

Vincent Manet

Méthode des éléments finis

*Vulgarisation des aspects mathématiques
et illustration de la méthode*

Vincent Manet — 2013 (Ceci est la version « livre » de ce document)

Ce document est sous licence Creative Commons 3.0 France :

- paternité ;
- pas d'utilisation commerciale ;
- partage des conditions initiales à l'identique ;

<http://creativecommons.org/licenses/by-nc-sa/3.0/deed.fr>



Introduction

Dans ce (de moins en moins court) document, plutôt à destination d'ingénieurs mécaniciens connaissant déjà la méthode des éléments finis, nous allons essayer de faire une présentation un peu plus théorique que ce qui leur est généralement proposé (et qui est quand même souvent de type « preuve par les mains », ce qui occulte trop de points).

Nous ne ferons appel qu'à des notions mathématiques de bases généralement déjà vues pour la plupart en taupe (ou en tout début de cycle d'ingé)... bien que des compléments que l'on peut qualifier d'élémentaires nous aient été demandés et aient été inclus.

Nous espérons, grâce à cette présentation théorique montrer toute la souplesse et la puissance de la méthode, afin de permettre au lecteur d'envisager d'autres simulations que celles qu'il a pu déjà réaliser par le passé.

But du document

Le but initial était de *présenter brièvement la théorie mathématique* derrière les éléments finis afin que les ingénieurs utilisant cette méthode puisse en envisager toutes les applications, ainsi que de *couvrir les aspects qui, selon nous, devraient être connus de tout ingénieur mécanicien impliqué ou intéressé par le calcul numérique.*

Toutefois, il s'envisage comme support de référence à plusieurs cours, cours qui ne portent pas sur tous les aspects traités dans ce document, et pendant lesquels les aspects pratiques sont plus développés (avec mise en situation sur machine).

Même si nous avons voulu rester le plus succinct possible, l'introduction de notions de proche en proche a conduit à un document fait aujourd'hui une certaine taille (par exemple, nous avons besoins des espaces de Sobolev, mais comment les introduire sans parler des espaces de Lebesgue, mais comment les introduire sans parler...).

Aussi le document a-t-il finalement été découpé en plusieurs parties : un survol des notions mathématiques, puis le traitement du problème continu constituent l'ossature théorique nécessaire à asseoir la MEF sur un socle solide. La discrétisation par éléments finis à proprement parler n'est abordé qu'ensuite, et d'ailleurs un seul chapitre suffirait à en faire le tour... sauf à entrer plus dans le détail concernant « ce qui fâche » : homogénéisation, non linéarité, dynamique, ce qui est fait dans des chapitres séparés.

Enfin, d'autres méthodes sont abordées car également très employées aujourd'hui. Aussi est-il indispensable selon nous d'en avoir entendu parlé et d'en connaître les principales notions (BEM, FEEC...).

En annexes, se trouve un petit fourre-tout comprenant des choses censées être maîtrisées depuis la taupe (mais qui parfois nous sont demandées) et les compléments qui alourdiraient encore les propos précédents.

Certaines notions (essentiellement de topologie) ne sont pas présentées dans ce document. Il nous a semblé que le lecteur devait avoir quelques souvenirs de ce qu'est un ouvert, un fermé, l'adhérence, la densité... Par ailleurs, leur nom peut être suffisamment évocateur pour se passer d'une définition formelle dans le contexte de ce document.

Attention, ce document n'est pas un document de mathématiques, il ne contient d'ailleurs aucune preuve. C'est, dans ces deux premières parties, un document de vulgarisation de notions mathématiques nécessaires à une bonne compréhension de la méthode des éléments finis.

Nous avons voulu réaliser un survol des notions importantes, mais malgré tout, afin de ne pas être parfois trop laconique, nous avons un peu débordé.

En fin de document, un petit index des noms propres permettra au lecteur de replacer les divers développements mentionnés dans l'histoire... Il se peut qu'il subsistent quelques erreurs, notamment au niveau des nationalités mentionnées, car il n'est pas toujours aisé de déterminer rapidement cette information (et nous ne connaissons pas toutes les biographies des personnes citées).

Ce document a été réalisé très rapidement, et de manière extrêmement hachée. Il comporte forcément encore beaucoup de fautes : merci de m'en faire part.

Démarche de l'ingénieur numéricien

En préambule à ce document, nous tenions à synthétiser la démarche complète de l'ingénieur numéricien :

- Modélisation / mise en équations – Construction du problème continu (système d'EDP).
- Analyse mathématique du problème posé – Existence, unicité, propriétés des solutions.
- Conception d'une méthode numérique – Construction d'un problème discrétisé.
- Analyse numérique – Questions de stabilité, convergence, précision.
- Algorithmique – Choix de méthodes de résolution en dimension finie.
- Mise en œuvre sur ordinateur – Programmation.
- Pre et Post Traitement (maillages / visualisation) – Interpolation, extrapolation, outils de la CAO.

Tous ces points ne seront évidemment pas abordés dans ce document !

Remerciements :

Nous n'avions pas prévu de réaliser une deuxième version aussi rapidement. Celle-ci existe suite aux sollicitations de Mathias Legrand. C'est lui qui a développé les macros nécessaires à l'amélioration très très nette de la qualité typographique (environnements pour les notes historiques, les théorèmes, lemmes...).

C'est également pourquoi coexistent aujourd'hui deux versions (mais issues du même code source) : l'une que nous appelons « version cours » (plus en accord avec ce que nous proposons en cours), et l'autre « version livre », plus proche d'un ouvrage.

Table des matières

Introduction	3
But du document	3
Démarche de l'ingénieur numéricien	4
Table des matières	5

I	ANNEXES
A	Interpolation et approximation 9
A.1	Quelques bases polynomiales 10
A.1.1	Motivation 10
A.1.2	Orthogonalité 11
A.1.3	Base naturelle 11
A.1.4	Polynômes de Lagrange 11
A.1.5	Polynômes d'Hermite 12
A.1.6	Polynômes de Legendre 12
A.1.7	Polynômes de Tchebychev 13
A.1.8	Polynômes de Laguerre 14
A.1.9	Polynômes de Bernstein 15
A.2	Interpolation polynomiale 16
A.2.1	Interpolation de Lagrange 16
A.2.2	Interpolation par Spline 16
A.2.3	Interpolation d'Hermite 16
A.3	Méthodes d'approximation 17
A.3.1	Courbe de Bézier 17
A.3.2	B-Spline 18
A.3.3	B-splines rationnelles non uniformes 18
B	Intégration numérique 21
B.1	Méthodes de Newton-Cotes 21
B.1.1	Méthode des rectangles 21
B.1.2	Méthode des trapèzes 21
B.1.3	Méthode de Simpson 22
B.2	Méthodes de quadrature de Gauß 23
C	Résolution des équations différentielles ordinaires 27
C.1	Résolution exacte des équations différentielles linéaires 27
C.1.1	Équation différentielle linéaire scalaire d'ordre 1 27
C.1.2	Équation différentielle du premier ordre à variables séparées 29
C.1.3	Équation différentielle linéaire d'ordre deux 30

C.2	Résolution numérique	32
C.2.1	Méthode d'Euler, Runge-Kutta ordre 1	32
C.2.2	Méthode de Runge-Kutta d'ordre 2	32
C.2.3	Méthode de Runge-Kutta d'ordre 4	33
C.2.4	Méthode de Newmark	33
D	Méthode de Newton Raphson	35
D.1	Présentation	35
D.2	Algorithme	36

Annexe A

Interpolation et approximation

L'*interpolation* est une opération consistant à approcher une courbe qui n'est connue que par la donnée d'un nombre fini de points (ou une fonction à partir de la donnée d'un nombre fini de valeurs).

Ainsi, l'interpolation numérique sert souvent à « faire émerger une courbe parmi des points ». Il s'agit de toutes les méthodes développées afin de mieux prendre en compte les erreurs de mesure, i.e. d'exploiter des données expérimentales pour la recherche de lois empiriques. Nous citerons par exemple la régression linéaire et la méthode des moindres carrés souvent bien maîtrisées. On demande à la solution du problème d'interpolation de *passer par les points prescrits*, voire, suivant le type d'interpolation, de vérifier des propriétés supplémentaires (de continuité, de dérivabilité, de tangence en certains points...). Toutefois, parfois on ne demande pas à ce que l'approximation passe exactement par les points prescrits. On parle alors plutôt d'approximation.

Histoire

Le jour du Nouvel An de 1801, l'astronome italien Giuseppe Piazzi a découvert l'astéroïde Cérès ; il a suivi sa trajectoire jusqu'au 14 février 18012. Durant cette année, plusieurs scientifiques ont tenté de prédire sa trajectoire sur la base des observations de Piazzi mais à cette époque, la résolution des équations non linéaires de Kepler de la cinématique est un problème très difficile. La plupart des prédictions sont erronées et le seul calcul suffisamment précis pour permettre au baron Franz Xaver von Zach de localiser à nouveau Cérès à la fin de l'année est celui de Gauß, (alors âgé de 24 ans). Gauß avait déjà réalisé l'élaboration des concepts fondamentaux en 1795, lorsqu'il avait 18 ans. Cependant, sa méthode des moindres carrés ne fut publiée qu'en 1809 dans le tome 2 de ses travaux sur la Mécanique céleste *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium*. Le mathématicien français Adrien-Marie Legendre a développé indépendamment la même méthode en 1805. Le mathématicien américain Robert Adrain a publié en 1808 une formulation de la méthode.

En 1829, Gauß a pu donner les raisons de l'efficacité de cette méthode : celle-ci est optimale à l'égard de bien des critères. Cet argument est maintenant connu sous le nom de théorème de Gauß-Markov. Ce théorème dit que dans un modèle linéaire dans lequel les erreurs ont une espérance nulle, sont non corrélées et dont les variances sont égales, le meilleur estimateur linéaire non biaisé des coefficients est l'estimateur des moindres carrés. Plus généralement, le meilleur estimateur linéaire non biaisé d'une combinaison linéaire des coefficients est son estimateur par les moindres carrés. On ne suppose pas que les erreurs possèdent une loi normale, ni qu'elles sont indépendantes mais seulement non corrélées, ni qu'elles possèdent la même loi de probabilité.



Piazzi



Gauß



Legendre



Adrain



Markov

L'approximation étant l'utilisation de méthodes permettant d'approcher une fonction mathématique par une suite de fonctions qui convergent dans un certain espace fonctionnel, on voit donc que

ce qui a été fait dans la deuxième partie ressort bien de cela : on cherche une fonction généralement notée u qui n'est pas connue explicitement mais solution d'une équation différentielle ou d'une équation aux dérivées partielles, et l'on cherche à construire une suite de problèmes plus simples, que l'on sait résoudre à chaque étape, et telle que la suite des solutions correspondantes converge vers la solution cherchée. L'approximation peut servir aussi dans le cas où la fonction considérée est connue : on cherche alors à la remplacer par une fonction plus simple, plus régulière ou ayant de meilleures propriétés. L'intégration numérique sera détaillée un peu plus au chapitre B. Ce sont les méthodes d'approximation numériques qui sont utilisées.

Pour en revenir à l'interpolation, la méthode des éléments finis est en elle-même une méthode d'interpolation (globale, basée sur des interpolations locales). On peut citer quelques méthodes d'interpolation telle que l'interpolation linéaire (dans laquelle deux points successifs sont reliés par un segment), l'interpolation cosinus (dans laquelle deux points successifs sont considérés comme les pics d'un cosinus. L'interpolation cubique ou spline (dans laquelle un polynôme de degré 3 passe par quatre points successifs : selon le type de continuité demandée plusieurs variantes existent) et de manière générale, l'interpolation polynomiale, est abordée ci-dessous.

Faisons d'emblée une mise en garde : la plus connue des interpolations polynomiale, l'interpolation lagrangienne (approximation par les polynômes de Lagrange, découverte initialement par Waring et redécouverte par Euler) peut fort bien diverger même pour des fonctions très régulières. C'est le phénomène de Runge : contrairement à l'intuition, l'augmentation du nombre de points d'interpolation ne constitue pas nécessairement une bonne stratégie d'approximation avec certaines fonctions (même infiniment dérivables).

Dans le cas où l'on travaille sur le corps des complexes, une méthode d'approximation d'une fonction analytique par une fonction rationnelle est l'approximant de Padé. Cela correspond à un développement limité qui approche la fonction par un polynôme. Tout comme les développements limités forment une suite appelée série entière, convergeant vers la fonction initiale, les approximants de Padé sont souvent vus comme une suite, s'exprimant sous la forme d'une fraction continue dont la limite est aussi la fonction initiale. En ce sens, ces approximants font partie de la vaste théorie des fractions continues. Les approximants offrent un développement dont le domaine de convergence est parfois plus large que celui d'une série entière. Ils permettent ainsi de prolonger des fonctions analytiques et d'étudier certains aspects de la question des séries divergentes. En théorie analytique des nombres, l'approximant permet de mettre en évidence la nature d'un nombre ou d'une fonction arithmétique comme celle de la fonction zêta de Riemann. Dans le domaine du calcul numérique, l'approximant joue un rôle, par exemple, pour évaluer le comportement d'une solution d'un système dynamique à l'aide de la théorie des perturbations. L'approximant de Padé a été utilisé pour la première fois par Euler pour démontrer l'irrationalité de e , la base du logarithme népérien. Une technique analogue a permis à Johann Heinrich Lambert de montrer celle de π .

A.1 Quelques bases polynomiales

A.1.1 Motivation

Lorsque l'on souhaite approximer une courbe par une autre, recourir aux polynômes semble une voie naturelle. Le théorème de Taylor (1715) montre qu'une fonction plusieurs fois dérivable au voisinage d'un point peut être approximée par une fonction polynôme dont les coefficients dépendent uniquement des dérivées de la fonction en ce point. Le *théorème d'approximation de Weierstrass* en analyse réelle dit que toute fonction continue définie sur un segment peut être approchée uniformément par des fonctions polynômes. Le théorème de Stone-Weierstrass généralise ce résultat aux fonctions continues définies sur un espace compact et à valeurs réelles, en remplaçant l'algèbre des polynômes par une algèbre de fonctions qui sépare les points et contient au moins une fonction constante non nulle.

L'interpolation polynomiale consiste donc à trouver un polynôme passant par un ensemble de

points donnés.

A.1.2 Orthogonalité

Une *suite de polynômes orthogonaux* est une suite infinie de polynômes $P_0(x), P_1(x), \dots$ à coefficients réels, dans laquelle chaque $P_n(x)$ est de degré n , et telle que les polynômes de la suite sont orthogonaux deux à deux pour un produit scalaire de fonctions donné. Le produit scalaire de fonctions le plus simple est l'intégrale du produit de ces fonctions sur un intervalle borné :

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx \quad (\text{A.1})$$

Plus généralement, on peut ajouter une fonction de poids $\varpi(x)$ dans l'intégrale. On notera bien que sur l'intervalle d'intégration $]a, b[$, la fonction poids W doit être à valeurs finies et strictement positives, et l'intégrale du produit de la fonction poids par un polynôme doit être finie (voir espaces L^p). Par contre, les bornes a et b peuvent être infinies. Il vient alors :

$$\langle f, g \rangle = \int_a^b f(x)g(x)\varpi(x)dx \quad (\text{A.2})$$

La norme associée est définie par $\|f\| = \sqrt{\langle f, f \rangle}$. Le produit scalaire fait de l'ensemble de toutes les fonctions de norme finie un espace de Hilbert. L'intervalle d'intégration est appelé *intervalle d'orthogonalité*.

A.1.3 Base naturelle

On rappelle que $1, X, \dots, X^n$ est une base de $K_n[X]$ en tant que polynômes échelonnés.

A.1.4 Polynômes de Lagrange

Connaissant $n+1$ points $(x_0, y_0), \dots, (x_n, y_n)$ d'abscisses distinctes, le *polynôme de Lagrange* est l'unique polynôme de degré n passant tous les points. Ce polynôme est trivialement défini par :

$$L(X) = \sum_{j=0}^n y_j \left(\prod_{i=0, i \neq j}^n \frac{X - x_i}{x_j - x_i} \right) \quad (\text{A.3})$$

Si on note :

$$L(X) = \sum_{j=0}^n y_j l_j(X) \quad (\text{A.4})$$

avec :

$$l_i(X) = \prod_{j=0, j \neq i}^n \frac{X - x_j}{x_i - x_j} = \frac{X - x_0}{x_i - x_0} \dots \frac{X - x_{i-1}}{x_i - x_{i-1}} \frac{X - x_{i+1}}{x_i - x_{i+1}} \dots \frac{X - x_n}{x_i - x_n} \quad (\text{A.5})$$

alors, on remarque que :

- l_i est de degré n pour tout i ;
- $l_i(x_j) = \delta_{i,j}, 0 \leq i, j \leq n$, i.e. $l_i(x_i) = 1$ et $l_i(x_j) = 0$ pour $j \neq i$

On en déduit immédiatement que $\forall i, L(x_i) = y_i$ qui est bien la propriété recherchée par construction.

A.1.5 Polynômes d'Hermite

Les *polynômes d'Hermite* sont définis comme suit :

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} \quad (\text{forme dite probabiliste}) \quad (\text{A.6})$$

ou :

$$\hat{H}_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \quad (\text{forme dite physique}) \quad (\text{A.7})$$

Les deux définitions sont liées par la propriété d'échelle suivante :

$$\hat{H}_n(x) = 2^{n/2} H_n(x\sqrt{2}) \quad (\text{A.8})$$

Les premiers polynômes d'Hermite sont les suivants : $H_0 = 1$, $H_1 = X$, $H_2 = X^2 - 1$, $H_3 = X^3 - 3X$, $H_4 = X^4 - 6X^2 + 3$, $H_5 = X^5 - 10X^3 + 15X$, $H_6 = X^6 - 15X^4 + 45X^2 - 15$. $\hat{H}_0 = 1$, $\hat{H}_1 = 2X$, $\hat{H}_2 = 4X^2 - 2$, $\hat{H}_3 = 8X^3 - 12X$, $\hat{H}_4 = 16X^4 - 48X^2 + 12$, $\hat{H}_5 = 32X^5 - 160X^3 + 120X$, $\hat{H}_6 = 64X^6 - 480X^4 + 720X^2 - 120$.

H_n est un polynôme de degré n . Ces polynômes sont orthogonaux pour la mesure μ de densité :

$$\frac{d\mu(x)}{dx} = \frac{e^{-x^2/2}}{\sqrt{2\pi}}. \quad (\text{A.9})$$

Ils vérifient :

$$\int_{-\infty}^{+\infty} H_n(x) H_m(x) e^{-x^2/2} dx = n! 2^n \sqrt{2\pi} \delta_{nm} \quad (\text{A.10})$$

où $\delta_{n,m}$ est le symbole de Kronecker. Ces polynômes forment donc une base orthogonale de l'espace de Hilbert $L^2(\mathbb{C}, \mu)$ des fonctions boréliennes telles que :

$$\int_{-\infty}^{+\infty} |f(x)|^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx < +\infty \quad (\text{A.11})$$

dans lequel le produit scalaire est donné par l'intégrale :

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(x) \overline{g(x)} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad (\text{A.12})$$

Des propriétés analogues sont vérifiées par les polynômes de Hermite sous leur forme physique.

A.1.6 Polynômes de Legendre

On appelle équation de Legendre l'équation :

$$\frac{d}{dx} \left[(1-x^2) \frac{dy}{dx} \right] + n(n+1)y = 0 \quad (\text{A.13})$$

On définit le *polynôme de Legendre* P_n par :

$$\frac{d}{dx} \left[(1-x^2) \frac{dP_n(x)}{dx} \right] + n(n+1)P_n(x) = 0, \quad P_n(1) = 1 \quad (\text{A.14})$$

La manière la plus simple de les définir est par la formule de récurrence de Bonnet : $P_0(x) = 1$, $P_1(x) = x$ et :

$$\forall n > 0, \quad (n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x) \quad (\text{A.15})$$

Les premiers polynômes de Legendre sont :

$$\begin{aligned}
P_0(x) &= 1 \\
P_1(x) &= x \\
P_2(x) &= \frac{1}{2}(3x^2 - 1) \\
P_3(x) &= \frac{1}{2}(5x^3 - 3x) \\
P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3) \\
P_5(x) &= \frac{1}{8}(63x^5 - 70x^3 + 15x)
\end{aligned} \tag{A.16}$$

Le polynôme P_n est de degré n . La famille $(P_n)_{n \leq \mathbb{N}}$ est une famille de polynômes à degrés étagés, elle est donc une base de l'espace vectoriel $\mathbb{R}_n[X]$. On remarquera la propriété suivante :

$$P_n(-x) = (-1)^n P_n(x) \tag{A.17}$$

qui donne en particulier $P_n(-1) = (-1)^n$ et $P_{2n+1}(0) = 0$. Les polynômes orthogonaux les plus simples sont les polynômes de Legendre pour lesquels l'intervalle d'orthogonalité est $[-1; 1]$ et la fonction poids est simplement la fonction constante de valeur 1 : ces polynômes sont orthogonaux par rapport au produit scalaire défini sur $\mathbb{R}[X]$ par :

$$\langle P, Q \rangle = \int_{-1}^{+1} P(x)Q(x)dx \quad \langle P_m, P_n \rangle = \int_{-1}^1 P_m(x)P_n(x)dx = 0 \quad \text{pour} \quad m \neq n \tag{A.18}$$

De plus, comme $(P_n)_{n \leq N}$ est une base de $\mathbb{R}_N[X]$, on a $P_{N+1} \in (\mathbb{R}_N[X])^\perp$:

$$\forall Q \in \mathbb{R}_N[X], \quad \int_{-1}^1 P_{N+1}(x)Q(x)dx = 0 \tag{A.19}$$

Le carré de la norme, dans $L^2([-1; 1])$, est :

$$\|P_n\|^2 = \frac{2}{2n+1}. \tag{A.20}$$

Ces polynômes peuvent servir à décomposer une fonction holomorphe, une fonction lipschitzienne ou à retrouver l'intégration numérique d'une fonction par la méthode de quadrature de Gauss-Legendre [chapitre ??].

A.1.7 Polynômes de Tchebychev

Les *polynômes de Tchebychev* servent pour la convergence des interpolations de Lagrange. Ils sont également utilisés dans le calcul de filtres de Tchebychev en électronique analogique.

Les polynômes de Tchebychev constituent deux familles de polynômes (notés T_n pour la première espèce et U_n pour la seconde) définis sur l'intervalle $[-1; 1]$ par les relations trigonométriques :

$$T_n(\cos(\theta)) = \cos(n\theta) \tag{A.21}$$

$$U_n(\cos(\theta)) = \frac{\sin((n+1)\theta)}{\sin \theta} \tag{A.22}$$

Ces deux suites sont définies par la relation de récurrence :

$$\forall n \in \mathbb{N}, \quad P_{n+2}(X) = 2X P_{n+1}(X) - P_n(X) \tag{A.23}$$

et les deux premiers termes :

$$T_0 = 1, T_1 = X \quad \text{pour la suite } T \quad (\text{A.24})$$

$$U_0 = 1, U_1 = 2X \quad \text{pour la suite } U \quad (\text{A.25})$$

Chacune de ces deux familles est une suite de polynômes orthogonaux par rapport à un produit scalaire de fonctions assorti d'une pondération spécifique.

Propriétés des polynômes de Tchebychev de première espèce

$$\forall n > 0, \quad T_n(x) = \frac{n}{2} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k \frac{(n-k-1)!}{k!(n-2k)!} (2x)^{n-2k} \quad (\text{A.26})$$

Les T_n sont orthogonaux pour le produit scalaire suivant :

$$\int_{-1}^1 \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}} dx = \begin{cases} 0 & \text{si } n \neq m \\ \pi & \text{si } n = m = 0 \\ \pi/2 & \text{si } n = m \neq 0. \end{cases} \quad (\text{A.27})$$

$$\forall n, \quad T_n(1) = 1, \quad \forall n, m \in \mathbb{N}, \quad \forall x \in \mathbb{R}, \quad T_n(T_m(x)) = T_{mn}(x) \quad (\text{A.28})$$

Les premiers polynômes de Tchebychev de première espèce sont $T_0 = 1$, $T_1 = x$, $T_2 = 2x^2 - 1$, $T_3 = 4x^3 - 3x$, $T_4 = 8x^4 - 8x^2 + 1$, $T_5 = 16x^5 - 20x^3 + 5x$, $T_6 = 32x^6 - 48x^4 + 18x^2 - 1$ et $T_7 = 64x^7 - 112x^5 + 56x^3 - 7x$.

Propriétés des polynômes de Tchebychev de seconde espèce

$$\forall n \geq 0, \quad U_n(x) = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k \binom{n-k}{k} (2x)^{n-2k} \quad (\text{A.29})$$

Les U_n sont orthogonaux pour le produit scalaire suivant :

$$\int_{-1}^1 U_n(x)U_m(x)\sqrt{1-x^2}dx = \begin{cases} 0 & \text{si } n \neq m \\ \pi/2 & \text{si } n = m \end{cases} \quad \text{et} \quad \forall n, \quad U_n(1) = n+1 \quad (\text{A.30})$$

Les premiers polynômes de Tchebychev de deuxième espèce sont $U_0 = 1$, $U_1 = 2x$, $U_2 = 4x^2 - 1$, $U_3 = 8x^3 - 4x$, $U_4 = 16x^4 - 12x^2 + 1$, $U_5 = 32x^5 - 32x^3 + 6x$, $U_6 = 64x^6 - 80x^4 + 24x^2 - 1$ et $U_7 = 128x^7 - 192x^5 + 80x^3 - 8x$.

A.1.8 Polynômes de Laguerre

Les *polynômes de Laguerre* apparaissent en mécanique quantique dans la partie radiale de la solution de l'équation de Schrödinger pour un atome à un électron. Ces polynômes sont les solutions de l'équation de Laguerre :

$$xy'' + (1-x)y' + ny = 0 \quad (\text{A.31})$$

qui est une équation différentielle linéaire du second ordre possédant des solutions non singulières si seulement si n est un entier positif.

Traditionnellement notés L_0, L_1, \dots ces polynômes forment une suite de polynômes qui peut être définie par la formule de Rodrigues :

$$L_n(x) = \frac{e^x}{n!} \frac{d^n}{dx^n} (e^{-x} x^n). \quad (\text{A.32})$$

Ils sont orthogonaux les uns par rapport aux autres pour le produit scalaire :

$$\langle f, g \rangle = \int_0^\infty f(x)g(x)e^{-x}dx. \quad (\text{A.33})$$

Cette propriété d'orthogonalité revient à dire que si X est une variable aléatoire distribuée exponentiellement avec la fonction densité de probabilité suivantes :

$$f(x) = \begin{cases} e^{-x} & \text{si } x > 0 \\ 0 & \text{si } x < 0 \end{cases} \quad (\text{A.34})$$

alors $E(L_n(X), L_m(X)) = 0$ si $n \neq m$. Les premiers polynômes de Laguerre sont $L_0 = 1, L_1 = -x + 1$ et :

$$L_2 = \frac{1}{2}(x^2 - 4x + 2), L_3 = \frac{1}{6}(-x^3 + 9x^2 - 18x + 6), L_4 = \frac{1}{24}(x^4 - 16x^3 + 72x^2 - 96x + 24) \quad (\text{A.35})$$

Il existe des polynômes de Laguerre généralisés dont l'orthogonalité peut être liée à une densité de probabilité faisant intervenir la fonction Gamma. Ils apparaissent dans le traitement de l'oscillateur harmonique quantique. Ils peuvent être exprimés en fonction des polynômes d'Hermite.

A.1.9 Polynômes de Bernstein

Les *polynômes de Bernstein* permettent de donner une démonstration constructive du théorème de Stone-Weierstrass. Dans le cadre de ce cours, nous les présentons surtout car ils sont utilisés dans la formulation générale des courbes de Bézier. Pour un degré n , il y a $n + 1$ polynômes de Bernstein B_0^n, \dots, B_n^n définis, sur l'intervalle $[0, 1]$ par :

$$B_i^n(u) = \binom{n}{i} u^i (1-u)^{n-i} \quad \text{où les } \binom{n}{i} \text{ sont les coefficients binomiaux.} \quad (\text{A.36})$$

Ces polynômes présentent quatre propriétés importantes :

— Partition de l'unité :

$$\sum_{i=0}^n B_i^n(u) = 1, \quad \forall u \in [0, 1] \quad (\text{A.37})$$

— Positivité :

$$B_i^n(u) \geq 0, \quad \forall u \in [0, 1], \quad \forall i \in 0, \dots, n \quad (\text{A.38})$$

— Symétrie :

$$B_i^n(u) = B_{n-i}^n(1-u), \quad \forall u \in [0, 1], \quad \forall i \in 0, \dots, n \quad (\text{A.39})$$

— Formule de récurrence :

$$B_i^n(u) = \begin{cases} (1-u)B_i^{n-1}(u), & i = 0 \\ (1-u)B_i^{n-1}(u) + uB_{i-1}^{n-1}(u), & \forall i \in 1, \dots, n-1, \\ uB_{i-1}^{n-1}(u), & i = n \end{cases} \quad \forall u \in [0, 1] \quad (\text{A.40})$$

On notera la grande ressemblance de ces polynômes avec la loi binomiale.

A.2 Interpolation polynomiale

A.2.1 Interpolation de Lagrange

Dans la version la plus simple (*interpolation lagrangienne*), on impose simplement que le polynôme passe par tous les points donnés. On obtient les polynômes de Lagrange tels que présentés juste avant. Le théorème de l'unisolvance (voir paragraphe ??) précise qu'il n'existe qu'un seul polynôme de degré n au plus défini par un ensemble de $n + 1$ points.

L'erreur d'interpolation lors de l'approximation d'une fonction f (donnée par les points $(x_i, y_i = f(x_i))$) par un polynôme de Lagrange p_n est donnée par une formule de type Taylor-Young : Si f est $n + 1$ fois continûment différentiable sur $I = [\min(x_0, \dots, x_n, x), \max(x_0, \dots, x_n, x)]$, alors :

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^n (x - x_i) \quad \text{avec } \xi \in I. \quad (\text{A.41})$$

Dans le cas particulier où $x_i = x_0 + ih$ (points uniformément répartis), il se produit en général une aggravation catastrophique de l'erreur d'interpolation, connue sous le nom de *phénomène de Runge* lorsqu'on augmente le nombre de points pour un intervalle $[x_0, x_n]$, donné (on a alors $\xi \in]-1, 1[$).

Pour limiter le *phénomène de Runge*, i.e. pour minimiser l'oscillation des polynômes interpolateurs, on peut utiliser les abscisses de Tchebychev au lieu de points équirépartis pour interpoler. Dans ce cas, on peut montrer que l'erreur d'interpolation décroît lorsque n augmente. On peut aussi préférer utiliser des splines pour approximer la fonction f (ce sont des polynômes par morceaux définis plus bas). Dans ce cas, pour améliorer l'approximation, on augmente le nombre de morceaux et non le degré des polynômes.

A.2.2 Interpolation par Spline

Une *spline* est une fonction définie par morceaux par des polynômes. Comme mentionné au dessus, la méthode des splines est souvent préférée à l'interpolation polynomiale, car on obtient des résultats similaires en se servant de polynômes ayant des degrés inférieurs, tout en évitant le phénomène de Runge. De plus, leur simplicité d'implémentation les rend très populaires et elles sont fréquemment utilisées dans les logiciels de dessin.

Une courbe spline est une fonction polynomiale par morceaux définie sur un intervalle $[a, b]$ divisé en sous intervalles $[t_{i-1}, t_i]$ tels que $a = t_0 < t_1 < \dots < t_{k-1} < t_k = b$. On la note $S : [a, b] \rightarrow \mathbb{R}$. Sur chaque intervalle $[t_{i-1}, t_i]$ on définit un polynôme $P_i : [t_{i-1}, t_i] \rightarrow \mathbb{R}$. Cela entraîne pour une spline à k intervalles : $S(t) = P_1(t), t_0 \leq t < t_1, S(t) = P_2(t), t_1 \leq t < t_2, \dots, S(t) = P_k(t), t_{k-1} \leq t \leq t_k$.

Le *degré de la spline* est défini comme étant celui du polynôme $P_i(t)$ de plus haut degré. Si tous les polynômes ont le même degré, on dit que la spline est uniforme. Dans le cas contraire, elle est non uniforme.

Tout polynôme étant C^∞ , la *continuité d'une spline* dépend de la continuité au niveau de la jointure des courbes polynômes. Si $\forall i$ tel que $1 \leq i \leq k$ et $\forall j$ tel que $0 \leq j \leq n$ l'égalité suivante est vérifiée :

$$P_i^{(j)}(t_i) = P_{i+1}^{(j)}(t_i). \quad (\text{A.42})$$

alors la spline est C^n

A.2.3 Interpolation d'Hermite

L'*interpolation d'Hermite* consiste à chercher un polynôme qui non seulement prend les valeurs fixées en les abscisses données, mais dont également la dérivée, donc la pente de la courbe, prend une valeur imposée en chacun de ces points. Naturellement, il faut pour cela un polynôme de degré supérieur au polynôme de Lagrange. On peut aussi imposer encore la valeur des dérivées secondes, troisièmes, etc. en chaque point. La démarche de l'interpolation newtonienne utilisant les différences divisées est particulièrement adaptée pour construire ces polynômes.

A.3 Méthodes d'approximation

A.3.1 Courbe de Bézier

Histoire

Pierre Bézier (ingénieur de l'École nationale supérieure d'arts et métiers en 1930 et de l'École supérieure d'électricité en 1931, il reçoit le titre de docteur en mathématiques de l'université de Paris en 1977) est connu pour son invention des courbes et surfaces de Bézier, couramment utilisées en informatique.

Entré chez Renault en 1933, il y fera toute sa carrière jusqu'en 1975 au poste de directeur des méthodes mécaniques. Il y conçoit, en 1945, des machines transferts pour la ligne de fabrication des Renault 4CV, et, en 1958, l'une des premières machines à commande numérique d'Europe, une fraiseuse servant aux maquettes. Sa préoccupation était de créer un moyen simple et puissant pour modéliser des formes et faciliter la programmation des machines à commande numérique. Le problème auquel il s'attaque est celui de la modélisation des surfaces en trois dimensions, les commandes numériques se contentant jusqu'alors de courbes en deux dimensions. La solution qu'il cherche est celle d'une interface intuitive accessible à tout utilisateur. Il décide de considérer classiquement les surfaces comme une transformation de courbes. Son exigence de s'adapter au dessinateur et non de contraindre le dessinateur à devenir calculateur, l'amène à une inversion géniale, déduire le calcul à partir du dessin et non le dessin à partir du calcul. Il invente alors la poignée de contrôle, curseur de déplacement des courbes d'un dessin informatisé transmettant automatiquement les variations de coordonnées au processeur. Ces poignées de contrôle sont toujours utilisées aujourd'hui.

Les courbes de Bézier sont des courbes polynomiales paramétriques. Elles ont de nombreuses applications dans la synthèse d'images et le rendu de polices de caractères (Pierre Bézier a travaillé sur les deux sujets). Ses recherches aboutirent à un logiciel, Unisurf, breveté en 1966. Il est à la base de tous les logiciels créés par la suite. Les concepts de CAO et de CFAO venaient de prendre forme. Ultérieurement, l'un des développeurs d'Apple, John Warnock, réutilise les travaux de Pierre Bézier pour élaborer un nouveau langage de dessin de polices : Postscript. Il crée ensuite en 1982, avec Charles M. Geschke, la société Adobe pour lancer un logiciel de dessin dérivé de ces résultats : Illustrator.

Notons que les splines existaient avant Bézier, mais leur défaut était de changer d'aspect lors d'une rotation de repère, ce qui les rendait inutilisables en CAO. Bézier partit d'une approche géométrique fondée sur la linéarité de l'espace euclidien et la théorie, déjà existante, du barycentre : si la définition est purement géométrique, aucun repère n'intervient puisque la construction en est indépendante. Les splines conformes aux principes de Bézier seront par la suite nommées B-splines.

Pour $n + 1$ points de contrôle ($\mathbf{P}_0, \dots, \mathbf{P}_n$) on définit une *courbe de Bézier* par l'ensemble des points :

$$b(t) = \sum_{i=0}^n B_i^n(t) \cdot \mathbf{P}_i \quad \text{avec} \quad t \in [0, 1] \quad (\text{A.43})$$

où les B_i^n sont les polynômes de Bernstein. La suite des points $\mathbf{P}_0, \dots, \mathbf{P}_n$ forme le *polygone de contrôle de Bézier*.

Chaque point de la courbe peut être vu alors comme un barycentre des $n + 1$ points de contrôle pondérés d'un poids égal au polynôme de Bernstein. Les principales propriétés des courbes de Bézier sont les suivantes :

- la courbe est à l'intérieur de l'enveloppe convexe des points de contrôle ;
- la courbe commence par le point \mathbf{P}_0 et se termine par le point \mathbf{P}_n , mais ne passe pas a priori par les autres points de contrôle ;
- $\mathbf{P}_0\mathbf{P}_1$ est le vecteur tangent à la courbe en \mathbf{P}_0 et $\mathbf{P}_{n-1}\mathbf{P}_n$ au point \mathbf{P}_n ;
- une courbe de Bézier est C^∞ ;
- la courbe de Bézier est un segment si et seulement si les points de contrôle sont alignés ;
- chaque restriction d'une courbe de Bézier est aussi une courbe de Bézier ;
- un arc de cercle (ni même aucun arc de courbe conique, en dehors du segment de droite) ne peut pas être décrit par une courbe de Bézier, quel que soit son degré ;

- le contrôle de la courbe est global : modifier un point de contrôle modifie toute la courbe, et non pas un voisinage du point de contrôle ;
- pour effectuer une transformation affine de la courbe, il suffit d'effectuer la transformation sur tous les points de contrôle.

A.3.2 B-Spline

Une *B-spline* est une combinaison linéaire de splines positives à support compact minimal. Les B-splines sont la généralisation des courbes de Bézier, elles peuvent être à leur tour généralisées par les NURBS. Étant donné $m + 1$ points t_i dans $[0, 1]$ tels que $0 \leq t_0 \leq t_1 \leq \dots \leq t_m \leq 1$, une courbe spline de degré n est une courbe paramétrique $\mathbf{S} : [0, 1] \rightarrow \mathbb{R}^d$, composée de fonctions *B-splines* de degré n :

$$\mathbf{S}(t) = \sum_{i=0}^{m-n-1} b_{i,n}(t) \cdot \mathbf{P}_i, \quad t \in [0, 1], \quad (\text{A.44})$$

où les P_i forment un polygone appelé *polygone de contrôle*. Le nombre de points composant ce polygone est égal à $m - n$. Les $m - n$ fonctions B-splines de degré n sont définies par récurrence sur le degré inférieur :

$$b_{j,0}(t) = \begin{cases} 1 & \text{si } t_j \leq t < t_{j+1} \\ 0 & \text{sinon} \end{cases} \quad (\text{A.45})$$

$$b_{j,n}(t) := \frac{t - t_j}{t_{j+n} - t_j} b_{j,n-1}(t) + \frac{t_{j+n+1} - t}{t_{j+n+1} - t_{j+1}} b_{j+1,n-1}(t). \quad (\text{A.46})$$

Quand les points sont équidistants, les B-splines sont dites uniformes. C'est le cas des courbes de Bézier qui sont des B-splines uniformes, dont les points t_i (pour i entre 0 et m) forment une suite arithmétique de 0 à 1 avec un pas constant $h = 1/m$, et où le degré n de la courbe de Bézier ne peut être supérieur à m).

Par extension, lorsque deux points successifs t_j et t_{j+1} sont confondus, on pose $0/0 = 0$: cela a pour effet de définir une discontinuité de la tangente, pour le point de la courbe paramétré par une valeur de t , donc d'y créer un sommet d'angle non plat. Toutefois il est souvent plus simple de définir ce B-spline étendu comme l'union de deux B-splines définis avec des points distincts, ces splines étant simplement joints par ce sommet commun, sans introduire de difficulté dans l'évaluation paramétrique ci-dessus des B-splines pour certaines valeurs du paramètre t . Mais cela permet de considérer alors tout polygone simple comme un B-spline étendu.

La forme des fonctions de base est déterminée par la position des points. La courbe est à l'intérieur de l'enveloppe convexe des points de contrôle. Une B-spline de degré n , $b_{i,n}(t)$ est non nulle dans l'intervalle $[t_i, t_{i+n+1}]$:

$$b_{i,n}(t) = \begin{cases} > 0 & \text{si } t_i \leq t < t_{i+n+1} \\ 0 & \text{sinon} \end{cases} \quad (\text{A.47})$$

En d'autres termes, déplacer un point de contrôle ne modifie que localement l'allure de la courbe. Par contre, les B-splines ne permettent pas de décrire un arc de courbe conique.

A.3.3 B-splines rationnelles non uniformes

Ces objets couramment nommés *NURBS*, pour Non-Uniform Rational Basis Splines, correspondent à une généralisation des B-splines car ces fonctions sont définies avec des points en coordonnées homogènes. Les coordonnées homogènes, introduites par Möbius, rendent les calculs possibles dans l'espace projectif comme les coordonnées cartésiennes le font dans l'espace euclidien. Ces

coordonnées homogènes sont largement utilisées en infographie ou en CAO car elles permettent la représentation de scènes en trois dimensions. Les NURBS parviennent à ajuster des courbes qui ne peuvent pas être représentées par des B-splines uniformes. Ils permettent même une représentation exacte de la totalité des arcs coniques ainsi que la totalité des courbes et surfaces polynomiales, avec uniquement des paramètres entiers ou rationnels si les NURBS passent par un nombre limité mais suffisant de points définis dans un maillage discret de l'espace.

Les fonctions NURBS de degré d sont définies par la formule doublement récursive de Cox-de Boor (formulation trouvée de manière indépendante par M.G. Cox en 1971 et C. de Boor en 1972) :

$$\begin{cases} b_{j,0}(t) = \begin{cases} 1 & \text{si } t_j \leq t < t_{j+1} \\ 0 & \text{sinon} \end{cases} \\ b_{j,d}(t) = \frac{t - t_j}{t_{j+d} - t_j} b_{j,d-1}(t) + \frac{t_{j+d+1} - t}{t_{j+d+1} - t_{j+1}} b_{j+1,d-1}(t) \end{cases} \quad (\text{A.48})$$

où les t_j sont des points. Lorsque plusieurs points t_j sont confondus, on peut encore poser $0/0 = 0$ comme pour les B-splines.

Annexe B

Intégration numérique

La méthode des éléments finis conduit à la discrétisation d'une formulation faible où la construction des matrices constitutives du système à résoudre nécessitent le calcul d'intégrales. Dans certains cas particuliers, ou en utilisant des codes de calcul formel, ces intégrations peuvent être réalisées de manière exacte. Cependant, dans la plupart des cas et dans la plupart des codes de calcul, ces intégrations sont calculées numériquement. On parle alors de méthodes d'intégration numérique et de formules de quadrature.

B.1 Méthodes de Newton-Cotes

Soit à calculer l'intégrale suivante :

$$I = \int_a^b f(x) dx \quad (\text{B.1})$$

L'idée consiste à construire un polynôme pour interpoler $f(x)$ et à intégrer ce polynôme. Plusieurs types de polynômes peuvent être utilisés pour cette interpolation. Les principales méthodes d'interpolations sont détaillées au chapitre A.

B.1.1 Méthode des rectangles

La *méthode des rectangles* consiste à interpoler $f(x)$ par un polynôme de degré 0, i.e. par la constante valant, selon les variantes de la méthode, soit $f(a)$, soit $f((a+b)/2)$. Comme cette approximation est très brutale, il est possible de subdiviser l'intervalle $[a; b]$ en plusieurs intervalles et d'appliquer la méthode sur chacun des intervalles, i.e. d'approcher f par une fonction en escalier. Si l'on subdivise l'intervalle $[a; b]$ en n intervalles égaux, il vient alors :

$$I \approx h \sum_{i=0}^{n-1} f(x_i) \quad (\text{B.2})$$

où $h = (b-a)/n$ est la longueur de chaque sous intervalle et $x_i = a + ih$ le point courant.

B.1.2 Méthode des trapèzes

La *méthode des trapèzes* consiste à interpoler $f(x)$ par un polynôme de degré 1, i.e. par la droite passant par les points $(a, f(a))$ et $(b, f(b))$. On obtient alors :

$$I \approx h \frac{f(a) + f(b)}{2} \quad (\text{B.3})$$

où $h = b - a$ est la longueur de l'intervalle, et l'erreur commise vaut $-\frac{h^3}{12} f''(w)$ pour un certain $w \in [a; b]$ (sous réserve que f soit 2 fois dérivable). L'erreur étant proportionnelle à f'' , la méthode

est dite d'ordre 2, ce qui signifie qu'elle est exacte (erreur nulle) pour tout polynôme de degré inférieur ou égale à 1.

Comme cette approximation peut sembler un peu brutale, il est possible de subdiviser l'intervalle $[a; b]$ en plusieurs intervalles et d'appliquer cette formule sur chacun des intervalles, i.e. d'approcher f par une fonction affine continue par morceaux. Si l'on subdivise l'intervalle $[a; b]$ en n intervalles égaux, il vient alors :

$$I \approx \frac{(b-a)}{n} \sum_{i=0}^n f(x_i) \quad (\text{B.4})$$

où $h = (b-a)/n$ est la longueur de chaque sous intervalle, $x_i = a + ih$ le point courant et l'erreur commise vaut $-\frac{h^3}{12n^2} f''(w)$ pour un certain $w \in [a; b]$.

Remarques.

- la méthode de Romberg permet d'accélérer la convergence de la méthode des trapèzes ;
- la méthode des trapèzes est une méthode de Newton-Cotes pour $n = 1$.

B.1.3 Méthode de Simpson

La méthode de Simpson consiste à interpoler $f(x)$ par un polynôme de degré 2, i.e. par la parabole passant par les points extrêmes $(a, f(a))$ et $(b, f(b))$ et le point milieu $(c, f(c))$ avec $c = (a+b)/2$. On obtient alors :

$$I \approx \frac{h}{6} (f(a) + f(b) + f(c)) \quad (\text{B.5})$$

où $h = b - a$ est la longueur de l'intervalle, et l'erreur commise vaut $-\frac{h^5}{25.90} f^{(4)}(w)$ pour un certain $w \in [a; b]$ (sous réserve que f soit 4 fois dérivable). L'erreur étant proportionnelle à $f^{(4)}$, la méthode est dite d'ordre 4, ce qui signifie qu'elle est exacte (erreur nulle) pour tout polynôme de degré inférieur ou égale à 3.

Comme dans le cas précédent, il est possible de subdiviser l'intervalle $[a; b]$ en plusieurs intervalles et d'appliquer cette formule sur chacun des intervalles. Si l'on subdivise l'intervalle $[a; b]$ en n intervalles égaux, avec n pair, il vient alors :

$$I \approx \frac{h}{3} \left(f(a) + f(b) + 2 \sum_{i=1}^{n/2-1} f(x_{2i}) + 4 \sum_{i=1}^{n/2} f(x_{2i-1}) \right) \quad (\text{B.6})$$

où $h = (b-a)/n$ est la longueur de chaque sous intervalle, $x_i = a + ih$ le point courant et l'erreur commise vaut $-\frac{nh^5}{180} f^{(4)}(w)$ pour un certain $w \in [a; b]$.

Remarques.

- la parabole interpolant f est trouvée en utilisant l'interpolation de Lagrange ;
- la méthode de Simpson est un cas particulier de celle de Newton-Cotes pour $n = 2$.

Méthode de Newton-Cotes

Les formules de Newton-Cotes se proposent également d'approximer l'intégrale I et découpant l'intervalle $[a; b]$ en n intervalles identiques. On posera donc encore une fois $h = (b-a)/n$ la longueur de chaque sous intervalle, et $x_i = a + ih$ le point courant. La formule est :

$$I \approx \sum_{i=0}^n \bar{\omega}_i f(x_i) \quad (\text{B.7})$$

où les $\bar{\omega}_i$ sont appelés *poids* ou *coefficients de la quadrature* et sont construits à partir des polynômes de Lagrange. La méthode de Newton-Cotes intègre exactement un polynôme de degré $n-1$ avec n points.

Remarque. Il est possible de construire une formule de Newton-Cotes de degré quelconque. Toutefois, une telle formule *n'est pas inconditionnellement stable*. C'est pourquoi, on se cantonnera aux plus bas degrés : $n = 0$ méthode du point médian (i.e. méthode des rectangle où la valeur est évaluée en milieu d'intervalle) ; $n = 1$ méthode des trapèzes ; $n = 2$ méthode de Simpson dite 1/3, i.e. celle présentée avant ; $n = 3$ méthode de Simpson 3/8 (il suffit de faire le calcul) ; $n = 4$ méthode de Boole. Lorsque le degré augmente, des instabilités apparaissent, dues au *phénomène de Runge*. En effet, avec certaines fonctions (même infiniment dérivables), l'augmentation du nombre n de points d'interpolation ne constitue pas nécessairement une bonne stratégie d'approximation. Carle Runge a montré qu'il existe des configurations où l'écart maximal entre la fonction et son interpolation augmente indéfiniment avec n . Pour remédier à cela on peut utiliser les abscisses de Tchebychev au lieu de points équirépartis pour interpoler, ou plus simplement utiliser des splines (i.e. des polynômes par morceaux), et donc augmenter le nombre de morceaux et non le degré des polynômes.

B.2 Méthodes de quadrature de Gauß

Le principe de la méthode reste le même que pour la méthode de Newton-Cotes, mais on va essayer d'améliorer un peu encore la qualité du résultat. Pour cela, on souhaite que :

$$I = \int_a^b \varpi(x) f(x) dx \approx \sum_{i=1}^n \varpi_i f(x_i) \quad (\text{B.8})$$

où $\varpi(x) : (a, b) \rightarrow \mathbb{R}$ est une *fonction de pondération*, qui peut assurer l'intégrabilité de f . Les ϖ_i sont appelés les *poids ou coefficients de quadrature (ou poids)*. Les x_i sont réels, distincts, uniques et sont les racines de polynômes orthogonaux (et non plus uniquement de Lagrange) pour le produit scalaire :

$$\langle f, g \rangle = \int_a^b f(x) g(x) \varpi(x) dx \quad (\text{B.9})$$

Ils sont appelés *points ou nœuds de Gauß*. Les poids et les nœuds sont choisis de façon à obtenir des degrés d'exactitude les plus grands possibles. Cette fois-ci (a, b) peut être *n'importe quel type d'intervalle (fermé, ouvert, fini ou non)*.

Intégration sur un intervalle type

Intervalle (a, b)	Fonction de pondération $\varpi(x)$	Famille de polynômes orthogonaux
$[-1; 1]$	1	Legendre
$] -1; 1[$	$(1-x)^\alpha (1+x)^\beta$, $\alpha, \beta > -1$	Jacobi
$] -1; 1[$	$\frac{1}{\sqrt{1-x^2}}$	Tchebychev (premier type)
$] -1; 1[$	$\sqrt{1-x^2}$	Tchebychev (second type)
\mathbb{R}^+	e^{-x}	Laguerre
\mathbb{R}	e^{-x^2}	Hermite

TABLE B.1: Polynômes et intégration

On rappelle que les nœuds sont déterminés comme les n racines du n ème polynôme orthogonal associé à la formule de quadrature. *Les méthodes de quadrature de Gauß intègrent exactement un polynôme de degré $2n - 1$ avec n points.*

Changement d'intervalle d'intégration

Si on intègre sur (a, b) au lieu de $(-1, 1)$, alors on fait un changement de variable. Finalement, on obtient l'approximation :

$$\frac{b-a}{2} \sum_{i=1}^n \varpi_i f\left(\frac{b-a}{2} x_i + \frac{a+b}{2}\right) \quad (\text{B.10})$$

Remarque.

- pour la méthode des éléments finis, l'intégration se déroule sur l'élément de référence, donc on n'a pas besoin de faire ce changement. Il est fait par la transformation affine entre l'élément considéré et l'élément de référence ;
- le nombre de points de Gauss ainsi que leurs positions sur l'élément sont donnés dans les documentations des logiciels.

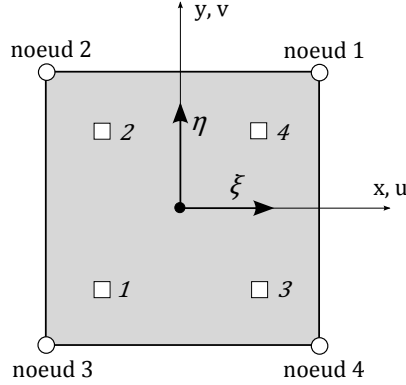


FIGURE B.1: Élément rectangulaire de référence Q_1 avec quatre points de Gauss

Intégration sur des carrés ou des cubes

Sur les carrés et les cubes, qui correspondent à ce qui nous intéresse en terme d'éléments de référence, on obtient les formules suivantes :

$$\int_{-1}^{-1} \int_{-1}^{-1} f(x, y) dx dy \approx \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \varpi_i \varpi_j f(x_i, x_j) \quad (\text{B.11})$$

$$\int_{-1}^{-1} \int_{-1}^{-1} \int_{-1}^{-1} f(x, y, z) dx dy dz \approx \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \sum_{k=1}^{n_z} \varpi_i \varpi_j \varpi_k f(x_i, x_j, x_k) \quad (\text{B.12})$$

où n_x , n_y et n_z sont les nombres de points de Gauss utilisés dans les directions x , y et z . Dans la pratique, on a souvent $n_x = n_y = n_z$.

Intégration sur un triangle ou un tétraèdre

Malheureusement, tous les éléments ne sont pas des segments, carrés ou cubes... on a également souvent à faire à des triangles et à des tétraèdres. Dans ce cas, on construit des formules spécifiques qui ne sont pas issues du cas monodimensionnel. L'élément triangulaire de référence est un triangle isocèle côtés égaux de longueur 1. L'angle droit est à l'origine du repère. La forme générale d'intégration est :

$$I = \int_{\hat{K}} f(x, y) dx dy \approx \sum_{i=1}^n \varpi_i f(x_i, y_i) \quad (\text{B.13})$$

Les positions et poids des n points de Gauss sont choisis afin d'intégrer exactement un polynôme de degré N . Le tout est listé dans le tableau B.2. Le même travail peut être fait sur un tétraèdre.

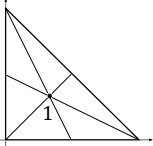
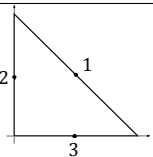
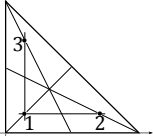
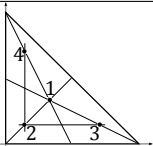
	n	x_i	y_i	$\bar{\omega}_i$	N
	1	1/3	1/3	1/2	1
	3	1/2	1/2	1/6	2
		0	1/2		
		1/2	0		
	3	1/6	1/6	1/6	2
		2/3	1/6		
		1/6	2/3		
	4	1/3	1/3	-27/96	4
		1/5	1/5	25/96	
		3/5	1/5	25/96	
		1/5	3/5	25/96	

TABLE B.2: Triangles et points de Gauß

Annexe C

Résolution des équations différentielles ordinaires

Résumé — La résolution exacte des équation différentielle fait partie des choses qui ont été demandées comme complément. Elles ne correspondent pas vraiment au but de ce document. Toutefois, le paragraphe sur la résolution numérique des équation différentielle nous a permis d'introduire des méthodes qui sont employées également dans la méthode des éléments finis (notamment la méthode de Newmark).

C.1 Résolution exacte des équations différentielles linéaires

Une *Équation différentielle linéaire* est de la forme suivante :

$$a_0(x)y + a_1(x)y' + a_2(x)y'' + \dots + a_n(x)y^{(n)} = f(x) \quad (\text{C.1})$$

où les coefficients $a_i(x)$ sont des fonctions numériques continues. Si les fonctions y dépendent d'une seule variable x , alors ces équation différentielle sont dites équation différentielle linéaires scalaires, si c'est un vecteur elles sont dites équation différentielle linéaires vectorielles ou système différentiel linéaire. Dans ce dernier cas, les a_i sont des applications linéaires. L'*ordre* d'une équation différentielle est le degré non nul le plus des a_i , ici n . La méthode générale consiste à résoudre d'abord l'*équation homogène*, i.e. sans second membre, i.e. $f(x) = 0$, qui donne une solution contenant une ou des « constantes d'intégration » que l'on identifie ensuite en appliquant la forme générale trouvée à l'équation avec second membre.

C.1.1 Équation différentielle linéaire scalaire d'ordre 1

Il s'agit du cas particulier :

$$a(x)y' + b(x)y = c(x) \quad (\text{C.2})$$

où a , b et c sont des fonctions (connues).

Coefficients constants

Si a et b sont des constantes, alors l'équation homogène précédente s'écrit sous la forme :

$$y' = ky \quad (\text{C.3})$$

avec $k \in \mathbb{R}$, et la solution est :

$$f(x) = Ce^{kx} \quad (\text{C.4})$$

où la constante $C \in \mathbb{R}$ est déterminée à l'aide des conditions initiales :

- si pour x_0 on a $f(x_0) = y_0$ alors $C = y_0 e^{-kx_0}$.
- si $c = 0$, alors la solution du problème est celle de l'équation homogène et on a fini le travail.

Ce type d'équation se retrouve :

- $k < 0$: modélisation de la décroissance radioactive dans un milieu homogène et fermé ;
- $k > 0$: modélisation de la croissance d'une population (modèle simplifié car en milieu fermé, cette croissance ne peut durer indéfiniment).

Si $c \neq 0$, il faut déterminer une solution particulière de l'équation avec second membre. On appliquera les mêmes techniques que dans le cas des coefficients non constants (voir ci-après).

Coefficients non constants

On réécrit l'équation homogène :

$$y' + \frac{b(x)}{a(x)}y = 0 \quad (\text{C.5})$$

sur un intervalle I où $a(x)$ ne s'annule pas. Ensuite, en notant $A(x)$, une primitive de la fonction $b(x)/a(x)$, il vient :

$$y' + A'(x)y = 0 \quad (\text{C.6})$$

puis :

$$y' e^{A(x)} + y \frac{b(x)}{a(x)} e^{A(x)} = 0 \quad (\text{C.7})$$

et :

$$y' e^{A(x)} + y A'(x) e^{A(x)} = 0 \quad (\text{C.8})$$

qui est de la forme $u'v + uv'$ et vaut $(uv)'$, d'où :

$$\frac{d}{dx}(y e^{A(x)}) = 0 \quad (\text{C.9})$$

soit :

$$y e^{A(x)} = C \quad (\text{C.10})$$

Les solutions sont alors les fonctions, définies sur I , de la forme :

$$y(x) = C e^{-A(x)} \quad (\text{C.11})$$

où encore une fois la constante $C \in \mathbb{R}$ est déterminée par les conditions initiales.

Solution particulière

Plusieurs cas se présentent :

- $c(x) = 0$: la solution du problème est celle de l'équation homogène et on a fini ;
- $c(x) \neq 0$: il faut déterminer une solution particulière de l'équation avec second membre. Cela n'est pas toujours simple car la forme de cette solution particulière varie en fonction de $c(x)$;
- $c(x) = C^{\text{te}}$: la quantité $f_0 = c/b$ est aussi une constante et l'ensemble des solutions est donc $f = y(x) + f_0$;
- $c(x)$ est une somme de fonctions : f_0 est alors la somme des solutions particulières obtenues pour chacun des termes de la somme constituant $c(x)$;

- $c(x)$ est un polynôme de degré n : f_0 est également un polynôme de degré n ;
- $c(x) = A \cos(\omega x + \varphi) + B \sin(\omega x + \varphi)$: on cherche f_0 comme combinaison linéaire de $\cos(\omega x + \varphi)$ et $\sin(\omega x + \varphi)$, i.e. sous la même forme que $c(x)$.

Dans le cas général, on utilise la méthode dite de la *variation des constantes* qui consiste à se ramener, par un changement de fonction variable, à un problème de calcul de primitive. On suppose que, sur l'intervalle d'étude, la fonction $a(x)$ ne s'annule pas¹. On connaît déjà la solution générale (C.11) de l'équation homogène. On décide d'étudier le cas de la fonction $z(x)$ définie par :

$$z(x) = k(x)e^{-A(x)} \quad (\text{C.12})$$

en substituant dans $y(x)$ la fonction $k(x)$ à la constante C , d'où le nom de la méthode. En reportant dans l'équation différentielle initiale, il vient :

$$a(x)k'(x) = c(x)e^{A(x)} \quad (\text{C.13})$$

qui est une équation différentielle dépendant cette fois de la fonction $k(x)$. Si on note $B(x)$ une primitive de la fonction $c(x)e^{A(x)}/a(x)$, l'ensemble des solutions est alors :

$$k(x) = B(x) + C \quad (\text{C.14})$$

et la solution générale s'écrit alors sous la forme :

$$f(x) = (B(x) + C)e^{-A(x)} \quad (\text{C.15})$$

soit, finalement :

$$f = \exp\left(-\int \frac{b(x)}{a(x)} dx\right) \left\{ C + \int \frac{c(x)}{a(x)} \exp\left(\int \frac{b(x)}{a(x)} dx\right) dx \right\} \quad (\text{C.16})$$

On peut évidemment être confronté à un calcul d'intégral qui n'est pas simple (ou même pas possible à l'aide des fonctions usuelles).

C.1.2 Équation différentielle du premier ordre à variables séparées

Une *équation différentielle d'ordre un à variables séparées* est une équation différentielle qui peut se mettre sous la forme :

$$y' = f(x)g(y) \quad (\text{C.17})$$

Dans un tel problème, on commence par chercher les *solutions régulières* qui sont les solutions telles que $g(y)$ n'est jamais nul. Comme $g(y) \neq 0$, on peut écrire l'équation sous la forme :

$$\frac{1}{g(y(x))} y'(x) = f(x) \quad (\text{C.18})$$

par rapport à la variable x , ce qui conduit à :

$$\int_{x_0}^x \frac{1}{g(y(u))} y'(u) du = \int_{x_0}^x f(u) du \quad (\text{C.19})$$

et qui après changement de variable, est de la forme :

$$\int_{y_0}^y \frac{1}{g(v)} dv = \int_{x_0}^x f(u) du \quad (\text{C.20})$$

1. Il est possible de résoudre sur plusieurs intervalles de type I et essayer de « recoller » les solutions.

L'hypothèse $g(y) \neq 0$ écarte certaines solutions particulières. Par exemple, si y_0 est un point d'annulation de g , alors la fonction constante égale à y_0 est une solution maximale de l'équation. Une telle solution, dite *solution singulière*, est donc telle que $g(y)$ est toujours nul. Si la seule hypothèse faite sur g est la continuité, il peut exister des solutions « hybrides » constituées du raccordement de solutions régulières et singulières. D'une manière générale, pour une solution donnée, la quantité $g(y)$ sera soit toujours nulle, soit jamais nulle.

Il existe un cas particulier, qui est celui de l'équation différentielle d'ordre un à variables séparées autonome qui s'écrit :

$$y' = g(y) \quad (\text{C.21})$$

c'est-à-dire que la relation formelle ne dépend pas de x . Dans ce cas, si $x \mapsto y_0(x)$ est une solution, les fonctions obtenues par translation de la variable, de la forme $x \mapsto y_0(x + A)$, sont également solutions. Il y a en outre une propriété de monotonie, au moins pour les solutions régulières : puisque g ne s'annule pas, il garde alors un signe constant.

C.1.3 Équation différentielle linéaire d'ordre deux

Les équations différentielles linéaires d'ordre deux sont de la forme :

$$a(x)y'' + b(x)y' + c(x)y = d(x) \quad (\text{C.22})$$

où $a(x)$, $b(x)$, $c(x)$ et $d(x)$ sont des fonctions. Si elles ne peuvent pas toutes être résolues explicitement, beaucoup de méthodes existent.

Équation différentielle homogène

Pour l'équation différentielle homogène ($d(x) = 0$), une somme de deux solutions est encore solution, ainsi que le produit d'une solution par une constante. L'ensemble des solutions est donc un espace vectoriel et contient notamment une solution évidente, la fonction nulle.

Équation différentielle homogène à coefficients constants

On cherche des solutions sous forme exponentielle $f(x) = e^{\lambda x}$. Une telle fonction sera solution de l'équation différentielle si et seulement si λ est solution de l'équation caractéristique de l'ED :

$$a\lambda^2 + b\lambda + c = 0 \quad (\text{C.23})$$

Comme pour toute équation du second degré, il y a trois cas correspondant au signe du discriminant Δ :

1. $\Delta > 0$ — L'équation caractéristique possède deux solutions λ_1 et λ_2 , et les solutions de l'équation différentielle sont engendrées par $f_1(x) = e^{\lambda_1 x}$ et $f_2(x) = e^{\lambda_2 x}$, i.e. de la forme :

$$f(x) = C_1 f_1(x) + C_2 f_2(x) \quad (\text{C.24})$$

les constantes réelles C_1 et C_2 étant définies par :

- les conditions initiales : en un point (instant) donné x_0 , on spécifie les valeurs de $y_0 = y(x_0)$ et $y'_0 = y'(x_0)$. Dans ce cas l'existence et l'unicité de la solution vérifiant ces conditions initiales sont garanties.
- les conditions aux limites : pour de nombreux problèmes physiques, il est fréquent de donner des conditions aux limites en précisant les valeurs y_1 et y_2 aux instants x_1 et x_2 . Il y a alors fréquemment existence et unicité des solutions, mais ce n'est pas toujours vrai.

2. $\Delta = 0$ — L'équation caractéristique possède une solution double λ , et les solutions de l'équation différentielle sont de la forme :

$$f(x) = (C_1x + C_2)e^{\lambda x} \quad (\text{C.25})$$

les constantes réelles C_1 et C_2 étant définies comme précédemment.

3. $\Delta < 0$ — L'équation caractéristique possède deux solutions λ_1 et λ_2 complexes conjuguées, et les solutions de \mathbb{R} dans \mathbb{C} de l'équation différentielle sont *engendrées* par $f_1(x) = e^{\lambda_1 x}$ et $f_2(x) = e^{\lambda_2 x}$, i.e. de la forme :

$$f(x) = C_1 f_1(x) + C_2 f_2(x) \quad (\text{C.26})$$

où cette fois C_1 et C_2 sont des complexes. Comme on cherche des solutions de \mathbb{R} dans \mathbb{R} , on note $\lambda_1 = u + iv$ (et donc $\lambda_2 = u - iv$), on exprime f_1 et f_2 , et on déduit que les fonctions g_1 et g_2 , à valeurs dans \mathbb{R} cette fois, sont encore solutions :

$$g_1(x) = \frac{1}{2}(f_1(x) + f_2(x)) = e^{ux} \cos(vx) \quad (\text{C.27})$$

$$g_2(x) = \frac{1}{2i}(f_1(x) - f_2(x)) = e^{ux} \sin(vx) \quad (\text{C.28})$$

et engendrent encore l'ensemble des solutions. On a donc les solutions sous la forme :

$$f(x) = e^{ux}(C_1 \cos(vx) + C_2 \sin(vx)) \quad (\text{C.29})$$

les constantes réelles C_1 et C_2 étant définies comme précédemment. Notons que $f(x)$ s'écrit également :

$$f(x) = qe^{ux} \cos(vx + r) \quad (\text{C.30})$$

avec q et r deux réels à déterminer comme précédemment. Cette forme est parfois plus pratique selon les problèmes.

Équation différentielle homogène à coefficients non constants

Si les fonctions $a(x)$, $b(x)$ et $c(x)$ ne sont pas constantes, alors il n'existe pas d'expression générale des solutions. C'est pour cette raison qu'au XIX^e siècle furent introduites de nombreuses fonctions spéciales, comme les fonctions de Bessel ou la fonction d'Airy, définies comme solutions d'équations qu'il est impossible de résoudre explicitement.

Toutefois, dès lors qu'une solution particulière non nulle de l'équation est connue, il est possible de la résoudre complètement. En effet, le théorème de Cauchy-Lipschitz affirme que l'ensemble des solutions de l'équation constitue un espace vectoriel de dimension deux. Résoudre l'équation différentielle revient donc à exhiber deux fonctions solutions non proportionnelles : elles formeront une base de l'espace des solutions. Une telle base est appelée *système fondamental de solutions*.

Solution particulière et traitement du second membre

On peut agir de la même manière que pour les équation différentielle d'ordre un, et les mêmes remarques s'appliquent. On résout l'équation homogène puis on cherche une solution de l'équation avec second membre pour les connaître toutes.

Les équation différentielle d'ordre deux correspondent typiquement en physique aux problèmes dynamiques. Même si dans le cas réel on n'a rarement des phénomènes linéaires, des hypothèses de petits mouvements permettent de s'y ramener. Si cette hypothèse de petits déplacements ne peut être vérifiée, on aura alors recourt à des techniques dites de linéarisation, ou à des méthodes numériques comme la méthode de Newmark (qui sera présentée un petit peu plus loin).

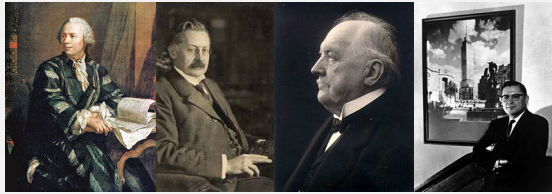
C.2 Résolution numérique

Dans le cas d'équation différentielle non linéaires, on passera forcément à une résolution numérique. Mais les méthodes numériques permettent évidemment aussi de résoudre numériquement les équations différentielles et les équations aux dérivées partielles.

Histoire

La première méthode numérique fut introduite en 1768 par Leonhard Euler. Depuis, un grand nombre de techniques ont été développées : elles se basent sur la discrétisation de l'intervalle d'étude en un certain nombre de pas. Suivant le type de formule utilisé pour approcher les solutions, on distingue les méthodes numériques à un pas ou à pas multiples, explicites ou implicites.

Il existe plusieurs critères pour mesurer la performance des méthodes numériques : la consistance d'une méthode indique que l'erreur théorique effectuée en approchant la solution tend vers zéro avec les pas. La stabilité indique la capacité à contrôler l'accumulation des erreurs d'arrondi. Ensemble elles assurent la convergence, i.e. la possibilité de faire tendre l'erreur globale vers zéro avec le pas.



Euler

Runge

Kutta

Newmark

L'idée générale est toujours la même : on approche la dérivée d'une fonction en un point par sa tangente (ce qui revient finalement à la définition de la dérivée). Pour une fonction $f(x)$, on écrit donc au point $x = a$ une relation de la forme :

$$f'(a) \approx \frac{f(b) - f(c)}{b - c} \quad (\text{C.31})$$

où b et c sont d'autres points. Par exemple, pour $c = a$ et $b = a + \varepsilon$ on obtient un schéma décentré à droite ; pour $c = a - \varepsilon$ et $b = a + \varepsilon$, on obtient un schéma centré.

C.2.1 Méthode d'Euler, Runge-Kutta ordre 1

Soit à résoudre l'équation différentielle suivante :

$$y' = f(t, y), \quad y(t_0) = y_0 \quad (\text{C.32})$$

D'après ce qui précède, on utilise une discrétisation de pas h , ce qui donne comme point courant $y_i = y_0 + ih$, et on fait l'approximation :

$$y' = \frac{y_{i+1} - y_i}{h} \quad (\text{C.33})$$

On obtient alors le schéma numérique :

$$y_{i+1} = y_i + hf(t, y_i) \quad (\text{C.34})$$

qui permet d'obtenir y_{i+1} uniquement en fonction de données au pas i . Cette méthode due à Euler, correspond également à la méthode de Runge-Kutta à l'ordre 1.

C.2.2 Méthode de Runge-Kutta d'ordre 2

La méthode de Runge-Kutta à l'ordre 2 est obtenue par amélioration de la méthode d'Euler en considérant le point milieu du pas h . Ainsi, on écrit cette fois :

$$y_{i+1} = y_i + h \cdot f\left(t + \frac{h}{2}, y_i + \frac{h}{2} f(t, y_i)\right) \quad (\text{C.35})$$

Mésalor me direz-vous, il manque des bouts... Les dérivées au milieu du pas d'intégration sont obtenues par :

$$y_{i+\frac{1}{2}} = y_i + \frac{h}{2} f(t, y_i) \quad \text{et} \quad y'_{i+\frac{1}{2}} = f\left(t + \frac{h}{2}, y_{i+\frac{1}{2}}\right) \quad (\text{C.36})$$

En réinjectant cela, on obtient sur le pas h complet :

$$y_{i+1} = y_i + h y'_{i+\frac{1}{2}} \quad (\text{C.37})$$

Notons qu'il s'agit du cas centré ($\alpha = 1/2$) de la formule plus générale :

$$y_{i+1} = y_i + h \left[\left(1 - \frac{1}{2\alpha}\right) f(t, y_i) + \frac{1}{2\alpha} f\left(t + \alpha h, y_i + \alpha h f(t, y_i)\right) \right] \quad (\text{C.38})$$

C'est une méthode d'ordre 2 car l'erreur est de l'ordre de h^3 .

C.2.3 Méthode de Runge-Kutta d'ordre 4

Aujourd'hui, le cas le plus fréquent est celui de l'ordre 4. L'idée est toujours d'estimer la pente de y , mais de façon plus précise. Pour cela, on ne prend plus la pente en un point (début ou milieu), mais on utilise la moyenne pondérée des pentes obtenues en 4 points du pas.

- $k_1 = f(t_i, y_i)$ est la pente au début de l'intervalle ;
- $k_2 = f\left(t_i + \frac{h}{2}, y_i + \frac{h}{2} k_1\right)$ est la pente au milieu de l'intervalle, en utilisant la pente k_1 pour calculer la valeur de y au point $t_i + h/2$ par la méthode d'Euler ;
- $k_3 = f\left(t_i + \frac{h}{2}, y_i + \frac{h}{2} k_2\right)$ est de nouveau la pente au milieu de l'intervalle, mais obtenue en utilisant la pente k_2 pour calculer y ;
- $k_4 = f(t_i + h, y_i + h k_3)$ est la pente en fin d'intervalle, avec la valeur de y calculée en utilisant k_3 .

On obtient finalement la discrétisation de Runge-Kutta à l'ordre 4 :

$$y_{i+1} = y_i + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4) \quad (\text{C.39})$$

La méthode est d'ordre 4, ce qui signifie que l'erreur commise à chaque étape est de l'ordre de h^5 , alors que l'erreur totale accumulée est de l'ordre de h^4 . Notons enfin que toutes ces formulations sont encore valables pour des fonctions à valeurs vectorielles.

C.2.4 Méthode de Newmark

La *méthode de Newmark (1959)* permet de résoudre numériquement des équations différentielles du second ordre. Elle convient, non seulement pour des systèmes différentiels linéaires, mais aussi pour des systèmes fortement non-linéaires avec une matrice de masse et une force appliquée qui peuvent dépendre à la fois de la position et du temps. Dans ce second cas, le calcul nécessite à chaque pas une boucle d'itération.

L'idée générale reste la même : on cherche à estimer les valeurs des dérivées (premières, secondes...) à l'instant t à partir des informations disponibles à l'instant précédent (au pas de temps précédent). Pour cela, on va recourir à un développement limité. Considérons l'équation de la dynamique :

$$M \ddot{x}(t) + C \dot{x}(t) + K x(t) = f(t) \quad (\text{C.40})$$

On fait un développement en série de Taylor :

$$u_{t+\Delta t} = u_t + \Delta t \dot{u}_t + \frac{\Delta t^2}{2} \ddot{u}_t + \beta \Delta t^3 \dddot{u}_t \quad \text{et} \quad \dot{u}_{t+\Delta t} = \dot{u}_t + \Delta t \ddot{u}_t + \gamma \Delta t^2 \dddot{u}_t \quad (\text{C.41})$$

et on fait l'hypothèse de linéarité de l'accélération à l'intérieur d'un pas de temps Δ_t :

$$\ddot{u} = \frac{\ddot{u}_{t+\Delta_t} - \ddot{u}_t}{\Delta_t} \quad (\text{C.42})$$

Les différents *schémas de Newmark* correspondent à des valeurs particulières de β et γ . Dans le cas $\beta = 0$ et $\gamma = 1/2$, on retombe sur le schéma des *différences finies centrées*.

La méthode de Newmark fonctionne également pour les problèmes non linéaires, mais dans ce cas la matrice de rigidité devra être réévaluée à chaque pas de temps (ainsi que celle d'amortissement dans les cas les plus tordus).

Annexe D

Méthode de Newton Raphson

Résumé — La méthode de Newton-Raphson est, dans son application la plus simple, un algorithme efficace pour trouver numériquement une approximation précise d'un zéro (ou racine) d'une fonction réelle d'une variable réelle.

D.1 Présentation

Histoire

La méthode de Newton fut décrite par Newton dans *De analysi per aequationes numero terminorum infinitas*, écrit en 1669 et publié en 1711 par William Jones. Elle fut à nouveau décrite dans *De metodis fluxionum et serierum infinitarum* (De la méthode des fluxions et des suites infinies), écrit en 1671, traduit et publié sous le titre *Methods of Fluxions* en 1736 par John Colson. Toutefois, Newton n'appliqua la méthode qu'aux seuls polynômes. Comme la notion de dérivée et donc de linéarisation n'était pas définie à cette époque, son approche diffère de l'actuelle méthode : Newton cherchait à affiner une approximation grossière d'un zéro d'un polynôme par un calcul polynomial.



Newton

Jones

Colson



Wallis



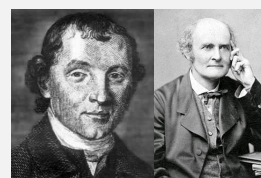
Raphson

Cette méthode fut l'objet de publications antérieures. En 1685, John Wallis en publia une première description dans *A Treatise of Algebra both Historical and Practical*. En 1690, Joseph Raphson en publia une description simplifiée dans *Analysis aequationum universalis*. Raphson considérait la méthode de Newton toujours comme une méthode purement algébrique et restreignait aussi son usage aux seuls polynômes. Toutefois, il mit en évidence le calcul récursif des approximations successives d'un zéro d'un polynôme au lieu de considérer

comme Newton une suite de polynômes.

C'est Thomas Simpson qui généralisa cette méthode au calcul itératif des solutions d'une équation non linéaire, en utilisant les dérivées (qu'il appelait fluxions, comme Newton). Simpson appliqua la méthode de Newton à des systèmes de deux équations non linéaires à deux inconnues, en suivant l'approche utilisée aujourd'hui pour des systèmes ayant plus de 2 équations, et à des problèmes d'optimisation sans contrainte en cherchant un zéro du gradient.

Arthur Cayley fut le premier à noter la difficulté de généraliser la méthode de Newton aux variables complexes en 1879, par exemple aux polynômes de degré supérieur à 3.



Simpson

Cayley

Sous sa forme moderne, l'algorithme se déroule comme suit : à chaque itération, la fonction dont on cherche un zéro est linéarisée en l'itéré (ou point) courant ; et l'itéré suivant est pris égal au zéro de la fonction linéarisée.

Cette description sommaire indique qu'au moins deux conditions sont requises pour la bonne marche de l'algorithme : la fonction doit être différentiable aux points visités (pour pouvoir y linéariser la fonction) et les dérivées ne doivent pas s'y annuler (pour que la fonction linéarisée ait un zéro) ; s'ajoute à ces conditions la contrainte forte de devoir prendre le premier itéré assez proche d'un zéro régulier de la fonction (i.e. en lequel la dérivée de la fonction ne s'annule pas), pour que la convergence du processus soit assurée.

L'intérêt principal de l'algorithme de Newton-Raphson est sa convergence quadratique locale. En termes imagés mais peu précis, cela signifie que le nombre de chiffres significatifs corrects des itérés double à chaque itération, asymptotiquement. Comme le nombre de chiffres significatifs représentables par un ordinateur est limité (environ 15 chiffres décimaux sur un ordinateur avec un processeur 32-bits), on peut simplifier grossièrement en disant que, soit il converge en moins de 10 itérations, soit il diverge. En effet, si l'itéré initial n'est pas pris suffisamment proche d'un zéro, la suite des itérés générée par l'algorithme a un comportement erratique, dont la convergence éventuelle ne peut être que le fruit du hasard (i.e. si l'un des itérés est par chance proche d'un zéro).

L'importance de l'algorithme de Newton-Raphson a incité les numériciens à étendre son application et à proposer des remèdes à ses défauts.

Par exemple, l'algorithme permet également de trouver un zéro d'une fonction de plusieurs variables à valeurs vectorielles, voire définie entre espaces vectoriels de dimension infinie ; la méthode conduit d'ailleurs à des résultats d'existence de zéro.

On peut aussi l'utiliser lorsque la fonction est différentiable dans un sens plus faible, ainsi que pour résoudre des systèmes d'inégalités non linéaires, des problèmes d'inclusion, d'ED ou EDP, d'inéquations variationnelles...

On a également mis au point des techniques de globalisation de l'algorithme, lesquelles ont pour but de forcer la convergence des suites générées à partir d'un itéré initial arbitraire (non nécessairement proche d'un zéro)

Dans les versions dites inexactes ou tronquées, on ne résout le système linéaire à chaque itération que de manière approchée.

Enfin, la famille des algorithmes de quasi-Newton (par exemple si l'on ne connaît pas l'expression analytique de la fonction dont on cherche une racine) propose des techniques permettant de se passer du calcul de la dérivée de la fonction.

Toutes ces améliorations ne permettent toutefois pas d'assurer que l'algorithme trouvera un zéro existant, quel que soit l'itéré initial.

Appliqué à la dérivée d'une fonction réelle, cet algorithme permet d'obtenir des points critiques (i.e. des zéros de la dérivée). Cette observation est à l'origine de son utilisation en optimisation sans ou avec contraintes.

D.2 Algorithme

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ la fonction dont on cherche à construire une bonne approximation d'un zéro. pour cela, on se base sur son développement de Taylor au premier ordre.

Partant d'un point x_0 que l'on choisit de préférence proche du zéro à trouver (en faisant des estimations grossières par exemple), on approche la fonction au premier ordre, autrement dit, on la considère à peu près égale à sa tangente en ce point :

$$f(x) \simeq f(x_0) + f'(x_0)(x - x_0) \quad (\text{D.1})$$

Pour trouver un zéro de cette fonction d'approximation, il suffit de calculer l'intersection de la droite tangente avec l'axe des abscisses, i.e. de résoudre :

$$0 = f(x_0) + f'(x_0)(x - x_0) \quad (\text{D.2})$$

On obtient alors un point x_1 qui en général a de bonnes chances d'être plus proche du vrai zéro de f que le point x_0 précédent. Par cette opération, on peut donc espérer améliorer l'approximation par itérations successives.

Cette méthode requiert que la fonction possède une tangente en chacun des points de la suite que l'on construit par itération. Cela est évidemment vrai si f est dérivable.

Formellement, on construit donc la suite :

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (\text{D.3})$$

à partir d'un point x_0 .

Bien que la méthode soit très efficace, certains aspects pratiques doivent être pris en compte :

- Avant tout, la méthode de Newton-Raphson nécessite que la dérivée soit effectivement calculée. Dans les cas où la dérivée est seulement estimée en prenant la pente entre deux points de la fonction, la méthode prend le nom de méthode de la sécante, moins efficace.
- Par ailleurs, si la valeur de départ est trop éloignée du vrai zéro, la méthode de Newton-Raphson peut entrer en boucle infinie sans produire d'approximation améliorée. À cause de cela, toute mise en œuvre de la méthode de Newton-Raphson doit inclure un code de contrôle du nombre d'itérations.