Équations différentielles (GM3)

Hasnaa Zidani

LMI - INSA Rouen

2022/2023 - CM 6&7

Problème de Cauchy: (EDO)

$$\begin{aligned} & y'(t) = f(t, y(t)) \quad \forall t \in [t_0, t_0 + T] \\ & y(t_0) = y_0 \end{aligned}$$

- Dans tout ce chapitre, on suppose que $f : [t_0, t_0 + T] \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ est continue
- On suppose aussi que pour tout $t \in [t_0, t_0 + T]$, la fonction $x \longmapsto f(t, x)$ est L_f-Lipschitz:

$$\forall t \in [t_0, t_0 + T], \quad \forall x \in \mathbb{R}^d, \quad |f(t, x) - f(t, y)| \leq L_f |x - y|.$$

Rappels

- D'après le théorème de Cauchy-Lipschitz, l'équation (EDO) admet une solution unique.
- Dans la suite, on note $X(\cdot)$ la solution exacte de (EDO).
  Si pour $m \in \mathbb{N}$, f est de classe C^m, alors X est de classe C^{m+1}.

Définition

Supposons que f est de classe C^m. Pour $0 \leq k \leq m$, on définit:

$$\begin{aligned}
f^{[0]}(t,y) &= f(t,y) \\
f^{[1]}(t,y) &= \frac{\partial f}{\partial t}(t,y) + \frac{\partial f}{\partial y}(t,y)\cdot f(t,y) \\
\vdots \quad & \vdots \\
f^{[k]}(t,y) &= \frac{\partial f^{[k-1]}}{\partial t}(t,y) + \frac{\partial f^{[k-1]}}{\partial y}(t,y)\cdot f(t,y)
\end{aligned}$$

Méthodes numériques pour la résolution des EDO

Proposition

On suppose que $f \in C^m([t_0, t_0+T]\times\mathbb{R}^d)$. Alors la solution X de (EDO) vérifie:

$$\forall 0 \leq k \leq m, \quad X^{(k+1)}(t) = f^{[k]}(t, X(t)).$$

Preuve par récurrence.

Idée de la démarche suivie pour calculer une approximation numérique de la solution de l'EDO

- D'abord, on va considérer une subdivision de $[t_0, t_0+T]$:

$$t_0 < t_1 < \cdots, t_i < t_{i+1} < \cdots < t_N = t_0 + T.$$

- On pose $h_i = t_{i+1} - t_i$
- Supposons que $f \in C^m([t_0, t_0+T] \times \mathbb{R}^d)$, la formule de Taylor nous donne:

$$\begin{aligned} X(t_{i+1}) &= X(t_i) + h_i X'(t_i) + \cdots + \frac{h_i^m}{m!} X^{(m)}(t_i) + O(h_i^{m+1}) \\ &= X(t_i) + h_i f^{[0]}(t_i, X(t_i)) + \cdots + \frac{h_i^m}{m!} f^{[m]}(t_i, X(t_i)) + O(h_i^{m+1}) \\ &= X(t_i) + h_i \widetilde{f}(t_i, X(t_i), h_i) + O(h_i^{m+1}) \end{aligned}$$

- La formule de Taylor suggère d'approcher $X(t_{i+1})$ par y_i

$$y_{i+1} = y_i + h_i F(t_i, y_i, h_i).$$

Forme générale d'une méthode à un pas

- Considérons une subdivision de $[t_0, t_0+T]$:

$$t_0 < t_1 < \cdots, t_i < t_{i+1} < \cdots < t_N = t_0 + T.$$

- Pour chaque $i = 0, 1, \cdots, N-1$, on pose $h_i = t_{i+1} - t_i$ et

$$h = \max_{0 \le i \le N-1} h_i.$$

- Si $h_i = h \quad \forall i = 0, 1, \cdots, N-1$, alors la discrétisation est dite **uniforme**, et dans ce cas, on a:

$$t_i = t_0 + ih, \quad h = \frac{T}{N}.$$

- A chaque instant t_i, on calcule la valeur y_i par la récurrence:

$$(S) \quad \begin{cases} y_0 \text{ donné}, \\ y_{i+1} = y_i + h_i F(t_i, y_i, h_i) \quad \text{pour } 0 \le i \le N-1. \end{cases}$$

Exemple 1 - Euler Explicite

- **Méthode d'Euler Explicite**: consiste à prendre $F(t, y, h) = f(t, y)$. Le schéma s'écrit alors

$$(EE) \quad \begin{cases} y_0 \text{ donné,} \\ y_{i+1} = y_i + h_i f(t_i, y_i) \quad \text{pour } 0 \leq i \leq N-1. \end{cases}$$

Interprétation géométrique: si on prend $y_i = X(t_i)$, la méthode (EE) consiste à approcher la valeur $X(t_{i+1})$ en utilisant la tangente à la courbe $X(\cdot)$ passant par $X(t_i)$.

Exemple 2 - Point Milieu

- **Méthode du Point Milieu**: consiste à approcher la valeur $X(t_{i+1})$ en utilisant la tangente à la courbe $X(\cdot)$ passant par $X(t_i + \frac{h_i}{2})$. On a donc

$$\begin{aligned} X(t_{i+1}) &\simeq X(t_i) + h_i X'(t_i + \frac{h_i}{2}) = X(t_i) + h_i f(t_i + \frac{h_i}{2}, X(t_i + \frac{h_i}{2})) \\ \text{et} \quad X(t_i + \frac{h_i}{2}) &\simeq X(t_i) + \frac{h_i}{2} f(t_i, X(t_i)) \end{aligned}$$

Cela conduit à:

$$(PM) \quad \begin{cases} y_0 \text{ donné,} \\ y_{i+1} = y_i + h_i f(t_i + \frac{h_i}{2}, y_i + \frac{h_i}{2} f(t_i, y_i)) \quad \text{pour } 0 \leq i \leq N-1. \end{cases}$$

Interprétation géométrique: L'idée est que la corde qui relie $X(t_{i+1})$ et $X(t_i)$ est mieux approchée par la tangente en $X\left(t_i + \frac{h_i}{2}\right)$:

$$\begin{aligned} X(t_{i+1}) &\simeq X(t_i) + h_i X'\left(t_i + \frac{h_i}{2}\right) \qquad \text{et} \quad X\left(t_i + \frac{h_i}{2}\right) \simeq X(t_i) + \frac{h_i}{2} f(t_i, X(t_i)) \\ &= X(t_i) + h_i f\left(t_i + \frac{h_i}{2}, X\left(t_i + \frac{h_i}{2}\right)\right) \end{aligned}$$

Exemple 3 - Euler Implicite

➤ **Méthode d'Euler Implicite**: consiste à approcher la valeur $X(t_{i+1})$ en utilisant la tangente à la courbe $X(\cdot)$ passant par $X(t_{i+1})$. Le schéma s'écrit alors

$$(EI) \qquad \begin{cases} y_0 \text{ donné}, \\ y_{i+1} = y_i + h_i f(t_{i+1}, y_{i+1}) \quad \text{pour } 0 \leq i \leq N-1. \end{cases}$$

Dans ce schéma, la fonction F est implicite et n'a pas de forme explicite générale.

Erreur de consistance - Ordre du schéma

$$(S) \qquad \begin{cases} y_0 \text{ donné}, \\ y_{i+1} = y_i + h_i F(t_i, y_i, h_i) \quad \text{pour } 0 \leq i \leq N-1. \end{cases}$$

Définition

On appelle **erreur de consistance** de la méthode (S) à l'instant t_i la quantité

$$\epsilon_i = X(t_{i+1}) - X(t_i) - h_i F(t_i, X(t_i), h_i)$$

Définition

Supposons que $f \in C^p([t_o, t_o + T] \times \mathbb{R}^d)$ avec $p \geq 1$.
On dit que (S) est **d'ordre** p s'il existe une constante $K > 0$ ne dépendant[a] pas de h

$$\sum_{i=0}^{N-1} |\epsilon_i| \leq K h^p, \qquad \text{avec } \; h = \max_{0 \leq i \leq N-1} h_i.$$

[a] la constante K dépend éventuellement de X et de F

Théorème

Soit $p \geq 1$. On suppose que $f \in C^p([t_o, t_o + T] \times \mathbb{R}^d)$ et que les fonctions $F, \dfrac{\partial F}{\partial h}, \cdots \dfrac{\partial^p F}{\partial h^p}$ existent et sont continues dans $[t_0, t_0 + T] \times \mathbb{R}^d \times [0, h]$.
Alors, une condition suffisante pour que (S) soit d'ordre p est :

$$\forall (t, y) \in [t_0, t_0 + T] \times \mathbb{R}^d, \qquad \begin{cases} F(t, y, 0) = f(t, y) \\ \dfrac{\partial F}{\partial h}(t, y, 0) = \dfrac{1}{2} f^{[1]}(t, y) \\ \vdots \\ \dfrac{\partial^{p-1} F}{\partial h^{p-1}}(t, y, 0) = \dfrac{1}{p} f^{[p-1]}(t, y) \end{cases}$$

- Le théorème donne une condition suffisante pour que le schéma (S) soit d'ordre p. On peut montrer que ces conditions sont aussi nécessaires (admis).

Exemples

- **Méthode d'Euler Explicite**
  On a bien $F(t, y, h) = f(t, y) = f^{[0]}(t, y)$. Donc le schéma (EE) est d'ordre 1.

- **Méthode du Point Milieu**
  Prenons $f \in C^2([t_0, t_0 + T] \times \mathbb{R}^d)$. Pour cette méthode, on a

$$F(t, y, h) = f(t + \frac{h}{2}, y + \frac{h}{2} f(t, y))$$

$$F(t, y, 0) = f(t, y)$$

$$\frac{\partial F}{\partial h}(t, y, 0) = \frac{1}{2}\frac{\partial f}{\partial t}(t, y) + \frac{1}{2}\frac{\partial f}{\partial y}(t, y) \cdot f(t, y) = \frac{1}{2} f^{[1]}(t, y)$$

Donc le schéma (PM) est d'ordre 2.

Définition

On dit que la méthode (S) est stable s'il existe une constante $M > 0$, indépendante de h, telle que pour toutes suites $(y_i)_i$, $(z_i)_i$ et $(\delta_i)_i$ vérifiant:

$$\begin{aligned} y_{i+1} &= y_i + h_i F(t_i, y_i, h_i) \\ z_{i+1} &= z_i + h_i F(t_i, z_i, h_i) + \delta_i \end{aligned}$$

on a:

$$\max_{0 \le i \le N} |z_i - y_i| \le M \left[|z_0 - y_0| + \sum_{i=0}^{N-1} |\delta_i| \right].$$

– Un algorithme stable est une algorithme pour lequel une petite perturbation des données n'entraîne qu'une petite perturbation sur la solution de l'EDO

Théorème

On suppose qu'il existe une constante $L > 0$ telle que

$$\forall t \in [t_0, t_0 + T],\ \forall y, z \in \mathbb{R}^d,\ \forall h \in [0, T], \quad |F(t, y, h) - F(t, z, h)| \le L|y - z|.$$

Alors la méthode (S) est stable.

Corollaire. Si f est Lipschitz, alors

- la méthode d'Euler Explicite (EE) est stable
- la méthode du Point Milieu (PM) est stable.

La preuve du théorème précédent repose sur le lemme suivant.

Lemme - Inégalité de Gronwall "lemme discret"

Soient θ_i et α_i deux suites de réels positifs vérifiant, pour $L > 0$

$$\forall i \geq 0, \quad \theta_{i+1} \leq (1 + h_i L)\theta_i + \alpha_i.$$

Alors, on a:

$$\forall i \geq 0, \quad \theta_i \leq e^{L(t_i - t_0)} + \sum_{j=0}^{i-1} e^{L(t_i - t_{j+1})}\alpha_j.$$

Convergence d'un schéma - Ordre de convergence

Définition

On dit que le schéma (S) est convergent si et seulement si

$$\max_{0 \leq i \leq N} |X(t_i) - y_i| \longrightarrow 0, \quad \text{lorsque} \quad h \to 0.$$

Théorème

Si la méthode (S) est stable et d'ordre 1, alors elle est convergente. De plus, si (S) est d'ordre p, alors on a la majoration suivante:

$$\max_{0 \leq i \leq N} |X(t_i) - y_i| \leq M(|x_0 - y_0| + h^p).$$

➢ Si f est Lipschitz, alors les méthodes d'Euler et du Point-Milieu sont convergentes.

Idée de construction de schémas à un pas d'ordre élevé

- La solutions de l'EDO vérifie:

$$X(t_{i+1}) = X(t_i) + \int_{t_i}^{t_{i+1}} f(t, X(t))\, dt$$

- Si on approche l'intégrale $\int_{t_i}^{t_{i+1}} f(t, X(t))\, dt$ par $h_i f(t_i, X(t_i))$, alors on obtient le schéma (EE).
- Pour avoir une méthode d'ordre plus élevé, on considère des instants intermédiaires

$$[t_i, t_{i+1}] \ni t_{ij} = t_i + \theta_j h_i \quad j \in \{1, \dots, r\}$$

avec $0 \leqslant \theta_1 < \cdots \leqslant \theta_r \leqslant 1$, et une approximation de la solution de l'EDO sous la forme suivante:

$$\left(\widetilde{S}\right) \begin{cases} y_{i+1} = y_i + h_i \sum_{j=1}^{r} c_j f(t_{ij}, y_{ij}) & 0 \leqslant i \leqslant N-1 \\ y_{ij} = y_i + h_i \sum_{k=1}^{r} a_{jk} f(t_{ik}, y_{ik}) & 1 \leqslant j \leqslant r \end{cases}$$

- La valeur y_i approche $X(t_i)$, tandis que les valeurs de y_{ij} sont des valeurs intermédiaires qui permettrait d'améliorer la précision de l'approximation.

Définition

- Le schéma $\left(\widetilde{S}\right)$ est un schéma de type Runge-Kutta. Il est déterminé par (θ_j), par (c_j) et par (a_{jk}).

- le schéma RK peut être représenté par le diagramme de Butcher avec $r \geq 1$):

$$\begin{array}{c|ccc} \theta_1 & a_{11} & \dots & a_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_r & a_{r1} & \dots & a_{rr} \\ \hline & c_1 & \dots & c_r \end{array}$$

➥ La représentation de Butcher

$$\begin{array}{c|ccc} \theta_1 & a_{11} & \dots & a_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_r & a_{r1} & \dots & a_{rr} \\ \hline & c_1 & \dots & c_r \end{array}$$

➥ La première colonne permet de définir les instants intermédiaires

$$t_{ij} = t_i + \theta_j h_i, \quad j = 1, \cdots, r$$

➥ La matrice supérieure permet de calculer les valeurs y_{ij}

$$y_{ij} = y_i + h_i \sum_{k=1}^{r} a_{jk} f(t_{ik}, y_{ik}) \qquad 1 \leqslant j \leqslant r$$

➥ Et enfin, la dernière ligne permet de calculer la valeur à l'instant t_{i+1}

$$y_{i+1} = y_i + h_i \sum_{j=1}^{r} c_j f(t_{ij}, y_{ij})$$

➥ Par exemple, le schéma d'EE est représenté par

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$$

➥ Un autre exemple est le suivant (pour $\alpha \in]0, 1]$):

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \alpha & \alpha & 0 \\ \hline & 1 - \frac{1}{2\alpha} & \frac{1}{2\alpha} \end{array} \quad \Longleftrightarrow \quad \begin{cases} t_{i1} = t_i, \quad t_{i2} = t_i + \alpha h_i, \\ y_{i1} = y_i, \\ y_{i2} = y_i + \alpha h_i f(t_{i1}, y_{i1}), \\ y_{i+1} = y_i + h_i \left[\left(1 - \frac{1}{2\alpha}\right) f(t_{i1}, y_{i1}) + \frac{1}{2\alpha} f(t_{i2}, y_{i2}) \right] \end{cases}$$

➢ Dans la suite, il sera utile d'utiliser une écriture du schéma sous une forme plus compacte. Pour cela, on pose

$$A = \begin{bmatrix} a_{11} & \dots & a_{1r} \\ \vdots & \ddots & \vdots \\ a_{r1} & \dots & a_{rr} \end{bmatrix} \qquad \text{et } \widetilde{Y} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{ir} \end{pmatrix}$$

➢ Le schéma (S) peut se réécrire sous la forme suivante:

$$(\widetilde{S}) \begin{cases} y_{i+1} = y_i + h_i \sum_{j=1}^{r} c_j f\left(t_{ij}, \widetilde{Y}_j\right) \\ \tilde{Y} = y_i \mathbf{1}_r + h_i A \begin{pmatrix} f\left(t_{i1}, \tilde{Y}_1\right) \\ \vdots \\ f\left(t_{ir}, \tilde{Y}_r\right) \end{pmatrix} \end{cases} \qquad \text{avec} \qquad \mathbf{1}_r = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Notons qu'un schéma RK peut être implicite. Sous quelles conditions est-ce que ce schéma est bien posée? autrement sous quelles conditions le système non-linéaire en $\widetilde{Y}$ admet-il une solution unique ?

Théorème

Soit $h^* > 0$ tel que

$$h^* \, \|A\| \, L_f < 1$$

avec L_f la constante de Lipschitz de f. Alors, on a:

(i) Le schéma RK admet une solution unique pour tout $h \leqslant h^*$.

(ii) Le schéma RK est stable pour tout $h \leqslant h^*$.

Idée de la preuve

(*i*) La condition $h^* \|A\| L_f < 1$ assure que l'application

$$G : u \in \mathbb{R}^r \mapsto y_i \mathbf{1}_r + hA \begin{pmatrix} f(t_{i1}, u_1) \\ \vdots \\ f(t_{ir}, u_r) \end{pmatrix}$$

est contractante. Donc l'équation

$$\tilde{Y} = G\left(\tilde{Y}\right) = y_i \mathbf{1} + hA \begin{pmatrix} f\left(t_{i1}, \tilde{Y}_1\right) \\ \vdots \\ f\left(t_{ir}, \tilde{Y}_r\right) \end{pmatrix}$$

admet une solution unique et le schéma RK est bien défini.

(*ii*) D'autre part, on peut vérifier que $|F(t, y, h) - F(t, z, h)| \leqslant M|y - z|$ dès que $0 < h \leqslant h^*$. Ce qui prouve que le schéma RK est stable si $0 < h \leqslant h^*$.

Quelques remarques

- La condition $h^* \|A\| L_f < 1$ n'est pas restrictive. En effet, on peut toujours adapter le pas h pour que cette condition soit vérifiée (et donc pour que la méthode soit stable).
- On rappelle que $\forall \varepsilon > 0$, il existe une norme matricielle $\|\cdot\|$ telle que

$$\|A\| \leqslant \rho(A) + \varepsilon.$$

  Donc, il suffit d'avoir $h^* \rho(A) L_f < 1$ pour que la méthode RK soit stable.
- Si la méthode RK est explicite avec

$$A = \begin{bmatrix} 0 & \dots & \dots & 0 \\ a_{21} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a_{r1} & \dots & a_{r(r-1)} & 0 \end{bmatrix},$$

  alors $\rho(A) = 0$ et la méthode est inconditionnellement stable.

Pour analyser la consistance, on note d'abord que le schéma Runge-Kutta est une méthode à un pas, avec

$$F(t,y,h) = \sum_{j=1}^{r} c_j f(t + \theta_j h \underbrace{y_{ij}(t,y,h)}_{\text{valeurs intermédiaires}})$$

avec $$y_{ij}(t,y,h) = y\mathbf{1}_r + hA \begin{pmatrix} f(t+\theta_1 h, y_{i1}) \\ \vdots \\ f(t+\theta_r h, y_{ir}) \end{pmatrix}$$

Théorème

Une méthode de Runge-Kutta est consistante si et seulement si

$$\sum_{j=1}^{r} c_j = 1$$

Preuve. On note que $F(t,y,0) = \left(\sum_{j=1}^{r} c_j\right) f(t,y)$. Or, on sait que le schéma est consistant (d'ordre 1) si $F(t,y,0) = f(t,y)$. Cette condition est vérifiée ssi $\sum_{j=1}^{r} c_j = 1$.

Si on applique le théorème général qui donne des conditions sous lesquelles un schéma à un pas est d'ordre p, on trouve que

- **la méthode RK est d'rdre 1** si $F(t,y,0) = f(t,y)$, ce qui est équivalent à

$$\sum_{j=1}^{r} c_j = 1.$$

- **la méthode RK est d'ordre 2** si

$$F(t,y,0) = f(t,y), \qquad \frac{\partial F}{\partial h}(t,y,0) = \frac{1}{2} f^{[1]}(t,y).$$

Pour simplifier, on choisit $\sum_{k=1}^{r} a_{jk} = \theta_j$, alors le schéma est d'ordre 2 si

$$\sum_{j=1}^{r} c_j = 1, \qquad \sum_{j=1}^{r} c_j \theta_j = \frac{1}{2}.$$

La forme générale d'un schéma RK permet d'avoir de la souplesse dans le choix des instants intermédiaires et dans le calcul des valeurs intermédiaires. Cela permet de construire des méthodes d'ordre élevé.

Théorème

Si $\sum_{j=1}^{r} c_j = 1$ *et si* $h \leqslant h^*$*, alors la méthode de Runge-Kutta est convergente.*

Preuve La condition $\sum_{j=1}^{r} c_j = 1$ assure que le schéma est consistant. De plus, la condition $h \leqslant h^*$ assure que le schéma est stable. D'après le théorème général de convergence, on conclut que le schéma est convergeant.