

---

NOTE DE COURS

---

Vincent ACARY

INRIA Rhône Alpes

# INTRODUCTION À L'ANALYSE NUMÉRIQUE

Version préliminaire – Février 2003



# Table des matières

|  |           |
|--|-----------|
| <b>Avis aux lecteurs</b>   | <b>1</b>  |
| <hr/>  |           |
| <b>Chapitre 1 Meilleure approximation et interpolation polynomiale</b>             | <b>3</b>  |
| 1.1 Introduction . . . . .   | 4         |
| 1.2 Interpolation de Lagrange . . . . .  | 5         |
| 1.2.1 Principes . . . . .  | 5         |
| 1.2.2 Majoration de l'erreur d'interpolation de Lagrange . . . . .                 | 6         |
| 1.2.3 Techniques de construction du polynôme d'interpolation de Lagrange . . . . . | 6         |
| 1.2.4 Stabilité et convergence de l'approximation de Lagrange . . . . .            | 9         |
| 1.2.5 Interpolation aux points de Chebyshev . . . . .                              | 12        |
| 1.3 Interpolation de Hermite . . . . .   | 14        |
| 1.4 Interpolation par fonctions splines . . . . .                                  | 16        |
| 1.4.1 Principes . . . . .  | 16        |
| 1.4.2 Caractérisation des splines cubiques . . . . .                               | 17        |
| 1.4.3 Construction d'une spline d'interpolation . . . . .                          | 17        |
| 1.4.4 Un résultat de convergence pour les splines cubiques . . . . .               | 19        |
| 1.5 Existence d'un polynôme de meilleure approximation. . . . .                    | 20        |
| 1.5.1 Cas général . . . . .  | 20        |
| 1.5.2 Exemple classiques d'espace de fonction $\mathcal{E}$ . . . . .              | 20        |
| 1.6 Meilleure approximation polynomiale au sens des moindres carrés . . . . .      | 22        |
| 1.6.1 Cas hilbertien. Polynômes orthogonaux . . . . .                              | 22        |
| 1.6.2 Approximation continue au sens de moindres carrés. . . . .                   | 24        |
| 1.6.3 Approximation discrète – Méthodes des moindres carrés. . . . .               | 25        |
| <b>Chapitre 2 Intégration numérique</b>  | <b>27</b> |
| 2.1 Introduction . . . . .   | 28        |
| 2.2 Méthodes d'intégration numérique composées . . . . .                           | 28        |
| 2.2.1 Principe et exemples usuels . . . . .  | 28        |

|  |   |           |
|--|---|-----------|
| 2.2.2  | Principe général de construction . . . . .  | 30        |
| 2.2.3  | Théorème de convergence des méthodes de quadrature composées. . . . .                 | 31        |
| 2.2.4  | Les méthodes de Newton-Cotes . . . . .  | 32        |
| 2.3  | Étude générale de l'erreur d'intégration numérique . . . . .                          | 32        |
| 2.3.1  | Ordre et erreur d'une méthode d'intégration numérique . . . . .                       | 33        |
| 2.3.2  | Noyau et théorème de Peano . . . . .  | 33        |
| 2.3.3  | Cas des méthodes de quadrature composées. . . . .                                     | 36        |
| 2.4  | Méthodes d'intégration numérique de Gauss . . . . .                                   | 38        |
| 2.4.1  | Principe . . . . .  | 38        |
| 2.4.2  | Cas général . . . . .   | 38        |
| 2.4.3  | Méthodes de Gauss usuelles . . . . .  | 40        |
| 2.4.4  | Méthodes de Gauss-Lobatto . . . . .   | 43        |
| <b>Chapitre 3 Méthodes itératives pour la résolution d'équations non linéaires</b> |   | <b>45</b> |
| 3.1  | Introduction . . . . .  | 46        |
| 3.2  | Une méthode topologique: la méthode de dichotomie . . . . .                           | 46        |
| 3.3  | Théorème du point fixe et applications contractantes . . . . .                        | 47        |
| 3.3.1  | Applications contractantes et point fixe . . . . .                                    | 47        |
| 3.3.2  | Théorème du point fixe . . . . .  | 48        |
| 3.3.3  | Estimation de la vitesse de convergence . . . . .                                     | 49        |
| 3.3.4  | Principe de construction des méthodes itératives métriques . . . . .                  | 49        |
| 3.4  | Méthodes itératives métriques – Cas $f : \mathbb{R} \mapsto \mathbb{R}$ . . . . .     | 49        |
| 3.4.1  | Points fixes répulsifs et points fixes attractifs . . . . .                           | 50        |
| 3.4.2  | Méthode de la corde . . . . .   | 51        |
| 3.4.3  | Méthode de Newton . . . . .   | 52        |
| 3.4.4  | Méthode de la sécante de Steffensen . . . . .   | 54        |
| 3.5  | Méthodes itératives métriques – Cas $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ . . . . . | 55        |
| 3.5.1  | Méthode Newton-Raphson . . . . .  | 55        |
| 3.6  | Cas particuliers où $f$ est un polynôme – Racines d'un polynôme . . . . .             | 56        |
| 3.6.1  | Procédé de localisation de Sturm . . . . .  | 56        |
| 3.6.2  | Méthode de Bairstow . . . . .   | 57        |
| 3.6.3  | Méthode d'Aitken . . . . .  | 59        |
| <b>Chapitre 4 Équations Différentielles Ordinaires (EDO) – Généralités</b>         |   | <b>61</b> |
| 4.1  | Introduction . . . . .  | 62        |
| 4.2  | Terminologie et définition . . . . .  | 62        |
| 4.2.1  | Système différentiel du premier ordre . . . . .                                       | 62        |

|   |  |           |
|---|--|-----------|
| 4.2.2   | Le problème de Cauchy . . . . .                                | 62        |
| 4.2.3   | Système différentiel du second ordre . . . . .                 | 63        |
| 4.2.4   | Système différentiel d'ordre $m$ . . . . .                     | 64        |
| 4.3   | Théorèmes d'existence et d'unicité . . . . .                   | 65        |
| 4.3.1   | Solutions locales, maximales, globales . . . . .               | 65        |
| 4.3.2   | Théorème d'existence de Cauchy-Peano . . . . .                 | 67        |
| 4.3.3   | Théorème d'unicité de Cauchy-Lipschitz . . . . .               | 68        |
| 4.3.4   | Influence des perturbations sur les données . . . . .          | 69        |
| 4.3.5   | Régularité des solutions . . . . .                             | 70        |
| 4.4   | Méthodes explicites de résolution . . . . .                    | 71        |
| 4.4.1   | Équations différentielles linéaires du premier ordre . . . . . | 71        |
| 4.4.2   | Équations différentielles linéaires du second ordre . . . . .  | 71        |
| <b>Chapitre 5 EDO – Méthodes numériques à un pas</b>        |  | <b>73</b> |
| 5.1   | Introduction . . . . .   | 74        |
| 5.2   | Exemples de méthodes numériques classiques . . . . .           | 75        |
| 5.2.1   | Principe de construction . . . . .                             | 75        |
| 5.2.2   | La méthode d'Euler progressive ou explicite . . . . .          | 75        |
| 5.2.3   | La méthode d'Euler rétrograde ou implicite . . . . .           | 76        |
| 5.2.4   | La $\theta$ -méthode . . . . .                                 | 76        |
| 5.3   | Étude générale des méthodes à un pas . . . . .                 | 76        |
| 5.3.1   | Consistance, stabilité et convergence . . . . .                | 76        |
| 5.3.2   | Condition de convergence des méthodes à un pas. . . . .        | 78        |
| 5.3.3   | Ordre d'une méthode à un pas . . . . .                         | 79        |
| 5.4   | Les méthodes de développement de Taylor . . . . .              | 80        |
| 5.5   | Les méthodes de Runge–Kutta . . . . .                          | 82        |
| 5.5.1   | Principe de construction . . . . .                             | 82        |
| 5.5.2   | Exemples . . . . .   | 84        |
| 5.5.3   | Stabilité des méthodes de Runge–kutta . . . . .                | 87        |
| 5.5.4   | Ordre des méthodes de Runge–kutta . . . . .                    | 87        |
| 5.6   | Contrôle du pas de temps . . . . .                             | 89        |
| 5.6.1   | Principe générale du contrôle . . . . .                        | 89        |
| 5.6.2   | Applications aux méthodes présentées. . . . .                  | 89        |
| <b>Chapitre 6 EDO – Méthodes numériques à pas multiples</b> |  | <b>91</b> |
| 6.1   | Introduction . . . . .   | 92        |
| 6.2   | Exemples classiques . . . . .                                  | 92        |

|       |   |     |
|-------|---|-----|
| 6.2.1 | La méthode de Nyström . . . . .                                       | 92  |
| 6.2.2 | La méthode de Milne . . . . .   | 93  |
| 6.3   | Étude générale d'une classe de méthodes à pas multiples . . . . .     | 93  |
| 6.3.1 | Erreur de consistance et ordre . . . . .                              | 93  |
| 6.3.2 | Stabilité . . . . .   | 93  |
| 6.4   | Les méthodes d'Adams-Bashforth à $r + 1$ pas ( $AB_{r+1}$ ) . . . . . | 94  |
| 6.4.1 | Principe . . . . .  | 94  |
| 6.4.2 | Erreur de consistance et ordre des méthodes $AB_{r+1}$ . . . . .      | 96  |
| 6.4.3 | Stabilité des méthodes $AB_{r+1}$ . . . . .                           | 97  |
| 6.5   | Les méthodes d'Adams-Moulton à $r + 1$ pas ( $AM_{r+1}$ ) . . . . .   | 97  |
| 6.5.1 | Principe . . . . .  | 97  |
| 6.5.2 | Exemple . . . . .   | 98  |
| 6.5.3 | Consistance et stabilité . . . . .                                    | 99  |
| 6.6   | Les méthodes prédicteur-correcteur <i>PECE</i> . . . . .              | 100 |
| 6.6.1 | Principe . . . . .  | 100 |
| 6.6.2 | Ordre et stabilité des schémas <i>PECE</i> . . . . .                  | 101 |
| 6.6.3 | Exemples . . . . .  | 101 |

|                      |  |  |                      |                      |          |
|----------------------|--|--|----------------------|----------------------|----------|
| <b>Bibliographie</b> | <b>Chapitre 6. EDO – Méthodes numériques à pas multiples</b> | <b>6.6. Les méthodes prédicteur-correcteur <i>PECE</i></b> | <b>BIBLIOGRAPHIE</b> | <b>BIBLIOGRAPHIE</b> | <b>i</b> |
|----------------------|--|--|----------------------|----------------------|----------|

|              |  |            |
|--------------|--|------------|
| <b>Index</b> |  | <b>iii</b> |
|--------------|--|------------|

|  |          |
|--|----------|
| <b>Annexe A Évaluation d'un polynôme grâce au schéma de Hörner</b> | <b>1</b> |
|--|----------|

|     |   |   |
|-----|---|---|
| A.1 | Schéma de Hörner sur la forme canonique . . . . . | 1 |
| A.2 | Schéma de Hörner sur la forme de Newton . . . . . | 2 |

|  |          |
|--|----------|
| <b>Annexe B Les polynômes de Chebyshev</b> | <b>5</b> |
|--|----------|

|     |  |   |
|-----|--|---|
| B.1 | Forme explicite des polynômes de Chebyshev . . . . . | 5 |
| B.2 | Racines des polynômes de Chebyshev . . . . .         | 6 |

# Avis aux lecteurs

Cette note de cours a servi de support à la seconde partie du module d'introduction à l'analyse numérique du DESS Technique et application de la Physique, option modélisation et simulation numérique de l'université Joseph Fourier de Grenoble.

Il s'agit d'une version préliminaire de ce qui aurait du devenir un polycopié. Il s'appuie largement sur les ouvrages classiques sur les sujets abordés, CROUZEIX & MIGNOT (1992) ; DEMAILLY (1996) ; SCHATZMAN (2001) auxquels je renvoie pour la plupart des démonstrations.

Loin d'être finalisée, cette note comporte encore de nombreuses erreurs de frappe et des imprécisions. Je serais reconnaissant aux lecteurs qui me feront parvenir leurs remarques pour améliorer cette note.



# Chapitre 1

## Meilleure approximation et interpolation polynomiale

### Contenu du Chapitre 1

---

|            |  |           |
|------------|--|-----------|
| <b>1.1</b> | <b>Introduction</b>  | <b>4</b>  |
| <b>1.2</b> | <b>Interpolation de Lagrange</b>                                       | <b>5</b>  |
| 1.2.1      | Principes  | 5         |
| 1.2.2      | Majoration de l'erreur d'interpolation de Lagrange                     | 6         |
| 1.2.3      | Techniques de construction du polynôme d'interpolation de Lagrange     | 6         |
| 1.2.4      | Stabilité et convergence de l'approximation de Lagrange                | 9         |
| 1.2.5      | Interpolation aux points de Chebyshev                                  | 12        |
| <b>1.3</b> | <b>Interpolation de Hermite</b>  | <b>14</b> |
| <b>1.4</b> | <b>Interpolation par fonctions splines</b>                             | <b>16</b> |
| 1.4.1      | Principes  | 16        |
| 1.4.2      | Caractérisation des splines cubiques                                   | 17        |
| 1.4.3      | Construction d'une spline d'interpolation                              | 17        |
| 1.4.4      | Un résultat de convergence pour les splines cubiques                   | 19        |
| <b>1.5</b> | <b>Existence d'un polynôme de meilleure approximation.</b>             | <b>20</b> |
| 1.5.1      | Cas général  | 20        |
| 1.5.2      | Exemple classiques d'espace de fonction $\mathcal{E}$                  | 20        |
| <b>1.6</b> | <b>Meilleure approximation polynomiale au sens des moindres carrés</b> | <b>22</b> |
| 1.6.1      | Cas hilbertien. Polynômes orthogonaux                                  | 22        |
| 1.6.2      | Approximation continue au sens de moindres carrés.                     | 24        |
| 1.6.3      | Approximation discrète – Méthodes des moindres carrés.                 | 25        |

---

## Notations

On notera  $\mathcal{E}$  un espace vectoriel de fonctions muni d'une norme  $\|\cdot\|_{\mathcal{E}}$ .

On note  $\mathcal{P}_n$  l'ensemble des polynômes algébriques à une variable de degré  $\leq n$ , et à coefficients réels. L'ensemble  $\mathcal{P}_n$  est un espace vectoriel de dimension  $n + 1$ .

On note  $\mathcal{C}^m([a, b])$ ,  $m \geq 0$  l'ensemble des fonctions  $m$  fois continûment différentiables d'un intervalle  $[a, b]$  de  $\mathbb{R}$ .

Si  $f$  est une fonction définie sur  $[a, b] \subset \mathbb{R}$  à valeurs dans  $\mathbb{R}$ , on notera la norme uniforme de  $f$  sur  $[a, b]$  de la manière suivante :

$$\|f\|_{\infty} = \sup_{x \in [a, b]} |f(x)|$$

### 1.1 Introduction

Nous allons nous intéresser à l'approximation d'une fonction numérique d'une variable réelle par un polynôme sur un intervalle de  $\mathbb{R}$ . Deux types de méthodes peuvent être employées :

1. l'*interpolation polynomiale* de la fonction en des points distincts. Le problème est le suivant :

Considérons l'espace des polynômes  $\mathcal{P}_n$ ,  $(n+1)$  points distincts  $\{x_0, x_1, \dots, x_n\}$ , "interpoler une fonction numérique  $f$  dans  $\mathcal{P}_n$ " signifie :

Trouver (existence? unicité?) un élément  $p_n$  de  $\mathcal{P}_n$  tel que

$$\forall i \in 0 \dots n, \quad p_n(x_i) = f(x_i) \tag{1.1.1}$$

2. la *meilleure approximation polynomiale* d'une fonction numérique  $f$  dans un espace vectoriel de fonctions  $\mathcal{E}$  tel que  $\mathcal{P}_n \subset \mathcal{E}$ , revient à chercher un élément de  $\mathcal{P}_n$ , noté  $p_n$ , "proche" de  $f$ , c'est à dire :

$$\|f - p_n\|_{\mathcal{E}} = \inf_{q \in \mathcal{P}_n} \|f - q\|_{\mathcal{E}} \tag{1.1.2}$$

En réalité, l'interpolation constitue une méthode d'approximation particulière, plus intuitive mais souvent moins performante. Dans ce chapitre, nous allons mettre l'accent sur l'interpolation de Lagrange (§ 1.2), de Hermite (§ 1.3) et l'interpolation par splines (§ 1.4). Dans un deuxième temps, nous allons présenter les techniques de meilleure approximation polynomiale (§ 1.5 et § 1.6).

## 1.2 Interpolation de Lagrange

### 1.2.1 Principes

Considérons un intervalle  $[a, b]$ , fermé et borné de  $\mathbb{R}$ , une fonction numérique réelle,  $f$  définie sur  $[a, b]$ . On a le théorème suivant :

**Théorème 1.2.1** Soit un ensemble de  $(n + 1)$  points *distincts* de  $[a, b]$  noté  $\{x_0, x_1, \dots, x_n\}$ , il existe un unique polynôme  $p_n \in \mathcal{P}_n$  tel que :

$$\forall i \in \{0, 1, \dots, n\}, \quad p_n(x_i) = f(x_i) \quad (1.2.1)$$

□

**Proposition 1.2.1** Le polynôme  $p_n$  répondant à la condition (1.2.1), s'exprime :

$$p_n(x) = \sum_{i=0}^n f(x_i)L_i(x), \quad (1.2.2)$$

où l'on a posé

$$L_i(x) = \prod_{j=0, j \neq i}^n \frac{x - x_j}{x_i - x_j} \quad (1.2.3)$$

*Preuve* : Pour montrer que le polynôme  $p_n$  défini par (1.2.2) vérifie les conditions (1.2.1), il suffit de constater que

$$L_i(x_j) = \delta_{ij} \quad (1.2.4)$$

où  $\delta_{ij}$  est le symbole de Kronecker<sup>1</sup>. Il résulte directement de (1.2.2) que  $p_n$  vérifie (1.2.1). De plus, les polynômes  $L_i$  appartiennent à  $\mathcal{P}_n$ , donc  $p_n$  appartient à  $\mathcal{P}_n$ .

Supposons qu'il existe maintenant une autre polynôme  $\hat{p}_n$  vérifiant (1.2.1). Posons  $q_n = \hat{p}_n - p_n$ ; on peut remarquer que  $q_n \in \mathcal{P}_n$  et que  $q_n(x_i) = 0$ . Le polynôme  $q_n$  est de degré  $\leq n$  et possède  $n + 1$  racines distinctes, donc le polynôme suivant

$$\pi_{n+1} = \prod_{j=0}^n (x - x_j) \quad (1.2.5)$$

qui est un polynôme de degré  $n + 1$  divise  $q_n$ ; ceci n'est possible que si  $q_n \equiv 0$ , d'où  $\hat{p}_n \equiv p_n$ .

□

**Définition 1.2.1** Le polynôme  $p_n$  s'appelle le *polynôme d'interpolation de Lagrange* de la fonction  $f$  relativement aux points  $\{x_i\}_{i=0, \dots, n}$ .

Les fonctions  $L_i(x)$  s'appellent les *polynômes de base de Lagrange* associés aux points  $\{x_i\}_{i=0, \dots, n}$ .

□

---

1.  $\delta_{ij} = 0$  si  $i \neq j$  et  $\delta_{ij} = 1$  si  $i = j$ .

**Remarques.** On peut considérer le problème suivant qui est le problème général d'interpolation de Lagrange :

Étant donné  $n+1$  point distincts  $\{x_i\}_{i=0,\dots,n}$ , et  $n+1$  valeurs  $\{y_i\}_{i=0,\dots,n}$ , déterminer  $p_n \in \mathcal{P}_n$  tel que :

$$\forall i \in \{0, 1, \dots, n\}, \quad p_n(x_i) = y_i \quad (1.2.6)$$

La démonstration du théorème (1.2.1) donne l'existence et l'unicité d'une solution à ce problème général d'interpolation de Lagrange. Il est intéressant de remarquer que le système (1.2.6) est en fait un système linéaire de  $(n+1)$  équations à  $(n+1)$  inconnues.

Posons  $p_n(x) = a_0 + a_1x + \dots + a_nx^n$ ,  $a = [a_0, \dots, a_n]^T$  et  $y = [y_0, \dots, y_n]^T$ . Le système d'équations (1.2.6) revient à écrire :

$$Va = y \quad (1.2.7)$$

avec la matrice carrée d'ordre  $(n+1)$   $V = (v_{ij})$ , composée des éléments  $v_{ij} = (x_i)^j$ . La matrice  $V$  n'est autre que la matrice de Vandermonde dont l'existence et l'unicité des solutions de (1.2.7) prouve qu'elle est inversible pour  $x_i \neq x_j, \forall i \neq j$ .

Nous verrons au paragraphe 1.2.3 qu'il existe des techniques plus efficaces pour résoudre le problème d'interpolation de Lagrange que la résolution du système linéaire (1.2.7) le plus souvent mal conditionné.

## 1.2.2 Majoration de l'erreur d'interpolation de Lagrange

**Théorème 1.2.2** On suppose que  $f$  est  $n+1$  fois dérivable sur  $[a, b]$ . Alors  $\forall x \in [a, b]$ , il existe un point  $\xi_x \in ]\min(x, x_i), \max(x, x_i)[$  tel que

$$f(x) - p_n(x) = \frac{1}{(n+1)!} \pi_{n+1}(x) f^{(n+1)}(\xi_x) \quad (1.2.8)$$

où  $f^{(n+1)}$  désigne la dérivée d'ordre  $(n+1)$  de  $f$ . □

*Preuve :* La preuve est basée sur l'application du Théorème de Rolle sur la fonction  $f(x) - p_n(x)$  et sur ces dérivées. On renvoie, pour plus de détails aux ouvrages suivants : (DEMAILLY, 1996), (CROUZEIX & MIGNOT, 1992) □

## 1.2.3 Techniques de construction du polynôme d'interpolation de Lagrange

L'évaluation du polynôme de Lagrange peut se faire au moyen de son expression dans la base de Lagrange (1.2.2) ou par la résolution du système linéaire (1.2.7). Dans la pratique numérique, ces deux méthodes sont peu efficaces. L'objet de ce paragraphe est de donner une technique de construction du polynôme de Lagrange basée sur la forme de Newton du polynôme d'interpolation de Lagrange.

**Forme de Newton du polynôme d'interpolation de Lagrange** On se donne une fonction numérique  $f$  définie sur  $[a, b]$  et  $\{x_i\}_{i=0..n}$  ( $n+1$ ) points distincts de cet intervalle.

On note  $p_k$  le polynôme d'interpolation de Lagrange de  $f$  relativement aux  $(k+1)$  points  $x_0, x_1, \dots, x_k$ ,  $k = 0, \dots, n$ .

Essayons de construire le polynôme  $p_n$  à partir de polynômes  $p_k$ . Nous avons tout d'abord :

$$p_0(x) = f(x_0) = \alpha_0 \quad (1.2.9)$$

Pour  $k \geq 1$ , le polynôme  $p_k - p_{k-1}$  s'annule aux points  $x_0, x_1, \dots, x_{k-1}$ ; il est donc de la forme :

$$p_k(x) - p_{k-1}(x) = \alpha_k(x - x_0)(x - x_1) \cdots (x - x_{k-1}) \quad (1.2.10)$$

Remarquons au passage que  $\alpha_k$  est le coefficient de  $x^k$  dans  $p_k$ .

**Définition 1.2.2** Le polynôme d'interpolation de Lagrange  $p_n$  relatif aux  $(n+1)$  points,  $\{x_i\}_{i=0..n}$ , s'écrit donc :

$$p_n(x) = \alpha_0 + \sum_{k=1}^n \alpha_k(x - x_0)(x - x_1) \cdots (x - x_{k-1}) \quad (1.2.11)$$

qui est appelé la *forme de Newton du polynôme d'interpolation*  $p_n$ .

Les polynômes suivants :

$$N_k(x) = (x - x_0)(x - x_1) \cdots (x - x_{k-1}) \quad (1.2.12)$$

sont appelés les *polynômes de base de Newton*.

□

**Méthode des différences divisées** L'expression du polynôme dans la base de Newton (1.2.11) ramène le calcul par récurrence de  $p_n$  à l'évaluation des  $\alpha_k$ . Cette remarque est à la base de la *méthode des différences divisées*

Introduisons une notation qui permet de généraliser l'expression des  $\alpha_k$ . Considérons les points,  $x_l, \dots, x_k$  et le polynôme d'interpolation de degré  $k-l$  relatifs à ces points. Notons, désormais,  $\delta_{[x_l, \dots, x_k]}$  le coefficient de ce polynôme en  $x^{k-l}$ . On retrouve, naturellement, que  $\alpha_k = \delta_{[x_0, \dots, x_k]}$ .

**Proposition 1.2.2**

$$\forall k \geq l, \quad \delta_{[x_l, \dots, x_k]} = \frac{\delta_{[x_{l+1}, \dots, x_k]} - \delta_{[x_l, \dots, x_{k-1}]}}{x_k - x_l} \quad (1.2.13)$$

et

$$\delta_{[x_i]} = f(x_i) \quad (1.2.14)$$

*Preuve* : Considérons la formule pour  $l = 0$ ,

$$\forall k \geq 1, \quad \delta_{[x_0, \dots, x_k]} = \frac{\delta_{[x_1, \dots, x_k]} - \delta_{[x_0, \dots, x_{k-1}]}}{x_k - x_0} \quad (1.2.15)$$

Notons pour  $k \geq 1$ ,  $p_{k-1}^*$  le polynôme de degré  $k - 1$  d'interpolation de  $f$  aux points  $x_1, x_2, \dots, x_k$ . Le coefficient de  $x^{k-1}$  est, par définition,  $\delta_{[x_1, \dots, x_k]}$ . Considérons le polynôme de degré  $k$  défini par

$$q_k(x) = \frac{(x - x_0)p_{k-1}^* - (x - x_k)p_{k-1}}{x_k - x_0} \quad (1.2.16)$$

coïncide avec  $f$  aux points  $x_0, x_1, \dots, x_k$  et donc  $q_k \equiv p_k$ . On obtient la formule (1.2.13) en identifiant dès lors les termes en  $x_k$  dans l'équation (1.2.16). La formule générale (1.2.15) se déduit de la même manière en considérant le polynôme d'interpolation de  $x_l, \dots, x_k$ . □

**Définition 1.2.3** La quantité  $\delta_{[x_0, \dots, x_k]}$  est appelée *différence divisée* de  $f$  aux points  $x_0, x_1, \dots, x_k$ . □

**Algorithme** La donnée des  $\delta_{[x_i]}$  est immédiate grâce à la formule (1.2.14). On calcule ensuite de "proche en proche" les différentes quantités  $\delta_{[x_l, \dots, x_k]}$  grâce à la formule (1.2.13). L'algorithme est illustré sur la figure 1.1.

**Remarque 1** Cet algorithme est aussi particulièrement efficace lorsqu'on ajoute un point d'interpolation supplémentaire. On remarque en effet qu'il suffit d'ajouter une nouvelle valeur de  $\delta_{[i, i]}$  dans la colonne de gauche et de recalculer les différentes différences divisées.

**Remarque 2** Afin de mieux comprendre le fonctionnement de cet algorithme, on peut essayer de donner le système linéaire à résoudre pour l'interpolation dans la base de Newton.

En appliquant la forme de Newton du polynôme pour tous les points d'interpolation  $x_i$ , on obtient le système suivant :

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & (x_1 - x_0) & 0 & \dots & 0 \\ 1 & (x_2 - x_0)(x_2 - x_1) & (x_2 - x_1) & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & (x_n - x_0) \dots (x_n - x_1) & \dots & \dots & (x_n - x_{n-1}) \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} \quad (1.2.17)$$

La matrice de ce système linéaire est triangulaire inférieure. Elle est inversible si tous les  $x_i$  sont distincts; il suffit pour s'en convaincre d'envisager le déterminant de cette matrice. Si on résout ce système en effectuant une descente sur les  $\alpha_i$  on peut voir apparaître l'algorithme de Newton.

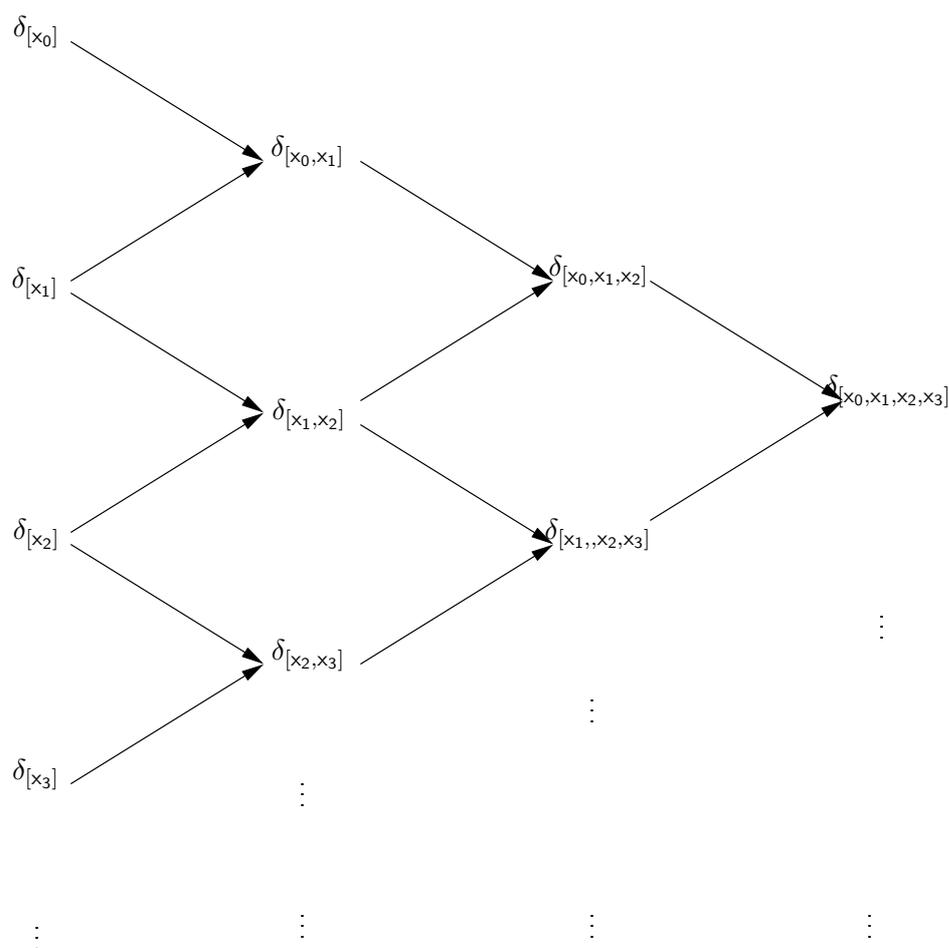


FIG. 1.1 – Méthode des différences divisées — Algorithme de Newton

### 1.2.4 Stabilité et convergence de l'approximation de Lagrange

Deux questions importantes peuvent être posées concernant l'interpolation de Lagrange :

1. La convergence uniforme (globale) de l'interpolation de Lagrange lorsque que le nombre de points d'interpolation tend vers l'infini :

$$\lim_{n \rightarrow +\infty} \|f - p_n\|_{\infty} ? \quad (1.2.18)$$

2. La stabilité de l'interpolation de Lagrange, c'est à dire, l'influence sur  $p_n$  des erreurs sur les valeurs  $f(x_i)$ .

Introduisons l'opérateur d'interpolation de Lagrange :

$$\begin{aligned} L_n : \mathcal{C}([a, b]) &\longmapsto \mathcal{P}_n \\ f &\longmapsto p_n \end{aligned} \quad (1.2.19)$$

Considérons des valeurs approchées de  $f(x_i)$  notées  $\hat{f}_i$ , on a

$$p_n(x) = \sum_{i=0}^n f(x_i) L_i(x) \quad (1.2.20)$$

$$\hat{p}_n(x) = \sum_{i=0}^n \hat{f}_i L_i(x) \quad (1.2.21)$$

$$p_n(x) - \hat{p}_n(x) = \sum_{i=0}^n (f(x_i) - \hat{f}_i) L_i(x) \quad (1.2.22)$$

et donc

$$|p_n(x) - \hat{p}_n(x)| \leq \sum_{i=0}^n |L_i(x)| \max_{0 \leq i \leq n} |f(x_i) - \hat{f}_i| \quad (1.2.23)$$

**Définition 1.2.4** Le nombre réel,

$$\Lambda_n = \left\| \sum_{i=0}^n |L_i(x)| \right\|_{\infty}, \quad (1.2.24)$$

s'appelle la *constante de Lebesgue* associée aux points  $\{x_i\}_{i=0, \dots, n}$  et l'intervalle  $[a, b]$ . □

On peut montrer que

$$\Lambda_n = \max_{\substack{g \in \mathcal{C}^0([a, b]) \\ g \neq 0}} \frac{\|L_n(g)\|_{\infty}}{\|g\|_{\infty}} \quad (1.2.25)$$

ce qui représente la norme de l'opérateur d'interpolation  $L_n$ .

**Proposition 1.2.3** La stabilité de l'interpolation de Lagrange est donnée par la formule suivante :

$$|p_n(x) - \hat{p}_n(x)| \leq \Lambda_n \max_{0 \leq i \leq n} |f(x_i) - \hat{f}_i| \quad (1.2.26)$$

La constante de Lebesgue s'interprète donc comme le coefficient d'amplification des erreurs dans le procédé d'interpolation.

*Preuve* : Immédiate à partir de (1.2.23) et de la définition 1.2.4. □

La constante de Lebesgue est également liée à l'erreur d'interpolation comme le montre le théorème suivant :

**Théorème 1.2.3**  $\forall f \in \mathcal{C}^0([a, b])$ , on a

$$\|f - p_n\|_{\infty} \leq (1 + \Lambda_n) E_n(f) \quad (1.2.27)$$

$$\text{avec } E_n(f) = d(f, \mathcal{P}_n) = \inf_{q \in \mathcal{P}_n} \|f - q\|_{\infty} \quad (1.2.28)$$

□

*Preuve* :  $\forall q \in \mathcal{P}_n$ , on a

$$f - p_n = f - q + q - p_n = f - q + L_n(q - p_n) = f - q + L_n(q - f) \quad (1.2.29)$$

d'où

$$\|f - p_n\|_\infty \leq \|f - q\|_\infty + \|L_n(f - q)\|_\infty \leq \|f - q\|_\infty(1 + \Lambda_n) \quad (1.2.30)$$

En prenant la borne inférieure de  $\|f - q\|_\infty$  pour  $q \in \mathcal{P}_n$ , on obtient le résultat.  $\square$

**Définition 1.2.5** La valeur  $E_n(f)$  est appelée *degré d'approximation* de la fonction  $f$  par les polynômes de  $\mathcal{P}_n$  au sens de la convergence uniforme.  $\square$

Examinons de plus près les différents termes de la majoration (1.2.27). La constante de Lebesgue ne dépend que du choix des points  $\{x_i\}_{0 \leq i \leq n}$  dans  $[a, b]$  tandis que le degré d'approximation,  $E_n(f)$  ne dépend que de  $f$ .

Afin de minimiser l'erreur d'interpolation pour un nombre de points fixé, nous avons intérêt à choisir les points  $\{x_i\}_{0 \leq i \leq n}$  qui minimise la constante de Lebesgue. Nous avons toutefois le résultat suivant :

**Proposition 1.2.4** Si l'on note  $\bar{\Lambda}_n$  la valeur inférieure de  $\Lambda_n$  pour tous les choix possibles de points  $\{x_i\}_{0 \leq i \leq n}$  dans  $[a, b]$ , on a

$$\bar{\Lambda}_n \underset{n \rightarrow +\infty}{\sim} \frac{2}{\pi} \ln(n) \quad (1.2.31)$$

Il est bon de remarquer que ce résultat a pour conséquence que

$$\lim_{n \rightarrow +\infty} \bar{\Lambda}_n = +\infty, \quad (1.2.32)$$

ce qui a pour conséquence qu'il existera toujours une fonction continue tel que  $L_n(f)$  ne converge pas uniformément vers  $f$  quand  $n \rightarrow +\infty$ . En d'autres termes, aucune garantie ne peut être donnée sur la convergence de l'interpolation de Lagrange même pour des fonctions continues.

**Phénomène de Runge** On donne dans ce paragraphe un exemple classique de fonction analytique pour laquelle l'interpolation de Lagrange ne converge pas. Ce phénomène est communément appelé le phénomène de Runge.

Considérons la fonction :

$$f_\alpha(x) = \frac{1}{x^2 + \alpha^2}, x \in [-1; 1] \quad (1.2.33)$$

où  $\alpha > 0$  est un paramètre.

On illustre sur la figure 1.2 le résultat de l'interpolation de Lagrange pour  $\alpha = 0.45$  et  $n + 1$  points équidistants :

$$x_i = a + \frac{(b-a)}{n}i \quad (1.2.34)$$

On constate que plus le nombre de points augmente plus le polynôme d'interpolation oscille.

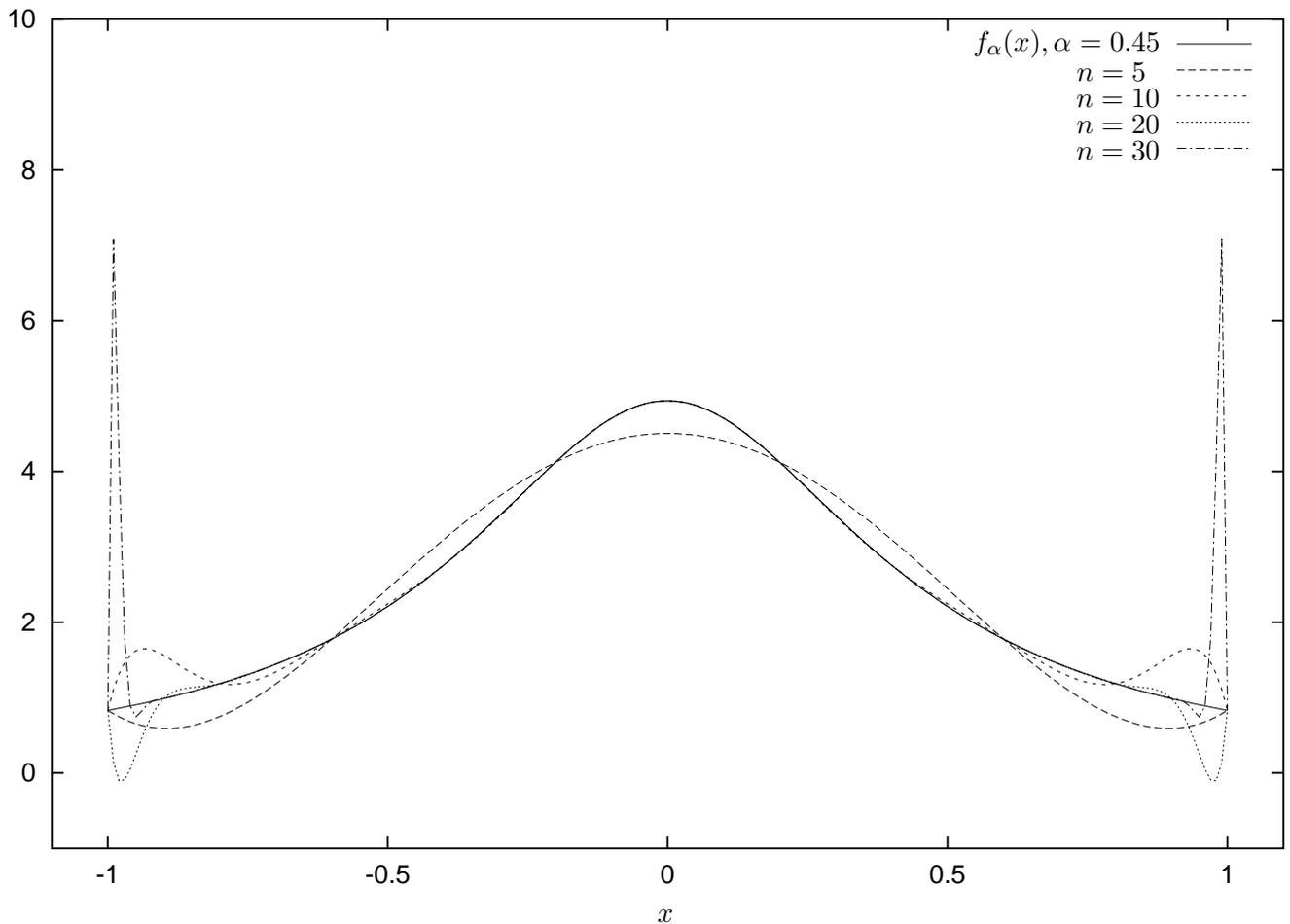


FIG. 1.2 – Phénomène de Runge

L'interpolation ne converge visiblement pas. Une étude plus fine peut montrer que suivant la valeur affectée à  $\alpha$  l'interpolation peut converger ou non.

### 1.2.5 Interpolation aux points de Chebyshev

D'une manière générale, il ne faut pas s'attendre à ce que l'interpolation de Lagrange converge. Cependant, il est possible d'améliorer son comportement en choisissant les points d'interpolation (quand cela est possible !) de façon à minimiser la constante de Lebesgue.

Par exemple, pour des points d'interpolation équi-distants, le constante de Lebesgue est estimée à :

$$\Lambda_n \underset{n \rightarrow +\infty}{\sim} \frac{2^{n+1}}{en \ln(n)}, \quad (1.2.35)$$

ce qui est loin d'être optimum par rapport à l'estimation (1.2.31).

On peut caractériser les points qui donne à la constante de Lebesgue sa valeur minimale  $\bar{\Lambda}_n$  mais le calcul est trop compliqué pour être efficace numériquement.

Une solution consiste à choisir pour points d'interpolation les zéros des polynômes de Chebyshev de degré  $(n + 1)$ , appelés communément les *points de Chebyshev*, soit :

$$x_i^{(n)} = \frac{a+b}{2} + \frac{b-a}{2} \cos\left(\frac{(2i+1)\pi}{2n+2}\right), i \in \{0, \dots, n\} \quad (1.2.36)$$

Dans ce cas, on a certes  $\Lambda_n > \bar{\Lambda}_n$  mais on conserve la propriété :

$$\Lambda_n \underset{n \rightarrow +\infty}{\sim} \frac{2}{\pi} \ln(n) \quad (1.2.37)$$

Un résultat intéressant concernant l'interpolation de Lagrange aux points de Chebyshev est qu'elle converge si  $f \in \mathcal{C}^1([a, b])$  (Résultat de Bernstein). On illustre ce résultat sur la Figure 1.3 en reprenant le précédent exemple.

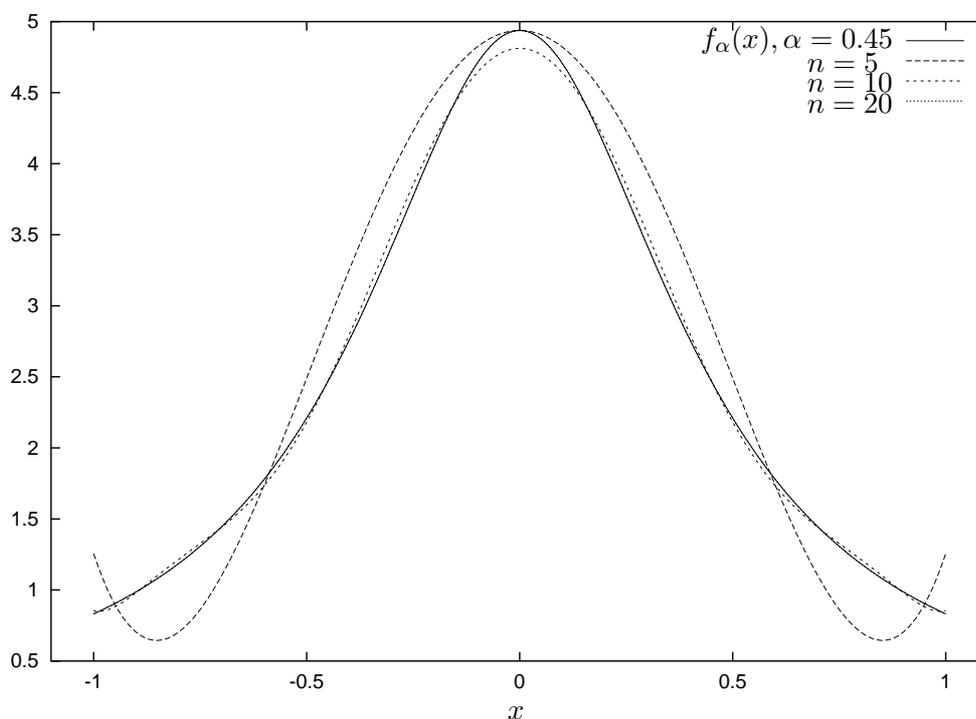


FIG. 1.3 – Interpolation de Lagrange aux points de Chebyshev

### 1.3 Interpolation de Hermite

**Principe.** L'interpolation de Hermite cherche à faire coïncider non seulement  $f$  et  $p_n$  aux points  $x_i$  mais aussi leurs dérivées jusqu'à l'ordre  $\alpha_i$ . On supposera donc que  $f$  est suffisamment dérivable.

**Théorème 1.3.1** Étant donnés  $\{x_i\}_{i=0\dots k}$ ,  $(k+1)$  points distincts de  $[a, b]$  et  $\{\alpha_i\}_{i=0\dots k}$   $(k+1)$  entiers naturels, on pose  $n = k + \alpha_0 + \dots + \alpha_n$ .

Étant donnée une fonction numérique  $f$  définie sur  $[a, b]$  admettant des dérivées d'ordre  $\alpha_i$  aux points  $x_i$ , il existe un unique polynôme  $p_n \in \mathcal{P}_n$  tel que

$$\forall (i, l), 0 \leq i \leq k, 0 \leq l \leq \alpha_i, \quad p_n^{(l)}(x_i) = f^{(l)}(x_i), \quad (1.3.1)$$

où  $f^{(l)}(x_i)$  désigne la dérivée d'ordre  $l$  de la fonction  $f$  au point  $x_i$ . □

*Preuve :* Les équations (1.3.1) forment un système linéaire de  $(n+1)$  équations à  $(n+1)$  inconnues. Pour prouver le théorème 1.3.1, il suffit de considérer le système linéaire homogène associé et de montrer qu'il n'a que le vecteur nul pour solution, c'est à dire que

$$p_n \in \mathcal{P}_n, \forall (i, l), 0 \leq i \leq k, 0 \leq l \leq \alpha_i, \quad p_n^{(l)}(x_i) = 0, \quad (1.3.2)$$

entraîne que  $p_n \equiv 0$ .

Les équations (1.3.2) entraîne que, pour  $i \in \{0, \dots, k\}$ , le point  $x_i$  est une racine d'ordre  $\alpha_i + 1$  de  $p_n$ . Le polynôme  $p_n$  peut donc s'écrire :

$$p_n = q(x) \prod_{i=0}^k (x - x_i)^{\alpha_i + 1}, \quad (1.3.3)$$

où  $q$  est un polynôme. Comme  $\sum_{i=0}^k (\alpha_i + 1) = n + 1$  et le degré de  $p_n$  est inférieur ou égale à  $n$ , nécessairement  $q \equiv 0$ , donc  $p_n \equiv 0$ . □

**Définition 1.3.1** Le polynôme  $p_n$  ainsi défini s'appelle le *polynôme d'interpolation de Hermite* associé aux points  $\{x_i\}_{i=0\dots k}$  et aux entiers  $\{\alpha_i\}_{i=0\dots k}$ . □

**Définition 1.3.2** On appelle *polynômes de base de Hermite* relatifs aux points  $\{x_i\}_{i=0\dots k}$  et aux entiers  $\{\alpha_i\}_{i=0\dots k}$  les polynômes appartenant  $\mathcal{P}_n$  définis de la manière suivante :

$$\forall (i, l), 0 \leq i \leq k, 0 \leq l \leq \alpha_i, \begin{cases} p_{il}(x_i) = 1 \\ p_{jm}(x_i) = 0, \forall (j, m) \neq (i, j) \end{cases} \quad (1.3.4)$$

□

A partir de ces polynômes, le polynôme d'interpolation de Hermite s'écrit

$$p_n(x) = \sum_{i=0}^k \sum_{l=0}^{\alpha_i} f^{(l)}(x_i) p_{il}(x) \quad (1.3.5)$$

**Construction des polynômes de base de Hermite** Posons

$$q_i(x) = \prod_{j=0, j \neq i}^k \left( \frac{x - x_j}{x_i - x_j} \right)^{\alpha_j + 1}, \quad (1.3.6)$$

on vérifie que les polynômes de base sont alors définis par les relations de récurrence suivantes :

$$\text{pour } \alpha_i, \quad p_{i\alpha_i}(x) = \frac{(x - x_i)^{\alpha_i}}{\alpha_i - i!} q_i(x) \quad (1.3.7)$$

et <sup>2</sup>

$$\text{pour } l = \alpha_i - 1, \dots, 1, 0, \quad p_{il}(x) = \frac{(x - x_i)^l}{l!} q_i(x) - \sum_{j=l+1}^{\alpha_i} C_j^l q_i^{(j-l)}(x_i) p_{ij}(x) \quad (1.3.8)$$

**Remarques** L'interpolation de Lagrange est un cas particulier de l'interpolation de Hermite pour  $k = 0$  et  $\alpha_i = 0, i = 1, \dots, k$

**Majoration de l'erreur d'interpolation**

**Théorème 1.3.2** On suppose que  $f \in \mathcal{C}^{n+1}([a, b])$ , alors

$$\forall x \in [a, b], \exists \xi_x \in ]\min(x, x_i), \max(x, x_i)[ \quad (1.3.9)$$

$$\text{tel que } f(x) - p_n(x) = \frac{1}{(n+1)!} \pi_{n+1}(x) f^{(n+1)}(\xi_x) \quad (1.3.10)$$

où l'on a posé

$$\pi_{n+1} = \prod_{j=0}^k (x - x_j)^{\alpha_j + 1} \quad (1.3.11)$$

□

*Preuve* : Voir (CROUZEIX & MIGNOT, 1992).

□

---

2.  $C_n^p = \frac{n!}{p!(n-p)!}$

## 1.4 Interpolation par fonctions splines

Les inconvénients de l'interpolation de Lagrange sont les suivants:

- manque de convergence de l'interpolation lorsque l'on augmente le nombre de point, ce qui se traduit le plus souvent par de fortes oscillations du polynôme d'interpolation,
- l'approximation n'est pas dérivable si l'on effectue une interpolation par morceaux.

Pour palier à ces deux inconvénients, nous allons maintenant décrire une interpolation de Hermite par morceaux bien particulière. C'est l'interpolation par *fonctions splines*.

On se bornera à la présentation des splines cubiques; pour plus de détails, on se reportera à (SCHATZMAN, 2001) et (QUARTERONI, 2000).

### 1.4.1 Principes

On se donne une nouvelle fois un ensemble de  $n + 1$  points distincts d'un intervalle  $[a, b] \subset \mathbb{R}$  et une fonction  $f : [a, b] \mapsto \mathbb{R}$ . On note  $f(x_i) = y_i$ .

Le principe de l'interpolation par splines est simple. L'idée est d'utiliser une interpolation par morceaux sur chaque subdivision  $[x_i, x_{i+1}]$  et de forcer le raccord des dérivées aux points  $x_i$ .

Plus précisément, donnons la définition suivante :

**Définition 1.4.1** Une fonction  $\sigma$  est appelée une spline d'interpolation relative aux points  $\{x_i, 0 \leq i \leq n\}$  si elle vérifie :

1.  $\sigma$  est de classe  $\mathcal{C}^1$  sur  $[a, b]$ , donc

$$\sigma'(x_i^+) = \sigma'(x_i^-), \forall i = \{1, \dots, n-1\}, \quad (1.4.1)$$

2.  $\sigma$  est de classe  $\mathcal{C}^\infty$  sur  $[x_i, x_{i+1}], \forall i = \{0, \dots, n-1\}$ ,
3.  $\sigma$  coïncide avec  $f$  aux points  $x_i$  soit

$$\forall i = \{0, \dots, n\}, \quad \sigma(x_i) = y_i \quad (1.4.2)$$

□

**Remarque** On peut étendre cette définition en supposant que  $\sigma$  est de classe  $\mathcal{C}^k$  sur  $[a, b]$ . On dit alors que  $\sigma$  est une spline d'ordre  $k$ . Cette définition ne détermine pas de manière unique une fonction spline pour la donnée des couples  $\{(x_i, y_i)\}$ . Pour cela, il faut ajouter une condition supplémentaire. Cette condition est la suivante :

**Définition 1.4.2** On dit qu'une spline  $\sigma$  associée aux points  $\{(x_i, y_i)\}$  a des oscillations qui sont minimales si

$$\int_a^b \sigma''(t)^2 dt \text{ est minimale.} \quad (1.4.3)$$

□

**Remarque** Le terme “spline” signifie “latte” en anglais. Une latte de bois souple était en effet utilisée par les dessinateurs industriels pour tracer des courbes dérivables passant par des points donnés (le dessin des congés par exemple). La quantité intégrée dans (1.4.3) peut être vue comme l'énergie élastique de flexion de la latte.

### 1.4.2 Caractérisation des splines cubiques

Le théorème suivant permet de caractériser une spline particulière :

**Théorème 1.4.1** Il existe une unique spline d'interpolation dont les oscillations sont minimales. Cette spline vérifie :

1. la restriction de  $\sigma$  à  $[x_i, x_{i+1}]$  est un polynôme de degré  $\leq 3$ ,  $\forall i = \{0, \dots, n-1\}$ ,
2.  $\sigma(x_0) = y_0, \sigma(x_n) = y_n$ ,
3.  $\sigma(x_i^+) = \sigma(x_i^-) = y_i, \forall i = \{1, \dots, n-1\}$ ,
4.  $\sigma'(x_i^+) = \sigma'(x_i^-), \forall i = \{1, \dots, n-1\}$ ,
5.  $\sigma''(x_0) = M_0, \sigma''(x_n) = M_n, (M_0, M_n) \in \mathbb{R}^2$  choisis arbitrairement,
6.  $\sigma''(x_i^+) = \sigma''(x_i^-), \forall i = \{1, \dots, n-1\}$ .

□

*Preuve* : Voir (SCHATZMAN, 2001).

□

Cette spline d'interpolation est communément appelée “spline cubique”.

### 1.4.3 Construction d'une spline d'interpolation

Posons  $M_i = \sigma''(x_i)$ . Sachant que  $\sigma \in \mathcal{P}_3$ , on a  $\sigma'' \in \mathcal{P}_1$ , et donc la formule d'interpolation de Lagrange nous donne :

$$\sigma''(x) = M_i \frac{x - x_{i+1}}{x_i - x_{i+1}} + M_{i+1} \frac{x - x_i}{x_{i+1} - x_i} \quad (1.4.4)$$

Le pas de la subdivision d'interpolation sera noté  $h_i = x_{i+1} - x_i$ .

En intégrant deux fois sur l'intervalle  $[x_i, x_{i+1}]$ , on obtient :

$$\sigma(x) = -M_i \frac{(x - x_{i+1})^3}{6h_i} + M_{i+1} \frac{(x - x_i)^3}{6h_i} + C_i(x - x_i) + \hat{C}_i \quad (1.4.5)$$

où  $C_i$  et  $\hat{C}_i$  sont deux constantes d'intégration déterminées par  $\sigma(x_i) = y_i$  et  $\sigma(x_{i+1}) = y_{i+1}$ , soit :

$$\hat{C}_i = y_i + M_i \frac{(x_i - x_{i+1})^3}{6h_i} = y_i - M_i \frac{(h_i)^2}{6} \quad (1.4.6)$$

$$C_i = \frac{y_{i+1} - y_i}{h_i} - \frac{h_i}{6}(M_{i+1} - M_i) \quad (1.4.7)$$

Sur l'intervalle  $[x_i, x_{i+1}]$ , la spline est alors donnée par :

$$\begin{aligned} \sigma(x) = & \frac{1}{6} \frac{(x_{i+1} - x)(x - x_i)(x - 2x_{i+1} + x_i)}{h_i} M_i \\ & - \frac{1}{6} \frac{(x_{i+1} - x)(x - x_i)(x - 2x_i + x_{i+1})}{h_i} M_{i+1} \\ & + \frac{y_{i+1} - y_i}{h_i} x + \frac{y_i x_{i+1} - x_i y_{i+1}}{h_i} \end{aligned} \quad (1.4.8)$$

et ses dérivées aux bords de l'intervalle :

$$\sigma'(x_i) = -\frac{1}{3} M_i (x_{i+1} - x_i) - \frac{1}{6} M_{i+1} (x_{i+1} - x_i) + \frac{y_{i+1} - y_i}{h_i} \quad (1.4.9)$$

$$\sigma'(x_{i+1}) = -\frac{1}{6} M_i (x_{i+1} - x_i) + \frac{1}{3} M_{i+1} (x_{i+1} - x_i) + \frac{y_{i+1} - y_i}{h_i}. \quad (1.4.10)$$

Le raccord des dérivées premières au point  $x_i$  impose :

$$\frac{1}{6} M_{i+1} (h_i) + \frac{1}{6} M_{i-1} (h_{i-1}) + \frac{1}{3} M_i (h_i + h_{i-1}) = \frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}} \quad (1.4.11)$$

Pour simplifier, on suppose que le pas de la subdivision est constant, c'est à dire

$$h = x_{i+1} - x_i = \frac{b - a}{n}, \forall i = \{0, \dots, n - 1\} \quad (1.4.12)$$

Le système d'équations (1.4.11) est alors équivalent au système tridiagonale suivant :

$$\begin{bmatrix} 2/3 & 1/6 & & & & \\ 1/6 & 2/3 & 1/6 & & & \\ & 1/6 & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ (0) & & & \ddots & \ddots & 1/6 \\ & & & & 1/6 & 2/3 \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_{n-1} \end{bmatrix} = \frac{1}{h^2} \begin{bmatrix} y_2 - 2y_1 + y_0 + M_0 h^2 \\ y_3 - 2y_2 + y_1 \\ \vdots \\ \vdots \\ y_{n-1} - 2y_{n-2} + y_{n-3} \\ y_n - 2y_{n-1} + y_{n-2} + M_n h^2 \end{bmatrix} \quad (1.4.13)$$

**Méthode de construction** La méthode de calcul des splines cubiques revient donc à écrire ce système tridiagonale, à le résoudre et à réinjecter les valeurs trouvées pour  $M_i$  dans l'expression de la spline donnée par l'équation (1.4.8).

**Remarque** Une seconde méthode de construction des splines peut être utilisée en considérant une base de l'ensemble de fonctions splines définies par le théorème 1.4.1. On considère pour cela  $n + 3$  fonctions de base  $\varphi_i$  de support  $[a, b]$  définies de la manière suivante :

$$\varphi_i(x_j) = \delta_{ij}, \quad \varphi'_i(x_0) = \varphi'_i(x_n) = 0, \quad \forall i, j = \{0, \dots, n\} \quad (1.4.14)$$

$$\varphi_{n+1}(x_j) = 0, \quad \varphi_{n+2}(x_j) = 0, \quad \forall j = \{0, \dots, n\} \quad (1.4.15)$$

$$\varphi'_{n+1}(x_0) = 1, \quad \varphi'_{n+1}(x_n) = 0, \quad (1.4.16)$$

$$\varphi'_{n+2}(x_0) = 0, \quad \varphi'_{n+2}(x_n) = 1 \quad (1.4.17)$$

La spline cubique est alors donnée par la formule suivante :

$$\sigma(x) = \sum_{i=0}^n ky_i \varphi_i(x) + y'_0 \varphi_{n+1}(x) + y'_n \varphi_{n+2}(x) \quad (1.4.18)$$

où  $y'_0, y'_n$  sont deux valeurs données.

#### 1.4.4 Un résultat de convergence pour les splines cubiques

**Théorème 1.4.2** Soit une fonction  $f \in C^4([a, b])$ , on note  $h = \max_i h_i$  et  $\beta = h / \min_i h_i$ . Soit  $\sigma$  la spline cubique interpolant  $f$ ; alors

$$\|f^{(r)} - \sigma^{(r)}\|_\infty \leq C_r h^{4-r} \|f^{(4)}\|_\infty, \quad r = 0, \dots, 3 \quad (1.4.19)$$

avec

$$C_0 = 5/384, \quad C_1 = 1/24, \quad C_2 = 3/8, \quad C_3 = (\beta + \beta^{-1})/2 \quad (1.4.20)$$

□

*Preuve* : Nous renvoyons à (SCHUMAKER, 1981).

□

Ce théorème permet de conclure que  $\sigma$  et ses dérivées première et seconde convergent uniformément vers  $f$  et ses dérivées quand  $h$  tend vers 0.

## 1.5 Existence d'un polynôme de meilleure approximation.

Comme on l'a dit dans l'introduction de ce chapitre, une autre méthode pour faire l'approximation d'une fonction numérique consiste à chercher un polynôme "proche" de la fonction au sens d'une norme donnée. Dans ce type de méthode, on relaxe la condition de faire coïncider le polynôme et la fonction en un nombre de points donnés au profit d'une minimisation de la distance entre le graphe du polynôme et de la fonction.

### 1.5.1 Cas général

Plaçons nous dans ce paragraphe dans un cadre relativement général. On considère un espace vectoriel  $\mathcal{E}$  de fonction définies sur un sous ensemble de  $\mathbb{R}$ , muni d'une norme  $\|\cdot\|_{\mathcal{E}}$ . On suppose, de plus, que  $\mathcal{P}_n \subset \mathcal{E}$ .

On a le théorème fondamental suivant :

**Théorème 1.5.1** Pour toute fonction  $f \in \mathcal{E}$ , il existe au moins un polynôme  $p_n \in \mathcal{P}_n$  tel que

$$\|f - p_n\|_{\mathcal{E}} = \inf_{q \in \mathcal{P}_n} \|f - q\|_{\mathcal{E}} \quad (1.5.1)$$

□

*Preuve* : voir (CROUZEIX & MIGNOT, 1992).

□

**Définition 1.5.1** Un polynôme  $p_n$  vérifiant (1.5.1) est appelé *polynôme de meilleure approximation* de  $f$  dans  $\mathcal{E}$  par un polynôme de degré  $n$ .

□

### 1.5.2 Exemple classiques d'espace de fonction $\mathcal{E}$

**Meilleure approximation**  $\mathcal{L}^p, 1 \leq p \leq \infty$ ,

Dans ce cas, l'espace de fonction  $\mathcal{E}$  choisi est l'ensemble  $\mathcal{L}^p(a, b)$  des fonctions mesurables  $f$  sur  $(a, b)$  telles que :

$$\|f\|_{\mathcal{L}^p(a,b)} = \left( \int_a^b |f(x)|^p, dx \right)^{1/p} < \infty, \quad \text{pour } p \neq \infty \quad (1.5.2)$$

$$\|f\|_{\mathcal{L}^\infty(\Omega)} = \sup_{x \in \omega} \text{ess.}|f(x)| < \infty \quad (1.5.3)$$

**Théorème 1.5.2** Pour  $p < +\infty$  et  $[a, b]$  un intervalle fermé borné de  $\mathbb{R}$ , la suite  $p_n$  de polynôme de meilleure approximation converge vers  $f$  soit ,

$$\forall p, 1 \leq p < +\infty, \quad \lim_{n \rightarrow +\infty} \|f - p_n\|_{\mathcal{L}^p} = 0 \quad (1.5.4)$$

*Preuve* : Ceci est une conséquence du théorème de Weierstrass. La condition sur l'intervalle  $[a, b]$  est essentielle. Pour plus de détails, voir (CROUZEIX & MIGNOT, 1992).

□

□

Ce théorème permet de montrer qu'en augmentant le degré du polynôme, on est assuré de la convergence uniforme du procédé d'approximation.

Signalons tout de même que l'unicité n'est pas toujours garantie. Considérons les exemples suivants :

1. Soit  $\mathcal{E} = \mathcal{L}^1(-1, 1)$ , on considère la fonction définie par :

$$\begin{cases} f(x) = 1, & x > 0 \\ f(x) = -1, & x \leq 0 \end{cases} \quad (1.5.5)$$

On vérifie facilement que  $f \in \mathcal{L}^1(-1, 1)$  et que  $\forall \alpha \in [-1, 1]$ ,  $p_0(x) = \alpha$  est un polynôme de meilleure approximation de  $f$  dans  $\mathcal{L}^1(-1, 1)$  de degré 0 et que  $\|f - p_0\|_{\mathcal{L}^1(-1, 1)} = 2$ . Cet exemple est illustré à la figure 1.4(a).

2. Soit  $\mathcal{E} = \mathcal{L}^\infty(-1, 1)$ , considérons une nouvelle fois la fonction  $f$  définie par (1.5.5). Le polynôme  $p_1(x) = \alpha x$ ,  $\forall \alpha \in [0, 2]$  réalise la meilleure approximation de  $f$  dans  $\mathcal{L}^\infty(-1, 1)$  de degré 1. Notons de plus que  $\|f - p_1\|_{\mathcal{L}^\infty(-1, 1)} = 1$ . Cet exemple est illustré à la figure 1.4(b).

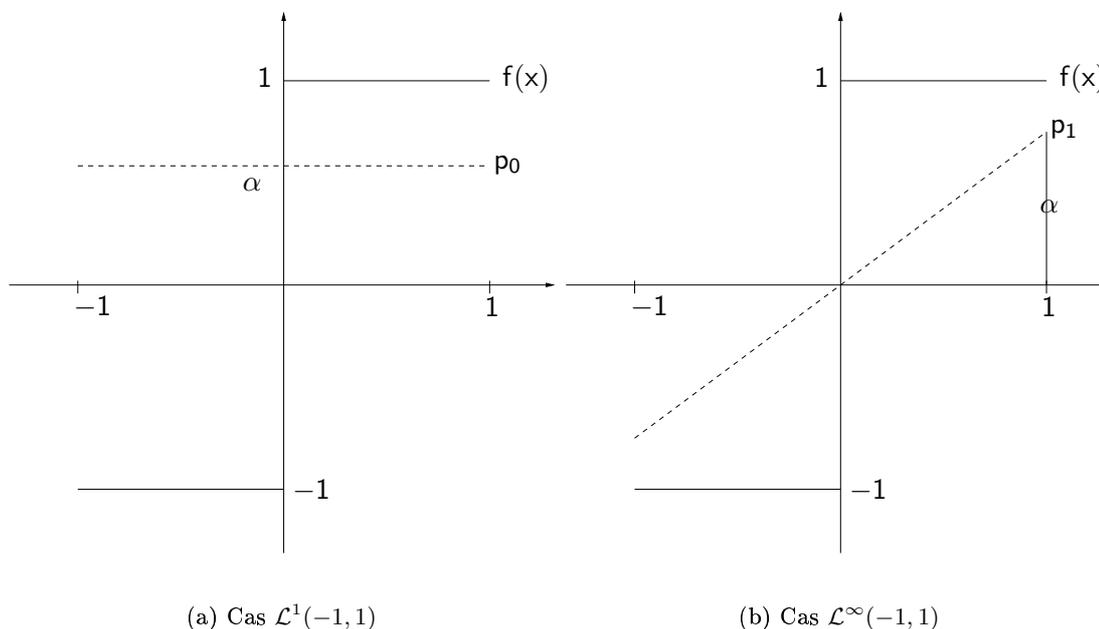


FIG. 1.4 – Contre-exemple à l'unicité de la meilleure approximation

Pour  $p = 1$ , on obtient un cas particulier très intéressant dans la pratique. On le connaît mieux sous le nom de meilleure approximation au sens de moindres carrés. Nous reviendrons en détail sur ce cas au § 1.6.

**Meilleure approximation uniforme (au sens de Chebyshev)**  $\mathcal{E} = \mathcal{C}^0([a, b])$  muni de  $\|\cdot\|_\infty$

La meilleure approximation uniforme, dite aussi au sens de Chebyshev est réalisée en choisissant pour l'espace de fonction  $\mathcal{E}$  l'ensemble de fonctions continues sur un intervalle  $[a, b]$ , soit  $\mathcal{C}^0([a, b])$  muni de la norme de la convergence uniforme  $\|\cdot\|_\infty$ .

Nous n'entrerons pas dans le détail de ce type d'approximation. Donnons simplement les résultats suivants :

**Théorème 1.5.3** Pour tout  $f \in \mathcal{C}^0([a, b])$ , il existe un unique polynôme  $q_n \in \mathcal{P}_n$  qui réalise le minimum de distance suivant :

$$\|f - q_n\|_\infty = d(f, \mathcal{P}_n) = \inf_{q \in \mathcal{P}_n} \|f - q\|_\infty \quad (1.5.6)$$

Si de plus  $[a, b]$  est un intervalle borné de  $\mathbb{R}$  alors

$$\lim_{n \rightarrow +\infty} \|f - p_n\|_\infty = 0 \quad (1.5.7)$$

□

## 1.6 Meilleure approximation polynomiale au sens des moindres carrés

### 1.6.1 Cas hilbertien. Polynômes orthogonaux

Dans ce paragraphe, nous nous plaçons dans le cadre général où  $\mathcal{E}$  est un espace de Hilbert de fonctions définies sur un sous-ensemble de  $\mathbb{R}$  qui inclut  $\mathcal{P}_n$ .

**Rappels – Espace de Hilbert** Un *espace de Hilbert* est un espace vectoriel normé complet dont la norme  $\|\cdot\|$  est déduite d'un produit scalaire  $\langle \cdot, \cdot \rangle$  de la manière suivante:

$$\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle} \quad (1.6.1)$$

On rappelle qu'une *norme* sur un espace vectoriel  $\mathcal{E}$  est une application  $\|\cdot\| : \mathcal{E} \rightarrow \mathbb{R}$  qui vérifie les trois propriétés suivantes :

(N1)  $\|x\| \geq 0$  &  $\|x\| = 0 \Leftrightarrow x = 0$

(N2)  $\|\lambda \cdot x\| = |\lambda| \|x\|$

(N3)  $\|x + y\| \geq \|x\| + \|y\|$  (inégalité du triangle)

Un *produit scalaire* est une application  $\langle \cdot, \cdot \rangle : \mathcal{E} \times \mathcal{E} \rightarrow \mathbb{R}$  qui vérifie :

(PS1)  $\langle x, x \rangle \geq 0$  &  $\langle x, x \rangle = 0 \Leftrightarrow x = 0$  (définie positive)

(PS2)  $\langle x, y \rangle = \langle y, x \rangle$  (symétrique)

(PS3)  $\langle \lambda x_1 + \mu x_2, y \rangle = \lambda \langle x_1, y \rangle + \mu \langle x_2, y \rangle$  (linéaire)

**Polynômes orthogonaux**

**Théorème 1.6.1** Il existe une suite de polynômes  $p_0, p_1, \dots, p_n$  et une seule vérifiant :

$$\begin{cases} \text{le degré de } p_n = n \\ p_n \text{ est monique (i.e, le coefficient de } x^n \text{ est 1)} \\ \forall q \in \mathcal{P}_{n-1}, \quad \langle q, p_n \rangle = 0 \end{cases} \quad (1.6.2)$$

□

*Preuve* : La preuve de l'existence est basée sur le procédé d'orthogonalisation de Gram–Schmidt sur la base canonique des polynômes,  $1, x, x^2, \dots, x^n$ . Pour plus de détails voir (CROUZEIX & MIGNOT, 1992).

□

Considérons les polynômes :

$$\hat{p}_n = \frac{p_n}{\|p_n\|} \quad (1.6.3)$$

Les polynômes  $\{\hat{p}_n\}$  forment une base orthonormée de l'espace de polynômes.

**Définition 1.6.1** Les polynômes  $\{p_n\}, \{\hat{p}_n\}$  s'appellent les *polynômes orthogonaux* (respectivement orthonormaux) pour le produit scalaire  $\langle \cdot, \cdot \rangle$ .

□

**Approximation dans une base de polynômes orthogonaux** L'intérêt des polynômes orthogonaux dans les espaces de Hilbert réside dans le théorème suivant :

**Théorème 1.6.2** Il existe un unique polynôme  $q_n \in \mathcal{P}_n$  tel que :

$$\|f - q_n\| = \inf_{q \in \mathcal{P}_n} \|f - q\| \quad (1.6.4)$$

Ce polynôme est donné par

$$q_n(x) = \sum_{i=0}^n \frac{\langle f, p_i \rangle}{\langle p_i, p_i \rangle} p_i(x) \quad (1.6.5)$$

et c'est l'unique polynôme tel que :

$$\forall q \in \mathcal{P}_n, \langle q, q_n \rangle = \langle f, q \rangle \quad (1.6.6)$$

□

*Preuve* : voir (CROUZEIX & MIGNOT, 1992).

□

Ce théorème donne l'unicité du polynôme de meilleure approximation et sa construction par projection sur la base de polynômes orthogonaux.

### 1.6.2 Approximation continue au sens de moindres carrés.

**Espace de fonction  $\mathcal{L}_\omega^2$**  Choisissons maintenant pour espace de fonction, l'ensemble de fonction de carrés mesurables, c'est à dire :

$$\mathcal{L}^2 = \{f \text{ tel que } \|f\|_{\mathcal{L}^2(a,b)} = \left( \int_a^b |f(x)|^2 dx \right)^{1/2} < \infty\} \quad (1.6.7)$$

Pour plus de généralité, on ajoute une fonction poids dans l'intégration  $\omega(x)$  dans l'intégration :

$$\mathcal{L}_\omega^2 = \{f \text{ tel que } \|f\|_{\mathcal{L}_\omega^2(a,b)} = \left( \int_a^b |f(x)|^2 \omega(x) dx \right)^{1/2} < \infty\} \quad (1.6.8)$$

On munit l'espace  $L_\omega^2$  d'un produit scalaire :

$$\langle f, g \rangle = \int_a^b f(x)g(x)\omega(x) dx \quad (1.6.9)$$

L'espace  $L_\omega^2$  est un espace de Hilbert où l'on peut construire une base de polynômes orthogonaux. Cette construction est basée sur le théorème suivant :

**Théorème 1.6.3** Pour  $\mathcal{E} = L_\omega^2$ , la suite de polynômes orthogonaux vérifie les équations de récurrence suivantes :

$$p_n(x) = (x - \lambda_n)p_{n-1}(x) - \mu_n p_{n-2}(x) \quad (1.6.10)$$

avec

$$\mu_n = \frac{\|p_{n-1}^2\|}{\|p_{n-2}^2\|}, \quad \lambda_n = \frac{\langle xp_{n-1}, p_{n-1} \rangle}{\|p_{n-1}^2\|} \quad (1.6.11)$$

□

Terminons enfin par un résultat qui nous permet de conclure sur la convergence uniforme des polynômes d'approximation vers la fonction  $f$  :

**Théorème 1.6.4** Si  $[a, b]$  est un intervalle borné, si  $q_n$  est le polynôme de meilleure approximation au sens de moindres carrés de degré  $n$  alors

$$\lim_{n \rightarrow +\infty} \|f - q_n\| = 0 \quad (1.6.12)$$

□

#### Exemples classiques de polynômes orthogonaux.

1.  $[a, b] = [-1, 1]$ ,  $\omega(x) = (1-x)^\alpha(1+x)^\beta$  avec  $\alpha > -1$  et  $\beta > -1$ . Les polynômes orthogonaux  $p_n$  s'appellent les *polynômes de Jacobi*. On a les cas particuliers suivants :

- (a)  $\alpha = \beta = 0$ , les polynômes s'appellent les *polynômes de Legendre*.

- (b)  $\alpha = \beta = -1/2$ , les polynômes s'appellent les *polynômes de Chebyshev de première espèce*.
- (c)  $\alpha = \beta = 1/2$ , les polynômes s'appellent les *polynômes de Chebyshev de seconde espèce*.
2.  $[a, b] = [0, +\infty]$ ,  $\omega(x) = \exp(-x)$ . Les polynômes orthogonaux  $p_n$  s'appellent les *polynômes de Laguerre*.
3.  $[a, b] = [-\infty, +\infty]$ ,  $\omega(x) = \exp(-x^2)$ . Les polynômes orthogonaux  $p_n$  s'appellent les *polynômes de Hermite*.

### 1.6.3 Approximation discrète – Méthodes des moindres carrés.

Lorsque la fonction  $f$  est donnée par un ensemble  $y_i = f(x_i)$  de  $(n+1)$  valeurs ou lorsqu'on donne simplement un "nuage de points"  $y_i, x_i$ , on ne peut calculer une norme continue entre un polynôme d'approximation et ces points. On préfère alors réaliser une approximation discrète des points aux moyens d'une méthode aux moindres carrés.

Le principe est toujours fonder sur une minimisation d'une distance entre ce nuage de points et un polynôme  $p_p$  de degré  $p$ .

Le produit scalaire se définit entre deux vecteurs  $x$  et  $y$  dans un cas plus général, si l'on affecte des poids  $\omega_i$  aux différents points :

$$\langle y, x \rangle = \sum_{i=0}^n \omega_i y_i x_i \quad (1.6.13)$$

La méthode des moindres carrés peut être vue comme une projection orthogonale ou alors comme une minimisation d'une distance, ce qui est bien sûr équivalent des que l'on dispose d'un produit scalaire. Choisissons de la présenter sous une approche de minimisation. Considérons la norme à minimiser entre les points et le polynôme :

$$S(a_0, a_1, \dots, a_n) = \sum_{i=0}^n \omega_i [y_i - (a_0 + a_1 x_i + \dots + a_n x_i^n)]^2 \quad (1.6.14)$$

Le minimum de cette norme est atteint lorsque les gradients de  $S$  par rapport à  $a_k$  s'annule, soit :

$$\frac{\partial S}{\partial a_k} = 0 \iff \sum_{i=0}^n 2\omega_i [y_i - (a_0 + a_1 x_i + \dots + a_n x_i^n)] (-x_i^k) = 0, \quad k \in 0, \dots, p \quad (1.6.15)$$

soit

$$\left[ \sum_{i=0}^n \omega_i x_i^k \right] a_0 + \left[ \sum_{i=0}^n \omega_i x_i^{k+1} \right] a_1 + \dots + \left[ \sum_{i=0}^n \omega_i x_i^{k+p} \right] a_p = \left[ \sum_{i=0}^n \omega_i y_i x_i^k \right], \quad k \in 0, \dots, p \quad (1.6.16)$$

On obtient donc un système de  $p$  équations à  $p$  inconnues suivant :

$$\begin{bmatrix} \sum_{i=0}^n \omega_i & \sum_{i=0}^n \omega_i x_i & \dots & \sum_{i=0}^n \omega_i x_i^p \\ \sum_{i=0}^n \omega_i x_i & \sum_{i=0}^n \omega_i x_i^2 & \dots & \sum_{i=0}^n \omega_i x_i^{p+1} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \sum_{i=0}^n \omega_i x_i^{2k} & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=0}^n \omega_i x_i^p & \sum_{i=0}^n \omega_i x_i^{p+1} & \dots & \sum_{i=0}^n \omega_i x_i^{2p} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ \vdots \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n \omega_i y_i \\ \sum_{i=0}^n \omega_i y_i x_i \\ \vdots \\ \vdots \\ \vdots \\ \sum_{i=0}^n \omega_i y_i x_i^p \end{bmatrix} \quad (16.17)$$

# Chapitre 2

## Intégration numérique

### Contenu du Chapitre 2

---

|            |   |           |
|------------|---|-----------|
| <b>2.1</b> | <b>Introduction</b>   | <b>28</b> |
| <b>2.2</b> | <b>Méthodes d'intégration numérique composées</b>             | <b>28</b> |
| 2.2.1      | Principe et exemples usuels                                   | 28        |
| 2.2.2      | Principe général de construction                              | 30        |
| 2.2.3      | Théorème de convergence des méthodes de quadrature composées. | 31        |
| 2.2.4      | Les méthodes de Newton-Cotes                                  | 32        |
| <b>2.3</b> | <b>Étude générale de l'erreur d'intégration numérique</b>     | <b>32</b> |
| 2.3.1      | Ordre et erreur d'une méthode d'intégration numérique         | 33        |
| 2.3.2      | Noyau et théorème de Peano                                    | 33        |
| 2.3.3      | Cas des méthodes de quadrature composées.                     | 36        |
| <b>2.4</b> | <b>Méthodes d'intégration numérique de Gauss</b>              | <b>38</b> |
| 2.4.1      | Principe  | 38        |
| 2.4.2      | Cas général   | 38        |
| 2.4.3      | Méthodes de Gauss usuelles                                    | 40        |
| 2.4.4      | Méthodes de Gauss-Lobatto                                     | 43        |

---

## 2.1 Introduction

Considérons un intervalle  $[a, b]$  de  $\mathbb{R}$ , une fonction  $f \in \mathcal{C}^0([a, b])$  et une fonction de poids  $\omega(x) > 0$  telles que :

$$f \omega \in \mathcal{L}^1(a, b). \quad (2.1.1)$$

Le but des méthodes d'intégration numérique est d'approcher l'intégrale suivante :

$$\int_a^b f(x)\omega(x) dx, \quad (2.1.2)$$

par une "formule de quadrature" de la forme :

$$\sum_{i=0}^k \lambda_i f(x_i), \text{ avec } \lambda_i, x_i \text{ réels.} \quad (2.1.3)$$

Les points  $x_i$  seront appelés les *points d'intégration* et les coefficients  $\lambda_i$  les *poids d'intégration* associés à ces points.

Nous considérerons principalement deux classes de méthodes d'intégration numérique :

1. les *méthodes d'intégration composées* lorsque  $[a, b]$  est borné et  $\omega(x) \equiv 1$  :

L'intervalle d'intégration  $[a, b]$  est alors subdivisé en  $k$  sous-intervalles  $[\alpha_i, \alpha_{i+1}]$ ,  $i = 0, \dots, k-1$  et l'on calcule chaque intégrale

$$\int_{\alpha_i}^{\alpha_{i+1}} f(x) dx, \quad (2.1.4)$$

en substituant  $f$  par un polynôme d'interpolation.

2. les *méthodes d'intégration de Gauss* pour des poids  $\omega(x) > 0$  et des intervalles  $[a, b]$  particuliers.

On approche  $f$  par un polynôme d'interpolation sur l'intervalle  $[a, b]$  en des points choisis de manière à obtenir une formule qui soit exacte pour des polynômes de degré le plus élevé possible. Les points d'interpolation sont alors les zéros des polynômes orthogonaux associés à une fonction de poids particulière.

Dans tout ce chapitre, on s'attachera avant tout à montrer comment peuvent être construites des formules de quadrature. Les questions de convergence des formules d'intégration numérique ne seront pas envisagées. Seuls quelques résultats seront donnés dans des cas simples. De même, il existe de nombreuses méthodes particulières d'intégration numérique dédiées à des cas particuliers. On pense par exemple aux fonctions périodiques ou aux intégrales indéfinies. Pour plus de détails, on renvoie aux ouvrages, (CROUZEIX & MIGNOT, 1992) et (DEMAILLY, 1996).

## 2.2 Méthodes d'intégration numérique composées

### 2.2.1 Principe et exemples usuels

Considérons un intervalle borné  $[a, b] \subset \mathbb{R}$ , nous cherchons à approcher l'intégrale :

$$\int_a^b f(x) dx. \quad (2.2.1)$$

La méthode consiste à décomposer l'intervalle  $[a, b]$  en  $k$  sous-intervalles  $[\alpha_i, \alpha_{i+1}]_{i=0, \dots, k-1}$  et à calculer chaque intégrale :

$$\int_{\alpha_i}^{\alpha_{i+1}} f(x) dx, \quad (2.2.2)$$

en substituant  $f$  par un polynôme d'interpolation.

Nous donnons dans les paragraphes qui suivent des exemples usuels de méthodes d'intégration numérique composées.

### 2.2.1.a Méthodes des rectangles

On choisit  $\xi_i \in [\alpha_i, \alpha_{i+1}]$ . On remplace  $f$  par le polynôme d'interpolation de Lagrange de degré 0 :

$$p_0(x) = f(\xi_i) \quad (2.2.3)$$

On a alors

$$\int_{\alpha_i}^{\alpha_{i+1}} f(x) dx \approx (\alpha_{i+1} - \alpha_i) f(\xi_i) \quad (2.2.4)$$

et par suite sur  $[a, b]$  :

$$\int_a^b f(x) dx \approx \sum_{i=0}^{k-1} (\alpha_{i+1} - \alpha_i) f(\xi_i) \quad (2.2.5)$$

qui est une somme de Riemann.

Les choix les plus courants pour  $\xi_i$  sont :

1.  $\xi_i = \alpha_i$ , on obtient la *méthode des rectangles à gauche* :

$$\int_a^b f(x) dx \approx \sum_{i=0}^{k-1} (\alpha_{i+1} - \alpha_i) f(\alpha_i) \quad (2.2.6)$$

2.  $\xi_i = \alpha_{i+1}$ , on obtient la *méthode des rectangles à droite* :

$$\int_a^b f(x) dx \approx \sum_{i=0}^{k-1} (\alpha_{i+1} - \alpha_i) f(\alpha_{i+1}) \quad (2.2.7)$$

### 2.2.1.b Méthode des trapèzes

Pour la *méthode des trapèzes*, on choisit le polynôme de Lagrange de degré 1 qui interpole  $f$  en  $\alpha_i$  et  $\alpha_{i+1}$ .

$$\int_{\alpha_i}^{\alpha_{i+1}} f(x) dx \approx \int_{\alpha_i}^{\alpha_{i+1}} p_1(x) dx = \frac{(\alpha_{i+1} - \alpha_i)}{2} [f(\alpha_i) + f(\alpha_{i+1})] \quad (2.2.8)$$

soit sur  $[a, b]$ ,

$$\int_a^b f(x) dx \approx T_k(f) = \sum_{i=0}^{k-1} \frac{(\alpha_{i+1} - \alpha_i)}{2} [f(\alpha_i) + f(\alpha_{i+1})] \quad (2.2.9)$$

### 2.2.1.c Méthode du point milieu

Pour la *méthode du point milieu*, on choisit le polynôme de Hermite de degré 1 au point milieu de l'intervalle, soit :

$$\alpha_{i+1/2} = \frac{(\alpha_{i+1} - \alpha_i)}{2}, \quad p_1(\alpha_{i+1/2}) = f(\alpha_{i+1/2}), \quad p_1'(\alpha_{i+1/2}) = f'(\alpha_{i+1/2}) \quad (2.2.10)$$

On a alors :

$$\int_{\alpha_i}^{\alpha_{i+1}} f(x) dx \approx \int_{\alpha_i}^{\alpha_{i+1}} p_1(x) dx = (\alpha_{i+1} - \alpha_i) f(\alpha_{i+1/2}) \quad (2.2.11)$$

### 2.2.2 Principe général de construction

Posons  $h_i = \alpha_{i+1} - \alpha_i$  et effectuons le changement de variable linéaire :

$$\nu = \frac{2x - \alpha_i - \alpha_{i+1}}{h_i}. \quad (2.2.12)$$

Si l'on note,

$$\varphi_i(\nu) = f(x) = f\left(\frac{\alpha_i + \alpha_{i+1} + \nu h_i}{2}\right), \quad (2.2.13)$$

l'intégrale de  $f$  peut alors s'écrire :

$$\int_{\alpha_i}^{\alpha_{i+1}} f(x) dx = \frac{h_i}{2} \int_{-1}^1 \varphi_i(\nu) d\nu \quad (2.2.14)$$

Grâce à ce simple changement de variable, on peut se ramener au calcul sur un intervalle fixe  $[-1, 1]$ .

Sur  $[-1, 1]$ , on se donne  $l+1$  points  $\{\tau_j\}_{j=0, \dots, l}$  et l'on choisit d'approcher  $\varphi$  par son polynôme d'interpolation de Lagrange  $p_\varphi$  de degré  $\leq l$  aux points  $\{\tau_j\}$ , soit :

$$\varphi(\nu) \approx p_\varphi(\nu) = \sum_{j=0}^l \varphi(\tau_j) L_j(\nu) \quad (2.2.15)$$

où la fonction  $L_j$  est la fonction de base de Lagrange :

$$L_j(\nu) = \prod_{k=0, k \neq j}^l \left( \frac{\nu - \tau_k}{\tau_l - \tau_k} \right) \quad (2.2.16)$$

L'intégrale de  $\varphi$  sur  $[-1, 1]$  est donc approchée par :

$$\boxed{\int_{-1}^1 \varphi(\nu) d\nu \approx 2 \sum_{j=0}^l \varphi(\tau_j) \omega_j \quad \text{avec } \omega_j = \frac{1}{2} \int_{-1}^1 L_j(\nu) d\nu} \quad (2.2.17)$$

En appliquant la formule (2.2.17) à  $\varphi_i$ , on obtient la formule d'intégration pour  $f$  :

$$\int_{\alpha_i}^{\alpha_{i+1}} f(x) dx \approx h_i \sum_{j=0}^l \omega_j f(\alpha_{ij}) \quad \text{avec } \alpha_{ij} = \frac{\alpha_i + \alpha_{i+1} + h_i \tau_j}{2} \quad (2.2.18)$$

On aurait pu traiter directement avec le polynôme d'interpolation de  $f$  aux points  $\alpha_{ij}$ . L'intérêt de cette écriture est de systématiser le calcul des poids  $\omega_j$  et de montrer qu'ils ne dépendent pas de  $i$ .

Si on revient à l'estimation de l'intégrale sur  $[a, b]$ , on obtient

$$\boxed{\int_a^b f(x) dx \approx \sum_{i=0}^k h_i \sum_{j=0}^l \omega_j f(\alpha_{ij}) \equiv T_{kl}(f)} \quad (2.2.19)$$

**Définition 2.2.1** La formule (2.2.17) s'appelle une *formule de quadrature élémentaire* et la formule (2.2.19) s'appelle une *formule de quadrature composée*

□

### 2.2.3 Théorème de convergence des méthodes de quadrature composées.

Un résultat de convergence des méthodes de quadrature composées est donné par le théorème suivant :

**Théorème 2.2.1** On suppose que la formule de quadrature élémentaire (2.2.17) est fixée et que  $\forall j \in \{0, \dots, l\}, \tau_j \in [-1; 1]$ ; alors, pour toute fonction  $f$  intégrable au sens de Riemann sur  $[a, b]$ , on a

$$T_{kl}(f) \text{ tends vers } \int_a^b f(x) dx \quad (2.2.20)$$

lorsque que la longueur  $\delta_k = \max_{0 \leq i \leq k} h_i$  tend vers 0.

□

*Preuve* : On a d'après (2.2.19)

$$T_{kl}(f) = \sum_{j=0}^l \omega_j I_{kj}(f) \text{ avec } I_{kj}(f) = \sum_{i=0}^k h_i f(\alpha_{ij}) \quad (2.2.21)$$

Comme  $\alpha_{ij} \in [\alpha_i, \alpha_{i+1}]$ , la quantité  $I_{kj}(f)$  est une somme de Riemann et donc

$$\lim_{\delta_k \rightarrow 0} I_{kj}(f) = \int_a^b f(x) dx \quad (2.2.22)$$

Il suffit enfin de remarquer que  $\sum_{j=0}^l \omega_j = 1$  pour terminer la preuve.

□

**Remarque** Les formules d'intégration composées sont exactes pour les polynômes de degré  $\leq l$ . Nous dirons dans la suite qu'elles sont d'ordre  $l$ .

**Remarque** D'une manière générale, il est impossible de démontrer la convergence des formules de quadrature élémentaire (2.2.17) lorsque  $l$  tend vers l'infini. Ceci n'est pas étonnant puisque l'interpolation de Lagrange ne converge pas forcément vers  $f$  en norme uniforme. Ceci justifie l'intérêt des formules de quadrature composées avec des interpolations polynomiales de degré faible. De plus, elle sont très intéressantes du point de vue de la mise en œuvre.

### 2.2.4 Les méthodes de Newton-Cotes

Dans la pratique, on choisit le plus souvent des points d'interpolations équidistants de la forme :

$$\tau_j = -1 + \frac{2}{l}j, \quad j = \{0, \dots, l\} \quad (2.2.23)$$

Les formules de quadrature élémentaire que l'on obtient, s'appellent les *formules d'intégration de Newton-Cotes*.

**Propriétés** Outre qu'elles soient exactes, par construction, pour les polynômes de degré  $\leq l$ , elles profitent de propriétés supplémentaires. Par raison de symétrie,  $\forall j \in \{0, \dots, l\}, \omega_{l_j} = \omega_j$ ; il en résulte que si  $\varphi$  est une fonction impaire sur  $[-1, 1]$ . Par conséquent, les formules de Newton-Cotes sont exactes :

- pour  $\varphi \in \mathcal{P}_l$  si  $l$  est impair,
- pour  $\varphi \in \mathcal{P}_{l+1}$  si  $l$  est pair.

#### Exemples usuels

1.  $l = 1$ ; on retrouve la méthode des trapèzes,
2.  $l = 2$ ; la *méthode de Simpson* est donnée par  $\omega_0 = \omega_2 = 1/6, \omega_1 = 2/3$ , soit

$$\int_{-1}^1 \varphi(\nu) d\nu = \frac{1}{3} [\varphi(-1) + 4\varphi(0) + \varphi(1)] \quad (2.2.24)$$

3.  $l = 4$ ; la *méthode de Boole-Villarceau* est donnée par  $\omega_0 = \omega_4 = 7/90, \omega_1 = \omega_3 = 16/45, \omega_2 = 2/15$ .

**Remarque** Les méthodes de Newton-Cotes sont utilisées jusqu'à  $l \leq 8$ . Au delà, les coefficients  $\omega_j$  deviennent "grands" et de signes mélangés ce qui rend les formules très sensibles aux erreurs d'arrondis.

## 2.3 Étude générale de l'erreur d'intégration numérique

L'étude générale de l'erreur d'intégration numérique est réalisée aux moyens du noyau de Peano d'une méthode d'intégration numérique. Nous ne donnerons pas le détail de cette méthode. Nous en rappellerons simplement les grandes lignes.

### 2.3.1 Ordre et erreur d'une méthode d'intégration numérique

**Définition 2.3.1** On appelle *ordre d'une méthode d'intégration* numérique, l'entier  $N$  tel que la méthode soit exacte pour tout polynôme de degré inférieur ou égale à  $N$ .

□

Considérons une méthode numérique d'intégration sous la forme suivante :

$$\int_{\alpha}^{\beta} f(x)\omega(x) dx \approx \sum_{i=0}^k \lambda_i f(x_i). \quad (2.3.1)$$

Considérons un intervalle  $[a, b]$  contenant les points  $\alpha, \beta$  ainsi que tous les points  $x_i$ . On définit l'erreur d'intégration de la manière suivante :

**Définition 2.3.2** L'erreur d'une méthode d'intégration numérique,  $E(f)$  se définit par :

$$E(f) = \int_{\alpha}^{\beta} f(x)\omega(x) dx - \sum_{i=0}^k \lambda_i f(x_i) \quad (2.3.2)$$

□

L'erreur d'une méthode d'intégration est une forme linéaire continue sur  $\mathcal{C}^0([a, b])$  pour la norme  $\|\cdot\|_{\infty}$

### 2.3.2 Noyau et théorème de Peano

Définissons maintenant ce que l'on appelle le noyau de Peano d'une méthode d'intégration numérique :

**Définition 2.3.3** Le noyau de Peano, noté  $K_N$ , d'une méthode d'intégration numérique d'ordre  $N$  donnée par la formule (2.3.1), est défini par :

$$K_N(t) = E(x \mapsto (x - t)_+^N) \quad (2.3.3)$$

où  $u_+$  représente la partie positive de  $u$ <sup>1</sup>.

□

Le théorème suivant dû à Peano donne l'erreur d'une méthode d'intégration numérique d'ordre  $N$  :

**Théorème 2.3.1** On suppose que le méthode numérique est d'ordre  $N$  et que  $f \in \mathcal{C}^{N+1}([a, b])$  alors

$$E(f) = \frac{1}{N!} \int_a^b K_N(t) f^{(N+1)}(t) dt \quad (2.3.4)$$

□

---

<sup>1</sup>.  $u_+ = \max(0, u)$  et avec le convention que  $u_+^0 = 1$  si  $u \geq 0$  et  $u_+^0 = 0$  si  $u < 0$

*Preuve* : La preuve est basée sur la formule de Taylor avec reste intégral de  $f$  à l'ordre  $N + 1$ , soit

$$\forall x \in [a, b], \quad f(x) = \sum_{k=0}^N \frac{1}{k!} f^{(k)}(a)(x-a)^k + \frac{1}{N!} \int_a^x (x-t)^N f^{(N+1)}(t) dt \quad (2.3.5)$$

qui peut s'écrire :

$$\forall x \in [a, b], \quad f(x) = q_N(x) + \frac{1}{N!} \int_a^b (x-t)_+^N f^{(N+1)}(t) dt \quad (2.3.6)$$

où  $q_N$  est un polynôme de degré  $\leq N$ .

Sachant que  $E(q_N(x)) = 0$ , l'erreur d'intégration s'évalue donc de la manière suivante :

$$E(f) = \frac{1}{N!} E \left( x \mapsto \int_a^b (x-t)_+^N f^{(N+1)}(t) dt \right) \quad (2.3.7)$$

En utilisant la linéarité de  $E$  et le théorème de Fubini; on établit le résultat du théorème. Pour plus de détails voir (CROUZEIX & MIGNOT, 1992). □

Deux corollaires peuvent être donnés à partir du théorème 2.3.1 :

**Corollaire 2.3.1** Sous les hypothèses du théorème 2.3.1, on a

$$|E(f)| \leq \frac{M_{N+1}}{N!} \int_a^b K_N(t) dt \quad (2.3.8)$$

avec

$$M_{N+1} = \|f^{(N+1)}\|_\infty \quad (2.3.9)$$

□

*Preuve* : La preuve est évidente en prenant la borne supérieure de  $f$  sur  $[a, b]$ . □

**Corollaire 2.3.2** Si l'on suppose de plus que  $K_N$  garde un signe constant sur  $[a, b]$ , alors :

$$\exists \xi \in [a, b], \quad E(f) = \frac{f^{(N+1)}(\xi)}{(N+1)!} E(x \mapsto x^{N+1}) \quad (2.3.10)$$

□

*Preuve* : Soit  $\omega$  une fonction positive intégrable sur  $] \alpha, \beta[$  tel que

$$\int_\alpha^\beta \omega(x) dx \quad (2.3.11)$$

converge. On a alors la formule de moyenne suivante :

$$\exists \xi \in [\alpha, \beta], \int_{\alpha}^{\beta} f(x)\omega(x) dx = f(\xi) \int_{\alpha}^{\beta} \omega(x) dx \quad (2.3.12)$$

En appliquant cette formule de moyenne pour  $\omega(x) = K_N(t)$  (ou  $\omega(t) = -K_N(t)$ ) si  $K_N$  est négatif), on obtient

$$E(f) = \frac{f^{(N+1)}(\xi)}{N!} \int_a^b K_N(t) dt \quad (2.3.13)$$

Cette dernière équation donne en particulier pour  $x \mapsto x^{N+1}$  :

$$E(x \mapsto x^{N+1}) = (N+1) \int_a^b K_N(t) dt, \quad (2.3.14)$$

qui combinée avec (2.3.13) donne le résultat. □

**Exemples de Noyau de Peano** Considérons la formule du rectangle :

$$\int_{\alpha}^{\beta} f(x) dx \approx (\beta - \alpha)f(\gamma), \gamma \in [\alpha, \beta] \quad (2.3.15)$$

dans le cas où  $\gamma$  n'est pas le point milieu. La méthode est d'ordre 0. Le noyau de Peano est alors

$$K_0(t) = E(x \mapsto (x-t)_+^0) \quad (2.3.16)$$

On a

$$\int_{\alpha}^{\beta} (x-t)_+^0 dt = \begin{cases} (\beta - \alpha) & \text{si } \alpha \geq t \\ (\beta - t) & \text{si } \alpha \leq t \leq \beta \\ 0 & \text{si } t \geq \beta \end{cases} \quad (2.3.17)$$

et

$$(\beta - \alpha)(\gamma - t)_+^0 = \begin{cases} \beta - \alpha & \text{si } t \leq \gamma \\ 0 & \text{si } t \geq \gamma \end{cases} \quad (2.3.18)$$

Le noyau de Peano de la méthode est donc donné pour  $\gamma \in [\alpha, \beta]$  par :

$$K_0(t) = \int_{\alpha}^{\beta} (x-t)_+^0 dt - (\beta - \alpha)(\gamma - t)_+^0 \quad (2.3.19)$$

$$= \begin{cases} 0 & \text{si } t \leq \alpha \\ \alpha - t & \text{si } \alpha \leq t \leq \gamma \\ t - \beta & \text{si } \gamma \leq t \leq \beta \\ 0 & \text{si } t \geq \beta \end{cases} \quad (2.3.20)$$

Il est représenté à la figure 2.1. Il est à noter que le noyau de Peano est pas de signe constant uniquement pour  $\gamma = \alpha$  et  $\gamma = \beta$

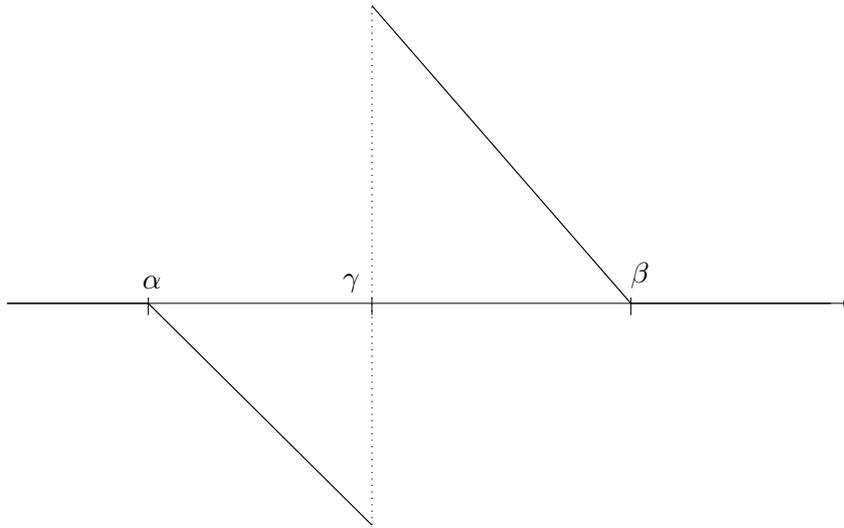


FIG. 2.1 – Noyau de Peano pour la méthode des rectangles

### 2.3.3 Cas des méthodes de quadrature composées.

Dans le cas des méthodes de quadrature composées, on a le théorème suivant :

**Théorème 2.3.2** On suppose que la formule de quadrature élémentaire est d'ordre  $N$  et que  $f \in \mathcal{C}^{(N+1)}([a, b])$ , alors

$$|E(f)| \leq C_N(\beta - \alpha)h^{N+1}M_{N+1} \quad (2.3.21)$$

avec

$$M_{N+1} = \|f^{(N+1)}\|_{\infty} \quad (2.3.22)$$

$$h = \max_{0 \leq i \leq n} h_i \quad (2.3.23)$$

$$C_N = \frac{1}{2^{2N+2}N!} \int_{-\infty}^{\infty} |k_N(t)| dt \quad (2.3.24)$$

où  $k_n$  est le noyau de Peano de la formule de quadrature élémentaire. □

*Preuve :* Voir (CROUZEIX & MIGNOT, 1992). □

Ce résultat est très intéressant car il donne la décroissance de l'erreur  $E(f)$  en fonction de la taille maximale de la subdivision  $h$ .

**Corollaire 2.3.3** Si l'on suppose de plus que le noyau de la formule élémentaire  $k_n$  garde un signe constant alors

$$\exists \xi \in [a, b], \quad E(f) = C_N^* f^{(N+1)}(\xi) \sum_{i=0}^k h_i^{(N+2)} \quad (2.3.25)$$

avec

$$C_N^* = \frac{1}{2^{2N+2}N!} \int_{-\infty}^{\infty} k_N(t) dt \quad (2.3.26)$$

$$= \frac{1}{2^{2N+2}(N+1)!} \left[ \int_{-1}^1 x^{N+1} dx - 2 \sum_{j=0}^l \omega_j \tau_j^{N+1} \right] \quad (2.3.27)$$

□

Le théorème suivant dû à Steffensen permet d'utiliser le dernier corollaire donnant la constante d'erreur pour des méthodes de Newton-Cotes :

**Théorème 2.3.3** Dans les cas des méthodes de Newton-Cotes, le noyau de Peano reste de signe constant.

□

**Exemples de calcul d'erreur** On donne dans le tableau 2.1 des estimations d'erreur pour les formules de Newton-Cotes.

| Méthode                 | $N$ | $C_N$           | $E(f)$   |
|-------------------------|-----|-----------------|--|
| Méthode du point milieu | 1   | $\frac{1}{3}$   | $\frac{1}{24}h^2(\beta - \alpha)f''(\xi)$        |
| Méthode des trapèzes    | 1   | $-\frac{2}{3}$  | $-\frac{1}{12}h^2(\beta - \alpha)f''(\xi)$       |
| Méthode de Simpson      | 3   | $-\frac{1}{15}$ | $-\frac{1}{2880}h^4(\beta - \alpha)f^{(4)}(\xi)$ |

TAB. 2.1 – Erreur d'intégration pour des formules de Newton-Cotes

## 2.4 Méthodes d'intégration numérique de Gauss

### 2.4.1 Principe

On considère une fois encore une formule de quadrature de la forme suivante :

$$\int_a^b f(x)\omega(x) dx \approx \sum_{i=0}^k \lambda_i f(x_i), \text{ avec } \lambda_i, x_i \text{ réels.} \quad (2.4.1)$$

Le principe des méthodes de Gauss consiste à déterminer les points d'intégration,  $x_i$ , appelés aussi points de Gauss, et les poids de l'intégration,  $\lambda_i$ , de sorte à maximiser l'ordre de la méthode d'intégration pour un nombre donné de points. A l'inverse des méthodes de quadrature composées, les points d'intégration ne sont pas donnés, *a priori*. Nous verrons que les théorèmes qui suivent nous montrent que ces points sont les racines des polynômes orthogonaux associés à une fonction de poids particulière.

### 2.4.2 Cas général

Commençons par un cadre assez général. On note  $\{x_0, \dots, x_{l-1}\}$ ,  $l$  points de l'intervalle  $[a, b]$  et  $\{x_l, \dots, x_k\}$  ( $k+1-l$ ) points extérieurs à  $[a, b]$ . On a le théorème suivant :

**Théorème 2.4.1** Étant donnés :

- deux entiers  $k$  et  $l$  tel que  $k+1 \geq l \geq 0$
- $k+1-l$  points extérieurs à  $[a, b]$ ,  $\{x_l, \dots, x_k\}$

Alors il existe une unique formule d'intégration de la forme (2.4.1) qui soit d'ordre  $k+l$ .

□

*Preuve : Unicité.* On suppose qu'il existe une formule d'intégration de la forme (2.4.1) qui soit d'ordre  $k+l$ . Considérons le polynôme suivant :

$$\pi_{k+1}(x) = \prod_{i=l}^k (x - x_i) \quad (2.4.2)$$

Ce polynôme garde un signe constant sur  $]a, b[$  puisque ces racines sont à l'extérieur. Considérons la fonction suivante :

$$\theta(x) = \varepsilon \omega(x) \pi_{k+1}(x) \quad (2.4.3)$$

En choisissant  $\varepsilon = \pm 1$ , le signe constant constant de  $\pi_{k+1}$  permet d'écrire que

$$\theta(x) > 0, \forall x \in ]a, b[ \quad (2.4.4)$$

et donc de considérer  $\theta$  comme une fonction de poids.

Soit le polynôme de degré inférieur ou égale à  $l$ ,

$$p_l(x) = \prod_{i=0}^{l-1} (x - x_i) \quad (2.4.5)$$

Pour polynôme  $q$  de degré inférieur ou égale à  $l - 1$ , on a

$$\int_a^b p_l(x)q(x)\theta(x) dx = \int_a^b q(x) \prod_{i=0}^k (x - x_i)\omega(x) dx = \sum_{i=0}^k \lambda_i q(x_i) \prod_{i=0}^k (x_i - x_i) = 0 \quad (2.4.6)$$

car la formule (2.4.1) est exacte pour le polynôme  $q(x) \prod_{i=0}^k (x_i - x_i)$ , qui est de degré  $\leq k + l - 1$ .

Pour résumé, on a la formule suivante :

$$\int_a^b p_l(x)q(x)\theta(x) dx = 0, \theta(x) > 0, \forall q \in \mathcal{P}_{l-1} \quad (2.4.7)$$

D'après (2.4.7), le polynôme  $p_l(x)$  est le  $(l + 1)$ ème polynôme orthogonal pour le poids  $\theta(x)$ . Il possède donc  $l$  racines distincts sur l'intervalle  $]a, b[$  que nous notons  $\{x_0, \dots, x_{l-1}\}$  et qui sont donc définies de manière unique.

En écrivant que la formule (2.4.1) est exacte pour les polynômes de base de Lagrange  $L_i$  définis sur les points  $\{x_0, \dots, x_{l-1}\}$ , on obtient que :

$$\lambda_i = \int_a^b L_i(x)\omega(x) dx \quad (2.4.8)$$

Les poids de la formule de quadrature sont donc définis de manière unique.

Les points d'intégration  $x_i$  ainsi que leurs poids associés  $\lambda_i$  étant définis de manière unique, on conclut à l'unicité de la méthode de la formule d'intégration.

*Existence.* Considérons à nouveau les  $l$  racines du  $(l + 1)$ ème polynôme orthogonal pour le poids  $\theta(x)$  sur  $]a, b[$ . Notons les  $\{x_0, \dots, x_{l-1}\}$ . Posons

$$\lambda_i = \int_a^b L_i(x)\omega(x) dx \quad (2.4.9)$$

on alors

$$\sum_{i=0}^k \lambda_i f(x_i) = \int_a^b p_f(x)\omega(x) dx \quad (2.4.10)$$

où  $p_f$  est le polynôme d'interpolation de  $f$  en  $x_i, 0 \leq i \leq k$ . La formule (2.4.1) est donc exacte pour les polynômes de degré  $\leq k$ .

Soit un polynôme  $\rho$  de degré  $k + l$ , en utilisant la division euclidienne des polynômes, il existe deux polynômes  $q$  de degré  $\leq l - 1$  et  $r$  de degré  $\leq k$  tel que

$$\rho(x) = q(x) \prod_{i=0}^k (x - x_i) + r(x) \quad (2.4.11)$$

Nous avons donc

$$\int_a^b \rho(x)\omega(x) dx = \int_a^b q(x) \prod_{i=0}^k (x - x_i)\omega(x) dx + \int_a^b r(x)\omega(x) dx \quad (2.4.12)$$

$$= \underbrace{\int_a^b q(x)p_l(x)\theta(x) dx}_{=0 \text{ (Orthogonalité)}} + \int_a^b r(x)\omega(x) dx \quad (2.4.13)$$

$$= \sum_{i=0}^k \lambda_i r(x_i) \text{ (Formule exacte pour } r) \quad (2.4.14)$$

$$= \sum_{i=0}^k \lambda_i f(x_i) \quad (2.4.15)$$

La formule est donc exacte pour les polynômes de degré  $k + l$ .

□

**Remarque 1** La démonstration du théorème précédent permet de construire la formule d'intégration numérique. Les principales étapes de la construction d'une formule d'intégration d'ordre  $k + l$  sont les suivantes :

1. Choisir  $k + l - 1$  points extérieurs à  $]a, b[$ , notés  $\{x_l, \dots, x_k\}$
2. Construire la fonction poids  $\theta(x)$  :

$$\theta(x) = \varepsilon\omega(x)\pi_{k+1}(x) \quad (2.4.16)$$

3. Déterminer le  $(l + 1)$ ème polynôme orthogonal  $p_l(x)$  par rapport à  $\theta$  sur  $[a, b]$ .
4. Déterminer les  $l$  racines de  $p_l(x)$  dans  $]a, b[$ , notées  $\{x_0, \dots, x_{l-1}\}$
5. Calculer les poids de la formule d'intégration :

$$\lambda_i = \int_a^b L_i(x)\omega(x) dx \quad (2.4.17)$$

**Remarque 2** On constate que la construction d'une formule d'intégration de Gauss peut s'avérer être assez compliquée. Pour des formules d'ordre faible ( $N < 10$ ), on préfère souvent trouver les points et les poids d'intégration en vérifiant l'ordre attendu de la méthode sur des polynômes de base.

Nous allons donner dans la suite deux exemples de méthodes de Gauss très utilisées pour leur relative simplicité.

### 2.4.3 Méthodes de Gauss usuelles

Les méthodes de Gauss usuelles sont construites en choisissant  $k = l - 1$ . C'est à dire que tous les points sont à l'intérieur de  $]a, b[$ . Nous avons le théorème suivant :

**Théorème 2.4.2** Il existe une unique formule d'intégration de la forme (2.4.1) (à  $k + 1$  points), d'ordre  $2k + 1$ ; elle obtenue en choisissant pour points d'intégration  $x_i$  les racines du  $(k + 2)$ ème polynôme orthogonal pour le poids  $\omega(x)$  sur  $(a, b)$  et en prenant pour poids d'intégration

$$\lambda_i = \int_a^b L_i(x)\omega(x) dx \quad (2.4.18)$$

De plus, on a que

$$\forall i = \{0, \dots, 2k + 1\}, \quad \lambda_i > 0 \quad (2.4.19)$$

□

*Preuve :* En utilisant le théorème 2.4.1, il ne reste qu'à démontrer que  $\lambda_i > 0$ . Pour cela, il suffit de considérer le carré des polynômes de base de Lagrange,  $L_i^2(x)$ . Ces polynômes sont de degré inférieur ou égal à  $2k$  et donc la formule d'intégration est exacte, soit :

$$\lambda_i = \int_a^b L_i^2(x)\omega(x) dx > 0 \quad (2.4.20)$$

□

Les formules ainsi obtenues s'appellent les *formules d'intégration de Gauss*

**Exemples de formules de Gauss usuelles** Suivant la fonction poids qui est choisie, on retrouve différentes formules d'intégration de Gauss. Les plus connues sont obtenues avec

- $\omega(x) = 1$  sur  $[a, b]$ ; il s'agit des *formules d'intégration de Gauss-Legendre*,<sup>2</sup>
- $\omega(x) = (1 - x^2)^{1/2}$  sur  $] - 1, 1[$ ; il s'agit des *formules d'intégration de Gauss-Chebyshev*,
- $\omega(x) = \exp(-x^2)$  sur  $] - \infty, +\infty[$ ; il s'agit des *formules d'intégration de Gauss-Hermite*,
- $\omega(x) = x^\alpha \exp(-x)$  sur  $]0, +\infty[$ ; il s'agit des *formules d'intégration de Gauss-Laguerre*,

La formule du point milieu correspond à la méthode de Gauss-Legendre pour  $k = 0$ .

**Formules explicites pour l'intégration de Gauss-Chebyshev** D'une manière générale, les points de Gauss,  $x_i$  et les coefficients  $\lambda_i$  ne sont pas donnés explicitement. Pour l'intégration de Gauss-Chebyshev, il est par contre possible de donner une formule explicite :

**Théorème 2.4.3** On montre que la méthode d'intégration de Gauss-Chebyshev s'écrit explicitement sous la forme suivante :

$$\int_{-1}^1 f(x)(1 - x^2)^{-1/2} dx \approx \frac{\pi}{k + 1} \sum_{i=0}^k f\left(\cos\left(\frac{(2i + 1)\pi}{2(k + 1)}\right)\right) \quad (2.4.21)$$

□

---

2. Ces formules sont parfois appelées les formules de Gauss lorsqu'il n'y a pas d'ambiguïté

*Preuve* : Les polynômes orthogonaux de Chebyshev sont donnés explicitement dans l'annexe B. Les racines du  $(k + 1)$ ème polynôme orthogonal sont

$$x_i = \cos\left(\frac{(2i + 1)\pi}{2(k + 1)}\right), \quad 0 \leq i \leq k \quad (2.4.22)$$

Il reste à démontrer que :

$$\lambda_i = \frac{\pi}{k + 1} \quad (2.4.23)$$

Pour cela, il faut démontrer les relations suivantes (en exercice) :

$$\lambda_i = \int_{-1}^1 \frac{T_{n+1}(x)(1 - x^2)^{-1/2}}{T'_{n+1}(x_i)(x - x_i)} dx, \quad x_i = \cos \theta_i, \quad \theta_i = \frac{(2i + 1)\pi}{2(k + 1)} \quad (2.4.24)$$

et en posant pour  $i$  fixé que

$$\alpha_i = \int_0^\pi \frac{\cos j\theta - \cos j\theta_i}{\cos \theta - \cos \theta_i} d\theta, \quad j = 0, 1, \dots \quad (2.4.25)$$

démontrer que

$$\alpha_{j+1} - 2 \cos(\theta_i) \alpha_j + \alpha_j = 0, \quad j = 1, 2, \dots \quad (2.4.26)$$

et calculer  $\alpha_n$ .

□

**Points de Gauss et coefficients pour la méthode de Gauss-Legendre** On donne dans le tableau suivant les points et les poids d'intégration pour la méthode de Gauss-Legendre.

| n | Points d'intégration $x_i$          | Poids d'intégration $\lambda_i$ | Ordre |
|---|-------------------------------------|---------------------------------|-------|
| 1 | 0                                   | 2                               | 1     |
| 2 | $\pm 1/\sqrt{3}$                    | 1                               | 3     |
| 3 | $\pm \sqrt{3/5}$                    | 5/9                             | 5     |
|   | 0                                   | 8/9                             |       |
| 4 | $\pm 1/35 \sqrt{525 - 70\sqrt{30}}$ | $1/36(18 + \sqrt{30})$          | 7     |
|   | $\pm 1/35 \sqrt{525 + 70\sqrt{30}}$ | $1/36(18 - \sqrt{30})$          |       |
| 5 | $\pm 1/21 \sqrt{245 - 14\sqrt{70}}$ | $1/900(322 + 13\sqrt{70})$      | 9     |
|   | $\pm 1/21 \sqrt{245 + 14\sqrt{70}}$ | $1/900(322 - 13\sqrt{70})$      |       |
|   | 0                                   | 128/225                         |       |

TAB. 2.2 – Points et poids de Gauss-Legendre

### 2.4.4 Méthodes de Gauss-Lobatto

Une seconde classe d'intégration numérique de Gauss est souvent rencontrée dans la pratique. Elle est donnée par le théorème suivant :

**Théorème 2.4.4** Il existe une unique formule d'intégration de la forme (2.4.1) (à  $k + 1$  points), d'ordre  $2k - 1$  telle que  $x_0 = a$  et  $x_k = b$ ; elle est obtenue en choisissant pour points d'intégration  $\{x_i\}_{0 \leq i \leq k-1}$  les racines du  $k$ ème polynôme orthogonal pour le poids  $\omega(x)(x - a)(b - x)$  sur  $]a, b[$  et en prenant pour poids d'intégration

$$\lambda_i = \int_a^b L_i(x)\omega(x) dx \quad (2.4.27)$$

De plus, on a que

$$\forall i \in \{0, \dots, 2k + 1\}, \quad \lambda_i > 0 \quad (2.4.28)$$

□

*Preuve :* Voir (CROUZEIX & MIGNOT, 1992).

□

Cette méthode est surtout utilisée avec  $\omega(x) = 1$ ; elle s'appelle alors la *formule d'intégration de Gauss-Lobatto*. Pour  $k = 1$ , on retrouve la méthode du trapèze et pour  $k = 2$  la méthode de Simpson. Cette méthode peut paraître moins avantageuse que les méthodes de Gauss-Legendre. Elle est surtout utilisée de manière composée. Elle donne dès lors des résultats comparables à la méthode de Gauss-Legendre pour un coût comparable.



## Chapitre 3

# Méthodes itératives pour la résolution d'équations non linéaires

### Contenu du Chapitre 3

---

|            |   |           |
|------------|---|-----------|
| <b>3.1</b> | <b>Introduction</b>   | <b>46</b> |
| <b>3.2</b> | <b>Une méthode topologique : la méthode de dichotomie</b>                                     | <b>46</b> |
| <b>3.3</b> | <b>Théorème du point fixe et applications contractantes</b>                                   | <b>47</b> |
| 3.3.1      | Applications contractantes et point fixe  | 47        |
| 3.3.2      | Théorème du point fixe  | 48        |
| 3.3.3      | Estimation de la vitesse de convergence   | 49        |
| 3.3.4      | Principe de construction des méthodes itératives métriques                                    | 49        |
| <b>3.4</b> | <b>Méthodes itératives métriques – Cas <math>f : \mathbb{R} \mapsto \mathbb{R}</math></b>     | <b>49</b> |
| 3.4.1      | Points fixes répulsifs et points fixes attractifs   | 50        |
| 3.4.2      | Méthode de la corde   | 51        |
| 3.4.3      | Méthode de Newton   | 52        |
| 3.4.4      | Méthode de la sécante de Steffensen   | 54        |
| <b>3.5</b> | <b>Méthodes itératives métriques – Cas <math>f : \mathbb{R}^n \mapsto \mathbb{R}^n</math></b> | <b>55</b> |
| 3.5.1      | Méthode Newton-Raphson  | 55        |
| <b>3.6</b> | <b>Cas particuliers où <math>f</math> est un polynôme – Racines d'un polynôme</b>             | <b>56</b> |
| 3.6.1      | Procédé de localisation de Sturm  | 56        |
| 3.6.2      | Méthode de Bairstow   | 57        |
| 3.6.3      | Méthode d'Aitken  | 59        |

---

### 3.1 Introduction

Dans ce chapitre, nous allons nous intéresser à la résolution d'équations du type :

$$f(x) = 0, \text{ avec } f : \mathbb{R}^n \mapsto \mathbb{R}^n \quad (3.1.1)$$

Naturellement, si  $f$  est une fonction linéaire, ce problème se ramène à la résolution d'un système linéaire homogène de dimension  $n$ . Dans ce chapitre, nous allons donner quelques méthodes de résolution de ces équations lorsque  $f$  n'est pas une fonction linéaire sur  $\mathbb{R}^n$ .

Il existe deux types de méthodes principales pour résoudre les systèmes d'équations non linéaires :

- Les méthodes topologiques. Dans ce type de méthode, on cherche à localiser une solution dans un ensemble. Plus la dimension de cet ensemble est petite, plus la solution est approchée de manière précise. La méthode de dichotomie est la méthode topologique la plus connue. Nous la présenterons au § Ces méthodes sont surtout employées lorsque  $f$  est fonction de  $\mathbb{R}$  dans  $\mathbb{R}$ . Pour des fonctions vectorielles, le nombre d'évaluation de  $f$  devient souvent prohibitif.
- Les méthodes métriques. Dans ce type de méthode, on cherche la solution comme la limite d'une suite le plus souvent donnée par le théorème du point fixe pour une fonction contractante. Le théorème du point fixe sera présenté de manière générale au § 3.3. Nous présenterons ensuite des méthodes itératives métriques dans le cas d'une fonction  $f : \mathbb{R} \mapsto \mathbb{R}$  au §

### 3.2 Une méthode topologique : la méthode de dichotomie

La méthode de dichotomie s'appuie sur les théorèmes des valeurs intermédiaires :

**Théorème 3.2.1** Soit un intervalle  $[a, b] \subset \mathbb{R}$  et une fonction  $f : [a, b] \mapsto \mathbb{R}$  continue sur  $[a, b]$  telle que :

$$f(a)f(b) < 0 \quad (3.2.1)$$

alors

$$\exists \alpha \in ]a, b[, f(\alpha) = 0 \quad (3.2.2)$$

□

**Principe** Le principe de la méthode de dichotomie consiste à fabriquer une suite d'intervalles  $I_n = [a_n, b_n]$  sur lesquels la fonction  $f$  vérifie la propriété (3.2.1). Cette suite d'intervalles est généralement construite par récurrence en découpant en deux (dichotomie) l'intervalle précédent et en vérifiant la propriété (3.2.1) au bord de chaque intervalle. La précision de l'évaluation d'une racine dépend alors de la longueur de l'intervalle.

**Algorithme 1** Méthode de dichotomie**Require:**  $I = [a, b]$  tel que  $f(a)f(b) < 0$ **Ensure:**  $\alpha \in [a_n, b_n]$  avec  $|a_n - b_n| < \varepsilon$ 

```

 $n \leftarrow 0$ 
 $a_0 \leftarrow a$ 
 $b_0 \leftarrow b$ 
while  $|a_n - b_n| > \varepsilon$  do
   $n \leftarrow n + 1$ 
   $c_n = \frac{a_{n-1} + b_{n-1}}{2}$ 
  if  $f(a_{n-1})f(c_n) > 0$  then
     $a_n \leftarrow c_n$ ;  $b_n \leftarrow b_{n-1}$ 
  end if
  if  $f(a_{n-1})f(c_n) < 0$  then
     $a_n \leftarrow a_{n-1}$ ;  $b_n \leftarrow c_n$ 
    if  $f(c_n)f(b_{n-1}) < 0$  then
      Il existe plusieurs racines sur  $[a, b]$ 
      Réinitialiser l'algorithme avec  $a = c_n$  et  $b = b_{n-1}$ 
    end if
  end if
end while

```

**Algorithme.** L'algorithme suivant peut être proposé :

### 3.3 Théorème du point fixe et applications contractantes

Dans cette partie, nous allons donner un théorème particulièrement important pour construire des méthodes itératives de recherche de solutions d'équations non-linéaires. Il s'agit du théorème du point fixe.

#### 3.3.1 Applications contractantes et point fixe

Nous allons commencer par la définition d'une application contractante :

**Définition 3.3.1** Soit  $(\mathcal{E}, d)$  un espace métrique complet, et  $g : \mathcal{E} \mapsto \mathcal{E}$  une application continue. On dit que  $g$  est une application strictement contractante dans  $\mathcal{E}$  si et seulement si

$$\exists \alpha, 0 \leq \alpha < 1, d(g(x), g(y)) \leq \alpha d(x, y), \forall x, y \in \mathcal{E} \quad (3.3.1)$$

On parle aussi pour  $g$  de contraction stricte.

□

Une fonction strictement contractante est une fonction  $k$ -lipschitzienne pour  $k < 1$ . Dans le cas où  $\mathcal{E}$  est l'ensemble de nombres réels  $\mathbb{R}$ , la propriété (3.3.1) est équivalente à :

$$\exists \alpha, 0 \leq \alpha < 1, |g(x) - g(y)| \leq \alpha|x - y|, \forall x, y \in \mathbb{R} \quad (3.3.2)$$

**Définition 3.3.2** On dit que  $a \in \mathcal{E}$  est un point fixe de  $g$  si

$$g(a) = a \quad (3.3.3)$$

□

### 3.3.2 Théorème du point fixe

Le théorème du point fixe s'énonce alors sous la forme suivante :

**Théorème 3.3.1** Soit une application  $g$  strictement contractante d'un espace métrique  $\mathcal{E}$  dans lui-même, alors

1. Il existe un unique point fixe de  $g$ , i.e.,  $g(x) = x$
2. Pour toute donnée initiale  $x_0 \in \mathbb{R}$ , la suite des itérées définie par

$$x_{n+1} = g(x_n) \quad (3.3.4)$$

converge vers  $x$ .

□

*Preuve :*

**Unicité du point fixe** Si  $g$  possède deux points fixes,  $a$  et  $b$  distincts alors

$$d(g(a), g(b)) = d(a, b) \neq 0 \quad (3.3.5)$$

Il y a donc contradiction avec le fait que  $g$  soit contractante.

**Existence du point fixe** Soit  $x_0 \in \mathcal{E}$  une donnée initiale, et  $x_n$  la suite des itérées définies par (3.3.4) on alors :

$$d(x_p, x_{p+1}) = d(g(x_p), g(x_{p-1})) \leq \alpha d(x_p, x_{p-1}) \quad (3.3.6)$$

soit par récurrence :

$$d(x_p, x_{p+1}) \leq \alpha^p d(x_0, x_1) \quad (3.3.7)$$

Pour tout entier  $q > p$ , on peut écrire :

$$d(x_p, x_q) \leq \sum_{l=p}^{q-1} d(x_l, x_{l+1}) \leq \left( \sum_{l=p}^{q-1} \alpha^l \right) d(x_0, x_1) \quad (3.3.8)$$

or

$$\left( \sum_{l=p}^{q-1} \alpha^l \right) \leq \left( \sum_{l=p}^{+\infty} \alpha^l \right) = \frac{\alpha^p}{1-\alpha} \quad (3.3.9)$$

d'où

$$d(x_p, x_q) \leq \frac{\alpha^p}{1-\alpha} d(x_0, x_1) \forall p < q \quad (3.3.10)$$

La suite  $x_p$  est donc une suite de Cauchy. Comme  $\mathcal{E}$  est un espace complet, elle converge vers un point  $x \in \mathcal{E}$ .

□

### 3.3.3 Estimation de la vitesse de convergence

Soit  $x$  le point fixe de  $g$ , on a

$$d(x_p, x) = d(g(x_{p-1}), g(x)) \leq \alpha d(x_{p-1}, x) \quad (3.3.11)$$

d'où par récurrence :

$$d(x_p, x) \leq \alpha^p d(x_0, x) \quad (3.3.12)$$

La convergence est donc exponentiellement rapide.

### 3.3.4 Principe de construction des méthodes itératives métriques

Grâce au théorème du point fixe, il devient possible de construire des méthodes itératives en remplaçant le problème  $f(x) = 0$  par un problème équivalent de la forme  $g(x) = x$  (par exemple  $g(x) = x - f(x)$  sur une partie fermée  $\Omega \subset \mathbb{R}^n$  (toute partie fermée de  $\mathbb{R}^n$  étant complète) à condition de respecter les conditions suivantes :

- $g(\Omega) \subset \Omega$ ,
- la solution  $x$  au problème  $f(x) = 0$  appartienne  $\Omega$ ,
- $g$  soit une application strictement contractante de  $\Omega$  dans  $\Omega$ .

## 3.4 Méthodes itératives métriques – Cas $f : \mathbb{R} \mapsto \mathbb{R}$

Dans ce paragraphe, on restreint le problème (3.1.1) à une fonction d'une seule variable réelle,  $f : \mathbb{R} \mapsto \mathbb{R}$ .

### 3.4.1 Points fixes répulsifs et points fixes attractifs

Soit  $I$  un intervalle fermé de  $\mathbb{R}$ , on suppose que  $g$  est une fonction de classe  $\mathcal{C}^1$  de  $I$  dans  $I$ . Puisque  $g$  est dérivable, la propriété de contractance (3.3.1) est équivalente à :

$$\forall x \in [a, b], |g'(x)| \leq \alpha, 0 \leq \alpha < 1 \quad (3.4.1)$$

Le fait de pouvoir borner la dérivée de  $g$  par une constante  $\alpha$  strictement inférieure à 1 apparaît comme essentielle.

Notons  $a \in I$  un point fixe de  $g$ . Différents cas se posent pour  $g$  quelconque :

1. **Premier cas**  $|g'(a)| < 1, g'(a) \neq 0$ . Soit  $\alpha < 1$  tel que  $|g'(a)| < \alpha < 1$ . Par continuité de  $g'$ , il existe un intervalle  $E = [a - h, a + h]$  sur lequel  $|g'| < \alpha$ , et donc  $g$  est contractante de rapport  $\alpha$  sur  $E$ . On a nécessairement  $g(E) \subset E$  et donc

$$\forall x - 0 \in [a - h, a + h], \lim_{p \rightarrow +\infty} x_p = a \quad (3.4.2)$$

On dir alors que  $a$  est un point fixe attractif et la convergence est au moins exponentiellement rapide.

2. **Second cas**  $g'(a) = 0$  Supposons de plus que  $g$  est de classe  $\mathcal{C}^2$  et que  $|g''| \leq M$  sur un intervalle  $E = [a - h, a + h]$  où  $|g'| < \alpha < 1$ . La formule de Taylor donne :

$$g(x) = g(a) + g'(a)(x - a) + g''(c) \frac{(x - a)^2}{2}, c \in ]a, x[ \quad (3.4.3)$$

$$= a + \frac{1}{2} g''(c)(x - a)^2 \quad (3.4.4)$$

d'où

$$|g(x) - a| \leq \frac{M}{2} |x - a|^2 \quad (3.4.5)$$

Par récurrence, on obtient que :

$$|x_p - a| \leq \frac{2}{M} \left( \frac{M}{2} |x_0 - a|^2 \right)^p \quad (3.4.6)$$

Pour peu que la valeur initiale soit choisie avec la condition  $|x_0 - a| \leq 1/5M$ , on obtient

$$|x_p - a| \leq \frac{2}{M} 10^{-2p} \quad (3.4.7)$$

ce qui signifie que le nombre de décimale exacte dans l'approximation de  $a$  double à chaque itération. On parle parfois de point fixe superattractif. Nous verrons que les méthodes de Newton consiste à se ramener à ce cas.

3. **Troisième cas**  $|g'(a)| > 1$  Comme

$$\lim_{x \rightarrow a} \left| \frac{g(x) - g(a)}{x - a} \right| = |g'(a)| > 1 \quad (3.4.8)$$

d'où

$$\exists h > 0, \forall x \in [a - h, a + h], |g(x) - g(a)| = |g(x) - a| > |x - a| \quad (3.4.9)$$

On dit dans cas que  $a$  est point fixe répulsif. Il faut lors envisagé la fonction réciproque de  $g$  sur un voisinage de  $a$  qui peut être définie car  $g'$  est de signe constant sur un voisinage de  $a$ . L'équation  $x = g(x)$  peut alors se réécrire comme  $x = g^{-1}(x)$  avec  $g^{-1}(a) = 1/g'(a)$ . On est donc ramené au premier cas.

4. **Quatrième cas**  $|g'(a)| = 1$  Dans cas, il est très difficile de conclure sur la qualité du point fixe.

### 3.4.2 Méthode de la corde

**Principe.** La méthode de la corde est la méthode itérative la plus simple que l'on peut proposer à partir du théorème du point fixe. Il s'agit de remplacer le problème

$$f(x) = 0 \quad (3.4.10)$$

par

$$g(x) = x, \quad \text{avec } g(x) = x - \lambda f(x), \lambda \neq 0 \quad (3.4.11)$$

où  $\lambda$  est une constante réelle.

On constate facilement qu'un point fixe de  $g$  correspond bien à une solution de l'équation  $f(x) = 0$ .

La méthode itérative est donc donnée par :

$$\boxed{x_{n+1} = x_n - \lambda f(x_n)} \quad (3.4.12)$$

**Convergence** Une condition suffisante pour que  $g$  est un point fixe dans  $[a, b]$  est que  $g([a, b]) \subset [a, b]$ , et  $g$  soit une application contractante sur  $[a, b]$ , soit, si on suppose  $f$  dérivable :

$$\forall x \in [a, b], |g'(x)| = |1 - \lambda f'(x)| \leq \alpha < 1 \quad (3.4.13)$$

soit

$$\forall x \in [a, b], 1 - \alpha \leq \lambda f'(x) \leq 1 + \alpha, \text{ avec } 0 \leq \alpha < 1 \quad (3.4.14)$$

Cette relation implique en particulier que  $f'$  ne change pas de signe sur  $[a, b]$  ( $f$  soit monotone sur  $[a, b]$ ) et que  $\lambda$  est du signe de  $f'(x)$ .

**Interprétation géométrique de la méthode de la corde** La droite passant par le point  $(x_n, f(x_n))$  de pente  $\mu$  a pour équation :

$$y = f(x_n) + \mu(x - x_n) \quad (3.4.15)$$

Cette droite coupe l'axe des abscisses en  $x^*$  qui a pour valeur :

$$x^* = x_n - \frac{f(x_n)}{\mu} \quad (3.4.16)$$

Si on suppose que  $\mu = 1/\lambda$ , on voit que le point suivant de l'itération  $x_{n+1}$  est obtenu comme l'intersection de la droite de pente  $1/\lambda$  passant par  $(x_n, f(x_n))$  et l'axe des abscisses.

Cette méthode est illustrée sur la figure pour  $\lambda = -1$

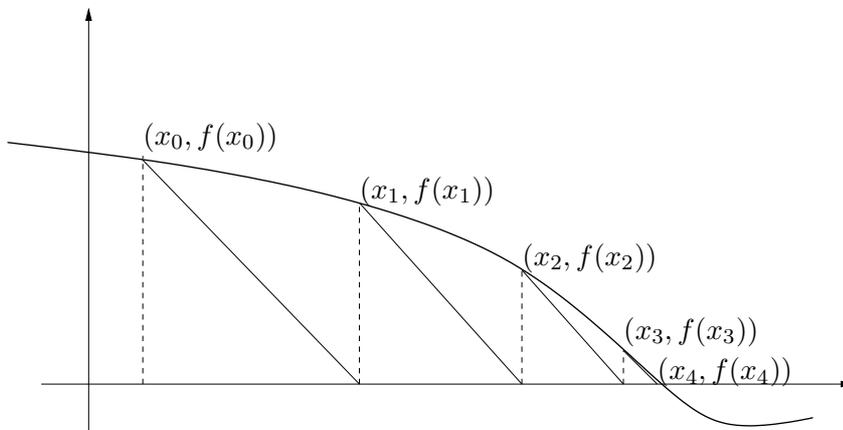


FIG. 3.1 – Méthode de la corde

L'inconvénient principale de la méthode de la corde est le choix de  $\lambda$  pour que l'inégalité (3.4.13) soit respectée. Il est bien clair que si  $f'(x) \neq 0$  on peut toujours trouver une valeur  $\lambda$  pour que cette inégalité soit respectée sur un voisinage de  $x$  et permettre la convergence. Le problème est que ce choix ne peut se faire qu'à tâtons. Nous allons donner dans la suite une méthode beaucoup plus efficace.

### 3.4.3 Méthode de Newton

**Principe** Une des difficultés principales de la méthode de la corde est le choix du paramètre  $\lambda$ . On sait que les itérations convergeront d'autant plus vite que le rapport  $k$  est petit (voir estimation de la vitesse de convergence). En étudiant l'équation (3.4.13), il semble qu'un choix judicieux consiste à remplacer la constante  $\lambda$  par  $\frac{1}{f'(x)}$  si  $f'$  ne s'annule pas sur l'intervalle  $[a, b]$ .

**Méthode itérative de Newton** Ce choix conduit donc à chercher le point de la fonction suivante :

$$g(x) = x - \frac{f(x)}{f'(x)} \quad (3.4.17)$$

qui conduit aux itérées suivantes :

$$\boxed{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}} \quad (3.4.18)$$

**Convergence de la méthode de Newton** On donne ici un théorème de convergence locale de la méthode de Newton :

**Théorème 3.4.1** Soit  $f$  une fonction de classe  $\mathcal{C}^2$  sur un intervalle  $[a, b] \subset \mathbb{R}$ . Supposons qu'il existe  $x \in [a, b]$  tel que  $f(x) = 0$  et  $f'(x) \neq 0$  alors il existe  $\epsilon > 0$  tel que quelque soit la condition initiale  $x_0 \in [x - \epsilon, x + \epsilon]$ , la suite de itérées de Newton soit bien définie, reste dans l'intervalle  $[x - \epsilon, x + \epsilon]$  et converge vers  $x$  quand  $n$  tend vers l'infini. □

*Preuve* : La fonction a itérée s'exprime de la manière suivante :

$$g(x) = x - \frac{f(x)}{f'(x)} \quad (3.4.19)$$

Si  $f$  est de classe  $\mathcal{C}^2$  alors

$$g'(x) = 1 - \frac{f'(x)f'(x) - f(x)f''(x)}{f'(x)f'(x)} = \frac{f(x)f''(x)}{f'(x)f'(x)} \quad (3.4.20)$$

On constate donc que  $g'(x) = 0$  et que  $x$  est donc un point fixe superattractif. En utilisant les propriétés des points superattractifs, on peut en déduire la convergence quadratique vers  $x$  des itérées sur un voisinage de  $x$ . □

**Interprétation géométrique** Si l'on effectue la même interprétation géométrique que précédemment on voit que l'itérée  $x_{n+1}$  s'obtient en réalisant l'intersection de la tangente à la courbe de  $f$  au point  $(x_n, f(x_n))$  avec l'axe des abscisses. Cette méthode est illustrée sur la figure 3.2

**Interprétation analytique** Une autre interprétation peut être donnée à la méthode de Newton. Elle se base sur le développement de Taylor de la fonction  $f$  au point  $x_0$  :

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \mathcal{O}(|x|^2) \quad (3.4.21)$$

Si l'on choisit de remplacer  $f$  par son développement au premier ordre, l'équation  $f(x)=0$  devient :

$$f(x_0) + f'(x_0)(x - x_0) = 0 \quad (3.4.22)$$

qui a pour solution :

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \quad (3.4.23)$$

Le principe de la méthode de Newton a itéré ce processus.

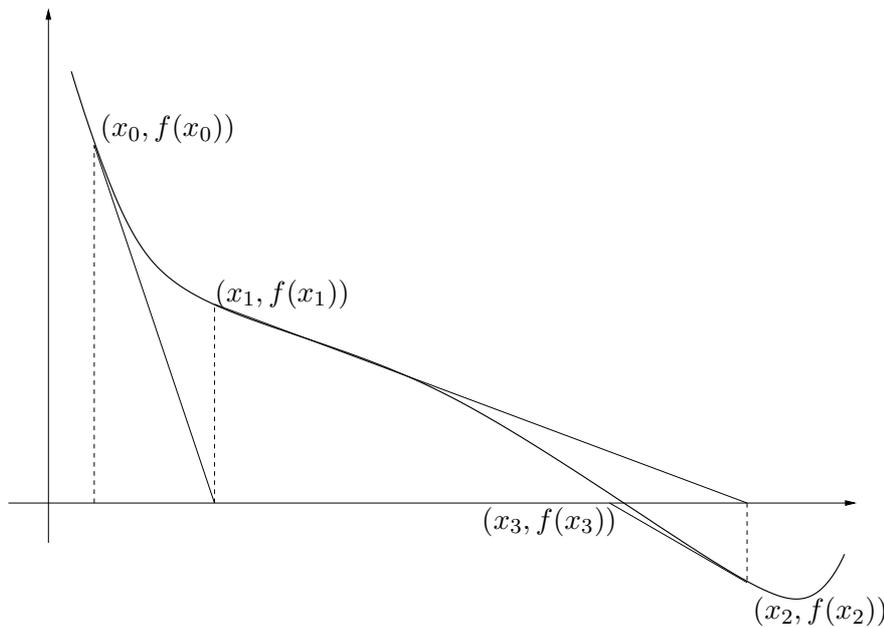


FIG. 3.2 – Méthode de Newton

### 3.4.4 Méthode de la sécante de Steffensen

**Principe** Le principe de la méthode de la sécante se base sur une approximation de la dérivée de  $f$  au point  $x_n$ . Dans la pratique, il n'est pas toujours évident de donner de manière simple la dérivée d'une fonction même si elle est connue analytiquement.

**Méthode itérative de la sécante** La dérivée au point  $x_n$  est généralement approché par :

$$\hat{f}'(x_n) = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} \quad (3.4.24)$$

et conduit aux itérées suivantes :

$$x_{n+1} = x_n - (x_n - x_{n-1}) \frac{f(x_n)}{f(x_n) - f(x_{n-1})} \quad (3.4.25)$$

La méthode de la sécante est une méthode à deux pas dans le sens où elle nécessite la connaissance des itérées  $x_n$  et  $x_{n-1}$  pour pouvoir calculer l'itérée  $x_{n+1}$ . Il est donc nécessaire de donner deux valeurs initiales  $x_0$  et  $x_1$  pour démarrer l'algorithme.

**Interprétation géométrique** L'interprétation géométrique est illustrée sur le figure 3.3. La méthode consiste à trouver à chaque itération l'intersection de l'axe des abscisses avec le droite de pente  $\hat{f}'(x_n)$  passant par les points  $(x_n, f(x_n))$  et  $(x_{n-1}, f(x_{n-1}))$ .

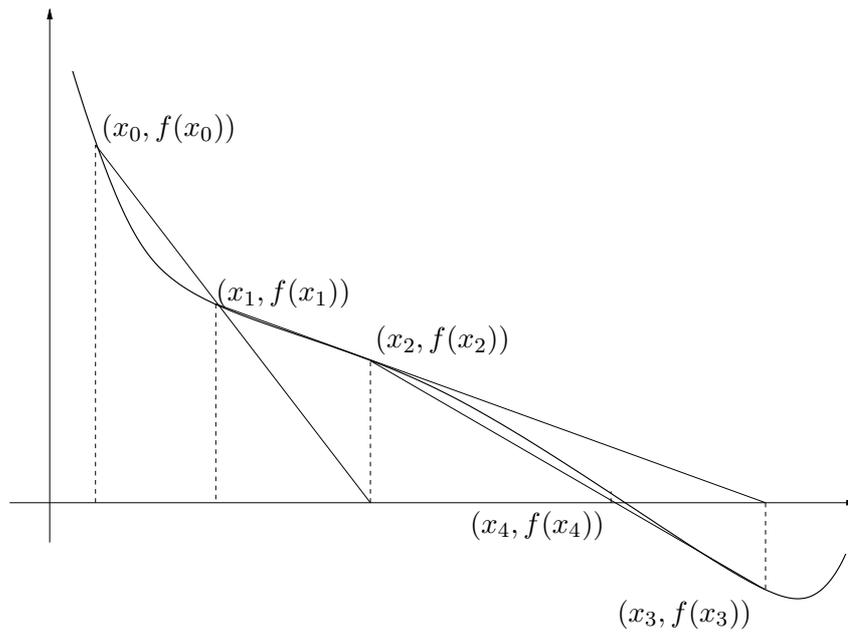


FIG. 3.3 – Méthode de la sécante

### 3.5 Méthodes itératives métriques – Cas $f : \mathbb{R}^n \mapsto \mathbb{R}^n$

Dans le cas d'une fonction vectorielle, les méthodes métriques se révèlent être le plus efficaces. Elles généralisent simplement les méthodes que l'on a vu dans le cas d'une variable réelle. Nous nous bornerons à présenter ici la méthode de Newton-Raphson.

#### 3.5.1 Méthode Newton-Raphson

**Principe.** La méthode de Newton peut être étendue au problème d'une fonction vectorielle de la manière suivante. On suppose que  $f$  est une fonction de classe  $\mathcal{C}^2$  d'un ouvert  $\Omega \subset \mathbb{R}^n$  dans  $\mathbb{R}^n$ . L'idée est d'approximer  $f$  par son développement linéaire de la manière suivante :

$$f(x) = f(a) + \nabla f(a)(x - a) + \mathcal{O}(\|x - a\|) \quad (3.5.1)$$

où  $\nabla f$  représente le gradient de  $f$  soit :

$$\nabla f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \cdots & \frac{\partial f_2}{\partial x_n} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \cdots & \cdots & \frac{\partial f_3}{\partial x_n} \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \quad (3.5.2)$$

Le gradient de  $f$  est une application linéaire de  $\mathbb{R}^n$  dans  $\mathbb{R}^n$ . Si cette application est inversible,

la méthode de Newton revient à itérer comme dans le cas à une seule variable la fonction suivante :

$$g(x) = x - \nabla f^{-1}(x).f(x) \quad (3.5.3)$$

**Méthode itérative de Newton-Raphson** La méthode de Newton-Raphson revient à effectuer les itérations suivantes :

$$x_{n+1} = x_n - \nabla f^{-1}(x_n).f(x_n) \quad (3.5.4)$$

### 3.6 Cas particuliers où $f$ est un polynôme – Racines d'un polynôme

Pour terminer ce chapitre, on donne quelques méthodes de recherche de racines d'un polynôme.

#### 3.6.1 Procédé de localisation de Sturm

Le procédé de localisation de Sturm permet de localiser le nombre de racines d'un polynôme sur un intervalle fermé de  $\mathbb{R}$ . Il ne permet, *a priori* de donner de valeurs précises des racines, mais il est très utile si l'on veut employer une méthode de dichotomie.

**Principe** Soit un polynôme,  $p \in \mathcal{P}_n$  ne possédant que des racines simples et un intervalle fermé de  $\mathbb{R}$   $[a, b]$ , on considère la suite de polynôme suivante :

$$p_0 = p \quad (3.6.1)$$

$$p_1 = p' \quad (3.6.2)$$

$$p_2 = q_2 p_1 - p_0 \text{ Opposé au reste de la division euclidienne} \quad (3.6.3)$$

$$p_{i+2} = q_{i+2} p_{i+1} - p_i \quad (3.6.4)$$

Cette suite est nécessairement finie. En effet, le degré de  $p_1$  est  $n - 1$  et à chaque division euclidienne de polynômes, on a le degré du reste qui est nécessairement inférieur au degré du diviseur. La suite  $p_i$  est donc de degré strictement décroissant. Au delà d'un rang  $m < n$ , les polynômes sont donc de degré 0.

Pour compter le nombre de racines dans  $[a, b]$ , il suffit de compter les changements de signe de la suite  $p_i$  au bord de l'intervalle. Le nombre de racines sera égal au nombre de changements de signe en  $a$  diminué du nombre de changement de signe en  $b$ .

**Exemple** Considérons le polynôme  $p(x) = x^3 - x$  sur  $[-2, 2]$ . La suite de polynômes de Sturm est la suivante :

$$p_0 = x^3 - x \quad (3.6.5)$$

$$p_1 = 3x^2 - 1 \quad (3.6.6)$$

$$p_2 = \frac{2}{3}x \quad (3.6.7)$$

$$p_3 = 1 \quad (3.6.8)$$

Les valeurs de la suite  $p_i$  au bord de l'intervalle sont les suivantes :

| $x$ | $p_0$ | $p_1$ | $p_2$ | $p_3$ |
|-----|-------|-------|-------|-------|
| -2  | -6    | 11    | -4/3  | 1     |
| 2   | 6     | 11    | 4/3   | 1     |

(3.6.9)

On constate donc que la suite  $p_i(-2)$  change trois fois de signes et que la suite  $p_i(2)$  n'en change pas. Le nombre de racines dans  $[-2, 2]$  est donc 3.

### 3.6.2 Méthode de Bairstow

**Principe** La méthode de Bairstow permet de calculer avec précision les racines réelles et complexes d'un polynôme. Le principe est le suivant. Soit un polynôme  $p \in \mathcal{P}_n$ , il peut s'écrire sous sa forme canonique de la manière suivante :

$$p(x) = a_0 + a_1x + \cdots + a_nx^n \quad (3.6.10)$$

Le principe de la méthode de Bairstow est de chercher un polynôme de degré 2 qui divise  $p$ . Les racines de ce polynôme de degré 2 seront donc des racines de  $p$ . Écrivons la division euclidienne de  $p$  par un polynôme de degré 2 :

$$p(x) = \underbrace{(x^2 + px + q)}_{\text{Diviseur}}(b_0 + b_1x + \cdots + b_{n-2}x^{n-2}) + \underbrace{Rx + S}_{\text{Reste}} \quad (3.6.11)$$

Pour un choix quelconque du couple  $(p, q)$ , il a a priori aucune raison que le reste s'annule. Il faut donc chercher le couple  $(p, q)$  pour que le reste s'annule. Ceci revient à résoudre le problème suivant :

$$\begin{cases} R(p, q) = 0 \\ S(p, q) = 0 \end{cases} \quad (3.6.12)$$

On reconnaît dans (3.6.12) un problème d'équations non linéaires de la forme  $F(p, q) = 0$ , qui peut être résolu par une méthode de Newton. L'itérée de la méthode de Newton peut s'écrire de la manière suivante :

$$(p, q)_{i+1} = (p, q)_i - \nabla^{-1}F((p, q)_i).F((p, q)_i) \quad (3.6.13)$$

Le gradient de  $F$  s'exprime sous la forme suivante :

$$\nabla F((p, q)) = \begin{bmatrix} \frac{\partial R}{\partial p} & \frac{\partial R}{\partial q} \\ \frac{\partial S}{\partial p} & \frac{\partial S}{\partial q} \end{bmatrix} \quad (3.6.14)$$

et son inverse, lorsqu'elle existe :

$$\nabla^{-1} F(p, q) = \frac{1}{\frac{\partial R}{\partial p} \frac{\partial S}{\partial q} - \frac{\partial S}{\partial p} \frac{\partial R}{\partial q}} \begin{bmatrix} \frac{\partial S}{\partial q} & -\frac{\partial R}{\partial q} \\ -\frac{\partial S}{\partial p} & \frac{\partial R}{\partial p} \end{bmatrix} \quad (3.6.15)$$

A chaque itération, il convient donc d'évaluer :

$$p_{i+1} = p_i - \frac{1}{\Delta_i} \cdot \left[ \frac{\partial S}{\partial q}(p_i, q_i) R(p_i, q_i) - \frac{\partial R}{\partial q}(p_i, q_i) S(p_i, q_i) \right] \quad (3.6.16)$$

$$q_{i+1} = q_i - \frac{1}{\Delta_i} \cdot \left[ -\frac{\partial S}{\partial p}(p_i, q_i) R(p_i, q_i) + \frac{\partial R}{\partial p}(p_i, q_i) S(p_i, q_i) \right] \quad (3.6.17)$$

$$\text{avec } \Delta_i = \frac{\partial R}{\partial p}(p_i, q_i) \frac{\partial S}{\partial q}(p_i, q_i) - \frac{\partial S}{\partial p}(p_i, q_i) \frac{\partial R}{\partial q}(p_i, q_i) \quad (3.6.18)$$

Pour cela calculons les expressions de  $R$  et  $S$  en fonction de  $p$  et  $q$ .

Le résultat de la division nous fournit les coefficients  $b_i, i = 0 \dots n-2$  de la manière suivante :

$$(I) \begin{cases} b_0 = a_0 \\ b_1 = a_1 - pb_0 \\ b_2 = a_2 - pb_1 - qb_0 \\ \vdots \\ b_{n-2} = a_{n-2} - pb_{n-3} - qb_{n-4} \end{cases} \quad (3.6.19)$$

Pour le reste, on pose

$$R = b_{n-1} \quad (3.6.20)$$

$$S = pb_{n-1} + b_n \quad (3.6.21)$$

avec

$$b_{n-1} = a_{n-1} - pb_{n-2} - qb_{n-3} \quad (3.6.22)$$

$$b_n = a_n - pb_{n-1} - qb_{n-2} \quad (3.6.23)$$

Il convient de calculer les dérivées de  $R$  et  $S$  par rapport à  $p$  et à  $q$ . Pour cela, il faut calculer les dérivées de  $b_k$  par rapport à  $p$  et à  $q$ . On pose

$$\frac{\partial b_k}{\partial p} = -c_{k-1}, i = 1 \dots n-1 \quad (3.6.24)$$

$$\frac{\partial b_n}{\partial p} = -c_{n-1} - b_{n-1} \quad (3.6.25)$$

En dérivant le tableau (I), on obtient :

$$(II) \left\{ \begin{array}{l} c_0 = b_0 \\ c_1 = b_1 - pc_0 \\ c_2 = b_2 - pc_1 - qc_0 \\ \vdots \\ c_{n-2} = b_{n-2} - pc_{n-3} - q_{n-4} \\ c_{n-1} = b_{n-1} - pc_{n-2} - q_{n-3} \end{array} \right. \quad (3.6.26)$$

La dérivation par rapport à  $q$  montrerait que :

$$\frac{\partial b_k}{\partial q} = -c_{k-2}, i = 2 \dots n-1 \quad (3.6.27)$$

$$(3.6.28)$$

On obtient alors pour les dérivées de  $R$  et  $S$  par rapport à  $p$  et  $q$ , les valeurs suivantes :

$$\frac{\partial R}{\partial p} = -\frac{\partial b_{n-1}}{\partial p} = -c_{n-2} \quad (3.6.29)$$

$$\frac{\partial R}{\partial q} = -\frac{\partial b_{n-1}}{\partial q} = -c_{n-3} \quad (3.6.30)$$

$$\frac{\partial S}{\partial p} = -\frac{\partial b_n + pb_{n-1}}{\partial p} = -c_{n-1} - pc_{n-2} \quad (3.6.31)$$

$$\frac{\partial S}{\partial q} = -\frac{\partial b_n + pb_{n-1}}{\partial q} = -c_{n-2} - pc_{n-3} \quad (3.6.32)$$

$$(3.6.33)$$

et par conséquent :

$$\begin{aligned} \Delta_i &= c_{n-2}^2 - c_{n-1}c_{n-3} \\ p_{i+1} &= p_i - \frac{b_n c_{n-3} - b_{n-1} c_{n-2}}{\Delta_i} \\ q_{i+1} &= q_i - \frac{b_{n-1} c_{n-1} - b_n c_{n-2}}{\Delta_i} \end{aligned} \quad (3.6.34)$$

**Algorithme** L'algorithme se déroule de ma manière suivante :

### 3.6.3 Méthode d'Aitken

#### 3.6.1 Note de Rédaction

A compléter

---

**Algorithme 2** Méthode de Bairstow

---

**Require:**  $a_i, i = 0 \dots n, \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \varepsilon$ Initialisation  $p_0 \leftarrow \alpha, q_0 \leftarrow \beta$  $erreur \leftarrow 10 * \varepsilon$ **while**  $erreur > \varepsilon$  **do** $i \leftarrow i + 1$ Calculer  $b_i, c_i$  grâce à (3.6.19) et (3.6.26)Calculer  $p_i$  et  $q_i$  grâce à (3.6.34) $erreur = \frac{|p_i - p_{i-1}| + |q_i - q_{i-1}|}{q_i + q_i}$ **end while**Calculer les racines de  $x^2 + p_i x + q_i = 0$ Reprendre l'algorithme avec le polynôme  $p(x) = \sum_{i=0}^{n-2} b_i x^i$ 

---

# Chapitre 4

## Équations Différentielles Ordinaires (EDO) – Généralités

### Contenu du Chapitre 4

---

|            |  |           |
|------------|--|-----------|
| <b>4.1</b> | <b>Introduction</b>                                  | <b>62</b> |
| <b>4.2</b> | <b>Terminologie et définition</b>                    | <b>62</b> |
| 4.2.1      | Système différentiel du premier ordre                | 62        |
| 4.2.2      | Le problème de Cauchy                                | 62        |
| 4.2.3      | Système différentiel du second ordre                 | 63        |
| 4.2.4      | Système différentiel d'ordre $m$                     | 64        |
| <b>4.3</b> | <b>Théorèmes d'existence et d'unicité</b>            | <b>65</b> |
| 4.3.1      | Solutions locales, maximales, globales               | 65        |
| 4.3.2      | Théorème d'existence de Cauchy-Peano                 | 67        |
| 4.3.3      | Théorème d'unicité de Cauchy-Lipschitz               | 68        |
| 4.3.4      | Influence des perturbations sur les données          | 69        |
| 4.3.5      | Régularité des solutions                             | 70        |
| <b>4.4</b> | <b>Méthodes explicites de résolution</b>             | <b>71</b> |
| 4.4.1      | Équations différentielles linéaires du premier ordre | 71        |
| 4.4.2      | Équations différentielles linéaires du second ordre  | 71        |

---

## 4.1 Introduction

## 4.2 Terminologie et définition

### 4.2.1 Système différentiel du premier ordre

Un système différentiel du premier ordre se définit de la manière suivante :

**Définition 4.2.1** Soit une fonction  $f : \mathbb{R}^{n+1} \mapsto \mathbb{R}^n$ , on appelle un système différentiel du premier ordre l'équation de la forme suivante :

$$y'(t) = f(t, y(t)) \quad (4.2.1)$$

où  $y'$  représente la dérivée de  $y$  par rapport à  $t$ .

□

Dans ce type de système, l'inconnue est la fonction  $y(t) : \mathbb{R} \mapsto \mathbb{R}^n$ . Le qualificatif ordinaire vient du fait que cette fonction ne dépend que d'une seule variable. Sinon, on parle d'équations aux dérivées partielles (EDP).

### 4.2.2 Le problème de Cauchy

Sans la donnée de conditions initiales consistantes, il est impossible de définir la notion de solution d'un système différentiel du premier ordre. C'est pourquoi, on introduit un problème standard qui est le problème de Cauchy :

**Définition 4.2.2** On appelle *problème de Cauchy* le problème suivant :

Étant donnés :

- un intervalle  $I_0 \subset \mathbb{R}$ ,
- une fonction  $f$ , définie et continue sur  $I_0 \times \mathbb{R}^n$  à valeurs dans  $\mathbb{R}^n$  :

$$f : I_0 \times \mathbb{R}^n \mapsto \mathbb{R}^n \quad (4.2.2)$$

$$(t, y) \mapsto f(t, y) \quad (4.2.3)$$

Trouver une fonction  $y \in \mathcal{C}^1(I_0)$  telle que

$$\begin{cases} y'(t) = f(t, y(t)), \forall t \in I_0, \forall y \in \mathbb{R}^n \\ y(t_0) = y_0, t_0 \in I_0, \quad \text{Condition initiale} \end{cases} \quad (4.2.4)$$

□

Outre la donnée d'un intervalle et d'une fonction  $f$ , le problème de Cauchy se caractérise par la donnée d'une condition dite condition de Cauchy ou condition initiale. La forme de cette condition est essentielle. On peut envisager, en effet, des problèmes où des conditions sont données

sur la dérivée de  $y'$ , ou encorte en deux points d'un intervalle. Il n'agit alors plus d'un problème de Cauchy.

Le problème de Cauchy peut se mettre sous une forme équivalente donnée par le théorème suivant :

**Théorème 4.2.1** Une fonction  $y : I_0 \mapsto \mathbb{R}^n$  est une solution du problème de Cauchy si et seulement si :

1. la fonction  $y$  est continue et  $\forall t \in I_0, (t, y(t)) \in I_0 \times \mathbb{R}^n$
- 2.

$$\forall t \in I_0, y(t) = y_0 + \int_{t_0}^t f(t, y(t)) \quad (4.2.5)$$

□

*Preuve* : La preuve est simple. Si  $y$  vérifie les deux hypothèses de ce théorème alors  $y$  est différentiable et  $y(t_0) = y_0$  et  $y'(t) = f(t, y(t))$ . Réciproquement, si les relations du problème de Cauchy (4.2.4) sont satisfaites, l'équation (4.2.5) de déduit directement par intégration.

□

La solution  $y$  du problème de Cauchy est parfois appelée l'intégrale du problème et la résolution de ce système est parfois appelée intégration.

### Remarques

1. Si  $f$  est une fonction uniquement de  $t$ , le problème de Cauchy se ramène à la recherche d'une primitive. Dans le cas unidimensionnel,  $y(t) \in \mathbb{R}$ , c'est également le cas lorsque l'on peut effectuer une séparation des variables, i.e. :

$$f(t, y) = g(t)h(y) \quad (4.2.6)$$

Dans ce cas, si  $h(y) \neq 0$ , on peut écrire le problème de Cauchy sous la forme suivante :

$$\int_{y_0}^y \frac{dy}{h(y)} = \int_{t_0}^t g(t)dt \quad (4.2.7)$$

2. Si  $f$  est une fonction linéaire de  $y$ , il existe alors de nombreuses méthodes analytiques pour résoudre le problème comme on le verra au § 4.4.
3. Si  $f$  est une fonction non linéaire de  $y$ , nous verrons comment ce problème peut être abordée numériquement.

### 4.2.3 Système différentiel du second ordre

On définit un système différentiel du second ordre de la manière suivante :

**Définition 4.2.3** Soit une fonction  $f : \mathbb{R}^{2n+1} \mapsto \mathbb{R}^n$  continue, on appelle un système différentiel du premier ordre l'équation de la forme suivante :

$$y''(t) = f(t, y(t), y'(t)) \quad (4.2.8)$$

□

Pour ce type de système, on se ramène à un système du premier ordre en effectuant le changement de variable suivant :

$$z(t) = \begin{bmatrix} y(t) \\ y'(t) \end{bmatrix} \quad (4.2.9)$$

et en posant une nouvelle fonction  $F : \mathbb{R}^{2n+1} \mapsto \mathbb{R}^{2n+1}$  telle que :

$$F(t, z(t)) = \begin{bmatrix} y'(t) \\ f(t, y) \end{bmatrix} \quad (4.2.10)$$

Le système différentiel (4.2.8) peut alors s'écrire sous la forme d'un système différentiel au premier ordre :

$$z'(t) = F(t, z(t)) \quad (4.2.11)$$

Il est important de noter que la dimension du problème (4.2.11) est double par rapport au problème initial du second ordre. Pour retrouver le problème de Cauchy, il convient donc de fournir  $z(t_0) = z_0$ , c'est à dire à la fois  $y(t_0)$  et  $y'(t_0)$ .

#### 4.2.4 Système différentiel d'ordre $m$

Un système différentiel d'ordre  $m$  peut être défini de la manière suivante :

**Définition 4.2.4** Soit une fonction  $f : \mathbb{R}^{m \times n+1} \mapsto \mathbb{R}^n$  continue, on appelle un système différentiel d'ordre  $m$  l'équation de la forme suivante :

$$y^{(m)}(t) = f(t, y(t), \dots, y^{(m-1)}(t)) \quad (4.2.12)$$

□

De la même manière que pour le système du second ordre, on peut se ramener à un système du premier ordre en posant le changement de variable :

$$z(t) = \begin{bmatrix} y(t) \\ \vdots \\ y^{(m-1)}(t) \end{bmatrix} \quad (4.2.13)$$

et la nouvelle fonction  $F$  :

$$F(t, z) = \begin{bmatrix} y'(t) \\ \vdots \\ y^{(m-2)}(t) \\ f(t, y(t), \dots, y^{(m-1)}(t)) \end{bmatrix} \quad (4.2.14)$$

Dans la suite, nous nous intéresserons exclusivement au problème du premier ordre sachant que l'on peut toujours s'y ramener grâce à un changement de variables adéquat.

### 4.3 Théorèmes d'existence et d'unicité

Cette partie aborde les questions fondamentales de l'existence et de l'unicité de solutions du problème de Cauchy.

#### 4.3.1 Solutions locales, maximales, globales

Avant de donner des théorèmes d'existence de solution, il est bon de préciser quelle type de solution on recherche. Pour cela, on introduit les définitions suivantes.

**Définition 4.3.1** On appelle solution locale du problème de Cauchy (4.2.4) la donnée d'un couple  $(I, y)$  où  $I$  est un intervalle de  $\mathbb{R}$  qui est voisinage de  $t_0$  dans  $I_0$  et où  $y$  est une fonction de classe  $\mathcal{C}^1$  sur  $I$  telle que :

$$y(t_0) = y_0, \quad \forall t \in I, y'(t) = f(t, y(t)) \quad (4.3.1)$$

□

**Définition 4.3.2** On dit que la solution locale  $(J, z)$  prolonge la solution locale  $(I, y)$  si on a

$$I \subset J, \quad \forall t \in I, y(t) = z(t) \quad (4.3.2)$$

Si de plus,  $I \neq J$  on dit que la solution  $(J, z)$  prolonge strictement la solution  $(I, y)$ .

□

**Définition 4.3.3** On dit que la solution locale  $(I, y)$  du problème de Cauchy (4.2.4) est maximale si il n'existe pas de solution locale qui la prolonge strictement.

□

**Définition 4.3.4** On dit que la solution  $(I, y)$  est une solution globale du problème de Cauchy (4.2.4) dans  $I_0$  si  $(I, y)$  est une solution locale et  $I = I_0$ .

□

**Exemples**

1. Considérons le problème de Cauchy suivant :

$$\begin{cases} y'(t) = -2ty^2(t), t \in \mathbb{R} \\ y(0) = 1 \end{cases} \quad (4.3.3)$$

Ce problème admet une solution globale<sup>1</sup> et une seule dans  $\mathbb{R}$  :

$$(\mathbb{R}, y(t) = \frac{1}{1+t^2}) \quad (4.3.4)$$

et il n'y a pas d'autre solution maximale.

2. Le problème de Cauchy suivant :

$$\begin{cases} y'(t) = 2ty^2(t), t \in \mathbb{R} \\ y(0) = 1 \end{cases} \quad (4.3.5)$$

admet une solution maximale et une seule dans  $\mathbb{R}$ , c'est  $(] - 1, 1[, y(t) = \frac{1}{1-t^2})$ . Il n'y a pas de solution globale. Si l'on choisit comme condition initiale à ce problème la donnée  $y(-2) = 0$ , il existe une solution maximale et une seule qui est  $(] - 5, 5[, y(t) = \frac{1}{5-t^2})$ . Il est bon de remarquer que l'ensemble sur lequel se définit les solutions d'un problème est dépendant aussi de la condition initiale.

3. Le problème de Cauchy suivant :

$$\begin{cases} y'(t) = -y^2(t), t \in \mathbb{R}_+ \\ y(0) = 1 \end{cases} \quad (4.3.6)$$

admet une solution globale et une seule dans  $[0, +\infty[$ , qui est  $(]0, +\infty[, y(t) = \frac{1}{1+t})$ . Sur  $\mathbb{R}$ , il n'y a pas de solution globale mais seulement une solution maximale sur  $] - 1, +\infty[$ .

4. Le problème de Cauchy suivant :

$$\begin{cases} y'(t) = -\sqrt[3]{y(t)}, t \in [0, +\infty[ \\ y(0) = 0 \end{cases} \quad (4.3.7)$$

admet comme solutions globales :

$$([0, +\infty[, y(t) = 0) \quad (4.3.8)$$

$$([0, +\infty[, y(t) = \sqrt{\frac{8t^3}{27}}) \quad (4.3.9)$$

$$([0, +\infty[, y(t) = -\sqrt{\frac{8t^3}{27}}) \quad (4.3.10)$$

Nous pouvons constater que les problèmes d'existence et d'unicité de solutions ne sont pas des questions triviales. Nous allons présenter dans la suite quelques uns des théorèmes fondamentaux d'existence et d'unicité des solutions moyennant des hypothèses sur  $f$ .

1. Remarquer simplement que  $(1/y)' = -y'/y^2$

### 4.3.2 Théorème d'existence de Cauchy-Peano

**Théorème 4.3.1 (Cauchy-Peano)** On suppose que  $f$  est continue dans un voisinage du point  $(t_0, y_0) \in I_0 \times \mathbb{R}^n$ , alors il existe un intervalle  $J_0$  voisinage de  $t_0$  dans  $I_0$  et une fonction  $y \in \mathcal{C}^1(J_0)$  tels que :

$$\begin{cases} y'(t) = f(t, y(t)), \forall t \in J_0, \\ y(t_0) = y_0 \end{cases} \quad (4.3.11)$$

□

Le théorème de Cauchy-Peano donne donc un résultat d'existence d'une solution locale  $(J_0, y)$ .

*Preuve* : Voir (CROUZEIX & MIGNOT, 1992).

□

La condition de continuité de  $f$  ne suffit pas pour passer à une solution globale. Nous verrons dans la suite qu'il faut imposer des conditions supplémentaires. Par contre, le résultat qui suit permet d'assurer l'existence d'une solution maximale.

**Théorème 4.3.2** Étant donnée une solution locale, il existe au moins (pas forcément unique) une solution maximale qui la prolonge.

□

*Preuve* : Pour la preuve, nous renvoyons à (CROUZEIX & MIGNOT, 1992).

□

On a pu remarquer sur les exemples du § précédent que la notion de solution est très dépendante de l'ensemble sur lequel on la cherche. On constate aussi que lorsque des solutions maximales non globales existent, cela est souvent dû à une singularité de  $y$ . Le théorème suivant précise cette intuition.

**Théorème 4.3.3** On suppose que  $f$  est continue sur  $I_0 \times \mathbb{R}^n$  et que  $I_0$  est de la forme  $[t_0, t_0 + T]$ ,  $[t_0, t_0 + T[$  ou  $[t_0, +\infty[$ ; alors si  $(I, y)$  est une solution maximale non globale du problème de Cauchy (4.2.4),  $I$  est de la forme  $[t_0, t_1[$  et  $y$  n'est pas borné sur  $I$ .

□

*Preuve* : Pour la preuve, nous renvoyons à (CROUZEIX & MIGNOT, 1992).

□

Enfin, donnons un résultat d'existence de solution globale :

**Théorème 4.3.4** On suppose que  $I_0 = [t_0, t_0 + T]$ , que  $f$  est continue sur  $I_0 \times \mathbb{R}^n$  et qu'il existe un produit scalaire de  $\mathbb{R}^n$  noté  $\langle \cdot, \cdot \rangle$  associé à une norme  $\|\cdot\|$  et une fonction  $l \in \mathcal{L}^1(I_0)$  tels que

$$\forall (t, y) \in I_0 \times \mathbb{R}^n, \quad \langle f(t, y), y \rangle \leq l(t)(1 + \|y\|^2) \quad (4.3.12)$$

alors le problème de Cauchy (4.2.4) admet au moins une solution globale.

□

*Preuve* : Pour la preuve, nous renvoyons à (CROUZEIX & MIGNOT, 1992). □

**Exemple.** Considérons le problème de Cauchy suivant :

$$\begin{cases} y'(t) = -y^3(t) + \sin(t)y^2(t) + 1 \text{ sur } I_0 = [0, +\infty[ \\ y(0) = y_0 \text{ donné dans } \mathbb{R} \end{cases} \quad (4.3.13)$$

Le théorème précédent peut s'appliquer avec le produit scalaire usuel de  $\mathbb{R}$ . En effet, on a

$$\forall (t, y) \in [0, +\infty[ \times \mathbb{R} \quad \langle f(t, y), y \rangle = -y^4 + \sin(t)y^3 + y \leq -y^4 + |y|^3 + |y| \leq 1 + |y| \quad (4.3.14)$$

et donc il existe une solution globale sur  $I_0 = [0, +\infty[$ .

### 4.3.3 Théorème d'unicité de Cauchy-Lipschitz

La notion d'unicité de la solution du problème de Cauchy peut être précisée de la façon suivante :

**Définition 4.3.5** On dit que le problème de Cauchy (4.2.4) admet une solution et une seule s'il admet une solution globale et si toute solution locale est la restriction de cette solution globale. □

Le théorème suivant est un théorème fondamental d'unicité de la solution :

**Théorème 4.3.5** On suppose que  $I_0$  est de la forme  $[t_0, t_0 + T]$  ou  $[t_0, t_0 + T[$  ou  $[t_0, +\infty[$ , que  $f$  est continue sur  $I_0 \times \mathbb{R}^n$  et qu'il existe une fonction  $l \in \mathcal{L}^1(I_0)$  telle que

$$\forall t \in I_0, \forall y, z \in \mathbb{R}^n, \langle f(t, y) - f(t, z), y - z \rangle \leq l(t) \|y - z\|^2 \quad (4.3.15)$$

alors le problème de Cauchy (4.2.4) admet une solution et une seule. □

*Preuve* : L'existence est basée sur le théorème 4.3.4. Pour la démonstration de l'unicité, on renvoie à (CROUZEIX & MIGNOT, 1992). □

Le Théorème de Cauchy-Lipschitz est une conséquence du théorème précédant pour des hypothèses plus fortes :

**Théorème 4.3.6 (Cauchy-Lipschitz)** On suppose que la fonction  $f$  est continue sur  $I_0 \times \mathbb{R}^n$  et qu'il existe un réel  $L$  tel que :

$$\forall (t, y), (t, z) \in I_0 \times \mathbb{R}^n, \|f(t, y) - f(t, z)\| \leq L \|y - z\| \quad (4.3.16)$$

alors le problème de Cauchy (4.2.4) admet une solution et une seule. □

L'hypothèse (4.3.16) revient simplement à ce que  $f$  soit L-lipschitzienne par rapport à  $y$ .

Donnons enfin un corollaire qui est une application directe de ce théorème. Il concerne le cas des équations différentielles linéaires.

**Corollaire 4.3.1 (Cas des équations différentielles linéaires)** Si la fonction est linéaire par rapport à  $y$ , c'est à dire qu'elle se met sous la forme :

$$f(t, y) = A(t)y(t) + b(t) \quad (4.3.17)$$

et que l'application  $t \mapsto A(t)$  est continue sur  $I_0$  dans l'ensemble des applications linéaires de  $\mathbb{R}^n$ ,  $\mathcal{L}(\mathbb{R}^n)$  et que  $b \in \mathcal{C}^0(I_0)$  alors le problème de Cauchy admet une solution et une seule.  $\square$

*Preuve :* La preuve se scinde en deux cas :

1. si  $I_0$  est compact, on a

$$f(t, y) - f(t, z) = A(t)(y - z) \quad (4.3.18)$$

En posant  $L = \max_{t \in I_0} \|A(t)\|_{\mathcal{L}(\mathbb{R}^n)}$  qui existe puisque  $I_0$  est compact, on obtient que  $f$  est L-lipschitzienne par rapport à  $y$ .

2. Si  $I_0$ , n'est pas compact, par exemple  $I_0 = [t_0, t_0 + T[$ , on envisage la suite de  $t_n = t_0 + T - 1/n$ . Il existe une solution unique globale sur  $[t_0, t_n]$  que l'on note  $([t_0, t_n], y_n)$ . On pose  $y(t) = \lim_{n \rightarrow +\infty} y - n(t)$ . Cette limite existe puisque pour  $t_n \geq t$ ,  $y_n(t)$  est constante. La solution  $(I_0, y)$  est donc l'unique solution du problème.  $\square$

#### 4.3.4 Influence des perturbations sur les données

Derrière la question de l'influence des perturbations sur les données sur une solution se cache la notion de stabilité du problème de Cauchy. On souhaite en effet qu'une perturbation sur les données du problème de Cauchy, plus précisément sur  $f$  et sur la condition initiale  $y_0$  n'entraîne que des variations sur la solution  $y(t)$  qui peuvent être maîtrisées, et les plus petites possibles.

Sans la stabilité du problème continue, il n'y a pas d'espoir de construire des méthodes numériques (discrètes) qui soient stables.

Afin de préciser cette notion, on envisage un problème de Cauchy dit perturbé :

Étant donné :

- un intervalle  $I_0 \subset \mathbb{R}$ ,
- une fonction  $f$ , définie et continue sur  $I_0 \times \mathbb{R}^n$  à valeurs dans  $\mathbb{R}^n$  :
- une fonction  $g$ , définie et continue de  $I_0$  dans  $\mathbb{R}^n$
- un élément  $\alpha \in \mathbb{R}^n$  et deux réels  $(\xi_1, \xi_2) \in \mathbb{R}^2$

Trouver une fonction  $y_\varepsilon \in \mathcal{C}^1(I_0)$  telle que

$$\begin{cases} y'_\varepsilon(t) = f(t, y_\varepsilon(t)) + \xi_0 g(t), \forall t \in I_0, \forall y \in \mathbb{R}^n \\ y_\varepsilon(t_0) = y_0 + \xi_1 \alpha, t_0 \in I_0, \quad \text{Condition initiale} \end{cases} \quad (4.3.19)$$

Le théorème suivant permet de mesurer l'écart entre une solution  $y$  du problème de Cauchy et une solution  $y_\varepsilon$  du problème de Cauchy perturbé.

**Théorème 4.3.7** Si  $f$  est continue et vérifie :

$$\exists l \in \mathcal{L}^1(I_0), \quad \forall t \in I_0, \forall y, z \in \mathbb{R}^n, \langle f(t, y) - f(t, z), y - z \rangle \leq l(t) \|y - z\|^2 \quad (4.3.20)$$

et  $g$  est continue alors le problème de Cauchy perturbé admet une solution unique  $y_\varepsilon$  et

$$\|y(t) - y_\varepsilon(t)\| \leq \|\xi_1 \alpha\| \exp(L(t)) + \int_{t_0}^t \exp(L(t) - L(s)) \|\xi_0 g(s)\| ds \quad (4.3.21)$$

avec

$$L(t) = \int_{t_0}^t l(s) ds \quad (4.3.22)$$

□

On remarque donc que la stabilité de l'opérateur est contrôlée par la fonction  $L(t)$  qui doit rester "faible" pour espérer avoir une bonne stabilité du problème.

### 4.3.5 Régularité des solutions

Rappelons qu'une fonction vectorielle est dite de classe  $\mathcal{C}^k$  si elle admet des dérivées partielles continues jusqu'à l'ordre  $k$ .

On donne le théorème suivant sur la régularité des solutions :

**Théorème 4.3.8** Si  $f : \mathbb{R}^{n+1} \mapsto \mathbb{R}^n$  est de classe  $\mathcal{C}^k$ , alors toute solution du problème de Cauchy est de classe  $\mathcal{C}^{k+1}$

□

*Preuve* : La preuve se fait par récurrence :

- $k = 0$ ,  $f$  est continue. Par hypothèse,  $y : I_0 \mapsto \mathbb{R}^n$  est dérivable donc continue. Par conséquent,  $y'(t) = f(t, y(t))$  est continue et donc  $y$  est de classe  $\mathcal{C}^1$ .
- On suppose que le résultat est vrai à l'ordre  $k-1$ , alors  $y$  est au moins de classe  $\mathcal{C}^k$ . Comme  $f$  est de classe  $\mathcal{C}^k$ , il s'ensuit que  $y'$  est de classe  $\mathcal{C}^k$  et donc que  $y$  est de classe  $\mathcal{C}^{k+1}$ .

□

**Calcul des dérivées successives d'une solution  $y$  en fonction de  $f$**  Le calcul des dérivées successives de  $y$  par rapport à  $t$  se fait grâce aux dérivées totales de  $f$  par rapport à  $t$ . En dérivant la relation  $y'(t) = f(t, y(t))$ , on obtient :

$$y''(t) = \frac{\partial f(t, y(t))}{\partial t} + \nabla_y f(t, y(t)) \frac{\partial y(t)}{\partial t} \quad (4.3.23)$$

$$= \frac{\partial f(t, y(t))}{\partial t} + \nabla_y f(t, y(t)) y'(t) \quad (4.3.24)$$

$$= \frac{\partial f(t, y(t))}{\partial t} + \nabla_y f(t, y(t)) f(t, y) \quad (4.3.25)$$

On note de manière conventionnelle la dérivée totale de  $f$  par rapport au temps :

$$f^{(1)}(t, y) = \frac{\partial f(t, y(t))}{\partial t} + \nabla_y f(t, y(t)) f(t, y) \quad (4.3.26)$$

On a donc

$$y''(t) = f^{(1)}(t, y) \quad (4.3.27)$$

et d'une manière plus générale on notera la dérivée à l'ordre  $k$  :

$$y^{(k)}(t) = f^{(k-1)}(t, y) \quad (4.3.28)$$

Si on dérive une nouvelle fois cette dernière expression, on obtient :

$$y^{(k+1)}(t) = f^{(k)}(t, y) \quad (4.3.29)$$

$$= \frac{\partial f^{(k-1)}(t, y(t))}{\partial t} + \nabla_y f^{(k-1)}(t, y(t)) y'(t) \quad (4.3.30)$$

$$= \frac{\partial f^{(k-1)}(t, y(t))}{\partial t} + \nabla_y f^{(k-1)}(t, y(t)) f(t, y) \quad (4.3.31)$$

$$(4.3.32)$$

Le calcul des dérivées totales de  $f$  par rapport à  $t$  peuvent donc se faire grâce à la récurrence suivante :

$$f^{(0)}(t, y) = f(t, y) \quad (4.3.33)$$

$$f^{(l)}(t, y) = \frac{\partial f^{(l-1)}(t, y)}{\partial t} + \nabla_y f^{(l-1)}(t, y) f(t, y) \quad (4.3.34)$$

$$y^{(l+1)}(t) = f^{(l)}(t, y) \quad (4.3.35)$$

## 4.4 Méthodes explicites de résolution

### 4.4.1 Équations différentielles linéaires du premier ordre

### 4.4.2 Équations différentielles linéaires du second ordre



# Chapitre 5

## EDO – Méthodes numériques à un pas

### Contenu du Chapitre 5

---

|            |   |           |
|------------|---|-----------|
| <b>5.1</b> | <b>Introduction</b>                               | <b>74</b> |
| <b>5.2</b> | <b>Exemples de méthodes numériques classiques</b> | <b>75</b> |
| 5.2.1      | Principe de construction                          | 75        |
| 5.2.2      | La méthode d'Euler progressive ou explicite       | 75        |
| 5.2.3      | La méthode d'Euler rétrograde ou implicite        | 76        |
| 5.2.4      | La $\theta$ -méthode                              | 76        |
| <b>5.3</b> | <b>Étude générale des méthodes à un pas</b>       | <b>76</b> |
| 5.3.1      | Consistance, stabilité et convergence             | 76        |
| 5.3.2      | Condition de convergence des méthodes à un pas.   | 78        |
| 5.3.3      | Ordre d'une méthode à un pas                      | 79        |
| <b>5.4</b> | <b>Les méthodes de développement de Taylor</b>    | <b>80</b> |
| <b>5.5</b> | <b>Les méthodes de Runge–Kutta</b>                | <b>82</b> |
| 5.5.1      | Principe de construction                          | 82        |
| 5.5.2      | Exemples  | 84        |
| 5.5.3      | Stabilité des méthodes de Runge–kutta             | 87        |
| 5.5.4      | Ordre des méthodes de Runge–kutta                 | 87        |
| <b>5.6</b> | <b>Contrôle du pas de temps</b>                   | <b>89</b> |
| 5.6.1      | Principe générale du contrôle                     | 89        |
| 5.6.2      | Applications aux méthodes présentées.             | 89        |

---

## 5.1 Introduction

Dans tout ce chapitre, on va considérer le problème de Cauchy dans  $\mathbb{R}$ , c'est à dire pour des fonctions à valeurs réelles. Les résultats et les méthodes qui sont donnés dans la suite peuvent généraliser sans problème dans  $\mathbb{R}^n$ . Rappelons la définition du problème de Cauchy.

**Définition 5.1.1** On appelle *problème de Cauchy* le problème suivant :

Étant donnés :

- un intervalle  $I_0 \subset \mathbb{R}$ ,
- une fonction  $f$ , définie et continue sur  $I_0 \times \mathbb{R}$  à valeurs dans  $\mathbb{R}$  :

$$f : I_0 \times \mathbb{R} \mapsto \mathbb{R} \quad (5.1.1)$$

$$(t, y) \mapsto f(t, y) \quad (5.1.2)$$

Trouver une fonction  $y \in \mathcal{C}^1(I_0)$  telle que

$$\begin{cases} y'(t) = f(t, y(t)), \forall t \in I_0, \forall y \in \mathbb{R}^n \\ y(t_0) = \eta, t_0 \in I_0, \quad \text{Condition initiale} \end{cases} \quad (5.1.3)$$

□

On supposera que  $I_0$  est de la forme  $[t_0, t_0+T]$  ou  $[t_0, t_0+T[$  avec  $T > 0$  ou encore  $[t_0, +\infty[$ . De plus, on supposera pour simplifier que  $f$  vérifie les conditions du théorème de Cauchy- Lipschitz, c'est à dire

$$\exists k, \forall (t, y)(t, z) \in I_0 \times \mathbb{R}, |f(t, y) - f(t, z)| \leq k|y - z| \quad (5.1.4)$$

Le but des méthodes numériques de résolution des EDO est approcher le problème de Cauchy continue en temps par un problème discrétisé. Pour cela, on se donne une subdivision de l'intervalle  $I_0$ , notée  $t_0 < t_1 < \dots < t_n < \dots < t_N$  et on construit une suite  $y_n, 0 \leq n \leq N$  d'approximation de valeurs  $y(t_n), 0 \leq n \leq N$ . On notera le pas de la discrétisation  $h_n = t_{n+1} - t_n$  et  $h = \max_{0 \leq n \leq N} h_n$ .

On désignera de manière équivalente une méthode numérique de résolution des EDO par un schéma numérique.

Terminons enfin par la définition des méthodes à un pas :

**Définition 5.1.2** On appelle méthode de résolution numérique à un pas une méthode qui s'écrit :

$$\begin{cases} y_{n+1} = y_n + h_n \Phi(t_n, y_n, h_n) \\ y_0 = \hat{\eta} \end{cases} \quad (5.1.5)$$

où  $\Phi$  est une fonction continue de  $I_0 \times \mathbb{R} \times [0, h^*]$  avec  $h^* > 0$ .

□

La méthode est dite à un pas car la valeur de l'approximation au pas  $t_{n+1}$  ne dépend que de la valeur de l'approximation au pas de temps précédent.

## 5.2 Exemples de méthodes numériques classiques

Afin de préciser les idées sur les méthodes numériques de résolution des EDO, donnons dans cette partie quelques exemples de méthodes classiques.

### 5.2.1 Principe de construction

Le principe de construction des méthodes simples s'appuie sur la remarque suivante. Comme on l'a vu au chapitre précédent, le problème de Cauchy (5.1.3) peut être écrit sous la forme intégrale sur l'intervalle  $[t_n, t_{n+1}]$ ,  $0 \leq n \leq N - 1$  :

$$\int_{t_n}^{t_{n+1}} y'(t) dt = y(t_{n+1}) - y(t_n) = \int_{t_n}^{t_{n+1}} f(t, y(t)) dt \quad (5.2.1)$$

soit

$$y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} f(t, y(t)) dt \quad (5.2.2)$$

La construction de méthodes numériques simples peut se faire en se donnant une approximation de l'intégrale

$$\int_{t_n}^{t_{n+1}} f(t, y(t)) dt \approx h_n \Phi(t_n, y_n, h_n) \quad (5.2.3)$$

où  $y_n \approx y(t_n)$  la suite des approximations de  $y$  en  $t_n$ . La suite des approximations est donc construite par :

$$\begin{cases} y_{n+1} = y_n + h_n \Phi(t_n, y_n, h_n) \\ y_0 = \hat{\eta} \end{cases} \quad (5.2.4)$$

où  $\hat{\eta}$  est une approximation numérique de la condition initiale  $\eta$ .

Nous verrons dans la suite que l'utilisation de formules de quadrature pour approcher l'intégrale (5.2.3) conduit à la construction générale des méthodes de Runge-Kutta. Donnons dans la suite des exemples élémentaires de méthodes numériques construites de la sorte.

### 5.2.2 La méthode d'Euler progressive ou explicite

La plus simple des méthodes est la méthode d'Euler progressive. L'intégrale (5.2.3) est approchée par une méthode des rectangles à gauche, soit :

$$\int_{t_n}^{t_{n+1}} f(t, y(t)) dt \approx h_n \Phi(t_n, y_n, h_n) = h_n f(t_n, y_n) \quad (5.2.5)$$

La méthode d'Euler progressive s'écrit donc :

$$\boxed{y_{n+1} = y_n + h_n f(t_n, y_n)} \quad (5.2.6)$$

avec la condition initiale

$$y_0 = \hat{\eta} \quad (5.2.7)$$

## Interprétation géométrique

### 5.2.3 La méthode d'Euler rétrograde ou implicite

Pour la méthode d'Euler rétrograde, l'intégrale (5.2.3) est approchée par une méthode des rectangles à droite :

$$\int_{t_n}^{t_{n+1}} f(t, y(t)) dt \approx h_n \Phi(t_n, y_n, h_n) = h_n f(t_{n+1}, y_{n+1}) \quad (5.2.8)$$

La méthode d'Euler rétrograde s'écrit donc :

$$\boxed{y_{n+1} = y_n + h_n f(t_{n+1}, y_{n+1})} \quad (5.2.9)$$

avec la condition initiale

$$y_0 = \hat{\eta} \quad (5.2.10)$$

Cette méthode est souvent la méthode d'Euler implicite. On constate en particulier que le terme recherché  $y_{n+1}$  apparaît aussi au second membre. Il est donc nécessaire de résoudre à chaque pas une équation du type :

$$F(y_{n+1}) = 0 \quad (5.2.11)$$

Si  $f$  est linéaire, il suffit alors de résoudre un système linéaire à chaque pas. Si  $f$  est non linéaire, cette équation peut être résolue par une méthode dédiée aux systèmes non linéaires (point fixe, Newton–Raphson, *etc.*...)

## Interprétation géométrique

### 5.2.4 La $\theta$ -méthode

La  $\theta$ -méthode est construite en généralisant les méthodes d'Euler de la façon suivante :

$$\boxed{y_{n+1} = y_n + h_n [(1 - \theta)f(t_n, y_n) + \theta f(t_{n+1}, y_{n+1})]}, \quad \theta \in [0, 1] \quad (5.2.12)$$

Pour  $\theta = 0$ , on retrouve la méthode d'Euler progressive et pour  $\theta = 1$ , on retrouve la méthode d'Euler rétrograde. Pour  $\theta = \frac{1}{2}$ , on remarque que l'intégrale est en fait approchée par une méthode des trapèzes, et l'on appelle usuellement cette méthode la méthode de Crank–Nicholson.

## 5.3 Étude générale des méthodes à un pas

### 5.3.1 Consistance, stabilité et convergence

Donnons quelques définitions utiles à l'étude des méthodes numériques à un pas.

**Définition 5.3.1** On appelle erreur à l'instant  $t_n$  de la méthode à un pas, notée  $e_n$  la valeur suivante :

$$e_n = |y_n - y(t_n)| \quad (5.3.1)$$

□

Cette définition est une définition globale de l'erreur qui mesure la précision de la méthode à un pas.

**Définition 5.3.2** On appelle l'erreur de consistance à l'instant  $t_n$  de la méthode à un pas, notée  $\epsilon_n$  la valeur suivante :

$$\epsilon_n = y(t_n + 1) - y(t_n) - h_n \Phi(t_n, y(t_n), h_n) \quad (5.3.2)$$

□

Dans cette définition, il faut prendre garde à ne pas confondre la valeur exacte de  $y(t_n)$  avec sa valeur approchée par la schéma  $y_n$ . L'erreur de consistance mesure à chaque pas l'erreur commise par le schéma par rapport à une valeur exacte à ce même temps. C'est une notion d'erreur locale où l'on mesure la déviation du schéma à chaque pas de temps.

**Définition 5.3.3** On dit que la méthode à un pas est consistante si pour toute solution du problème de Cauchy, on a

$$\lim_{h \rightarrow 0} \sum_0^{N-1} |\epsilon_n| = \lim_{h \rightarrow 0} \sum_0^{N-1} |y(t_n + 1) - y(t_n) - h_n \Phi(t_n, y(t_n), h_n)| = 0 \quad (5.3.3)$$

□

Une autre notion importante est la stabilité du schéma. Elle permet en particulier de mesurer les effets des erreurs d'arrondis.

**Définition 5.3.4** On dit qu'une méthode à un pas est stable si il existe une constante  $M$  indépendante de  $h$  telle que pour toutes suites  $y_n$  et  $\hat{y}_n$  vérifiant :

$$y_{n+1} = y_n + h_n \Phi(t_n, y_n, h_n) \quad (5.3.4)$$

$$\hat{y}_{n+1} = \hat{y}_n + h_n \Phi(t_n, \hat{y}_n, h_n) + \rho_n \quad (5.3.5)$$

on ait

$$\max_{0 \leq n \leq N} |y_n - \hat{y}_n| \leq M \left( |y_0 - \hat{y}_0| + \sum_0^N |\rho_n| \right) \quad (5.3.6)$$

□

Si  $\hat{y}_n$  est la suite des approximations de  $y(t_n)$  perturbée par des erreurs, la notion de stabilité implique que ces erreurs sur les données et sur le calcul n'impliquent que de faibles erreurs sur les résultats. Cette propriété est absolument nécessaire du point de vue de la mise en œuvre numérique. La suite  $\rho_n$  représente les erreurs qui peuvent être introduites à chaque pas de temps (Erreur de troncature, d'évaluation, etc) et  $|y_0 - \hat{y}_0|$  représente l'erreur sur la donnée initiale.

Enfin nous terminons ces définitions par la notion de convergence d'une méthode à un pas :

**Définition 5.3.5** On dit qu'une méthode numérique à un pas est convergente si

$$\max_{0 \leq n \leq N} e_n = \max_{0 \leq n \leq N} |y(t_n) - y_n| \text{ tend vers } 0 \quad (5.3.7)$$

lorsque  $h$  tends vers 0 et  $\hat{\eta}$  tends vers  $\eta$ .

□

### 5.3.2 Condition de convergence des méthodes à un pas.

Nous avons défini dans le § précédent deux notions fondamentales pour l'étude des méthodes à un pas : la stabilité et la consistance. La stabilité garantit que les erreurs ne sont pas amplifiées par le schéma numérique. La consistance garantit que le schéma ne s'éloigne pas trop de la solution exacte à chaque pas. Les deux propriétés réunies permettent de conclure à la convergence de la méthode comme nous allons le voir dans le théorème suivant :

**Théorème 5.3.1** Si  $f$  est une fonction qui vérifie les hypothèses du théorème de Cauchy-Lipschitz, si de plus la méthode à un pas est stable et consistante, alors la méthode est convergente.

□

*Preuve :* Soit  $y_n$  la suite donnée par la méthode à un pas :

$$y_{n+1} = y_n + h_n \Phi(t_n, y_n, h_n), \quad (5.3.8)$$

on considère l'erreur de consistance de la méthode à un pas :

$$\epsilon_n = y(t_{n+1}) - y(t_n) - h_n \Phi(t_n, y(t_n), h_n) \quad (5.3.9)$$

La méthode est consistante donc :

$$\lim_{h \rightarrow 0} \sum_0^N |\epsilon_n| = 0 \quad (5.3.10)$$

Soit  $z_n = y(t_n)$  la suite donnée par les valeurs de  $y$  aux points  $t_n$ , la définition de l'erreur de consistance donne :

$$z_{n+1} = z_n + h_n \Phi(t_n, z_n, h_n) + \epsilon_n \quad (5.3.11)$$

que l'on peut prendre comme suite de valeurs perturbées  $\rho_n$  pour la définition de stabilité. Si la méthode est stable, on a

$$\max_{0 \leq n \leq N} |z_n - y_n| = \max_{0 \leq n \leq N} |y(t_n) - y_n| \leq M \left( |\eta - \hat{\eta}| + \lim_{h \rightarrow 0} \sum_0^N |\epsilon_n| \right) \quad (5.3.12)$$

sachant que

$$\lim_{h \rightarrow 0} \sum_0^N |\epsilon_n| = 0 \quad (5.3.13)$$

$$\lim_{h \rightarrow 0} |\eta - \hat{\eta}| = 0 \quad (5.3.14)$$

on obtient que :

$$\lim_{h \rightarrow 0} \max_{0 \leq n \leq N} |y(t_n) - y_n| = 0 \quad (5.3.15)$$

La méthode est donc convergente. □

L'étude de la convergence des méthodes numériques pour les EDO est le plus souvent ramené à l'étude de la consistance et de la stabilité. Donnons des conditions de consistance et de stabilité des méthodes numériques.

**Théorème 5.3.2** Une condition nécessaire et suffisante pour qu'une méthode à un pas soit consistante est que

$$\Phi(t, y, 0) = f(t, y) \quad (5.3.16)$$

□

**Théorème 5.3.3** Une condition nécessaire et suffisante pour qu'une méthode à un pas soit stable est qu'il existe une constante  $\Lambda$  telle que :

$$\forall t \in [t_0, t_0 + T], \forall y, z \in \mathbb{R}, \forall h \in [0, h^*], |\Phi(t, y, h) - \Phi(t, z, h)| \leq \Lambda |y - z| \quad (5.3.17)$$

La constante de stabilité est alors donnée par

$$M = \exp(\Lambda T) \quad (5.3.18)$$

□

### 5.3.3 Ordre d'une méthode à un pas

Afin de préciser la notion de consistance, on introduit la notion d'ordre d'une méthode à un pas :

**Définition 5.3.6** On dit qu'une méthode à un pas est d'ordre  $p$  si il existe un réel  $K$  indépendant de  $y$  et de  $\Phi$  tel que :

$$\sum_0^N |\epsilon_n| \leq Kh^{p+1} \quad (5.3.19)$$

□

Si l'on introduit cette définition pour un schéma stable dans la théorème de convergence, on constate que la précision est d'autant meilleur que l'ordre est élevé. Le but des paragraphes suivants est de construire des méthodes d'ordre plus élevé pour un pas de discrétisation donné.

### 5.3.1 Note de Rédaction

Ajouter le calcul de l'erreur globale et le choix du pas de temps (DEMAILLY, 1996).

Avant cela donnons un théorème qui fournit une condition pour qu'un schéma soit d'ordre voulu :

**Théorème 5.3.4** On suppose que  $f$  est  $(p-1)$  fois différentiable dans  $I_0 \times \mathbb{R}$  et que les fonctions  $\Phi, \frac{\partial \Phi}{\partial h}, \dots, \frac{\partial^{p-1} \Phi}{\partial h^{p-1}}$  existent et soit continues dans  $I_0 \times \mathbb{R} \times [0, h^*]$ , alors une condition nécessaire et suffisante pour que la méthode soit d'ordre  $p$  est que

$$\forall (t, y) \in I_0 \times \mathbb{R}, \frac{\partial^l \Phi}{\partial h^l}(t, y, 0) = \frac{1}{l+1} f^{(l)}(t, y), \quad 0 \leq l \leq p-1 \quad (5.3.20)$$

où  $f^{(l)}(t, y)$  désigne la dérivée totale par rapport à  $t$ , c'est à dire  $f^{(l)}(t, y) = \frac{\partial f^{(l-1)}(t, y)}{\partial t} + \nabla_y f^{(l-1)}(t, y) f(t, y)$

□

Ce dernier théorème conduit à la construction des méthodes de Taylor que nous allons exposer dans la suite.

### Ordre de la méthode d'Euler progressive

#### 5.3.2 Note de Rédaction

A compléter

## 5.4 Les méthodes de développement de Taylor

Dans les méthodes de développement de Taylor, on choisit  $\phi$  de la manière la plus simple pour respecter les conditions (5.3.20), soit :

$$\Phi(t, y, h) = f(t, y) + \frac{h}{2} f^{(1)}(t, y) + \dots + \frac{h^{p-1}}{p!} f^{(p-1)}(t, y) \quad (5.4.1)$$

Supposons en effet que  $f$  est de classe  $\mathcal{C}^p$ , comme nous l'avons vu au chapitre précédent, une solution  $y$  du problème de Cauchy est de classe  $\mathcal{C}^{p+1}$  et sa dérivée d'ordre  $k$  s'écrit :

$$y^{(k)}(t) = f^{(k-1)}(t, y) \quad (5.4.2)$$

Le développement de Taylor de  $y$  d'ordre  $p$  autour de  $t_n$  conduit donc à

$$y(t_n + h_n) = y(t_n) + \sum_{k=1}^p \frac{1}{k!} h_n^k y^{(k)}(t_n) + o(h_n^p) \quad (5.4.3)$$

$$= y(t_n) + \sum_{k=1}^p \frac{1}{k!} h_n^k f^{(k-1)}(t_n, y) + o(h_n^p) \quad (5.4.4)$$

La méthode de Taylor d'ordre  $p$  consiste à remplacer  $y$  par son développement de Taylor à l'ordre  $p$  :

$$\left\{ \begin{array}{l} y_{n+1} = y_n + \sum_{k=1}^p \frac{1}{k!} h_n^k f^{(k-1)}(t_n, y_n) \end{array} \right. \quad (5.4.5)$$

On constate que d'après les définitions des méthodes à un pas, cet algorithme correspond au choix :

$$\Phi(t, y, h) = \sum_{k=1}^p \frac{1}{k!} h^k f^{(k-1)}(t, y) \quad (5.4.6)$$

Pour se convaincre que la méthode de Taylor est effectivement d'ordre  $p$ , il suffit de calculer l'erreur de consistance :

$$\epsilon_n = y(t_{n+1}) - y(t_n) - \sum_{k=1}^p \frac{1}{k!} h_n^k f^{(k-1)}(t_n, y(t_n)) \quad (5.4.7)$$

$$= y(t_n + h_n) - y(t_n) - \sum_{k=1}^p \frac{1}{k!} h_n^k y^{(k)}(t_n) \quad (5.4.8)$$

La formule de Taylor à l'ordre  $p+1$  donne :

$$\epsilon_n = \frac{1}{(p+1)!} h_n^{p+1} y^{(p+1)}(t_n) + o(h_n^{p+1}) \quad (5.4.9)$$

$$= \frac{1}{(p+1)!} h_n^{p+1} f^{(p)}(t_n, y_n) + o(h_n^{p+1}) \quad (5.4.10)$$

On peut noter que pour  $p=1$ , on retrouve la méthode d'Euler progressive.

**Remarques** La méthode souffre de deux inconvénients majeurs qui font qu'elle est rarement utilisée pour  $p \geq 2$  :

1. Le calcul des dérivées totales  $f^{(k)}(t_n, y_n)$  est souvent complexe et coûteux en temps de calcul. De plus, ces dérivées ne sont pas forcément connues même si l'on se doute que l'évolution est souvent suffisamment régulière.
2. La méthode suppose une forte régularité de  $f$ . Dans la pratique, même si  $f$  est théoriquement régulière, la méthode est limitée par des forts gradients qui rendent le problème raide.

## 5.5 Les méthodes de Runge–Kutta

Les méthodes de Runge–Kutta sont sûrement les méthodes à un pas les plus largement répandues pour la résolution numérique des EDO. Nous allons en présenter les principes de constructions et en donner quelques exemples.

### 5.5.1 Principe de construction

L'idée des méthodes de Runge–Kutta est de calculer par récurrence les points  $(t_{n+1}, y_{n+1})$  à chaque itération en utilisant des points intermédiaires  $(t_{n,i}, y_{n,i})$ .

Pour cela on définit des points intermédiaires dans l'intervalle  $[t_n, t_{n+1}]$  de la manière suivante :

$$t_{n,i} = t_n + c_i h_n, 1 \leq i \leq q, c_i \in [0, 1] \quad (5.5.1)$$

A chacun de ces points intermédiaires, on associe la pente  $p_{n,i}$  définie par :

$$p_{n,i} = f(t_{n,i}, y_{n,i}) \quad (5.5.2)$$

Si  $y$  est une solution du problème de Cauchy, on a grâce au changement de variable  $t = t_n + uh_n$

$$y(t_{n,i}) = y(t_n) + \int_{t_n}^{t_{n,i}} f(t, y(t)) dt \quad (5.5.3)$$

$$= y(t_n) + h_n \int_0^{c_i} f(t_n + uh_n, y(t_n + uh_n)) du \quad (5.5.4)$$

De même, on a

$$y(t_{n+1}) = y(t_n) + h_n \int_0^1 f(t_n + uh_n, y(t_n + uh_n)) du \quad (5.5.5)$$

Pour calculer les points  $(t_{n,i}, y_{n,i})$ , on se donne une formule de quadrature pour chaque  $i = 1, 2, \dots, q$  :

$$\int_0^{c_i} g(u) du \approx \sum_{1 \leq j \leq i} a_{ij} g(c_j), \quad (Q_i) \quad (5.5.6)$$

De même pour calculer, le point  $(t_{n+1}, y_{n+1})$  on se donne une formule de quadrature :

$$\int_0^1 g(u) du \approx \sum_{1 \leq j \leq q} b_j g(c_j), \quad (Q) \quad (5.5.7)$$

En appliquant ces méthodes à  $g(u) = f(t_n + uh_n, y(t_n + uh_n))$ , il vient :

$$y(t_{n,i}) \approx y(t_n) + h_n \sum_{1 \leq j \leq i} a_{ij} f(t_n + c_j h_n, y_n + c_j h_n) \quad (5.5.8)$$

$$\approx y(t_n) + h_n \sum_{1 \leq j \leq i} a_{ij} f(t_{n,j}, y_{n,j}) \quad (5.5.9)$$

$$\approx y(t_n) + h_n \sum_{1 \leq j \leq i} a_{ij} p_{n,j} \quad (5.5.10)$$

$$y(t_{n+1}) \approx y(t_n) + h_n \sum_{1 \leq j \leq q} b_j f(t_n + c_j h_n, y_n + c_j h_n) \quad (5.5.11)$$

$$\approx y(t_n) + h_n \sum_{1 \leq j \leq q} b_j p_{n,j} \quad (5.5.12)$$

La méthode de Runge–Kutta correspondante est donc définie par la récurrence suivante :

$$\left\{ \begin{array}{l} \left\{ \begin{array}{l} t_{n,i} = t_n + c_i h_n, \\ y_{n,i} = y_n + h_n \sum_{1 \leq j \leq i} a_{ij} p_{n,j} \\ p_{n,j} = f(t_{n,j}, y_{n,j}) \end{array} \right\} \text{ pour } 1 \leq i \leq q \\ t_{n+1} = t_n + h_n, \\ y_{n+1} = y_n + h_n \sum_{1 \leq j \leq q} b_j p_{n,j} \end{array} \right. \quad (5.5.13)$$

Les méthodes de Runge–Kutta sont définies entièrement par les données :

1. des points intermédiaires, c'est à dire des  $c_i, 1 \leq i \leq q$
2. de la formule de quadrature (5.5.6), c'est à dire des  $a_{ij}, 1 \leq i \leq q, 1 \leq j \leq i$
3. de la formule de quadrature (5.5.7), c'est à dire des  $b_j, 1 \leq j \leq q$

On les représente conventionnellement par le tableau 5.1. La méthode de Runge–Kutta est dite explicite si les termes diagonaux des la matrices  $A = [a_{ij}]$  sont nuls. Dans les autres cas, il s'agit de méthodes explicites où ils faut résoudre une équation non linéaire pour obtenir  $y_{n,i}$ . On ne considérera dans la suite que des méthodes de Runge–Kutta explicite.

On suppose de plus généralement que les formules de quadrature (5.5.6) et (5.5.7) sont au moins d'ordre 0. Cette implique que

$$c_i = \sum_{1 \leq j \leq i} a_{ij}, \quad \sum_{1 \leq j \leq q} b_j = 1 \quad (5.5.14)$$

En particulier, on aura toujours pour les méthodes de Runge–kutta explicite :

$$c_1 = 0, \quad t_{n,1} = t_n, \quad y_{n,1} = y_n, \quad p_{n,1} = f(t_n, y_n) \quad (5.5.15)$$

|             |           |            |            |          |              |          |
|-------------|-----------|------------|------------|----------|--------------|----------|
| $(Q_1)$     | $c_1$     | $a_{11}$   | $0$        | $\dots$  | $0$          | $0$      |
| $(Q_2)$     | $c_2$     | $a_{21}$   | $a_{22}$   | $\dots$  | $0$          | $0$      |
| $\vdots$    | $\vdots$  | $\vdots$   | $\ddots$   | $\vdots$ | $\vdots$     |          |
| $(Q_{q-1})$ | $c_{q-1}$ | $a_{q-11}$ | $a_{q-12}$ | $\dots$  | $a_{q-1q-1}$ | $0$      |
| $(Q_q)$     | $c_q$     | $a_{q1}$   | $a_{q2}$   | $\dots$  | $a_{qq-1}$   | $a_{qq}$ |
| $(Q)$       |           | $b_1$      | $b_2$      | $\dots$  | $\dots$      | $b_q$    |

TAB. 5.1 – Tableau de représentation d'une méthode de Runge-Kutta

### 5.5.2 Exemples

**Exemple 1 : Méthode explicite avec  $q = 1$**  Le seul choix possible pour une méthode explicite est  $c_1 = 0, a_{11} = 0, b_1 = 1$  soit le tableau suivant :

|     |     |
|-----|-----|
| $0$ | $0$ |
|     | $1$ |

La méthode est donc donnée par

$$\left\{ \begin{array}{l} \left\{ \begin{array}{l} t_{n,1} = t_n, \\ y_{n,1} = y_n \\ p_{n,1} = f(t_{n,1}, y_{n,1}) \end{array} \right. \\ t_{n+1} = t_n + h_n, \\ y_{n+1} = y_n + h_n p_{n,1} = y_n + h_n f(t_n, y_n) \end{array} \right. \quad (5.5.16)$$

On reconnaît donc la méthode d'Euler progressive.

**Exemple 2 : Méthode explicite avec  $q = 2$**  Pour  $q = 2$ , on considère un tableau de la forme

|          |                         |                     |                        |
|----------|-------------------------|---------------------|------------------------|
| $0$      | $0$                     | $0$                 |                        |
| $\alpha$ | $\alpha$                | $0$                 | où $\alpha \in ]0, 1]$ |
|          | $1 - \frac{1}{2\alpha}$ | $\frac{1}{2\alpha}$ |                        |

Dans ce cas, l'algorithme s'écrit :

$$\left\{ \begin{array}{l} \left\{ \begin{array}{l} p_{n,1} = f(t_n, y_n) \\ t_{n,2} = t_n + \alpha h_n, \\ y_{n,2} = y_n + \alpha h_n p_{n,1} \\ p_{n,2} = f(t_{n,2}, y_{n,2}) \end{array} \right. \\ t_{n+1} = t_n + h_n, \\ y_{n+1} = y_n + h_n \left[ \left(1 - \frac{1}{2\alpha}\right) p_{n,1} + \left(\frac{1}{2\alpha}\right) p_{n,2} \right] \end{array} \right. \quad (5.5.17)$$

Sous forme condensée, l'algorithme peut s'écrire :

$$y_{n+1} = y_n + h_n \left[ \left(1 - \frac{1}{2\alpha}\right) f(t_n, y_n) + \left(\frac{1}{2\alpha}\right) f\left(t_n + \alpha h_n, y_n + \alpha h_n f(t_n, y_n)\right) \right] \quad (5.5.18)$$

Cette forme condensée n'est généralement pas utilisée car elle est moins efficace que l'algorithme précédent que ne nécessite que deux évaluations de  $f$  au lieu de 3.

Les cas particuliers suivants sont particulièrement utilisés :

- $\alpha = \frac{1}{2}$ , on trouve alors la méthode du point milieu :

$$\boxed{y_{n+1} = y_n + h_n f\left(t_n + \frac{1}{2}h_n, y_n + \frac{1}{2}h_n f(t_n, y_n)\right)} \quad (5.5.19)$$

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

$$\int_0^1 g(u) du \approx g\left(\frac{1}{2}\right) \quad (5.5.20)$$

- $\alpha = 1$ , on trouve alors la méthode de Heun :

$$\boxed{y_{n+1} = y_n + h_n \left[ \frac{1}{2}f(t_n, y_n) + \frac{1}{2}f(t_{n+1}, y_n + h_n f(t_n, y_n)) \right]} \quad (5.5.21)$$

qui est basée sur la méthode d'intégration des trapèzes :

$$\int_0^1 g(u) du \approx \frac{1}{2}(g(0) + g(1)) \quad (5.5.22)$$

**Exemple 3 : Méthode explicite classique de Runge–Kutta avec  $q = 4$**  Il s'agit de la méthode classique lorsque l'on parle de "la" méthode de Runge–kutta. Elle est donnée par le

tableau suivant :

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

L'algorithme correspondant s'écrit sous la forme suivant :

$$\left( \begin{array}{l} p_{n,1} = f(t_n, y_n) \\ t_{n,2} = t_n + \frac{1}{2}h_n, \\ y_{n,2} = y_n + \frac{1}{2}h_n p_{n,1} \\ p_{n,2} = f(t_{n,2}, y_{n,2}) \\ y_{n,3} = y_n + \frac{1}{2}h_n p_{n,2} \\ p_{n,3} = f(t_{n,2}, y_{n,3}) \quad \text{noter que } t_{n,3} = t_{n,2} \\ t_{n,4} = t_n + h_n \quad \text{noter que } t_{n,3} = t_{n+1} \\ y_{n,4} = y_n + h_n p_{n,3} p_{n,4} = f(t_{n,4}, y_{n,4}) \\ t_{n+1} = t_n + h_n, \\ y_{n+1} = y_n + h_n \left[ \frac{1}{6}p_{n,1} + \frac{2}{6}p_{n,2} + \frac{2}{6}p_{n,3} + \frac{1}{6}p_{n,4} \right] \end{array} \right) \quad (5.5.23)$$

On verra plus loin que c'est une méthode d'ordre 4 qui s'appuie sur les formules d'intégration suivante :

$$(Q_2) \quad \int_0^{\frac{1}{2}} g(u) du \approx \frac{1}{2}g(0) \quad \text{rectangle à gauche} \quad (5.5.24)$$

$$(Q_3) \quad \int_0^{\frac{1}{2}} g(u) du \approx \frac{1}{2}g\left(\frac{1}{2}\right) \quad \text{rectangle à droite} \quad (5.5.25)$$

$$(Q_4) \quad \int_0^1 g(u) du \approx \frac{1}{2}g\left(\frac{1}{2}\right) \quad \text{point milieu} \quad (5.5.26)$$

$$(Q) \quad \int_0^1 g(u) du \approx \frac{1}{6}g(0) + \frac{2}{6}g\left(\frac{1}{2}\right) + \frac{2}{6}g\left(\frac{1}{2}\right) + \frac{1}{6}g(1) \quad \text{Simpson} \quad (5.5.27)$$

$$(5.5.28)$$

### 5.5.3 Stabilité des méthodes de Runge–kutta

Pour résumé, les méthodes de Runge–Kutta sont des méthodes données par :

$$y_{n+1} = y_n + h_n \Phi(t_n, y_n, h_n) \quad (5.5.29)$$

$$\text{avec } \Phi(t_n, y_n, h_n) = \sum_{1 \leq j \leq q} b_j p_{n,j} \quad (5.5.30)$$

$$(5.5.31)$$

La fonction  $\Phi(t, y, h)$  est donnée de manière explicite par :

$$\phi(t, y, h) = \sum_{1 \leq j \leq q} b_j f(t + c_j h, y_j) \quad (5.5.32)$$

$$y_i = y + h \sum_{1 \leq j \leq i} a_{ij} f(t + c_j h, y_j) \quad (5.5.33)$$

On constate que les méthodes de Runge-Kutta sont bien des méthodes à un pas.

Pour démontrer la stabilité des méthodes de Runge-Kutta il faut reprendre les conditions du théorème 5.3.3 et les démontrer pour  $\Phi$ . En supposant une fois de plus que  $f$  est  $k$ -lipschitzienne par rapport à  $y$ , on aboutit au théorème suivant :

**Théorème 5.5.1** Les méthodes de Runge–Kutta sont stables, avec une constante de stabilité égale à

$$S = \exp(\Lambda T) \quad (5.5.34)$$

$$\text{avec } \Lambda = k \sum_{1 \leq j \leq q} |b_j| (1 + (\alpha k h) + \dots + (\alpha k h)^{j-1}), \quad \alpha = \max_i \left( \sum_{1 \leq j \leq i} |a_{ij}| \right) \quad (5.5.35)$$

□

*Preuve* : On renvoie à (DEMAILLY, 1996).

□

**Remarque** Dans le cas fréquent où les coefficients,  $b_j$  sont positives, on alors :

$$\Lambda = k (1 + (\alpha k h) + \dots + (\alpha k h)^{j-1}), \quad (5.5.36)$$

Si de plus les coefficients  $a_{ij}$  sont positifs, alors on a  $\alpha = \max_i c_i$ . dès lors lorsque  $h$  est petit devant  $1/\alpha k$ , la constante de stabilité est donc  $e^{kT}$  ce qui est pratiquement la borne inférieure pour cette constante. Les méthodes de Runge–Kutta sont généralement très stables.

### 5.5.4 Ordre des méthodes de Runge–kutta

La détermination de l'ordre des méthodes Runge-Kutta se fait par l'intermédiaire par l'application du théorème 5.3.4. Il convient donc de calculer les dérivées successives de  $\Phi(t, y, 0)$  par rapport à  $h$  à partir de l'expression (5.5.32).

Ces calculs sont généralement longs et pénibles, on peut tout de même les réaliser à titre d'exemples pour les premier ordre.

**Ordre  $\geq 1$**  La condition du théorème 5.3.4 porte sur  $\Phi(t, y, 0)$  or

$$\Phi(t, y, 0) = \sum_{1 \leq j \leq q} b_j f(t, y) = f(t, y) \quad (5.5.37)$$

et donc la condition est bien vérifiée. Les méthodes de Runge–kutta sont donc au moins d'ordre  $\geq 1$ .

**Ordre  $\geq 2$**  La condition du théorème 5.3.4 porte sur  $\frac{\partial \Phi(t, y, h)}{\partial h}$  qu'il faut calculer :

$$\frac{\partial \Phi(t, y, h)}{\partial h} = \sum_{1 \leq j \leq q} b_j \left( c_j \frac{\partial f(t + c_j h, y_j)}{\partial t} + \frac{\partial f(t + c_j h, y_j)}{\partial y} \frac{\partial y_j}{\partial h} \right), \quad (5.5.38)$$

$$\frac{\partial y_i}{\partial h} = \left( \sum_{1 \leq j \leq i} a_{ij} f(t + c_j h, y_j) \right) + h \left( \sum_{1 \leq j \leq i} a_{ij} \left( c_j \frac{\partial f(t + c_j h, y_j)}{\partial t} + \frac{\partial f(t + c_j h, y_j)}{\partial y} \frac{\partial y_j}{\partial h} \right) \right) \quad (5.5.39)$$

soit en  $h = 0$ , on obtient :

$$\left. \frac{\partial y_i}{\partial h} \right|_{h=0} = \sum_{1 \leq j \leq i} a_{ij} f(t, y) = c_i f(t, y) \quad (5.5.40)$$

$$\left. \frac{\partial \Phi(t, y, h)}{\partial h} \right|_{h=0} = \sum_{1 \leq j \leq q} b_j c_j \left( \frac{\partial f(t, y)}{\partial t} + \frac{\partial f(t, y)}{\partial y} f(t, y) \right), \quad (5.5.41)$$

$$= \left( \sum_{1 \leq j \leq q} b_j c_j \right) f^{(1)}(t, y) \quad (5.5.42)$$

On constate que la méthode de Runge–Kutta est au moins d'ordre  $\geq 2$  si et seulement si

$$\left( \sum_{1 \leq j \leq q} b_j c_j \right) = \frac{1}{2} \quad (5.5.43)$$

On peut remarquer que le choix qui a été fait dans l'exemple 2 du § 5.5.2 était le seul possible pour obtenir une méthode d'ordre 2.

**Ordre  $\geq 3$  et Ordre  $\geq 4$**  Nous n'allons pas faire les calculs explicites des conditions d'ordre de méthodes de Runge–kutta. Nous allons simplement donné un théorème qui rassemble ces conditions jusqu'à l'ordre 4 :

**Théorème 5.5.2** La méthode de Runge–Kutta définie par le tableau 5.1 des coefficients  $c_j, b_j, a_{ij}$  est

- d'ordre  $\geq 2$  si et seulement si

$$\left( \sum_{1 \leq j \leq q} b_j c_j \right) = \frac{1}{2} \quad (5.5.44)$$

- d'ordre  $\geq 3$  si et seulement si

$$\left( \sum_{1 \leq j \leq q} b_j c_j \right) = \frac{1}{2}, \quad \left( \sum_{1 \leq j \leq q} b_j c_j^2 \right) = \frac{1}{3}, \quad \left( \sum_{i,j} b_j a_{ij} c_j \right) = \frac{1}{6}, \quad (5.5.45)$$

- d'ordre  $\geq 4$  si et seulement si

$$\left( \sum_{1 \leq j \leq q} b_j c_j \right) = \frac{1}{2}, \quad \left( \sum_{1 \leq j \leq q} b_j c_j^2 \right) = \frac{1}{3}, \quad \left( \sum_{1 \leq j \leq q} b_j c_j^3 \right) = \frac{1}{4}, \quad (5.5.46)$$

$$\left( \sum_{i,j} b_j a_{ij} c_j \right) = \frac{1}{6}, \quad \left( \sum_{i,j} b_j a_{ij} c_j^2 \right) = \frac{1}{12}, \quad \left( \sum_{i,j} b_i c_i a_{ij} c_j \right) = \frac{1}{8}, \quad (5.5.47)$$

$$\left( \sum_{i,j,k} b_i a_{ij} a_{jk} c_k \right) = \frac{1}{12}, \quad (5.5.48)$$

□

## 5.6 Contrôle du pas de temps

### 5.6.1 Principe générale du contrôle

### 5.6.2 Applications aux méthodes présentées.



## Chapitre 6

# EDO – Méthodes numériques à pas multiples

### Contenu du Chapitre 6

---

|            |  |            |
|------------|--|------------|
| <b>6.1</b> | <b>Introduction</b> . . . . .  | <b>92</b>  |
| <b>6.2</b> | <b>Exemples classiques</b> . . . . .   | <b>92</b>  |
| 6.2.1      | La méthode de Nyström . . . . .  | 92         |
| 6.2.2      | La méthode de Milne . . . . .  | 93         |
| <b>6.3</b> | <b>Étude générale d’une classe de méthodes à pas multiples</b> . . . . .                         | <b>93</b>  |
| 6.3.1      | Erreur de consistance et ordre . . . . .   | 93         |
| 6.3.2      | Stabilité . . . . .  | 93         |
| <b>6.4</b> | <b>Les méthodes d’Adams-Bashforth à <math>r + 1</math> pas (<math>AB_{r+1}</math>)</b> . . . . . | <b>94</b>  |
| 6.4.1      | Principe . . . . .   | 94         |
| 6.4.2      | Erreur de consistance et ordre des méthodes $AB_{r+1}$ . . . . .                                 | 96         |
| 6.4.3      | Stabilité des méthodes $AB_{r+1}$ . . . . .  | 97         |
| <b>6.5</b> | <b>Les méthodes d’Adams-Moulton à <math>r + 1</math> pas (<math>AM_{r+1}</math>)</b> . . . . .   | <b>97</b>  |
| 6.5.1      | Principe . . . . .   | 97         |
| 6.5.2      | Exemple . . . . .  | 98         |
| 6.5.3      | Consistance et stabilité . . . . .   | 99         |
| <b>6.6</b> | <b>Les méthodes prédictor-correcteur <math>PECE</math></b> . . . . .                             | <b>100</b> |
| 6.6.1      | Principe . . . . .   | 100        |
| 6.6.2      | Ordre et stabilité des schémas $PECE$ . . . . .  | 101        |
| 6.6.3      | Exemples . . . . .   | 101        |

---

## 6.1 Introduction

On s'intéresse toujours à la résolution du problème de Cauchy par l'intermédiaire d'un schéma discret qui fournit des valeurs approchées  $y_n$  pour une subdivision de  $I_0$ ,  $t_n, 0 \leq n \leq N$ .

Nous avons vu que dans les méthodes à un pas, la valeur de  $y_n$  était uniquement déduite de la valeur approchée  $y_n$  à l'instant précédent. L'idée des méthodes à pas multiples (dites aussi multipas) est d'utiliser des valeurs approchées sur un certain nombre d'instant précédents. Plus précisément, donnons la définition d'une méthode à pas multiples :

**Définition 6.1.1** On appelle une méthode numérique à  $r + 1$  pas toute méthode numérique pouvant se mettre sous la forme :

$$y_{n+1} = \Psi(t_n, y_n, h_n, t_{n-1}, y_{n-1}, h_{n-1}, \dots, t_{n-1}, t_{n-r}, y_{n-r}, h_{n-r}) \quad (6.1.1)$$

□

L'intérêt de ces méthodes vient du fait que l'on peut obtenir un ordre élevé de précision avec un coût de calcul plus faible que les méthodes complexes à un pas de type Runge–Kutta. L'idée est de substituer le calcul des pentes sur des valeurs intermédiaires à un pas par les résultats des approximations au pas de temps précédent. L'un des problèmes essentiels dans ces méthodes est qu'il faut cependant, s'assurer que la stabilité numérique reste suffisamment bonne.

Avant de présenter quelques types de méthodes à pas multiples, il est important de noter que le démarrage d'une méthode à pas multiples n'est pas aussi simple que les méthodes à un pas. Le premier point  $(y_0, t_0)$  étant en général donné l'algorithme ne peut démarrer que si les valeurs  $y_i, 1 \leq i \leq r$  ont déjà été calculé. Ceci se fait généralement à l'aide de méthode à un pas comme les méthodes de Runge–kutta d'ordre égal à celui de la méthode à pas multiples utilisées.

## 6.2 Exemples classiques

On donne dans cette partie deux exemples classiques de méthodes à pas multiples.

### 6.2.1 La méthode de Nyström

La méthode de Nyström à pas constant se définit par l'algorithme suivant :

$$\begin{cases} y_{n+1} = y_{n-1} + 2hf_n, n \leq 1 \\ t_{n+1} = t_n + h_n \\ f_{n+1} = f(y_{n+1}, t_{n+1}) \end{cases} \quad (6.2.1)$$

Il s'agit d'une méthode a deux pas  $t_n, t_{n-1}$  dont il est possible de prouver qu'elle d'ordre 2.

### 6.2.2 La méthode de Milne

La méthode de Milne à pas constant se définit par l'algorithme suivant :

$$\begin{cases} y_{n+1} = y_{n-3} + h \left[ \frac{8}{3}f_n - \frac{4}{3}f_{n-1} + \frac{8}{3}f_{n-2} \right], n \leq 3 \\ t_{n+1} = t_n + h_n \\ f_{n+1} = f(y_{n+1}, t_{n+1}) \end{cases} \quad (6.2.2)$$

Il s'agit d'une méthode à quatre pas dont il est possible de prouver qu'elle d'ordre 4.

## 6.3 Étude générale d'une classe de méthodes à pas multiples

On donne dans cette partie les définitions de consistance, d'ordre et de stabilité des méthodes à pas multiples. Il s'agit principalement d'étendre les définitions vues au chapitre précédent.

### 6.3.1 Erreur de consistance et ordre

**Définition 6.3.1** Soit  $y$  une solution exacte du problème de Cauchy, on appelle erreur de consistance de la méthode à  $r + 1$  pas, notée  $\epsilon_n$  la valeur suivante :

$$\epsilon_n = y(t_{n+1}) - \Psi(t_n, y(t_n), h_n, t_{n-1}, y(t_{n-1}), h_{n-1}, \dots, t_{n-1}, t_{n-r}, y(t_{n-r}), h_{n-r}) \quad (6.3.1)$$

□

**Définition 6.3.2** On dit que la méthode à pas multiples est consistante si

$$\lim_{h \rightarrow 0} \sum_{n=0}^{N-1} |\epsilon_n| = 0 \quad (6.3.2)$$

□

**Définition 6.3.3** On dit que la méthode à pas multiples est d'ordre  $p$ , si il existe un réel  $K$  tel que

$$\sum_{n=0}^{N-1} |\epsilon_n| \leq Kh^p \quad (6.3.3)$$

□

### 6.3.2 Stabilité

**Définition 6.3.4** On dit qu'une méthode à  $r + 1$  pas est stable si toutes suites  $y_n$  et  $\hat{y}_n$  vérifiant :

$$y_{n+1} = y_n + h_n \Phi(t_{n-i}, y_{n-i}, h_{n-i}; 0 \leq i \leq r), \quad r \leq n \leq N \quad (6.3.4)$$

$$\hat{y}_{n+1} = \hat{y}_n + h_n \Phi(t_{n-i}, \hat{y}_{n-i}, h_{n-i}; 0 \leq i \leq r) + \rho_n, \quad r \leq n \leq N \quad (6.3.5)$$

il existe il réel  $M$  tel que :

$$\max_{0 \leq n \leq N} |y_n - \hat{y}_n| \leq M \left( \max_{0 \leq n \leq r} |y_n - \hat{y}_n| + \sum_{r \leq n \leq N} |\rho_n| \right) \quad (6.3.6)$$

□

## 6.4 Les méthodes d'Adams-Bashforth à $r + 1$ pas ( $AB_{r+1}$ )

### 6.4.1 Principe

Si  $y$  est un solution du problème de Cauchy, on écrit :

$$y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} f(t, y(t)) dt \quad (6.4.1)$$

Supposons que pour  $0 \leq i \leq r$ , on est déjà calculé les valeurs approchées de  $y$  aux instants  $t_{n-i}$ ,  $y_{n-i}$  et les valeurs approchées des pentes  $f_{n-i} = f(t_{n-i}, y_{n-i})$ . L'idée de la méthode d'Adams-Bashforth est d'approximer la  $f$  entre  $t_n$  et  $t_{n+1}$  par son polynôme d'interpolation aux points  $(t_{n-i}, y_{n-i})$ .

Notons ce polynôme  $p_{n,r}(t)$ . C'est un polynôme de degré,  $r$  qui s'exprime d'après la définition de l'interpolation de Lagrange :

$$p_{n,r}(t) = \sum_{i=0}^r f_{n-i} L_{n,i,r}(t) \quad (6.4.2)$$

où  $L_{n,i,r}$  sont les polynômes de base de Lagrange donnés par :

$$L_{n,i,r}(t) = \prod_{j=0, j \neq i}^r \frac{t - t_{n-j}}{t_{n-i} - t_{n-j}} \quad (6.4.3)$$

L'équation (6.4.1) peut donc être approchée par :

$$y(t_{n+1}) \approx y(t_n) + \int_{t_n}^{t_{n+1}} \sum_{i=0}^r f_{n-i} L_{n,i,r}(t) dt \quad (6.4.4)$$

$$\approx y(t_n) + \sum_{i=0}^r \int_{t_n}^{t_{n+1}} L_{n,i,r}(t) dt f_{n-i} \quad (6.4.5)$$

$$\approx y(t_n) + h_n \sum_{i=0}^r b_{n,i,r} f_{n-i} \quad (6.4.6)$$

avec

$$b_{n,i,r} = \frac{1}{h_n} \int_{t_n}^{t_{n+1}} L_{n,i,r}(t) dt \quad (6.4.7)$$

L'algorithme d'Adams-Bashforth à  $r + 1$  pas, noté en abrégé  $AB_{r+1}$  s'écrit donc :

$$\begin{cases} b_{n,i,r} = \frac{1}{h_n} \int_{t_n}^{t_{n+1}} L_{n,i,r}(t) dt \\ y_{n+1} = y_n + h_n \sum_{i=0}^r b_{n,i,r} f_{n-i} \\ t_{n+1} = t_n + h_n \\ f_{n+1} = f(t_{n+1}, y_{n+1}) \end{cases} \quad (6.4.8)$$

### Remarques

1. Les coefficients  $b_{n,i,r}$  ne dépendent que de la subdivision. Ils peuvent donc être calculés à l'avance si la subdivision est connue à l'avance par exemple si le pas est constant. Dans ce dernier cas, il ne dépendent plus de  $n$  et on les note  $b_{i,r}$ . On donne dans le tableau 6.1 les valeurs de ces coefficients pour des valeurs faibles  $r$ .
2. Un des intérêts de la méthode, outre sa relative simplicité est le gain en temps de calcul. A la différence des méthodes à un pas de type Runge-Kutta, on ne fait qu'une évaluation de  $f$  par incrément de temps. On se sert des pentes déjà calculées au pas précédent plutôt que de calculer des pentes intermédiaires.

| $r$ | $b_{0,r}$ | $b_{1,r}$ | $b_{2,r}$ | $b_{3,r}$ | $b_{4,r}$ | $\beta_r = \sum_{0 \leq i \leq r}  b_{i,r} $ |
|-----|-----------|-----------|-----------|-----------|-----------|--|
| 0   | 1         |           |           |           |           | 1  |
| 1   | 3/2       | -1/2      |           |           |           | 2  |
| 2   | 23/12     | -16/21    | 5/12      |           |           | 3,66..                                       |
| 3   | 55/24     | -59/24    | 37/24     | -9/24     |           | 6,66..                                       |
| 4   | 1901/120  | -1387/360 | 109/30    | -637/360  | 251/720   | 12,24..                                      |

TAB. 6.1 – Coefficient  $b_{i,r}$  des méthodes de Adams-Bashforth pour de faibles valeurs de  $r$  et un pas constant

### Exemples

- $r = 0$ , le polynôme d'interpolation  $p_{n,0}$  est donc un polynôme de degré 0 qui interpole  $f$  en  $t_n$  soit :

$$p_{n,0}(t) = f(t_n, y_n) = f_n \quad (6.4.9)$$

L'algorithme est donc le suivant :

$$y_{n+1} = y_n + h_n f_n \quad (6.4.10)$$

et l'on reconnaît bien sur l'algorithme de Euler progressive.

- $r = 1$ , le polynôme d'interpolation  $p_{n,1}$  est donc un polynôme de degré 1 qui interpole  $f$  en  $(t_n, f_n)$  et  $(t_{n-1}, f_{n-1})$  soit :

$$p_{n,1}(t) = \frac{t - t_{n-1}}{t_n - t_{n-1}} f_n - \frac{t - t_n}{t_n - t_{n-1}} f_{n-1} \quad (6.4.11)$$

$$= f_n + \frac{f_n - f_{n-1}}{t_n - t_{n-1}} (t - t_n) \quad (6.4.12)$$

Le calcul des coefficients se fait en intégrant ce polynôme sur  $[t_n, t_{n+1}]$  :

$$\int_{t_n}^{t_{n+1}} p_{n,1}(t) dt = \int_{t_n}^{t_{n+1}} f_n + \frac{f_n - f_{n-1}}{t_n - t_{n-1}} (t - t_n) dt \quad (6.4.13)$$

$$= h_n f_n + \frac{f_n - f_{n-1}}{h_{n-1}} \left[ \frac{1}{2} (t - t_n)^2 \right]_{t_n}^{t_{n+1}} \quad (6.4.14)$$

$$= h_n f_n + \frac{f_n - f_{n-1}}{h_{n-1}} \frac{1}{2} h_n^2 \quad (6.4.15)$$

$$= h_n \left[ \left( 1 + \frac{h_n}{2h_{n-1}} \right) f_n - \frac{h_n}{2h_{n-1}} f_{n-1} \right] \quad (6.4.16)$$

on a donc

$$b_n = 1 + \frac{h_n}{2h_{n-1}}, \quad b_{n-1} = \frac{h_n}{2h_{n-1}} \quad (6.4.17)$$

L'algorithme s'écrit donc :

$$AB_2 \begin{cases} y_{n+1} = y_n + h_n \left[ \left( 1 + \frac{h_n}{2h_{n-1}} \right) f_n - \frac{h_n}{2h_{n-1}} f_{n-1} \right] \\ t_{n+1} = t_n + h_n \\ f_{n+1} = f(t_{n+1}, y_{n+1}) \end{cases} \quad (6.4.18)$$

Dans le cas où le pas est constant égal à  $h$ , on retrouve bien :

$$y_{n+1} = y_n + h_n \left[ \frac{3}{2} f_n - \frac{1}{2} f_{n-1} \right] \quad (6.4.19)$$

$$(6.4.20)$$

## 6.4.2 Erreur de consistance et ordre des méthodes $AB_{r+1}$

Le théorème suivant caractérise l'ordre et donc la consistance des méthodes d'Adams-Bashforth.

**Théorème 6.4.1** Les méthodes d'Adams-Bashforth  $AB_{r+1}$  sont d'ordre  $r + 1$ .

□

*Preuve* : Voir (CROUZEIX & MIGNOT, 1992)

□

### 6.4.3 Stabilité des méthodes $AB_{r+1}$

La stabilité est donnée par le théorème suivant :

**Théorème 6.4.2** On suppose que  $f$  est  $k$ -lipschitzienne en  $y$  et que les sommes  $\sum_{0 \leq i \leq r} |b_{i,n,r}|$  sont majorées indépendamment de  $n$  par une constante  $\beta_r$  :

$$\beta_r = \max_n \sum_{0 \leq i \leq r} |b_{i,n,r}| \quad (6.4.21)$$

alors la méthode d'Adams-Bashforth  $AB_{r+1}$  est stable et la constante de stabilité est donnée par

$$M = \exp(\beta_r k T) \quad (6.4.22)$$

□

*Preuve :* Voir (CROUZEIX & MIGNOT, 1992)

□

**Remarque** Pour que la méthode amplifie les moins possibles les erreurs, on souhaite que la constante de stabilité soit la plus faible possible. Cette constante dépend de  $k$  et de  $T$  qui sont des données du problème et de la constante majorant les coefficients,  $\beta_r$ . Cette constante est donnée pour de faibles valeurs de  $r$  dans le tableau 6.1. On constate qu'elle croit très vite avec  $r$  et ceci est résultat général. Il s'ensuit que les méthodes sont de moins en moins stable lorsque l'on augmente l'ordre. Ce défaut de stabilité constituent le principal défaut des méthodes d'Adams-Bashforth  $AB_{r+1}$  qui sont rarement utilisées pour de grandes valeurs de  $r$ .

Nous allons voir dans la suite des méthodes pour pallier ce défaut.

## 6.5 Les méthodes d'Adams-Moulton à $r + 1$ pas ( $AM_{r+1}$ )

### 6.5.1 Principe

Le principe de construction des méthodes d'Adams-Moulton est identique à celui des méthodes d'Adams-Bashforth. La différence majeure est que l'on réalise une interpolation en incluant le point  $(t_{n+1}, f_{n+1})$ .

On considère le polynôme  $p_{n,r}^*$  de degré  $r + 1$  qui interpole les points  $(t_{n-i}, f_{n-i})$ ,  $-1 \leq i \leq r$  :

$$p_{n,r}^*(t) = \sum_{i=-1}^r f_{n-i} L_{n,i,r}(t) \quad (6.5.1)$$

où  $L_{n,i,r}$  sont les polynômes de base de Lagrange donnés par :

$$L_{n,i,r}(t) = \prod_{j=-1, j \neq i}^r \frac{t - t_{n-j}}{t_{n-i} - t_{n-j}} \quad (6.5.2)$$

L'équation (6.4.1) peut donc être approchée par :

$$y(t_{n+1}) \approx y(t_n) + \int_{t_n}^{t_{n+1}} \sum_{i=-1}^r f_{n-i} L_{n,i,r}(t) dt \quad (6.5.3)$$

$$\approx y(t_n) + \sum_{i=-1}^r \int_{t_n}^{t_{n+1}} L_{n,i,r}(t) dt f_{n-i} \quad (6.5.4)$$

$$\approx y(t_n) + h_n \sum_{i=-1}^r b_{n,i,r}^* f_{n-i} \quad (6.5.5)$$

avec

$$b_{n,i,r}^* = \frac{1}{h_n} \int_{t_n}^{t_{n+1}} L_{n,i,r}(t) dt \quad (6.5.6)$$

L'algorithme d'Adams-Moulton à  $r + 1$  pas, noté en abrégé  $AM_{r+1}$  s'écrit donc :

$$\begin{cases} b_{n,i,r}^* = \frac{1}{h_n} \int_{t_n}^{t_{n+1}} L_{n,i,r}(t) dt \\ y_{n+1} - h_n b_{n,-1,r}^* f(t_{n+1}, y_{n+1}) = y_n + h_n \sum_{i=0}^r b_{n,i,r}^* f_{n-i} \\ t_{n+1} = t_n + h_n \\ f_{n+1} = f(t_{n+1}, y_{n+1}) \end{cases} \quad (6.5.7)$$

### Remarques

1. Le calcul des  $b_{n,i,r}^*$  peut la aussi être calculés à l'avance si le pas de temps est constant et égal à  $h$ . On donne les valeurs de ces coefficients pour de faibles valeurs de  $r$  dans le tableau 6.2
2. Contrairement aux méthodes d'Adams-Bashforth, les méthodes  $AM_{r+1}$  sont des méthodes implicites. Il est donc nécessaire de résoudre une équation, éventuellement non linéaire à chaque incrément. En règle général, on utilise une méthode itérative métrique basée sur le théorème du point fixe. Pour avoir une application contractante, dans l'hypothèse où  $f$  est  $k$ -lipschitzienne, il est nécessaire que le pas de discrétisation maximum,  $h_{max}$  vérifie :

$$h_{max} \leq \frac{1}{k\gamma_r^*}, \quad \gamma_r^* = \max_n |b_{n,i,r}^*| \quad (6.5.8)$$

### 6.5.2 Exemple

Considérons l'exemple le plus simple des méthodes d'Adams-Moulton pour  $r = 0$ . Le polynôme d'interpolation  $p_{n,0}^*$  interpole les points  $t_n, f_n$  et  $t_{n+1}, f_{n+1}$ . Il s'écrit donc :

$$p_{n,0}^*(t) = f_n + \frac{f_{n+1} - f_n}{t_{n+1} - t_n} (t - t_n) \quad (6.5.9)$$

| $r$ | $b_{-1,r}^*$ | $b_{0,r}^*$ | $b_{1,r}^*$ | $b_{2,r}^*$ | $b_{3,r}^*$ | $b_{4,r}^*$ | $\beta_r^* = \sum_{-1 \leq i \leq r}  b_{i,r}^* $ | $\beta_{r+1}$ |
|-----|--------------|-------------|-------------|-------------|-------------|-------------|---|---------------|
| 0   | 1/2          | 1/2         |             |             |             |             | 1   | 2             |
| 1   | 5/12         | 2/3         | -1/12       |             |             |             | 1, 16..   | 3, 66..       |
| 2   | 9/24         | 19/24       | -5/24       | 1/24        |             |             | 1, 41..   | 6, 66..       |
| 3   | 251/720      | 323/360     | -11/30      | 53/360      | -19/720     |             | 1, 73..   | 12, 24..      |
| 3   | 95/288       | 1447/1440   | -133/240    | 241/720     | -173/1440   | 3/160       | 2, 34..   | 22, 14..      |

TAB. 6.2 – Coefficient  $b_{i,r}^*$  des méthodes de Adams-Moulton pour de faibles valeurs de  $r$  et un pas constant

Son intégration conduit donc à

$$\int_{t_n}^{t_{n+1}} p_{n,0}^*(t) dt = f_n h_n + \frac{f_{n+1} - f_n}{h_n} \frac{1}{2} h_n^2 \quad (6.5.10)$$

$$= \frac{1}{2} h_n [f_{n+1} + f_n] \quad (6.5.11)$$

On a donc

$$b_{n,-1,0} = \frac{1}{2}, \quad b_{n,0,0} = \frac{1}{2} \quad (6.5.12)$$

La méthode s'écrit donc :

$$y_{n+1} = y_n + \frac{1}{2} h_n [f_{n+1} + f_n] \quad (6.5.13)$$

On reconnaît la méthode de Crank-Nicholson.

### 6.5.3 Consistance et stabilité

**Théorème 6.5.1** Les méthodes d'Adams-Moulton  $AM_{r+1}$  sont d'ordre  $r + 2$ . □

*Preuve* : Voir (CROUZEIX & MIGNOT, 1992) □

**Théorème 6.5.2** Sous réserve que la condition de convergence (6.5.8) de la méthode itérative soit remplie, la méthode d'Adams-Moulton à  $r + 1$  pas  $AM_{r+1}$  est stable. Sa constante de stabilité est donnée par :

$$M = \exp\left(\frac{\beta_r^* k T}{1 - \gamma_r^* k h_{max}}\right) \quad (6.5.14)$$

où

$$\beta_r^* = \max_n \sum_{-1 \leq i \leq r} |b_{i,n,r}^*| \quad (6.5.15)$$

□

### Remarques

1. Pour  $h_{max} \ll \frac{1}{k\gamma_r^*}$ , on retrouve que  $M \approx \exp(\beta_r^* kT)$ . La comparaison a un ordre identique entre les méthodes d'Adams-Bashforth et les méthodes d'Adams-Moulton montrent que ces dernières sont beaucoup plus stables.

## 6.6 Les méthodes prédictor-correcteur PECE

L'idée des schémas *PECE* est de construire des schémas aussi stables que les schémas d'Adams-Moulton mais en évitant les problèmes liés à la résolution d'une équation non-linéaire à) chaque incrément.

### 6.6.1 Principe

Le principe de construction des méthodes de prédiction-correction est le suivant. On se donne une prédiction de la valeur  $y_{n+1}$ , notée  $py_{n+1}$  donnée par un schéma explicite. Cette valeur est ensuite utilisée dans un schéma implicite, comme le schéma d'Adams-Moulton pour réaliser une correction. Plus précisément, on obtient l'algorithme suivant :

$$\left\{ \begin{array}{ll} (P) & \text{prédiction} \quad py_{n+1} = \dots \text{ (à partir des } y_{n-i}, f_{n-i}, 0 \leq i \leq r \text{ schéma explicite )} \\ (E) & \text{évaluation} \quad pf_{n+1} = f(t_{n+1}, py_{n+1}) \\ (C) & \text{correction} \quad y_{n+1} = y_n + h_n \sum_{i=0}^r b_{n,i,r}^* f_{n-i} + h_n b_{n,-1,r}^* pf_{n+1}, \quad AM_{r+1} \\ (E) & \text{évaluation} \quad f_{n+1} = f(t_{n+1}, y_{n+1}) \end{array} \right. \quad (6.6.1)$$

On constate bien qu'à travers cette algorithmme que le schéma obtenue est explicite.

### Remarques

1. Nous avons utiliser un schéma d'Adams-Moulton pour réaliser l'étape de correction. On peut cependant utiliser n'importe quel schéma implicite.
2. le démarrage de ce type de méthode nécessite comme pour toutes les méthodes à pas multiples le calcul préalable des points  $y_i, f_i, 0 \leq i \leq r$ .

3. Le coût de calcul des méthodes de prédiction-correction basée sur un schéma d'Adams-Moulton est environ le double de celui d'une méthode d'Adams-Bashforth de même ordre. Nous verrons par contre que sa stabilité est bien meilleur. Ce coût de calcul reste cependant moindre que celui des méthodes Runge-Kutta sophistiquées d'ordre  $\geq 3$

### 6.6.2 Ordre et stabilité des schémas PECE

L'étude de l'erreur de consistance pour les schémas PECE conduit aux résultats suivants :

1. L'influence du choix prédicteur est nettement moins important que celui du correcteur.
2. Si l'on choisit un schéma prédicteur d'ordre  $r+1$  et un correcteur d'ordre  $r+1$ , par exemple  $AM_{r+1}$ , le schéma PECE est alors d'ordre  $r+2$ .

Une étude de la stabilité montre que l'on conserve les propriétés de stabilité des schémas d'Adams-Moulton pour une constante de stabilité équivalente.

### 6.6.3 Exemples

**Prédicteur d'ordre 1 (Euler) et correcteur d'ordre 2 (Crank-Nicholson)** L'algorithme est donné de la manière suivante :

$$\left\{ \begin{array}{l} (P) \quad py_{n+1} = y_n + h_n f_n \\ (E) \quad pf_{n+1} = f(t_{n+1}, py_{n+1}) \\ (C) \quad y_{n+1} = y_n + \frac{1}{2} h_n [f_n + pf_{n+1}] \\ (E) \quad f_{n+1} = f(t_{n+1}, y_{n+1}) \end{array} \right. \quad (6.6.2)$$

Cet algorithme coïncide avec la méthode de Heun, qui n'est autre qu'une méthode de Runge-Kutta définie par :

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

**Prédicteur d'ordre 2 (Nyström à pas constant) et correcteur d'ordre 3 ( $AM_{r+3}$ )**

$$\left\{ \begin{array}{l} (P) \quad py_{n+1} = y_{n-1} + 2hf_n \\ (E) \quad pf_{n+1} = f(t_{n+1}, py_{n+1}) \\ (C) \quad y_{n+1} = y_n + h \left[ \frac{5}{12}pf_{n+1} + \frac{8}{12}f_n - \frac{1}{12}f_{n-1} \right] \\ (E) \quad f_{n+1} = f(t_{n+1}, y_{n+1}) \end{array} \right. \quad (6.6.3)$$

# Bibliographie

- E. BUSVELLE. *Notes de cours d'analyse numérique élémentaire : DEs équations algébriques aux équations différentielles*. Université de Bourgogne, Départements de Mathématiques, 2002.
- M. CROUZEIX & A.L. MIGNOT. *Analyse numérique des équations différentielles*. Collection mathématiques appliquées pour la maîtrise. Masson, Paris., 2ème édition, 1992. édition révisée et augmentée.
- R. DAUTRAY & J.L. LIONS. *Analyse mathématique et calcul numérique pour les sciences et la technique*. Collection du Commissariat à l'Energie Atomique, série scientifique. Masson, 1988.
- J.P. DEMAILLY. *Analyse numérique et équations différentielles*. Collection Grenoble sciences — Presses universitaires de Grenoble, 1996. 309 p.
- J.-M. FERRARD. *Maths et Maple*. Dunod, Paris, 1998.
- A. FORTIN. *Analyse numérique pour ingénieurs*. Presses internationales polytechniques, 2ème édition, 2001.
- E. HAIRER, S.P. NORSETT & G. WANNER. *Solving Ordinary Differential Equations I. Nonstiff Problems*. Springer, 1993.
- E. HAIRER, S.P. NORSETT & G. WANNER. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*. Springer, 1996.
- A. QUARTERONI. *Méthodes numériques pour le calcul scientifique*. Springer-Verlag France, 2000.
- M. SCHATZMAN. *Analyse numérique — Une approche mathématique*. Dunod, Paris, 2ème édition édition, 2001. édition revue et augmentée.
- L. SCHUMAKER. *Splines Functions: Basic Theory*. Wiley, New York, 1981.
- E. WEINSTEIN. <http://mathworld.wolfram.com/>, 2002.



# Index

- constante de Lebesgue, 10
- degré d'approximation, 11
- différence divisée, 8
- erreur d'une méthode d'intégration numérique, 32
- espace de Hilbert, 21
- fonctions splines, 16
- forme de Newton du polynôme d'interpolation, 7
- formule de quadrature
  - élémentaire, 30
  - composée, 30
- formules d'intégration
  - de Gauss, 41
  - de Gauss-Chebychev, 41
  - de Gauss-Hermite, 41
  - de Gauss-Laguerre, 41
  - de Gauss-Legendre, 41
  - de Newton-Cotes, 31
- Interpolation
  - de Hermite, 14
  - de Lagrange, 5
  - par fonctions splines, 16
- interpolation polynomiale, 4
- méthode
  - des différences divisées, 7
- méthode d'intégration
  - de Boole-Villarceau, 32
  - de Simpson, 32
  - des rectangles à droite, 29
  - des rectangles à gauche, 29
  - des trapèzes, 29
  - du point milieu, 29
- méthodes d'intégration
  - composées, 28
  - de Gauss, 28
- meilleure approximation polynomiale, 4
- norme, 21
- ordre d'une méthode d'intégration, 32
- points de Chebyshev, 13
- polynôme, 4
  - d'interpolation de Hermite, 14
  - d'interpolation de Lagrange, 5
  - de base de Hermite, 14
  - de base de Lagrange, 5
  - de base de Newton, 7
  - de meilleure approximation, 19
- polynômes
  - de Chebyshev
    - de première espèce, 24
    - de seconde espèce, 24
  - de Hermite, 24
  - de Jacobi, 23
  - de Laguerre, 24
  - de Legendre, 23
- polynômes orthogonaux, 22
- produit scalaire, 21
- schéma de Hörner, 1

## Annexe A

# Évaluation d'un polynôme grâce au schéma de Hörner

Considérons un polynôme  $p_n$ , de degré inférieur ou égal à  $n$ , à coefficients réels, i.e.,  $p_n \in \mathcal{P}_n$ . L'objet du *schéma de Hörner* est d'évaluer ce polynôme en un point de la droite réelle  $z$ . Une méthode naïve consiste à effectuer l'algorithme 3

---

**Algorithme 3** Calculer  $p_n(z) = \sum_{i=0}^n a_i z^i$  — Méthode naïve

---

**Require:**  $n, z, a_i, 0 \leq i \leq n$

**Ensure:**  $p = \sum_{i=0}^n a_i z^i$

$p \leftarrow a_0$

$q \leftarrow 1$

**for**  $i = 1$  to  $n$  **do**

$q \leftarrow q * z$

$p \leftarrow p + a_i q$

**end for**

---

On constate que l'on effectue à chaque itération, deux multiplications et une addition. Nous allons donner dans la suite une méthode plus efficace pour évaluer le polynôme.

### A.1 Schéma de Hörner sur la forme canonique

Considérons la forme canonique du polynôme  $p_n$ , évalué au point  $z$ :

$$p_n(z) = a_0 + a_1 z + a_2 z^2 + \dots + a_n z^n \tag{A.1.1}$$

On remarque que ce polynôme peut se factoriser sous la forme suivante :

$$p_n(z) = a_0 + z [a_1 + a_2z + \dots + a_n z^{n-1}] \quad (\text{A.1.2})$$

$$= a_0 + z [a_1 + z [a_2 + \dots + a_n z^{n-2}]] \quad (\text{A.1.3})$$

...

$$= a_0 + z [a_1 + z [a_2 + z [\dots + z a_n] \dots]] \quad (\text{A.1.4})$$

A partir de cette remarque, on peut définir l'algorithme de Hörner qui s'écrit de la manière suivante :

---

**Algorithme 4** Calculer  $p_n(z) = \sum_{i=0}^n a_i z^i$  — Schéma de Hörner

---

**Require:**  $n, z, a_i, 0 \leq i \leq n$

**Ensure:**  $p(z) = \sum_{i=0}^n a_i z^i$

$p \leftarrow a_n$

**for**  $i = n - 1$  **to**  $0$  **do**

$p \leftarrow zp + a_i$

**end for**

---

On remarque dès lors qu'on effectue plus qu'une addition et une multiplication à chaque étape. Hormis le gain substantiel en temps de calcul, le schéma de Hörner a de bonnes propriétés concernant les erreurs d'arrondis.

## A.2 Schéma de Hörner sur la forme de Newton

De la même manière, le schéma de Hörner peut être étendu à la forme de Newton du polynôme d'interpolation. Considérons le polynôme  $p_n \in \mathcal{P}_n$  qui interpole  $f$  au point  $\{x_i\}_{0 \leq i \leq n}$ . La forme de Newton du polynôme peut être écrite :

$$\begin{aligned} p_n(z) &= \alpha_0 + \alpha_1(z - x_0) + \alpha_2(z - x_0)(z - x_1) + \dots + \alpha_n(z - x_0)(z - x_1) \dots (z - x_{n-1}) \\ &= \alpha_0 + (z - x_0) [\alpha_1 + \alpha_2(z - x_1) + \dots + \alpha_n(z - x_1) \dots (z - x_{n-1})] \\ &= \alpha_0 + (z - x_0) [\alpha_1(z - x_1) [\alpha_2 + \dots + \alpha_n(z - x_2) \dots (z - x_{n-1})]] \end{aligned} \quad (\text{A.2.1})$$

A partir de cette écriture, on peut écrire l'algorithme de Hörner pour la forme de Newton du polynôme d'interpolation :

Il est bon de remarquer que les  $\alpha_i$  sont le résultat de la méthode des différences divisées. Ces deux algorithmes peuvent être avantageusement couplés.

---

**Algorithme 5** Calculer  $p_n(z) = \sum_{i=0}^n a_i z^i$  — Schéma de Hörner, Forme de Newton

---

**Require:**  $n, z, \alpha_i, 0 \leq i \leq n$

**Ensure:**  $p(z) = \sum_{i=0}^n a_i z^i$

$p \leftarrow \alpha_n$

**for**  $i = n - 1$  to  $0$  **do**

$p \leftarrow (z - x_i)p + \alpha_i$

**end for**

---



## Annexe B

# Les polynômes de Chebyshev

### B.1 Forme explicite des polynômes de Chebyshev

Posons

$$\forall \theta \in \mathbb{R}, T_n(\cos \theta) = \cos(n\theta) \quad (\text{B.1.1})$$

ou de manière équivalente pour  $x = \cos \theta$ ,

$$\forall x \in [-1; 1], T_n(x) = \cos(n \arccos x) \quad (\text{B.1.2})$$

On constate facilement que :

$$T_0(x) = 1, \quad T_1(x) = x \quad (\text{B.1.3})$$

De plus, on a les relations suivantes :

$$T_{n+1}(x) = \cos((n+1) \arccos x) \quad (\text{B.1.4})$$

$$= \cos(\arccos x) \cos(n \arccos x) - \sin(\arccos x) \sin(n \arccos x) \quad (\text{B.1.5})$$

$$= xT_n(x) - \sin(\arccos x) \sin(n \arccos x) \quad (\text{B.1.6})$$

$$T_{n-1}(x) = \cos((n-1) \arccos x) \quad (\text{B.1.7})$$

$$= \cos(-\arccos x) \cos(n \arccos x) - \sin(-\arccos x) \sin(n \arccos x) \quad (\text{B.1.8})$$

$$= xT_n(x) + \sin(\arccos x) \sin(n \arccos x) \quad (\text{B.1.9})$$

qu conduisent à la relation de récurrence :

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \quad (\text{B.1.10})$$

On en déduit que pour  $n \geq 1$ ,  $T_n$  est un polynôme de degré  $n$ . Le coefficient de terme de plus haut degré est  $2^{n-1}$ .

Par ailleurs, en posant  $\theta = \arccos x$ , on peut évaluer l'intégrale suivante :

$$\int_{-1}^1 T_n(x)T_k(x)(1-x^2)^{-1/2} dx = \int_0^\pi \cos(n\theta) \cos(k\theta) d\theta = 0, \text{ si } k \neq n \quad (\text{B.1.11})$$

Les polynômes  $T_n$  sont donc les polynômes orthogonaux pour la fonction poids  $\omega(x) = (1 - x^2)^{1/2}$ . En posant,

$$p_n(x) = 2^{1-n} T_n(x) \quad (\text{B.1.12})$$

les polynômes  $p_n$  sont donc les polynômes de Chebyshev.

La norme des polynômes de Chebyshev est donnée par :

$$\int_{-1}^1 T_n(x) T_k(x) (1 - x^2)^{-1/2} dx = \begin{cases} \pi & \text{si } k = 0 \\ \frac{\pi}{2} & \text{sinon.} \end{cases} \quad (\text{B.1.13})$$

**Remarque** Les polynômes de Chebyshev peuvent être obtenus directement par le procédé d'orthogonalisation de Gram-Schmidt :

$$p_0(x) = 1 \quad (\text{B.1.14})$$

$$p_1(x) = x - \frac{\int_{-1}^1 x(1 - x^2)^{-1/2} dx}{\int_{-1}^1 (1 - x^2)^{-1/2} dx} \quad (\text{B.1.15})$$

$$= x - \frac{[-(1 - x^2)^{-1/2}]_{-1}^1}{[\sin^{-1} x]_{-1}^1} = x \quad (\text{B.1.16})$$

$$p_2(x) = \left[ x - \frac{\int_{-1}^1 x^3(1 - x^2)^{-1/2} dx}{\int_{-1}^1 x^2(1 - x^2)^{-1/2} dx} \right] x - \left[ \frac{\int_{-1}^1 x^2(1 - x^2)^{-1/2} dx}{\int_{-1}^1 (1 - x^2)^{-1/2} dx} \right] \cdot 1 \quad (\text{B.1.17})$$

$$= [x - 0]x - \frac{\pi/2}{\pi} = x^2 - \frac{1}{2} \quad (\text{B.1.18})$$

$$\dots \quad (\text{B.1.19})$$

## B.2 Racines des polynômes de Chebyshev

Les racines sur  $[-1, 1]$  du polynôme de Chebyshev sont données par :

$$p_n(x) = 0 \Leftrightarrow T_n(x) = 0 \quad (\text{B.2.1})$$

$$\cos(n \arccos x) = 0 \quad (\text{B.2.2})$$

$$n \arccos x = \frac{\pi}{2} + k\pi, k \in \mathbb{N} \quad (\text{B.2.3})$$

$$\arccos x = \frac{\pi(1 + 2k)}{2n}, k \in \mathbb{N} \quad (\text{B.2.4})$$

$$x = \cos\left(\frac{\pi(1 + 2k)}{2n}\right), 0 \leq k \leq n - 1, k \in \mathbb{N} \quad (\text{B.2.5})$$

Les racines des polynômes de Chebyshev sont utilisées comme point d'interpolation polynomiale et comme point d'intégration dans les techniques de Gauss-Chebyshev.