

Pour plus de livres visitez notre site web :

biblio-scientifique.com






*Bibliothèque
scientifique*

biblio-scientifique.com

André
Fortin


$$\begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & x_1^3 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & x_2^3 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \dots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}$$

Analyse numérique pour ingénieurs

Quatrième édition

PRESSES INTERNATIONALES
POLYTECHNIQUE



Extrait de la publication

André
Fortin

Professeur
à l'Université
Laval

Analyse numérique pour ingénieurs

Quatrième édition



PRESSES INTERNATIONALES
POLYTECHNIQUE

Analyse numérique pour ingénieurs, quatrième édition

André Fortin

Couverture : Cyclone Design

Pour connaître nos distributeurs et nos points de vente, veuillez consulter notre site Web à l'adresse suivante : www.polymtl.ca/pub

Courriel des Presses internationales Polytechnique : pip@polymtl.ca

Nous reconnaissons l'aide financière du gouvernement du Canada par l'entremise du Fonds du livre du Canada pour nos activités d'édition.

Gouvernement du Québec – Programme de crédit d'impôt pour l'édition de livres – Gestion SODEC.

Tous droits réservés

© Presses internationales Polytechnique, 2011

On ne peut reproduire ni diffuser aucune partie du présent ouvrage, sous quelque forme ou par quelque procédé que ce soit, sans avoir obtenu au préalable l'autorisation écrite de l'éditeur.

Dépôt légal : 4^e trimestre 2011

Bibliothèque et Archives nationales du Québec

Bibliothèque et Archives Canada

ISBN 978-2-553-01622-6 (version imprimée)

ISBN 978-2-553-01624-0 (version pdf)

Imprimé au Canada

À mon épouse Marie
et mes fils Michel
et Jean-Philippe

Une pensée spéciale
pour mon père et
pour Line et Marc
ainsi que pour ma mère
qui nous a quittés

Avant-propos à la quatrième édition

L'analyse numérique et les méthodes numériques en général poursuivent leur essor considérable amorcé depuis plusieurs années. La vaste majorité des facultés de génie offrent au moins un cours d'introduction à cette discipline, suivi très souvent d'un second cours plus avancé.

Ce manuel reflète mon expérience comme professeur d'analyse numérique aux ingénieurs, d'abord à l'École Polytechnique de Montréal et, par la suite, à l'Université Laval à Québec. Chaque année, plus de 500 étudiants de ces deux institutions suivent un tel cours qui propose un survol des principales méthodes numériques élémentaires et couvre plus particulièrement les sujets suivants :


- analyse d'erreurs ;
- racines d'une équation algébrique ;
- systèmes d'équations linéaires et non linéaires ;
- méthodes itératives et systèmes dynamiques ;
- interpolation ;
- différentiation et intégration numériques ;
- équations différentielles ordinaires.

L'approche pédagogique de ce manuel repose toujours sur une compréhension profonde des méthodes plutôt que sur l'aspect calculatoire. Cela signifie que les exemples choisis cherchent avant tout à illustrer différents aspects des méthodes et à souligner leurs avantages et leurs inconvénients. Cette approche est justifiée en partie par le fait que de plus en plus d'ingénieurs utilisent des outils logiciels commerciaux. L'objectif de ce manuel est donc de faire des étudiants des utilisateurs intelligents, en ce sens qu'ils sauront exactement à quoi s'attendre de chaque méthode et qu'ils seront en mesure de valider leurs résultats.

Le *prix francophone du livre et de la technologie*, ou *prix Roberval*, décerné par l'Université de Compiègne en France, est venu récompenser mes efforts en 1996. Ce fut une belle récompense, mais il demeure que rien ne vaut les commentaires des principaux intéressés, les étudiants. Bien entendu, on ne peut plaire à tous, et cet ouvrage ne fait pas exception, mais j'ai quand même senti un accueil largement favorable. C'est un encouragement à poursuivre le travail entrepris, à améliorer la présentation et à rechercher d'autres exemples et applications.

C'est encore le but visé par cette quatrième édition qui ne contient pas de modifications majeures par rapport à la troisième. Quelques sections ont été réécrites, dans l'espoir d'en améliorer la présentation et d'en faciliter la

compréhension. C'est le cas notamment pour la section sur la transformée de Fourier rapide et celle sur la méthode de Runge-Kutta-Fehlberg. J'ai modifié la numérotation des définitions, exemples, remarques, etc. Ainsi, la remarque 1.2 précédera l'exemple 1.3 et suivra forcément la définition 1.1. Le lecteur devrait s'y retrouver plus facilement dans le texte.

Certains exercices plus élaborés sont maintenant identifiés par le symbole  et nécessitent l'emploi de l'ordinateur. Pour résoudre ces exercices, la plupart des méthodes décrites sont disponibles sous forme de programmes en langage Matlab à l'adresse Internet suivante :

www.giref.ulaval.ca/~afortin/

Ces programmes constituent un complément fort utile pour explorer les possibilités et limites des différentes méthodes présentées. L'aide en ligne permettra au lecteur de reprendre certains des exemples décrits dans ce manuel et ainsi de s'initier à l'utilisation des différentes méthodes qui y sont décrites. On peut également s'en servir comme outil pour d'éventuels travaux pratiques en laboratoire ou pour des devoirs.

En terminant, j'aimerais remercier toutes les personnes qui ont contribué à la réalisation de ce manuel. Mme Carole Burney-Vincent, M. Gilles Savard M. et M. Robert Roy de l'École Polytechnique de Montréal ont patiemment lu et commenté plusieurs chapitres de la première édition.

Plusieurs personnes ont contribué de près ou de loin aux éditions subséquentes. À l'Université Laval, M. Michel Fortin m'a fortement incité à inclure de nouvelles sections, notamment sur les NURBS, tandis que messieurs Roger Pierre et José Urquiza ont eu la patience de relire et de commenter plusieurs des nouveaux ajouts. Je note aussi la contribution de M. Robert Guénette qui m'a proposé quelques nouveaux sujets ainsi que de nombreux exercices.

Enfin, je ne peux passer sous silence l'appui inconditionnel de mon épouse Marie et de mes fils Michel et Jean-Philippe qui ont dû, entre autres choses, subir mes absences fréquentes lors de la rédaction et de la mise en pages finale de cet ouvrage. Lorsque j'en ai commencé la rédaction, je ne me serais jamais douté que mes deux fils auraient à l'utiliser eux-mêmes dans l'un de leurs cours...

Que chacun et chacune veuillent bien trouver ici l'expression de ma plus profonde reconnaissance.

André Fortin

Table des matières

1	Analyse d'erreurs	1
1.1	Introduction	1
1.2	Erreurs de modélisation	4
1.3	Représentation des nombres sur ordinateur	7
1.3.1	Représentation des entiers signés	9
1.3.2	Représentation des nombres réels	11
1.3.3	Erreurs dues à la représentation	12
1.4	Norme IEEE-754	14
1.4.1	Exceptions	16
1.4.2	Nombres non normalisés	17
1.5	Arithmétique flottante	18
1.5.1	Opérations élémentaires	19
1.5.2	Opérations risquées	22
1.5.3	Évaluation des polynômes	26
1.6	Erreurs de troncature	27
1.6.1	Développement de Taylor en une variable	28
1.6.2	Développement de Taylor en plusieurs variables	34
1.6.3	Propagation d'erreurs dans le cas général	35
1.7	Évaluation de la fonction e^x	38
1.8	Exercices	42
2	Équations non linéaires	49
2.1	Introduction	49
2.2	Méthode de la bisection	50
2.3	Méthodes des points fixes	54
2.3.1	Convergence de la méthode des points fixes	58
2.3.2	Interprétation géométrique	62
2.3.3	Extrapolation d'Aitken	65
2.4	Méthode de Newton	67
2.4.1	Interprétation géométrique	69
2.4.2	Analyse de convergence	69
2.4.3	Cas des racines multiples	73
2.5	Méthode de la sécante	77
2.6	Applications	80
2.6.1	Modes de vibration d'une poutre	81
2.6.2	Premier modèle de viscosité	83
2.7	Exercices	86

3	Systèmes d'équations algébriques	95
3.1	Introduction	95
3.2	Systèmes linéaires	95
3.3	Opérations élémentaires sur les lignes	100
3.3.1	Multiplication d'une ligne par un scalaire	101
3.3.2	Permutation de deux lignes	102
3.3.3	Opération $(\vec{l}_i \leftarrow \vec{l}_i + \lambda \vec{l}_j)$	103
3.4	Élimination de Gauss	104
3.5	Décomposition LU	108
3.5.1	Principe de la méthode	108
3.5.2	Décomposition de Crout	109
3.5.3	Décomposition LU et permutation de lignes	115
3.5.4	Factorisation de Choleski	119
3.5.5	Les systèmes tridiagonaux	122
3.6	Calcul de la matrice inverse A^{-1}	124
3.7	Effets de l'arithmétique flottante	127
3.8	Conditionnement d'une matrice	132
3.9	Systèmes non linéaires	142
3.10	Applications	149
3.10.1	Calcul des tensions dans une ferme	149
3.10.2	Deuxième modèle de viscosité	152
3.10.3	Réseau de distribution d'eau	154
3.11	Exercices	157
4	Méthodes itératives et systèmes dynamiques discrets	167
4.1	Introduction	167
4.2	Application quadratique	167
4.3	Méthodes des points fixes : cas complexe	176
4.4	Rappels sur les valeurs et vecteurs propres	181
4.4.1	Méthode des puissances	183
4.4.2	Méthode des puissances inverses	186
4.5	Méthodes des points fixes en dimension n	187
4.6	Méthodes itératives pour les systèmes linéaires	195
4.6.1	Méthode de Jacobi	197
4.6.2	Méthode de Gauss-Seidel	202
4.7	Exercices	205
5	Interpolation	207
5.1	Introduction	207
5.2	Matrice de Vandermonde	209
5.3	Interpolation de Lagrange	210
5.4	Polynôme de Newton	214
5.5	Erreur d'interpolation	221
5.6	Splines cubiques	229
5.6.1	Courbes de la forme $y = f(x)$	230
5.6.2	Splines paramétrées	238
5.7	Krigeage	242
5.7.1	Effet pépité	249
5.7.2	Courbes paramétrées	253

5.7.3	Cas multidimensionnel	256
5.8	Transformée de Fourier discrète	258
5.9	Introduction aux NURBS	270
5.9.1	B-splines	270
5.9.2	Génération de courbes	276
5.9.3	B-splines rationnelles non uniformes	276
5.9.4	Construction des coniques	278
5.9.5	Construction des surfaces	280
5.10	Exercices	282
6	Différentiation et intégration numériques	291
6.1	Introduction	291
6.2	Différentiation numérique	291
6.2.1	Dérivées d'ordre 1	292
6.2.2	Dérivées d'ordre supérieur	298
6.3	Extrapolation de Richardson	302
6.4	Intégration numérique	304
6.4.1	Formules de Newton-Cotes simples et composées	304
6.4.2	Méthode de Romberg	319
6.4.3	Quadratures de Gauss-Legendre	322
6.4.4	Intégration à l'aide des splines	330
6.5	Applications	332
6.5.1	Courbe des puissances classées	332
6.5.2	Puissance électrique d'un ordinateur	332
6.6	Exercices	334
7	Équations différentielles	341
7.1	Introduction	341
7.2	Méthode d'Euler explicite	343
7.3	Méthodes de Taylor	350
7.4	Méthodes de Runge-Kutta	354
7.4.1	Méthodes de Runge-Kutta d'ordre 2	354
7.4.2	Méthode de Runge-Kutta d'ordre 4	357
7.4.3	Contrôle de l'erreur	360
7.5	Méthodes à pas multiples	363
7.6	Systèmes d'équations différentielles	370
7.7	Équations d'ordre supérieur	373
7.8	Stabilité absolue	375
7.8.1	Quelques mots sur les méthodes implicites	380
7.9	Méthodes de tir	384
7.9.1	Équations linéaires	384
7.9.2	Équations non linéaires	394
7.10	Méthodes des différences finies	397
7.11	Applications	401
7.11.1	Problème du pendule	401
7.11.2	Systèmes de masses et de ressorts	405
7.11.3	Attracteur étrange de Lorenz	408
7.11.4	Défi du golfeur	411
7.12	Exercices	416

Réponses aux exercices du chapitre 1 425

Réponses aux exercices du chapitre 2 431

Réponses aux exercices du chapitre 3 438

Réponses aux exercices du chapitre 4 447

Réponses aux exercices du chapitre 5 449

Réponses aux exercices du chapitre 6 457

Réponses aux exercices du chapitre 7 464

Chapitre 1

Analyse d'erreurs

1.1 Introduction

Les cours traditionnels de mathématiques nous familiarisent avec des théories et des méthodes qui permettent de résoudre de façon *analytique* un certain nombre de problèmes. C'est le cas notamment des techniques d'intégration et de résolution d'équations algébriques ou différentielles. Bien que l'on puisse proposer plusieurs méthodes pour résoudre un problème donné, celles-ci conduisent à un même résultat, normalement exempt d'erreur.

C'est ici que l'analyse numérique se distingue des autres champs plus classiques des mathématiques. En effet, pour un problème donné, il est possible d'utiliser plusieurs techniques de résolution qui résultent en différents algorithmes.¹ Ces algorithmes dépendent de certains paramètres qui influent sur la précision du résultat. De plus, on utilise en cours de calcul des approximations plus ou moins précises. Par exemple, on peut remplacer une dérivée par une différence finie de façon à transformer une équation différentielle en une équation algébrique. Le résultat final et son ordre de précision dépendent des choix que l'on fait.

Une partie importante de l'analyse numérique consiste donc à contenir les effets des erreurs ainsi introduites, qui proviennent de trois sources principales :

- les erreurs de modélisation ;
- les erreurs de représentation sur ordinateur ;
- les erreurs de troncature.

Les erreurs de modélisation, comme leur nom l'indique, proviennent de l'étape de mathématisation du phénomène physique auquel on s'intéresse. Cette étape consiste à faire ressortir les causes les plus déterminantes du phénomène observé et à les mettre sous forme d'équations (différentielles le plus souvent). Si le phénomène observé est très complexe, il faut simplifier et négliger ses composantes qui paraissent moins importantes ou qui rendent la résolution numérique trop difficile. C'est ce que l'on appelle les *erreurs de modélisation*.

La deuxième catégorie d'erreurs est liée à l'utilisation de l'ordinateur. En effet, la représentation sur ordinateur (généralement binaire) des nombres

1. Le mot algorithme vient du mathématicien arabe Al-Khuwārizmī (VIII^e siècle après J.-C.) qui fut l'un des premiers à utiliser une séquence de calculs simples pour résoudre certaines équations quadratiques. Il est l'un des pionniers de l'al-jabr (l'algèbre).

introduit souvent des erreurs. Même infimes au départ, ces erreurs peuvent s'accumuler lorsqu'on effectue un très grand nombre d'opérations. Par exemple, la fraction $\frac{1}{3}$ n'a pas de représentation binaire finie, pas plus qu'elle ne possède de représentation décimale finie. On ne pourra donc pas représenter exactement cette fraction, ce qui introduit une erreur. Ces erreurs se propagent au fil des calculs et peuvent compromettre la précision des résultats.

Enfin, les *erreurs de troncature* proviennent principalement de l'utilisation du développement de Taylor, qui permet par exemple de remplacer une équation différentielle par une équation algébrique. Le développement de Taylor est le principal outil mathématique du numéricien. Il est primordial d'en maîtriser l'énoncé et ses conséquences.

Ce chapitre traite donc principalement d'erreurs numériques, et non des inévitables erreurs de programmation qui font, hélas, partie du quotidien du numéricien. Il devrait permettre au lecteur de mieux gérer les erreurs au sein des processus numériques afin d'être en mesure de mieux interpréter les résultats. Introduisons d'abord un peu de terminologie qui nous permettra éventuellement de quantifier les erreurs.

Définition 1.1

Soit x , un nombre, et x^* , une approximation de ce nombre. L'*erreur absolue* est définie par :

$$\Delta x = |x - x^*| \quad (1.1)$$

Définition 1.2

Soit x , un nombre, et x^* , une approximation de ce nombre. L'*erreur relative* est définie par :

$$E_r(x^*) = \frac{|x - x^*|}{|x|} = \frac{\Delta x}{|x|} \simeq \frac{\Delta x}{|x^*|} \quad (1.2)$$

De plus, en multipliant par 100 %, on obtient l'*erreur relative en pourcentage*.

En pratique, il est difficile d'évaluer les erreurs absolue et relative, car on ne connaît généralement pas la valeur exacte de x et l'on n'a que x^* . C'est pourquoi on utilise l'approximation $\Delta x/|x^*|$ pour l'erreur relative. Dans le cas de quantités mesurées expérimentalement dont on ne connaît que la valeur approximative, on dispose souvent d'une borne supérieure pour l'erreur absolue qui dépend de la précision des instruments de mesure utilisés. Cette borne est quand même appelée erreur absolue, alors qu'en fait on a :

$$|x - x^*| \leq \Delta x$$

ce qui peut également s'écrire :

$$x^* - \Delta x \leq x \leq x^* + \Delta x \quad (1.3)$$

et que l'on note parfois $x = x^* \pm \Delta x$. On peut interpréter ce résultat en disant que l'on a estimé la valeur exacte x à partir de x^* avec une incertitude de Δx de part et d'autre.

L'erreur absolue donne une mesure quantitative de l'erreur commise et l'erreur relative en mesure l'importance. Par exemple, si l'on fait usage d'un chronomètre dont la précision est de l'ordre du dixième de seconde, l'erreur absolue est bornée par 0,1 s. Mais est-ce une erreur importante ? Dans le contexte d'un marathon d'une durée approximative de 2 h 20 min, l'erreur relative liée au chronométrage est très faible :

$$\frac{0,1}{2 \times 60 \times 60 + 20 \times 60} = 0,000\,0119$$

et ne devrait pas avoir de conséquence sur le classement des coureurs. Par contre, s'il s'agit d'une course de 100 m d'une durée d'environ 10 s, l'erreur relative est beaucoup plus importante :

$$\frac{0,1}{10,0} = 0,01$$

soit 1 % du temps de course. Avec une telle erreur, on ne pourra vraisemblablement pas distinguer le premier du dernier coureur. Cela nous amène à parler de précision et de *chiffres significatifs* au sens de la définition suivante.

Définition 1.3

Si l'erreur absolue vérifie :

$$\Delta x \leq 0,5 \times 10^m$$

alors le chiffre correspondant à la m^{e} puissance de 10 est dit *significatif* et tous ceux à sa gauche, correspondant aux puissances de 10 supérieures à m , le sont aussi. On arrête le compte au dernier chiffre non nul. Il existe une exception à la règle. Si le chiffre correspondant à la m^{e} puissance de 10 est nul **ainsi** que tous ceux à sa gauche, on dit qu'il n'y a aucun chiffre significatif. Inversement, si un nombre est donné avec n chiffres significatifs, on commence à compter à partir du premier chiffre non nul à gauche et l'erreur absolue est inférieure à 0,5 fois la puissance de 10 correspondant au dernier chiffre significatif.

Remarque 1.4

En pratique, on cherchera à déterminer une borne pour Δx aussi petite que possible et donc la valeur de m la plus petite possible. ♦

Exemple 1.5

On obtient une approximation de π ($x = \pi$) au moyen de la quantité $\frac{22}{7}$ ($x^* = \frac{22}{7} = 3,142\,857\cdots$). On en conclut que :

$$\Delta x = \left| \pi - \frac{22}{7} \right| = 0,001\,26\cdots \simeq 0,126 \times 10^{-2}$$

Puisque l'erreur absolue est plus petite que $0,5 \times 10^{-2}$, le chiffre des centièmes est significatif et on a en tout 3 chiffres significatifs (3,14). ■

Exemple 1.6

Si l'on retient 3,1416 comme approximation de π , on a :

$$\Delta x = |\pi - 3,1416| \simeq 0,73 \times 10^{-5}$$

et l'erreur absolue est inférieure à $0,5 \times 10^{-4}$. Le chiffre correspondant à cette puissance de 10 (6) est significatif au sens de la définition, ainsi que tous les chiffres situés à sa gauche. Il est à remarquer que le chiffre 6 dans 3,1416 est significatif même si la quatrième décimale de π est un 5 (3,141 59...). L'approximation 3,1416 possède donc 5 chiffres significatifs. ■

Exemple 1.7

On a mesuré le poids d'une personne et trouvé 90,567 kg. On vous assure que l'appareil utilisé est suffisamment précis pour que tous les chiffres fournis soient significatifs. En vertu de la remarque précédente, puisque le dernier chiffre significatif correspond aux millièmes (milligrammes), cela signifie que :

$$\Delta x \leq 0,5 \times 10^{-3} \text{ kg}$$

En pratique, on conclut que $\Delta x = 0,5 \times 10^{-3} \text{ kg}$. ■

1.2 Erreurs de modélisation

La première étape de la résolution d'un problème, et peut-être la plus délicate, consiste à modéliser le phénomène observé. Il s'agit en gros d'identifier tous les facteurs internes et externes qui influent (ou que l'on soupçonne d'influer) sur les résultats. Dans le cas d'un phénomène physique, on fait l'inventaire des forces en présence : gravitationnelle, de friction, électrique, etc. On a par la suite recours aux lois de conservation de la masse, de l'énergie, de la quantité de mouvement et à d'autres principes mathématiques pour traduire l'influence de ces différents facteurs sous forme d'équations. Le plus souvent, on obtient des équations différentielles ou aux dérivées partielles.

L'effort de modélisation produit en général des systèmes d'équations complexes qui comprennent un grand nombre de variables inconnues. Pour réussir à les résoudre, il faut simplifier certaines composantes et négliger les moins importantes. On fait alors une première erreur de modélisation.

De plus, même bien décomposé, un phénomène physique peut être difficile à mettre sous forme d'équations. On introduit alors un modèle qui décrit au mieux son influence, mais qui demeure une approximation de la réalité. On commet alors une deuxième erreur de modélisation. Illustrons cette démarche à l'aide d'un exemple.

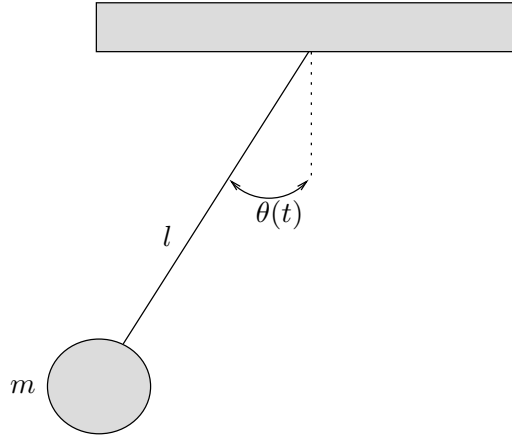


Figure 1.1 – Problème du pendule

Exemple 1.8

Le problème du pendule est connu depuis très longtemps (voir par exemple Simmons, réf. [33]). Une masse m est suspendue à une corde de longueur l (fig. 1.1). Au temps $t = 0$, on suppose que l'angle $\theta(0)$ entre la corde et la verticale est θ_0 et que sa vitesse angulaire $\theta'(0)$ est θ'_0 . Les forces en présence sont, d'une part, la gravité agissant sur la masse et la corde et, d'autre part, la friction de l'air agissant sur tout le système.

Suivant la deuxième loi de Newton, la force due à l'accélération tangentielle $ml\theta''(t)$ est équilibrée par la composante tangentielle de l'accélération gravitationnelle $mg \sin(\theta(t))$ et par la force de friction F_f qui s'oppose au mouvement. On a alors :

$$ml\theta''(t) = -mg \sin(\theta(t)) - F_f$$

Pour connaître comment se comporte la force de friction F_f en fonction de $\theta(t)$, il faut recourir à des mesures expérimentales, qui démontrent que la friction est à peu près proportionnelle à la vitesse $l\theta'(t)$ avec un coefficient de proportionnalité c_f . Il est important de remarquer que cette loi est approximative et que le coefficient c_f est obtenu avec une précision limitée. Tout cela entraîne des erreurs de modélisation.

On obtient une équation différentielle du second ordre :

$$\begin{cases} \theta''(t) &= \frac{-c_f \theta'(t)}{m} - \frac{g \sin(\theta(t))}{l} \\ \theta(0) &= \theta_0 \\ \theta'(0) &= \theta'_0 \end{cases} \quad (1.4)$$

L'équation différentielle 1.4 est non linéaire et l'on démontre qu'elle ne possède pas de solution analytique simple.

Une brève analyse montre qu'il est raisonnable de négliger la friction de l'air, car les vitesses de mouvement sont faibles. Il s'agit encore d'une erreur de modélisation, qui paraît acceptable à cette étape de l'analyse. Si les résultats se révèlent insatisfaisants, il faudra revenir en arrière et identifier, parmi les

Chapitre 2

Équations non linéaires

2.1 Introduction

Le numéricien est souvent confronté à la résolution d'équations algébriques de la forme :

$$f(x) = 0 \tag{2.1}$$

et ce, dans toutes sortes de contextes. Introduisons dès maintenant la terminologie qui nous sera utile pour traiter ce problème.

Définition 2.1

Une valeur de x solution de $f(x) = 0$ est appelée une *racine* ou un *zéro* de la fonction $f(x)$ et est notée r .

Nous avons tous appris au secondaire comment résoudre l'équation du second degré :

$$ax^2 + bx + c = 0$$

dont les deux racines sont :

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Certains ont également vu comment calculer les racines d'une équation du troisième ordre et se souviennent que la formule est beaucoup plus complexe. On peut aussi obtenir une formule générale pour le quatrième degré. Par contre, on ignore souvent qu'il n'existe pas de formule permettant de trouver les racines des polynômes de degré plus grand ou égal à 5. Non pas que les mathématiciens ne l'aient pas encore trouvée, mais Abel¹ et par la suite Galois² ont démontré que cette formule n'existe pas.

Puisqu'il n'existe pas de formule générale pour des fonctions aussi simples que des polynômes, il est peu probable que l'on puisse résoudre analytiquement l'équation 2.1 dans tous les cas qui nous intéressent. Il faudra donc recourir aux

1. Le mathématicien norvégien Niels Henrik Abel (1802-1829) fut à l'origine de la première démonstration de cet important résultat.

2. Le mathématicien Évariste Galois (1811-1832) fut tué dans un duel à l'âge de 21 ans, non sans avoir eu le temps d'apporter une contribution considérable à la théorie des groupes.

méthodes numériques. Dans ce qui suit, nous présentons plusieurs techniques de résolution, chacune ayant ses avantages et ses inconvénients. Nous tâcherons de les mettre en évidence de façon à tirer le meilleur parti de chacune des méthodes proposées.

Il faudra également se souvenir des enseignements du chapitre précédent pour éviter de développer des algorithmes numériquement instables.

2.2 Méthode de la bisection

La méthode de la bisection repose sur une idée toute simple : en général, de part et d'autre d'une solution de l'équation 2.1, une fonction continue $f(x)$ change de signe et passe du positif au négatif ou vice versa (fig. 2.1). De toute évidence, ce n'est pas toujours le cas puisque la fonction $f(x)$ peut aussi être tangente à l'axe des x . Nous reviendrons plus loin sur ces situations particulières (fig. 2.2).

Supposons pour l'instant qu'il y ait effectivement un changement de signe autour d'une racine r de $f(x)$. Nous nous occuperons des cas pathologiques un peu plus tard. Soit $[x_1, x_2]$, un intervalle ayant un changement de signe, c'est-à-dire :

$$f(x_1) \times f(x_2) < 0 \quad (2.2)$$

On pose alors :

$$x_m = \frac{x_1 + x_2}{2}$$

qui est bien sûr le point milieu de l'intervalle. Il suffit alors de déterminer, entre les intervalles $[x_1, x_m]$ et $[x_m, x_2]$, celui qui possède encore un changement de signe. La racine se trouvera forcément dans cet intervalle. À la première itération de la figure 2.1, ce serait l'intervalle $[x_m, x_2]$, tandis qu'à la deuxième itération ce serait $[x_1, x_m]$. Cela nous amène à l'algorithme suivant.

Algorithme 2.2: Algorithme de la bisection

1. Étant donné un intervalle $[x_1, x_2]$ pour lequel $f(x)$ possède un changement de signe
2. Étant donné ϵ_a , le critère d'arrêt, et N , le nombre maximal d'itérations
3. Poser :

$$x_m = \frac{x_1 + x_2}{2}$$

4. Si $\frac{|x_2 - x_1|}{2|x_m|} < \epsilon_a$:
 - convergence atteinte
 - écrire la racine x_m
 - écrire $f(x_m)$
 - arrêt
5. Écrire $x_1, x_2, x_m, f(x_1), f(x_2), f(x_m)$
6. Si $f(x_1) \times f(x_m) < 0$, alors $x_2 = x_m$
7. Si $f(x_m) \times f(x_2) < 0$, alors $x_1 = x_m$

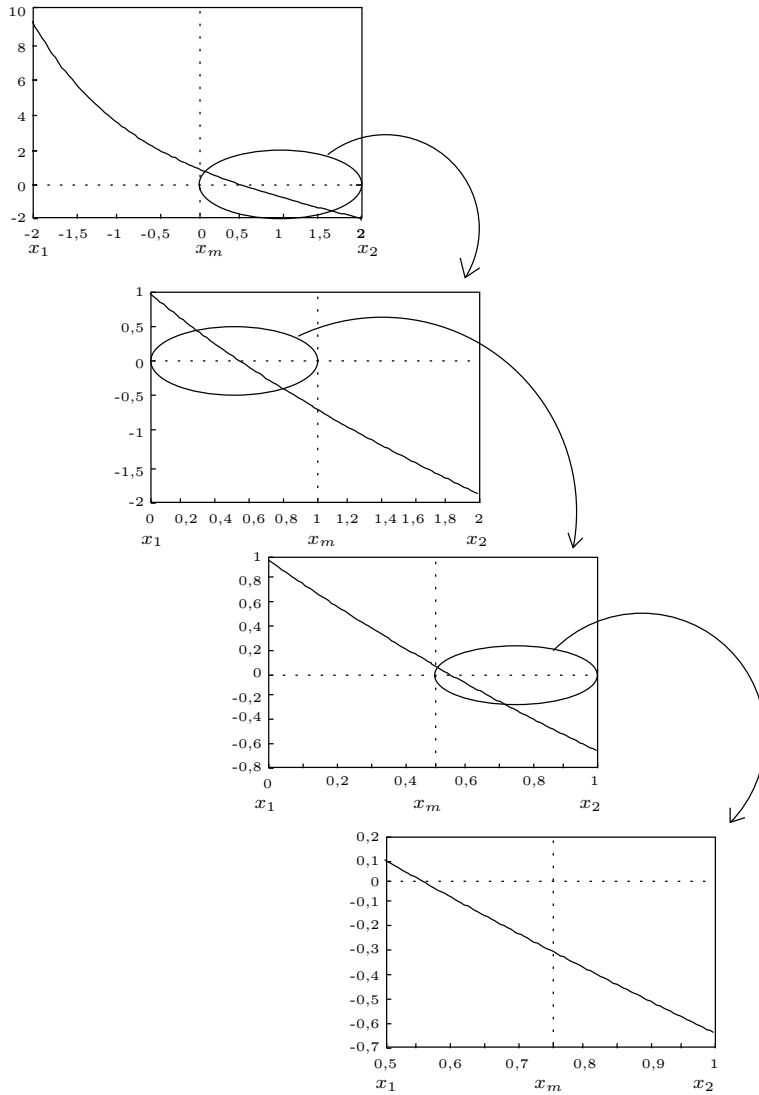


Figure 2.1 – Méthode de la bisection : $f(x) = e^{-x} - x$

- 8. Si le nombre maximal d'itérations N est atteint :
 - convergence non atteinte en N itérations
 - arrêt

9. Retour à l'étape 3



L'expression :

$$\frac{|x_2 - x_1|}{2|x_m|}$$

est une approximation de l'erreur relative. En effet, à l'étape 3 de l'algorithme de la bisection, la racine recherchée est soit dans l'intervalle $[x_1, x_m]$ ou dans l'intervalle $[x_m, x_2]$, qui sont tous deux de longueur $(x_2 - x_1)/2$, ce qui constitue une borne supérieure de l'erreur absolue. En divisant par x_m , on obtient une approximation assez fiable de l'erreur relative.

Remarque 2.3

Dans l'algorithme précédent, il faut prendre garde au cas où la racine recherchée est 0. Il y a alors risque de division par 0 au cours de l'évaluation de l'erreur relative. Ce cas est toutefois rare en pratique. ♦

Remarque 2.4

Il est parfois utile d'introduire un critère d'arrêt sur la valeur de $f(x)$, qui doit tendre également vers 0. ♦

Exemple 2.5

La fonction $f(x) = x^3 + x^2 - 3x - 3$ possède un zéro dans l'intervalle $[1, 2]$. En effet :

$$f(1) \times f(2) = -4,0 \times 3,0 = -12,0 < 0$$

On a alors $x_m = 1,5$ et $f(1,5) = -1,875$. L'intervalle $[1,5, 2]$ possède encore un changement de signe, ce qui n'est pas le cas pour l'intervalle $[1, 1,5]$. Le nouvel intervalle de travail est donc $[1,5, 2]$, dont le point milieu est $x_m = 1,75$. Puisque $f(1,75) = 0,171\,87$, on prendra l'intervalle $[1,5, 1,75]$ et ainsi de suite. Le tableau suivant résume les résultats.

Méthode de la bisection : $f(x) = x^3 + x^2 - 3x - 3$						
x_1	x_2	x_m	$f(x_1)$	$f(x_2)$	$f(x_m)$	Erreur absolue liée à x_m
1,0	2,0	1,5	-4,0	3,0	-1,875	0,5
1,5	2,0	1,75	-1,875	3,0	+0,171 87	0,25
1,5	1,75	1,625	-1,875	0,171 87	-0,943 35	0,125
1,625	1,75	1,6875	-0,943 35	0,171 87	-0,409 42	0,0625
1,6875	1,75	1,718 75	-0,409 42	0,171 87	-0,124 78	0,031 25



On remarque aisément que la longueur de l'intervalle entourant la racine est divisée par deux à chaque itération. Cette constatation permet de déterminer à l'avance le nombre d'itérations nécessaire pour obtenir une certaine erreur absolue Δr sur la racine r . Soit $L = x_2 - x_1$, la longueur de l'intervalle de départ. Après une itération, le nouvel intervalle est de longueur $\frac{L}{2}$ et après n itérations la longueur de l'intervalle est $L/2^n$. Si l'on veut connaître la valeur de n nécessaire pour avoir :

$$\frac{L}{2^n} < \Delta r$$

il suffit de résoudre cette équation en fonction de n et l'on trouve la condition :

$$n > \frac{\ln\left(\frac{L}{\Delta r}\right)}{\ln 2} \quad (2.3)$$

Il est clair que, sur le plan pratique, on doit prendre pour valeur de n le plus petit entier vérifiant cette condition. On a aussi tout intérêt à bien cerner la racine recherchée et à prendre, dès le départ, un intervalle de longueur aussi petite que possible.

Exemple 2.6

Dans l'exemple précédent, $L = 2,0 - 1,0$. Si l'on veut une erreur absolue plus petite que $0,5 \times 10^{-2}$, ce qui revient à s'assurer que le chiffre des centièmes est significatif, il faut effectuer au moins :

$$\frac{\ln\left(\frac{1,0}{0,5 \times 10^{-2}}\right)}{\ln 2} = 7,64 \text{ itérations}$$

On fera donc 8 itérations pour s'assurer de cette précision. On peut aisément vérifier qu'après 8 itérations l'erreur maximale liée à x_m est de 0,003 906 25 et que la véritable erreur est 0,001 582. ■

Exemple 2.7

On souhaite calculer $\sqrt{2}$ avec une calculatrice dotée seulement des 4 opérations élémentaires. Cela revient à résoudre :

$$x^2 - 2 = 0$$

Cette fonction présente un changement de signe dans l'intervalle $[1, 2]$. L'algorithme de la bisection donne les résultats suivants.

Méthode de la bisection : $f(x) = x^2 - 2$						
x_1	x_2	x_m	$f(x_1)$	$f(x_2)$	$f(x_m)$	$(x_m)_2$
1,0	2,0	1,5	-1,0	2,0	+0,25	1,1
1,0	1,5	1,25	-1,0	0,25	-0,4375	1,01
1,25	1,5	1,375	-0,4375	0,25	-0,1094	1,011
1,375	1,5	1,4375	-0,1094	0,25	+0,066 41	1,0111
1,375	1,4375	1,4062	-0,1094	0,066 41	-0,022 46	1,011 01
1,4062	1,4375	1,4219	-0,022 46	0,066 41	+0,021 73	1,011 011
1,4062	1,4219	1,4141	-0,022 46	0,021 73	-0,000 43	1,011 010 1

Chapitre 3

Systèmes d'équations algébriques

3.1 Introduction

Les systèmes d'équations algébriques jouent un rôle très important en ingénierie. On peut classer ces systèmes en deux grandes familles : les systèmes *linéaires* et les systèmes *non linéaires*.

Ici encore, les progrès de l'informatique et de l'analyse numérique permettent d'aborder des problèmes de taille prodigieuse. On résout couramment aujourd'hui des systèmes de plusieurs centaines de milliers d'inconnues. On rencontre ces applications en mécanique des fluides et dans l'analyse de structures complexes. On peut par exemple calculer l'écoulement de l'air autour d'un avion ou l'écoulement de l'eau dans une turbine hydraulique complète. On peut également analyser la résistance de la carlingue d'un avion à différentes contraintes extérieures et en vérifier numériquement la solidité.

Ces calculs complexes requièrent des méthodes évoluées comme les méthodes d'éléments finis (voir Reddy, réf. [31]). On obtient généralement des systèmes d'équations non linéaires de taille considérable, qu'on doit résoudre à l'aide de méthodes efficaces qui minimisent le temps de calcul et l'espace-mémoire requis.

Dans ce chapitre, nous allons aborder les principales méthodes de résolution des systèmes linéaires, à savoir la méthode d'élimination de Gauss et la décomposition LU . L'effet des erreurs dues à l'arithmétique flottante sera également étudié et nous introduirons le concept de *conditionnement* d'une matrice.

Par la suite, nous verrons comment résoudre les systèmes non linéaires au moyen d'une suite de systèmes linéaires. C'est ce que nous appelons la linéarisation du problème.

3.2 Systèmes linéaires

De façon générale, la résolution d'un système d'équations linéaires consiste à trouver un vecteur $\vec{x} = [x_1 \ x_2 \ x_3 \ \cdots \ x_n]^T$ (\vec{x} dénotera toujours un vecteur colonne et l'indice supérieur T désignera sa transposée) solution de :

$$\begin{array}{ccccccccc}
 a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\
 a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & b_2 \\
 a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & b_3 \\
 & & & & \vdots & & & & & = & \vdots \\
 a_{n1}x_1 & + & a_{n2}x_2 & + & a_{n3}x_3 & + & \cdots & + & a_{nn}x_n & = & b_n
 \end{array} \tag{3.1}$$

On peut utiliser la notation matricielle, qui est beaucoup plus pratique et surtout plus compacte. On écrit alors le système précédent sous la forme :

$$A\vec{x} = \vec{b} \tag{3.2}$$

où A est la matrice :

$$\begin{bmatrix}
 a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\
 a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\
 a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn}
 \end{bmatrix}$$

et où $\vec{b} = [b_1 \ b_2 \ b_3 \ \cdots \ b_n]^T$ est le *membre de droite*. Bien entendu, la matrice A et le vecteur \vec{b} sont connus. Il reste à déterminer le vecteur \vec{x} . Le problème 3.1 (ou 3.2) est un système de n équations et n inconnues. En pratique, la valeur de n varie considérablement et peut s'élever jusqu'à plusieurs centaines de milliers. Dans ce chapitre, nous nous limitons à des systèmes de petite taille, mais les stratégies développées sont valides pour des systèmes de très grande taille. Notons toutefois que le coût de la résolution croît rapidement avec n .

Remarque 3.1

Dans la plupart des cas, nous traitons des *matrices non singulières* ou *inversibles*, c'est-à-dire dont la matrice inverse existe. Nous ne faisons pas non plus de révision systématique de l'algèbre linéaire élémentaire que nous supposons connue. Ainsi, la solution de l'équation 3.2 peut s'écrire :

$$\vec{x} = A^{-1}\vec{b}$$

et la discussion peut sembler close. Nous verrons cependant que le calcul de la matrice inverse A^{-1} est plus difficile et plus long que la résolution du système linéaire de départ. ♦

Exemple 3.2

Considérons le système linéaire suivant :

$$\begin{array}{rcl}
 2x_1 + 3x_2 & = & 8 \\
 3x_1 + 4x_2 & = & 11
 \end{array}$$

Pour le résoudre, on peut utiliser la méthode classique qui consiste à éliminer les équations une à une par *substitution successive*. Dans un premier temps, on isole x_1 de la première équation :

$$x_1 = \frac{8 - 3x_2}{2}$$

que l'on substitue dans la deuxième équation :

$$3 \left(\frac{8 - 3x_2}{2} \right) + 4x_2 = 11$$

ou encore :

$$12 - \frac{9x_2}{2} + 4x_2 = 12 - 0,5x_2 = 11$$

On déduit alors facilement que $x_2 = 2$ et par la suite que $x_1 = 1$. ■

Il est théoriquement possible d'étendre la substitution successive à des systèmes de grande taille. Il est cependant difficile de transcrire cette façon de faire sous forme d'algorithme (qui peut par la suite être programmé dans un langage informatique quelconque). Il est donc préférable de recourir à d'autres méthodes pour simplifier le système d'équations.

On peut d'abord se demander quels types de systèmes linéaires sont faciles à résoudre, et ce, même s'ils sont de grande taille. Le cas le plus simple est sans doute celui des *systèmes diagonaux*, c'est-à-dire dont la matrice A n'a de coefficients non nuls que sur la diagonale.

Exemple 3.3

Le système suivant :

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 9 \end{bmatrix}$$

est très facile à résoudre. Il suffit de considérer séparément chaque ligne. On obtient ainsi $\vec{x} = \begin{bmatrix} 2 & 1 & 3 \end{bmatrix}^T$. On voit tout de suite comment résoudre le cas général :

$$x_i = \frac{b_i}{a_{ii}} \quad \text{pour } i = 1, 2, \dots, n$$

On remarque de plus que le système a une solution unique seulement si tous les termes diagonaux sont non nuls. Hélas, on rencontre rarement des systèmes diagonaux en pratique et il faudra travailler un peu plus pour s'attaquer aux applications. ■

Le deuxième type de système simple est le *système triangulaire inférieur* ou *supérieur*.

Définition 3.4

Une matrice est dite *triangulaire inférieure* (ou *supérieure*) si tous les a_{ij} (ou tous les a_{ji}) sont nuls pour $i < j$. Une matrice triangulaire inférieure a la forme type :

$$\begin{bmatrix} a_{11} & 0 & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1\ 1} & a_{n-1\ 2} & a_{n-1\ 3} & \cdots & a_{n-1\ n} & 0 \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn-1} & a_{nn} \end{bmatrix}$$

Une matrice triangulaire supérieure est tout simplement la transposée d'une matrice triangulaire inférieure.

Les systèmes triangulaires sont également faciles à résoudre. Il suffit en effet de commencer par l'équation qui se trouve à la pointe du triangle (la première pour une matrice triangulaire inférieure et la dernière pour une matrice triangulaire supérieure) et de résoudre une à une les équations. On parle de *descente triangulaire* ou de *remontée triangulaire*, selon le cas.

Exemple 3.5

La descente triangulaire du système :

$$\begin{bmatrix} 3 & 0 & 0 \\ 1 & 2 & 0 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 7 \\ 14 \end{bmatrix}$$

consiste à résoudre la première équation :

$$x_1 = \frac{b_1}{a_{11}} = \frac{9}{3} = 3$$

Puisque x_1 est maintenant connue, on peut déterminer x_2 :

$$x_2 = \frac{b_2 - a_{21}x_1}{a_{22}} = \frac{7 - (1)(3)}{2} = 2$$

La dernière équation s'écrit :

$$x_3 = \frac{b_3 - a_{31}x_1 - a_{32}x_2}{a_{33}} = \frac{14 - (3)(3) - (2)(2)}{1} = 1$$

■

De l'exemple précédent, on peut rapidement déduire le cas général pour la descente triangulaire :

$$\begin{aligned} x_1 &= b_1/a_{11} \\ x_i &= \frac{\left(b_i - \sum_{k=1}^{i-1} a_{ik}x_k\right)}{a_{ii}} \quad \text{pour } i = 2, 3, \dots, n \end{aligned} \quad (3.3)$$

Pour la remontée triangulaire, on a :

$$\begin{aligned} x_n &= b_n/a_{nn} \\ x_i &= \frac{\left(b_i - \sum_{k=i+1}^n a_{ik}x_k\right)}{a_{ii}} \quad \text{pour } i = n-1, n-2, \dots, 2, 1 \end{aligned} \quad (3.4)$$

Remarque 3.6

Les équations 3.3 et 3.4 sont valides si les a_{ii} sont tous non nuls. Dans le cas contraire, la matrice A n'est pas inversible et, donc, le système $A\vec{x} = \vec{b}$ n'a pas une solution unique. On se souvient en effet que le déterminant d'une matrice triangulaire inférieure (ou supérieure) est :

$$\det A_{\text{triangulaire}} = \prod_{i=1}^n a_{ii} \quad (3.5)$$

En d'autres mots, le déterminant est le produit des termes de la diagonale de A . Le produit est donc non nul seulement si aucun des a_{ii} n'est nul. ♦

Les matrices triangulaires sont primordiales pour la résolution des systèmes linéaires. Dans les sections qui suivent, nous voyons comment ramener un système linéaire quelconque à un ou plusieurs systèmes triangulaires. Nous abordons essentiellement deux méthodes dites *directes* au sens de la définition suivante.

Définition 3.7

Une méthode de résolution d'un système linéaire est dite *directe* si la solution du système peut être obtenue par cette méthode en un nombre fini et prédéterminé d'opérations.

Autrement dit, les méthodes directes permettent d'obtenir le résultat après un nombre connu de multiplications, divisions, additions et soustractions. On peut alors en déduire le temps de calcul nécessaire à la résolution (qui peut être très long si n est grand). Les méthodes directes s'opposent sur ce point aux méthodes dites *itératives*, qui peuvent converger en quelques itérations, converger en un très grand nombre d'itérations ou même diverger, selon le

cas. Nous présentons quelques exemples de méthodes itératives à la fin du chapitre 4.

Les deux principales méthodes directes sont la *méthode d'élimination de Gauss* et la *décomposition LU*. Il s'agit en fait d'une seule et même méthode puisque la méthode d'élimination de Gauss est un cas particulier de décomposition *LU*. La stratégie de résolution est basée sur la question suivante : *quelles opérations sont permises sur les lignes du système 3.1 pour le ramener à un système triangulaire ?* Ou encore : pour ramener un système linéaire quelconque à un système triangulaire, quels sont les coups permis, c'est-à-dire *ceux qui ne changent pas la solution du système de départ ?* C'est à ces questions que nous répondons dans la section suivante.

3.3 Opérations élémentaires sur les lignes

Revenons au système :

$$A\vec{x} = \vec{b} \quad (3.6)$$

et voyons comment on peut le transformer sans en modifier la solution. La réponse est toute simple. On peut toujours multiplier (à gauche de chaque côté) les termes de cette relation par une matrice W *inversible*; la solution n'est pas modifiée puisque l'on peut remultiplier par W^{-1} pour revenir au système de départ. Ainsi :

$$WA\vec{x} = W\vec{b}$$

possède la même solution que le système 3.6.

Remarque 3.8

Ce résultat n'est plus vrai si la matrice W n'est pas inversible. On ne peut plus en effet revenir en arrière si la matrice W^{-1} n'existe pas. ♦

Exemple 3.9

Nous avons vu que la solution du système :

$$\begin{bmatrix} 3 & 0 & 0 \\ 1 & 2 & 0 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 7 \\ 14 \end{bmatrix}$$

est $\vec{x} = [3 \quad 2 \quad 1]^T$. Si l'on multiplie ce système par la matrice inversible :

$$W = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{bmatrix}$$

on obtient le nouveau système :

$$\begin{bmatrix} 3 & 0 & 0 \\ 5 & 4 & 0 \\ 14 & 10 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 23 \\ 65 \end{bmatrix}$$

Chapitre 4

Méthodes itératives et systèmes dynamiques discrets

4.1 Introduction

Nous voyons dans ce chapitre comment de simples méthodes itératives, telles les méthodes des points fixes, peuvent mener à des systèmes au comportement complexe. On pourrait croire, à la suite du chapitre 2, que la discussion sur la convergence d'une méthode des points fixes s'arrête lorsqu'on a déterminé si le point fixe est attractif ou répulsif. Nous allons pousser cette discussion beaucoup plus loin et tâcher d'étudier un certain nombre de phénomènes intéressants rencontrés dans l'étude des systèmes dynamiques. Il ne s'agit pas de faire une analyse mathématique profonde de la théorie des systèmes dynamiques, mais bien de démontrer que des méthodes itératives simples peuvent résulter en des systèmes complexes.

4.2 Application quadratique

Nous reprenons ici une partie du travail de Feigenbaum (réf. [16]) sur l'application quadratique. Cette application remarquablement simple conduit à un comportement de nature universelle.

Considérons la méthode itérative :

$$\begin{cases} x_0 & \text{donné} \\ x_{n+1} &= \lambda x_n(1 - x_n) \end{cases} \quad (4.1)$$

qui est en fait une méthode des points fixes (voir l'équation 2.5) appliquée à la fonction :

$$g(x) = \lambda x(1 - x)$$

Le paramètre λ est appelé à varier, si bien que le comportement de l'algorithme 4.1 sera très différent suivant la valeur de λ .

Tout d'abord, il est facile de montrer que la fonction $g(x)$ est une application de l'intervalle $[0, 1]$ dans lui-même ($g(x) \in [0, 1]$ si $x \in [0, 1]$) seulement pour :

$$0 < \lambda < 4$$

En effet, le maximum de $g(x)$ est atteint en $x = \frac{1}{2}$ et vaut $\frac{\lambda}{4}$. Nous nous limitons donc à ces valeurs de λ , qui sont de loin les plus intéressantes. En premier lieu, il convient de déterminer les points fixes de $g(x)$ et de vérifier s'ils sont attractifs ou répulsifs. Bien entendu, cela dépendra de λ . Les points fixes sont les solutions de :

$$x = g(x) = \lambda x(1 - x)$$

On constate immédiatement que 0 est une solution de cette équation et est donc un point fixe. Si l'on suppose que $x \neq 0$ et que l'on divise chaque côté de l'égalité par x , on obtient :

$$1 = \lambda(1 - x)$$

ce qui entraîne que :

$$x = r^* = \frac{\lambda - 1}{\lambda} \quad (4.2)$$

est un autre point fixe. En fait, 0 et r^* sont les deux seuls points fixes de $g(x)$. Voyons maintenant s'ils sont attractifs. Pour ce faire, il faut calculer la dérivée de $g(x)$, à savoir :

$$g'(x) = \lambda(1 - 2x) \quad (4.3)$$

On a donc d'une part :

$$g'(0) = \lambda$$

ce qui signifie que 0 sera un point fixe attractif si :

$$|g'(0)| < 1$$

c'est-à-dire si :

$$0 < \lambda < 1 = \lambda_1$$

puisqu'on ne considère pas les valeurs négatives de λ . D'autre part :

$$g'(r^*) = \lambda(1 - 2r^*) = \lambda \left(1 - 2 \left(\frac{\lambda - 1}{\lambda} \right) \right) = 2 - \lambda$$

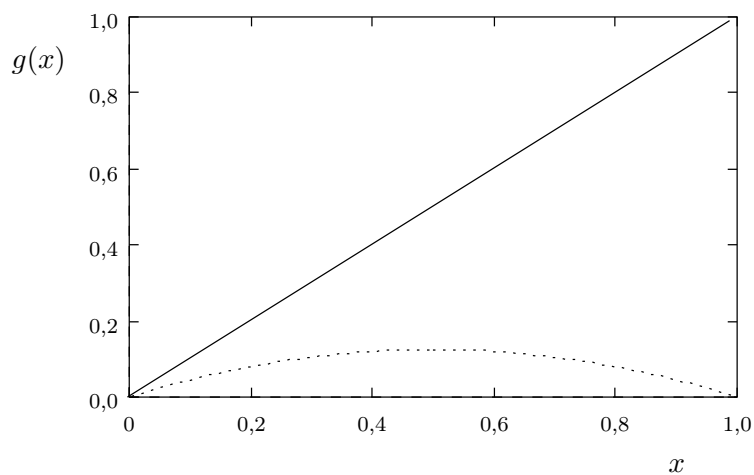
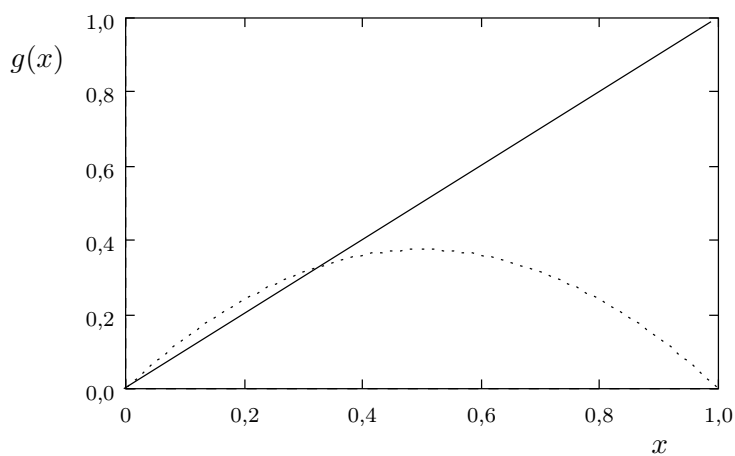
Le point fixe r^* est donc attractif si :

$$|2 - \lambda| < 1$$

ou encore si :

$$\begin{array}{ccccccc} -1 & < & 2 - \lambda & < & 1 \\ -3 & < & -\lambda & < & -1 \\ 1 & < & \lambda & < & 3 & = & \lambda_2 \end{array}$$

On note λ_2 la borne supérieure de cet intervalle, soit $\lambda_2 = 3$. On en conclut que r^* est attractif pour ces valeurs de λ . On remarque de plus que, lorsque $\lambda = 1$, $r^* = 0$ et il n'y a alors qu'un seul point fixe. On montre que ce point fixe est attractif même si $g'(0) = 1$ en vertu de l'équation 4.3. La convergence est cependant très lente. La situation est illustrée aux figures 4.1 et 4.2, où la fonction $g(x)$ est tracée pour deux valeurs de λ , soit 0,5 et 1,5. On voit dans le premier cas ($\lambda = 0,5$) que la pente de $g(x)$ est inférieure à 1 en $x = 0$ et qu'il n'y a pas d'autre point fixe dans l'intervalle $[0, 1]$. Par contre, pour $\lambda = 1,5$,

**Figure 4.1** – Application quadratique : $\lambda = 0,5$ **Figure 4.2** – Application quadratique : $\lambda = 1,5$

il y a bien deux points fixes $x = 0$ et $x = r^*$, dont seul r^* est attractif, car $g'(r^*) < 1$.

Vérifions tout cela avec quelques exemples. Prenons d'abord $\lambda = 0,5$. On a alors $r^* = -1$ qui ne peut être attractif puisqu'il est l'extérieur de l'intervalle $[0, 1]$. À partir de $x_0 = 0,9$ par exemple, on trouve les itérations suivantes.

Application quadratique : $\lambda = 0,5$			
n	x_n	n	x_n
0	0,900 0000	6	0,001 2902
1	0,045 0000	7	0,000 6426
2	0,021 4875	8	0,000 3219
3	0,010 5128	9	0,000 1609
4	0,005 2011	10	0,000 0804
5	0,002 5871	\vdots	\vdots

Ces itérations convergent rapidement vers le point fixe 0. Si l'on prend maintenant $\lambda = 0,95$, toujours à partir de $x_0 = 0,9$, on obtient les itérations suivantes.

Application quadratique : $\lambda = 0,95$			
n	x_n	n	x_n
0	0,900 0000	6	0,046 7527
1	0,085 5000	7	0,042 3386
2	0,074 2802	8	0,038 5187
3	0,065 3245	9	0,035 1833
4	0,058 0044	10	0,032 2482
5	0,051 9079	\vdots	\vdots

Ces dernières convergent vers 0, mais beaucoup plus lentement. Cela tient au fait que le taux de convergence $g'(0) = \lambda$ vaut 0,5 dans le premier cas et 0,95 dans le second cas. La convergence est donc plus rapide pour $\lambda = 0,5$. Pour s'assurer de la convergence vers 0, il faudrait faire beaucoup plus que 10 itérations. Par exemple, pour $\lambda = 0,95$, on trouverait $x_{200} = 0,114\,1385 \times 10^{-5}$.

Passons maintenant à $\lambda = 1,5$, pour lequel $r^* = \frac{1}{3}$. L'analyse a démontré que, dans ce cas, 0 est répulsif puisque $g'(0) = 1,5$, mais que r^* est attractif. À partir cette fois de $x_0 = 0,1$, on obtient les itérations suivantes.

Application quadratique : $\lambda = 1,5$			
n	x_n	n	x_n
0	0,100 0000	7	0,042 3386
1	0,085 5000	8	0,038 5187
2	0,074 2802	9	0,035 1833
3	0,065 3245	10	0,032 2482
4	0,058 0044		
5	0,051 9079	\vdots	\vdots
6	0,046 7527	20	0,333 3313

Les itérations convergent donc vers $r^* = \frac{1}{3}$, un résultat qui confirme l'analyse précédente. On obtiendrait des résultats similaires pour des valeurs de λ situées dans l'intervalle $]1, 3[$. Notons cependant que la valeur de r^* varie avec λ .

La question fondamentale est maintenant la suivante : que se passe-t-il si l'on prend des valeurs de λ supérieures à 3 ? On pourrait s'attendre à ce que les itérations de l'algorithme 4.1 divergent. Heureusement, ce n'est pas le cas, mais pour expliquer ce comportement il nous faut élargir la notion de convergence. Jusqu'à maintenant, nous n'avons parlé de convergence que vers un point (fixe). Or, il arrive qu'un algorithme converge vers autre chose qu'un point fixe. On parle alors d'un *attracteur* (voir par exemple Gulick, réf. [21]).

Définition 4.1

Un ensemble $A \subset R^n$ est dit un *attracteur* d'une application :

$$g : V \rightarrow R^n$$

où V est un sous-ensemble de R^n , si les conditions suivantes sont respectées :

1. Si $x \in A$, alors $g(x) \in A$.
2. Il existe un voisinage U de A tel que, si $x_0 \in U$, la suite $x_{n+1} = g(x_n)$ converge vers A .
3. L'ensemble A est indécomposable.

Remarque 4.2

La définition qui précède indique que, pour que A soit un attracteur, il faut que tout point x de l'ensemble A soit projeté sur un autre point de A par l'application $g(x)$. C'est bien sûr le cas d'un point fixe qui est envoyé sur lui-même. La deuxième condition traduit le fait que, si l'on part d'un point x_0 suffisamment près de A , les itérations de l'algorithme des points fixes s'approchent de plus en plus de l'ensemble A . On étend ainsi aux attracteurs la notion de bassin d'attraction introduite pour les points fixes. Mentionnons enfin que le sens précis que l'on doit donner au mot « indécomposable » varie d'un auteur à l'autre. Disons simplement qu'on ne peut rien enlever à l'ensemble A pour que les deux premières propriétés restent vraies. ♦

Remarque 4.3

On ne considère pour l'instant que le cas $n = 1$ des applications de R dans R . Le cas général sera traité un peu plus loin. Un point fixe est donc un attracteur (voir le chapitre 2) s'il existe un intervalle I contenant ce point fixe pour lequel $g(x) \in I$, $\forall x \in I$, et qui vérifie :

$$|g'(x)| < 1 \quad \forall x \in I$$

♦

Cette définition d'un attracteur est quelque peu imprécise, mais elle suffit aux besoins de l'exposé. Prenons par exemple $\lambda = 3,1$ et observons ce qui se passe. À partir de $x_0 = 0,5$, on obtient les itérations suivantes.

Application quadratique : $\lambda = 3,1$			
n	x_n	n	x_n
1	0,775 0000	14	0,557 4733
2	0,540 5625	15	0,764 7601
3	0,769 8995	16	0,557 6964
4	0,549 1781	17	0,764 6804
5	0,767 5026	18	0,557 8271
6	0,553 1711	19	0,764 6336
7	0,766 2357	20	0,557 9039
8	0,555 2674	\vdots	\vdots
9	0,765 5310	47	0,764 5665
10	0,556 4290	48	0,558 0140
11	0,765 1288	49	0,764 5665
12	0,557 0907	50	0,558 0140
13	0,764 8960		

On remarque immédiatement un comportement surprenant. Les itérations oscillent entre les valeurs de 0,5580 et de 0,7645. Il n'y a donc pas convergence au sens habituel. En fait, les itérations paires convergent vers environ 0,558 014 et les itérations impaires, vers 0,764 566. Pour comprendre ce qui se passe, il suffit de constater que les itérations paires et impaires correspondent aux itérations de la fonction composée :

$$g_1(x) = g(g(x)) = \lambda g(x)(1 - g(x)) = \lambda(\lambda x(1 - x))(1 - \lambda x(1 - x))$$

c'est-à-dire :

$$g_1(x) = \lambda^2 x(1 - x)(1 - \lambda x + \lambda x^2)$$

Pour déterminer les points fixes de la fonction $g_1(x)$, il suffit de résoudre :

$$x = g_1(x) = \lambda^2 x(1 - x)(1 - \lambda x + \lambda x^2)$$

Il est clair que tout point fixe de $g(x)$ est un point fixe de $g_1(x)$. Le point r^* donné par l'équation 4.2 ainsi que 0 sont donc des points fixes de $g_1(x)$, mais nous savons qu'ils sont répulsifs pour $\lambda > 3$. Il existe cependant d'autres points fixes de $g_1(x)$ qui ne sont pas des points fixes de $g(x)$. Après avoir divisé l'équation précédente par x de chaque côté, quelques manipulations algébriques nous amènent à résoudre l'équation :

$$\lambda^3 x^3 - 2\lambda^3 x^2 + \lambda^2(1 + \lambda)x + (1 - \lambda^2) = 0$$

dont les trois racines sont r^* et :

$$r_1^{(2)}, r_2^{(2)} = \left(\frac{1}{2} + \frac{1}{2\lambda} \right) \pm \frac{1}{2\lambda} \sqrt{(\lambda - 3)(\lambda + 1)} \quad (4.4)$$

On montre alors que :

$$r_1^{(2)} = g(r_2^{(2)}) \quad \text{et} \quad r_2^{(2)} = g(r_1^{(2)})$$

Chapitre 5

Interpolation

5.1 Introduction

Ce chapitre ainsi que le chapitre suivant qui porte sur la dérivation et l'intégration numériques sont très étroitement reliés puisqu'ils tendent à répondre à diverses facettes d'un même problème. Ce problème est le suivant : à partir d'une fonction $f(x)$ connue seulement en $(n+1)$ points de la forme $((x_i, f(x_i)))$ pour $i = 0, 1, 2, \dots, n$, peut-on construire une approximation de $f(x)$, et ce, pour tout x ? Les x_i sont appelés *abscisses* ou *nœuds d'interpolation* tandis que les couples $((x_i, f(x_i)))$ pour $i = 0, 1, 2, \dots, n$ sont les *points de collocation* ou *points d'interpolation* et peuvent provenir de données expérimentales ou d'une table. En d'autres termes, si l'on ne connaît que les points de collocation $(x_i, f(x_i))$ d'une fonction, peut-on obtenir une approximation de $f(x)$ pour une valeur de x différente des x_i ? La figure 5.1 résume la situation.

Sur la base des mêmes hypothèses, nous verrons, au chapitre suivant, comment évaluer les dérivées $f'(x), f''(x) \dots$ de même que :

$$\int_{x_0}^{x_n} f(x) dx$$

Il s'agit d'un *problème d'interpolation*, dont la solution est relativement simple. Il suffit de construire un polynôme de degré suffisamment élevé dont la courbe passe par les points de collocation. On parle alors du *polynôme de collocation* ou *polynôme d'interpolation*. Pour obtenir une approximation des dérivées ou de l'intégrale, il suffit de dériver ou d'intégrer le polynôme de collocation. Il y a cependant des éléments fondamentaux qu'il est important d'étudier. En premier lieu, il convient de rappeler certains résultats cruciaux relatifs aux polynômes, que nous ne démontrons pas.

Théorème 5.1

Un polynôme de degré n dont la forme générale est :

$$p_n(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n \quad (5.1)$$

avec $a_n \neq 0$ possède, tenant compte des multiplicités, très exactement n racines qui peuvent être réelles ou complexes conjuguées. Rappelons que r est une racine de $p_n(x)$ si $p_n(r) = 0$. ★

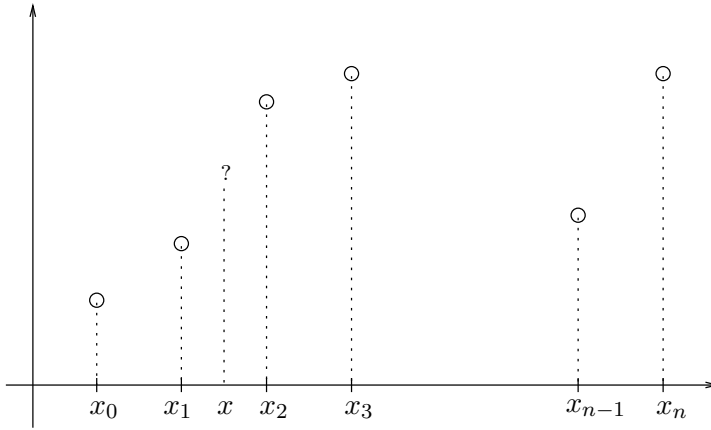


Figure 5.1 – Problème d'interpolation

Corollaire 5.2

Par $(n + 1)$ points de collocation d'abscisses distinctes $((x_i, f(x_i))$ pour $i = 0, 1, 2, \dots, n$), on ne peut faire correspondre qu'un et un seul polynôme de degré n .

Démonstration :

On procède par l'absurde et l'on suppose l'existence de 2 polynômes de degré n , notés $p(x)$ et $q(x)$, et qui passent tous les deux par les $(n + 1)$ points de collocation donnés. On considère ensuite la différence $P(x) = p(x) - q(x)$ qui est également un polynôme de degré au plus n . Ce polynôme vérifie :

$$P(x_i) = p(x_i) - q(x_i) = f(x_i) - f(x_i) = 0$$

et ce, pour i allant de 0 à n . Le polynôme $P(x)$ posséderait donc $(n + 1)$ racines, ce qui est impossible en vertu du théorème précédent. ♠

Définition 5.3

L'unique polynôme de degré n passant par les points $(x_i, f(x_i))$ pour $i = 0, 1, 2, \dots, n$, est appelé l'interpolant de $f(x)$ de degré n aux abscisses (nœuds) x_0, x_1, \dots, x_n .

Remarque 5.4

Rien n'oblige à ce que le coefficient a_n de l'interpolant soit différent de 0. L'interpolant passant par les $n + 1$ points d'interpolation peut donc être de degré inférieur à n . Si on choisit par exemple 10 points sur une droite, l'interpolant sera quand même de degré 1. ♦

Il reste à assurer l'existence de l'interpolant, ce que nous ferons tout simplement en le construisant au moyen de méthodes diverses qui feront l'objet des prochaines sections.

5.2 Matrice de Vandermonde

Le problème d'interpolation consiste donc à déterminer l'unique polynôme de degré n passant par les $(n+1)$ points de collocation $((x_i, f(x_i))$ pour $i = 0, 1, 2, 3, \dots, n$). Selon le théorème précédent, il ne saurait y en avoir deux. Il reste maintenant à le construire de la manière la plus efficace et la plus générale possible. Une première tentative consiste à déterminer les inconnues a_i du polynôme 5.1 en vérifiant directement les $(n+1)$ équations de collocation :

$$p_n(x_i) = f(x_i) \quad \text{pour } i = 0, 1, 2, \dots, n$$

ou encore :

$$a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3 + \dots + a_n x_i^n = f(x_i)$$

qui est un système linéaire de $(n+1)$ équations en $(n+1)$ inconnues. Ce système s'écrit sous forme matricielle :

$$\begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} \quad (5.2)$$

Remarque 5.5

La matrice de ce système linéaire porte le nom de *matrice de Vandermonde*. On peut montrer que le conditionnement de cette matrice augmente fortement avec la taille $(n+1)$ du système. De plus, comme le révèlent les sections qui suivent, il n'est pas nécessaire de résoudre un système linéaire pour calculer un polynôme d'interpolation. *Cette méthode est donc rarement utilisée.* ♦

Exemple 5.6

On doit calculer le polynôme passant par les points $(0, 1)$, $(1, 2)$, $(2, 9)$ et $(3, 28)$. Étant donné ces 4 points, le polynôme recherché est tout au plus de degré 3. Ses coefficients a_i sont solution de :

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

dont la solution (obtenue par décomposition LU) est $[1 \ 0 \ 0 \ 1]^T$. Le polynôme recherché est donc $p_3(x) = 1 + x^3$. ■

Les sections suivantes proposent des avenues différentes et plus efficaces pour calculer le polynôme de collocation.

5.3 Interpolation de Lagrange

L'interpolation de Lagrange est une façon simple et systématique de construire un polynôme de collocation. Étant donné $(n + 1)$ points $((x_i, f(x_i)))$ pour $i = 0, 1, 2, \dots, n$, on suppose un instant que l'on sait construire $(n + 1)$ polynômes $L_i(x)$ de degré n et satisfaisant les conditions suivantes :

$$\begin{cases} L_i(x_i) = 1 & \forall i \\ L_i(x_j) = 0 & \forall j \neq i \end{cases} \quad (5.3)$$

Cela signifie que le polynôme $L_i(x)$ de degré n prend la valeur 1 en x_i et s'annule à tous les autres points de collocation. Nous verrons comment construire les $L_i(x)$ un peu plus loin. Dans ces conditions, la fonction $L(x)$ définie par :

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

est un polynôme de degré n , car chacun des $L_i(x)$ est de degré n . De plus, ce polynôme passe par les $(n + 1)$ points de collocation et est donc le polynôme recherché. En effet, il est facile de montrer que selon les conditions 5.3 :

$$\begin{aligned} L(x_j) &= f(x_j) L_j(x_j) + \sum_{i=0, i \neq j}^n f(x_i) L_i(x_j) \\ &= f(x_j) + 0 = f(x_j) \quad \forall j \end{aligned}$$

Le polynôme $L(x)$ passe donc par tous les points de collocation. Puisque ce polynôme est unique, $L(x)$ est bien le polynôme recherché. Il reste à construire les fonctions $L_i(x)$. Suivons une démarche progressive.

Polynômes de degré 1

Il s'agit de déterminer le polynôme de degré 1 dont la courbe (une droite) passe par les deux points $(x_0, f(x_0))$ et $(x_1, f(x_1))$. On doit donc construire deux polynômes $L_0(x)$ et $L_1(x)$ de degré 1 qui vérifient :

$$\begin{cases} L_0(x_0) = 1 \\ L_0(x_1) = 0 \end{cases} \quad \begin{cases} L_1(x_0) = 0 \\ L_1(x_1) = 1 \end{cases}$$

Le polynôme $L_0(x)$ doit s'annuler en $x = x_1$. On pense immédiatement au polynôme $(x - x_1)$ qui s'annule en $x = x_1$, mais qui vaut $(x_0 - x_1)$ en $x = x_0$. Pour s'assurer d'une valeur 1 en $x = x_0$, il suffit d'effectuer la division appropriée afin d'obtenir :

$$L_0(x) = \frac{(x - x_1)}{(x_0 - x_1)}$$

Un raisonnement similaire pour $L_1(x)$ donne :

$$L_1(x) = \frac{(x - x_0)}{(x_1 - x_0)}$$

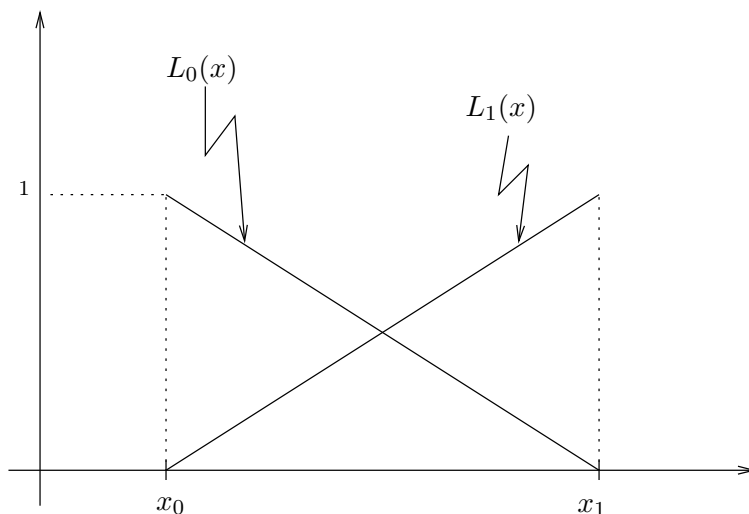


Figure 5.2 – Polynômes de Lagrange de degré 1 : $L_0(x)$ et $L_1(x)$

Ces deux fonctions sont illustrées à la figure 5.2. Le polynôme de degré 1 est donc :

$$p_1(x) = f(x_0)L_0(x) + f(x_1)L_1(x)$$

Exemple 5.7

L'équation de la droite passant par les points $(2, 3)$ et $(5, -6)$ est :

$$3 \frac{(x-5)}{(2-5)} + (-6) \frac{(x-2)}{(5-2)} = -(x-5) - 2(x-2) = -3x + 9$$

■

Polynômes de degré 2

Si l'on cherche le polynôme de degré 2 passant par les points $(x_0, f(x_0))$, $(x_1, f(x_1))$ et $(x_2, f(x_2))$, on doit construire trois fonctions $L_i(x)$. Le raisonnement est toujours le même. La fonction $L_0(x)$ s'annule cette fois en $x = x_1$ et en $x = x_2$. On doit forcément avoir un coefficient de la forme :

$$(x - x_1)(x - x_2)$$

qui vaut $(x_0 - x_1)(x_0 - x_2)$ en $x = x_0$. Pour satisfaire la condition $L_0(x_0) = 1$, il suffit alors de diviser le coefficient par cette valeur et de poser :

$$L_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}$$

Cette fonction vaut bien 1 en x_0 et 0 en x_1 et x_2 . De la même manière, on obtient les fonctions $L_1(x)$ et $L_2(x)$ définies par :

$$L_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} \quad \text{et} \quad L_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

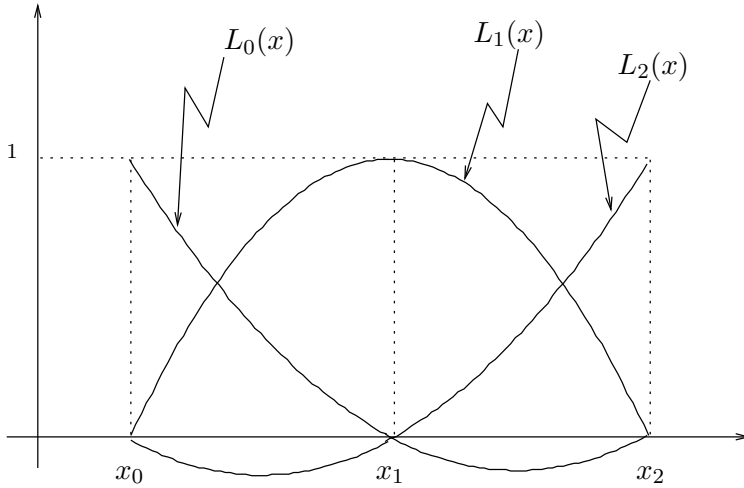


Figure 5.3 – Polynômes de Lagrange de degré 2 : $L_0(x)$, $L_1(x)$ et $L_2(x)$

Ces trois fonctions sont à leur tour illustrées à la figure 5.3.

Exemple 5.8

La parabole passant par les points $(1, 2)$, $(3, 7)$, $(4, -1)$ est donnée par :

$$\begin{aligned} p_2(x) &= 2 \frac{(x-3)(x-4)}{(1-3)(1-4)} + 7 \frac{(x-1)(x-4)}{(3-1)(3-4)} + (-1) \frac{(x-1)(x-3)}{(4-1)(4-3)} \\ &= \frac{(x-3)(x-4)}{3} - \frac{7(x-1)(x-4)}{2} - \frac{(x-1)(x-3)}{3} \end{aligned}$$

■

Polynômes de degré n

On analyse le cas général de la même façon. La fonction $L_0(x)$ doit s'annuler en $x = x_1, x_2, x_3, \dots, x_n$. Il faut donc introduire la fonction :

$$(x-x_1)(x-x_2)(x-x_3) \cdots (x-x_n)$$

qui vaut :

$$(x_0-x_1)(x_0-x_2)(x_0-x_3) \cdots (x_0-x_n)$$

en $x = x_0$. On a alors, après division :

$$L_0(x) = \frac{(x-x_1)(x-x_2)(x-x_3) \cdots (x-x_n)}{(x_0-x_1)(x_0-x_2)(x_0-x_3) \cdots (x_0-x_n)}$$

On remarque qu'il y a n facteurs de la forme $(x-x_i)$ dans cette expression et qu'il s'agit bien d'un polynôme de degré n . Pour la fonction $L_1(x)$, on pose :

$$L_1(x) = \frac{(x-x_0)(x-x_2)(x-x_3) \cdots (x-x_n)}{(x_1-x_0)(x_1-x_2)(x_1-x_3) \cdots (x_1-x_n)}$$

Chapitre 6

Différentiation et intégration numériques

6.1 Introduction

Le contenu de ce chapitre prolonge celui du chapitre 5 sur l'interpolation. À peu de choses près, on y manie les mêmes outils d'analyse. Dans le cas de l'interpolation, on cherchait à évaluer une fonction $f(x)$ connue seulement en quelques points. Dans le présent chapitre, le problème consiste à obtenir des approximations des différentes dérivées de cette fonction de même que de :

$$\int_{x_0}^{x_n} f(x)dx$$

On parle alors de *dérivation numérique* et d'*intégration numérique*. On fait face à ce type de problèmes lorsque, par exemple, on connaît la position d'une particule à intervalles de temps réguliers et que l'on souhaite obtenir sa vitesse. On doit alors effectuer la dérivée de la position connue seulement en quelques points. De même, l'accélération de cette particule nécessite le calcul de la dérivée seconde.

Si, à l'inverse, on connaît la vitesse d'une particule à certains intervalles de temps, on obtient la distance parcourue en intégrant la vitesse dans l'intervalle $[x_0, x_n]$.

Nous avons vu au chapitre précédent que la fonction $f(x)$ peut être convenablement estimée à l'aide d'un polynôme de degré n avec une certaine erreur. En termes concis :

$$f(x) = p_n(x) + E_n(x) \tag{6.1}$$

où $E_n(x)$ est le terme d'erreur d'ordre $(n + 1)$ donné par la relation 5.21. L'expression 6.1 est à la base des développements de ce chapitre.

6.2 Différentiation numérique

On peut aborder la différentiation numérique d'au moins deux façons. La première approche consiste à utiliser le développement de Taylor et la seconde

est fondée sur l'égalité 6.1. Nous utiliserons un mélange des deux approches, ce qui nous permettra d'avoir un portrait assez complet de la situation.

Commençons d'abord par l'équation 6.1. Si l'on dérive de chaque côté de l'égalité, on obtient successivement :

$$\begin{aligned} f'(x) &= p'_n(x) + E'_n(x) \\ f''(x) &= p''_n(x) + E''_n(x) \\ f'''(x) &= p'''_n(x) + E'''_n(x) \\ \vdots &= \vdots \end{aligned} \tag{6.2}$$

Ainsi, pour évaluer la dérivée d'une fonction connue aux points $((x_i, f(x_i))$ pour $i = 0, 1, 2, \dots, n$), il suffit de dériver le polynôme d'interpolation passant par ces points. De plus, *le terme d'erreur associé à cette approximation de la dérivée est tout simplement la dérivée de l'erreur d'interpolation*. Ce résultat est vrai quel que soit l'ordre de la dérivée.

Remarque 6.1

Bien qu'en théorie on soit en mesure d'estimer les dérivées de tout ordre, sur le plan pratique, on dépasse rarement l'ordre 4. Cela s'explique par le fait que la différentiation numérique est un procédé numériquement instable. ♦

6.2.1 Dérivées d'ordre 1

Commençons par faire l'approximation des dérivées d'ordre 1, ce qui revient à évaluer la pente de la fonction $f(x)$. Tout comme pour l'interpolation, nous avons le choix entre plusieurs polynômes de degré plus ou moins élevé. De ce choix dépendent l'ordre et la précision de l'approximation. Nous avons rencontré un problème semblable dans le cas de l'interpolation : si un polynôme de degré n est utilisé, on obtient une approximation d'ordre $(n + 1)$ de la fonction $f(x)$ (voir la relation 5.26).

Il est également utile de se rappeler que l'erreur d'interpolation s'écrit :

$$E_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!} [(x - x_0)(x - x_1) \cdots (x - x_n)] \tag{6.3}$$

pour un certain ξ compris dans l'intervalle $[x_0, x_n]$. En dérivant l'expression précédente, tout en tenant compte de la dépendance de ξ envers x , on obtient une relation pour la dérivée de l'erreur d'interpolation :

$$\begin{aligned} E'_n(x) &= \frac{f^{(n+2)}(\xi(x))\xi'(x)}{(n+1)!} [(x - x_0)(x - x_1) \cdots (x - x_n)] \\ &\quad + \frac{f^{(n+1)}(\xi(x))}{(n+1)!} [(x - x_0)(x - x_1) \cdots (x - x_n)]' \end{aligned}$$

La dérivée du produit apparaissant dans le deuxième terme de droite est plus délicate. Cette dérivée débouche sur une somme de produits où, tour à tour,

l'un des facteurs $(x - x_i)$ est manquant. Il est facile de se convaincre, en reprenant ce développement avec $n = 2$ par exemple, que l'on obtient :

$$\begin{aligned} E'_n(x) &= \frac{f^{(n+2)}(\xi(x))\xi'(x)}{(n+1)!}[(x-x_0)(x-x_1)\cdots(x-x_n)] \\ &\quad + \frac{f^{(n+1)}(\xi(x))}{(n+1)!} \left(\sum_{k=0}^n \prod_{j=0(j \neq k)}^n (x-x_j) \right) \end{aligned} \quad (6.4)$$

On peut simplifier cette expression quelque peu complexe en choisissant l'un ou l'autre des points d'interpolation. En effet, en $x = x_i$, le premier terme de droite s'annule, faisant disparaître la dérivée de $\xi(x)$, qui est inconnue. De la somme, il ne reste qu'un seul terme puisque tous les autres contiennent un facteur $(x - x_i)$ et s'annulent. Il reste :

$$E'_n(x_i) = \frac{f^{(n+1)}(\xi(x_i))}{(n+1)!} \left(\prod_{j=0(j \neq i)}^n (x_i - x_j) \right)$$

Si l'on suppose de plus que les x_i sont également distancés, c'est-à-dire :

$$x_{i+1} - x_i = h$$

ce qui signifie que $x_i - x_j = (i - j)h$, on obtient :

$$E'_n(x_i) = \frac{f^{(n+1)}(\xi_i)h^n}{(n+1)!} \left(\prod_{j=0(j \neq i)}^n (i - j) \right) \quad (6.5)$$

où ξ_i est simplement une notation différente de $\xi(x_i)$. En particulier, si $i = 0$, on trouve :

$$E'_n(x_0) = \frac{f^{(n+1)}(\xi_0)h^n}{(n+1)!} \left(\prod_{j=0(j \neq 0)}^n (-j) \right) = \frac{f^{(n+1)}(\xi_0)h^n}{(n+1)!} \left(\prod_{j=1}^n (-j) \right)$$

c'est-à-dire :

$$E'_n(x_0) = \frac{(-1)^n h^n f^{(n+1)}(\xi_0)}{(n+1)} \quad (6.6)$$

pour un certain ξ_0 compris dans l'intervalle $[x_0, x_n]$.

Remarque 6.2

L'équation 6.5 montre que, si l'on utilise un polynôme d'interpolation de degré n (c'est-à-dire d'ordre $(n+1)$), la dérivée de ce polynôme évaluée en $x = x_i$ est une approximation d'ordre n de $f'(x_i)$. ♦

Définition 6.3

Aux points d'interpolation, on a :

$$f'(x_i) = p'_n(x_i) + E'_n(x_i) \quad (6.7)$$

Le terme $p'_n(x_i)$ dans l'équation 6.7 est une *formule aux différences finies* ou plus simplement une formule aux différences. Nous proposons plus loin plusieurs formules aux différences finies pour évaluer les différentes dérivées de $f(x)$. Elles se distinguent principalement par le degré du polynôme et par les points d'interpolation retenus.

Approximations d'ordre 1

Si l'on choisit le polynôme de degré 1 passant par les points $(x_0, f(x_0))$ et $(x_1, f(x_1))$, on a, grâce à la formule d'interpolation de Newton :

$$p_1(x) = f(x_0) + f[x_0, x_1](x - x_0)$$

et donc :

$$f'(x) = p'_1(x) + E'_1(x) = f[x_0, x_1] + E'_1(x) \quad (6.8)$$

En vertu de la relation 6.6 avec $n = 1$ et puisque $(x_1 - x_0) = h$, on arrive à :

$$f'(x_0) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} + E'_1(x_0) = \frac{f(x_1) - f(x_0)}{h} + \frac{(-1)^1 h^1 f^{(2)}(\xi_0)}{2}$$

qui peut encore s'écrire :

$$f'(x_0) = \frac{f(x_1) - f(x_0)}{h} - \frac{h f^{(2)}(\xi_0)}{2} \quad \text{pour } \xi_0 \in [x_0, x_1] \quad (6.9)$$

qui est la *différence avant d'ordre 1*. On l'appelle différence avant car, pour évaluer la dérivée en $x = x_0$, on cherche de l'information vers l'avant (en $x = x_1$).

De la même manière, si l'on évalue l'équation 6.8 en $x = x_1$, la relation 6.5 avec $(i = 1)$ donne :

$$\begin{aligned} f'(x_1) &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} + E'_1(x_1) \\ &= \frac{f(x_1) - f(x_0)}{h} + \frac{h^1 f^{(2)}(\xi_1)}{2!} \left(\prod_{j=0(j \neq 1)}^1 (1 - j) \right) \end{aligned}$$

ou encore :

$$f'(x_1) = \frac{f(x_1) - f(x_0)}{h} + \frac{h f^{(2)}(\xi_1)}{2} \quad \text{pour } \xi_1 \in [x_0, x_1] \quad (6.10)$$

qui est la *différence arrière d'ordre 1*.

On constate ainsi que la même différence divisée est une approximation de la dérivée à la fois en $x = x_0$ et en $x = x_1$. On remarque cependant que le terme d'erreur est différent aux deux endroits.

Approximations d'ordre 2

Passons maintenant aux polynômes de degré 2. Soit les points $(x_0, f(x_0))$, $(x_1, f(x_1))$ et $(x_2, f(x_2))$. Le polynôme de degré 2 passant par ces trois points est :

$$p_2(x) = f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1)$$

dont la dérivée est :

$$p'_2(x) = f[x_0, x_1] + f[x_0, x_1, x_2](2x - (x_0 + x_1))$$

Lorsque x prend successivement les valeurs x_0 , x_1 et x_2 , il est facile de montrer que l'on obtient des approximations d'ordre 2 de la dérivée.

Formules de différences d'ordre 2 pour $f'(x)$	
$f'(x_0) = \frac{-f(x_2) + 4f(x_1) - 3f(x_0)}{2h} + \frac{h^2 f'''(\xi_0)}{3}$ <p style="text-align: center;"><i>Différence avant d'ordre 2</i></p>	(6.11)
$f'(x_1) = \frac{f(x_2) - f(x_0)}{2h} - \frac{h^2 f'''(\xi_1)}{6}$ <p style="text-align: center;"><i>Différence centrée d'ordre 2</i></p>	
$f'(x_2) = \frac{3f(x_2) - 4f(x_1) + f(x_0)}{2h} + \frac{h^2 f'''(\xi_2)}{3}$ <p style="text-align: center;"><i>Différence arrière d'ordre 2</i></p>	

Les termes d'erreur de ces formules aux différences finies découlent tous de la relation 6.5 lorsqu'on pose successivement $i = 0, 1$ et 2 . Pour $i = 0$, on peut utiliser directement l'équation 6.6. Les points ξ_0 , ξ_1 et ξ_2 sont situés quelque part dans l'intervalle $[x_0, x_2]$ et sont inconnus (voir les exercices de fin de chapitre).

Remarque 6.4

Toutes ces formules aux différences sont d'ordre 2. Les mentions avant, centrée et arrière renvoient au point où l'on calcule la dérivée et aux points utilisés pour la calculer. Ainsi, la différence avant est évaluée en x_0 sur la base des valeurs situées vers l'avant, soit en x_1 et en x_2 . La différence arrière fixe la dérivée en $x = x_2$ avec l'appui des valeurs de la fonction en x_0 et en x_1 . La différence centrée, quant à elle, fait intervenir des valeurs situées de part et d'autre de x_1 .

La figure 6.1 illustre les différentes possibilités. Pour les différences d'ordre 1, on estime la dérivée par la pente du segment de droite joignant les points $(x_0, f(x_0))$ et $(x_1, f(x_1))$. Dans le cas des différences d'ordre 2, on

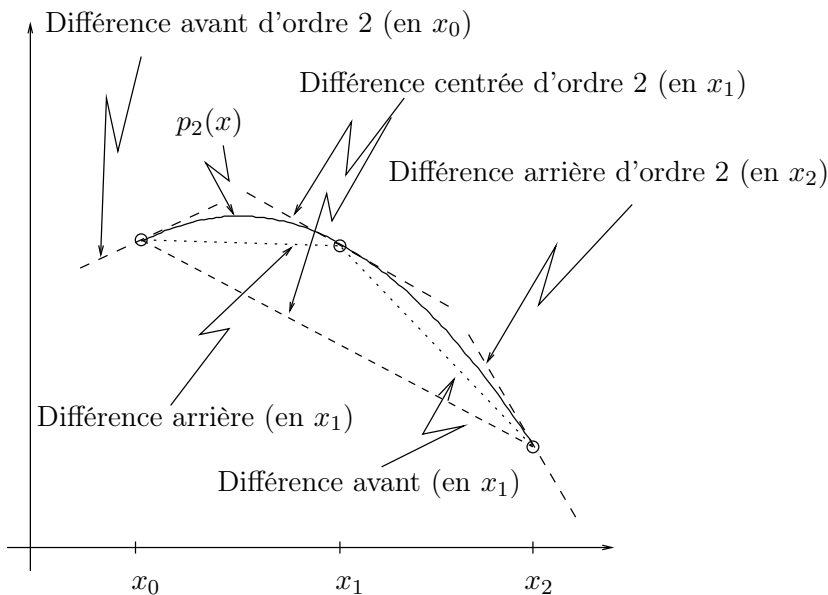


Figure 6.1 – Interprétation géométrique des formules aux différences

détermine un polynôme de degré 2 dont la pente en x_0 , en x_1 et en x_2 donne respectivement les différences avant, centrée et arrière. ♦

On peut aussi convenir de toujours évaluer la dérivée en x . Dans ce cas, on utilise les valeurs de $f(x+h)$ et de $f(x+2h)$ pour la différence avant et les valeurs de $f(x+h)$ et de $f(x-h)$ pour la différence centrée. En ce qui concerne le terme d'erreur, on ne retient que son ordre. Les tableaux suivants résument la situation.

Formules de différences finies d'ordre 1 pour $f'(x)$	
$f'(x) = \frac{f(x+h) - f(x)}{h} + O(h)$ <p style="text-align: center;"><i>Différence avant d'ordre 1</i></p>	(6.12)
$f'(x) = \frac{f(x) - f(x-h)}{h} + O(h)$ <p style="text-align: center;"><i>Différence arrière d'ordre 1</i></p>	

Chapitre 7

Équations différentielles

7.1 Introduction

La résolution numérique des équations différentielles est probablement le domaine de l'analyse numérique où les applications sont les plus nombreuses. Que ce soit en mécanique des fluides, en transfert de chaleur ou en analyse de structures, on aboutit souvent à la résolution d'équations différentielles, de systèmes d'équations différentielles ou plus généralement d'équations aux dérivées partielles.

Le problème du pendule abordé au chapitre 1 trouvera ici une solution numérique qui sera par la suite analysée et comparée à d'autres solutions approximatives ou quasi analytiques. Parmi leurs avantages, les méthodes numériques permettent d'étudier des problèmes complexes pour lesquels on ne connaît pas de solution analytique, mais qui sont d'un grand intérêt pratique.

Dans ce chapitre comme dans les précédents, les diverses méthodes de résolution proposées sont d'autant plus précises qu'elles sont d'ordre élevé. Nous amorçons l'exposé par des méthodes relativement simples ayant une interprétation géométrique. Elles nous conduiront progressivement à des méthodes plus complexes telles les méthodes de Runge-Kutta d'ordre 4, qui permettent d'obtenir des résultats d'une grande précision. Nous considérons principalement les équations différentielles avec conditions initiales, mais nous ferons une brève incursion du côté des équations différentielles avec conditions aux limites par le biais des méthodes de tir et de différences finies.

Nous prenons comme point de départ la formulation générale d'une équation différentielle d'ordre 1 avec condition initiale. La tâche consiste à déterminer une fonction $y(t)$ solution de :

$$\begin{cases} y'(t) &= f(t, y(t)) \\ y(t_0) &= y_0 \end{cases} \quad (7.1)$$

La variable indépendante t représente très souvent (mais pas toujours) le temps. La variable dépendante est notée y et dépend bien sûr de t . La fonction f est pour le moment une fonction quelconque de deux variables que nous supposons suffisamment différentiable. La condition $y(t_0) = y_0$ est la condition initiale et en quelque sorte l'état de la solution au moment où l'on commence à s'y intéresser. Il s'agit d'obtenir $y(t)$ pour $t \geq t_0$, si l'on cherche une so-

lution analytique, ou une approximation de $y(t)$, si l'on utilise une méthode numérique.

Définition 7.1

L'équation différentielle 7.1 est dite *d'ordre 1*, car seule la dérivée d'ordre 1 de la variable dépendante $y(t)$ est présente. Si des dérivées de $y(t)$ d'ordre 2 apparaissaient dans l'équation différentielle 7.1, on aurait une équation d'ordre 2, et ainsi de suite.

Commençons par présenter quelques exemples d'équations différentielles avec condition initiale.

Exemple 7.2

Soit l'équation différentielle du premier ordre :

$$\begin{cases} y'(t) = t \\ y(0) = 1 \end{cases} \quad (7.2)$$

Voilà certainement l'un des exemples les plus simples que l'on puisse imaginer. En intégrant de chaque côté, on obtient :

$$\int y'(t) dt = \int t dt$$

c'est-à-dire :

$$y(t) = \frac{t^2}{2} + C$$

où C est une constante. Cette dernière expression est la *solution générale de l'équation différentielle* en ce sens qu'elle satisfait $y'(t) = t$, quelle que soit la constante C . Pour déterminer la constante C , il suffit d'imposer la condition initiale :

$$y(0) = 1 = C$$

La *solution particulière* est alors :

$$y(t) = \frac{t^2}{2} + 1$$

qui vérifie à la fois l'équation différentielle et la condition initiale. ■

Exemple 7.3

Soit l'équation différentielle :

$$\begin{cases} y'(t) = ty(t) \\ y(1) = 2 \end{cases} \quad (7.3)$$

Il ne suffit pas dans ce cas d'intégrer les deux côtés de l'équation pour obtenir la solution. On doit d'abord séparer les variables en écrivant par exemple :

$$\frac{dy}{dt} = t y(t) \quad \text{qui devient} \quad \frac{dy}{y} = t dt$$

Les variables étant séparées, on peut maintenant faire l'intégration :

$$\ln y = \frac{t^2}{2} + C \text{ ou encore } y(t) = Ce^{\frac{t^2}{2}}$$

qui est la solution générale. On obtient la solution particulière en imposant la condition initiale :

$$y(1) = 2 = Ce^{\frac{1}{2}}$$

ce qui signifie que :

$$C = 2e^{-\frac{1}{2}}$$

et donc que la solution particulière est :

$$y(t) = 2e^{\frac{(t^2-1)}{2}}$$

■

Les ouvrages de Simmons (réf. [33]) et de Derrick et Grossman (réf. [11]) contiennent d'autres exemples d'équations différentielles. Notre propos concerne plutôt les méthodes numériques de résolution de ces équations différentielles. À cet égard, nous suivons l'approche de Burden et Faires (réf. [5]), notamment en ce qui concerne la notion d'erreur de troncature locale qui indique l'ordre de précision de la méthode utilisée.

Avec les outils numériques de résolution d'équations différentielles, il n'est plus possible d'obtenir une solution pour toutes les valeurs de la variable indépendante t . On obtient plutôt une approximation de la solution analytique seulement à *certaines valeurs de t notées t_i* et distancées d'une valeur $h_i = t_{i+1} - t_i$. Dans la plupart des méthodes présentées, cette distance est constante pour tout i et est notée h . On appelle h le *pas de temps*.

Remarque 7.4

On note $y(t_i)$ la *solution analytique* de l'équation différentielle 7.1 en $t = t_i$. On note y_i la *solution approximative* en $t = t_i$ obtenue à l'aide d'une méthode numérique. ♦

7.2 Méthode d'Euler explicite

La méthode d'Euler explicite est de loin la méthode la plus simple de résolution numérique d'équations différentielles ordinaires. Elle possède une belle interprétation géométrique et son emploi est facile. Toutefois, elle est relativement peu utilisée en raison de sa faible précision. On la qualifie d'explicite car elle ne nécessite pas de résolution d'équation non linéaire contrairement à la méthode d'Euler dite implicite que nous verrons plus loin (voir la section 7.8.1).

Reprenons l'équation différentielle 7.1 et considérons plus attentivement la condition initiale $y(t_0) = y_0$. Le but est maintenant 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 :

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

et est illustrée à la figure 7.1. En $t = t_1$, on a :

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

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

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

Il est important de noter que, le plus souvent, $y_1 \neq y(t_1)$. Cette inégalité n'a rien pour étonner, mais elle a des conséquences sur la suite du raisonnement. En effet, si l'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 on possède 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 (fig. 7.1) :

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

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 :

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

Remarque 7.5

Le développement qui précède met en évidence une propriété importante des méthodes numériques de résolution des équations différentielles. En effet, l'erreur introduite à la première itération a des répercussions sur les calculs de la deuxième itération, ce qui signifie que les erreurs se propagent d'une itération à l'autre. Il en résulte de façon générale que l'erreur :

$$|y(t_n) - y_n|$$

augmente légèrement avec n . ♦

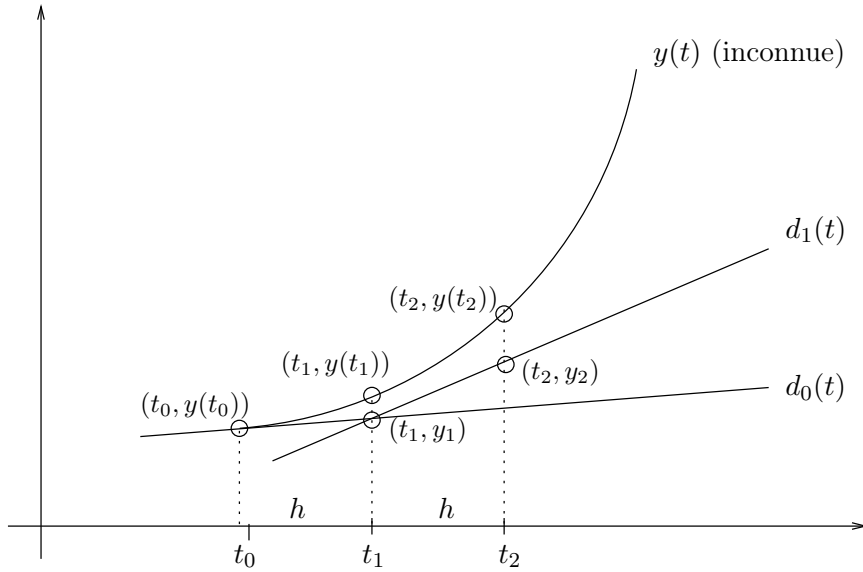


Figure 7.1 – Méthode d'Euler explicite

On en arrive donc à l'algorithme suivant.

Algorithme 7.6 : Méthode d'Euler explicite

1. Étant donné un pas de temps h , une condition initiale (t_0, y_0) et un nombre maximal d'itérations N
2. Pour $0 \leq n \leq N$:

$y_{n+1} = y_n + hf(t_n, y_n)$
 $t_{n+1} = t_n + h$
 Écrire t_{n+1} et y_{n+1}
3. Arrêt

▲

Exemple 7.7

Soit l'équation différentielle (voir Fortin et Pierre, réf. [18]) :

$$y'(t) = -y(t) + t + 1$$

et la condition initiale $y(0) = 1$. On a donc $t_0 = 0$ et $y_0 = 1$ et l'on prend un pas de temps $h = 0,1$. De plus, on a :

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

On peut donc utiliser la méthode d'Euler explicite et obtenir successivement des approximations de $y(0,1), y(0,2), y(0,3) \dots$, notées $y_1, y_2, y_3 \dots$. Le premier pas de temps produit :

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

Le deuxième pas de temps 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 ensuite à :

$$\begin{aligned} y_3 &= y_2 + hf(t_2, y_2) = 1,01 + 0,1f(0,2, 1,01) \\ &= 1,01 + 0,1(-1,01 + 0,2 + 1) = 1,029 \end{aligned}$$

Le tableau suivant rassemble les résultats des dix premiers pas de temps. La solution analytique de cette équation différentielle est $y(t) = e^{-t} + t$, ce qui permet de comparer les solutions numérique et analytique et de constater la croissance de l'erreur. On peut aussi comparer les résultats à la figure 7.2.

Méthode d'Euler explicite : $y'(t) = -y(t) + t + 1$			
t_i	$y(t_i)$	y_i	$ y(t_i) - y_i $
0,0	1,000 000	1,000 000	0,000 000
0,1	1,004 837	1,000 000	0,004 837
0,2	1,018 731	1,010 000	0,008 731
0,3	1,040 818	1,029 000	0,011 818
0,4	1,070 302	1,056 100	0,014 220
0,5	1,106 531	1,090 490	0,016 041
0,6	1,148 812	1,131 441	0,017 371
0,7	1,196 585	1,178 297	0,018 288
0,8	1,249 329	1,230 467	0,018 862
0,9	1,306 570	1,287 420	0,019 150
1,0	1,367 879	1,348 678	0,019 201

■

Les résultats précédents nous amènent à parler de précision et donc d'erreur. La figure 7.2 montre une légère différence entre la solution numérique et la solution analytique. On peut se demander comment se comporte cette erreur en fonction du pas de temps h . La définition qui suit aidera à apporter une réponse. Elle s'applique à la plupart des méthodes étudiées dans ce chapitre.

Définition 7.8

Une méthode de résolution d'équations différentielles est dite à *un pas* si elle est de la forme :

$$y_{n+1} = y_n + h\phi(t_n, y_n) \tag{7.4}$$

où ϕ est une fonction quelconque. Une telle relation est appelée *équation aux différences*. La méthode est à *un pas* si, pour obtenir la solution en $t = t_{n+1}$, on doit utiliser la solution numérique au temps t_n seulement. On désigne *méthodes à pas multiples* les méthodes qui exigent également la solution numérique aux temps $t_{n-1}, t_{n-2}, t_{n-3} \dots$.

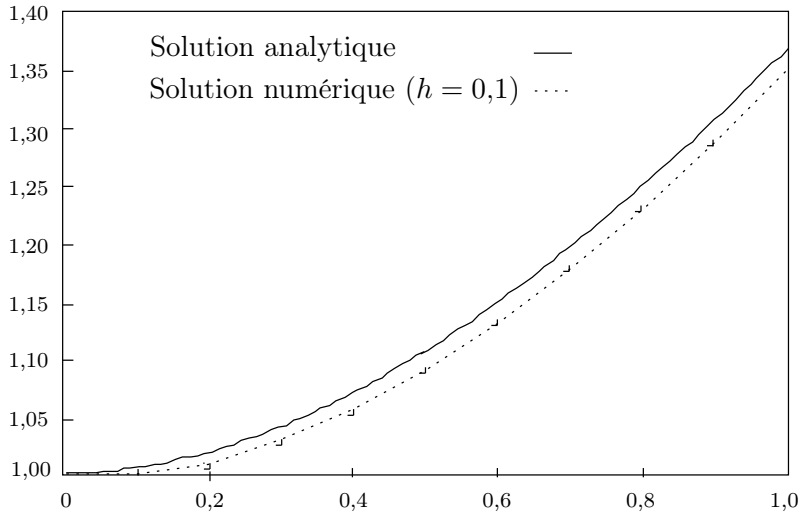


Figure 7.2 – Méthode d'Euler explicite : $y'(t) = -y(t) + t + 1$ pour $y(0) = 1$

La méthode d'Euler explicite est bien sûr une méthode à un pas où $\phi(t, y) = f(t, y)$. Dans ce chapitre, l'attention est principalement portée sur les méthodes à un pas. Nous pouvons maintenant définir l'ordre de convergence de ces méthodes.

Définition 7.9

On dira qu'un schéma à un pas converge à l'ordre p si :

$$\max_{1 \leq n \leq N} |y(t_n) - y_n| = O(h^p) \quad (7.5)$$

où N est le nombre total de pas de temps.

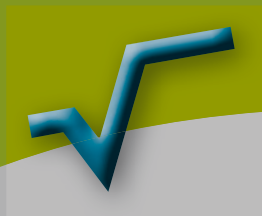
L'ordre de convergence d'une méthode à un pas dépend de l'erreur commise à chaque pas de temps via l'erreur de troncature locale que nous allons maintenant définir.

Définition 7.10

L'erreur de troncature locale au point $t = t_n$ est définie par :

$$\tau_{n+1}(h) = \frac{y(t_{n+1}) - y(t_n)}{h} - \phi(t_n, y(t_n)) \quad (7.6)$$

L'erreur de troncature locale mesure la précision avec laquelle la solution analytique vérifie l'équation aux différences 7.4.



Analyse numérique pour ingénieurs

Quatrième édition

Depuis plusieurs années, l'analyse numérique connaît un essor considérable et la plupart des facultés de sciences et de génie offrent au moins un cours d'introduction à cette discipline. La maîtrise de cet outil extrêmement performant est devenue indispensable dans la formation scientifique en général, et en particulier dans celle des ingénieurs, puisqu'elle permet d'aborder et de résoudre des problèmes dont la solution est inimaginable par les méthodes analytiques classiques. Ce livre couvre notamment l'analyse d'erreurs, les racines d'équations algébriques, les systèmes d'équations linéaires et non linéaires, les techniques d'interpolation, la différentiation et l'intégration numériques ainsi que les systèmes d'équations différentielles ordinaires. L'auteur met l'accent sur la compréhension profonde des méthodes proposées plutôt que sur la programmation, en présentant chaque thème à l'aide d'exemples, de figures, de tableaux et d'applications.

Ce livre s'adresse aux étudiants en sciences et en génie ainsi qu'aux ingénieurs et aux scientifiques qui désirent acquérir des connaissances et des habiletés de base dans le domaine de l'analyse numérique.

André Fortin est professeur au Département de mathématiques et de statistique de l'Université Laval à Québec. Il a enseigné au Département de mathématiques et de génie industriel de l'École Polytechnique de Montréal de 1984 à 2000. Il est également directeur du Groupe interdisciplinaire de recherche en éléments finis (GIREF) qui réunit des chercheurs provenant principalement de l'Université Laval et de l'École Polytechnique. Ce centre s'intéresse aux aspects théoriques et pratiques de la modélisation numérique et de ses applications industrielles.

ISBN : 978-2-553-01622-6



9 782553 016226



PRESSES INTERNATIONALES
POLYTECHNIQUE

www.polymtl.ca/pub

Extrait de la publication