$$

Une méthode de résolution approchée du problème (1) consiste d'abord à effectuer une subdvision de l'intervalle de définition de y

$$[t_0, t_0+T] = \bigcup_{n=0}^{N-1} [t_n^N, t_{n+1}^N], \quad N \in \mathbf{N}^*$$

puis à construire une suite $(y_n^N)_{0 \leq n \leq N}$ telle que y_n^N approche $y(t_n^N)$ pour tout $n \in \{0, .., N\}$.

Dans la suite de la construction, pour des raisons de clarté, les indices N seront omis. De plus, on se restreindra au cas scalaire $m = 1$ et aux méthodes à pas de temps constant, à savoir :

$$\forall n \in \{0, ..., N\}, \quad t_n = t_0 + n\Delta T \quad \text{avec} \quad \Delta T = \frac{T}{N}$$

3.2 Méthode d'Euler explicite

Définition : On appelle méthode de résolution approchée d'Euler explicite (à pas de temps constant) du problème (1), la construction pour tout $N \in \mathbf{N}^*$ de la suite finie $(y_n)_{0\leq n\leq N}$ telle que

$$\begin{cases} y_0 \in \mathbf{R}^m \text{ fixé,} \\ y_{n+1} = y_n + \Delta T f(t_n, y_n) \quad 0 \leq n \leq N-1. \end{cases} \tag{2}$$

Dans toute la suite, on désignera par e_n l'erreur commise en remplaçant $y(t_n)$ par y_n :

$$e_n = y_n - y(t_n).$$

Remarque : La méthode d'Euler revient en fait à approcher l'intégrale $\int_{t_n}^{t_{n+1}} y'(t)dt$ par la valeur $(t_{n+1} - t_n)f(t_n, y_n)$, c'est à dire à utiliser la méthode de quadrature élémentaire des rectangles à gauche.

3.3 Convergence de la méthode d'Euler explicite

Théorème 1. *: la méthode d'Euler explicite définie par la relation (2) vérifie l'inégalité suivante valable pour tout $n \in \{0, ..., N\}$ et toute solution y du problème (1) :*

$$||e_n|| \leq \frac{e^{nL\Delta T} - 1}{L} \omega(\Delta T, y') + e^{nL\Delta T} ||e_0||$$

où ω désigne le module de continuité d'une fonction :

$$\omega(\delta, g) = \max_{\{(t,t')\in[t_0,t_0+T]^2 \ / \ |t'-t|\leq\delta\}} ||g(t') - g(t)||$$

et L la constante de Lipschitz de f par rapport à la seconde variable.

Démonstration. : On commence par démontrer deux Lemmes techniques très utiles :

Lemme 1. *soit $(z_n)_{n\in\mathbf{N}}$ une suite réelle telle que*

$$\begin{cases} z_0 \geq 0, \\ z_{n+1} \leq Az_n + B \quad (n \in \mathbf{N}) \end{cases}$$

où A et B sont deux réels tels que $A \geq 1$ et $B \geq 0$. On a l'inégalité suivante valable pour tout $n \in \mathbf{N}$:

$$z_n \leq e^{nQ} z_0 + \frac{e^{nQ} - 1}{Q} B$$

avec $Q = A-1$ et la convention

$$\frac{e^{nQ}-1}{Q} = n \quad \text{lorsque } Q = 0.$$

Démonstration. On démontre ce résultat en appliquant n fois la propriété de la suite $(z_n)_{n\in\mathbf{N}}$:

$$\begin{cases} z_n \leq A^n z_0 + \dfrac{A^n-1}{A-1}B & \text{lorsque } A > 1, \\ z_n \leq z_0 + nB & \text{lorsque } A = 1 \end{cases}$$

et en remarquant que

$$A = Q + 1 \leq e^Q \quad \square$$

Lemme 2. *soit $(z_n)_{n\in\mathbf{N}}$ une suite réelle et $(h_n, \alpha_n)_{n\in\mathbf{N}} \in (\mathbf{R}_+^2)^{\mathbf{N}}$ tels que :*

$$\begin{cases} z_0 \geq 0, \\ z_{n+1} \leq (1+h_n)z_n + \alpha_n. \end{cases}$$

En notant

$$t_0 = 0 \quad \text{et} \quad t_n = \sum_{i=0}^{n-1} h_i \quad (n \in \mathbf{N}^*),$$

on a l'inégalité suivante valable pour tout $n \in \mathbf{N}$:

$$z_n \leq e^{t_n} z_0 + \sum_{i=0}^{n-1} e^{(t_n - t_{i+1})} \alpha_i.$$

Démonstration. : Il suffit d'adapter la démonstration précédente $\square$

Pour démontrer l'estimation de l'erreur de la méthode d'Euler, on considère une fonction y solution du problème (1). On a :

$$\begin{aligned} y(t_{n+1}) &= y(t_n) + \int_{t_n}^{t_{n+1}} y'(t)dt \\ &= y(t_n) + \Delta T f(t_n, y(t_n)) + \int_{t_n}^{t_{n+1}} (y'(t) - y'(t_n))dt \\ &= y(t_n) + \Delta T f(t_n, y(t_n)) + \varepsilon_n \end{aligned}$$

en notant $\varepsilon_n = \int_{t_n}^{t_{n+1}} (y'(t) - y'(t_n))dt$. On peut rapprocher cette égalité de la définition (2) de la suite des approximations :

$$y_{n+1} = y_n + \Delta T f(t_n, y_n)$$

En soustrayant ces deux égalités et en utilisant la définition du module de continuité, on obtient

$$||y(t_{n+1}) - y_{n+1}|| \leq ||y(t_n) - y_n|| + L\Delta T||y(t_n) - y_n|| + \Delta T\,\omega(\Delta T, y').$$

On se trouve alors dans les conditions d'application du Lemme 1. Il vient :

$$||y(t_n) - y_n|| = ||e_n|| \leq e^{nL\Delta T}||e_0|| + \frac{e^{nL\Delta T} - 1}{L\Delta T}\Delta T\,\omega(\Delta T, y')$$

et le résultat annoncé est bien démontré □

Le corollaire suivant peut être déduit de la convergence de la méthode d'Euler :

Corollaire 1. *soit la suite de fonctions $(u_N)_{N\in\mathbf{N}}$ définies sur $[t_0, t_0+T]$ par interpolation affine entre les points $(t_n)_{0\leq n\leq N}$ pour lesquels $u_N(t_n) = y_n$ où la famille $(y_n)_{0\leq n\leq N}$ est construite suivant la méthode d'Euler (2) avec $y_0 \in \mathbf{R}^m$ fixé. Alors u_N converge dans $L^\infty([t_0, t_0+T], \mathbf{R}^m)$ vers l'unique solution y de (1) telle que $y(t_0) = y_0$.*

Démonstration. : Soit N fixé et $t \in [t_0, t_0 + T[$. On note n l'unique entier compris entre 0 et $N-1$ tel que $t \in [t_n, t_{n+1}[$.On a alors l'estimation suivante :

$$\begin{aligned} ||u_N(t) - y(t)|| &\leq ||u_N(t) - y_n|| + ||y_n - y(t_n)|| + ||y(t_n) - y(t)|| \\ &\leq ||u_N(t) - y_n|| + ||e_n|| + \omega(\Delta T, y). \end{aligned}$$

u_N étant affine sur $[t_n, t_{n+1}]$, on a

$$||u_N(t) - y_n|| \leq ||y_{n+1} - y_n|| = \Delta T||f(t_n, y_n)||$$

Il suffit donc pour conclure à l'uniforme convergence de u_N vers y de montrer que y_n reste borné indépendamment de n et N. Pour cela, on remarque d'abord que

$$\begin{aligned} ||y_{n+1}|| \leq ||y_n|| + \Delta T||f(t_n, y_n)|| &\leq ||y_n|| + \Delta T(||f(t_n, 0)|| + L||y_n||) \\ &\leq (1 + \Delta T L)||y_n|| + \Delta T \max_{t\in[t_0, t_0+T]} ||f(t, 0)|| \end{aligned}$$

et on utilise ensuite le Lemme 1 :

$$||y_n|| \leq e^{TL}||y_0|| + \frac{e^{TL} - 1}{L} \max_{t\in[t_0, t_0+T]} ||f(t, 0)||$$

□

L'estimation de l'erreur donnée dans le Théorème 1, peu utilisable en pratique (car dépendant de la fonction y' a priori inconnue) peut être reformulée ou améliorée avec des hypothèses supplémentaires sur f :

Proposition 1. *Les notations et hypothèses du Théorème 8 sont conservées. On note K un compact de* $\mathbf{R}^m$ *tel que* $y([t_0, t_0+T]) \subset K$ *et* $Q = [t_0, t_0+T] \times K$. *Alors :*

$$||e_n|| \leq C\frac{e^{nL\Delta T}-1}{L} + e^{nL\Delta T}||e_0||$$

avec

$$C = L\Delta T \sup_{(t,z)\in Q} ||f(t,z)|| + \sup_{z\in K} \omega(\Delta T, t \mapsto f(t,z))$$

Si de plus $f \in C^1(Q, \mathbf{R}^m)$, *on a pour tout* $n \in \{0,...,N\}$

$$||e_n|| \leq \Delta T \int_{t_0}^{t_n} e^{L(t_n-z)}||y''(z)||dz + e^{nL\Delta T}||e_0||$$

et

$$||e_n|| \leq \frac{\Delta T}{2} \max_{(t,z)\in Q} ||f^{[1]}(t,z)||\frac{e^{nL\Delta T}-1}{L} + e^{nL\Delta T}||e_0||.$$

Démonstration. : La première inégalité revient seulement à déterminer l'expression de $\omega(\Delta T, y')$ en fonction de f. On a pour tout $(s,s') \in [t_0, t_0+T]^2$:

$$\begin{aligned} y'(s')-y'(s) &= f(s',y(s'))-f(s,y(s)) \\ &= f(s',y(s'))-f(s',y(s))+f(s',y(s))-f(s,y(s)) \end{aligned}$$

et ainsi

$$\begin{aligned} |s'-s| \leq \Delta T \Rightarrow ||y'(s')-y'(s)|| &\leq L||y(s')-y(s)|| + \omega(\Delta T, t \mapsto f(t,y(s))) \\ &\leq L\Delta T \sup_{\xi\in[s,s']} ||y'(\xi)|| + \omega(\Delta T, t \mapsto f(t,y(s))) \\ &\leq L\Delta T \sup_{(t,z)\in Q} ||f(t,z)|| + \sup_{z\in K} \omega(\Delta T, t \mapsto f(t,z)) \end{aligned}$$

Pour démontrer la deuxième inégalité, on remarque tout d'abord que

$$\forall n \in \{0,...,N-1\}, \quad y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} y'(s)ds = y(t_n) + \Delta T f(t_n, y(t_n)) + \alpha_n$$

avec

$$\alpha_n = \int_{t_n}^{t_{n+1}} (\int_{t_n}^{s} y''(z)dz)ds$$

En raisonnant alors exactement comme dans la démonstration du Théorème 8 (avec α_n à la place de ε_n et en utilisant cette fois le Lemme 2), il vient

$$\begin{aligned}||e_n|| - e^{nL\Delta T}||e_0|| \leq \sum_{i=0}^{n-1} e^{L(t_n - t_{i+1})}|\alpha_i| &\leq \sum_{i=0}^{n-1} e^{L(t_n - t_{i+1})} \int_{t_i}^{t_{i+1}} (\int_{t_i}^{s} ||y''(z)||dz)ds \\ &\leq \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} (\int_{t_i}^{t_{i+1}} e^{L(t_n - z)}||y''(z)||dz)ds \\ &\leq \Delta T \int_{t_0}^{t_n} e^{L(t_n - z)}||y''(z)||dz.\end{aligned}$$

La dernière inégalité se déduit de la même façon que la précédente en utilisant la majoration de α_n suivante :

$$||\alpha_n|| \leq \frac{\Delta T^2}{2} \max_{s \in [t_0, t_0+T]} ||y''(s)|| \leq \frac{\Delta T^2}{2} \max_{(t,z)\in Q} ||f^{[1]}(t,z)||$$

□

3.4 Méthode d'Euler implicite

Il existe une version implicite de la méthode d'Euler qui peut s'avérer à l'usage mieux conditionnée que la version explicite (voir exemple ultérieurement).

Définition : On appelle méthode de résolution approchée d'Euler implicite (à pas de temps variable) de l'équation (1), la construction valable pour tout $N \in \mathbf{N}^*$ tel que $\Delta T L < 1$ d'une suite finie $(y_n)_{0 \leq n \leq N}$ vérifiant

$$\begin{cases} y_0 \in \mathbf{R}^m \text{ fixé,} \\ y_{n+1} = y_n + \Delta T f(t_n, y_{n+1}) \quad 0 \leq n \leq N-1. \end{cases} \tag{7}$$

Théorème 2. *: la méthode d'Euler implicite est convergente au sens suivant : si on note* $L_1 = \dfrac{L}{1 - L\Delta T}$*, on a une estimation de l'erreur commise :*

$$\forall n \in \{0, ..., N\}, \quad ||e_n|| \leq \frac{e^{L_1(t_n - t_0)} - 1}{L_1} \omega(\Delta T y') + e^{L_1(t_n - t_0)}||e_0||$$

si f est seulement continue. Si f est continûment dérivable, on a de plus

$$||e_n|| \leq \frac{\Delta T}{1 - L\Delta T} \int_{t_0}^{t_n} e^{L_1(t_n - s)}||y''(s)||ds + e^{L_1(t_n - t_0)}||e_0||.$$

Démonstration. La démonstration de ce théorème est réalisée en exercice (voir feuille de TD).

3.5 Cas général d'une méthode à un pas

On s'intéresse à présent à étudier les propriétés générales des méthodes à un pas (dont la méthode d'Euler est un cas particulier).

Pour rappel une méthode de résolution approchée du problème (1) à un pas consiste d'abord à effectuer une subdvision de l'intervalle de définition de y

$$[t_0, t_0 + T] = \bigcup_{n=0}^{N-1} [t_n^N, t_{n+1}^N], \quad N \in \mathbf{N}^*$$

puis à construire une suite $(y_n^N)_{0 \leq n \leq N}$ telle que y_n^N approche $y(t_n^N)$ pour tout $n \in \{0, .., N\}$.

Dans la suite de la construction, pour des raisons de clarté, les indices N seront omis. De plus, on se restreindra au cas scalaire $m = 1$ et aux méthodes à pas de temps constant, à savoir :

$$\forall n \in \{0, ..., N\}, \quad t_n = t_0 + n\Delta T \quad \text{avec} \quad \Delta T = \frac{T}{N}$$

Définition 1. *On appelle méthode de résolution approchée explicite à un pas (et à pas de temps constant) du problème (1), la construction pour tout $N \in \mathbf{N}^*$ d'une suite finie $(y_n)_{0 \leq n \leq N}$ telle que*

$$\begin{cases} y_0 \in \mathbf{R} \text{ fixé}, \\ y_{n+1} = y_n + \Delta T \Phi(t_n, y_n, \Delta T) \quad 0 \leq n \leq N-1, \end{cases} \tag{3}$$

où Φ désigne une fonction continue de $[t_0, t_0 + T] \times \mathbf{R} \times [0, T]$ dans $\mathbf{R}$.

La méthode d'Euler explicite est donc un cas particulier d'une méthode à un pas où :

$$\forall (t, y, h) \in [t_0, t_0 + T] \times \mathbf{R} \times [0, T], \quad \Phi(t, y, h) = f(t, y)$$

3.6 Consistance, stabilité, convergence et ordre d'une méthode à un pas

On définit ci-dessous les différentes propriétés importantes éventuellement satisfaites par une méthode à un pas.

Définition 2. *On dit qu'une méthode à un pas de type (3) est consistante avec problème (1) si pour toute solution y de celui-ci, on a*

$$\lim_{N\to+\infty}\sum_{n=0}^{N-1}|\varepsilon_n| = 0$$

où

$$\varepsilon_n = y(t_{n+1}) - y(t_n) - \Delta T\Phi(t_n, y(t_n), \Delta T)$$

Définition 3. *On dit qu'une méthode à un pas de type (3) est stable si pour tout entier $N \in \mathbf{N}^*$ et pour toute famille $(z_n)_{0\le n\le N}$ solution de la relation de récurrence perturbée par le réel ε_n :*

$$\begin{cases} z_0 \in \mathbf{R} \text{ fixé}, \\ z_{n+1} = z_n + \Delta T\Phi(t_n, z_n, \Delta T) + \varepsilon_n \quad 0 \le n \le N-1, \end{cases}$$

on a la relation

$$\max_{0\le n\le N-1}|z_n - y_n| \le M|z_0 - y_0| + M'\sum_{n=0}^{N-1}|\varepsilon_n|$$

où M et M' sont des constantes indépendantes de y_0, z_0 et N.

Définition 4. *On dit qu'une méthode à un pas de type (3) est convergente si pour toute solution y du problème (1)*

$$\lim_{\substack{N\to+\infty \\ y_0\to y(t_0)}} \max_{0\le n\le N}|y(t_n) - y_n| = 0.$$

Définition 5. *On dit qu'une méthode à un pas de type (3) est d'ordre $p \in \mathbf{N}^*$ si Φ et f sont p-fois continûment différentiables sur leur ensemble de définition et si pour toute solution y du problème (1), il existe une constante $C > 0$ indépendante de N telle que*

$$\forall N \in \mathbf{N}^*, \quad \sum_{n=0}^{N-1}|\varepsilon_n| \le C(\Delta T)^p$$

où

$$\varepsilon_n = y(t_{n+1}) - y(t_n) - \Delta T\Phi(t_n, y(t_n), \Delta T) \quad (0 \le n \le N-1).$$

Remarque : La notion de consistance revient en fait à s'assurer que la méthode considérée est cohérente (ou consistante) avec le problème initial. La stabilité signifie que les éventuelles erreurs numériques commises dans l'évaluation des termes de la

suite d'approximations sont contrôlables. Enfin, la notion d'ordre est reliée avec la vitesse de convergence d'une méthode si celle-ci est stable.

Les sous-paragraphes suivants vont s'efforcer de donner des conditions nécessaires et/ou suffisantes sur Φ permettant de vérifier ces différentes propriétés. A noter que les deux lemmes vus dans la démonstration de convergence de la méthode d'Euler seront en particulier réutilisés.

Les démonstrations de ces théorèmes ne sont pas exigibles et sont présentées à titre indicatif, sauf la démonstration de la convergence (théorème 5).

3.6.1 Consistance d'une méthode explicite à un pas

Théorème 3. *: une méthode explicite à un pas de type (3) est consistante si et seulement si*

$$\forall (t,y) \in [t_0, t_0+T] \times \mathbf{R}, \quad \Phi(t,y,0) = f(t,y).$$

Démonstration. : Soit $\varepsilon_n = y(t_{n+1}) - y(t_n) - \Delta T \Phi(t_n, y(t_n), \Delta T) \quad (0 \leq n \leq N-1)$. Par le théorème des acroissements finis, on sait qu'il existe $c_n \in]t_n, t_{n+1}[$ tel que

$$\varepsilon_n = \Delta T f(c_n, y(c_n)) - \Delta T \Phi(t_n, y(t_n), \Delta T) = \Delta T(\alpha_n + \beta_n)$$

avec

$$\begin{cases} \alpha_n = f(c_n, y(c_n)) - \Phi(c_n, y(c_n), 0), \\ \beta_n = \Phi(c_n, y(c_n), 0) - \Phi(t_n, y(t_n), \Delta T). \end{cases}$$

Comme la fonction $\tilde{\Phi} : \begin{pmatrix} [t_0, t_0+T] \times [0,T] \to \mathbf{R} \\ (t,h) \mapsto \Phi(t, y(t), h) \end{pmatrix}$ est continue sur l'ensemble compact $[t_0, t_0+T] \times [0,T]$, donc uniformément continue, on peut affirmer que

$$\forall \varepsilon > 0, \exists N_0 \in \mathbf{N}^*, \forall N \geq N_0, \forall n \in \{0, ..., N-1\}, |\beta_n| \leq \varepsilon.$$

Ainsi, pour $N \geq N_0$,

$$\left| \sum_{n=0}^{N-1} |\varepsilon_n| - \sum_{n=0}^{N-1} \Delta T |\alpha_n| \right| \leq \sum_{n=0}^{N-1} \Delta T |\beta_n| \leq \varepsilon T.$$

De plus, par définition de l'intégrale de Rieman,

$$\lim_{N \to +\infty} \sum_{n=0}^{N-1} \Delta T |\alpha_n| = \int_{t_0}^{t_0+T} |f(t, y(t)) - \Phi(t, y(t), 0)| dt.$$

Au vu de la définition et des remarques précédentes, une condition nécessaire et suffisante de consistance de la méthode étudiée est donc que pour toute solution y de (1) on ait

$$\forall t \in [t_0, t_0+T], \quad f(t, y(t)) = \Phi(t, y(t), 0). \tag{4}$$

Or, pour tout couple $(t^*, y^*) \in [t_0, t_0+T] \times \mathbf{R}$, il existe par le théorème de Cauchy-Lipschitz une (unique) solution y définie sur $[t_0, t_0+T]$ du problème de Cauchy suivant :

$$\begin{cases} y'(t) = f(t, y(t)), & t \in [t_0, t_0+T], \\ y(t^*) = y^*. \end{cases}$$

En écrivant la relation (4) en t^* pour cette fonction, également solution d'un problème de type (1), on obtient :

$$\Phi(t^*, y^*, 0) = f(t^*, y^*)$$

ce qui est bien le résultat recherché □

3.6.2 Stabilité d'une méthode explicite à un pas

Théorème 4. *: pour qu'une méthode explicite à un pas de type (3) soit stable, il suffit que la fonction Φ soit Lipschitzienne en y, à savoir qu'il existe $\Lambda > 0$ tel que*

$$\forall (t, y_1, y_2, h) \in [t_0, t_0+T] \times \mathbf{R}^2 \times [0, T], \quad |\Phi(t, y_2, h) - \phi(t, y_1, h)| \leq \Lambda |y_2 - y_1|$$

On peut alors prendre comme constantes de stabilité $M = M' = e^{\Lambda T}$.

Démonstration. : En conservant les notations de la définition de la stabilité, on peut écrire sous les hypothèses du théorème :

$$\begin{aligned} |y_{n+1} - z_{n+1}| &= |y_n - z_n + \Delta T\,(\Phi(t_n, y_n, \Delta T) - \Phi(t_n, z_n, \Delta T)) - \varepsilon_n| \\ &\leq (1 + \Lambda \Delta T)|y_n - z_n| + |\varepsilon_n| \end{aligned}$$

Grâce au Lemme 2, on a alors

$$\begin{aligned} |y_n - z_n| &\leq e^{n\Lambda\Delta T}|y_0 - z_0| + \sum_{i=0}^{n-1} e^{(n-1-i)\Lambda\Delta T}|\varepsilon_i| \\ &\leq e^{\Lambda T}|y_0 - z_0| + e^{\Lambda T}\sum_{i=0}^{N-1}|\varepsilon_i| \end{aligned}$$

ce qui clôt la démonstration □

Remarque : Étant sous forme d'exponentielles, les constantes de stabilité peuvent malheureusement être très grandes (voir exemples ultérieurs). Cette estimation ne peut en fait être améliorée dans le cas général (voir l'exemple trivial où $y' = \lambda y$ avec $\lambda > 0$).

3.6.3 Convergence d'une méthode explicite à un pas

Théorème 5. *: si une méthode explicite à un pas est stable et consistante, alors elle est convergente.*

Démonstration. : On pose $\varepsilon_n = y(t_{n+1}) - y(t_n) - \Delta T \Phi(t_n, y(t_n), \Delta T) \quad (0 \leq n \leq N-1)$. La famille $(z_n)_{0 \leq n \leq N-1}$ solution de la relation de récurrence perturbée

$$\begin{cases} z_0 = y(t_0), \\ z_{n+1} = z_n + \Delta T \Phi(t_n, z_n, \Delta T) + \varepsilon_n \quad (0 \leq n \leq N-1) \end{cases}$$

vérifie facilement par construction même :

$$\forall n \in \{0, ..., N\}, \quad z_n = y(t_n).$$

La méthode considérée étant stable et consistante, il vient respectivement

$$\max_{0 \leq n \leq N-1} |y_n - y(t_n)| \leq M |y_0 - y(t_0)| + M' \sum_{n=0}^{N-1} |\varepsilon_n|$$

et

$$\lim_{N \to +\infty} \sum_{n=0}^{N-1} |\varepsilon_n| = 0.$$

En regroupant ces deux résultats, on a bien démontré la convergence de la méthode, à savoir :

$$\lim_{\substack{N \to +\infty \\ y_0 \to y(t_0)}} \max_{0 \leq n \leq N} |y(t_n) - y_n| = 0$$

□

Le corollaire suivant permet de déterminer une condition suffisante de convergence d'une méthode simplement à partir d'informations sur Φ :

Corollaire 2. *si la fonction Φ associée à une méthode explicite à un pas est Lispchitzienne par rapport à la variable y et si*

$$\forall (t, y) \in [t_0, t_0 + T] \times \mathbf{R} \quad \Phi(t, y, 0) = f(t, y),$$

alors la méthode est convergente.

Démonstration. : Il suffit d'utiliser les Théorèmes 3, 4 et 5 □

Remarque : On retrouve avec ce corollaire que la méthode d'Euler explicite est bien convergente.

3.6.4 Ordre d'une méthode explicite à un pas

Théorème 6. *: une méthode explicite à un pas de type (3) est d'ordre $p \in \mathbf{N}^*$ si et seulement si Φ et f sont p-fois continûment différentiables sur leur ensemble de définition et si*

$$\forall l \in \{0,...,p-1\}, \quad \forall (t,y) \in [t_0, t_0+T] \times \mathbf{R}, \ \frac{\partial^l}{\partial h^l}\Phi(t,y,0) = \frac{1}{l+1} f^{[l]}(t,y), \quad (5)$$

où $f^{[l]}(t,y)$ désigne la l-ième dérivée totale de f suivant les caractéristiques du problème (1) :

$$\begin{cases} f^{[0]}(t,y) = f(t,y), \\ f^{[k+1]}(t,y) = \partial_t f^{[k]}(t,y) + \nabla_y f^{[k]}(t,y) \cdot f(t,y) \quad (0 \le k \le p-1). \end{cases}$$

Démonstration. : Condition suffisante : on remarque tout d'abord que si $f \in C^p([t_0, t_0+T] \times \mathbf{R}, \mathbf{R})$, alors toute solution y de (1) appartient à $C^{p+1}([t_0, t_0+T], \mathbf{R})$ et vérifie

$$\forall l \in \{0,...,p\}, \quad \forall t \in [t_0, t_0+T], \quad y^{(l+1)}(t) = f^{[l]}(t, y(t)).$$

On écrit alors la relation de Taylor-Lagrange à l'ordre p pour la fonction $h \mapsto \Phi(t_n, y(t_n), h)$ entre 0 et ΔT :

$$\Phi(t_n, y(t_n), \Delta T) = \sum_{l=0}^{p-1} \frac{(\Delta T)^l}{l!} \frac{\partial^l}{\partial h^l}\Phi(t_n, y(t_n), 0) + \frac{(\Delta T)^p}{p!} \frac{\partial^p}{\partial h^p}\Phi(t_n, y(t_n), \lambda_n)$$

($\lambda_n \in]0, \Delta T[$) et à l'ordre $p+1$ pour la fonction y entre t_n et t_{n+1} :

$$\begin{aligned} y(t_{n+1}) - y(t_n) &= \sum_{k=1}^{p} \frac{(\Delta T)^k}{k!} y^{(k)}(t_n) + \frac{(\Delta T)^{p+1}}{(p+1)!} y^{(p+1)}(c_n) \qquad (c_n \in]t_n, t_{n+1}[) \\ &= \sum_{k=1}^{p} \frac{(\Delta T)^k}{k!} f^{[k-1]}(t_n, y(t_n)) + \frac{(\Delta T)^{p+1}}{(p+1)!} y^{(p+1)}(c_n). \end{aligned}$$

En soustrayant ces deux égalités après un changement d'indices ($k = l+1$), il vient

$$\begin{aligned} \varepsilon_n &= y(t_{n+1}) - y(t_n) - \Delta T \Phi(t_n, y(t_n), \Delta T) \\ &= \sum_{l=0}^{p-1} \frac{(\Delta T)^{l+1}}{l!} \left[\frac{f^{[l]}(t_n, y(t_n))}{l+1} - \frac{\partial^l \Phi(t_n, y(t_n), 0)}{\partial h^l} \right] \\ &+ \frac{(\Delta T)^{p+1}}{p!} \left[\frac{y^{(p+1)}(c_n)}{p+1} - \frac{\partial^p}{\partial h^p}\Phi(t_n, y(t_n), \lambda_n) \right] \\ &= \frac{(\Delta T)^{p+1}}{p!} \left[\frac{y^{(p+1)}(c_n)}{p+1} - \frac{\partial^p}{\partial h^p}\Phi(t_n, y(t_n), \lambda_n) \right] \end{aligned}$$

grâce à la relation (5). Ainsi, avec la régularité supposée de Φ et f

$$\exists C > 0, \quad \forall n \in \{0, ..., N-1\}, \quad |\varepsilon_n| \leq C(\Delta T)^{p+1}.$$

En sommant ces N inégalités, on trouve bien que la méthode est d'ordre p. Condition nécessaire : on raisonne par l'absurde en supposant que la méthode est d'ordre p et que la relation (4) est violée à partir d'un certain $l \in \{0, ..., p-1\}$. En écrivant les mêmes égalités de Taylor-Lagrange que précédemment mais cette fois à l'ordre $l+1$ et $l+2$ respectivement, il vient

$$\varepsilon_n = \frac{(\Delta T)^{l+1}}{l!}\Big[\frac{f^{[l]}(t_n, y(t_n))}{l+1} - \frac{\partial^l}{\partial h^l}\Phi(t_n, y(t_n), 0)\Big] + O((\Delta T)^{l+2})$$

puis en sommant

$$\sum_{n=0}^{N-1}\Big|\frac{\varepsilon_n}{(\Delta T)^l}\Big| = \frac{1}{l!}\sum_{n=0}^{N-1}\Delta T|\psi(t_n)| + O(\Delta T)$$

où ψ désigne la fonction continue $\psi : \begin{pmatrix} [t_0, t_0+T] \to \mathbf{R} \\ t \mapsto \dfrac{f^{[l]}(t, y(t))}{l+1} - \dfrac{\partial^l}{\partial h^l}\Phi(t, y(t), 0) \end{pmatrix}$.

On déduit alors par passage à la limite dans les deux membres de la dernière égalité lorsque ΔT tend vers 0 que

$$0 = \frac{1}{l!}\int_{t_0}^{t_0+T} |\psi(s)|ds$$

ce qui implique immédiatement que pour toute fonction y solution de (1)

$$\forall t \in [t_0, t_0+T], \quad \psi(t) = \frac{f^{[l]}(t, y(t))}{l+1} - \frac{\partial^l}{\partial h^l}\Phi(t, y(t), 0) = 0.$$

En raisonnant comme dans la fin de la démonstration du Théorème 3, on en déduit facilement que

$$\forall (t^*, y^*) \in [t_0, t_0+T] \times \mathbf{R} \quad \frac{\partial^l}{\partial h^l}\Phi(t^*, y^*, 0) = \frac{f^{[l]}(t^*, y^*)}{l+1}$$

ce qui achève la démonstration par l'absurde □

Remarque : On retrouve facilement avec ce théorème que la méthode d'Euler est exactement d'ordre 1.

Cette présentation générale des propriétés des méthodes à un pas va permettre à présent dans le dernier paragraphe de construire des méthodes d'ordre plus élevé que les méthodes d'Euler, et donc plus précises (méthode de Runge Kutta).

3.7 Résolution approchée : les méthodes de Runge-Kutta

La famille des méthodes de Runge-Kutta explicites construites dans ce paragraphe (et à laquelle appartient la méthode d'Euler explicite) est un exemple important de méthodes explicites à un pas.

3.7.1 Définition

On suppose pour simplifier que $m = 1$ et que le pas de temps est constant :

$$\begin{cases} \Delta T = \dfrac{T}{N}, \\ t_n = t_0 + n\Delta T \quad (0 \leq n \leq N). \end{cases}$$

Définition : Soit $q \in \mathbf{N}^*$ et $(a_{i,j})_{1\leq j<i\leq q}$, $(b_i)_{1\leq i\leq q}$ et $(c_i)_{1\leq i\leq q}$ trois familles de coefficients réels (avec $c_i \in [0,1]$). On appelle méthode de résolution approchée de Runge-Kutta explicite à un pas (et à pas de temps constant) du problème (1), la construction pour tout $N \in \mathbf{N}^*$ de la suite finie $(y_n)_{0\leq n\leq N}$ telle que

$$\begin{cases} y_0 \in \mathbf{R}^m \text{ fixé}, \\ t_{n,j} = t_n + c_j \Delta T, \quad 1 \leq j \leq q, \\ y_{n,i} = y_n + \Delta T \displaystyle\sum_{j=1}^{i-1} a_{i,j} f(t_{n,j}, y_{n,j}), \quad 1 \leq i \leq q, \\ y_{n+1} = y_n + \Delta T \displaystyle\sum_{j=1}^{q} b_j f(t_{n,j}, y_{n,j}), \quad 0 \leq n \leq N-1. \end{cases} \tag{6}$$

On représente conventionnellement une méthode de Runge-Kutta par le tableau de ses coefficients rangés de la manière suivante :

$$\begin{array}{c|cccc} c_1 & 0 & & & \\ c_2 & a_{2,1} & 0 & & \\ c_3 & a_{3,1} & a_{3,2} & 0 & \\ \vdots & \vdots & \vdots & \ddots & \ddots \\ c_q & a_{q,1} & a_{q,2} & \dots & a_{q-1,q}\ 0 \\ \hline & b_1 & b_2 & \dots & b_{q-1} b_q \end{array}$$

Remarque : Il est également possible de définir des méthodes de Runge-Kutta implicites. Dans ce cas, le tableau des coefficients n'est plus triangulaire inférieur strict mais rectangulaire.

3.7.2 Interprétation

Pour mieux comprendre la définition des méthodes de Runge-Kutta, les formules (6) sont à rapprocher de celles valables pour toute solution exacte y du problème (1) :

$$\begin{cases} y(t_{n,i}) = y(t_n) + \Delta T \displaystyle\int_0^{c_i} f(t_n + u\Delta T, y(t_n + u\Delta T))du, \quad 1 \leq i \leq q, \\ y(t_{n+1}) = y(t_n) + \Delta T \displaystyle\int_0^{1} f(t_n + u\Delta T, y(t_n + u\Delta T))du \quad 0 \leq n \leq N-1. \end{cases}$$

Le principe des méthodes de Runge-Kutta consiste donc à approcher successivement par une méthode de quadrature, $y(t_{n,i})$ pour tout $i \in \{1, ..., q\}$, puis $y(t_{n+1})$ à l'aide des précédentes valeurs calculées.

Ainsi, les familles de coefficients $(a_{i,j})_{1 \leq j < i \leq q}$ et $(b_i)_{1 \leq i \leq q}$ sont associées aux méthodes de quadrature :

$$\begin{cases} \displaystyle\int_0^{c_i} g(t)dt \simeq \sum_{j=1}^{i-1} a_{i,j} g(c_j), \quad 1 \leq i \leq q, \\ \displaystyle\int_0^{1} g(t)dt \simeq \sum_{j=1}^{q} b_j g(c_j). \end{cases}$$

On obtient ainsi deux premières conditions (qui seront supposées vérifiées dans toutes la suite) sur les familles de coefficients pour que les méthodes de quadrature soient au moins d'ordre 0 (c'est à dire exactes sur les constantes) :

$$\forall i \in \{1, ..., q\}, \quad c_i = \sum_{j=1}^{i-1} a_{i,j} \tag{7}$$

et

$$1 = \sum_{j=1}^{q} b_j. \tag{8}$$

En particulier, ceci impose que $c_1 = 0$ (et aussi par conséquent $t_{n,1} = t_n$ et $y_{n,1} = y_n$).

3.7.3 Stabilité des méthodes de Runge-Kutta explicites

Les méthodes de Runge-Kutta explicites sont un cas particulier de méthodes explicites à un pas correspondant à une fonction Φ égale à :

$$\Phi(t,y,h) = \sum_{j=1}^{q} b_j f(t + c_j h, y_j)$$

où la famille $(y_i)_{1\leq i\leq q}$ est définie pout tout (t,y,h) par

$$y_i = y + h\sum_{j=1}^{i-1} a_{i,j} f(t + c_j h, y_j), \quad 1 \leq i \leq q. \tag{9}$$

On a alors le résultat suivant :

Proposition 2. *Les méthodes de Runge-Kutta explicites sont stables et la constante de stabilité $M = e^{\Lambda T}$ est donnée par les relations (10) et (11).*

Démonstration. : Grâce au Théorème 4, il suffit de montrer que Φ est Lipschitzienne par rapport à la variable y. Pour cela, soit

$$\alpha = \max_{1\leq i\leq q} \sum_{j=1}^{i-1} |a_{i,j}|. \tag{10}$$

Si $(y_i)_{1\leq i\leq q}$ et $(z_i)_{1\leq i\leq q}$ sont deux familles construites suivant la formule (9) (à partir de y et z respectivement), on montre aisément par récurrence l'inégalité

$$|y_i - z_i| \leq \left(1 + (\alpha L h) + \ldots + (\alpha L h)^{i-1}\right)|y - z|$$

où L désigne la constante de Lispschitz de f par rapport à sa deuxième variable. On peut alors écrire pour tout $(t,h) \in [t_0, t_0 + T] \times [0, \Delta T]$:

$$|\Phi(t,y,h) - \Phi(t,z,h)| \leq \sum_{j=1}^{q} |b_j| L |y_j - z_j| \leq \Lambda |y - z|$$

où

$$\Lambda = L\sum_{j=1}^{q} |b_j| \left(1 + (\alpha L \Delta T) + \ldots + (\alpha L \Delta T)^{j-1}\right), \tag{11}$$

ce qui achève la démonstration □

Remarque : On peut estimer le coefficient Λ lorsque le pas de temps ΔT est petit et lorsque les coefficients b_j sont tous positifs : en effet, dans ce cas et grâce à la relation (8), on a

$$\Lambda \leq L(1+(\alpha L\Delta T)+\ldots+(\alpha L\Delta T)^{q-1}) = L\frac{1-(\alpha L\Delta T)^q}{1-\alpha L\Delta T} \simeq L$$

si $\Delta T \ll \frac{1}{\alpha L}$.

3.7.4 Convergence et ordre des méthodes de Runge-Kutta explicites

Théorème 7. *: toute méthode de Runge-Kutta explicite vérifiant (6) (7) et (8) est convergente et d'ordre* 1 *(si f est C^1). Une condition nécessaire et suffisante pour qu'elle soit d'ordre* 2 *(si f est C^2) est que ses coefficients vérifient en outre :*

$$\sum_{j=1}^{q} b_j c_j = \frac{1}{2}. \tag{12}$$

Démonstration. La preuve de ce théorème est demandée en exercice (voir feuille de TD). □

Remarque : Les conditions nécessaires et suffisantes sur les coefficients pour obtenir des méthodes de Runge-Kutta d'ordre supérieur à 2 deviennent rapidement très complexes à exprimer. Elles sont données par les solutions d'un système polynomial à plusieurs indéterminées. Le recours à un logiciel de calcul formel comme Maple s'avère alors très utile : voir [GH] pour un exemple de détermination de méthodes de Runge-Kutta d'ordre 3 par résolution du système associé avec la méthode du résultant.

Exemples

(i) $q = 1$: on a nécessairement $b_1 = 1$ et la méthode se réduit à :

$$y_{n+1} = y_n + (t_{n+1} - t_n) f(t_n, y_n).$$

On retrouve la méthode d'Euler explicite étudiée au paragraphe précédent.

(ii) $q = 2$: pour tout $\alpha \in [0,1]$, on construit des méthodes de Runge-Kutta d'ordre 2 avec le tableau

$$\begin{array}{c|cc} 0 & 0 & \\ \alpha & \alpha & 0 \\ \hline & 1-\frac{1}{2\alpha} & \frac{1}{2\alpha} \end{array}$$

En particulier, lorsque $\alpha = 1$, la méthode (dite de Heun ou d'Euler Cauchy) s'écrit

$$y_{n+1} = y_n + \frac{\Delta T}{2}\left[f(t_n, y_n) + f\left(t_n + \Delta T, y_n + \Delta T f(t_n, y_n)\right)\right]$$

et lorsque $\alpha = \frac{1}{2}$, la méthode (dite d'Euler modifié ou du point milieu) s'écrit

$$y_{n+1} = y_n + \Delta T f\left(t_n + \frac{\Delta T}{2}, y_n + \frac{\Delta T}{2} f(t_n, y_n)\right).$$

(iii) $q = 4$: un exemple de méthode de Runge-Kutta fréquemment utilisé dans la pratique est le suivant :

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{2}{6} & \frac{2}{6} & \frac{1}{6} \end{array}$$

On peut montrer que cette méthode est d'ordre 4 (voir par exemple [CrMi]).

Remarque : Lorsque $5 \leq q \leq 7$ (respectivement $8 \leq q$), on montre que les méthodes de Runge-Kutta explicites sont forcément d'ordre inférieur à $q-1$ (respectivement $q-2$).