



Département Sciences du Numérique

Équations différentielles ordinaires

Joseph Gergaud

17 septembre 2021

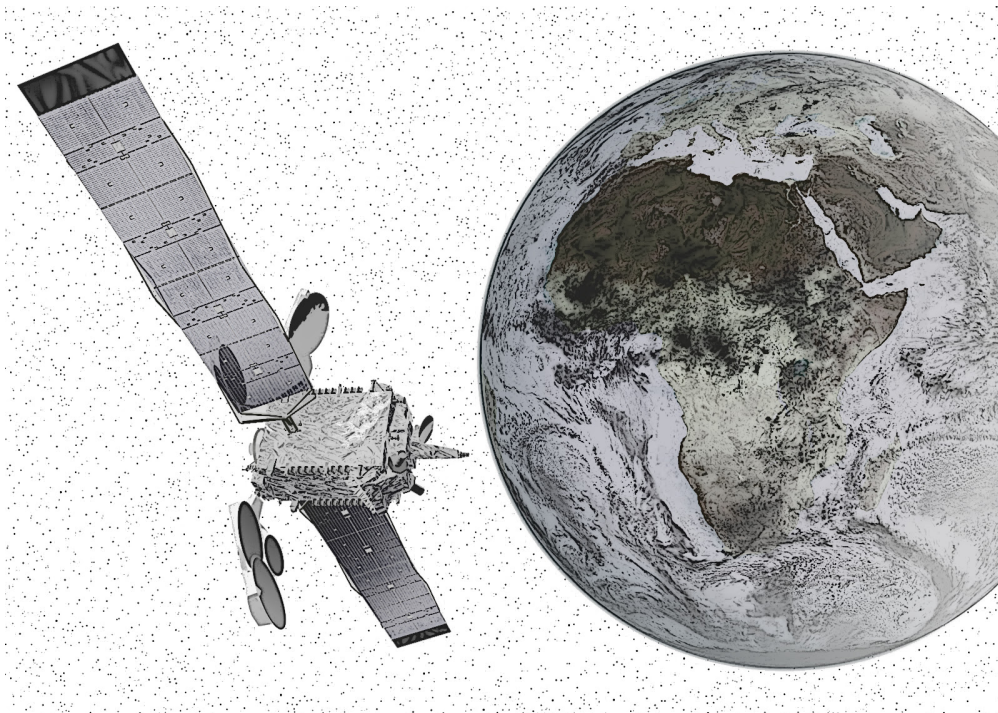


Table des matières

Chapitre 1. Introduction	1
1.1 Objectifs	1
1.2 Terminologie	4
1.3 Qu'est-ce qu'une solution ?	5
1.3.1 Solution classique	5
1.4 Plan du cours	7
Chapitre 2. Équations différentielles linéaires	8
2.1 Introduction	8
2.2 Équations différentielles linéaires homogènes autonomes	8
2.2.1 Approche élémentaire	8
2.2.2 Exponentielle de matrice	9
2.2.3 Calcul de l'exponentielle de matrice	12
2.2.4 Forme des solutions	14
2.3 Équations linéaires	16
2.3.1 Introduction	16
2.3.2 Existence et unicité de solution	16
2.3.3 Résolvante	17
2.3.4 Équations différentielles linéaires avec second membre	19
Chapitre 3. Théorie des équations différentielles	22
3.1 Existence	22
3.2 Dépendances par rapports aux données	29
3.2.1 Introduction	29
3.2.2 Continuité	30
3.2.3 Dérivée	32
Chapitre 4. Intégration numérique, les méthodes de Runge-Kutta	35
4.1 Introduction	35
4.2 Exemples	35
4.2.1 Exemple 1	35
4.2.2 Modèle de Lorenz	36
4.2.3 Exemple de Roberston	37
4.3 Définitions et exemples	38
4.4 Méthodes de Runge-Kutta explicite	40
4.4.1 Définition	40
4.4.2 Ordre	40
4.4.3 Convergence	43
4.5 Erreurs d'arrondi	48
4.6 Contrôle du pas	49
4.6.1 Introduction	49
4.6.2 Extrapolation de Richardson	49

4.6.3	Méthode de Runge-Kutta emboîtées	50
4.7	Les méthodes de Runge-Kutta implicites	54
4.8	Exercices	56
Chapitre 5.	Sortie dense, discontinuités, dérivées	58
5.1	Sortie dense	58
5.1.1	Objectif	58
5.1.2	Calcul de la sortie dense	58
5.1.3	Détection d'évènements	59
5.1.4	Intégration d'équations différentielles à second membre discontinues ..	60
5.2	Calcul de la dérivée	61
5.2.1	Exemple	61
5.2.2	Différences finies externes	61
5.2.3	Équation variationnelle	61
5.2.4	Différentiation interne de Bock (IND)	62
5.2.5	Exemple	62
Chapitre 6.		66
6.1	Espace de Banach	66
6.2	Théorèmes de points fixes	66
6.3	Topologie	66
6.4	Développement de Taylor	67
Bibliographie		69
Index		71
Index		71

Introduction

1.1 Objectifs

L'objectif de ce cours est l'étude mathématique, algorithmique et numérique des systèmes différentielles à condition initiale aussi appelé problème de Cauchy

$$(IVP)^1 \begin{cases} \dot{x}(t) = f(t, x(t)) \\ x(t_0) = x_0, \end{cases}$$

où $\dot{x}(t) = \frac{dx}{dt}(t)$ et f est une fonction

$$\begin{aligned} f: \Omega \subset \mathbf{R} \times \mathbf{R}^n &\longrightarrow \mathbf{R}^n \\ (s, y) &\longmapsto f(s, y), \end{aligned}$$

Ω ouvert et $(t_0, x_0) \in \Omega$. Le fait que Ω soit un **ouvert** est un point essentiel.

Remarque 1.1.1. Ici x est une fonction d'un intervalle ouvert I de \mathbf{R} contenant t_0 à valeur dans \mathbf{R}^n :

$$\begin{aligned} x: I &\longrightarrow \mathbf{R}^n \\ t &\longmapsto x(t), \end{aligned}$$

et $\dot{x}(t) = \frac{dx}{dt}(t)$.

Définition 1.1.1

L'équation $\dot{x}(t) = f(t, x(t))$ s'appelle une équation différentielle.

Exemple 1.1.1 (Circuit RLC). On considère le circuit de la figure 1.1 et on note $i(t) = \dot{q}(t)$. Le bilan des tensions conduit à l'équation différentielle linéaire du deuxième ordre :

$$L\ddot{q}(t) + R\dot{q}(t) + \frac{q(t)}{C} = 0. \tag{1.1}$$

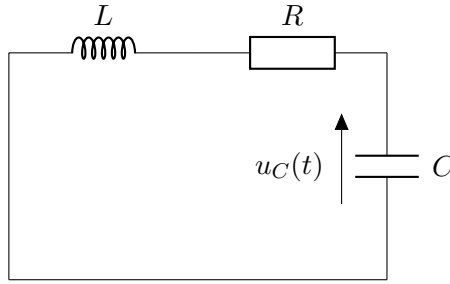
On pose $x_1(t) = q(t)$ et $x_2(t) = \dot{q}(t)$ alors l'équation 1.1 est équivalente au système d'équation

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = -\frac{R}{L}x_2(t) - \frac{1}{LC}x_1(t). \end{cases}$$

ici f s'écrit

$$\begin{aligned} f: \mathbf{R} \times \mathbf{R}^2 &\longrightarrow \mathbf{R}^2 \\ (s, y) &\longmapsto f(s, y) = \begin{pmatrix} y_2 \\ -\frac{R}{L}y_2 - \frac{1}{LC}y_1 \end{pmatrix}. \end{aligned}$$

1. Initial Value Problem.

FIGURE 1.1 – *Circuit RLC.*

On peut aussi écrire $f(s, y) = Ay$ avec

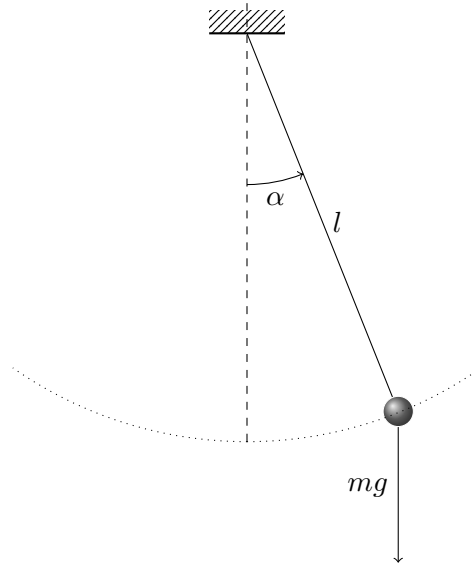
$$\begin{pmatrix} 0 & 1 \\ -\frac{1}{LC} & -\frac{R}{L} \end{pmatrix},$$

et le système est dit linéaire, à coefficients constants et sans second membre. \square

Exemple 1.1.2 (Pendule simple). On considère le pendule de la figure 1.2. Les principes physiques de la mécanique classique donnent comme équation qui régit l'évolution du mouvement

$$ml^2\ddot{\alpha}(t) + mlg \sin(\alpha(t)) = 0,$$

où $\ddot{\alpha}(t)$ désigne la dérivée seconde de l'angle α par rapport au temps t .

FIGURE 1.2 – *Pendule simple.*

On prend ici comme variable d'état qui décrit le système $x(t) = (x_1(t), x_2(t)) = (\alpha(t), \dot{\alpha}(t))$. Le système différentiel du premier ordre que l'on obtient s'écrit alors

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = -\frac{g}{l} \sin(x_1(t)) \\ x_1(0) = x_{0,1} = \alpha_0 \\ x_2(0) = x_{0,2} = \dot{\alpha}_0 \end{cases}$$

Cette équation s'écrit

$$\begin{cases} \dot{x}(t) = f(t, x(t)) \\ x(0) = x_0, \end{cases}$$

avec

$$\begin{aligned} f: \mathbf{R} \times \mathbf{R}^2 &\longrightarrow \mathbf{R}^2 \\ (s, y) &\longmapsto f(s, y) = \begin{pmatrix} y_2 \\ -\frac{g}{l} \sin(y_1) \end{pmatrix}. \end{aligned}$$

Ici le système est non linéaire car la fonction \sin n'est pas linéaire. Si on fait l'approximation des petits angles $\sin x_1 \approx x_1$, le système devient linéaire à coefficient constant et sans second membre : $\dot{x}(t) = Ax(t)$, avec

$$A = \begin{pmatrix} 0 & 1 \\ -\frac{g}{l} & 0 \end{pmatrix}.$$

□

Remarque 1.1.2. Dans les deux exemples précédents, la fonction f ne dépend pas du premier argument s . On peut donc aussi écrire l'équation différentielle sous la forme $\dot{x}(t) = g(x(t))$ avec


$$\begin{aligned} g: \mathbf{R}^2 &\longrightarrow \mathbf{R}^2 \\ (y) &\longmapsto g(y) = f(s, y). \end{aligned}$$

 **Exercice 1.1.3.** On considère l'équation différentielle

$$(IVP) \begin{cases} \dot{x}_1(t) = x_1(t) + x_2(t) + \sin t \\ \dot{x}_2(t) = -x_1(t) + 3x_2(t) \\ x_1(0) = -9/25, x_2(0) = -4/25. \end{cases}$$

1. Écrivez la fonction f permettant d'écrire le système différentiel $\dot{x}(t) = f(t, x(t))$.
2. Écrivez $f(s, y) = Ay + b(s)$. On donnera A , matrice $(2, 2)$ et la fonction b .

□

 **Exercice 1.1.4** (Modèle de Lorenz (effet papillon)). On considère l'équation différentielle

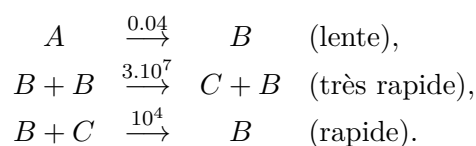
$$(IVP) \begin{cases} \dot{x}_1(t) = -\sigma x_1(t) + \sigma x_2(t) \\ \dot{x}_2(t) = -x_1(t)x_3(t) + rx_1(t) - x_2(t) \\ \dot{x}_3(t) = x_1(t)x_2(t) - bx_3(t) \\ x_1(0) = -8, x_2(0) = 8, x_3(0) = r - 1, \end{cases}$$

avec $\sigma = 10, r = 28, b = 8/3$.

1. Écrivez la fonction f permettant d'écrire le système différentiel $\dot{x}(t) = f(t, x(t))$.
2. Peut-on écrire ici $f(s, y) = Ay$, A matrice constante ?

□

 **Exercice 1.1.5** (exemple de Roberston). On considère la réaction chimique



Le système différentiel associé à cette réaction chimique est donnée par

$$(IVP) \begin{cases} \dot{x}_1(t) = & -0.04x_1(t) + 10^4x_2(t)x_3(t) \\ \dot{x}_2(t) = & 0.04x_1(t) - 10^4x_2(t)x_3(t) & -3.10^7x_2^2(t) \\ \dot{x}_3(t) = & & 3.10^7x_2^2(t) \\ x_1(0) = 1, x_2(0) = 0, x_3(0) = 0, \end{cases}$$

1. Écrivez la fonction f permettant d'écrire le système différentiel $\dot{x}(t) = f(t, x(t))$.
2. Peut-on écrire ici $f(s, y) = Ay$, A matrice constante? □

Remarque 1.1.3. On peut très bien écrire f de la façon suivante (par exemple pour le second exemple)

$$\begin{aligned} f: \mathbf{R} \times \mathbf{R}^2 &\longrightarrow \mathbf{R}^2 \\ (a, b) &\longmapsto f(a, b) = \begin{pmatrix} b_2 \\ -\frac{g}{l} \sin(b_1) \end{pmatrix}, \end{aligned}$$

ou encore

$$\begin{aligned} f: \mathbf{R} \times \mathbf{R}^2 &\longrightarrow \mathbf{R}^2 \\ (t, x) &\longmapsto f(t, x) = \begin{pmatrix} x_2 \\ -\frac{g}{l} \sin(x_1) \end{pmatrix}. \end{aligned}$$

Dans la suite on prendra toujours comme argument de f , les variables t et x , ceci afin de ne pas multiplier les notations. C'est le contexte qui fera la différence entre les significations de x : dans la définition de la fonction f c'est une variable de \mathbf{R}^n et dans l'équation différentielle $\dot{x}(t) = f(t, x(t))$, c'est une fonction d'un intervalle ouvert de \mathbf{R} à valeurs dans \mathbf{R}^n .

Remarque 1.1.4. Nous ne traiterons ni les équations à retard[7]

$$(DDE)^a \begin{cases} \dot{x}(t) = f(t, x(t), x(t - \tau)) \\ x(t) = f(t) \quad \forall t \in [t_0 - \tau, t_0], \end{cases}$$

ni les équations différentielles algébriques[8]

$$(DAE)^b \begin{cases} \dot{x}(t) = f(x(t), z(t)) \\ \psi(x(t), z(t)) = 0 \\ x(t_0) = x_0, z(t_0) = z_0. \end{cases}$$

^a. Delay Differential Equation.

^b. Differential Algebraic Equation.

1.2 Terminologie

- On appellera équation différentielle ou système dynamique toute équation $\dot{x}(t) = f(t, x(t))$ (c'est-à-dire sans la condition initiale).
- Si la fonction f ne dépend pas explicitement du temps, c'est-à-dire que l'équation différentielle s'écrit $\dot{x}(t) = f(x(t))$, on dit que l'équation différentielle est autonome.
- L'équation différentielle est dite linéaire² si elle s'écrit $\dot{x}(t) = A(t)x(t) + b(t)$.
 - $x : I \longrightarrow \mathbf{R}^n$;
 - $A : I \longrightarrow \mathcal{L}(\mathbf{R}^n, \mathbf{R}^n) \equiv \mathcal{M}_n(\mathbf{R})$;

2. En français, on devrait dire équation affine.

— $b : I \longrightarrow \mathbf{R}^n$.

- Elle est dite linéaire et homogène ou sans terme constant si elle s'écrit $\dot{x}(t) = A(t)x(t)$;
- Elle est dite linéaire à coefficients constants si elle s'écrit $\dot{x}(t) = Ax(t) + b$.
- On appelle dimension de l'équation différentielle $\dot{x} = f(t, x(t))$ la dimension n de $x(t)$.
- On appelle équation différentielle d'ordre m une équation qui s'écrit

$$x^{(m)}(t) = g(t, x(t), \dots, x^{(m-1)}(t)).$$

1.3 Qu'est-ce qu'une solution ?

La première question qui se pose est de savoir ce que l'on entend par une solution de (IVP).

1.3.1 Solution classique

Définition 1.3.1

[Définition classique] On suppose f définie et continue sur un ouvert Ω de \mathbf{R}^{n+1} à valeurs dans \mathbf{R}^n . On appelle solution classique de (IVP) tout couple (I, x) , I intervalle ouvert de \mathbf{R} , contenant t_0 et $x : I \rightarrow \mathbf{R}^n$ dérivable en tout point et vérifiant

1. $(t, x(t)) \in \Omega, \forall t \in I$
2. $\dot{x}(t) = f(t, x(t)), \forall t \in I$
3. $x(t_0) = x_0$.


Une solution est aussi appelée courbe intégrale de l'équation différentielle.

Remarque 1.3.1. Si f est continue (respectivement C^k) et (I, x) est une solution alors x est C^1 (respectivement C^{k+1}).

 **Exercice 1.3.1.** Vérifiez que la fonction

$$\begin{aligned} \varphi : \mathbf{R} &\longrightarrow \mathbf{R}^2 \\ t &\longmapsto \begin{pmatrix} -\frac{1}{25}(13 \sin t + 9 \cos t) \\ -\frac{1}{25}(3 \sin t + 4 \cos t) \end{pmatrix} \end{aligned}$$

est solution du système différentiel à condition initiale de l'exercice 1.1.3. □

 **Exercice 1.3.2.** On considère le problème de Cauchy définie sur $\Omega = \mathbf{R} \times \mathbf{R}$

$$(IVP1) \begin{cases} \dot{x}(t) = -x^2(t) \\ x(t_0) = x_0, \end{cases}$$

où (t_0, x_0) est fixé. Vérifiez que l'on a les solutions :

- Si $x_0 = 0$,

$$\begin{aligned} \varphi : I] -\infty, +\infty[&\longrightarrow \mathbf{R} \\ t &\longmapsto \varphi(t) = 0 \end{aligned}$$

- Si $x_0 > 0$,

$$\begin{aligned} \varphi : I] t_0 - 1/x_0, +\infty[&\longrightarrow \mathbf{R} \\ t &\longmapsto \varphi(t) = \frac{x_0}{(t-t_0)x_0 + 1}; \end{aligned}$$

- Si $x_0 < 0$,

$$\begin{aligned} \varphi: \quad I =]-\infty, t_0 - 1/x_0[&\longrightarrow \mathbf{R} \\ t &\longmapsto \varphi(t) = \frac{x_0}{(t-t_0)x_0+1}. \end{aligned}$$

Ces résultats sont visualisés sur les figures 1.3 et 1.4.

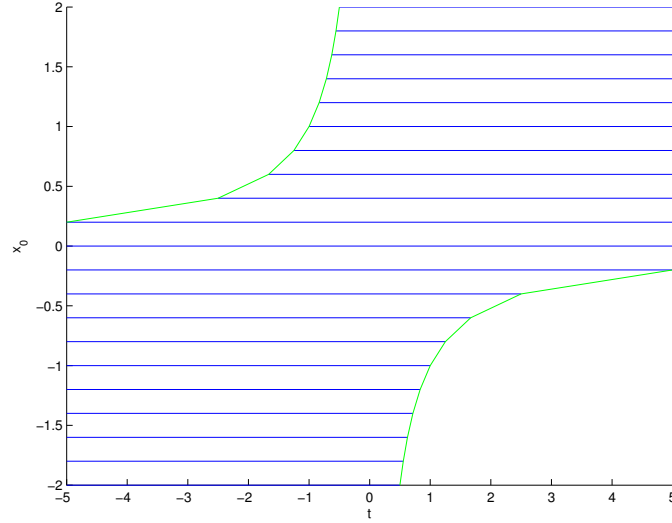


FIGURE 1.3 – Visualisation de l'ensemble $]\omega_-(0, x_0), \omega_+(0, x_0)[\times x_0$ pour l'exemple $\dot{x}(t) = -x^2(t), x(0) = x_0$.

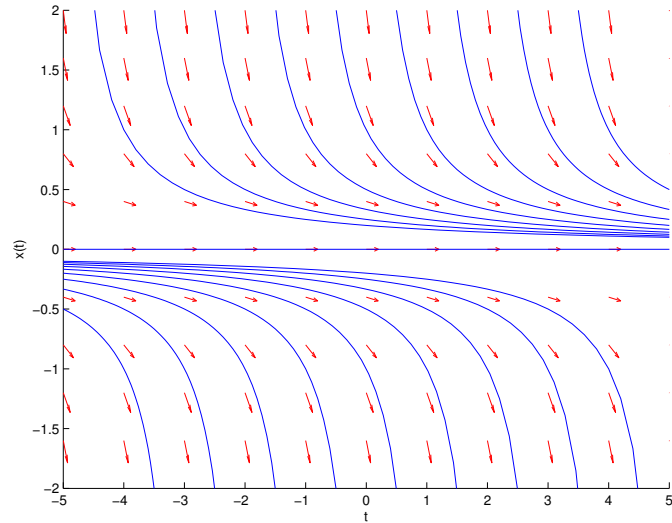


FIGURE 1.4 – Solutions pour le problème $\dot{x}(t) = -x^2(t), x(t_0) = x_0$, les courbes intégrales ne peuvent se couper, elles forment une partition de l'ouvert $\Omega = \mathbf{R}^2$.

□

Théorème 1.3.2

On suppose f continue et Ω ouvert. Alors (I, x) est une solution de (IVP) si et seulement si $t_0 \in I$, $(t, x(t)) \in \Omega$ pour tout t dans I , et

$$x(t) = x_0 + \int_{t_0}^t f(s, x(s)) ds.$$

► Si (I, x) est une solution alors x est C^1 , par suite

$$x(t) = c + \int_{t_0}^t \dot{x}(s) ds.$$

Le résultat est alors immédiat.

Réciproquement si

$$x(t) = x_0 + \int_{t_0}^t f(s, x(s)) ds,$$

alors, $x(t_0) = x_0$ et $\dot{x}(t) = f(t, x(t))$. ■

1.4 Plan du cours

Maintenant que l'on a défini le problème traité, nous allons dans ce cours étudier tout d'abord les aspects mathématiques qui comprendront les équations différentielles ordinaires linéaires, le théorème d'existence de solution de Cauchy-Lipschitz et la dépendance de la solution par rapport aux données. Dans une deuxième partie, nous verrons l'intégration numérique via les méthodes de Runge-Kutta.

Ce cours n'est bien sûr qu'une introduction au domaine. Concernant la deuxième partie par exemple une bonne référence est donnée par les 3 livres du professeur E. Hairer et al. [7, 8, 6] qui font plus de 300 pages chacun !

Équations différentielles linéaires

2.1	Introduction	8
2.2	Équations différentielles linéaires homogènes autonomes	8
2.2.1	Approche élémentaire	8
2.2.2	Exponentielle de matrice	9
2.2.3	Calcul de l'exponentielle de matrice	12
2.2.4	Forme des solutions	14
2.3	Équations linéaires	16
2.3.1	Introduction	16
2.3.2	Existence et unicité de solution	16
2.3.3	Résolvante	17
2.3.4	Équations différentielles linéaires avec second membre	19

2.1 Introduction

Dans tout ce chapitre I sera un intervalle de \mathbf{R} et \mathbf{K} sera le corps des réels ou des complexes. L'objectif de ce chapitre est l'étude mathématique des équations différentielles linéaires

$$\dot{x}(t) = A(t)x(t) + b(t),$$

où

- $x : I \longrightarrow \mathbf{K}^n$;
- $A : I \longrightarrow \mathcal{L}(\mathbf{K}^n, \mathbf{K}^n) \equiv \mathcal{M}_n(\mathbf{K})$;
- $b : I \longrightarrow \mathbf{K}^n$.

Ce chapitre est fortement inspiré de l'ouvrage de Frédéric Jean [9]. Nous commencerons par le cas des équations linéaires homogènes et autonomes $\dot{x}(t) = Ax(t)$, puis nous étudierons le cas non autonome $\dot{x}(t) = A(t)x(t)$ et le cas générale des équations linéaires $\dot{x}(t) = A(t)x(t) + b(t)$.

2.2 Équations différentielles linéaires homogènes autonomes

2.2.1 Approche élémentaire

Dans cette sous section le corps \mathbf{K} sera le corps des réels \mathbf{R} .

Exemple 2.2.1. On considère l'équation différentielle ordinaire linéaire scalaire

$$(IVP1) \begin{cases} \dot{x}(t) = \lambda x(t) \\ x(t_0) = x_0, \end{cases}$$

où λ est un réel et x est une fonction de \mathbf{R} dans \mathbf{R} . On sait que la solution de cette équation, qui est unique (cf. le théorème 2.3.2 ci-après), est donnée par

$$x(t) = e^{\lambda(t-t_0)}x_0.$$

On en déduit que cette solution est définie sur $I = \mathbf{R}$ et que l'on a comme comportement asymptotique

- Si $\lambda < 0$ alors $\lim_{t \rightarrow +\infty} x(t) = 0$;
- Si $\lambda = 0$ alors $x(t) = x_0$;
- Si $\lambda > 0$
 - Si $x_0 < 0$ alors $\lim_{t \rightarrow +\infty} x(t) = -\infty$;
 - Si $x_0 = 0$ alors $x(t) = 0$;
 - Si $x_0 > 0$ alors $\lim_{t \rightarrow +\infty} x(t) = +\infty$.

□

Exemple 2.2.2. Considérons maintenant un système de deux équations différentielles

$$(IVP2) \begin{cases} \dot{x}_1(t) = \lambda_1 x_1(t) \\ \dot{x}_2(t) = \lambda_2 x_2(t) \\ x_1(t_0) = x_{0,1} \\ x_2(t_0) = x_{0,2}. \end{cases}$$

La solution est alors donnée par

$$\begin{cases} x_1(t) = e^{\lambda_1(t-t_0)}x_{0,1} \\ x_2(t) = e^{\lambda_2(t-t_0)}x_{0,2}. \end{cases}$$

et le comportement asymptotique est

- si $\lambda_1 < 0$ et $\lambda_2 < 0$ alors $\lim_{t \rightarrow +\infty} x(t) = 0$;
- si $\lambda_1 < 0$ et $\lambda_2 > 0$ et $x_{0,2} \neq 0$ alors $|x_1(t)| \rightarrow 0$ et $|x_2(t)| \rightarrow +\infty$, et donc $\|x(t)\| \rightarrow +\infty$, quand $t \rightarrow +\infty$;
- ...

□

2.2.2 Exponentielle de matrice

L'espace vectoriel normé $(\mathcal{M}_n(\mathbf{K}), \|\cdot\|)$, est un espace de Banach. On considère ici une norme qui vérifie $\|AB\| \leq \|A\|\|B\|$. La série $\sum_{k=0}^{+\infty} \frac{A^k}{k!}$ est alors normalement convergente (6) car

$$\sum_{k=0}^{+\infty} \frac{\|A^k\|}{k!} \leq \sum_{k=0}^{+\infty} \frac{\|A\|^k}{k!} = e^{\|A\|}.$$

Définition 2.2.1

[Exponentielle de matrice] On appelle exponentiel de matrice l'application

$$\begin{aligned} \exp: \mathcal{M}_n(\mathbf{K}) &\longrightarrow \mathcal{M}_n(\mathbf{K}) \\ A &\longmapsto \exp(A) = e^A = \sum_{k=0}^{+\infty} \frac{A^k}{k!}. \end{aligned}$$

Théorème 2.2.2

L'exponentielle de matrice a les propriétés suivantes :

1. $e^0 = I$;
2. si $A = \text{diag}(\lambda_1, \dots, \lambda_n)$ alors

$$\exp(A) = \begin{pmatrix} e^{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{\lambda_n} \end{pmatrix}; \quad (2.1)$$

3. si P est inversible on a

$$\exp(PAP^{-1}) = P \exp(A) P^{-1}; \quad (2.2)$$

4. si A et B sont deux matrices qui commutent alors

$$\exp(A + B) = \exp(A) \exp(B); \quad (2.3)$$

5. pour tout α et β , $e^{(\alpha+\beta)A} = e^{\alpha A} e^{\beta A}$;
6. pour toute matrice A , e^A est inversible et

$$(\exp(A))^{-1} = \exp(-A); \quad (2.4)$$

7. pour toute matrice A , l'application $t \rightarrow e^{tA}$ est C^∞ et

$$\frac{d}{dt} e^{tA} = A e^{tA} = e^{tA} A. \quad (2.5)$$

- 1. Évident.
 2. Il suffit d'écrire la définition.
 3. On a

$$\begin{aligned} \exp(PAP^{-1}) &= PIP^{-1} + PAP^{-1} + \dots + \frac{PA^k P^{-1}}{k!} + \dots \\ &= P \left(I + A + \dots + \frac{A^k}{k!} + \dots \right) P^{-1}. \end{aligned}$$

4. Soit A, B deux matrices carrés qui commutent. Par définition on a

$$e^{A+B} = I + (A+B) + \dots + \frac{(A+B)^m}{m!} + \dots$$

Mais A et B commutent, la formule du binôme permet alors d'écrire

$$(A + B)^m = \sum_{k=0}^m C_m^k A^k B^{m-k}.$$

Mais les séries définissant e^A et e^B étant normalement convergentes, nous avons

$$\begin{aligned} e^A e^B &= \left(I + A + \cdots + \frac{A^k}{k!} + \cdots \right) \left(I + B + \cdots + \frac{B^k}{k!} + \cdots \right) \\ &= I + (A + B) + \cdots \left(\sum_{k=0}^m \frac{A^k}{k!} \frac{B^{m-k}}{(m-k)!} \right) + \cdots \\ &= I + (A + B) + \cdots \frac{1}{m!} \left(\sum_{k=0}^m \frac{m!}{k!(m-k)!} A^k B^{m-k} \right) + \cdots \end{aligned}$$

d'où le résultat.

5. C'est immédiat car αA et βA commutent.
6. Il suffit d'écrire $e^{A-A} = e^A e^{-A} = I$.
7. Posons

$$\begin{aligned} f: \mathbf{R} &\longrightarrow \mathcal{M}_n(\mathbf{K}) \\ t &\longmapsto e^{tA}. \end{aligned}$$

La série définissant l'application $f(t)$ est normalement convergente, par suite elle est uniformément convergente sur tout intervalle compacte I de \mathbf{R} . La série des termes dérivées s'écrit

$$A \left(I + A + \cdots + \frac{t^k A^k}{(k)!} + \cdots \right) = A e^{tA}.$$

Cette série est normalement convergente. On en déduit qu'elle est égale à $f'(t)$. ■

Exemple 2.2.3. Si maintenant nous considérons le cas du système différentiel

$$(IVP3) \begin{cases} \dot{x}(t) = \Lambda x(t) \\ x(t_0) = x_0, \end{cases}$$

avec

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}.$$

La solution est alors

$$x(t) = \begin{pmatrix} e^{(t-t_0)\lambda_1} x_{0,1} \\ \vdots \\ e^{(t-t_0)\lambda_n} x_{0,n} \end{pmatrix} = \begin{pmatrix} e^{(t-t_0)\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{(t-t_0)\lambda_n} \end{pmatrix} x_0 = e^{(t-t_0)\Lambda} x_0.$$

Le comportement asymptotique est alors

- si tous les λ_i sont strictement négatifs alors $\lim_{t \rightarrow +\infty} x(t) = 0$;

- si tous les λ_i sont négatifs ou nuls alors la solution est bornée quand $t \rightarrow +\infty$;
- si au moins un λ_i est strictement positif et que $x_{0,i} \neq 0$ alors $\|x(t)\| \rightarrow +\infty$, quand $t \rightarrow +\infty$.

□

Exemple 2.2.4. Soit maintenant A une matrice diagonalisable, $A = P\Lambda P^{-1}$ et le système différentiel à valeur initiale

$$(IVP4) \begin{cases} \dot{x}(t) = Ax(t) \\ x(t_0) = x_0, \end{cases}$$

Posons $z(t) = P^{-1}x(t)$, alors $z(t)$ est solution du système différentielle à valeur initiale

$$(IVP5) \begin{cases} \dot{z}(t) = P^{-1}\dot{x}(t) = P^{-1}P\Lambda P^{-1}x(t) = \Lambda z(t) \\ z(t_0) = P^{-1}x_0. \end{cases}$$

On a donc $z(t) = e^{(t-t_0)\Lambda}P^{-1}x_0$ et $x(t) = Pz(t) = (Pe^{(t-t_0)\Lambda}P^{-1})x_0$. Par suite le comportement asymptotique est caractérisé par les valeurs propres de la matrice A . □

Théorème 2.2.3

L'unique solution du système différentielle linéaire, autonome et homogène à valeur initial

$$(IVP6) \begin{cases} \dot{x}(t) = Ax(t) \\ x(t_0) = x_0, \end{cases}$$

est

$$x(t) = e^{(t-t_0)A}x_0. \quad (2.6)$$

► Soit x vérifiant (2.6), la propriété (2.5) implique alors que

$$\dot{x}(t) = \frac{d}{dt}(e^{(t-t_0)A}x_0) = Ae^{(t-t_0)A}x_0 = Ax(t).$$

Comme on a $x(t_0) = x_0$, c'est bien une solution.

Supposons maintenant que x soit une solution de (IVP6). Posons $z(t) = e^{-(t-t_0)A}x(t)$, alors

$$\dot{z}(t) = -Ae^{-(t-t_0)A}x(t) + e^{-(t-t_0)A}Ax(t) = 0,$$

car A et $e^{-(t-t_0)A}$ commutent. Par suite $z(t) = z(t_0) = x_0$ pour tout t et $x(t) = e^{(t-t_0)A}z(t) = e^{(t-t_0)A}x_0$. ■

2.2.3 Calcul de l'exponentielle de matrice



Exercice 2.2.5. À partir de la définition calculer e^{tA} pour

1.

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix},$$

2.

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

3.

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

4.

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

□

Remarque 2.2.1. Si A est une matrice réelle diagonalisable dans \mathbf{C} , $e^A = Pe^{\Lambda}P^{-1}$ est aussi une matrice réelle.



Exercice 2.2.6. Calculer

$$\exp \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix}$$

□



Exercice 2.2.7. Soit $a \in \mathbf{R}$, à partir de la diagonalisation dans \mathbf{C} , calculer

$$\exp \begin{pmatrix} 0 & -a \\ a & 0 \end{pmatrix}$$

□

Dans le cas général, pour calculer l'exponentiel d'une matrice, on va utiliser sa décomposition de Jordan [1]. On note respectivement $P(\lambda)$ et $m(\lambda)$ les polynômes caractéristique et minimale de $A \in \mathcal{M}_n(\mathbf{C})$

$$P(\lambda) = \prod_{i=1}^r (\lambda - \lambda_i)^{p_i}$$

$$m(\lambda) = \prod_{i=1}^r (\lambda - \lambda_i)^{m_i},$$

et on rappelle que l'on a les propriétés suivantes

1. $\Gamma_i = \text{Ker}_{\mathbf{C}}(A - \lambda_i I)^{p_i} = \text{Ker}_{\mathbf{C}}(A - \lambda_i I)^{m_i}$;
2. $\dim \Gamma_i = p_i$;
3. Les Γ_i sont invariants par A ;
4. La restriction de A à Γ_i s'écrit $A_{\Gamma_i} = \lambda_i I_{\Gamma_i} + N_i$, où N_i est un endomorphisme de Γ_i nilpotent d'ordre inférieur ou égal à p_i ($N_i^{p_i} = 0$).

Par suite, si P est la matrice de passage à une base formée d'une base de Γ_1, \dots , d'une base

de Γ_r , on a $P^{-1}AP = \Lambda + N$ avec

$$\Lambda = \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_1 & & \\ & & & \ddots & \\ & & & & \lambda_r \\ & & & & & \ddots \\ & & & & & & \lambda_r \end{pmatrix}, \quad N = \begin{pmatrix} N_1 & & \\ & \ddots & \\ & & N_r \end{pmatrix},$$

où λ_i est présent p_i fois dans Λ et où N est nilpotent. Pour aller plus loin nous allons utiliser la décomposition sous la forme de Jordan.

Théorème 2.2.4

[Décomposition de Jordan] Soit A une matrice carré complexe, alors il existe une matrice P inversible telle que

$$P^{-1}AP = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{pmatrix},$$

où J_i est de dimension (p_i, p_i) et s'écrit

$$J_i = \begin{pmatrix} J_{i,1} & & \\ & \ddots & \\ & & J_{i,e_i} \end{pmatrix}, \quad J_{i,k} = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}, \quad J_{i,k} \text{ matrice de dimension } (n_{i,k}, n_{i,k})$$

avec

1. $e_i = \dim(\text{Ker}(A - \lambda_i I))$;
2. m_i est la dimension du plus grand bloc de J_i .

► cf. [1]. ■

En conclusion, nous avons

$$\exp(tA) = P e^{tJ} P^{-1} = \begin{pmatrix} e^{tJ_{1,1}} & & \\ & \ddots & \\ & & e^{tJ_{r,e_r}} \end{pmatrix}, \quad e^{tJ_{i,k}} = e^{t\lambda_i} \begin{pmatrix} 1 & t & \cdots & \frac{t^{n_{i,k}-1}}{(n_{i,k}-1)!} \\ & \ddots & \ddots & t \\ & & & 1 \end{pmatrix}$$

2.2.4 Forme des solutions

Théorème 2.2.5

Soit A une matrice carrée complexe. Toute solution de $\dot{x}(t) = Ax(t)$ s'écrit

$$x(t) = \sum_{i=1}^r e^{t\lambda_i} \sum_{k=0}^{m_i-1} t^k v_{i,k} \quad (2.7)$$

où $v_{i,k} \in \Gamma_i$.

Remarque 2.2.2. • Le terme en λ_i est polynomiale en t lorsque $m_i > 1$, et constant lorsque $m_i = 1$ (ce dernier cas se produit si et seulement si les ordres de multiplicités algébrique et géométrique de λ_i coïncident).

• Fixons $x(0) = x_0 = v_1 + \dots + v_r$, avec $v_i \in \Gamma_i$. Cette décomposition de x_0 existe et est unique car $\mathbf{C}^n = \Gamma_1 \oplus \dots \oplus \Gamma_r$. Alors en écrivant les dérivées successives en 0 dans la formule (2.7) et en utilisant le fait que $\frac{d^k x}{dt^k}(0) = A^k x_0$, on obtient

$$v_{i,k} = \frac{1}{k!} N^k v_i$$

où N est la matrice nilpotente de la décomposition $\Lambda + N$ de $P^{-1}AP$.

Considérons maintenant le cas d'une matrice à coefficients dans \mathbf{R} et d'une solution dans \mathbf{R}^n . On peut bien évidemment considérer A comme une matrice complexe, prendre comme condition initiale $x_0 + i0$, appliquer le théorème ??, puis prendre la partie réelle. Mais on peut profiter du fait que si une valeur propre est complexe, sa valeur conjuguée est aussi une valeur propre et on peut décomposer \mathbf{R}^n comme la somme directe

$$\mathbf{R}^n = E_1 \oplus \dots \oplus E_q,$$

où $E_i = \Gamma_{\lambda_i} \cap \mathbf{R}^n$ si λ est une valeur propre réelle et $E_i = (\Gamma_{\lambda_i} \oplus \Gamma_{\bar{\lambda}_i}) \cap \mathbf{R}^n$ si λ_i est une valeur propre complexe. On obtient ainsi le

Théorème 2.2.6

Si A est une matrice réelle, l'ensemble des solutions réelles s'écrit

$$x(t) = \sum_{i=1}^q e^{t\alpha_i} \sum_{0 \leq k \leq m_i-1} t^k (\cos(\beta_i t) a_{i,k} + \sin(\beta_i t) b_{i,k}),$$

où $\lambda_i = \alpha_i + i\beta_i$ et les $a_{i,k}, b_{i,k}$ sont dans E_i .

Corollaire 2.2.7

Si A est une matrice réelle diagonalisable dans \mathbf{C} , alors toute solution de $\dot{x}(t) = Ax(t)$ s'écrit

$$x(t) = \sum_{i=1}^q e^{t\alpha_i} (\cos(\beta_i t) a_i + \sin(\beta_i t) b_i)$$

avec $\lambda_1, \dots, \lambda_s$, valeurs propres réelles de A , $\lambda_{s+1}, \bar{\lambda}_{s+1}, \dots, \lambda_q, \bar{\lambda}_q$, valeurs propres complexes conjuguées de A , $\lambda_j = \alpha_j + i\beta_j$ et $a_j + ib_j$ vecteur propre de A associée à λ_j

2.3 Équations linéaires

2.3.1 Introduction

On s'intéresse dans cette section aux équations différentielles linéaires à condition initiale

$$(IVP7) \begin{cases} \dot{x}(t) = A(t)x(t) + b(t) \\ x(t_0) = x_0, \end{cases}$$

Les fonctions $A : I \subset \mathbf{R} \rightarrow \mathcal{M}_n(\mathbf{K})$ et $b : I \rightarrow \mathbf{K}^n$ seront toujours supposées de classe $C^k, k \geq 0$.

2.3.2 Existence et unicité de solution

Théorème 2.3.1

On suppose que A et b sont $C^k, k \geq 0$, $t_0 \in I$, intervalle ouvert de \mathbf{R} et que $x_0 \in \mathbf{K}^n$, alors le problème de Cauchy (IVP7) admet une solution et une seule.

► Ceci est aussi démontré au corollaire (3.3.1). Démontrons tout d'abord l'existence. On considère un intervalle compact $J \subset I$ qui contient t_0 . Nous allons démontrer que l'application

$$\begin{aligned} T: E = (C^0(J, \mathbf{K}^n), \|\cdot\|_\infty) &\longrightarrow E \\ x(\cdot) &\longmapsto T(x(\cdot)) = x_0 + \int_{t_0}^t (A(s)x(s) + b(s))ds, \end{aligned}$$

admet un point fixe. Le théorème (1.1.3.1) permettra alors de conclure que le problème de Cauchy admet une solution sur tout intervalle compact $J \subset I$ et donc sur tout I . Pour démontrer que l'application T admet un point fixe, nous allons utiliser le théorème du point fixe qui dit que si T , application d'un espace de Banach dans lui même, admet un itéré T^p qui est contractant pour $p \geq 1$, alors T admet un point fixe (cf. 6.6.2). Remarquons tout d'abord que T est bien une application de E dans E et que E est un espace de Banach. Considérons maintenant deux éléments de $E, x_1(\cdot)$ et $x_2(\cdot)$. Nous avons

$$\begin{aligned} \|(T(x_1(\cdot)) - T(x_2(\cdot)))(t)\| &= \left\| \int_{t_0}^t (A(s)(x_1(s) - x_2(s)))ds \right\| \\ &= \int_{t_0}^t \|A(s)(x_1(s) - x_2(s))\|ds \\ &\leq \|A(\cdot)\|_\infty \int_{t_0}^t \|x_1(s) - x_2(s)\|ds \\ &\leq \|A(\cdot)\|_\infty \|x_1(\cdot) - x_2(\cdot)\|_\infty (t - t_0), \end{aligned}$$

où $\|x(\cdot)\|_\infty = \max_{t \in J} \|x(t)\|$. Comme $J = [a, b]$, car c'est un intervalle compact de \mathbf{R} , on a

$$\|T(x_1(\cdot)) - T(x_2(\cdot))\|_\infty \leq \|A(\cdot)\|_\infty \|x_1(\cdot) - x_2(\cdot)\|_\infty (b - a).$$

Ceci montre en particulier que T est continue. Montrons maintenant par récurrence que l'on a

$$\|(T^p(x_1(\cdot)) - T^p(x_2(\cdot)))(t)\| \leq \|A(\cdot)\|_\infty^p \frac{(t - t_0)^p}{p!} \|x_1(\cdot) - x_2(\cdot)\|_\infty \quad (2.8)$$

$$\text{et } \|T^p(x_1(\cdot)) - T^p(x_2(\cdot))\|_\infty \leq \frac{\|A(\cdot)\|_\infty^p (b - a)^p}{p!} \|x_1(\cdot) - x_2(\cdot)\|_\infty \quad (2.9)$$

L'assertion est vraie pour $p = 1$. Supposons la vraie pour p et montrons là pour $p + 1$.

$$\begin{aligned}
\|(T^{p+1}(x_1(\cdot)) - T^{p+1}(x_2(\cdot)))(t)\| &= \|(T(T^p x_1(\cdot)) - T(T^p x_2(\cdot)))(t)\| \\
&= \left\| \int_{t_0}^t (A(s)(T^p x_1(s) - T^p x_2(s))) ds \right\| \\
&\leq \|A(\cdot)\|_\infty \int_{t_0}^t \|T^p x_1(s) - T^p x_2(s)\| ds \\
&\leq \|A(\cdot)\|_\infty \int_{t_0}^t \|A(\cdot)\|_\infty^p \frac{(s - t_0)^p}{p!} \|x_1(\cdot) - x_2(\cdot)\|_\infty ds \\
&\leq \|A(\cdot)\|_\infty^{p+1} \frac{(t - t_0)^{p+1}}{(p+1)!} \|x_1(\cdot) - x_2(\cdot)\|_\infty
\end{aligned}$$

En prenant le maximum sur $t \in J$ on obtient (2.9).

Il nous suffit donc de choisir l'entier p de façon à avoir

$$\frac{\|A(\cdot)\|_\infty^p (b - a)^p}{p!} < 1$$

pour conclure. ■

2.3.3 Résolvante

On considère ici l'équation linéaire homogène

$$\dot{x}(t) = A(t)x(t). \quad (2.10)$$

Théorème 2.3.2

L'ensemble des solutions de l'équation différentielle linéaire et homogène (2.10) \mathcal{E} est un espace vectoriel de dimension n .

► Le fait que \mathcal{E} soit un espace vectoriel est immédiat. Considérons maintenant l'application

$$\begin{aligned}
L_{t_0}: \mathbf{K}^n &\longrightarrow \mathcal{E} \\
x_0 &\longmapsto x(\cdot, t_0, x_0)
\end{aligned}$$

où $x(\cdot, t_0, x_0)$ est l'unique solution de l'équation différentielle (2.10) vérifiant $x(t_0) = x_0$. Il est évident que cette application est linéaire. Le théorème 2.3.2 d'existence et d'unicité de solution implique que cette application est une bijection, d'où le résultat en ce qui concerne la dimension. ■

Définition 2.3.3

On appelle résolvante de l'équation différentielle linéaire et homogène $\dot{x}(t) = A(t)x(t)$ l'application

$$\begin{aligned}
R(t, t_0): \mathbf{K}^n &\longrightarrow \mathbf{K}^n \\
x_0 &\longmapsto x(t, t_0, x_0).
\end{aligned}$$

Théorème 2.3.4

1. On a $R(t, t_0).x_0 = x(t, t_0, x_0)$.
2. Si le système est autonome on a $R(t, t_0) = e^{(t-t_0)A}$.
3. Pour tout t_0 fixé, $R(., t_0)$ est la solution du problème de Cauchy

$$(IVP8) \begin{cases} \dot{X}(t) = A(t)X(t) \\ X(t_0) = I_n. \end{cases}$$

4. Pour tout t_0, t_1 et t_2 dans I on a

$$R(t_2, t_0) = R(t_2, t_1) \times R(t_1, t_0)$$

5. Pour tout t_0, t_1 dans I on a $R(t_0, t_1) = (R(t_1, t_0))^{-1}$.
6. Si $A(.)$ est C^k , alors $R(., t_0)$ est C^{k+1} .

- 1. Le théorème d'existence et d'unicité de solution et la nature de l'équation différentielle $\dot{x}(t) = A(t)x(t)$, implique que $R(t, t_0)$ est une application linéaire et bijective, L'application $R(t, t_0)$ est donc un isomorphisme de \mathbf{K}^n sur \mathbf{K}^n et on a $x(t, t_0, x_0) = R(t, t_0)x_0$.
2. Évident.
3. Dans le problème (IVP8), $X(t)$ est une matrice (n, n) et I_n est la matrice identité d'ordre n . La propriété provient immédiatement du fait que

$$\frac{\partial}{\partial t}(R(t, t_0)x_0) = A(t)(R(t, t_0)x_0),$$

et du fait que $R(t_0, t_0) = I_n$ par définition.

4. Il s'agit tout simplement de la composée de deux applications linéaires.
5. Il suffit de remarquer que $R(t_0, t_1) \times R(t_1, t_0) = R(t_0, t_0) = I_n$.
6. Cela provient des propriétés des solutions d'une équation différentielle.

■

Théorème 2.3.5

L'application $R : I \times I \rightarrow \mathcal{M}_n(\mathbf{K})$ qui à (t, s) associe $R(t, s)$ est continue et de classe C^{k+1} si $A(.)$ est de classe C^k .

- On peut écrire $R(t, s) = R(t, t_0) \times (R(s, t_0))^{-1}$. Or l'application $R(., t_0) : I \rightarrow \mathcal{M}_n(\mathbf{K})$ est continue car c'est la solution de (IVP8). On sait de plus que l'application

$$\begin{aligned} \Phi: \mathcal{M}_n(\mathbf{K}) \times \mathcal{M}_n(\mathbf{K}) &\longrightarrow \mathcal{M}_n(\mathbf{K}) \\ (A, B) &\longmapsto AB \end{aligned}$$

est bilinéaire et continue et que l'application

$$\begin{aligned} inv: GL(\mathbf{K}^n) &\longrightarrow \mathcal{M}_n(\mathbf{K}) \\ A &\longmapsto A^{-1} \end{aligned}$$

est continue.

On en déduit par composition des fonctions que R est continue.

Si $A(\cdot)$ est C^k , alors $R(\cdot, t_0)$ est C^{k+1} . Comme Φ et inv sont C^∞ , on a, toujours par composition R qui est C^{k+1} . ■

 **Exercice 2.3.1.** Soit $R(t, t_0)$ la résolvante de l'équation différentielle linéaire

$$\dot{x}(t) = A(t)x(t).$$

On note $\Delta(t) = \det R(t, t_0)$.

1. On considère l'application

$$\begin{aligned} \det: GL_n(\mathbf{R}) &\longrightarrow \mathbf{R} \\ A &\longmapsto \det(A). \end{aligned}$$

Montrer que $\det'(A) \cdot H = \det(A) \operatorname{trace}(A^{-1}H)$

2. Montrer que $\Delta(t)$ est solution du problème de Cauchy

$$(IVP) \begin{cases} \dot{\Delta}(t) = \operatorname{trace}(A(t))\Delta(t) \\ \Delta(t_0) = 1. \end{cases}$$

3. En déduire que

$$\det(R(t, t_0)) = \exp \left(\int_{t_0}^t \operatorname{trace}(A(s)) ds \right).$$

□

Corollaire 2.3.6

[Liouville] Si pour tout t , $\operatorname{trace}(A(t)) = 0$, alors $\det(R(t, t_0)) = 1$ pour tout t .

On se place ici dans le cas où le corps \mathbf{K} est le corps des réels. Une conséquence du corollaire précédent est que si la trace de l'opérateur $A(t)$ est constamment nulle, alors l'équation différentielle conserve le volume d'un domaine de \mathbf{R}^n . En effet, si on note Γ_0 un domaine de \mathbf{R}^n et Γ_t son transport de t_0 à t par l'équation différentielle, c'est-à-dire

$$\Gamma_t = \{R(t, t_0)x_0 \in \mathbf{R}^n, x_0 \in \Gamma_0\} = R(t, t_0)\Gamma_0,$$

alors en utilisant la formule du changement de variable dans une intégrale multiple on a

$$\begin{aligned} \operatorname{Vol}(\Gamma_t) &= \int_{\Gamma_t} dx = \int_{\Gamma_0} \left| \det \left(\frac{\partial x}{\partial x_0}(t, t_0, x_0) \right) \right| dx_0 \\ &= \int_{\Gamma_0} |\det(R(t, t_0))| dx_0 \\ &= |\det(R(t, t_0))| \operatorname{Vol}(\Gamma_0). \end{aligned}$$

Donc, si $\operatorname{trace}(A(t)) = 0$ pour tout t , $\operatorname{Vol}(\Gamma_t) = \operatorname{Vol}(\Gamma_0)$.

2.3.4 Équations différentielles linéaires avec second membre

Théorème 2.3.7

La solution du problème de Cauchy linéaire

$$(IVP9) \begin{cases} \dot{x}(t) = A(t)x(t) + b(t) \\ x(t_0) = x_0, \end{cases}$$

s'écrit

$$x(t) = R(t, t_0)x_0 + \int_{t_0}^t R(t, s)b(s)ds. \quad (2.11)$$

► Nous allons appliquer la méthode de la variation de la constante. Posons $z(t) = R(t, t_0)v(t)$, nous avons alors

$$\begin{aligned} \dot{z}(t) &= \left(\frac{d}{dt} R(t, t_0) \right) v(t) + R(t, t_0) \dot{v}(t) \\ &= A(t)R(t, t_0)v(t) + R(t, t_0) \dot{v}(t) \\ &= A(t)z(t) + R(t, t_0) \dot{v}(t). \end{aligned}$$

Il suffit alors de poser $R(t, t_0)\dot{v}(t) = b(t)$, soit $\dot{v}(t) = R(t_0, t)b(t)$, pour que z vérifie l'équation linéaire. Si on pose maintenant $v(t) = \int_{t_0}^t R(t_0, s)b(s)ds$, on obtient pour z une solution qui vérifie $z(t_0) = 0$. En conclusion, si on prend

$$x(t) = R(t, t_0)x_0 + z(t),$$

nous aurons bien $x(t_0) = x_0$ et

$$\begin{aligned} \dot{x}(t) &= A(t)R(t, t_0)x_0 + \dot{z}(t) \\ &= A(t)R(t, t_0)x_0 + \frac{d}{dt}(R(t, t_0)v(t)) \\ &= A(t)R(t, t_0)x_0 + A(t)R(t, t_0)v(t) + R(t, t_0)\dot{v}(t) \\ &= A(t)(R(t, t_0)x_0 + R(t, t_0)v(t)) + b(t) \\ &= A(t)x(t) + b(t) \end{aligned}$$

Il suffit maintenant d'écrire

$$\begin{aligned} x(t) &= R(t, t_0)x_0 + R(t, t_0) \int_{t_0}^t R(t_0, s)b(s)ds \\ &= R(t, t_0)x_0 + \int_{t_0}^t R(t, t_0)R(t_0, s)b(s)ds \\ &= R(t, t_0)x_0 + \int_{t_0}^t R(t, s)b(s)ds, \end{aligned}$$

pour conclure. ■



Exercice 2.3.2. ¹ On considère le système contrôlé suivant (pendule inversé linéarisé où on

contrôle le couple moteur et avec $g = l$)

$$\ddot{\theta}(t) - \theta(t) = u(t),$$

avec $\theta(0) = 1$ et $\dot{\theta}(0) = -2$.

1. Écrire le système sous la forme

$$(IVP) \begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ x(0) = x_0. \end{cases}$$

2. Calculer e^{tA} à l'aide de la définition.

3. On considère le contrôle en boucle ouverte $u(t) = 3e^{-2t}$. Résoudre (IVP) .

4. Résoudre le système

$$\begin{cases} \dot{x}(t) = Ax(t) \\ x(0) = x_0. \end{cases}$$

avec comme condition initiale $\theta(0) = 0$ et $\dot{\theta}(0) = \varepsilon$. En déduire la solution de (IVP) avec $\theta(0) = 1$, $\dot{\theta}(0) = -2 + \varepsilon$ et toujours pour le contrôle $u(t) = 3e^{-2t}$.

5. Commentaire.

6. On prend maintenant $u(t) = -\alpha\theta(t) - \beta\dot{\theta}(t)$. Quelle relation doivent vérifier les constantes α et β afin que $x_e = 0$ soit un point d'équilibre asymptotiquement stable? \square

Théorie des équations différentielles

3.1	Existence	22
3.2	Dépendances par rapports aux données	29
3.2.1	Introduction	29
3.2.2	Continuité	30
3.2.3	Dérivée	32

3.1 Existence

On s'intéresse au problème de Cauchy suivant

$$(IVP) \begin{cases} \dot{x}(t) = f(t, x(t)) \\ x(t_0) = x_0, \end{cases}$$

où Ω est un ouvert, $(t_0, x_0) \in \Omega$ et

$$f: \begin{array}{ccc} \Omega \subset \mathbf{R} \times \mathbf{R}^n & \longrightarrow & \mathbf{R}^n \\ (t, x) & \longmapsto & f(t, x) \end{array} .$$

Définition 3.1.1 – Fonction localement lipschitzienne

L'application $f : \Omega \subset \mathbf{R} \times \mathbf{R}^n \rightarrow \mathbf{R}^n$, Ω ouvert, est localement lipschitzienne par rapport à la variable x si et seulement si pour tout $(t_0, x_0) \in \Omega$ il existe un voisinage $V \in \mathcal{V}(t_0, x_0)$ et une constante $k \geq 0$ tels que

$$\forall (t, x_1) \in V, \forall (t, x_2) \in V, \|f(t, x_1) - f(t, x_2)\| \leq k \|x_1 - x_2\|.$$

Théorème 3.1.2

Si f est différentiable par rapport à x et si l'application

$$\begin{array}{ccc} \frac{\partial f}{\partial x}: & \Omega & \longrightarrow \mathcal{L}(\mathbf{R}^n, \mathbf{R}^n) \\ & (t, x) & \longmapsto \frac{\partial f}{\partial x}(t, x) \end{array}$$

est continue alors f est localement lipschitzienne.

► C'est un corollaire du théorème des accroissements finis (cf. le corollaire 1.3.6 page 17 de [13]). ■

Théorème 3.1.3 – Théorème de Cauchy-Lipschitz

Soit $f : \Omega \rightarrow \mathbf{R}^n$, Ω ouvert de $\mathbf{R} \times \mathbf{R}^n$, f continue et localement lipschitzienne par rapport à x alors pour tout $(t_0, x_0) \in \Omega$, il existe une unique solution locale au problème de Cauchy

$$(IVP) \begin{cases} \dot{x}(t) = f(t, x(t)) \\ x(t_0) = x_0. \end{cases}$$

Remarque 3.1.1. Par unique solution locale on entend : si (I_1, x_1) et (I_2, x_2) sont deux solutions de (IVP) alors elles coïncident sur $I_1 \cap I_2$.

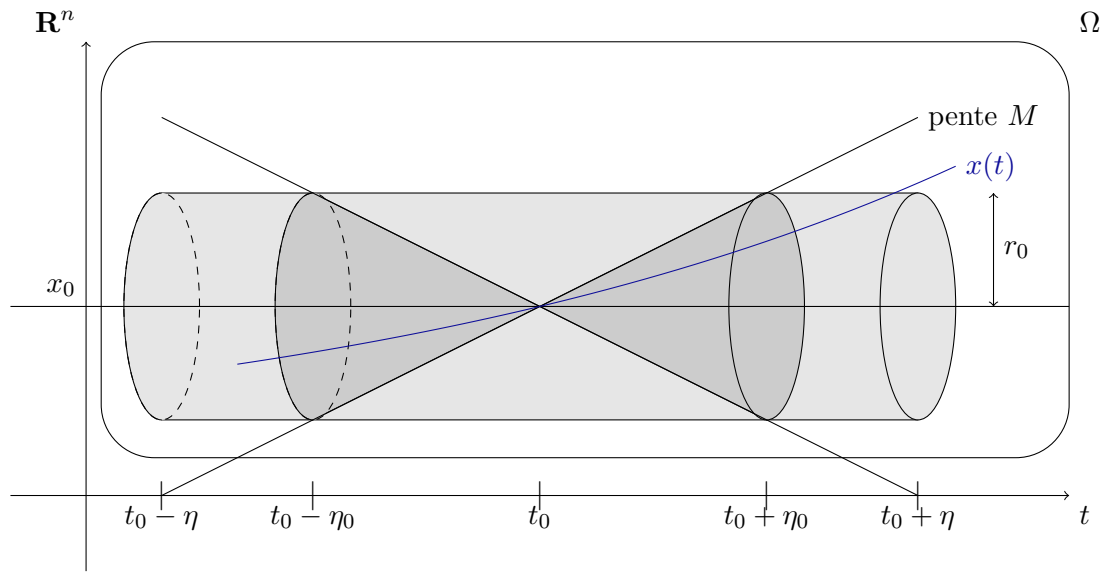


FIGURE 3.1 – *Cylindre de sécurité*; $x(t)$ est dans le cylindre de sécurité $[t_0 - \eta_0, t_0 + \eta_0] \times B_f(x_0, r_0)$, mais n'est pas dans le cylindre initial $[t_0 - \eta, t_0 + \eta] \times B_f(x_0, r_0)$. Ceci implique que $\|\varphi(t, T(x)(t))\| \leq M$ pour tout $t \in [t_0 - \eta_0, t_0 + \eta_0]$.

► Nous allons appliquer le théorème du point fixe (6.6.2) à l'application

$$\begin{aligned} T: E &\longrightarrow E \\ x &\longmapsto T(x), \end{aligned}$$

définie par

$$T(x)(t) = x_0 + \int_{t_0}^t f(s, x(s)) ds.$$

Il nous faut pour cela construire tout d'abord l'espace métrique complet E , puis montrer que T est bien définie et qu'il existe un entier p pour lequel T^p est une contraction.



Exercice 3.1.1. Afin de bien comprendre la définition de l'application T donner les premiers itérés $x^{(k+1)} = T(x^{(k)})$ dans le cas du problème de Cauchy $\dot{x}(t) = Ax(t)$, $x(t_0) = x_0$ avec $x^{(0)}(t) = x_0$ pour tout t . \square

Définissons donc tout d'abord E . Comme f est localement lipschitzienne, il existe $V_f = [t_0 - \eta, t_0 + \eta] \times B_f(x_0, r_0)$ voisinage de (t_0, x_0) sur lequel f est k -lipschitzienne. f est

continue, donc $\sup_{V_f} \|f(t, x)\| = M$. Si $M = 0$ alors le théorème est évident. Si $M \neq 0$, on pose $\eta_0 = \min(\eta, r_0/M)$ et $C = [t_0 - \eta_0, t_0 + \eta_0] \times B_f(x_0, r_0)$; C s'appelle le cylindre de sécurité (cf. la Fig. 3.1). Posons maintenant $E = \mathcal{C}^0([t_0 - \eta_0, t_0 + \eta_0], B_f(x_0, r_0))$ et d la distance sur E définie par $d(x_1, x_2) = \sup_{[t_0 - \eta_0, t_0 + \eta_0]} \|x_1(t) - x_2(t)\|$. (E, d) est alors un espace métrique complet.

Montrons maintenant que l'application T est bien définie. x est continue, f est continue donc $x_0 + \int_{t_0}^t f(s, x(s))ds \in \mathbf{R}^n$ existe. Il nous reste à vérifier que cette quantité est bien dans $B_f(x_0, r_0)$; c'est ici que nous allons utiliser le cylindre de sécurité. Mais ceci est immédiat car

$$\begin{aligned} \|T(x)(t) - x_0\| &= \left\| \int_{t_0}^t f(s, x(s))ds \right\| \leq \int_{t_0}^t \|f(s, x(s))\|ds \\ &\leq \int_{t_0}^t Mds = M(t - t_0) \leq M(r_0/M) = r_0. \end{aligned}$$

Pour démontrer la contraction de T^p , constatons tout d'abord que

$$\begin{aligned} \|T(x_1)(t) - T(x_2)(t)\| &\leq \int_{t_0}^t \|f(s, x_1(s)) - f(s, x_2(s))\|ds \\ &\leq \int_{t_0}^t k\|x_1(s) - x_2(s)\|ds \\ &\leq k(t - t_0)\|x_1 - x_2\|. \end{aligned}$$

Par récurrence on a alors

$$\begin{aligned} \|T^p(x_1)(t) - T^p(x_2)(t)\| &\leq \frac{k^p(t - t_0)^p}{p!} \|x_1 - x_2\| \\ d(T^p(x_1), T^p(x_2)) &\leq \frac{k^p \eta_0^p}{p!} d(x_1, x_2). \end{aligned}$$

Il suffit donc de prendre p tel que $\frac{k^p \eta_0^p}{p!} < 1$. ■

Nous avons donc montré l'existence et l'unicité de la solution dans E . Il reste donc à montrer pour l'unicité que toute solution définie sur $[t_0 - \eta_0, t_0 + \eta_0]$ est à valeur dans la boule $B_f(x_0, r_0)$. Raisonnons par l'absurde et supposons qu'il existe une solution v telle que

$$A = \{t \in [t_0, t_0 + \eta_0], \|v(t) - x_0\| > r_0\} \neq \emptyset$$

Posons $t_1 = \inf\{t \in A\}$, on a $t_0 < t_1 < t_0 + \eta_0$, $\|v(t_1) - x_0\| = r_0$ et $v(t) \in B_f(x_0, r_0)$, $\forall t \in [t_0, t_1]$. Par suite

$$r_0 = \|v(t_1) - x_0\| \leq \int_{t_0}^{t_1} \|f(s, v(s))\|ds \leq M(t_1 - t_0) \leq r_0$$

Donc $r_0 = M(t_1 - t_0)$ et $t_1 = t_0 + r_0/M \geq t_0 + \eta_0$ d'où la contradiction.

Nous allons maintenant montrer l'unicité d'une solution dite maximale. Pour cela rappelons la

Définition 3.1.4

[ensemble ordonné inductif] Un ensemble ordonné est dit inductif si toute partie totalement ordonnée est majorée.

Rappelons aussi le

Lemme 3.1.1 (lemme de Zorn). *Tout ensemble ordonné inductif admet un élément maximal.*

► cf. page 24 de [15]. ■

Théorème 3.1.5

L'ensemble des solutions du problème de Cauchy est ordonnée par la relation d'ordre

$$(I_1, x_1) \leq (I_2, x_2) \iff \begin{cases} I_1 \subset I_2 \\ x_2|_{I_1} = x_1 \end{cases}$$

et si cet ensemble est non vide alors il est inductif.

► La relation d'ordre est évidente. Montrons que toute partie totalement ordonnée $(I_\alpha, x_\alpha)_{\alpha \in A}$ admet un élément maximal. Posons $I = \cup_{\alpha \in A} I_\alpha$ et $x : I \rightarrow \mathbf{R}^n$ la fonction définie par $x|_{I_\alpha} = x_\alpha$. x est bien définie, en effet soit t dans I , alors il existe I_α qui contient t et $x(t) = x_\alpha(t)$ ne dépend pas de α , si $t \in I_\alpha \cap I_\beta$ alors on a $(I_\alpha, x_\alpha) \leq (I_\beta, x_\beta)$ ou $(I_\beta, x_\beta) \leq (I_\alpha, x_\alpha)$, et donc $x_\alpha(t) = x_\beta(t)$. On a bien évidemment $(I_\alpha, x_\alpha) \leq (I, x)$ il reste donc à montrer que (I, x) est bien une solution.

- I est un intervalle. Soit donc $(s, t) \in I^2$. Alors il existe I_α et I_β qui contiennent respectivement s et t . Mais on a soit $I_\alpha \subset I_\beta$, soit $I_\beta \subset I_\alpha$, donc soit $[s, t] \subset I_\alpha \subset I$, soit $[s, t] \subset I_\beta \subset I$.
- $t_0 \in I$ et $x(t_0) = x_0$.
- Pour tout $t \in I$, il existe α tel que $t \in I_\alpha$ et $x(t) = x_\alpha(t)$, par suite $(t, x(t)) \in \Omega$.
- Pour tout $t \in I$, il existe α tel que I_α contient t et (I_α, x_α) est une solution. Donc $x(t) = x_\alpha(t) = x_0 + \int_{t_0}^t f(s, x(s)) ds$.

■

Corollaire 3.1.6

Sous les hypothèses du théorème de Cauchy-Lipschitz, toute solution se prolonge en une solution maximale (I, x) , cette solution maximale est unique et I est un intervalle ouvert.

► L'existence d'une solution maximale vient de l'application du lemme de Zorn et l'unicité du théorème d'existence et d'unicité de Cauchy-Lipschitz. Pour montrer que I est un intervalle ouvert il suffit de voir que si $I =]t_0 - \eta, t_0 + \eta_0]$ on peut la prolonger en $t_0 + \eta_0$ grâce au théorème de Cauchy-Lipschitz. ■

Remarque 3.1.2. Le résultat d'existence et d'unicité d'une solution maximale implique que les courbes solutions $x(t, t_0, x_0)$ ne peuvent se couper, elles sont confondues ou sans point commun, elles forment une partition de l'ouvert Ω (cf. Fig. 3.3).

Exemple 3.1.2. On considère le problème de Cauchy définie sur $\Omega = \mathbf{R} \times \mathbf{R}$

$$(IVP1) \begin{cases} \dot{x}(t) = -x^2(t) \\ x(t_0) = x_0. \end{cases}$$

Pour (t_0, x_0) fixé on a une unique solution maximale définie sur $I =]\omega_-(t_0, x_0), \omega_+(t_0, x_0)[$ qui dépend de (t_0, x_0) :

- Si $x_0 = 0$ alors $I =]-\infty, +\infty[; \omega_-(t_0, 0) = -\infty, \omega_+(t_0, 0) = +\infty$ et $x(t) = 0$;
- Si $x_0 > 0$ alors $I =]t_0 - 1/x_0, +\infty[$ et

$$x(t) = \frac{x_0}{(t - t_0)x_0 + 1};$$

- Si $x_0 < 0$ alors $I =]-\infty, t_0 - 1/x_0[$ et

$$x(t) = \frac{x_0}{(t - t_0)x_0 + 1}.$$

Ces résultats sont visualisés sur les figures 3.2 et 3.3.

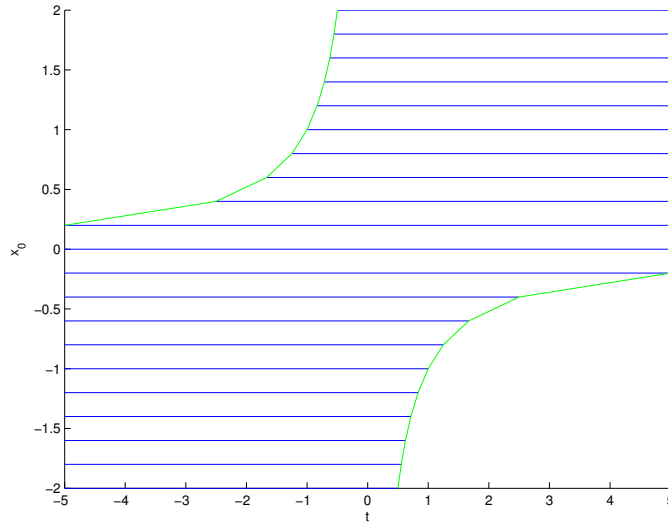


FIGURE 3.2 – Visualisation de l'ensemble $]\omega_-(0, x_0), \omega_+(0, x_0)[\times x_0$ pour l'exemple $\dot{x}(t) = -x^2(t), x(0) = x_0$.

□

Corollaire 3.1.7

Soit $f : I \times \mathbf{R}^n \rightarrow \mathbf{R}^n$, I intervalle ouvert, continue. On suppose qu'il existe une fonction continue $k : I \rightarrow \mathbf{R}^+$ telle que pour tout t fixé l'application f soit localement lipschitzienne par rapport à x de rapport $k(t)$. Alors la solution maximale est globale, c'est-à-dire définie sur I .

► Soit $[t_0 - \eta_1, t_0 + \eta_2] \subset I$. Dans la démonstration du théorème de Cauchy-Lipschitz, prenons $r_0 = +\infty, E = C^0([t_0 - \eta_1, t_0 + \eta_2], \mathbf{R}^n, d)$ et $K = \max_{[t_0 - \eta_1, t_0 + \eta_2]} k(t)$. $f(t, x)$ est alors lipschitzienne de rapport K sur cet intervalle et si p est tel que $(K^p/p!) \max(\eta_1, \eta_2)^p < 1$,

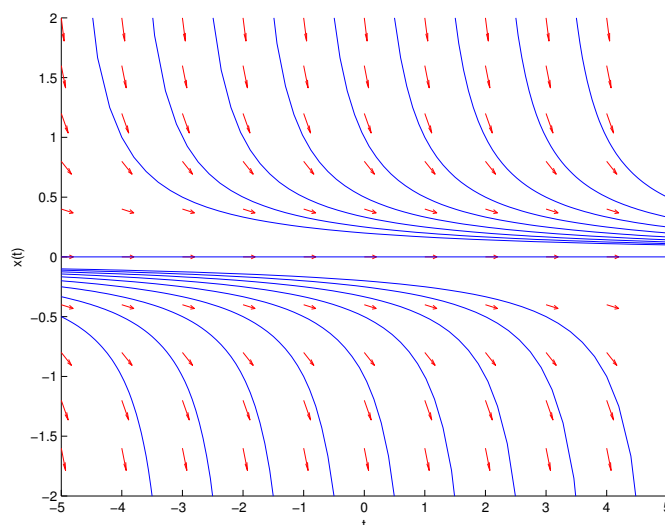


FIGURE 3.3 – Solutions pour le problème $\dot{x}(t) = -x^2(t), x(t_0) = x_0$, les courbes intégrales ne peuvent se couper, elles forment une partition de l'ouvert $\Omega = \mathbf{R}^2$.

alors T^p est une contraction. ■

Corollaire 3.1.8

On considère le système de Cauchy linéaire

$$(IVP2) \begin{cases} \dot{x} = A(t)x(t) + b(t) \\ x(t_0) = x_0. \end{cases}$$

Si $A(t)$ et $b(t)$ sont définis sur \mathbf{R} et continus alors on a existence et unicité de la solution sur $I = \mathbf{R}$. Le résultat est en particulier vérifié si le système est autonome.

► Il suffit de prendre $k(t) = \|A(t)\|$. En effet on a

$$\|f(t, x_1) - f(t, x_2)\| \leq \|A(t)(x_1 - x_2)\| \leq \|A(t)\| \|x_1 - x_2\|.$$
■

Remarque 3.1.3. Le résultat est encore vrai si $A(t)$ est localement intégrable.

Théorème 3.1.9 – Théorème d'explosion

On suppose $f : \Omega \subset \mathbf{R} \times \mathbf{R}^n \rightarrow \mathbf{R}^n$, Ω ouvert, continue dérivable par rapport à x et telle que l'application

$$\begin{aligned} \frac{\partial f}{\partial x} : \quad \Omega &\longrightarrow \mathcal{L}(\mathbf{R}^n, \mathbf{R}^n) \\ (t, x) &\longmapsto \frac{\partial f}{\partial x}(t, x) \end{aligned}$$

soit continue. Soit K un compact contenu dans Ω et $(t_0, x_0) \in K$, alors il existe η_1 et $\eta_2 > 0$ tels que pour tout $t \notin]t_0 - \eta_1, t_0 + \eta_2[$, $(t, x(t, t_0, x_0))$ soit extérieur à K .

► Nous allons montrer l'existence de η_2 .

Si $\omega_+(t_0, x_0) = +\infty$ alors le résultat est évident car K étant compact, on a $K \subset [t_1, t_2] \times B_f(x_0, R)$ et donc pour tout $t > t_2$, on aura $(t, x(t, t_0, x_0)) \notin K$. Supposons donc maintenant que $\omega_+ = \omega_+(t_0, x_0) \in \mathbf{R}$. Posons $F = \mathbb{C}_{\mathbf{R}^{n+1}}\Omega$ et $\rho = d(K, F) > 0$ et définissons le compact

$$K_{\rho/2} = \{(t, x) \in \mathbf{R} \times \mathbf{R}^n, d((t, x), K) \leq \rho/2\}.$$

On a $K \subset K_{\rho/2} \subset \Omega$. Considérons maintenant un point $(t_1, x_1) \in K$, alors l'ensemble $\Omega(t_1, x_1) = \{(t, x) \in \Omega, |t - t_1| < \eta \text{ et } \|x - x_1\| < a\}$ avec $\eta^2 + a^2 \leq \rho^2/4$ est un ouvert inclu dans $K_{\rho/2}$. Or $K_{\rho/2}$ est un compact, par suite pour tout $(t, x) \in \Omega(t_1, x_1)$, $\|f(t, x)\| \leq M$ et $\|\frac{\partial f}{\partial x}(t, x)\| \leq k$. On peut alors dans la démonstration du théorème de Cauchy-Lipschitz prendre $r_0 = +\infty$ et $\eta_0 = \eta$, ces quantités étant indépendantes de (t_1, x_1) . Montrons alors que $t_2 = \omega_+ - \eta = t_0 + \eta_2$ répond à la question. Raisonnons par l'absurde et supposons qu'il existe $\bar{t} \in]t_2, \omega_+]$ tel que $x(\bar{t}, t_0, x_0)$ soit dans K . Prenons $t_1 = \bar{t}$ et $x_1 = x(\bar{t}, t_0, x_0)$, la solution du système de Cauchy

$$(IVP3) \begin{cases} \dot{x}(t) = f(t, x(t)) \\ x(t_1) = x_1 \end{cases}$$

existe alors sur $]t_1 - \eta, t_1 + \eta[$. Par conséquent la fonction définie par

$$x(t) = \begin{cases} x(t, t_0, x_0) & \text{si } t \in]\omega_-, t_1[, \\ x(t, t_1, x_1) & \text{si } t \in [t_1, t_1 + \eta[\end{cases}$$

est une solution du problème de Cauchy de départ et $\bar{t} + \eta > t_2 + \eta > \omega_+$, ce qui est impossible. ■

Corollaire 3.1.10

Si dans le théorème précédent $\Omega = \mathbf{R} \times \mathbf{R}^n$ et $\omega_+(t_0, x_0) \in \mathbf{R}$ alors on a $\|x(t, t_0, x_0)\| \rightarrow +\infty$ quand $t \rightarrow \omega_+(t_0, x_0)$ (cf Fig. 3.3).



Exercice 3.1.3. On considère le problème de Cauchy définie sur $\Omega = \mathbf{R} \times \mathbf{R}$

$$(IVP4) \begin{cases} \dot{x}(t) = \sqrt{|x(t)|} \\ x(0) = 0. \end{cases}$$

1. Vérifier que $\varphi_1(t) = 0$ et $\varphi_2(t) = \frac{t|t|}{4}$ pour tout t dans \mathbf{R} sont solutions de (IVP)4.

2. Soit $a > 0$, vérifier que

$$\varphi_a(t) \begin{cases} 0 & \text{si } t \leq a \\ \varphi_2(t - a) & \text{si } t > a \end{cases}$$

pour tout t dans \mathbf{R} est solution de (IVP)??.

3. Quelle hypothèse du théorème d'existence de Cauchy Lipschitz n'est pas vérifiée ici? □

Si on suppose seulement que f est continue, alors on peut montrer qu'il existe une solution locale.

Théorème 3.1.11 – Théorème de Peano

On suppose que $f : \Omega \rightarrow \mathbf{R}^n$ est continue, que $\|f(t, x)\|$ est borné par A sur l'ensemble $D = \{(t, x), t_0 \leq t \leq t_f \text{ et } \|x - x_0\| \leq b\}$. Si $t_f - t_0 \leq b/A$, alors il existe une sous suite des polygones d'Euler qui converge vers une solution du problème de Cauchy.

► cf. [7] page 42. ■

On peut cependant perdre l'unicité. Par exemple le problème

$$(IVP5) \begin{cases} \dot{x}(t) = \sqrt{|x(t)|} \\ x(0) = 0, \end{cases}$$

admet les solutions $x_1(t) = 0$ et $x_2(t) = \frac{t|t|}{4}$ ainsi que

$$x_a(t) = \begin{cases} 0 & \text{si } t \leq a \\ x_2(t - a) & \text{si } t > a \end{cases}$$

(cf. la Fig. 3.4).

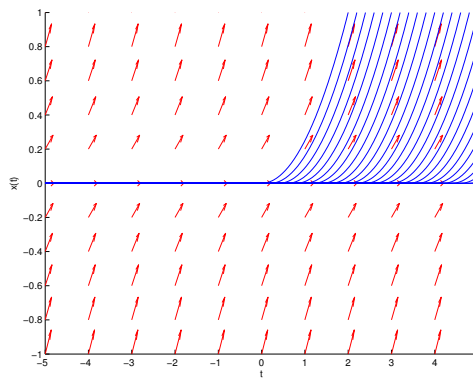


FIGURE 3.4 – Solutions pour le problème $\dot{x}(t) = \sqrt{|x(t)|}$, $x(0) = 0$.

3.2 Dépendances par rapports aux données

3.2.1 Introduction

On s'intéresse ici au problème de Cauchy

$$(IVP6) \begin{cases} \dot{x}(t) = f(t, x(t), \lambda) \\ x(0) = x_0, \end{cases}$$

avec

- $f : \Omega \subset \mathbf{R} \times \mathbf{R}^n \times \mathbf{R}^p \rightarrow \mathbf{R}^n$ continue ;
- Ω ouvert ;
-

$$\frac{\partial f}{\partial x} : \begin{array}{ccc} \Omega & \longrightarrow & \mathcal{L}(\mathbf{R}^n, \mathbf{R}^n) \\ (t, x, \lambda) & \longmapsto & \frac{\partial f}{\partial x}(t, x, \lambda) \end{array}$$

continue.

Sous ces hypothèses le problème de Cauchy

$$(IVP7)_\lambda \begin{cases} \dot{x}(t) = f(t, x(t), \lambda) \\ x(t_0) = x_0, \end{cases}$$

admet une solution locale maximale pour tout $(t_0, x_0, \lambda_0) \in \Omega$

$$\begin{aligned} x:]\omega_-(t_0, x_0, \lambda_0), \omega_+(t_0, x_0, \lambda_0)[&\longrightarrow \mathbf{R}^n \\ t &\longmapsto x(t, t_0, x_0, \lambda_0). \end{aligned}$$

Posons $\mathcal{O} = \{(t, t_0, x_0, \lambda_0) \in \mathbf{R} \times \Omega, \omega_-(t_0, x_0, \lambda_0) < t < \omega_+(t_0, x_0, \lambda_0)\}$, l'objectif de cette section est d'étudier les propriétés de continuité et de dérivabilité de la fonction

$$\begin{aligned} x: \mathcal{O} &\longrightarrow \mathbf{R}^n \\ (t, t_0, x_0, \lambda_0) &\longmapsto x(t, t_0, x_0, \lambda_0). \end{aligned}$$

3.2.2 Continuité

Nous allons dans un premier temps considérer la dépendance par rapport au paramètre λ . Posons ici $\mathcal{O} = \{(t, \lambda) \in \mathbf{R} \times \mathbf{R}^p, \omega_-(\lambda) < t < \omega_+(\lambda)\}$. On a alors le

Théorème 3.2.1

Sous les hypothèses de la sous-section précédente

1. \mathcal{O} est un ouvert ;
2. $x : \mathcal{O} \rightarrow \mathbf{R}^n$ est continue.

Pour démontrer ce théorème, nous avons besoin du

Lemme 3.2.1. *Soit $u(t)$ une fonction de \mathbf{R} à valeurs dans \mathbf{R} continue et α et β deux réels strictement positifs. On suppose que*

$$u(t) \leq \int_{t_0}^t (\alpha u(s) + \beta) ds,$$

alors on a

$$u(t) \leq \frac{\beta}{\alpha} (e^{\alpha(t-t_0)} - 1).$$

► Posons $v(t) = \int_{t_0}^t (\alpha v(s) + \beta) ds$, $v(t)$ est la solution du système de Cauchy

$$(IVP8) \begin{cases} \dot{x}(t) = \alpha x(t) + \beta \\ x(t_0) = 0. \end{cases}$$

Donc $v(t) = (\beta/\alpha)(e^{\alpha(t-t_0)} - 1)$. Considérons maintenant la suite $v_0(t) = u(t), \dots, v_i(t) = \int_{t_0}^t (\alpha v_{i-1}(s) + \beta) ds$. Cette suite converge uniformément sur tout compact vers $v(t)$ la solution du problème à valeur initiale linéaire (IVP8). Montrons par récurrence que l'on a toujours $v_i(t) \leq \int_{t_0}^t (\alpha v_i(s) + \beta) ds$ et $v_{i+1}(t) \geq v_i(t)$. Nous aurons donc montrer ainsi que

$$\frac{\beta}{\alpha} (e^{\alpha(t-t_0)} - 1) = v_\infty(t) \geq v_0(t) = u(t)$$

Ceci est évident pour $i = 0$. Supposons donc l'inégalité vraie pour i , alors par hypothèse de

réurrence

$$v_{i+1}(t) = \int_{t_0}^t (\alpha v_i(s) + \beta) ds \geq v_i(t).$$

Par suite on a, puisque α et β sont strictement positifs

$$\alpha v_{i+1}(t) + \beta \geq \alpha v_i(t) + \beta.$$

Et donc

$$\int_{t_0}^t (\alpha v_{i+1}(s) + \beta) ds \geq v_{i+1}(t).$$

■

► du théorème. Le fait que \mathcal{O} soit un ouvert est non trivial. Considérons $(t^*, \lambda^*) \in \mathcal{O}$ et supposons pour fixer les idées que $t^* \geq t_0$ (le cas $t^* \leq t_0$ se traite de la même manière). Comme $x(t^*, \lambda^*)$ existe, on a $t^* < \omega_+(\lambda^*)$. On peut donc choisir $t^* < t_1 < \omega_+(\lambda^*)$ et $x(t, \lambda^*)$ est définie pour tout $t \in [t_0, t_1]$. Définissons alors l'ensemble $\mathcal{C} = \{(t, x(t, \lambda^*), \lambda^*), t_0 \leq t \leq t_1\} \subset \Omega$ qui est un compact car c'est l'image du compact $[t_0, t_1]$ par l'application continue $t \rightarrow (t, x(t, \lambda^*), \lambda^*)$. \mathcal{C} est un compact de \mathbf{R}^{n+1} , donc il existe un coefficient de Lebesgue $\varepsilon > 0$ du recouvrement Ω (cf. (6.6.3.1)), c'est-à-dire tel que pour tout $(t, x, \lambda^*) \in \mathcal{C}$, $B((t, x, \lambda^*), \varepsilon) \subset \Omega$. Par suite il existe a et b strictement positifs tels que l'ensemble $K = \{(t, x, \lambda) \in \mathbf{R} \times \mathbf{R}^n \times \mathbf{R}^p, t_0 \leq t \leq t_1, \|x - x(t, \lambda^*)\| \leq a \text{ et } \|\lambda - \lambda^*\| \leq b\} \subset \Omega$.

De plus K est un fermé borné, donc un compact, et $f(t, x, \lambda)$ est uniformément continue sur K :

$$\forall \varepsilon, \exists \eta, \forall (t_1, x_1, \lambda_1) \in K, \forall (t_2, x_2, \lambda_2) \in K, \|(t_1, x_1, \lambda_1) - (t_2, x_2, \lambda_2)\| < \eta \Rightarrow \|f(t_1, x_1, \lambda_1) - f(t_2, x_2, \lambda_2)\| < \varepsilon$$

Considérons maintenant $x(t, t_0, x_0, \lambda)$ pour $\|\lambda - \lambda^*\| < b$. Le théorème d'explosion (3.1) implique que $x(\cdot, t_0, x_0, \lambda)$ quitte le compact K quand $t \rightarrow \omega_+(\lambda)$. Posons alors $t_2 = \inf\{t, x(t, t_0, x_0, \lambda) \notin K\}$. On a bien évidemment $t_0 < t_2 \leq t_1$ par définition de K . Nous allons montrer que $t_2 = t_1$.

$$\|x(t, t_0, x_0, \lambda) - x(t, t_0, x_0, \lambda^*)\| \leq \int_{t_0}^t \|f(s, x(s, t_0, x_0, \lambda), \lambda) - f(s, x(s, t_0, x_0, \lambda^*), \lambda^*)\| ds$$

Or

$$\begin{aligned} & \|f(s, x(s, t_0, x_0, \lambda), \lambda) - f(s, x(s, t_0, x_0, \lambda^*), \lambda^*)\| \\ & \leq \|f(s, x(s, t_0, x_0, \lambda), \lambda) - f(s, x(s, t_0, x_0, \lambda^*), \lambda)\| + \|f(s, x(s, t_0, x_0, \lambda^*), \lambda) - f(s, x(s, t_0, x_0, \lambda^*), \lambda^*)\| \\ & \leq k \|x(t, t_0, x_0, \lambda) - x(t, t_0, x_0, \lambda^*)\| \text{ car } f(t, x, \lambda) \text{ est Lipschitz par rapport à } x \\ & + \varepsilon \text{ uniforme continuité.} \end{aligned}$$

Par suite

$$\|x(t, t_0, x_0, \lambda) - x(t, t_0, x_0, \lambda^*)\| \leq \int_{t_0}^t (k \|x(t, t_0, x_0, \lambda) - x(t, t_0, x_0, \lambda^*)\| + \varepsilon) ds.$$

Donc en posant $u(t) = \|x(t, t_0, x_0, \lambda) - x(t, t_0, x_0, \lambda^*)\|$ dans le lemme (3.2.1) on obtient

$$u(t) = \|x(t, t_0, x_0, \lambda) - x(t, t_0, x_0, \lambda^*)\| \leq \frac{\varepsilon}{k} (e^{k(t-t_0)} - 1) = \varepsilon_2. \quad (3.1)$$

Prenons maintenant $\rho_2 \leq b, \varepsilon_2 < a$ et $\|\lambda - \lambda^*\| < \rho_2$, alors $(t_2, x(t_2, t_0, x_0, \lambda^*), \lambda^*) \in \partial K$ par définition de t_2 , mais

- $\|\lambda - \lambda^*\| < \rho_2 < b$;
- $\|x(t, t_0, x_0, \lambda) - x(t, t_0, x_0, \lambda^*)\| \leq \varepsilon_2 < a$.

Il faut donc que $t_2 = t_1$. En conclusion pour $t^* \geq t_0$, il existe $t_2 > t^*$ et $\rho_2 > 0$ tels que pour tout t , $t_0 \leq t \leq t_2$ et tout λ , $\|\lambda - \lambda^*\| \leq \rho_2$, $x(t, t_0, x_0, \lambda)$ existe et donc $(t, \lambda) \in \mathcal{O}$. Par une démonstration identique on peut construire $t_1 < t^*$ et donc un ouvert $]t_1, t_2[\times B(\lambda^*, \rho) \subset \mathcal{O}$. On en déduit que \mathcal{O} est bien un ouvert. Quant à la continuité, elle provient immédiatement de (3.1). ■

Théorème 3.2.2

Sous les hypothèses de la sous-section précédente

1. les fonctions $\omega_- : (t_0, x_0, \lambda_0) \rightarrow \omega_-(t_0, x_0, x_0)$ et $\omega_+ : (t_0, x_0, \lambda_0) \rightarrow \omega_+(t_0, x_0, x_0)$ sont respectivement semi-continue inférieurement et semi-continue supérieurement ;
2. \mathcal{O} est un ouvert ;
3. $x : \mathcal{O} \rightarrow \mathbf{R}^n$ est continue ;
4. si f est indépendante de λ , $(t_0, x_0) \in \Omega \subset \mathbf{R} \times \mathbf{R}^n$ et si $\omega_-(t_0, x_0) < a < b < \omega_+(t_0, x_0)$, alors pour tout $r > 0$, il existe $s > 0$ et $M \geq 0$ tels que pour tout (t_1, x_1) vérifiant $t_1 \in [a, b]$ et $\|x_1 - x(t_1, t_0, x_0)\| \leq s$, on ait $\omega_-(t_1, x_1) < a < b < \omega_+(t_1, x_1)$ et pour tout $t \in [a, b]$

$$\begin{aligned} \|x(t, t_1, x_1) - x(t, t_0, x_0)\| &\leq r \\ &\leq M|t_1 - t_0| + M\|x_1 - x_0\| \end{aligned}$$

► cf. [14] page 344 ou [10] ■

Remarque 3.2.1. On suppose f indépendante de λ , $f : \Omega \in \mathbf{R} \times \mathbf{R}^n \rightarrow \mathbf{R}^n$, et on considère $t_0 < t_1$. On pose

$$\begin{aligned} \mathcal{O}_0 &= \{x_0 \in \mathbf{R}^n, (t_0, x_0) \in \Omega, t_1 < \omega_+(t_0, x_0)\} \\ \mathcal{O}_1 &= \{x_1 \in \mathbf{R}^n, (t_1, x_1) \in \Omega, \omega_-(t_1, x_1) < t_0\}. \end{aligned}$$

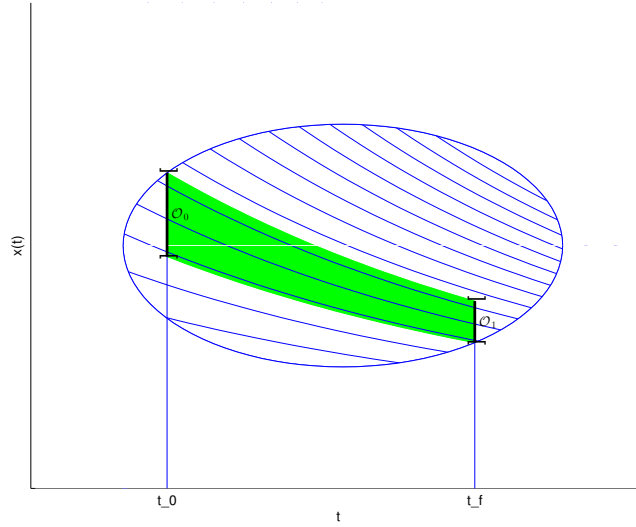
Dire que $x_0 \in \mathcal{O}_0$, c'est dire que l'intervalle de définition de $x(t, t_0, x_0)$ contient t_1 , le point $x(t_1, t_0, x_0)$ appartient alors à \mathcal{O}_1 . Ces ensembles \mathcal{O}_0 et \mathcal{O}_1 sont des ouverts et l'application

$$\begin{aligned} R(t_0, t_1): \quad \mathcal{O}_0 &\longrightarrow \mathcal{O}_1 \\ x_0 &\longmapsto x(t_1, t_0, x_0) \end{aligned}$$

est un homéomorphisme (cf. la figure 3.5).

3.2.3 Dérivée

On suppose dans cette sous-section que $\Lambda = \mathbf{R}^p$.

FIGURE 3.5 – Homéomorphisme entre \mathcal{O}_0 et \mathcal{O}_1 .**Théorème 3.2.3**

Soit $f : \Omega \subset \mathbf{R} \times \mathbf{R}^n \times \mathbf{R}^p \rightarrow \mathbf{R}^n$, Ω ouvert, f continue. On suppose que f admet des dérivées partielles continues par rapport à x et λ

$$\frac{\partial f}{\partial x} : \Omega \rightarrow \mathcal{M}_n(\mathbf{R}) \quad \frac{\partial f}{\partial \lambda} : \Omega \rightarrow \mathcal{M}_{(n,p)}(\mathbf{R}).$$

Alors, la fonction $x : \mathcal{O} \rightarrow E$ est de classe C^1 . Elle est de classe C^{k+1} , $1 \leq k \leq \infty$, si f est de classe C^k . De plus

1. La fonction $\frac{\partial x}{\partial x_0}(\cdot, t_0, x_0, \lambda_0) :]\omega_-(t_0, x_0, \lambda_0), \omega_+(t_0, x_0, \lambda_0)[\rightarrow \mathcal{M}_n(\mathbf{R})$ est la solution au problème de Cauchy linéaire

$$(VAR9)^a \begin{cases} \dot{X}(t) = A(t)X(t) \\ X(t_0) = I_n, \end{cases}$$

où $A(t) = \frac{\partial f}{\partial x}(t, x(t, t_0, x_0, \lambda_0), \lambda_0) \in \mathcal{M}_n(\mathbf{R})$ et I_n est la matrice identité d'ordre n .

2. La fonction $\frac{\partial x}{\partial t_0}(\cdot, t_0, x_0, \lambda_0) :]\omega_-(t_0, x_0, \lambda_0), \omega_+(t_0, x_0, \lambda_0)[\rightarrow \mathbf{R}^n$ est la solution au problème de Cauchy linéaire

$$(VAR10) \begin{cases} \dot{x}(t) = A(t)x(t) \\ x(t_0) = -f(t_0, x_0, \lambda_0). \end{cases}$$

3. La fonction $\frac{\partial x}{\partial \lambda_0}(\cdot, t_0, x_0, \lambda_0) :]\omega_-(t_0, x_0, \lambda_0), \omega_+(t_0, x_0, \lambda_0)[\rightarrow \mathcal{M}_{n,p}(\mathbf{R})$ est la solution au problème de Cauchy linéaire

$$(VAR11) \begin{cases} \dot{X}(t) = A(t)X(t) + B(t) \\ X(t_0) = 0, \end{cases}$$

avec $B(t) = \frac{\partial f}{\partial \lambda}(t, x(t; t_0, x_0, \lambda_0), \lambda_0) \in \mathcal{M}_{n,p}(\mathbf{R})$.

a. Équations variationnelles.

► Supposons que l'on ait la dérivabilité alors, puisque nous avons par définition

$$x(t, t_0, x_0, \lambda_0) = x_0 + \int_{t_0}^t f(s, x(s, t_0, x_0, \lambda_0), \lambda_0) ds,$$

nous pouvons écrire grâce aux théorèmes classiques du calcul intégral

$$\frac{\partial x}{\partial x_0}(t, t_0, x_0, \lambda_0) = I + \int_{t_0}^t \frac{\partial f}{\partial x}(s, x(s, t_0, x_0, \lambda_0), \lambda_0) \frac{\partial x}{\partial x_0}(s, t_0, x_0, \lambda_0) ds.$$

D'où le résultat.

Les autres formules s'obtiennent de façon similaire. Pour une démonstration complète du théorème nous renvoyons à [14] page 350 ou [10] ■

Intégration numérique, les méthodes de Runge-Kutta

4.1	Introduction	35
4.2	Exemples	35
4.2.1	Exemple 1	35
4.2.2	Modèle de Lorenz	36
4.2.3	Exemple de Roberston	37
4.3	Définitions et exemples	38
4.4	Méthodes de Runge-Kutta explicite	40
4.4.1	Définition	40
4.4.2	Ordre	40
4.4.3	Convergence	43
4.5	Erreurs d'arrondi	48
4.6	Contrôle du pas	49
4.6.1	Introduction	49
4.6.2	Extrapolation de Richardson	49
4.6.3	Méthode de Runge-Kutta emboîtées	50
4.7	Les méthodes de Runge-Kutta implicites	54
4.8	Exercices	56

4.1 Introduction

L'objectif de cette partie est d'étudier les algorithmes numériques utilisés dans les programmes actuels, en particulier les programmes ODE45 de `Matlab`[\[12\]](#) et DOPRI5 du professeur Hairer [\[7\]](#).

4.2 Exemples

4.2.1 Exemple 1

$$(IVP) \begin{cases} \dot{x}_1(t) = x_1(t) + x_2(t) + \sin t \\ \dot{x}_2(t) = -x_1(t) + 3x_2(t) \\ x_1(0) = -9/25 \\ x_2(0) = -4/25. \end{cases}$$

La solution est

$$\begin{aligned} x_1(t) &= (-1/25)(13 \sin t + 9 \cos t) \\ x_2(t) &= (-1/25)(3 \sin t + 4 \cos t). \end{aligned}$$

Sur cet exemple on constate que si on utilise les valeurs des paramètres par défauts dans les codes d'intégrations numériques on peut très rapidement obtenir des résultats faux. Par exemple sur la Fig. 4.1, la solution calculée pour les valeurs par défauts des paramètres RelTol et AbsTol de ODE45 n'est pas du tout périodique.

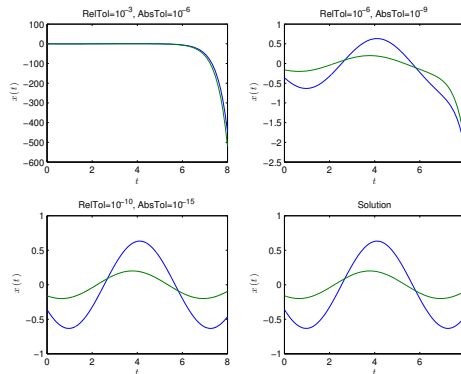


FIGURE 4.1 – Solutions calculées avec ODE45, RelTol=10⁻³ et AbsTol=10⁻⁶, RelTol=10⁻⁶ et AbsTol=10⁻⁹, RelTol=10⁻¹⁰ et AbsTol=10⁻¹⁵ et solution exacte

4.2.2 Modèle de Lorenz

$$(IVP) \begin{cases} \dot{x}_1(t) = -\sigma x_1(t) + \sigma x_2(t) \\ \dot{x}_2(t) = -x_1(t)x_3(t) + rx_1(t) - x_2(t) \\ \dot{x}_3(t) = x_1(t)x_2(t) - bx_3(t) \\ x_1(0) = -8 \\ x_2(0) = 8 \\ x_3(0) = r - 1, \end{cases}$$

avec $\sigma = 10, r = 28, b = 8/3$. La solution est visualisée sur la figure 4.2.

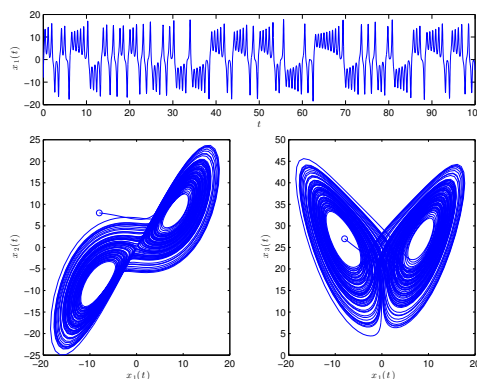


FIGURE 4.2 – Solutions pour le problème de Lorenz.

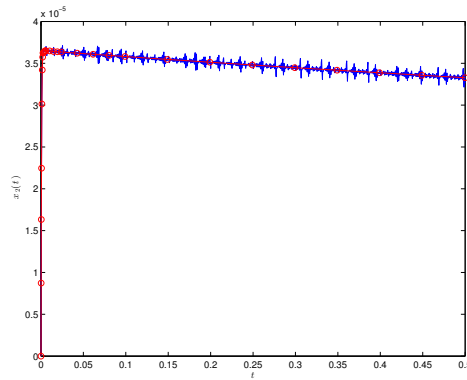
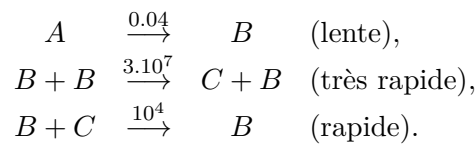


FIGURE 4.3 – Solutions numérique pour le problème raide obtenues par une méthode classique (ODE45) et par une méthode pour les équations différentielles raides (ODE15S)

4.2.3 Exemple de Roberston

On considère la réaction chimique



Le système différentiel associé à cette réaction chimique, qui est un exemple de problème raide¹, est donnée par

$$(IVP) \begin{cases} \dot{x}_1(t) = -0.04x_1(t) + 10^4x_2(t)x_3(t) \\ \dot{x}_2(t) = 0.04x_1(t) - 10^4x_2(t)x_3(t) - 3 \cdot 10^7x_2^2(t) \\ \dot{x}_3(t) = 3 \cdot 10^7x_2^2(t) \\ x_1(0) = 1 \\ x_2(0) = 0 \\ x_3(0) = 0, \end{cases}$$

et la solution calculée par les deux programmes ODE45 et ODE15S sont visualisées sur la figure 4.3.

Exemple 4.2.1 (Orbite d'Arenstorf). On considère deux corps de masses $(1-\mu)$ et μ en rotation circulaire dans le plan, et un troisième corps de masse négligeable dont on souhaite étudier le mouvement $x(t) = (x_1(t), x_2(t))^T$ en fonction de l'attraction des 2 autres corps dans le même plan. Les équations du mouvement sont

$$(IVP) \begin{cases} \ddot{x}_1(t) = x_1(t) + 2\dot{x}_2(t) - \mu' \frac{x_1(t)+\mu}{r_1^3(t)} - \mu \frac{x_1(t)-\mu'}{r_2^3(t)} \\ \ddot{x}_2(t) = x_2(t) - 2\dot{x}_1(t) - \mu' \frac{x_2(t)}{r_1^3(t)} - \mu \frac{x_2(t)}{r_2^3(t)} \\ r_1(t) = \sqrt{(x_1(t)+\mu)^2 + x_2^2(t)} \\ r_2(t) = \sqrt{(x_1(t)-\mu')^2 + x_2^2(t)} \\ x_1(0) = 0.994 \\ \dot{x}_1(0) = 0 \\ x_2(0) = 0 \\ \dot{x}_2(0) = -2.00158510637908252240537862224, \end{cases}$$

1. stiff problem en anglais.

avec $\mu = 0.012277471$ et $\mu' = 1 - \mu$. La solution est alors périodique de période $t_f = T = 17.0652165601579625588917206249$. Elle est visualisée sur la figure 4.4.

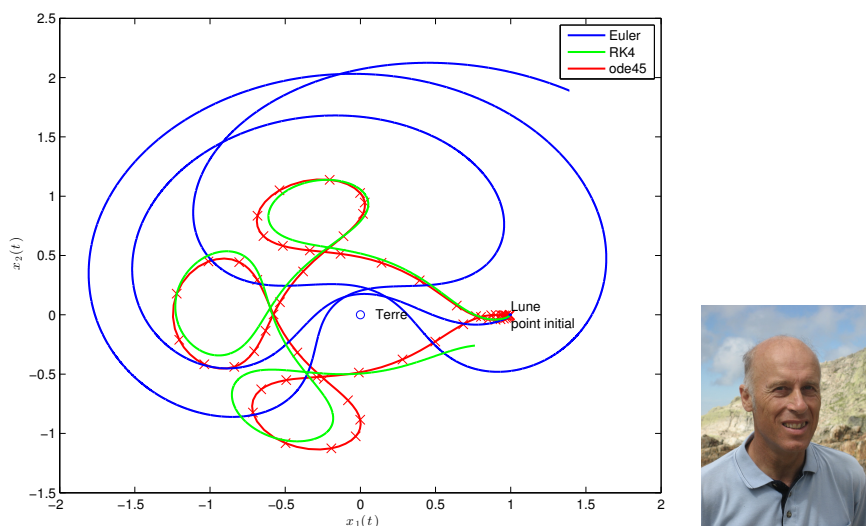


FIGURE 4.4 – Orbite d'Arenstorf calculée avec Euler (24000 pas équidistants), RK4 (6000 pas équidistants) et ODE45 (64 pas variables), cf. [7] page 130. Le professeur Ernst Hairer est né en 1949 et est le père de Martin Hairer, médaille Fields 2014.

□

4.3 Définitions et exemples

On désire calculer la solution sur l'intervalle $I = [t_0, t_f]$ du problème de Cauchy

$$(IVP) \begin{cases} \dot{x}(t) = f(t, x(t)) \\ x(t_0) = x_0. \end{cases}$$

On considère pour cela une subdivision $t_0 < t_1 < \dots < t_N = t_f$ de I . On note $h_i = t_{i+1} - t_i, i = 0, N - 1$, les pas, et $h_{max} = \max_i(h_i)$. Nous allons calculer successivement les valeurs approchées x_1, \dots, x_N de $x(t_1), \dots, x(t_N)$.

Définition 4.3.1 – Méthodes à un pas explicite

On appelle méthode à un pas explicite toute méthode pour laquelle la valeur de x_{i+1} est calculée de façon explicite en fonction de t_i, x_i et h_i :

$$x_{i+1} = x_i + h_i \Phi(t_i, x_i, h_i). \quad (4.1)$$

Remarque 4.3.1. Pour simplifier les notations, nous n'écrirons dans la suite que le premier pas

$$x_1 = x_0 + h \Phi(t_0, x_0, h).$$

Définition 4.3.2 – Schéma d'Euler (1768)

On appelle méthode d'Euler explicite le schéma

$$x_1 = x_0 + hf(t_0, x_0). \quad (4.2)$$

Remarque 4.3.2. Le schéma d'Euler explicite est tout simplement une approximation de l'intégrale $\int_{t_0}^{t_1} f(s, x(s))ds$ par $hf(t_0, x_0)$.

Exemple 4.3.1.

$$(IVP) \begin{cases} \dot{x}(t) = t^2 + x^2(t) \\ x(-1.5) = -1.4 \end{cases}$$

□

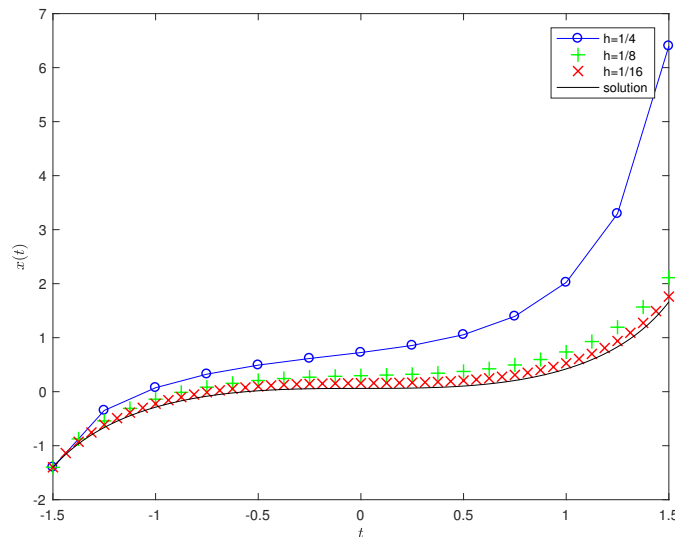


FIGURE 4.5 – Schéma d'Euler, $h = 1/2, 1/4, 1/8$.

L'idée évidente pour améliorer la précision numérique est d'approcher cette intégrale par une formule de quadrature ayant un ordre plus élevé. Si on exploite, pour améliorer l'approximation de l'intégrale, le point milieu, nous obtenons

$$x(t_1) \approx x_0 + hf(t_0 + \frac{h}{2}, x(t_0 + \frac{h}{2})).$$

Mais on ne connaît pas la valeur de $x(t_0 + \frac{h}{2})$, d'où l'idée d'approximer cette quantité par un pas d'Euler : $x(t_0 + \frac{h}{2}) \approx x_0 + \frac{h}{2}f(t_0, x_0)$. Nous obtenons ainsi le schéma de Runge.

Définition 4.3.3 – Schéma de Runge (1895)

$$x_1 = x_0 + hf(t_0 + \frac{h}{2}, x_0 + \frac{h}{2}f(t_0, x_0)). \quad (4.3)$$

4.4 Méthodes de Runge-Kutta explicite

4.4.1 Définition

Définition 4.4.1 – Méthode de Runge-Kutta explicite

On appelle méthode de Runge-Kutta explicite à s étages, la méthode définie par le schéma

$$\begin{aligned} k_1 &= f(t_0, x_0) \\ k_2 &= f(t_0 + c_2 h, x_0 + h a_{21} k_1) \\ &\vdots \\ k_s &= f(t_0 + c_s h, x_0 + h \sum_{i=1}^{s-1} a_{si} k_i) \\ x_1 &= x_0 + h \sum_{i=1}^s b_i k_i \end{aligned} \quad (4.4)$$

où les coefficients c_i, a_{ij} et b_i sont des constantes qui définissent précisément le schéma. On supposera toujours dans la suite que $c_1 = 0$ et $c_i = \sum_{j=1}^{i-1} a_{ij}$ pour $i = 2, \dots, s$.

On représente en pratique ce schéma par le tableau de Butcher[3], cf. la table 4.1.

c_1					
c_2	a_{21}				
c_3	a_{31}	a_{32}			
\vdots	\vdots	\vdots	\ddots		
c_s	a_{s1}	a_{s2}	\dots	a_{ss-1}	
	b_1	b_2	\dots	b_{s-1}	b_s



TABLE 4.1 – Tableau de Butcher (né en 1933).

Exemple 4.4.1. On considère par exemple les schémas de la table ??

□



Exercice 4.4.2. On considère le schéma de Heun, cf. la table 4.2.

1. Écrire le schéma de Runge-Kutta correspondant.
2. Donner explicitement $\Phi(t_0, x_0, h)$.

□

4.4.2 Ordre

Rappelons tout d'abord les notations de Landau

Définition 4.4.2 – Notations de Landau

1. L'équation (4.5) ci-après signifie qu'il existe un voisinage U de 0 et il existe une

$\begin{array}{c c} 0 & \\ \hline & 1 \end{array}$	$\begin{array}{c cc} 0 & & \\ \hline 1/2 & 1/2 & \\ \hline & 0 & 1 \end{array}$	$\begin{array}{c cc} 0 & & \\ \hline 1/3 & 1/3 & \\ \hline 2/3 & 0 & 2/3 \\ \hline & 1/4 & 0 & 3/4 \end{array}$
Euler (ordre 1)	Runge (ordre 2)	Heun (ordre3)
$\begin{array}{c ccc} 0 & & & \\ \hline 1/2 & 1/2 & & \\ \hline 1/2 & 0 & 1/2 & \\ \hline 1 & 0 & 0 & 1 \\ \hline & 1/6 & 2/6 & 2/6 & 1/6 \end{array}$	$\begin{array}{c ccc} 0 & & & \\ \hline 1/3 & 1/3 & & \\ \hline 2/3 & -1/3 & 1 & \\ \hline 1 & 1 & -1 & 1 \\ \hline & 1/8 & 3/8 & 3/8 & 1/8 \end{array}$	
La méthode RK4 (ordre 4)	RK4, règle 3/8 (ordre 4)	

TABLE 4.2 – Schémas de Runge-Kutta classiques.

constante C telle que pour tout $h \in U$ on a $\|e(h)\| \leq C|h|^p$.

$$e(h) = O(h^p) \quad (4.5)$$

2. L'équation (4.6) ci-après signifie qu'il existe une fonction $\varepsilon(h)$ à valeurs réelles telle que $\|e(h)\| = |h|^p \varepsilon(h)$ avec $\varepsilon(h) \rightarrow 0$ quand $h \rightarrow 0$.

$$e(h) = o(h^p) \quad (4.6)$$

Définition 4.4.3 – Ordre

On dit d'une méthode à un pas est d'ordre $p \geq 1$ si, pour tout problème de Cauchy avec f suffisamment dérivable, l'erreur sur un pas, appelée erreur locale satisfait

$$e(h) = x_1 - x(t_1, t_0, x_0) = O(h^{p+1}), \quad (4.7)$$

où x_1 est la valeur calculée par le schéma et $x(t_1, t_0, x_0)$ est la valeur exacte du problème de Cauchy avec la valeur initiale $x(t_0) = x_0$.

Remarque 4.4.1. Attention, un schéma d'ordre p a une erreur locale en $O(h^{p+1})$. Nous verrons au corollaire (4.4.3) que c'est l'erreur globale $x_N - x(t_f, t_0, x_0)$ qui est en $O(h_{max}^p)$.

Exemple 4.4.3. Le schéma d'Euler explicite est d'ordre $p = 1$ car par définition de la dérivée on a

$$\begin{aligned} x(t_1) &= x(t_0 + h) = x_0 + h\dot{x}(t_0) + O(h^2) = x_0 + hf(t_0, x_0) + O(h^2) \\ &= x_1 + O(h^2). \end{aligned}$$

□

Nous allons maintenant étudier les relations que doivent vérifier les coefficients a_{ij} , b_i et c_i pour qu'un schéma de Runge-Kutta explicite à 2 étages soit d'ordre 2. Pour cela nous allons

comparer les développements de Taylor à l'ordre 2 de x_1 et de $x(t_1, t_0, x_0)$. Pour x_1 , nous avons

$$\begin{aligned} k_1 &= f(t_0, x_0) \\ k_2 &= f(t_0 + c_2 h, x_0 + a_{21} h k_1) \\ x_1 &= x_0 + h(b_1 k_1 + b_2 k_2) \\ &= x_0 + h(b_1 f(t_0, x_0) + b_2 f(t_0 + c_2 h, x_0 + a_{21} h f(t_0, x_0))) \\ &= x_0 + h(b_1 + b_2) f(t_0, x_0) + h^2 b_2 (c_2 f_t(t_0, x_0) + a_{21} f_x(t_0, x_0) f(t_0, x_0)) + O(h^3). \end{aligned}$$

Quand à la solution exacte sur un pas, nous avons

$$\begin{aligned} x(t_0 + h, t_0, x_0) &= x_0 + h \dot{x}(t_0) + \frac{h^2}{2} \ddot{x}(t_0) + O(h^3) \\ &= x_0 + h f(t_0, x_0) + \frac{h^2}{2} (f_t(t_0, x_0) + f_x(t_0, x_0) f(t_0, x_0)) + O(h^3). \end{aligned}$$

Par suite en identifiant les termes des coefficients des puissances de h et en utilisant $c_2 = a_{21}$, on peut voir que le schéma sera d'ordre deux si et seulement si on a

$$b_1 + b_2 = 1 \quad \text{et} \quad b_2 a_{21} = \frac{1}{2}.$$

Si on considère la méthode de Runge de la table 4.2, on voit qu'elle est d'ordre 2.



Exercice 4.4.4. On considère le schéma de Runge-Kutta explicite à 3 étages

$$\begin{array}{c|cc} 0 & & \\ c_2 & a_{21} & \\ c_3 & a_{31} & a_{32} \\ \hline & b_1 & b_2 & b_3 \end{array} \quad \text{avec} \quad c_i = \sum_{j=1}^s a_{ij},$$

1. Vérifier que, dans le cas d'un système autonome, les relations que doivent vérifier les coefficients pour avoir un schéma d'ordre 3 sont

$$\begin{cases} b_1 + b_2 + b_3 = 1 \\ b_2 a_{21} + b_3 a_{31} + b_3 a_{32} = \frac{1}{2} \\ b_2 a_{21}^2 + b_3 (a_{31}^2 + 2a_{31} a_{32} + a_{32}^2) = \frac{1}{3} \\ b_3 a_{32} a_{21} = \frac{1}{6} \end{cases}$$

□

Théorème 4.4.4

Les méthodes de Heun et de Runge-Kutta RK4 de la table 4.2 sont respectivement d'ordre 3 et 4.

► cf. [7].

■

Remarque 4.4.2. On démontre que :

- $p = 4$ est possible pour $s = 4$ et que l'on a 8 conditions ;
- pour avoir $p = 5$, il faut que $s \geq 6$;

- pour avoir $p = 8$, il faut $s \geq 11$ et il y a 200 conditions ;
- pour $p = 10$, il y a 1205 conditions.

4.4.3 Convergence

Définition 4.4.5 – Erreur de consistance

Soit $x(., t_0, x_0)$ la solution du problème de Cauchy à valeur initiale $x(t_0) = x_0$. L'erreur de consistance e_{i+1} est l'erreur locale en t_{i+1} obtenue par le schéma à un pas en partant de la solution à l'instant t_i (cf. Fig. 4.6).

$$e_{i+1} = x(t_{i+1}, t_0, x_0) - x_{i+1} = \int_{t_i}^{t_{i+1}} f(s, x(s)) ds - h_i \Phi(t_i, x(t_i, t_0, x_0), h_i). \quad (4.8)$$

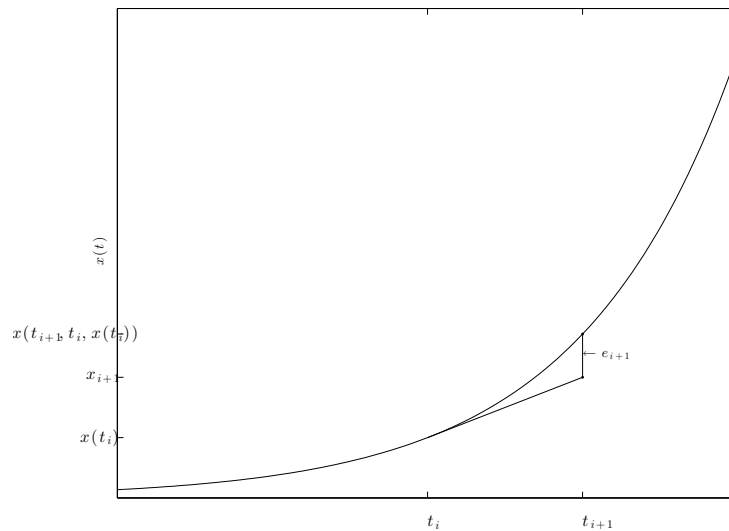


FIGURE 4.6 – Erreur de consistance du schéma d'Euler.

Définition 4.4.6 – Méthode consistante

On dit qu'une méthode à un pas est consistante avec l'équation différentielle $\dot{x}(t) = f(t, x(t))$ si pour toute solution x de l'équation différentielle on a

$$\sum_{i=0}^{N-1} \|e_i\| = \sum_{i=0}^{N-1} \|x(t_{i+1}, t_0, x_0) - h_i \Phi(t_i, x(t_i, t_0, x_0), h_i)\| \rightarrow 0, \quad (4.9)$$

lorsque $h_{\max} = \max_i(h_i) \rightarrow 0$.

Définition 4.4.7 – Méthode stable

On dit qu'une méthode à un pas est stable si et seulement si il existe une constante S

indépendante de h telle que pour toutes suites

$$x_{i+1} = x_i + h_i \Phi(t_i, x_i, h_i) \quad 0 \leq i \leq N-1 \quad (4.10)$$

$$\tilde{x}_{i+1} = \tilde{x}_i + h_i \Phi(t_i, \tilde{x}_i, h_i) + \varepsilon_{i+1} \quad 0 \leq i \leq N-1, \quad (4.11)$$

on ait

$$\max_{0 \leq i \leq N} \|\tilde{x}_i - x_i\| \leq S(\|\tilde{x}_0 - x_0\| + \sum_{i=1}^N \|\varepsilon_i\|). \quad (4.12)$$

Définition 4.4.8 – Méthode convergente

Une méthode à un pas est convergente si pour toute solution exacte $x(\cdot, t_0, x_0)$, la suite $(x_i)_i$ définie par $x_{i+1} = x_i + h_i \Phi(t_i, x_i, h_i)$ vérifie

$$\max_{0 \leq i \leq N} \|x(t_i, t_0, x_0) - x_i\| \rightarrow 0, \quad (4.13)$$

lorsque $x_0 \rightarrow x(t_0)$ et $h_{\max} \rightarrow 0$.

Remarque 4.4.3. La quantité $\max_{0 \leq i \leq N} \|x(t_i, t_0, x_0) - x_i\|$ dans l'équation (4.13) est l'erreur globale.

Calculons cette erreur globale. Posons $\tilde{x}_i = x(t_i, t_0, x_0)$. Par définition de l'erreur de consistance on a

$$\tilde{x}_{i+1} = \tilde{x}_i + h_i \Phi(t_i, \tilde{x}_i, h_i) + e_{i+1}.$$

Si la méthode est stable, nous en déduisons que

$$\max_{0 \leq i \leq N} (\|x(t_i, t_0, x_0) - x_i\|) \leq S(\|\tilde{x}_0 - x_0\| + \sum_{1 \leq i \leq N} \|e_i\|).$$

Nous en déduisons le théorème

Théorème 4.4.9

Si la méthode est stable et consistante, alors, elle est convergente.

Corollaire 4.4.10

Si la méthode est stable, consistante, d'ordre p et si $\tilde{x}_0 = x_0$ alors l'erreur globale est en $O(h^p)$.

► Il suffit d'écrire (cf. Fig 4.7) que

$$\begin{aligned}
 \max_{0 \leq i \leq N} \|x(t_i, t_0, x_0) - x_i\| &\leq S \sum_{0 \leq i \leq N-1} \|e_{i+1}\| \\
 &\leq S \sum_{0 \leq i \leq N-1} Ch_i^{p+1} \\
 &\leq SC(h_{\max})^p \sum_{0 \leq i \leq N-1} h_i \\
 &\leq SC(t_f - t_0)(h_{\max})^p
 \end{aligned}$$

■

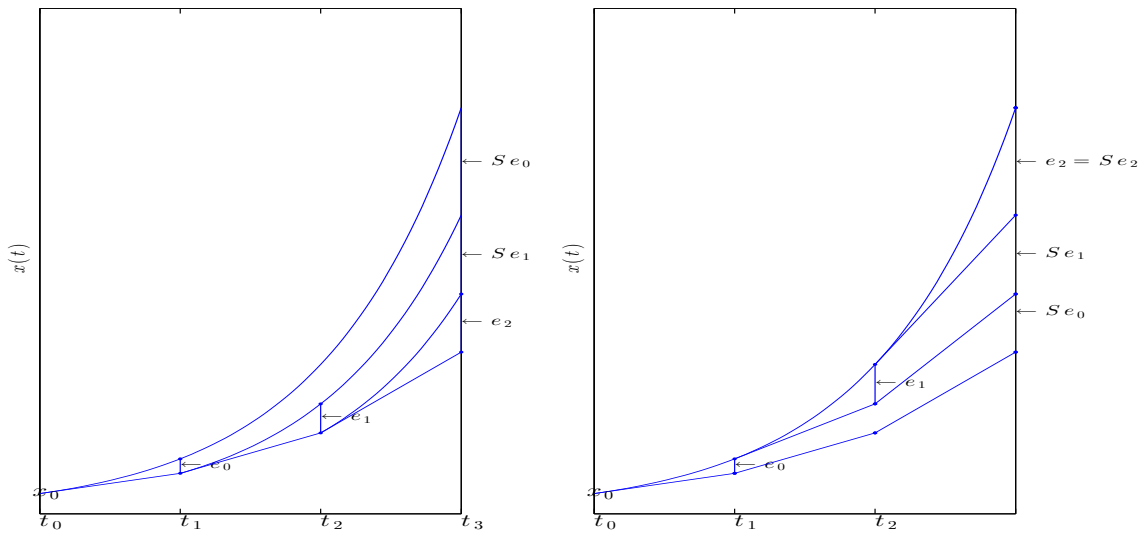


FIGURE 4.7 – Convergence, attention l'erreur de consistance et la propagation des erreurs sont visualisées sur le graphique de droite et non de gauche.

Théorème 4.4.11

Une méthode à un pas est consistante si et seulement si

$$\forall (t, x) \in [t_0, t_f] \times \mathbf{R}^n, \Phi(t, x, 0) = f(t, x).$$

► Par définition, nous avons

$$e_{i+1} = x(t_{i+1}, t_0, x_0) - x(t_i, t_0, x_0) - h_i \Phi(t_i, x(t_i, t_0, x_0), h_i).$$

Le théorème des accroissements finis appliqué à $x(t)$ sur l'intervalle $[t_i, t_{i+1}]$ implique qu'il existe $c_i \in]t_i, t_{i+1}[$ tel que

$$x(t_{i+1}, t_0, x_0) - x(t_i, t_0, x_0) = h_i \dot{x}(c_i) = h_i f(c_i, x(c_i)).$$

Par suite, nous pouvons écrire

$$\begin{aligned} e_{i+1} &= h_i(f(c_i, x(c_i)) - \Phi(t_i, x(t_i, t_0, x_0), h_i)) = h_i(\alpha_i + \beta_i) \\ \alpha_i &= f(c_i, x(c_i)) - \Phi(c_i, x(c_i), 0) \\ \beta &= \Phi(c_i, x(c_i), 0) - \Phi(t_i, x(t_i, t_0, x_0), h_i). \end{aligned}$$

Or la fonction $(t, h) \rightarrow \Phi(t, x(t), h)$ est continue. Elle est donc uniformément continue sur le compact $[t_0, t_f] \times [0, \delta]$, et donc pour tout $\varepsilon > 0$, il existe $\eta > 0$, tels que $h_{\max} \leq \eta (\Rightarrow |c_i - t_i| < \eta) \Rightarrow \|\beta_i\| < \varepsilon$.

$$\begin{aligned} \left| \sum_{i=0}^{N-1} (\|e_{i+1}\| - h_i \|\alpha_i\|) \right| &\leq \sum_{i=0}^{N-1} \left| \|e_{i+1}\| - h_i \|\alpha_i\| \right| \\ &\leq \sum_{i=0}^{N-1} \|e_{i+1} - h_i \alpha_i\| \\ &\leq \sum_{i=0}^{N-1} \|h_i \beta_i\| \leq \varepsilon(t_f - t_0) \end{aligned}$$

Par suite, si les limites existent, on a $\lim_{h_{\max} \rightarrow 0} \sum_{i=1}^N \|e_i\| = \lim_{h_{\max} \rightarrow 0} \sum_{i=0}^{N-1} h_i \|\alpha_i\|$. Or, par définition de l'intégrale de Riemann, nous avons

$$\int_{t_0}^{t_f} \|f(t, x(t)) - \Phi(t, x(t), 0)\| dt = \lim_{h_{\max} \rightarrow 0} \sum_i h_i \|f(\xi_i, x(\xi_i)) - \Phi(\xi_i, x(\xi_i), 0)\|,$$

$\xi_i \in [t_i, t_{i+1}]$. Donc la méthode est consistante si et seulement si

$$\begin{aligned} \int_{t_0}^{t_f} \|f(t, x(t)) - \Phi(t, x(t), 0)\| dt &= 0 \\ \iff f(t, x(t)) &= \Phi(t, x(t), 0). \end{aligned}$$

■

Théorème 4.4.12 – Condition suffisante de stabilité

Pour que la méthode à un pas explicite soit stable il suffit que Φ soit Lipchitzienne en x : il existe Λ tel que pour tout $t \in [t_0, t_f]$, $h \geq 0$ et pour tout $(x_1, x_2) \in \mathbf{R}^2$ on ait

$$\|\Phi(t, x_1, h) - \Phi(t, x_2, h)\| \leq \Lambda \|x_1 - x_2\|.$$

On peut alors prendre comme constante de stabilité $S = e^{\Lambda(t_f - t_0)}$.

Lemme 4.4.1 (Lemme de Gronwall). Soient $(h_i)_i, (\theta_i)_i$ et $(\varepsilon_i)_i$ des suites telles que

$$\theta_{i+1} \leq (1 + \Lambda h_i) \theta_i + |\varepsilon_{i+1}|, \quad (4.14)$$

alors

$$\theta_i \leq e^{\Lambda(t_i - t_0)} \theta_0 + \sum_{k=0}^{i-1} e^{\Lambda(t_i - t_{k+1})} |\varepsilon_{k+1}|. \quad (4.15)$$

► Démontrons le résultat par récurrence. Pour $i = 0$, l'inégalité (4.15) devient $\theta_0 \leq \theta_0$. Pour $i = 1$, l'inégalité (4.14) s'écrit $\theta_1 \leq (1 + \Lambda h_0)\theta_0 + |\varepsilon_1|$. Il suffit alors de remarquer que $(1 + \Lambda h_0) \leq e^{\Lambda(t_1 - t_0)}$ pour conclure.

Supposons donc maintenant que l'assertion soit vraie pour $i - 1$ et montrons là pour i .

$$\begin{aligned} \theta_{i+1} &\leq (1 + \Lambda h_i)\theta_i + |\varepsilon_{i+1}| \\ &\leq e^{\Lambda(t_{i+1} - t_i)}(e^{\Lambda(t_i - t_0)}\theta_0 + \sum_{k=0}^{i-1} e^{\Lambda(t_i - t_{k+1})}|\varepsilon_{k+1}|) + |\varepsilon_{i+1}| \\ &\leq e^{\Lambda(t_{i+1} - t_0)}\theta_0 + \sum_{k=0}^i e^{\Lambda(t_{i+1} - t_{k+1})}|\varepsilon_{k+1}|. \end{aligned}$$

■

► du théorème 4.4.3. On considère les deux suites

$$\begin{aligned} x_{i+1} &= x_i + h_i \Phi(t_i, x_i, h_i) \\ \tilde{x}_{i+1} &= \tilde{x}_i + h_i \Phi(t_i, \tilde{x}_i, h_i) + \varepsilon_{i+1}. \end{aligned}$$

Comme Φ est Lipchitzienne par rapport à la variable x , nous avons

$$\|x_{i+1} - \tilde{x}_{i+1}\| \leq \|x_i - \tilde{x}_i\| + h_i \Lambda \|x_i - \tilde{x}_i\| + \|\varepsilon_{i+1}\|$$

Posons $\theta_i = \|x_i - \tilde{x}_i\|$, le lemme de Gronwall 4.4.1 implique alors que

$$\theta_i \leq e^{\Lambda(t_i - t_0)}\theta_0 + \sum_{k=0}^{i-1} \|\varepsilon_{k+1}\|.$$

Mais $t_i - t_0 \leq t_f - t_0$ et $t_i - t_{k+1} \leq t_f - t_0$, ce qui donne

$$\max_{0 \leq i \leq N} (\theta_i) \leq e^{\Lambda(t_f - t_0)}(\theta_0 + \sum_{k=0}^{N-1} \|\varepsilon_k + 1\|).$$

■

En conclusion lorsque le pas est constant, si nous notons $nfe = Ns$ le nombre total d'évaluations du deuxième membre $f(t, x)$ de l'équation différentielle, l'erreur globale en t_f pour une méthode à un pas explicite d'ordre p stable et consistante s'écrit

$$\begin{aligned} err &= Ch^p \\ &= C(t_f - t_0)^p s^p (nfe)^{-p} \\ &= C(p)(nfe)^{-p}. \end{aligned}$$

Soit, en passant au logarithme en base 10

$$\log_{10}(err) = \log_{10}(C(p)) - p \log_{10}(nfe).$$

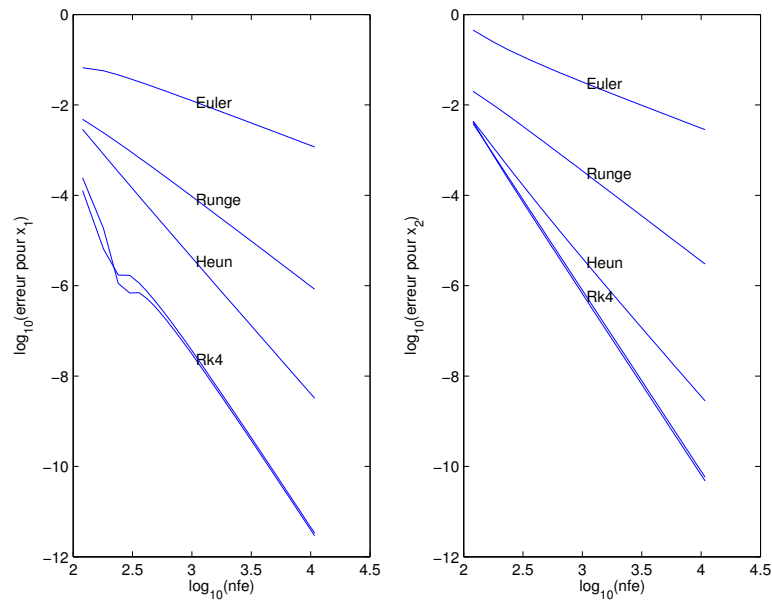


FIGURE 4.8 – Erreur globale, pour chaque composante, en fonction du nombre d'évaluations, pour l'équation de Van der Pol (E. Hairer, S.P. Nørsett and G. Wanner, Tome I, page 140).

Ceci est illustré, pour l'équation Van der Pol (cf. page 140 de [7])

$$(IVP) \begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = (1 - x_1^2(t))x_2(t) - x_1(t) \\ x_1(0) = 2.00861986087484313650940188 \\ x_2(0) = 0, \end{cases}$$

avec $t_f = T = 6.6632868593231301896996820305$, sur la Fig. 4.8 où la pente de la droite est $-p$.

4.5 Erreurs d'arrondi

On considère sur $[0, 1]$ l'équation différentielle ordinaire

$$(IVP) \begin{cases} \dot{x}(t) = x(t) \\ x(0) = 1. \end{cases}$$

Si on considère le schéma de Runge on a

1. Calcul 1

$$\begin{aligned} k_1 &= x_0 \\ k_2 &= x_0 + (h/2)x_0 \\ x_1 &= x_0 + h(x_0 + (h/2)x_0) \end{aligned}$$

2. Calcul 2 $x_1 = (1 + h + h^2/2)x_0$

Donc $e_i \sim h^3 x_i/6$ d'où $e_i \leq Ch^3$ avec $C = e/6$ sur $[0, 1]$. Par suite

$$h_{opt} \geq \left(\frac{\text{macheps}}{2e/6} \right)^{1/3} = 6.26 \cdot 10^{-6}$$

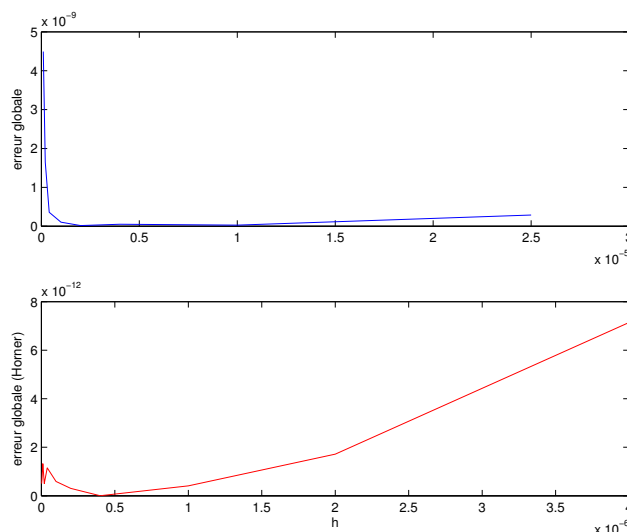


FIGURE 4.9 – Erreurs globales en fonction du pas h pour le schéma de Runge ; en haut $x_1 = (1 + h + h^2/2)x_0$ et en bas $x_1 = x_0 + h(x_0 + (h/2)x_0)$, $h_{opt} = 6.26 \cdot 10^{-6}$, [4] page 218.

4.6 Contrôle du pas

4.6.1 Introduction

La question qui se pose maintenant est le choix du pas ; sachant qu'un calcul à pas constant est en général inefficace (cf. Fig 4.4). L'idée est de choisir les pas afin que l'erreur locale soit partout environ égale à une tolérance Tol fournie par l'utilisateur. Il faut pour cela avoir une estimation de cette erreur locale.

4.6.2 Extrapolation de Richardson

Pour simplifier la présentation, nous supposons ici que la dimension de l'équation différentielle est $n = 1$.

L'idée de Richardson [11] est de calculer 2 pas d'une longueur h d'une méthode de Runge-Kutta d'ordre p . En partant de x_0 , on obtient ainsi les valeurs de x_1 et de x_2 . On calcule ensuite un grand pas de longueur $2h$ et on obtient une nouvelle quantité \bar{x}_2 . L'erreur locale en $t_1 = t_0 + h$ est $e_1 = x(t_0 + h) - x_1 = Ch^{p+1} + O(h^{p+2})$. Quant-à l'erreur en $t_2 = t_0 + 2h$ pour la première

méthode elle est composée de deux parties

$$\begin{aligned}
x(t_2, t_0, x_0) - x_2 &= (x(t_2, t_0, x_0) - x(t_2, t_1, x_1)) + (x(t_2, t_1, x_1) - x_2) \\
&= (x(t_2, t_1, x(t_1, t_0, x_0)) - x(t_2, t_1, x_1)) + e_2 \\
&= \frac{\partial x}{\partial x_1}(t_2, t_1, x_1)e_1 + e_1\varepsilon(e_1) + e_2 \\
&= (I + \int_{t_1}^{t_2} \frac{\partial f}{\partial x}(s, t_1, x_1) \frac{\partial x}{\partial x_1}(s, t_1, x_1) ds)e_1 + e_1\varepsilon(e_1) + e_2 \\
&= (I + h \frac{\partial \phi}{\partial x}(t_1, x_1) + O(h^2))e_1 + e_2 \\
&= (I + O(h))Ch^{p+1} + Ch^{p+1} + O(h^{p+2}) \\
&= 2Ch^{p+1} + O(h^{p+2}).
\end{aligned}$$

Pour la deuxième méthode on a

$$\bar{e}_2 = x(t_0 + 2h, t_0, x_0) - \bar{x}_2 = C((2h)^{p+1}) + O(h^{p+2}).$$

Nous en déduisons alors l'estimation de l'erreur

$$\begin{aligned}
x(t_0 + 2h, t_0, x_0) - x_2 &= \frac{1}{2^p}(x(t_0 + 2h, t_0, x_0) - \bar{x}_2) + O(h^{p+1}) \\
&= \frac{x_2 - \bar{x}_2}{2^p - 1} + O(h^{p+2}).
\end{aligned}$$

Remarque 4.6.1. Si on pose $\hat{x}_2 = x_2 + \frac{x_2 - \bar{x}_2}{2^p - 1}$, l'erreur local est en $O(h^{p+2})$. Nous pouvons donc ainsi construire à partir d'une méthode d'ordre p , une méthode d'ordre $p + 1$. Cette méthode est implémenté en particulier dans le code ODEX des professeurs Hairer et Wanner[7].

4.6.3 Méthode de Runge-Kutta emboîtées

On considère deux schémas de Runge-Kutta à s étages d'ordre p et $\hat{p} = p - 1$ emboîtés, c'est-à-dire avec les mêmes coefficient a_{ij} . Nous pourrions alors par différence estimer l'erreur locale

$$\begin{aligned}
e_1 \sim x_1 - \hat{x}_1 &= (x_1 - x(t_1, t_0, x_0)) + (x(t_1, t_0, x_0) - \hat{x}_1) = O(h^{p+1}) + O(h^{\hat{p}+1}) = O(h^{\hat{p}+1}) \\
&= Ch^{\hat{p}+1}.
\end{aligned}$$

Exemple 4.6.1 (Méthode de Runge-Kutta emboîtées RK2(1)).

0		
1/2	1/2	
	0	1
	1	0

$$\begin{aligned}
k_1 &= f(t_0, x_0) \\
k_2 &= f(t_0 + h/2, x_0 + h/2k_1) \\
x_1 &= x_0 + hk_2 \quad \text{Runge} \\
\hat{x}_1 &= x_0 + hk_1 \quad \text{Euler}
\end{aligned}$$

estimation de l'erreur locale : $\|x_1 - \hat{x}_1\| = h\|k_2 - k_1\|$

□

Exemple 4.6.2 (Méthode de Runge-Kutta emboîtées RK4(3)).

0					
1/3	1/3				
2/3	-1/3	1			
1	1	-1	1		
1	1/8	3/8	3/8	1/8	
	1/12	1/2	1/4	0	1/6

$$k_1 = f(t_0, x_0)$$

$$k_2 = f(t_0 + h/3, x_0 + (h/3)k_1)$$

$$k_3 = f(t_0 + 2h/3, x_0 + h(-k_1/3 + k_2))$$

$$k_4 = f(t_0 + h, x_0 + h(k_1 - k_2 + k_3))$$

$$x_1 = x_0 + (h/8)(k_1 + 3k_2 + 3k_3 + k_4) O(h^5)$$

$$\hat{x}_1 = x_0 + (h/12)(k_1 + 6k_2 + 3k_3 + 2f(t_0 + h, x_1)) O(h^4)$$

□

L'idée est alors d'accepter le pas si $\|x_1 - \hat{x}_1\| < Tol$. La norme choisie est en générale

$$\|x_1 - \hat{x}_1\| = \sqrt{\frac{1}{n} \sum_i \left(\frac{x_{i1} - \hat{x}_{i1}}{1 + \max(|x_{i0}|, |x_{i1}|)} \right)^2}.$$

On peut aussi écrire ceci $err < 1$, où

$$err = \sqrt{\frac{1}{n} \sum_i \left(\frac{x_{i1} - \hat{x}_{i1}}{sci} \right)^2}$$

et $sci = Tol + Tol \max(|x_{i0}|, |x_{i1}|)$. En pratique on prend $sci = Atol_i + Rtol_i \max(|x_{i0}|, |x_{i1}|)$.

- Remarque 4.6.2.**
- Si $Rtol_i = 0$, on a pour cette composante une erreur locale absolue.
 - Si $Atol_i = 0$ on a pour cette composante une erreur locale relative.
 - Dans ODE45 $Rtol$ est un scalaire et $Atol$ est un vecteur.
 - Dans DOPRI5 $Rtol$ et $Atol$ sont des vecteurs.

Il nous reste maintenant à déterminer la mise à jour du pas. Comme $\|x_1 - \hat{x}_1\| \approx Ch^{\hat{p}+1}$, et que l'on veut $Ch_{opt}^{\hat{p}+1} = Tol$, on peut estimer que $C = \frac{Tol}{h_{opt}^{\hat{p}+1}}$. On en déduit que $h_{opt} = 0.9h \left(\frac{Tol}{\|x_1 - \hat{x}_1\|} \right)^{1/(\hat{p}+1)} = 0.9h \left(\frac{1}{err} \right)^{1/(\hat{p}+1)}$. Le coefficient 0.9 est un coefficient de sécurité. Afin d'éviter de plus des variations brusques du pas on imposera en pratique que $0.2 \leq h_{opt} \leq 5h$. En conclusion on obtient l'algorithme suivant.

Algorithme 4.1. [Contrôle du pas]

Calculer les $sc_i = Atol_i + Rtol_i \max(|x_{i0}|, |x_{i1}|)$

Calculer

$$err = \|x_1 - \hat{x}_1\| = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_{i1} - \hat{x}_{i1}}{sc_i} \right)^2}$$

```

si  $err < 1$  alors
    Le pas est accepté
fin si
Calculer le nouveau pas :  $h = h \min(5, \max(0.2, 0.9^{\hat{p}+1} \sqrt{1/err}))$ 
si  $t + h > t_f$  alors
     $h = t_f - t$ 
fin si

```

Quand au pas initiale, nous donnons ci-après l'algorithme implémenté dans DOPRI5.

Algorithme 4.2. [Pas initial DOPRI5] Ici $sc_i = Atol + Rtol|x_i|$ et $\|x\| = \sqrt{\sum_{i=1}^n (\frac{x_i}{sc_i})^2}$.

Calcul d'un premier pas d'Euler

```

 $k_0 = f(t_0, x_0)$ ,  $d_0 = \|x_0\|$ ,  $d_1 = \|k_0\|$ 
 $h_0 = 0.01 \frac{d_0}{d_1}$  {le pas est petit par rapport à la solution}
si  $d_0 < 10^{-5}$  ou  $d_1 < 10^{-5}$  alors
     $h_0 = 10^{-6}$ 

```

fin si

$h_0 = \min(h_{max}, h_0)$

$x_1 = x_0 + h_0 k_0$ {pas d'Euler}

Estimation de la dérivée seconde

```

 $k_1 = f(t_0 + h_0, x_1)$ 
 $d_2 = \|k_1 - k_0\|/h_0$ 
si  $\max(d_1, d_2) < 10^{-15}$  alors
     $h_1 = \max(10^{-6}, 10^{-3}h_0)$ 

```

sinon

$h_1 = (0.01 / \max(d_1, d_2))^{1/p}$ { h_1 tel que $h_1^{\hat{p}+1} \max(d_1, d_2) = 0.01$ }

fin si

$h = \min(100h_0, h_1, h_{max})$

Illustrons maintenant cet algorithme sur l'équation du Brusselator

$$(IVP) \begin{cases} \dot{x}_1(t) = 1 + x_1^2(t)x_2(t) - 4x_1(t) \\ \dot{x}_2(t) = 3x_1(t) - x_1^2(t)x_2(t) \\ x_1(0) = 1.5 \\ x_2(0) = 3, \end{cases}$$

La Fig. 4.10 donne pour $t_f = 20$ la solution calculée, les pas acceptés et rejetés et les erreurs commises.

Exemple 4.6.3 (Méthode de Runge-Kutta emboîtées RK3(2)). Paires de Bogacki-Shampine utilisés dans ODE23 de Matlab.

0				
1/2	1/2			
3/4	0	3/4		
	2/9	1/3	4/9	
	7/24	1/4	1/3	1/8

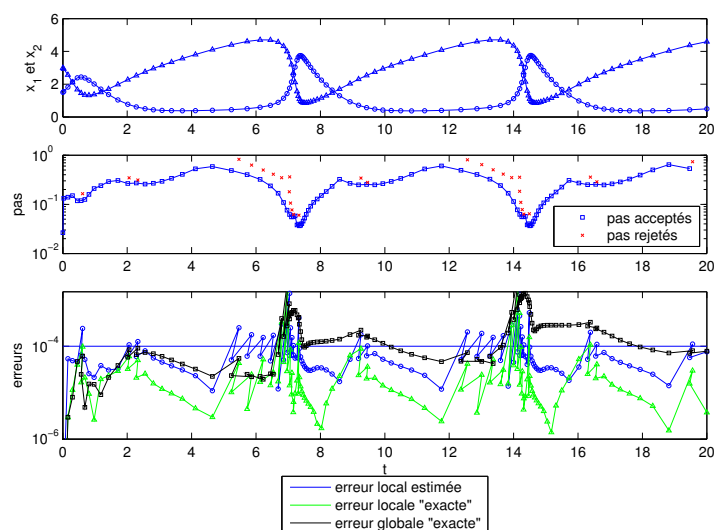


FIGURE 4.10 – Solutions, pas acceptés et rejetés, erreurs locales estimées, "exactes" et erreurs globales "exactes". L'intégration est effectuée avec rk34 (règle 3/8 pour l'ordre 4) et les solutions "exactes" sont calculées avec ode45, E. Hairer, S.P. Nørsett and G. Wanner, Tome I, page 170.

$$k_1 = f(t_0, x_0)$$

$$k_2 = f(t_0 + h/2, x_0 + h/2k_1)$$

$$k_3 = f(t_0 + 3/4h, x_0 + 3/4hk_2)$$

$$x_1 = x_0 + h(2/9k_1 + 3/9k_2 + 4/9k_3) \quad O(h^4)$$

$$\hat{x}_1 = x_0 + h(7/24k_1 + 1/4k_2 + 1/3k_3 + 1/8f(t_1, x_1)) \quad O(h^3)$$

estimation de l'erreur locale : $\|x_1 - \hat{x}_1\| = (h/72)\| -5k_1 + 6k_2 + 8k_3 - 9f(t_1, x_1) \|$ □

Exemple 4.6.4 (Méthode de Runge-Kutta emboîtées RK5(4)). Paires de Dormand Prince utilisé dans ODE45 et DOPRI5.

0						
$\frac{1}{5}$	$\frac{1}{5}$					
$\frac{3}{10}$	$\frac{3}{40}$	$\frac{9}{40}$				
$\frac{4}{5}$	$\frac{44}{45}$	$-\frac{56}{15}$	$\frac{32}{9}$			
$\frac{8}{9}$	$\frac{19372}{6561}$	$-\frac{25360}{2187}$	$\frac{64448}{6561}$	$-\frac{212}{729}$		
1	$\frac{9017}{3168}$	$-\frac{355}{33}$	$\frac{46732}{5247}$	$\frac{49}{176}$	$-\frac{5103}{18656}$	
1	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$
	$\frac{5179}{57600}$	0	$\frac{7571}{16695}$	$\frac{393}{640}$	$-\frac{92097}{339200}$	$\frac{187}{2100} - \frac{1}{40}$

□

4.7 Les méthodes de Runge-Kutta implicites

Définition 4.7.1 – Méthode de Runge-Kutta implicite

On appelle méthode de Runge-Kutta implicite à s étages, la méthode définie par le schéma

$$\begin{cases} k_i = f(t_0 + c_i h, x_0 + h \sum_{j=1}^s a_{ij} k_j) & \text{pour } i = 1, \dots, s \\ x_1 = x_0 + h \sum_{i=1}^s b_i k_i \end{cases} \quad (4.16)$$

où les coefficients c_i, a_{ij} et b_i sont des constantes qui définissent précisément le schéma.

On supposera toujours dans la suite que $c_i = \sum_{j=1}^s a_{ij}$ pour $i = 1, \dots, s$.

Remarque 4.7.1. La méthode est implicite car sur chaque pas, il faut résoudre un système d'équations non linéaires à ns équations et ns inconnues.

$$(S) \begin{cases} k_i = f(t_0 + c_i h, x_0 + h \sum_{j=1}^s a_{ij} k_j) & \text{pour } i = 1, \dots, s \end{cases}$$

On représente en pratique ce schéma par le tableau de Butcher, cf. la table 4.3.

c_1	a_{11}	\dots	a_{1s}
\vdots	\vdots	\vdots	\ddots
c_s	a_{s1}	\dots	a_{ss}
	b_1	\dots	b_s

TABLE 4.3 – *Tableau de Butcher.*

Exemple 4.7.1 (Euler implicite).

$$x_1 = x_0 + hf(t_1, x_1)$$

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

□

Exemple 4.7.2 (Trapèze).

$$x_1 = x_0 + \frac{h}{2}(f(t_0, x_0) + f(t_1, x_1))$$

$$\begin{array}{c|cc} 0 & & \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$$

□

Exemple 4.7.3 (Point milieu).

$$x_1 = x_0 + h(f(t_0 + h/2, (x_0 + x_1)/2))$$

$$\begin{array}{c|c} 1/2 & 1/2 \\ \hline & 1 \end{array}$$

□

Exemple 4.7.4 (Gauss implicite à deux étage d'ordre 4).

$$\begin{array}{c|cc} 1/2 - \sqrt{3}/6 & 1/4 & 1/4 - \sqrt{3}/6 \\ 1/2 + \sqrt{3}/6 & 1/4 + \sqrt{3}/6 & 1/4 \\ \hline & 1/2 & 1/2 \end{array}$$

□

Théorème 4.7.2

On suppose $f : \mathbf{R} \times \mathbf{R}^n \rightarrow \mathbf{R}^n$ continue et Lipchitzienne par rapport à la deuxième variable x de constante L . Si

$$h < \frac{1}{L \max_i (\sum_j |a_{ij}|)}$$

alors il existe une unique solution à (S) (cf. (4.16)). Si $f(t, x)$ est C^p , alors les k_i sont aussi C^p .



Exercice 4.7.5. Démonstration du théorème (4.7)

1. Démontrer l'existence et l'unicité de la solution de (S) (utiliser le théorème du point fixe).
2. Démontrer que les k_i sont C^p si f est C^p (utiliser le théorème des fonctions implicites).

□

Nous allons maintenant voir le lien entre les méthodes de Runge-Kutta implicites et les méthodes de collocation. Ces dernières méthodes consistent à chercher sur chaque intervalle d'une discrétisation en temps un polynôme u de degré s dont les dérivées coïncident² en s points donnés à $f(t_0 + c_i h, u(t_0 + c_i h))$.

Définition 4.7.3 – Méthode de collocation

Soit $s \geq 1$ et c_1, \dots, c_s , s points distincts de $[0, 1]$, on définit l'unique polynôme $u(t)$ de degré s vérifiant

$$\begin{aligned} u(t_0) &= x_0 \\ u'(t_0 + c_i h) &= f(t_0 + c_i h, u(t_0 + c_i h)). \end{aligned}$$

La méthode de collocation consiste alors à calculer x_1 par

$$x_1 = u(t_0 + h).$$

Théorème 4.7.4

[[5, 16]] La méthode de collocation est équivalente à une méthode de Runge-Kutta implicite à s étages avec

$$a_{ij} = \int_0^{c_i} l_j(t) dt \quad b_j = \int_0^1 l_j(t) dt \quad i, j = 1, \dots, s,$$

2. co-locate en anglais.

où $l_j(t)$ est le polynôme de Lagrange.


$$l_j(t) = \prod_{k \neq j} \frac{t - c_k}{c_j - c_k}.$$

Remarque 4.7.2. Pour résoudre sur chaque pas le système d'équations non linéaires, on peut soit utiliser un point fixe (cf. la démonstration du théorème (4.7)), soit utiliser un algorithme de type Newton. Nous allons donner ici le cas le plus simple d'un algorithme du point fixe, sachant qu'en pratique, il est préférable d'utiliser un algorithme du type Newton (cf. IV.8 page 118 de [8]).


Algorithme 4.3. Initialisation

$N =$ nombre de pas
 $fp\epsilon ps =$ epsilon pour le point fixe
 $fpitermax =$ nombre maximum d'itérations pour le point fixe
 $h = (t_f - t_0)/N$
pour $l = 1$ à N **faire**
 Initialiser les $k_i = f(t_0 + c_i h, x_0)$
 $normprog = 1$; $nbiter = 0$
 tant que $(normprog > fp\epsilon ps)$ et $(nbiter < fpitermax)$ **faire**
 Calculer les $newk_i = f(t_0 + c_i h, x_0 + h \sum_{j=1}^s a_{ij} k_j)$ pour $i = 1, \dots, s$
 Calculer $normprog = ||(newk_1 - k_1; \dots; newk_s - k_s)||$
 $k_i := newk_i$ pour $i = 1, \dots, s$
 $nbiter = nbiter + 1$
 fin tant que
 $t_0 = t_0 + h$
 $x_0 = x_0 + h \sum_{j=1}^s b_j k_j$
fin pour

4.8 Exercices

 **Exercice 4.8.1.** 1. Appliquez un pas d'une méthode de Runge-Kutta explicite à s étages à l'équation $\dot{x}(t) = x(t)$, $x(0) = 1$ et montrez que x_1 est un polynôme en h de degré s .

2. En déduire que l'ordre d'une méthode de Runge-Kutta explicite ne peut être plus grand que le nombre d'étages : $p \leq s$. □

 **Exercice 4.8.2.** On rappelle qu'un système non autonome peut toujours s'écrire sous forme autonome. En effet si on considère le problème à valeur initiale

$$(IVP) \begin{cases} \dot{x}(t) = f(t, x(t)) \\ x(t_0) = x_0, \end{cases}$$

et si on pose $t = t_0 + \tau$, $\tau \in [0, t_f - t_0]$, on a

$$(IVP) \iff \begin{cases} \frac{dt}{d\tau}(\tau) = 1 \\ \frac{dx}{d\tau}(\tau) = f(t(\tau), x(\tau)) \\ \tau(0) = t_0 \\ x(0) = x_0, \end{cases} \iff \begin{cases} \frac{dz}{d\tau}(\tau) = \tilde{f}(z(\tau)) \\ z(0) = \begin{pmatrix} t_0 \\ x_0 \end{pmatrix}, \end{cases}$$

avec

$$z(\tau) = \begin{pmatrix} t(\tau) \\ x(\tau) \end{pmatrix} \quad \text{et} \quad \tilde{f}(z(\tau)) = \begin{pmatrix} 1 \\ f(t(\tau), x(\tau)) \end{pmatrix}.$$


1. Écrire ce que devient le schéma de Runge-Kutta explicite à s étages dans le cas autonome, c'est-à-dire avec la variable $z : \tilde{k}_1 = \dots$

2. On pose $g_1 = z_0$ et $g_i = z_0 + h \sum_{j=1}^{i-1} a_{ij} \tilde{f}(g_j)$ pour $i = 1, \dots, s$. Écrire ce que donne ce schéma de Runge-Kutta explicite avec ces notations.

3. En considérant la première composante des g_i et de x_1 , montrer que l'on doit avoir

$$\sum_{i=1}^s b_i = 1 \quad \text{et} \quad \sum_{j=1}^{i-1} a_{ij} = c_i, \forall i = 2, \dots, s$$

□

 **Exercice 4.8.3.** ³ On considère l'équation différentielle de Cauchy du deuxième ordre et autonome (on considère pour simplifier ici que $x(t) \in \mathbf{R}$)

$$(IVP) \begin{cases} \ddot{x}(t) = f(x(t)) \\ x(t_0) = x_0 \\ \dot{x}(t_0) = \dot{x}_0. \end{cases}$$

1. On pose $z(t) = \dot{x}(t)$ et $u(t) = \begin{pmatrix} x(t) & z(t) \end{pmatrix}^T$. Écrire le système (IVP) sous la forme d'un système à condition initiale du premier ordre.

2. On considère maintenant le schéma de Nyström suivant

$$\begin{cases} k_1 = f(x_0) \\ k_2 = f(x_0 + h\alpha_1 z_0 + (h^2/2)\alpha_2 k_1) \\ \Phi_1(u_0, h) = \Phi_1(x_0, z_0, h) = z_0 + (h/2)(\gamma_1 k_1 + \gamma_2 k_2) \\ \Phi_2(u_0, h) = \Phi_2(x_0, z_0, h) = (1/2)(\delta_1 k_1 + \delta_2 k_2) \\ u_1 = \begin{pmatrix} x_1 \\ z_1 \end{pmatrix} = u_0 + h\Phi(u_0, h) = \begin{pmatrix} x_0 \\ z_0 \end{pmatrix} + h \begin{pmatrix} \Phi_1(x_0, z_0, h) \\ \Phi_2(x_0, z_0, h) \end{pmatrix} \end{cases}$$

Montrer que la méthode est consistante si et seulement si $\delta_1 + \delta_2 = 2$.

3. Donner le développement limité à l'ordre 2 de $k_2 : k_2 = f(x_0) + \dots$.

4. Donner les relations que doivent vérifier les coefficients pour que la méthode soit d'ordre 3.

□

3. sujet 36 page 83, Calcul différentiel et équations différentielles, Azé, Constans, Hiriart-Urruty, Dunod, 2002.

Sortie dense, discontinuités, dérivées

5.1 Sortie dense

5.1.1 Objectif

Les intégrateurs numériques ne donnent les résultats qu'aux points de la subdivision $t_0 < t_1 < \dots < t_N$ (celle-ci étant déterminée *a priori* pour les méthodes à pas fixes et *a posteriori* pour les méthodes à pas variables), or l'utilisateur peut vouloir calculer la solution en tout point de l'intervalle $[t_0, t_f]$ ou sur une grille très fine. L'objectif de ce qui est appelé dans la littérature la sortie dense¹ est de répondre à ce besoin. La sortie dense est en outre nécessaire :

- pour faire le tracé de la solution (cf. la trajectoire tracée dans le cas de l'utilisation d'ODE45 sur la figure 4.4) ;
- pour détecter un événement, on intègre le système différentiel jusqu'à l'instant T où $\psi(x(T)) = 0$ (cf. la sous-section 5.1.3 ci-après) ;
- pour intégrer les équations différentielles à second membre discontinue (cf. la sous section 5.1.4 ci-après).

On désire donc sur un pas $[t_0, t_1]$ avoir une approximation peu coûteuse de la solution $x(t) = x(t_0 + \theta h)$, $\theta \in [0, 1]$, c'est-à-dire qui ne nécessite pas d'évaluation supplémentaire du second membre de l'équation différentielle, et de même ordre que la solution aux points de la subdivision t_0, \dots, t_N . Cette approximation sera en pratique un polynôme $u(\theta) \approx x(t_0 + \theta h)$ de degré $p^* = p - 1$. Si on choisit correctement ce polynôme, nous pouvons avoir une erreur d'approximation en $O(\theta^p)$. En ajoutant cette erreur à l'erreur globale qui est en $O(h^p)$ pour un schéma d'ordre p , nous obtiendrons bien ce que nous voulons.

5.1.2 Calcul de la sortie dense

Une solution simple pour un schéma de Runge-Kutta d'ordre 4 est de considérer l'interpolation de l'Hermite, c'est-à-dire de déterminer l'unique polynôme de degré 3 tel que

$$\begin{aligned} u(0) &= x_0 & \dot{u}(0) &= f(t_0, x_0)h \\ u(1) &= x_1 & \dot{u}(1) &= f(t_1, x_1)h. \end{aligned}$$

On obtient ainsi le polynôme suivant

$$u(\theta) = (1 - \theta)x_0 + \theta x_1 + \theta(\theta - 1)((1 - 2\theta)(x_1 - x_0) + (\theta - 1)hf(t_0, x_0) + \theta hf(t_1, x_1)).$$

Remarque 5.1.1. • La solution approximée ainsi est C^1 ;

- pour calculer le polynôme, il faut calculer $f(t_1, x_1)$, mais comme cette valeur est calculée au pas suivant, il ne s'agit pas d'une évaluation supplémentaire de f .

1. dense output en anglais.

Dans le cas général, on suppose que l'on part d'une méthode de Runge-Kutta d'ordre p à s étages. On rajoute $s^* - s$ étage et on cherche le polynôme $u(\theta)$ sous la forme

$$u(\theta) = x_0 + \sum_{i=1}^{s^*} b_i(\theta) k_i.$$

En générale on prend $s^* \geq s + 1$ avec $k_{s+1} = f(t_1, x_1)$. On cherche alors les polynômes $b_i(\theta)$ tels que

$$u(\theta) - x(t_0 + \theta h) = O(h^{p^*+1}).$$

Exemple 5.1.1. Pour la méthode RK4, $s = s^* = 4$, on peut prendre

$$\begin{aligned} b_1(\theta) &= \theta - \frac{3\theta^2}{2} + \frac{2\theta^3}{3} \\ b_2(\theta) &= b_3(\theta) = \theta^2 - \frac{2\theta^3}{3} \\ b_4(\theta) &= -\frac{\theta^2}{2} + \frac{2\theta^3}{3}. \end{aligned}$$

$u(\theta)$ est alors continue sur $[t_0, t_f]$, mais non C^1 . □

Exemple 5.1.2. Pour la paire de Dorman-Prince (DOPRI5 et ODE45) on trouvera les valeurs des $b_i(\theta)$ page 192 de [7]. □

Remarque 5.1.2. La sortie dense permet aussi d'obtenir une approximation polynomiale des dérivées $u^{(k)}(\theta)$ et donc des $x^{(k)}(t_0 + \theta h)$:

$$h^{-k} u^{(k)}(\theta) - x^{(k)}(t_0 + \theta h) = O(h^{p^*+1-k}).$$

5.1.3 Détection d'évènements

L'objectif est de déterminer l'instant t tel que $\psi(t, x(t)) = 0 \in \mathbf{R}$.

Algorithme 5.1. *si* $\psi(t_i, x_i)\psi(t_{i+1}, x_{i+1}) \leq 0$ *alors*

Déterminer $\tau = t_i + \theta h_i$ *et* x_τ , $\theta \in [0, 1]$ *tel que* $g(\theta) = \psi(t_i + \theta h, u(\theta)) = 0$ *{* $u(\theta)$ *est la sortie dense sur* $[t_i, t_{i+1}]$.*}*

$t_{i+1} = t_i + \theta h_i$

$x_{i+1} = u(t_i + \theta h_i)$

fin si

Exemple 5.1.3 (Balle qui rebondie). `>> help ballode`

BALLODE Run a demo of a bouncing ball.

This is an example of repeated event location, where the initial conditions are changed after each terminal event. This demo computes ten bounces with calls to ODE23. The speed of the ball is attenuated by 0.9 after each bounce. The trajectory is plotted using the output function ODEPLOT.

See also ODE23, ODE45, ODESET, ODEPLOT, FUNCTION_HANDLE.

`>> orbitode`

□

Exemple 5.1.4 (Problème aux trois corps restreints). `>> help orbitode`

ORBITODE Restricted three body problem.

This is a standard test problem for non-stiff solvers stated in Shampine and Gordon, p. 246 ff. The first two solution components are coordinates of the body of infinitesimal mass, so plotting one against the other gives the orbit of the body around the other two bodies. The initial conditions have been chosen so as to make the orbit periodic. Moderately stringent tolerances are necessary to reproduce the qualitative behavior of the orbit. Suitable values are $1e-5$ for RelTol and $1e-4$ for AbsTol.

ORBITODE runs a demo of event location where the ability to specify the direction of the zero crossing is critical. Both the point of return to the initial point and the point of maximum distance have the same event function value, and the direction of the crossing is used to distinguish them.

The orbit of the third body is plotted using the output function ODEPHAS2.

L. F. Shampine and M. K. Gordon, Computer Solution of Ordinary Differential Equations, W.H. Freeman & Co., 1975.

See also ODE45, ODE23, ODE113, ODESET, ODEPHAS2, FUNCTION_HANDLE.

`>> orbitode`

This is an example of event location where the ability to specify the direction of the zero crossing is critical. Both the point of return to the initial point and the point of maximum distance have the same event function value, and the direction of the crossing is used to distinguish them.

Calling ODE45 with event functions active...

Note that the step sizes used by the integrator are NOT determined by the location of the events, and the events are still located accurately.

□

5.1.4 Intégration d'équations différentielles à second membre discontinues

$$(IVP) \begin{cases} \dot{x}(t) = t^2 + 2x^2(t) & \text{si } (t + 0.05)^2 + (x(t) + 0.15)^2 \leq 1 \\ \dot{x}(t) = 2t^2 + 3x^2(t) - 2 & \text{si } (t + 0.05)^2 + (x(t) + 0.15)^2 > 1 \end{cases}$$

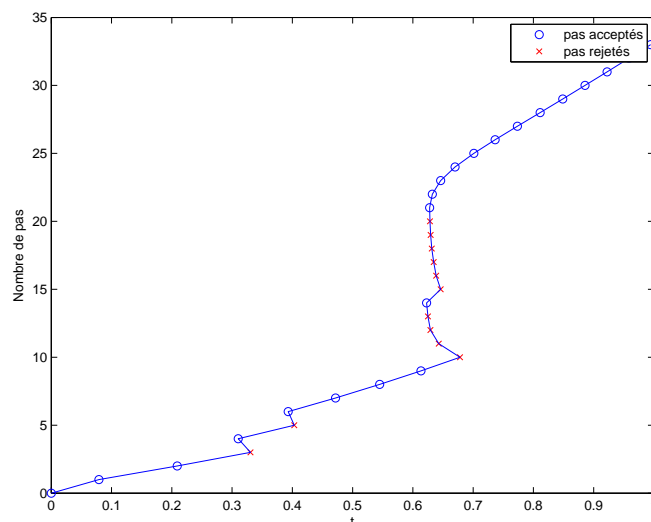


FIGURE 5.1 – Utilisation du pas variable, programme `myode43.m`, E. Hairer, S.P. Nørsett and G. Wanner, Tome I, page 197.

5.2 Calcul de la dérivée

5.2.1 Exemple

On considère le problème à valeur initiale suivant (Brusselator)[7], page 201.

$$(IVP) \begin{cases} \dot{x}_1 = 1 + x_1^2 x_2 - (\lambda + 1)x_1 \\ \dot{x}_2 = \lambda x_1 - x_1^2 x_2 \\ x_1(0) = 1.3 \\ x_2(0) = \lambda \end{cases}$$

où $t_f = 20$.

5.2.2 Différences finies externes

$$\frac{\partial x}{\partial x_{0j}}(t_f, x_0) \approx \frac{1}{\delta}(x(t_f, x_0 + \delta e_j) - x(t_f, x_0)),$$

où e_1, \dots, e_n désigne la base canonique de \mathbf{R}^n .

5.2.3 Équation variationnelle

$\frac{\partial x}{\partial x_{0j}}(t_f, x_0)$ est la solution $X_j(t_f)$ du système de Cauchy

$$(EQ)_j \begin{cases} \dot{x}(t) = f(t, x(t)) \\ \dot{X}_j(t) = \frac{\partial f}{\partial x}(t, x(t)) X_j(t) \\ x(0) = x_0 \\ X_j(0) = e_j \end{cases}$$

5.2.4 Différentiation interne de Bock (IND)

Nous allons approximer par différences finies le deuxième membre de l'équation variationnelle [2]. On a

$$f(t, x(t) + \delta X_j(t)) = f(t, x(t)) + \frac{\partial f}{\partial x}(t, x(t))\delta X_j(t) + o(\delta X_j(t))$$

On approxime alors dans les équations variationnelles

$$\frac{\partial f}{\partial x}(t, x(t))X_j(t) \approx \frac{1}{\delta}(f(t, x(t) + \delta X_j(t)) - f(t, x(t))).$$

On résout donc

$$(IND)_j \begin{cases} \dot{x}(t) = f(t, x(t)) \\ \dot{X}_j(t) = \frac{1}{\delta}(f(t, x(t) + \delta X_j(t)) - f(t, x(t))) \\ x(0) = x_0 \\ X_j(0) = e_j. \end{cases}$$

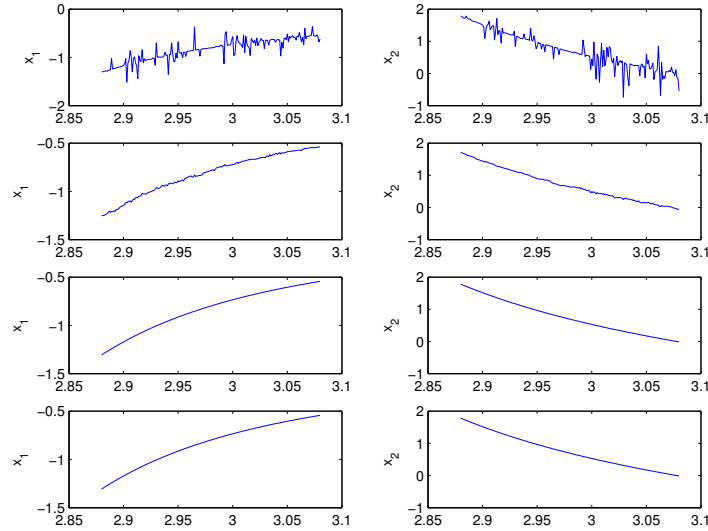
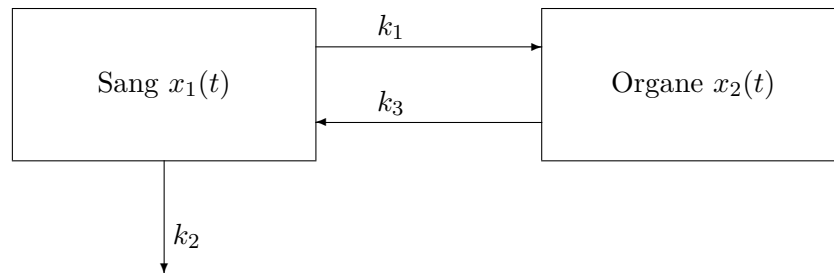


FIGURE 5.2 – Dérivée de la solution par rapport à λ $\frac{\partial x_1(t_f, \lambda)}{\partial \lambda}$ et $\frac{\partial x_2(t_f, \lambda)}{\partial \lambda}$ pour (de haut en bas) : les différences finies avant ($\delta\lambda = 4Tol$), les différences finies avant ($\delta\lambda = \sqrt{Tol}$), la différentiation interne de Bock ($\delta\lambda = \sqrt{macheps}$) et l'équation variationnelle. On utilise le code ODE45 de MATLAB avec $Atol = Rtol = Tol = 10^{-4}$.

5.2.5 Exemple

5.2.5.1 Modèle à compartiments



5.2.5.2 Données

Les concentrations dans le sang sont mesurées à différents instants :

t_i	x_{i1}	t_i	x_{i1}
0.25	215.6	3.00	101.2
0.50	189.2	4.00	88.0
0.75	176.0	6.00	61.6
1.00	162.8	12.00	22.0
1.50	138.6	24.00	4.4
2.00	121.0	48.00	0.0

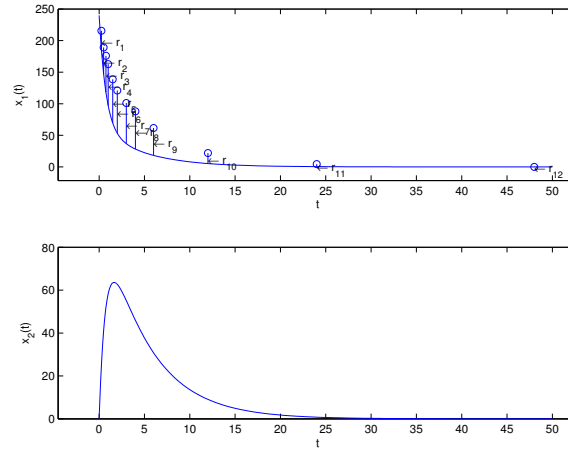
5.2.5.3 Modèle Mathématique

Le système d'équations différentielles décrivant le modèle est le suivant :

$$(EDO) \begin{cases} \dot{x}(t) = \begin{pmatrix} -k_1 - k_2 & k_3 \\ k_1 & -k_3 \end{pmatrix} x(t) = A_k x(t) \\ x_1(0) = c_0 \\ x_2(0) = 0 \end{cases}$$

5.2.5.4 Problème aux moindres carrées

$$(P) \begin{cases} \text{Min} & f(\beta) = \frac{1}{2} \sum_{i=1}^n (x_{i1} - x_1(t_i; \beta))^2 = \frac{1}{2} \|r(\beta)\|^2 \\ \beta = {}^t(c_0, k_1, k_2, k_3) \in \mathbf{R}^4 \end{cases}$$

FIGURE 5.3 – Données, courbe pour $\beta = (0.5, 0.55, 0.5, 240)$ et résidus

5.2.5.5 Calcul de la dérivée de la fonction résidus

L'utilisation d'un algorithme à pas variables pour calculer les résidus impose d'utiliser les équations variationnelles pour calculer les dérivées de x_1 par rapport aux paramètres.

$$(VAR) \begin{cases} \dot{x}(t) = A_k x(t) \\ x_1(0) = c_0 \\ x_2(0) = 0 \\ \dot{X} = A_k X(t) \\ X(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \dot{Z}(t) = A_k Z(t) + B(t) \\ Z(0) = 0_{(3,3)} \end{cases}$$

avec

$$B(t) = \begin{pmatrix} -x_1(t) & -x_1(t) & x_2(t) \\ x_1(t) & 0 & x_2(t) \end{pmatrix}$$

5.2.5.6 Solution

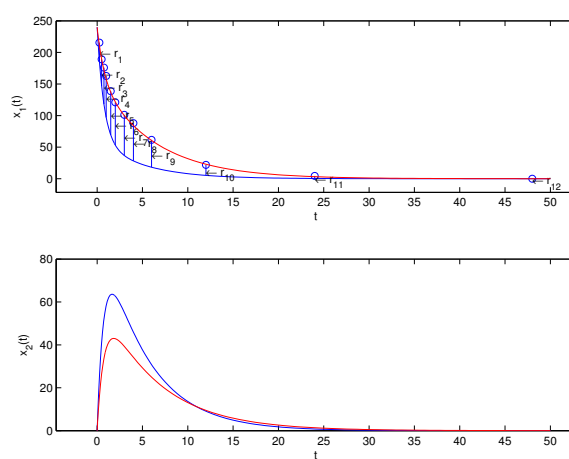


FIGURE 5.4 – Données, courbe pour $\beta = (0.5, 0.55, 0.5, 240)$ et résidus

6.1 Espace de Banach

On rappelle qu'un espace de Banach E est un espace vectoriel normé complet.

6.2 Théorèmes de points fixes

Théorème 6.2.1

Soit (E, d) un espace métrique complet et $f : E \rightarrow E$ une application contractante (c'est-à-dire qu'il existe $0 \leq \rho < 1$ tel que pour tout $(x, y) \in E^2$, $d(f(x), f(y)) \leq \rho d(x, y)$), alors f admet un point fixe unique $x = f(x)$.

► cf. [15] ■

Théorème 6.2.2

Soit (E, d) un espace métrique complet et $f : E \rightarrow E$ une application, on suppose qu'il existe $p > 0$ telle que f^p soit contractante, alors f admet un point fixe unique $x = f(x)$.

► Le théorème du point fixe précédent appliqué à $g = f^p$ implique que g a un unique point fixe $x = g(x) = f^p(x)$. Par suite $f(x) = f(f^p(x)) = g(f(x))$. Donc $f(x)$ est aussi un point fixe de g . L'unicité du point fixe de g implique alors que $x = f(x)$.
Supposons maintenant que f possède 2 points fixes x_1 et x_2 . On a $x_1 = f(x_1) = f^p(x_1)$ et $x_2 = f^p(x_2)$. Donc x_1 et x_2 sont des points fixes de g . L'unicité du point fixe de g implique alors que $x_1 = x_2$. ■

6.3 Topologie

Lemme 6.3.1 (Coefficient de Lebesgue d'un recouvrement). Soient X un espace métrique compact et $(O_i)_{i \in I}$ un recouvrement ouvert de X , alors il existe $\varepsilon > 0$, appelé coefficient de Lebesgue du recouvrement, tel que pour tout $x \in X$, $\exists i \in I$, $B(x, \varepsilon) \subset O_i$.

► Pour tout $x \in X$, il existe $i \in I$ et $\eta_x > 0$, tel que $B(x, \eta_x) \subset O_i$. Comme $X \subset \bigcup_{x \in X} B(x, \eta_x/2)$ et que X est compact il existe J fini tel que $X \subset \bigcup_{j \in J} B(x_j, \eta_{x_j}/2)$. Posons alors $\varepsilon = \min_{j \in J} (\eta_{x_j}/2) > 0$, alors pour tout $x \in X$ il existe $j \in J$ tel que $d(x, x_j) < \eta_{x_j}/2$, par suite si $x' \in B(x, \varepsilon)$ on a $d(x', x_j) \leq d(x', x) + d(x, x_j) < \varepsilon + \eta_{x_j}/2 < \eta_{x_j}$. Et donc il existe $i \in O_i$ tel que $B(x, \varepsilon) \subset B(x_j, \eta_{x_j}) \subset O_i$. ■

6.4 Développement de Taylor

On rappelle ici les développements de Taylors. Soit $f : \Omega \subset E \rightarrow F$, une fonction, Ω ouvert, E et F espaces vectoriels normés. On suppose f de classe C^k . On rappelle que

$$\begin{aligned}
 f'(x) &\in \mathcal{L}(E, F) \\
 f''(x) &\in \mathcal{L}(E, \mathcal{L}(E, F)) \equiv \mathcal{L}_2(E, F) \\
 f''(x): E \times E &\longrightarrow F \\
 (h_1, h_2) &\longmapsto f''(x).(h_1, h_2) \\
 f^{(k)}(x) &\in \mathcal{L}_k(E, F) \\
 f^{(k)}(x): E^k &\longrightarrow F \\
 (h_1, \dots, h_k) &\longmapsto f^{(k)}(x).(h_1, \dots, h_k)
 \end{aligned} \tag{6.1}$$

Si $E = \mathbf{R}$, on identifie $f^{(k)}(x) \equiv f^{(k)}(x).(1, \dots, 1) \in F$.

Proposition 6.7.1

Si $x : [t_0, t_f] \rightarrow F$ est C^k et $(k+1)$ fois différentiable sur $]t_0, t_f[$ et telle que $\|x^{(k+1)}(t)\| \leq M, \forall t_0 < t < t_f$ alors

$$x(t_0 + h) = x(t_0) + x'(t_0).h + \dots + x^{(k)}(t_0) \frac{h^k}{k!} + r_k(t),$$

avec

$$\|r_k(t)\| \leq \frac{M}{(k+1)!} h^{k+1} = \mathcal{O}(h^{k+1})$$

Théorème 6.7.2

Soit $f : \Omega \subset E \rightarrow F$ une fonction C^{k+1} et x_0 et $x_0 + h$ deux points de Ω tels que $[x_0, x_0 + h] \subset \Omega$ alors

$$f(x_0 + h) = f(x_0) + f'(x_0).h + \frac{1}{2} f''(x_0).(h, h) + \dots + \frac{1}{k!} f^{(k)}(x_0).(\underbrace{h, \dots, h}_{k \text{ fois}}) + r_k(h),$$

avec

$$\|r_k(h)\| \leq \sup_{0 < t < 1} \|f^{(k+1)}(x_0 + th)\| \frac{\|h\|^{k+1}}{(k+1)!} \leq M \frac{\|h\|^{k+1}}{(k+1)!} = \mathcal{O}(\|h\|^{k+1})$$

Bibliographie

- [1] Patrick Altibelli and Luc Giraud. *Cours d'algèbre linéaire : valeurs propres et vecteurs propres de matrices, Décompositions canoniques de matrices et applications*. Université de Toulouse, INPT-ENSEEIH, département Informatique et Mathématiques Appliquées, 2, rue Camichel - B.P. 7122 - 31071 Toulouse Cedex FRANCE, 2008. \leftrightarrow 13 et 14.
- [2] H.G. Bock. Numerical treatment of inverse problems in chemical reaction kinetics. In K.H. Hebert, P. Deuffhard, and W. Jäger, editors, *Modelling of chemical reaction systems*, volume 18 of *Springer series in Chem. Phys.*, pages 102–125, 1981. \leftrightarrow 62.
- [3] J.C. Butcher. *Numerical Methods For Ordinary Differential Equations*. John Wiley & Sons, 2003. \leftrightarrow 40.
- [4] Jean-Pierre Demailly. *Analyse numérique et équations différentielles*. Collection Grenoble Sciences. Presses Universitaires de Grenoble, 1996. \leftrightarrow 49.
- [5] A. Guillou and J.L. Soulé. La résolution numérique des problèmes différentiels aux conditions initiales par les méthodes de collocation. *R.I.R.O.*, (R-3) :17–44, 1969. \leftrightarrow 55.
- [6] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*, volume 31 of *Springer Serie in Computational Mathematics*. Springer-Verlag, second edition edition, 2005. \leftrightarrow 7.
- [7] E. Hairer, S.P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I, Nonstiff Problems*, volume 8 of *Springer Serie in Computational Mathematics*. Springer-Verlag, second edition, 1993. \leftrightarrow 4, 7, 29, 35, 38, 42, 48, 50, 59 et 61.
- [8] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II, Stiff and Differential-Algebraic Problems*, volume 14 of *Springer Serie in Computational Mathematics*. Springer-Verlag, second edition, 1996. \leftrightarrow 4, 7 et 56.
- [9] Frédéric Jean. *Stabilité et Commande des Systèmes Dynamiques : Cours et exercices*. Les Presses de l'ENSTA, 2011. ISBN : 978-2-7225-0936-8. \leftrightarrow 8.
- [10] L. Pontriaguine. *Équations différentielles ordinaires*. Mir, 1975. \leftrightarrow 32 et 34.
- [11] L.F. Richardson. The deferred approach to the limit. *Philosophical Transactions of the Royal Society of London, Series A*, 226 :299–349, 1927. \leftrightarrow 49.
- [12] L.F. Shampine and M.W. Reichelt. The matlab ode suite.
http://www.mathworks.com/access/helpdesk/help/pdf_doc/otherdocs/ode_suite.pdf.
 \leftrightarrow 35.
- [13] C. Wagschal. *Dérivation, Intégration*. Hermann, 1999. \leftrightarrow 22.
- [14] C. Wagschal. *Fonctions holomorphes, Équations différentielles*. Hermann, 2003. \leftrightarrow 32 et 34.
- [15] C. Wagschal. *Topologie et analyse fonctionnelle*. Hermann, 2003. \leftrightarrow 25 et 66.
- [16] K. Wright. Some relationships between implicit runge-kutta collocation and lanczos τ methods, and their stability properties. *BIT*, 10 :217–227, 1970. \leftrightarrow 55.

Index

équations différentielles discontinues, [60](#)
équations variationnelles, [33](#)

détection d'événements, [59](#)
développement de Taylor, [67](#)

effet papillon, [36](#)
erreur de consistance, [43](#)
erreur locale, [41](#)
extrapolation de Richardson, [49](#)

fonction localement lipschitzienne, [22](#)

Landau, [40](#)
lemme de Gronwall, [46](#)

méthode à un pas, [38](#)
Méthode de Runge-Kutta, [40](#)
méthode explicite, [38](#)
méthode stable, [43](#)
Méthode consistante, [43](#)
Méthode convergente, [44](#)
méthode de collocation, [55](#)
Méthode de Runge-Kutta implicite, [54](#)
modèle de Lorentz, [36](#)

orbite d'Arenstorf, [37](#)
ordre, [41](#)

problème raide, [37](#)

Roberston, [37](#)

schéma d'Euler, [39](#)
schéma de Runge, [39](#)
sortie dense, [58](#)

tableau de Butcher, [40](#), [54](#)
théorème d'explosion, [27](#)
Théorème de Cauchy-Lipschitz, [23](#)
théorème de Peano, [29](#)
théorème de point fixe, [66](#)