

Équations différentielles

Étude théorique et schémas numériques

Jawad SALHI

Faculté des Sciences et Techniques - Errachidia

A.U. 2022/2023

Plan

- 1 Introduction
- 2 Théorie de Cauchy-Lipschitz
 - Enoncé du théorème principal
 - Démonstration du théorème de Cauchy-Lipschitz
- 3 Approximation numérique des EDO
 - La notion de schéma numérique
 - Schémas numériques classiques (à un pas)
 - Problème raide et notion de stabilité absolue
 - Convergence des schémas explicites à un pas
 - Convergence des schémas implicites à un pas
 - Schémas de Runge-Kutta
- 4 Notes bibliographiques et remarques

Définitions de base

Definition

- On appelle équation différentielle d'ordre 1 les équations de la forme

$$y'(t) = f(t, y(t)), \quad \forall t \in I \quad (1)$$

où I est un intervalle de $\mathbb{R}$, l'inconnue $y : I \rightarrow \mathbb{R}^m$ est une fonction vectorielle de la variable t et $f : I \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ est une fonction donnée.

- On appelle équation différentielle d'ordre p les équations de la forme

$$y^{(p)}(t) = f\left(t, y(t), y'(t), \dots, y^{(p-1)}(t)\right), \quad \forall t \in I$$

où I est un intervalle de $\mathbb{R}$, $y : I \rightarrow \mathbb{R}^m$ est une fonction vectorielle de la variable t et $f : I \times (\mathbb{R}^m)^p \rightarrow \mathbb{R}^m$.

REMARQUE : Toute équation différentielle d'ordre p peut se réécrire comme une équation différentielle d'ordre 1. En effet, en posant

$$x_1(t) = y(t), x_2(t) = y'(t), x_3(t) = y''(t), \dots, x_p(t) = y^{(p-1)}(t)$$

le problème devient

$$x_1'(t) = x_2(t), x_2'(t) = x_3(t), \dots, x_{p-1}'(t) = x_p(t)$$

et

$$x_p'(t) = f(t, x_1(t), x_2(t), \dots, x_p(t)).$$

Ce qui s'écrit encore

$$X'(t) = F(t, X(t)),$$

en posant

$$X(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_p(t) \end{pmatrix} \quad \text{et} \quad F(t, X) = \begin{pmatrix} x_2 \\ x_3 \\ \vdots \\ f(t, x_1, x_2, \dots, x_p) \end{pmatrix}$$

Attention : ici chacun des $x_i(t) = y^{(i-1)}(t)$ est de dimension m , donc au final le système est bien de dimension $X(t) \in \mathbb{R}^{m \times p}$.

Definition

Soit $I \subset \mathbb{R}$ un intervalle ouvert de $\mathbb{R}$ et $f : I \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ une application. On appelle **solution** de l'équation différentielle (1), tout couple (J, y) où $J \subset I$ est un sous-intervalle de I et y une fonction dérivable définie sur J telle que

$$\forall t \in J, y'(t) = f(t, y(t)).$$

Remarque

Dans le cas général, il est tout à fait possible que l'intervalle J ne soit pas égal à I et qu'on ne puisse pas faire mieux. La notion de solution maximale permet de donner un sens précis à "on ne peut pas faire mieux".

Definition (Solution maximale)

On dit que (J, y) est une solution maximale de (1) s'il n'existe pas de solution $(\tilde{J}, \tilde{y})$ vérifiant $J \subsetneq \tilde{J}$ et $\tilde{y}|_J = y$.

En utilisant le lemme de Zorn, on peut montrer le résultat suivant, sans aucune hypothèse sur l'équation considérée.

Proposition

*Pour toute solution (J, y) de (1), il existe **au moins une** solution maximale $(\tilde{J}, \tilde{y})$ qui prolonge (J, y) , c'est-à-dire telle que $J \subset \tilde{J}$ et $\tilde{y}|_J = y$.*

Definition (Solution globale)

Toute solution (J, y) de (1) définie sur l'intervalle $J = I$ tout entier est dite **globale**.

Le problème de Cauchy : définition

Il arrive qu'on ne recherche pas toutes les solutions d'une EDO mais seulement celles qui vérifient certaines conditions, dites *conditions initiales de Cauchy* ou tout simplement *conditions de Cauchy*.

Definition (Problème de Cauchy)

Étant donné un intervalle I , une fonction $f : I \times \mathbb{R}^m \rightarrow \mathbb{R}^m$, $t_0 \in I$ et $y_0 \in \mathbb{R}^m$, résoudre le **problème de Cauchy**

$$\begin{cases} y'(t) = f(t, y(t)), & \forall t \in I \\ y(t_0) = y_0 \end{cases} \quad (2)$$

signifie déterminer toutes les fonctions $y : I \rightarrow \mathbb{R}^m$ solutions de l'équation différentielle et vérifiant $y(t_0) = y_0$.

Une fonction y qui satisfait (2) est appelée **intégrale de l'équation différentielle**.

Le théorème fondamental de cette section est le suivant :

Théorème (Cauchy-Lipschitz, forme faible)

Si la fonction $f : I \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ est de classe C^1 alors pour toute donnée de Cauchy $(t_0, y_0) \in I \times \mathbb{R}^m$, il existe un intervalle $J \subset I$ contenant t_0 tel qu'il existe dans J une unique solution du problème de Cauchy associé. En particulier, pour toute telle donnée, il existe une unique solution maximale associée et toute autre solution vérifiant la condition de Cauchy est une restriction de cette solution maximale.

En réalité, ce théorème est encore vrai sous des hypothèses plus faibles.

Definition

Une fonction continue $f : I \times \mathbb{R}^m \mapsto \mathbb{R}^m$ est dite **localement Lipschitzienne** par rapport à la variable d'état (ou à la seconde variable), si pour tout $(t_0, y_0) \in I \times \mathbb{R}^m$, il existe $C_{t_0, y_0} > 0$ et un voisinage U de (t_0, y_0) dans $I \times \mathbb{R}^m$ tel que $\forall t \in I, \forall y_1, y_2 \in \mathbb{R}^m$, tels que $(t, y_1) \in U$ et $(t, y_2) \in U$, on a :

$$\|F(t, y_1) - F(t, y_2)\| \leq C_{t_0, y_0} \|y_1 - y_2\|. \quad (3)$$

Grâce au théorème des accroissements finis, on vérifie que toute fonction de classe C^1 est localement lipschitzienne par rapport à sa deuxième variable. C'est pourquoi le théorème suivant est bien plus fort que le précédent.

Théorème (Cauchy-Lipschitz, forme forte)

Le théorème de Cauchy-Lipschitz est encore vrai si f est continue et localement lipschitzienne par rapport à la variable d'état.

La démonstration de ce théorème peut se faire de plusieurs manières. Elles partent toutes de la constatation que (J, y) est solution du problème de Cauchy si et seulement si $t_0 \in J$ et si y est une fonction continue sur J qui vérifie

$$y(t) = y_0 + \int_{t_0}^t f(s, y(s)) ds, \forall t \in J. \quad (4)$$

Il s'agit donc de résoudre l'équation intégrale (4) dans l'espace fonctionnel $C^0(J, \mathbb{R}^m)$, c'est pourquoi il est naturel que les preuves du théorème utilisent de façon fondamentale les grands théorèmes de l'analyse fonctionnelle : ou bien le théorème d'Ascoli (méthode de compacité) ou bien le théorème du point fixe de Banach.

Le lemme de Gronwall

Un outil central dans tous les problèmes d'équations différentielles est le lemme de Gronwall qui permet de déduire des bornes sur les solutions à partir d'inégalité intégrales qu'elles vérifient.

Lemme (de Gronwall)

Soit $[a, b[\subset \mathbb{R}$, $C \in \mathbb{R}$ et z, φ deux fonctions continues sur $[a, b[$ à valeurs réelles. On suppose que

- φ est positive.
- L'inégalité suivante est vérifiée :

$$z(t) \leq C + \int_a^t \varphi(s)z(s) ds, \forall a \leq t < b.$$

Alors, on a l'estimation

$$z(t) \leq C \exp \left(\int_a^t \varphi(s) ds \right), \forall a \leq t < b.$$

Unicité

Dans le cas où f est localement Lipschitzienne par rapport à sa seconde variable, on peut démontrer l'unicité d'une éventuelle solution en utilisant le Lemme de Gronwall.

En effet, soient $(J_1, y_1), (J_2, y_2)$ deux solutions du même problème de Cauchy en t_0 . On veut montrer que y_1 et y_2 sont égales sur $J_0 := J_1 \cap J_2$. Pour cela on introduit l'ensemble

$$S = \{t \in J_0, \text{ tel que } y_1(s) = y_2(s), \forall s \in [t_0, t]\},$$

où $[t_0, t]$ est remplacé par $[t, t_0]$ si $t < t_0$.

Cet ensemble est non vide car il contient t_0 (y_1 et y_2 vérifient la même donnée de Cauchy à l'instant t_0). On va montrer que

$$S \cap [t_0, +\infty[= J_0 \cap [t_0, +\infty[$$

(la même idée montrerait l'égalité de $S \cap]-\infty, t_0]$ et de $J_0 \cap]-\infty, t_0]$).

Unicité

Raisonnons par l'absurde et supposons que $S \cap [t_0, +\infty[\neq J_0 \cap [t_0, +\infty[$. On pose alors $t^* = \sup(S)$. On a $t^* \geq t_0$ et $t^* \in J_0$. En effet, si ce n'était pas le cas on aurait $t^* \in \partial J_0$ et alors $y_1 = y_2$ sur $[t_0, \sup J_0[$ et donc $y_1 = y_2$ sur $J_0 \cap [t_0, +\infty[$ par continuité de y_1 et y_2 . Ceci contredit l'hypothèse. Par ailleurs, par continuité de y_1 et y_2 , on sait que $y_1(t^*) = y_2(t^*) = \tilde{y}$.

Soit L une constante de Lipschitz de f sur le compact $K = [t^*, t^* + 1] \times \bar{B}(\tilde{y}, 1)$. Par continuité, il existe $\delta > 0$ tel que $t^* + \delta \in J_0$ et tel que

$$y_i(t) \in \bar{B}(\tilde{y}, 1), \forall t \in [t^*, t^* + \delta], \forall i = 1, 2.$$

Par ailleurs, comme y_1 et y_2 vérifient l'équation on a

$$y_i(t) = y_i(t^*) + \int_{t^*}^t f(s, y_i(s)) ds.$$

Unicité

Par soustraction, on trouve

$$|y_1(t) - y_2(t)| \leq \int_{t^*}^t |f(s, y_1(s)) - f(s, y_2(s))| ds, \quad \forall t \in [t^*, t^* + \delta].$$

Comme y_1 et y_2 prennent leur valeurs dans K , on en déduit

$$|y_1(t) - y_2(t)| \leq L \int_{t^*}^t |y_1(s) - y_2(s)| ds, \quad \forall t \in [t^*, t^* + \delta].$$

Le lemme de Gronwall donne alors $y_1(t) = y_2(t)$ pour tout $t \in [t^*, t^* + \delta]$. Ceci montre que $t^* + \delta$ est dans S et contredit donc la définition de t^* .

Existence : la preuve par le théorème du point fixe

Soit $J = [t_0 - \alpha, t_0 + \alpha]$ un intervalle inclus dans I (on rappelle qu'on a supposé que I est ouvert) contenant t_0 . On pose $y^0(t) = y_0$ pour tout $t \in J$ et on construit, par récurrence, la suite de fonctions

$$y^{n+1}(t) = y_0 + \int_{t_0}^t f(s, y^n(s)) ds, \quad \forall t \in J.$$

Ceci revient à définir $y^{n+1} = \Phi(y^n)$ où $\Phi : \mathcal{C}^0(J, \mathbb{R}^m) \mapsto \mathcal{C}^0(J, \mathbb{R}^m)$ est l'application qui à y associe

$$\Phi(y)(t) = y_0 + \int_{t_0}^t f(s, y(s)) ds, \quad \forall t \in J.$$

Résoudre l'équation (4) revient à trouver un point fixe de l'application Φ . Comme J est compact on peut munir $E = \mathcal{C}^0(J, \mathbb{R}^d)$ de la norme infinie, ce qui en fait un espace complet. On peut donc espérer appliquer le théorème du point fixe de Banach à cette fonction. Pour cela, il faudrait montrer que Φ est contractante.

Existence : la preuve par le théorème du point fixe

Comme on ne possède aucune information globale sur f , il se peut que $\|f(s, y)\|$ soit très grand quand $\|y\|$ est grand et il y a donc aucune chance que nous arrivions à montrer que Φ est contractante sur E .

On va donc essayer d'appliquer le théorème sur le sous-espace fermé $F = \mathcal{C}^0(J, \bar{B}(y_0, R))$ de E (qui est donc bien complet). Pour cela, on peut jouer sur les paramètres α et R pour faire en sorte que $\Phi(F) \subset F$ et que Φ soit contractante.

- Fixons une valeur $\alpha_0 > 0$ et un nombre $R_0 > 0$ tels que le compact $K_0 = [t_0 - \alpha_0, t_0 + \alpha_0] \times \bar{B}(y_0, R_0)$ soit inclus dans l'ouvert U sur lequel (3) est vraie.
- On note maintenant $M = \sup_{[t_0 - \alpha_0, t_0 + \alpha_0] \times \bar{B}(y_0, R_0)} \|F\|$. Ainsi, pour toute fonction $y \in \mathcal{C}^0([t_0 - \alpha_0, t_0 + \alpha_0], \bar{B}(y_0, R_0))$ on a :

$$\|\Phi(y)(t) - y_0\| \leq \left\| \int_{t_0}^t F(s, y(s)) ds \right\| \leq |t - t_0| M.$$

Existence : la preuve par le théorème du point fixe

Si on veut s'assurer que $\Phi(y)(t)$ reste dans la boule $\bar{B}(y_0, R_0)$, il faut se restreindre à un intervalle $[t_0 - \alpha, t_0 + \alpha]$ avec $0 < \alpha \leq \alpha_0$ choisi pour que

$$\alpha M \leq R_0$$

Ainsi, l'espace $F_\alpha = \mathcal{C}^0([t_0 - \alpha, t_0 + \alpha], \bar{B}(y_0, R_0))$ est laissé fixe par Φ dès que (I.4) est vérifiée.

- Essayons maintenant d'étudier le caractère contractant de Φ sur un tel espace. Soient $y, z \in F_\alpha$, on a

$$\begin{aligned} \|\Phi(y)(t) - \Phi(z)(t)\| &\leq \left| \int_{t_0}^t \|F(s, y(s)) - F(s, z(s))\| ds \right| \\ &\leq C_{t_0, y_0} |t - t_0| \|y - z\|_\infty \end{aligned}$$

et donc

$$\|\Phi(y) - \Phi(z)\|_\infty \leq C_{t_0, y_0} \alpha \|y - z\|_\infty.$$

Existence : la preuve par le théorème du point fixe

Ainsi, Φ sera contractante dès que $\alpha C_{t_0, y_0} < 1$.

- En conclusion, on va choisir $0 < \alpha \leq \alpha_0$ telle que

$\alpha < \min\left(\frac{R_0}{M}, \frac{1}{C_{t_0, y_0}}\right)$. La fonction Φ laisse alors invariant le sous-espace fermé $F_\alpha \subset E$ et elle est contractante dans cet espace. D'après le théorème du point fixe de Banach, il existe donc une unique solution $y \in F_\alpha$ à l'équation (4) et ainsi $([t_0 - \alpha, t_0 + \alpha], y)$ est une solution du problème de Cauchy considéré. C'est également l'unique solution sur cet intervalle qui prend ses valeurs dans la boule $\bar{B}(y_0, R_0)$.

Remarque

La méthode de point fixe ne permet pas réellement, en général, le calcul effectif des solutions (ou d'une approximation) des équations différentielles.

Introduction

Nous allons maintenant définir un nombre de méthodes d'approximation numérique, que nous appellerons aussi des schémas numériques. Nous aborderons ensuite l'étude mathématique de ces schémas et leur convergence vers la solution du problème de Cauchy de départ. Cette étude portera sur deux notions essentielles, la stabilité du schéma et sa consistance. On parlera aussi d'estimation d'erreur et l'on appliquera tout cela à plusieurs grandes familles de schémas numériques.

On se donne un problème de Cauchy générique de la forme :

$$\begin{cases} y'(t) = f(t, y(t)), & \forall t \in I = [0, T] \\ y(0) = y_0 \end{cases} \quad (5)$$

que l'on supposera satisfaire de bonnes hypothèses assurant existence, unicité et régularité de la solution. Dans l'hypothèse réaliste où il est impossible d'en donner une solution analytique, si l'on veut disposer d'informations quantitatives sur la solution, il faut donc définir des procédés d'approximation effectivement calculables (à la main dans le passé et sur ordinateur aujourd'hui). Cela implique en particulier que de tels procédés ne fassent intervenir qu'un nombre fini d'inconnues scalaires, alors qu'une fonction fait naturellement intervenir une infinité non dénombrable de valeurs scalaires.

- On commence donc par discrétiser, c'est-à-dire remplacer du continu par du discret, l'intervalle $I = [0, T]$ en y plaçant $N + 1$ points $t_n = nh$, $n = 0, \dots, N$, appelés points de discrétisation, uniformément espacés de $h = T/N$ (h est appelé pas de la discrétisation), pour un entier $N > 0$ donné. En particulier $t_0 = 0$ est l'instant initial et $t_N = T$ l'instant final.
- L'approximation numérique du problème de Cauchy consiste à construire une suite indexée par $N \in \mathbb{N}^*$ de valeurs $y_0^N, \dots, y_N^N$ censées approcher les valeurs exactes de la solution aux points de discrétisation, $y(t_0), \dots, y(t_N)$, en un sens que l'on précisera plus loin. Comme $y(t_0) = y_0$ est connu, on prendra (presque) toujours $y_0^N = y_0$ et, même si cela peut créer de l'ambiguïté, on notera généralement y_n^N par y_n (mais il faut garder à l'esprit la dépendance par rapport à N).
- Un schéma numérique est la donnée d'une telle construction, nous en verrons de nombreux exemples.

Les schémas qu'on va construire permettent de calculer (explicitement ou implicitement) y_{n+1} à partir de $y_n, y_{n-1}, \dots, y_{n-k}$ et il est donc possible de calculer successivement $y_1, y_2, \dots$, en partant de y_0 par une formule de récurrence de la forme

$$\begin{cases} y_{n+1} = \Phi(y_{n+1}, y_n, y_{n-1}, \dots, y_{n-k}), \forall n \in \mathbb{N} \\ y_0 = y(0). \end{cases} \quad (6)$$

Definition

- Une méthode est dite explicite si la valeur y_{n+1} peut être calculée directement à l'aide des valeurs précédentes $y_k, k \leq n$ (ou d'une partie d'entre elles). Une méthode est dite implicite si y_{n+1} n'est définie que par une relation implicite faisant intervenir la fonction f .
- Une méthode numérique pour l'approximation du problème de Cauchy (5) est dite à un pas si pour tout $n \in \mathbb{N}$, y_{n+1} ne dépend que de y_n et éventuellement de lui-même. Autrement, on dit que le schéma est une méthode multi-pas (ou à pas multiples).

Une manière de construire un schéma numérique utilise les techniques d'intégration numérique ou de quadrature. En intégrant l'EDO (1) sur un intervalle $[t_n, t_{n+1}]$, on obtient

$$y(t_{n+1}) - y(t_n) = \int_{t_n}^{t_{n+1}} y'(t) dt = \int_{t_n}^{t_{n+1}} f(t, y(t)) dt. \quad (7)$$

Ainsi, en introduisant une subdivision

$$0 = t_0 < t_1 < t_2 < \cdots < t_N = T$$

notre problème implique que

$$\begin{cases} y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} f(t, y(t)) dt, & \forall 0 \leq n \leq N-1 \\ y(0) = y_0 \end{cases} \quad (8)$$

On verra que tout choix d'une méthode de quadrature pour approcher

$$I_n = \int_{t_n}^{t_{n+1}} f(t, y(t)) dt$$

donne une méthode numérique de résolution des équations différentielles.

→ On procédera ici aussi en deux temps : approximation de l'intégrale par une formule de quadrature faisant intervenir les valeurs exactes en des points de discrétisation (si possible), puis remplacement de toutes les valeurs exactes par des valeurs approchées potentielles.

Schéma d'Euler progressif

La méthode d'Euler progressif (ou schéma d'Euler explicite) correspond à une approximation de I_n par la méthode des rectangles à gauche. On obtient donc :

$$I_n \approx hf(t_n, y(t_n)).$$

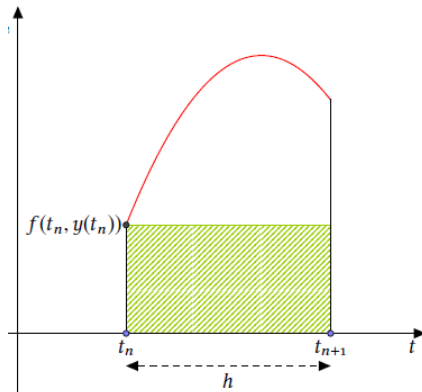


Schéma d'Euler progressif

En remplaçant maintenant dans (8) et dans l'approximation précédente les valeurs exactes $y(t_n)$ par des valeurs que l'on espère approchées y_n , il vient :

$$\begin{cases} y_{n+1} = y_n + hf(t_n, y_n), & \forall 0 \leq n \leq N-1 \\ y_0 = y(0). \end{cases} \quad (9)$$

Remarque

*Il s'agit d'un **schéma à 1 pas explicite** car il permet d'expliciter y_{n+1} en fonction de y_n .*

Schéma d'Euler progressif : Mise en œuvre avec Python

La méthode d'Euler progressif est implémentée dans le code suivant :

```
####Schéma d'Euler progressif:
def euler_progressif(f,tt,y0):
    h = tt[1]-tt[0]
    yy = [y0]
    for i in range(len(tt)-1):
        yy.append(yy[i]+h*f(tt[i],yy[i]))
    return yy
```

- Les nœuds d'intégration $[t_0, t_1, \dots, t_N]$ sont contenus dans le vecteur tt .
- Les valeurs $[y_0, y_1, \dots, y_N]$ sont contenus dans le vecteur yy .
- f est une fonction python qui contient la fonction mathématique $f(t, y(t))$ dépendant des variables t et y .

Remarque

Comme pour tous les algorithmes de ce cours, nous ne ferons que ré-implémenter des fonctions déjà existantes en Python.

Schéma d'Euler progressif : Mise en œuvre avec Python

En utilisant le code ci-dessus on résout le problème de Cauchy : trouver la fonction $y : I \subset \mathbb{R} \rightarrow \mathbb{R}$ définie sur l'intervalle $I = [0, 3]$ telle que

$$\begin{cases} y'(t) = y(t), & \forall t \in I \\ y(0) = 1. \end{cases} \quad (10)$$

On compare la solution approchée obtenue avec la méthode d'Euler progressive avec la solution exacte qui est la fonction $y(t) = \exp(t)$. Pour bien visualiser les erreurs, on considère seulement 8 sous-intervalles.

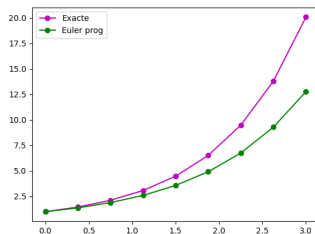


Schéma d'Euler progressif : Mise en œuvre avec Python

```

1  ### L'importation
2  import numpy as np
3  import matplotlib.pyplot as plt
4  ####Schéma d'Euler progressif:
5  def euler_progressif(f,tt,y0):
6      h = tt[1]-tt[0]
7      yy = [y0]
8      for i in range(len(tt)-1):
9          yy.append(yy[i]+h*f(tt[i],yy[i]))
10     return yy
11  # INITIALISATION
12  N = 8
13  t0 = 0
14  y0 = 1
15  tfinal = 3
16  def f(t,y):
17      return y
18  def sol_exacte(t):
19      return np.exp(t)
20  # CALCUL
21  h = (tfinal-t0)/N
22  tt = [t0+i*h for i in range(N+1) ]
23  y_exacte = [sol_exacte(t) for t in tt]
24  y_app = euler_progressif(f,tt,y0)
25  plt.plot(tt,y_exacte,'m-o',label='Exacte')
26  plt.plot(tt,y_app,'g-o',label='Euler prog')
27  plt.legend()
28  plt.show()

```

Schéma d'Euler rétrograde

Le même procédé appliqué avec la méthode des rectangles à droite, donne

$$I_n \approx hf(t_{n+1}, y(t_{n+1})).$$

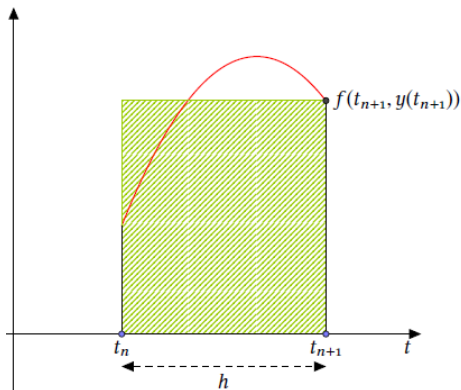


Schéma d'Euler rétrograde

On obtient le schéma dit d'Euler rétrograde (ou implicite) :

$$\begin{cases} y_{n+1} = y_n + hf(t_{n+1}, y_{n+1}), & \forall 0 \leq n \leq N-1 \\ y_0 = y(0). \end{cases} \quad (11)$$

Le schéma d'Euler rétrograde est un **schéma à 1 pas implicite** car il ne permet pas d'expliciter directement y_{n+1} en fonction de y_n lorsque la fonction f n'est pas triviale. En effet, on a dans ce cas :

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

Si nous voulons calculer y_{n+1} , nous définissons la fonction

$$g(x) = x - hf(t_{n+1}, x) - y_n$$

et nous cherchons un zéro de $g(x)$ en prenant par exemple une méthode de Newton.

Schéma d'Euler rétrograde

Ainsi nous pouvons poser $x_0 = y_n$ et $x_{m+1} = x_m - \frac{g(x_m)}{g'(x_m)}$, $m = 0, 1, \dots$

Puisque $g'(x) = 1 - h \frac{\partial f}{\partial x}(t_{n+1}, x)$, nous obtenons donc dans ce cas le schéma :

$$\begin{aligned} x_0 &= y_n, \\ x_{m+1} &= x_m - \frac{x_m - hf(t_{n+1}, x_m) - y_n}{1 - h \frac{\partial f}{\partial x}(t_{n+1}, x_m)}, \quad m = 0, 1, \dots \end{aligned}$$

Nous savons que

$$\lim_{m \rightarrow \infty} x_m = y_{n+1}$$

pour autant que f soit suffisamment régulière et que x_0 soit suffisamment proche de y_{n+1} , ce qui est le cas si le pas h est suffisamment petit.

Schéma d'Euler rétrograde : Mise en œuvre avec Python

La méthode d'Euler rétrograde est implémentée dans le code suivant :

```
####Schéma d'Euler rétrograde:
def euler_retrograde(f,tt,y0):
    h = tt[1]-tt[0]
    yy = [y0]
    for i in range(len(tt)-1):
        temp = fsolve( lambda x: x-h*f(tt[i+1],x)-yy[i] , yy[i] )
        yy.append(temp[0])
    return yy
```

Remarque

la fonction fsolve du module scipy.optimize requiert deux paramètres : une fonction et un point de départ. Elle renvoie deux valeurs : le zéro approché et l'estimation d'erreur.

Schéma d'Euler rétrograde : Mise en œuvre avec Python

En utilisant le code ci-dessus on résout le problème de Cauchy : trouver la fonction $y : I \subset \mathbb{R} \rightarrow \mathbb{R}$ définie sur l'intervalle $I = [0, 1]$ telle que

$$\begin{cases} y'(t) = 2ty(t), & \forall t \in I \\ y(0) = 1. \end{cases} \quad (12)$$

On compare la solution approchée obtenue avec la méthode d'Euler rétrograde avec la solution exacte qui est la fonction $y(t) = \exp(t^2)$.

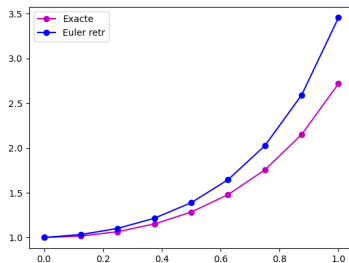


Schéma d'Euler rétrograde : Mise en œuvre avec Python

```

1  ### L'importation
2  import numpy as np
3  import matplotlib.pyplot as plt
4  from scipy.optimize import fsolve
5  #####Schéma d'Euler rétrograde:
6  def euler_retrograde(f,tt,y0):
7      h = tt[1]-tt[0]
8      yy = [y0]
9      for i in range(len(tt)-1):
10         temp = fsolve( lambda x: x-h*f(tt[i+1],x)-yy[i] , yy[i] )
11         yy.append(temp[0])
12     return yy
13  # INITIALISATION
14  N = 8
15  t0 = 0
16  y0 = 1
17  tfinal = 1
18  def f(t,y):
19      return 2*t*y
20  def sol_exacte(t):
21      return np.exp(t**2)
22  # CALCUL
23  h = (tfinal-t0)/N
24  tt = [t0+i*h for i in range(N+1) ]
25  y_exacte = [sol_exacte(t) for t in tt]
26  y_app = euler_retrograde(f,tt,y0)
27  plt.plot(tt,y_exacte,'m-o',label='Exacte')
28  plt.plot(tt,y_app,'b-o',label='Euler retr')
29  plt.legend()
30  plt.show()

```

Schéma d'Euler modifié

Si l'on approche I_n par la méthode du point milieu, on trouve l'approximation :

$$I_n \approx hf\left(t_n + \frac{h}{2}, y\left(t_n + \frac{h}{2}\right)\right).$$

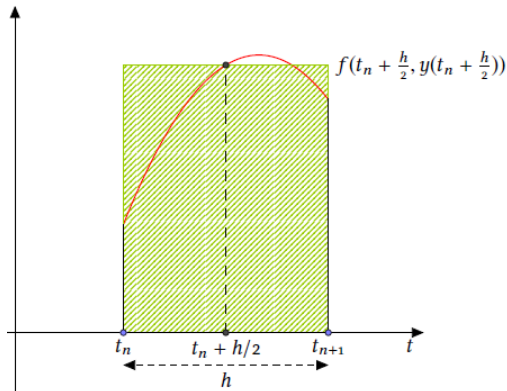


Schéma d'Euler modifié

Cela ne permet pas de construire directement un schéma numérique suivant la recette habituelle, puisque la valeur $y(t_n + \frac{h}{2})$ ne correspond pas à un point de discrétisation. Néanmoins, on peut reprendre pour cette valeur intermédiaire l'idée du schéma d'Euler et dire que :

$$y(t_n + \frac{h}{2}) \approx y(t_n) + \frac{h}{2}y'(t_n) = y(t_n) + \frac{h}{2}f(t_n, y(t_n)).$$

On obtient ainsi un schéma appelé schéma d'Euler modifié (ou schéma du point milieu) :

$$\begin{cases} y_{n+1} = y_n + hf(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)), & \forall 0 \leq n \leq N-1 \\ y_0 = y(0), \end{cases} \quad (13)$$

qui est un **schéma à 1 pas explicite**.

Schéma d'Euler modifié

Remarque

En tant que méthode de quadrature, la méthode du point milieu, qui est aussi une méthode des rectangles, est nettement plus précise que les méthodes des rectangles à gauche et à droite quand $h \rightarrow 0$. On peut espérer que le schéma numérique que l'on en tire soit plus performant que les schémas d'Euler, ce qui en fait se révélera bien être le cas.

Schéma du trapèze ou de Crank-Nicolson

En approchant I_n par la méthode des trapèzes, on obtient l'approximation :

$$I_n \approx \frac{h}{2} \left(f(t_n, y(t_n)) + f(t_{n+1}, y(t_{n+1})) \right).$$

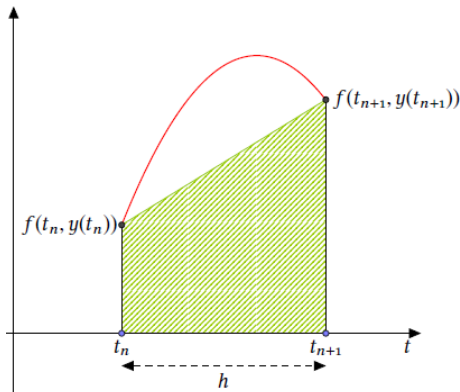


Schéma du trapèze ou de Crank-Nicolson

On obtient ainsi un schéma appelé schéma du trapèze ou de Crank-Nicolson

$$\begin{cases} y_{n+1} = y_n + \frac{h}{2} \left(f(t_n, y_n) + f(t_{n+1}, y_{n+1}) \right), & \forall 0 \leq n \leq N-1 \\ y_0 = y(0). \end{cases} \quad (14)$$

Il s'agit à nouveau d'un **schéma à 1 pas implicite** car il ne permet pas d'expliciter directement y_{n+1} en fonction de y_n lorsque la fonction f n'est pas triviale. En fait, ce schéma fait la moyenne des schémas d'Euler progressif et rétrograde.

Schéma de Heun

Pour éviter le calcul implicite de y_{n+1} dans le schéma du trapèze, nous pouvons utiliser une prédiction d'Euler progressive et remplacer le y_{n+1} dans le terme $f(t_{n+1}, y_{n+1})$ par $y_n + hf(t_n, y_n)$. Nous avons construit ainsi un nouveau schéma appelé schéma de Heun. Plus précisément, la méthode de Heun s'écrit :

$$\begin{cases} y_{n+1} = y_n + \frac{h}{2} \left(f(t_n, y_n) + f(t_{n+1}, y_n + hf(t_n, y_n)) \right), & \forall 0 \leq n \leq N-1 \\ y_0 = y(0). \end{cases} \quad (15)$$

Il s'agit à nouveau d'un **schéma à 1 pas explicite**.

Stabilité absolue

- Dans tout ce qui précède, on a considéré la résolution du problème de CAUCHY sur des intervalles bornés. Dans ce cadre, le nombre N de sous-intervalles ne tend vers l'infini que quand h tend vers zéro.
- Cependant, il existe de nombreuses situations dans lesquelles le problème de Cauchy doit être intégré sur des intervalles en temps très grands ou même infini. Dans ce cas, même pour h fixé, N tend vers l'infini. On s'intéresse donc à des méthodes capables d'approcher la solution pour des intervalles en temps arbitrairement grands, même pour des pas de temps h «assez grands».

Definition (A-stabilité)

On considère un problème de CAUCHY (5) dont la solution exacte vérifie $\lim_{t \rightarrow +\infty} y(t) = 0$. Soit h fixé et considérons la limite $T \rightarrow +\infty$ (ainsi $N \rightarrow +\infty$). On dit que la méthode est A-stable si $\lim_{n \rightarrow \infty} y_n = 0$.

Étude de la A-stabilité des schémas d'Euler

Étudions la A-stabilité des schémas d'Euler introduits ci-dessus :

Exemple

Soit $\beta > 0$ un nombre réel positif et considérons le problème de Cauchy

$$\begin{cases} y'(t) = -\beta y(t), & \forall t > 0 \\ y(0) = y_0 \end{cases} \quad (16)$$

où $y_0 \neq 0$ est une valeur donnée. Sa solution est $y(t) = \exp(-\beta t)y_0$ et $\lim_{t \rightarrow +\infty} y(t) = 0$.

Étude de la A-stabilité des schémas d'Euler

- Le schéma d'Euler progressif est

$$y_{n+1} = (1 - \beta h)y_n, \quad n = 0, 1, 2, \dots$$

et par suite

$$y_n = (1 - \beta h)^n y_0, \quad n = 0, 1, 2, \dots \quad (17)$$

Par conséquent, $\lim_{n \rightarrow +\infty} y_n = 0$ si et seulement si

$$|1 - \beta h| < 1,$$

ce qui a pour effet de limiter h à

$$h < \frac{2}{\beta}. \quad (18)$$

La condition (18) est appelée **condition de stabilité**.

Étude de la A-stabilité des schémas d'Euler

- Le schéma d'Euler rétrograde devient dans le cadre de notre exemple :

$$(1 + \beta h)y_{n+1} = y_n, \quad n = 0, 1, 2, \dots$$

et par suite :

$$y_n = \left(\frac{1}{1 + \beta h} \right)^n y_0, \quad n = 0, 1, 2, \dots$$

Dans ce cas, nous voyons que pour tout $h > 0$, nous avons

$$\lim_{n \rightarrow \infty} y_n = 0.$$

Le schéma d'Euler rétrograde est donc toujours stable, sans condition sur h .

Euler progressif vs Euler rétrograde

Remarque

A première vue, il semble que le schéma d'Euler progressif soit préférable au schéma d'Euler rétrograde puisque ce dernier n'est pas explicite. Cependant, la méthode d'Euler implicite est inconditionnellement A-stable. C'est aussi le cas de nombreuses autres méthodes implicites. Cette propriété rend les méthodes implicites attractives, bien qu'elles soient plus coûteuses que les méthodes explicites.

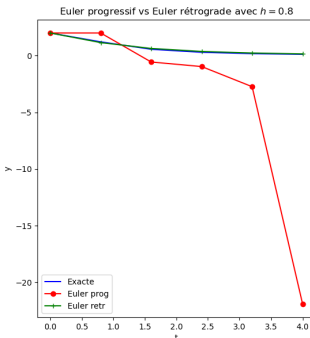
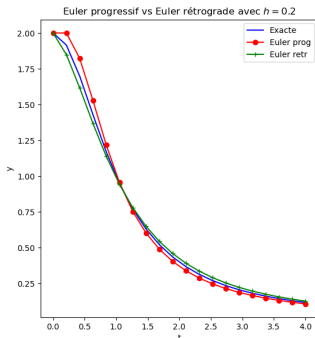
Exemple

Considérons le problème

$$\begin{cases} y'(t) = -ty^2(t), & \forall t > 0, \\ y(0) = 2, \end{cases} \quad (19)$$

dont la solution exacte est $y(t) = \frac{2}{1+t^2}$.

Euler progressif vs Euler rétrograde



La figure ci-dessus donne une comparaison entre la solution exacte et celles obtenues par les méthodes d'Euler progressive et rétrograde. En fait, le schéma d'Euler progressif est plus facile à utiliser, mais il n'est pas toujours stable comme on peut le voir sur cet exemple avec le pas $h = 0.8$.

Schémas explicites à un pas

Il s'agit des schémas qui s'écrivent sous la forme générique :

$$\begin{cases} y_{n+1} = y_n + hF(t_n, y_n, h), \\ y_0 = y^h. \end{cases} \quad (20)$$

où F est une fonction définie et continue sur $[0, T] \times \mathbb{R}^m \times [0, 1]$ à valeurs dans $\mathbb{R}^m$, que l'on sait écrire explicitement. On remarque qu'on autorise ici une erreur sur la donnée initiale, (c'est-à-dire $y^h \approx y(0)$).

Definition

Le schéma (20) est dit convergent si, pour toute donnée initiale $y(0)$ du problème continu,

$$\sup_{0 \leq n \leq N} \|y_n - y(t_n)\| = 0 \text{ lorsque } h \rightarrow 0 \text{ et } y^h \rightarrow y(0). \quad (21)$$

Schémas explicites à un pas

Remarque

- *Si la méthode est convergente, en choisissant h assez petit, et y^h assez voisin de $y(0)$, on obtient une bonne approximation de $y(t_n) \approx y_n$.*
- *Il semblerait plus naturel de poser directement $y^h = y(0)$ dans notre schéma. Cependant, dans la pratique, si $y(0)$ est réel, il ne pourra pas toujours être pris en compte de façon exacte à cause des arrondis. Une analyse correcte nécessite donc de supposer $y(0) \neq y^h$.*
- *La notion de convergence que l'on vient d'introduire est bien entendue valable pour des schémas généraux, et pas seulement les schémas à un pas explicites.*

La convergence d'un schéma est liée à deux notions indépendantes l'une de l'autre : la stabilité et la consistance.

Stabilité par rapport aux erreurs

Definition

Le schéma (20) est stable s'il existe une constante C indépendante de N telle que, pour toute suite de vecteurs $(\eta_n)_{0 \leq n \leq N}$, les suites $(y_n)_{0 \leq n \leq N}$ et $(z_n)_{0 \leq n \leq N}$ de $\mathbb{R}^m$ définies respectivement par

$$y_0 \in \mathbb{R}^m \text{ et } y_{n+1} = y_n + hF(t_n, y_n, h) \text{ pour } 0 \leq n \leq N-1$$

et

$$z_0 = y_0 + \eta_0 \text{ et } z_{n+1} = z_n + hF(t_n, z_n, h) + \eta_{n+1} \text{ pour } 0 \leq n \leq N-1$$

sont telles que

$$\max_{0 \leq n \leq N} \|z_n - y_n\| \leq C \sum_{n=0}^N \|\eta_n\|. \quad (22)$$

Proposition

S'il existe une constante $M > 0$ telle que pour tous y et z dans $\mathbb{R}^m$, $h \in [0, 1]$ et $t \in [0, T]$, on ait

$$\|F(t, y, h) - F(t, z, h)\| \leq M\|y - z\|, \quad (23)$$

alors le schéma (20) est stable.

Pour démontrer ce résultat, on va utiliser le lemme de Gronwall discret.

Lemme (Gronwall discret, version 1)

Soit $(u_n)_{n \geq 0}$ une suite de réels. On suppose qu'il existe deux réels λ et μ , avec $\lambda \neq 0$, $1 + \lambda \geq 0$, et $u_0 + \frac{\mu}{\lambda} \geq 0$, tels que :

$$\forall n \geq 0, u_{n+1} \leq (1 + \lambda)u_n + \mu.$$

Alors pour tout $n \geq 1$, on a : $u_n + \frac{\mu}{\lambda} \leq (u_0 + \frac{\mu}{\lambda})e^{\lambda n}$.

Démonstration.

La suite $v_n = u_n + \frac{\mu}{\lambda}$ vérifie $v_{n+1} \leq (1 + \lambda)v_n$. Par récurrence sur n et en utilisant le fait que $1 + \lambda \geq 0$, on montre que $v_n \leq (1 + \lambda)^n v_0$. Et comme $1 + \lambda \leq e^\lambda$ et $v_0 \geq 0$, on obtient le résultat annoncé. $\square$

Lemme (Gronwall discret, version 2)

Soient $(u_n)_{n \geq 0}$ et $(\mu_n)_{n \geq 0}$ deux suites de réels avec $u_0 \geq 0$ et $\mu_n \geq 0$. On suppose qu'il existe un réel λ avec $\lambda \neq 0$ et $1 + \lambda \geq 0$, tels que :

$\forall n \geq 0, u_{n+1} \leq (1 + \lambda)u_n + \mu_n$. Alors pour tout $n \geq 1$, on a :

$$u_n \leq e^{\lambda n} u_0 + \sum_{k=0}^{n-1} e^{\lambda k} \mu_{n-k-1}.$$

Démonstration.

Par récurrence sur n . $\square$

Consistance

La suite y_n est construite de façon à vérifier l'égalité

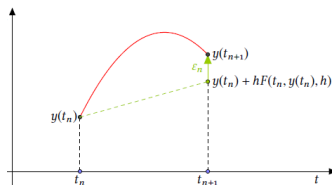
$$y_{n+1} - y_n - hF(t_n, y_n, h) = 0.$$

Definition

On appelle erreur de consistance (ou erreur de discrétisation locale) du schéma (20) la quantité $\epsilon_n \in \mathbb{R}^m$ définie par

$$\epsilon_n = y(t_{n+1}) - y(t_n) - hF(t_n, y(t_n), h),$$

où y est une solution de l'EDO.



Consistance

Definition

Le schéma (20) est dit consistant avec l'EDO (1) si pour toute solution y de (1), on a :

$$\lim_{h \rightarrow 0} \sum_{n=0}^{N-1} \|\epsilon_n\| = 0.$$

Explication : Un schéma est donc consistant si la somme des erreurs de consistance sur tous les instants de discrétisation tend vers 0 avec h , ce pour toute solution y de l'EDO.

Proposition

On suppose l'application F continue sur $[0, T] \times \mathbb{R}^m \times [0, 1]$. Le schéma (20) est consistant si et seulement si, pour tout $(t, y) \in I \times \mathbb{R}^m$, on a :

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

Convergence

Théorème

Si le schéma (20) est stable et consistant alors il est convergent.

Remarque

Etant donné un schéma stable, donné a priori par une fonction continue F , si l'on pose $g(t, y) = F(t, y, 0)$, on voit que ce schéma est automatiquement consistant avec l'EDO $y' = g(t, y)$ et converge donc vers les solutions du problème de Cauchy correspondant. On peut donc dire grossièrement qu'un schéma stable converge, mais qu'il ne converge vers ce que l'on veut que s'il est consistant. Sinon, on est entrain d'approcher une autre EDO.

Ordre d'un schéma, estimation d'erreur

Savoir qu'un schéma numérique est convergent, c'est bien, mais on aimerait savoir aussi à quelle vitesse cette convergence a lieu. En d'autres termes, on souhaite quantifier la qualité de l'approximation. On introduit d'abord une définition qui raffine celle de la consistance.

Definition

Le schéma (20) est dit d'ordre au moins $p \in \mathbb{N}^*$ si, pour toute solution y de l'EDO (1), il existe une constante C indépendante de h telle que

$$\forall N, \sum_{n=0}^{N-1} \|\epsilon_n\| \leq Ch^p.$$

Il est d'ordre p s'il n'est en outre pas d'ordre au moins $p + 1$.

Le résultat qui suit est un simple raffinement du théorème de convergence dans le cas d'un schéma d'ordre p .

Théorème

On suppose que le schéma (20) est stable et d'ordre $p \geq 1$ et qu'il existe une constante $C > 0$ telle que $\|y(0) - y_0\| \leq Ch^p$. Alors il existe $\tilde{C} > 0$ telle qu'on ait l'estimation d'erreur suivante

$$\max_{0 \leq n \leq N} \|y(t_n) - y_n\| \leq \tilde{C}h^p.$$

On a la condition suffisante suivante qui est celle utilisée dans la pratique.

Proposition

Si pour toute solution y , il existe une constante C' indépendante de h telle que l'erreur de consistance du schéma vérifie

$$\forall n \leq N, \|\epsilon_n\| \leq C'h^{p+1},$$

alors le schéma est d'ordre au moins p . Si l'on a de plus $\|\epsilon_n\| \geq C''h^{p+1}$ pour tout $n \leq N$, avec $C'' > 0$ indépendante de h pour au moins une solution y de l'EDO, alors le schéma est d'ordre p .

Exercice :

- ① On suppose la fonction f est globalement lipschitzienne par rapport à y . Montrer que le schéma d'Euler progressif et le schéma d'Euler modifié sont stables.
- ② On suppose la fonction f est continue par rapport à (t, y) . Montrer que le schéma d'Euler progressif et le schéma d'Euler modifié sont consistants.
- ③ En déduire que le schéma d'Euler progressif et le schéma d'Euler modifié sont convergents (si f est globalement lipschitzienne par rapport à y et continue par rapport à (t, y)).
- ④ Montrer que le schéma d'Euler progressif est d'ordre un dès que les solutions de l'EDO (1) sont de classe C^2 .
- ⑤ Montrer que le schéma d'Euler modifié est d'ordre deux dès que les solutions de l'EDO (1) sont de classe C^3 .

Remarque : Notons que l'ordre est obtenu sous une hypothèse de régularité de la solution. Si l'on utilise la méthode d'Euler modifiée sur une EDO dont les solutions ne sont pas de classe C^3 , il ne faut pas espérer voir un gain d'estimation d'erreur par rapport à la méthode d'Euler tout court.

Schémas implicites à un pas

On a introduit déjà quelques schémas implicites (Euler rétrograde, Crank-Nicolson). Avant de se demander s'ils convergent ou non et à quoi ils peuvent bien servir, il faut d'abord s'interroger sur leur caractère bien défini.

Considérons donc un schéma implicite générique à un pas de la forme :

$$\begin{cases} y_{n+1} = y_n + h\Phi(t_{n+1}, y_{n+1}, t_n, y_n, h), \\ y_0 = y^h. \end{cases} \quad (24)$$

supposé résoudre le problème de Cauchy pour l'EDO (1). À chaque pas de temps, il s'agit de résoudre l'équation en général non linéaire

$$z = \varphi(z), \text{ avec } \varphi(z) = y_n + h\Phi(t_{n+1}, z, t_n, y_n, h). \quad (25)$$

Un tel z est appelé un point fixe de φ , qui est une application de $\mathbb{R}^m$ dans $\mathbb{R}^m$. En fait, l'existence et l'unicité d'un point fixe sont assurées si la fonction φ est contractante.

Definition

Soit (E, d) un espace métrique et φ une application de E dans lui-même. On dit que φ est **contractante** s'il existe $k \in [0, 1[$ tel que pour tout $x_1, x_2 \in E$,

$$d(\varphi(x_1), \varphi(x_2)) \leq kd(x_1, x_2).$$

En effet, on a le théorème de point fixe de Picard ou de Banach suivant :

Théorème

*Soit (E, d) un espace métrique **complet** et φ une application contractante de E dans lui-même. Alors φ admet un point fixe unique x^* . De plus, pour tout $x_0 \in E$, la suite récurrente $(x_p)_{p \in \mathbb{N}}$ définie par $x_{p+1} = \varphi(x_p)$ pour tout $p \geq 0$ converge vers le point fixe x^* .*

Revenons au schéma implicite général à un pas, lequel est défini par la donnée d'une fonction $\Phi : [0, T] \times \mathbb{R}^m \times [0, T] \times \mathbb{R}^m \times [0, 1] \rightarrow \mathbb{R}^m$. Notons $\Phi(s, z, t, y, h)$ l'image d'un élément générique par cette fonction et φ l'application $z \mapsto y + h\Phi(s, z, t, y, h)$ pour s, t, y, h fixés.

Proposition

Si Φ est globalement lipschitzienne par rapport à z , alors il existe $h_0 > 0$ indépendant de (s, t, y) tel que l'application φ soit contractante pour tout $h \leq h_0$.

Comme $\mathbb{R}^m$ est complet pour la distance induite par n'importe quelle norme, on en déduit le

Corollaire

Le schéma implicite (24) est bien défini pour tout $h \leq h_0$.

Remarque

- Pour un schéma implicite, on ne prendra donc pas la variable h dans l'intervalle un peu arbitraire $[0, 1]$ comme précédemment, mais dans $[0, h_0]$ ce qui ne change essentiellement rien.
- Dans le cas du schéma d'Euler implicite, où $\Phi(s, z, t, y, h) = f(s, z)$, on voit que la condition est satisfaite dès que la fonction f est globalement lipschitzienne par rapport à sa deuxième variable. Par conséquent, le schéma d'Euler implicite est bien défini pour $h < \frac{1}{L}$, où L désigne la constante de Lipschitz de f . Bien sûr, si on ne connaît pas L , il est difficile de deviner ce que « h suffisamment petit» signifie quantitativement parlant.

Intéressons-nous maintenant à la convergence du schéma implicite vers la solution exacte quand le pas de discrétisation h tend vers 0. On se placera toujours dans l'hypothèse que h est suffisamment petit pour que le schéma soit bien défini, c'est-à-dire $hM < 1$ où M désigne la constante de Lipschitz de Φ par rapport à z .

- Le schéma (24) est stable s'il existe une constante C indépendante de N telle que, pour toute suite de vecteurs $(\eta_n)_{0 \leq n \leq N}$, les suites $(y_n)_{0 \leq n \leq N}$ et $(z_n)_{0 \leq n \leq N}$ de $\mathbb{R}^m$ définies respectivement par

$$y_0 \in \mathbb{R}^m \text{ et } y_{n+1} = y_n + h\Phi(t_{n+1}, y_{n+1}, t_n, y_n, h) \text{ pour } 0 \leq n \leq N-1$$

et

$$z_0 = y_0 + \eta_0 \text{ et } z_{n+1} = z_n + h\Phi(t_{n+1}, z_{n+1}, t_n, z_n, h) + \eta_{n+1}, 0 \leq n \leq N-1$$

sont telles que

$$\max_{0 \leq n \leq N} \|z_n - y_n\| \leq C \sum_{n=0}^N \|\eta_n\|. \quad (26)$$

- L'erreur de consistance du schéma (24) est la quantité :

$$\epsilon_n = y(t_{n+1}) - y(t_n) - h\Phi(t_{n+1}, y(t_{n+1}), t_n, y(t_n), h),$$

où y est une solution de l'EDO. Le schéma (24) est consistant si pour

toute solution y de (1), on a : $\lim_{h \rightarrow 0} \sum_{n=0}^{N-1} \|\epsilon_n\| = 0$.

On a par exactement la même démonstration que dans le cas explicite :

Théorème

Un schéma implicite (24) stable et consistant est convergent.

En termes de conditions suffisantes de stabilité et de convergence, on a des résultats également très semblables.

Proposition

Soit Φ une fonction globalement lipschitzienne de constante M par rapport à y et z et soit $0 < h_0 < \frac{1}{M}$. Alors pour tout $h \leq h_0$, le schéma (24) est stable. Si Φ est continue par rapport à l'ensemble de ses arguments et que $\Phi(t, y, t, y, 0) = f(t, y)$ pour tous (t, y) , alors le schéma est consistant.

Exemple :

- Dans le cas du schéma d'Euler implicite, $\Phi(s, z, t, y, h) = f(s, z)$, on a donc sous les hypothèses du théorème de Cauchy-Lipschitz global que $M = L$ et le schéma est bien défini et stable pour $h \leq h_0 < \frac{1}{L}$, et il est consistant. Il est donc convergent.
- Passons à la question d'ordre de la méthode d'Euler implicite. On suppose donc la fonction f de classe C^1 de telle sorte que toute solution y soit C^2 . On calcule l'erreur de consistance

$$\begin{aligned}\epsilon_n &= y(t_{n+1}) - y(t_n) - hf(t_{n+1}, y(t_{n+1})) \\ &= -\left(y(t_n) - y(t_{n+1}) - (-h)y'(t_{n+1})\right),\end{aligned}$$

d'où comme $t_n = t_{n+1} - h$, par l'inégalité de Taylor-Lagrange

$$\epsilon_n \leq \frac{h^2}{2} \max_{[0, T]} \|y''\|, \text{ et la méthode d'Euler implicite est donc d'ordre 1.}$$

On pourra à titre d'exercice traiter le cas du schéma de Crank-Nicolson qui se trouve être d'ordre 2.

Méthodes de Runge-Kutta : Principe général

L'idée est de calculer par récurrence les points (t_n, y_n) en utilisant des points intermédiaires $(t_{n,i}, y_{n,i})$ avec

$$t_{n,i} = t_n + c_i h, \quad 1 \leq i \leq q, c_i \in [0, 1].$$

A chacun de ces points on associe la pente correspondante

$$p_{n,i} = f(t_{n,i}, y_{n,i}).$$

Soit y une solution exacte de l'équation. On a :

$$\begin{aligned} y(t_{n,i}) &= y(t_n) + \int_{t_n}^{t_{n,i}} f(t, y(t)) dt \\ &= y(t_n) + h \int_0^{c_i} f(t_n + sh, y(t_n + sh)) ds \end{aligned}$$

grâce au changement de variable $t = t_n + sh$.

De même :

$$y(t_{n+1}) = y(t_n) + h \int_0^1 f(t_n + sh, y(t_n + sh)) ds.$$

On se donne alors pour chaque $i = 1, 2; \dots, q$ une méthode d'intégration approchée

$$(M_i) \quad \int_0^{c_i} g(t) dt \approx \sum_{1 \leq j < i} a_{i,j} g(c_j),$$

ces méthodes pouvant être a priori différentes. On se donne également une méthode d'intégration approchée sur $[0, 1]$:

$$(M) \quad \int_0^1 g(t) dt \approx \sum_{1 \leq j \leq q} b_j g(c_j),$$

En appliquant ces méthodes d'intégrations à $g(s) = f(t_n + sh, y(t_n + sh))$, il vient :

$$y(t_{n,i}) \approx y(t_n) + h \sum_{1 \leq j < i} a_{ij} f(t_{n,j}, y(t_{n,j})),$$

$$y(t_{n+1}) \approx y(t_n) + h \sum_{1 \leq j \leq q} b_j f(t_{n,j}, y(t_{n,j})).$$

La méthode Runge-Kutta correspondante est définie par l'algorithme

$$\left\{ \begin{array}{l} t_{n,i} = t_n + c_i h_n \\ y_{n,i} = y_n + h_n \sum_{1 \leq j < i} a_{ij} p_{n,j} \\ p_{n,i} = f(t_{n,i}, y_{n,i}) \\ t_{n+1} = t_n + h_n \\ y_{n+1} = y_n + h_n \sum_{1 \leq j \leq q} b_j p_{n,j} \end{array} \right. \quad 1 \leq i \leq q$$

On la représente conventionnellement par le tableau

(M_1)	c_1	0	0	$\dots$	0	0
(M_2)	c_2	a_{21}	0	$\dots$	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		0	0
(M_q)	c_q	a_{q1}	a_{q2}	$\dots$	a_{qq-1}	0
(M)		b_1	b_2	$\dots$	b_{q-1}	b_q

où les méthodes d'intégration approchées correspondent aux lignes. On pose par convention $a_{ij} = 0$ pour $j \geq i$.

Remarque

On supposera toujours que les méthodes d'intégrations (M_i) et (M) sont d'ordre au moins 0, c'est-à-dire :

$$c_i = \sum_{1 \leq j < i} a_{i,j} \text{ et } 1 = \sum_{1 \leq j \leq q} b_j.$$

En particulier, on aura toujours :

$$c_1 = 0, \quad t_{n,1} = t_n, \quad y_{n,1} = y_n, \quad p_{n,1} = f(t_n, y_n).$$

Exemples

- Pour $q = 1$, le seul choix possible est $\frac{0}{1} \mid \frac{0}{1}$

On ici $c_1 = 0$, $a_{11} = 0$, $b_1 = 1$. L'algorithme est donné par :

$$\begin{cases} p_{n,1} = f(t_n, y_n) \\ t_{n+1} = t_n + h_n \\ y_{n+1} = y_n + h_n p_{n,1} \end{cases}$$

Il s'agit de la méthode d'Euler.

- Pour $q = 2$, on considère les tableaux de la forme :

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \alpha & \alpha & 0 \\ \hline & 1 - \frac{1}{2\alpha} & \frac{1}{2\alpha} \end{array} \text{ où } \alpha \in]0, 1].$$

L'algorithme s'écrit ici :

$$\left\{ \begin{array}{l} p_{n,1} = f(t_n, y_n) \\ t_{n,2} = t_n + \alpha h_n \\ y_{n,2} = y_n + \alpha h_n p_{n,1} \\ p_{n,2} = f(t_{n,2}, y_{n,2}) \\ t_{n+1} = t_n + h_n \\ y_{n+1} = y_n + h_n \left(\left(1 - \frac{1}{2\alpha}\right) p_{n,1} + \frac{1}{2\alpha} p_{n,2} \right) \end{array} \right.$$

ou encore sous forme condensée :

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

- ❶ Pour $\alpha = \frac{1}{2}$, on retrouve la méthode du point milieu

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

qui est basé sur la méthode d'intégration du point milieu :

$$(M) \quad \int_0^1 g(t) dt \approx g\left(\frac{1}{2}\right).$$

- ❷ Pour $\alpha = 1$, on retrouve la méthode de Heun :

$$y_{n+1} = y_n + h_n \left(\frac{1}{2} f(t_n, y_n) + \frac{1}{2} f(t_{n+1}, y_n + h_n f(t_n, y_n)) \right),$$

qui repose sur la méthode d'intégration des trapèzes :

$$(M) \quad \int_0^1 g(t) dt \approx \frac{1}{2} (g(0) + g(1)).$$

Méthode de Runge-Kutta classique RK4

Il s'agit de la méthode définie par le tableau ($q = 4$) :

0	0	0	0	0
1	1			
$\frac{1}{2}$	$\frac{1}{2}$	0	0	0
$\frac{1}{2}$	2	1		
$\frac{1}{2}$	0	$\frac{1}{2}$	0	0
1	0	0	1	0
	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

Méthode de Runge-Kutta classique RK4

L'algorithme correspondant s'écrit :

$$\left\{ \begin{array}{l} p_{n,1} = f(t_n, y_n) \\ t_{n,2} = t_n + \frac{1}{2}h_n \\ y_{n,2} = y_n + \frac{1}{2}h_n p_{n,1} \\ p_{n,2} = f(t_{n,2}, y_{n,2}) \\ y_{n,3} = y_n + \frac{1}{2}h_n p_{n,2} \\ p_{n,3} = f(t_{n,2}, y_{n,3}) \quad \text{noter que } t_{n,3} = t_{n,2} \\ t_{n+1} = t_n + h_n \quad \text{noter que } t_{n,4} = t_{n+1} \\ y_{n,4} = y_n + h_n p_{n,3} \\ p_{n,4} = f(t_{n+1}, y_{n,4}) \\ y_{n+1} = y_n + h_n \left(\frac{1}{6}p_{n,1} + \frac{2}{6}p_{n,2} + \frac{2}{6}p_{n,3} + \frac{1}{6}p_{n,4} \right). \end{array} \right.$$

Méthode de Runge-Kutta classique RK4

Dans ce cas les méthodes d'intégrations (M_i) et (M) utilisées sont respectivement :

$$(M_2) \quad \int_0^{\frac{1}{2}} g(t) dt \approx \frac{1}{2}g(0) : \text{rectangle à gauche}$$

$$(M_3) \quad \int_0^{\frac{1}{2}} g(t) dt \approx \frac{1}{2}g\left(\frac{1}{2}\right) : \text{rectangle à droite}$$

$$(M_4) \quad \int_0^1 g(t) dt \approx g\left(\frac{1}{2}\right) : \text{point milieu}$$

$$(M) \quad \int_0^1 g(t) dt \approx \frac{1}{6}g(0) + \frac{2}{6}g\left(\frac{1}{2}\right) + \frac{2}{6}g\left(\frac{1}{2}\right) + \frac{1}{6}g(1) : \text{Simpson.}$$

Remarque

On peut montrer (avec pas mal de calculs...) que si $y \in C^4$ alors le schéma RK4 est consistant et convergent d'ordre 4.

Notes bibliographiques et remarques

- ① Les méthodes que nous avons présentées dans ce chapitre pour résoudre numériquement des systèmes différentiels d'ordre 1 sont des **méthodes à un pas** (Euler, Runge-Kutta). Ces méthodes permettent de calculer y_{n+1} à partir de y_n . Les **méthodes multipas** utilisent les valeurs $y_n, y_{n-1}, y_{n-2}, \dots$ pour calculer y_{n+1} . Citons par exemple les méthodes d'Adams, les méthodes de Nyström et de Milne-Simpson...

- J.P. Demailly, *Analyse numérique et équations différentielles (3ième édition)*, EDP sciences, (2006).

- ② Comme nous l'avons vu, la résolution numérique de certaines équations différentielles nécessite un pas anormalement petit. On dit alors que le problème est **raide** (*stiff* en anglais). Il existe un certain nombre de méthodes pour résoudre ce genre de problèmes, voir par exemple :

- E. Hairer, G. Wanner, *Solving Ordinary Differential Equations II : Stiff and Differential-Algebraic Problems*, Springer-Verlag Berlin Heidelberg (1996).

Bibliographie

- ✓ **K. Atkinson, W. Han, D.E. Stewart**, *Numerical Solution of Ordinary Differential Equations*, John Wiley & Sons (2011).
- ✓ **J.C. Butcher**, *Numerical methods for ordinary differential equations*, John Wiley & Sons, (2016).
- ✓ **M. Crouzeix et A. Mignot**, *Analyse numérique des équations différentielles*, Collection Mathématiques appliquées pour la maîtrise, Masson (1984).
- ✓ **M. Crouzeix et A. Mignot**, *Exercices d'analyse numérique des équations différentielles*, Masson (1986).
- ✓ **F. Filbet**, *Analyse numérique : algorithme et étude mathématique*, Dunod (2009).
- ✓ **A. Quarteroni, R. Sacco et F. Saleri**, *Méthodes Numériques : Algorithmes, analyse et applications*, Springer-Verlag Italia, Milano, (2007).

Merci pour votre attention