Introduction aux Équations Différentielles

Kévin Santugini-Repiquet

Kévin Santugini

Introduction
aux
Équations Différentielles

Théorie et Méthodes numériques

Cours

Enseirb-Matmeca

2021–2022

Sommaire

Sommaire **1**

1 Existence et unicité des solutions **9**

2 Étude de la stabilité **24**

3 Résolution numérique d'une équation différentielle **37**

A Prérequis **80**

Table des matières **89**

Introduction et motivations

Le but de ce cours est d'introduire les outils de base pour l'étude des équations différentielles et de leurs solutions, *i.e.*, les équations de type

$$\boldsymbol{x}^{(m)}(t) = \boldsymbol{f}(t, \boldsymbol{x}(t), \boldsymbol{x}'(t), \boldsymbol{x}''(t), \dots, \boldsymbol{x}^{(m-1)}(t)),$$

où l'inconnue $\boldsymbol{x}$ est une fonction d'une seule variable réelle à valeurs dans $\mathbb{R}^d$ et où $\boldsymbol{f}$ est une fonction connue d'un ouvert $\Omega \subset \mathbb{R} \times (\mathbb{R}^d)^m$ à valeurs dans $\mathbb{R}^d$. Une telle équation différentielle est dite d'ordre m car elle fait intervenir la m^{e} dérivée de $\boldsymbol{x}$. Nous verrons dans ce chapitre, à la §2, que toute équation différentielle d'ordre $m \geqslant 1$ est équivalente à un système d'équations différentielles d'ordre 1. Pour cette raison, la plupart des résultats mathématiques seront énoncés pour des systèmes d'équations différentielles d'ordre 1, *i.e.*, des équations différentielles de la forme :

$$\boldsymbol{x}'(t) = \boldsymbol{f}(t, \boldsymbol{x}(t)).$$

Les équations différentielles sont utilisées, entre autres, pour modéliser des systèmes physiques d'évolution [1] ayant un nombre fini de degrés de liberté, voir quelques exemples §1. Une fois qu'un système physique est modélisé par une équation différentielle, il reste à étudier le comportement de ses solutions. Excepté pour certaines équations différentielles académiques, il est rare de pouvoir expliciter ces solutions exactes. En l'absence de solutions exactes, le comportement des solutions peut être étudié soit avec des outils théoriques soit avec l'outil de la simulation numérique. Dans ce cours, nous nous concentrerons sur ces outils. L'aspect modélisation, obtenir l'équation différentielle gouvernant un système physique, ne sera pas abordé.

Avant d'introduire ces outils, nous allons, dans ce chapitre introductif, donner plusieurs exemples d'équations différentielles.

1 Quelques exemples d'équations différentielles

La physique et la mécanique offrent de nombreux exemples d'équations différentielles. Nous citons ici quelques unes de ces équations différentielles.

1. Systèmes dont l'état dépend du temps. Les physiciens emploieraient le mot "dynamique".

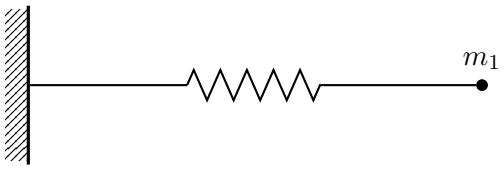
1.1 Exemples de mécanique du point

La seconde loi de Newton pour les systèmes fermés donne l'accélération d'un point mobile de masse m en fonction de la force qui est appliquée à ce point.

$$m\boldsymbol{X}''(t) = \boldsymbol{F}(t, \boldsymbol{X}(t), \boldsymbol{X}'(t)),$$

où la force $\boldsymbol{F}$ est une fonction connue. Il est fréquent que la force $\boldsymbol{F}$ ne dépende que de sa deuxième variable et dérive d'un potentiel $\phi\colon \mathbb{R}^3 \to \mathbb{R}$, *i.e.* : $\boldsymbol{F}(t, \boldsymbol{x}, \boldsymbol{v}) = -\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})$. Un exemple de force dérivée d'un potentiel est la force élastique.

Exemple 1.1 (Mouvement d'une masse mobile attachée à un ressort). On considère le mouvement unidimensionnel d'un point mobile de masse m attaché à un ressort de raideur k



$$mx''(t) + kx(t) = 0.$$

La force est dérivée du potentiel $\phi(x) = kx^2/2$.

Exemple 1.2 (Mouvement d'une masse dans un champ gravitationnel). Le mouvement d'une masse m dans un champ gravitationnel crée par une masse centrale M est donnée par

$$\boldsymbol{x}''(t) = -\frac{\mathcal{G}M\boldsymbol{x}(t)}{|\boldsymbol{x}|^3}.$$

où $\mathcal{G}$ est la constante universelle de la gravitation qui vaut $6.67{\cdot}10^{-11}\mathrm{kg}^{-1}\,\mathrm{m}^3\,\mathrm{s}^{-2}$. La force est dérivée du potentiel

$$\phi(\boldsymbol{x}) = -\frac{\mathcal{G}M}{|\boldsymbol{x}|}.$$

1.2 Exemple de mécanique du solide indéformable

Le mouvement d'un solide indéformable se décompose en un mouvement de translation du centre de gravité et un mouvement de rotation autour du centre de gravité. Si $\mathbf{I}$ est la matrice d'inertie du solide indéformable dans une base attachée au solide lui-même, et si $\boldsymbol{\omega}$ est le vecteur de rotation instantanée exprimée dans une base attachée au solide indéformable, alors l'équation différentielle régissant le mouvement de rotation du solide est donnée par :

$$\mathbf{I}\boldsymbol{\omega}' + \boldsymbol{\omega} \wedge (\mathbf{I}\boldsymbol{\omega}) = \mathcal{M}_0, \tag{1.1}$$

où $\mathcal{M}_0$ est le couple de force exercée sur le solide exprimé dans la même base que $\boldsymbol{\omega}$. En utilisant cette formule (ou une version simplifiée qui tient compte de l'invariance éventuelle par rotation autour de l'axe vertical), on peut calculer les équations du mouvement d'une toupie.

Exemple 1.3 (Mouvement d'une toupie, matrices de rotation). Considérons le mouvement d'une toupie de base circulaire de rayon R et de hauteur H. On suppose que la masse volumique de la toupie est constante et vaut ρ. Soit g la gravité terrestre. Soit $M = \pi R^2 h/3$ la masse de la toupie. On suppose aussi l'absence de frottement entre la pointe de la toupie et le plan horizontal. La matrice de rotation $\mathbf{R}(t)$ et le vecteur de rotation instantanée $\boldsymbol{omega}(t)$ satisfont le système suivant :

$$\mathbf{I}\boldsymbol{\omega}' + \boldsymbol{\omega} \wedge (\mathbf{I}\boldsymbol{\omega}) = \begin{bmatrix} 0 \\ 0 \\ -3H/4 \end{bmatrix} \wedge \mathbf{R}^{-1}(t) \begin{bmatrix} 0 \\ 0 \\ Mg \end{bmatrix}, \tag{1.2a}$$

et

$$\dot{\mathbf{R}}(t) = \mathbf{R}(t) \begin{bmatrix} 0 & -\omega_3 & \omega_1 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}, \tag{1.2b}$$

où $(\omega_1, \omega_2, \omega_3)$ sont les composantes du vecteur de rotation instantannée $\boldsymbol{\omega}$ exprimé dans le base mobile attachée au solide et où la matrice d'inertie dans une base attachée au solide est :

$$\mathbf{I} = M \begin{bmatrix} \frac{3}{80}H^2 + \frac{3}{20}R^2 & 0 & 0 \\ 0 & \frac{3}{80}H^2 + \frac{3}{20}R^2 & 0 \\ 0 & 0 & \frac{3}{10}R^2 \end{bmatrix}.$$

Il s'agit d'un système d'EDO d'ordre 1 faisant intervenir 12 fonctions scalaires. Mais certaines sont redondantes, en effet, comme $\mathbf{R}(t)$ doit être une rotation, ses 9 composantes ne sont pas indépendantes.

En exprimant la rotation $\mathbf{R}(t)$ et les vecteurs de rotation instantanée en fonction des angles d'Euler, on peut obtenir un système d'EDO d'ordre 2 ne faisant intervenir que trois fonctions scalaires.

Exemple 1.4 (Mouvement d'une toupie, angles d'Euler). Considérons le mouvement d'une toupie de base circulaire de rayon R et de hauteur H. On suppose que la masse volumique de la toupie est constante et vaut ρ. Soit g la gravité terrestre. Soit $M = \pi R^2 H/3$ la masse de la toupie. On suppose aussi l'absence de frottement entre la pointe de la toupie et le plan horizontal. Soit ϕ, η et θ les angles d'Euler : précession, nutation et rotation propre. Ces angles vérifient le système (non-linéaire) d'équations différentielles suivant :

$$\ddot{\eta} - I_3\dot{\theta}\dot{\phi}\sin\eta + (I_1 - I_3)\dot{\phi}^2 \sin\eta\cos\eta = -\frac{3}{4}MgH\sin\eta, \tag{1.3a}$$

$$-\ddot{\phi}\sin\eta - \dot{\phi}\dot{\eta}\cos\eta + (I_1 - I_3)\dot{\eta}\dot{\phi}\cos\eta - I_3\dot{\eta}\dot{\theta} = 0, \tag{1.3b}$$

$$\ddot{\theta} + \ddot{\phi}\cos\eta - \dot{\phi}\dot{\eta}\sin\eta = 0. \tag{1.3c}$$

où

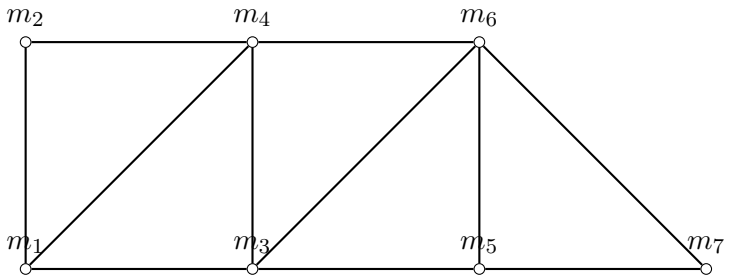
$$I_1 = \frac{3}{80}Mh^2 + \frac{3}{20}MR^2, \qquad I_3 = \frac{3}{10}MR^2.$$

Un inconvénient de ces équations est que le système devient singulier quand l'angle de nutation η tend vers 0.

1.3 Exemple en Élasticité

Les équations différentielles permettent aussi de modéliser le comportement de solides déformables. Par exemple, on peut modéliser le comportement d'un treillis de barres élastiques.

Exemple 1.5 (Treillis de barres). Considérons un ensemble de barres élastiques de masse nulle connectant N masses mobiles ponctuelles m_i, $1 \leqslant i \leqslant N$. Notons $\boldsymbol{x}_i$ la position de la masse m_i.



Notons $\mathcal{N}_i$ l'ensemble des indices j pour lesquels le point $\boldsymbol{x}_j$ est connecté au point $\boldsymbol{x}_i$ par une barre. Quand j est dans $\mathcal{N}_i$, notons k_{ij} la raideur de la barre connectant les points $\boldsymbol{x}_i$ et $\boldsymbol{x}_j$; et notons ℓ_{ij} la longueur au repos de cette barre. Alors pour tout indice i,

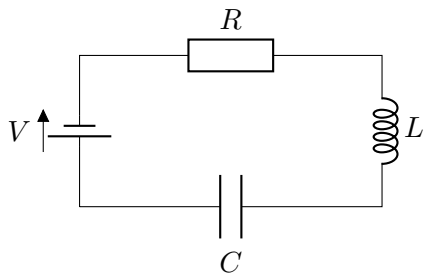
$$m_i \boldsymbol{x}_i''(t) = -\sum_{j\in\mathcal{N}_i} k_{ij}(\ell_{ij} - \|\boldsymbol{x}_i - \boldsymbol{x}_j\|)\frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}.$$

En pratique, ce système sera linéarisé pour les petits déplacements.

1.4 Exemple de modélisation d'un circuit électrique RLC

Les EDO sont aussi utiles pour modéliser le comportement des circuits électriques.

Exemple 1.6 (Circuit électrique). Considérons le circuit électrique suivant :



Il s'agit d'un circuit RLC classique. L'équation différentielle modélisant ce circuit électrique est

$$Lq''(t) + Rq'(t) + \frac{1}{C}q(t) = V(t).$$

où q est la charge du condensateur, et l'intensité électrique i dans le circuit est égal à $q'(t)$.

1.5 Exemple de désintégration atomique

Pour dater les roches (et même des peintures), on peut comparer les concentrations entre divers radioisotopes de demi-vie différentes.

Exemple 1.7 (Désintégration de radio-isotopes). L'uranium 234 a une demi-vie de $t_{1/2,U_{234}} = 3.45 \times 10^5$ ans. L'équation donnant la quantité d'uranium 234 au cours du temps (si on suppose qu'il n'y a dans l'échantillon aucun précurseur de l'uranium 234) est

$$Q'_{U_{234}}(t) = -\frac{\ln(2)}{t_{1/2,U_{234}}} Q_{U_{234}}(t).$$

où t est donnée en années. La désintégration de l'uranium 234 crée du Thorium 230 lui-même radio-actif de demi-vie $t_{1/2,Th_{230}} = 7.6 \times 10^4$ ans et dont la désintégration crée du Radium 226. L'équation donnant la quantité de Thorium 230 au cours du temps est

$$Q'_{Th_{230}}(t) = \frac{\ln(2)}{t_{1/2,U_{234}}} Q_{U_{234}}(t) - \frac{\ln(2)}{t_{1/2,Th_{230}}} Q_{Th_{230}}(t).$$

1.6 Exemple de dynamique des populations

Les lois de la physique ne sont pas les seuls exemples d'équations différentielles. Elles sont aussi utilisées en dynamique des populations :

Exemple 1.8 (Modèle prédateur proie). Si on souhaite prédire l'évolution de la population de plusieurs espèces, par exemple lapins et renards, Il est

possible d'utiliser un modèle prédateur-proie. Une modélisation possible de l'évolution dans le temps de ces deux espèces est donné par

$$\begin{aligned} \ell'(t) &= \alpha\ell^2(t) - \beta\ell(t)r(t), \\ r'(t) &= \gamma r^2(t) + \delta\ell(t)r(t), \end{aligned}$$

où $r(t)$ représente la densité de population des renards à l'instant t et $\ell(t)$ représente la densité de population de lapins à l'instant t. Ici α, β, γ et δ sont des paramètres qu'il conviendra de déterminer en comparant les prédictions du modèle à l'observation sur le terrain.

1.7 Exemple de Cinétique chimique

La chimie, en particulier la cinétique chimique, offre de nombreux exemples d'équations différentielles :

Exemple 1.9 (Cinétique chimique). Considérons une réaction chimique $\alpha A + \beta B \to \gamma C$ et notons $\nu_A(t)$, $\nu_B(t)$ et $\nu_C(t)$ les concentrations molaires des espèces $[A]$, $[B]$ et $[C]$ à l'instant t. Pour prédire la vitesse à laquelle la réaction chimique se produit, les chimistes utilisent des équations différentielles. Pour une réaction éléméntaire non réversible, un modèle de cinétique possible est

$$\frac{\nu_A'(t)}{\alpha} = \frac{\nu_B'(t)}{\beta} = -\frac{\nu_C'(t)}{\gamma} = -K\nu_A\nu_B$$

où K est un paramètre appelé « coefficient de vitesse ».

1.8 Exemple de Mécanique du vol

Si on considère la trajectoire d'un avion dans un plan vertical en faisant l'approximation d'un Terre plate (*i.e.* $\boldsymbol{g}$ constant), et en négligent le vent, alors les équations de la mécanique du vol sont :

$$\begin{aligned} \dot{X} &= V\cos(\gamma), \\ \dot{h} &= V\sin(\gamma), \\ m\dot{V} &= -mg\sin(\gamma) + T\cos(\varepsilon), \\ mV\dot{\gamma} &= -mg\cos(\gamma) + T\sin(\varepsilon), \\ \dot{m} &= -\beta. \end{aligned} \tag{1.4}$$

où

- m est la masse de l'avion (la masse d'un avion dépend du temps).
- X est la position horizontale de l'avion.
- h est l'altitude de l'avion
- V est la magnitude du vecteur vitesse de l'avion.

— γ est l'angle entre le vecteur vitesse de l'avion et l'axe horizontal.
— α est l'angle d'attaque, angle entre l'axe du fuselage et le vecteur vitesse.
— ε est l'angle entre le vecteur vitesse de l'avion et la direction de poussée.
— L est la portance qui est une fonction connue de V, de h et de α.
— D est la trainée qui est une fonction connue de V, de h et de α.
— π est le paramètre de contrôle du moteur.
— β est la masse de carburant consommée par unité de temps et est une fonction connue de V, h et π.
— T est la poussée et est une fonction connue de V, h et π.

Ici, les variables m, X, h, V et γ sont les inconnues du système d'EDO. Et les variables π (paramètre moteur), ε (angle de poussée), et α (angle d'attaque) sont les paramètres de contrôle sur lequel le pilote peut agir.

En mécanique du vol, le but n'est pas de résoudre les équations différentielles pour calculer une trajectoire mais de calculer les paramètre de contrôle permettant d'arriver à destination (le plus rapidement ou le plus économiquement) ou permettant de suivre une certaine trajectoire. C'est ce que l'on appelle un problème de contrôle ou de contrôle optimal.

2 Équations différentielles d'ordre m

Nous allons voir dans cette section que toute équation différentielle d'ordre m est équivalente à une équation différentielle d'ordre 1. Considérons l'équation différentielle

$$\boldsymbol{x}^{(m)}(t) = f(t, \boldsymbol{x}(t), \boldsymbol{x}'(t), \boldsymbol{x}''(t), \dots, \boldsymbol{x}^{(m-1)}(t)), \tag{2.1}$$

où l'inconnue $\boldsymbol{x}$ est une fonction à valeurs dans $\mathbb{R}^d$ et $\boldsymbol{f}$ est une fonction à valeurs dans $\mathbb{R}^d$ définie sur un ouvert Ω de $\mathbb{R} \times \mathbb{R}^d$. On pose

$$\begin{aligned} \boldsymbol{F} : \mathbb{R} \times (\mathbb{R}^d)^m &\to (\mathbb{R}^d)^m \\ (t, \boldsymbol{y}_{(0)}, \dots, \boldsymbol{y}_{(m-2)}, \boldsymbol{y}_{(m-1)}) &\mapsto (\boldsymbol{y}_{(1)}, \dots, \boldsymbol{y}_{(m-1)}, f(t, \boldsymbol{y}_{(0)}, \dots, \boldsymbol{y}_{(m-2)}, \boldsymbol{y}_{(m-1)}) \end{aligned}$$

On considère alors l'équation différentielle

$$\boldsymbol{Y}'(t) = \boldsymbol{F}(t, \boldsymbol{Y}). \tag{2.2}$$

Si $\boldsymbol{x}$ est solution de (2.1), alors $\boldsymbol{Y} = (\boldsymbol{x}, \boldsymbol{x}', \dots, \boldsymbol{x}^{(m-1)})$ est solution de (2.2). Réciproquement, si $\boldsymbol{Y} = (\boldsymbol{y}_{(0)}, \dots, \boldsymbol{y}_{(m-1)})$ est solution de (2.2), alors $\boldsymbol{y}_{(0)}$ est solution de (2.1).

Comme toute équation différentielle d'ordre m peut se mettre sous la forme d'une équation différentielle d'ordre 1, nous énoncerons par la suite, les théorèmes généraux sur les équations différentielles dans le cas particulier des équations différentielles d'ordre 1.

Chapitre 1

Existence et unicité des solutions

Dans ce chapitre, nous commençons par étudier la plus simple des équations différentielles : $x' = ax$. De cette étude, nous en déduisons le lemme de Grönwall. En utilisant ce lemme, nous démontrons ensuite le théorème de Cauchy-Lipshitz sur l'existence et l'unicité des solutions aux équations différentielles. Enfin, nous donnons plusieurs méthodes permettant la résolution exacte d'équations différentielles.

1.1 Définitions et vocabulaire

Dans cette section, nous allons donner les définitions et le vocabulaire mathématique propre aux équations différentielles. Dans toute cette section, Ω est un ouvert de $\mathbb{R} \times \mathbb{R}^d$, f est une fonction de $\Omega \subset \mathbb{R} \times \mathbb{R}^d$ à valeurs dans $\mathbb{R}^d$ et on considère l'équation différentielle

$$\boldsymbol{x}' = f(t, \boldsymbol{x}). \tag{1.1.1}$$

Si $d = 1$, l'équation différentielle est dite scalaire.

Commençons par préciser la notion de solution à une équation différentielle.

Définition 1.1

On dit que $(I, \boldsymbol{x})$, avec $I \subset \mathbb{R}$ et $\boldsymbol{x} \colon I \to \mathbb{R}^d$ est solution de l'équation différentielle (1.1.1) si

- Pour tout t dans I, $(t, \boldsymbol{x}(t))$ est dans Ω.
- $\boldsymbol{x}$ est de classe $\mathcal{C}^1$ sur I.
- Pour tout t dans I, $\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t))$.

Remarquer qu'une solution à une équation différentielle est toujours définie sur un intervalle. Par example, la fonction inverse $t \mapsto 1/t$ vérifie l'équation $x'(t) = -x(t)^2$. Cette fonction admet deux branches, une sur $\mathbb{R}^{+,*}$ et une sur $\mathbb{R}^{-,*}$. Ainsi, $(\mathbb{R}^{+,*}, t \mapsto 1/t)$ et $(\mathbb{R}^{-,*}, t \mapsto 1/t)$ sont solutions de $x' = -x^2$ mais $(\mathbb{R}^*, t \mapsto 1/t)$ ne l'est pas car $\mathbb{R}^*$ n'est pas un intervalle.

Si on adjoint des conditions initiales à une équation différentielle, on parle de problème de Cauchy.

Définition 1.2: Problème de Cauchy

Soit $(t_0, \boldsymbol{x}_0)$ dans Ω. Le système composé de l'équation différentielle (1.1.1) et de la condition initiale $\boldsymbol{x}(t_0) = \boldsymbol{x_0}$ est appelé problème de Cauchy.

$$\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t)), \tag{1.1.2a}$$
$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0. \tag{1.1.2b}$$

On dit que $(I, \boldsymbol{x})$, avec I intervalle de $\mathbb{R}$ et $\boldsymbol{x} \colon I \to \mathbb{R}^d$ est solution du problème de Cauchy (1.1.2) si $(I, \boldsymbol{x})$ est solution de l'équation différentielle (1.1.2a) et si les conditions initiales (1.1.2b) sont satisfaites.

Si l'expression d'une solution à un problème de Cauchy admet plusieurs branches, seule la restriction de cette fonction à la branche contenant l'instant initial est solution du problème de Cauchy. Regardons sur un exemple. Soit le problème de Cauchy, $x' = -x^2$, $x(1/2) = 2$. La solution à ce problème de Cauchy est la branche $(\mathbb{R}^{+,*}, t \mapsto 1/t)$ car l'instant initial $1/2$ appartient à $\mathbb{R}^{+,*}$. Si on regarde le problème de Cauchy $x' = -x^2$, $x(-1/2) = -2$, alors la solution est la branche $(\mathbb{R}^{-,*}, t \mapsto 1/t)$ car l'instant initial est dans $\mathbb{R}^{-,*}$.

Si $(I, \boldsymbol{x})$ est solution d'un problème de Cauchy (1.1.2), et si J est un intervalle inclus dans I, alors $(J, \boldsymbol{x}|_J)$ est aussi solution de (1.1.2). En pratique, nous ne nous intéresserons qu'aux solutions dites maximales :

Définition 1.3. *Une solution $(I, \boldsymbol{x})$ de (1.1.2) est dite maximale si elle ne peut être prolongée par une solution de (1.1.2) définie sur un intervalle $\tilde{I} \supsetneq I$. I.E., si pour tout $(\tilde{I}, \tilde{\boldsymbol{x}})$ solution de (1.1.2),*

$$\left(I \subset \tilde{I} \text{ et } \tilde{\boldsymbol{x}}|_I = \boldsymbol{x}\right) \implies \tilde{I} = I \text{ et } \tilde{\boldsymbol{x}} = \boldsymbol{x}.$$

1.2 Équations différentielles linéaires d'ordre 1

Commençons par considérer l'une des plus simples équations différentielles :

$$x'(t) = ax(t), \tag{1.2.1}$$

où a appartient à $\mathbb{R}$. Nous recherchons toutes les fonctions x de classe $\mathcal{C}^1$ définies sur un intervalle $I \subset \mathbb{R}$ qui satisfont (1.2.1). La somme de deux solutions de (1.2.1) est une solution. Le produit d'une solution de (1.2.1) par un scalaire est aussi une solution. L'ensemble des solutions de (1.2.1) forment un espace vectoriel. L'équation différentielle (1.2.1) est dite linéaire. Elle est d'ordre 1 car elle ne fait intervenir que la dérivée d'ordre 1. Et elle est dite à coefficients constants car a est dans $\mathbb{R}$ et ne dépend donc pas du temps t.

Nous connaissons toutes les solutions de cette équation différentielle.

Proposition 1.4. *Soit A dans $\mathbb{R}$, $t \mapsto A\exp(at)$ est solution de (1.2.1). Réciproquement, si $x : I \mapsto \mathbb{R}$ est solution de (1.2.1), il existe A dans $\mathbb{R}$ tel que pour tout t dans I, $x(t) = A\exp(at)$.*

Démonstration. Il est aisé de vérifier que $t \mapsto A\exp(at)$ est solution de (1.2.1). Il faut maintenant démontrer que toutes les solutions de (1.2.1), sont de cette forme. Soit x solution de (1.2.1). Posons

$$\begin{aligned} y\colon I &\to \mathbb{R}, \\ t &\mapsto x(t)\exp(-at). \end{aligned}$$

On montre que $y'(t)$ est nulle pour tout t dans I et donc que la fonction y est constante sur I. Notons A cette constante. □

Si on adjoint une condition initiale à cette équation différentielle, alors on peut calculer A en fonction de la condition initiale : le problème de Cauchy

$$x'(t) = ax(t), \tag{1.2.2a}$$
$$x(t_0) = x_0, \tag{1.2.2b}$$

admet $t \mapsto x_0\exp(a(t-t_0))$ comme unique solution. *I.E.*, ici $A = x_0\exp(-at_0)$.

Considérons maintenant la même équation différentielle lorsque a n'est plus une constante et dépend du temps :

$$x'(t) = a(t)x(t), \tag{1.2.3a}$$
$$x(t_0) = x_0 \tag{1.2.3b}$$

où a est une fonction continue sur un intervalle de $\mathbb{R}$. Cette équation différentielle est linéaire à coefficients variables. En employant les mêmes techniques qu'au lemme 1.4, on démontre que l'unique solution au problème de Cauchy (1.2.3) est :

$$x(t) = x_0\exp\left(\int_{t_0}^{t} a(s)\mathrm{d}s\right). \tag{1.2.4}$$

Considérons maintenant la version non homogène de l'EDO (1.2.3) obtenue en rajoutant un terme non nul au second membre :

$$x'(t) = a(t)x(t) + g(t), \tag{1.2.5a}$$

$$x(t_0) = x_0 \tag{1.2.5b}$$

où a et g sont des fonctions continues sur un intervalle de $\mathbb{R}$. Inspirée par la solution de (1.2.3), on remplace la constante x_0 dans (1.2.4) par une fonction dépendant de t. *I.E.*, nous cherchons une solution de la forme

$$x(t) = \lambda(t) \exp\left(\int_{t_0}^{t} a(s)\mathrm{d}s\right). \tag{1.2.6}$$

On appelle cette méthode « méthode de variation de la constante ». Elle est ainsi nommée car la constante dans l'expression (1.2.4) a été remplacée par la fonction λ. Le problème de Cauchy (1.2.5) est alors équivalent à

$$\lambda'(s) = \exp\left(-\int_{t_0}^{s} a(\sigma)\mathrm{d}\sigma\right) g(s), \tag{1.2.7a}$$

$$\lambda(t_0) = x_0. \tag{1.2.7b}$$

Donc, l'unique solution du problème de Cauchy (1.2.5) est

$$x(t) = x_0 \exp\left(\int_{t_0}^{t} a(s)\mathrm{d}s\right) + \int_{t_0}^{t} \exp\left(\int_{s}^{t} a(\sigma)\mathrm{d}\sigma\right) g(s)\mathrm{d}s. \tag{1.2.8}$$

1.2.1 Lemme de Grönwall

La même technique que celle utilisée pour la démonstration de la Proposition 1.4 permet de démontrer le lemme de Grönwall. Ce lemme nous sera utile pour démontrer l'unicité dans le théorème de Cauchy-Lipschitz à la section 1.3.

Lemme 1.5: Lemme de Grönwall

Soit t_0 et t dans $\mathbb{R}$. Alors les deux propriétés suivantes sont vraies :

1. Soit a une fonction continue sur $[t_0, t]$ à valeurs dans $\mathbb{R}$. Soit $r\colon [t_0, t] \to \mathbb{R}$ une fonction continue sur $[t_0, t]$, de classe $\mathcal{C}^1$ sur $]t_0, t[$, et telle que pour tout $t_0 < s < t$, on a $r'(s) \leqslant a(s)r(s)$. Alors,
$$r(t) \leqslant r(t_0) \exp\left(\int_{t_0}^{t} a(s)\mathrm{d}s\right). \tag{1.2.9}$$
2. Soit a continue de $[t_0, t]$ à valeurs dans $\mathbb{R}^+$. Soit $\boldsymbol{x}\colon [t_0, t] \to \mathbb{R}^d$ une fonction continue sur $[t_0, t]$, de classe $\mathcal{C}^1$ sur $]t_0, t[$, et telle que pour tout $t_0 < s < t$, on a $\|\boldsymbol{x}'(s)\| \leqslant a(s)\|\boldsymbol{x}(s)\|$. Alors,
$$\|\boldsymbol{x}(t)\| \leqslant \|\boldsymbol{x}(t_0)\| \exp\left(\int_{t_0}^{t} |a(s)|\mathrm{d}s\right). \tag{1.2.10}$$

Démonstration. 1. Il suffit de calculer la dérivée de l'application $t \mapsto r(t)\exp(-\int_{t_0}^{t} a(s)\mathrm{d}s)$ et d'observer qu'elle est négative sur $]t_0, t[$.

2. Soit $t > t_0$. Supposons d'abord que $\boldsymbol{x}$ ne s'annule pas sur $]t_0, t[$, alors $r : t \mapsto \|\boldsymbol{x}\|$ est de classe $\mathcal{C}^1$ sur $]t_0, t[$ et vérifie $r'(s) = \frac{(\boldsymbol{x}(s)|\boldsymbol{x}'(s))}{\|\boldsymbol{x}(s)\|}$ donc $r'(s) \leqslant a(s)r(s)$. On applique la première partie du lemme de Grönwall et on obtient (1.2.10).
Si $\boldsymbol{x}$ s'annule sur $]t_0, t[$, alors l'ensemble $\{s : t_0 \leqslant s \leqslant t, \boldsymbol{x}(s) = 0\}$ est non vide et est borné. On pose $\hat{t} = \sup\{s : t_0 \leqslant s \leqslant t, \boldsymbol{x}(s) = 0\}$. On a $\boldsymbol{x}(\hat{t}) = 0$ et $\boldsymbol{x}$ ne s'annule pas sur $]\hat{t}, t[$. On applique le résultat précédent en remplaçant t_0 par $\hat{t}$ et on obtient

$$\|\boldsymbol{x}(t)\| \leqslant \|\boldsymbol{x}(\hat{t})\| \exp\left(\int_{\hat{t}}^{t} a(s)\mathrm{d}s\right) = 0. \qquad \square$$

1.3 Théorème de Cauchy-Lipschitz

Nous pouvons maintenant énoncer le théorème de Cauchy-Lipschitz maximal.

Théorème 1.6: Cauchy-Lipschitz

Soit Ω ouvert de $\mathbb{R} \times \mathbb{R}^d$. Soit $f\colon \Omega \to \mathbb{R}^d$ de classe $\mathcal{C}^1$ sur Ω. Alors le problème de Cauchy (1.1.2) admet une unique [a] solution maximale $(I, \boldsymbol{x})$. L'intervalle d'existence maximale I est un ouvert.

a. Unicité est à comprendre au sens suivant. Toute solution $(\hat{I}, \hat{\boldsymbol{x}})$ de (1.1.2) vérifie $\hat{I} \subset I$, et pour tout t dans $\hat{I}$, on a $\hat{\boldsymbol{x}}(t) = \boldsymbol{x}(t)$.

Démonstration. Nous nous contentons de donner l'idée de la preuve sans rentrer dans tous les détails techniques.

Unicité Soient $(I, \boldsymbol{x})$ et $(\hat{I}, \hat{\boldsymbol{x}})$ deux solutions à (1.1.2). Nous souhaitons démontrer que $\boldsymbol{x}$ et $\hat{\boldsymbol{x}}$ coïncident sur $I \cap \hat{I}$. Par l'absurde, si ce n'est pas le cas, alors il existe t dans $\hat{I} \cap I$ tel que $\boldsymbol{x}(t) \neq \hat{\boldsymbol{x}}(t)$. Il y a deux cas possibles : soit $t < t_0$, soit $t > t_0$. Ces deux cas se traitent exactement de la même manière. Nous ne considérerons que le cas $t > t_0$. Posons $\tilde{s} = \inf\{s > t_0, \boldsymbol{x}(s) \neq \hat{\boldsymbol{x}}(s)\}$. On a $\boldsymbol{x}(s) = \hat{\boldsymbol{x}}(s)$ pour s dans $[t_0, \tilde{s}[$, donc par continuité $\boldsymbol{x}(\tilde{s}) = \hat{\boldsymbol{x}}(\tilde{s})$. Soit $\varepsilon > 0$ tel que $A = [\tilde{s}, \tilde{s} + \varepsilon] \times \overline{B(\boldsymbol{x}(\tilde{s}), \varepsilon)} \subset \Omega$. Soit $\delta > 0$ tel que $\delta < \varepsilon$ et tel que pour tout s dans $[\tilde{s}, \tilde{s} + \delta]$,

$$\|\hat{\boldsymbol{x}}(s) - \hat{\boldsymbol{x}}(\tilde{s})\| \leqslant \varepsilon, \qquad \|\boldsymbol{x}(s) - \boldsymbol{x}(\tilde{s})\| \leqslant \varepsilon.$$

Posons $K = \sup_A \|\frac{\partial f}{\partial \boldsymbol{x}}\|$. On a pour s dans $[\hat{s}, \hat{s} + \delta]$,

$$\begin{aligned}\|\boldsymbol{x}'(s) - \hat{\boldsymbol{x}}'(s)\| &= \|f(s, \boldsymbol{x}(s)) - f(s, \hat{\boldsymbol{x}}(s))\|,\\ &\leqslant K\|\boldsymbol{x}(s) - \hat{\boldsymbol{x}}(s)\|.\end{aligned}$$

On applique le lemme de Grönwall, lemme 1.5, et on obtient que $\hat{\boldsymbol{x}}(s) = \boldsymbol{x}(s)$ pour tout s dans $[\tilde{s}, \tilde{s}+\delta]$, ce qui contredit la définition de $\tilde{s}$.

Existence d'une solution locale L'idée est de se placer sur un petit intervalle. On choisit $a > 0$ et $b > 0$ tel que $A = [t_0 - a, t_0 + a] \times \overline{B(\boldsymbol{x}_0, b)} \subset \Omega$. On pose $M = \sup_{(t,\boldsymbol{x})\in A} \|f(t,\boldsymbol{x})\|$, $K = \sup_{(t,\boldsymbol{x})\in A} \|\frac{\partial f}{\partial \boldsymbol{x}}\|$ et $r = \min(a, b/M)$. On pose alors pour t dans $[t_0 - a, t_0 + a]$

$$\boldsymbol{x}_{n+1}(t) = \boldsymbol{x}_0 + \int_{t_0}^{t} f(s, \boldsymbol{x}_n(s))\mathrm{d}s \quad \text{Pour tout } n \geqslant 1 \tag{1.3.1}$$

On démontre alors par récurrence que $\boldsymbol{x}_n$ est bien définie sur $]t_0 - r, t_0 + r[$, que $\|\boldsymbol{x}_n(t) - \boldsymbol{x}_0\| \leqslant aM \leqslant b$ pour t dans $]t_0 - r, t_0 + r[$ et que $\|\boldsymbol{x}_{n+1}(t) - \boldsymbol{x}_n(t)\| \leqslant K^n(t-t_0)^n/n!$. On réutilise (1.3.1) et on obtient que $\|\boldsymbol{x}'_{n+1}(t) - \boldsymbol{x}'_n(t)\| \leqslant K\|\boldsymbol{x}_{n+1}(t) - \boldsymbol{x}_n(t)\| \leqslant K^{n+1}(t-t_0)^n/n!$. Donc la série de terme $\boldsymbol{x}_{n+1} - \boldsymbol{x}_n$ et celle de terme $\boldsymbol{x}'_{n+1} - \boldsymbol{x}'_n$ convergent normalement donc les suites $(\boldsymbol{x}_n)_{n\in\mathbb{N}*}$ et $(\boldsymbol{x}'_n)_{n\in\mathbb{N}*}$ convergent uniformément sur $[t_0 - r, t_0 + r]$. La limite $\boldsymbol{x}$ est de classe $\mathcal{C}^1$, est solution de l'équation différentielle (1.1.2a) et vérifie les conditions initiales (1.1.2b).

Existence d'une solution maximale : On considère l'ensemble des solutions $\{(\hat{I}_i, \boldsymbol{x}_i)\}$ du problème de Cauchy (1.1.2), on pose $I = \bigcup_i I_i$ et $\boldsymbol{x}(t) = \boldsymbol{x}_i(t)$ pour tout t dans I_i. Le résultat d'unicité précédent garantit que la définition de $\boldsymbol{x}$ est cohérente. □

L'unicité des solutions du problème de Cauchy a pour conséquence l'observation suivante.

Corollaire 1.7. *Deux solutions maximales d'une équation différentielle vérifiant les hypothèses du théorème de Cauchy-Lipschitz sont soit identiques, soit ne se croisent jamais.*

En particulier, pour les équations différentielles scalaires, on a

Corollaire 1.8. *Soit Ω un ouvert de $\mathbb{R} \times \mathbb{R}$. Soit $f : \Omega \to \mathbb{R}$ vérifiant les hypothèses du théorème de Cauchy-Lipschitz. Soit (I_1, x_1) et (I_2, x_2) deux solutions à l'équation différentielle scalaire $x'(t) = f(t, x(t))$ avec $I_1 \cap I_2 \neq \varnothing$. Soit $\hat{t}$ dans $I_1 \cap I_2$, tel que $x_1(\hat{t}) < x_2(\hat{t})$, alors $x_1 < x_2$ sur $I_1 \cap I_2$.*

Ce corollaire est très pratique quand on considère les solutions stationnaires à une équation différentielle.

Corollaire 1.9

Soit $x' = f(x)$ une équation différentielle scalaire autonome. On appelle solution stationnaire toute constante c tel que $f(c) = 0$. Soit

(I, x) la solution du problème de Cauchy

$$x'(t) = f(x(t)),$$
$$x(t_0) = x_0.$$

Soit c une solution stationnaire.

1. Si $x_0 > c$, alors, pour tout t dans I, $x(t) > c$.
2. Si $x_0 < c$, alors, pour tout t dans I, $x(t) < c$.
3. Si $f(x_0) > 0$, alors x est croissante sur I.
4. Si $f(x_0) < 0$, alors x est décroissante sur I.

1.4 Résolution des équations différentielles autonomes

Dans cette section, on s'intéresse à la résolution exacte des équations différentielles autonomes scalaires, *i.e.*, aux équations différentielles dans lesquelles $f\colon \mathbb{R} \to \mathbb{R}$ ne dépend pas explicitement de t et vérifie les hypothèses du théorème de Cauchy-Lipschitz. Considérons le problème de Cauchy

$$x' = f(x), \tag{1.4.1a}$$
$$x(t_0) = x_0, \tag{1.4.1b}$$

On souhaite ramener ce problème de Cauchy au calcul d'une primitive. Soit (I, x) la solution maximale. Soit t dans I. On souhaite calculer $x(t)$. On distingue deux cas :

1er cas : Il existe $\hat{s}$ dans $[t_0, t]$ telle que $f(x(\hat{s})) = 0$. La fonction constante $a = x(\hat{s})$ est alors solution et l'unicité dans le théorème de Cauchy-Lipschitz implique que pour tout t dans I, $x(t) = a$.

2e cas : La fonction $s \mapsto f(x(s))$ ne s'annule jamais pour s dans $[t_0, t]$. On divise alors les deux côtés de (1.4.1a) par $f(x)$ puis en intégrant sur $[t_0, t]$. On obtient alors

$$\int_{t_0}^{t} \frac{x'(s)}{f(x(s))} \mathrm{d}s = t - t_0.$$

Puis, en effectuant le changement de variable $u = x(s)$, on obtient donc pour tout t dans I

$$\int_{x_0}^{x(t)} \frac{\mathrm{d}u}{f(u)} = t - t_0.$$

On ne peut pas aller plus loin sans connaître l'expression exacte de f. Nous allons donc continuer sur des exemples.

Exemple 1.4.1. On considère le problème de Cauchy suivant.

$$\begin{aligned} x' &= \exp(x) \\ x(t_0) &= x_0 \end{aligned}$$

On divise alors les deux côtés de l'équation par $\exp(x)$ puis on intègre :

$$\begin{aligned} \int_{t_0}^{t} \frac{x'(s)}{\exp(x(s))}\mathrm{d}s &= t - t_0, \\ \int_{x_0}^{x(t)} \exp(-u)\mathrm{d}u &= t - t_0, \\ \exp(-x_0) - \exp(-x(t)) &= t - t_0, \\ x(t) &= -\ln(\exp(-x_0) - (t - t_0)). \end{aligned}$$

Puis, on récupère l'intervalle d'existence $I =]-\infty, t_0 + \exp(-x_0)[$.

Exemple 1.4.2. On considère le problème de Cauchy suivant.

$$x' = x^2 \tag{1.4.2a}$$
$$x(t_0) = x_0 \tag{1.4.2b}$$

On obtient

$$\begin{aligned} \int_{t_0}^{t} \frac{x'(s)}{x^2(s)}\mathrm{d}s &= t - t_0, \\ \int_{x(t_0)}^{x(t)} \frac{\mathrm{d}u}{u^2} &= t - t_0, \\ \frac{1}{x_0} - \frac{1}{x(t)} &= t - t_0, \\ x(t) &= \frac{1}{\frac{1}{x_0} - (t - t_0)}. \end{aligned}$$

On remarque que le dénominateur de $1/(1/x_0 - (t - t_0))$ s'annule en $t^* = 1/x_0 + t_0$. Seule une des deux branches de $t \mapsto 1/(1/x_0 - (t - t_0))$ peut représenter la solution puisqu'une solution à une équation différentielle existe toujours sur un intervalle. L'intervalle d'existence est donc soit $]-\infty, t^*[$, soit $]t^*, +\infty,[$. Pour discerner lequel de ces deux intervalles est l'intervalle d'existence, le plus simple est de se rappeler que t_0 doit forcément appartenir à cet intervalle. Donc la solution du problème de Cauchy (1.4.2) est

$$\begin{aligned} &I = \mathbb{R} && x = t \mapsto 0 \quad \text{si } x_0 = 0 \\ &I =]-\infty, t_0 + \frac{1}{x_0}[&& x = t \mapsto \frac{1}{\frac{1}{x_0} - (t - t_0)} \quad \text{si } x_0 > 0, \\ &I =]t_0 + \frac{1}{x_0}, +\infty[&& x = t \mapsto \frac{1}{\frac{1}{x_0} - (t - t_0)} \quad \text{si } x_0 < 0 \end{aligned}$$

On peut résoudre les équations différentielles à variables séparable, *i.e.*, les équations différentielles de type $x' = f(x)g(t)$ avec cette même technique. On divise les deux côtés par $f(x)$ puis on intègre sur $[t_0, t]$. On obtient $\int_{x_0}^{x(t)} \mathrm{d}u/f(u) = \int_{t_0}^{t} g(s)\mathrm{d}s$.

1.5 Résolution des systèmes linéaires d'ordre 1

Dans cette section, nous considérons la résolution exacte de systèmes linéaires d'équations différentielles.

1.5.1 Le cas des coefficient constants

Soit d un entier supérieur ou égal à 2. Soit $\mathbf{A} \in \mathcal{M}_d(\mathbb{R})$ une matrice carré de taille d. On considère le problème de Cauchy suivant :

$$\boldsymbol{x}'(t) = \mathbf{A}\boldsymbol{x}(t), \tag{1.5.1a}$$
$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0. \tag{1.5.1b}$$

D'après le théorème de Cauchy-Lipschitz, ce problème de Cauchy admet une unique solution. On vérifie que $(\mathbb{R}, t \mapsto \exp((t-t_0)\mathbf{A})\boldsymbol{x}_0)$ est cette solution.

Soit g une fonction $\mathcal{C}^1$ sur $\mathbb{R}$ à valeurs dans $\mathbb{R}^d$. Supposons que l'on souhaite résoudre

$$\boldsymbol{x}'(t) = \mathbf{A}\boldsymbol{x}(t) + \boldsymbol{g}(t), \tag{1.5.2a}$$
$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0. \tag{1.5.2b}$$

D'après le théorème de Cauchy-Lipschitz, ce problème de Cauchy admet une unique solution. On cherche alors une solution sous la forme $t \mapsto \exp((t-t_0)A)\boldsymbol{\lambda}(t)$ où $\boldsymbol{\lambda}$ est une fonction $\mathcal{C}^1$ à valeurs dans $\mathbb{R}^d$. On obtient alors que (1.5.2) est équivalent à

$$\boldsymbol{\lambda}'(t) = \exp(-(t-t_0)\mathbf{A})\boldsymbol{g}(t)$$
$$\boldsymbol{\lambda}(t_0) = \boldsymbol{x}_0$$

Ce qui donne

$$\boldsymbol{\lambda}(t) = \boldsymbol{x}_0 + \int_{t_0}^{t} \exp(-(s-t_0)\mathbf{A})\boldsymbol{g}(s)\mathrm{d}s.$$

Donc, la solution $\boldsymbol{x}$ du système (1.5.2) est

$$\boldsymbol{x}(t) = \exp((t-t_0)\mathbf{A})\boldsymbol{x}_0 + \int_{t_0}^{t} \exp((t-s)\mathbf{A})\boldsymbol{g}(s)\mathrm{d}s. \tag{1.5.3}$$

1.5.2 Le cas des coefficients non constants

Considérons maintenant le cas où $\mathbf{A}$ dépend du temps.

$$\boldsymbol{x}'(t) = \mathbf{A}(t)\boldsymbol{x}(t) + \boldsymbol{g}(t), \tag{1.5.4a}$$
$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0. \tag{1.5.4b}$$

où $\mathbf{A}$ est une fonction de classe $\mathcal{C}^1$ définie sur un intervalle I de $\mathbb{R}$, à valeurs dans $\mathcal{M}_d(\mathbb{R})$. D'après le théorème de Cauchy-Lipschitz, ce problème de Cauchy admet une unique solution.

On ne connait pas la solution exacte pour cette équation dans le cas général même quand $\boldsymbol{g}$ est la fonction nulle. En particulier, on ne peut généraliser [1] la formule (1.2.4).

Cependant, on peut exprimer la solution du problème non homogène, *i.e.*, quand $\boldsymbol{g} \neq \mathbf{0}$, en fonction des solutions au problème homogènes, *i.e.*, quand $\boldsymbol{g} = \mathbf{0}$. Notons $\boldsymbol{w}_i$ la solution du problème de Cauchy

$$\boldsymbol{w}_i'(t) = \mathbf{A}(t)\boldsymbol{w}_i(t), \tag{1.5.5a}$$
$$\boldsymbol{w}_i(t_0) = \boldsymbol{w}_{i,0}, \tag{1.5.5b}$$

où $(\boldsymbol{w}_{i,0})_{1\leqslant i\leqslant d}$ forment une base de $\mathbb{R}^d$. Notons $\mathbf{W}_0$ la matrice dans $\mathcal{M}_d(\mathbb{R})$ dont la i^{e} colonne est $\boldsymbol{w}_{i,0}$ pour tout entier i, $1 \leqslant i \leqslant d$. Notons $\mathbf{W}$ l'application à valeurs dans $\mathcal{M}_d(\mathbb{R})$ définie par :

$$\boldsymbol{W}(t) = \begin{bmatrix} \begin{bmatrix} \\ \boldsymbol{w}_1(t) \\ \\ \end{bmatrix} & \dots & \begin{bmatrix} \\ \boldsymbol{w}_i(t) \\ \\ \end{bmatrix} & \dots & \begin{bmatrix} \\ \boldsymbol{w}_d(t) \\ \\ \end{bmatrix} \end{bmatrix}$$

La i^{e} colonne de $\mathbf{W}(t)$ est le vecteur $\boldsymbol{w}_i(t)$. La matrice $\mathbf{W}$ est appelée le Wronskien du système (1.5.5). L'unicité des solutions donnée par le théorème de Cauchy-Lipschitz implique que les $\boldsymbol{w}_i(t)$ forment une partie libre de $\mathbb{R}^d$ donc que $\mathbf{W}(t)$ est inversible quel que soit t. De plus, on a $\mathbf{W}'(t) = \mathbf{A}(t)\mathbf{W}(t)$ et $\mathbf{W}(t_0) = \mathbf{W}_0$. On cherche maintenant une solution à (1.5.4) sous la forme

$$\boldsymbol{x} : t \mapsto \mathbf{W}(t)\boldsymbol{\lambda}(t) \tag{1.5.6}$$

L'équation (1.5.4) est alors équivalente à

$$\boldsymbol{\lambda}'(t) = \mathbf{W}^{-1}(t)\boldsymbol{g}(t), \tag{1.5.7a}$$
$$\boldsymbol{\lambda}(t_0) = \mathbf{W}_0^{-1}\boldsymbol{x}_0, \tag{1.5.7b}$$
$$\boldsymbol{x}(t) = \mathbf{W}(t)\boldsymbol{\lambda}(t). \tag{1.5.7c}$$

1. On peut le faire seulement si pour tout t_1 et t_2, les matrices $\mathbf{A}(t_1)$ et $\mathbf{A}(t_2)$ commutent.

1.6 Résolution des équations linéaires scalaires d'ordre 2 à coefficients constants

Comme les équations différentielles scalaires linéaires d'ordre 2 sont équivalentes à un système linéaires de deux équations différentielles d'ordre 1, nous pouvons utiliser les techniques de la section précédente pour les résoudre.

1.6.1 Le cas homogène

Nous considérons l'équation différentielle

$$x''(t) + ax'(t) + bx(t) = 0. \tag{1.6.1}$$

Cette équation est dite homogène car le second membre est nulle. Elle est équivalente à

$$\begin{bmatrix} x(t) \\ x'(t) \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix} \begin{bmatrix} x(t) \\ x'(t) \end{bmatrix} \tag{1.6.2}$$

Pour résoudre cette équation, il suffit de calculer l'exponentielle de la matrice

$$t\mathbf{A} = t \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix}$$

Pour calculer cette exponentielle, il est utile de diagonaliser cette matrice. Le polynôme caractéristique de la matrice $\mathbf{A}$ est $X^2 + aX + b$. Soit r_1 et r_2 les racines de ce polynôme. On distingue deux cas :

1. Si $r_1 \neq r_2$, la matrice $\mathbf{A}$ est diagonalisable. Donc $\exp(tA)$ peut se mettre sous la forme

   $$\mathbf{P} \begin{bmatrix} \exp(r_1 t) & 0 \\ 0 & \exp(r_2 t) \end{bmatrix} \mathbf{P}^{-1},$$

   où P est la matrice de passage dont les colonnes sont les vecteurs propres de A. Toutes les solutions à (1.6.2) sont de la forme

   $$\begin{bmatrix} x(t) \\ x'(t) \end{bmatrix} = \mathbf{P} \begin{bmatrix} \exp(r_1 t) & 0 \\ 0 & \exp(r_2 t) \end{bmatrix} \mathbf{P}^{-1} \begin{bmatrix} x(0) \\ x'(0) \end{bmatrix}$$

   On ne s'intéresse qu'à la première composante et on en déduit que toute solution x à (1.6.2) est de la forme

   $$x(t) = \alpha \exp(r_1 t) + \beta \exp(r_2 t),$$

   où α et β sont des constantes.

2. Si $r_1 = r_2$, la matrice n'est pas diagonalisable. Posons $r = r_1 = r_2$. On peut par un changement de variable mettre $t\mathbf{A}$ sous la forme de Jordan :
$$t\mathbf{P}\begin{bmatrix} r & 1 \\ 0 & r \end{bmatrix}\mathbf{P}^{-1}$$
Et l'exponentielle de cette matrice vaut
$$\mathbf{P}\begin{bmatrix} \exp(rt) & t\exp(rt) \\ 0 & \exp(rt) \end{bmatrix}\mathbf{P}^{-1}$$

Et on en déduit que toute solution x à (1.6.2) est de la forme
$$x(t) = \alpha\exp(rt) + \beta t\exp(rt)$$
où α et β sont des constantes.

1.6.2 Cas non homogène

On ajoute maintenant un second membre et on considère l'équation
$$x''(t) + ax'(t) + bx(t) = g(t). \tag{1.6.3}$$
On va se ramener à la méthode de la section 1.5. Cette équation est équivalente à
$$\begin{bmatrix} x(t) \\ x'(t) \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix}\begin{bmatrix} x(t) \\ x'(t) \end{bmatrix} + \begin{bmatrix} 0 \\ g(t) \end{bmatrix} \tag{1.6.4}$$

Soient $t \mapsto \omega_1(t)$ et $t \mapsto \omega_2(t)$ les solutions élémentaires[2] à (1.6.1). Les solutions élémentaires à (1.6.1) sont $[\omega_1, \omega_1']^\top$ et $[\omega_2, \omega_2']^\top$. Nous définissons alors la matrice Wronskienne :
$$\mathbf{W} = \begin{bmatrix} \omega_1 & \omega_2 \\ \omega_1' & \omega_2' \end{bmatrix}$$
Et on recherche les solutions à (1.6.4) sous la forme $t \mapsto \mathbf{W}(t)\boldsymbol{\lambda}(t)$ où $\boldsymbol{\lambda}(t) = [\lambda_1(t), \lambda_2(t)]^\top$ et on obtient que le système (1.6.4) est équivalent à
$$\mathbf{W}(t)\boldsymbol{\lambda}'(t) = \begin{bmatrix} 0 \\ g(t) \end{bmatrix}$$
Soit
$$\omega_1(t)\lambda_1'(t) + \omega_2(t)\lambda_2'(t) = 0, \tag{1.6.5a}$$
$$\omega_1'(t)\lambda_1'(t) + \omega_2'(t)\lambda_2'(t) = g(t). \tag{1.6.5b}$$

2. Soit $\omega_1 = t \mapsto \exp(r_1 t)$ et $\omega_2 = t \mapsto \exp(r_2 t)$ si le polynôme $X^2 + aX + b$ admet deux racines distinctes r_1 et r_2, soit $\omega_1 = t \mapsto \exp(rt)$ et $\omega_2 = t \mapsto t\exp(rt)$ si le polynôme $X^2 + aX + b$ admet une unique racine double.

Si on ajoute les conditions initiales

$$x(t_0) = x_0 \qquad\qquad x'(t_0) = x_1,$$

à (1.6.3) pour obtenir un problème de Cauchy alors la condition initiale sur $\boldsymbol{\lambda}$ est

$$\mathbf{W}(t_0)\begin{bmatrix}\lambda_1(t_0)\\ \lambda_2(t_0)\end{bmatrix} = \begin{bmatrix}x_0\\ x_1\end{bmatrix}$$

Soit sous forme de système :

$$\omega_1(t_0)\lambda_1(t_0) + \omega_2(t_0)\lambda_2(t_0) = x_0, \tag{1.6.6a}$$
$$\omega_1'(t_0)\lambda_1(t_0) + \omega_2'(t_0)\lambda_2(t_0) = x_1. \tag{1.6.6b}$$

Une fois les λ_i calculées, il reste juste à poser $x(t) = \lambda_1(t)\omega_1(t) + \lambda_2(t)\omega_2(t)$. On aura aussi $x'(t) = \lambda_1(t)\omega_1'(t) + \lambda_2(t)\omega_2'(t)$.

1.6.3 Généralisation à un ordre entier m quelconque

Il est possible de généraliser la résolution de l'équation différentielle (1.6.1). En effet, considérons une équation différentielle linéaire homogène d'ordre m :

$$x^{(m)}(t) + \sum_{i=0}^{m-1} a_i x^{(i)}(t) = 0. \tag{1.6.7}$$

On pose alors le polynôme de degré m :

$$P(X) = X^m + \sum_{i=0}^{m-1} a_i X^i.$$

alors si on note par k le nombre de racines distinctes de P, par r_i la i^{e} racine de P et par μ_i sa multiplicité dans P alors toute solution de (1.6.7) est de la forme

$$x(t) = \sum_{i=1}^{k}\sum_{j=0}^{\mu_i-1} \alpha_{i,j} t^j \exp(r_i t)$$

où les $\alpha_{i,j}$ sont des scalaires. Comme à l'ordre 2, la formule ne peut pas être généralisée si les a_i ne sont pas des constantes.

Par contre, si les solutions à une équation différentielle homogène linéaire d'ordre m à coefficients non constants sont connues, on peut, comme à la section §1.6.2, ramener le cas non homogène à un problème de calcul de primitives.

Proposition 1.10. *Soit t_0 dans $\mathbb{R}$. Soient $(a_j)_{0\leqslant j\leqslant m-1}$ des fonctions de classe $\mathcal{C}^1$ sur $\mathbb{R}$ à valeurs dans $\mathbb{R}$. Soit $(w_i)_{1\leqslant i\leqslant m}$ une base de solutions de l'EDO homogène*

$$x^{(m)}(t) + \sum_{j=0}^{m-1} a_j(t)x^{(j)}(t) = 0 \tag{1.6.8}$$

On note $w_{i,j} = w_i^{(j)}(t_0)$. Soit g de classe $\mathcal{C}^1$ sur $\mathbb{R}$. Alors l'unique solution du problème de Cauchy

$$x^{(m)}(t) + \sum_{j=0}^{m-1} a_j(t)x^{(j)}(t) = g(t)$$
$$x^{(j)}(t_0) = x_j \quad \textit{pour tout } 0 \leqslant j \leqslant m-1$$

s'écrit

$$x(t) = \sum_{i=1}^{m} \lambda_i(t) w_i(t)$$

où les λ_i forment l'unique solution de

$$\sum_{i=1}^{m} w_i^{(j)}(t)\lambda_i'(t) = 0 \quad \textit{pour tout } 0 \leqslant j \leqslant m-2. \tag{1.6.9a}$$
$$\sum_{i=1}^{m} w_i^{(m-1)}(t)\lambda_i'(t) = g(t). \tag{1.6.9b}$$

avec pour conditions initiales

$$\sum_{i=1}^{m} \lambda_i(t_0) w_i^{(j)}(t_0) = x_j \quad \textit{pour tout } 0 \leqslant j \leqslant m-1. \tag{1.6.9c}$$

Conclusion

Dans ce chapitre, nous avons énoncé le théorème de Cauchy-Lipshitz sur l'existence-unicité des solutions maximales à une équation différentielle ordinaire. Nous avons aussi montré quelques méthodes pour calculer des solutions exactes aux EDO. En général, sauf dans de rares exceptions, nous ne pourrons pas calculer des solutions exactes aux EDO qui modélisent des problèmes réels. Quand il est impossible de calculer des solutions exactes, il faut soit établir des propriétés des solutions alors qu'on ne les a pas calculées explicitement, soit calculer des solutions numériques. Dans le chapitre 2, nous verrons comment étudier la propriété de stabilité d'une solution particulière d'une EDO sans calculer toutes les solutions de l'EDO. Dans le chapitre 3, nous verrons comment calculer des solutions numériques.

Références

[1] Serge Lang. *Real and Functional Analysis*. Third. Springer, 1993.

[2] Lev S. Pontryagin. *Ordinary Differential Equations*. Addison-Wesley, 1962.

Chapitre 2

Étude de la stabilité

Un système mécanique ou électrique peut avoir été conçu pour atteindre un certain régime. Par example, un circuit électrique peut être conçu pour émettre un signal sinusoïdal. Un système mécanique peut être conçu pour rester immobile, en équilibre [1]. Mais vérifier qu'une fonction, représentant le régime recherché, est bien solution de l'équation différentielle modélisant le système n'est pas suffisant : en pratique, il est impossible de fixer la position initiale d'une particule ou l'intensité initiale d'un courant électrique avec une précision infinie. Une erreur à l'instant initial sera toujours présente. Cette petite erreur va-t-elle s'amplifier au cours du temps et pousser le système en dehors de son régime souhaité ? Ou au contraire, va-t-elle rester bornée ou même diminuer au cours du temps et devenir négligeable. Dans le premier cas, la solution de l'équation différentielle sera dite instable, dans le deuxième cas elle sera dite stable [2].

Jusqu'ici, nous avons toujours considéré des problèmes de Cauchy où les conditions initiales exactes étaient connues. D'un point de vue mathématique, nous allons considérer un problème de Cauchy ou nous avons perturbé les conditions initiales

$$\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t)),, \tag{2.0.1a}$$
$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 + \boldsymbol{\varepsilon}, \tag{2.0.1b}$$

où $\boldsymbol{\varepsilon}$ est petit. La solution perturbée va-t-elle rester proche de la solution non perturbée ? Va-t-elle s'en approcher ? Ou au contraire va-t-elle s'en éloigner ? Un cas particulier important est celui des solutions stationnaires, *i.e.* quand $f(\boldsymbol{x}_0, t) = 0$ pour tout t. Vous avez probablement déjà rencontré ce problème dans vos cours de mécanique et de physique quand vous étudiez la stabilité

1. Par example, certains téléscopes spatiaux sont placés à proximité du point de Lagrange L2 à $1,5$ million de kilomètres derrière la Terre dans l'axe Soleil-Terre. Les points de Lagrange sont les points d'équilibre gravitationnels. Ils sont au nombre de 5.

2. Nous donnerons une définition mathématique précise de la notion de stabilité plus loin dans ce chapitre.

des « points d'équilibre ». C'est le même problème mais les mathématiciens parleront d'étude de la stabilité des « solutions stationnaires ».

2.1 Flot et dépendance de la solution par rapport aux conditions initiales

Afin de connaître la dépendance des solutions d'une équation différentielle en fonction des conditions initiales, nous introduisons le flot, une fonction représentant toutes les solutions à une EDO et où la dépendance par rapport aux conditions initiales t_0 et $\boldsymbol{x}_0$ est explicite. Soit $f\colon \Omega \to \mathbb{R}^d$ est de classe $\mathcal{C}^1$ sur un ouvert Ω de $\mathbb{R} \times \mathbb{R}^d$. Considérons le problème de Cauchy

$$\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t)), \tag{2.1.1a}$$
$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{2.1.1b}$$

On appelle flot de f l'application φ telle que pour tout $(t_0, \boldsymbol{x}_0)$ dans Ω, l'application $t \mapsto \varphi(t; t_0, \boldsymbol{x}_0)$ est l'unique solution de (2.1.1). Le domaine de définition du flot est noté Σ. Il s'agit d'un sous-ensemble de $\mathbb{R} \times \Omega$.

Le résultat suivant donne la dépendance de la solution de (2.1.1a) en fonction des conditions initiales.

Proposition 2.1. *Le domaine de définition Σ du flot φ est définie sur un ouvert Σ de $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$. Si f est de classe $\mathcal{C}^k$, alors son flot φ est aussi de classe $\mathcal{C}^k$ sur Σ. De plus,*

1. *Pour tout $(t; t_0, \boldsymbol{x}_0)$ dans Σ,*

$$\frac{\partial \varphi}{\partial t}(t; t_0, \boldsymbol{x}_0) = f(t, \varphi(t; t_0, \boldsymbol{x}_0)).$$

2. *La fonction*

$$\begin{aligned} \boldsymbol{y}\colon \Sigma &\to \mathbb{R}^d \\ (t; t_0, \boldsymbol{x}_0) &\mapsto \frac{\partial \varphi}{\partial t_0}(t; t_0, \boldsymbol{x}_0) \end{aligned}$$

est l'unique solution du problème de Cauchy suivant :

$$\frac{\partial \boldsymbol{y}}{\partial t}(t; t_0, \boldsymbol{x}_0) = \frac{\partial f}{\partial \boldsymbol{x}}(t, \varphi(t; t_0, \boldsymbol{x}_0)) \cdot \boldsymbol{y}(t; t_0, \boldsymbol{x}_0), \tag{2.1.2a}$$
$$\boldsymbol{y}(t_0; t_0, \boldsymbol{x}_0) = -f(t_0, \boldsymbol{x}_0). \tag{2.1.2b}$$

3. *La fonction*

$$\begin{aligned} \mathbf{J}\colon \Sigma &\to \mathcal{M}_d(\mathbb{R}), \\ t &\mapsto \frac{\partial \varphi}{\partial \boldsymbol{x}_0}(t; t_0, \boldsymbol{x}_0), \end{aligned}$$

est l'unique solution du problème de Cauchy suivant :

$$\frac{\partial \mathbf{J}}{\partial t}(t;t_0,\boldsymbol{x}_0) = \frac{\partial f}{\partial \boldsymbol{x}}(t,\varphi(t;t_0,\boldsymbol{x}_0))\mathbf{J}(t;t_0,\boldsymbol{x}_0), \tag{2.1.3a}$$

$$\mathbf{J}(t_0;t_0,\boldsymbol{x}_0) = \mathbf{I}_d. \tag{2.1.3b}$$

En particulier,

$$\frac{\partial \varphi}{\partial t_0}(t;t_0,\boldsymbol{x}_0) + \frac{\partial \varphi}{\partial \boldsymbol{x}_0}(t;t_0,\boldsymbol{x}_0) f(t_0,\boldsymbol{x}_0) = 0. \tag{2.1.4}$$

Démonstration. La partie difficile est de prouver que le flot φ est de classe $\mathcal{C}^k$. La preuve la plus courte utilise le théorème des fonctions implicites. Le lecteur intéressé pourra la trouver dans [1, chap. XIV, Théorème 4.3 et 5.2]. Une fois prouvé que le flot est au moins $\mathcal{C}^1$, il suffit de dériver

$$\frac{\partial \varphi}{\partial t}(t;t_0,\boldsymbol{x}_0) = f(t,\varphi(t;t_0,\boldsymbol{x}_0)$$

suivant t_0 ou suivant $\boldsymbol{x}_0$ en appliquant la règle de dérivation en chaînes. Le calcul de la condition initiale pour la dérivée suivant $\boldsymbol{x}_0$ est trivial. L'égalité (2.1.4) peut être retrouvé en dérivant l'égalité suivante

$$\varphi(t;t_0,\boldsymbol{x}_0) = \varphi(t;\hat{t},\varphi(\hat{t};t_0,\boldsymbol{x}_0)).$$

suivant la variable $\hat{t}$ en $\hat{t} = t_0$. On déduit de (2.1.4) la condition initiale pour la dérivée suivant t_0. □

2.2 Stabilité, attractivité et stabilité asymptotique

2.2.1 Définitions

Pour la notion de « stabilité au sens de Lyapounov », nous regardons comment la solution perturbée se comporte (dans le futur, pour $t \geqslant t_0$) quand la condition initiale de la solution perturbée tend vers la condition initiale de la solution non perturbée. À la figure 2.1, nous avons représenté un voisinage tubulaire d'épaisseur constante ε de la solution non perturbée $\boldsymbol{x}$. La solution non perturbée $\boldsymbol{x}$ est dite stable au sens de Lyapounov si pour tout voisinage tubulaire de $\boldsymbol{x}$, toute solution perturbée dont la condition initiale est « suffisamment proche » de $\boldsymbol{x}_0$ reste pour tout temps futur dans ce voisinage tubulaire. Nous donnons maintenant une définition rigoureuse de cette notion :

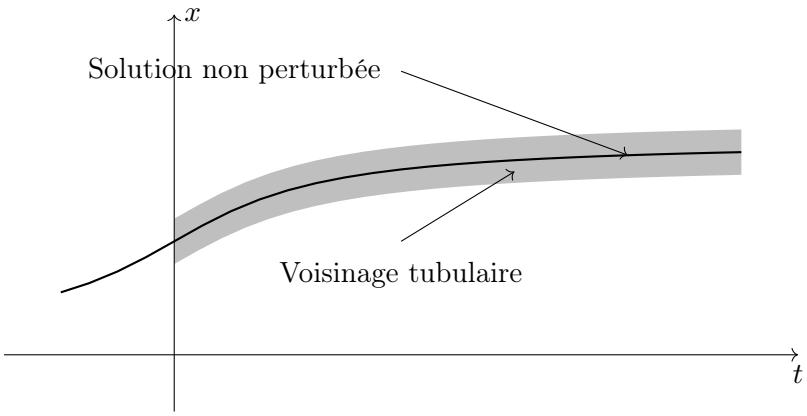


FIGURE 2.1 – Stabilité au sens de Lyapounov : la solution perturbée restera-t-elle dans le voisinage tubulaire si la perturbation est suffisamment petite ?

Définition 2.2: Stabilité au sens de Lyapounov

Soit Ω un ouvert de $\mathbb{R}\times\mathbb{R}^d$. Soit $(t_0,\boldsymbol{x}_0)$ dans Ω. Soit $f\colon\Omega\to\mathbb{R}^d$ de classe $\mathcal{C}^1$. Une solution $([t_0,+\infty[,\boldsymbol{x})$ de $\boldsymbol{x}'(t)=f(t,\boldsymbol{x}(t))$, $\boldsymbol{x}(t_0)=\boldsymbol{x}_0$ est dite stable au sens de Lyapounov si pour tout $\varepsilon>0$, il existe $\delta>0$ tel que pour tout $\tilde{\boldsymbol{x}}_0$ dans $\mathbb{R}^d$ tel que $\|\tilde{\boldsymbol{x}}_0-\boldsymbol{x}_0\|\leqslant\delta$, alors $[t_0,+\infty[\times\{t_0\}\times\{\tilde{\boldsymbol{x}}_0\}\subset\Sigma$ et

$$\sup_{t\in[t_0,+\infty[}\|\varphi(t;t_0,\tilde{\boldsymbol{x}}_0)-\boldsymbol{x}(t)\|\leqslant\varepsilon$$

où $\varphi\colon\Sigma\to\mathbb{R}^d$ est le flot de l'EDO $\boldsymbol{x}'=f(t,\boldsymbol{x}(t))$ définie en §2.1.

Pour la notion d'attractivité, nous nous intéressons au comportement de la solution perturbée quand le temps t tend vers $+\infty$. Nous souhaitons établir si la solution perturbée se rapproche infiniment de la solution non perturbée au fur et à mesure que le temps s'écoule. Si c'est le cas, la solution non perturbée sera dite attractive. Donnons une définition rigoureuse à cette notion :

Définition 2.3: Attractivité

Soit Ω un ouvert de $\mathbb{R}\times\mathbb{R}^d$. Soit $(t_0,\boldsymbol{x}_0)$ dans Ω. Soit $f\colon\Omega\to\mathbb{R}^d$ de classe $\mathcal{C}^1$. Une solution $([t_0,+\infty[,\boldsymbol{x})$ de $\boldsymbol{x}'(t)=f(t,\boldsymbol{x}(t))$, $\boldsymbol{x}(t_0)=\boldsymbol{x}_0$ est dite attractive s'il existe $\delta>0$ tel que

$$\|\tilde{\boldsymbol{x}}_0-\boldsymbol{x}_0\|\leqslant\delta\implies\lim_{t\to+\infty}\|\varphi(t;t_0,\tilde{\boldsymbol{x}}_0)-\boldsymbol{x}(t)\|=0$$

où φ est le flot de $x'=f(t,x)$ définie en §2.1.

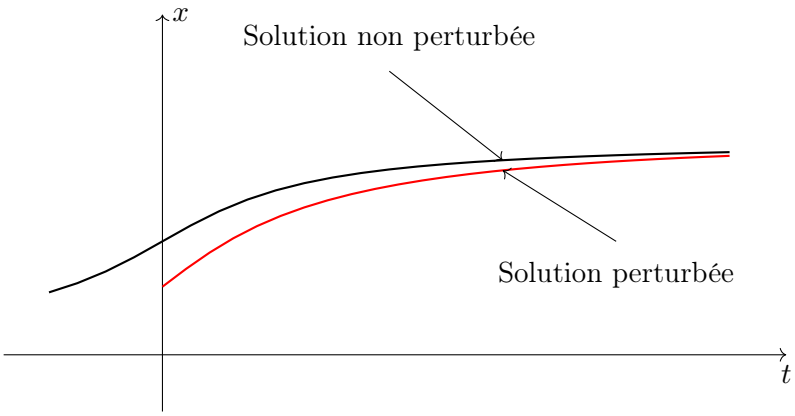


FIGURE 2.2 – Attractivité : la solution perturbée se rapproche-t-elle de la solution non perturbée pour peu que la perturbation soit suffisamment petite ?

Enfin, nous introduisons la notion de stabilité asymptotique qui n'est que la simple « intersection » des notions de stabilité au sens de Lyapounov et d'attractivité :

Définition 2.4: Stabilité asymptotique

Soit Ω un ouvert de $\mathbb{R} \times \mathbb{R}^d$. Soit $(t_0, \boldsymbol{x}_0)$ dans Ω. Soit $f : \Omega \to \mathbb{R}^d$ de classe $\mathcal{C}^1$. Une solution $([t_0, +\infty[, \boldsymbol{x})$ de $\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t))$, $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ est dite asymptotiquement stable si elle est attractive et stable au sens de Lyapounov. Elle est dite instable si elle n'est ni stable au sens de Lyapounov ni attractive.

2.3 Stabilité par étude du spectre

Avant d'énoncer notre premier théorème, nous rappelons la définition d'une valeur propre semi-simple :

Définition 2.5. *Soit* $\mathbf{A}$ *une matrice. Une valeur propre* λ *de* $\mathbf{A}$ *est dite semi-simple si et seulement si le noyau de* $(\mathbf{A} - \lambda\mathbf{I})^2$ *est égal au noyau de* $\mathbf{A} - \lambda\mathbf{I}$.

En particulier, si la matrice $\mathbf{A}$ est diagonalisable, toutes ses valeurs propres sont semi-simples.

Nous avons le théorème suivant qui lie la stabilité de la solution stationnaire nulle d'un système linéaire d'EDO à l'étude du spectre d'une matrice :

Théorème 2.6

Soit $\mathbf{A}$ dans $\mathcal{M}_d(\mathbb{R})$. La solution stationnaire $(\mathbb{R}, \mathbf{0})$ de l'EDO

$$\boldsymbol{x}' = \mathbf{A}\boldsymbol{x}.$$

est

1. Asymptotiquement stable si et seulement si toutes les valeurs propres λ_i de $\mathbf{A}$ vérifient
$$\Re(\lambda_i) < 0.$$
2. Stable au sens de Lyapounov si stable si et seulement si toutes les valeurs propres λ_i de $\mathbf{A}$ vérifient
$$\Re(\lambda_i) \leqslant 0.$$
et si toutes les valeurs propres non semi-simples de $\mathbf{A}$ vérifient
$$\Re(\lambda_i) < 0.$$
3. Instable si une des valeurs propre de $\mathbf{A}$ a une partie réelle strictement positive ou si une des valeurs propre non semi simple de $\mathbf{A}$ a une partie réelle nulle.

Démonstration. Nous allons nous contenter de donner une ébauche de la preuve. La preuve repose sur l'existence de la forme de Jordan [3], voir Annexe A.1.3. L'idée est de mettre la matrice $\mathbf{A}$ sous la forme de Jordan en employant une matrice de passage $\mathbf{P}$. Alors, $\mathbf{J} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ où la matrice $\mathbf{J}$ est sous forme de Jordan. Nous posons $\boldsymbol{y}(t) = \mathbf{P}\boldsymbol{x}(t)$, la fonction $\boldsymbol{y}$ vérifie alors $\boldsymbol{y}' = \mathbf{J}\boldsymbol{y}$. La stabilité de la solution stationnaire nulle pour l'EDO $\boldsymbol{x}' = \mathbf{A}\boldsymbol{x}$ est équivalente à la stabilité de la solution stationnaire nulle pour l'EDO $\boldsymbol{y}' = \mathbf{J}\boldsymbol{y}$. Les solutions de cette EDO sont de la forme $\boldsymbol{y}(t) = \exp(t\mathbf{J})\boldsymbol{y}_0$. Le calcul explicite de $\exp(t\mathbf{J})$ permet de conclure. □

Regardons la Proposition 2.1. Nous remarquons qu'à l'ordre 1, la perturbation d'une solution stationnaire d'une EDO autonome vérifie $\tilde{\boldsymbol{x}} - \boldsymbol{x} \approx \mathbf{J}(\tilde{\boldsymbol{x}}_0 - \boldsymbol{x}_0)$ où la matrice $\mathbf{J}$ est caractérisée par l'équation (2.1.3). Il devient alors naturel de se demander s'il existe un lien entre la stabilité de la solution nulle pour l'EDO linéaire (2.1.3a) et la stabilité de la solution stationnaire $\boldsymbol{x}_0$ pour l'EDO initiale. La réponse est positive pour les EDO autonomes et est donné par ce théorème :

3. Si vous ne connaissez pas la forme de Jordan, vous pouvez vérifier la preuve dans le cas particulier où $\mathbf{A}$ est diagonalisable.

Théorème 2.7

Soit U un ouvert de $\mathbb{R}^d$. Soit $f\colon U \to \mathbb{R}^d$ une fonction de classe $\mathcal{C}^1$. Soit $\boldsymbol{x}_0$ dans U telle que $f(\boldsymbol{x}_0) = \mathbf{0}$. Soit $\mathrm{J}f(\boldsymbol{x}_0)$ la jacobienne de f en $\boldsymbol{x}_0$.

1. Si toutes les valeurs propres λ_i de la matrice $\mathrm{J}f(\boldsymbol{x}_0)$ vérifient $\Re(\lambda_i) < 0$, alors la solution stationnaire $(\mathbb{R}, \boldsymbol{x}_0)$ de l'EDO $\boldsymbol{x}' = f(\boldsymbol{x})$ est asymptotiquement stable.
2. S'il existe une valeur propre λ_i de la matrice $\mathrm{J}f(\boldsymbol{x}_0)$ dont la partie réelle est strictement positive alors la solution stationnaire $(\mathbb{R}, \boldsymbol{x}_0)$ de l'EDO $\boldsymbol{x}' = f(\boldsymbol{x})$ est instable
3. Si toutes les valeurs propres de $\mathrm{J}f(\boldsymbol{x}_0)$ ont une partie réelle négative ou nulle et si au moins une des valeurs propre de $\mathrm{J}f(\boldsymbol{x}_0)$ a une partie réelle nulle, alors on ne peut pas conclure.

Démonstration. **La preuve est technique dans le cas général. Le lecteur pourra supposer que $\mathrm{J}f(\boldsymbol{x}_0)$ est diagonale ou diagonalisable. Ou même se placer dans le cas scalaire.** Tout d'abord, nous pouvons, sans perte de généralité, supposer que $\boldsymbol{x}_0$ est le vecteur nul $\mathbf{0}$. Pour le voir, il suffit de remplacer $\boldsymbol{x}$ par $\boldsymbol{x} - \boldsymbol{x}_0$, puis de remplacer la fonction f par $\boldsymbol{x} \mapsto f(\boldsymbol{x} + \boldsymbol{x}_0)$.

Pour faire la preuve dans le cas le plus général, il faut passer par la forme de Jordan de $\mathrm{J}f(\mathbf{0})$, voir Théorème A.10. Plus exactement, nous allons utiliser le corollaire A.11. On en déduit que pour tout $\rho > 0$, il existe une base de $\mathbb{R}^d$ dans laquelle la jacobienne est de la forme $\mathbf{D}+\mathbf{B}_\rho$ où $\mathbf{D}$ diagonale, et $\mathbf{B}_\rho$ triangulaire inférieure vérifiant $\|\mathbf{B}_\rho\|_2 \leqslant \rho$. La matrice diagonale $\mathbf{D}$ a comme éléments diagonaux les valeurs propres λ_i de $\mathrm{J}f(\mathbf{0})$. Soit $\mathbf{P}_\rho$ la matrice de passage tel que

$$\mathbf{D} + \mathbf{B}_\rho = \mathbf{P}_\rho^{-1} \mathrm{J}f(\mathbf{0}) \mathbf{P}_\rho.$$

On pose $\boldsymbol{y}(t) = \mathbf{P}_\rho^{-1}\boldsymbol{x}(t)$, alors $\boldsymbol{y}$ est solution de $\boldsymbol{y}' = g_\rho(\boldsymbol{y})$ avec $g(\boldsymbol{y}) = \mathbf{P}_\rho^{-1} f(\mathbf{P}\boldsymbol{y})$. De plus, $\mathbf{0}$ est aussi solution stationnaire de l'équation $\boldsymbol{y}' = g_\rho(\boldsymbol{y})$. Et la solution stationnaire $\mathbf{0}$ est stable au sens de Lyapounov, respectivement attractive, pour l'EDO $\boldsymbol{x}' = f(\boldsymbol{x})$ si et seulement si $\mathbf{0}$ est stable au sens de Lyapounov, respectivement attractive, pour l'EDO $\boldsymbol{y}' = g_\rho(\boldsymbol{y})$. Par construction, $\mathrm{J}g_\rho(\boldsymbol{y}_0) = \mathbf{D} + \mathbf{B}_\rho$ a les mêmes valeurs propres que $\mathrm{J}f(\mathbf{0})$.

Nous allons maintenant distinguer les deux cas du théorème. Supposons maintenant que toutes les valeurs propres λ_i de la matrice $\mathrm{J}f(\mathbf{0})$ aient une partie réelle strictement négatives. Nous posons $\rho = \min_i(-\Re(\lambda_i))/3$. Pour éviter d'alourdir les notations, l'indice ρ dans $\mathbf{B}_\rho$ et dans g_ρ sera élidé : nous écrirons simplement $\mathbf{B}$ et g. Soit $\eta > 0$ tel que pour tout $\boldsymbol{y}$ vérifiant $\|\boldsymbol{y}\|_2 \leqslant \eta$:

$$\|g(\boldsymbol{y}) - \mathrm{J}g(\mathbf{0})\boldsymbol{y}\|_2 \leqslant \rho\|y\|_2.$$

Soit $\boldsymbol{y}_0$ dans $\mathbb{R}^d$ tel que $\|\boldsymbol{y}_0\|_2 < \eta$. Soit $\boldsymbol{y}$ la solution du problème de Cauchy

$$\boldsymbol{y}'(t) = g(\boldsymbol{y}(t)), \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0.$$

avec $\|\boldsymbol{y}_0\| < \eta$. Posons alors

$$t^* = \sup\{t > 0 : \forall s < t,\ \|\boldsymbol{y}(s)\| < \eta\}.$$

On a $t^* > 0$. Faisons alors le produit scalaire de l'EDO $\boldsymbol{y}'(t) = g_\rho(\boldsymbol{y}(t))$ par $\overline{\boldsymbol{y}(t)}$ et prenons en la partie réelle. Nous obtenons pour tout t dans $[0, t^*[$:

$$\begin{aligned}
\frac{\mathrm{d}\|\boldsymbol{y}(t)\|_2^2}{\mathrm{d}t} &= 2\Re\big(\boldsymbol{y}^\top(t)(\mathbf{D}+\mathbf{B})\boldsymbol{y}(t)\big) + 2\Re\big((g(\boldsymbol{y}(t)) - \mathrm{J}g(\mathbf{0})\boldsymbol{y}(t))\cdot\boldsymbol{y}(t)\big),\\
&= 2\sum_{i=1}^{d}\Re(\lambda_i)|y_i(t)|^2 + 2\Re\big(\boldsymbol{y}^\top(t)\mathbf{B}\boldsymbol{y}(t)\big) + 2\Re\big((g(\boldsymbol{y}(t)) - \mathrm{J}g(\mathbf{0})\boldsymbol{y}(t))\cdot\boldsymbol{y}(t)\big),\\
&\leqslant -2\min_i(-\Re(\lambda_i))\|\boldsymbol{y}\|_2^2 + 4\rho\|\boldsymbol{y}\|_2^2,\\
&\leqslant -2\rho\|\boldsymbol{y}\|_2^2.
\end{aligned}$$

Cela implique que $t \mapsto \|\boldsymbol{y}(t)\|_2$ est décroissante. Donc $t^* = +\infty$ et $\mathbf{0}$ est une solution stationnaire stable au sens de Lyapounov de l'EDO $\boldsymbol{y}' = g(\boldsymbol{y})$. De plus, par le lemme de Grönwall 1.5 :

$$\|\boldsymbol{y}(t)\|^2 \leqslant \|\boldsymbol{y}_0\|^2 \exp(-2\rho t).$$

Donc, $\mathbf{0}$ est une solution stationnaire attractive de l'EDO $\boldsymbol{y}' = g(\boldsymbol{y})$. Pour le deuxième cas, lorsque une des valeurs propres λ_i a une partie réelle strictement positive, il faut procéder selon le même principe. □

2.4 Fonctions de Lyapounov

Les fonctions de Lyapounov permettent d'étudier la stabilité des solutions d'une équation différentielle. Une fonction de Lyapounov est une fonction $\mathcal{L}$ de $(t, \boldsymbol{x})$ telle que si $\boldsymbol{x}$ est solution de l'équation différentielle (2.0.1a), alors $t \mapsto \mathcal{L}(t, \boldsymbol{x}(t))$ est décroissante. Les fonctions de Lyapounov représentent ce que les physiciens appellent des « puits de potentiel ».

Soit $\mathcal{L}$ une fonction de classe $\mathcal{C}^1$ définie sur $\mathbb{R}\times\mathbb{R}^d$ à valeurs dans $\mathbb{R}$. Soit $\boldsymbol{x}$ vérifiant $\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t))$. Nous avons en appliquant la règle de dérivation en chaîne :

$$\frac{\mathrm{d}\mathcal{L}(t,\boldsymbol{x}(t))}{\mathrm{d}t} = \frac{\partial\mathcal{L}}{\partial t}(t,\boldsymbol{x}(t)) + \nabla_{\boldsymbol{x}}\mathcal{L}(t,\boldsymbol{x}(t))\cdot f(t,\boldsymbol{x}(t)). \tag{2.4.1}$$

où $\nabla_{\boldsymbol{x}}\mathcal{L}(t,\boldsymbol{x})$ est le gradient par rapport aux composantes de $\boldsymbol{x}$, *i.e.*, le vecteur $[\frac{\partial\mathcal{L}}{\partial x_1},\dots,\frac{\partial\mathcal{L}}{\partial x_d}]^\top$.

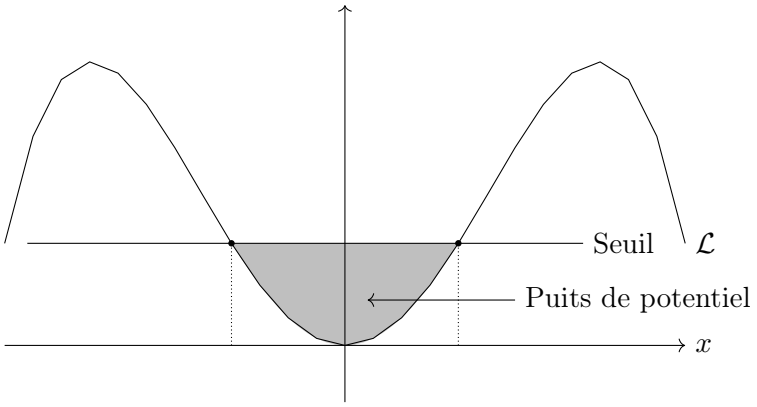


FIGURE 2.3 – Exemple de fonction de Lyapounov ne dépendant pas de t.

Définition 2.8: Fonctions de Lyapounov

Soit f une fonction de classe $\mathcal{C}^1$ définie sur $\mathbb{R}\times\mathbb{R}^d$. Soit $\boldsymbol{x}_0$ dans $\mathbb{R}^d$ tel que pour tout t dans $\mathbb{R}$, $f(t,\boldsymbol{x}_0)=0$. On dit que $\mathcal{L}$ est une fonction de Lyapounov pour la solution stationnaire $\boldsymbol{x}_0$ de l'EDO $\boldsymbol{x}'=f(t,\boldsymbol{x}(t))$ si les trois conditions suivantes sont vérifiées :

1. La fonction $\mathcal{L}$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}\times\mathbb{R}^d$ à valeur dans $\mathbb{R}$.
2. Pour tout $(t,\boldsymbol{x})$ dans $\mathbb{R}\times\mathbb{R}^d$:

$$\frac{\partial\mathcal{L}}{\partial t}(t,\boldsymbol{x})+\nabla_{\boldsymbol{x}}\mathcal{L}(t,\boldsymbol{x})\cdot f(t,\boldsymbol{x})\leqslant 0. \tag{2.4.2a}$$

   Ce qui est équivalent à montrer que pour toute solution $\boldsymbol{x}$ de l'EDO $\boldsymbol{x}'=f(t,\boldsymbol{x}(t))$:

$$\frac{\mathrm{d}\mathcal{L}(t,\boldsymbol{x}(t))}{\mathrm{d}t}\leqslant 0 \tag{2.4.2b}$$

3. Il existe $g\colon\mathbb{R}^d\to\mathbb{R}$ continue et $\delta>0$ telle que pour tout t dans $\mathbb{R}$, pour tout $\boldsymbol{x}$ dans $B(\boldsymbol{x}_0,\delta)\backslash\{\boldsymbol{x}_0\}$:

$$\mathcal{L}(t,\boldsymbol{x})\geqslant g(\boldsymbol{x})>g(\boldsymbol{x}_0)=\mathcal{L}(t,\boldsymbol{x}_0).$$

L'existence d'une fonction de Lyapounov est suffisante pour obtenir la stabilité au sens de Lyapounov :

Théorème 2.9

Soit f une fonction de classe $\mathcal{C}^1$ définie sur $\mathbb{R}\times\mathbb{R}^d$. Soit $\boldsymbol{x}_0$ dans $\mathbb{R}^d$ tel que pour tout t dans $\mathbb{R}$, $f(t,\boldsymbol{x}_0)=0$. Si la solution stationnaire

$\boldsymbol{x}_0$ de l'EDO $\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t))$ admet une fonction de Lyapounov alors $(\mathbb{R}, \boldsymbol{x}_0)$ est solution stationnaire stable au sens de Lyapounov de l'EDO $\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t))$.

Démonstration. Soit $\varepsilon > 0$ suffisamment petit pour que $\boldsymbol{x}_0$ soit un minimum strict de g sur la boule de centre $\boldsymbol{x}_0$ et de rayon ε.

$$M = \inf_{\boldsymbol{y}\in\mathbb{R}^{\boldsymbol{d}}:\|\boldsymbol{y}-\boldsymbol{x}_0\|=\varepsilon} (g(\boldsymbol{y})).$$

Comme g est continue, M est atteint sur la sphère de centre $\boldsymbol{x}_0$ et de rayon ε. Donc $M > g(\boldsymbol{x}_0)$. Comme $\mathcal{L}(t_0, \boldsymbol{x}_0) = g(\boldsymbol{x}_0) < M$, et que $\mathcal{L}$ est continue, il existe $\delta > 0$, $\delta < \varepsilon$ tel que pour tout $\tilde{\boldsymbol{x}}_0$ dans la boule ouverte de centre $\boldsymbol{x}_0$ et de rayon δ, $\mathcal{L}(t_0, \tilde{\boldsymbol{x}}_0) < M$. Donc, pour tout $\tilde{\boldsymbol{x}}_0$ dans la boule ouverte de centre $\boldsymbol{x}_0$ et de rayon δ, pour tout t dans $[t_0, +\infty[$

$$g(\varphi(t; t_0, \tilde{\boldsymbol{x}}_0)) \leqslant \mathcal{L}(t, \varphi(t; t_0, \tilde{\boldsymbol{x}}_0)) \leqslant \mathcal{L}(t_0, \tilde{\boldsymbol{x}}_0) < M.$$

Donc, pour tout $t \geqslant t_0$, $\varphi(t; t_0, \tilde{\boldsymbol{x}}_0)$ est dans la boule de centre $\boldsymbol{x}_0$ et de rayon ε. En effet, si ce n'était pas le cas, il existerait $\hat{t}$ tel que $\varphi(\hat{t}; t_0, \tilde{\boldsymbol{x}}_0)$ appartienne à la sphère de centre $\boldsymbol{x}_0$ et de rayon ε et alors par définition de M, nous aurions $g(\varphi(\hat{t}; t_0, \tilde{\boldsymbol{x}}_0)) \geqslant M$, ce qui serait une contradiction. □

Il est aussi possible d'utiliser cette méthode pour étudier l'attractivité d'une solution stationnaire à une EDO. Pour des raisons de simplicité dans l'énoncé du théorème, nous nous restreignons dans ce cas aux EDO autonomes.

Définition 2.10: Fonctions de Lyapounov strictes

Soit f une fonction de classe $\mathcal{C}^1$ définie sur $\mathbb{R}^d$. Soit $\boldsymbol{x}_0$ dans $\mathbb{R}^d$ tel que $f(\boldsymbol{x}_0) = 0$. On dit que $\mathcal{L}$ est une fonction de Lyapounov stricte pour la solution stationnaire $\boldsymbol{x}_0$ de l'EDO $\boldsymbol{x}' = f(\boldsymbol{x}(t))$ si les trois conditions suivantes sont vérifiées :

1. La fonction $\mathcal{L}$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}^d$ à valeur dans $\mathbb{R}$.
2. Pour tout $\boldsymbol{x}$ dans $\mathbb{R}^d \backslash \{\boldsymbol{x}_0\}$:
$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}) \cdot f(\boldsymbol{x}) < 0.$$
3. Il existe $\delta > 0$ tel que pour tout $\boldsymbol{x}$ dans $B(\boldsymbol{x}_0, \delta) \backslash \{\boldsymbol{x}_0\}$:
$$\mathcal{L}(\boldsymbol{x}) > \mathcal{L}(\boldsymbol{x}_0).$$

Si les fonctions de Lyapounov permettent de conclure quant à la stabilité au sens de Lyapounov, les fonctions de Lyapounov strictes permettent de conclure quant à la stabilité asymptotique :

Théorème 2.11

Soit f une fonction de classe $\mathcal{C}^1$ définie sur $\mathbb{R}^d$. Soit $\boldsymbol{x}_0$ dans $\mathbb{R}^d$ tel que $f(\boldsymbol{x}_0) = 0$. Si la solution stationnaire $\boldsymbol{x}_0$ de l'EDO $\boldsymbol{x}'(t) = f(\boldsymbol{x}(t))$ admet une fonction de Lyapounov stricte alors $(\mathbb{R}, \boldsymbol{x}_0)$ est solution asymptotiquement stable de l'EDO $\boldsymbol{x}'(t) = f(\boldsymbol{x}(t))$.

Démonstration. Les fonctions de Lyapounov strictes étant des fonctions au sens de Lyapounov, nous avons par le théorème 2.9 que la solution stationnaire $\boldsymbol{x}_0$ est stable au sens de Lyapounov. Il reste à montrer qu'elle est aussi attractive. Comme elle est stable au sens de Lyapounov, il existe $\hat{\delta} > 0$ tel que pour tout $\tilde{\boldsymbol{x}}$ dans la boule ouverte de centre $\boldsymbol{x}_0$ et de rayon $\hat{\delta}$, l'unique solution du problème de Cauchy perturbé $\tilde{\boldsymbol{x}}'(t) = f(\tilde{\boldsymbol{x}}(t))$, $\tilde{\boldsymbol{x}}(t_0) = \tilde{\boldsymbol{x}}_0$ satisfait

$$\sup_{t \geqslant 0} \|\tilde{\boldsymbol{x}}(t) - \boldsymbol{x}_0\|_{\mathbb{R}^d} < \delta/2.$$

On suppose dorénavant que $\tilde{\boldsymbol{x}}_0$ est toujours dans la boule ouverte de centre $\boldsymbol{x}_0$ et de rayon $\hat{\delta}$.

$$\tilde{\boldsymbol{x}}'(t) = f(\tilde{\boldsymbol{x}}(t)), \qquad \tilde{\boldsymbol{x}}(t_0) = \tilde{\boldsymbol{x}}_0.$$

Alors, l'application, $t \mapsto \mathcal{L}(\tilde{\boldsymbol{x}}(t))$ est décroissante et minorée par $\mathcal{L}(\boldsymbol{x}_0)$ donc convergente. Raisonnons par l'absurde. Si la limite est différente de $\mathcal{L}(\boldsymbol{x}_0)$ alors comme $\mathcal{L}$ est continue en $\boldsymbol{x}_0$, il existe $\eta > 0$, $\eta \leqslant \delta/2$ tel que pour tout $t > 0$ $\tilde{\boldsymbol{x}}(t)$ vérifie

$$\eta \leqslant \|\tilde{\boldsymbol{x}}(t) - \boldsymbol{x}_0\| \leqslant \delta/2.$$

On a alors pour tout $t \geqslant t_0$

$$\frac{\mathrm{d}\mathcal{L}(\tilde{\boldsymbol{x}}(t))}{\mathrm{d}t} \leqslant \sup_{\eta \leqslant \|\boldsymbol{x} - \boldsymbol{x}_0\| \leqslant \delta/2} (\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}) \cdot f(\boldsymbol{x})).$$

La borne supérieure est atteinte car elle est prise sur un compact et car $\mathcal{L}$ est continue. Notons $-r$ la valeur de cette borne supérieure. Nous avons $r > 0$ et

$$\mathcal{L}(\tilde{\boldsymbol{x}}(t)) \leqslant \mathcal{L}(\tilde{\boldsymbol{x}_0}) - r(t - t_0)$$

ce qui implique que

$$\lim_{t \to +\infty} \mathcal{L}(\tilde{\boldsymbol{x}}(t)) = -\infty.$$

C'est impossible. Donc la limite de l'application $t \mapsto \mathcal{L}(\tilde{\boldsymbol{x}}(t))$ quand t tend vers $+\infty$ est $\mathcal{L}(\boldsymbol{x}_0)$. Comme $\mathcal{L}(\boldsymbol{x}_0)$ est un minimum local strict de $\mathcal{L}$, nous en déduisons que $\tilde{\boldsymbol{x}}(t)$ converge vers $\boldsymbol{x}_0$ quand t tend vers $+\infty$. □

L'énergie mécanique d'un système physique est souvent une fonction de Lyapounov. Ses minima locaux stricts seront des points d'équilibres stables.

Exemple 2.4.1. Soit ϕ un potentiel, *i.e.* une fonction de $\mathbb{R}^3$ à valeurs dans $\mathbb{R}$. Soit $\boldsymbol{x}_0$ un minimum local strict de ϕ. Considérons alors l'EDO :

$$m\boldsymbol{x}''(t) = -\nabla\phi(\boldsymbol{x}(t)).$$

Nous pouvons mettre cette EDO sous forme d'un système d'ordre 1 :

$$\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{p}\end{bmatrix}' = \begin{bmatrix}\boldsymbol{p}/m\\ -\nabla\phi(\boldsymbol{x})\end{bmatrix}$$

Alors la fonction $\mathcal{L}$ définie par

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{p}) = \frac{\|\boldsymbol{p}\|^2}{2m} + \phi(\boldsymbol{x}).$$

est une fonction de Lyapounov pour le point d'équilibre $(\boldsymbol{x}_0, \mathbf{0})$: position $\boldsymbol{x}_0$, vitesse nulle à l'instant initiale.

Quand la fonction de Lyapounov est une valeur conservatrice, *i.e.*, est constante au cours du temps, et ne dépend pas explicitement du temps, alors la solution stationnaire est stable au sens de Lyapounov mais n'est pas attractive.

Théorème 2.12. *Soit f une fonction de classe $\mathcal{C}^1$ définie sur $\mathbb{R}^d$. Soit $\boldsymbol{x}_0$ dans $\mathbb{R}^d$ tel que $f(\boldsymbol{x}_0) = 0$. S'il existe une fonction de Lyapounov $\mathcal{L}$ pour la solution stationnaire $\boldsymbol{x}_0$ de l'EDO $\boldsymbol{x}' = f(\boldsymbol{x}(t))$, conservatrice, et ne dépendant pas explicitement du temps, i.e., une fonction $\mathcal{L}$ vérifiant :*

1. *La fonction $\mathcal{L}$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}^d$ à valeurs dans $\mathbb{R}$.*
2. *Pour tout $\boldsymbol{x}$ dans $\mathbb{R}^d\backslash\{\boldsymbol{x}_0\}$:*

$$\nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}) \cdot f(\boldsymbol{x}) = 0.$$

3. *Il existe $\delta > 0$ tel que pour tout $\boldsymbol{x}$ dans $B(\boldsymbol{x}_0,\delta)\backslash\{\boldsymbol{x}_0\}$:*

$$\mathcal{L}(\boldsymbol{x}) > \mathcal{L}(\boldsymbol{x}_0).$$

Alors, $\boldsymbol{x}_0$ est une solution stationnaire, stable au sens de Lyapounov, mais non attractive de l'EDO $x'(t) = f(x(t))$.

Démonstration. Comme les hypothèses du théorème 2.9 sont vérifiées, $\boldsymbol{x}_0$ est une solution stationnaire, stable au sens de Lyapounov. Si $\tilde{\boldsymbol{x}}_0$ appartient à $B(\boldsymbol{x}_0,\delta)\backslash\{\boldsymbol{x}_0\}$, alors $\mathcal{L}(\tilde{\boldsymbol{x}}_0) > \mathcal{L}(\boldsymbol{x}_0)$. Donc, pour tout $t > t_0$:

$$\mathcal{L}(\tilde{\boldsymbol{x}}(t)) = \mathcal{L}(\tilde{\boldsymbol{x}}_0) > \mathcal{L}(\boldsymbol{x}_0).$$

Or, $\mathcal{L}$ est continue. Donc, il est impossible que $\tilde{\boldsymbol{x}}(t)$ converge vers $\boldsymbol{x}_0$ quand t tend vers $+\infty$. □

Conclusion

Dans ce chapitre, nous avons donné deux méthodes pour étudier la stabilité des solutions stationnaires d'une EDO : l'étude du spectre du linéarisé et les fonctions de Lyapounov. Pour un problème non linéaire, seules les fonctions de Lyapounov permettent de démontrer la stabilité au sens de Lyapounov sans démontrer la stabilité asymptotique. Pour une équation différentielle provenant de la physique ou de la mécanique, on peut utiliser l'énergie mécanique ou une autre grandeur conservée comme fonction de Lyapounov.

Références

[1] Serge Lang. *Real and Functional Analysis*. Third. Springer, 1993.

[2] Lev S. Pontryagin. *Ordinary Differential Equations*. Addison-Wesley, 1962.

[3] David A. Sanchez. *Ordinary Differential Equations and Stability Theory : an Introduction*. Dover Publications, Inc., 1968.

Chapitre 3

Résolution numérique d'une équation différentielle

Les équations différentielles étant couramment utilisées pour modéliser le comportement de systèmes physiques, mécaniques ou autres, il est important de pouvoir prévoir le comportement des solutions de ces équations différentielles. Nous avons vu que dans certains cas, il est possible de donner la solution exacte mais ce cas de figure reste limité aux équations différentielles simples et académiques.

La plupart du temps, il est difficile, voire impossible d'obtenir les solutions exactes d'une équation différentielle. Dans ce cas, pour pouvoir observer le comportement desdites solutions, on peut recourir à la résolution numériques des EDO. C'est l'objet de ce chapitre.

3.1 Discrétisation d'une équation différentielle

Nous souhaitons procéder à une simulation numérique pour obtenir une solution numérique du problème de Cauchy suivant :

$$\boldsymbol{x}'(t) = \boldsymbol{f}(t, \boldsymbol{x}(t)), \tag{3.1.1a}$$

$$\boldsymbol{x}(a) = \boldsymbol{x}_0. \tag{3.1.1b}$$

où $\boldsymbol{f} \colon \mathbb{R} \times \mathbb{R}^d \supset \Omega \to \mathbb{R}^d$.

La solution recherchée est une fonction. Un ordinateur ne pouvant traiter que des données discrètes, il faut d'abord discrétiser l'équation différentielle. Pour simuler numériquement la solution $\boldsymbol{x}$ sur l'intervalle $[a, b]$, on commence par se donner une suite réelle $(t_k)_{1 \leqslant k \leqslant n}$ vérifiant

$$a = t_0 < t_1 < t_2 < \ldots < t_n < \ldots < t_N = b,$$

La quantité $h_n = t_{n+1} - t_n$ est appelée n^{e} pas de temps. Si tous les h_n sont égaux, on parlera de pas constant. Dans le cas contraire, il s'agit de pas variable.

Notre but est d'obtenir une suite finie $(\boldsymbol{x}_n^h)_{0\leqslant n\leqslant N}$ appartenant à $(\mathbb{R}^d)^{N+1}$ vérifiant

$$\boldsymbol{x}_k^h \approx \boldsymbol{x}(t_k).$$

Pour ce faire, on cherche une relation de récurrence qui nous donne $\boldsymbol{x}_{n+1}^h$ en fonction des $\boldsymbol{x}_k^h$, des t_k pour $k \leqslant n$, et de h_n. Une telle relation de récurrence est appelée schéma ou méthode (de résolution numérique). Lorsque $\boldsymbol{x}_{n+1}^h$ dépend uniquement de $\boldsymbol{x}_n$, de t_n et de h_n, on parlera de méthode à un pas. Si $\boldsymbol{x}_{n+1}^h$ dépend de $\boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \ldots, \boldsymbol{x}_{n-m+1}$, de $t_n, t_{n-1}, \ldots, t_{n-m+1}$ et de h_n, on parlera de méthode à m pas, voir §3.4 pour les méthodes multipas.

Dans cette section, nous introduisons deux méthodes à un pas : la méthode d'Euler explicite et la méthode d'Euler implicite.

3.1.1 Les méthodes d'Euler

Algorithme 3.1.1: Méthode d'Euler-Explicite

Entrées :

Fonction f.
Condition initiale $\boldsymbol{x}_0$, t_0.
Temps initial t_0.
Temps final t_f.
Nombre de pas de temps N.

Algorithme :

$h := (t_f - t_0)/N$.
Pour n allant de 0 à $N-1$:
$\boldsymbol{x}_{n+1} := \boldsymbol{x}_n + hf(t_n, \boldsymbol{x}_n)$
Fin Pour

Il existe deux méthodes d'Euler. La méthode d'Euler explicite et la méthode d'Euler implicite.

```
def EulerExplicite(f,t0,x0,tf,h) :
    nbiter=int(math.ceil((tf-t0)/h))
    h=(tf-t0)/nbiter
    x=x0;
    t=t0;
    for i in range(0,nbiter) :
        x=x+h*f(t,x);
        t=t+h
    return x;
```

Code 3.1 – Code Euler Explicite en Python

La méthode d'Euler explicite est la plus simple des méthodes de résolution numérique d'une EDO. On l'obtient à partir du développement de Taylor d'ordre 1 de $\boldsymbol{x}$ en t_n. Si $\boldsymbol{f}$ est de classe $\mathcal{C}^2$

$$\begin{aligned}\boldsymbol{x}(t_{n+1}) &= \boldsymbol{x}(t_n) + h_n\boldsymbol{x}'(t_n) + O(|h_n|^2)\\ &= \boldsymbol{x}(t_n) + h_n\boldsymbol{f}(t_n, \boldsymbol{x}(t_n)) + O(|h_n|^2)\\ &\approx \boldsymbol{x}(t_n) + h_n\boldsymbol{f}(t_n, \boldsymbol{x}(t_n)).\end{aligned} \tag{3.1.2}$$

Ce qui nous permet d'introduire :

Définition 3.1 (Méthode/Schéma d'Euler explicite)**.**

$$\boldsymbol{x}^h_{n+1} = \boldsymbol{x}^h_n + (t_{n+1} - t_n)\boldsymbol{f}(t_n, \boldsymbol{x}^h_n). \tag{3.1.3}$$

La méthode d'Euler explicite est comme son nom l'indique une méthode explicite. En effet, la valeur de $\boldsymbol{x}^h_{n+1}$ est donnée par une formule explicite. Géométriquement, voir figure 3.1, le schéma d'Euler explicite revient à suivre la tangente à l'unique solution exacte à (3.1.1a) passant par $t_n, \boldsymbol{x}^h_n$. La méthode d'Euler-explicite à pas constant en pseudo-code est donnée à l'algorithme 3.1.1. Une implémentation en Python de la méthode d'Euler Explicite est donnée au Code 3.1.

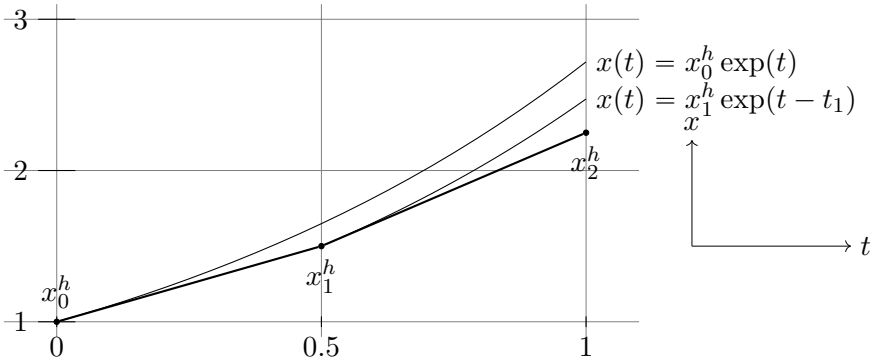


FIGURE 3.1 – Euler Explicite à $x' = x, x(0) = 1$ avec pas $h = 1/2$

Pour obtenir la méthode d'Euler implicite, on effectue le développement de Taylor de la solution exacte $\boldsymbol{x}$ à l'ordre 1 au point t_{n+1}

$$\boldsymbol{x}(t_{n+1}) = \boldsymbol{x}(t_n) + (t_{n+1} - t_n)f(t_{n+1}, \boldsymbol{x}(t_{n+1})) + O(|h_n|^2). \tag{3.1.4}$$

Si on oublie le $O(|h_n|^2)$ et on remplace $\boldsymbol{x}(t_n)$ par $\boldsymbol{x}^h_n$, on obtient la méthode d'Euler implicite :

Définition 3.2 (Schéma/Méthode d'Euler implicite)**.**

$$\boldsymbol{x}^h_{n+1} = \boldsymbol{x}^h_n + h_n f(t_{n+1}, \boldsymbol{x}^h_{n+1}). \tag{3.1.5}$$

Algorithme 3.1.2: Méthode d'Euler-Implicite

Entrées :

Fonction f.
Condition initiale $\boldsymbol{x}_0$, t_0.
Temps initial t_0.
Temps final t_f.
Nombre de pas de temps N.

Algorithme :

$h := (t_f - t_0)/N$.
Pour n allant de 0 à $N-1$:
On pose $\boldsymbol{x}_{n+1}$ solution de $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + hf(t_{n+1}, \boldsymbol{x}_{n+1})$
Fin Pour

La méthode d'Euler implicite est comme son nom l'indique une méthode implicite. En effet pour pouvoir avancer d'un pas de temps et calculer $\boldsymbol{x}_{n+1}$ à partir de $(\boldsymbol{x}_n, t_n, h_n)$, il faut résoudre une équation. En pratique, cette résolution se fait numériquement. Pour cela, on peut utiliser une méthode de point fixe ou une méthode de Newton (voir Cours d'analyse numérique). La méthode d'Euler Implicite a été écrite sous forme de pseudo-code à l'Algorithme 3.1.2.

Une implémentation en Python de la méthode d'Euler Implicite est donnée au Code 3.2. Dans cette implémentation, nous avons utilisé la méthode de Newton pour résoudre l'équation (3.1.3). Nous aurions aussi pu utiliser une méthode de point fixe ou une autre méthode de résolution d'une équation. Dans une implémentation robuste, il faudrait au minimum laisser le choix à l'utilisateur de la méthode de résolution et des différents paramètres de tolérances. Il faudrait aussi mieux traiter les échecs de la méthode de résolution. Dans cette implémentation, nous nous contentons d'imprimer un message d'erreur, ce qui ne serait pas acceptable dans un code devant être robuste. La gestion acceptable des erreurs dans un programme étant à la fois plutôt difficile, et plus un problème informatique qu'un problème algorithmique, nous renvoyons le lecteur intéressé à des ouvrages spécialisés.

Géométriquement, dans la méthode d'Euler implicite, le point $(t_n, \boldsymbol{x}_n)$ est sur la tangente à la solution exacte de (3.1.1a) passant par $(t_{n+1}, \boldsymbol{x}_{n+1})$, voir figure 3.2.

Pour une trajectoire fermée, comme par exemple la solution de l'équation

$$x' = -y, \qquad y' = x,$$

les méthodes d'Euler ne sont pas du tout satisfaisantes. En effet, la trajectoire des solutions exacte à cette équation est toujours un cercle. Les trajectoires de la solution exacte et des solutions numériques sont représentées sur la Figure 3.3. La solution numérique calculée par la méthode d'Euler explicite

```
import numpy as np;
import numpy.linalg as LA;
def ImplicitEuler(f,fJac,t0,x0,tf,h,**kwargs) :
    x=x0;
    t=t0;
    tolfactor=1e-4;
    maxiternewton=1000;
    alpha=math.sqrt(np.finfo(float).eps);
    sz=np.size(np.asarray(x));
    x=np.asarray(x).reshape(sz);
    A=np.zeros([sz,sz]);
    for i in range(0,nbiter) :
        xm1=x;
        tol=tolfactor*h*LA.norm(f(t,xm1));
        t=t+h;
        k=0;
        condition=True;
        while condition:
            A=-h*fJac(t,x);
            for l in range(0,sz) :
                A[l,l]=A[l,l]+1;
            ## Newton Iterate
            xNewtonm1=x;
            x=x-LA.solve(A,x-xm1-h*f(t,x));
            k=k+1;
            condition = ((k<maxiternewton) and \
                                (np.linalg.norm(x-xNewtonm1)>tol));
        if np.linalg.norm(x-xNewtonm1)>tol :
            ##Terrible error handling
            print('Newton failure. tol=',
                      np.linalg.norm(x-xNewtonm1),' > ',
                      tol, 'i=', i, 'nbtiter=', nbiter);
    if sz==1 :
        return x[0];
    else :
        return x;
```

Code 3.2 – Code Euler Implicite en Python

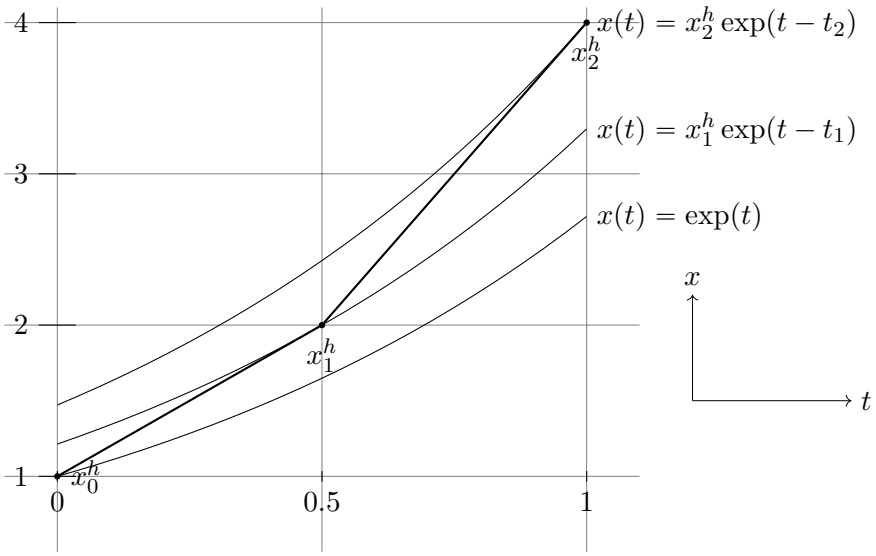


FIGURE 3.2 – Euler Implicite à $x' = x, x(0) = 1$ avec pas $h = 1/2$

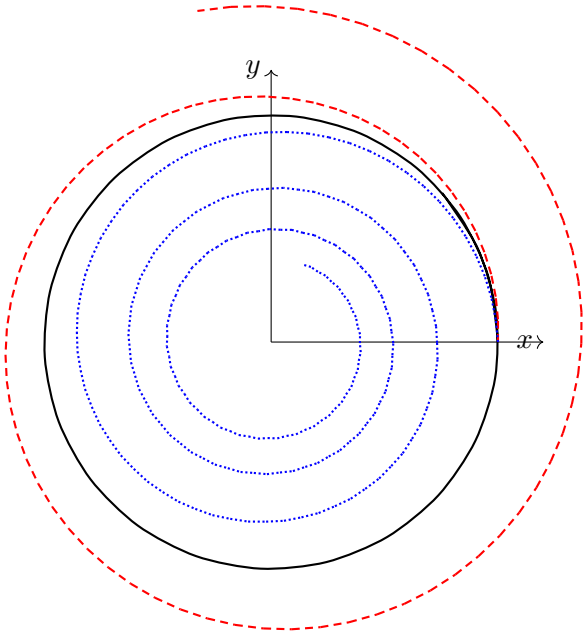


FIGURE 3.3 – Comparaison trajectoire solution exacte (cercle noir trait continu), solution Euler Explicite (spirale divergente, tirets rouge) et Euler implicite (spirale convergent, pointillées bleu) pour $z' = iz$, $z(0) = 1$ avec $h = 0.1$ et $T_f = 8$.

est une spirale qui part vers l'infini. La solution numérique calculée par la méthode d'Euler explicite est une spirale qui s'écrase vers le centre du cercle. Ces deux phénomènes se produisent très rapidement.

3.1.2 Ordre, consistance et analyse de l'erreur

Soit $\boldsymbol{x}'(t) = f(t, \boldsymbol{x})$ une EDO avec $f\colon \mathbb{R} \times \mathbb{R}^d \supset \Omega \to \mathbb{R}^d$ régulière. Une méthode de résolution numérique à 1 pas est définie par la donnée d'une fonction $\Phi_f\colon \Omega \times \mathbb{R} \to \mathbb{R}^d$. La méthode à un pas est alors donnée par la formule de récurrence

$$\boldsymbol{x}_{n+1}^h = \Phi_f(t_n, \boldsymbol{x}_n^h, h).$$

Si Φ_f est connue explicitement, on parlera de méthode explicite. Si on ne connait pas Φ_f explicitement mais qu'on ne connait qu'une fonction Ψ_f telle que $\Psi_f(\Phi_f(t, \boldsymbol{x}, h), t, \boldsymbol{x}, h) = 0$, on parlera de méthode implicite car à chaque étape, pour calculer $\boldsymbol{x}_{n+1}^h$, il faudra résoudre

$$\Psi_f(\boldsymbol{x}_{n+1}^h, t_n, \boldsymbol{x}_n^h, h) = 0.$$

Comme au §2.1, on note $\varphi_f(t; t_0, \boldsymbol{x}_0)$ la valeur en t de l'unique solution exacte à l'EDO $\boldsymbol{x}'(t) = f(t, \boldsymbol{x})$ qui vaut x_0 en t_0.

Nous pouvons maintenant donner les définitions de consistance d'une méthode numérique.

Définition 3.3: Consistance

Une méthode Φ_f est dite consistante si et seulement si pour tout $(t, \boldsymbol{x})$ dans Ω

$$\|\Phi_f(t, \boldsymbol{x}, h) - \varphi_f(t + h; t, \boldsymbol{x})\| = o(|h|).$$

Les méthodes d'Euler implicite et d'Euler explicite sont consistantes. Le minimum que l'on puisse demander à une méthode de résolution numérique d'une EDO est d'être consistante mais c'est rarement suffisant.

Pour aller au delà de la notion de consistance, nous introduisons la notion d'ordre d'une méthode numérique.

Définition 3.4: Ordre

Une méthode Φ_f est dite d'ordre p si et seulement si pour tout $(t, \boldsymbol{x})$ dans Ω

$$\|\Phi_f(t, \boldsymbol{x}, h) - \varphi_f(t + h; t, \boldsymbol{x})\| = O(|h|^{p+1}).$$

Les méthodes d'Euler explicite et d'Euler implicite sont d'ordre 1.

Nous allons maintenant introduire deux méthodes d'ordre 2. Commençons par effectuer le développement de Taylor de $\boldsymbol{x}$ en $t + h/2$ à l'ordre 2.

Nous avons

$$\boldsymbol{x}(t+h) = \boldsymbol{x}(t+\frac{h}{2}) + \frac{h}{2}\boldsymbol{x}'(t+\frac{h}{2}) + \frac{h^2}{8}\boldsymbol{x}''(t+\frac{h}{2}) + O(h^3)$$
$$\boldsymbol{x}(t) = \boldsymbol{x}(t+\frac{h}{2}) - \frac{h}{2}\boldsymbol{x}'(t+\frac{h}{2}) + \frac{h^2}{8}\boldsymbol{x}''(t+\frac{h}{2}) + O(h^3)$$

Nous ajoutons les deux quantités et obtenons que

$$\boldsymbol{x}(t+\frac{h}{2}) = \frac{\boldsymbol{x}(t)+\boldsymbol{x}(t+h)}{2} + O(h^2).$$

Si au contraire, nous retranchons l'une à l'autre, et remplaçons $\boldsymbol{x}'(t+h/2)$ par $f(t+h/2, \boldsymbol{x}(t+h/2))$, nous obtenons

$$\begin{aligned}\boldsymbol{x}(t+h) &= \boldsymbol{x}(t) + hf(t+h/2, \boldsymbol{x}(t+h/2)) + O(h^3)\\ &= \boldsymbol{x}(t) + hf(t+h/2, \frac{\boldsymbol{x}(t)+\boldsymbol{x}(t+h)}{2} + O(h^2)) + O(h^3)\\ &= \boldsymbol{x}(t) + hf(t+h/2, \frac{\boldsymbol{x}(t)+\boldsymbol{x}(t+h)}{2}) + O(h^3)\end{aligned} \tag{3.1.6}$$

On en déduit la méthode implicite suivante :

Définition 3.5 (Méthode du point milieu implicite)**.**

$$\boldsymbol{x}_{n+1}^h = \boldsymbol{x}_n^h + hf(t_n + h/2, \frac{\boldsymbol{x}_n + \boldsymbol{x}_{n+1}^h}{2}). \tag{3.1.7}$$

Il s'agit d'une méthode d'ordre 2. Nous pouvons en déduire une autre méthode d'ordre 2. En effet,

$$f(t+h/2, \frac{\boldsymbol{x}(t)+\boldsymbol{x}(t+h)}{2}) = \frac{f(t,\boldsymbol{x}(t)) + f(t+h, \boldsymbol{x}(t+h))}{2} + O(h^2).$$

Nous injectons ce développement asymptotique en h dans (3.1.6), et obtenons

$$\boldsymbol{x}(t+h) = \boldsymbol{x}(t) + h\frac{f(t,\boldsymbol{x}(t)) + f(t+h, \boldsymbol{x}(t+h))}{2} + O(h^3). \tag{3.1.8}$$

On en déduit la méthode implicite suivante :

Définition 3.6 (Méthode de Crank-Nicolson ou des trapèzes implicites)**.**

$$\boldsymbol{x}_{n+1}^h = \boldsymbol{x}_n^h + h\frac{f(t_n, \boldsymbol{x}_n^h) + f(t_n+h, \boldsymbol{x}_{n+1}^h)}{2}. \tag{3.1.9}$$

Il s'agit aussi d'une méthode d'ordre 2. Pour voir immédiatement que ces méthodes étaient d'ordre 2, nous avons utilisé un critère plus pratique que la définition

Proposition 3.7. Méthodes explicites : *Soit Φ_f une méthode explicite de résolution numérique d'une EDO. Cette méthode est d'ordre p si et seulement si pour toute solution $(I, \boldsymbol{x})$ de l'EDO $\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t))$, on a*

$$\|\boldsymbol{x}(t+h) - \Phi_f(t, \boldsymbol{x}(t), h)\| = O(h^{p+1}).$$

Méthodes implicites : *Soit Ψ_f une méthode implicite de résolution numérique d'une EDO. Cette méthode est d'ordre p si et seulement si pour toute solution $(I, \boldsymbol{x})$ de l'EDO $\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t))$, et tout t dans I, on a*

$$\|\Psi_f(\boldsymbol{x}(t+h), t, \boldsymbol{x}(t), h)\| = O(h^{p+1}),$$
$$\left\| \left(\frac{\partial \Psi_f}{\partial \boldsymbol{x}_{n+1}^h} \right)^{-1} (\boldsymbol{x}(t), t, \boldsymbol{x}(t), h) \right\| = O(1).$$

où $\frac{\partial \Psi_f}{\partial \boldsymbol{x}_{n+1}^h}$ est la jacobienne de Ψ_f par rapport à la première variable vectorielle (ou par rapport aux d premières variables scalaires).

Démonstration. C'est immédiat pour une méthode explicite. Pour une méthode implicite,c'est une conséquence directe du théorème des fonctions implicites. □

La condition sur la dérivée partielle de Ψ_f dans la proposition 3.7 est nécessaire. Sinon, on pourrait poser $\Psi_f = h^{p+1}\Psi_f$ et montrer que toute méthode implicite est d'ordre p, ce qui est évidemment faux.

Les conditions d'ordres portent sur ce que l'on appelle l'erreur locale, *i.e.* l'erreur entre la solution exacte et la solution numérique après une unique itération en temps. Nous sommes interessés par l'erreur globale, l'erreur commise après avoir itéré N fois, *i.e.*, par la norme de $\boldsymbol{x}_N^h - \boldsymbol{x}(T)$. Y a-t-il un lien entre l'erreur locale et l'erreur globale ? Peut-on déduire une majoration de l'erreur globale en fonction des différentes erreurs locales ? La réponse est positive. Mais avant de donner un résultat théorique, nous allons l'observer numériquement.

Pour illustrer l'importance de la notion d'ordre d'une méthode de résolution numérique d'une EDO, nous appliquons plusieurs de ces méthodes au problème de Cauchy $x' = \cos(t)x$, $x(0) = 1$. La solution exacte est connue : il s'agit de la fonction $x\colon t \mapsto \exp(\sin(t))$. Nous allons utiliser les méthodes d'Euler Explicite, d'Euler Implicite, du point Milieu Implicite et une méthode d'ordre 4 nommée RK4. Cette dernière méthode sera introduite à la §3.2.1 et qui est d'ordre 4. Nous regardons l'erreur absolue commise par ces méthodes en $t = 1$ pour différents pas de temps. Ces erreurs sont données à la Table 3.1. Nous avons aussi représentés ces erreurs en fonction du pas à la figure 3.1 en échelle loglog. Les courbes d'erreurs dans cette échelle sont des droites. Si on calcule la pente de ces courbes, on remarque que les pentes

$1/h$	Euler Explicite	Euler Implicite	Pt Mil. Impl.	RK4
8	$3.95 \cdot 10^{-2}$	$3.77 \cdot 10^{-2}$	$3.22 \cdot 10^{-3}$	$2.37 \cdot 10^{-6}$
16	$1.96 \cdot 10^{-2}$	$1.92 \cdot 10^{-2}$	$8.04 \cdot 10^{-4}$	$1.47 \cdot 10^{-7}$
32	$9.75 \cdot 10^{-3}$	$9.64 \cdot 10^{-3}$	$2.01 \cdot 10^{-4}$	$9.15 \cdot 10^{-9}$
64	$4.86 \cdot 10^{-3}$	$4.84 \cdot 10^{-3}$	$5.02 \cdot 10^{-5}$	$5.70 \cdot 10^{-10}$
128	$2.43 \cdot 10^{-3}$	$2.42 \cdot 10^{-3}$	$1.25 \cdot 10^{-5}$	$3.56 \cdot 10^{-11}$
256	$1.21 \cdot 10^{-3}$	$1.22 \cdot 10^{-3}$	$3.14 \cdot 10^{-6}$	$2.22 \cdot 10^{-12}$
512	$6.06 \cdot 10^{-4}$	$6.18 \cdot 10^{-4}$	$7.84 \cdot 10^{-7}$	$1.39 \cdot 10^{-13}$
1 024	$3.03 \cdot 10^{-4}$	$3.26 \cdot 10^{-4}$	$1.96 \cdot 10^{-7}$	$1.78 \cdot 10^{-15}$
2 048	$1.52 \cdot 10^{-4}$	$1.98 \cdot 10^{-4}$	$4.90 \cdot 10^{-8}$	$7.11 \cdot 10^{-15}$

TABLE 3.1 – Comparaison de l'erreur globale d'Euler Explicite, d'Euler Implicite, du Point-Milieu Implicite et de RK4 en $t = 1$ pour l'EDO $x' = \cos(t)x$.

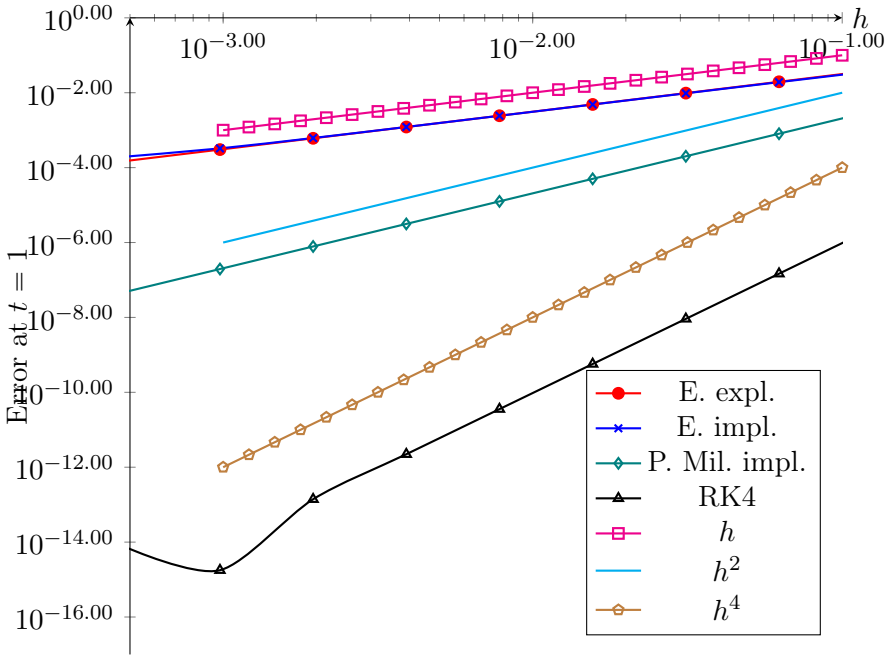


FIGURE 3.4 – Comparaison de l'erreur globale d'Euler Explicite, d'Euler Implicite et du Point-Milieu Implicite en $t = 1$ pour l'EDO $x' = \cos(t)x$ en échelle log log.

pour les Méthodes d'Euler valent 1, que celle pour la méthode du Point Milieu Implicite vaut 2 et que celle pour la méthode RK4 vaut 4. Pour toutes les méthodes, la pente de la courbe d'erreur en échelle loglog est l'ordre. Nous venons d'observer numériquement que l'erreur globale d'une méthode d'ordre p est en $O(|h|^p)$. Cette représentation en échelle loglog est très utile pour vérifier l'ordre d'une méthode. Quand la pente de l'erreur numérique en échelle loglog numériquement s'avère inférieure à l'ordre théorique d'une méthode de résolution numérique d'une EDO, c'est le plus souvent un signe que l'implémentation de la méthode est incorrect.

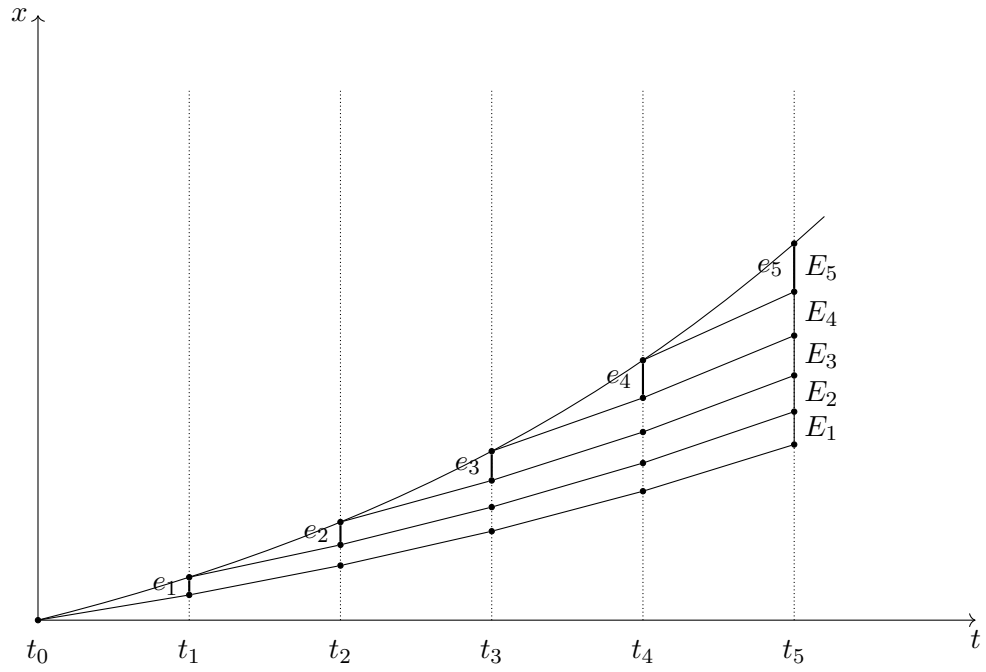


FIGURE 3.5 – Analyse de l'erreur : estimation de l'erreur globale

Revenons à l'analyse théorique de l'erreur. Nous allons démontré ce que nous avons observé numériquement : que l'erreur globale d'une méthode d'ordre p est en $O(|h|^p)$. Comme on peut le voir sur la figure 3.5, l'erreur globale n'est pas la somme des erreurs locales commises à chaque itération. En effet, l'erreur locale e_n commise à la n^{e} itération peut s'amplifier aux itérations suivantes. Pour estimer l'erreur globale, nous avons le théorème suivant :

Théorème 3.8: Cauchy(1824)

Soit $\boldsymbol{x}(t)$ la solution de $\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t))$, $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ sur t_0, T. On

pose
$$U_\varepsilon = \{(t, \boldsymbol{y}) \in \Omega, \text{ t.q. } t_0 \leqslant t \leqslant T, \|\boldsymbol{x}(t) - \boldsymbol{y}\| \leqslant \epsilon\}.$$
On suppose qu'une des deux conditions suivantes est vérifiée :

— Soit il existe $\varepsilon > 0$, $h_{\max} > 0$ et $L > 0$ tels que pour tout $(t, \boldsymbol{y})$ dans U_ε, $(t, \hat{\boldsymbol{y}})$ dans U_ε et $h \leqslant h_{\max}$,
$$\|\Phi_f(t, \boldsymbol{y}, h) - \Phi_f(t, \hat{\boldsymbol{y}}, h)\|) \leqslant (1 + Lh)\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|. \tag{3.1.10}$$

— Soit il existe $\varepsilon > 0$, et $L > 0$ tels que pour tout $(t, \boldsymbol{y})$ dans U_ε, $(t, \hat{\boldsymbol{y}})$ dans U_ε,
$$\|f(t, \boldsymbol{y}) - f(t, \hat{\boldsymbol{y}})\|) \leqslant L\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|. \tag{3.1.11}$$

On suppose de plus, qu'il existe $C > 0$ telle que pour tout $(t, \boldsymbol{y})$ et tout $h \leqslant h_{\max}$, on a
$$\|\varphi_f(t + h; t, \boldsymbol{y}) - \Phi_f(t, \boldsymbol{y}, h)\| \leqslant Ch^{p+1}. \tag{3.1.12}$$

Alors, nous avons
$$\|\boldsymbol{x}(t_n) - \boldsymbol{x}_n\| \leqslant h^p \frac{C}{L}(\exp(L(t_n - t_0)) - 1). \tag{3.1.13}$$

Démonstration. Pour simplifier la preuve, nous allons supposer que $\Omega = \mathbb{R} \times \mathbb{R}^d$ et que les inégalités sont valables globalement. Nous faisons la démonstration uniquement dans le cas où c'est la condition (3.1.10) qui est vérifiée [1]. Il suffit de vérifier que le résultat est vrai pour $n = N$. Nous posons pour tout $0 \leqslant j \leqslant N$, voir figure 3.5,
$$\begin{aligned} \boldsymbol{x}_j^{h,j} &= \boldsymbol{x}(t_j), \\ \boldsymbol{x}_{n+1}^{h,j} &= \Phi_f(t_n, \boldsymbol{x}_n^{h,j}, h) \text{ pour } j \leqslant n \leqslant N - 1. \end{aligned}$$

On a $\boldsymbol{x}_n^h = \boldsymbol{x}_n^{h,0}$. On pose pour $1 \leqslant n \leqslant N$:
$$e_n = \|\boldsymbol{x}_n^{h,n-1} - \boldsymbol{x}_n^{h,n}\|, \qquad E_n = \|\boldsymbol{x}_N^{h,n} - \boldsymbol{x}_N^{h,n-1}\|.$$

Les e_n sont les erreurs locales commises à l'itération n. Les E_n sont les erreurs

1. L'autre cas, lorsque c'est la condition (3.1.11) qui est vérifiée, se fait de manière similaire en utilisant le lemme de Grönwall.

e_n une fois amplifiées à l'iteration N. On a

$$\begin{aligned}\|\boldsymbol{x}(t_N) - \boldsymbol{x}_N^h\| &= \|\boldsymbol{x}_N^{h,N} - \boldsymbol{x}_N^{h,0}\| \\ &\leqslant \sum_{n=1}^{N} \|\boldsymbol{x}_N^{h,n} - \boldsymbol{x}_N^{h,n-1}\| \\ &\leqslant \sum_{n=1}^{N} E_n.\end{aligned}$$

On a aussi

$$\begin{aligned}e_n &= \|\boldsymbol{x}_n^{h,n-1} - \boldsymbol{x}(t_n)\| \\ &\leqslant Ch^{p+1}.\end{aligned}$$

Or

$$E_n \leqslant (1 + Lh)^{N-n} e_n \leqslant \exp(Lh(N-n))e_n.$$

Donc,

$$\begin{aligned}\|\boldsymbol{x}(t_N) - \boldsymbol{x}_N^h\| &\leqslant Ch^{p+1} \sum_{n=1}^{N} \exp(Lh(N-n)) \\ &\leqslant Ch^{p+1} \sum_{j=0}^{N-1} \exp(Lhj) \\ &\leqslant Ch^{p+1} \frac{\exp(LhN) - 1}{\exp(Lh) - 1} \\ &\leqslant Ch^{p+1} \frac{\exp(L(t_N - t_0)) - 1}{\exp(Lh) - 1} \\ &\leqslant C\frac{h^{p+1}}{Lh}(\exp(L(t_N - t_0)) - 1) \\ &\leqslant C\frac{h^p}{L}(\exp(L(t_N - t_0)) - 1).\end{aligned}$$

□

3.1.3 La A-stabilité

Nous introduisons maintenant le concept de A-stabilité. On considère l'EDO scalaire $x'(t) = \lambda x$ avec $\Re(\lambda) \leqslant 0$. On connaît l'expression de la solution exacte. On a $x(t) = \exp(\lambda(t-t_0))x(t_0)$. En particulier la solution x reste bornée sur $[t_0, +\infty[$. Si on applique une méthode de résolution numérique à cette EDO, on souhaite que la suite des itérées x_n reste bornée elle-aussi. C'est le concept de A-stabilité.

Définition 3.9: A-stabilité

Soit $f \mapsto \Phi_f$ une méthode de résolution numérique des EDO. On suppose qu'appliquée à l'EDO $x'(t) = \lambda x$, la méthode donne x_{n+1} comme fonction de x_n et de $a = \lambda h$ où h est le pas de la méthode. Le domaine de A-stabilité est

$$S = \{\lambda h \in \mathbb{C} \text{ t.q. la suite des itérées } (x_n) \text{ reste bornée}\}.$$

La méthode est dite inconditionnellement A-stable si le domaine de stabilité S contient $\mathbb{C}^- = \{z \in \mathbb{C}, \Re(z) \leqslant 0\}$.

On va appliquer ce concept à la méthode d'Euler explicite et d'Euler implicite. Pour la méthode d'Euler explicite appliquée à $x' = \lambda x$, on obtient $x_n = (1 + \lambda h)^n x_0$, le domaine de stabilité S de la méthode d'Euler explicite est l'ensemble des z tel que $|1 + z| \leqslant 1$. On a donc S. La méthode d'Euler explicite n'est donc pas inconditionnellement A-stable. Elle est A-stable sous la condition $|1 + \lambda h| \leqslant 1$.

On va maintenant appliquer le concept de A-stabilité à la méthode d'Euler implicite. Pour la méthode d'Euler implicite appliquée à $x' = \lambda x$, on obtient $x_n = x_0/(1 - \lambda h)^n$. Le domaine de stabilité S de la méthode d'Euler implicite est l'ensemble des z dans $\mathbb{C}$ tel que $|1-z| \geqslant 1$. Comme $\mathbb{C}^-$ est inclus dans S, la méthode d'Euler implicite est inconditionnellement A-stable.

Le théorème de Cauchy (1824), Théorème 3.8 nous dit qu'une méthode d'ordre 1 convergera à l'ordre 1 sous des hypothèses faibles. En particulier, ce théorème ne fait pas intervenir la notion de A-stabilité. Malgré cela, la notion de A-stabilité reste très importante. En effet, le Théorème 3.8 ne regarde que la limite quand le pas de temps h tend vers zéro. Or, en pratique, nous ne souhaitons évidemment pas faire tendre le pas de temps vers zéro dans nos simulations. Le concept de A-stabilité permet de donner un ordre de grandeur au pas de temps le plus grand pour avoir une solution numérique qui n'est pas complètement absurde.

Regardons comment faire sur une équation autonome non linéaire : $\boldsymbol{x}'(t) = f(\boldsymbol{x}(t))$. Soit $\boldsymbol{x}_n$ la n^{e} itération par une méthode de résolution numérique. Notons $\mathrm{J}f(\boldsymbol{x}_n)$ la jacobienne de f en $\boldsymbol{x}_n$. Soit $(\lambda_i)_{1 \leqslant i \leqslant d}$ les valeurs propres de cette jacobienne. Il faut choisir un pas de temps h_n tel que pour tout i dans $[\![1, d]\!]$ tel que $\Re(\lambda_i) \leqslant 0$, on ait $h_n \lambda_i$ qui appartienne au domaine de A-stabilité de la méthode numérique. Attention, il ne s'agit pas d'un théorème, seulement d'une règle heuristique. Pour une équation non autonome $\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t))$, c'est le même principe mais les notations changent, au lieu de regarder les valeurs propres de la jacobienne de f (qui n'est même pas une matrice carrée), on regarde celles de la jacobienne de f à t fixé, qui est noté $\frac{\partial f}{\partial \boldsymbol{x}}(t_n, \boldsymbol{x}_n)$.

L'intérêt d'une méthode inconditionnellement stable est que le pas de

temps peut être choisi en fonction de la tolérance de l'erreur que l'on souhaite atteindre. Pour une méthode non inconditionnellement stable, il faudra en plus veiller à rester dans le domaine de stabilité de la méthode.

3.2 Méthodes de Runge-Kutta explicites

Les méthodes de Runge-Kutta explicites [2] sont des méthodes explicites employées pour la résolution numérique des équations différentielles ordinaires (EDO). Étant explicites, leur mise en oeuvre est aisée et leur coût très raisonnable. De plus, ces méthodes peuvent, par un choix judicieux des coefficients, atteindre un ordre élevé. Dans ce chapitre, nous montrons comment obtenir, à partir de la notion d'arbre, les conditions nécessaires et suffisantes pour qu'une méthode de Runge-Kutta soit d'ordre p.

3.2.1 La méthode de Runge-Kutta 4

Algorithme 3.2.1: Méthode de Runge-Kutta 4

Entrées :
Fonction f.
Condition initiale $\boldsymbol{x}_0$, t_0.
Temps initial t_0.
Temps final t_f.
Nombre de pas de temps N.

Algorithme :
$h \coloneqq (t_f - t_0)/N$.
Pour n allant de 0 à $N-1$:
$\boldsymbol{k}_1 \coloneqq f(t_n, \boldsymbol{x}_n)$
$\boldsymbol{k}_2 \coloneqq f(t_n + h/2, \boldsymbol{x}_n + h\boldsymbol{k}_1/2)$
$\boldsymbol{k}_3 \coloneqq f(t_n + h/2, \boldsymbol{x}_n + h\boldsymbol{k}_2/2)$
$\boldsymbol{k}_4 \coloneqq f(t_n + h, \boldsymbol{x}_n + h\boldsymbol{k}_3)$
$\boldsymbol{x}_{n+1} \coloneqq \boldsymbol{x}_n + \frac{h}{6}(k_1 + 2\boldsymbol{k}_2 + 2\boldsymbol{k}_3 + \boldsymbol{k}_4)$
Fin Pour

Commençons par un exemple de méthode de Runge-Kutta très connue : la méthode de Runge-Kutta 4, abrégé en méthode RK4. Pour obtenir $\boldsymbol{x}_{n+1}$ à partir de $\boldsymbol{x}_n$ de t_n et de t_{n+1}, on applique la formule suivante suivant :

$$\boldsymbol{k}_1 = \boldsymbol{f}(t_n, \boldsymbol{x}_n),$$

2. Il existe des méthodes de Runge-Kutta implicites mais nous ne les aborderons pas dans ce cours. Aussi, dans ce cours, nous désigneronstoujours par « méthodes de Runge-Kutta » les méthodes de Runge-Kutta explicites.

```
def RK4(f,t0,x0,tf,h) :
    nbiter=int(math.ceil((tf-t0)/h))
    h=(tf-t0)/nbiter
    x=x0;
    t=t0;
    for i in range(0,nbiter) :
        k1=f(t,x);
        k2=f(t+h/2.0,x+h*k1/2.0);
        k3=f(t+h/2.0,x+h*k2/2.0);
        k4=f(t+h,x+h*k3);
        x=x+h*(k1/6.0+k2/3.0+k3/3.0+k4/6.0);
        t=t+h
    return x;
```

Code 3.3 – Code Runge-Kutta 4 en Python

$$
\begin{aligned}
\boldsymbol{k}_2 &= \boldsymbol{f}(t_n + \frac{h}{2}, \boldsymbol{x}_n + \frac{h}{2}\boldsymbol{k}_1),\\
\boldsymbol{k}_3 &= \boldsymbol{f}(t_n + \frac{h}{2}, \boldsymbol{x}_n + \frac{h}{2}\boldsymbol{k}_2),\\
\boldsymbol{k}_4 &= \boldsymbol{f}(t_n + h, \boldsymbol{x}_n + h\boldsymbol{k}_3),\\
\boldsymbol{x}_{n+1} &= \boldsymbol{x}_n + \frac{h}{6}(\boldsymbol{k}_1 + 2\boldsymbol{k}_2 + 2\boldsymbol{k}_3 + \boldsymbol{k}_4),
\end{aligned}
$$

où $h = t_{n+1} - t_n$. La méthode de Runge-Kutta 4 est écrite sous forme de pseudo-code à l'Algorithme 3.2.1. Un exemple d'implémentation de Runge-Kutta 4 est donnée au Code 3.3. La méthode de Runge-Kutta 4 est d'ordre 4, d'où son nom. Justifier le choix des divers coefficients dans la méthode de Runge-Kutta 4 n'est pas chose aisée. En effet, à ce stade, ces coefficients doivent vous paraître complètement arbitraires. Quant à calculer théoriquement l'ordre de cette méthode, cela semble être une tâche très fastidieuse.

Par contre, nous pouvons aisément vérifier numériquement l'ordre de la méthode RK-4 en appliquant cette méthode. D'ailleurs, nous l'avons déjà fait à la Table 3.1 et à la Figure 3.4 où nous avons représenté l'erreur commise par la méthode RK4 avec la solution exacte dans une échelle log log.

Nous expliquerons la provenance des coefficients de la méthode de Runge-Kutta 4 et comment calculer son ordre théorique aux sections §3.2.3 et §3.2.4.

3.2.2 Forme générale des Méthodes de Runge-Kutta

La forme générale des méthodes de Runge-Kutta explicites est donnée par :

Définition 3.10: Forme générale des méthodes de Runge-Kutta

Une méthode de Runge-Kutta est donnée par la donnée d'un entier $s \geqslant 1$, de s coefficients $(b_j)_{1\leqslant j\leqslant s}$, de s coefficients $(c_i)_{1\leqslant i\leqslant s}$, et de $s(s-1)$ coefficients $(a_{i,j})_{1\leqslant i\leqslant s, 1\leqslant j\leqslant i-1}$. Le calcul de $\boldsymbol{x}_{n+1}$ en fonction de t_n, du pas de temps h et de $\boldsymbol{x}_n$ est donnée par les formules suivantes :

$$\begin{aligned}
\boldsymbol{k}_1 &= \boldsymbol{f}(t_n, \boldsymbol{x}_n),\\
\boldsymbol{k}_2 &= \boldsymbol{f}(t_n + c_2 h, \boldsymbol{x}_n + a_{21} h \boldsymbol{k}_1),\\
\boldsymbol{k}_3 &= \boldsymbol{f}(t_n + c_3 h, \boldsymbol{x}_n + a_{31} h \boldsymbol{k}_1 + a_{32} h \boldsymbol{k}_2),\\
\ldots &= \ldots\\
\boldsymbol{k}_i &= \boldsymbol{f}(t_n + c_i h, \boldsymbol{x}_n + \sum_{j<i} a_{ij} h \boldsymbol{k}_j),\\
\ldots &= \ldots\\
\boldsymbol{k}_s &= \boldsymbol{f}(t_n + c_s h, \boldsymbol{x}_n + \sum_{j<n} a_{sj} h \boldsymbol{k}_j),\\
\boldsymbol{x}_{n+1} &= \boldsymbol{x}_n + h \sum_{j=1}^{s} b_j \boldsymbol{k}_j.
\end{aligned}$$

L'entier s est appelé nombre d'étages de la méthode de Runge-Kutta.

Pour la méthode de Runge-Kutta 4, l'ordre 4 est égal au nombre d'étages 4 mais c'est un cas particulier. Par exemple, 6 étages au moins sont nécessaires pour qu'une méthode de Runge-Kutta soit d'ordre 5.

Une méthode de Runge-Kutta est complètement caractérisée par ses coefficients (a_{ij}), (b_j) et (c_i). On représente couramment une méthode de Runge-Kutta par un tableau, appelé tableau de Butcher, contenant lesdits coefficients.

$$\begin{array}{c|c} c_i & a_{ij} \\ \hline & b_j \end{array} \qquad \begin{array}{c|cccc} 0 & & & & \\ c_2 & a_{2,1} & & & \\ \vdots & \vdots & \ddots & & \\ c_s & a_{s,1} & \ldots & a_{s,s-1} & \\ \hline & b_1 & \ldots & \ldots & b_s \end{array}$$

Voir les tableaux de Butcher, Table 3.2, pour des méthodes de Runge-Kutta d'ordre faible et les tableaux de Butcher, Table 3.3, pour la représentation en tableau de la méthode RK4 et d'une autre méthode d'ordre 4 : la règle des 3/8.

Nous allons étudier la A-stabilité des méthodes de Runge-Kutta.

Proposition 3.11. *Les itérées x_n obtenues en employant une méthode de*

0	
	1

Euler explicite

0		
1	1	
	1/2	1/2

Méthode de Heun

0		
1/2	1/2	
	0	1

Méthode du point milieu

TABLE 3.2 – Méthodes de Runge-Kutta d'ordre $\leqslant 2$

0				
1/2	1/2			
1/2	0	1/2		
1	0	0	1	
	1/6	1/3	1/3	1/6

La Méthode RK4

0				
1/3	1/3			
2/3	−1/3	1		
1	1	−1	1	
	1/8	3/8	3/8	1/8

La Règle des 3/8

TABLE 3.3 – Deux méthodes de Runge-Kutta d'ordre 4

Runge-Kutta à s étages appliquée à l'EDO $x' = \lambda x$ avec un pas h vérifient

$$x_{n+1} = P(\lambda h)x_n$$

où P est un polynôme de degré au plus s, appelée fonction de stabilité de la méthode de Runge-Kutta. Une méthode de Runge-Kutta n'est donc jamais inconditionnellement A-stable.

Avant de se lancer dans le calcul de l'ordre, nous allons ramener le cas général à celui d'une équation autonome où $\boldsymbol{f}$ ne dépend pas explicitement du temps. Cela nous permettra d'établir une relation entre les c_j et les a_{ij}. Pour rendre l'équation autonome, il suffit d'ajouter une ligne au système d'équations. Nous considérons la fonction $\boldsymbol{X}$ à valeurs dans $\mathbb{R} \times \mathbb{R}^d$ définie par $\boldsymbol{X}(t) = (t, \boldsymbol{x}(t))$. La fonction $\boldsymbol{X}$ est solution de l'équation différentielle autonome

$$\begin{aligned} \boldsymbol{X}'(t) &= F(\boldsymbol{X}(t)) = \big(1, \boldsymbol{f}(X_0(t), \boldsymbol{X}(t))\big), \\ \boldsymbol{X}(t_0) &= (t_0, \boldsymbol{x}_0), \end{aligned} \tag{3.2.1}$$

où nous notons $(X_0, \hat{\boldsymbol{X}}(t)) \coloneqq \boldsymbol{X}(t)$. Appliquons une méthode de Runge-Kutta à cette équation autonome. Nous obtenons

$$\boldsymbol{K}_i = \boldsymbol{F}(\boldsymbol{X}_n + h \sum_{j<i} a_{ij} \boldsymbol{K}_j),$$

soit si on pose $\boldsymbol{K}_i = (k_{0i}, \boldsymbol{k_i})$

$$k_{0i} = 1,$$

$$\begin{aligned}\boldsymbol{k}_j &= \boldsymbol{F}\big(t_n + h\sum_{j<i} a_{ij}k_{0j}, \boldsymbol{x}_n + h\sum_{j<i} a_{ij}\boldsymbol{k}_j\big),\\ &= \boldsymbol{f}\big(t_n + h(\sum_{j<i} a_{ij}), \boldsymbol{x}_n + h\sum_{j<i} a_{ij}\boldsymbol{k}_j\big).\end{aligned}$$

Nous posons donc

$$c_i = \sum_j a_{ij} \tag{3.2.2}$$

pour tout $i \leqslant s$. Cela garantit que la méthode de Runge-Kutta sur l'équation autonome (3.1.1a) et celle sur sur l'équation non autonome (3.2.1) coïncident. Par la suite et ce sans perte de généralité, nous ne considérerons que des équations différentielles autonomes.

3.2.3 Calcul inélégant de l'ordre d'une méthode de Runge-Kutta

Nous souhaitons calculer l'ordre p d'une méthode de Runge-Kutta ou plus exactement établir les conditions que doivent vérifier les a_{ij}, les b_j et les c_i pour qu'une méthode de Runge-Kutta soit d'ordre p.

Pour calculer l'ordre d'une méthode de Runge-Kutta, le principe est simple : il suffit de calculer le développement de Taylor de la solution exacte à l'ordre désiré p, de calculer celui de la solution numérique au même ordre, puis de comparer les coefficients. La méthode est d'ordre p si les développements sont identiques jusqu'à l'ordre p. Calculons le développement de Taylor à l'ordre 3 de la solution exacte $\boldsymbol{x}$.

$$\begin{aligned}\boldsymbol{x}'(t) &= \boldsymbol{f}(\boldsymbol{x}(t)),\\ \boldsymbol{x}''(t) &= \mathrm{D}\boldsymbol{f}(\boldsymbol{x}(t))(\boldsymbol{f}(\boldsymbol{x}(t))),\\ \boldsymbol{x}'''(t) &= \mathrm{D}^2\boldsymbol{f}(\boldsymbol{x}(t))(\boldsymbol{f}(\boldsymbol{x}(t)), \boldsymbol{f}(\boldsymbol{x}(t))) + \mathrm{D}\boldsymbol{f}(\boldsymbol{x}(t))(\mathrm{D}\boldsymbol{f}(\boldsymbol{x}(t))(\boldsymbol{f}(\boldsymbol{x}(t)))),\end{aligned}$$

donc

$$\begin{aligned}\boldsymbol{x}(t+h) = \boldsymbol{x}(t) &+ h\boldsymbol{f}(\boldsymbol{x}(t)) + \frac{(h)^2}{2}\mathrm{D}\boldsymbol{f}(\boldsymbol{x}(t))(\boldsymbol{f}(\boldsymbol{x}(t)))\\ &+ \frac{(h)^3}{6}\mathrm{D}^2\boldsymbol{f}(\boldsymbol{x}(t))(\boldsymbol{f}(\boldsymbol{x}(t)), \boldsymbol{f}(\boldsymbol{x}(t)))\\ &+ \frac{(h)^3}{6}\mathrm{D}\boldsymbol{f}(\boldsymbol{x}(t))(\mathrm{D}\boldsymbol{f}(\boldsymbol{x}(t))(\boldsymbol{f}(\boldsymbol{x}(t)))) + O((h)^4)\end{aligned} \tag{3.2.3}$$

Prenons maintenant une méthode à 3 étages [3] et calculons son développement de Taylor à l'ordre 3 en fonction du pas h :

$\boldsymbol{k}_1 = \boldsymbol{f}(\boldsymbol{x}_n)$,

3. Il est tout à fait possible et même très courant que le nombre d'étages soit différent de l'ordre.

$$
\begin{aligned}
\boldsymbol{k}_2 &= \boldsymbol{f}(\boldsymbol{x}_n) + ha_{21}\mathrm{D}\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{f}(\boldsymbol{x}_n)) + h^2\frac{a_{21}^2}{2}\mathrm{D}^2\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{f}(\boldsymbol{x}_n),\boldsymbol{f}(\boldsymbol{x}_n)) + O(h^3)\\
\boldsymbol{k}_3 &= \boldsymbol{f}(\boldsymbol{x}_n) + ha_{31}\mathrm{D}\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{f}(\boldsymbol{x}_n)) + ha_{32}\mathrm{D}\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{k}_2)\\
&\quad + \frac{h^2}{2}\mathrm{D}^2\boldsymbol{f}(\boldsymbol{x}_n)(a_{31}\boldsymbol{f}(\boldsymbol{x}_n) + a_{32}\boldsymbol{k}_2, a_{31}\boldsymbol{f}(\boldsymbol{x}_n) + a_{32}\boldsymbol{k}_2) + O(h^3)\\
&= \boldsymbol{f}(\boldsymbol{x}_n) + ha_{31}\mathrm{D}\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{f}(\boldsymbol{x}_n)) + ha_{32}\mathrm{D}\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{f}(\boldsymbol{x}_n))\\
&\quad + \frac{(h)^2}{2}(a_{32}+a_{31})^2\mathrm{D}^2\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{f}(\boldsymbol{x}_n),\boldsymbol{f}(\boldsymbol{x}_n))\\
&\quad + h^2a_{32}a_{21}\mathrm{D}\boldsymbol{f}(\boldsymbol{x}_n)(\mathrm{D}\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{f}(\boldsymbol{x}_n))) + O(h^3).
\end{aligned}
$$

d'où

$$
\begin{aligned}
\boldsymbol{x}_{n+1} = \boldsymbol{x}_n &+ (b_1+b_2+b_3)h\boldsymbol{f}(\boldsymbol{x}_n) + (b_2c_2+b_3c_3)h^2\mathrm{D}\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{f}(\boldsymbol{x}_n))\\
&+ \frac{b_2c_2^2+b_3c_3^2}{2}h^3\mathrm{D}^2\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{f}(\boldsymbol{x}_n),\boldsymbol{f}(\boldsymbol{x}_n))\\
&+ (b_3a_{32}c_2)h^3\mathrm{D}\boldsymbol{f}(\boldsymbol{x}_n)(\mathrm{D}\boldsymbol{f}(\boldsymbol{x}_n)(\boldsymbol{f}(\boldsymbol{x}_n))).
\end{aligned}
\tag{3.2.4}
$$

En comparant (3.2.3) et (3.2.4), nous obtenons une condition d'ordre 1, une d'ordre 2, et deux d'ordre 3 :

$$
\begin{aligned}
b_1+b_2+b_3 &= 1\\
b_2c_2+b_3c_3 &= \frac{1}{2}\\
b_2c_2^2+b_3c_3^2 &= \frac{1}{3}\\
b_3a_{32}c_2 &= \frac{1}{6}
\end{aligned}
$$

Si le principe du calcul de l'ordre est simple, l'exécution, elle, est fastidieuse. Le nombre de calculs requis devient rapidement prohibitif lorsque l'ordre désiré augmente. Cependant, ce calcul est utile car il nous permet de supputer la forme générale de ces deux développements. Dans chacun de ces deux développements, les termes qui apparaissent à l'ordre p sont des expressions, à un facteur scalaire près, de type

$$\mathrm{D}^{k_1}\boldsymbol{f}(\boldsymbol{x}(t))(\mathrm{D}^{k_{12}}\boldsymbol{f}(\boldsymbol{x}(t))(\ldots,\ldots,\ldots),\ldots,\mathrm{D}^{k_{1k_1}}\boldsymbol{f}(\boldsymbol{x}(t))(\ldots)),)$$

où la somme de tout les ordres de dérivation vaut $p-1$. Pour qu'une méthode de Runge-Kutta soit d'ordre p, il suffit que les coefficients scalaires présents devant ces expressions différentielles coïncident jusqu'à l'ordre p.

Et nous pouvons faire une première remarque : le terme

$$\mathrm{D}\boldsymbol{f}(\boldsymbol{x}(t))(\mathrm{D}\boldsymbol{f}(\boldsymbol{x}(t))(\mathrm{D}\boldsymbol{f}(\boldsymbol{x}(t))(\ldots(\mathrm{D}\boldsymbol{f}(\boldsymbol{x}(t))(\boldsymbol{f}(\boldsymbol{x}(t))))))))$$

avec $p-1$ dérivations apparaît dans le développement de la solution exacte avec un coefficient non nul à l'ordre p. Ce même terme ne peut apparaître

dans le développement de la solution de Runge-Kutta qu'au bout de p étages. Nous avons donc le résultat suivant ;

Lemme 3.12. *L'ordre p d'une méthode de Runge-Kutta est inférieure à son nombre d'étages : $p \leqslant s$.*

Pour aller plus loin, il va falloir expliciter ces deux développements. À cette fin, nous utiliserons la notion d'arbre.

3.2.4 Arbres et expressions différentielles

Nous souhaitons obtenir une expression explicite du facteur scalaire pour chaque terme dans les deux développements : celui de la solution numérique et celui de la solution exacte. Pour ce faire, à chacune des expressions de type $D^k \boldsymbol{f}(\dots)$, nous associons un arbre, voir figure 3.6. Pour un arbre A

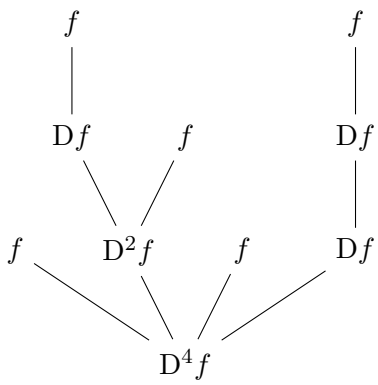


FIGURE 3.6 – Arbre pour $\mathrm{D}^4\boldsymbol{f}(\boldsymbol{f}, \mathrm{D}^2\boldsymbol{f}(\mathrm{D}\boldsymbol{f}(\boldsymbol{f}), \boldsymbol{f}), \boldsymbol{f}, \mathrm{D}\boldsymbol{f}(\mathrm{D}\boldsymbol{f}(\boldsymbol{f})))$

et $\boldsymbol{f}$ une fonction suffisamment régulière, nous notons $A(\boldsymbol{f}, \boldsymbol{y})$ l'expression différentielle associée, voir Figure 3.6.

Définition 3.13. *On appelle ordre d'un arbre son nombre de nœuds.*

À chaque arbre d'ordre p, correspond un coefficient devant $h^p A(\boldsymbol{f}, \boldsymbol{x})$ dans le développement de Taylor en t de $\boldsymbol{x}(t+h)$ exacte ainsi qu'un coefficient devant $h^p A(\boldsymbol{f}, \boldsymbol{x})$ dans le développement de Taylor de la solution numérique. Une méthode de Runge-Kutta est d'ordre p dans le cas général si ces deux coefficients coïncident pour tous les arbres d'ordre inférieure ou égal à p. Nous n'allons pas donner l'expression de ces deux coefficients mais donner leur multiple par un facteur commun.

Soit A un arbre. Soit $\mathcal{N}(A)$ l'ensemble des noeuds de A. Soit P dans $\mathcal{N}(A)$, on note $\tilde{\mathcal{N}}(A, P)$ l'ensemble des noeuds de A qui descendent de P (P

Ordre	arbre	condition
1		$\sum_i b_i = 1$
2		$\sum_i b_i c_i = \frac{1}{2}$
3		$\sum_{j<i} b_i a_{ij} c_j = \frac{1}{6}$
3		$\sum_i b_i c_i^2 = \frac{1}{3}$
4		$\sum_i b_i c_i^3 = \frac{1}{4}$
4		$\sum_{j<i} b_i c_i a_{ij} c_j = \frac{1}{8}$
4		$\sum_{j<i} b_i a_{ij} c_j^2 = \frac{1}{12}$
4		$\sum_{k<j<i} b_i a_{ij} a_{jk} c_k = \frac{1}{24}$

TABLE 3.4 – Conditions d'ordre 1 à 4

compris). On note $R(A)$ le nœud racine de A. Si $P \in \mathcal{N}(A)\backslash\{R(A)\}$, on note $\mathcal{F}(P)$, le nœud parent de P dans A. On pose

$$\beta(A) = \prod_{P\in\mathcal{N}(A)} \frac{1}{\mathrm{card}(\tilde{\mathcal{N}}(A,P))}$$

Soit une méthode de Runge-Kutta à s étages de coefficients $(c_i)_{1\leqslant i\leqslant s}$, $(b_j)_{1\leqslant j\leqslant s}$ et $(a_{ij})_{\substack{1\leqslant i\leqslant s\\ 1\leqslant j\leqslant i-1}}$. On pose

$$\alpha(A,(c_i),(b_j),(a_{ij})) = \sum_{i_R=1}^{s} \sum_{\substack{(i_P)_{P\in\mathcal{N}(A)\backslash\{R(A)\}}\\ 1\leqslant i_P<i_{\mathcal{F}(P)}}} b_{i_R} \prod_{P\in\mathcal{N}(A)\backslash\{R(A)\}} a_{i_{\mathcal{F}(P)}i_P}.$$

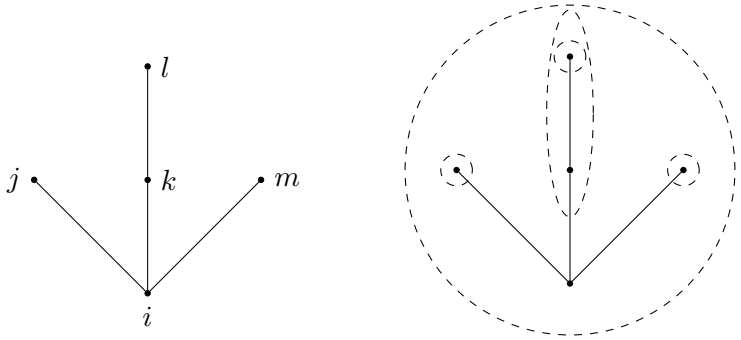
Théorème 3.14. *Une méthode de Runge-Kutta à s étages de coefficients $(c_i)_{1\leqslant i\leqslant s}$, $(b_j)_{1\leqslant j\leqslant s}$ et $(a_{ij})_{\substack{1\leqslant i\leqslant s\\ 1\leqslant j\leqslant i-1}}$ est d'ordre (au moins) p si pour tout arbre d'ordre inférieur ou égal à p*

$$\alpha\big(A,(c_i)_{1\leqslant i\leqslant s}(b_j)_{1\leqslant j\leqslant s}(a_{ij})_{\substack{1\leqslant i\leqslant s\\ 1\leqslant j\leqslant i-1}}\big) = \beta(A). \tag{3.2.5}$$

Voici un exemple. On considère l'arbre

Ordre	arbre	condition
5		$\sum_i b_i c_i^4 = \frac{1}{5}$
5		$\sum_{j<i} b_i c_i^2 a_{ij} c_j = \frac{1}{10}$
5		$\sum_{j<i} b_i c_i a_{ij} c_j^2 = \frac{1}{15}$
5		$\sum_{j<i,k<i} b_i a_{ij} a_{ik} c_j c_k = \frac{1}{20}$
5		$\sum_{j<i} b_i a_{ij} c_j^3 = \frac{1}{20}$
5		$\sum_{k<j<i} b_i a_{ij} a_{jk} c_k^2 = \frac{1}{60}$
5		$\sum_{k<j<i} b_i a_{ij} c_j a_{jk} c_k = \frac{1}{40}$
5		$\sum_{k<j<i} b_i c_i a_{ij} a_{jk} c_k = \frac{1}{30}$
5		$\sum_{l<k<j<i} b_i a_{ij} a_{jk} a_{kl} c_l = \frac{1}{120}$

TABLE 3.5 – Les 9 conditions d'ordre 5



On a labellisé l'arbre de gauche avec des indices i, j, k, l, m.

$$\sum_{\substack{i,j,k,l,m \\ j,k,l,m<i \\ l<k}} b_i a_{ij} a_{ik} a_{kl} a_{im} = \sum_{\substack{i,k \\ k<i}} b_i c_i^2 a_{ik} c_k$$

Dans le deuxième arbre, on a dessiné pour chaque noeud non feuille, l'ensemble des noeuds descendants. Le noeud racine a 5 descendants, un autre noeud en a 2, les autres sont des noeuds feuilles donc le terme de droite dans l'égalité (3.2.5) vaut 1/10 donc une condition d'ordre 5 pour une méthode de Runge-Kutta est

$$\sum_{\substack{i,j \\ j<i}} b_i c_i^2 a_{ij} c_j = \frac{1}{5 \cdot 2 \cdot 1 \cdot 1 \cdot 1}.$$

Nous avons explicité les conditions d'ordre 5 et moins dans les tables 3.4 et 3.5. Nous pouvons voir qu'il y a une condition d'ordre 1, une condition d'ordre 2, deux conditions d'ordre 3, quatre conditions d'ordre 4 et neuf conditions d'ordre 5. Pour qu'une méthode soit d'ordre 5, il faut que les 17 relations dans ces deux tables soient vérifiées.

3.2.5 Limites des méthodes de Runge-Kutta

Nous avons vu qu'il existait des méthodes de Runge-Kutta d'ordre élevée avec des méthodes. Mais calculer les coefficients pour qu'une une méthode de Runge-Kutta soit d'ordre élevée n'est pas chose facile. De plus, les méthodes de Runge-Kutta nécessitent plusieurs évaluations de la fonction f, à chaque itération et il arrive que l'évaluation de la fonction f soit très coûteuse informatiquement. Il existe d'autres familles de méthodes numériques pour lesquelles la montée en ordre est beaucoup plus aisée. Nous les abordons à la section §3.4. Enfin, les méthodes de Runge-Kutta que nous avons présentées n'ont pas de propriétés de « presque » conservation à long terme de certaines grandeurs physiques conservatives comme l'énergie ou le moment cinétique. Nous verrons quelques méthodes qui ont ces propriétés à la section §3.3.

```
def Verlet(f,t0,q0,p0,tf,h) :
    q=q0;
    p=p0;
    t=t0;
    for i in range(0,nbiter) :
        pundemi=p+h*f(t,q)/2.0;
        q=q+h*pundemi;
        p=pundemi+h*f(t+h,q)/2.0;
        t=t+h;
    return [q,p];
```

Code 3.4 – Code Verlet en Python

3.3 Méthodes de Newmark et de Störmer-Verlet

Dans cette section, nous introduisons les méthodes de Störmer-Verlet et de Newmark. Ces méthodes sont très employées en mécanique des structures.

3.3.1 La méthode de Störmer-Verlet

Cette méthode est souvent appelée méthode de Verlet ou méthode de Störmer. Elle est principalement utilisée pour des équations d'ordre 2 dans lesquels la dérivée d'ordre 1 n'apparaît pas et dont le second membre dérive d'un potentiel

$$\boldsymbol{r}''(t) = f(\boldsymbol{r}), \tag{3.3.1}$$

où $f = -\nabla E$, où $E\colon \mathbb{R}^d \to \mathbb{R}$ est une énergie de potentiel.

La mécanique offre de nombreux exemples de telles équations. En particulier, en mécanique du point, le mouvement d'un point mobile de masse m soumis à un potentiel qui ne dépend que de la position de ce point. L'EDO (3.3.1) est équivalente au système :

$$\boldsymbol{r}' = \boldsymbol{v} \tag{3.3.2a}$$
$$\boldsymbol{v}' = f(t, \boldsymbol{r}). \tag{3.3.2b}$$

Le principe de la méthode de Störmer-Verlet est de décaler la position où on discrétise les $\boldsymbol{x}'_n$ d'un demi-pas de temps par rapport aux positions où on discrétise les $\boldsymbol{x}$. Si on est en mécanique du point, cela revient à calculer les vitesses en $(t_n + t_{n+1})/2$ et les positions (et donc les accélérations) en t_n.

Définition 3.15: Méthode de Störmer-Verlet

Pour résoudre une EDO du type (3.3.2). La méthode de Störmer-Verlet est donnée par la formule d'itération suivante :

$$\boldsymbol{v}_{n+1/2} = \boldsymbol{v}_{n-1/2} + hf(t_n, \boldsymbol{r}_n),$$

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + h\boldsymbol{v}_{n+1/2}.$$

où $t_{n+1/2} = (t_n + t_{n+1})/2$. L'initialisation de $\boldsymbol{v}_{1/2}$ à partir $\boldsymbol{v}_0$ et de $\boldsymbol{x}_0$ se fait par la formule

$$\boldsymbol{v}_{1/2} = \boldsymbol{v}_0 + \frac{h}{2} f(t_0, \boldsymbol{r}_0),$$

La méthode de Störmer-Verlet est d'ordre 2. Pour le voir, il suffit de faire un développement de Taylor de $\boldsymbol{x}(t_{n+1})$ et de $\boldsymbol{x}(t_n)$ en $t_{n+1/2}$; puis un développement de Taylor de $\boldsymbol{v}(t_{n-1/2})$ et de $\boldsymbol{v}(t_{n+1/2})$ en t_n. Une implémentation en Python de la méthode de Störmer-Verlet est donnée au Code 3.4. La méthode de Störmer-Verlet peut aussi être donnée sans faire intervenir la vitesse. En effet, elle est équivalente à la formule de récurrence :

$$\boldsymbol{r}_{n+1} - 2\boldsymbol{r}_n + \boldsymbol{r}_{n-1} = h^2 f(t_n, \boldsymbol{r}_n), \tag{3.3.3a}$$

avec l'initialisation

$$\boldsymbol{r}_1 = \boldsymbol{r}_0 + h\boldsymbol{v}_0 + \frac{h^2}{2} f(t_0, \boldsymbol{r}_0). \tag{3.3.3b}$$

La méthode de Störmer-Verlet peut aussi être écrite en explicitant aussi les valeurs de la vitesse aux temps t_n.

$$\boldsymbol{v}_{n+1/2} = \boldsymbol{v}_n + \frac{h}{2} f(t_n, \boldsymbol{r}_n), \tag{3.3.4a}$$

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + h\boldsymbol{v}_{n+1/2}, \tag{3.3.4b}$$

$$\boldsymbol{v}_{n+1} = \boldsymbol{v}_{n+1/2} + \frac{h}{2} f(t_{n+1}, \boldsymbol{r}_{n+1}). \tag{3.3.4c}$$

Cette formulation est pratique pour l'étude théorique de la méthode de Störmer-Verlet.

L'intérêt de la méthode de Störmer-Verlet est qu'elle maintient stable l'énergie mécanique même sur les temps longs. C'est une propriété importante car de nombreuses EDO provenant de la mécanique ou de la physique conservent certaines grandeurs physiques au cours du temps. Nous allons voir sur un exemple. Regardons un problème de mécanique céleste : le mouvement de la Terre soumis à l'attraction gravitationnelle du Soleil. Nous négligeons le mouvement du soleil et faisons tous les calculs en 2 dimensions. Nous devons résoudre l'équation différentielle

$$\boldsymbol{r}' = \boldsymbol{v}, \tag{3.3.5a}$$

$$\boldsymbol{v}' = -\frac{\mathcal{G}(M_T)}{\|\boldsymbol{r}\|^3}\boldsymbol{r}, \tag{3.3.5b}$$

où la masse du Soleil $M_S = 1.989 \cdot 10^{30}$kg, la masse de la Terre $M_T = 5.9722 \cdot 10^{24}$kg et la constante universelle de la gravitation $\mathcal{G} = 6.673 \cdot$

T/h	RK4 $\boldsymbol{r}_h - \boldsymbol{r}_{\text{exact}}$ m	RK4 $\boldsymbol{v}_h - \boldsymbol{v}_{\text{exact}}$ m s^{-1}	Verlet $\boldsymbol{r}_h - \boldsymbol{r}_{\text{exact}}$ m	Verlet $\boldsymbol{v}_h - \boldsymbol{v}_{\text{exact}}$ m s^{-1}
64	$3.09 \cdot 10^{6}$	$6.43 \cdot 10^{-1}$	$3.12 \cdot 10^{9}$	$6.38 \cdot 10^{2}$
128	$1.57 \cdot 10^{5}$	$3.30 \cdot 10^{-2}$	$7.81 \cdot 10^{8}$	$1.60 \cdot 10^{2}$
256	$8.69 \cdot 10^{3}$	$1.83 \cdot 10^{-3}$	$1.95 \cdot 10^{8}$	$4.00 \cdot 10^{1}$
512	$5.08 \cdot 10^{2}$	$1.07 \cdot 10^{-4}$	$4.88 \cdot 10^{7}$	$1.00 \cdot 10^{1}$
1 024	$3.06 \cdot 10^{1}$	$6.49 \cdot 10^{-6}$	$1.22 \cdot 10^{7}$	$2.50 \cdot 10^{0}$
2 048	$1.88 \cdot 10^{0}$	$3.99 \cdot 10^{-7}$	$3.05 \cdot 10^{6}$	$6.26 \cdot 10^{-1}$
4 096	$1.15 \cdot 10^{-1}$	$2.43 \cdot 10^{-8}$	$7.63 \cdot 10^{5}$	$1.56 \cdot 10^{-1}$

TABLE 3.6 – Comparaison de l'erreur (en position et en impulsion)) globale de Störmer-Verlet et de RK4 après une période orbitale.

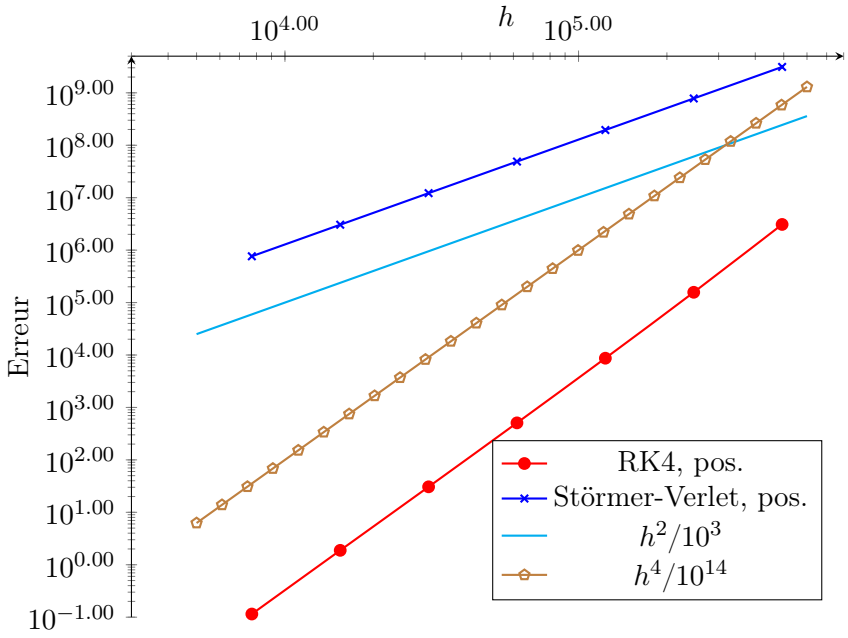


FIGURE 3.7 – Comparaison de l'erreur de position de RK4 et de Störmer-Verlet après une période orbitale en m en échelle log log.

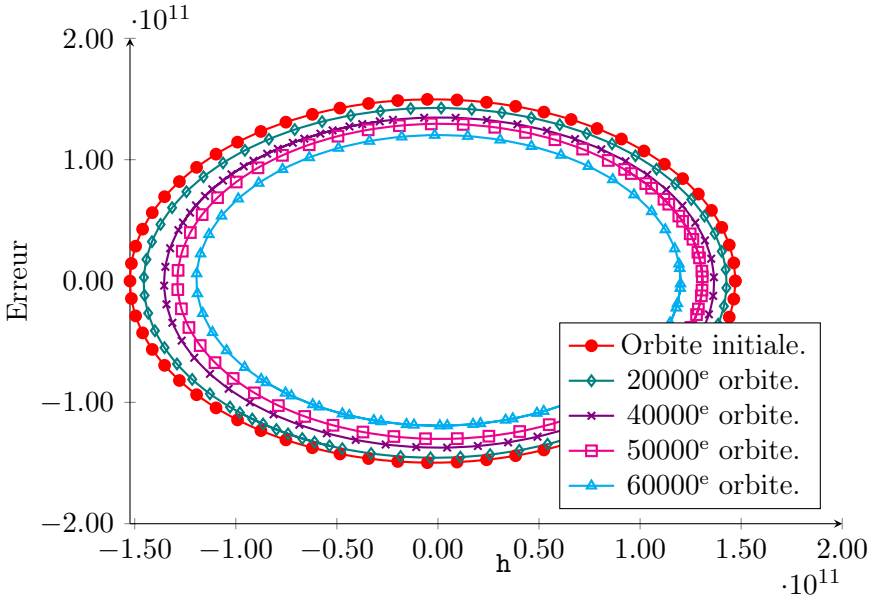


FIGURE 3.8 – Orbite de la Terre par RK4 pour 64 pas de temps par période de révolution.

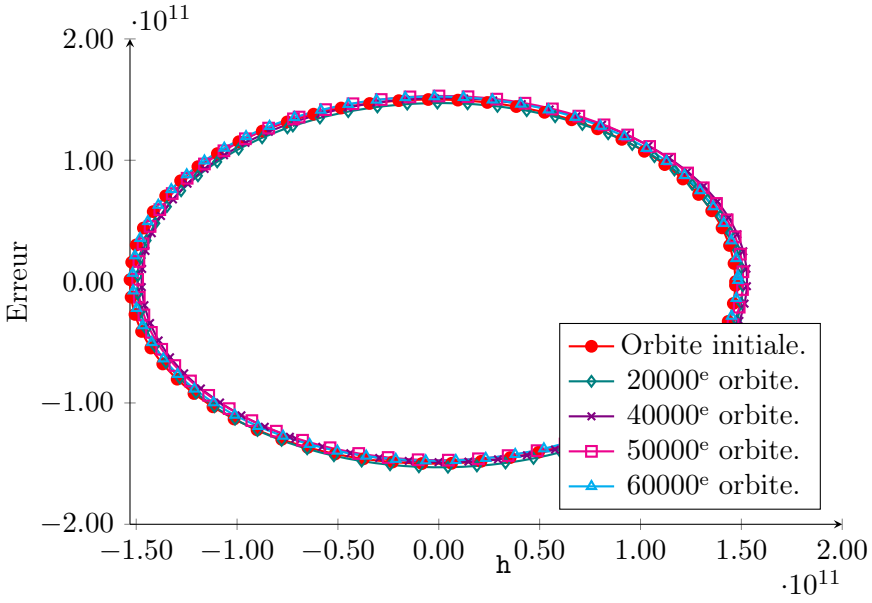


FIGURE 3.9 – Orbite de la Terre par Störmer-Verlet pour 64 pas de temps par période de révolution.

$10-11\mathrm{m}^3\,\mathrm{s}^{-2}\,\mathrm{kg}^{-1}$. Les conditions initiales sont :

$$\boldsymbol{r}_0 = \begin{bmatrix} 1.471 \cdot 10^{11}\mathrm{m} \\ 0\mathrm{m} \end{bmatrix}, \qquad \boldsymbol{v}_0 = \begin{bmatrix} 0\mathrm{m\,s}^{-1} \\ 3.03 \cdot 10^{4}\mathrm{m\,s}^{-1} \end{bmatrix}.$$

La trajectoire de la Terre est périodique. C'est une ellipse et la période de révolution terrestre est de $T = 31596130\mathrm{s}$. Nous pouvons donc mesurer numériquement l'ordre de la méthode RK4 et de la méthode de Störmer-Verlet en représentant l'erreur après une periode en fonction du pas de temps suivant une échelle loglog, voir Figure 3.7 et Table 3.6. Pour un temps court, la méthode RK4 est plus précise.

Regardons maintenant l'évolution des orbites pendant environ 60000 révolutions terrestres avec 64 pas de temps par période de révolution terrestre T. Pour voir quelque chose, nous ne traçons qu'une révolution terrestre toutes les 20000, voir Figure 3.8 et 3.9. La méthode de Störmer-Verlet a conservé le caractère périodique de la solution, même sur 60000 ans, alors que la méthode de Runge-Kutta 4 spirale lentement mais inexorablement vers le Soleil jusqu'à ce que la condition de A-stabilité ne soit plus vérifiée après environ 70000 ans pour 64 pas de temps par période de révolution terrestre, ce qui place alors la Terre sur une trajectoire hyperbolique. Ainsi la méthode de Störmer-Verlet, bien que d'ordre inférieur à la méthode RK-4, exhibe un bien meilleur comportement pour les temps longs.

Il est instructif, de comparer l'évolution de l'énergie mécanique et du moment cinétique au cours du temps pour les deux méthodes, voir la Tables 3.7 où on indique pour certaines orbites l'erreur absolue commise par les méthodes RK4 et de Störmer-Verlet. Pour la méthode RK4, l'erreur commise par RK4 sur ces deux grandeurs physiques évolue lentement. Mais cette évolution est inexorable. Cela ne se voit presque pas sur une seule révolution terrestre mais après plusieurs dizaines de milliers de révolution terrestre, les erreurs successives se sont accumulées. L'erreur sur le moment cinétique est multipliée par 10^6 en 70000 périodes de révolution terrestre. Celle sur l'énergie par 10^5 durant la même période. Pour la méthode de Störmer-Verlet, nous observons que l'erreur sur l'énergie mécanique reste stable même après 70000 révolutions autour du Soleil. La méthode de Störmer-Verlet ne conserve pas l'énergie mécanique mais l'erreur sur l'énergie mécanique reste bornée, même sur des temps très longs. On dit que la méthode de Störmer-Verlet conserve presque l'énergie mécanique globalement en temps. Pour les problèmes à force centrale comme le problème à deux corps en mécanique céleste, la méthode de Störmer-Verlet conserve le moment cinétique. Cela se voit assez

Orbite n°	RK4 ΔE J	Verlet ΔE J	RK4 $\Delta \mathcal{C}$ $\mathrm{kg\,m^2\,s^{-1}}$	Verlet $\Delta \mathcal{C}$ $\mathrm{kg\,m^2\,s^{-1}}$
1	$4.26 \cdot 10^{27}$	$5.06 \cdot 10^{29}$	$2.14 \cdot 10^{34}$	$9.67 \cdot 10^{24}$
2	$8.53 \cdot 10^{27}$	$5.06 \cdot 10^{29}$	$4.27 \cdot 10^{34}$	$2.42 \cdot 10^{25}$
10	$4.26 \cdot 10^{28}$	$5.05 \cdot 10^{29}$	$2.14 \cdot 10^{35}$	$1.45 \cdot 10^{25}$
100	$4.27 \cdot 10^{29}$	$5.06 \cdot 10^{29}$	$2.14 \cdot 10^{36}$	$1.93 \cdot 10^{25}$
1 000	$4.30 \cdot 10^{30}$	$5.05 \cdot 10^{29}$	$2.15 \cdot 10^{37}$	$2.27 \cdot 10^{26}$
5 000	$2.22 \cdot 10^{31}$	$5.06 \cdot 10^{29}$	$1.11 \cdot 10^{38}$	$1.53 \cdot 10^{27}$
10 000	$4.64 \cdot 10^{31}$	$5.05 \cdot 10^{29}$	$2.30 \cdot 10^{38}$	$5.22 \cdot 10^{26}$
20 000	$1.02 \cdot 10^{32}$	$5.06 \cdot 10^{29}$	$4.99 \cdot 10^{38}$	$1.13 \cdot 10^{27}$
30 000	$1.73 \cdot 10^{32}$	$5.06 \cdot 10^{29}$	$8.26 \cdot 10^{38}$	$3.63 \cdot 10^{27}$
40 000	$2.66 \cdot 10^{32}$	$5.05 \cdot 10^{29}$	$1.24 \cdot 10^{39}$	$4.98 \cdot 10^{27}$
50 000	$4.05 \cdot 10^{32}$	$5.06 \cdot 10^{29}$	$1.83 \cdot 10^{39}$	$6.68 \cdot 10^{27}$
60 000	$6.63 \cdot 10^{32}$	$5.05 \cdot 10^{29}$	$2.81 \cdot 10^{39}$	$7.12 \cdot 10^{27}$
70 000	$1.02 \cdot 10^{35}$	$5.06 \cdot 10^{29}$	$1.75 \cdot 10^{40}$	$8.60 \cdot 10^{27}$

TABLE 3.7 – Erreur absolue entre Énergie (E) mécanique exacte et Énergie mécanique calculée de la Terre dans le système Soleil-Terre avec 64 pas de temps par période de révolution pour les méthodes RK4 et de Störmer-Verlet. Même chose pour le moment cinétique ($\mathcal{C}$). Valeur exacte de l'énergie mécanique : $-2.65 \cdot 10^{33}$J. Valeur exacte du moment cinétique suivant $\boldsymbol{e}_z$: $2.66 \cdot 10^{40}\mathrm{kg\,m^2\,s^{-1}}$.

facilement à partir de (3.3.4).

$$\begin{aligned}\boldsymbol{r}_{n+1} \wedge \boldsymbol{v}_{n+1} &= \boldsymbol{r}_{n+1} \wedge (\boldsymbol{v}_{n+1/2} - h\frac{\mathcal{G}M_S}{|\boldsymbol{r}_{n+1}|^3}\boldsymbol{r}_{n+1}) \\ &= \boldsymbol{r}_{n+1} \wedge \boldsymbol{v}_{n+1/2} = (\boldsymbol{r}_n + h\boldsymbol{v}_{n+1/2}) \wedge \boldsymbol{v}_{n+1/2} \\ &= \boldsymbol{r}_n \wedge \boldsymbol{v}_{n+1/2} = \boldsymbol{r}_n \wedge (\boldsymbol{v}_n - h\frac{\mathcal{G}M_S}{|\boldsymbol{r}_n|^3}\boldsymbol{r}_n) \\ &= \boldsymbol{r}_n \wedge \boldsymbol{v}_n.\end{aligned}$$

Ainsi l'erreur mesurée sur le moment cinétique pour cette est de l'ordre du « epsilon-machine », cette erreur est multipliée par 1000 sur 70000 révolutions. Il ne peut s'agir que de la conséquence des erreurs d'arrondis puisque la méthode de Störmer-Verlet conserve le moment cinétique. Pour quelle raison la méthode de Störmer-Verlet se comporte-t-elle si bien pour les temps longs ? En particulier, pourquoi l'erreur sur l'énergie reste-t-elle stable alors que la méthode de Störmer-Verlet ne conserve pas exactement l'énergie mécanique. C'est parce que la méthode de Störmer-Verlet est une méthode symplectique. Les méthodes symplectiques constituent une famille de méthodes numériques de résolution d'une EDO. Ces méthodes se comportent bien sur les temps longs pour les problèmes issues de la mécanique analytique. La signification mathématique précise du mot « symplectique » dans ce contexte dépasse de très loin la portée de ce cours. Nous renvoyons le lecteur intéressé à l'excellent et très complet[2].

3.3.2 Méthodes de Newmark

Les méthodes de Newmark sont employées en mécanique des structures pour résoudre les équations du type :

$$\mathbf{M}\boldsymbol{r}'' + \mathbf{C}\boldsymbol{r}' + \mathbf{K}\boldsymbol{r} = \boldsymbol{F}$$

où $\mathbf{M}$ est une matrice symétrique définie positive et $\mathbf{K}$ une matrice symétrique semi-définie positive. Nous allons définir la méthode de Newmark pour une EDO plus générale.

$$\boldsymbol{r}'' = F(\boldsymbol{r}, \boldsymbol{r}'). \tag{3.3.6}$$

Définition 3.16: Les Méthodes de Newmark

Soient β et γ deux paramètres réels donnés. Pour résoudre numériquement l'EDO (3.3.6), la méthode de Newmark est définie par la relation de récurrence

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + h(\boldsymbol{v}_n + \frac{h}{2}((1-2\beta)F(\boldsymbol{r}_n, \boldsymbol{v}_n) + 2\beta F(\boldsymbol{r}_{n+1}, \boldsymbol{v}_{n+1}))), \tag{3.3.7a}$$

$$\boldsymbol{v}_{n+1} = \boldsymbol{v}_n + h((1-\gamma)F(\boldsymbol{r}_n, \boldsymbol{v}_n) + \gamma F(\boldsymbol{r}_{n+1}, \boldsymbol{v}_{n+1})). \tag{3.3.7b}$$

Le plus souvent, la méthode de Newmark est implicite. Si $\beta = 0$ et si F dépend uniquement de la position $\boldsymbol{r}$, alors la méthode de Newmark est explicite. Si en plus, $\gamma = 1/2$, alors on retrouve la méthode de Störmer-Verlet sous la forme (3.3.4).

L'ordre des méthodes de Newmark est 1 si $\gamma \neq 1/2$ et 2 si $\gamma = 1/2$. Aussi, presque toujours, on prendra $\gamma = 1/2$. On rencontre dans la littérature plusieurs valeurs du paramètre β. Un des choix courants est de prendre $\beta = 1/4$ (accélération constante). Comme la méthode de Störmer-Verlet, lorsque $\gamma = 1/2$, les méthodes de Newmark « conservent presque » l'énergie mécanique et du moment cinétique globalement en temps. En particulier, l'erreur sur l'énergie mécanique reste stable même sur les temps infinis.

3.4 Méthodes multipas

Jusqu'à présent, hormis dans la formule multipas de la méthode de Störmer-Verlet, Eq (3.3.3), toutes les méthodes que nous avons considérées sont des méthodes à un pas : la valeur de la solution numérique à l'itérée $n+1$ dépend seulement de la valeur à l'itérée n, de t_n et du pas de temps $h = t_{n+1} - t_n$. Dans cette section nous introduisons les méthodes multipas. Une méthode sera dite à m pas si la valeur de la solution numérique à l'itérée $n+1$, *i.e.*, $\boldsymbol{x}_{n+1}$, dépend de $\boldsymbol{x}_n, \ldots, \boldsymbol{x}_{n-m+1}$, de $t_n, \ldots, t_{n-m+1}$, et du pas $h = t_{n+1} - t_n$ qui dans ce document sera pris indépendent de n. Utiliser des méthodes multipas, au lieu de méthodes à un pas, a des avantages et des inconvénients. Parmi leurs avantages, nous verrons qu'il est beaucoup plus aisé de monter en ordre avec des méthodes multipas, *i.e.*, de concevoir des méthodes multipas d'ordre élevé. De plus, le nombre d'évaluations requises de f par itération est, à ordre donné, en général, beaucoup plus petit pour les méthodes multipas que celui requis par les méthodes de Runge-Kutta. Parmi les inconvénients, il y a la nécessité d'initialiser les premières itérées à l'aide d'autre méthode, ainsi que l'existence possible de solutions numériques non physiques. Dans ce document, nous nous limitons aux méthodes multipas linéaires.

3.4.1 Méthodes multipas linéaires et ordre

Les méthodes multipas linéaires sont les plus simples méthodes multipas. Elles ont pour forme générale :

$$\boldsymbol{x}^h_{n+1} = \sum_{j=0}^{m-1} a_j \boldsymbol{x}^h_{n-j} + h \sum_{j=-1}^{m-1} b_j \boldsymbol{f}_{n-j}, \tag{3.4.1}$$

où $\boldsymbol{f}_{n-j} = f(t_{n-j}, \boldsymbol{x}^h_{n-j})$. Une méthode multipas linéaire est ainsi entièrement définie par son nombre de pas m et ses coefficients a_j et b_j. Si $b_{-1} = 0$, il s'agit d'une méthode multipas explicite. Dans le cas contraire, lorsque $b_{-1} \neq 0$, la méthode est implicite. On remarque que si une méthode multipas

linéaire est explicite, chaque itération ne nécessite qu'une unique évaluation de f en $\boldsymbol{x}_n^h$.

Calculer les conditions nécessaires et suffisantes pour qu'une méthode multipas linéaire soit d'ordre p est beaucoup plus aisé pour les méthodes multipas linéaires que pour les méthodés de Runge.

Proposition 3.17: Ordre des méthodes multipas linéaires

Une méthode multipas linéaire (3.4.1) est d'ordre p si et seulement si :

$$\sum_{j=0}^{m-1} a_j = 1, \tag{3.4.2a}$$

$$\sum_{j=0}^{m-1} a_j(-j)^k + k\sum_{j=-1}^{m-1} b_j(-j)^{k-1} = 1 \quad \text{pour tout } 1 \leqslant k \leqslant p. \tag{3.4.2b}$$

Démonstration. On injecte la solution exacte $\boldsymbol{x}$ dans (3.4.1). On remplace $f(t_{n-j}, \boldsymbol{x}(t_{n-j}))$ par $\boldsymbol{x}'(t_{n-j})$ puis on effectue le développement de Taylor en t_n. On obtient

$$\begin{aligned}
\sum_{k=0}^{p} \frac{h^k}{k!}\boldsymbol{x}^{(k)}(t_n) &= \sum_{j=0}^{m-1} a_j \sum_{k=0}^{p} \frac{(-jh)^k}{k!}\boldsymbol{x}^{(k)}(t_n) \\
&\quad + h\sum_{j=-1}^{m-1} b_j \sum_{k=0}^{p-1} \frac{(-j)^k h^k}{k!}\boldsymbol{x}^{(k+1)}(t_n) + O(h^{p+1}), \\
&= \sum_{k=0}^{p} \frac{h^k}{k!} \sum_{j=0}^{m-1} a_j(-j)^k \boldsymbol{x}^{(k)}(t_n) \\
&\quad + \sum_{j=-1}^{m-1} b_j \sum_{k=1}^{p} \frac{(-j)^{k-1}h^k}{(k-1)!}\boldsymbol{x}^{(k)}(t_n) + O(h^{p+1}), \\
&= \sum_{k=0}^{p} \frac{h^k}{k!} \sum_{j=0}^{m-1} a_j(-j)^k \boldsymbol{x}^{(k)}(t_n) \\
&\quad + \sum_{k=1}^{p} \frac{kh^k}{k!} \sum_{j=-1}^{m-1} b_j(-j)^{k-1}\boldsymbol{x}^{(k)}(t_n) + O(h^{p+1}).
\end{aligned}$$

On identifie les termes en $h^k\boldsymbol{x}^{(k)}(t_n)/k!$ et on obtient (3.4.2). □

Pour pouvoir appliquer une méthode multipas, nous avons besoin de m conditions initiales $\boldsymbol{x}_0^h, \ldots, \boldsymbol{x}_{m-1}^h$. Le problème de Cauchy de départ ne nous donne que la valeur de $\boldsymbol{x}_0^h$. Il sera donc nécessaire d'initialiser les valeurs $\boldsymbol{x}_j^h$ pour j compris entre 1 et $m-1$. Pour ce faire, nous devrons employer une

$1/h$	AB2/EE	AB2/Naïve $x_1 = x_0$
8	$1.00 \cdot 10^{-3}$	$2.72 \cdot 10^{-1}$
16	$1.35 \cdot 10^{-5}$	$1.40 \cdot 10^{-1}$
32	$2.68 \cdot 10^{-5}$	$7.13 \cdot 10^{-2}$
64	$1.05 \cdot 10^{-5}$	$3.60 \cdot 10^{-2}$
128	$3.10 \cdot 10^{-6}$	$1.80 \cdot 10^{-2}$
256	$8.36 \cdot 10^{-7}$	$9.04 \cdot 10^{-3}$
512	$2.16 \cdot 10^{-7}$	$4.53 \cdot 10^{-3}$
1 024	$5.50 \cdot 10^{-8}$	$2.26 \cdot 10^{-3}$
2 048	$1.39 \cdot 10^{-8}$	$1.13 \cdot 10^{-3}$

TABLE 3.8 – Comparaison de l'erreur globale d'Adams-Bashforth-2 avec x_1 initialisé par Euler Explicite, et par $x_1 = x_0$, au temps $t = 1$ pour l'EDO $x' = \cos(t)x$.

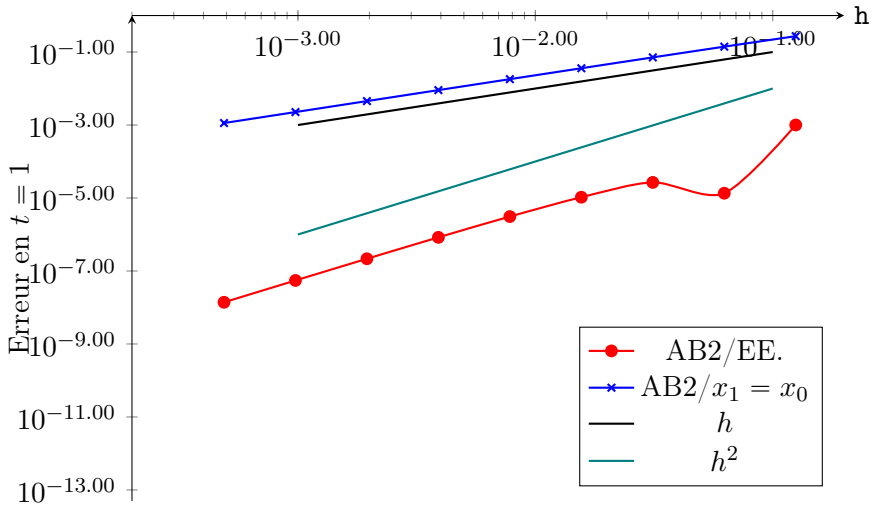


FIGURE 3.10 – Comparaison de l'erreur globale d'Adams-Bashforth-2 avec x_1 initialisé par Euler Explicite, et par $x_1 = x_0$, au temps $t = 1$ pour l'EDO $x' = \cos(t)x$.

méthode numérique de résolution d'EDO à au plus j pas pour calculer une valeur initiale de $\boldsymbol{x}_j^h$. Ainsi, une méthode à un pas sera nécessaire pour initialiser $\boldsymbol{x}_1^h$. Une méthode à au plus 2 pas le sera pour initialiser $\boldsymbol{x}_2^h$. En général, on utilisera une unique méthode à un pas pour réaliser cette initialisation. Pour garder l'ordre p d'une méthode à m pas, il sera nécessaire de connaître les valeurs $(\boldsymbol{x}_j^h)_{1\leqslant j\leqslant m-1}$ avec une erreur d'ordre p au plus. Si une méthode multipas est d'ordre p, il faudra donc utiliser des méthodes à un pas d'ordre au moins $p-1$ pour initialiser les m premières valeurs de $\boldsymbol{x}^h$ si l'on souhaite conserver cet ordre p. En effet, comme on utilise la méthode à 1 pas pour les $m-1$ premières itérées, l'erreur commise au bout de m itérées est du même ordre que l'erreur locale [4]. Cela peut s'observer numériquement à la Table 3.8 et sur la Figure 3.10 où nous avons employé une méthode multipas d'ordre 2, la méthode d'Adams-Bashforth, voir §3.4.2, en initialisant x_1 soit avec Euler Explicite, soit naïvement en posant $x_1 = x_0$. Seule la première version est d'ordre 2. La deuxième version est seulement d'ordre 1.

3.4.2 Les méthodes d'Adams

Dans cette section, nous introduisons les méthodes d'Adams-Bashforth et les méthodes d'Adams-Moulton qui sont des méthodes multipas. Les méthodes d'Adams-Bashforth sont des méthodes explicites et les méthodes d'Adams-Moulton sont des méthodes implicites.

Les méthodes d'Adams sont obtenues en partant de la formule :

$$\boldsymbol{x}(t_{n+1}) - \boldsymbol{x}(t_n) = \int_{t_n}^{t_{n+1}} f(s, \boldsymbol{x}(s))\mathrm{d}s. \tag{3.4.3}$$

Dans la méthode d'Adams-Bashforth à m pas, on interpole $f(s, \boldsymbol{x}(s))$ en $t_n, \ldots, t_{n-m+1}$ par un polynôme de degré $m-1$ noté P_m^{AB}. On remplace les $\boldsymbol{x}(t_j)$ par $\boldsymbol{x}_j^h$ dans la formule puis on injecte le polynôme dans le second membre de l'égalité (3.4.3). et on obtient la méthode d'Adams-Bashforth à m pas, noté ABm.

Nous allons appliquer ce procédé avec $m = 1$. Nous obtenons $P_1^{AB}(s) = f(t_n, \boldsymbol{x}_n)$. On obtient alors la méthode AB1 :

$$\boldsymbol{x}_{n+1}^h = \boldsymbol{x}_n^h + hf(t_n, \boldsymbol{x}_n^h). \tag{AB1}$$

On reconnaît la méthode d'Euler explicite.

Appliquons maintenant ce procédé avec $m = 2$. Nous avons

$$P_2^{AB}(s) = f(t_n, \boldsymbol{x}_n) + \frac{s - t_n}{h}(f(t_n, \boldsymbol{x}_n) - f(t_{n-1}, \boldsymbol{x}_{n-1})).$$

4. On perd un ordre quand on passe de l'erreur locale à l'erreur globale car il faut itérer environ T_{final}/h fois. On ne perd pas d'ordre lorque le nombre d'itérations à calculer est indépendant du pas de temps. Ainsi, pour une méthode à 1 pas d'ordre $\hat{p}$, l'erreur commise sur les itérées 1 à $m-1$ est en $O(h^{\hat{p}+1})$.

```
def AdamsBashforth2(f,phif,t0,x0,h):
    nbiter=int(math.ceil((tf-t0)/h))
    h=(tf-t0)/nbiter
    [x1,g,gg]=phif(f,t0,x0,h,nbiter=1)
    xm1=x0;
    tm1=t0;
    x=x1;
    t=t0+h;
    vf=f(t0,x0);
    for i in range(1,nbiter) :
        vfm1=vf;
        vf=f(t,x);
        tmpx=x;
        tmpt=t;
        x=x+h*(3.0/2.0*vf-1.0/2.0*vfm1);
        t=t+h;
        xm1=tmpx;
        tm1=tmpt;
    return x;
```

Code 3.5 – Méthode Adams-Bashforth 2 en Python

On obtient alors la méthode AB2 :

$$\boldsymbol{x}_{n+1}^h = \boldsymbol{x}_n^h + h\left(\frac{3}{2}f(t_n, \boldsymbol{x}_n) - \frac{1}{2}f(t_{n-1}, \boldsymbol{x}_{n-1})\right). \tag{AB2}$$

Une implémentation en Python de la méthode d'Adams-Bashforth est donnée au Code 3.5. Remarquer la fonction phif en argument qui sert à passer la méthode numérique à un pas pour initialiser x_1.

Dans la méthode d'Adams-Moulton à m pas, on interpole $f(s, \boldsymbol{x}(s))$ en $t_{n+1}, t_n, \ldots, t_{n-m+1}$ par un polynôme de degré m noté P_m^{AM}. On remplace les $\boldsymbol{x}(t_j)$ par $\boldsymbol{x}_j^h$ dans la formule puis on injecte le polynôme dans le second membre de l'égalité (3.4.3). Et on obtient la méthode d'Adams-Moulton à m pas, noté AMm.

Appliquons maintenant ce procédé avec $m = 1$, nous obtenons

$$P_1^{AM}(s) = f(t_{n+1}, \boldsymbol{x}_{n+1}) + \frac{s - t_{n+1}}{h}(f(t_{n+1}, \boldsymbol{x}_{n+1}) - f(t_n, \boldsymbol{x}_n)).$$

On obtient alors la méthode AM1 :

$$\boldsymbol{x}_{n+1}^h = \boldsymbol{x}_n^h + h\frac{f(t_{n+1}, \boldsymbol{x}_{n+1}) + f(t_n, \boldsymbol{x}_n)}{2}. \tag{AM1}$$

On reconnaît la méthode des trapèzes implicites.

On cherche maintenant à calculer la méthode AM2. Le polynôme P_2^{AM}

est donné par

$$P_2^{AM}(s) = f(t_{n+1}, \boldsymbol{x}_{n+1}) + \frac{s - t_{n+1}}{h}(f(t_{n+1}, \boldsymbol{x}_{n+1}) - f(t_n, \boldsymbol{x}_n) + (s - t_n)\frac{f(t_{n+1}, \boldsymbol{x}^h_{n+1}) - 2f(t_n, \boldsymbol{x}^h_n) + f(t_{n-1}, \boldsymbol{x}^h_{n-1})}{2h}).$$

On en déduit la méthode AM2.

$$\boldsymbol{x}^h_{n+1} = \boldsymbol{x}^h_n + h(\frac{5}{12}f(t_{n+1}, \boldsymbol{x}^h_{n+1}) + \frac{2}{3}f(t_n, \boldsymbol{x}^h_n) - \frac{1}{12}f(t_{n-1}, \boldsymbol{x}^h_{n-1})). \quad \text{(AM2)}$$

La méthode d'Adams-Bashforth à m pas est d'ordre m. La méthode d'Adams-Moulton à m pas est d'ordre $m + 1$.

3.4.3 Les méthodes BDF

Les méthodes BDF (backward differentiation formula) sont des méthodes implicites. Pour les construire, on part de l'égalité

$$\boldsymbol{x}'(t_{n+1}) = f(t_{n+1}, \boldsymbol{x}(t_{n+1})). \tag{3.4.4}$$

Pour construire la méthode BDF-m. On interpole les points $(t_{n-m+1}, \boldsymbol{x}^h_{n-m+1})$, …, $(t_n, \boldsymbol{x}^h_n)$, $(t_{n+1}, \boldsymbol{x}^h_{n+1})$ par un polynôme P_m^{BDF} de degré m. On remplace alors le côté gauche de l'égalité (3.4.4) par la dérivée de P_m^{BDF} en t_{n+1}. Nous allons construire les méthodes BDF-1 et BDF-2.

Pour la méthode BDF-1, nous avons

$$P_1^{BDF}(t) = \boldsymbol{x}^h_{n+1} + \frac{t - t_{n+1}}{h}(\boldsymbol{x}^h_{n+1} - \boldsymbol{x}^h_n).$$

La dérivée de P_1^{BDF} en t_{n+1} vaut $(\boldsymbol{x}^h_{n+1} - \boldsymbol{x}^h_n)/h$. Nous remplaçons le côté gauche de (3.4.4) par cette quantité et obtenons la méthode BDF-1 :

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}^h_n + hf(t_{n+1}, \boldsymbol{x}^h_{n+1}). \quad \text{(BDF-1)}$$

On reconnaît la méthode d'Euler implicite. On sait qu'il s'agit d'une méthode d'ordre 1.

Pour la méthode BDF-2, nous avons

$$P_2^{BDF}(t) = \boldsymbol{x}^h_{n+1} + \frac{t - t_{n+1}}{h}(\boldsymbol{x}^h_{n+1} - \boldsymbol{x}^h_n + \frac{t - t_n}{2h}(\boldsymbol{x}^h_{n+1} - 2\boldsymbol{x}^h_n + \boldsymbol{x}^h_{n-1})).$$

La dérivée de P_2^{BDF} en t_{n+1} vaut $(3\boldsymbol{x}^h_{n+1} - 4\boldsymbol{x}^h_n + \boldsymbol{x}^h_{n-1})/(2h)$. Nous remplaçons le côté gauche de (3.4.4) par cette quantité et obtenons la méthode BDF-2 :

$$\boldsymbol{x}^h_{n+1} = \frac{4}{3}\boldsymbol{x}^h_n - \frac{1}{3}\boldsymbol{x}^h_{n-1} + \frac{2h}{3}f(t_{n+1}, \boldsymbol{x}^h_{n+1}). \quad \text{(BDF-2)}$$

La méthode BDF à m pas est d'ordre m.

3.4.4 Stabilité et barrières de Dahlquist

Le concept de A-stabilité, exposé d'abord pour les méthodes à un pas, peut aussi s'appliquer aux méthodes multipas. Mais pour les méthodes multipas, la situation est plus compliqué. En effet, pour les méthodes à un pas, en prenant un pas de temps suffisamment (excessivement) petit, une méthode consistante convergera toujours, cf Théorème 3.8. Mais pour les méthodes multipas, il n'est même pas garanti qu'un pas de temps suffisamment petit permette d'obtenir une méthode stable.

Pour voir le problème intuitivement, considérons une EDO linéaire scalaire d'ordre 1, $x'(t) = \lambda x(t)$. L'ensemble des solutions exactes dorment un espace vectoriel de dimension 1. Considérons maintenant l'ensemble des solutions numériques obtenues par une méthode multipas linéaires d'ordre m :

$$\boldsymbol{x}_{n+1}^h = \sum_{j=0}^{m-1} a_j \boldsymbol{x}_{n-j}^h + h \sum_{j=-1}^{m-1} b_j \lambda \boldsymbol{x}_{n-j}^h.$$

L'ensemble des solutions numériques forment un sous-espace vectoriel de dimension m de l'espace des suites réelle. Autrement dit, nous essayons d'approcher un ensemble de solutions exacte qui forme une droite d'un espace par un ensemble de solutions numériques qui forment un sous-espace vectoriel de dimension m. Le sous-espace vectoriel des solutions numériques, est la somme directe d'une droite de solutions physique et d'un sous-espace vectoriel de dimension $m-1$ de solutions dites parasites.

Regardons sur un exemple. Pour la méthode (BDF-2),

$$(1 - \frac{2h\lambda}{3})\boldsymbol{x}_{n+1}^h = \frac{4}{3}\boldsymbol{x}_n^h - \frac{1}{3}\boldsymbol{x}_{n-1}^h$$

L'ensemble des solutions numériques est donc, voir §A.2, est

$$\mathbb{R}\big(\alpha^n\big)_{n\in\mathbb{N}} + \mathbb{R}\big(\beta^n\big)_{n\in\mathbb{N}}$$

où

$$\alpha = \frac{2 + \sqrt{1+2\lambda h}}{3 - 2\lambda h}, \qquad \beta = \frac{2 - \sqrt{1+2\lambda h}}{3 - 2\lambda h}.$$

Les solutions exactes de $x' = \lambda x$ sont $t \mapsto A\exp(\lambda t)$, où A est dans $\mathbb{R}$. De plus, $\alpha = 1 + \lambda h + o(h)$. Aussi, la droite $\mathbb{R}(\alpha^n)_{n\in\mathbb{N}}$ est l'ensemble des solutions numériques physiques. Et, la droite $\mathbb{R}(\beta^n)_{n\in\mathbb{N}}$ l'ensemble des solutions numériques parasites. Pour que les composantes parasites de la solution soient suffisamment petites pour être négligés, il faut d'une part que le choix des valeurs initiales $x_1^h, \ldots, x_{m-1}^h$ soit compatible avec l'équation de départ ; et d'autre part que l'amplitude des composantes parasites de la solution numériques n'augmente pas au cours du temps.

Aussi, nous introduisons le concept de 0-stabilité. Considérons l'EDO $x' = 0$, la solution est évidemment une constante. Si on applique une méthode multipas linéaire à cette équation, on obtient $x^h_{n+1} = \sum_{j=0}^{m-1} a_j x^h_{n-j}$. La solution d'une telle suite est donnée par une combinaison linéaire de $(n^k \zeta_i^n)$, où les ζ_i sont les racines complexes du polynôme $\zeta^m - \sum_{j=0}^{m-1} a_j \zeta^{m-1-j}$ et où k est un entier inférieur à la multiplicité de ζ_i dans ce polynôme. Si une des racines ζ_i a un module supérieur à 1 alors x^h_{n+1} va croître exponentiellement. Pour cette raison, on introduit un concept de stabilité pour les méthodes multipas linéaires.

Définition 3.18: 0-stabilité

Soit une méthode multipas linéaire donnée par (3.4.1). On note

$$\rho(\zeta) = \zeta^m - \sum_{j=0}^{m-1} a_j \zeta^{m-1-j}. \tag{3.4.5}$$

La méthode multipas linéaire donnée par (3.4.1) est dite 0-stable si

1. Les racines du polynôme ρ sont de module inférieur ou égal à 1.
2. Les racines du polynôme ρ dont le module vaut 1 sont simples.

Une méthode qui n'est pas 0-stable est complètement inutilisable. Quel que soit le pas de temps, aussi petit qu'il soit, la nome de la solution calculée par une méthode qui n'est pas 0-stable va croître extrêmement rapidement et dépasser le plus grand flottant représentable en machine. Diminuer le pas de temps ne fera qu'exacerber le problème. Nous verrons qu'une méthode multipas linéaire, a besoin d'être 0-stable pour être convergente.

Commençons par donner la définition de convergence d'une méthode multipas linéaire :

Définition 3.19 (Convergence)**.** *Une méthode multipas linéaire* (3.4.1) *à m pas est dite convergente si pour tout Ω dans $\mathbb{R} \times \mathbb{R}^d$, tout $f : \Omega \to \mathbb{R}^d$ de classe $\mathcal{C}^1(\Omega)$, tout $(t_0, \boldsymbol{x}_0)$ dans Ω, et tout $(\boldsymbol{x}^h_j)_{0 \leqslant j \leqslant m-1}$ vérifiant pour tout entier j compris entre 0 et $m-1$*

$$\lim_{h \to 0} \|\boldsymbol{x}^h_j - \boldsymbol{x}(t_0 + jh)\| = 0,$$

où $(I, \boldsymbol{x})$ est l'unique solution de $\boldsymbol{x}'(t) = f(t, \boldsymbol{x}(t))$, $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$, alors pour tout T dans I tel que $T \geqslant t_0$

$$\lim_{h \to 0} \sup_{\{n \in \mathbb{N} : t_0 + nh \leqslant T\}} \|\boldsymbol{x}^h_n - \boldsymbol{x}(t_0 + nh)\| = 0.$$

Puis, comme pour les méthodes à un pas, nous pouvons compléter cette notion de convergence par la notion de convergence d'ordre p

Définition 3.20 (Convergence d'ordre p)**.** *Une méthode multipas linéaire* (3.4.1) *à m pas est dite convergente d'ordre p si pour tout Ω dans $\mathbb{R}\times\mathbb{R}^d$, tout $f\colon\Omega\to\mathbb{R}^d$ de classe $\mathcal{C}^{p+1}(\Omega)$, tout $(t_0,\boldsymbol{x}_0)$ dans Ω, tout $C_0>0$, tout T dans I tel que $T\geqslant t_0$, il existe une constante $C>0$ telle que pour tout $(\boldsymbol{x}_j^h)_{j=0\leqslant j\leqslant m-1}$ vérifiant pour tout entier j compris entre 0 et $m-1$*

$$\|\boldsymbol{x}_j^h-\boldsymbol{x}(t_0+jh)\|\leqslant C_0h^p,$$

où $(I,\boldsymbol{x})$ est l'unique solution de $\boldsymbol{x}'(t)=f(t,\boldsymbol{x}(t))$, $\boldsymbol{x}(t_0)=\boldsymbol{x}_0$; alors, on a pour tout n dans $\mathbb{N}$ tel que $t+nh$ appartienne à $[t_0,T]\subset I$:

$$\|\boldsymbol{x}_n^h-\boldsymbol{x}(t_0+nh)\|\leqslant Ch^p.$$

Le théorème suivant lie la convergence des méthodes multipas linéaires à la 0-stabilité, condition qui n'existait pas pour les méthodes à un pas.

Théorème 3.21: Convergence des méthodes multipas linéaires.

Soit $p\geqslant 1$. Une méthode multipas linéaire (3.4.1) est convergente d'ordre p si et seulement si elle est 0-stable et d'ordre supérieur ou égal à p.

Démonstration. Voir la preuve du [3, chap. III.4, Théorème 4.5]. □

Évidemment, comme pour les méthodes à un pas, ce résultat n'est valable que lorsque le pas de temps h tend vers 0. Il ne dit rien sur la stabilité de la méthode multipas pour un pas de temps donné. Pour cela, nous avons besoin du concept de A-stabilité, mais appliqué aux méthodes multipas.

Considérons l'EDO $x'=\lambda x$. Si on applique une méthode multipas linéaire à cette équation, on obtient $(1-b_{-1}\lambda h)x_{n+1}^h=\sum_{j=0}^{m-1}(a_j+b_j\lambda h)x_{n-j}^h$. Cela nous permet de généraliser l'A-stabilité, voir Définition 3.9, aux méthodes multipas linéaires.

Définition 3.22: A-stabilité

Soit une méthode multipas linéaire donnée par (3.4.1). On pose

$$\begin{aligned}\pi(\zeta,z)&=\rho(\zeta)-z\sigma(\zeta)\\&=\zeta^m-\sum_{j=0}^{m-1}a_j\zeta^{m-1-j}-z\sum_{j=-1}^{m-1}b_j\zeta^{m-1-j}.\end{aligned}\tag{3.4.6}$$

Le domaine d'A-stabilité, noté S, d'une méthode multipas linéaire donnée par (3.4.1) est l'ensemble des z dans $\mathbb{C}$ tel que

1. Toutes les racines ζ_i de $\zeta\mapsto\pi(\zeta,z)$ vérifient $|\zeta_i|\leqslant 1$.

2. Toutes les racines ζ_i de $\zeta \mapsto \pi(\zeta, z)$ dont le module vaut 1 sont simples.

Une méthode multipas linéaire est dite inconditionnellement A-stable si $\mathbb{C}^- \subset S$.

Pour qu'une méthode multipas soit utilisable, il est nécessaire qu'elle soit 0-stable. Pour en plus pouvoir choisir le pas de temps uniquement selon la tolérance sur l'erreur et non en fonction de critères de stabilité, il faut utiliser une méthode multipas linéaire inconditionnellement A-stable. Mais deux théorèmes, appelées barrières de Dahlquist, établissent un certain nombres de contraintes sur ces méthodes multipas linéaires. Tout d'abord la première barrière de Dahlquist limite l'ordre des méthodes multipas linéaires 0-stable en fonction du nombre de pas.

Théorème 3.23: Première barrière de Dahlquist

L'ordre p d'une méthode multipas linéaire 0-stable à m pas satisfait :

$$p \leqslant m + 2 \text{ si } m \text{ est pair.}$$
$$p \leqslant m + 1 \text{ si } m \text{ est impair.}$$

De plus, si $b_{-1} \leqslant 0$, et donc en particulier lorsque la méthode multipas linéaire est explicite, on a

$$p \leqslant m$$

Démonstration. Voir la preuve du [3, chap. III.3, Théorème 3.5]. □

Enfin, la deuxième barrière de Dahlquist nous indique qu'il n'existe pas de méthode multipas linéaire inconditionnellement A-stable d'ordre arbitraire.

Théorème 3.24: Deuxième barrière de Dahlquist

Il n'existe aucune méthode multipas linéaire explicite inconditionnellement A-stable. Les méthodes multipas linéaires inconditionnellement A-stable sont au maximum d'ordre 2.

Démonstration. Voir la preuve du [4, chap V.1, Theorem 1.4]. □

Pour s'affranchir des barrières de Dahlquist, il est nécessaire d'utiliser des méthodes multipas non linéaires, comme les méthode de Runge-Kutta multipas. Ces méthodes dépassent la portée de ce cours.

Conclusion

Dans ce chapitre, nous avons introduit le principe de la discrétisation d'une EDO. Après avoir exposé les méthodes d'Euler, nous avons étudié les méthodes de Runge-Kutta, puis les méthodes multipas. Ce chapitre consacré à la résolution numérique des EDO peut sembler long. En réalité, il effleure à peine la surface des méthodes numériques existantes. Pour des raisons de temps, il nous a fallu faire l'impasse sur de nombreuses méthodes.

En particulier, nous n'avons exposé ni les méthodes à pas variables, ni les méthodes symplectiques (hormis la méthode de Verlet qui est symplectique), ni les méthodes de Runge-Kutta implicites, ni les méthodes de Runge-Kutta multipas. Les méthodes symplectiques sont des méthodes qui préservent approximativement certaines quantités conservatrices (comme l'énergie ou le moment cinétique) sur des temps longs. Nous avons vu leur utilité en comparant Verlet et Runge-Kutta 4 sur une EDO de mécanique céleste sur les temps longs. Les méthodes à pas variables sont des algorithmes pour adapter le pas de temps à la solution : on souhaite en effet utiliser un petit pas de temps là ou la solution varie beaucoup et un grand pas de temps là où elle est presque constante.

Pour aller au delà de ce cours, si un jour vous avez besoin plus tard de choisir des méthodes plus sophistiquées que celles présentées dans ce polycopié, le lecteur est invité à lire les ouvrages cités dans la bibliographie de ce chapitre. En particulier, les ouvrages d'Ernst Hairer et de Gerhard Wanner, spécialistes de la résolution numériques des équations différentielles raides, sont très exhaustifs, en particulier sur les méthodes symplectiques qui sont à privilégier pour les problèmes en temps longs dans lesquels certaines grandeurs physiques ou mécaniques sont conservées.

Références

[1] John C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 2016. isbn : 9781119121527.

[2] Ernst Hairer, Christian Lubich et Gerhard Wanner. *Geometric Numerical Integration : Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2013. isbn : 9783662050187.

[3] Ernst Hairer, P. Nørsett Syvert et Gerhard Wanner. *Solving Ordinary Differential Equations I : Nonstiff Problems*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2008. isbn : 9783540566700.

[4] Ernst Hairer et Gerhard Wanner. *Solving Ordinary Differential Equations II : Stiff and Differential - Algebraic Problems*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2013. isbn : 9783662099476.

Annexe A

Prérequis

Dans ce chapitre, nous rappelons ou introduisons des outils mathématiques nécessaires à certaines preuves de ce polycopié.

A.1 Manipulations matricielles

A.1.1 Décomposition de Schur

On peut toujours par un changement de variable orthonormale transformer une matrice en une matrice triangulaire supérieure :

Proposition A.1. *Soit* $\mathbf{A}$ *une matrice carré appartenant à* $\mathcal{M}_d(\mathbb{C})$*. Alors il existe une matrice orthonormal* $\mathbf{Q}$ *dans* $\mathcal{O}_d(\mathbb{C})$*, i.e., vérifiant* $\mathbf{Q}^*\mathbf{Q} = \mathbf{I}_d$*, tel que* $\mathbf{Q}^*\mathbf{A}\mathbf{Q}$ *est une matrice triangulaire supérieur.*

Démonstration. Cela se fait par récurrence sur d. Le résultat est évidemment vrai pour $d = 1$. Supposons le vrai pour un certain d dans $\mathbb{N}^*$. Soit $\mathbf{A}$ dans $\mathcal{M}_{d+1}(\mathbb{C})$. Alors, par le théorème de d'Alembert, le polynôme caractéristique de $\mathbf{A}$ admet au moins une racine complexe λ_1 dans $\mathbb{C}$. Soit $\boldsymbol{e}_1$ un vecteur propre de $\mathbf{A}$ pour la valeur propre λ_1. Soit $\mathbf{Q}_1$ une matrice orthonormale dont la première colonne est $\boldsymbol{e}_1$. Alors

$$\mathbf{Q}_1^*\mathbf{A}\mathbf{Q}_1 = \begin{bmatrix} \lambda_1 & * & \dots & * \\ 0 & & & \\ \vdots & & \mathbf{C} & \\ 0 & & & \end{bmatrix}$$

La matrice $\mathbf{C}$ appartient à $\mathcal{M}_d(\mathbb{C})$. Donc, d'après l'hypothèse de récurrence, il existe $\mathbf{R}$ dans $\mathcal{O}_d(\mathbb{C})$ tel que $\mathbf{R}^*\mathbf{C}\mathbf{R}$ soit une matrice triangulaire supérieur de $\mathcal{M}_d(\mathbb{C})$. On pose

$$\mathbf{Q}_2 = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \mathbf{R} & \\ 0 & & & \end{bmatrix}$$

et $\mathbf{Q} = \mathbf{Q}_1\mathbf{Q}_2$. □

De ce résultat, simple en apparence on peut déduire des résultats importants sur les matrices symétriques, hermitiennes et normales. Nous commençons par rappeler la définition des termes symétriques, normales et hermitiennes.

Définition A.2. *Soit* $\mathbf{A}$ *dans* $\mathcal{M}_d(\mathbb{C})$*. On dit que* $\mathbf{A}$ *est une matrice hermitienne si et seulement si* $\mathbf{A}^* = \mathbf{A}$*. On dit que* $\mathbf{A}$ *est normale si et seulement si* $\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*$*.*

Soit $\mathbf{B}$ *dans* $\mathcal{M}_d(\mathbb{R})$*. On dit que* $\mathbf{B}$ *est une matrice symétrique si et seulement si* $\mathbf{B}^\top = \mathbf{B}$*.*

Il est clair que toute matrice symétrique en tant que matrice réelle est hermitienne en tant que matrice complexe.

Nous avons ce résultat de diagonalisabilité pour les matrices hermitiennes :

Proposition A.3. *Pour toute matrice* $\mathbf{A}$ *dans* $\mathcal{M}_d(\mathbb{C})$ *hermitienne, il existe* $\mathbf{Q}$ *dans* $\mathcal{O}_d(\mathbb{C})$ *tel que* $\mathbf{Q}^*\mathbf{A}\mathbf{Q} = \mathbf{D}$ *où* $\mathbf{D}$ *est une matrice diagonale. De plus, les coefficients diagonaux de* $\mathbf{D}$ *sont tous réels.*

Démonstration. Soit $\mathbf{A}$ une matrice hermitienne. Soit $\mathbf{Q}$ dans $\mathcal{O}_d(\mathbb{C})$ tel que $\mathbf{Q}^*\mathbf{A}\mathbf{Q}$ soit une matrice triangulaire supérieure. Alors, $\mathbf{Q}^*\mathbf{A}\mathbf{Q}$ est aussi une matrice hermitienne. Une matrice qui est à la fois hermitienne et triangulaire supérieure est forcément diagonale. Et ses éléments diagonaux sont forcément réels. On note $\mathbf{D}$ cette matrice. □

Pour les matrices normales, on obtient un résultat similaire, mais dans ce cas, il n'y aucune garantie que les valeurs propres soient réelles.

Proposition A.4. *Pour toute matrice normale* $\mathbf{A}$ *dans* $\mathcal{M}_d(\mathbb{C})$*, il existe* $\mathbf{Q}$ *dans* $\mathcal{O}_d(\mathbb{C})$ *tel que* $\mathbf{Q}^*\mathbf{A}\mathbf{Q} = \mathbf{D}$ *où* $\mathbf{D}$ *est une matrice diagonale appartenant à* $\mathcal{M}_d(\mathbb{C})$*.*

Démonstration. Soit $\mathbf{A}$ une matrice normal. Soit $\mathbf{Q}$ dans $\mathcal{O}_d(\mathbb{C})$ tel que $\mathbf{Q}^*\mathbf{A}\mathbf{Q}$ soit une matrice triangulaire supérieure. Alors, $\mathbf{R} = \mathbf{Q}^*\mathbf{A}\mathbf{Q}$ est aussi une matrice normale. Donc $\mathbf{R}^*\mathbf{R} = \mathbf{R}\mathbf{R}^*$. En particulier,

$$(\mathbf{R}^*\mathbf{R})_{11} = |r_{11}|^2, \qquad (\mathbf{R}\mathbf{R}^*)_{11} = |r_{11}|^2 + \sum_{j=2}^{i-1} |r_{1j}|^2.$$

Donc, pour tout $j > 1$, $r_{1j} = 0$. On procède par récurrence en recommençant avec $(\mathbf{R}^*\mathbf{R})_{kk}$, lorsque k parcourt $[\![2, d]\!]$. On obtient que tous les éléments non diagonaux de $\mathbf{R}$ sont nuls. □

Il existe une version de A.3 pour les matrices symétriques :

Proposition A.5. *Pour toute matrice* $\mathbf{A}$ *dans* $\mathcal{M}_d(\mathbb{R})$ *symétrique, il existe* $\mathbf{Q}$ *dans* $\mathcal{O}_d(\mathbb{R})$ *tel que* $\mathbf{Q}^*\mathbf{A}\mathbf{Q} = \mathbf{D}$ *où* $\mathbf{D}$ *est une matrice diagonale. De plus, les coefficients diagonaux de* $\mathbf{D}$ *sont tous réels.*

Démonstration. Comme $\mathbf{A}$ est hermitienne en tant que matrice complexe, on obtient que toutes les valeurs propres de $\mathbf{A}$ sont réelles. On peut reprendre la preuve de la décomposition de Schur complexe en choisissant à chaque étape un vecteur propre réel, et en complétant à chaque étape par une base orthonormale faite de vecteurs propres réels. □

Il existe aussi des versions plus ou moins compliquées des Propositions A.1 et A.4 pour les matrices réelles. Nous ne les donnerons pas dans ce document.

A.1.2 Théorème de Cayley-Hamilton et trigonalisation par blocs

Le théorème de Cayley-Hamilton dont nous donnerons l'énoncé précis plus loin, nous dit que le polynôme caractéristique d'une matrice annule cette même matrice. Il est possible de démontrer le théorème directement par le calcul brutal en développant tous les termes mais ce n'est pas aisé, aussi nous allons utiliser une voie moins calculatoire. Nous allons commencer par un lemme :

Lemme A.6. *Soit* $(\mathbf{A}_n)_{1\leqslant n\leqslant d}$, d *matrices triangulaires supérieures appartenant à* $\mathcal{M}_d(\mathbb{C})$. *Si, de plus, pour tout entier* n *entre* 1 *et* d, *le* n^e *coefficient diagonal de* $\mathbf{A}_n$ *est nul alors*

$$\prod_{n=1}^{d} \mathbf{A}_n = \mathbf{0}.$$

Démonstration. Pour tout n entre 1 et d, on pose $\mathbf{C}_n$ égale à $\prod_{r=n}^{d} \mathbf{A}_r = \mathbf{0}$. Par récurrence, nous allons montrer que pour tout n entre 1 et d que les lignes n à d de $\mathbf{C}_n$ ne contiennent que des 0. On initialise à $n = d$. La d^{e} ligne de $\mathbf{A}_d$ ne contient que des 0 car $\mathbf{A}_d$ est triangulaire supérieure, carré de taille d, et que son d^{e} coefficient diagonal est nul. Supposons l'hypothèse de récurrence vérifiée en $n+1$ avec n compris entre 1 et $d-1$. Alors, posons $\mathbf{C}_n = \mathbf{A}_n\mathbf{C}_{n+1}$. Nous notons $c_{n;i,j}$ le coefficient à la ligne i et à la colonne j de $\mathbf{C}_n$. Pour $\mathbf{A}_n$, nous notons ce coefficient $a_{n;i,j}$. Soit i compris entre n et d, et j entre 1 et d, on a

$$\begin{aligned} c_{n;i,j} &= \sum_{k=1}^{d} a_{n;i,k}c_{n+1;k,j}, \\ &= \sum_{k=1}^{n} a_{n;i,k}c_{n+1;k,j} \text{ car } c_{n+1,k,j} = 0 \text{ quand } k \geqslant n+1, \\ &= 0 \text{ car } a_{n+1,i,k} = 0 \text{ quand } k \leqslant n \text{ car } i \geqslant n. \end{aligned}$$

Si on prend $n = 1$, on obtient que les lignes 1 à d de $\mathbf{C}_1$ sont nulles. Donc, la matrice $\mathbf{C}_1$ est nulle. □

Nous sommes maintenant prêts pour démontrer le théorème de Cayley-Hamilton :

Théorème A.7 (Cayley-Hamilton)**.** *Soit* $\mathbf{A}$ *une matrice dans* $\mathbb{C}$*. Soit* $P(X)$ *le polynôme défini par* $P(X) = \det(X\mathbf{I} - \mathbf{A})$*. Alors* $P(\mathbf{A}) = \mathbf{0}$*.*

Démonstration. Grâce à la décomposition de Schur, voir §A.1.1, on peut, sans perte de généralité, supposer que $\mathbf{A}$ est une matrice triangulaire supérieure. On factorise $P(X) = \prod_{i=1}^{d}(X - \lambda_i)$ où λ_i est le i^{e} élément diagonal de $\mathbf{A}$ Alors

$$\prod_{i=1}^{d}(\lambda_i\mathbf{I} - \mathbf{A})$$

vérifie les hypothèses du Lemme A.6 donc est la matrice nulle. □

Du théorème de Cayley-Hamilton, on déduit

Corollaire A.8. *Soit* $\mathbf{A}$ *une matrice de* $\mathcal{M}_d(\mathbb{K})$*. Soit* $P(X)$ *le polynôme caractéristique de* $\mathbf{A}$*. Soit* $\prod_{k=1}^{m}(X - \lambda_i)^{r_k}$ *la factorisation de* P*. Alors,*

$$\mathbb{C}^d = \bigoplus_{k=1}^{m} \operatorname{Ker}((\mathbf{A} - \lambda_i\mathbf{I}_d)^{r_k}).$$

Démonstration. Ce corollaire est valable pour les matrices réelles et pour les matrices complexes. C'est la conséquence directe du fait que $\mathbb{R}[X]$ et $\mathbb{C}[X]$ sont des anneaux principaux et du théorème de Bézout. L'étude de ces propriétés algébriques dépassant de loin le cadre de cette annexe, nous admettrons ce résultat. □

Une conséquence du théorème de Cayley-Hamilton et de son corollaire est l'existence d'une trigonalisation par blocs.

Proposition A.9 (Trigonalisation par bloc)**.** *Soit* $\mathbb{K}$ *Soit* $\mathbf{A}$ *une matrice appartenant à* $\mathcal{M}(\mathbb{K})$ *où* $\mathbb{K}$ *est soit* $\mathbb{R}$ *soit* $\mathbb{C}$*. Soit* $P(X)$ *le polynôme caractéristique de* $\mathbf{A}$*. Soit* $\prod_{k=1}^{m}(X - \lambda_i)^{r_k}$ *la factorisation de* P*.*

Alors $\mathbf{A}$ *est semblable dans* $\mathcal{M}_d(\mathbb{K})$ *à une matrice diagonale par blocs de la forme*

$$\begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C}_{m-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C}_m \end{bmatrix}$$

où $\mathbf{C}_k$ *admet comme unique valeur propre* λ_k *et est de taille* r_k*. En particulier, les matrices* $\mathbf{C}_k - \lambda_k\mathbf{I}_{r_k}$ *sont nilpotentes.*

De plus, si $\mathbb{K}$ est $\mathbb{C}$, alors on peut imposer que les différents blocs $\mathbf{C}_k$ soient triangulaires supérieurs. Dans ce cas, les éléments diagonaux de $\mathbf{C}_k$ sont forcément tous égaux à λ_k.

Démonstration. On choisit une base adaptée à la somme directe donnée au Corollaire A.8. Chaque terme de la somme directe est stable par application de $\mathbf{A}$. On obtient que $\mathbf{A}$ est semblable à une matrice diagonale par bloc. Pour tout bloc, $\mathbf{C}_k - \lambda_k \mathbf{I}$ est nilpotente donc $\mathbf{C}_k$ admet pour unique valeur propre λ_k. Donc, le polynôme caractéristique de $\mathbf{C}_k$ est une puissance de $X - \lambda_k$. Le polynôme caractéristique de la matrice $\mathbf{A}$ est le produit des polynômes caractéristiques des matrices $\mathbf{C}_k$. Donc, la taille du bloc $\mathbf{C}_k$ est forcément égale à r_k. Enfin, si on travaille dans les matrices complexe, on peut appliquer le décomposition de Schur sur chaque bloc pour obtenir des blocs triangulaires supérieurs. Le k^{e} bloc triangulaire supérieur admettant comme unique valeur propre λ_k, ses éléments diagonaux sont forcément tous égaux à λ_k. □

A.1.3 Forme de Jordan

Dans cette section nous allons montrer que toute matrice appartenant à $\mathcal{M}_d(\mathbb{R})$ est semblable à ce qu'on appelle une matrice de Jordan.

On appelle bloc de Jordan une matrice de la forme :

$$\mathbf{J}_{d,\lambda} = \begin{bmatrix} \lambda & 0 & \dots & \dots & \dots & 0 \\ 1 & \lambda & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & \lambda & 0 \\ 0 & \dots & \dots & 0 & 1 & \lambda \end{bmatrix}$$

On appelle matrice de Jordan une matrice dont les blocs diagonaux sont des blocs de Jordan. Ainsi $\mathbf{J}$ est une matrice de Jordan s'il existe r, $d_1, \dots, d_r$ et $\lambda_1, \dots, \lambda_r$ tel que

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{d_1,\lambda_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{d_2,\lambda_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_{d_{r-1},\lambda_{r-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_{d_r,\lambda_r} \end{bmatrix}$$

Théorème A.10 (Forme de Jordan)**.** *Toute matrice carré complexe $\mathbf{A}$ est semblable à une matrice de Jordan appelée forme de Jordan de $\mathbf{A}$. I.E., pour toute matrice $\mathbf{A}$ dans $\mathcal{M}_{,}d(\mathbb{C})$, il existe une matrice $\mathbf{P}$ dans $\mathcal{M}_d(\mathbb{C})$, inversible tel que $\mathbf{J} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ soit une matrice de Jordan.*

La forme de Jordan d'une matrice est unique à permutation des blocs près.

Démonstration. **La preuve est technique. Il n'est pas nécessaire de la lire. Elle est juste fournie pour satisfaire la curiosité éventuelle du lecteur. Seul le résultat du théorème sera utilisé.** On part de la trigonalisation par blocs, voir Proposition A.9. Il suffit donc de démontrer le théorème pour chaque bloc trigonal, *i.e.*, pour chaque matrice $\mathbf{B}_\lambda$ n'admettant qu'une unique valeur propre λ dans $\mathbb{C}$. Puis, en posant $\mathbf{K} = \mathbf{B}_\lambda - \lambda\mathbf{I}$, on s'aperçoit qu'il suffit de prouver le théorème pour les matrices nilpotentes. Pour cela, il faut pour toute matrice nilpotente construire une base dans laquelle la matrice nilpotente est une matrice de Jordan. L'idée basique serait de choisir un vecteur $\boldsymbol{e}_{1,0}$ puis de poser $\boldsymbol{e}_{1,j}$ égal à $\mathbf{K}^j\boldsymbol{e}_{1,0}$ ce qui donnerait un bloc de Jordan. Puis d'essayer de recommencer le procédé à partir d'un autre vecteur $\boldsymbol{e}_{2,0}$ pour obtenir un deuxième bloc de Jordan, et ainsi de suite. L'idée n'est pas compliquée mais la réalisation est délicate. En effet, décrire comment on choisit les différents vecteurs $\boldsymbol{e}_{i,0}$ n'est pas évident dans le cas le plus général.

Soit $\mathbf{K}$ une matrice nilpotente de taille $\hat{d}$. Pour tout entier j entre 0 et $\hat{d}$, on pose $E_j = \mathrm{Ker}(K^j)$. Les E_j sont des sous-espaces vectoriels imbriqués : E_j est inclus dans E_{j+1}. De plus, on a $\mathbb{R}^{\hat{d}} = E_{\hat{d}}$ et $E_0 = \{\mathbf{0}\}$. Pour tout j entre 0 et $\hat{d}$, soit F_j un supplémentaire de $\mathbf{K}E_{j+1} + E_{j-1}$ dans E_j. Nous allons montrer que

$$\mathbb{R}^{\hat{d}} = \bigoplus_{j=1}^{d}\left(\bigoplus_{\ell=0}^{j-1}\mathbf{K}^\ell F_j\right). \tag{A.1.1}$$

Montrons d'abord que la somme est directe. Soient $\boldsymbol{x}_{j,\ell}$ appartenant à F_j tels que $\sum_{j=1}^{d}\sum_{\ell=0}^{j-1}\mathbf{K}^\ell\boldsymbol{x}_{j,\ell} = 0$. Supposons qu'il existe (j,ℓ) tel que $\boldsymbol{x}_{j,\ell} \neq 0$. Posons $m = \max_{\{(j,\ell):\boldsymbol{x}_{j,\ell}\neq 0\}}(j-\ell)$. Forcément, $m \leqslant d$. Alors, on multiplie la somme par $\mathbf{K}^{m-1}$. On obtient $\sum_{j=m}^{d}\mathbf{K}^{j-1}\boldsymbol{x}_{j,j-m} = 0$. Soit n le plus petit entier j tel que $\boldsymbol{x}_{j,j-m}$ soit non nulle. Donc, $\sum_{j=n}^{d}\mathbf{K}^{j-n}\boldsymbol{x}_{j,j-m}$ appartient à E_{n-1}. Donc, $\boldsymbol{x}_{n,n-m}$ appartient à $E_{n-1} + \mathbf{K}E_{n+1}$ qui est un supplémentaire de F_n, donc le vecteur $\boldsymbol{x}_{n,n-m}$ est nul. Donc, pour tout (j,ℓ), $\boldsymbol{x}_{j,\ell}$ est nul. Il reste à montrer que cette somme directe est égal à E. Par récurrence, on montre que pour tout n dans $[\![0,d]\!]$:

$$E_n = \left(\bigdotplus_{j=n}^{d}\mathbf{K}^{j-n}F_j\right) + E_{n-1}, \qquad E = \left(\bigdotplus_{j=n}^{d}\bigdotplus_{\ell=0}^{j-n}\mathbf{K}^\ell F_j\right) + E_{n-1}.$$

On initialise à n égal à d, on a $E_d = F_d \oplus E_{d-1}$ car $\mathbf{K}E_{d+1} = \mathbf{K}E_{d+1}$ est inclus dans E_{d-1}. De plus, $E = E_d$ donc la deuxième égalité est vraie aussi pour $n = d$. Pour la récurrence, supposons les deux égalités ci-dessus vraies

pour un certain n vérifiant $n \geqslant 2$. Alors, par définition de F_{n-1}

$$\begin{aligned} E_{n-1} &= F_{n-1} + \left(\bigplus_{j=n}^{d} \mathbf{K}^{j-n+1} F_j \right) + \mathbf{K} E_n + E_{n-2}, \\ &= \left(\bigplus_{j=n-1}^{d} \mathbf{K}^{j-(n-1)} F_j \right) + E_{n-2}. \end{aligned}$$

De plus, on a

$$\begin{aligned} \mathbb{R}^{\hat{d}} &= \left(\bigplus_{j=n}^{d} \bigplus_{\ell=0}^{j-n} \mathbf{K}^{\ell} F_j \right) + E_{n-1} \\ &= \left(\bigplus_{j=n}^{d} \bigplus_{\ell=0}^{j-n} \mathbf{K}^{\ell} F_j \right) + \left(\bigplus_{j=n-1}^{d} \mathbf{K}^{j-n+1} F_j \right) + E_{n-2}, \\ &= \left(\bigplus_{j=n-1}^{d} \bigplus_{\ell=0}^{j-n+1} \mathbf{K}^{\ell} F_j \right) + E_{n-2}. \end{aligned}$$

La récurrence est terminée. En prenant $n = 1$, et en utilisant le résultat de somme directe précédent, nous avons au final démontré (A.1.1). Pour tout j tel que F_j est non vide, on pose $d_j = \dim(F_j)$ et on choisit une base $(\boldsymbol{e}_{j,k})_{1 \leqslant k \leqslant d_j}$ de F_j. Soit ℓ un entier compris entre 0 et $j-1$. Alors, il est évident que $(\mathbf{K}^{\ell} \boldsymbol{e}_{j,k})_{1 \leqslant k \leqslant d_j}$ forment une partie génératrice de $\mathbf{K}^{\ell} F_j$. Il reste à démontrer qu'il s'agit aussi d'une partie libre. Soit $(\lambda_k)_{1 \leqslant k \leqslant d_j}$ tel que

$$\sum_{k=1}^{d_j} \lambda_k \mathbf{K}^{\ell} \boldsymbol{e}_{j,k} = 0.$$

Donc $\sum_{k=1}^{d_j} \lambda_k \boldsymbol{e}_{j,k}$ appartient à E_ℓ. Mais cette somme appartient aussi à F_j qui est en somme directe avec E_{j-1} qui contient E_ℓ, donc $\sum_{k=1}^{d_j} \lambda_k \boldsymbol{e}_{j,k}$ est nul. Et comme les $\boldsymbol{e}_{j,k}$ forment une base de F_j, cela implique que tous les λ_k sont nuls.

Si $\mathbf{P}$ est la matrice de passage de la base canonique à la base $(\mathbf{K}^{\ell} \boldsymbol{e}_{j,k})_{\ell,k,j}$ où j va de 1 à $\hat{d}$, k va de 1 à d_j, et ℓ va de 0 à $d-1$; alors $\mathbf{P}^{-1}\mathbf{K}\mathbf{P}$ a la forme de Jordan. La taille des différents blocs de Jordan peut se déduire des dimensions des sous-espaces vectoriels $\mathrm{Ker}((\mathbf{A} - \lambda \mathbf{I})^j)$. En effet, si on appelle $k_{j,\lambda}$ le nombre de blocs de Jordan de taille j et de valeur propre λ alors

$$\begin{aligned} k_{d,\lambda} &= \dim(E_{d,\lambda}) - \dim(E_{d-1,\lambda}), \\ k_{j,\lambda} &= \dim(E_{j,\lambda}) - \dim(E_{j-1,\lambda}) - \sum_{\ell=j+1}^{d} k_{\ell,\lambda}, \end{aligned}$$

pour j compris entre 1 et $d-1$. Cela nous donne l'unicité. □

On en déduit le corollaire suivant :

Corollaire A.11. *Soit $\varepsilon \neq 0$. Toute matrice $\mathbf{A}$ appartenant à $\mathcal{M}_d(\mathbb{C})$ est semblable à une matrice de Jordan appartenant à $\mathcal{M}_d(\mathbb{C})$ où tous les 1 sur la première sous-diagonale dans les blocs de Jordan ont été remplacée par des ε.*

Démonstration. On met d'abord la matrice $\mathbf{A}$ sous forme de Jordan. Puis,on effectue le changement de variable suivant pour chaque bloc de Jordan :

$$\begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \varepsilon & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \ddots & \varepsilon^{j-1} & 0 \\ 0 & \dots & \dots & 0 & 0 & \varepsilon^{j} \end{bmatrix} \begin{bmatrix} \lambda & 0 & \dots & \dots & \dots & 0 \\ 1 & \lambda & 0 & \dots & \dots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & 1 & \lambda & 0 \\ 0 & \dots & \dots & 0 & 1 & \lambda \end{bmatrix} \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \varepsilon^{-1} & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \ddots & \varepsilon^{-(j-1)} & 0 \\ 0 & \dots & \dots & 0 & 0 & \varepsilon^{j} \end{bmatrix}$$

□

A.2 Suites récurrentes linéaires d'ordre m

Une suite définie par la relation de récurrence à :

$$u_{n+1} = a_{m-1}u_n + \ldots + a_0 u_{n-m+1} \tag{A.2.1}$$

et par la donnée de m conditions initiales $(u_i)_{0\leqslant i\leqslant m-1}$, est appelée une suite récurrente linéaire d'ordre m à coefficients constants. L'ensemble des suites vérifiant la relation (A.2.1) forment un espace vectoriel. Une suite vérifiant cette relation de récurrence étant univoquement caractérisée par la donnée de ses m premières valeurs, cet espace vectoriel est de dimension m. Pour trouver toutes les suites vérifiant la relation de récurrence (A.2.1), il suffit donc de trouver m suites linéairement indépendantes satisfaisant cette relation.

Pour calculer l'expression explicite de la valeur de u_n, on regarde les racines du polynôme

$$P(X) = X^m - a_{m-1}X^{m-1} - \ldots - a_0. \tag{A.2.2}$$

Si toutes les racines ζ_j de ce polynôme sont simples, alors toute suite satisfaisant les relations (A.2.1) est combinaison linéaire des m suites $(\zeta_j^n)_{n\in\mathbb{N}}$. *I.E.*, si la suite u satisfait (A.2.1), alors il existe des constantes $(b_j)_{1\leqslant j\leqslant m}$ tel que pour tout n dans $\mathbb{N}$:

$$u_n = \sum_{j=1}^{m} b_j \zeta_j^n \tag{A.2.3}$$

Supposons maintenant que le polynôme P, qui est de degré m, défini en (A.2.2) admet $\hat{m}$ racines multiples ζ_j. Soit r_j la multiplicité de ζ_j. On a

$\sum_j r_j = m$. Toute suite satisfaisant les relations (A.2.1) est combinaison linéaire des m suites $(n^s\zeta_j^n)_{n\in\mathbb{N}}$ lorsque j va de 1 à $\hat{m}$ et s va de 0 à r_j-1. *I.E.*, si la suite u satisfait (A.2.1), alors il existe des constantes $(b_{j,s})_{1\leqslant j\leqslant\hat{m},0\leqslant s\leqslant r_j-1}$ tel que pour tout n dans $\mathbb{N}$:

$$u_n = \sum_{j=1}^{\hat{m}} \sum_{s=0}^{r_j-1} b_{j,s} n^s \zeta_j^n \tag{A.2.4}$$

Table des matières

Sommaire **1**
1 Quelques exemples d'équations différentielles 2
1.1 Exemples de mécanique du point 3
1.2 Exemple de mécanique du solide indéformable 3
1.3 Exemple en Élasticité 5
1.4 Exemple de modélisation d'un circuit électrique RLC . 5
1.5 Exemple de désintégration atomique 6
1.6 Exemple de dynamique des populations 6
1.7 Exemple de Cinétique chimique 7
1.8 Exemple de Mécanique du vol 7
2 Équations différentielles d'ordre m 8

1 Existence et unicité des solutions **9**
1.1 Définitions et vocabulaire 9
1.2 Équations différentielles linéaires d'ordre 1 10
1.2.1 Lemme de Grönwall 12
1.3 Théorème de Cauchy-Lipschitz 13
1.4 Résolution des équations différentielles autonomes 15
1.5 Résolution des systèmes linéaires d'ordre 1 17
1.5.1 Le cas des coefficient constants 17
1.5.2 Le cas des coefficients non constants 18
1.6 Résolution des équations linéaires scalaires d'ordre 2 à coefficients constants 19
1.6.1 Le cas homogène 19
1.6.2 Cas non homogène 20
1.6.3 Généralisation à un ordre entier m quelconque 21

2 Étude de la stabilité **24**
2.1 Flot et dépendance de la solution par rapport aux conditions initiales 25
2.2 Stabilité, attractivité et stabilité asymptotique 26
2.2.1 Définitions 26
2.3 Stabilité par étude du spectre 28
2.4 Fonctions de Lyapounov 31

3 Résolution numérique d'une équation différentielle 37
3.1 Discrétisation d'une équation différentielle 37
3.1.1 Les méthodes d'Euler 38
3.1.2 Ordre, consistance et analyse de l'erreur 43
3.1.3 La A-stabilité 49
3.2 Méthodes de Runge-Kutta explicites 51
3.2.1 La méthode de Runge-Kutta 4 51
3.2.2 Forme générale des Méthodes de Runge-Kutta 52
3.2.3 Calcul inélégant de l'ordre d'une méthode de Runge-Kutta 55
3.2.4 Arbres et expressions différentielles 57
3.2.5 Limites des méthodes de Runge-Kutta 60
3.3 Méthodes de Newmark et de Störmer-Verlet 61
3.3.1 La méthode de Störmer-Verlet 61
3.3.2 Méthodes de Newmark 67
3.4 Méthodes multipas 68
3.4.1 Méthodes multipas linéaires et ordre 68
3.4.2 Les méthodes d'Adams 71
3.4.3 Les méthodes BDF 73
3.4.4 Stabilité et barrières de Dahlquist 74

A Prérequis 80
A.1 Manipulations matricielles 80
A.1.1 Décomposition de Schur 80
A.1.2 Théorème de Cayley-Hamilton et trigonalisation par blocs 82
A.1.3 Forme de Jordan 84
A.2 Suites récurrentes linéaires d'ordre m 87

Table des matières 89

Cours d'Équations différentielles

Auteur du polycopié : Kévin SANTUGINI

Filière Mathématiques et Mécanique, Semestre 5, UE M5-B, AM 105

Programme : Le but du cours d'Équations Différentielles (EDO) est d'apprendre les outils de bases qui permettent d'étudier le comportement des solutions d'équations différentielles. Après une brève introduction contenant des exemples d'équation différentielles provenant de la physique, nous aborderons trois grands chapitres : l'existence-unicité des solutions et le calcul des solutions exactes, la stabilité des solutions à une EDO, et les méthodes de résolution numérique des EDO.

Solutions exactes des équations différentielle

1. Définitions : Équations différentielles(EDO), Solutions d'une EDO, Problèmes de Cauchy.
2. Existence-Unicité des solutions. Lemme de Grönwall. Théorème de Cauchy-Lipschitz.
3. Résolution exacte pour
   - les EDO scalaires d'ordre 1 à variables séparables,
   - les systèmes linéaires d'EDO homogènes et à coefficients constants
   - les EDO linéaires scalaires d'ordre m, homogènes et à coefficients constants.
4. Méthodes de variation des constantes. Application aux systèmes linéaires d'EDO d'ordre 1 non homogènes. Wronskien. Application aux EDO linéaires scalaires d'ordre m.

Stabilité des solutions d'une EDO

1. Flot. Dépendance des solutions par rapport aux conditions initiales.
2. Définition de la stabilité au sens de Lyapounov, de l'attractivité et de la stabilité asymptotique.
3. Étude de la stabilité de la solution nulle d'un système linéaire d'EDO.
4. Étude de la stabilité d'une solution stationnaire d'une EDO non linéaire par étude du spectre de son linéarisé.
5. Fonctions de Lyapounov. Fonctions de Lyapounov strictes.

Méthodes numériques

1. Discrétisation des ODE. Méthodes d'Euler explicite et implicite. Autres méthodes à un pas.
2. Définitions : Consistance, Ordre et Convergence.
3. Définitions : Domaine de A-stabilité, A-stabilité inconditionnelle pour méthodes à 1 pas.
4. Méthodes de Runge-Kutta explicites :
   - Méthode du Point-Milieu explicite. Méthode des trapèzes explicites.
   - Méthode de Runge-Kutta 4. Méthode des Trois-Huitièmes.
   - Arbres orientés et calcul de l'ordre
   - Fonction de stabilité et Domaine de A-stabilité d'une méthode de Runge-Kutta.
5. Méthodes de Newmark et de Störmer-Verlet.
   - Calcul de l'ordre.
   - Conservation du moment cinétique d'un problème à force centrale par la méthode de Störmer-Verlet.
   - Observation du comportement en temps long de ces méthodes.
6. Méthodes multipas linéaires :
   - Méthodes d'Adams-Bashforth, d'Adams-Moulton et BDF.
   - Critères d'Ordre des méthodes multipas linéaires.
   - Dimension de l'espace vectoriel de toutes les solutions numériques obtenues par une méthode multipas linéaire pour une EDO linéaire homogène d'ordre 1. Solutions parasites.
   - 0-stabilité et A-stabilité des méthodes multipas.
   - Barrières de Dahlquist.

École nationale supérieure d'électronique, informatique, télécommunications, mathématiques et mécanique de Bordeaux
1 Avenue du Dr Albert Schweitzer
33400 Talence
France