


Equations Différentielles II

STEP, MINES ParisTech*

22 octobre 2019 (#2f387a0)

Table des matières

Introduction	2
Méthodes à un pas	2
Principe	2
Exemples	2
Définition implicite de Φ	3
Analyse d'erreur	4
Erreur de troncature locale	5
Consistance	5
Condition suffisante	5
Consistance du schéma d'Euler explicite	5
Stabilité	6
Définition	6
Condition suffisante	6
Convergence	7
Définition	7
Théorème de Lax	7
Relaxation de la stabilité	7
Erreurs d'arrondi et pas optimal	7
Exercices	8
Consistance de schémas	8
Convergence de schémas	8
Corrections	8
Consistance de schémas	8
Convergence de schémas	8
Références	8

*Ce document est un des produits du projet  **boisgera/CDIS**, initié par la collaboration de (S)ébastien Boisgérault (CAOR), (T)homas Romary et (E)milie Chautru (GEOSCIENCES), (P)auline Bernard (CAS), avec la contribution de Gabriel Stoltz (Ecole des Ponts ParisTech, CERMICS). Il est mis à disposition selon les termes de la licence Creative Commons “attribution – pas d'utilisation commerciale – partage dans les mêmes conditions” 4.0 internationale.

Introduction

Méthodes à un pas

Principe

Pour approximer les solutions d'une équation différentielle sur un intervalle $[0, \bar{t}]$, les méthodes numériques à un pas se basent sur la représentation intégrale

$$x(t) = x_0 + \int_{t_0}^t f(s, x(s)) ds = x_0 + \sum_{j=0}^{J-1} \int_{t_j}^{t_{j+1}} f(s, x(s)) ds$$

où $t_0 < t_1 < \dots < t_J$ avec $t_J = \bar{t}$. L'idée est d'approximer les intégrales $\int_{t_j}^{t_{j+1}} f(s, x(s))$ sur des intervalles $[t_j, t_{j+1}]$ suffisamment petits.

Dans la suite, on note x^j l'approximation au temps t_j de la valeur exacte $x(t_j)$ et $\Delta t_j = t_{j+1} - t_j$ le j ème pas de temps. L'idée est de calculer récursivement

$$x^{j+1} = x^j + \Delta t_j \Phi_{\Delta t_j}(t_j, x^j)$$

où $\Phi_{\Delta t_j}(t_j, x^j)$ doit donc approximer

$$\frac{1}{t_{j+1} - t_j} \int_{t_j}^{t_{j+1}} f(s, x(s)) .$$

Les différentes méthodes de quadrature, i.e. d'approximation de l'intégrale, peuvent donc être mises à profit. La difficulté ici est que seule la valeur initiale $f(t_j, x(t_j))$ de f est connue (ou du moins estimée) à l'itération j , par $f(t_j, x^j)$. On distingue donc les méthodes *explicites* où $\Phi_{\Delta t_j}(t_j, x^j)$ est écrite directement explicitement en fonction de la valeur initiale x^j , et les méthodes *implicites* où cette expression n'est connue qu'implicitement et des étapes intermédiaires de calcul sont nécessaires.

Exemples

1. Méthodes explicites:

- Euler explicite: l'intégrale est approxmée par l'aire d'un rectangle déterminé par la valeur initiale de f à *gauche de l'intervalle*, i.e.

$$x^{j+1} = x^j + \Delta t_j f(t_j, x^j) .$$

- méthode de Heun : l'intégrale est approxmée par l'aire d'un trapèze déterminé par la valeur initiale de f et une approximation de sa valeur finale, i.e.,

$$x^{j+1} = x^j + \frac{\Delta t_j}{2} \left(f(t_j, x^j) + f(t_{j+1}, x^j + \Delta t_j f(t_j, x^j)) \right) .$$

— schéma de Runge–Kutta d’ordre 4:

$$\begin{cases} F_1 = f(t_j, x^j) \\ F_2 = f\left(t_j + \frac{\Delta t_j}{2}, x^j + \frac{\Delta t_j}{2} F_1\right) \\ F_3 = f\left(t_j + \frac{\Delta t_j}{2}, x^j + \frac{\Delta t_j}{2} F_2\right) \\ F_4 = f(t_j + \Delta t_j, x^j + \Delta t_j F_3), \end{cases}$$

et on pose

$$x^{j+1} = x^j + \Delta t \frac{F_1 + 2F_2 + 2F_3 + F_4}{6}.$$

2. Méthodes implicites:

— Euler implicite : l’intégrale est approximée par l’aire d’un rectangle déterminé par la valeur finale de f à droite de l’intervalle, i.e.

$$x^{j+1} = x^j + \Delta t_j f(t_{j+1}, x^{j+1}).$$

— méthode des trapèzes (ou Crank–Nicolson) : l’intégrale est approximée par l’aire du trapèze déterminé par les valeurs initiales et finales de f , i.e.

$$x^{j+1} = x^j + \frac{\Delta t_j}{2} \left(f(t_j, x^j) + f(t_{j+1}, x^{j+1}) \right).$$

— méthode du point milieu : l’intégrale est approximée par l’aire d’un rectangle déterminé par une approximation de la valeur de f au milieu de l’intervalle, i.e.

$$x^{j+1} = x^j + \Delta t_j f\left(\frac{t_j + t_{j+1}}{2}, \frac{x^j + x^{j+1}}{2}\right).$$

On peut bien sûr construire des méthodes plus compliquées et plus précises pour des méthodes de Runge–Kutta d’ordre supérieur (explicites ou implicites).

Définition implicite de Φ

Dans les schémas implicites, l’application $\Phi_{\Delta t}$ est définie de manière implicite. Par exemple, pour le schéma d’Euler, on a :

$$\Phi_{\Delta t_j}(t_j, x^j) = f\left(t_j + \Delta t_j, x^j + \Delta t_j \Phi_{\Delta t_j}(t_j, x^j)\right).$$

Il faut donc s’assurer que Φ est bien définie, c’est-à-dire qu’il existe bien x^{j+1} tel que

$$x^{j+1} = x^j + \Delta t_j f(t_{j+1}, x^{j+1}).$$

Pour cela, nous pouvons voir x^{j+1} comme le point fixe de l’application F_j définie par

$$F_j(x) = x^j + \Delta t_j f(t_{j+1}, x).$$

à x^j , Δt_j , t_{j+1} fixés. L'existence (et l'unicité) de ce point fixe peut alors être démontrée par le théorème de point fixe de Banach. Si $x \mapsto f(t_{j+1}, x)$ est Lipschitzienne, c'est-à-dire s'il existe L_j tel que

$$|f(t_{j+1}, x_a) - f(t_{j+1}, x_b)| \leq L_j |x_a - x_b| \quad \forall (x_a, x_b) \in \mathbb{R}^n \times \mathbb{R}^n ,$$

alors $F_j : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est contractante pour un pas de temps Δt_j suffisamment petit puisque

$$|F_j(x_a) - F_j(x_b)| \leq \Delta t_j L_j |x_a - x_b| .$$

Puisque \mathbb{R}^n est complet, on déduit par le théorème du point fixe que x^{j+1} existe bien et est bien défini.

En pratique, on peut utiliser la méthode itérative de construction de se point fixe donnée par la preuve du théorème pour approcher x^{j+1} . Une stratégie est de partir de la valeur donnée par le schéma d'Euler explicite

$$x^{j,0} = x^j + \Delta t_j f(t_j, x^j)$$

et affiner ensuite par l'algorithme du point fixe en itérant

$$x^{j,k+1} = F_j(x^{j,k})$$

jusqu'à ce que l'évolution relative

$$\frac{x^{j,k+1} - x^{j,k}}{x^{j,0}}$$

devienne inférieure à un seuil choisi par l'utilisateur. Puisque la suite $(x^{j,k})_{k \in \mathbb{N}}$ est de Cauchy, on sait que cette algorithme s'arrête en un nombre fini d'itérations.

Un tel schéma est plus lourd en terme de calculs qu'un algorithme explicite mais il apporte en général plus de stabilité et permet souvent d'utiliser un pas plus grand. EXO ?

Analyse d'erreur

L'objectif de l'analyse d'erreur *a priori* est de donner une estimation de l'erreur commise par la méthode numérique en fonction des paramètres du problème (temps d'intégration, pas de temps, propriétés de f). L'idée générale est de remarquer qu'à chaque pas de temps, on commet une erreur d'intégration locale (erreur de troncature dans la discrétisation de l'intégrale, à laquelle s'ajoutent souvent des erreurs d'arrondi), et que ces erreurs locales s'accumulent. Le contrôle de cette accumulation demande l'introduction d'une notion de stabilité adéquate, alors que les erreurs locales sont liées à une notion de consistance. L'alliance de stabilité et de consistance donne une propriété de convergence qui est souhaitée lors de l'implémentation de méthodes numériques.

Erreur de troncature locale

L'erreur de troncature locale est l'erreur résiduelle que l'on obtiendrait si on appliquait le schéma numérique à la solution exacte. En d'autres termes, c'est l'erreur due à l'approximation de l'intégrale et aux erreurs d'arrondi de l'ordinateur. Elle est ainsi définie comme

$$\eta^{j+1} := \frac{x(t_{j+1}) - x(t_j) - \Delta t_j \Phi_{\Delta t_j}(t_j, x(t_j))}{\Delta t_j}.$$

Consistance

On note $\Delta t = \max_{0 \leq j \leq J-1} \Delta t_j$ le pas de temps maximal. On dit qu'une méthode numérique est *consistante* si

$$\lim_{\Delta t \rightarrow 0} \left(\max_{1 \leq j \leq J} |\eta^j| \right) = 0,$$

et qu'elle est *consistante d'ordre p* s'il existe une constante C telle que, pour tout $0 \leq j \leq J-1$,

$$|\eta^{j+1}| \leq C(\Delta t_j)^p.$$

L'erreur de consistance se calcule souvent par des développements de Taylor des solutions lorsque celles-ci sont suffisamment régulières, et C s'exprime alors comme une borne des dérivées des solutions. En fait, on remarque que lorsque f est continue, la solution est C^1 (par définition de nos solutions). Mais puisque $\dot{x}(t) = f(t, x(t))$, \dot{x} hérite de la régularité de f : si f est C^k alors les solutions x sont C^{k+1} .

Condition suffisante

Consistance du schéma d'Euler explicite

L'erreur de troncature s'écrit

$$\eta^{j+1} = \frac{x(t_j + \Delta t_j) - \left(x(t_j) + \Delta t_j f(t_j, x(t_j)) \right)}{\Delta t_j}.$$

Or, si f est C^1 , alors x est C^2 et par application la formule de Taylor avec reste intégral, on a

$$x(t_j + \Delta t_j) = x(t_j) + \Delta t_j f(t_j, x(t_j)) + \Delta t_j^2 \int_0^1 \ddot{x}(t_j + s\Delta t_j)(1-s)ds,$$

en utilisant $\dot{x}(t_j) = f(t_j, x(t_j))$. Ceci donne donc

$$|\eta^{j+1}| \leq \Delta t_j \int_0^1 \ddot{x}(t_j + s\Delta t_j)(1-s)ds \leq \frac{\Delta t_j}{2} \max_{t \in [t_j, t_{j+1}]} \|\ddot{x}(t)\| \leq \frac{\Delta t_j}{2} \max_{t \in [0, T]} \|\ddot{x}(t)\|.$$

Le schéma d'Euler explicite est donc consistant d'ordre 1 avec

$$C = \frac{\max_{t \in [0, T]} \|\ddot{x}(t)\|}{2}.$$

Notons qu'en utilisant $\dot{x}(t) = f(t, x(t))$,

$$\ddot{x}(t) = \partial_t f(t, x(t)) + \partial_x f(t, x(t)) \cdot f(t, x(t)),$$

et on peut exprimer C en fonction de bornes sur x et sur les dérivées de f .

Stabilité

La notion de stabilité quantifie la robustesse de l'approximation numérique par rapport à l'accumulation des erreurs locales et perturbations.

Définition

On dit qu'une méthode numérique est *stable* s'il existe une constante $S(T) > 0$ (indépendante des Δt_j) telle que, pour toutes suites $x = \{x^j\}_{1 \leq j \leq J}$ et $z = \{z^j\}_{1 \leq j \leq J}$ vérifiant

$$\begin{cases} x^{j+1} = x^j + \Delta t_j \Phi_{\Delta t_j}(t_j, x^j), \\ z^{j+1} = z^j + \Delta t_j \Phi_{\Delta t_j}(t_j, z^j) + \Delta t_j \delta^{j+1}, \end{cases}$$

on ait

$$\max_{1 \leq j \leq J} |x^j - z^j| \leq S(T) \left(|x^0 - z^0| + \sum_{j=1}^J \Delta t_{j-1} |\delta^j| \right).$$

Condition suffisante

Si les $\Phi_{\Delta t_j}$ sont Lipschitziennes en x , c'est-à-dire il existe $L > 0$ tel que pour tout $0 \leq j \leq J$, et tout $(x_a, x_b) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$|\Phi_{\Delta t_j}(t_j, x_a) - \Phi_{\Delta t_j}(t_j, x_b)| \leq L|x_a - x_b|$$

alors le schéma est stable avec $S(T) = e^{LT}$.

Démonstration: On a alors

$$|x^{j+1} - z^{j+1}| \leq \Delta t_j |\delta^{j+1}| + (1 + \Delta t_j L) |x^j - z^j| \leq e^{\Delta t_j L} |x^j - z^j|$$

puisque $1 + x \leq e^x$ pour tout $x \in \mathbb{R}$. Par récurrence, on montre alors que pour tout $1 \leq j \leq J$,

$$|x^j - z^j| \leq e^{(t_j - t_0)L} |x^0 - z^0| + \sum_{k=1}^j e^{(t_j - t_k)L} \Delta t_{k-1} |\delta^k|.$$

Il s'ensuit que

$$|x^j - z^j| \leq e^{TL} \left(|x^0 - z^0| + \sum_{k=1}^j \Delta t_{k-1} |\delta^k| \right),$$

ce qui donne le résultat. ■

Convergence

La combinaison de consistance et de stabilité donne une propriété dite de *convergence* qui dit que l'erreur commise par le schéma par rapport à la vraie solution converge vers 0 lorsque le pas de temps converge vers 0.

Définition

Soit $\Delta t = \max_{0 \leq j \leq J-1} \Delta t_j$. Un schéma numérique est *convergent* si

$$\lim_{\Delta t \rightarrow 0} \max_{1 \leq j \leq J} |x^j - x(t_j)| = 0$$

lorsque $x^0 = x(t_0)$. S'il existe $p \in \mathbb{N}_{>0}$ et $C > 0$ (indépendent de Δt) tel que

$$\max_{1 \leq j \leq J} |x^j - x(t_j)| \leq C(\Delta t)^p$$

on dit que le schéma est *convergent à l'ordre p*.

Théorème de Lax

Une méthode stable et consistante (à l'ordre p) est convergente (à l'ordre p).

Démonstration: Notons $z^j = x(t_j)$. On remarque que

$$z^{j+1} = z^j + \Delta t_j \Phi_{\Delta t_j}(t_j, z^j) + \Delta t_j \eta^{j+1},$$

où η est l'erreur de consistance. D'après la propriété de stabilité, on a donc

$$|x^j - x(t_j)| \leq S(T) \sum_{j=1}^J \Delta t_{j-1} |\eta^j|,$$

et par consistance

$$|x^j - x(t_j)| \leq S(T) C \sum_{j=1}^J \Delta t_{j-1} (\Delta t_{j-1})^p \leq C S(T) T (\Delta t)^p. \blacksquare$$

Relaxation de la stabilité

Relaxation du théorème avec Lipschitz dans compact plutôt que global ?

Erreurs d'arrondi et pas optimal

A chaque itération, lorsque la machine calcule x^{j+1} , elle commet des erreurs d'arrondi de l'ordre de la précision machine. La solution obtenue est donc donnée par

$$\hat{x}^{j+1} = \hat{x}^j + \Delta t_j \Phi_{\Delta t_j}(t_j, \hat{x}^j) + \varepsilon^{j+1}$$

au lieu de

$$x^{j+1} = x^j + \Delta t_j \Phi_{\Delta t_j}(t_j, x^j).$$

La stabilité nous donne alors l'écart

$$|x^j - \hat{x}^j| \leq$$

Exercices

Consistance de schémas

Montrer que :

1. le schéma d'Euler implicite est consistant d'ordre 1 (en supposant que le pas est suffisamment petit pour que le schéma soit défini)
2. le schéma de Heun est consistant d'ordre 2.
3. le schéma du point milieu est consistant d'ordre 2.
4. la méthodes des trapèzes est consistante d'ordre 2.

Convergence de schémas

Supposons f uniformément Lipschitzien par rapport à x , c'est-à-dire qu'il existe L_f tel que pour tout $t \geq 0$, pour tout $(x_a, x_b) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$|f(t, x_a) - f(t, x_b)| \leq L_f |x_a - x_b|.$$

Montrer que le schéma de Heun et d'Euler implicite sont convergents.

Corrections

Consistance de schémas

Convergence de schémas

Références