GRAVEJAT Philippe

Agrégation

4/4/26

Chapitre VI : Méthodes de résolution approchée des équations différentielles ordinaires

Introduction

Nous nous intéressons à la résolution approchée d'une équation différentielle ordinaire du premier ordre :

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

Afin de simplifier la présentation, nous nous limitons au cas où la fonction f est définie sur un sous-ensemble de $\mathbb{R}^2$ de la forme $I \times U$, où I est un intervalle de $\mathbb{R}$ et U désigne un sous-ensemble ouvert de $\mathbb{R}$, et à valeurs réelles. Les différentes méthodes de résolution approchée que nous décrirons dans la suite s'étendent sans difficulté majeure au cas où la fonction f est définie sur un sous-ensemble de $\mathbb{R} \times \mathbb{R}^n$ de la forme $I \times U$, où U désigne désormais un sous-ensemble ouvert de $\mathbb{R}^n$, et à valeurs vectorielles. De suite, elles s'étendent aussi à des équations différentielles ordinaires d'ordre strictement supérieur à un.

Nous nous plaçons de plus dans un cadre où le problème de Cauchy associé à l'équation différentielle $y'(t) = f(t, y(t))$ est bien posé (au moins localement). Etant donné un nombre $t_0 \in I$, ceci signifie qu'il existe une unique solution $y \in \mathcal{C}^1(I_0, U)$ sur un intervalle $I_0 \subset I$ contenant t_0 pour la donnée initiale $y(t_0) = y_0 \in U$, et que cette solution dépend de manière continue de la valeur y_0 de la donnée initiale. Rappelons que le théorème de Cauchy-Lipschitz garantit le caractère bien posé de l'équation différentielle $y'(t) = f(t, y(t))$.

Théorème de Cauchy-Lipschitz : Soit I un intervalle ouvert de $\mathbb{R}$ et U un sous-

ensemble ouvert de $\mathbb{R}$. Considérons une fonction $f \in \mathcal{C}^0(I \times U, \mathbb{R})$ qui est localement lipschitzienne par rapport à $y \in U$, c'est-à-dire telle que :

$\forall (t_0, y_0) \in I \times U$, $\exists (\delta_0, \ell_0) \in (\mathbb{R}_+^*)^2$ t.q. $\exists K_0 \in \mathbb{R}_+$ t.q. $\forall t \in I$, $\forall (y, y') \in U^2$, $|t - t_0| < \delta_0$ et $\max\{|y - y_0|, |y' - y_0|\} < \ell_0 \Rightarrow |f(t,y) - f(t,y')| \leq K_0 |y - y'|$.

Etant donnés deux nombres $t_0 \in I$ et $y_0 \in U$, il existe un nombre strictement positif ε_0 tel qu'il existe une unique solution $y \in \mathcal{C}^1([t_0 - \varepsilon_0, t_0 + \varepsilon_0], U)$ de l'équation différentielle :

$\forall t \in (t_0 - \varepsilon_0, t_0 + \varepsilon_0)$, $y'(t) = f(t, y(t))$,

pour la donnée initiale $y(t_0) = y_0$. De plus, si deux suites $(t_0^{(n)})_{n \in \mathbb{N}} \in I^{\mathbb{N}}$ et $(y_0^{(n)})_{n \in \mathbb{N}} \in U^{\mathbb{N}}$ satisfont :

$$t_0^{(n)} \xrightarrow[n \to +\infty]{} t_0 \quad \text{et} \quad y_0^{(n)} \xrightarrow[n \to +\infty]{} y_0,$$

alors, les solutions $y^{(n)} \in \mathcal{C}^1([t_0 - \varepsilon^{(n)}, t_0 + \varepsilon^{(n)}], U)$ de l'équation différentielle pour les données initiales $y^{(n)}(t_0^{(n)}) = y_0^{(n)}$ satisfont :

$$\varliminf_{n \to +\infty} \varepsilon^{(n)} \geq \varepsilon_0 \quad \text{et} \quad \max_{t \in (t_0 - \varepsilon_0, t_0 + \varepsilon_0)} |y^{(n)}(t) - y(t)| \xrightarrow[n \to +\infty]{} 0.$$

Lorsque cela s'avèrera nécessaire, nous nous placerons de plus dans le cas où les solutions de l'équation différentielle sont globales, c'est-à-dire définies sur tout l'intervalle I. Ceci n'exclut pas de traiter des équations différentielles pour lesquelles l'unicité des solutions n'est pas garantie, mais cela facilite l'analyse de la convergence des méthodes numériques proposées puisque leur limite est alors définie de manière unique (et si nécessaire, globalement).

Les méthodes de résolution approchée auxquelles nous nous intéressons reposent sur un processus de discrétisation. Etant donnée une solution $y \in \mathcal{C}^1([t_0, t_0 + T], U)$ du problème de Cauchy pour une donnée initiale $y(t_0) = y_0$, nous commençons par introduire une subdivision $t_0 < t_1 < \dots < t_N = t_0 + T$ de son intervalle de définition $[t_0, t_0 + T]$, et nous cherchons à déterminer des

valeurs approchées $(y_n)_{1 \leq n \leq N}$ des valeurs exactes $(y(t_n))_{1 \leq n \leq N}$ prises par la solution y aux temps $(t_n)_{1 \leq n \leq N}$.

Les méthodes de résolution approchée que nous décrirons seront des méthodes à un pas. Le calcul de la valeur approchée y_{n+1} ne dépendra que de la valeur approchée y_n. Les méthodes d'Euler explicite et implicite, et du point milieu sont des exemples de méthodes à un pas. Nous chercherons à déterminer sous quelles conditions ces méthodes sont convergentes, c'est-à-dire telles que l'approximation ainsi obtenue converge vers la solution exacte lorsque le pas de la subdivision:

$$h = \max_{0 \leq n \leq N-1} |y_{n+1} - y_n|$$

tend vers 0. En particulier, nous introduirons les notions de consistance et de stabilité qui garantissent la convergence de ces méthodes. Nous analyserons également l'ordre de ces méthodes afin de déterminer leur précision, et l'influence des erreurs d'arrondis.

Nous détaillerons ensuite le cas des méthodes de Runge-Kutta, avant de conclure par une présentation succincte des difficultés à surmonter lorsque les problèmes de Cauchy étudiés ne sont plus bien posés mathématiquement ou numériquement, ou encore bien conditionnés.

Nous renvoyons aux ouvrages de Michel Crouzeix et Alain Mignot intitulé "Analyse numérique des équations différentielles", et de Jean-Pierre Demailly intitulé "Analyse numérique et équations différentielles" pour de plus amples détails au sujet des diverses méthodes de résolution approchée des équations différentielles ordinaires.

I Étude générale des méthodes à un pas

Soit I, un intervalle ouvert de $\mathbb{R}$. Nous considérons une fonction $f \in \mathcal{C}^0(I$

$\mathbb{R}, \mathbb{R}$), et nous nous intéressons à la résolution approchée du problème de Cauchy :

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

où les nombres t_0 et t_0+T appartiennent à l'intervalle I, et la valeur y_0 de la donnée initiale est réelle.

Afin de simplifier la présentation qui suit, nous supposons que la fonction f est globalement lipschitzienne par rapport à $y \in \mathbb{R}$, à savoir qu'il existe un nombre positif K tel que :

$$\forall t \in I, \ \forall (y, y') \in \mathbb{R}^2, \quad |f(t, y') - f(t, y)| \leq K |y - y'|.$$

Ceci garantit qu'il existe une unique solution $y \in \mathcal{C}^1([t_0, t_0+T], \mathbb{R})$ du problème de Cauchy précédent, qui dépend de plus de manière continue de la donnée initiale y_0. Cette hypothèse simplifie la preuve ou la forme de certaines des estimations d'erreur que nous allons établir. Notons cependant que la plupart de ces estimations demeurent valables sous des hypothèses plus faibles telles que le caractère localement lipschitzien par rapport à $y \in \mathbb{R}$ de la fonction f.

Afin de déterminer une valeur approchée de la solution y, nous introduisons une subdivision :

$$t_0 < t_1 < \dots < t_N = t_0 + T,$$

du segment $[t_0, t_0+T]$. Nous notons :

$$\forall n \in [\![0, N-1]\!], \quad h_n = t_{n+1} - t_n,$$

le pas entre les temps t_n et t_{n+1}, et :

$$h_{max} = \max_{0 \leq n \leq N-1} h_n,$$

le pas maximal de la subdivision. Notre objectif est de déterminer des valeurs approchées $(y_n)_{0 \leq n \leq N}$ de la solution $(y(t_n))_{0 \leq n \leq N}$ aux temps $(t_n)_{0 \leq n \leq N}$. Dans ce but, nous introduisons les méthodes numériques à un pas, qui sont définies de la façon suivante.

1. Définition et exemples

Définition : Une méthode numérique à un pas pour la résolution approchée du problème de Cauchy :

$$\text{(P)} \begin{cases} \forall t \in [t_0, t_0+T], \ y'(t) = f(t, y(t)), \\ y(t_0) = y_0, \end{cases}$$

est définie par la donnée d'une fonction $\Phi \in \mathcal{C}^0([t_0, t_0+T] \times \mathbb{R}^2, \mathbb{R})$ et la formule de récurrence pour les valeurs approchées $(y_n)_{0 \le n \le N}$ aux temps $(t_n)_{0 \le n \le N}$:

$$\forall n \in [\![0, N-1]\!], \ y_{n+1} = y_n + h_n \Phi(t_n, y_n, h_n).$$

La terminologie de méthodes numériques à un pas s'explique par le fait que le calcul de la valeur approchée y_{n+1} au temps t_{n+1} ne dépend que de la valeur approchée y_n au temps t_n (et du pas h_n).

Il est possible d'affaiblir les hypothèses sur la fonction Φ qui peut n'être définie qu'au voisinage du triplet $(t_0, y_0, 0)$. Il est néanmoins nécessaire que son ensemble de définition recoupe les valeurs prises par les couples $(t, y(t))$ pour $t \in [t_0, t_0+T]$, où y désigne l'unique solution du problème de Cauchy considéré.

Exemples : (i) *Méthode d'Euler explicite (ou progressive)*

Cette méthode correspond à la fonction Φ définie par :

$$\forall (t, y, h) \in [t_0, t_0+T] \times \mathbb{R}^2, \ \Phi(t, y, h) = f(t, y).$$

La suite $(y_n)_{0 \le n \le N}$ des valeurs approchées au temps $(t_n)_{0 \le n \le N}$ est donnée par la formule de récurrence :

$$\forall n \in [\![0, N-1]\!], \ y_{n+1} = y_n + h_n f(t_n, y_n).$$

Cette suite est toujours bien définie sous les hypothèses générales que nous avons imposées pour la fonction f.

(ii) *Méthode d'Euler implicite (ou rétrograde)*

Cette méthode correspond à la fonction Φ définie par :

$\forall (t,y,h) \in [t_0, t_0+T] \times \mathbb{R}^2$, $\Phi(t,y,h) = f\left(t+h, \left[\mathrm{Id} - h f(t+h, \cdot)\right]^{-1}(y)\right)$.

La suite $(y_n)_{0 \le n \le N}$ des valeurs approchées aux temps $(t_n)_{0 \le n \le N}$ est donnée par la formule de récurrence implicite :

$\forall\, 0 \le n \le N-1$, $y_{n+1} = y_n + h_n f(t_{n+1}, y_{n+1})$.

Il n'est pas clair que cette suite soit bien définie. Le calcul de la valeur y_{n+1} passe par la résolution d'une équation de type point fixe, ce qui ne s'avère possible que si la fonction f satisfait des hypothèses ad hoc. Notons que ces hypothèses reviennent à supposer que l'application $x \mapsto x - h f(t+h, x)$ est un homéomorphisme de $\mathbb{R}$ sur $\mathbb{R}$, qui dépend de manière continue des paramètres t et h (auquel cas la fonction Φ est bien définie et continue sur $[t_0, t_0+T] \times \mathbb{R}^2$).

(iii) Méthode du point milieu (ou d'Euler modifié)

Cette méthode correspond à la fonction Φ définie par :

$\forall (t,y,h) \in [t_0, t_0+h] \times \mathbb{R}^2$, $\Phi(t,y,h) = f\left(t + \frac{h}{2}, y + \frac{h}{2} f(t,y)\right)$.

La suite $(y_n)_{0 \le n \le N}$ des valeurs approchées aux temps $(t_n)_{0 \le n \le N}$ est donnée par l'algorithme récursif :

$\forall\, 0 \le n \le N-1$, $\begin{cases} y_{n+\frac{1}{2}} = y_n + \frac{h_n}{2} f(t_n, y_n), \\ y_{n+1} = f\left(t_n + \frac{h_n}{2}, y_{n+\frac{1}{2}}\right). \end{cases}$

Cette suite est toujours bien définie sous les hypothèses générales que nous avons imposées pour la fonction f.

Nous chercherons dans la suite à établir sous quelles conditions une méthode numérique à un pas est convergente, à savoir qu'elle vérifie la propriété :

$$\max_{0 \le n \le N} |y(t_n) - y_n| \longrightarrow 0,$$

lorsque le pas maximal h max de la subdivision tend vers 0. Ceci passe par l'introduction des notions de consistance et de stabilité.

2. Consistance d'une méthode numérique à un pas

Considérons la méthode numérique à un pas associée à la fonction $\Phi \in C^0([t_0, t_0+T] \times \mathbb{R}^d, \mathbb{R}^d)$.

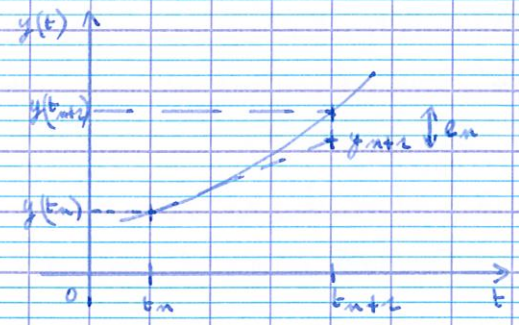
Définition : Soit $0 \leq n \leq N-1$. Considérons une solution $y \in C^1([t_0, t_0+T], \mathbb{R}^d)$ de l'équation différentielle :

$\forall\, t \in [t_0, t_0+T]$, $y'(t) = f(t, y(t))$.

L'erreur de consistance ε_n au temps t_n relative à la solution y est définie par :

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

L'erreur de consistance au temps t_n mesure donc l'écart entre la valeur exacte au temps t_{n+1} donnée par $y(t_{n+1})$ et la valeur approchée fournie par la méthode numérique à un pas à partir de la valeur exacte $y(t_n)$ au temps t_n. Autrement dit, elle mesure l'erreur faite au $n^{ième}$ pas lorsque l'équation différentielle est remplacée par le schéma numérique à un pas.



Cette erreur n'est pas égale à l'erreur finale entre la valeur exacte $y(t_n)$ au temps t_n et sa valeur approchée y_n, laquelle est donnée par la différence $y(t_n) - y_n$. Cependant, il semble clair que cette erreur sera au moins de l'ordre de $\sum_{k=0}^{n-1} \varepsilon_k$, quantité qu'il est donc important de pouvoir contrôler.

Définition : Une méthode numérique à un pas est consistante si et seulement si quelle que soit la solution $y \in C^1([t_0, t_0+T] \times \mathbb{R}^d, \mathbb{R}^d)$ de l'é-

quation différentielle :

$\forall\, t \in [t_0, t_0+T],\ y'(t) = f(t, y(t))$,

les erreurs de consistance $(e_n)_{0 \le n \le N-1}$ aux temps $(t_n)_{0 \le n \le N}$ relatives à la solution y satisfont :

$$\sum_{n=0}^{N-1} |e_n| \xrightarrow[h_{max} \to 0]{} 0.$$

Une méthode numérique à un pas est donc consistante lorsqu'il est possible de contrôler l'accumulation des erreurs de consistance quelle que soit la solution y considérée. Nous avons le critère de consistance suivant pour une méthode numérique à un pas.

Théorème : La méthode numérique à un pas est consistante avec l'équation différentielle $y'(t) = f(t, y(t))$ si et seulement si :

$\forall\, (t, y) \in [t_0, t_0+T],\ \Phi(t, y, 0) = f(t, y)$.

Preuve :

Soit $y \in \mathcal{C}^1([t_0, t_0+T], \mathbb{R})$, une solution de l'équation différentielle :

$\forall\, t \in [t_0, t_0+T],\ y'(t) = f(t, y(t))$.

Pour $0 \le n \le N-1$, l'erreur de consistance e_n au temps t_n relative à la solution y est donnée par :

$e_n = y(t_{n+1}) - y(t_n) - h_n \Phi(t_n, y(t_n), h_n)$.

Par le théorème des accroissements finis, il existe un nombre $t_n \le \sigma_n \le t_{n+1}$ tel que :

$y(t_{n+1}) - y(t_n) = h_n y'(\sigma_n) = h_n f(\sigma_n, y(\sigma_n))$,

De sorte que :

$e_n = h_n \left[f(\sigma_n, y(\sigma_n)) - \Phi(\sigma_n, y(\sigma_n), 0) \right] + h_n \left[\Phi(\sigma_n, y(\sigma_n), 0) - \Phi(t_n, y(t_n), h_n) \right]$.

A ce stade, nous pouvons supposer que :

$h_{max} \le 1$.

Notons alors que la fonction $(t, h) \mapsto \Phi(t, y(t), h)$ est continue sur le compact $[t_0, t_0+T] \times [0, 1]$, donc uniformément continue sur cet

ensemble. Etant donné un nombre strictement positif ε, il existe donc un nombre strictement positif δ tel que :

$\forall (t,t') \in [t_0, t_0+T]^2$, $\forall (h,h') \in [0,L]^2$, $|t-t'| \leq \delta$ et $|h-h'| \leq \delta \Rightarrow$

$|\Phi(t', y(t'), h') - \Phi(t, y(t), h)| \leq \varepsilon$.

Lorsque $h_{max} \leq \delta$, nous concluons donc que :

$\sum_{n=0}^{N-1} h_n |\Phi(\sigma_n, y(\sigma_n), 0) - \Phi(t_n, y(t_n), h_n)| \leq \sum_{n=0}^{N-1} h_n \varepsilon = T\varepsilon$

Notons alors que la fonction $t \mapsto f(t, y(t)) - \Phi(t, y(t), 0)$ est continue sur $[t_0, t_0+T]$, de sorte que par le théorème des sommes de Riemann :

$\sum_{n=0}^{N-1} h_n |f(\sigma_n, y(\sigma_n)) - \Phi(\sigma_n, y(\sigma_n), 0)| \xrightarrow[h_{max} \to 0]{} \int_{t_0}^{t_0+T} dt \, |f(t, y(t)) - \Phi(t, y(t), 0)|$

En conclusion, nous avons établi que :

$\sum_{n=0}^{N-1} |\varepsilon_n| \xrightarrow[h_{max} \to 0]{} \int_{t_0}^{t_0+T} |f(t, y(t)) - \Phi(t, y(t), 0)| \, dt$

La consistance de la méthode numérique est donc équivalente au fait que

$\int_{t_0}^{t_0+T} |f(t, y(t)) - \Phi(t, y(t), 0)| \, dt = 0$,

quelle que soit la solution y considérée, soit au fait que :

$\forall t \in [t_0, t_0+T]$, $\Phi(t, y(t), 0) = f(t, y(t))$.

Par le théorème de Cauchy-Lipschitz, quels que soient $(t, y) \in [t_0, t_0+T] \times \mathbb{R}$, il existe une solution y telle que :

$y(t) = y$,

ce qui suffit à établir l'équivalence recherchée.

Exemples : (i) Méthode d'Euler explicite

Cette méthode est consistante puisque :

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

(ii) Méthode d'Euler implicite

Cette méthode est consistante (lorsqu'elle est bien définie), puisque :

$\forall (t,y) \in [t_0, t_0+T] \times \mathbb{R}$, $\Phi(t, y, 0) = f(t, Id^{-1}(y)) = f(t, y)$.

(iii) Méthode du point milieu

Cette méthode est consistante puisque :

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

3. Stabilité d'une méthode numérique à un pas

Considérons la méthode numérique à un pas associée à la fonction $\bar{F} \in \mathcal{C}^0([t_0, t_0+T] \times \mathbb{R}^2, \mathbb{R})$. En pratique, le calcul des valeurs approchées y_n est entaché d'erreurs d'arrondis. Il est crucial de contrôler ces erreurs afin que les valeurs approchées y_n soient significatives.

Définition: La méthode numérique à un pas est stable si et seulement s'il existe un nombre positif S tel que, quelles que soient les suites $(y_n)_{0 \le n \le N} \in \mathbb{R}^{N+1}$, $(y'_n)_{0 \le n \le N} \in \mathbb{R}^{N+1}$ et $(\delta_n)_{0 \le n \le N-1} \in \mathbb{R}^N$ telles que:

$$\forall\, 0 \le n \le N-1, \begin{cases} y_{n+1} = y_n + h_n \bar{F}(t_n, y_n, h_n), \\ y'_{n+1} = y'_n + h_n \bar{F}(t_n, y'_n, h_n) + \delta_n, \end{cases}$$

nous avons l'inégalité:

$$\max_{0 \le n \le N} |y'_n - y_n| \le S \left[|y'_0 - y_0| + \sum_{n=0}^{N-1} |\delta_n| \right].$$

Le nombre S est alors une constante de stabilité de la méthode.

La notion de stabilité est intrinsèque à la méthode numérique considérée en sens où elle ne dépend pas de l'équation différentielle $y'(t) = f(t, y(t))$ à résoudre. Elle garantit qu'une petite erreur sur la donnée initiale y_0 et de petites erreurs d'arrondis à chaque étape de la méthode ne conduisent pas à une erreur finale incontrôlable. Nous pouvons établir le critère suivant pour la stabilité d'une méthode numérique.

Théorème: Si la fonction $\bar{F}$ est globalement lipschitzienne par rapport à la variable $y \in \mathbb{R}$, c'est-à-dire s'il existe un nombre positif K tel que:

$$\forall\, t \in [t_0, t_0+T],\ \forall (y, y') \in \mathbb{R}^2,\ \forall h \in \mathbb{R},\ |\bar{F}(t, y', h) - \bar{F}(t, y, h)| \le K|y'-y|$$

alors, la méthode numérique à un pas est stable. De plus, le nombre $S = e^{KT}$ est une constante de stabilité de la méthode.

Preuve :

Considérons des suites $(y_n)_{0\le n\le N} \in \mathbb{R}^{N+2}$, $(y'_n)_{0\le n\le N} \in \mathbb{R}^{N+2}$ et $(\delta_n)_{0\le n\le N-2}$ telles que :

$$\forall\, 0\le n\le N-2,\ \begin{cases} y_{n+2} = y_n + h_n \Phi(t_n, y_n, h_n), \\ y'_{n+2} = y'_n + h_n \Phi(t_n, y'_n, h_n) + \delta_n. \end{cases}$$

Nous avons alors :

$$\forall\, 0\le n\le N-2,\ y'_{n+2} - y_{n+2} = y'_n - y_n + h_n\left[\Phi(t_n, y'_n, h_n) - \Phi(t_n, y_n, h_n)\right] + \delta_n$$

$$\Rightarrow |y'_{n+2} - y_{n+2}| \le (1 + h_n K)\,|y'_n - y_n| + |\delta_n|$$

$$\le e^{K h_n}\,|y'_n - y_n| + |\delta_n|$$

$$= e^{K(t_{n+2} - t_n)}\,|y'_n - y_n| + |\delta_n|$$

$$\Rightarrow e^{-K t_{n+2}}\,|y'_{n+2} - y_{n+2}| \le e^{-K t_n}\,|y'_n - y_n| + e^{-K t_{n+2}}|\delta_n|.$$

De suite que :

$$\sum_{\ell=0}^{n}\left(e^{-K t_{\ell+2}}\,|y'_{\ell+2} - y_{\ell+2}| - e^{-K t_\ell}\,|y'_\ell - y_\ell|\right] \le \sum_{\ell=0}^{n} e^{-K t_{\ell+2}}\,|\delta_\ell|.$$

Il s'ensuit que :

$$e^{-K t_{n+2}}\,|y'_{n+2} - y_{n+2}| \le e^{-K t_0}\,|y'_0 - y_0| + \sum_{\ell=0}^{n} e^{-K t_{\ell+2}}\,|\delta_\ell|.$$

Il vient alors :

$$\forall\, 0\le n\le N-2,\ |y'_{n+2} - y_{n+2}| \le e^{KT}\left[|y'_0 - y_0| + \sum_{\ell=0}^{n}|\delta_\ell|\right],$$

ce qui assure l'inégalité de stabilité pour une constante de stabilité $S = e^{KT}$.

Exemples : Rappelons que la fonction f est globalement lipschitzienne par rapport à $y \in \mathbb{R}$, s'il existe un nombre positif K tel que :

$$\forall\, t\in I,\ \forall (y, y') \in \mathbb{R}^2,\ |f(t, y') - f(t, y)| \le K|y - y'|.$$

(i) Méthode d'Euler explicite

Comme

$$\forall\, t\in[t_0, t_0+T],\ \forall (y, y') \in \mathbb{R}^2,\ \forall\, h\in\mathbb{R},\ |\Phi(t, y', h) - \Phi(t, y, h)| \le K|y - y'|,$$

la méthode est stable pour une constante de stabilité $S = e^{KT}$.

(ii) Méthode d'Euler implicite

Rappelons que la fonction Φ est définie par :

$\forall (t,y,h) \in (t_0, t_0+T) \times \mathbb{R}^2$, $\Psi(t,y,h) = f\left(t+h, \left(Id - h f(t+h,.)\right)^{-1}(y)\right)$

et supposons que, pour des nombres $t \in (t_0, t_0+T]$ et $h \in \mathbb{R}$ fixés, l'application $Id - h f(t+h,.)$ soit un homéomorphisme de $\mathbb{R}$ sur $\mathbb{R}$. Dans ce cas, la fonction $y \mapsto \Psi(t,y,h)$ est bien définie. De plus, nous pouvons écrire :

$$\forall (y,y') \in \mathbb{R}^2, \quad |\Psi(t,y',h) - \Psi(t,y,h)| \leq K \left| \left(Id - h f(t+h,.)\right)^{-1}(y') - \left(Id - h f(t+h,.)\right)^{-1}(y) \right|$$

Soit alors $z' = \left(Id - h f(t+h,.)\right)^{-1}(y')$ et $z = \left(Id - h f(t+h,.)\right)^{-1}(y)$

Il vient :

$$\begin{aligned} |y'-y| &= |z' - h f(t+h, z') - z + h f(t+h, z)| \\ &\geq |z'-z| - |h| \, |f(t+h, z') - f(t+h, z)| \\ &\geq |z'-z| - K|h| \, |z'-z|. \end{aligned}$$

Lorsque $|h| < \frac{1}{K}$, nous en déduisons que :

$$|z'-z| \leq \frac{|y'-y|}{1-K|h|},$$

Ce qui conduit à l'inégalité :

$$\forall (y,y') \in \mathbb{R}^2, \quad |\Psi(t,y',h) - \Psi(t,y,h)| \leq \frac{K}{1-K|h|} |y'-y|.$$

Notons qu'en pratique, nous ne nous intéressons qu'à des valeurs de h petites, de sorte que nous pouvons introduire une notion de stabilité où le pas maximal h_{max} est borné par le nombre $\frac{1}{K}$. Pour cette notion de stabilité, la méthode d'Euler implicite s'avère stable (lorsqu'elle est bien définie) pour une constante de stabilité $e^{\frac{KT}{1-Kh_{max}}}$, qui est du même ordre que celle de la méthode d'Euler explicite lorsque le pas maximal h_{max} tend vers 0.

(iii) Méthode du point milieu

Rappelons que la fonction Ψ est définie par :

$$\forall (t,y,h) \in (t_0, t_0+T] \times \mathbb{R}^2, \quad \Psi(t,y,h) = f\left(t+\frac{h}{2}, y + \frac{h}{2} f(t,y)\right)$$

Il s'ensuit que :

$$\forall (t,h) \in (t_0, t_0+T] \times \mathbb{R}, \ \forall (y,y') \in \mathbb{R}^2, \quad |\Psi(t,y',h) - \Psi(t,y,h)| \leq K \left| y' + \frac{h}{2} f(t,y') - y - \frac{h}{2} f(t,y) \right|$$

$$\leq K|y'-y|\left[1+\frac{K|h|}{2}\right]$$

Comme pour la méthode d'Euler implicite, nous pouvons introduire une notion de stabilité pour laquelle le pas maximal h_{max} est donné, par exemple, par 1. La méthode du point milieu s'avère alors stable pour une constante de stabilité $e^{KT\left(1+\frac{Kh_{max}}{2}\right)}$, qui est du même ordre que celle de la méthode d'Euler explicite lorsque le pas maximal h_{max} tend vers 0.

4. Convergence d'une méthode numérique à un pas

Considérons la méthode numérique à un pas associée à la fonction $\Phi \in \mathcal{C}^0([t_0, t_0+T] \times \mathbb{R}^2, \mathbb{R})$.

Définition: La méthode numérique à un pas est convergente si et seulement, quelle que soit une solution $y \in \mathcal{C}^1([t_0, t_0+T], \mathbb{R})$ de l'équation différentielle:

$\forall t \in [t_0, t_0+T]$, $y'(t) = f(t, y(t))$,

les suites $(y_n)_{0\leq n\leq N} \in \mathbb{R}^{N+1}$ telles que :

$\forall 0 \leq n \leq N-1$, $y_{n+1} = y_n + h_n \Phi(t_n, y_n, h_n)$,

satisfont à la convergence :

$$\max_{0\leq n\leq N} |y(t_n) - y_n| \to 0,$$

lorsque $y(t_0) \to y_0$ et le pas maximal $h_{max} = \max_{0\leq n\leq N-1} h_n \to 0$.

Cette notion garantit que la méthode numérique à un pas considérée fournit une approximation valable d'une solution $y \in \mathcal{C}^1([t_0, t_0+T], \mathbb{R})$ de l'équation différentielle considérée, puisque l'erreur globale définie par:

$$E_N = \max_{0\leq n\leq N} |y(t_n) - y_n|,$$

est aussi petite que souhaitée lorsque le pas maximal h_{max} est suffisamment petit, et les données initiales $y(t_0)$ et y_0, suffisamment proches.

Les notions de consistance et de stabilité introduites précédemment permettent de garantir la convergence d'une méthode numérique à un pas, comme l'exprime le résultat suivant.

Théorème : Si la méthode numérique à un pas est consistante et stable, alors elle est convergente.

Preuve :

Considérons une solution $y \in \mathcal{C}^1([t_0, t_0+T], \mathbb{R})$ de l'équation différentielle

$\forall t \in [t_0, t_0+T]$, $y'(t) = f(t, y(t))$,

et désignons par ε_n l'erreur de consistance au temps t_n relative à cette solution qui est définie par :

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

Etant donnée une suite $(y_n)_{0 \le n \le N}$ telle que :

$\forall\ 0 \le n \le N-1$, $y_{n+1} = y_n + h_n \Phi(t_n, y_n, h_n)$,

la stabilité de la méthode garantit l'existence d'un nombre positif S tel que :

$$\max_{0 \le n \le N} |y(t_n) - y_n| \le S \left[|y(t_0) - y_0| + \sum_{n=0}^{N-1} |\varepsilon_n| \right].$$

La consistance de la méthode quantifie quant à elle que :

$$\sum_{n=0}^{N-1} |\varepsilon_n| \xrightarrow[h_{\max} \to 0]{} 0,$$

ce qui suffit à achever la preuve.

Ce théorème fondamental ramène la validité d'une méthode numérique à un pas aux seules notions de consistance et de stabilité dont nous avons pu observer qu'elles étaient faciles à caractériser. Nous en déduisons en particulier le corollaire suivant.

Corollaire : Supposons que :

(i) $\forall (t, y) \in [t_0, t_0+T]$, $\Phi(t, y, 0) = f(t, y)$

(ii) La fonction Φ est globalement lipschitzienne par rapport à la variable $y \in \mathbb{R}$, c'est-à-dire qu'il existe un nombre positif K

tel que :

$\forall t \in [t_0, t_0+h]$, $\forall (y,y') \in \mathbb{R}^2$, $\forall h \in \mathbb{R}$, $|\Phi(t,y,h) - \Phi(t,y',h)| \leq K|y-y'|$

La méthode numérique à un pas est alors convergente.

Preuve :

Admise

Exemples : (i) Méthode d'Euler explicite

La méthode est convergente par le corollaire précédent.

(ii) Méthode d'Euler implicite

La méthode est convergente sous les conditions de stabilité du paragraphe précédent.

(iii) Méthode du point milieu

La méthode est convergente sous les conditions de stabilité du paragraphe précédent.

5. Ordre d'une méthode numérique à un pas

Considérons la méthode numérique à un pas associée à la fonction $\Phi \in \mathcal{C}^0([t_0, t_0+T] \times \mathbb{R}^2, \mathbb{R})$.

Définition : Soit $p \in \mathbb{N}^*$. La méthode numérique à un pas est d'ordre (au moins égal à) p si et seulement s'il existe un nombre positif C tel que, quelle que soit la solution $y \in \mathcal{C}^{p+1}([t_0, t_0+T], \mathbb{R})$ de l'équation différentielle :

$\forall t \in [t_0, t_0+T]$, $y'(t) = f(t, y(t))$,

les erreurs de consistance $(\varepsilon_n)_{0 \leq n \leq N-1}$ aux temps $(t_n)_{0 \leq n \leq N-1}$ relatives à la solution y satisfont :

$$\sum_{n=0}^{N-1} |\varepsilon_n| = \sum_{n=0}^{N-1} \left| y(t_{n+1}) - y(t_n) - h_n \Phi(t_n, y(t_n), h_n) \right| \leq C h_{max}^p$$

où le pas maximal h_{max} est défini par :

$$h_{max} = \max_{0 \leq n \leq N-1} h_n .$$

Remarque : La méthode numérique est (parfois) dite d'ordre exactement égal à p si et seulement si elle est d'ordre (au moins égal à) p, mais pas d'ordre (au moins égal à) $p+1$. Les deux définitions précédentes cohabitent dans la littérature.

La notion d'ordre d'une méthode numérique à un pas permet de caractériser la précision qu'elle est capable d'atteindre.

Théorème : Soit $p \in \mathbb{N}^*$. Si la méthode numérique à un pas est d'ordre (au moins égal à) p et stable, alors, il existe des nombres positifs C et S tels que, quelle que soit la solution $y \in \mathcal{C}^{p+1}([t_0, t_0+T], \mathbb{R})$ de l'équation différentielle :

$$\forall t \in (t_0, t_0+T], \quad y'(t) = f(t, y(t)),$$

les valeurs approchées $(y_n)_{0 \leq n \leq N}$ données par la méthode numérique satisfont :

$$\max_{0 \leq n \leq N} |y(t_n) - y_n| \leq S\left[|y(t_0) - y_0| + C\, h_{max}^p\right],$$

où $h_{max} = \max_{0 \leq n \leq N-1} h_n$.

Preuve :

La stabilité de la méthode garantit l'existence d'un nombre positif S tel que :

$$\max_{0 \leq n \leq N} |y(t_n) - y_n| \leq S\left[|y(t_0) - y_0| + \sum_{n=0}^{N-1} |\varepsilon_n|\right].$$

Le théorème découle alors de l'estimation précédente de la somme des erreurs de consistance $(\varepsilon_n)_{0 \leq n \leq N-1}$.

En pratique, l'erreur initiale $|y(t_0) - y_0|$ est marginale. L'erreur de la méthode numérique est donc gouvernée par son ordre p à travers le terme $C\, h_{max}^p$. Améliorer cet ordre permet d'augmenter la précision de

la méthode.

Afin de caractériser l'ordre d'une méthode numérique, nous introduisons les notations suivantes.

Définitions: Soit $p \in \mathbb{N}^*$. Supposons que la fonction f soit de classe $\mathcal{C}^p$ sur $I \times \mathbb{R}$.
Les fonctions $\left(f^{[q]}\right)_{0 \leq q \leq p}$ sont définies sur l'ensemble $I \times \mathbb{R}$ par les formules de récurrence:

$$\forall (t,y) \in I \times \mathbb{R}, \begin{cases} f^{[0]}(t,y) = f(t,y) \\ \forall\, 0 \leq q \leq p-1, \quad f^{[q+1]}(t,y) = \partial_t f^{[q]}(t,y) + \partial_y f^{[q]}(t,y) f(t,y) \end{cases}$$

Les fonctions $\left(f^{[q]}\right)_{0 \leq q \leq p}$ sont définies de manière à satisfaire aux identités suivantes.

Lemme: Soit $p \in \mathbb{N}^*$. Supposons que la fonction f soit de classe $\mathcal{C}^p$ sur $I \times \mathbb{R}$ et considérons une solution $y \in \mathcal{C}^{p+1}([t_0, t_0+T], \mathbb{R})$ de l'équation différentielle :

$$\forall t \in [t_0, t_0+T], \quad y'(t) = f(t, y(t)).$$

Les dérivées successives de la fonction y satisfont alors les formules:

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

Preuve :
Immédiate par récurrence sur l'entier $0 \leq q \leq p$.

Nous pouvons alors établir la caractérisation suivante de l'ordre d'une méthode numérique à un pas.

Théorème: Soit $p \in \mathbb{N}^*$. Supposons que la fonction f est de classe $\mathcal{C}^p$ sur $I \times \mathbb{R}$ et que la fonction Φ est de classe $\mathcal{C}^p$ sur $[t_0, t_0+T] \times \mathbb{R}^2$. La méthode numérique à un pas est d'ordre (au moins égal à) p si et seulement si :

$$\forall\, 0 \leq q \leq p-1, \quad \forall (t,y) \in [t_0, t_0+T] \times \mathbb{R}, \quad \partial_2^q \Phi(t,y,0) = \frac{1}{q+1} f^{[q]}(t,y).$$

Preuve :

Soit $y \in \mathcal{C}^{p+1}([t_0, t_0+T], \mathbb{R})$ une solution de l'équation différentielle :

$\forall\, t \in [t_0, t_0+T]$, $y'(t) = f(t, y(t))$.

Pour $0 \le n \le N-1$, l'erreur de consistance e_n au temps t_n relative à la solution y est donnée par :

$$e_n = y(t_{n+1}) - y(t_n) - h_n \Phi(t_n, y(t_n), h_n).$$

Introduisons les fonctions $(\theta_q)_{0 \le q \le p-1}$ définies par les formules :

$$\forall\, 0 \le q \le p-1,\ \forall (t,y) \in [t_0, t_0+T] \times \mathbb{R},\quad \theta_q(t,y) = \frac{f^{(q)}(t,y)}{(q+1)!} - \frac{\partial_h^q \Phi(t,y,0)}{q!}$$

Comme $y(t_{n+1}) = y(t_n + h_n)$, la formule de Taylor avec reste intégral permet d'établir que :

$$e_n = \sum_{q=0}^{p-1} h_n^{q+1} \left[\frac{y^{(q+1)}(t_n)}{(q+1)!} - \frac{\partial_h^q \Phi(t_n, y(t_n), 0)}{q!} \right] + \underset{h_n \to 0}{O}\left(h_n^{p+1}\right),$$

De sorte que, par le lemme précédent,

$$e_n = \sum_{q=0}^{p-1} h_n^{q+1}\, \theta_q(t_n, y(t_n)) + \underset{h_n \to 0}{O}\left(h_n^{p+1}\right).$$

Vus des identités du théorème, nous savons que :

$$\forall\, 0 \le q \le p-1,\ \forall (t,y) \in [t_0, t_0+T] \times \mathbb{R},\quad \theta_q(t,y) = 0,$$

De sorte qu'il existe un nombre positif C tel que :

$$\forall\, 0 \le n \le N-1,\quad |e_n| \le C\, h_n^{p+1} \le C\, h_n\, h_{\max}^p.$$

Il s'ensuit que :

$$\sum_{n=0}^{N-1} |e_n| \le C\, T\, h_{\max}^p,$$

Et la méthode numérique est bien d'ordre (au moins égal à) p.

Réciproquement, il existe un nombre positif C tel que :

$$\sum_{n=0}^{N-1} \left| \sum_{q=0}^{p-1} h_n^{q+1}\, \theta_q(t_n, y(t_n)) + \underset{h_n \to 0}{O}\left(h_n^{p+1}\right) \right| \le C\, h_{\max}^p.$$

La preuve des identités du théorème est alors par récurrence sur l'entier $0 \le q \le p-1$. Au rang $q = 0$, l'inégalité précédente conduit au fait que :

$$\sum_{n=0}^{N-1} \left| h_n\, \theta_0(t_n, y(t_n)) \right| \le C\, h_{\max}.$$

Et la limite $h_{\max} \to 0$, le théorème sur les sommes de Riemann garantit que :

$$\int_{t_0}^{t_0+T} |\theta_0(t, y(t))|\, dt = 0,$$

puis, par continuité de la fonction $t \mapsto \theta_0(t, y(t))$, il vient :

$$\forall\, t \in [t_0, t_0+T],\quad \theta_0(t, y(t)) = 0.$$

Cette identité est vraie pour toute solution y de l'équation différentielle considérée. Le théorème de Cauchy-Lipschitz assure alors que :

$\forall (t,y) \in [t_0, t_0+T] \times \mathbb{R}$, $\theta_0(t,y) = 0$,

soit l'identité :

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

Il s'ensuit alors que :

$$\sum_{n=0}^{N-1} \left| \sum_{q=1}^{p-1} h_n^{q+1} \theta_q(t_n, y(t_n)) + \underset{h_n \to 0}{O}\left(h_n^{p+1}\right) \right| \leq C h_{max}^p,$$

Et la même preuve s'applique aux fonctions θ_1, puis, θ_2, et jusqu'à θ_p.

Ceci achève la preuve du théorème.

Nous déduisons du théorème précédent le corollaire suivant.

Corollaire : Supposons que la fonction f est de classe $\mathcal{C}^1$ sur $I \times \mathbb{R}$ et que la fonction Φ est de classe $\mathcal{C}^1$ sur $[t_0, t_0+T] \times \mathbb{R}^2$. La méthode numérique à un pas est d'ordre (au moins égal à) 1 si et seulement si elle est consistante.

Preuve :

Ces deux propriétés sont en effet équivalentes au fait que :

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

Exemples : (i) Méthode d'Euler explicite

Cette méthode est d'ordre exactement égal à 1 puisque :

$$\forall (t,y) \in [t_0, t_0+T] \times \mathbb{R}, \begin{cases} \Phi(t,y,0) = f(t,y) \\ \text{et } \partial_h \Phi(t,y,0) = 0 \neq \frac{1}{2}\left(\partial_t f(t,y) + \partial_y f(t,y) f(t,y)\right) \end{cases}$$

(en général)

(ii) Méthode d'Euler implicite

Cette méthode est d'ordre exactement égal à 1 puisque :

$$\forall (t,y) \in [t_0, t_0+T] \times \mathbb{R}, \begin{cases} \Phi(t,y,0) = f(t,y) \\ \text{et } \partial_h \Phi(t,y,0) = \partial_t f(t,y) + \partial_y f(t,y) f(t,y) \\ \neq \frac{1}{2}\left(\partial_t f(t,y) + \partial_y f(t,y) f(t,y)\right) \text{ (en général)} \end{cases}$$

(iii) Méthode du point milieu

Cette méthode est d'ordre exactement égal à 2 puisque :

$$\forall (t,y) \in [t_0, t_0+T] \times \mathbb{R}, \quad \begin{cases} \Phi(t,y,0) = f(t,y) \\ \partial_h \Phi(t,y,0) = \frac{1}{2}\left[\partial_t f(t,y) + f(t,y)\,\partial_y f(t,y)\right] \\ \partial_h^2 \Phi(t,y,0) = \frac{1}{4}\left[\partial_t^2 f(t,y) + 2 f(t,y)\,\partial_{ty}^2 f(t,y) + f(t,y)^2\,\partial_y^2 f(t,y)\right] \\ \qquad \neq \frac{1}{3}\left[\partial_t^2 f(t,y) + \partial_t f(t,y)\,\partial_y f(t,y) + 2 f(t,y)\,\partial_{ty}^2 f(t,y) + f(t,y)\left(\partial_y f(t,y)\right)^2 + f(t,y)^2 \partial_y^2 f(t,y)\right] \end{cases}$$

6. Influence des erreurs d'arrondis

Considérons la méthode numérique à un pas associée à la fonction $\Phi \in \mathcal{C}^0([t_0, t_0+T] \times \mathbb{R}^2, \mathbb{R})$ et supposons qu'elle est d'ordre (au moins égal à) $p \in \mathbb{N}^*$.

Rappelons que cette méthode revient à calculer les valeurs approchées $(y_n)_{0 \le n \le N}$ aux temps $(t_n)_{0 \le n \le N}$ données par les formules :

$\forall n \in [\![0, N-1]\!]$, $y_{n+1} = y_n + h_n \Phi(t_n, y_n, h_n)$.

En pratique, ce calcul induit des erreurs d'arrondis $(\varepsilon_n)_{0 \le n \le N-1}$ sur la valeur de $\left(\Phi(t_n, y_n, h_n)\right)_{n \in \mathbb{N}}$, et $(\delta_n)_{0 \le n \le N-1}$ sur la valeur de $(y_{n+1})_{0 \le n \le N-1}$ de sorte que les valeurs approchées $(\tilde{y}_n)_{0 \le n \le N}$ satisfont :

$\forall n \in [\![0, N-1]\!]$, $\tilde{y}_{n+1} = \tilde{y}_n + h_n \Phi(t_n, \tilde{y}_n, h_n) + h_n \varepsilon_n + \delta_n$.

Nous avons alors l'estimation suivante de l'erreur entre les valeurs approchées $\tilde{y}_n$ et y_n.

Lemme : Supposons que la méthode numérique est stable et qu'il existe des nombres positifs ε et δ tels que :

$\forall\, 0 \le n \le N-1$, $|\varepsilon_n| \le \varepsilon$ et $|\delta_n| \le \delta$.

Il existe un nombre positif S tel que :

$$\max_{0 \le n \le N} |\tilde{y}_n - y_n| \le S\left[|\tilde{y}_0 - y_0| + T\varepsilon + N\delta\right].$$

Preuve :

La stabilité de la méthode fournit l'existence d'un nombre positif S tel que

$$\max_{0\le n\le N} |\tilde{y}_n - y_n| \le S\left(|\tilde{y}_0 - y_0| + \sum_{n=0}^{N-1}\left(|h_n \varepsilon_n| + |\delta_n|\right)\right)$$
$$\le S\left(|\tilde{y}_0 - y_0| + T\varepsilon + N\delta\right)$$

Lorsque la méthode numérique est stable et d'ordre (au moins égal à) p, l'erreur totale commise par rapport aux valeurs exactes $\left(y(t_n)\right)_{0\le n\le N}$ de la solution est donc égale à :

$$\max_{0\le n\le N} |\tilde{y}_n - y(t_n)| \le \max_{0\le n\le N} |\tilde{y}_n - y_n| + \max_{0\le n\le N} |y_n - y(t_n)|$$
$$\le S\left(|\tilde{y}_0 - y_0| + |y_0 - y(t_0)| + T\varepsilon + N\delta + C h_{max}^p\right)$$

En supposant de plus que le pas est constant, c'est-à-dire que :

$\forall\ 0\le n\le N-1,\ h_n = \frac{T}{N} = h_{max}$,

il vient :

$$\max_{0\le n\le N} |\tilde{y}_n - y(t_n)| \le S\left(|\tilde{y}_0 - y_0| + |y_0 - y(t_0)| + T\varepsilon\right) + S\frac{T\delta}{h_{max}} + SC h_{max}^p$$

L'erreur commise est ainsi minimale lorsque :

$$h_{max} = \left(\frac{T\delta}{pC}\right)^{\frac{1}{p+1}}.$$

Elle est alors de l'ordre de :

$$\max_{0\le n\le N} |\tilde{y}_n - y(t_n)| \le S\left(|\tilde{y}_0 - y_0| + |y_0 - y(t_0)| + T\varepsilon + C'(T\delta)^{\frac{p}{p+1}}\right)$$

Cette limitation sur le choix du pas maximal h_{max} est cohérente, d'une part, avec le fait que les erreurs d'arrondis limitent la précision possible, et d'autre part, avec la propriété que plus le pas maximal h_{max} est petit, plus l'entier N est grand, et plus les erreurs d'arrondis s'additionnent.

En pratique, choisir un pas maximal h_{max} très petit ne garantit donc pas une précision plus grande sur la solution approchée en raison des erreurs d'arrondis. Il existe des méthodes de contrôle de choix du pas afin d'optimiser à la fois le temps de calcul et la précision de la solution approchée. Nous renvoyons à l'ouvrage de Michel Crouzeix et Alain Mignot "Analyse numérique des équations différentielles" pour de plus amples détails sur ce sujet.

II Méthodes de Runge - Kutta

1. Principe et exemples

Considérons une solution $y \in \mathcal{C}^1([t_0, t_0+T], \mathbb{R})$ de l'équation différentielle :
$\forall t \in [t_0, t_0+T]$, $y'(t) = f(t, y(t))$.
Etant donné une subdivision $t_0 < t_1 < \dots < t_N = t_0+T$ de l'intervalle $[t_0, t_0+T]$, nous cherchons à déterminer des valeurs approchées $(y_n)_{0 \le n \le N}$ des valeurs exactes $(y(t_n))_{0 \le n \le N}$ prises par la solution y aux temps $(t_n)_{0 \le n \le N}$. Dans ce but, nous pouvons utiliser l'identité :
$\forall\ 0 \le n \le N-1$, $y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} f(t, y(t))\, dt$.
Calculer les valeurs approchées $(y_n)_{0 \le n \le N}$ se réduit donc à calculer de manière approchée des intégrales. Il est alors naturel d'introduire des formules de quadrature pour le calcul de ces intégrales.

Plus précisément, nous pouvons introduire une formule de quadrature élémentaire :
$$\int_0^1 g(x)\, dx \simeq \sum_{j=1}^{q} b_j\, g(c_j),$$
dans laquelle les points $(c_j)_{1 \le j \le q}$ appartiennent au segment $[0,1]$.
Notons alors :
$\forall\ 0 \le n \le N-1$, $\forall\ 1 \le j \le q$, $t_{n,j} = t_n + c_j h_n$.
Le changement de variables $t = t_n + \theta h_n$ conduit à la formule de quadrature :
$$\int_{t_n}^{t_{n+1}} f(t, y(t))\, dt \simeq \sum_{j=1}^{q} b_j h_n f(t_{n,j}, y(t_{n,j})).$$
Il reste encore à déterminer la valeur des nombres $(y(t_{n,j}))_{1 \le j \le q}$.
Rappelons que ces nombres satisfont aussi identités :
$\forall\ 1 \le j \le q$, $y(t_{n,j}) = \int_{t_n}^{t_{n,j}} f(t, y(t))\, dt$.
Nous pouvons donc utiliser de nouvelles formules de quadrature élémentaires pour le calcul de valeurs approchées $(y_{n,j})_{1 \le j \le q}$ des valeurs exactes $(y(t_{n,j}))_{1 \le j \le q}$.

Quitte à renuméroter, nous pouvons supposer que :

$0 \leq c_1 \leq c_2 \leq \dots \leq c_q \leq 1$.

Dans ce cas, nous pouvons définir des méthodes de quadrature élémentaires, de sorte que :

$$\forall\, 1 \leq j \leq q, \quad \int_0^{c_j} g(t)\,dt \simeq \sum_{i=1}^{j} a_{ji}\, g(c_i),$$

chacune de ces méthodes pouvant être distincte des autres. Nous avons alors le calcul approché :

$$\int_{t_n}^{t_{n,j}} f(t, y(t))\,dt \simeq \sum_{i=1}^{j} h_n\, a_{ji}\, f\left(t_{n,i}, y(t_{n,i})\right).$$

Ceci permet de définir des valeurs approchées $(y_{n,j})_{1 \leq j \leq q}$, puis de définir à chaque étape la valeur approchée y_{n+1} de la valeur approchée y_n.

Ce principe conduit à la définition suivante des méthodes de Runge-Kutta

Définition : Soit $(c_j)_{1 \leq j \leq q} \in [0,1]^q$ tel que : $c_1 \leq \dots \leq c_q$. Etant donnés des coefficients réels $(b_j)_{1 \leq j \leq q}$ et $(a_{j,i})_{1 \leq i \leq j \leq q}$, la méthode de Runge-Kutta de résolution de l'équation différentielle :

$$\forall\, t \in (t_0, t_0 + T], \quad y'(t) = f(t, y(t))$$

est définie par les formules de récurrence suivantes pour le calcul des valeurs approchées $(y_n)_{0 \leq n \leq N}$ aux temps $(t_n)_{0 \leq n \leq N}$:

$$\forall\, n \in [\![0, N-1]\!], \begin{cases} \forall\, 1 \leq j \leq q, \begin{cases} t_{n,j} = t_n + c_j h_n \\ y_{n,j} = y_n + h_n \sum_{i=1}^{j} a_{ji} f(t_{n,i}, y_{n,i}) \end{cases} \\ y_{n+1} = y_n + h_n \sum_{j=1}^{q} b_j f(t_{n,j}, y_{n,j}) \end{cases}$$

Notation : Soit $(c_j)_{1 \leq j \leq q} \in (0,1]^q$ tel que : $c_1 \leq \dots \leq c_q$. La méthode de Runge-Kutta associée aux coefficients réels $(b_j)_{1 \leq j \leq q}$ et $(a_{j,i})_{1 \leq i \leq j \leq q}$ est représentée par le tableau :

$$\begin{array}{c|ccc} c_1 & a_{11} & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ c_q & a_{q1} & \cdots & a_{qq} \\ \hline & b_1 & \cdots & b_q \end{array}$$

Exemples : (i) Méthode d'Euler explicite

C'est la méthode de Runge - Kutta associée au tableau :

0	0
	1

Cette méthode correspond à l'application de la méthode de quadrature des rectangles à gauche.

(ii) Méthode d'Euler implicite

C'est la méthode de Runge - Kutta associée au tableau :

1	1
	1

Cette méthode correspond à l'application de la méthode de quadrature des rectangles à droite.

(iii) Méthode du point milieu

C'est la méthode de Runge - Kutta associée au tableau :

0	0	0
$\frac{1}{2}$	$\frac{1}{2}$	0
	0	1

Cette méthode correspond à l'application de la méthode de quadrature du point milieu.

(iv) Méthode de Heun

C'est la méthode de Runge - Kutta associée au tableau :

0	0	0
1	1	0
	$\frac{1}{2}$	$\frac{1}{2}$

Cette méthode correspond à l'application de la méthode de quadrature des trapèzes.

(v) Méthode de Runge - Kutta classique

C'est la méthode de Runge - Kutta associée au tableau :

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

Cette méthode correspond à l'application de la méthode de quadrature de Simpson.

Nous pouvons remarquer que les méthodes de Runge - Kutta sont explicites dès lors que les coefficients $(a_{j,i})_{1 \leq i \leq j \leq q}$ satisfont :
$\forall\, 1 \leq j \leq q,\ a_{j,j} = 0$.
Nous nous limitons dans la suite de ce cours à ce cas particulier, même si certains des résultats qui suivent s'étendent aux méthodes de Runge - Kutta implicites.

2. Convergence et ordre

Les méthodes de Runge - Kutta sont des méthodes numériques à un pas.

Théorème : La méthode de Runge - Kutta associée aux points $(c_j)_{1 \leq j \leq q}$, et aux coefficients réels $(b_j)_{1 \leq j \leq q}$ et $(a_{j,i})_{1 \leq i < j \leq q}$ est la méthode numérique à un pas associée à la fonction $\Phi \in \mathcal{C}^0\left([t_0, t_0+T] \times \mathbb{R}^2, \mathbb{R}\right)$ donnée par la formule :
$\forall (t, y, h) \in [t_0, t_0+T] \times \mathbb{R}^2,\ \Phi(t, y, h) = \sum_{j=1}^{q} b_j f(t + c_j h, y_j)$,
où les points $(y_j)_{1 \leq j \leq q}$ sont donnés par les formules :
$y_1 = y$ et $\forall\, 2 \leq j \leq q,\ y_j = y + h \sum_{i=1}^{j-1} a_{j,i} f(t + c_i h, y_i)$.

Preuve :

Immédiat.

Il est donc possible de leur appliquer les résultats généraux quant à la consistance, la stabilité et la convergence de ces méthodes.

Lemme: La méthode de Runge - Kutta associée aux points $(c_j)_{1 \le j \le q}$, et aux coefficients réels $(b_j)_{1 \le j \le q}$ et $(a_{ji})_{1 \le i < j \le q}$ est consistante si et seulement si :

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

Preuve:

En effet, nous savons que :

$\forall (t,y) \in [t_0, t_0+T] \times \mathbb{R}$, $\Phi(t,y,0) = \left(\sum_{j=1}^{q} b_j \right) f(t,y)$.

L'équivalence résulte alors de la condition de consistance des méthodes numériques à un pas.

En ce qui concerne la stabilité, nous pouvons établir le résultat suivant.

Lemme: Supposons que la fonction f est globalement lipschitzienne par rapport à y, c'est-à-dire qu'il existe un nombre positif K tel que :

$\forall t \in [t_0, t_0+T]$, $\forall (y,y') \in \mathbb{R}^2$, $|f(t,y) - f(t,y')| \le K |y - y'|$.

La méthode de Runge - Kutta associée aux points $(c_j)_{1 \le j \le q}$, et aux coefficients réels $(b_j)_{1 \le j \le q}$ et $(a_{ji})_{1 \le i < j \le q}$ est stable dès lors que le pas maximal h_{max} est borné par $\frac{1}{K}$.

Preuve:

Il suffit de vérifier que la fonction Φ est globalement lipschitzienne par rapport à la variable $y \in \mathbb{R}$. Soit $(t,h) \in [t_0, t_0+T] \times [-h_{max}, h_{max}]$. Alors pouvons calculer :

$$\forall (y,y') \in \mathbb{R}^2, \quad |\Phi(t,y',h) - \Phi(t,y,h)| \le K \sum_{j=1}^{q} |b_j| \, |y'_j - y_j|,$$

où

$$\forall 1 \le j \le q, \quad y_j^{(i)} = y^{(i)} + h \sum_{i=1}^{j-1} a_{ji} f(t + c_i h, y_i).$$

Au rang 1, nous avons :

$|y'_1 - y_1| = |y' - y|$.

Supposons alors que : $\forall\, 1 \leq \ell \leq j-1,\ |y'_\ell - y_\ell| \leq |y' - y| \sum_{k=0}^{\ell-1} (K|h|\alpha)^k$,

où $\alpha = \max_{2 \leq j \leq q} \sum_{i=1}^{j-1} |a_{ji}|$. Au rang j, il vient alors :

$$|y'_j - y_j| \leq |y' - y| + K|h|\alpha \max_{1 \leq i \leq j-1} |y'_i - y_i|$$
$$\leq |y' - y| \sum_{i=0}^{j-1} (K|h|\alpha)^i,$$

ce qui achève la récurrence. Il s'ensuit que :

$$\forall (y, y') \in \mathbb{R}^2,\ |\Phi(t, y', h) - \Phi(t, y, h)| \leq K \sum_{j=1}^{q} |b_j| \sum_{i=0}^{j-1} (K|h|\alpha)^i |y' - y|$$
$$\leq \Lambda |y' - y|,$$

où $\Lambda = K \sum_{j=1}^{q} |b_j| \sum_{i=0}^{j-1} (K h_{\max} \alpha)^i$. La méthode de Runge-Kutta est donc stable sous la condition $h_{\max} \leq 1$.

Remarque : La constante de stabilité de la méthode de Runge-Kutta dépend du pas maximal $h_{\max}$. D'après le lemme précédent, elle est de l'ordre de $e^{\Lambda T}$, où $\Lambda = K \sum_{j=1}^{q} |b_j| \sum_{i=0}^{j-1} (K h_{\max} \alpha)^i$.

Lorsque le pas maximal $h_{\max}$ tend vers 0, nous vérifions que :

$$\Lambda \underset{h_{\max} \to 0}{\longrightarrow} K \sum_{j=1}^{q} |b_j|$$

Lorsque les coefficients $(b_j)_{1 \leq j \leq q}$ sont positifs, ce qui est naturel en pratique, et sous la condition de consistance $\sum_{j=1}^{q} b_j = 1$, nous obtenons :

$$\Lambda \underset{h_{\max} \to 0}{\longrightarrow} K.$$

Cette limite est la meilleure possible puisqu'elle correspond au cas explicite d'une fonction f linéaire. Les méthodes de Runge-Kutta présentent donc le précieux avantage d'être très stable (au moins lorsque le pas maximal $h_{\max}$ est suffisamment petit).

Les deux lemmes précédents permettent d'établir la convergence d'une méthode de Runge-Kutta.

Théorème : Supposons que la fonction f est globalement lipschitzienne par rapport à $y \in \mathbb{R}$, c'est qu'il existe un nombre positif K tel que :

$\forall t \in [t_0, t_0+T]$, $\forall (y,y') \in \mathbb{R}^2$, $|f(t,y) - f(t,y')| \leq K |y-y'|$,

et considérons des points $(c_{ij})_{1 \leq i \leq j \leq q} \in [0,1]^{\frac{q}{2}}$, et des coefficients réels $(b_j)_{1 \leq j \leq q}$ et $(a_{ji})_{1 \leq i < j \leq q}$ tels que:

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

La méthode de Runge - Kutta associée est convergente.

Preuve:

Immédiat.

Exemples: Les méthodes de Heun et de Runge - Kutta classiques sont convergentes.

Il est enfin permis d'étendre les résultats généraux sur l'ordre des méthodes numériques à un pas aux méthodes de Runge - Kutta.

Théorème: Une méthode de Runge - Kutta associée aux points $(c_{ij})_{1 \leq i \leq j \leq q}$ et aux coefficients réels $(b_j)_{1 \leq j \leq q}$ et $(a_{ji})_{1 \leq i < j \leq q}$ sont:

(i) d'ordre (au moins égal à) 1 si et seulement si:

$$\sum_{j=1}^{q} b_j = 1$$

(ii) d'ordre (au moins égal à) 2 si et seulement si:

$$\sum_{j=1}^{q} b_j = 2 \sum_{j=1}^{q} b_j c_j = 2 \sum_{j=1}^{q} \sum_{i=1}^{j-1} b_j a_{ji} = 1$$

Preuve:

(i) est une conséquence directe du résultat de consistance précédent. En ce qui concerne (ii), rappelons que:

$\forall t \in [t_0, t_0+T]$, $\forall (y,h) \in \mathbb{R}^2$, $\Phi(t,y,h) = \sum_{j=1}^{q} b_j f(t+c_j h, y_j)$,

où $\forall 1 \leq j \leq q$, $y_j = y + h \sum_{i=1}^{j-1} a_{ji} f(t+c_i h, y_i)$. Il vient donc:

$$\partial_h \Phi(t,y,h) = \sum_{j=1}^{q} b_j \Big(c_j \partial_t f(t+c_j h, y_j) + \partial_y f(t+c_j h, y_j) \times \sum_{i=1}^{j-1} \big(a_{ji} f(t+c_i h, y_i) + h a_{ji} \partial_t f(t+c_i h, y_i) \big) \Big)$$

De suite que:

$$\partial_h \Phi(t,y,0) = \sum_{j=1}^{q} b_j \Big(c_j \partial_t f(t,y) + \partial_y f(t,y) \sum_{i=1}^{j-1} a_{ji} f(t,y) \Big),$$

ce qui conduit à l'équivalence de (ii) par le résultat général sur l'ordre d'une méthode numérique à un pas.

En pratique, les méthodes de quadrature élémentaires sous-jacentes aux méthodes de Runge-Kutta sont au moins d'ordre 0, ce qui revient à dire que:
$\sum_{j=1}^{q} b_j = 1$ et $\forall\, 1 \le j \le q,\ c_j = \sum_{i=1}^{j-1} a_{ji}$.
Dans ce cas, il est possible d'étendre le théorème précédent de la façon suivante.

Théorème: Considérons des points $(c_j)_{1\le j\le q} \in [0,1]^q$ et des coefficients réels $(b_j)_{1\le j\le q}$ et $(a_{ji})_{1\le i<j\le q}$ tels que:
$\sum_{j=1}^{q} b_j = 1$ et $\forall\, 1 \le j \le q,\ c_j = \sum_{i=1}^{j-1} a_{ji}$.
La méthode de Runge-Kutta associée est d'ordre (au moins égal à) un. Elle est d'ordre (au moins égal à):
(i) 2 si et seulement si: $\sum_{j=1}^{q} b_j c_j = \frac{1}{2}$.
(ii) 3 si et seulement si: $\sum_{j=1}^{q} b_j c_j = \frac{1}{3}$ et $\sum_{j=1}^{q} b_j c_j^2 = \frac{1}{3} \sum_{i=1}^{j-1} \sum_{j=1}^{q} a_{ji} b_j c_i = \frac{1}{6}$.

Preuve:
Elle est similaire à celle du théorème précédent.

Exemples: (i) Méthode de Heun
La méthode de Heun est d'ordre exactement égal à 2.

(ii) Méthode de Runge-Kutta classique
La méthode de Runge-Kutta classique est d'ordre exactement égal à 4. Elle présente l'avantage d'un ordre élevé, d'une complexité algorithmique faible et d'une grande stabilité, raisons pour lesquelles elle est souvent employée en pratique.

III Difficultés numériques de la résolution approchée des équations différentielles ordinaires

Considérons le problème de Cauchy :

$$(\mathcal{C}) \begin{cases} \forall t \in (t_0, T+t_0], \quad y'(t) = f(t, y(t)) \\ y(t_0) = y_0 \end{cases}$$

où $y_0 \in \mathbb{R}$ et $f : (t_0, t_0+T] \times \mathbb{R} \to \mathbb{R}$. Afin de résoudre ce problème, nous pouvons appliquer les méthodes numériques à un pas que nous avons introduites précédemment, et les résultats de convergence que nous avons démontrés garantissent alors que nous obtiendrons des solutions approchées raisonnables de la solution de ce problème de Cauchy.

Ces résultats reposent néanmoins sur des hypothèses précises sur la fonction f, lesquelles ne sont pas toujours vérifiées. De plus, même dans le cas où ces hypothèses sont vérifiées, les estimations d'erreur fournies par les résultats de convergence précédents peuvent être très mauvaises, parce que les constantes (par exemple, de stabilité) qui apparaissent dans ces estimations sont très grandes.

La mise en pratique des méthodes numériques à un pas de haute ainsi à un certain nombre de difficultés numériques que nous allons maintenant décrire.

1. Problèmes mathématiquement et numériquement bien posés

Définition : Soit $f \in \mathcal{C}^0([t_0, t_0+T] \times \mathbb{R}, \mathbb{R})$. Le problème de Cauchy

$$(\mathcal{C}) \begin{cases} \forall t \in (t_0, t_0+T], \quad y'(t) = f(t, y(t)), \\ y(t_0) = y_0 \end{cases}$$

est mathématiquement bien posé si et seulement si :

(i) il existe une unique solution $y \in \mathcal{C}^1([t_0, t_0+T], \mathbb{R})$ pour

chaque donnée initiale $y_0 \in \mathbb{R}$.

(ii) la solution y dépend de manière continue de la donnée initiale, autrement dit, l'application $y_0 \mapsto y(t)$ est continue pour chaque $t \in [t_0, t_0+T]$.

Le théorème de Cauchy - Lipschitz (global) garantit qu'un problème de Cauchy est mathématiquement bien posé.

Théorème de Cauchy - Lipschitz (global): Si la fonction $f \in \mathcal{C}^0([t_0, t_0+T] \times \mathbb{R}, \mathbb{R})$ est globalement lipschitzienne par rapport à la variable $y \in \mathbb{R}$, c'est-à-dire s'il existe un nombre positif k tel que:

$\forall t \in [t_0, t_0+T]$, $\forall (y, y') \in \mathbb{R}^2$, $|f(t,y') - f(t,y)| \leq k|y'-y|$,

alors, le problème de Cauchy:

$$(\mathcal{C}) \begin{cases} \forall t \in [t_0, t_0+T], \ y'(t) = f(t, y(t)), \\ y(t_0) = y_0, \end{cases}$$

est mathématiquement bien posé.

Preuve:

Voir ci-dessus.

Certains problèmes de Cauchy ne sont pas mathématiquement bien posés.

Exemple: Le problème de Cauchy

$$\begin{cases} \forall t \in \mathbb{R}_+, \ y'(t) = 2\sqrt{|y(t)|} \\ y(0) = 0 \end{cases}$$

admet les solutions:

(i) $\forall t \in \mathbb{R}_+, \ y(t) = 0$

(ii) $\forall a \in \mathbb{R}_+, \ \forall t \in \mathbb{R}_+, \ y(t) = \begin{cases} 0 & \text{si } 0 \leq t \leq a, \\ (t-a)^2 & \text{sinon,} \end{cases}$

Il n'y a pas d'unicité de la solution.

Définition : Soit $f \in \mathcal{C}^0([t_0, t_0+T] \times \mathbb{R}, \mathbb{R})$. Le problème de Cauchy

$$(\mathcal{C}) \begin{cases} \forall t \in [t_0, t_0+T], \ y'(t) = f(t, y(t)), \\ y(t_0) = y_0 \end{cases}$$

est numériquement bien posé si et seulement si la continuité d'une solution y par rapport à la donnée initiale y_0 est suffisamment bonne pour que la solution ne soit pas perturbée par une erreur initiale ou une erreur d'arrondi sur la fonction f.

Cette définition n'est pas très précise. La continuité de la solution y par rapport à la donnée initiale y_0 se mesure par la taille des constantes qui apparaissent dans le lemme de Gronwall. Ces constantes doivent être suffisamment petites par rapport à la précision des calculs. En particulier, cette définition dépend de la précision possible.

Il existe des problèmes de Cauchy qui sont mathématiquement bien posés, mais numériquement mal posés.

Exemple : Le problème de Cauchy :

$$(\mathcal{C}) \begin{cases} \forall t \in [0, 20], \ y'(t) = 3\, y(t) - 1, \\ y(0) = \frac{1}{3}, \end{cases}$$

a pour solution :

$\forall t \in [0, 20], \ y(t) = t + \frac{1}{3}$.

Par le théorème de Cauchy-Lipschitz (global), il est mathématiquement bien posé. Cependant, la solution associée à la donnée initiale $y_\varepsilon(0) = \frac{1}{3} + \varepsilon$ pour $\varepsilon \neq 0$ est égale à :

$\forall t \in [0, 20], \ y_\varepsilon(t) = t + \frac{1}{3} + \varepsilon e^{3t}$.

Nous observons en particulier que :

$y_\varepsilon(20) - y(20) = \varepsilon e^{30} \approx 10^{13} \varepsilon$

Ce problème est donc numériquement mal posé pour une précision des calculs de l'ordre de 10^{-20}, mais redevient mal-

numériquement bien posé pour une précision de l'ordre de 10^{-20}.

Moins qu'il n'est pas possible de résoudre numériquement un problème de Cauchy numériquement mal posé puisque les erreurs d'arrondi sont alors incontrôlables.

2. Problèmes bien conditionnés et problèmes raides

Même lorsqu'un problème de Cauchy est mathématiquement et numériquement bien posés, il peut s'avérer difficile d'en déterminer une solution approchée en un temps raisonnable.

Définition: Soit $f \in \mathcal{C}^0([t_0, t_0+T] \times \mathbb{R}^n, \mathbb{R}^n)$. Le problème de Cauchy

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

est bien conditionné si les méthodes numériques classiques peuvent en donner la solution en un temps raisonnable.

Cette définition n'est pas très précise. Les méthodes numériques classiques dont il est question se réduisent en général aux méthodes numériques à un pas explicites (et parfois implicites). Le temps de calcul dépend du matériel informatique utilisé.

Exemple: Le problème de Cauchy:

$$(C) \begin{cases} \forall t \in [0, 1], \ y'(t) = -250\, y(t) + 30, \\ y(0) = \frac{1}{5}, \end{cases}$$

a pour unique solution:

$\forall t \in [0,1], \ y(t) = \frac{1}{5}$.

Par le théorème de Cauchy-Lipschitz (global), il est mathématiquement bien posé. De plus, l'unique solution y_ε associée à la donnée initiale $y_\varepsilon(0) = \frac{1}{5} + \varepsilon$ pour $\varepsilon > 0$ est égale à:

$\forall t \in [0,2],\ y_\varepsilon(t) = \frac{1}{5} + \varepsilon e^{-250t}$,

De sorte que ce problème de Cauchy est numériquement bien posé. Cependant, la méthode d'Euler explicite avec un pas constant $h = \frac{1}{N}$, où $N \in \mathbb{N}^*$, fournit des valeurs approchées $(y_n)_{0 \le n \le N}$ de la solution y aux temps $\left(t_n = \frac{n}{N}\right)_{0 \le n \le N}$ données par la formule de récurrence :

$\forall\ 0 \le n \le N-1,\ y_{n+1} = y_n + \frac{1}{N}\left(-250\, y_n + 30\right)$,

soit par les formules :

$\forall\ 0 \le n \le N,\ y_n = \frac{1}{5} + \left(1 - \frac{250}{N}\right)^n \left(y_0 - \frac{1}{5}\right)$.

L'erreur ainsi commise est de l'ordre de :

$$\max_{0 \le n \le N} \left| y_n - y(t_n) \right| = \left(\frac{250}{N} - 1\right)^N \left| y_0 - \frac{1}{5} \right|.$$

Lorsque $N \le 75$, quantité qui peut s'avérer très grande lorsque N est par exemple de l'ordre de 50. Ce problème de Cauchy est ainsi mal conditionné pour la méthode d'Euler explicite. Au contraire, la méthode d'Euler implicite avec le même pas constant $h = \frac{1}{N}$ fournit des valeurs approchées $(y_n)_{0 \le n \le N}$ de la solution y aux temps $(t_n)_{0 \le n \le N}$ données par la formule de récurrence :

$\forall\ 0 \le n \le N-1,\ y_{n+1} = y_n + \frac{1}{N}\left(-250\, y_{n+1} + 30\right)$,

soit par les formules :

$\forall\ 0 \le n \le N,\ y_n = \frac{1}{5} + \frac{y_0 - \frac{1}{5}}{\left(1 + \frac{250}{N}\right)^n}$.

L'erreur ainsi commise est de l'ordre de :

$$\max_{0 \le n \le N} \left| y_n - y(t_n) \right| = \left| y_0 - \frac{1}{5} \right|.$$

Le problème de Cauchy est donc bien conditionné pour la méthode d'Euler implicite.

Notons qu'il existe des problèmes de Cauchy qui sont mathématiquement et numériquement bien posés, mais mal conditionnés pour toutes les méthodes numériques classiques. Ce sont des problèmes raides. Nous renvoyons à l'ouvrage de Michel Crouzeix et Alain Mignot « Analyse numérique des équations différentielles » pour de plus amples détails sur ce sujet.