

INTRODUCTION À L'ANALYSE NUMÉRIQUE

Résumé du cours

Jean-Antoine Désidéri, INRIA

Version compilée le 7 septembre 2009

Table des matières

1 Compléments d'Algèbre Linéaire	7
1.1 Disques de Gershgorin	7
1.2 Théorème de Bendixon	8
2 Approximation Polynomiale	11
2.1 Formule(s) de Taylor	11
2.2 Interpolation de Lagrange	12
2.3 Forme de Newton	12
2.4 Polynômes orthogonaux	13
2.4.1 Polynômes de Legendre	14
2.4.2 Polynômes de Tchebychev	15
2.5 Approximation par Moindres Carrés	15
2.6 Meilleure Approximation	16
2.7 Oscillations de Gibbs	17
3 Différentiation et Intégration Numériques	19
3.1 Introduction	19
3.2 Différentiation Numérique	20
3.2.1 Dérivée première	20
3.2.2 Dérivées d'ordre supérieur :	20
3.3 Intégration Numérique	21
3.3.1 Règles d'intégration élémentaires	21
3.3.2 Règles d'intégration composées	22
3.3.3 Règles d'intégration de Gauss	23
4 Equations Différentielles	27
4.1 Rappels	27
4.1.1 Equation différentielle d'ordre n	27
4.1.2 Forme canonique	27
4.1.3 Equation différentielle linéaire "homogène"	28
4.1.4 Equation différentielle linéaire homogène à coefficients constants	28
4.1.5 Equation différentielle linéaire non homogène.	29
4.1.6 Analogie équations différentielles/équations aux différences	30
4.2 Intégration numérique	31
4.2.1 Intégration numérique par développement de Taylor	32
4.2.2 Méthodes de Runge-Kutta	35
4.2.3 Méthodes Multipas	37
4.2.4 Méthodes prédicteur-correcteur	39
4.2.5 Notion de stabilité	40
4.2.6 Un exemple d'analyse de méthode	44

Nota Bene

Les chapitres 1, 2 et 3 qui suivent ne constituent pas un cours mais rassemblent les principaux résultats (généralement sans démonstration, ni illustration) d'un premier cours sur l'approximation numérique des fonctions, de leurs dérivées et de leurs intégrales, et préparent au chapitre 4 portant sur la résolution numérique des équations différentielles.

Notations d'ensembles :

\mathbb{N}	les entiers naturels
\mathbb{R}	les réels
\mathbb{C}	les complexes
$\mathcal{C}^n(]a, b[)$	les fonctions de $]a, b[$ dans \mathbb{R} n fois continûment dérivables

Chapitre 1

Compléments d'Algèbre Linéaire

1.1 Disques de Gershgorin

Définition 1.1 “Disques de Gershgorin”. Etant donné une matrice carrée A dont les éléments $\{a_{jk}\}$ ($j = 1, 2, \dots, n$; $k = 1, 2, \dots, n$) sont des complexes, on associe à chaque ligne j le disque fermé D_j dont le centre est l'élément diagonal a_{jj} et le rayon la somme des modules des éléments extra-diagonaux $R_j = \sum_{k=1; k \neq j}^n |a_{jk}|$.

Théorème 1.1 “Théorème de Gershgorin”. Avec les notations précédentes :
Toute valeur propre λ de la matrice A appartient à l'un au moins des disques de Gershgorin, i.e. :

$$\forall \lambda \in Sp(A), \exists j \text{ tel que } \lambda \in D_j.$$

De manière équivalente, ce théorème précise que le spectre de A , noté $Sp(A)$, est globalement inclus dans la réunion des disques de Gershgorin, ce qui permet d'exclure la région du plan complexe qui est extérieure à cette réunion :

$$Sp(A) \subset \bigcup_{j=1}^n D_j.$$

Exemple 1.1 La matrice réelle

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 1 & 5 & 3 \\ 2 & 4 & 7 \end{pmatrix}$$

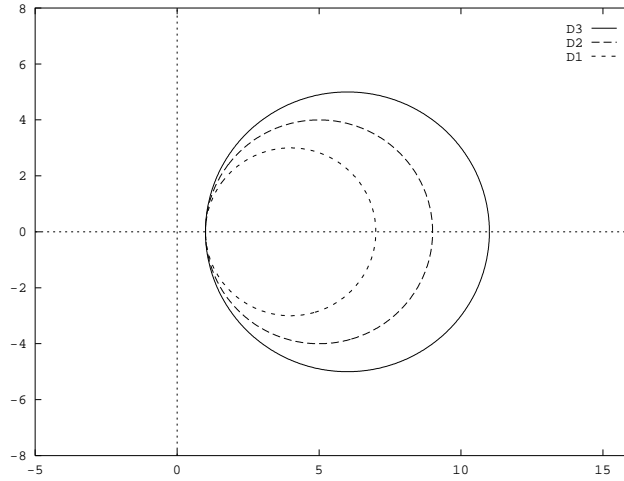
est à “diagonale dominante” stricte car chaque élément diagonal dépasse strictement la somme (des valeurs absolues) des éléments extra-diagonaux de la même ligne. En construisant les disques de Gershgorin,

D_1 : disque de centre 4 de rayon $2+1=3$

D_2 : disque de centre 5 de rayon $1+3=4$

D_3 : disque de centre 7 de rayon $2+4=6$

on constate ici que $D_1 \subset D_2 \subset D_3$. Par conséquent, $Sp(A) \subset D_3$ et en particulier $\lambda = 0$ n'est pas une valeur propre de A . La matrice A est donc inversible.



1.2 Théorème de Bendixon

Définition 1.2 “Matrice Adjointe”. Etant donné une matrice carrée A dont les éléments sont complexes, la matrice adjointe de A est notée A^* et s’obtient par transposition et conjugaison de la matrice A :

$$A^* = \bar{A}^T$$

(En d’autres termes, si on pose $A = \{a_{jk}\}$ et $A^* = \{a^*_{jk}\}$ ($j = 1, 2, \dots, N$, $k = 1, 2, \dots, N$), alors $\forall j, \forall k, a^*_{jk} = \bar{a}_{kj}$.)

En particulier, si les éléments de A sont réels, $A^* = A^T$ (matrice transposée).

Définition 1.3 “Matrice Hermitienne”. Une matrice carrée H dont les éléments sont complexes est dite “hermitienne” si elle est égale à sa matrice adjointe :

$$H = H^*$$

En particulier, une matrice réelle hermitienne S est une matrice réelle-symétrique ($S^T = S$).

Définition 1.4 “Matrice Unitaire”. Une matrice U est dite unitaire si elle est inversible et si son inverse est égal à son adjoint, ou de manière équivalente ssi :

$$U^* U = U U^* = I.$$

Ceci signifie que les vecteurs colonnes de la matrice U sont orthonormés (orthogonaux et normalisés) ; ils forment donc une base orthonormale. En particulier, une matrice réelle unitaire O est une matrice orthogonale ($O^T O = O O^T = I$).

Théorème 1.2 Toute matrice hermitienne H est diagonalisable par une transformation unitaire et ses valeurs propres sont réelles :

$$H = U \Lambda U^* \quad (U^* U = U U^* = I, \Lambda \text{ diagonale et réelle})$$

De plus, si on définit pour tout vecteur non-nul $z \in \mathbb{C}^N$, le “quotient de Rayleigh” suivant

$$R(z) = \frac{z^* H z}{z^* z}$$

(où z^* est le vecteur ligne \bar{z}^T), et si on note λ_{min} et λ_{max} la plus petite et la plus grande valeurs propres de H , on a :

$$\lambda_{min} = \min_{z \in \mathbb{C}^N, z \neq 0} R(z), \quad \lambda_{max} = \max_{z \in \mathbb{C}^N, z \neq 0} R(z),$$

le minimum (resp. maximum) étant atteint lorsque le vecteur z est précisément un vecteur propre non-nul de H associé à la valeur propre λ_{min} (resp. λ_{max}).

Théorème 1.3 (Corollaire.) Toute matrice réelle-symétrique S est diagonalisable par une transformation orthogonale et ses valeurs propres sont réelles :

$$S = O D O^T \quad (O, D \text{ réelles, } O^T O = O O^T = I, D \text{ diagonale})$$

De plus, si on définit pour tout vecteur non-nul $x \in \mathbb{R}^N$, le “quotient de Rayleigh” suivant

$$R(x) = \frac{x^T S x}{x^T x}$$

et si on note λ_{min} et λ_{max} la plus petite et la plus grande valeurs propres de S , on a :

$$\lambda_{min} = \min_{x \in \mathbb{R}^N, x \neq 0} R(x), \quad \lambda_{max} = \max_{x \in \mathbb{R}^N, x \neq 0} R(x),$$

le minimum (resp. maximum) étant atteint lorsque le vecteur x est précisément un vecteur propre non-nul de S associé à la valeur propre λ_{min} (resp. λ_{max}).

Théorème 1.4 (Bendixson.) Soit A une matrice carrée quelconque dont les éléments sont complexes. On pose :

$$H_1 = \frac{A + A^*}{2}, \quad H_2 = \frac{A - A^*}{2i}$$

de sorte que H_1 et H_2 sont des matrices hermitiennes et

$$A = H_1 + i H_2.$$

Soient a, b, c, d les réels suivants :

$$\begin{aligned} a &= \lambda_{min}(H_1), \quad b = \lambda_{max}(H_1) \\ c &= \lambda_{min}(H_2), \quad d = \lambda_{max}(H_2). \end{aligned}$$

Alors, pour toute valeur propre λ de la matrice A , on a :

$$\begin{aligned} a &\leq \Re(\lambda) \leq b \\ c &\leq \Im(\lambda) \leq d. \end{aligned}$$

Dans le cas d’une matrice réelle $A = \bar{A}$, les matrices H_1 et iH_2 sont les parties symétrique et anti-symétrique de A :

$$H_1 = \frac{A + A^T}{2} \stackrel{\text{déf}}{=} S, \quad iH_2 = \frac{A - A^T}{2} \stackrel{\text{déf}}{=} \Sigma,$$

$$S^T = S, \quad \Sigma^T = -\Sigma.$$

Le rectangle $[a, b] \times [c, d]$ est alors symétrique par rapport à l’axe des imaginaires, comme le schématise la figure ci-dessous :

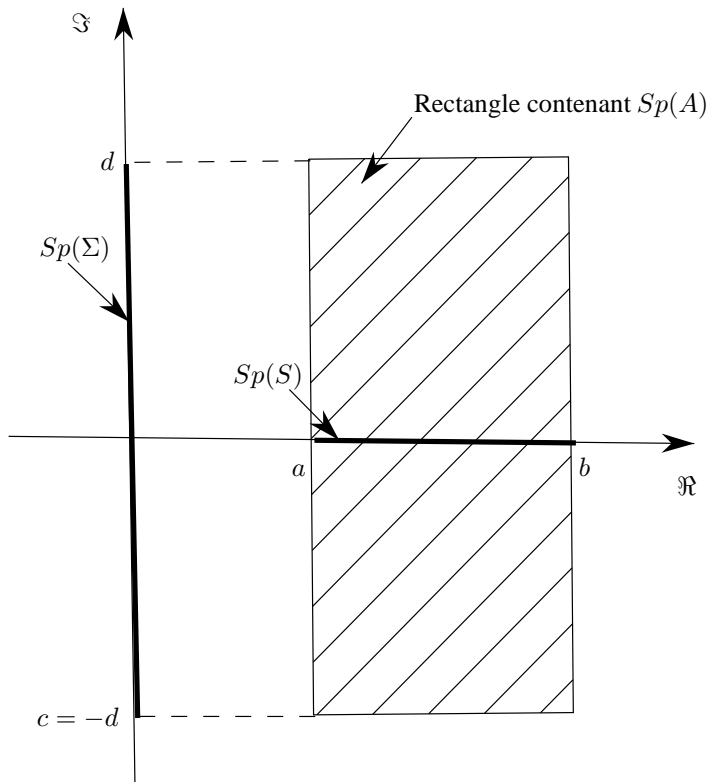


FIG. 1.1 – Illustration du théorème de Bendixon dans le cas d'une matrice réelle

Chapitre 2

Approximation Polynomiale

2.1 Formule(s) de Taylor

Théorème 2.1 Si $f \in C^m([0, T])$ on a

$$f(x) = \sum_{k=0}^{m-1} \frac{x^k}{k!} f^{(k)}(0) + \frac{1}{(m-1)!} \int_0^x (x-y)^{m-1} f^{(m)}(y) dy$$

pour tout $x \in [0, T]$. (Reste intégral).

Définition 2.1 Soient u et v des fonctions de la variable réelle x définies dans un voisinage de $x = a$ (a est un réel ou l'un des symboles $+\infty$, $-\infty$ ou ∞). De plus, v est à valeurs dans \mathbb{R} .

(1) "Grand O". On dit que

$$u(x) = O(v(x))$$

quand $x \rightarrow a$, ssi le rapport $u(x)/v(x)$ reste borné dans cette limite, i.e. il existe un réel positif B et un voisinage \mathcal{V}_a de a tels que :

$$\forall x \in \mathcal{V}_a : \|u(x)\| \leq B |v(x)|.$$

(2) "Petit o". On dit que

$$u(x) = o(v(x))$$

quand $x \rightarrow a$, ssi le rapport $u(x)/v(x)$ tend vers 0 dans cette limite.

Théorème 2.2 (Mêmes hypothèses sur f .) On a

$$f(x) - \sum_{k=0}^{m-1} \frac{x^k}{k!} f^{(k)}(0) = O(x^m)$$

quand $x \rightarrow 0$.

Théorème 2.3 (Mêmes hypothèses sur f .) On a

$$f(x) - \sum_{k=0}^m \frac{x^k}{k!} f^{(k)}(0) = o(x^m)$$

quand $x \rightarrow 0$. (Attention la sommation va jusqu'à m .)

Théorème 2.4 (“Thm de la moyenne”) Soient ϕ et $\psi \in C([a, b])$, à valeurs dans \mathbb{R} , et $\phi \geq 0$. On a :

$$\exists c \in [a, b] \text{ t.q. } \int_a^b \psi(t) \phi(t) dt = \psi(c) \int_a^b \phi(t) dt.$$

Théorème 2.5 (Mêmes hypothèses sur f .) On suppose de plus que f est à valeurs dans \mathbb{R} . Alors :

$$\forall x \in [0, T], \exists c \in [0, x], \text{ t.q. } f(x) - \sum_{k=0}^{m-1} \frac{x^k}{k!} f^{(k)}(0) = \frac{x^m}{m!} f^{(m)}(c).$$

2.2 Interpolation de Lagrange

Théorème 2.6 Soit f une fonction de la variable réelle x , à valeurs dans \mathbb{R} et définie sur l’intervalle $[a, b]$; soit n un entier et $\{x_0, x_1, \dots, x_n\}$ $n+1$ points distincts de $[a, b]$; il existe un polynôme et un seul $P_n(x)$ de degré au plus égal à n tel que :

$$\forall i \in \mathbb{N} \text{ et } i \leq n : P_n(x_i) = f(x_i)$$

Définition 2.2 Le polynôme $P_n(x)$ est appelé “polynôme d’interpolation de Lagrange de la fonction f aux points $\{x_0, x_1, \dots, x_n\}$.” (Le “PIL”.)

Expression du PIL :

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

où :

$$L_i(x) = \frac{(x-x_0)(x-x_1)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_n)}{(x_i-x_0)(x_i-x_1)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_n)} = \prod_{j=0, \dots, n \text{ et } j \neq i} \frac{x-x_j}{x_i-x_j}$$

Théorème 2.7 Mêmes hypothèses, mais on suppose de plus que $f \in C^{n+1}([a, b])$. Alors :

$$\forall x \in [a, b], \exists c \in]a, b[\text{ t.q. } : f(x) - P_n(x) = (x-x_0)(x-x_1)\dots(x-x_n) \frac{f^{(n+1)}(c)}{(n+1)!}$$

(La quantité $e_n(x) = f(x) - P_n(x)$ est appelée “erreur d’interpolation” de f au point x .)

2.3 Forme de Newton

Définition 2.3 (“Différences divisées”) Soient y_0, y_1, \dots des symboles représentant des abscisses toutes distinctes du domaine de définition $[a, b]$ de la fonction f . On introduit récursivement les définitions suivantes :

$$f[y_0] = f(y_0) \quad (\forall y_0 \in [a, b])$$

$$f[y_0, y_1] = \frac{f[y_1] - f[y_0]}{y_1 - y_0} \quad (\forall y_0, y_1 \in [a, b] \text{ distincts})$$

$$f[y_0, y_1, y_2] = \frac{f[y_1, y_2] - f[y_0, y_1]}{y_2 - y_0} \quad (\forall y_0, y_1, y_2 \in [a, b] \text{ distincts})$$

⋮

$$f[y_0, y_1, \dots, y_{k+1}] = \frac{f[y_1, y_2, \dots, y_{k+1}] - f[y_0, y_1, \dots, y_k]}{y_{k+1} - y_0} \quad (\forall y_0, y_1, \dots, y_{k+1} \in [a, b] \text{ distincts})$$

La quantité $f[y_0, y_1, \dots, y_k]$ ainsi calculée est appelée “ k -ème différence divisée de f aux points y_0, y_1, \dots, y_k ”.

A partir d'une suite particulière de valeurs deux à deux distinctes $\{x_i\}$, on peut construire ainsi une "Table des Différences Divisées" comme suit :

x_i	$f[\cdot]$	$f[\cdot, \cdot]$	$f[\cdot, \cdot, \cdot]$	$f[\cdot, \cdot, \cdot, \cdot]$
x_0	$f[x_0]$			
x_1	$f[x_1]$	$f[x_0, x_1]$		
x_2	$f[x_2]$	$f[x_1, x_2]$	$f[x_0, x_1, x_2]$	
x_3	$f[x_3]$	$f[x_2, x_3]$	$f[x_1, x_2, x_3]$	$f[x_0, x_1, x_2, x_3]$

Théorème 2.8 Le polynôme d'interpolation de Lagrange de la fonction f aux points $\{x_0, x_1, \dots, x_n\}$, $P_n(x)$, peut se mettre sous la forme suivante dite "forme de Newton du PIL basée sur les points $\{x_0, x_1, \dots, x_{n-1}\}$ " :

$$\begin{aligned}
 P_n(x) &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \dots \\
 &\quad + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1)\dots(x - x_{n-1}) \\
 &= \sum_{i=0}^n f[x_0, x_1, \dots, x_i] \prod_{j=0}^{i-1} (x - x_j)
 \end{aligned}$$

où par convention $\prod_{j=0}^{-1} (x - x_j) = 1$.

Théorème 2.9 Soit $\xi \in [a, b]$ une valeur particulière de x . On pose

$$a_i = f[x_0, x_1, \dots, x_i]$$

et on construit la suite $\{b_i\}$ comme suit :

$$\begin{cases} b_n = a_n \\ b_i = a_i + (\xi - x_i) b_{i+1} \end{cases} \quad (i = n - 1, n - 2, \dots, 1, 0)$$

Alors :

$$b_0 = P_n(\xi).$$

2.4 Polynômes orthogonaux

Dans cette section, w est une fonction définie-continue et à valeurs strictement positives sur $]a, b[$. De plus, l'intégrale

$$\int_a^b w(x) dx$$

existe.

Définition 2.4 ("Produit Scalaire") Etant donné deux fonctions continues sur $[a, b]$, on définit leur produit scalaire comme suit :

$$\langle u, v \rangle_w = \int_a^b u(x) v(x) w(x) dx.$$

Définition 2.5 (“Orthogonalité”) Deux fonctions u et v définies-continues sur $[a, b]$ sont dites orthogonales ssi :

$$\langle u, v \rangle_w = 0.$$

Théorème 2.10 Il existe une suite infinie de polynômes (deux à deux) orthogonaux $\{\pi_i(x)\}$ tels que $\deg(\pi_i(x)) = i, \forall i$. Cette suite est unique à une “normalisation” près.

Construction : (“Processus d’orthogonalisation de Gram-Schmidt”) On cherche $\pi_i(x)$ sous la forme

$$\pi_i(x) = K_i (x^i + c_{i,i-1} \pi_{i-1}(x) + c_{i,i-2} \pi_{i-2}(x) + \dots + c_{i,0} \pi_0(x))$$

Les constantes $c_{i,i-1}, c_{i,i-2}, \dots, c_{i,0}$ sont déterminées de manière unique par les conditions :

$$\langle \pi_i, \pi_{i-1} \rangle_w = \langle \pi_i, \pi_{i-2} \rangle_w = \dots = \langle \pi_i, \pi_0 \rangle_w = 0$$

et la constante (non-nulle) de normalisation K_i est arbitraire.

Théorème 2.11 La famille $\{\pi_0(x), \pi_1(x), \dots, \pi_k(x)\}$ forment une base des polynômes de degré au plus égal à k .

Théorème 2.12 Si p est un polynôme de degré strictement inférieur à k , alors p est orthogonal à π_k .

Théorème 2.13 Les k zéros du polynôme π_k sont simples, réels et appartiennent à l’intervalle ouvert $]a, b[$.

Théorème 2.14 Les polynômes orthogonaux satisfont la relation de récurrence à trois niveaux suivante :

$$\pi_{i+1}(x) = A_i (x - B_i) \pi_i(x) - C_i \pi_{i-1}(x) \quad (i = 0, 1, \dots, k - 1)$$

où :

$$\begin{aligned} A_i &= K_{i+1}/K_i \\ \pi_{-1}(x) &= 0 \text{ (conventionnellement)} \\ B_i &= \langle x \pi_i, \pi_i \rangle_w / S_i \\ C_i &= A_i S_i / (A_{i-1} S_{i-1}) \\ S_i &= \langle \pi_i, \pi_i \rangle_w. \end{aligned}$$

2.4.1 Polynômes de Legendre

Notation : $\mathcal{L}_k(x)$. Ces polynômes correspondent au cas où $[a, b] = [-1, 1]$ et

$$w(x) = 1$$

de sorte qu’ici :

$$\langle u, v \rangle = \int_{-1}^1 u(x) v(x) dx.$$

Ils satisfont la relation de récurrence suivante :

$$\mathcal{L}_{k+1}(x) = \frac{(2k+1)x\mathcal{L}_k(x) - k\mathcal{L}_{k-1}(x)}{k+1}.$$

k	$\mathcal{L}_k(x)$
0	1
1	x
2	$(3/2)(x^2 - 1/3)$
3	$(5/2)[x^3 - (3/5)x]$
4	$(35/8)[x^4 - (6/7)x^2 + 3/35]$

2.4.2 Polynômes de Tchebychev

Notation : $T_k(x)$. Ces polynômes correspondent au cas où $[a, b] = [-1, 1]$ et

$$w(x) = \frac{1}{\sqrt{1-x^2}}$$

de sorte qu'ici :

$$\langle u, v \rangle = \int_{-1}^1 \frac{u(x)v(x)}{\sqrt{1-x^2}} dx.$$

Ils satisfont la relation de récurrence suivante :

$$T_{k+1}(x) = 2x T_k(x) - T_{k-1}(x).$$

On a également :

$$\forall k \in \mathbb{N}, \forall \theta \in \mathbb{R} : T_k(\cos \theta) = \cos k\theta.$$

k	$T_k(x)$	x^k
0	1	T_0
1	x	T_1
2	$2x^2 - 1$	$(T_0 + T_2)/2$
3	$4x^3 - 3x$	$(3T_1 + T_3)/2^2$
4	$8x^4 - 8x^2 + 1$	$(3T_0 + 4T_2 + T_4)/2^3$
5	$16x^5 - 20x^3 + 5x$	$(10T_1 + 5T_3 + T_5)/2^4$
6	$32x^6 - 48x^4 + 18x^2 - 1$	$(10T_0 + 15T_2 + 6T_4 + T_6)/2^5$
7	$64x^7 - 112x^5 + 56x^3 - 7x$	$(35T_1 + 21T_3 + 7T_5 + T_7)/2^6$

2.5 Approximation par Moindres Carrés

Théorème 2.15 Soit $f \in C([a, b])$. Il existe un polynôme p^* de degré au plus égal à k qui réalise un minimum global de la quantité :

$$E(p) = \|f - p\|^2 = \langle f - p, f - p \rangle_w = \int_a^b (f(x) - p(x))^2 w(x) dx;$$

ce polynôme s'exprime de la manière suivante dans la base orthogonale $\{\pi_i(x)\}$:

$$p^*(x) = \sum_{i=0}^k \alpha_i^* \pi_i(x)$$

où :

$$\alpha_i^* = \frac{\langle f, \pi_i \rangle_w}{S_i}$$

(où $S_i = \langle \pi_i, \pi_i \rangle_w$).

2.6 Meilleure Approximation

Soit f une fonction (au moins) définie-continue sur $[a, b]$. On souhaiterait identifier le choix des points d'interpolation $\{x_i\}$ ($i \leq n$) pour lequel le PIL, $P_n(x)$, réalise la meilleure approximation de f au sens de la norme infinie, c'est-à-dire le minimum de la quantité :

$$e(x_0, x_1, \dots, x_n) = \max_{x \in [a, b]} |f(x) - P_n(x)|$$

La solution de ce problème n'est pas connue en général. Cependant, lorsque $f \in C^{n+1}([a, b])$, la forme connue de l'erreur d'interpolation $e_n(x) = f(x) - P_n(x)$ suggère d'introduire dans ce cas la définition suivante :

Définition 2.6 (“Meilleure Approximation”). On appelle meilleure approximation de f par un polynôme de degré au plus égal à n , le PIL de f , $P_n(x)$, associé aux points d'interpolation $\{x_0, x_1, \dots, x_n\}$ pour lesquels la quantité suivante est minimale :

$$\max_{x \in [a, b]} |(x - x_0)(x - x_1) \dots (x - x_n)|.$$

Théorème 2.16 La meilleure approximation de f par un polynôme de degré au plus égal à n est réalisée par le choix suivant des points d'interpolation :

$$x_i = \frac{b+a}{2} + \frac{b-a}{2} \xi_i \quad (i = 0, 1, \dots, n)$$

où $\xi_0, \xi_1, \dots, \xi_n$ sont les zéros du polynôme de Tchebychev de degré $n+1$, à savoir :

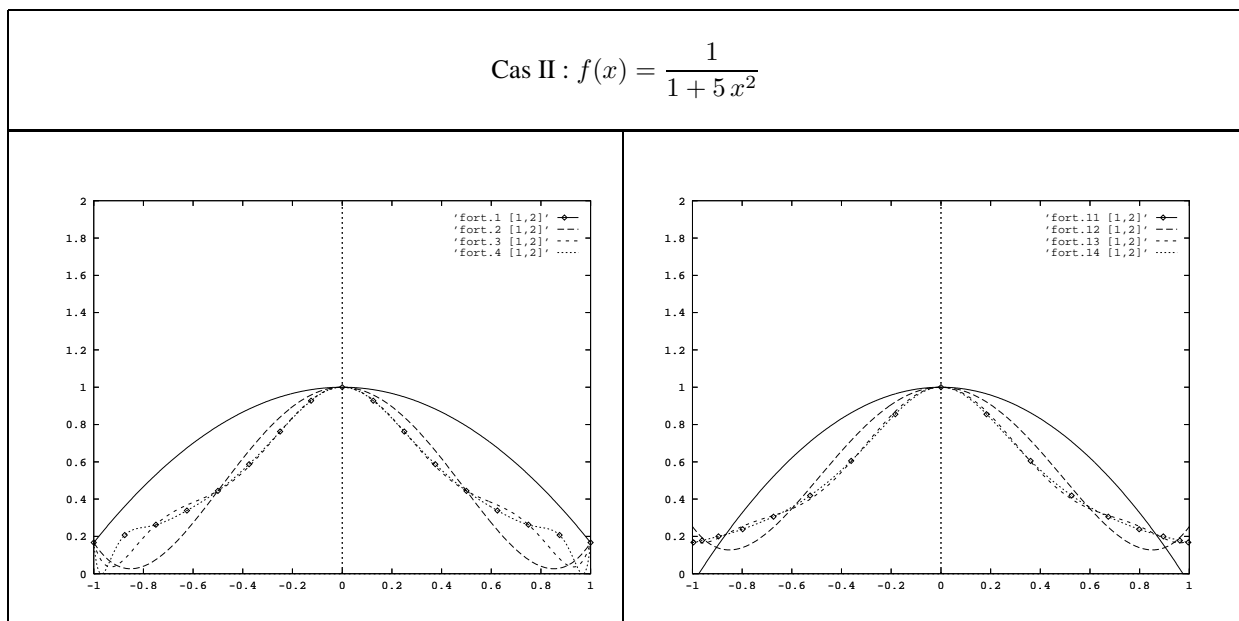
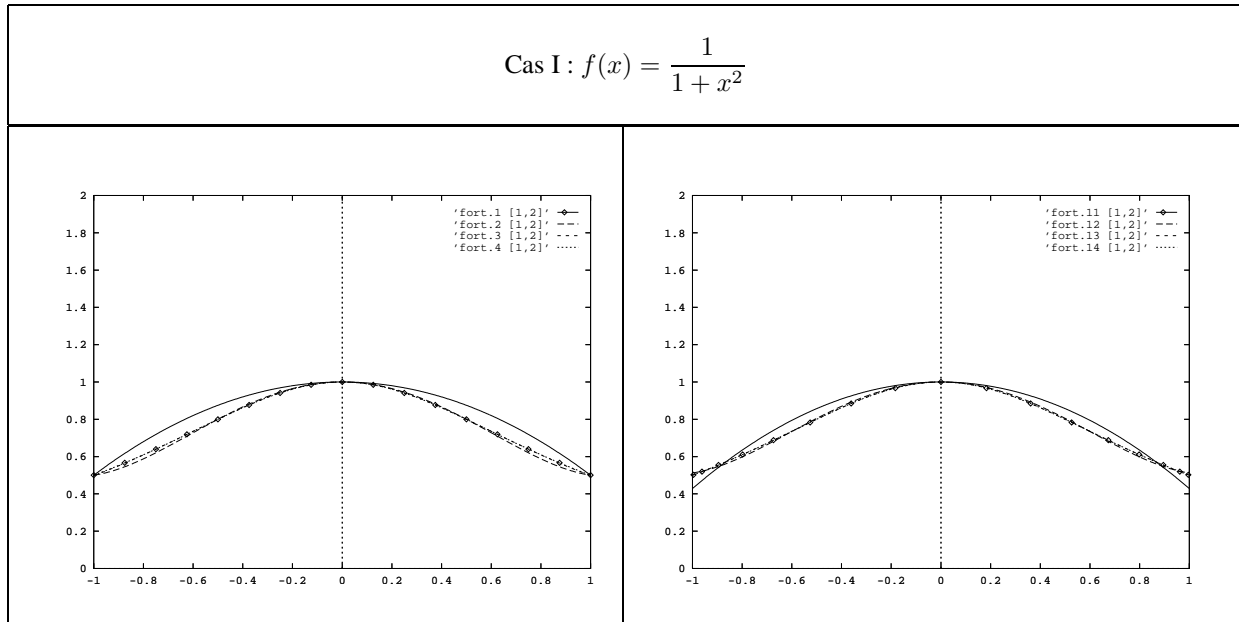
$$\xi_i = \cos \frac{(2i+1)\pi}{2(n+1)} \quad (i = 0, 1, \dots, n).$$

2.7 Oscillations de Gibbs

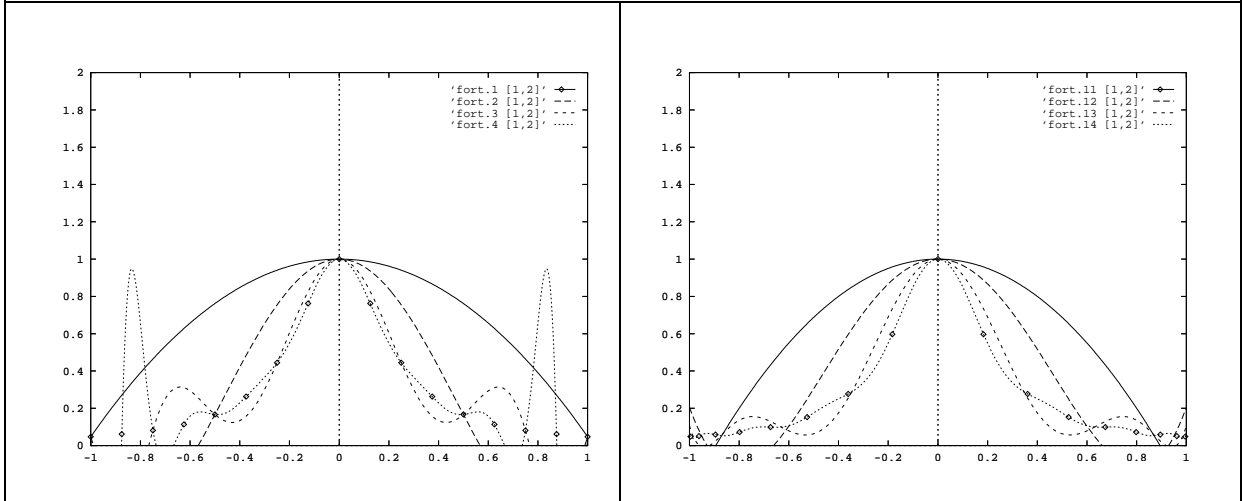
On examine l'approximation d'une fonction f sur $[-1,1]$ par une suite de PIL de degré $n = 2, 4, 8$ et 16 . Les points d'interpolation (au nombre de 3, 5, 9 et 17) sont :

(a) uniformément répartis (à gauche), et (b) les zéros de T_{n+1} (à droite).

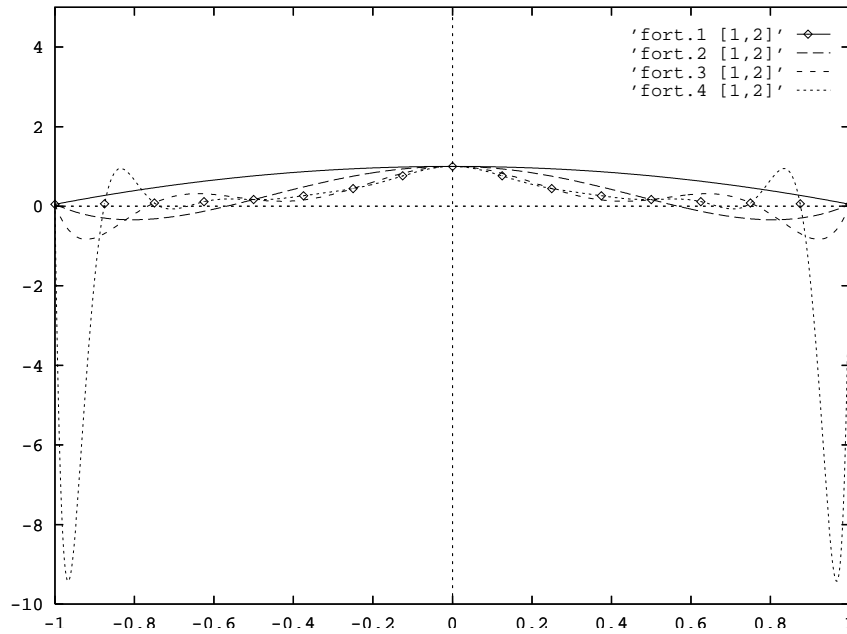
Les \diamond qui représentent les 17 valeurs interpolées pour $n = 16$, servent aussi à visualiser la fonction f .



$$\text{Cas III : } f(x) = \frac{1}{1 + 20x^2}$$



Agrandissement du cas III.(a) :



Chapitre 3

Différentiation et Intégration Numériques

3.1 Introduction

Ce chapitre traite de l'approximation numérique d'une quantité de la forme

$$\mathcal{L}(f)$$

où $f \in \mathcal{C}^k([a, b])$ ($k \geq n + 1$) (au moins) et \mathcal{L} est un opérateur linéaire. En particulier, on s'intéresse aux problèmes généraux suivants :

- approximation de dérivées telles que $f'(a)$, $f''(a)$, etc : différentiation numérique ;
- approximation d'intégrales définies telles que $\int_a^b f(x) dx$, $\int_a^b f(x) w(x) dx$,
 $\frac{1}{\pi} \int_0^{2\pi} f(x) \sin(k\pi x) dx$, etc : intégration numérique.

La méthode générale consiste à d'abord approcher f par un polynôme d'interpolation :

$$f(x) = P_n(x) + e_n(x).$$

On considérera presque toujours une interpolation de Lagrange aux points $\{x_0, x_1, \dots, x_n\}$:

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

$$L_i(x) = \prod_{j=0,1,\dots,n; j \neq i} \frac{x - x_j}{x_i - x_j}$$

$$e_n(x) = \prod_{i=0}^n (x - x_i) \frac{f^{n+1}(\xi)}{(n+1)!}$$

(Exceptionnellement on considérera plutôt une interpolation hermitienne.) Par linéarité de l'opérateur \mathcal{L} on a :

$$\mathcal{L}(f) = \mathcal{L}(P_n) + E_n$$

où

$$\mathcal{L}(P_n) = \sum_{i=0}^n a_i f(x_i)$$

est l'approximation de $\mathcal{L}(f)$ dans laquelle interviennent les coefficients ou "poids" suivants :

$$a_i = \mathcal{L}(L_i)$$

et

$$E_n = \mathcal{L}(e_n)$$

est l'erreur commise. Cette erreur s'analysera cas par cas en faisant une hypothèse de régularité plus ou moins forte sur f .

D'un point de vue algorithmique, on note que dans l'approximation $\mathcal{L}(P_n)$, les poids a_i ne dépendent pas de f , mais seulement du type et du degré de l'interpolation et de la localisation des points d'interpolation : en ce sens, on dit que la formule est une "règle" (de différentiation, d'intégration, etc).

3.2 Différentiation Numérique

Il convient de noter que la différentiation numérique est une opération instable en général ; la justification des formules qui suivent repose donc sur une hypothèse de régularité aussi, et suivant le cas, plus forte que $f \in \mathcal{C}^{n+2}([a, b])$.

3.2.1 Dérivée première

En appliquant la méthode générale au cas où :

$$\mathcal{L}(f) = f'(x_k)$$

on obtient :

$$f'(x_k) = P'_n(x_k) + E_n(x_k)$$

$$E_n(x_k) = e'_n(x_k) = \left(\prod_{i / 0 \leq i \leq n, i \neq k} (x_k - x_i) \right) \frac{f^{n+1}(\xi_k)}{(n+1)!}$$

L'étudiant est invité à identifier les valeurs de n , k et $\{x_i\}$ pour lesquelles la formule ci-dessus fournit les résultats suivants :

$$\begin{aligned} f'(a) &= \frac{f(a+h) - f(a)}{h} - \frac{h}{2} f''(c_1) \\ &= \frac{-3f(a) + 4f(a+h) - f(a+2h)}{2h} + \frac{h^2}{3} f'''(c_2) \\ &= \frac{f(a+h) - f(a-h)}{2h} - \frac{h^2}{6} f'''(c_3) \end{aligned}$$

3.2.2 Dérivées d'ordre supérieur :

Les formules d'approximation des dérivées d'ordre supérieur s'obtiennent généralement par manipulation directe de développements limités (formule de Taylor avec reste de Lagrange). On obtiendra par exemple, les deux formules

suivantes :

$$\begin{aligned} f''(a) &= \frac{f(a-h) - 2f(a) + f(a+h)}{h^2} - \frac{h^2}{12} f^{(4)}(c_4) \\ &= \frac{f(a+2h) - 2f(a+h) + f(a)}{h^2} - f'''(c_5)h \end{aligned}$$

3.3 Intégration Numérique

3.3.1 Règles d'intégration élémentaires

Ce sont des formules d'approximation de l'intégrale définie

$$I(f) = \int_a^b f(x) dx$$

dans lesquelles on utilise un nombre fixe de valeurs de $f(x)$.

La méthode générale donne :

$$I(f) = I(P_n) + E_n$$

où :

$$I(P_n) = \int_a^b P_n(x) dx$$

est l'approximation numérique de l'intégrale, et

$$E_n = \int_a^b e_n(x) dx,$$

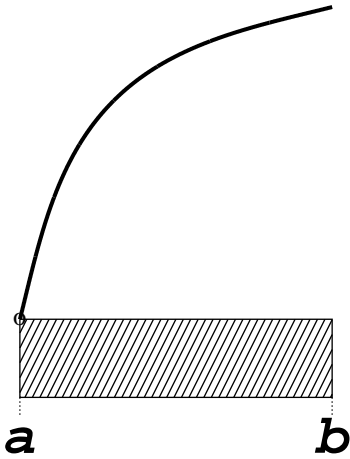
est l'erreur d'intégration.

Notons que l'intégration numérique est une opération stable : il suffit que l'interpolation soit précise pour que l'intégration le soit. En effet, on a l'implication :

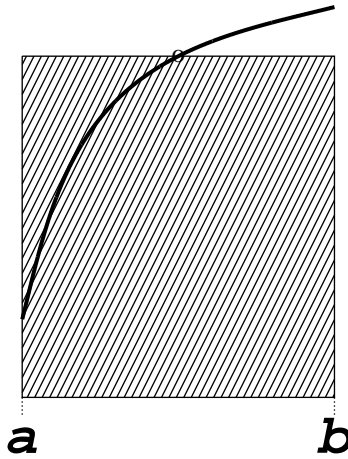
$$\exists \varepsilon > 0 \text{ tel que } \forall x \in [a, b], |e_n(x)| \leq \varepsilon \implies |E_n| \leq \varepsilon(b-a).$$

Pour différents choix du type et du degré de l'interpolation et des points d'interpolation, on obtient différentes "règles d'intégration" classiques :

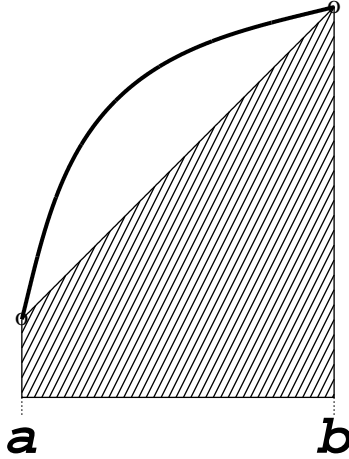
Règle du Rectangle
 $I(P_n) = (b-a) f(a)$
 $E_n = \frac{1}{2} f'(c)(b-a)^2$



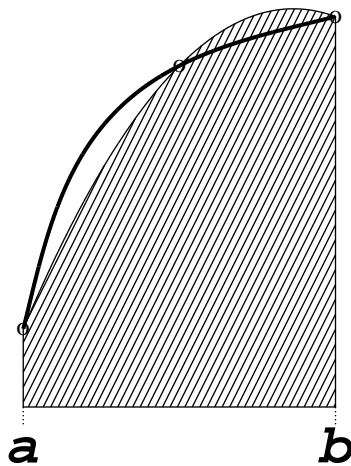
Règle du Point Milieu
 $I(P_n) = (b-a) f\left(\frac{a+b}{2}\right)$
 $E_n = \frac{1}{24} f''(c)(b-a)^3$



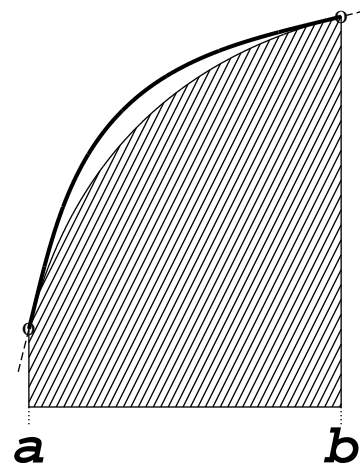
Règle du Trapèze
 $I(P_n) = (b-a) [f(a) + f(b)] / 2$
 $E_n = -\frac{1}{12} f''(c)(b-a)^3$



Règle de Simpson
 $I(P_n) = \frac{1}{6} (b-a) \times$
 $[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)]$
 $E_n = -\frac{1}{90} f''''(c) \left(\frac{b-a}{2}\right)^5$



Règle du Trapèze Corrigée
 $I(P_n) = (b-a) [f(a) + f(b)] / 2$
 $+ \frac{1}{12} (b-a)^2 [f'(a) - f'(b)]$
 $E_n = \frac{1}{720} f''''(c) (b-a)^5$



Nota Bene : Dans le cas de la règle du trapèze corrigée, on a utilisé une interpolation hermitienne.

3.3.2 Règles d'intégration composées

Il s'agit encore d'approcher numériquement l'intégrale définie

$$I(f) = \int_a^b f(x) dx$$

mais ici, on subdivise d'abord l'intervalle $[a, b]$:

$$x_i = a + ih, \quad h = \frac{b-a}{N}, \quad i = 0, 1, \dots, N,$$

et on note :

$$f_i = f(x_i), \quad f_{i-\frac{1}{2}} = f\left[a + \left(i - \frac{1}{2}\right)h\right].$$

On applique alors une règle d'intégration élémentaire à chaque sous-intervalle $[x_{i-1}, x_i]$ et on somme les équations obtenues pour obtenir la règle d'intégration composée (dans laquelle le terme d'erreur est simplifié au moyen du théorème de la moyenne).

Règle composée du rectangle :

$$I(f) = h \sum_{i=1}^N f_{i-1} + \frac{(b-a)f'(\eta_1)}{2} h$$

Règle composée du point central :

$$I(f) = h \sum_{i=1}^N f_{i-\frac{1}{2}} + \frac{(b-a)f''(\eta_2)}{24} h^2$$

Règle composée du trapèze :

$$I(f) = h \sum_{i=1}^{N-1} f_i + \frac{h}{2} (f_0 + f_N) - \frac{(b-a)f'''(\eta_3)}{12} h^2$$

Règle composée de Simpson :

$$I(f) = \frac{h}{6} \left[f_0 + 2 \sum_{i=1}^{N-1} f_i + f_N + 4 \sum_{i=1}^N f_{i-\frac{1}{2}} \right] - \frac{(b-a)f^{(4)}(\eta_4)}{180} \left(\frac{h}{2}\right)^4$$

Règle composée et "corrigée" du trapèze :

$$I(f) = h \sum_{i=1}^{N-1} f_i + \frac{h}{2} (f_0 + f_N) + \frac{h^2}{12} [f'(a) - f'(b)] + \frac{(b-a)f^{(4)}(\eta_5)}{720} h^4$$

3.3.3 Règles d'intégration de Gauss

Pour aboutir aux formules des deux sections précédentes, on a supposé implicitement que la fonction f à intégrer était continûment dérivable sur $[a, b]$ au moins une fois et généralement plusieurs fois. Or on sait bien que l'intégrale définie

$$I(f) = \int_a^b f(x) dx$$

existe sous des hypothèses bien moins fortes. Par exemple, il est suffisant que :

$$\left\{ \begin{array}{l} f \in \mathcal{C}^0([a, b]), \text{ et} \\ \text{dans un voisinage de } x = a^+ : f(x) = O\left(\frac{1}{(x-a)^\alpha}\right), \alpha < 1, \text{ et} \\ \text{dans un voisinage de } x = b^- : f(x) = O\left(\frac{1}{(b-x)^\beta}\right), \beta < 1. \end{array} \right.$$

Plaçons-nous donc dans le cadre où ces hypothèses sont vérifiées et choisissons une "fonction poids" $w(x)$ de la forme :

$$w(x) = \frac{\rho(x)}{(x-a)^\alpha (b-x)^\beta}$$

où la fonction $\rho(x)$ est continue sur $[a, b]$, à valeurs strictement positives et suffisamment simple pour que les intégrales définies

$$\int_a^b x^i w(x) dx$$

($i \in \mathbb{N}$) dont les hypothèses garantissent l'existence, aient des valeurs connues. Supposons enfin que la fonction quotient

$$g(x) = \frac{f(x)}{w(x)}$$

(dont on sait qu'elle est bornée) soit une fonction "régulière" c'est-à-dire au moins continue sur l'intervalle fermé $[a, b]$ et de préférence de classe C^∞ ($[a, b]$).

Ainsi $I(f)$ apparaît comme l'intégrale pondérée suivante :

$$I(f) = J_w(g) = \int_a^b g(x) w(x) dx .$$

En introduisant sur les fonctions continues sur $[a, b]$ le produit scalaire suivant :

$$u, v \in C^0([a, b]) : \langle u, v \rangle_w = \int_a^b u(x) v(x) w(x) dx ,$$

on peut aussi écrire

$$I(f) = \langle g, 1 \rangle_w .$$

La technique de "quadrature gaussienne" pour approcher $I(f)$ consiste alors à effectuer les étapes suivantes :

1. Construire la suite $\{\pi_0(x), \pi_1(x), \dots, \pi_{n+1}(x)\}$ des polynômes orthogonaux au sens du produit scalaire défini ci-dessus.
2. Choisir les $n + 1$ zéros du polynôme $\pi_{n+1}(x)$ (dont on sait qu'ils appartiennent tous à $]a, b[$) comme points d'interpolation :

$$x_0 = \xi_0, x_1 = \xi_1, \dots, x_n = \xi_n ,$$

et construire le polynôme d'interpolation $P_n(x)$ de g :

$$g(x) \approx P_n(x) = g(x_0) L_0(x) + g(x_1) L_1(x) + \dots + g(x_n) L_n(x)$$

où

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

Alors :

$$\begin{aligned} I(f) = J_w(g) \approx J_w(P_n) &= A_0 g(\xi_0) + A_1 g(\xi_1) + \dots + A_n g(\xi_n) \\ &= a_0 f(\xi_0) + a_1 f(\xi_1) + \dots + a_n f(\xi_n) \end{aligned}$$

où :

$$A_i = J_w(L_i) = \int_a^b L_i(x) w(x) dx ; a_i = \frac{A_i}{w(x_i)}$$

Conséquences :

- D'un point de vue algorithmique, si $a, b, w(x)$ ont des valeurs "standard", les polynômes $\{\pi_i(x)\}$ et $\{L_i(x)\}$ sont connus à l'avance indépendamment de f . De même pour les coefficients A_i (ou a_i) de la formule finale que l'on trouve dans une table d'ouvrage spécialisé. Seules les valeurs de g (ou de f) sont à calculer (= notion de règle).

- Du point de vue de la précision de l'approximation, le résultat suivant s'applique :

Théorème 3.1 Si $g \in C^{2n+2}([a, b])$, l'erreur commise sur l'intégrale de f peut se mettre sous la forme :

$$E_n = c_n \times \frac{g^{2n+2}(\xi)}{(2n+2)!}$$

où $c_n = \int_a^b [\pi_{n+1}(x)/K_{n+1}]^2 w(x) dx$, K_{n+1} est le coefficient du terme en x^{n+1} de $\pi_{n+1}(x)$ et $\xi \in]a, b[$.

En particulier, la formule est exacte lorsque g est un polynôme quelconque de degré au plus égal à $2n + 1$.

**Zéros des polynômes de Legendre, ξ_i , et poids de Gauss, A_i ,
associés à la formule de quadrature en ces points**

Le tableau suivant est tiré de Conte & de Boor. Il donne les zéros $\{\xi_i\}$ des premiers polynômes de Legendre et les poids $\{A_i\}$ qui interviennent dans la formule de quadrature de Gauss associée à ces points :

Points et Poids de Gauss pour $n = 1, 2, 3, 4$.

n	$\{\xi_i\} (i = 0, 1, \dots, n)$	$\{A_i\} (i = 0, 1, \dots, n)$
1	$\xi_1 = -\xi_0 = 0.57735027$	$A_1 = A_0 = 1.$
2	$\xi_1 = 0$ $\xi_2 = -\xi_0 = .77459667$	$A_1 = .88888889$ $A_2 = A_0 = .55555556$
3	$\xi_2 = -\xi_1 = .33998104$ $\xi_3 = -\xi_0 = .86113631$	$A_2 = A_1 = .65214515$ $A_3 = A_0 = .34785485$
4	$\xi_2 = 0$ $\xi_4 = -\xi_0 = .90617985$ $\xi_3 = -\xi_1 = .53846931$	$A_2 = .56888889$ $A_4 = A_0 = .23692689$ $A_3 = A_1 = .47862867$

(Des données plus complètes sont fournies par les ouvrages donnant des tables numériques.)

Ces données sont utilisables lorsque la famille de polynômes orthogonaux est bien celle des polynômes de Legendre. Ceci exige que le produit scalaire soit défini comme suit :

$$\langle u, v \rangle_w = \int_{-1}^1 u(x) v(x) dx.$$

En d'autres termes, les choix suivants sont imposés :

$$a = -1, b = 1, w(x) = 1.$$

Le cas échéant, on doit donc au préalable, se ramener à ce cas par changement de variable, et écrire $I(f)$ sous la forme :

$$I(f) = \int_{-1}^1 g(t) dt$$

Alors la formule de quadrature gaussienne s'écrit comme suit :

$$I(f) \approx A_0 g(\xi_0) + A_1 g(\xi_1) + \dots + A_k g(\xi_n)$$

où les $\{A_i\}$ et les $\{\xi_i\}$ sont fournis ci-dessus.

A titre d'illustration, considérons l'exemple suivant tiré de Conte & De Boor :

Soit à calculer une approximation de l'intégrale

$$I = \int_0^1 e^{-x^2} dx .$$

L'intégrande n'ayant aucune singularité, on peut choisir la fonction poids

$$w(x) = 1$$

de sorte que $g = f$ est bien une fonction régulière. De plus, afin que $a = -1$ et $b = 1$, on se ramène à l'intervalle d'intégration $[a, b] = [-1, 1]$ par le changement de variable $x = (t + 1)/2$, de sorte que :

$$g(t) = (1/2) \exp [-(t + 1)^2/4] .$$

Dans le cas de $n = 4$, on prend comme points d'intégration les 5 zéros du polynôme de Legendre de degré 5, $\mathcal{L}_5(x)$; ceux-ci sont fournis par le tableau précédent, et la formule de quadrature gaussienne s'écrit :

$$\begin{aligned} I &\approx \sum_{i=0}^4 A_i g(\xi_i) = (0.56888889) [g(0)] + (.23692689) [g(.90617985) + g(-.90617985)] \\ &\quad + (.47862867) [g(.53846931) + g(-.53846931)] \\ &= .74682413 \end{aligned}$$

D'après Conte et De Boor, les 8 chiffres après la virgule de cette approximation sont exacts, et pour atteindre une précision comparable il aurait fallu subdiviser l'intervalle en 20 dans le cas d'une application de la règle de Simpson, et en 2800 dans le cas d'une application de la règle du trapèze, alors qu'ici, seulement 5 évaluations de la fonction ont suffi. (L'étudiant est invité à vérifier ce résultat à partir des expressions connues des termes d'erreur.)

Chapitre 4

Equations Différentielles

4.1 Rappels

4.1.1 Equation différentielle d'ordre n

($n \in \mathbb{N}^*$)

On dit que la fonction $y = f(x)$ est une solution de l'équation différentielle d'ordre n

$$y^{(n)} = \phi(x, y, y', \dots, y^{(n-1)}) \quad (*)$$

dans laquelle ϕ est une fonction régulière de ses arguments, sur l'intervalle $[a, b]$, ssi :

- $f \in \mathcal{C}^n([a, b])$, et
- $\forall x \in]a, b[$, (*) est vérifiée lorsqu'on substitue $f(x)$ à y , $f'(x)$ à y' , ..., $f^{(n)}(x)$ à $y^{(n)}$.

4.1.2 Forme canonique

En posant

$$Y = \begin{pmatrix} y \\ y' \\ \vdots \\ y^{(n-1)} \end{pmatrix} \in \mathbb{R}^n$$

on ramène (*) à une équation différentielle d'ordre 1 dans laquelle l'inconnue est une application de \mathbb{R} dans \mathbb{R}^n

$$\begin{cases} Y : \mathbb{R} \longrightarrow \mathbb{R}^n \\ Y' = \Phi(x, Y) \end{cases} \quad (**)$$

où

$$\Phi(x, Y) = \begin{pmatrix} y' \\ y'' \\ \vdots \\ y^{(n-1)} \\ \phi(x, y, y', \dots, y^{(n-1)}) \end{pmatrix}$$

Remarque : Pour définir complètement un problème différentiel, il convient d'adjoindre à l'équation différentielle un “système compatible et complet de conditions” qui peuvent être :

- “initiales”

Exemple : Equation du pendule :

$$\begin{cases} J\theta'' + mg \sin \theta = 0 \\ \theta(0) = \theta_0, \theta'(0) = 0 \end{cases} \quad (n = 2, 2 \text{ conditions})$$

(Plus généralement le problème :

$$\begin{cases} (*) \\ y(a) = y_a, y'(a) = y'_a, \dots, y^{(n-1)}(a) = y_a^{(n-1)} \end{cases}$$

ou bien

$$\begin{cases} (**) \\ Y(a) = Y_a \end{cases}$$

est appelé “problème de Cauchy”)

- “aux limites”

Exemple :

$$\begin{cases} y'' = \phi(x, y, y') \\ y(0) = \alpha, y(1) = \beta \end{cases} \quad (n = 2, 2 \text{ conditions})$$

4.1.3 Equation différentielle linéaire “homogène”

:
Définitions :

(*) est linéaire “homogène” lorsque ϕ est linéaire par rapport à $y, y', \dots, y^{(n-1)}$, ou de manière équivalente, (***) est linéaire “homogène” lorsque $\Phi(x, Y)$ est linéaire par rapport à Y

$$\Phi(x, Y) = A(x)Y$$

NB : Cette définition n’exclut pas la possibilité que ϕ (ou Φ) dépende de x .

L’équation différentielle peut alors s’écrire sous la forme :

$$\mathcal{L}y \stackrel{\text{déf}}{=} a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \dots + a_0(x)y = 0$$

Théorème : L’ensemble des solutions est un espace vectoriel de dimension n ,

\iff

Si $y_1(x), y_2(x), \dots, y_n(x)$ sont n solutions indépendantes de l’équation différentielle (*) supposée linéaire, alors la solution générale de cette équation est de la forme :

$$y(x) = C_1 y_1(x) + C_2 y_2(x) + \dots + C_n y_n(x)$$

où C_1, C_2, \dots, C_n sont des constantes arbitraires.

4.1.4 Equation différentielle linéaire homogène à coefficients constants

:
On rajoute à l’hypothèse de linéarité celle d’indépendance de x . (*) peut alors s’écrire sous la forme

$$\mathcal{L}y \stackrel{\text{déf}}{=} a_n y^{(n)} + a_{n-1} y^{(n-1)} + \dots + a_0 y = 0$$

où les coefficients a_n, a_{n-1}, \dots, a_0 sont ici des constantes, et (***) devient :

$$Y' = AY$$

où

$$A = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ -\frac{a_0}{a_n} & -\frac{a_1}{a_n} & \dots & \dots & -\frac{a_{n-1}}{a_n} \end{pmatrix}$$

= matrice constante

Dans ce cas on cherche des solutions particulières de la forme :

$$y(x) = e^{rx} \quad (r \in \mathbb{C})$$

Par simple substitution dans l'équation différentielle on obtient la condition sur r pour que $y(x)$ soit une solution. Cette condition s'écrit après simplification

$$P(r) \stackrel{\text{déf}}{=} a_n r^n + a_{n-1} r^{n-1} + \dots + a_0 = 0$$

ce qui constitue l'équation caractéristique.

Lorsque les solutions r_1, r_2, \dots, r_n en sont distinctes, on obtient n solutions indépendantes $y_1(x), y_2(x), \dots, y_n(x)$ avec $y_i(x) = e^{r_i x}$ qui forment une base de l'espace vectoriel de dimension n des solutions. La "solution générale" est une combinaison linéaire arbitraire de ces solutions particulières.

Dans le cas contraire, notons r_1, r_2, \dots, r_k ($k < n$) les racines distinctes et $\alpha_1, \alpha_2, \dots, \alpha_k$ leurs multiplicités respectives de sorte que :

$$\alpha_1 + \alpha_2 + \dots + \alpha_k = n$$

On peut montrer (Exo.) que les fonctions :

$$\begin{aligned} &e^{r_1 x}, x e^{r_1 x}, x^2 e^{r_1 x}, \dots, x^{\alpha_1 - 1} e^{r_1 x}, \\ &e^{r_2 x}, x e^{r_2 x}, x^2 e^{r_2 x}, \dots, x^{\alpha_2 - 1} e^{r_2 x}, \\ &\quad \vdots \\ &e^{r_k x}, x e^{r_k x}, x^2 e^{r_k x}, \dots, x^{\alpha_k - 1} e^{r_k x}, \end{aligned}$$

sont en nombre n , sont indépendantes, et sont solutions de (*), et forment donc une base de l'espace vectoriel de dimension n des solutions. La "solution générale" est une combinaison linéaire arbitraire de ces solutions particulières.

Exo. : Etablir le lien entre les solutions de l'équation caractéristique et les valeurs propres de la matrice A .

4.1.5 Equation différentielle linéaire non homogène.

Si

$$\mathcal{L}y \stackrel{\text{déf}}{=} a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \dots + a_0(x)y = 0$$

est une équation différentielle linéaire homogène, l'équation

$$\mathcal{L}y = b(x) \quad (***)$$

où b est une fonction donnée, est dite "linéaire non homogène".

Si $y_0(x)$ est une solution particulière,

$$\mathcal{L}y_0(x) = b(x)$$

et si $y_1(x), y_2(x), \dots, y_n(x)$ sont n solutions indépendantes de l'équation linéaire homogène associée,

$$\mathcal{L}y_1(x) = \mathcal{L}y_2(x) = \dots = \mathcal{L}y_n(x) = 0$$

alors la solution générale de (***) est donnée par :

$$y(x) = y_0(x) + C_1 y_1(x) + C_2 y_2(x) + \dots + C_n y_n(x)$$

c'est-à-dire :

$$\begin{aligned} & \text{solution générale de l'équation linéaire non-homogène} \\ & = \\ & \text{solution particulière de l'équation linéaire non-homogène} \\ & + \\ & \text{solution générale de l'équation linéaire homogène associée} \end{aligned}$$

4.1.6 Analogie équations différentielles/équations aux différences

Etant donnée une suite $\{y_i\}$ ($i \in \mathbb{N}$, $y_i \in \mathbb{R}$), on définit l'opérateur de différence avancée Δ par :

$$\Delta y_i = y_{i+1} - y_i$$

de sorte que :

$$\begin{aligned} \Delta^2 y_i &= \Delta(\Delta y_i) = \Delta y_{i+1} - \Delta y_i = y_{i+2} - 2y_{i+1} + y_i \\ \Delta^3 y_i &= \Delta(\Delta^2 y_i) = \Delta^2 y_{i+1} - \Delta^2 y_i = y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i \\ & \vdots \\ \Delta^n y_i &= \Delta \Delta^{n-1} y_i = \Delta^{n-1} y_{i+1} - \Delta^{n-1} y_i = y_{i+n} - n y_{i+n-1} + \frac{n(n-1)}{2} y_{i+n-2} + \dots + (-1)^n y_i \end{aligned}$$

Exo. : Vérifier les formules.

On appelle équation aux différences d'ordre n toute équation de la forme :

$$\Delta^n y_i = \phi(i, y_i, \Delta y_i, \dots, \Delta^{n-1} y_i)$$

dans laquelle la suite $\{y_i\}$ est l'inconnue, et qui doit être satisfaite pour tout i .

Cas particulier : Equation linéaire homogène à coefficients constants.

Après quelques manipulations, l'équation s'écrit :

$$a_n y_{i+n} + a_{n-1} y_{i+n-1} + \dots + a_0 y_i = 0 \quad (a_n \neq 0)$$

dont on cherche des solutions particulières de la forme

$$y_i = r^i \quad (r \in \mathbb{C})$$

Après substitution et simplification, la condition sur r devient :

$$P(r) \stackrel{\text{déf}}{=} a_n r^n + a_{n-1} r^{n-1} + \dots + a_0 = 0$$

qui constitue l'équation caractéristique.

Si les racines r_1, r_2, \dots, r_n sont distinctes, on obtient ainsi n solutions indépendantes et la solution générale en découle :

$$y_i = C_1 r_1^i + C_2 r_2^i + \dots + C_n r_n^i$$

Exo. : Par analogie avec les équations différentielles linéaires homogènes, trouver n solutions indépendantes dans le cas où les racines ne sont pas distinctes, et donner la forme de la solution générale.

Equation non homogène :
Une telle équation s'écrit :

$$a_n y_{n+n} + a_{n-1} y_{i+n-1} + \dots + a_0 y_i = b_i \quad (a_n \neq 0)$$

dans laquelle la suite $\{b_i\}$ ($i \in \mathbb{N}, b_i \in \mathbb{R}$) est donnée.

Alors si \tilde{y}_i est une solution particulière,

$$a_n \tilde{y}_{i+n} + a_{n-1} \tilde{y}_{i+n-1} + \dots + a_0 \tilde{y}_i = b_i$$

la solution générale est de la forme :

$$y_i = \tilde{y}_i + z_i$$

où z_i est la solution générale de l'équation linéaire homogène associée généralement de la forme :

$$z_i = C_1 r_1^i + C_2 r_2^i + \dots + C_n r_n^i$$

A nouveau :

solution générale de l'équation linéaire non-homogène
=
solution particulière de l'équation linéaire non-homogène
+
solution générale de l'équation linéaire homogène associée

4.2 Intégration numérique

Remarque préliminaire : nous avons vu qu'une équation différentielle d'ordre n générale :

$$y^{(n)} = \phi(x, y, y', \dots, y^{(n-1)})$$

pouvait par le changement de variable

$$Y = \begin{pmatrix} y \\ y' \\ \vdots \\ y^{(n-1)} \end{pmatrix}$$

se ramener à la "forme canonique" suivante :

$$Y' = \Phi(x, Y)$$

(pour une définition appropriée de $\Phi(x, Y)$), c'est-à-dire à une équation différentielle du 1er ordre où cette fois-ci l'inconnue est une application $Y(x)$ de \mathbb{R} dans \mathbb{R}^n .

On rappelle que l'on distingue alors deux types de problème :

(a) Problèmes de Cauchy : le vecteur Y est spécifié initialement, disons en $x = x_0$; pour ces problèmes il est suffisamment général de traiter le cas $n = 1$.

(b) Problèmes aux limites : k ($k < n$) composantes du vecteur Y sont spécifiés en $x = a$, et $n - k$ en $x = b$; ces problèmes peuvent s'approcher par exemple par des méthodes de "différences finies" ou des "méthodes de tir".

Dans ce cours, on se limite à l' ETUDE DES PROBLEMES DE CAUCHY. Les méthodes numériques que l'on présente s'appliquent aussi bien au cas où l'inconnue est un vecteur qu'au cas scalaire. Pour ces raisons, il est suffisamment général de traiter le cas où l'équation est du 1er ordre :

$$\begin{cases} y' = \phi(x, y) \\ y(a) = y_0 \end{cases}$$

à intégrer sur $[a, b]$. La solution exacte de ce "problème de Cauchy" sera notée $y(x)$.

4.2.1 Intégration numérique par développement de Taylor

On suppose que la fonction $\phi(x, y)$ est suffisamment différentiable par rapport à ses variables x et y et que la solution exacte du problème, $y(x)$, est de classe $\mathcal{C}^{(k+1)}([a, b])$. Un développement limité à l'ordre k de la solution $y(x)$ autour d'un point quelconque $x_0 \in [a, b]$ s'écrit :

$$y(x) = y_0 + (x - x_0)y'(x_0) + \frac{(x - x_0)^2}{2!}y''(x_0) + \dots + \frac{(x - x_0)^k}{k!}y^{(k)}(x_0) + \frac{(x - x_0)^{k+1}}{(k + 1)!}y^{(k+1)}(\xi)$$

où $\xi \in]x_0, x[$.

Or

$$\begin{cases} y'(x) = \phi(x, y(x)) \\ y''(x) = \phi_x(x, y(x)) + \phi_y(x, y(x))y'(x) \\ = (\phi_x + \phi_y\phi')(x, y(x)) \\ \stackrel{\text{déf}}{=} \phi'(x, y(x)) \\ y'''(x) = (\phi_{xx} + \phi_{xy}y') + (\phi_{yx} + \phi_{yy}y')\phi + \phi_y(\phi_x + \phi_yy') \\ = \phi_{xx} + 2\phi_{xy}\phi + \phi_{yy}\phi^2 + \phi_x\phi_y + \phi_y^2\phi \\ \stackrel{\text{déf}}{=} \phi''(x, y(x)) \\ \text{etc ...} \end{cases}$$

On obtient donc les dérivées successives de y en fonction de x et de $y(x)$ en prenant les dérivées totales de $\phi(x, y(x))$ par rapport à x :

$$\begin{cases} y'(x) = \phi(x, y(x)) \\ y''(x) = \phi'(x, y(x)) \\ y'''(x) = \phi''(x, y(x)) \\ \vdots \\ \text{etc ...} \end{cases}$$

On cherche à faire "avancer" la solution "pas à pas". Le pas, ou incrément spatial, est noté $h = \Delta x$. On est conduit à introduire l'opérateur suivant :

$$T_k(x, y) \stackrel{\text{déf}}{=} \phi(x, y) + \frac{h}{2!}\phi'(x, y) + \frac{h^2}{3!}\phi''(x, y) + \dots + \frac{h^{k-1}}{k!}\phi^{(k-1)}(x, y)$$

de sorte que pour la solution exacte on a :

$$y(x_0 + h) = y_0 + hT_k(x_0, y_0) + E$$

où

$$E = \frac{h^{k+1}}{(k + 1)!}y^{(k+1)}(\xi)$$

On est amené à considérer l'algorithme d'intégration approché suivant :

ALGORITHME DE TAYLOR D'ORDRE k :

1. Initialisation : Choisir un pas $h = \frac{b-a}{N}$. Poser $x_0 = a$, $x_n = a + nh$, $x_N = b$ de sorte que $y(x_n) = y(a + nh)$.

2. Intégration : Calculer des approximations y_n de $y(x_n)$ en appliquant la formule de récurrence suivante :

$$y_{n+1} = y_n + hT_k(x_n, y_n) \\ (n = 0, 1, 2, \dots, N - 1)$$

La quantité

$$E_{loc} \stackrel{\text{déf}}{=} \frac{h^{k+1}}{(k+1)!} y^{(k+1)}(\xi)$$

est appelé "erreur locale". En général, $E_{loc} \neq E$, car en principe on a accumulé des erreurs aux pas précédents et $y_n \neq y(x_n)$. E_{loc} est au contraire l'erreur hypothétique qui serait commise sur y_{n+1} si l'information provenant du pas n , c'est-à-dire y_n , était exacte. Nous allons voir comment cette erreur locale s'accumule en une "erreur globale" (ou "erreur de troncature", ou "erreur de discrétisation")

$$E_{glo} \stackrel{\text{déf}}{=} e_{n+1} = y(x_{n+1}) - y_{n+1}$$

c'est-à-dire l'écart qui existe vraiment entre la valeur exacte et la valeur calculée.

A cette fin, on considère le cas particulier très important qui correspond à $k = 1$. On obtient la

METHODE D'EULER :

$$y_{n+1} = y_n + h\phi(x_n, y_n)$$

pour laquelle l'erreur locale se réduit à :

$$E_{loc} = \frac{h^2}{2} y''(\xi) = O(h^2)$$

Nous allons estimer l'erreur de troncature de cette méthode

$$e_n = y(x_n) - y_n$$

On a :

$$y_{n+1} = y_n + h\phi(x_n, y_n)$$

par définition de la méthode, et

$$y(x_{n+1}) = y(x_n) + h\phi(x_n, y(x_n)) + \frac{h^2}{2} y''(\xi_n)$$

où le dernier terme est l'erreur locale. Par soustraction :

$$e_{n+1} = e_n + h(\phi(x_n, y(x_n)) - \phi(x_n, y_n)) + \frac{h^2}{2} y''(\xi_n) \\ = e_n + h\phi_y(x_n, \bar{y}_n) e_n + \frac{h^2}{2} y''(\xi_n)$$

où \bar{y}^n est une valeur intermédiaire satisfaisant

$$\text{Min}(y_n, y(x_n)) < \bar{y}_n < \text{Max}(y_n, y(x_n))$$

(théorème des accroissements finis).

On considère le cas “normal” où les fonctions qui interviennent sont sans pathologie spéciale et sont en particulier continues. Elle admettent alors des bornes supérieures sur l’intervalle d’étude. Pour fixer les idées, on suppose que

$$|\phi_y(x, y)| < A \quad \text{et} \quad |y''(x)| < B$$

Alors :

$$|e_n + 1| \leq |e_n| + hA|e_n| + B\frac{h^2}{2}$$

et

$$|e_{n+1}| \leq (1 + hA)|e_n| + \frac{Bh^2}{2}$$

Ceci donne :

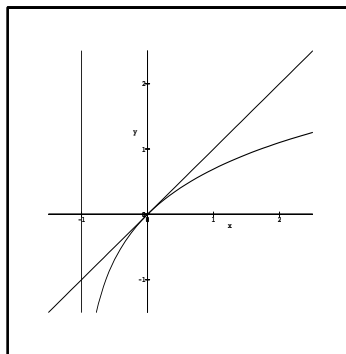
$$\begin{aligned} |e_1| &\leq \frac{Bh^2}{2} \quad \text{car} \quad e_0 = y(a) - y_0 = 0 \\ |e_2| &\leq (1 + hA)|e_1| + \frac{Bh^2}{2} \\ &\leq (1 + hA)\frac{Bh^2}{2} + \frac{Bh^2}{2} \\ &\leq [1 + (1 + hA)]\frac{Bh^2}{2} \\ |e_3| &\leq (1 + hA)|e_2| + \frac{Bh^2}{2} \\ &\leq (1 + hA)[1 + (1 + hA)]\frac{Bh^2}{2} + \frac{Bh^2}{2} \\ &\leq [1 + (1 + hA) + (1 + hA)^2]\frac{Bh^2}{2} \\ &\text{etc...} \end{aligned}$$

On aboutit à :

$$\begin{aligned} |e_n| &\leq [1 + (1 + hA) + \dots + (1 + hA)^{n-1}] \frac{Bh^2}{2} \\ |e_n| &\leq \underbrace{\frac{(1 + hA)^n - 1}{(1 + hA) - 1}}_{hA} \times \frac{Bh^2}{2} \\ |e_n| &\leq \frac{B}{2A} [(1 + hA)^n - 1] \times h \end{aligned}$$

NOTER QUE la sommation a fait apparaître un h au dénominateur, ce qui a pour effet de diminuer d’une unité l’exposant de h .

Remarque : La tangente à l’origine de la courbe d’équation $y = \log(1 + x)$ est la droite $y = x$:



Courbe $y = \log(1 + x)$, son asymptote et sa tangente à l’origine.

Etant donné la concavité uniformément négative de cette courbe, on en conclut que :

$$\forall x > 0, \quad \log(1 + x) \leq x \quad \text{et} \quad (1 + x)^n \leq e^{nx} .$$

Ceci nous permet d’obtenir l’inégalité suivante :

$$(1 + hA)^n \leq e^{nhA} = e^{A(x_n - x_0)}$$

Finalement :

$$|e_n| \leq \frac{B}{2A} [e^{A(x_n - x_0)} - 1] \times h$$

Le terme $[e^{A(x_n - x_0)} - 1]$ est borné lorsqu'on raffine la discrétisation ($h \rightarrow 0$), et cette majoration nous permet de conclure :

CONCLUSION : L'erreur locale est en $O(h^2)$; son accumulation produit une erreur globale en $O(h)$. On dit que la méthode d'Euler est du 1er ordre.

Remarques :

- La borne que nous venons d'établir ne constitue pas nécessairement une estimation précise de l'erreur, ce n'est qu'une borne.
- D'une manière générale nous dirons qu'une méthode est précise à l'ordre k si l'erreur locale E_{loc} est de la forme :

$$E_{loc} = \text{const. } h^{k+1} y^{(k+1)}(\xi)$$

et nous admettons que ceci implique que l'erreur globale ou erreur de troncature

$$E_{glo} = e_n = y(x_n) - y_n = O(h^k)$$

ce qui est vrai de l'algorithme de Taylor d'ordre k , outil théorique auquel on compare les autres méthodes pour mesurer leur précision.

On voit que la méthode d'Euler qui est du 1er ordre, converge lentement puisqu'il faut grosso modo doubler le nombre de pas de discrétisation pour diviser l'erreur par 2. On est tenté d'utiliser une méthode d'ordre plus élevé et notamment l'algorithme de Taylor d'ordre k , pour lequel l'erreur globale qui est en $O(h^k)$ est asymptotiquement divisée par 2^k par doublement du nombre d'intervalles de discrétisation. Cependant, cet algorithme est souvent très compliqué à mettre en oeuvre à cause de la nécessité d'évaluer formellement les dérivées successives $\phi, \phi', \phi'', \dots$. Pour cette raison, bien qu'il garde un intérêt théorique essentiel car il nous a permis de définir la notion d'ordre de précision, on lui préfère en pratique des méthodes de Runge-Kutta, ou des méthodes multipas.

4.2.2 Méthodes de Runge-Kutta

En remplacement de l'algorithme de Taylor d'ordre 2 on propose la formule de récurrence suivante :

$$y_{n+1} = y_n + a k_1 + b k_2$$

où

$$\begin{aligned} k_1 &= h\phi(x_n, y_n) \\ k_2 &= h\phi(x_n + \alpha h, y_n + \beta k_1) \end{aligned}$$

et a, b, α et β sont des constantes réglables.

Un développement de Taylor de k_2/h nous donne :

$$\begin{aligned} k_2/h &= \phi(x_n + \alpha h, y_n + \beta k_1) \\ &= \phi(x_n, y_n) + \alpha h \phi_x + \beta k_1 \phi_y + \frac{\alpha^2 h^2}{2} \phi_{xx} + \alpha \beta h k_1 \phi_{xy} + \frac{\beta^2 k_1^2}{2} \phi_{yy} + O(h^3) \end{aligned}$$

En effectuant les substitutions correspondantes, on obtient :

$$\begin{aligned} y_{n+1} &= y_n + (a + b)h\phi + bh^2(\alpha\phi_x + \beta\phi\phi_y) \\ &\quad + bh^3 \left(\frac{\alpha^2}{2} \phi_{xx} + \alpha\beta\phi\phi_{xy} + \frac{\beta^2}{2} \phi^2 \phi_{yy} + O(h^4) \right) \end{aligned}$$

Ce développement est à comparer à celui (déjà obtenu) de la solution exacte :

$$y(x_{n+1}) = y(x_n) + h\phi(x_n, y_n) + \frac{h^2}{2}(\phi_x + \phi\phi_y) + \frac{h^3}{6}(\phi_{xx} + 2\phi\phi_{xy} + \phi_{yy}\phi^2 + \phi_x\phi_y + \phi_y^2\phi) + O(h^4)$$

On voit que l'on est amené à choisir les constantes de telle sorte que

$$\begin{aligned} a + b &= 1 \\ b\alpha &= b\beta = \frac{1}{2} \end{aligned}$$

ce qui permet d'identifier les trois premiers termes. On constate qu'aucun réglage des coefficients ne permet d'identifier le terme suivant.

On a donc plusieurs solutions. On choisit généralement :

$$a = b = \frac{1}{2} \quad \alpha = \beta = 1$$

ce qui donne la méthode suivante :

METHODE DE RUNGE-KUTTA D'ORDRE 2 :

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

où

$$\begin{aligned} k_1 &= h\phi(x_n, y_n) \\ k_2 &= h\phi(x_n + h, y_n + k_1) \end{aligned}$$

En examinant à nouveau les développements limités de y_{n+1} et de $y(x_{n+1})$ on voit que l'erreur locale (c'est-à-dire ce que vaudrait la différence $y(x_{n+1}) - y_{n+1}$ si l'information provenant du pas n était exacte, c'est-à-dire si on avait $y_n = y(x_n)$) est de la forme :

$$E_{loc} = O(h^3)$$

Il en résulte que

$$|e_n| = |y(x_n) - y_n| = O(h^2)$$

(erreur globale), et la méthode est du second-ordre.

Généralisation : En utilisant la même technique d'ajustement de coefficients, on peut obtenir une famille de méthodes précises au troisième-ordre et une famille de méthodes précises au quatrième-ordre. La plus utilisée est la suivante :

METHODE DE RUNGE-KUTTA D'ORDRE 4.

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

où

$$\begin{aligned} k_1 &= h\phi(x_n, y_n) \\ k_2 &= h\phi\left(x_n + \frac{h}{2}, y_n + \frac{1}{2}k_1\right) \\ k_3 &= h\phi\left(x_n + \frac{h}{2}, y_n + \frac{1}{2}k_2\right) \\ k_4 &= h\phi(x_n + h, y_n + k_3) \end{aligned}$$

L'erreur locale est ici en $O(h^5)$ et l'erreur globale $O(h^4)$; il en résulte que la méthode est du 4ème ordre. Cette précision accrue est au prix de quatre évaluations de la fonction ϕ par pas.

4.2.3 Méthodes Multipas

On a vu dans la méthode de Runge-Kutta qu'une précision globale du 4ème ordre (erreur locale = $0(h^5)$) pouvait être atteinte par une formule combinant judicieusement quatre approximations de la fonction $\varphi(x) \stackrel{\text{déf}}{=} \phi(x, y(x))$. Ces approximations obtenues lors d'étapes intermédiaires du passage de (x_n, y_n) à (x_{n+1}, y_{n+1}) sont inutilisées par la suite, ce qui est une source d'inefficacité.

Dans une méthode multipas au contraire, on cherche à construire une formule utilisant (en même nombre) les valeurs de φ déjà calculées aux pas précédents.

Pour cela on remarque qu'en intégrant l'équation différentielle :

$$y' = \phi(x, y(x)) = \varphi(x)$$

de x_n à x_{n+1} on obtient :

$$y_{n+1} - y_n = \int_{x_n}^{x_{n+1}} \underbrace{\phi(x, y(x))}_{\varphi(x)} dx$$

ce qui équivaut à :

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} \varphi(x) dx$$

On peut donc utiliser la procédure habituelle d'approximation numérique des intégrales.

Supposons que les n premiers pas de l'intégration numérique aient déjà été effectués de sorte que l'on dispose des approximations

$$y_0 \text{ (donné)}, y_1, y_2, \dots, y_n$$

de la fonction inconnue $y(x)$ aux points

$$x_0 = a, x_1 = a + h, x_2 = a + 2h, \dots, x_n = nh$$

(où h est le pas), mais aussi de la fonction $\varphi(x) = \phi(x, y(x)) = y'(x)$:

$$\varphi_0 = \phi(x_0, y_0), \varphi_1 = \phi(x_1, y_1), \varphi_2 = \phi(x_2, y_2), \dots, \varphi_n = \phi(x_n, y_n)$$

En vue de l'intégration numérique de $\varphi(x)$ sur l'intervalle $[x_n, x_{n+1}]$ approchons cette fonction par le polynôme d'interpolation de Lagrange $P_m(x)$ ($m \in \mathbb{N}^*$ fixé) aux $(m+1)$ points de discrétisation précédant x_{n+1} c'est-à-dire : $x_n, x_{n-1}, x_{n-2}, \dots, x_{n-m}$ (en pratique on prendra $m = 3$) :

$$P_m(x) = \varphi_n + \varphi[x_n, x_{n-1}](x - x_n) + \varphi[x_n, x_{n-1}, x_{n-2}](x - x_n)(x - x_{n-1}) \\ + \dots + \varphi[x_n, x_{n-1}, \dots, x_{n-m}](x - x_n)(x - x_{n-1}) \dots (x - x_{n-m+1})$$

Pour simplifier l'écriture on posera

$$\phi_k = \varphi(x_k) = \phi(x_k, y_k)$$

de sorte que

$$\begin{aligned} \varphi_n &= \phi_n \\ \varphi[x_n, x_{n-1}] &= \frac{\phi_n - \phi_{n-1}}{h} \\ \varphi[x_n, x_{n-1}, x_{n-2}] &= \frac{\phi_n - 2\phi_{n-1} + \phi_{n-2}}{2h^2} \\ \varphi[x_n, x_{n-1}, x_{n-2}, x_{n-3}] &= \frac{\phi_n - 3\phi_{n-1} + 3\phi_{n-2} - \phi_{n-3}}{6h^3} \\ &\vdots \end{aligned}$$

On est donc conduit à définir la méthode d'intégration par l'équation :

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} P_m(x) dx$$

On obtient donc dans le cas général :

$$\begin{aligned}
 y_{n+1} &= y_n \\
 &+ \varphi_n \cdot h \\
 &+ \varphi[x_n, x_{n+1}] \int_{x_n}^{x_{n+1}} (x - x_n) dx \\
 &+ \varphi[x_n, x_{n-1}, x_{n-2}] \int_{x_n}^{x_{n+1}} (x - x_n)(x - x_{n-1}) dx \\
 &+ \dots \\
 &+ \varphi[x_n, x_{n-1}, \dots, x_{n-m}] \int_{x_n}^{x_{n+1}} (x - x_n)(x - x_{n-1}) \dots (x - x_{n-m+1}) dx
 \end{aligned}$$

On fait le changement de variable $x = x_n + \theta h$ ($0 \leq \theta \leq 1$) afin d'obtenir :

$$\begin{aligned}
 y_{n+1} = y_n + h \left\{ \varphi_n + h \varphi[x_n, x_{n-1}] \int_0^1 \theta d\theta \right. \\
 + h^2 \varphi[x_n, x_{n-1}, x_{n-2}] \int_0^1 \theta(\theta + 1) d\theta \\
 + h^3 \varphi[x_n, x_{n-1}, x_{n-2}, x_{n-3}] \int_0^1 \theta(\theta + 1)(\theta + 2) d\theta \\
 \left. + \dots + h^m \varphi[x_n, x_{n-1}, \dots, x_{n-m}] \int_0^1 \theta(\theta + 1) \dots (\theta + m - 1) d\theta \right\}
 \end{aligned}$$

On a $\int_0^1 \theta d\theta = \frac{1}{2}$, $\int_0^1 \theta(\theta + 1) d\theta = \frac{5}{6}$, $\int_0^1 \theta(\theta + 1)(\theta + 2) d\theta = \frac{9}{4}$, \dots .

Cas Particulier : METHODE MULTIPAS A 4 PAS :

Dans le cas particulier où $m = 3$, y_{n+1} dépend des valeurs de y aux 4 pas précédents. On obtient :

$$\begin{aligned}
 y_{n+1} = y_n + h \left\{ \varphi_n + \frac{\phi_n - \phi_{n-1}}{1} \cdot \frac{1}{2} + \frac{\phi_n - 2\phi_{n-1} + \phi_{n-2}}{2} \frac{5}{6} \right. \\
 \left. + \frac{\phi_n - 3\phi_{n-1} + 3\phi_{n-2} - \phi_{n-3}}{6} \frac{9}{4} \right\}
 \end{aligned}$$

soit

$$y_{n+1} = y_n + \frac{h}{24} (55\phi_n - 59\phi_{n-1} + 37\phi_{n-2} - 9\phi_{n-3})$$

Erreur locale - Précision

Si toutes les informations sur $[x_0, x_n]$ étaient exactes, on aurait :

$$\varphi(x) = y'(x) = P_m(x) + \frac{y^{(5)}(\xi)}{4!} (x - x_n)(x - x_{n-1}) \dots (x - x_{n-3})$$

d'où par intégration

$$y(x_{n+1}) = y_n + \underbrace{\int_{x_n}^{x_{n+1}} P_m(x) dx}_{y_{n+1}} + \underbrace{\int_{x_n}^{x_{n+1}} \frac{y^{(5)}(\xi)}{4!} (x - x_n) \dots (x - x_{n-3}) dx}_{E_{loc}}$$

ce qui fournit l'erreur locale :

$$E_{loc} = \int_{x_n}^{x_{n+1}} \frac{y^{(5)}(\xi)}{4!} (x - x_n) \dots (x - x_{n-3}) dx$$

Or sur $[x_n, x_{n+1}]$ tous les facteurs $(x - x_n), \dots, (x - x_{n-3})$ sont positifs de sorte que la formule de la moyenne s'applique :

$$E_{loc} = \frac{y^{(5)}(\eta)}{4!} \int_{x_n}^{x_{n+1}} (x - x_n) \cdots (x - x_{n-3}) dx = \frac{h^5 y^{(5)}(\eta)}{4!} \int_0^1 \theta(\theta + 1)(\theta + 2)(\theta + 3) d\theta$$

Finalement :

$$E_{loc} = \frac{251}{720} h^5 y^{(5)}(\eta) = O(h^5)$$

L'erreur globale est donc en $O(h^4)$. La méthode est précise au 4ème ordre.

Son inconvénient majeur réside dans le fait qu'elle ne peut s'appliquer qu'à partir du 4ème pas, les 3 premières approximations devant être obtenues par une autre méthode. L'autre inconvénient réside dans la nécessité de stocker plusieurs valeurs de ϕ ce qui peut être prohibitif dans le cas vectoriel ($y \in \mathbb{R}^p$, ϕ ou $\varphi \in \mathbb{R}^p$, $p \gg 1$). Son avantage principal est que l'on évalue qu'une seule fois la fonction ϕ par pas.

4.2.4 Méthodes prédicteur-correcteur

Dans les méthodes multipas examinées dans la section précédente, le polynôme d'interpolation de Lagrange de $\varphi(x) = y'(x) = \phi(x, y(x))$ est basé sur l'information "en amont" du point calculé c'est-à-dire sur les valeurs aux points $x_n, x_{n-1}, \dots, x_{n-m}$ pour un certain m .

Pour des raisons liées à la précision mais aussi à la stabilité, une notion que nous examinerons dans une section ultérieure, il est souvent préférable de construire un polynôme dont la forme dépend aussi de valeur en x_{n+1} .

Par exemple si dans la formule :

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} \phi(x, y(x)) dx$$

on approche l'intégrale par la formule du trapèze on obtient :

$$y_{n+1} = y_n + \frac{h}{2} [\phi(x_n, y_n) + \phi(x_{n+1}, y_{n+1})]$$

Ce qui constitue une équation "implicite" c'est-à-dire dans laquelle l'inconnue y_{n+1} apparaît à gauche mais aussi à droite, et généralement nonlinéairement, et qu'il faut donc résoudre par rapport à cette inconnue, directement ou itérativement. Dans ce dernier cas, au pas n , on cherche à construire une suite $y_{n+1}^{(0)}, y_{n+1}^{(1)}, \dots, y_{n+1}^{(k)}, \dots$, qui converge vers la solution y_{n+1} de l'équation précédente.

On peut appliquer la méthode d'Euler pour définir le premier élément de la suite $y_{n+1}^{(0)}$. C'est le

Prédicteur :

$$y_{n+1}^{(0)} = y_n + h\phi(x_n, y_n)$$

Aux itérations suivantes, on définit le

Correcteur : ($k \geq 1$) :

$$y_{n+1}^{(k)} = y_n + \frac{h}{2} [\phi(x_n, y_n) + \phi(x_{n+1}, y_{n+1}^{(k-1)})]$$

L'itération au correcteur est prolongée jusqu'à satisfaction d'un critère de convergence qui peut être de la forme :

$$\text{arrêt si : } \left| \frac{y_{n+1}^{(k)} - y_{n+1}^{(k-1)}}{y_{n+1}^{(k)}} \right| < \varepsilon$$

où ε est un nombre petit précisant la tolérance.

Erreur locale A convergence l'erreur locale est celle de la formule d'intégration du trapèze c'est-à-dire de la forme : $-\frac{h^3}{12}y'''(\xi) = O(h^3)$. Par accumulation, l'erreur globale est donc en $O(h^2)$: il en résulte que la méthode est donc du second-ordre.

Convergence du correcteur

L'itération intervenant au correcteur est de la forme :

$$u_k = G(u_{k-1})$$

où $u_k = y_{n+1}^{(k)}$ et $G(u_k) = \text{const.} + \frac{h}{2}\phi(x_{n+1}, u_{k-1})$.

Le "théorème du point fixe" qui n'est pas au programme de ce cours, nous révèle qu'une condition suffisante pour qu'une telle itération converge est que la fonction G satisfasse une hypothèse de contraction du type :

$$\exists K < 1 \quad \text{tel que} \quad \forall u, \quad |G'(u)| \leq K < 1$$

Ici $G'(u) = \frac{h}{2} \frac{\partial \phi}{\partial y}(x_{n+1}, u)$. Il en résulte le théorème suivant :

Théorème : Si $\phi(x, y)$ et $\frac{\partial \phi}{\partial y}$ sont continues en x et y sur l'intervalle fermé $[a, b]$, l'itération intervenant au correcteur converge pourvu que h soit suffisamment petit pour que l'implication suivante soit vraie :

$$|y - y_{n+1}| \leq |y_{n+1}^{(0)} - y_{n+1}| \implies \left| \frac{\partial \phi}{\partial y}(x_{n+1}, y) \frac{h}{2} \right| < 1$$

4.2.5 Notion de stabilité

Afin d'effectuer une analyse simple et complète dans un cas particulier, on commence par définir une méthode d'Adams-Bashforth du second-ordre plus simple.

Une méthode d'Adams-Bashforth du second-ordre

$$y(x_{n+1}) = y(x_{n-1}) + \int_{x_{n-1}}^{x_{n+1}} \underbrace{y'(x)}_{\phi(x, y(x))} dx$$

Or la formule d'intégration du point central nous donne :

$$\begin{aligned} \int_{x_{n-1}}^{x_{n+1}} y'(x) dx &= 2hy'(x_n) + \frac{(2h)^3(y')''(\xi_n)}{24} \\ &= 2h \underbrace{y'(x_n)}_{\phi_n} + \frac{y'''(\xi_n)}{3} h^3 \end{aligned}$$

D'où :

$$y(x_{n+1}) = y(x_{n-1}) + 2h\phi_n + \frac{y'''(\xi_n)}{3} h^3$$

Ceci nous conduit à définir la

METHODE D'ADAMS-BASHFORTH D'ORDRE 2 :

$$y_{n+1} = y_{n-1} + 2h\phi_n$$

où $\phi_n = \phi(x_n, y_n)$.

erreur locale $E_{loc} = \frac{y'''(\xi_n)}{3} h^3$.

erreur globale $E_{glo} = |e(x_n)| = |y(x_n) - y_n| = O(h^2)$.

Intégration numérique d'un exemple

A titre d'exemple, on résout numériquement le problème différentiel suivant :

$$\begin{cases} y' = -2y + 1 & (x \in [0, +\infty]) \\ y(0) = 1 \end{cases}$$

dont la solution exacte est connue,

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

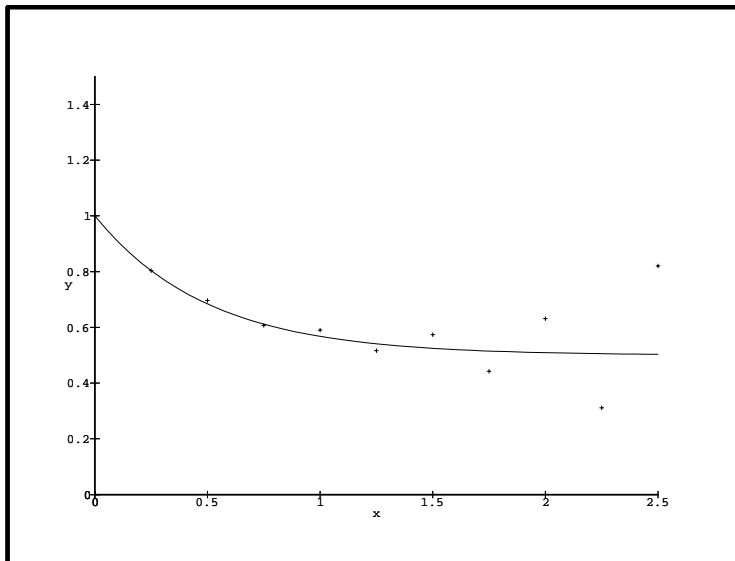
par la méthode d'Adams-Bashforth de la section précédente :

$$\begin{aligned} y_{n+1} &= y_{n-1} + 2h\phi_n \\ &= y_{n-1} + 2h(-2y_n + 1) \end{aligned}$$

que l'on applique pour $n = 1, 2, \dots$ après avoir spécifié non seulement $y_0 = 1$ mais aussi $y_1 = y(h)$.

On constate que la solution numérique initialement proche de la solution exacte, oscille autour de celle-ci et s'en éloigne de plus en plus quand x croît :

c'est le phénomène d'instabilité



Exemple d'intégration numérique instable par la méthode d'Adams-Bashforth (trait continu : solution exacte ; points : solution numérique pour $h = 0.25$).

Analyse du phénomène et définition de la stabilité

Dans ce cas particulier linéaire, la suite des valeurs calculées au moyen du schéma numérique, $\{y_n\}$ satisfait l'équation récurrente linéaire non-homogène suivante

$$y_{n+1} + 4hy_n - y_{n-1} = 2h$$

ainsi que les "conditions initiales" suivantes :

$$y_0 = 1, \quad y_1 = y(h)$$

Or, on connaît une solution particulière de l'équation non-homogène : $\tilde{y}_n = 1/2$. On est donc amené à poser :

$$y_n = \frac{1}{2} + z_n$$

de sorte que la suite $\{z_n\}$ ainsi définie est solution de l'équation linéaire homogène associée :

$$z_{n+1} + 4hz_n - z_{n-1} = 0$$

avec les conditions initiales suivantes :

$$z_0 = \frac{1}{2}, \quad z_1 = y(h) - \frac{1}{2}$$

On sait que

$$z_n = C_1 r_1^n + C_2 r_2^n$$

où r_1, r_2 sont les racines de l'équation caractéristique :

$$r^2 + 4hr - 1 = 0$$

c'est-à-dire :

$$\begin{cases} r_1 = -2h + \sqrt{1 + 4h^2} = 1 - 2h + 0(h^2) \\ r_2 = -2h - \sqrt{1 + 4h^2} = -(1 + 2h) + 0(h^2) \end{cases}$$

et les constantes C_1 et C_2 telles que les conditions initiales sont satisfaites.

Soit $x_n = nh$ fixé. On a :

$$\begin{aligned} \text{Log } r_1^n = n \text{ Log } r_1 &= \frac{x_n}{h} \text{Log}(1 - 2h + 0(h^2)) \\ &= \frac{x_n}{h} [-2h + 0(h^2)] \quad (h \rightarrow 0) \\ &= -2x_n + 0(h) \quad (h \rightarrow 0) \end{aligned}$$

et

$$\begin{aligned} r_1^n &= e^{-2x_n} e^{0(h)} \\ &= e^{-2x_n} [1 + 0(h)] \end{aligned}$$

Les coefficients C_1 et C_2 étant déterminés par les conditions initiales, le terme $\frac{1}{2} + C_1 r_1^n$ approche lorsque $h \rightarrow 0$ la solution exacte : c'est le mode principal.

Par contre,

$$\begin{aligned} \text{Log}|r_2|^n &= \frac{x_n}{h} \text{Log}(1 + 2h + 0(h^2)) \\ &= \frac{x_n}{h} [2h + 0(h^2)] \\ &= +2x_n + 0(h) \end{aligned}$$

d'où

$$|r_2|^n = e^{2x_n} [1 + 0(h)]$$

et, comme $r_2 < 0$,

$$r_2^n = (-|r_2|)^n = (-1)^n e^{2x_n} [1 + 0(h)]$$

Le terme $C_2 r_2^n$ n'approche aucun terme de la solution exacte : c'est un "mode parasite".

Dans ce cas particulier qui illustre le phénomène d'instabilité, on constate que pour h fixé, le mode parasite croît avec x pour devenir prépondérant par rapport au mode principal.

Exo. : Comment explique-t-on que la solution numérique oscille autour de la solution exacte ?

On remarque qu'au départ on avait une équation différentielle du 1er ordre dont la solution générale était constituée d'un seul mode (une seule exponentielle). Afin d'obtenir une méthode précise au second-ordre, on a introduit une formule à trois niveaux ($n+1$, n et $n-1$) représentée par une équation aux différences du 2nd ordre ; on a donc accru artificiellement l'ordre, en approchant une équation différentielle du 1er ordre par une équation aux différences du 2nd ordre ; la solution générale de cette dernière est la superposition d'un mode principal approchant la solution exacte et d'un mode parasite qui dans le cas particulier examiné est "instable". Ceci conduit à poser la définition suivante :

Définition (Stabilité)

On dit qu'une méthode numérique appliquée à une EDO linéaire est stable pour un certain pas h fixé, si les modes parasites présents dans la solution numérique et superposés aux modes principaux sont atténués par l'intégration (lorsque x ou n croît), et instable dans le cas contraire.

En général, les méthodes usuelles sont conditionnellement stable, c'est-à-dire stables lorsque le pas h est suffisamment petit. Mais il existe aussi des méthodes inconditionnellement stables (enfin d'autres, inutilisables, inconditionnellement instables).

Evaluation pratique de la stabilité

On examinera l'application de la méthode proposée à une EDO linéaire. Pour un schéma numérique linéaire, la solution satisfait une équation récurrente linéaire généralement non homogène à $\alpha + 1$ niveaux :

$$a_\alpha y_{n+\alpha} + a_{\alpha-1} y_{n+\alpha-1} + \dots + a_0 y_n = b_n$$

que l'on associe à l'équation récurrente linéaire homogène suivante :

$$a_\alpha z_{n+\alpha} + a_{\alpha-1} z_{n+\alpha-1} + \dots + a_0 z_n = 0$$

dont la solution générale est

$$z_n = c_1 (r_1)^n + c_2 (r_2)^n + \dots + c_\alpha (r_\alpha)^n$$

où $r_1, r_2, \dots, r_\alpha$ sont les racines complexes de l'équation caractéristique :

$$a_\alpha r^\alpha + a_{\alpha-1} r^{\alpha-1} + \dots + a_0 = 0$$

On imposera la condition suivante de STABILITE FORTE

$$\forall i, \quad |r_i| < 1$$

En général, cette condition équivaut à une restriction sur le pas d'avancement h du type

$$h < h_{\max}$$

La méthode est alors conditionnellement stable. Cependant, pour certaines équations différentielles, il peut exister des méthodes d'intégration pour lesquelles la condition de stabilité (SF) est satisfaite $\forall h$; on dit alors que la méthode est "inconditionnellement stable". (De telles méthodes sont généralement "implicites".)

4.2.6 Un exemple d'analyse de méthode

On se propose de résoudre numériquement le problème différentiel suivant :

$$\begin{aligned}y' &= \phi(x, y) = -y \quad (x \in \mathbb{R}_+) \\ y(0) &= 1\end{aligned}$$

par la méthode prédicteur/correcteur suivante :

RUNGE-KUTTA 2 LINEARISE :

$$\begin{cases} \text{Prédicteur : } \bar{y}_{n+1/2} = y_n + \frac{h}{2} \phi(x_n, y_n) \\ \text{Correcteur : } y_{n+1} = y_n + h \phi(x_n + \frac{h}{2}, \bar{y}_{n+1/2}) \end{cases}$$

Il s'agit d'analyser la précision et la stabilité de la méthode.

Analyse

$$\begin{aligned}\bar{y}_{n+1/2} &= y_n - \frac{h}{2} y_n = \left(1 - \frac{h}{2}\right) y_n \\ y_{n+1} &= y_n - h \bar{y}_{n+1/2} = \left[1 - h \left(1 - \frac{h}{2}\right)\right] y_n \\ y_{n+1} &= \left(1 - h + \frac{h^2}{2}\right) y_n\end{aligned}$$

Précision

On évalue d'abord l'erreur locale, c'est-à-dire ce que vaudrait la quantité $y(x_{n+1}) - y_{n+1}$ si on avait $y(x_n) = y_n$.

$$\begin{aligned}y(x_{n+1}) &= y(x_n + h) = y(x_n) + y'(x_n)h + y''(x_n)\frac{h^2}{2} + y'''(x_n)\frac{h^3}{6} + \dots \\ &= y(x_n) - hy(x_n) + \frac{h^2}{2}y(x_n) - \frac{h^3}{6}y(x_n) + \dots\end{aligned}$$

où on a utilisé l'équation différentielle $y' = -y$ pour obtenir

$$y'' = (y')' = (-y)' = y, \quad y''' = (y'')' = (y)' = -y, \quad \text{etc...}$$

On voit donc que dans le cas où $y(x_n) = y_n$ on a :

$$y(x_{n+1}) = y_{n+1} + O(h^3)$$

L'erreur locale est donc $O(h^3)$; l'erreur globale $O(h^2)$: la méthode est du second-ordre.

Stabilité

L'équation aux différences est ici seulement à 2 niveaux

$$y_{n+1} - \left(1 - h + \frac{h^2}{2}\right) y_n = 0$$

Son équation caractéristique associée est donc du 1er ordre :

$$r - \left(1 - h + \frac{h^2}{2}\right) = 0$$

d'où l'unique racine :

$$r_1 = 1 - h \left(1 - \frac{h}{2}\right)$$

Ici r_1 est un réel et la condition de stabilité forte

$$|r_1| < 1$$

se réduit à

$$-1 < r_1 < 1$$

Soit

$$-1 < 1 - h \left(1 - \frac{h}{2}\right) < 1$$

ce qui équivaut à

$$\underline{0 < h < 2}$$

Bibliographie

Quelques références

- *Mathématiques de l'ingénieur* : de très nombreux ouvrages dont :
 - J. Bass, Mathématiques, Tome I, II, III,...Masson.
- *Interpolation, approximation, intégration, équations différentielles* : tout ouvrage introductif d'analyse numérique. En particulier :
 - S. D. Conte et Carl de Boor, "Elementary Numerical Analysis - An algorithmic Approach", McGraw-Hill.
 - P. Henrici, "Discrete Variable Methods in ODE's", John Wiley.
 - M. Crouzeix, A. L. Mignot, "Analyse numérique des équations différentielles", Masson.
- *Analyse matricielle, optimisation* :
 - P. G. Ciarlet, Introduction à l'analyse matricielle et à l'optimisation, Masson.
 - P. Lascaux, R. Théodor, Analyse Numérique matricielle appliquée à l'art de l'ingénieur, Masson.

Ainsi que de très nombreux autres ouvrages que l'on trouve aux rubriques : Analyse Numérique, Approximation, Calcul Matriciel, Equations différentielles, Méthodes Directes/Itératives, Optimisation, Programmation Dynamique, Contrôle Optimal, etc.