

République Algérienne Démocratique et populaire  
Ministère de l'enseignement Supérieur et de la Recherche Scientifique  
Université 08 Mai 1945 Guelma  
Faculté de Mathématiques et Informatique  
Et Science de La Matière  
Département de Mathématiques



**Polycopie de Cours**  
Présentés aux Etudiants  
**Licence Académique en mathématiques**  
**Option : Analyse**  
**Par**  
**Dr. Noura Tabouche**  
**Intitulé**

***Introduction à l'analyse numérique***

**Septembre 2016**

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Résolution numérique de systèmes d'équations</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.1.1 Solution théorique . . . . .	4
1.1.2 Esprit des méthodes numériques . . . . .	5
1.2 Méthodes directes . . . . .	5
1.2.1 Méthode d'élimination de Gauss . . . . .	5
1.2.2 Méthode de factorisation LU . . . . .	11
1.3 Méthodes itératives . . . . .	13
1.3.1 Introduction . . . . .	13
1.3.2 Principe des méthodes itératives . . . . .	14
1.3.3 Le choix de la décomposition . . . . .	14
1.3.4 Méthode de Jacobi . . . . .	15
1.3.5 Méthode de Gauss-Seidel . . . . .	16
1.4 Exercices . . . . .	19
1.4.1 Réponses . . . . .	20
<b>2 Calcul numérique des valeurs propres d'une matrice</b>	<b>24</b>
2.1 Rappels. Définitions . . . . .	25
2.1.1 Rappels élémentaires sur les matrices . . . . .	25
2.1.2 Décomposition de SCHUR réelle . . . . .	27
2.2 Méthode de la puissance . . . . .	28
2.2.1 Principe de la méthode et algorithme . . . . .	28
2.3 Méthode de la puissance inverse . . . . .	30
2.4 Méthode de Jacobi pour matrices réelles symétriques . . . . .	32
2.5 Exercices . . . . .	35
2.5.1 Réponses . . . . .	35
<b>3 Interpolation numérique</b>	<b>39</b>
3.1 Généralités . . . . .	39
3.1.1 Position du problème . . . . .	39
3.2 Interpolation polynômiale . . . . .	40

---

3.2.1	Problème . . . . .	40
3.2.2	Interpolation de LAGRANGE . . . . .	42
3.2.3	Interpolation dans la base de Newton . . . . .	45
3.3	Exercices . . . . .	48
3.3.1	Réponses . . . . .	49
<b>4</b>	<b>Intégration numérique</b>	<b>51</b>
4.1	Méthode du rectangle . . . . .	52
4.2	Méthode du trapèze . . . . .	53
4.3	Méthode de Simpson . . . . .	53
4.4	Formule des Trapèzes composée . . . . .	54
4.5	Formule de Simpson composée . . . . .	57
4.6	Exercices . . . . .	58
4.6.1	Réponses . . . . .	58
<b>5</b>	<b>Résolution numérique des équations différentielles ordinaires</b>	<b>60</b>
5.1	Introduction . . . . .	60
5.2	Préliminaires . . . . .	61
5.2.1	Problèmes de conditions aux limites . . . . .	61
5.2.2	Problèmes de conditions initiales (problème de Cauchy) . . . . .	62
5.2.3	Existence et unicité des solutions . . . . .	63
5.3	Principe d'intégration approchée par des méthodes numériques . . . . .	63
5.4	Méthodes numériques à un pas . . . . .	65
5.4.1	<b>Méthode d'EULER</b> . . . . .	65
5.4.2	<b>Méthode de Taylor</b> . . . . .	67
5.4.3	<b>Méthode de Runge- kutta</b> . . . . .	70
5.4.4	<b>Méthode de Runge - kutta d'ordre 4 :</b> . . . . .	72
5.5	Exercices . . . . .	72
5.5.1	Réponses . . . . .	73
<b>6</b>	<b>Résolution des équations non linéaires</b>	<b>74</b>
6.1	Introduction . . . . .	74
6.2	Généralités sur les méthodes de résolution . . . . .	74
6.3	Séparation des racines . . . . .	76
6.3.1	Méthode graphique . . . . .	77
6.3.2	Méthodes algébriques . . . . .	77
6.4	Approximation des racines . . . . .	77
6.4.1	Méthode de dichotomie ( ou méthode de la bisection) . . . . .	77
6.4.2	Méthode de Newton-Raphson . . . . .	80
6.4.3	Méthode de la sécante . . . . .	82
6.4.4	Méthode du point fixe . . . . .	84
6.5	Exercices . . . . .	88
6.5.1	Réponses . . . . .	89

# Introduction

La modélisation des problèmes "réels " observés conduit souvent à un modèle mathématique mis en équation et dont on cherche à calculer la solution.

Néanmoins, les méthodes analytiques de résolution de problèmes mathématiques s'adaptent à un nombre limité d'équations algébriques ; notamment, à des problèmes linéaires. En effet, en considérant des hypothèses plus ou moins justifiées il s'ensuit des modèles très simplifiés représentant les phénomènes observés qui sont en réalité très complexes.

La plupart du temps les solutions dégagées à partir des équations simplifiées ne représentent le phénomène que dans le domaine où les hypothèses "simplifiées" utilisées ont un sens. En plus, lorsque certaines valeurs des paramètres sont ignorées la solution obtenue n'a pas toujours un grand rapport avec l'observation.

Il est alors naturel de songer à prendre en compte dans les équations tous les termes pour avoir une approche plus complète du phénomène étudié ; cependant, la résolution analytique échoue. Il convient de noter que sans moyens de calculs appropriés, il peut être difficile de résoudre un problème, même si l'on possède un moyen de résolution ( citons par exemple l'inversion d'une matrice de grande taille). Ces difficultés augmentent de plus en plus lorsqu'un phénomène est représenté d'une façon satisfaisante par des équations non linéaires.

Actuellement, le calcul automatique s'est considérablement développé et beaucoup de méthodes de calcul sont disponibles permettant :

- Le traitement de non-linéarités des équations ainsi que leur résolution ;
- la représentation du phénomène étudié à l'aide de fonctions d'une série de mesure ;
- la réduction du temps de calcul.

L'objet de l'analyse numérique est de concevoir et fournir des méthodes de calcul permettant la résolution numérique d'un problème mathématique.

Le présent polycopié est organisé en six chapitres. Pour une revue relativement exhaustive, on pourra consulter les ouvrages mis en bibliographie en fin de ce polycopié.

A la fin de chaque chapitre nous avons proposé des exercices.avec quelques corrigés.

# Chapitre 1

## Résolution numérique de systèmes d'équations

### 1.1 Introduction

Soit le système algébrique linéaire de  $n$  équations à  $n$  inconnues :

$$\sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, \dots, n \quad (1.1)$$

ou encore

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n \end{cases} \quad (1.1')$$

$x_1, x_2, \dots, x_n$  sont les inconnues du système.

Le système (1.1) peut s'écrire sous forme matricielle

$$A.x = b \quad (1.2)$$

avec

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix}.$$

La notation précédente est équivalente à (l'écriture la plus utilisée) :

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} & b_n \end{pmatrix}$$

appelée matrice augmentée du système.

Si  $A$  est inversible, le système (1.2) a une solution unique qui s'écrit :

$$x = A^{-1}b \quad (1.3)$$

Si  $A$  est singulière ( i.e.  $\det A = 0$ ) le système n'admet pas de solution.

### 1.1.1 Solution théorique

On peut mettre (1.3) sous la forme

$$x = \frac{\text{com}(A)}{\det A} b, \quad \text{car } A^{-1} = \frac{\text{com}(A)}{\det A}.$$

où  $\text{com}(A)$  est la comatrice de  $A$  donnée par  $\text{com}(A) = (\text{cof } A)^t$  telle que  $\text{cof } A = [A_{ij}]_{(n \times n)}$  (matrice des cofacteurs de  $A$ ) avec  $A_{ij} = (-1)^{i+j} \det(a_{ij})$ ,  $a_{ij}$  étant une sous-matrice déduite de la matrice  $A$  par suppression de la  $i^{\text{ème}}$  ligne et la  $j^{\text{ème}}$  colonne.

Ceci constitue une méthode très inefficace car il y a plus de calculs que nécessaires et plus d'espace mémoire pour moins de précisions

### Méthode de Cramer

Si  $\det A \neq 0$  alors le système (1.2) a exactement une solution donnée par

$$x_j = \frac{\det A_j}{\det A}, \quad j = 1, 2, \dots, n.$$

Où  $A_j$  la matrice obtenue en remplaçant la  $j^{\text{ème}}$  colonne de  $A$  par  $b$ .

On devra donc calculer  $(n + 1)$  déterminants, pour chaque déterminant  $n.n!$  opérations soit  $(n + 1).n.n!$  opérations et effectuer  $n$  divisions soit au total de l'ordre de  $n^3.(n - 1)!$  opérations.

Ces méthodes sont donc lourdes du point de vue programmation du fait qu'elles demandent un temps de calcul considérable, une place considérable en

mémoire, qu'elles soient imprécises en raison de l'influence d'erreurs d'arrondis et du fait qu'elles soient difficiles à programmer, car le programme de calcul d'un déterminant est difficile à écrire.

### 1.1.2 Esprit des méthodes numériques

Le choix des méthodes de résolution dépend :

- de la taille du système
- du type et de la structure de la matrice  $A$  (forme et symétrie) intervenant dans le problème.

D'une façon générale on distingue deux types de méthodes :

**Les méthodes directes :** conduisent à un système équivalent au système initial en un nombre fini d'étapes mais plus facile à résoudre, généralement on se ramène à un système diagonal ou triangulaire. et mis à part les erreurs d'arrondis, ces méthodes conduisent à la solution exacte. Cependant, elles ne sont applicables que lorsque  $n < 50$  environ.

**Les méthodes indirecte ou itératives :** consistent à fournir la solution en un nombre infini d'opérations, par approximations successives et sans modifier le système initial. Ces méthodes s'appliquent pour résoudre des systèmes de grandes tailles.

## 1.2 Méthodes directes

Les méthodes directes ont pour principe de transformer la matrice augmentée (le système) en une autre matrice (système équivalent), le nouveau système ayant la même solution que l'original.

### 1.2.1 Méthode d'élimination de Gauss

La méthode d'élimination de Gauss a pour but de transformer le système  $Ax = b$  par un système équivalent (c'est-à-dire ayant la même solution) de la forme  $Ux = \hat{b}$ , où  $U$  est une matrice triangulaire supérieure et  $\hat{b}$  est un second membre convenablement modifié.

La méthode comporte  $(n-1)$  étapes. On notera  $a_{ij}^{(k)}$  l'élément  $a_{ij}$  à l'étape  $k$ .

**Étape 1 :**

- On transforme  $A$  en une matrice dont les termes sous diagonaux de la première colonne sont nuls.

- Pour éliminer le terme  $a_{21}$  on multiplie la ligne  $l_1$  par  $\left(\frac{-a_{21}}{a_{11}}\right)$

$$l_2^{(1)} \leftarrow l_2 + \left(\frac{-a_{21}}{a_{11}}\right) l_1$$

on obtient ainsi

$$\begin{aligned} a_{2j}^{(1)} &= a_{2j} - \left(\frac{a_{21}}{a_{11}}\right) a_{1j} \\ b_2^{(1)} &= b_2 - \left(\frac{a_{21}}{a_{11}}\right) b_1 \end{aligned}$$

- Pour éliminer le terme  $a_{31}$  on multiplie  $l_1$  par  $\left(\frac{-a_{31}}{a_{11}}\right)$

$$l_3^{(1)} \leftarrow l_3 + \left(\frac{-a_{31}}{a_{11}}\right) l_1$$

on obtient ainsi

$$\begin{aligned} a_{3j}^{(1)} &= a_{3j} - \left(\frac{a_{31}}{a_{11}}\right) a_{1j} \\ b_3^{(1)} &= b_3 - \left(\frac{a_{31}}{a_{11}}\right) b_1 \end{aligned}$$

- D'une manière générale pour éliminer tout les termes  $a_{i1}$  on utilise la transformation

$$l_i^{(1)} \leftarrow l_i + \left(\frac{-a_{i1}}{a_{11}}\right) l_1$$

$$\begin{aligned} a_{ij}^{(1)} &= a_{ij} - \left(\frac{a_{i1}}{a_{11}}\right) a_{1j} \\ b_i^{(1)} &= b_i - \left(\frac{a_{i1}}{a_{11}}\right) b_1 \end{aligned}$$

- A la fin de la première étape le système d'équations aura la forme suivante :

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & & a_{2n}^{(1)} & b_2^{(1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & a_{n2}^{(1)} & a_{n3}^{(1)} & \dots & a_{nn}^{(1)} & b_n^{(1)} \end{pmatrix}$$

**Étape 2 :**

Nous éliminons ensuite les termes sous-diagonaux de la seconde colonne.

Comme à la première étape pour éliminer le terme  $a_{32}$  on doit utiliser la transformation élémentaire

$$l_3^{(2)} \leftarrow l_3^{(1)} + \begin{pmatrix} -a_{32}^{(1)} \\ a_{22}^{(1)} \end{pmatrix} l_2^{(1)}$$

on obtient ainsi

$$\left\{ \begin{array}{l} a_{32}^{(2)} = a_{32}^{(1)} - \begin{pmatrix} a_{32}^{(1)} \\ a_{22}^{(1)} \end{pmatrix} a_{22}^{(1)} = 0 \\ a_{33}^{(2)} = a_{33}^{(1)} - \begin{pmatrix} a_{32}^{(1)} \\ a_{22}^{(1)} \end{pmatrix} a_{23}^{(1)} \\ \dots\dots\dots \\ a_{3n}^{(2)} = a_{3n}^{(1)} - \begin{pmatrix} a_{32}^{(1)} \\ a_{22}^{(1)} \end{pmatrix} a_{2n}^{(1)} \\ b_3^{(2)} = b_3^{(1)} - \begin{pmatrix} a_{32}^{(1)} \\ a_{22}^{(1)} \end{pmatrix} b_2^{(1)} \end{array} \right. \Leftrightarrow \begin{pmatrix} a_{3j}^{(2)} = a_{3j}^{(1)} - \begin{pmatrix} a_{32}^{(1)} \\ a_{22}^{(1)} \end{pmatrix} a_{2j}^{(1)} \\ b_3^{(2)} = b_3^{(1)} - \begin{pmatrix} a_{32}^{(1)} \\ a_{22}^{(1)} \end{pmatrix} b_2^{(1)} \end{pmatrix}$$

En continuant de la même façon que l'étape 1 on obtient la fin de la seconde étape.

Le système d'équations aura la forme suivante :

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \dots & a_{2n}^{(1)} & b_2^{(1)} \\ 0 & 0 & a_{33}^{(2)} & \dots & a_{3n}^{(2)} & b_3^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & a_{n3}^{(2)} & \dots & a_{nn}^{(2)} & b_n^{(2)} \end{bmatrix}$$

**Étape k :**

À la  $k^{ème}$  étape, nous éliminons les termes sous-diagonaux de la  $k^{ième}$  colonne

$$l_{k+1}^{(k)} \leftarrow l_{k+1}^{(k-1)} - \begin{pmatrix} a_{k+1,k}^{(k-1)} \\ a_{kk}^{(k-1)} \end{pmatrix} l_k^{(k-1)},$$

on obtient ainsi

$$\left\{ \begin{array}{l} a_{k+1,k}^{(k)} = a_{k+1,k}^{(k-1)} - \left( \frac{a_{k+1,k}^{(k-1)}}{a_{kk}^{(k-1)}} \right) a_{kk}^{(k-1)} = 0 \\ a_{k+1,k+1}^{(k)} = a_{k+1,k+1}^{(k-1)} - \left( \frac{a_{k+1,k}^{(k-1)}}{a_{kk}^{(k-1)}} \right) a_{k,k+1}^{(k-1)} \\ \dots\dots\dots \\ a_{k+1,n+1}^{(k)} = a_{k+1,j}^{(k-1)} - \left( \frac{a_{k+1,k}^{(k-1)}}{a_{kk}^{(k-1)}} \right) a_{k,n+1}^{(k-1)} \end{array} \right. \Leftrightarrow \begin{array}{l} a_{k+1j}^{(k)} = a_{k+1,j}^{(k-1)} - \left( \frac{a_{k+1,k}^{(k-1)}}{a_{kk}^{(k-1)}} \right) a_{k,j}^{(k-1)} \\ b_{k+1}^{(k)} = b_{k+1}^{(k-1)} - \left( \frac{a_{k+1,k}^{(k-1)}}{a_{kk}^{(k-1)}} \right) b_1 \end{array}$$

A la fin de la dernière étape on obtient alors le système triangulaire supérieur suivant :

$$\begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \dots & \dots & a_{2n}^{(2)} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & a_{nn}^{(n-1)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1^{(1)} \\ b_2^{(2)} \\ \vdots \\ \vdots \\ b_n^{(n-1)} \end{pmatrix} \quad (1.4)$$

**Remarque** Les opérations précédentes supposent que les termes  $a_{kk}^{(k)}$  appelés pivots sont non nuls.

**Exemple 1.2.1** Soit le système d'équations représenté par la matrice augmentée suivante :

$$\begin{bmatrix} 2 & 1 & 2 & 10 \\ 6 & 4 & 0 & 26 \\ 8 & 5 & 1 & 35 \end{bmatrix}$$

$$\begin{cases} l_2 \leftarrow l_2 - \left(\frac{6}{2}\right) l_1 \\ l_3 \leftarrow l_3 - \left(\frac{8}{2}\right) l_1 \end{cases}$$

$$\begin{bmatrix} 2 & 1 & 2 & 10 \\ 0 & 1 & -6 & -4 \\ 0 & 1 & -7 & -5 \end{bmatrix}$$

$$l_3 \leftarrow l_3 - \left(\frac{1}{1}\right) l_2$$

$$\begin{bmatrix} 2 & 1 & 2 & 10 \\ 0 & 1 & -6 & -4 \\ 0 & 0 & -1 & -1 \end{bmatrix}$$

la solution du système est

$$x_3 = 1, \quad x_2 = 2, \quad x_1 = 3$$

### Remarque

- La méthode de Gauss nécessite  $\frac{2n^3}{3}$  opérations pour un système de taille  $n$ .
- Dans la méthode de Gauss si un pivot est nul on permutera l'équation correspondante avec l'une de ses suivantes ( permutation de 2 lignes ), donc pour  $\overline{i = k + 1, n}$ , de manière à mettre un pivot non nul en position  $k$ .

Si une telle permutation est impossible la matrice est alors singulière.

- Si le pivot de valeur proche de zéro c'est-à-dire pivot petit, les quantités  $\frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}$  deviennent plus grandes que 1 et amplifions les erreurs de troncature; dans ce cas il y peut y avoir perte de précision, et on opte pour la modification de l'algorithme de Gauss, d'où les méthodes suivantes :

**Pivot partiel** On intervertit les lignes à chaque étape de façon à placer en pivot le terme de coefficient le plus élevé de la ligne.

C'est la méthode du pivot partiel, à la  $k$ ième étape, le pivot est l'élément :

$$a_{i,k}^{(k)} = \max_{p=k, \dots, n} |a_{p,k}^{(k)}|$$

**Pivot total** On intervertit les lignes et les colonnes de façon à placer en pivot le terme de coefficient le plus élevé de la matrice : C'est la méthode du pivot total, à la  $k$ -ième étape, le pivot est l'élément

$$a_{i,j}^{(k)} = \max_{p,q=k, \dots, n} |a_{p,q}^{(k)}|.$$

**Exemple 1.2.2**  $A = \begin{pmatrix} 1 & 3 & 3 \\ 1 & 1 & 0 \\ 3 & 2 & 6 \end{pmatrix}$ ,  $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$   $b = \begin{pmatrix} 0 \\ 1 \\ 11 \end{pmatrix}$ ,

$k = 1$  :  $\max(1, 1, 3) = 3$  on permute, par exemple, les lignes 1 et 3 on obtient

$$\begin{pmatrix} \boxed{3} & 2 & 6 & \vdots & 11 \\ 1 & 1 & 0 & \vdots & 1 \\ 1 & 3 & 3 & \vdots & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 3 & 2 & 6 & \vdots & 11 \\ 0 & \frac{1}{3} & -2 & \vdots & \frac{-8}{3} \\ 0 & \frac{7}{3} & 1 & \vdots & \frac{-11}{3} \end{pmatrix}$$

$k = 2 : \max(\frac{1}{3}, \frac{7}{3}) = \frac{7}{3}$ . On permute les lignes 2 et 3.

$$\begin{pmatrix} 3 & 2 & 6 & \vdots & 11 \\ 0 & \frac{7}{3} & 1 & \vdots & \frac{-11}{3} \\ 0 & \frac{1}{3} & -2 & \vdots & \frac{-8}{3} \end{pmatrix} \rightarrow \begin{pmatrix} 3 & 2 & 6 & \vdots & 11 \\ 0 & \frac{7}{3} & 1 & \vdots & \frac{-11}{3} \\ 0 & 0 & \frac{-15}{7} & \vdots & \frac{-15}{7} \end{pmatrix}$$

Il s'ensuit que :

$$Ax = b \Leftrightarrow \begin{cases} x_3 = 1 \\ x_2 = \frac{3}{7} \left( \frac{-11}{3} - x_3 \right) \\ x_1 = \frac{1}{3} (11 - 2x_2 - 6x_3) \end{cases} \Leftrightarrow \begin{cases} x_3 = 1 \\ x_2 = -2 \\ x_1 = 3 \end{cases}$$

**Exemple 1.2.3** Soit le système  $\begin{cases} x_1 + 3x_2 + 3x_3 = -2 \\ 2x_1 + 2x_2 + 5x_3 = 7 \\ 3x_1 + 2x_2 + 6x_3 = 12 \end{cases}$

Posons :

$$(A:b) = \begin{pmatrix} 1 & 3 & 3 & \vdots & -2 \\ 2 & 2 & 5 & \vdots & 7 \\ 3 & 2 & 6 & \vdots & 12 \end{pmatrix}$$

$k = 1 : \max(|a_{i,j}^{(1)}|, i = 1, 2, 3 \quad j = 1, 2, 3) = 6$ . La ligne du pivot total sera alors  $L_3$ . En permutant les lignes 1 et 3 on obtient :

$$\begin{pmatrix} 3 & 2 & 6 & \vdots & 12 \\ 2 & 2 & 5 & \vdots & 7 \\ 1 & 3 & 3 & \vdots & -2 \end{pmatrix}$$

La colonne du pivot total est la colonne 3, et on permute les colonnes 1 et 3 :

$$\begin{pmatrix} \boxed{6} & 2 & 3 & \vdots & 12 \\ 5 & 2 & 2 & \vdots & 7 \\ 3 & 3 & 1 & \vdots & -2 \end{pmatrix} \rightarrow \begin{pmatrix} 6 & 2 & 3 & \vdots & 12 \\ 0 & \frac{1}{3} & \frac{-1}{2} & \vdots & -3 \\ 0 & 2 & \frac{-1}{2} & \vdots & -8 \end{pmatrix}$$

$k = 2 : \max(|a_{i,j}^{(2)}|, i = 2, 3 \quad j = 2, 3) = 2$ .

La ligne du pivot total est alors  $L_3$ . Et donc on permute les lignes 2 et 3 :

$$\begin{pmatrix} 6 & 2 & 3 & \vdots & 12 \\ 0 & \boxed{2} & \frac{-1}{2} & \vdots & -8 \\ 0 & \frac{1}{3} & \frac{-1}{2} & \vdots & -3 \end{pmatrix} \rightarrow \begin{pmatrix} 6 & 2 & 3 & \vdots & 12 \\ 0 & 2 & \frac{-1}{2} & \vdots & -8 \\ 0 & 0 & \frac{-5}{12} & \vdots & \frac{-5}{3} \end{pmatrix}$$

Du fait de la permutation précédentes des colonnes 1 et 3 on obtient le système équivalent suivant :

$$\left\{ \begin{array}{l} 6x_3 + 2x_2 + 3x_1 = 12 \\ 2x_2 - \frac{1}{2}x_1 = -8 \\ -\frac{5}{12}x_1 = \frac{-5}{3} \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} x_3 = 1 \\ x_2 = -3 \\ x_1 = 4 \end{array} \right.$$

**Remarque** Notons bien à chaque permutation de colonnes les inconnues changent de places.

### 1.2.2 Méthode de factorisation LU

La méthode  $LU$  consiste à transformer la matrice  $A$  en un produit de deux matrices triangulaires inférieures et l'autre supérieure.

$$A = LU$$

où  $L$  est une matrice triangulaire inférieure.

$U$  est une matrice triangulaire supérieure.

Une fois la décomposition effectuée il faut résoudre deux systèmes à matrices triangulaires

$$Ax = b \Leftrightarrow L \underbrace{Ux}_y = b$$

On résoudra d'abord le système

$$Ly = b$$

une fois  $y$  trouvé on résoudra le second système

$$Ux = y$$

**Remarque** La méthode impose en général que la diagonale de  $U$  soit des unités .

**Exemple 1.2.4** Résoudre le système

$$\begin{bmatrix} 2 & 1 & 2 & 10 \\ 6 & 4 & 0 & 26 \\ 8 & 5 & 1 & 35 \end{bmatrix}$$

$$\begin{pmatrix} 2 & 1 & 2 \\ 6 & 4 & 0 \\ 8 & 5 & 1 \end{pmatrix} = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} 1 & U_{12} & U_{13} \\ 0 & 1 & U_{23} \\ 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} L_{11} & L_{11}U_{12} & L_{11}U_{13} \\ L_{21} & L_{22} + L_{21}U_{12} & L_{21}U_{13} + L_{22}U_{23} \\ L_{31} & L_{32} + L_{31}U_{12} & L_{33} + L_{31}U_{13} + L_{32}U_{23} \end{pmatrix}$$

$$\Rightarrow (1) \begin{cases} L_{11} = 2 \\ L_{21} = 6 \\ L_{31} = 8 \end{cases} \quad (2) \begin{cases} L_{11}U_{12} = 1 \\ L_{22} + L_{21}U_{12} = 4 \\ L_{32} + L_{31}U_{12} = 5 \end{cases} \quad (3) \begin{cases} L_{11}U_{13} = 2 \\ L_{21}U_{13} + L_{22}U_{23} = 0 \\ L_{33} + L_{31}U_{13} + L_{32}U_{23} = 1 \end{cases}$$

$$\begin{cases} L_{11} = 2 \\ L_{21} = 6 \\ L_{31} = 8 \end{cases}$$

$$(2) \Leftrightarrow \begin{cases} L_{11}U_{12} = 1 \Rightarrow U_{12} = \frac{1}{2} \\ L_{22} + L_{21}U_{12} = 4 \Rightarrow L_{22} = 1 \\ L_{32} + L_{31}U_{12} = 5 \Rightarrow L_{32} = 1 \end{cases}$$

$$(3) \Leftrightarrow \begin{cases} L_{11}U_{13} = 2 \Rightarrow U_{13} = 1 \\ L_{21}U_{13} + L_{22}U_{23} = 0 \Rightarrow U_{23} = -6 \\ L_{33} + L_{31}U_{13} + L_{32}U_{23} = 1 \Rightarrow L_{33} = -1 \end{cases}$$

$$L = \begin{pmatrix} 2 & 0 & 0 \\ 6 & 1 & 0 \\ 8 & 1 & -1 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & \frac{1}{2} & 1 \\ 0 & 1 & -6 \\ 0 & 0 & 1 \end{pmatrix}$$

$$Ly = b \Rightarrow \begin{cases} 2y_1 = 10 \\ 6y_1 + y_2 = 26 \\ 8y_1 + y_2 - y_3 = 35 \end{cases} \Rightarrow \begin{cases} y_1 = 5 \\ y_2 = -4 \\ y_3 = 1 \end{cases}$$

$$Ux = y \Rightarrow \begin{cases} x_1 + \frac{1}{2}x_2 + x_3 = 5 \\ x_2 - 6x_3 = -4 \\ x_3 = 1 \end{cases}$$

$$\begin{cases} x_1 + \frac{1}{2}x_2 + x_3 = 5 \\ x_2 - 6x_3 = -4 \\ x_3 = 1 \end{cases} \Rightarrow \begin{cases} x_1 = 3 \\ x_2 = 2 \\ x_3 = 1 \end{cases}$$

## 1.3 Méthodes itératives

### 1.3.1 Introduction

Les méthodes itératives consistent à construire une suite de vecteurs  $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})^t$  convergente vers le vecteur solution exacte  $x = (x_1, x_2, \dots, x_n)^t$  du système (2.2) pour tout vecteur initial  $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^t$ .

### 1.3.2 Principe des méthodes itératives

Soit le système d'équations suivant :

$$A x = b.$$

Ecrivons la matrice  $A$  sous la forme suivante

$$A = M - N.$$

La matrice  $M$  doit être inversible. alors le système peut s'écrire :

$$\begin{aligned} (M - N) x &= b \Rightarrow M x = N x + b \\ x &= M^{-1} N x + M^{-1} b \end{aligned}$$

La méthode itérative associée à l'égalité précédente consiste à partir d'un vecteur initial  $x^{(0)}$  à générer la suite  $x^{(1)}, x^{(2)}, \dots, x^{(k+1)}$  de la manière suivante :

$$\begin{aligned} x^{(1)} &= M^{-1} N x^{(0)} + M^{-1} b \\ x^{(2)} &= M^{-1} N x^{(1)} + M^{-1} b \\ &\dots\dots\dots \\ x^{(k+1)} &= M^{-1} N x^{(k)} + M^{-1} b. \end{aligned}$$

Cette suite d'égalités peut être représentée par la relation itérative suivante :

$$x^{(k+1)} = T x^{(k)} + V,$$

avec  $T = M^{-1} N$ ,  $V = M^{-1} b$

### 1.3.3 Le choix de la décomposition

Les différentes méthodes itératives ne diffèrent que par le choix de la décomposition de la matrice  $A$

Soit  $A$  la matrice

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

Définissons les matrices suivantes

$$\begin{cases} d_{ii} = a_{ii} : \forall i \\ d_{ij} = 0 : \forall i \neq j \end{cases} \iff D = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

$$\begin{cases} l_{ij} = -a_{ij} : \forall i > j \\ 0 : \forall i \leq j \end{cases} \iff L = \begin{pmatrix} 0 & 0 & \dots & 0 \\ -a_{21} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ -a_{n1} & -a_{n2} & \dots & 0 \end{pmatrix}$$

$$\begin{cases} u_{ij} = -a_{ij} : \forall i < j \\ 0 : \forall i \geq j \end{cases} \iff U = \begin{pmatrix} 0 & -a_{12} & \dots & -a_{1n} \\ 0 & 0 & \dots & -a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

A partir de ces matrices on a  $A = D - L - U$ , selon les différentes possibilités d'écrire cette égalité sous la forme  $A = M - N$ .

### 1.3.4 Méthode de Jacobi

La matrice  $A$  étant décomposée en

$$A = M - N = D - (L + U) \dots (D = M)$$

Le système  $A x = b$  peut s'écrire donc sous la forme

$$D x = (L + U) x + b.$$

La méthode de Jacobi s'écrit donc

$$x = D^{-1} (L + U) x + D^{-1} b.$$

La matrice  $H_J = D^{-1} (L + U)$  est appelée "matrice de Jacobi" associée à  $A$ .

On obtient encore

$$x^{(k+1)} = D^{-1} (L + U) x^{(k)} + D^{-1} b.$$

Dans la méthode de Jacobi alors, pour une donnée  $x^{(0)}$ , on calcule  $x^{(k+1)}$  selon la formule

$$x_i^{(k+1)} = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ j \neq i}} \frac{a_{ij}}{a_{ii}} x_j^k, \quad \text{avec } a_{ii} \neq 0; \quad i = 1, \dots, n. \quad (1.5)$$

La méthode de Jacobi suppose des pivots  $a_{ii}$  non nuls sinon un simple échange de ligne suffit à résoudre le problème (pour avoir  $D$  inversible).

### 1.3.5 Méthode de Gauss-Seidel

La matrice  $A$  étant décomposée en :

$$A = M - N = (D - L) - U$$

Le système  $Ax = b$  peut s'écrire donc sous la forme

$$(D - L) x = U x + b$$

La méthode de Gauss-Seidel s'écrit donc

$$x = (D - L)^{-1} U x + (D - L)^{-1} b.$$

La matrice  $H_{GS} = (D - L)^{-1} U$  est appelée la matrice de Gauss-Seidel associée à  $A$

Soit :

$$x^{(k+1)} = (D - L)^{-1} U x^{(k)} + (D - L)^{-1} b.$$

L'inverse de la matrice  $D - L$  peut être compliqué à calculer, alors on décompose de la manière suivante :

$$\begin{aligned} D x^{(k+1)} &= L x^{(k+1)} + U x^{(k)} + b \Rightarrow \\ x^{(k+1)} &= D^{-1} L x^{(k+1)} + D^{-1} U x^{(k)} + D^{-1} b. \end{aligned}$$

Dans la méthode de Gauss-Seidel alors, pour une donnée  $x^{(0)}$ , on calcule  $x^{(k+1)}$  selon la formule

$$x_i^{(k+1)} = \frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{k+1} - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} x_j^k, \quad i = 1, \dots, n. \quad (1.6)$$

**Critère d'arrêt**

1. En pratique les itérations s'arrêtent quand

$$\|x^{(k+1)} - x^{(k)}\| \leq \varepsilon, \quad \varepsilon \text{ tolérance fixée}$$

2. Pour  $\|T\| < 1$  si on prends  $x^{(0)} = V$ ,  
on aura  $x^{(1)} = Tx^{(0)} + V \Rightarrow x^{(1)} - x^{(0)} = TV$  d'où

$$\|x - x^{(k)}\| \leq \frac{\|x^{(k+1)} - x^{(k)}\|}{1 - \|T\|} \leq \frac{\|T\|^{k+1} \|V\|}{1 - \|T\|}$$

et en posant

$$\frac{\|T\|^{k+1} \|V\|}{1 - \|T\|} \leq \varepsilon$$

ce qui nous permet d'estimer le nombre suffisant d'itérations  $k$  pour approximer  $x$  avec la précision donnée  $\varepsilon$ .

**Etude de convergence**

**Définition 1.3.1** .Une matrice  $A$  est dite à diagonale strictement dominante si :

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}|, \quad \forall i. \in \{1, \dots, n\}$$

ou

$$|a_{jj}| > \sum_{i=1, i \neq j}^n |a_{ij}|, \quad \forall j. \in \{1, \dots, n\}$$

Cette définition signifie que le terme diagonal  $a_{ii}$  de la matrice  $A$  est nettement dominant puisque sa valeur absolue est plus grande que la somme des valeurs absolues de tous les autres termes de la ligne

**Proposition 1.3.1** Si la matrice  $A$  est à diagonale strictement dominante, les méthodes de JACOBI et de GAUSS-SEIDEL convergent.

La méthode de GAUSS-SEIDEL est plus rapide que celle de JACOBI.

**Exemple 1.3.1** Considérons le système linéaire

$$\begin{pmatrix} 4 & 2 & 1 \\ -1 & 2 & 0 \\ 2 & 1 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 9 \end{pmatrix}$$

*mis sous la forme*

$$\begin{cases} x = 1 - \frac{y}{2} - \frac{z}{4}, \\ y = 1 + \frac{x}{2} \\ z = \frac{9}{4} - \frac{x}{2} - \frac{y}{4}. \end{cases}$$

Soit  $x^{(0)} = (0, 0, 0)$  le vecteur initial.

- En calculant les itérés avec la méthode de JACOBI en considérant la suite

$$\begin{cases} x_{n+1} = 1 - \frac{1}{2}y_n - \frac{1}{4}z_n, \\ y_{n+1} = 1 + \frac{1}{2}x_n \\ z_{n+1} = \frac{9}{4} - \frac{1}{2}x_n - \frac{1}{4}y_n. \end{cases}$$

*on trouve*

$$x^{(1)} = \begin{pmatrix} 1 \\ 1 \\ \frac{9}{4} \end{pmatrix}, \quad x^{(2)} = \begin{pmatrix} \frac{-1}{16} \\ \frac{3}{2} \\ \frac{3}{2} \end{pmatrix}, \quad x^{(3)} = \begin{pmatrix} \frac{-1}{8} \\ \frac{-1}{32} \\ \frac{61}{32} \end{pmatrix}, \quad x^{(4)} = \begin{pmatrix} \frac{5}{128} \\ \frac{15}{16} \\ \frac{265}{128} \end{pmatrix}.$$

La suite  $x^{(k)}$  converge vers  $(0, 1, 2)$  la solution du système au bout de quatre itérations.

-En calculant les itérés avec la méthode GAUSS-SEIDEL en considérant la suite

$$\begin{cases} x_{n+1} = 1 - \frac{1}{2}y_n - \frac{1}{4}z_n, \\ y_{n+1} = 1 + \frac{1}{2}x_{n+1} \\ z_{n+1} = \frac{9}{4} - \frac{1}{2}x_{n+1} - \frac{1}{4}y_{n+1}. \end{cases}$$

*on trouve*

$$x^{(1)} = \begin{pmatrix} 1 \\ \frac{3}{2} \\ \frac{11}{8} \end{pmatrix}, \quad x^{(2)} = \begin{pmatrix} \frac{-3}{32} \\ \frac{61}{64} \\ \frac{527}{256} \end{pmatrix}, \quad x^{(3)} = \begin{pmatrix} \frac{9}{1024} \\ \frac{2047}{2048} \\ \frac{16349}{8192} \end{pmatrix},$$

La suite  $x^{(k)}$  converge vers  $(0, 1, 2)$  la solution du système au bout de trois itérations.

## 1.4 Exercices

**Exercice 1** Soit le système d'équations linéaires suivant

$$(S) \begin{cases} 3x_1 - 2x_2 + x_3 = 2 \\ 2x_1 + x_2 + x_3 = 0 \\ 4x_1 - 3x_2 + 2x_3 = 4 \end{cases}$$

1) Ecrire la forme matricielle  $Ax = b$  de ce système et le résoudre à l'aide de la méthode de Gauss.

2) Calculer le déterminant de la matrice  $A$ .

3) Factoriser la matrice  $A$  du système en produit  $LU$ , puis résoudre ce système.

**Exercice 2** Résoudre le système d'équations linéaires suivant

$$\begin{cases} 3x_1 + x_2 - x_3 = 2 \\ x_1 + 5x_2 + 2x_3 = 17 \\ 2x_1 - x_2 - 6x_3 = -18 \end{cases}$$

par la méthode de Jacobi, ensuite par la méthode de Gauss-Seidel avec  $x^{(0)} = (0, 0, 0)^T$ .

**Exercice 3** Soit le systèmes d'équations linéaires suivant

$$\begin{cases} 2x_1 - x_2 + x_3 = 1 \\ -x_1 + 2x_2 - x_3 = 2 \\ x_1 - x_2 + 2x_3 = 3 \end{cases}$$

1) Ecrire ce système sous la forme matricielle  $Ax = b$  et montrer que ce système admet une solution unique.

2) Peut-on décomposer la matrice  $A$  sous la forme  $A = L L^T$ , où  $L$  est une matrice triangulaire inférieure ?

3) Si la réponse est oui, résoudre ce système par la méthode de Cholesky, en déduire la valeur du déterminant de  $A$ .

**Exercice 4** Soit le systèmes d'équations linéaires suivant

$$(S) \begin{cases} 4x_1 + x_2 + 2x_3 = 4 \\ 3x_1 + 5x_2 + x_3 = 7 \\ x_1 + x_2 + 3x_3 = 3 \end{cases}$$

Donner le nombre suffisant d'itérations pour que les deux méthodes ( Jacobi et Gauss-Seidel ) donnent une solution approchée de  $(S)$  à  $10^{-2}$  près.

### 1.4.1 Réponses

**Exercice 1**

1) Le système  $(S)$  sous sa forme matricielle

$$\begin{bmatrix} 3 & -2 & 1 \\ 2 & 1 & 1 \\ 4 & -3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}$$

avec la matrice  $A = \begin{bmatrix} 3 & -2 & 1 \\ 2 & 1 & 1 \\ 4 & -3 & 2 \end{bmatrix}$ ,  $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$  et  $b = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}$ .

$\det A = 5$  alors le système  $(S)$  a une solution admet une solution unique.

$$\begin{bmatrix} 3 & -2 & 1 & \vdots & 2 \\ 2 & 1 & 1 & \vdots & 0 \\ 4 & -3 & 2 & \vdots & 4 \end{bmatrix} \begin{matrix} \\ L_2 - \frac{2}{3}L_1 \\ \text{et } L_3 - \frac{4}{3}L_1 \end{matrix} \begin{bmatrix} 3 & -2 & 1 & \vdots & 2 \\ 0 & \frac{7}{3} & \frac{1}{3} & \vdots & -\frac{4}{3} \\ 0 & -\frac{1}{3} & \frac{2}{3} & \vdots & \frac{4}{3} \end{bmatrix}$$

$$L_3 + \frac{1}{7}L_2 \begin{bmatrix} 3 & -2 & 1 & \vdots & 2 \\ 0 & \frac{7}{3} & \frac{1}{3} & \vdots & -\frac{4}{3} \\ 0 & 0 & \frac{15}{21} & \vdots & \frac{24}{21} \end{bmatrix} \Rightarrow x_3 = \frac{8}{5}, \quad x_2 = -\frac{2}{5} \text{ et } x_1 = -\frac{2}{5}$$

La solution du système est  $x = (-\frac{2}{5}, -\frac{2}{5}, \frac{8}{5})^T$ .

$$2) \det A = \det \begin{bmatrix} 3 & -2 & 1 \\ 0 & \frac{7}{3} & \frac{1}{3} \\ 0 & -\frac{1}{3} & \frac{2}{3} \end{bmatrix} = \det \begin{bmatrix} 3 & -2 & 1 \\ 0 & \frac{7}{3} & \frac{1}{3} \\ 0 & 0 & \frac{15}{21} \end{bmatrix} = 5.$$

3)

$$\begin{aligned}
A &= LU = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} 1 & U_{12} & U_{13} \\ 0 & 1 & U_{23} \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} L_{11} & L_{11}U_{12} & L_{11}U_{13} \\ L_{21} & L_{21}U_{12} + L_{22} & L_{21}U_{13} + L_{22}U_{23} \\ L_{31} & L_{31}U_{12} + L_{32} & L_{31}U_{13} + L_{32}U_{23} + L_{33} \end{bmatrix} = \begin{bmatrix} 3 & -2 & 1 \\ 2 & 1 & 1 \\ 4 & -3 & 2 \end{bmatrix}
\end{aligned}$$

ce qui donne  $L_{11} = 3$ ,  $L_{11}U_{12} = -2 \implies U_{12} = -\frac{2}{3}$ ,  $L_{21} = 2$   
 et  $L_{21}U_{12} + L_{22} = 1 \implies L_{22} = \frac{7}{3}$ ,

$$L_{31} = 4 \text{ et } L_{31}U_{12} + L_{32} = -3 \implies L_{32} = -\frac{1}{3}.$$

$$L_{11}U_{13} = 1 \implies U_{13} = \frac{1}{3} \text{ et } L_{21}U_{13} + L_{22}U_{23} = 1 \implies U_{23} = \frac{1}{7}.$$

$$L_{31}U_{13} + L_{32}U_{23} + L_{33} = 2 \implies L_{33} = \frac{5}{7}.$$

$$A = LU = \begin{bmatrix} 3 & 0 & 0 \\ 2 & \frac{7}{3} & 0 \\ 4 & -\frac{1}{3} & \frac{5}{7} \end{bmatrix} \begin{bmatrix} 1 & -\frac{2}{3} & \frac{1}{3} \\ 0 & 1 & \frac{1}{7} \\ 0 & 0 & 1 \end{bmatrix}$$

$$Ly = b \implies \begin{cases} 3y_1 = 2 \\ 2y_1 + \frac{7}{3}y_2 = 0 \\ 4y_1 - \frac{1}{3}y_2 + \frac{5}{7} = 4 \end{cases}$$

ce qui donne  $y_1 = \frac{2}{3}$ ,  $y_2 = -\frac{4}{7}$ ,  $y_3 = \frac{8}{5}$ .

Ensuite, on résoud  $UX = Y$

$$\begin{bmatrix} 1 & -\frac{2}{3} & \frac{1}{3} \\ 0 & 1 & \frac{1}{7} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ -\frac{4}{7} \\ \frac{8}{5} \end{bmatrix}$$

d'où la solution  $x = (x_1, x_2, x_3)^T = (-\frac{2}{5}, -\frac{2}{5}, \frac{8}{5})^T$ .

### Exercice 2

Indication on pose  $A = D - (L + U)$ ,  $Ax = b \implies x^{-1}D^{-1}(L + U)x + D^{-1}b$

$$\text{avec } H_j = D^{-1}(L + U) = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & -\frac{1}{6} \end{bmatrix} \left( \begin{bmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ -2 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 & 1 \\ 0 & 0 & -2 \\ 0 & 0 & 0 \end{bmatrix} \right),$$

$\rho(H_j) < 1$ .

On définit la suite  $x^{(k)}$  par

$$\begin{cases} x^{(k)} = D^{-1}(L + U) x^{(k-1)} + D^{-1}b \\ x^{(0)} = (0, 0, 0) \end{cases}$$

Il s'ensuit que

$$\forall 1 \leq i \leq 3, \quad x_i^{(k)} = \frac{1}{a_{ii}} \left( b_i - \sum_{\substack{j=1 \\ i \neq j}}^3 a_{ij} x_j^{(k-1)} \right),$$

où

$$\begin{cases} x_1^{(k)} = \frac{1}{3}(2 - x_2^{(k-1)} + x_3^{(k-1)}) \\ x_2^{(k)} = \frac{1}{5}(17 - x_1^{(k-1)} - 2x_3^{(k-1)}) \\ x_3^{(k)} = -\frac{1}{6}(-18 - 2x_1^{(k-1)} + x_2^{(k-1)}). \end{cases}$$

A l'aide d'un calcul simple et en faisant varier  $k = 0, \dots, 7$  ( 7 itérations) on déduit que la  $(x_1^{(7)}, x_2^{(7)}, x_3^{(7)})^T$  converge vers la solution exacte  $(1, 2, 3)^T$ .

Pour la méthode de Gauss-Seidel on pose

$$\begin{aligned} A &= (D - L) - U \\ Ax &= b \implies (D - L)x = Ux + b \\ x &= (D - L)^{-1}Ux + (D - L)^{-1}b \end{aligned}$$

On définit la suite  $x^{(k)}$  par

$$\begin{cases} x^{(k)} = (D - L)^{-1}Ux^{(k-1)} + (D - L)^{-1}b \\ x^{(0)} = (0, 0, 0) \end{cases}, \quad \rho(H_{GS}) < 1.$$

Il s'ensuit que

$$\forall 1 \leq i \leq 3, \quad x_i^{(k)} = \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k)} - \sum_{j=i+1}^3 a_{ij} x_j^{(k-1)} \right)$$

où

$$\begin{cases} x_1^{(k)} = \frac{1}{3}(2 - x_2^{(k)} + x_3^{(k-1)}) \\ x_2^{(k)} = \frac{1}{5}(17 - x_1^{(k)} - 2x_3^{(k-1)}) \\ x_3^{(k)} = -\frac{1}{6}(-18 - 2x_1^{(k)} + x_2^{(k)}). \end{cases}$$

Au bout de 6 itérations en faisant  $k = 0, \dots, 6$  on trouve  $x^{(6)} = (0, 9914, 2, 0057, 2, 9962)^T$  qui converge vers la solution exacte  $(1, 2, 3)^T$ .

#### Exercice 4

$$H_J = D^{-1}(L + U) = \begin{bmatrix} 0 & -\frac{1}{4} & -\frac{1}{2} \\ -\frac{3}{5} & 0 & -\frac{1}{5} \\ -\frac{1}{3} & -\frac{1}{3} & 0 \end{bmatrix} \text{ avec } D^{-1} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

et  $\|H_J\|_\infty = \frac{4}{5} < 1$  méthode de Jacobi converge.

Posons

$$c = D^{-1}b = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 4 \\ \frac{7}{5} \\ 1 \end{bmatrix}$$

$$x = H_J x + c.$$

On sait que

$$\|x - x^{(k)}\| \leq \frac{\|H\|^k}{1 - \|H\|} \|x^{(1)} - x^{(0)}\|.$$

Puisque  $x^{(k)} = H_J x^{(k-1)} + D^{-1}b \implies x^{(1)} - x^{(0)} = H_J c$

et

$$\|x - x^{(k)}\|_{\infty} \leq \frac{\|H_J\|^{k+1} \|c\|}{1 - \|H_J\|} = \frac{\left(\frac{4}{5}\right)^{k+1} \frac{7}{5}}{1 - \frac{4}{5}} \leq 10^{-2} \implies k = 29.$$

De même on a  $H_{GS} = (D - L)^{-1}U$  et  $x^{(k)} = H_{GS} x^{(k-1)} + c$  avec

$$D - L = \begin{bmatrix} 4 & 0 & 0 \\ 3 & 5 & 0 \\ 1 & 1 & 3 \end{bmatrix} \quad \text{et} \quad (D - L)^{-1} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ \frac{-3}{20} & \frac{1}{5} & 0 \\ \frac{-1}{30} & \frac{-1}{5} & \frac{1}{3} \end{bmatrix}$$

$$\text{Alors } H_{GS} = \begin{bmatrix} 0 & \frac{-1}{4} & \frac{-1}{2} \\ 0 & \frac{3}{20} & \frac{1}{10} \\ 0 & \frac{1}{30} & \frac{2}{15} \end{bmatrix} \quad \text{et} \quad \|H_{GS}\|_{\infty} = \frac{3}{4} < 1 \text{ GS est CV.}$$

$$c = (D - L)^{-1}b = \begin{bmatrix} 1 \\ \frac{4}{5} \\ \frac{2}{5} \end{bmatrix} \quad \text{et} \quad \|c\|_{\infty} = 1,$$

$$\|x - x^{(k)}\|_{\infty} \leq \frac{\|H_{GS}\|^{k+1} \|c\|}{1 - \|H_{GS}\|} = \left(\frac{3}{4}\right)^{k+1} \frac{1}{1 - \frac{3}{4}} \leq 10^{-2} \implies k = 20.$$

## Chapitre 2

# Calcul numérique des valeurs propres d'une matrice

Le calcul des valeurs et vecteurs propres est indispensable dans plusieurs applications. Notamment, dans la théorie de la stabilité des systèmes dynamiques, en physique, mécanique, chimie....

Dans ce chapitre nous allons aborder le problème du calcul de valeurs et vecteurs propres d'une matrice qui s'avère être un problème ardu. En effet, déterminer les valeurs propres d'une matrice revient à calculer les racines de son polynôme caractéristique. Or, il ne suffit pas de factoriser ce polynôme afin d'obtenir les racines, d'après les résultats de GALOIS, N. ABEL et P. RUFFINI il n'existe pas toujours de possibilité d'exprimer les racines d'un polynôme de degré supérieur ou égal à cinq à partir des coefficients du polynôme et d'opérations élémentaires. En fin de compte, pour le calcul de valeurs propres d'une matrice il est nécessaire d'avoir recours aux méthodes itératives.

Nous commençons par donner quelques rappels et définitions qui seront utiles par la suite.

## 2.1 Rappels. Définitions

### 2.1.1 Rappels élémentaires sur les matrices

- On note  $A = (a_{ij})$  une matrice avec  $1 \leq i \leq m$  (lignes) et  $1 \leq j \leq n$  (colonnes).
- On appelle transposé de  $A$  et l'on note  $A^t = (a_{ji})$ .
- La matrice adjointe est donnée par  $A^* = (\overline{a_{ji}})$ .
- Une matrice carrée  $A$  d'ordre  $n$  est inversible s'il existe une matrice, notée  $A^{-1}$  telle que  $A A^{-1} = A^{-1} A = I_n$ , où  $I_n$  est la matrice identité.
- Soit  $A$  une matrice carrée réelle, alors
  - $A$  est symétrique si  $A^t = A$  ;
  - $A$  est orthogonale si  $A A^t = A^t A = I_n$ .
- Soit  $A$  une matrice carrée complexe, alors
  - $A$  est hermitienne si  $A^* = A$  ;
  - $A$  est unitaire si  $A A^* = A^* A = I_n$  ;
  - $A$  est normale si  $A A^* = A^* A$ .
- On dit qu'une matrice  $A$  réelle symétrique est définie positive si pour tout  $u \in \mathbb{R}^n$ ,
  - $u \neq 0$ ,  $u^t A u > 0$ , la matrice  $A$  est dite semi-définie positive si pour tout  $u \in \mathbb{R}^n$ ,  $u^t A u \geq 0$ .

### Normes matricielles

**Définition 2.1.1** Une norme matricielle sur l'espace vectoriel  $M_n$  des matrices carrées d'ordre  $n$ , est une application de  $M_n$  vers  $\mathbb{R}_+$  qui vérifie pour tous  $A, B \in M_n$ ,  $\alpha \in \mathbb{R}$  (ou  $\mathbb{C}$ )

$$\begin{aligned} \|A\| &= 0 \Leftrightarrow A = 0 \\ \|\alpha A\| &= |\alpha| \|A\| \\ \|A + B\| &\leq \|A\| + \|B\| \\ \|A B\| &\leq \|A\| \|B\| \end{aligned}$$

**Définition 2.1.2** ( Norme matricielle subordonnée) Soit  $\|\cdot\|$  une norme vectorielle, on appelle norme matricielle subordonnée (à cette norme vectorielle)

l'application de  $M_n$  vers  $\mathbb{R}_+$  définie par

$$\|A\| = \sup_{x \in V, x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{x \in V, \|x\| \leq 1} \|Ax\| = \sup_{x \in V, \|x\|=1} \|Ax\|.$$

Les normes matricielles subordonnées suivantes sont couramment utilisées

$$\begin{aligned} \|A\|_\infty &= \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| && \text{"maximum sur les lignes"} \\ \|A\|_1 &= \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}| && \text{"maximum sur les colonnes"} \\ \|A\|_2 &= \sqrt{\rho(A^t A)} \end{aligned}$$

où  $\rho(A^t A)$  est le rayon spectral de la matrice symétrique semi-définie positive  $A^t A$ .

**Définition 2.1.3** On appelle valeur propre de  $A$ , toute racine  $\lambda_i \in \mathbb{C}$  du polynôme caractéristique  $p(\lambda) = \det(A - \lambda I_n)$ . La multiplicité algébrique  $m_i \in \mathbb{N}^*$  est l'ordre de multiplicité de  $\lambda_i$  en tant que racine de  $p$  (ie : multiplicité dans le polynôme caractéristique). Si on note  $r \leq n$  le nombre de racines distinctes du polynôme  $p$ , on a

$$p(\lambda) = \prod_{i=1}^r (\lambda - \lambda_i)^{m_i}, \quad \text{où } r \text{ nombre de racines distinctes de } p, (r \leq n).$$

-Un vecteur propre  $v_i \in \mathbb{C}^n \setminus \{0\}$ , associé à la valeur propre  $\lambda_i$ , vérifie

$$Av_i = \lambda_i v_i.$$

On a donc  $v_i \in \ker(A - \lambda_i I_n) \setminus \{0\}$  et on appelle multiplicité géométrique de  $\lambda_i$  l'entier naturel  $g_i = \dim(\ker(A - \lambda_i I_n))$ ,  $1 \leq i \leq r$ .

**Remarque 2.1.1** Une valeur propre est dite simple si sa multiplicité algébrique est 1. Sinon, c'est une valeur propre multiple.

**Définition 2.1.4** L'ensemble des valeurs propres de  $A$  est appelé le spectre de  $A$  et est noté  $\sigma(A)$  ou  $\text{spec}(A)$ .

$$\sigma(A) = \text{spec}(A) = \{\lambda \in \mathbb{C} / \det(A - \lambda I_n) = 0\} = \{\lambda_1, \lambda_2, \dots, \lambda_n\}.$$

**Définition 2.1.5** On appelle rayon spectral de la matrice  $A$  le maximum du module de ces valeurs propres. On le note  $\rho(A)$ .

$$\rho(A) = \max \{|\lambda| / \lambda \in \text{spec}(A)\} = \max_{1 \leq i \leq n} \{|\lambda_1|, |\lambda_2|, \dots, |\lambda_n|\}.$$

### Matrices diagonalisables, matrices semblables

La proposition suivante fait l'objet de rappel de quelques résultats reliant la multiplicité des valeurs propres à la réduction des matrices.

**Proposition 2.1.1** *Soit  $A$  une matrice carrée d'ordre  $n$ , on a les propriétés suivantes :*

1- *Les multiplicités vérifient les relations :*

$$\text{pour } 1 \leq i \leq r : g_i \leq m_i, \quad \sum_{i=1}^r m_i = n \quad \text{et} \quad \sum_{i=1}^r g_i \leq n.$$

2- *La matrice  $A$  est diagonalisable si et seulement si  $\sum_{i=1}^r g_i = n$ .*

3- *Si toutes les valeurs propres  $\lambda_i \in \mathbb{C}$ ,  $1 \leq i \leq n$ , sont simples (i.e.  $m_i = g_i = 1$  et  $r = n$ ), alors la matrice est diagonalisable (dans  $\mathbb{C}$ ).*

4- *La matrice  $A$  n'est pas diagonalisable si et seulement si il existe au moins une valeur propre  $\lambda_i$  telle que  $m_i > g_i \geq 1$ .*

Dans ce cas, on dit que la matrice respectivement la valeur propre est déficiente.

**Proposition 2.1.2** *Soit  $P$  une matrice carrée d'ordre  $n$  inversible. On dit que  $B = P^{-1}AP$  et  $A$  sont des matrices semblables. Alors  $A$  et  $B$  ont les mêmes valeurs propres et  $v$  est un vecteur propre de  $A$  si et seulement si  $P^{-1}v$  est un vecteur propre de  $B$ .*

### 2.1.2 Décomposition de Schur réelle

La décomposition de SCHUR réelle permet de traiter tous les cas de matrice réelle  $A$  diagonalisable ou déficiente. D'où l'objet de la proposition suivante :

**Proposition 2.1.3** (*Décomposition de SCHUR réelle*) Soit  $A$  une matrice carrée réelle, alors il existe une matrice réelle et orthogonale  $U$  et une matrice triangulaire supérieure par blocs  $T$  tels que  $U^t A U = T$ . Sur la diagonale de  $T$  se trouvent des blocs  $1 \times 1$  correspondant aux valeurs propres réelles de  $A$  et des blocs  $2 \times 2$  correspondant aux valeurs propres complexes conjuguées de  $A$ .

### Exemple 2.1.1

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

$$= UTU^t.$$

Ici  $\text{spec}(A) = \{\lambda_1 = 1, \lambda_2 = 1 + i, \lambda_3 = 1 - i\}$ , donc  $r = 3$

et  $m_2 = 2 > g_1 = 1$ ,  $m_2 = m_3 = g_2 = g_3 = 1$ . La matrice est défective car  $\lambda_1$  est une valeur propre défective.

## 2.2 Méthode de la puissance

la méthode de la puissance est la méthode la plus simple pour le calcul des valeurs propres d'une matrice  $A$ . Cette méthode permet de fournir une approximation de la valeur propre du plus grand module (valeur propre dominante) et un vecteur propre associé. On obtient donc le rayon spectral d'une matrice.

### 2.2.1 Principe de la méthode et algorithme

La méthode des puissances appelée aussi méthode de la puissance itérée part du principe suivant : en appliquant un grand nombre de fois la matrice sur un vecteur arbitraire  $x^{(0)}$  normalisé, on obtient des vecteurs successifs qui se rapproche de la direction du vecteur propre associé à la plus grande valeur propre en module.

Soit  $A$  une matrice carrée diagonalisable, on suppose que ces valeurs propres sont ordonnées comme suit :  $|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$  et  $v_1, v_2, \dots, v_n$  les vecteurs propres associés.

La méthode consiste à calculer les suites des itérées définies par

$$y^{(k)} = Ax^{(k-1)}, \quad x^{(k)} = \frac{y^{(k)}}{\|y^{(k)}\|_2} \quad \text{et} \quad \tilde{\lambda}^{(k)} = (x^{(k)})^t Ax^{(k)}$$

Analysons les propriétés de la suite  $(x^{(k)})_{k \in \mathbb{N}}$ . Une simple récurrence sur l'indice  $k$  donne

$$\forall k \in \mathbb{N}, \quad x^{(k)} = \frac{A^k x^{(0)}}{\|A^k x^{(0)}\|_2} \quad (2.1)$$

et l'on voit la présence des puissances de la matrice  $A$  ce qui justifie le nom de cette méthode. l'ensemble  $\{v_1, \dots, v_n\}$  de vecteurs propres de  $A$  forment une base, alors le vecteur  $x^{(0)}$  peut se décomposer de la façon suivante :

$$x^{(0)} = \sum_{i=1}^n \alpha_i v_i$$

notons que  $Av_i = \lambda_i v_i \Rightarrow A^k v_i = \lambda_i^k v_i$  et l'on obtient, en supposant le coefficient  $\alpha_1 \neq 0$  et pour tout  $k \in \mathbb{N}$

$$A^k x^{(0)} = \sum_{i=1}^n \alpha_i (A^k v_i) \quad (2.2)$$

$$= \sum_{i=1}^n \alpha_i \lambda_i^k v_i \quad (2.1)$$

$$= \lambda_1^k \left( \alpha_1 v_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \dots + \alpha_n \left(\frac{\lambda_n}{\lambda_1}\right)^k v_n \right) \quad (2.2)$$

$$= \alpha_1 \lambda_1^k \left( v_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1}\right)^k v_i \right).$$

Comme  $|\lambda_1| > |\lambda_i|$ ,  $\forall i \neq 1$ , on a  $\left|\frac{\lambda_i}{\lambda_1}\right| < 1$  pour  $2 \leq i \leq n$ , la composante le long de  $v_1$  du vecteur  $x^{(k)}$  augmente en module avec l'entier  $k$ , par contre les composantes suivant les autres directions  $v_i, i = 2, \dots, n$ , diminuent. Les vecteurs  $\{v_i\}_{i=1, \dots, n}$  sont tels que  $\|v_i\|_2 = 1$ , il s'ensuit que

$$\left\| \sum_{i=2}^n \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1}\right)^k v_i \right\|_2 \leq \sum_{i=2}^n \left| \frac{\alpha_i}{\alpha_1} \right| \left| \frac{\lambda_i}{\lambda_1} \right|^k \leq \left| \frac{\lambda_2}{\lambda_1} \right|^k \sum_{i=2}^n \left| \frac{\alpha_i}{\alpha_1} \right| = C \left| \frac{\lambda_2}{\lambda_1} \right|^k, \quad (2.3)$$

en l'on déduit de (2.1), de la dernière égalité de (2.2) et de (2.3) que

$$\forall k \in \mathbb{N}, \quad x^{(k)} = \frac{\alpha_1 \lambda_1^k (v_1 + r_k)}{\|\alpha_1 \lambda_1^k (v_1 + r_k)\|_2}$$

où la suite de vecteurs  $(r_k)_{k \geq 2}$  a pour limite le vecteur nul quand  $k \rightarrow +\infty$ . Par conséquent, le vecteur  $x^{(k)}$  devient peu à peu colinéaire avec le vecteur propre  $v_1$  associé à la valeur propre dominante  $\lambda_1$  quand  $k \rightarrow +\infty$  d'autant plus rapidement que  $\left| \frac{\lambda_2}{\lambda_1} \right|$  est petit. La suite  $\forall k \in \mathbb{N}, (x^{(k)})^t A x^{(k)} = \lambda^{(k)}$ , converge donc vers la valeur propre  $\lambda_1$ .

Nous présentons maintenant, l'algorithme de la méthode de la puissance

ALGORITHME 1

$x^{(0)} = (x_1^{(0)}, \dots, x_n^{(0)}) \in \mathbb{R}^n$  vecteur arbitraire donné,  $k = 0$

**répéter**

$$y^{(k+1)} = Ax^{(k)}$$

$$x^{(k+1)} = \frac{y^{(k+1)}}{\|y^{(k+1)}\|_2}$$

$$\tilde{\lambda}^{(k+1)} = (x^{(k)})^t A x^{(k)}$$

$$k = k + 1$$

**jusqu'à** [test d'arrêt].

Le test d'arrêt est souvent basé sur  $|\tilde{\lambda}^{(k+1)} - \tilde{\lambda}^{(k)}| \leq \epsilon$ , c'est-à-dire que l'itération s'arrête dès que la différence entre deux estimations de la valeur propre est suffisamment petite.

## 2.3 Méthode de la puissance inverse

Cette méthode permet d'approcher la valeur propre de la matrice  $A$  la plus proche d'un nombre  $\mu$  donné n'appartenant pas à son spectre.

On applique la méthode de la puissance à la matrice  $(A - \mu I_n)^{-1}$  dont les vecteurs propres sont  $\{v_1, \dots, v_n\}$  ceux de la matrice  $A$  et les valeurs propres sont  $(\lambda_i - \mu)^{-1}$ ,  $1 \leq i \leq n$ .

Il suffit d'adapter l'algorithme1 pour obtenir :

ALGORITHME 2

Méthode de la puissance inverse

$x^{(0)} \in \mathbb{R}^n$  donné,  $k = 0$

**répéter**

$$y^{(k+1)} = (A - \mu I_n)^{-1} x^{(k)}$$

$$x^{(k+1)} = \frac{y^{(k+1)}}{\|y^{(k+1)}\|}$$

$$\tilde{\lambda}^{(k+1)} = (x^{(k)})^t A x^{(k)}$$

$k = k + 1$

**jusqu'à** [test d'arrêt]

**Remarque**

On calcule une décomposition  $LU$  de la matrice  $A - \mu I_n$  afin de résoudre le système

$$(A - \mu I_n) y^{(k+1)} = x^{(k)}.$$

**Exemple 2.3.1** Soit  $A = \begin{pmatrix} -3 & 0 & 0 \\ 17 & 13 & -7 \\ 16 & 14 & -8 \end{pmatrix}$ , avec  $\text{spec}(A) = \{\lambda_1 = 6, \lambda_2 = -3, \lambda_3 = -1\}$

. En appliquant la méthode de la puissance avec

$$x^{(0)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{on obtient} \quad x^{(1)} = \begin{pmatrix} -0.0938417 \\ 0.7194529 \\ 0.6881724 \end{pmatrix}, \quad x^{(10)} = \begin{pmatrix} 0.0002303 \\ 0.7070492 \\ 0.7071643 \end{pmatrix}$$

et  $\tilde{\lambda}^{(1)} = 4.1986301$ ,  $\tilde{\lambda}^{(10)} = 6.0036631$ .

Si l'on applique la méthode de la puissance inverse, avec  $\mu = 0$ , et en partant de

$$x^{(0)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{on obtient} \quad x^{(1)} = \begin{pmatrix} -0.13938466 \\ 0.6270597 \\ 0.7664063 \end{pmatrix}, \quad x^{(10)} = \begin{pmatrix} 0.0000151 \\ 0.4472257 \\ -0.8944211 \end{pmatrix}$$

et  $\tilde{\lambda}^{(1)} = 0.5242718$ ,  $\tilde{\lambda}^{(10)} = -1.0000478$ .

## 2.4 Méthode de Jacobi pour matrices réelles symétriques

Soit  $A = (a_{ij})_{1 \leq i, j \leq n}$  une matrice réelle carrée symétrique. La méthode de Jacobi consiste à construire une suite de matrice  $A = A_0, A_1, \dots, A_k, \dots$ , qui converge vers une matrice diagonale  $D$ . Avec

$$A_{k+1} = J_k^t A_k J_k,$$

où  $J_k$  est une rotation de Jacobi (voir la définition ci-après) et  $J_k^t J_k = I_n$ .

Si l'on pose  $J = J_0 J_1 \dots J_{k-1}$ , on aura une approximation de  $D \simeq J^t A J$ , respectivement une approximation de  $A \simeq J D J^t$ . Comme  $J$  est encore une matrice orthogonale,  $D$  contient une approximation de toutes les valeurs propres de  $A$  et les colonnes de  $J$  forment une base orthonormée de vecteurs propres.

Une rotation de Jacobi est définie grâce à une matrice de Givens : Les matrices de rotation de Givens sont des matrices orthogonales qui permettent, d'annuler certains coefficients d'un vecteur ou d'une matrice. Pour un couple d'indices  $i$  et  $j$  vérifiant  $1 \leq i < j \leq n$ , et un nombre  $\theta$  donnés, on définit la matrice de Givens comme suit :

$$G(i, j, \theta) = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & & & & & & \vdots \\ \vdots & & \ddots & & & & & 0 \\ 0 & & & c & & -s & & \vdots \\ \vdots & & & & \ddots & & & 0 \\ 0 & & & s & & c & & \vdots \\ \vdots & & & & & & 1 & 0 \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 & 1 \end{pmatrix}$$

avec  $c = \cos \theta$  et  $s = \sin \theta$ . Pour  $x \in \mathbb{R}^n$ ,  $G(i, j, \theta) x$  représente une rotation d'angle  $\theta$  dans le plan  $\Pi(e_i, e_j)$ .

Le produit  $G(i, j, \theta)^t A$  n'affecte que les  $i^e$  et  $j^e$  lignes de  $A$ .

Le  $AG(i, j, \theta)A$  n'affecte que les  $i^e$  et  $j^e$  colonnes de  $A$ .

La matrice de Givens est construite à partir d'une valeur de  $\theta$  qui permet d'annuler les éléments  $a_{ij}$  et  $a_{ji}$ .

L'angle  $\theta$  vérifie

$$\cot g(2\theta) = \frac{a_{ii} - a_{jj}}{2a_{ij}} = \tau$$

De cette relation on peut déduire les valeurs de  $c$  et  $s$  grâce à l'équation

$$t = \frac{s}{c} = tg \theta, \quad t^2 + 2\tau t - 1 = 0.$$

L'algorithme 5 suivant remplace la matrice  $A$  par une matrice  $\tilde{A}$  telle que  $\tilde{a}_{ij} = \tilde{a}_{ji} = 0$ .

ALGORITHME 3

Rotation de Jacobi pour annuler les éléments de coordonnées  $\{i, j\}$

**Si**  $|a_{ij}| > \epsilon$  **alors**

$$\tau = \frac{a_{ii} - a_{jj}}{2a_{ij}}$$

$$t = \frac{\text{signe}(\tau)}{|\tau| + \sqrt{1 + \tau^2}}$$

$$c = \frac{1}{\sqrt{1 + t^2}} \quad \text{et} \quad s = cte$$

$$\tilde{A} = G(i, j, \theta)^t A G(i, j, \theta).$$

L'algorithme 4 qui suit donne la convergence vers une matrice diagonale contenant les valeurs propres, et une matrice  $J$  contenant les vecteurs propres de  $A$ .

ALGORITHME 4

Algorithme de Jacobi pour valeurs et vecteurs propres d'une matrice symétrique

Seuil  $\sigma > 0$  donné,  $J = I_n$ ,  $k = 0$

**répéter**

Choisir  $\{i, j\}$  et en déduire  $J_k = G(i, j, \theta)$  par l'algorithme 3

$$J = J J_k$$

$$k = k + 1$$

**jusqu'à**  $\max_{1 \leq k < l \leq n} |a_{kl}| < \sigma$ .

**Remarque 2.4.1** Dans l'algorithme original de Jacobi, on choisit les coordonnées  $\{i, j\}$  de façon à avoir

$$|a_{ij}| = \max_{1 \leq k < l \leq n} |a_{kl}|.$$

Ceci permet d'avoir une bonne vitesse de convergence. Cependant, pour  $n$  assez grand, la recherche du maximum parmi les  $\frac{n^2-n}{2}$  éléments en dehors de la diagonale prends du temps. En pratique on parcourt  $A$  simplement ligne par ligne par exemple .

**Exemple 2.4.1** Soit  $A = \begin{pmatrix} 2 & 1 & -3 \\ 1 & -6 & 3 \\ -3 & 3 & 0 \end{pmatrix}$  et l'on fixe  $\sigma = 10^{-5}$ .

On a

$$\text{spec}(A) = \{-7.7155084, -0.5472851, 4.2627936\}$$

$$\text{Pour } J_0 = \begin{pmatrix} 0.8112422 & 0 & 0.5847103 \\ 0 & 1 & 0 \\ -0.5847103 & 0 & 0.8112422 \end{pmatrix},$$

on annule  $a_{13} = a_{31} = -3$ .

$$A_1 = J_0^t A J_0 = \begin{pmatrix} 4.1622777 & -0.9428887 & 2.2210^{-16} \\ -0.9428887 & -6 & 3.0184368 \\ 2.2210^{-16} & 3.0184368 & -2.1622777 \end{pmatrix}.$$

$$\text{Après 5 itérations, on a } J_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.9999904 & -0.0043757 \\ 0 & 0.0043757 & 0.9999904 \end{pmatrix} \text{ et}$$

$$A_5 = J_4^t A_4 J_4 = \begin{pmatrix} 4.2627928 & 0.0029793 & -0.0000130 \\ 0.0029793 & -7.7155077 & 6.94 \cdot 10^{-18} \\ -0.0000130 & 1.2210^{-16} & -0.5472851 \end{pmatrix}.$$

$$\text{En plus } J = J_1 J_2 J_3 J_4 = \begin{pmatrix} 0.7766476 & -0.2219207 & 0.5895504 \\ -0.1060891 & 0.8764440 & 0.4696712 \\ -0.6209376 & -0.4273139 & 0.6571448 \end{pmatrix}.$$

## 2.5 Exercices

**Exercice1** Soit la matrice :

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Déterminer ses valeurs et vecteurs propres

- 1) Par la méthode de la puissance itérée
- 2) Par la méthode de Jacobi.

**Exercice2** Soit la matrice

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Déterminer sa plus grande valeurs propre ainsi que le vecteur propre associé par la méthode de la puissance itérée.

### 2.5.1 Réponses

**Exercice1**

- 1) Résumons un peu la méthode :

En pratique, on se donne un vecteur initial  $x^{(0)}$  (de norme égale à 1 ou quelconque),

et on calcule les  $x^{(k)}$  par la relation  $x^{(k+1)} = A x^{(k)}$ .

De plus, on préfère ramener à 1 la valeur de l'une des composantes de ces vecteurs ( par exemple la 1<sup>ère</sup>).

La composante correspondante du vecteur itéré suivant donne le rapport à la valeur précédente et tend vers la valeur propre de plus grand module.

A présent on veut trouver les valeurs propres suivantes, on applique la méthode de déflation qui consiste à créer la matrice

$$A_1 = A - \lambda_1 \frac{v_1 (x^{(1)})^T}{(x^{(1)})^T v_1}$$

où le vecteur  $x^{(1)}$  est le vecteur propre de la matrice transposée de  $A$ , correspondant à la valeur propre  $\lambda_1$ .

On se donne un vecteur initial  $x^{(0)} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$  à partir duquel on construit la

suite de vecteurs :

$x^{(1)} = A x^{(0)}$ ,  $x^{(2)} = A x^{(1)}$ , ..., reporté dans le tableau ci-dessous

$\lambda$	1	5	$\frac{13}{5}$	$\frac{14}{13}$	$\frac{121}{41}$	$\frac{365}{121}$	...
$x^{(k)}$	1	1	1	1	1	1	
	0	2	$\frac{4}{5}$	$\frac{14}{13}$	$\frac{40}{41}$	$\frac{122}{121}$	$\frac{364}{365}$ .....
	0	0	0	0	0	0	

Les composantes de chaque vecteur de ce tableau ont été divisées par la première composante correspondant à chaque vecteur.

La suite des vecteurs  $x^{(k)}$  tend vers le vecteur  $v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$  et que la suite des

valeurs de  $\lambda$  converge vers la valeur  $\lambda_1 = 3$ .

La matrice  $A$  considérée étant symétrique, le vecteur propre de sa transposée, correspondant à la valeur propre  $\lambda_1$  sera  $x^{(1)} = v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ .

Construisons la matrice  $A_1$  qui est telle que  $A_1 = A - \lambda_1 \frac{v_1(x^{(1)})^T}{(x^{(1)})^T v_1} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & -1 \end{bmatrix}$ .

On se donne à nouveau un vecteur  $x^{(0)} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$  et on procède de la même

manière qu'avec la matrice  $A$  :

$\lambda$	$\frac{-1}{2}$	-1	-1	....
$x^{(k)}$	1	1	1	1
	0	-1	-1	....
	0	0	0	

On voit que cette suite de vecteurs converge vers le vecteur propre :  $v_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

et que la valeur propre qui lui correspond est  $\lambda_2 = -1$ .

On recommence le processus en définissons une matrice  $A_2$  à partir de  $A$ ,  $x^{(2)} = v_2$  :

$$A_2 = A_1 - \lambda_2 \frac{v_2 (x^{(2)})^T}{(x^{(2)})^T v_2} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Il est clair que la valeur propre de cette matrice est  $\lambda_3 = -1$  et que le vecteur

propre correspondant est  $v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

2) Le plus grand élément non diagonal de la matrice  $A$  étant l'élément  $a_{12}$ , l'angle  $\theta$  sera tel que :

$$\operatorname{tg}(2\theta) = \frac{1}{\cot g(2\theta)} = \frac{2a_{12}}{a_{11} - a_{22}}.$$

Comme  $a_{11} = a_{22}$ ,  $\theta = \frac{\pi}{4}$ , la matrice de transformation  $G_{12}$  est

$$G_{12} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

d'où

$$A_1 = G_{12} A G_{12} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

matrice diagonale. Les valeurs propres de  $A_1$ , et par conséquent de  $A$  sont les éléments de la diagonale :  $\lambda_1 = 3$  et  $\lambda_2 = \lambda_3 = -1$ .

Quand aux vecteurs propres, puisque le calcul s'est opéré en une seule étape, ils sont données directement par les colonnes de la matrice  $G$ .

On a donc :

$$V_1 = G_{12} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad V_2 = G_{12} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

et

$$V_3 = G_{12} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

**Exercice2**

En partant du vecteur initial  $x^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ , l'algorithme de la puissance  $x^{(k)} = Ax^{(k-1)}$  fournit les valeurs suivantes

$k$	0	1	2	3	4	5	6	7
$x^{(k)}$	1	2	5	14	41	122	365	1094
	0	-1	-4	-13	-40	-121	-364	-1093
$\lambda$					2.928	2.975	2.991	2.9973

Le rapport de 2 composantes homologues successives  $\left(\frac{x_i^{(k)}}{x_i^{(k-1)}}\right)$  est donné dans la dernière ligne, dans la dernière colonne,

il converge vers la valeur 3, qui serait la plus grande des valeurs propres.

Le vecteur  $x^{(k)}$  semble tendre vers une limite proportionnelle au vecteur  $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$  qui serait le vecteur propre  $v_1$  associé à la valeur propre  $\lambda_1 = 3$ .

**Remarque :** Comme il est expliqué dans l'algorithme, il vaut mieux normaliser  $x^{(k)} = \frac{x^{(k)}}{\|x^{(k)}\|}$  dès qu'il est calculé avec l'inconvénient que le calcul est très lourd à la main.

Je vous donne des résultats calculés à l'aide de Scilab

$k$	0	1	2	...	5	6
$v^{(k)}$	1	0.8944272	0.7808688	...	0.7100107	0.7080761
	0	-0.4472136	-0.6246950	...	-0.7041909	-0.7061361

Ainsi la valeur approchée de la valeur propre est obtenue par la formule

$$\lambda \simeq v^{(6)T} A v^{(6)} = 2.9999962.$$

Cette approximation est plus précise que celle calculée précédemment.

# Chapitre 3

## Interpolation numérique

### 3.1 Généralités

#### 3.1.1 Position du problème

Soit une fonction  $f$  définie sur un intervalle  $[a, b]$  de  $\mathbb{R}$  et dont la valeur numérique n'est connue qu'en  $(n + 1)$  points de  $[a, b]$ ; soit  $\{x_0, x_1, \dots, x_n\}$  cet ensemble de points tels que  $x_i \in [a, b]$ ,  $\forall i \in \{0, \dots, n\}$  et  $f(x_i)$  la valeur de la fonction  $f$  en un point  $x_i$ .

**Principe** L'interpolation est une méthode qui permet d'obtenir une valeur approchée  $\phi(x)$  de la fonction  $f(x)$  en un point  $x \in [a, b]$  et ceci de telle sorte que la condition suivante soit vérifiée :

$$\phi(x_i) = f(x_i) \quad \forall i \in \{0, \dots, n\}.$$

La fonction  $\phi$  ainsi déterminée est appelée fonction d'interpolation de la fonction  $f$  sur le maillage  $\{x_0, x_1, \dots, x_n\}$ . En principe la fonction d'interpolation est une combinaison linéaire de fonctions simples

- polynômes
- fonctions circulaires
- exponentielles
- fractions rationnelles

**Remarque 3.1.1** *On peut trouver un problème d'interpolation ayant un caractère plus général : En  $(n + 1)$  points  $x_i \in [a, b]$  on connaît les valeurs :  $f(x_i)$ ,  $f'(x_i)$ ,  $\dots$ ,  $f^{(\nu_i)}(x_i)$  la fonction d'interpolation  $\phi$  est alors telle que :*

$\phi(x_i) = f(x_i)$ ,  $\phi'(x_i) = f'(x_i)$ , ...,  $\phi^{(\nu_i)}(x_i) = f^{(\nu_i)}(x_i)$  c'est le problème de l'interpolation d'Hermite.

## 3.2 Interpolation polynômiale

### 3.2.1 Problème

Déterminer un polynôme  $P_n(x)$  de degré  $\leq n$  tel que :

$$P_n(x_i) = f(x_i), \quad \forall i \in \{0, \dots, n\}.$$

Question : Un tel polynôme existe-t-il ? Si oui, est-il unique ?

#### Solution théorique du problème

On choisit comme fonction de base :

$$\phi_0(x) = 1, \quad \phi_1(x) = x, \dots, \phi_n(x) = x^n.$$

La fonction d'interpolation devient alors le polynôme d'interpolation  $P_n(x)$  de degré  $\leq n$  tel que :  $P_n(x) = a_0 + a_1x + \dots + a_nx^n$  où  $a_0, \dots, a_n$  sont des constantes à déterminer. A partir de la condition

$$P_n(x_i) = f(x_i), \quad i = \overline{0, n},$$

les coefficients  $a_0, \dots, a_n$  sont solution du système linéaire :

$$\sum_{k=0}^n a_k x_i^k = f(x_i), \quad \forall i \in \{0, \dots, n\} \quad (S)$$

On écrit le système (S) sous la forme matricielle suivante

$$\underbrace{\begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix}}_A \underbrace{\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}}_X = \underbrace{\begin{pmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}}_B$$

Pour prouver que le système  $AX = B$  possède une solution unique  $X$  il suffit d'établir que le déterminant de la matrice  $A$  est non nul.

**Lemme 3.2.1** Soit  $x_0, x_1, \dots, x_n$  ( $n + 1$ ) points distincts alors

$$\Delta_n = \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix} = \prod_{\substack{i>j \\ j=0}}^n (x_i - x_j) = \prod_{j=0}^{n-1} \left[ \prod_{i=j+1}^n (x_i - x_j) \right]$$

$\Delta_n$  est appelé le déterminant de Vandermonde.

**Remarque 3.2.1** Notons que pour  $n = 1$  on a

$$\Delta_1 = \begin{vmatrix} 1 & x_0 \\ 1 & x_1 \end{vmatrix} = \prod_{\substack{i>j \\ j=0}}^1 (x_i - x_j) = (x_1 - x_0).$$

Par un raisonnement par récurrence on peut montrer le lemme i.e le déterminant ( du type VanderMonde) du système (S) est :

$$\Delta_n = \prod_{\substack{i>j \\ j=0}}^n (x_i - x_j)$$

et est par conséquent non nul si les points  $x_i$  sont tous distincts.

**Conclusion**  $\Delta_n \neq 0$  car les points  $x_i; i = \overline{0, n}$ , sont distincts. Donc le système (S) admet une solution unique  $(a_0 \ a_1 \ \dots \ a_n)^t$ . Ainsi le polynôme d'interpolation  $P_n$  de degré  $n$  de la fonction  $f$  existe et est unique si les points d'interpolation sont tous distincts i.e :  $\exists! P_n$  tel que  $P_n(x_i) = f(x_i), \forall i = \overline{0, n}$ .

**Exemple 3.2.1** Calculons le polynôme passant par les points  $(0, 1), (1, 2), (2, 9)$  et  $(3, 28)$ . Etant donné ces 4 points le polynôme recherché est tout au plus de degré 3. Ces coefficients  $a_i$  sont solution de

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 9 \\ 28 \end{pmatrix}$$

dont la solution (obtenue par décomposition LU) est  $(1 \ 0 \ 0 \ 1)^t$ . Le polynôme recherché est donc  $P_3(x) = 1 + x^3$ .

**Remarque 3.2.2** *En général le système linéaire permettant de déterminer les coefficients  $a_k$  est mal conditionné. Il n'est donc pas indiqué de calculer un polynôme d'interpolation  $P_n$  en résolvant un système linéaire. Cette méthode est donc rarement utilisée. Il est par contre classique de l'exprimer sous les formes suivantes :*

- polynôme d'interpolation de LAGRANGE
- polynôme d'interpolation de NEWTON

### 3.2.2 Interpolation de LAGRANGE

**Définition 3.2.1** *Soient  $x_0, x_1, \dots, x_n$  ( $n+1$ ) points distincts. Le polynôme de LAGRANGE au points  $x_i$  est un polynôme de degré  $n$  noté  $L_i(x)$  et vérifiant la condition suivante :*

$$\forall i \in \overline{0, n} : L_i(x_j) = \begin{cases} 0 & ; i \neq j \\ 1 & ; i = j \end{cases}$$

Cela signifie que le polynôme  $L_i(x)$  prend la valeur 1 en  $x_i$  et s'annule en tous les autres points de collocation.

**Théorème 3.2.1** *Pour tout  $i \in \overline{0, n}$  on a*

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

**Preuve.** On a par définition

$$L_i(x_j) = 0, \quad \forall i \neq j.$$

Par conséquent, pour tout  $i \in \overline{0, n}$  les racines de  $L_i(x)$  sont :  $x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ . Le polynôme  $L_i(x)$  s'écrit donc comme suit :

$$L_i(x) = K(x - x_0) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n) = k \prod_{\substack{j=0 \\ j \neq i}}^n (x - x_j),$$

où  $k$  est une constante à déterminer.

D'autre part  $L_i(x_i) = 1$ , d'où  $k \prod_{\substack{i \neq j \\ j=0}}^n (x_i - x_j) = 1$

$$\Rightarrow k = \frac{1}{\prod_{\substack{i \neq j \\ j=0}}^n (x_i - x_j)}. \text{ Finalement, on trouve } L_i(x)_{0 \leq i \leq n} = \frac{\prod_{\substack{i \neq j \\ j=0}}^n (x - x_j)}{\prod_{\substack{i \neq j \\ j=0}}^n (x_i - x_j)} = \prod_{\substack{i \neq j \\ j=0}}^n \frac{(x - x_j)}{(x_i - x_j)}.$$

■

À titre d'illustration, on représente sur la figure (3.1) les graphes sur l'intervalle  $[-1, 1]$  des polynômes de LAGRANGE qu'on note  $l_i$ ,  $i = 0, 4$  associés aux noeuds  $-1$ ,  $-\frac{1}{2}$ ,  $0$ ,  $\frac{1}{2}$  et  $1$ .

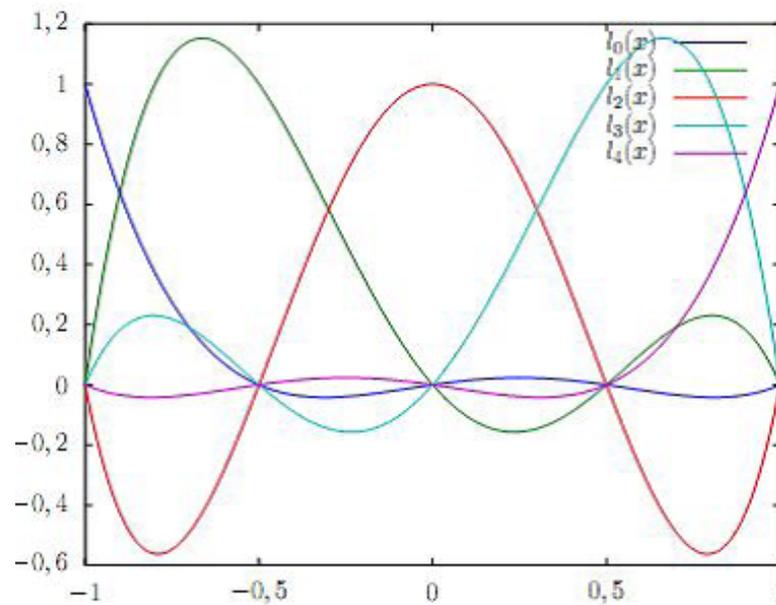


FIG.3.1-

Revenons à la résolution du problème  $P_n(x_i) = f(x_i)$  qui consiste à construire  $P_n(x)$  vérifiant les conditions ci-dessus. Ce polynôme est de la forme

$$P_n(x) = \sum_{i=0}^n f(x_i) L_i(x).$$

$P_n(x)$  est bien de degré inférieur ou égal à  $n$ . et

$$P_n(x_j) = \sum_{i=0}^n f(x_i) L_i(x_j) = f(x_j) L_j(x_j) = f(x_j); \quad j = 0, 1, \dots, n.$$

Ainsi, Le polynôme d'interpolation de LAGRANGE est défini par

$$P_n(x) = \sum_{i=0}^n f(x_i) \prod_{\substack{j=0 \\ j \neq i}}^n \frac{(x - x_j)}{(x_i - x_j)}.$$

**Exemple 3.2.2** On cherche à construire le polynôme d'interpolation de LAGRANGE qui en  $-1$  vaut 8, en 0 vaut 3 et en 1 vaut 6.

On a

$$\begin{aligned} P_2(x) &= f(x_0) \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + f(x_1) \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + f(x_2) \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} \\ &= 8 \frac{x(x - 1)}{2} + 3 \frac{(x + 1)(x - 1)}{-1} + 6 \frac{(x + 1)x}{2} = 4x^2 - x + 3. \end{aligned}$$

### Erreur d'interpolation

On admettra le théorème suivant sans démonstration

**Théorème 3.2.2** Soient  $x_0, x_1, \dots, x_n$ ,  $(n + 1)$  points distincts de  $[a, b]$

et  $f \in C^{n+1}([a, b])$ . Alors, pour chaque point  $x \in [a, b]$ , il existe  $\alpha \in [a, b]$  tel que :

$$f(x) - P_n(x) = e_n(x) = \frac{f^{(n+1)}(\alpha)}{(n+1)!} \prod_{j=0}^n (x - x_j), \text{ où } P_n(x) \text{ est le polynôme d'interpolation de } f.$$

Posons  $M_{n+1} = \max_{x \in [a, b]} |f^{(n+1)}(x)|$  alors

$$|e_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \prod_{j=0}^n (x - x_j); \quad x \in [a, b].$$

$e_n(x)$  est appelée erreur d'interpolation au point  $x$ .

**Exemple 3.2.3** Déterminons le polynôme d'interpolation de Lagrange de la fonction

$f(x) = \sqrt{x}$  sur les points  $x_0 = 100$ ,  $x_1 = 121$  et  $x_2 = 144$ .

Avec quelle précision peut-on calculer  $\sqrt{115}$  ?

Calculons  $P_2(x)$  :

$$\begin{aligned} P_2(x) &= f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x) \\ &= 10L_0(x) + 11L_1(x) + 12L_2(x) \end{aligned}$$

où

$$\begin{aligned} L_0(x) &= \frac{(x-121)(x-144)}{(100-121)(100-144)} = \frac{132}{7} - \frac{265}{924}x + \frac{1}{924}x^2 \\ L_1(x) &= \frac{(x-100)(x-144)}{(121-100)(121-144)} = \frac{-4800}{161} + \frac{244}{483}x - \frac{1}{483}x^2 \\ L_2(x) &= \frac{(x-100)(x-121)}{(144-100)(144-121)} = \frac{275}{23} - \frac{221}{1012}x + \frac{1}{1012}x^2 \end{aligned}$$

Donc  $P_2(x) = \frac{660}{161} + \frac{727}{10626}x - \frac{1}{10626}x^2$

$|e_2(115)| \leq \frac{M_3}{3!} |(115-100)(115-121)(115-144)|$  où

$$M_3 = \max_{[100,144]} |f'''(x)| = \frac{3}{8}10^{-5}$$

(avec  $f'''(x) = \frac{3}{8}x^{-\frac{5}{2}}$ ) d'où  $|e_2(115)| \leq \frac{3}{8 \cdot 3!}10^{-5} (2610) \approx 0,163 \times 10^{-2}$ .

### 3.2.3 Interpolation dans la base de Newton

#### Les différences divisées de Newton

**Définition 3.2.2** Soit  $f$  une fonction définie sur  $[a, b]$  et  $x_0, x_1, \dots, x_n$

$(n+1)$  points distincts de  $[a, b]$ . On définit les différences divisées de  $f$  aux points  $x_0, x_1, \dots, x_n$  par les relations de récurrence :

$$f[x_0] = f(x_0), \quad \forall i \in \overline{0, n} : f[x_i] = f(x_i)$$

$$f[x_0, x_1] = \frac{f[x_0] - f[x_1]}{x_0 - x_1} = \frac{f(x_0) - f(x_1)}{x_0 - x_1}$$

$$f[x_0, x_1, x_2] = \frac{f[x_0, x_1] - f[x_1, x_2]}{x_0 - x_2}$$

$$f[x_0, x_1, \dots, x_{n-1}, x_n] = \frac{\dots \dots \dots f[x_0, x_1, \dots, x_{n-1}] - f[x_1, x_2, \dots, x_n]}{x_0 - x_n},$$

$f[x_0, x_1, \dots, x_{n-1}, x_n]$  s'appelle la différence divisée d'ordre  $n$  de la fonction  $f$  aux points  $x_0, x_1, \dots, x_{n-1}, x_n$ .

### Construction de la table des différences divisées

Nous nous contentons de construire une table s'arrêtant aux quatrièmes différences divisées, les autres s'obtiendraient de la même manière.

$x_i$	$f(x_i)$	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$	$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}]$	$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}, x_{i+4}]$
$x_0$	$f[x_0]$				
$x_1$	$f[x_1]$	$f[x_0, x_1]$			
$x_2$	$f[x_2]$	$f[x_1, x_2]$	$f[x_0, x_1, x_2]$		
$x_3$	$f[x_3]$	$f[x_2, x_3]$	$f[x_1, x_2, x_3]$	$f[x_0, x_1, x_2, x_3]$	
$x_4$	$f[x_4]$	$f[x_3, x_4]$	$f[x_2, x_3, x_4]$	$f[x_1, x_2, x_3, x_4]$	$f[x_0, x_1, x_2, x_3, x_4]$

On verra dans ce qui suit que seules les valeurs encadrées, sur la diagonale principale de la table, interviennent dans l'expression du polynôme d'interpolation.

### Formule de Newton

**Théorème 3.2.3** Soit  $f : [a, b] \rightarrow \mathbb{R}$  fonction définie sur  $[a, b]$  et  $x_0, \dots, x_n$ ,  $n + 1$  points distincts de  $[a, b]$ . Le polynôme qui interpole  $f$  sur  $\{x_0, x_1, \dots, x_n\}$  est donné par :

$$\begin{aligned}
 P_n(x) &= f[x_0] + f[x_0, x_1](x - x_0) + \dots + f[x_0, x_1, \dots, x_{n-1}, x_n](x - x_0) \dots (x - x_{n-1}) \\
 &= f[x_0] + \sum_{i=1}^n f[x_0, x_1, \dots, x_i] \prod_{j=0}^{i-1} (x - x_j)
 \end{aligned} \tag{3.1}$$

Cette forme est appelée forme de Newton d'interpolation pour les différences divisées.

**Preuve.** Voir [12, Th 5.3] ■

**Exemple 3.2.4** Trouvons le polynôme d'interpolation de Newton pour la fonction  $f(x) = \sin x$  en les 3 points  $x_i = \frac{\pi}{2}i$  avec  $i = 0, \dots, 2$

$x_i$	$f(x_i)$	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$
0	0		
$\frac{\pi}{2}$	1	$\frac{2}{\pi}$	
$\pi$	0	$-\frac{2}{\pi}$	$-\frac{4}{\pi^2}$

$$\begin{aligned}
 P_2(x) &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) \\
 &= 0 + \frac{2}{\pi}(x - 0) - \frac{4}{\pi^2}(x - 0)\left(x - \frac{\pi}{2}\right) \\
 &= \frac{2}{\pi}x - \frac{4}{\pi^2}x\left(x - \frac{\pi}{2}\right) \\
 &= -\frac{4}{\pi^2}x(x - \pi).
 \end{aligned}$$

### Remarque

1- La formule (3.1) est la forme la plus utilisée pour calculer le polynôme d'interpolation, car elle nécessite le moins de calcul pour obtenir numériquement ses coefficients.

2- Dans la formule (3.1) posons  $x = x_n$ . On obtient alors :

$$P_n(x_n) = f[x_0] + f[x_0, x_1](x_n - x_0) + \dots + f[x_0, x_1, \dots, x_{n-1}, x_n](x_n - x_0) \dots (x_n - x_{n-1}).$$

Cette formule est connue sous le nom de " l'identité des différences divisées".

3- Considérons encore la forme (3.1) du polynôme d'interpolation. Le polynôme  $P_n(x)$  peut s'écrire aussi sous la forme suivante :

$$P_n(x) = a_0 + a_1x + \dots + a_nx^n. \quad (3.2)$$

Par identification des coefficients correspondants à (3.1) et (3.2), on obtient :

$$f[x_0, x_1, \dots, x_{n-1}, x_n] = a_n.$$

D'où la différence divisée d'ordre  $n$  d'un polynôme de degré  $n$  est constante. Par conséquent, la différence divisée d'ordre  $n + 1$  d'un polynôme de degré  $n$  est nulle.

4- Outre l'avantage d'être aisément calculable, l'algorithme de Newton peut être allongé ou réduit. Si l'on veut, par exemple, tenir compte de  $p$  points supplémentaires, on a juste à les ajouter à la suite de la table des différences divisées et de compléter les différences divisées. Si l'on veut au contraire, négliger les  $q$  points, il suffit d'arrêter la table des différences divisées aux  $(n + 1 - q)$  points.

**Exemple 3.2.5** Reprenons l'exemple précédent et calculons le polynôme d'interpolation de Newton de la même fonction en les 4 points  $x_i = \frac{\pi}{2}i$  avec  $i = 0, \dots, 3$  i.e on a juste ajouté le point  $x = \frac{3\pi}{2}$ . Il suffit de calculer une différence divisée en plus, i.e ajouter une ligne au tableau :

$x_i$	$f(x_i)$	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$	$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}]$
0	0			
$\frac{\pi}{2}$	1	$\frac{2}{\pi}$		
$\pi$	0	$-\frac{2}{\pi}$	$-\frac{4}{\pi^2}$	
$\frac{3\pi}{2}$	-1	$-\frac{2}{\pi}$	0	$\frac{8}{3\pi^2}$

$$\begin{aligned}
 P_3(x) &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) \\
 &\quad + f[x_0, x_1, x_2, x_3](x - x_0)(x - x_1)(x - x_2) \\
 &= P_2(x) + f[x_0, x_1, x_2, x_3](x - x_0)(x - x_1)(x - x_2) \\
 &= -\frac{4}{\pi^2}x(x - \pi) + \frac{8}{3\pi^2}x(x - \frac{\pi}{2})(x - \pi) \\
 &= \frac{8}{3\pi^3}x(x^2 - 3\pi x + 2\pi^2).
 \end{aligned}$$

### 3.3 Exercices

#### Exercice 1

Construire le polynôme d'interpolation de Lagrange de la fonction  $f(x) = \sin(\pi x)$  sur les points  $\{x_0 = 0, x_1 = \frac{1}{6}, x_2 = \frac{1}{2}\}$ . En déduire une valeur approchée de  $\sin \frac{\pi}{5}$  et calculer l'erreur d'interpolation au point  $x = \frac{1}{5}$ .

#### Exercice 2

Avec quelle précision peut-on calculer  $\sqrt{80}$  à l'aide de la formule de Lagrange pour la fonction  $f(x) = \sqrt{x}$  si on considère les points  $x_0 = 64, x_1 = 81,$  et  $x_2 = 100$ .

#### Exercice 3

On considère la fonction définie sur l'intervalle  $[0, 0.4]$  par la table de valeurs :

$x$	0	0.1	0.2	0.3	0.4
$f(x)$	0	0.097554	0.190428	0.278919	0.363304

1) Calculer à l'aide de la méthode d'interpolation de Lagrange, une valeur approchée de  $f(0.15)$ .

**Exercice 4**

1) Calculer les différences divisées relatives au tableau suivant

$i$	1	2	3	4
$x_i$	0	1	2	4
$y_i$	1	1	2	5

2) Déterminer le polynôme de Newton relatif au tableau de la question 1.

**Exercice 5**

On considère la fonction  $f(x) = \sqrt{x}$

1) Calculer les différences divisées de  $f$  liées aux points  $x_0 = 7$ ,  $x_1 = 9$ ,  $x_2 = 11$ ,  $x_3 = 13$ ,  $x_4 = 15$ .

2) Déterminer une approximation de  $f(8)$  à l'aide du polynôme de Newton.

3) Calculer l'erreur exacte en se basant respectivement sur le polynôme de Newton  $p_1(x)$ ,  $p_2(x)$  et  $p_3(x)$ . ensuite comparer les résultats avec l'estimation de l'erreur donnée par la formule de Newton suivante

$$(e_n(x) \simeq f[x_0, x_1, \dots, x_{n-1}, x_n, x_{n+1}](x - x_0)(x - x_1)\dots(x - x_n)).$$

**3.3.1 Réponses****Exercice 1**

$P_2(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x)$ , avec

$$L_0(x) = \frac{(x - \frac{1}{6})(x - \frac{1}{2})}{(0 - \frac{1}{6})(0 - \frac{1}{2})} = 1 - 8x + 12x^2,$$

$$L_1(x) = \frac{(x - 0)(x - \frac{1}{2})}{(\frac{1}{6} - 0)(\frac{1}{6} - \frac{1}{2})} = 9x - 18x^2,$$

$$L_2(x) = \frac{(x - 0)(x - \frac{1}{6})}{(\frac{1}{2} - 0)(\frac{1}{2} - \frac{1}{6})} = -x + 6x^2.$$

D'où,  $P_2(x) = \frac{7}{2}x - 3x^2$ ,  $x \in [0, \frac{1}{2}]$ .

La valeur approchée de  $\sin \frac{\pi}{5}$  :  $\sin \frac{\pi}{5} \simeq P_2(\frac{1}{5})$  (car  $\sin \frac{\pi}{5} = f(\frac{1}{5})$ ). par conséquent  $\sin \frac{\pi}{5} \simeq \frac{7}{10} - \frac{3}{25} = 0,58$ .

L'erreur d'interpolation au point  $x = \frac{1}{5}$  : on sait que

$$\left| e_2\left(\frac{1}{5}\right) \right| \leq \frac{M_3}{3!} \left| \left(\frac{1}{5} - x_0\right) \left(\frac{1}{5} - x_1\right) \left(\frac{1}{5} - x_2\right) \right|$$

où  $M_3 = \max_{x \in [0, \frac{1}{2}]} |f^{(3)}(x)| = \max_{x \in [0, \frac{1}{2}]} |-\pi^3 \cos \pi x| = \pi^3$ ,  $\left| e_2\left(\frac{1}{5}\right) \right| \leq \frac{\pi^3}{6} \frac{1}{500} = 1,0335 \cdot 10^{-2}$ .

**Exercice 2**

Pour qu'on puisse trouver la précision demandée, on utilise l'estimation de l'erreur de l'interpolation suivante

$$|f(x) - P_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \prod_{i=0}^n (x - x_i),$$

avec  $M_{n+1} = \max_{x \in [a,b]} |f^{(n+1)}(x)|$ .

Pour  $n = 2$  :  $a = 64$  et  $b = 100$  on a

$M_3 = \max_{x \in [64,100]} |f^{(3)}(x)| = f^{(3)}(64) = \frac{3}{8}8^{-5} = 3 \times 8^{-6}$  avec  $f^{(3)} = \frac{3}{8}x^{-\frac{5}{2}}$  qui est une fonction décroissante pour tout  $x > 0$ .

Par conséquent pour  $x = 80$  on a

$$\left| \sqrt{80} - P_2(80) \right| \leq \frac{8^{-6}}{2} |(100 - 80)(81 - 80)(64 - 80)| = \frac{160}{8^6} \simeq 0,00061 \leq 10^{-3}.$$

On peut alors calculer une valeur approchée de  $\sqrt{80}$  qui est  $P_2(80)$  avec une précision de  $10^{-3}$ .

**Exercice 4**

Indication En faisant un tableau de différences divisées comme vu au cours le polynôme d'interpolation de Newton s'écrit

$$\begin{aligned} P(x) &= f[x_0] + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2] \\ &\quad + (x - x_0)(x - x_1)(x - x_2)f[x_0, x_1, x_2, x_3] \\ &= 1 + (x - 0)0 + (x - 0)(x - 1)\frac{1}{2} + (x - 0)(x - 1)(x - 2)\left(\frac{-1}{12}\right) \\ &= 1 - \frac{2}{3}x + \frac{3}{4}x^2 - \frac{1}{12}x^3. \end{aligned}$$

# Chapitre 4

## Intégration numérique

### Principe

En général, il peut être difficile, voir impossible de trouver la valeur exacte d'une intégrale

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

Sauf dans le cas où on connaît explicitement une primitive de la fonction  $f$ , ou alors l'expression de  $f$  permettant par exemple une intégration par parties ou un changement de variables approprié.

L'idée consiste à remplacer  $f$  par une approximation  $\tilde{f}$  et calculer  $\int_a^b \tilde{f}(x) dx$

au lieu  $\int_a^b f(x) dx$  .i.e  $\int_a^b f(x) dx \simeq \int_a^b \tilde{f}(x) dx$  cette formule s'appelle formule de quadrature ou formule d'intégration numérique.

Si  $f$  est de classe  $C^0$  sur  $[a, b]$ , l'erreur de quadrature

$$E(f) = \left| \int_a^b f(x) dx - \int_a^b \tilde{f}(x) dx \right|$$

satisfait

$$E(f) = \left| \int_a^b f(x) - \tilde{f}(x) dx \right| \leq \int_a^b |f(x) - \tilde{f}(x)| dx \leq (b-a) \max_{x \in [a,b]} |f(x) - \tilde{f}(x)| .$$

L'approximation  $\tilde{f}$  doit être facilement intégrable, ce qui est le cas si, par exemple  $\tilde{f}$  est un polynôme.

Un choix naturel consiste à prendre  $\tilde{f}$  le polynôme d'interpolation de LAGRANGE de  $f$  sur un ensemble de  $n + 1$  noeuds distincts  $x_i$ ,  $i = \overline{0, n}$  on aura

$$I = \int_a^b f(x) dx \simeq \tilde{I} = \int_a^b \tilde{f}(x) dx = \sum_{i=0}^n \left( f(x_i) \int_a^b L_i(x) dx \right) \quad \text{où} \quad L_i(x) = \prod_{\substack{j=0 \\ i \neq j}}^n \frac{x - x_j}{x_i - x_j}$$

Il s'agit d'un cas particulier de la formule de quadrature suivante

$$\tilde{I} = \sum_{i=0}^n \alpha_i f(x_i)$$

qui est une somme pondérée des valeurs de  $f$  aux points  $x_i$  : on dit que ces points sont les noeuds de la formule de quadrature et que les nombres  $\alpha_i \in \mathbb{R}$  sont les coefficients ou encore les poids.

## 4.1 Méthode du rectangle

◆ **Formule de base** La formule du rectangle est obtenue en remplaçant  $f$  par une constante égale à la valeur de  $f$  en  $x_0 = a$ , polynôme qui interpole  $f$  en le point  $(a, f(a))$  et donc de degré 0, ce qui donne

$$I = \int_a^b f(x) dx \simeq \int_a^b \tilde{f}(x) dx = \int_a^b f(a) dx = (b - a) f(a)$$

C'est à dire  $I \simeq (b - a) f(a)$

◆ **Erreur** Si  $f \in C^1([a, b])$  alors il existe  $\delta \in ]a, b[$  tel que  $f(x) = f(a) + (x - a)f'(\delta)$  et l'erreur de quadrature est majorée par

$$\begin{aligned} |E| &= \left| \int_a^b f(x) - \tilde{f}(x) dx \right| = \left| \int_a^b f(x) - f(a) dx \right| \\ &= \left| \int_a^b (x - a) f'(\delta) dx \right| = \frac{(b - a)^2}{2} |f'(\delta)| \leq \frac{(b - a)^2}{2} M_1, \end{aligned}$$

avec  $M_1 = \max_{x \in [a, b]} |f'(x)|$ .

## 4.2 Méthode du trapèze

◆ **Formule de base** La formule du trapèze est obtenue en remplaçant  $f$  par le segment qui relie  $(a, f(a))$  à  $(b, f(b))$  polynôme qui interpole  $f$  en les points  $(a, f(a))$  et  $(b, f(b))$  et donc de degré 1, on obtient une fonction de classe  $C^2$  sur  $[a, b]$  et  $P_1(x)$  le polynôme qui interpole  $f$  aux points  $\{x_0 = a, x_1 = b\}$ . Alors

$$I = \int_a^b f(x) dx \simeq \int_a^b \tilde{f}(x) dx = \int_a^b \frac{f(b) - f(a)}{b - a} (x - a) + f(a) dx = \frac{b - a}{2} (f(a) + f(b)).$$

◆ **Erreur** Si  $f \in C^2([a, b])$  alors il existe  $\delta \in ]a, b[$  tel que

$f(x) - \tilde{f}(x) = \frac{f''(\delta)}{2}(x - a)(x - b)$  et l'erreur de quadrature est majorée par

$$\begin{aligned} |E| &= \left| \int_a^b f(x) - \tilde{f}(x) dx \right| = \left| \int_a^b \frac{f''(\delta)}{2}(x - a)(x - b) dx \right| \\ &\leq \frac{1}{2} |f''(\delta)| \int_a^b (x - a)(x - b) dx = \frac{1}{2} |f''(\delta)| \frac{(b - a)^3}{6} \\ &\leq \frac{(b - a)^3}{12} \max_{[a, b]} |f''(x)|. \end{aligned}$$

## 4.3 Méthode de Simpson

◆ **Formule de base** La formule de Simpson est obtenue en remplaçant  $f$  par la parabole qui interpole  $f$  en  $a$ , en  $b$  et en  $\frac{a+b}{2}$  donc un polynôme de degré 2, on a

$$\begin{aligned} I &= \int_a^b f(x) dx \simeq \int_a^b \tilde{f}(x) dx \\ &= \int_a^b \frac{(x - \frac{a+b}{2})(x - b)}{(a - \frac{a+b}{2})(a - b)} f(a) + \frac{(x - a)(x - b)}{(\frac{a+b}{2} - a)(\frac{a+b}{2} - b)} f\left(\frac{a+b}{2}\right) + \frac{(x - a)(x - \frac{a+b}{2})}{(b - a)(b - \frac{a+b}{2})} f(b) dx \\ &= \frac{b - a}{6} \left( f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right). \end{aligned}$$

◆ **Erreur** Si  $f \in C^4([a, b])$ , l'erreur de quadrature est majorée par

$$|E| \leq \max_{[a, b]} |f^{(4)}(x)| \frac{(b - a)^5}{2880}.$$

**Exemple 4.3.1** Approximons l'intégrale  $I = \int_0^1 \log(1+x) dx$  par la méthode de Simpson puis calculons l'erreur d'intégration. On a :

$$\begin{aligned} I &= \frac{1-0}{6} \left[ f(0) + 4f\left(\frac{1}{2}\right) + f(1) \right] \\ &= \frac{1}{6} \left[ 0 + 4 \log\left(\frac{3}{2}\right) + \log 2 \right] \\ &= 0.38583 \end{aligned}$$

L'erreur d'intégration :

$$\begin{aligned} |E| &\leq \frac{M_4}{2880} (b-a)^5, \text{ où } M_4 = \max_{[0,1]} |f^{(4)}(x)| = \left| \frac{-6}{(1+x)^4} \right| = 6 \\ \text{d'où } |E| &\leq \frac{6}{2880} = 0,2 \times 10^{-2} \end{aligned}$$

- La valeur exacte de  $I = \int_0^1 \log(1+x) dx = 0.38629$

- La valeur approchée de  $I$  par la méthode du trapèze :

$$I \simeq \frac{b-a}{2} [f(a) + f(b)] = \frac{1}{2} [0 + \log 2] = 0.34657$$

L'erreur d'intégration :

$$\begin{aligned} |E| &\leq \frac{(b-a)^3}{12} M_2; \quad M_2 = \max_{[0,1]} |f''(x)| = \max_{[0,1]} \left| \frac{-1}{(1+x)^2} \right| = 1 \\ \Rightarrow |E| &\leq \frac{1}{12} = 0.8 \times 10^{-1} \end{aligned}$$

Conclusion : La méthode de Simpson est plus précise que la méthode du trapèze.

## 4.4 Formule des Trapèzes composée

◆ Soit  $\{x_0, x_1, \dots, x_n\}$  une partition de  $[a, b]$ ,

où  $x_0 = a$ ,  $x_1 = a + h$ , ...,  $x_{n-1} = x_0 + (n-1)h$ ,  $x_n = b$  et  $h = \frac{b-a}{n}$ , on sait que

$$I = \int_a^b f(x) dx = \int_a^{x_1} f(x) dx + \int_{x_1}^{x_2} f(x) dx + \dots + \int_{x_{n-1}}^b f(x) dx = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f(x) dx.$$

En appliquant la formule du trapèze à chaque intervalle  $[x_{i-1}, x_i]$ ,  $i \in \overline{1, n}$  on obtient :

$$\int_{x_{i-1}}^{x_i} f(x) dx \simeq \frac{x_i - x_{i-1}}{2} [f(x_{i-1}) + f(x_i)], \text{ avec } x_i - x_{i-1} = h = \frac{b-a}{n}$$

c'est-à-dire

$$\int_a^b f(x) dx \simeq \frac{h}{2} \sum_{i=1}^n [f(x_{i-1}) + f(x_i)]$$

ce qui donne

$$\int_a^b f(x) dx \simeq \frac{b-a}{n} \left[ \frac{f(x_0)}{2} + f(x_1) + \dots + f(x_{n-1}) + \frac{f(x_n)}{2} \right].$$

Cette formule est appelée "formule des Trapèzes composées".

#### ◆ Erreur

On sait que l'erreur d'intégration  $E_i$  dans l'intervalle  $[x_{i-1}, x_i]$  est donnée par : l'intervalle

$$E_i = -\frac{(x_i - x_{i-1})^3}{12} f''(\alpha_i); \quad \alpha_i \in [x_{i-1}, x_i]$$

d'autre part :  $\sum_{i=1}^n E_i = E$  i.e :

$$\begin{aligned} E &= -\frac{h^3}{12} \sum_{i=1}^n f''(\alpha_i) \Rightarrow |E| \leq \frac{h^3}{12} \sum_{i=1}^n |f''(\alpha_i)| \leq \frac{h^3 \times n \times M_2}{12}, \text{ où } M_2 = \max_{[a,b]} |f''(x)| \\ &\Rightarrow |E| \leq \frac{(b-a)^3}{12n^2} M_2. \end{aligned}$$

Illustration de la règle du trapèze composée voir la figure 4.1.

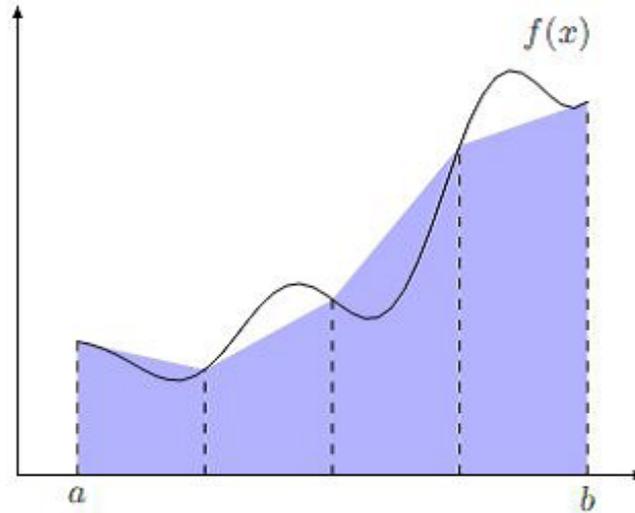


FIG 4.1 : trapèze composée à 4 sous intervalles sur  $[a, b]$

La valeur approchée de l'intégrale  $I$  correspond à l'aire colorée.

**Exemple 4.4.1** Approximons l'intégrale  $I = \int_0^1 \log(1+x) dx$  par la méthode du Trapèze composée avec  $n = 3$  et estimons l'erreur d'intégration associée.

On a :

$$\begin{aligned} I &\simeq \frac{1-0}{3} \left[ \frac{f(0)}{2} + f\left(\frac{1}{3}\right) + f\left(\frac{2}{3}\right) + \frac{f(1)}{2} \right] \\ &\Rightarrow I \simeq \frac{1}{3} \left[ \log\left(1 + \frac{1}{3}\right) + \log\left(1 + \frac{2}{3}\right) + \frac{1}{2} \log 2 \right] \\ I &\simeq 0,38169. \end{aligned}$$

L'erreur d'intégration :

$$\begin{aligned} |E| &\leq \frac{(1-0)^3}{108} M_2, \quad \text{où } M_2 = \max_{[0,1]} |f''(x)| = 1. \\ &\Rightarrow |E| \leq \frac{1}{108} = 0,92 \times 10^{-2}. \end{aligned}$$

## 4.5 Formule de Simpson composée

◆ On décompose l'intervalle  $[a, b]$  en  $n$  sous intervalles de largeur  $h = \frac{b-a}{n}$  avec  $n \geq 1$ . En introduisant les noeuds de quadrature  $x_i = a + i\frac{h}{2}$ ,  $i = 0, 1, \dots, 2n$  (i.e; chaque sous-intervalle  $[x_{2i}, x_{2i+2}]$  a pour largeur  $h$  et donc  $x_{2i+1}$  est son point milieu), la formule de Simpson composée est

$$\begin{aligned} I &= \int_a^b f(x)dx = \sum_{i=0}^{n-1} \int_{x_{2i}}^{x_{2i+2}} \tilde{f}(x)dx = \sum_{i=0}^{n-1} \frac{x_{2i+2} - x_{2i}}{6} (f(x_{2i}) + f(x_{2i+1}) + f(x_{2i+2})) \\ &= \frac{h}{6} \sum_{i=0}^{n-1} (f(x_{2i}) + f(x_{2i+1}) + f(x_{2i+2})) \\ &= \frac{h}{6} \left( f(a) + f(b) + \sum_{i=1}^{n-1} f(x_{2i}) + 4 \sum_{i=0}^{n-1} f(x_{2i+1}) \right) \\ &= \frac{h}{6} \left( f(a) + f(b) + \sum_{i=1}^{n-1} f(a + ih) + 4 \sum_{i=0}^{n-1} f\left(a + \left(i + \frac{1}{2}\right)h\right) \right). \end{aligned}$$

◆ **Erreur** Si  $f \in C^4([a, b])$  alors l'erreur de quadrature est majorée par

$$E_{[a,b]} = \sum_{i=0}^{n-1} E_{[x_{2i}, x_{2i+2}]} \leq n \frac{h^5}{2880} \max_{[a,b]} |f^{(4)}(x)| = \frac{b-a}{2880} h^4 \max_{[a,b]} |f^{(4)}(x)|.$$

Illustration de la règle de Simpson composée voir figure 4.2 :

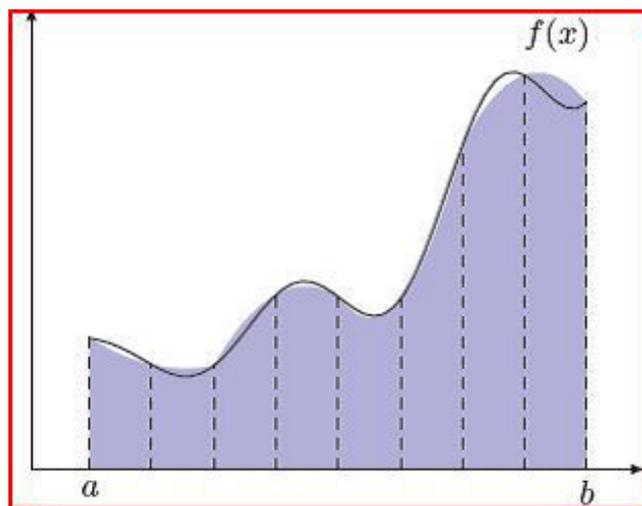


FIG 4.2 : Règle de Simpson composée à 4 sous-intervalles sur  $[a, b]$

La valeur approchée de l'intégrale  $I$  correspond à l'aire colorée.

## 4.6 Exercices

### Exercice 1

Calculer à 0.01 près l'intégrale  $I = \int_1^3 \frac{dx}{1+x}$  par la méthode des trapèzes généralisée et comparer le résultat avec la valeur exacte.

### Exercice 2

Trouver le nombre  $n$  de subdivisions nécessaires de l'intervalle d'intégration  $[-\pi, \pi]$ , pour évaluer à  $0.5 \cdot 10^{-3}$  près, grâce à la méthode de Simpson,

l'intégrale  $\int_{-\pi}^{\pi} \cos x \, dx$

### Exercice 3

Calculer une valeur approchée de l'intégrale définie

$$I = \int_0^1 \sqrt{1+x} \, dx$$

par la formule de Simpson à  $10^{-6}$  près, puis comparer le résultat trouvé avec la valeur exacte de l'intégrale.

### 4.6.1 Réponses

#### Exercice 1

On sait que

$$|E_T| \leq \frac{(b-a)^3}{12n^2} M_2 = \frac{b-a}{12n^2} h^2 M_2 \leq 0,01$$

$$h^2 \leq \frac{0,01 \times 12}{M_2(b-a)}, \text{ avec } M_2 = \max_{[1,3]} |f''(x)|$$

$f(x) = \frac{1}{x+1}$ ,  $f''(x) = \frac{2}{(1+x)^3}$  fonction décroissante son max est atteint pour la valeur 1 et  $h^2 \leq \frac{0,01 \times 12}{2 \times 0,25} = 0,24 \implies h \leq 0,48$ .

On prendra  $h = 0,4$  par suite  $n = \frac{b-a}{h} = \frac{2}{0,4} = 5$ .

$x_0 = 1 \rightarrow f_0 = 0,5$ ,  $x_1 = 1,4 \rightarrow f_1 = 0,416$ ,  $x_2 = 1,8 \rightarrow f_2 = 0,357$ ,

$x_3 = 2,2 \rightarrow f_3 = 0,312$ ,  $x_4 = 2,6 \rightarrow f_4 = 0,277$ ,

et  $x_5 = 3 \rightarrow f_5 = 0,25$ .

$$I = h \left[ \frac{f_0 + f_n}{2} + \sum_{i=1}^{n-1} f_i \right] = 0,4 \left[ \frac{0,5 + 0,25}{2} + 0,416 + 0,357 + 0,312 + 0,277 \right] = 0,694.$$

$I_T = 0,694 \pm 0,01$ . La valeur exacte  $\int_1^3 \frac{dx}{1+x} = [\ln(1+x)]_1^3 = 0,693$ .

**Exercice 2**

Le pas d'intégration est  $h = \frac{b-a}{n} = \frac{2\pi}{n}$ . D'autre part l'erreur théorique de la méthode de Simpson est donnée par

$$E(h) = -\frac{b-a}{180} h^4 f^{(4)}(\xi) = -\frac{2\pi}{180} \left(\frac{2\pi}{n}\right)^4 \cos \xi, \quad \text{avec } \xi \in [a, b].$$

Par conséquent

$$|E(h)| \leq \left| -\frac{2\pi}{180} \left(\frac{2\pi}{n}\right)^4 \right|.$$

Ainsi pour que  $|E(h)| \leq 0,5 \cdot 10^{-3}$ , il suffit que  $n$  vérifie  $\left| \frac{\pi}{90} \frac{16\pi^4}{n^4} \right| \leq 0,5 \cdot 10^{-3}$

donc  $n^4 \geq \frac{1}{0,5 \cdot 10^{-3}} \frac{\pi}{90} 16\pi^4$  ainsi  $n \geq 18,6$ .

On prendra par exemple  $n = 20$ , car pour la méthode de Simpson le nombre de subdivisions de  $[a, b]$  doit toujours être pair.

# Chapitre 5

## Résolution numérique des équations différentielles ordinaires

### 5.1 Introduction

Les équations différentielles, souvent obtenues lors de la modélisation des phénomènes physiques, peuvent être divisées en deux grandes familles :

Les équations différentielles ordinaires et les équations aux dérivées partielles. Dans ce dernier cas les équations sont caractérisées par le fait que la variable dépendante (ou les variables dépendantes) est (ou sont) fonction de plusieurs variables indépendantes.

L'autre catégorie d'équations différentielles sont les équations différentielles ordinaires qui sont caractérisées par le fait que la variable dépendante (ou les variables dépendantes) ne dépend (ou ne dépendent) que d'une seule variable indépendante et que sa dérivée soit alors, une dérivée totale.

Un exemple de ce genre d'équations est l'équation qui gouverne l'évolution, en fonction du temps  $t$ , de la vitesse  $v$  d'une masse  $m$  dans le champ de gravité terrestre  $g$  en chute amortie par un coefficient d'amortissement  $k$  :

$$\frac{dv}{dt} = g - \frac{k}{m} \times v,$$

ou bien encore, l'équation de la position angulaire  $\theta$  d'un pendule, de longueur

$l$ , qui oscille dans un plan vertical :

$$\frac{d^2\theta}{dt^2} + \frac{g}{l} \times \sin \theta = 0.$$

Il existe d'autres modèles basés sur des équations différentielles ordinaires qui sont extrêmement courants tel que :

En cinétique chimique, dynamique des populations et en météorologie.

Quand ces équations sont linéaires ou qu'elles peuvent raisonnablement être rendues linéaires, souvent des solutions analytiques peuvent être obtenues pour ces équations.

Toutefois, quand ces équations sont complexes ou non linéaires, les méthodes analytiques échouent et une approche numérique s'avère être une solution. L'objectif de ce chapitre est de présenter quelques méthodes numériques pour la résolution des équations différentielles ordinaires.

## 5.2 Préliminaires

Soit une équation différentielle de forme générale

$$F(y, y', y'', \dots, y^{(n)}, t) = 0, \quad t \in [t_0, T]. \quad (5.1)$$

On considère deux types de problèmes :

### 5.2.1 Problèmes de conditions aux limites

On connaît alors :

$$\left\{ \begin{array}{l} y(t_0) = y_0 \\ y'(t_0) = y'_0 \\ \vdots \\ y^{(n)}(t_0) = y_0^{(n)} \end{array} \right. \quad \left\{ \begin{array}{l} y(T) = y_T \\ y'(T) = y'_T \\ \vdots \\ y^{(n)}(T) = y_T^{(n)} \end{array} \right. .$$

Ces problèmes se traitent comme des problèmes d'équations aux dérivées partielles.

### 5.2.2 Problèmes de conditions initiales (problème de Cauchy)

On connaît ici :

$$\left\{ \begin{array}{l} y(t_0) = y_0 \\ y'(t_0) = y'_0 \\ \vdots \\ \vdots \\ y^{(n-1)}(t_0) = y_0^{(n-1)} \end{array} \right.$$

En général, une équation de type (5.1), se ramène à un système d'équations différentielles du 1er ordre; en effet

$$F(y, y', y'', \dots, y^{(n)}, t) = 0 \iff y^{(n)} = \varphi(y, y', y'', \dots, y^{(n-1)}, t)$$

Posons

$$u_1 = y$$

et

$$\begin{array}{lll} u'_1 = u_2 & \iff & \\ u'_2 = u_3 & \iff & u_3 = y'' \\ \vdots & \vdots & \vdots \\ u'_{n-1} = u_n & \iff & u_n = y^{(n-1)} \end{array}$$

et

$$u'_n = y^{(n)} = \varphi(y, y', \dots, y^{(n-1)}, t) = \varphi(u_1, \dots, u_n, t),$$

d'où en convenant des notations suivantes :

$$U(t) = \begin{pmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_n(t) \end{pmatrix} ; \quad \Phi(t, U) = \begin{pmatrix} u_2 \\ \vdots \\ u_n \\ \varphi(U, t) \end{pmatrix}, \quad U, \Phi \in \mathbb{R}^n.$$

On a donc ramené (5.1) à un système de la forme :

$$\left\{ \begin{array}{l} \frac{dU}{dt} = \Phi(t, U) \\ U(t_0) \text{ donné,} \end{array} \right. \quad \text{où } U(t_0) = \{y_0, y'_0, \dots, y_0^{(n-1)}\} \quad (5.2)$$

### 5.2.3 Existence et unicité des solutions

Dans toute la suite du chapitre, on ne s'intéressera qu'aux équations différentielles du premier ordre de conditions initiales du type :

$$\begin{cases} y'(t) = \frac{dy}{dt} = f(t, y(t)) & , y \in \mathbb{R}, \quad t \in [t_0, T] \\ y(t_0) = y_0 \end{cases} \quad (5.3)$$

La variable indépendante  $t$  représente très souvent le temps.

**Remarque 5.2.1** *On a vu que toute équation différentielle de conditions initiales d'ordre  $n$ , se ramène à un système d'équations différentielles du premier ordre et d'autre part les formules établies par une équation du premier ordre restent valables pour un système d'équations différentielles à condition de remplacer les fonctions scalaires par des fonctions vectorielles.*

#### Théorème(Cauchy-Lipshitz)

On suppose que la fonction  $f$  vérifie les conditions suivantes :

- $f$  est continue par rapport à ses deux variables
- $f(t, y(t))$  est Lipschitzienne en  $y$  uniformément par rapport à  $t$ , c'est-à-dire :  
 $\exists L > 0$  tel que  $|f(t, y) - f(t, z)| \leq L |y - z| \quad \forall t \in [t_0, T], \forall y, z \in \mathbb{R}$ .

Alors dans ce cas le problème (5.3) admet une solution unique passant par  $y_0$  pour  $t = t_0$ .

## 5.3 Principe d'intégration approchée par des méthodes numériques

Soit à résoudre le problème différentiel :

$$(P) \begin{cases} \frac{dy}{dt} = f(t, y(t)) & , y \in \mathbb{R}, \quad t \in [t_0, T] \\ y(t_0) = y_0 \end{cases}$$

1) On va subdiviser l'intervalle  $[t_0, T]$  en  $n$  parties égales par l'intermédiaire de points  $t_n$ , en général équidistants, qui déterminent le pas d'intégration  $h = t_{n+1} - t_n$ .

2) Effectuer le  $(n+1)^{ième}$  pas d'intégration consiste donc à passer de la valeur approchée  $y_n$  (approximation de la valeur exacte  $y(t_n)$ ) supposée calculée, à la valeur approchée suivante

$y_{n+1}$  (approximation de  $y(t_{n+1})$ ), (voir figure 5.1).

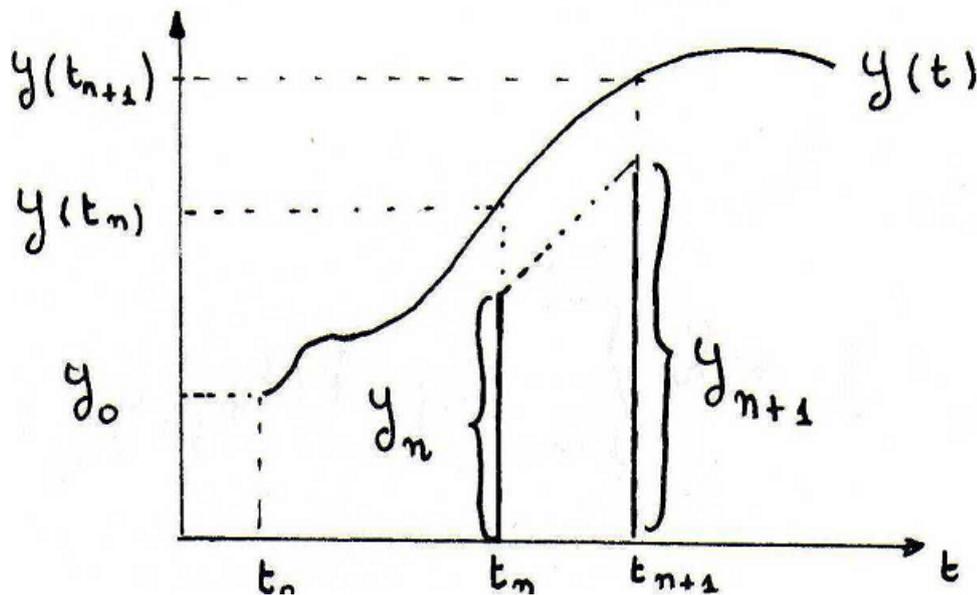


FIG. 5.1

On obtiendra alors la solution approchée sous forme de table numérique.

On distingue deux classes d'algorithmes (méthodes) numériques :

a) Méthodes à un pas (à pas séparés)

Ils permettent de calculer  $y_{n+1}$  en n'utilisant que la valeur  $y_n$ .

b) Méthodes à pas multiples (à pas liés)

Ils permettent d'obtenir  $y_{n+1}$  en utilisant les valeurs  $y_n, \dots, y_{n-k}$  ( $k$  fixé) ;  
pour initialiser ce type d'algorithme on effectue  $k$  pas d'une méthode à pas séparés.

## 5.4 Méthodes numériques à un pas

**Définition 5.4.1** *Les méthodes numériques à un pas sont Les méthodes de résolution numérique qui peuvent s'écrire sous la forme :*

$$y_{n+1} = y_n + h_n \Phi(t_n, y_n), \quad 0 \leq n \leq N$$

où  $\Phi : [t_0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  est une fonction que l'on supposera continue.

### 5.4.1 Méthode d'EULER

La méthode d'Euler est la méthode à un pas associée à la fonction

$$\Phi(t, y, h) = f(t, y).$$

Reprenons le problème  $(P)$ , le but est d'obtenir une approximation de la solution en  $t = t_1 = t_0 + h$ . Avant d'effectuer la première itération, il faut déterminer dans quelle direction on doit avancer à partir du point  $(t_0, y_0)$  pour obtenir le point  $(t_1, y_1)$ , qui est une approximation du point  $(t_1, y(t_1))$ . Nous n'avons pas l'équation de la courbe  $y(t)$ , mais nous en connaissons la pente  $y'(t)$  en  $t = t_0$ . En effet, l'équation différentielle assure que :

$$y'(t_0) = f(t_0, y(t_0)) = f(t_0, y_0).$$

On peut donc suivre la droite passant par  $(t_0, y_0)$  et de pente  $f(t_0, y_0)$  l'équation de cette droite, notée  $d_0(t)$ , est :

$$d_0(t) = y_0 + f(t_0, y_0)(t - t_0)$$

En  $t = t_1$ , on a :

$$d_0(t_1) = y_0 + f(t_0, y_0)(t_1 - t_0) = d_0(t) = y_0 + hf(t_0, y_0) = y_1$$

En d'autres termes,  $d_0(t_1)$  est proche de la solution analytique  $y(t_1)$ , c'est à dire

$$y(t_1) \simeq y_1 = d_0(t_1) = y_0 + hf(t_0, y_0).$$

Si on souhaite faire une deuxième itération et obtenir une approximation de  $y(t_2)$ , on peut refaire l'analyse précédente à partir du point  $(t_1, y_1)$ . On remarque cependant que la pente de la solution analytique en  $t = t_1$  est :

$$y'(t_1) = f(t_1, y(t_1)).$$

On ne connaît pas exactement  $y(t_1)$ , mais nous possédons l'approximation  $y_1$  de  $y(t_1)$ . On doit alors utiliser l'expression :

$$y'(t_1) = f(t_1, y(t_1)) \simeq f(t_1, y_1)$$

et construire la droite :

$$d_1(t) = y_1 + f(t_1, y_1)(t - t_1)$$

qui permettra d'estimer  $y(t_2)$ . On constate que l'erreur commise à la première itération est réintroduite dans les calculs de la deuxième itération. On a alors :

$$y(t_2) \simeq y_2 = d_1(t_2) = y_1 + hf(t_1, y_1).$$

Ceci conduit à l'Algorithme d'Euler :

$$\begin{cases} y_{n+1} = y_n + hf(t_n, y_n), & 0 \leq n \leq N \\ t_{n+1} = t_n + h \end{cases}$$

**Remarque 5.4.1** *L'erreur introduite à la première itération se répercute sur les calculs de la deuxième itération, ce qui signifie que les erreurs se propagent d'une itération à l'autre. Il en résulte que l'erreur :*

$$|y(t_i) - y_i|$$

*augmente légèrement avec  $i$ .*

**Exemple 1** : soit l'équation différentielle :

$$\begin{cases} y'(t) = -y(t) + t + 1 \\ y(0) = 1 \end{cases}$$

On a donc  $t_0 = 0$  et  $y(0) = 1$ , et on prend un pas de temps  $h = 0.1$ . De plus, on a :

$$f(t, y) = -y + t + 1$$

En utilisant la méthode d'Euler on obtient successivement des approximations de  $y(0.1)$ ,  $y(0.2)$ ,  $y(0.3)$ ...., notées  $y_1$ ,  $y_2$ ,  $y_3$ ....la première itération donne :

$$y_1 = y_0 + hf(t_0, y_0) = 1 + 0.1f(0, 1) = 1 + 0.1(-1 + 0 + 1) = 1$$

la deuxième itération fonctionne de manière similaire :

$$y_2 = y_1 + hf(t_1, y_1) = 1 + 0.1f(0.1, 1) = 1 + 0.1(-1 + 0.1 + 1) = 1.01$$

on parvient à :

$$\begin{aligned} y_3 &= y_2 + hf(t_2, y_2) = 1.01 + 0.1f(0.2, 1.01) \\ &= 1.01 + 0.1(-1.01 + 0.2 + 1) \\ &= 1.029 \end{aligned}$$

la tableau suivant rassemble les résultats des dix premières itérations.

La solution analytique, de cette équation différentielle est :  $y(t) = \exp(-t) + t$ .

Le tableau suivant rassemble les résultats des 10 premières itérations et permet de comparer les solutions numériques et analytiques et regarder la croissance de l'erreur.

$t_i$	$y(t_i)$	$y_i$	$ y(t_i) - y_i $
0.0	1.000000	1.000000	0.000000
0.1	1.004837	1.000000	0.004837
0.2	1.018731	1.010000	0.008731
0.3	1.040818	1.029000	0.011818
0.4	1.070302	1.056100	0.014220
0.5	1.106531	1.090490	0.016041
0.6	1.148812	1.131441	0.017371
0.7	1.196585	1.178297	0.018288
0.8	1.249329	1.230467	0.018862
0.9	1.306570	1.287420	0.019150
1.0	1.367879	1.348678	0.019201

### 5.4.2 Méthode de Taylor

Le développement de Taylor autorise une généralisation immédiate de la méthode d'Euler, qui permet d'obtenir des algorithmes dont l'erreur de troncature locale est d'ordre plus élevé. Nous nous limitons cependant à la méthode de Taylor de second ordre. On cherche, au temps  $t = t_n$ , une approximation de la solution en  $t = t_{n+1}$ . On a :

$$\begin{aligned} y(t_{n+1}) &= y(t_n + h) \\ &= y(t_n) + y'(t_n)h + \frac{y''(t_n)h^2}{2} + o(h^3) \end{aligned}$$

En se servant de l'équation différentielle (P), on trouve

$$y(t_{n+1}) = y(t_n) + f(t_n, y(t_n))h + \frac{f'(t_n, y(t_n))h^2}{2} + 0(h^3)$$

**Remarque 5.4.2** Dans la relation précédente, on voit apparaître la dérivée de la fonction  $f(t, y(t))$  par rapport au temps. On sait que :

$$f'(t, y(t)) = \frac{\partial f(t, y(t))}{\partial t} + \frac{\partial f(t, y(t))}{\partial y} y'(t)$$

c'est à dire

$$f'(t, y(t)) = \frac{\partial f(t, y(t))}{\partial t} + \frac{\partial f(t, y(t))}{\partial y} f'(t; y(t))$$

on obtient donc

$$\begin{aligned} y(t_{n+1}) &= y(t_n) + hf(t_n, y(t_n)) & (5.4) \\ &+ \frac{h^2}{2} \left( \frac{\partial f(t_n, y(t_n))}{\partial t} + \frac{\partial f(t_n, y(t_n))}{\partial y} f(t_n; y(t_n)) \right) \\ &+ 0(h^3). \end{aligned}$$

En négligeant les termes d'ordre supérieur ou égal à 3, on en arrive à poser

$$\begin{aligned} y(t_{n+1}) &\simeq y(t_n) + hf(t_n, y(t_n)) & (5.5) \\ &+ \frac{h^2}{2} \left( \frac{\partial f(t_n, y(t_n))}{\partial t} + \frac{\partial f(t_n, y(t_n))}{\partial y} f(t_n, y(t_n)) \right) \end{aligned}$$

qui sera à la base de la méthode de Taylor. Ceci conduit à l'algorithme de Taylor d'ordre 2

$$\begin{aligned} y_{n+1} &= y_n + hf(t_n, y_n) \\ &+ \frac{h^2}{2} \left( \frac{\partial f(t_n, y_n)}{\partial t} + \frac{\partial f(t_n, y_n)}{\partial y} f(t_n, y_n) \right) \\ \text{avec } t_{n+1} &= t_n + h \end{aligned}$$

Dans cet algorithme, on a remplacé la solution analytique  $y(t_n)$  par son approximation  $y_n$  dans (5.5). On en conclut que les erreurs se propagent d'une itération à une autre.

**Exemple 5.4.1** Soit l'équation différentielle déjà résolu par la méthode d'Euler :

$$y'(t) = -y(t) + t + 1 \quad y(0) = 1$$

Dans ce cas :

$$f(t, y) = -y(t) + t + 1$$

et

$$\frac{\partial f}{\partial t} = 1 \text{ et } \frac{\partial f}{\partial y} = -1$$

On aura :

$$y_{n+1} = y_n + hf(-y(t) + t + 1) + \frac{h^2}{2}(1 + (-1)(-y(t) + t + 1))$$

La première itération de la méthode de Taylor d'ordre 2 donne (avec  $h = 0.1$ )

$$y_1 = 1 + 0.1(-1 + 0 + 1) + \frac{(0.1)^2}{2}(1 + (-1)(-1 + 0 + 1)) = 1.005$$

Une deuxième itération donne

$$\begin{aligned} y_2 &= 1.005 + 0.1(-1.005 + 0.1 + 1) \frac{(0.1)^2}{2}(1 + (-1)(-1.005 + 0.1 + 1)) \\ &= 1.019025 \end{aligned}$$

Les résultats sont donnés dans le tableau suivant

$t_i$	$y(t_i)$	$y_i$	$ y(t_i) - y_i $
0.0	1.000000	1.000000	0.000000
0.1	1.004837	1.005000	0.000163
0.2	1.018731	1.019025	0.000294
0.3	1.040818	1.041218	0.000400
0.4	1.070302	1.070802	0.000482
0.5	1.106531	1.107075	0.000544
0.6	1.148812	1.149404	0.000592
0.7	1.196585	1.197210	0.000625
0.8	1.249329	1.249975	0.000646
0.9	1.306570	1.307228	0.000658
1.0	1.367879	1.368541	0.000662

**Remarque 5.4.3** On remarque que l'erreur est plus petite avec la méthode de Taylor d'ordre 2 qu'avec la méthode d'Euler.

L'algorithme de Taylor d'ordre  $p$  est :

$$y_{n+1} = y_n + \sum_{k=1}^p \frac{h^k}{k!} f^{(k-1)}(t_n, y_n)$$

avec  $t_{n+1} = t_n + h$

**Remarque 5.4.4** *En pratique, la méthode de Taylor pour  $p \geq 2$  est presque inutilisable pour les raisons suivantes : Le calcul des quantités  $f^{(k)}$  est très complexe et très couteux en temps machine. En plus la méthode suppose a priori que  $f$  soit très régulière ; les erreurs risquent donc de ne pas pouvoir être contrôlées si certaines dérivées de  $f$  présentent des discontinuités. Un moyen pour surmonter cette difficulté : les méthodes de Runge-kutta.*

### 5.4.3 Méthode de Runge- kutta

Les méthodes de Runge-kutta sont inspirées des méthodes de Taylor du même ordre.

#### Méthode de Runge - kutta d'ordre 2 :

On a vu que le développement de la méthode de Taylor passe par la relation (4)

$$y(t_{n+1}) = y(t_n) + hf(t_n, y(t_n)) + \frac{h^2}{2} \left( \frac{\partial f(t_n, y(t_n))}{\partial t} + \frac{\partial f(t_n, y(t_n))}{\partial y} f(t_n, y(t_n)) \right) + O(h^3). \quad (5.6)$$

Le but est de remplacer cette dernière relation par une expression équivalente possédant le même ordre de précision ( $O(h^3)$ ). Soit la forme :

$$y(t_{n+1}) = y(t_n) + a_1 hf(t_n, y(t_n)) + a_2 hf(t_n + a_3 h, y(t_n) + a_4 h) \quad (5.7)$$

où on doit déterminer les paramètres  $a_1, a_2, a_3$  et  $a_4$  de telle sorte les expressions (5.6) et (5.7) aient toutes deux une erreur en  $O(h^3)$ . On ne trouve par ailleurs aucune dérivée partielle dans cette expression. Pour  $y$  arriver, on doit recourir au développement de Taylor en deux variables autour du point  $(t_n, y(t_n))$ . On a ainsi :

$$f(t_n + a_3 h, y(t_n) + a_4 h) = f(t_n, y(t_n)) + a_3 h \frac{\partial f(t_n, y(t_n))}{\partial t} + a_4 h \frac{\partial f(t_n, y(t_n))}{\partial y} + O(h^2)$$

la relation (5.7) devient alors :

$$y(t_{n+1}) = y(t_n) + (a_1 + a_2) hf(t_n, y(t_n)) + a_2 a_3 h^2 \frac{\partial f(t_n, y(t_n))}{\partial t} + a_2 a_4 h^2 \frac{\partial f(t_n, y(t_n))}{\partial y} + O(h^3) \quad (5.8)$$

On voit immédiatement que les expressions (5.2) et (5.4) sont du même ordre. Pour déterminer les coefficients  $a_i$ , il suffit de comparer ces deux expressions terme à terme

◆ coefficients respectifs de  $f(t_n, y(t_n))$  :

$$h = (a_1 + a_2)h$$

◆ coefficients respectifs de  $\frac{\partial f(t_n, y(t_n))}{\partial t}$  :

$$\frac{h^2}{2} = a_2 a_3 h^2$$

◆ coefficients respectifs de  $\frac{\partial f(t_n, y(t_n))}{\partial y}$  :

$$\frac{h^2}{2} f(t_n, y(t_n)) = a_2 a_4 h^2.$$

On obtient ainsi un système non linéaire de 3 équations comprenant 4 inconnues :

$$\begin{cases} 1 = & (a_1 + a_2) \\ \frac{1}{2} = & a_2 a_3 \\ \frac{f(t_n, y(t_n))}{2} = & a_2 a_4 \end{cases} \quad (5.9)$$

le système (5.9) est insoluble car il y a moins d'équation que d'inconnues. D'où les variantes de la méthode de Runge-kutta d'ordre 2. Voici deux variantes les plus couramment utilisées.

### Méthode d'Euler modifiée

$$\begin{aligned} a_1 &= a_2 = \frac{1}{2}, \\ a_3 &= 1 \text{ et } a_4 = f(t_n, y(t_n)) \end{aligned}$$

Ainsi on obtient l'algorithme de la méthode d'Euler modifiée

$$\begin{aligned} \hat{y} &= y_n + hf(t_n, y_n) \\ y_{n+1} &= y_n + \frac{h}{2} f(t_n, y_n) + f(t_{n+1}, \hat{y}) \\ t_{n+1} &= t_n + h \end{aligned}$$

**Méthode du point milieu**

$$a_1 = 0, a_2 = 1, a_3 = \frac{1}{2}$$

$$\text{et } a_4 = \frac{f(t_n, y(t_n))}{2}$$

En remplaçant ces valeurs des coefficients  $a_i$  dans l'équation (5.7) on obtient l'algorithme suivant :

$$k_1 = hf(t_n, y_n)$$

$$y_{n+1} = y_n + h\left(f\left(t_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right)\right)$$

$$t_{n+1} = t_n + h.$$

**Remarque 5.4.5** On constate que la fonction  $f(t, y)$  est évaluée au point milieu de l'intervalle  $[t_n, t_{n+1}]$ , d'où le nom de la méthode.

**5.4.4 Méthode de Runge - kutta d'ordre 4 :**

L'algorithme de la méthode Runge - kutta d'ordre 4 est :

$$k_1 = hf(t_n, y_n)$$

$$k_2 = hf\left(t_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right)$$

$$k_3 = hf\left(t_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right)$$

$$k_4 = hf(t_n + h, y_n + k_3)$$

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

$$t_{n+1} = t_n + h.$$

**5.5 Exercices****Exercice 1**

Soit le problème de Cauchy

$$(1) \begin{cases} y' = y + t \\ y(0) = 1 \end{cases}$$

Approcher à  $10^{-3}$  la solution du problème 1 en  $t = 1$  à l'aide de la méthode d'Euler en subdivisant l'intervalle  $[0, 1]$  en 10 parties égales.

Sachant que la solution exacte de (1) est  $y(t) = 2 \exp(t) - t - 1$ .

**Exercice 2**

Soit le problème de Cauchy

$$(1) \begin{cases} y' = y - \frac{2t}{y} \\ y(0) = 1 \end{cases}$$

dont la solution exacte est  $y = \sqrt{2x + 1}$ .

Approcher la solution de (1) en  $t = 0.2$  en effectuant les calculs avec 6 décimales à l'aide des méthodes d'Euler modifié et de Runge-Kutta d'ordre 2.

**5.5.1 Réponses**

**Exercice 1**

Par l'algorithme d'Euler  $0 \leq n \leq 9$  et  $h = 0,1$  : on trouve  $y(1) \simeq 3,187$  et la solution exacte de l'équation (1) est donnée par l'équation  $y(t) = 2 \exp(t) - t - 1$ . Ce qui donne  $y(1) \simeq 3,437$ . L'approximation calculée est donc très grossière.

**Exercice 2**

La solution exacte étant  $y = \sqrt{2x + 1}$ . On considère l'intervalle  $[t_0, t_1]$  avec  $t_0 = 0, t_1 = t_0 + h = 0,2$  et  $h = 0,2$ .

Méthode de Runge-Kutta d'ordre 2 :  $\hat{y} = y_0 + h f(t_0, y_0)$

avec  $f(t_0, y_0) = y_0 - \frac{2t_0}{y_0} = 1$  donc  $\hat{y} = 1,2$ .

$f(t_1, y_0 + h f(t_0, y_0)) = f(0,2, 1,2) = 1,2 - \frac{2 \times 0,2}{1,2} = 0,866667$ .

$y_1 = y_0 + \frac{h}{2} (f(t_0, y_0) + f(t_1, \hat{y})) = 1 + \frac{0,2}{2} (1 + 0,866667) = 1,1866667$ .

Ainsi  $y(0,2) \simeq y_1 = 1,1866667$ . La valeur exacte étant

$y(0,2) = \sqrt{1,4} = 1,183216$ .

L'erreur commise est  $|y(0,2) - y_1| = 0,003451 < 0,5 \cdot 10^{-2}$ .

D'où  $y(0,2) = 1,19 \pm 0,01$ , ainsi  $y_1$  approche  $y(0,2)$  avec 3 chiffres significatifs exacts.

# Chapitre 6

## Résolution des équations non linéaires

### 6.1 Introduction

Beaucoup de problèmes réels se ramènent à la résolution d'une équation de la forme

$$f(x) = 0. \quad (6.1)$$

Ce problème intervient notamment dans l'étude générale d'une fonction d'une variable (réelle), pour laquelle des solutions exactes ne sont pas explicitement connues. Même dans le cas d'une équation algébrique, nous rappelons qu'il n'existe pas de méthode de résolution générale à partir du degré cinq.

Alors, on fait appel aux méthodes numériques de résolution qui sont, en général, des méthodes itératives fournissant la solution sous forme d'approximations successives.

### 6.2 Généralités sur les méthodes de résolution

Les méthodes numériques pour approcher une racine  $\hat{x}$  de  $f$  consistent à :

1. localiser grossièrement la racine de  $f$  en procédant à l'étude du graphe ou à des évaluations qui sont souvent de type graphique ; on note  $x_0$  cette solution grossière ;
2. Ensuite, on procède à construire, à partir de  $x_0$  une suite  $x_1, x_2, x_3, \dots$

telle que

$$\lim_{n \rightarrow +\infty} x_n = \hat{x} \quad \text{où} \quad f(\hat{x}) = 0.$$

On dit alors que la méthode est convergente.

**Définition 6.2.1** *Méthode itérative à deux niveaux*

On appelle méthode itérative à deux niveaux un procédé de calcul de la forme  $x_{n+1} = \phi(x_n)$ ,  $n = 0, 1, 2, \dots$

dans lequel on part d'une valeur donnée  $x_0$  pour calculer  $x_1$ , puis à partir de  $x_1$  on calcule  $x_2$  etc. La formule est dite formule de récurrence. Le procédé est appelé convergent si la suite  $(x_n)$  est convergente. Il est bien évident qu'une méthode itérative n'est utile que s'il y a convergence vers les valeurs cherchées.

**Remarque 6.2.1** *On peut envisager des méthodes itératives multi-niveaux, comme par exemple les schémas à trois niveaux dans lesquels on part de deux valeurs données  $x_0$  et  $x_1$  pour calculer  $x_2$ , puis à l'aide de  $x_1$  et  $x_2$  on calcule  $x_3$  etc.. ;*

**Ordre de convergence**

**Définition 6.2.2** *On dit qu'une méthode convergente est d'ordre  $p \geq 1$  s'il existe une constante  $\alpha$  telle que :  $|\hat{x} - x_{k+1}| \leq \alpha |\hat{x} - x_k|^p$  ou encore*

$$\lim_{k \rightarrow +\infty} \frac{|x_{k+1} - \hat{x}|}{|x_k - \hat{x}|^p} = \alpha.$$

Si  $p = 1$  et  $0 < \alpha < 1$  la convergence est dite linéaire, si  $p = 2$  la convergence est dite quadratique..

Si  $p > 1$  ou  $p = 1$  et  $\alpha = 0$  alors la convergence est dite super-linéaire et si  $\alpha = 1$  elle est dite sous-linéaire.

**Remarque 6.2.2** *Pour obtenir l'ordre d'un processus itératif de la forme*

$$x_{n+1} = \phi(x_n)$$

*on utilise un développement de TAYLOR au voisinage de  $x$*

$$x_{n+1} - x = \phi(x_n) - x = \phi(x_n) - \phi(x),$$

(en effet  $f(x) = 0$  peut être ramenée à la forme  $x - \phi(x)$  il suffit de poser  $\phi(x) = x + f(x)$ ) or :

$$\phi(x_n) = \phi(x) + (x_n - x) \phi'(x) + \frac{(x_n - x)^2}{2!} \phi''(x) + \dots + \frac{(x_n - x)^k}{k!} \phi^{(k)}(x) \dots$$

donc

$$x - x_{n+1} = \sum_{i=1}^{\infty} \frac{(x - x_n)^i}{i!} \phi^{(i)}(x)$$

ainsi si :

$$\phi^{(i)}(x) = 0, \quad \forall i = 1, 2, \dots, (p-1)$$

le processus itératif est d'ordre  $p$ .

**Remarque 6.2.3** La convergence d'un algorithme itératif de la forme

$$x_{n+1} = \phi(x_n)$$

est d'autant plus rapide que son ordre est grand.

### Critère d'arrêt

Il est important de disposer d'un critère pratique pour interrompre le processus itératif d'approximation de  $\hat{x}$  au bout d'un certain nombre fini d'opérations.

Nous présentons quelques tests d'arrêt :

$$|x_n - x_{n-1}| < c \times \varepsilon$$

$$|x_n - x_{n-1}| < \varepsilon \times (1 + |x_n|)$$

$$|x_n - x_{n-1}| < c' \times \varepsilon \times |x_n|$$

$$|f(x_n)| < \varepsilon$$

Avec  $\varepsilon$  est l'erreur maximale tolérée,  $c$  et  $c'$  sont des constantes définies par l'utilisateur.

## 6.3 Séparation des racines

**Définition 6.3.1** Une racine  $\hat{x}$  est dite séparable si l'on peut trouver des voisinages de  $x$  qui ne contiennent pas d'autres racines que  $\hat{x}$ . La racine  $\hat{x}$  est séparée si on peut trouver des nombres  $a$  et  $b$  tels que :  $a < x < b$ .

On utilise les méthodes suivantes pour séparer les racines d'une équation.

### 6.3.1 Méthode graphique

On localise grossièrement les racines de l'équation  $f(x) = 0$  en considérant par exemple :

- le graphe de  $y = f(x)$  et on cherche son intersection avec l'axe  $Ox$ .
- les graphes de  $y_1 = f_1(x)$  et  $y_2 = f_2(x)$  où  $f(x) = f_1(x) - f_2(x)$ , puis, on cherche les points d'intersection des graphes de  $f_1$  et  $f_2$ , dont les abscisses sont exactement les racines de l'équation  $f(x) = 0$ .

### 6.3.2 Méthodes algébriques

Soit une fonction continue  $f : [a, b] \rightarrow \mathbb{R}$ . On utilise le corollaire du théorème de Rolle :

Entre deux racines successives  $x_0$  et  $x_1$  de  $f'(x)$  il y a une seule racine de  $f(x) = 0$

si  $f(x_0).f(x_1) < 0$ .

Notons que cela nécessite la connaissance des racines de  $f'(x) = 0$ .

Utilisons le résultat précédent :

- Si  $f(a).f(b) < 0$   $\left\{ \begin{array}{l} f(x) = 0 \text{ a au moins une racine dans } [a, b], \\ f(x) = 0 \text{ a une seule dans } [a, b] \text{ si } f'(x) \neq 0 \text{ dans } ]a, b[. \end{array} \right.$

- Si  $f(a).f(b) > 0$   $\left\{ \begin{array}{l} f(x) = 0 \text{ n'a pas de racine dans } [a, b], \\ \text{ou } f(x) = 0 \text{ a un nombre pair de racines dans } [a, b] \end{array} \right.$

## 6.4 Approximation des racines

### 6.4.1 Méthode de dichotomie ( ou méthode de la bisection)

#### Principe

La méthode de dichotomie consiste à construire une suite de sous intervalles emboîtés  $[a_n, b_n]$

contenant une racine de l'équation  $f(x) = 0$ , outil de base le théorème des valeurs intermédiaires.

La fonction  $f$  est continue sur l'intervalle  $[a, b]$ , vérifie  $f(a).f(b) < 0$ , alors, d'après le théorème des valeurs intermédiaires, il existe au moins un point  $\hat{x} \in ]a, b[$  tel que  $f(\hat{x}) = 0$ .

Le principe est le suivant : posons  $a_1 = a$ ,  $b_1 = b$ , on note la valeur initiale  $x_1 = \frac{a_1+b_1}{2}$  milieu de l'intervalle  $[a_1, b_1]$ , et on évalue  $f$  en ce point.

- a) Si  $f(x_1) = 0$ , le point  $x_1$  est la solution  $\hat{x}$  et le problème est résolu.
- b) Si  $f(a_1).f(x_1) < 0$ , alors on a  $\hat{x} \in ]a_1, x_1[$  et on posera  $a_2 = a_1$  et  $b_2 = x_1$
- c) Si  $f(b_1).f(x_1) < 0$ , alors on a  $\hat{x} \in ]x_1, b_1[$  et on posera  $a_2 = x_1$  et  $b_2 = b_1$

Dans les deux derniers cas, on réitère le processus en partant de la valeur :  $x_2 = \frac{a_2+b_2}{2}$  et ainsi de suite...

Ce procédé permet de construire de manière récurrente la suite

$$x_1 = \frac{a_1 + b_1}{2}, x_2 = \frac{a_2 + b_2}{2}, \dots, x_n = \frac{a_n + b_n}{2}, \dots$$

qui converge vers la racine  $\hat{x}$  de l'équation  $f(x) = 0$ .

### Critère d'arrêt de la méthode

Avec la méthode de la dichotomie, les itérations s'achèvent à la  $m$ -ème étape quand la longueur de l'intervalle contenant la solution est inférieure à une tolérance  $\varepsilon$  fixée.

Posons  $L_0 = b - a$  la longueur de l'intervalle à la première itération, clairement, après  $m$  itérations la longueur de l'intervalle est égale à  $L_m = \frac{L_0}{2^m}$

$$\begin{aligned} |x_m - \hat{x}| \leq L_m \leq \varepsilon &\Rightarrow \frac{L_0}{2^m} \leq \varepsilon \Rightarrow \log\left(\frac{L_0}{2^m}\right) \leq \log(\varepsilon) \Rightarrow \log(L_0) - \log(2^m) \leq \log(\varepsilon) \\ &\Rightarrow \log(2^m) \geq \log(L_0) - \log(\varepsilon) \Rightarrow m \log(2) \geq \log(L_0) - \log(\varepsilon) \end{aligned}$$

$$\Rightarrow m \geq \frac{\log(L_0) - \log(\varepsilon)}{\log(2)} \Rightarrow m \geq \frac{\log\left(\frac{L_0}{\varepsilon}\right)}{\log(2)}$$

Notons que cette inégalité est générale elle ne dépend pas du choix de la fonction  $f$ .

**Exemple 6.4.1** Si on désire calculer une approximation  $x_k$  de  $\hat{x}$  avec  $n$  décimales exactes, il suffit de poser  $\frac{b-a}{2^k} \leq 0.5 \cdot 10^{-n}$ . Ce qui revient à ce qu'on aille dans les itérations jusqu'à ce que  $k$  vérifie l'inégalité :

$$k \geq \frac{\log \frac{b-a}{0.5 \cdot 10^{-n}}}{\log 2}.$$

**Exemple 6.4.2** Soit  $f(x) = -39 - 43x + 39x^2 - 5x^3$ . On cherche à estimer  $x \in [1; 5]$  tel que  $f(x) = 0$ . (Voir figure (6.1)).

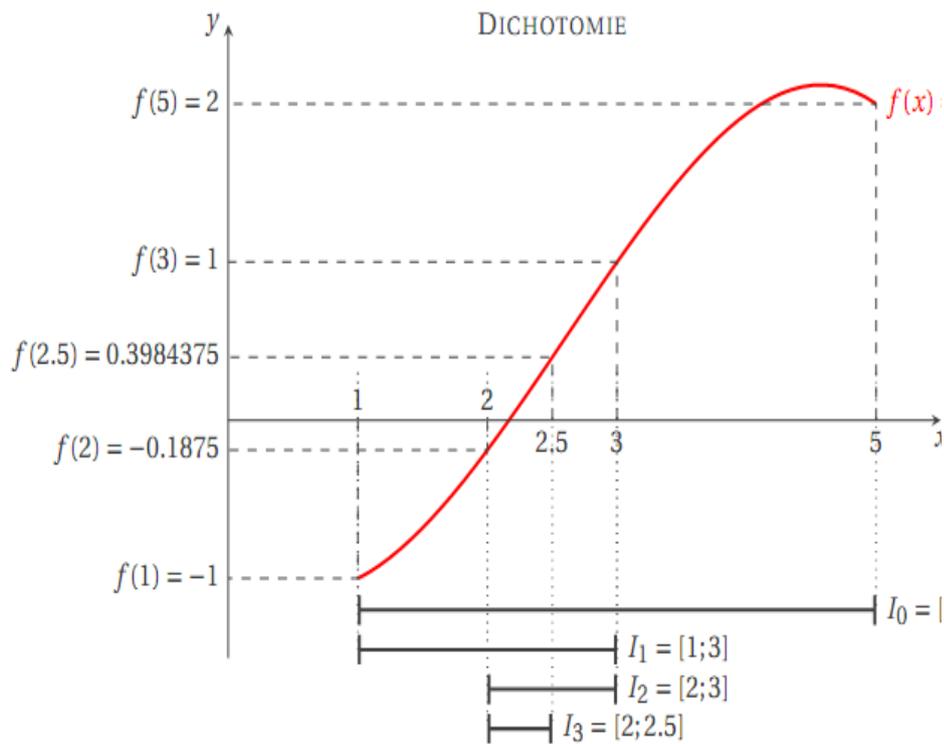


FIGURE 6.1

**Etude de convergence**

**Proposition 6.4.1** Soit  $f$  une fonction réelle continue sur  $[a, b]$ , vérifiant

$$f(a) \cdot f(b) < 0$$

et soit  $\hat{x} \in ]a, b[$  l'unique solution de l'équation  $f(x) = 0$ . Alors, la suite  $(x_n)_{n \in \mathbb{N}}$  construite par la méthode de dichotomie converge vers  $\hat{x}$  et on a l'estimation

$$\forall n \in \mathbb{N}, \quad |\hat{x} - x_n| \leq \frac{b - a}{2^n}$$

**Ordre de convergence**

La convergence de la méthode de dichotomie est linéaire, en effet

$$\lim_{k \rightarrow +\infty} \frac{|x_{k+1} - \hat{x}|}{|x_k - \hat{x}|} = \frac{1}{2} < 1.$$

La méthode de dichotomie est simple, malheureusement, assez lente. Cependant, cette lenteur est compensée par le fait que l'on est assuré d'obtenir toujours une suite qui converge vers une racine de  $f(x) = 0$  dans  $[a, b]$ . On dit que c'est une méthode globalement convergente.

On peut l'utiliser pour obtenir une approximation grossière (mais raisonnable) de la racine recherchée servant d'initialisation à une méthode d'ordre plus élevé dont la convergence n'est que locale, comme la méthode de NEWTON-RAPHSON (voir dans la section qui suit).

**6.4.2 Méthode de Newton-Raphson**

La méthode de Newton-Raphson, encore appelée méthode de la tangente ou encore méthode de Newton. Consiste à construire une suite  $(x_n)_{n \in \mathbb{N}}$ ; telle que l'itéré  $x_{n+1}$  est obtenue comme l'intersection de la tangente à la courbe de  $f(x)$  au point  $(x_n; f(x_n))$  avec l'axe des abscisses.

Cette tangente a pour équation :

$$y = f'(x_n)(x_{n+1} - x_n) + f(x_n)$$

En posant  $y = 0$  on obtient la formule de Newton

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

En effet, soit  $\hat{x}$  une racine recherchée et  $x_0$  une estimation à priori de  $\hat{x}$ , supposons également que la fonction  $f$  est de classe  $\mathcal{C}^2$  au voisinage de  $\hat{x}$ , le développement de Taylor d'ordre 1 donne

$$f(\hat{x}) = f(x_0) + f'(x_0)(\hat{x} - x_0) + \frac{f''(\xi)}{2}(\hat{x} - x_0)^2 \text{ avec } \xi \in ]\hat{x}, x_0[$$

comme  $f(\hat{x}) = 0$  on obtient :

$$0 = f(x_0) + f'(x_0)(\hat{x} - x_0) + \frac{f''(\xi)}{2}(\hat{x} - x_0)^2 \quad (*)$$

En négligeant le reste :

$$\frac{f''(\xi)}{2}(\hat{x} - x_0)^2$$

L'équation (\*) peut être approximée par :

$$0 \simeq f(x_0) + f'(x_0)(\hat{x} - x_0) \Rightarrow \hat{x} \simeq x_0 - \frac{f(x_0)}{f'(x_0)} = x_1$$

En répétant le processus on obtient la formule de récurrence :

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

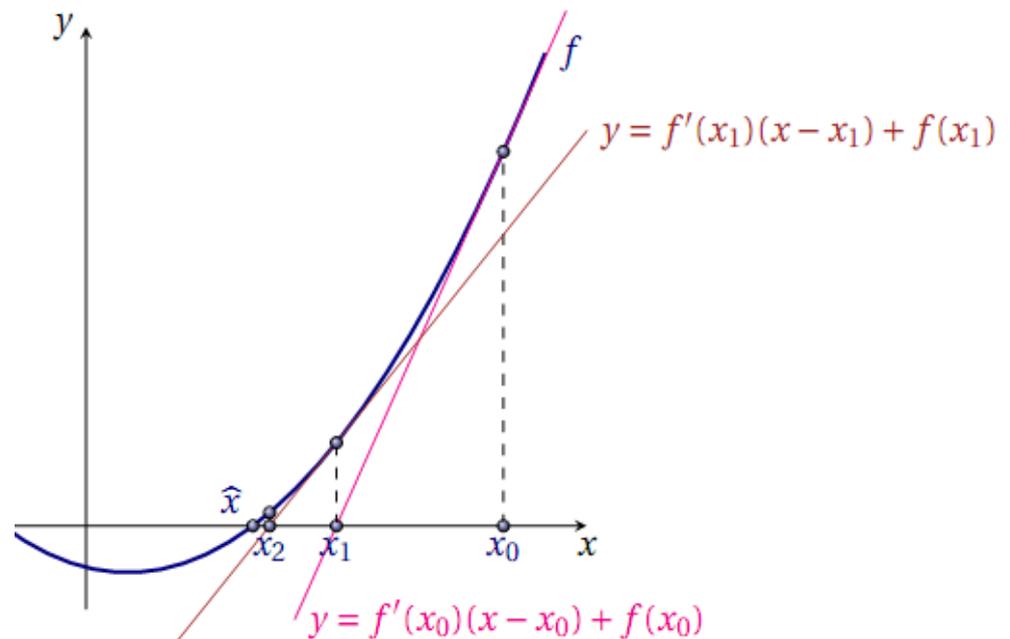


FIGURE 6.2 : Construction des premiers itérés de la méthode de Newton-Raphson

On remarque que la méthode de Newton nécessite à chaque itération l'évaluation des deux fonctions  $f$  et  $f'$  au point courant  $x_n$ . Cet effort est compensé par le fait que cette méthode est de convergence quadratique si la racine recherchée est simple. En effet, on a :

$$\phi(x) = x - \frac{f(x)}{f'(x)}$$

et

$$\phi' = \frac{f(x) \cdot f''(x)}{f'(x)^2} = 0 \quad \text{en } x$$

donc l'algorithme est d'ordre 2.

**Remarque 6.4.1** *Si la racine  $x$  est de multiplicité  $m$ , nous avons l'algorithme suivant :*

$$\forall n \in \mathbb{N}, \quad x_{n+1} = x_n - m \frac{f(x_n)}{f'(x_n)}.$$

La plus grande difficulté dans l'utilisation de la méthode de NEWTON-RAPHSON réside dans le caractère local de sa convergence. Si l'initialisation  $x_0$  est trop éloignée de la racine, la méthode peut très bien diverger. Pour cette raison, il est courant dans les applications de l'associer à une méthode d'encadrement, comme la méthode de dichotomie, cette dernière permettant d'approcher, bien que lentement, la racine recherchée de manière à fournir une bonne initialisation pour la méthode de NEWTON-RAPHSON.

En conclusion, la méthode de NEWTON-RAPHSON est la méthode de choix en termes de vitesse de convergence, puisqu'elle converge de manière quadratique pour un  $x_0$  assez proche de  $x$ . Par ailleurs, elle nécessite pour cela que la dérivée de la fonction  $f$  puisse être évaluée en tout point donné. Si cela n'est pas le cas, on utilisera la méthode de la sécante (voir la prochaine section), dont la vitesse de convergence est moindre mais ne requiert pas que la dérivée de  $f$  existe.

### 6.4.3 Méthode de la sécante

La méthode de la sécante peut être considérée comme une quasi-méthode de NEWTON-RAPHSON, dans laquelle on aurait remplacé la valeur de  $f'(x_k)$  par le taux d'accroissement de  $f$  sur un petit intervalle. C'est l'une des méthodes que l'on peut employer lorsque la dérivée de  $f$  est compliquée, voir impossible, à calculer ou encore coûteuse à évaluer.

Plus précisément,

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

et l'équation de la sécante aux points  $x_{n-1}$  et  $x_n$  est :

$$y = f(x_n) + (x - x_n) \frac{(f(x_n) - f(x_{n-1}))}{(x_n - x_{n-1})}$$

Ainsi au lieu d'utiliser la tangente au point  $x_n$  nous allons utiliser la sécante joignant le point  $(x_{n-1}, f(x_{n-1}))$  au point  $(x_n, f(x_n))$  pour en déduire  $x_{n+1}$ .

Ce dernier est obtenu comme intersection de la sécante passant par les points d'abscisse  $x_{n-1}$  et  $x_n$  et de l'axe des abscisses, i.e. on cherche  $x$  solution du système .

$$\begin{cases} y = f(x_n) + (x - x_n) \frac{(f(x_n) - f(x_{n-1}))}{(x_n - x_{n-1})} \\ y = 0 \end{cases}$$

ce qui donne

$$x = x_n - \frac{(x_n - x_{n-1})}{(f(x_n) - f(x_{n-1}))} f(x_n).$$

D'où l' algorithme correspondant est l'algorithme de la sécante

$$\begin{cases} x_0, x_1 & \text{donnés} \\ \text{Pour} & n = 0, \dots \\ \text{POSER} & x_{n+1} = x_n - \frac{(x_n - x_{n-1})}{(f(x_n) - f(x_{n-1}))} f(x_n) \end{cases}$$

### Critère d'arrêt

Le processus s'achève dès que l'on ait  $|x_{n+1} - x_n| \leq \varepsilon$ ,  $\varepsilon$  tolérance fixée.

### Convergence

**Théorème 6.4.1** *Soit  $f$  est une fonction de classe  $C^2(I)$ ,  $I$  étant un voisinage de  $\hat{x}$ , l'unique racine de  $f$  dans  $I$ . Alors, il existe un  $\eta > 0$  tel que si  $x_0, x_1$  sont choisis dans l'intervalle  $]\hat{x} - \eta, \hat{x} + \eta[$  alors la suite  $(x_k)$  converge vers  $\hat{x}$  à l'ordre  $p = \frac{(1+\sqrt{5})}{2} \simeq 1,618$  (le nombre d'or).*

Nous donnos une brève preuve du théorème : nous aurons besoin, d'abord, du lemme suivant :

**Lemme 6.4.1** Soit  $a_k$  une suite de réels positifs telle que

$$a_{k+1} \leq a_k a_{k-1}$$

Alors il existe  $c > 0$  tel que

$$a_k \leq c a_0^{p^k}.$$

où  $p = \frac{1+\sqrt{5}}{2}$  est le nombre d'or.

**Preuve** Posons  $r_k = \ln(a_k)$ . Nous avons

$$r_{k+1} \leq r_k + r_{k-1}$$

Il est facile de prouver que  $r_k$  est majorée par le terme général de la suite

$f_{k+1} = f_k + f_{k-1}$ ,  $f_0 = r_0$ ,  $f_1 = 1$  (suite de Fibonacci), suite dont le terme général vaut

$$f_k = a p^k + b \frac{1}{p^k}.$$

Le résultat en découle et d'où la preuve du théorème.

## 6.4.4 Méthode du point fixe

### Principe

Cette méthode que nous allons étudier utilise le fait que le problème  $f(x) = 0$  peut se ramener au problème équivalent  $g(x) - x = 0$  pour lequel on a le résultat suivant

**Théorème 6.4.2** Soit  $[a, b]$  un intervalle non vide de  $\mathbb{R}$  et  $g$  une application continue de  $[a, b]$  dans lui-même. Alors, il existe un point  $x^* \in [a, b]$ , appelé point fixe de la fonction  $g$ , vérifiant  $g(x^*) = x^*$ .

**Preuve** Posons  $f(x) = g(x) - x$ . On a alors  $f(a) = g(a) - a \geq 0$  et  $f(b) = g(b) - b \leq 0$ , puisque  $g(x) \in [a, b]$  pour tout  $x \in [a, b]$ . Par conséquent, la fonction  $f$ , continue sur  $[a, b]$ , est telle que  $f(a) f(b) \leq 0$ . Le théorème des valeurs intermédiaires assure alors l'existence d'un point  $x^* \in [a, b]$  tel que  $0 = f(x^*) = g(x^*) - x^*$ .

**Définition** (Méthode de point fixe)

Supposons que  $x^* \in \mathbb{R}$  soit un point fixe de  $g$ . La méthode de point fixe consiste en la construction d'une suite  $(x_k)_{k \in \mathbb{N}}$  définie par récurrence comme suit

$$\begin{cases} x_0 & \text{donné,} \\ x_{k+1} = g(x_k) & \forall k \in \mathbb{N}. \end{cases}$$

Naturellement une telle suite n'est pas forcément convergente. Par contre, si elle converge, c'est-à-dire si la suite  $x_k$  a une limite que nous notons  $x^*$ , et si  $g$  est continue, alors cette limite est nécessairement un point fixe de  $g$  puisque

$$x^* = \lim_{k \rightarrow \infty} x_{k+1} = \lim_{k \rightarrow \infty} g(x_k) = g(\lim_{k \rightarrow \infty} x_k) = g(x^*).$$

**Etude de convergence**

**Théorème 6.4.3** *Soit une fonction  $g : [a, b] \rightarrow \mathbb{R}$ . On se donne  $x_0 \in [a, b]$  et on considère la suite  $x_{k+1} = g(x_k)$  pour  $k \geq 0$ . Si les deux conditions suivantes sont satisfaites :*

1. *condition de stabilité :  $g(x) \in [a, b]$  pour tout  $x \in [a, b]$*
2. *condition de contraction stricte : il existe  $K \in [0, 1[$  tel que*

$$|g(x) - g(y)| < K |x - y| \text{ pour tout } x, y \in [a, b],$$

alors

- $g$  est continue
- $g$  a un et un seul point fixe  $x^* \in [a, b]$
- la suite  $x_{k+1} = g(x_k)$  converge vers  $x^*$  pour tout choix de  $x_0 \in [a, b]$ .

**Preuve**

**Continuité** La condition de contraction stricte entraîne que  $g$  est continue puisque, si on prend une suite  $(y_k)_{k \in \mathbb{N}} \in [a, b]$  qui converge vers un élément

$$x \text{ de } [a, b], \text{ alors nous avons } |g(x) - g(y_n)| \leq K |x - y_n|$$

et par suite  $\lim_{k \rightarrow \infty} g(y_k) = g(x)$ .

**Existence** Prouvons l'existence d'un point fixe de  $g$ .

La fonction  $f(x) = g(x) - x$  est continue dans  $[a, b]$  et, grâce à la condition de stabilité, on a  $f(a) = g(a) - a \geq 0$  et  $f(b) = g(b) - b \leq 0$ . En appliquant le théorème des valeurs intermédiaires, on en déduit que  $f$  a au moins une racine dans  $[a, b]$ , i.e.  $g$  a au moins un point fixe dans  $[a, b]$ .

**Unicité** L'unicité du point fixe découle de la condition de contraction stricte. En effet, si on avait deux points fixes distincts  $x_1^*$  et  $x_2^*$ , alors

$$|x_1^* - x_2^*| = |g(x_1^*) - g(x_2^*)| \leq K |x_1^* - x_2^*| < |x_1^* - x_2^*|$$

ce qui est impossible. d'où  $x_1^* = x_2^*$ .

**Convergence** Prouvons que la suite  $x_k$  converge vers l'unique point fixe  $x^*$  quand  $k$  tend vers  $+\infty$  pour toute donnée initiale  $x_0 \in [a, b]$ . On a

$$0 \leq |x_{k+1} - x^*| = |g(x_k) - g(x^*)| \leq K |x_k - x^*|$$

où  $K < 1$  est la constante de contraction. En itérant  $k + 1$  fois cette relation on obtient

$$|x_{k+1} - x^*| \leq K^{k+1} |x_0 - x^*|,$$

i.e, pour tout  $k \geq 0$

$$\frac{|x_{k+1} - x^*|}{|x_0 - x^*|} \leq K^{k+1}.$$

En passant à la limite quand  $k$  tend vers  $+\infty$  on obtient  $|x_{k+1} - x^*|$  tend vers 0.

Nous avons un résultat plus précis si  $g$  est  $C^1$ .

**Théorème 6.4.4** Soit  $g$  une fonction continue sur l'intervalle  $[a, b]$  qui vérifie

$$\forall x \in [a, b] : g(x) \in [a, b] \Leftrightarrow a \leq g(x) \leq b, \quad \forall x \in [a, b]$$

Alors la fonction  $g$  admet un point fixe dans l'intervalle  $[a, b]$ .

Si de plus la fonction  $g$  est de classe  $C^1([a, b])$  et qu'il existe une constante  $\rho \in ]0, 1[$  tel que :

$$|g'(x)| \leq \rho \quad \forall x \in [a, b]$$

Alors  $g$  admet un point fixe unique dans  $[a, b]$  et la suite  $x_k$  converge vers  $x^*$  pour tout  $x_0 \in [a, b]$ . De plus, nous avons

$$\lim_{k \rightarrow +\infty} \frac{x_{k+1} - x^*}{x_k - x^*} = g'(x^*).$$

**Preuve**

La dernière hypothèse implique que  $g$  est contractante stricte, c'est une conséquence du théorème des accroissement finis. En effet, pour tout  $x, y$  de  $[a, b]$ , il existe un  $c \in [a, b]$  tel que

$$g(x) - g(y) = g'(c)(x - y),$$

il en découle que

$$|g(x) - g(y)| < \rho |x - y|.$$

Les hypothèses du théorème 6.4.3 sont satisfaites, nous sommes assurés de l'existence et de l'unicité du point fixe  $x^*$  de  $g$  ainsi que la convergence de la suite  $x_k$ .

Pour établir le résultat de vitesse de convergence, En vertu du théorème des accroissement finis il existe, pour tout entier naturel  $k$ , un réel  $c_k$  tel que  $x_k < c_k < x^*$  et on a

$$x_{k+1} - x^* = g(x_k) - g(x^*) = g'(c_k)(x_k - x^*).$$

La suite  $x_k$  converge vers  $x^*$  ainsi  $c_k$  converge vers  $x^*$  et puis que  $g$  est  $C^1([a, b])$ , il s'ensuit de la dernière inégalité

$$\lim_{k \rightarrow +\infty} \frac{x_{k+1} - x^*}{x_k - x^*} = \lim_{k \rightarrow +\infty} g'(c_k) = g'(x^*).$$

**Critère d'arrêt**

Soit  $g : [a, b] \rightarrow \mathbb{R}$  satisfaisant aux hypothèses du théorème 6.4.4.

On désire avoir une estimation du nombre d'opérations nécessaires pour atteindre une précision donnée  $\varepsilon$ .

Soit  $x_k$  la suite des approximations définie par

$$x_{k+1} = g(x_k), \quad k = 0, 1, \dots$$

Pour  $k, n \in \mathbb{N}, n > k$  :

$$|x_k - x_n| = |g(x_{k-1}) - g(x_{n-1})| \leq \rho |x_{k-1} - x_{n-1}|$$

et par récurrence

$$|x_k - x_n| \leq \rho |x_{k-1} - x_{n-1}| \leq \rho^2 |x_{k-2} - x_{n-2}| \dots \leq \rho^k |x_0 - x_{n-k}| \leq \rho^n |b - a|.$$

En passant à la limite sur  $n$ , on obtient :

$$|x_k - x^*| \leq \rho^n (b - a).$$

Cette relation fournit donc une estimation de l'erreur commise à l'étape  $n$  du calcul et un moyen de trouver le nombre d'opérations nécessaires pour atteindre une précision donnée  $\varepsilon$ .

$$\begin{aligned} \rho^n |b - a| \leq \varepsilon &\Rightarrow \rho^n \leq \frac{\varepsilon}{|b - a|} \Rightarrow n \log(\rho) \leq \log(\varepsilon) - \log(b - a) \\ &\Rightarrow n \geq \frac{\log(\varepsilon) - \log(b - a)}{\log(\rho)}. \end{aligned}$$

**Remarque 6.4.2** Pour évaluer l'erreur de la  $n$ -ième approximation  $x_n$  on peut utiliser

la formule :  $|x^* - x_n| \leq |x_n - x_{n-1}| \leq \varepsilon$ .

## 6.5 Exercices

### Exercice 1

A l'aide de la méthode de dichotomie, résoudre l'équation  $x^3 + 2x - 7 = 0$  avec deux décimales exactes.

### Exercice 2

Calculer la racine de l'équation  $x^3 - 2x^2 - 5 = 0$

par la méthode de Newton avec une précision de 5 C.S.E.

Une fois la racine obtenue, calculer  $|e_n|$  et  $\left| \frac{e_n}{e_{n-1}} \right|$ .

Obtenir expérimentalement le taux de convergence de la méthode.

### Exercice 3

Trouver la racine positive comprise dans l'intervalle  $]1, 1.7[$  de l'équation :

$$x^4 - 2x - 4 = 0$$

avec une précision de deux décimales exactes, en utilisant la méthode de la sécante.

### Exercice 4

Utiliser la méthode du point fixe pour trouver la racine de l'équation

$$x^3 - x - 1 = 0$$

comprise dans l'intervalle  $]1, 2[$ , avec deux décimales exactes.

### 6.5.1 Réponses

#### Exercice 1

L'intervalle de séparation des racines exactes peut être trouver en construisant les courbes représentatives des fonctions  $y_1 = x^3$  et  $y_2 = -2x + 7$

( $f(x) = y_1 - y_2$ ) et chercher ensuite le point d'intersection de ces deux courbes. Vous trouvez que le seul point d'intersection se trouve dans  $]1, 2[$  vous pouvez prendre  $a = 1$  et  $b = 2$ .

$f(1) = -4 < 0$ ,  $f(2) = 5 > 0$  d'après le théorème des valeurs intermédiaires il y a existence de la solution de l'équation donnée dans  $]1, 2[$ , on notera cette solution  $\hat{x}$ .

Calculons le nombre d'itérations pour avoir la précision  $\varepsilon = 0.5 \cdot 10^{-2}$ .

On sait que  $k \geq \frac{\log \frac{b-a}{\varepsilon}}{\log 2} = \frac{\log \frac{1}{0.5 \cdot 10^{-2}}}{\log 2} \simeq 7,6$ , on prendra  $k = 8$ .

Soit  $x_1 = \frac{1+2}{2} = \frac{3}{2} = 1.5$  avec  $f(x_1) = -0.625 < 0$ , ainsi la racine cherchée se trouve dans  $]1.5, 2[$ .

On prend  $x_2 = \frac{1.5+2}{2} = 1.75$  avec  $f(x_2) = 1.313 > 0$  d'où la racine  $\hat{x}$  cherchée est dans  $]1.5, 1.75[$ .

On prend  $x_3 = \frac{1.5+1.75}{2} = 1.625$  avec  $f(x_3) = 0.296 > 0$  d'où la racine  $\hat{x}$  cherchée est dans  $]1.5, 1.625[$ .

On prend  $x_4 = \frac{1.5+1.625}{2} = 1.5625$  avec  $f(x_4) = -0.176 < 0$  d'où la racine  $\hat{x}$  cherchée est dans  $]1.5625, 1.625[$ .

On prend  $x_5 = \frac{1.562+1.625}{2} = 1.594$  avec  $f(x_5) > 0$  d'où la racine  $\hat{x}$  cherchée est dans  $]1.5625, 1.594[$ .

On prend  $x_6 = \frac{1.562+1.594}{2} = 1.578$  avec  $f(x_6) > 0$  d'où la racine  $\hat{x}$  cherchée est dans  $]1.5625, 1.578[$ .

On prend  $x_7 = \frac{1.562+1.578}{2} = 1.57$  avec  $f(x_7) > 0$  d'où la racine  $\hat{x}$  cherchée est dans  $]1.5625, 1.57[$ .

On prend  $x_8 = \frac{1.562+1.57}{2} = 1.566$  avec  $f(x_8) < 0$  d'où la racine  $\hat{x}$  cherchée est dans  $]1.566, 1.57[$ .

On voit que la racine cherchée calculée avec deux décimales exactes est

$$x = 1.57 \pm 0.01.$$

#### Exercice 2

On a  $f(0) = -5$  et  $f(10) = 795$  par le théorème des V.I la racine  $\hat{x}$  de l'équation donnée se trouve dans  $]0, 10[$ .

Réduisant l'intervalle obtenu, étant donné que  $f(3) = 4 > 0$ ,

d'après le T.V.I  $\exists \hat{x} \in ]0, 3[$  tel que  $f(\hat{x}) = 0$ .

Il faut noter que dans cet intervalle toutes les conditions de la méthode de Newton sont réunies.

Nous pouvons adopter comme approximation initiale  $x_0 = 3$ .

La suite est générée par la formule de Newton  $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ .

Ainsi on obtient  $x_1 = 2.73333$ .

La précision est de 5 C.S.E alors le dernier C.S.E est du rang -4 et donc  $\varepsilon = 0.5 \cdot 10^{-4}$ , puis on travaille avec 5 décimales on obtient

$x_1 = 2.73334$ ,  $x_2 = 2.69163$ ,  $x_3 = 2.69064$ ,  $x_4 = 2.69064$ ,

on s'arrête là car  $|x_4 - x_3| = 7.43988 \cdot 10^{-6} < \varepsilon$ . Donc  $\hat{x} = x_4$ .

	$e_n$	$e_n^2$	$\frac{e_n}{e_{n-1}}$	$\frac{e_n}{e_{n-1}^2}$
$x_0$	0.30936	$0.957 \cdot 10^{-2}$	/	/
$x_1$	$0.4269 \cdot 10^{-1}$	$0.18224 \cdot 10^{-2}$	0.13799	0.446
$x_2$	$0.98 \cdot 10^{-3}$	$0.96 \cdot 10^{-6}$	$0.22996 \cdot 10^{-1}$	0.5377
$x_3$	0	0	0	0
$x_4$	0	0	/	/

On a  $e_n \rightarrow 0.98 \cdot 10^{-3} \simeq 0$  et  $\frac{e_n}{e_{n-1}^2} \rightarrow 0.5377$ , on a bien  $\lim \frac{|x_{k+1} - \hat{x}|}{|x_k - \hat{x}|^2} = C < +\infty$ .

On conclut que la convergence est quadratique.

# Bibliographie

- [1] O. Axelsson [1994], Iterative Solution Methods. Cambridge University Press, Cambridge.
- [2] J. Barager [1977], Introduction à l'analyse numérique, Hermann.
- [3] N. Ben Salah, Z. Tourki [2005], Introduction aux méthodes numériques appliquées et à leur programmation. Centre de publication universitaire, Tunis.
- [4] C. Brezinski, M. Redivo-Zaglia [2006], Méthodes numériques itératives. Algèbre linéaire et non linéaire. Ellipses, Paris.
- [5] G. Charet [1975], Cours d'analyse numérique, Sedes- Informatique.
- [6] P.G. Ciarlet [1982], Introduction à l'analyse numérique et à l'optimisation, Masson, Paris.
- [7] Crouzeix, Mignot, Analyse numérique des équations différentielles, Masson, Paris.
- [8] J.P. Demailly [1991], Analyse numérique et équations différentielles. Presses Universitaires de Grenoble.
- [9] E. Durand [1972], Solutions numériques des équations algébriques, tomes 1 et 2, Masson, Paris.
- [10] G. Faccanoni, Analyse numérique, Recueil d'exercices corrigés et aide mémoire, [http:// faccanoni.univ-tln.fr/enseignement.html](http://faccanoni.univ-tln.fr/enseignement.html)
- [11] M.Fellah, N.H. Allal, Exercices corrigés en Analyse numérique élémentaire, Office des publications Universitaires.[2003].
- [12] A. Fortin [1995], Analyse numérique pour ingénieurs, éditions de l'école polytechnique de Montréal.
- [13] G. Hacques [1972], Algorithmes numériques, cours- tome 1- Collection U. Colin.

- [14] R. Herbin, Cours d'Analyse numérique, <https://www.cmi.univ-mrs.fr/~herbin/anamat.html>
- [15] F. Jędrzejwski [2005], Introduction aux méthodes numériques, deuxième édition, Springer-Verlag France Paris.
- [16] G. Koepfler, Notes de cours, Méthodes numériques, [www.math-info.univ-paris5.fr/~gk/MN\\_S6](http://www.math-info.univ-paris5.fr/~gk/MN_S6).
- [17] M. Lakrib [2003], Cours d'analyse numérique, Office des publications universitaires, Alger.
- [18] P. Lascaux, R.Théodor [2005], Analyse numérique matricielle appliquée à l'art de l'ingénieur. 2.Méthodes itératives. Dunod, Paris.
- [19] G. Legendre [2016], Introduction à l'analyse numérique et au calcul scientifique, [https://www.ceremade.dauphine.fr/~legendre/enseignement/cours\\_ananum\\_dauphine.pdf](https://www.ceremade.dauphine.fr/~legendre/enseignement/cours_ananum_dauphine.pdf)
- [20] Legras [1971], Initiation à l'analyse numérique, Dunod.
- [21] J-L. Matret [2001], Aide mémoire Analyse Numérique, <http://luminath.univ-mrs.fr/~jlm/cours/analnum>
- [22] P. Pelletier, Techniques numériques appliquées au calcul scientifique, Masson.
- [23] A. Quarteroni, R.Sacco, F.Saleri [2000], Méthodes numériques pour le calcul scientifique. Programme en MATLAB. Springer-Verlag France Paris.
- [24] P. Spiteri [2000], Introduction à l'analyse numérique, Polycopié INP Toulouse, ENSEEIHT.
- [25] M. R. Temam [1970], Analyse numérique, Collection SUP le mathématicien. Presses Universitaires de France Université Paris-Sud, Centre d'Orsay.