Analyse numérique (notes de cours)

Pierre del Castillo

Janvier 2024

Contents

1 Introduction

De nombreux problèmes issus de la physique, de la chimie, de la biologie, de l'économie et des finances conduisent à l'élaboration de modèles mathématiques. L'exploitation de ces modèles nécessite fréquemment d'avoir recours à l'utilisation de méthodes numériques. L'analyse numérique d'un problème se décompose alors en quatre étapes.
1. Modélisation. Obtention d'un modèle mathématiques du problème considéré. Ce modèle peut être constitué d'une équation différentielle ou d'une équation aux dérivées partielles, d'un système d'équations non linéaires, ...
2. Ces modèles sont très difficiles à résoudre mathématiquement, voire impossible à résoudre de façon exacte.
Il est alors nécessaire d'envisager un choix de méthodes numériques afin d'étudier et d'exploiter le modèle.
3. Programmation.
4. Exécution du calcul numérique et interprétation des résultats.
Dans de très nombreux cas, il est impossible de résoudre de façon exacte une équation différentielle ou de calculer une intégrale.
L'objectif du cours d'analyse numérique est de déterminer des méthodes pour calculer la valeur numérique (valeur approchée) d'une intégrale, ou d'une équation. Certaines de ces méthodes seront implémentées sur ordinateur lors des travaux pratiques.
Ce cours est essentiellement subdivisé en trois parties :

- Interpolation polynomiale.
- Intégration numérique.
- Résolution de l'équation $f(x) = 0$.

Dans la première partie, on abordera le problème de l'interpolation polynômiale. Etant donné $n+1$ points distincts $x_0 < \cdots < x_n$ d'un intervalle $[a, b]$, et $f : [a, b] \to I\!R$ une fonction donnée, on cherche un polynôme P de degré le plus petit possible satisfaisant

$$P(x_i) = f(x_i) \quad i = 0, \cdots, n. \tag{1.1}$$

On montrera qu'il existe un unique polynôme de degré inférieur ou égal à n satisfaisant (1.1).
Dans un second temps, on montrera qu'il est possible de calculer P par récurrence sur n par la méthode des différences divisées.
Puis on abordera le problème d'interpolation de Hermite. Etant donné $(y_i)_{0\leq i\leq n}$ et $(z_i)_{0\leq i\leq n}$, on cherche un polynôme P de degré le plus petit possible satisfaisant les conditions

$$P(x_i) = y_i \quad i = 0, \cdots, n, P'(x_i) = z_i \quad i = 0, \cdots, n. \tag{1.2}$$

On établira qu'il existe un unique polynôme de degré inférieur ou égal à $2n+1$ satisfaisant (1.2).
Dans un troisième temps, on montrera qe l'on peut choisir les (x_i) de manière à minimiser l'erreur d'interpolation $e(x) := |f(x) - P(x)|$.
Dans la seconde partie, on déduira de la méthode d'interpolation une méthode afin de déterminer la valeur approchée d'une intégrale d'une fonction d'une variable (formules de Newton-côtes). On présentera les méthodes dites des rectangles, des trapèzes ainsi que la méthode de Simpson. On étudiera dans chaque cas l'erreur obtenue en ayant recours à la méthode de Péano.
La troisième partie est consacrée à la résolution de l'équation $f(x) = 0$. Dans un premier temps, on rappellera la méthode de dichotomie, puis on abordera des méthodes de type point fixe, reposant sur le très important théorème du point fixe. En particulier, on étudiera la méthode des approximations successives ainsi que la méthode de la corde. On montrera que la vitesse de convergence de ces méthodes est "géométrique". La dernière partie sera consacrée à la méthode de Newton. On montrera notamment que la vitesse de convergence de cette méthode est quadratique et que par conséquent, elle est la plus efficace pourvu qu'elle converge.

2 Interpolation polynômiale

2.1 Rappels

2.1.1 Théorèmes de Rolle et des accroissements finis

On rappelle les énoncés de deux théorèmes importants en analyse réelle, le théorème de Rolle et des acroissements finis.

Théorème 2.1 *(Rolle) Soit f de $[a, b]$ dans $\mathbb{R}$, continue sur $[a, b]$, dérivable sur $]a, b[$ telle que $f(a) = f(b)$. Alors il existe $c \in]a, b[$ tel que*

$$f'(c) = 0.$$

Démonstration. Cas 1. f est constante sur $[a,b]$. Alors pour tout $x \in]a,b[$, $f'(x) = 0$.
Cas 2. La fonction f est non constante sur $[a,b]$. La fonction f étant continue sur $[a,b]$, l'image de $[a,b]$ par f est un intervalle fermé borné de $\mathbb{R}$ (l'image d'un compact connexe par une application continue est un compact connexe de $\mathbb{R}$). Etant donné que $f(a) = f(b)$ et que f est non constante sur $[a,b]$, elle admet un maximum ou un minimum sur $[a,b]$ atteint en un point c distinct de a et de b. En ce point, on a $f'(c) = 0$, ce qui achève la preuve du théorème.

Du théorème de Rolle, on déduit l'important théorème suivant :

Théorème 2.2 *(Acroissements finis)*
Soit $f : [a,b] \to \mathbb{R}$ continue sur $[a,b]$ et dérivable $]a,b[$.
Alors il existe $c \in]a,b[$ tel que

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

Preuve du théorème des accroissements finis
L'équation de la droite passant par $A(a, f(a))$ et $B(b, f(b))$ est donnée par

$$y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

On considère la fonction auxiliaire définie sur $[a,b]$ par

$$F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

La fonction F satisfait les hypothèses du théorème de Rolle puisque f est continue sur $[a,b]$ et dérivable sur $]a,b[$. On a clairement $F(a) = 0$). D'autre part, on a

$$F(b) = f(b) - f(a) - \frac{f(b) - f(a)}{b - a}(b - a) = f(b) - f(a) - (f(b) - f(a)) = 0.$$

Les hypothèses du théorème de Rolle sont satisfaites, on en déduit qu'il existe $c \in]a,b[$ tel que

$$F'(c) = 0 = f'(c) - \frac{f(b) - f(a)}{b - a},$$

d'où le résultat.

Le théorème des accroissement finis permet de conclure sur la monotonie d'une fonction à partir de sa dérivée.

Corollaire 2.3 *Soit $f : [a, b] \to \mathbb{R}$ une fonction continue sur $[a, b]$ et dérivable sur $]a, b[$. Alors on a $f'(x) \geq 0$ sur $]a, b[$ si et seulement si f est croissante sur $[a, b]$.*

Démonstration Montrons que si $f'(x) \geq 0$ sur $]a, b[$, f est croissante sur $[a, b]$. Soit $(u, v) \in [a, b]^2$, $u \leq v$. D'après le théorème 2.2, il existe $c \in]u, v[$ tel que

$$f(v) - f(u) = f'(c)(v - u).$$

Il en résulte aussitôt que $f(u) \leq f(v)$ ($u \leq v$ et $f'(c) \geq 0$), donc f est croissante sur $[a, b]$.
Réciproquement, soient $x_0 \in]a, b[$ et $h > 0$ tels que $x_0 + h \in]a, b[$. Comme f est croissante, on a $f(x_0 + h) - f(x_0) > 0$ et donc

$$\lim_{h \to 0^+} \frac{f(x_0 + h) - f(x_0)}{h} = f'(x_0) \geq 0.$$

2.1.2 Le théorème des valeurs intermédiaires

On rappelle l'important théorème suivant, dit théorème des valeurs intermédiaires ;

Théorème 2.4 *Soit f une fonction définie sur $[a, b]$, continue sur $[a, b]$ telle que $f(a).f(b) \leq 0$. Alors il existe $c \in [a, b]$ tel que $f(c) = 0$.*

On déduit du théorème 2.4 la proposition suivante (deuxième formule de la moyenne, version discrète, utile en intégration) :

Proposition 2.5 *Soient $f : [a, b] \to \mathbb{R}$ une fonction continue et $(g_i)_{0 \leq i \leq n}$, $n + 1$ nombres positifs (ou négatifs). Soient (x_i), $n + 1$ points distincts de $[a, b]$.*
Alors, il existe $\zeta \in [a, b]$ tel que

$$\sum_{i=0}^{n} f(x_i) g_i = f(\zeta) \sum_{i=0}^{n} g_i. \tag{2.1}$$

Preuve
On suppose ici $g_i \geq 0$ pour tout i. Si f est constante, le résultat est trivialement vrai. Supposons f non constante sur $[a, b]$.
Considérons la fonction $\psi : x \mapsto \sum_{i=0}^{n} (f(x_i) - f(x)) g_i$. Comme la fonction f est continue sur $[a, b]$, ψ admet un minimum et un maximum atteints respectivement en $\bar{x}$ et $\hat{x}$. On a alors

$$\psi(\bar{x}) \geq 0 \quad \text{et} \quad \psi(\hat{x}) \leq 0.$$

La fonction ψ est continue sur $[a,b]$, elle satisfait les hypothèses du théorème des valeurs intermédiaires. On déduit du théorème des valeurs intermédiaires qu'il existe $\zeta \in [a,b]$ tel que $\psi(\zeta) = 0$, ce qui achève la preuve de la proposition.

2.1.3 Racines d'un polynôme

Le théorème suivant sera fréquemment utilisé dans la suite du cours.

Théorème 2.6 *Soit $P \in \mathbb{C}[X]$ de degré $n \geq 1$. On suppose qu'il existe $\alpha \in \mathbb{C}$ tel que $P(\alpha) = 0$.*
Alors, il existe un polynôme Q de degré $n-1$ tel que

$$P(z) = (z-\alpha)Q(z).$$

Preuve
On effectue la division euclidienne de P par $z-\alpha$. On déduit qu'il existe un polynôme Q de degré $n-1$ et $C \in \mathbb{C}$ tel que

$$P(z) = Q(z)(z-\alpha) + C.$$

On a

$$P(\alpha) = C = 0,$$

d'où la conclusion du théorème.

Remarque 2.7 *Il résulte du théorème 2.6 qu'un polynôme de degré inférieur ou égal à n admettant $n+1$ racines est le polynôme nul.*

On rappelle également ici le très important théorème dû à d'Alembert et à Gauss.

Théorème 2.8 *(Alembert-Gauss)Un polynôme à coefficients complexes non constant admet au moins une racine dans $\mathbb{C}$. Par conséquent, s'il est de degré $n \neq 0$, il admet exactement n racines.*

Preuve Soient $n \in \mathbb{N}^*$ et $P(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0$, avec $a_n \neq 0$. On considère la fonction f définie sur $\mathbb{C}$ à valeurs réelles $f(z) := |P(z)|$. Posons

$$m = \inf_{z \in \mathbb{C}} |P(z)|.$$

Comme le degré de P est supérieur ou égal à 1, on a $|P(z)| \to +\infty$ quand $|z| \to +\infty$ et on peut se ramener à un disque fermé pour chercher le minimum de f sur $\mathbb{C}$. Comme f est continue et que les compacts de $\mathbb{C}$ sont les sous-ensembles fermés et bornés de $\mathbb{C}$, la fonction f admet un minimum (noté c_0)

sur $\mathbb{C}$ atteint en z_0.
Effectuons le changement de variable $u = z - z_0$. On a alors

$$P(z) = P(u + z_0) = c_0 + c_1 u + \cdots + c_n u^n.$$

Supposons que P n'admette pas de racine dans $\mathbb{C}$, autrement dit que

$$c_0 \neq 0.$$

Soit $p \in \mathbb{N}^*$, le plus petit indice tel que $c_p \neq 0$. On a

$$P(u + z_0) = c_0(1 + \frac{c_p}{c_0}u^p + \cdots + \frac{c_n}{c_0}u^n).$$

Le nombre complexe $-\frac{c_0}{c_p}$ admet une racine p-ième, autrement dit, il existe $\lambda \in \mathbb{C}$ telle que $\lambda^p = -\dfrac{c_0}{c_p}$. Effectuons alors le changement de variable $u = \lambda v$. On obtient

$$P(\lambda v + z_0) = c_0 \left(1 - v^p + \cdots + \frac{c_n}{c_0}\lambda^n v^n\right),$$

ou encore

$$P(\lambda v + z_0) = c_0(1 - v^p + v^p \epsilon(v))$$

où $\epsilon(v)$ tend vers 0 quand v tend vers 0.
Il existe $v_0 \in \mathbb{C}$ tel que $|1 - v_0^p + v_0^p \epsilon(v_0)| < 1$. Il suffit de prendre $v_0 > 0$ assez proche de 0 pour obtenir cette inégalité. Ainsi, on obtient :

$$|P(\lambda v_0 + z_0)| < |c_0|$$

ce qui contredit le fait que c_0 est le minimum de f. Donc P admet au moins une racine dans $\mathbb{C}$. Il résulte alors du théorème 2.6 qu'il admet exactement n racines.

2.1.4 Formules de Taylor

1. Formule de Taylor avec reste de Young.

Théorème 2.9 *Soit f une fonction définie sur I, n fois dérivable au point d'abscisse $x = a \in I$. Alors f admet un développement limité d'ordre n en a et de plus, on a pour $x \in I$*

$$f(x) = f(a) + f'(a)(x - a) + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n + (x - a)^n \epsilon(x - a), \quad (2.2)$$

où $\lim_{x \to a} \epsilon(x - a) = 0$.

Preuve On peut sans perdre en généralités établir le résultat en $x = 0$. Il suffit de poser $g(x) = f(x + a)$. La preuve est donc établie en $x = 0$.
La formule est vraie pour $n = 1$. Il a été établi en analyse appliquée que f est dérivable au point $x = 0$ **si et seulement si** elle admet un développement limité d'ordre 1 en ce point.
La preuve de la formule de Taylor-Young s'obtient par récurrence sur n en appliquant le théorème des accroissements finis entre 0 et x à la fonction ϕ définie par

$$\phi(x) := f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k. \tag{2.3}$$

En effet, supposons le résultat vrai au rang $n-1$ ($n \geq 2$) et considérons la fonction ϕ définie en (2.3). On a $\phi(0) = 0$ et d'autre part, appliquant le théorème des accroissements finis entre 0 et x, on obtient qu'il existe un nombre $c_x \in]x, 0[\cup]0, x[$ tel que

$$\phi(x) - \phi(0) = \phi'(c_x)x. \tag{2.4}$$

Mais la fonction f' est $n-1$ fois dérivable et on peut lui appliquer l'hypothèse de récurrence. On a donc

$$f'(x) = f'(0) + f''(0)x + \cdots + \frac{f^{(n)}(0)}{(n-1)!} x^{n-1} + x^{n-1}\epsilon(x),$$

où $\lim_{x\to 0} \epsilon(x) = 0$. On déduit alors de (2.3) et (2.4) que

$$f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k = \phi'(c_x)x = x c_x^{n-1}\epsilon(c_x).$$

On a alors

$$x c_x^{n-1}\epsilon(c_x) = x^n \frac{c_x^{n-1}}{x^{n-1}}\epsilon(c_x).$$

On pose $\epsilon_1(x) = \frac{c_x^{n-1}}{x^{n-1}}\epsilon(c_x)$. Compte tenu de la définition de c_x (en particulier du fait que c_x tend vers 0 quand x tend vers 0), on en déduit que

$$f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k = x^n \epsilon_1(x),$$

avec $\epsilon_1(x)$ qui tend vers 0 quand x tend vers 0. Le résultat est donc vrai au rang n, ce qui achève la preuve du théorème.

2. Formule de Taylor avec reste intégral.

Théorème 2.10 *Soient f une fonction de classe C^{n+1} sur I et $a \in I$. On a l'égalité*

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x), \tag{2.5}$$

où

$$R_n(x) = \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)dt = \int_0^1 \frac{(1-t)^n}{n!} f^{(n+1)}(a+t(x-a))(x-a)^{n+1}dt. \tag{2.6}$$

On donne ici la preuve de la formule de Taylor avec reste intégral. Elle repose sur le lemme suivant :

Lemme

Soit v une fonction définie sur I de classe C^{n+1}. On a l'égalité

$$\frac{d}{dx}[v(x) + (1-x)v'(x) + \cdots + \frac{(1-x)^n}{n!}v^{(n)}(x)] = \frac{(1-x)^n}{n!}v^{(n+1)}(x) \tag{2.7}$$

Démonstration du lemme Effectuons une récurence sur n.
Si $n = 0$, l'égalité est satisfaite. Supposons l'égalité satisfaite au rang n et montrons qu'elle est vraie au rang $n+1$. On a

$$\begin{aligned}
&\frac{d}{dx}[v(x) + (1-x)v'(x) + \cdots + \frac{(1-x)^n}{n!}v^{(n)}(x) + \frac{(1-x)^{n+1}}{(n+1)!}v^{(n+1)}(x)] \\
&= \frac{d}{dx}[v(x) + (1-x)v'(x) + \cdots + \frac{(1-x)^n}{n!}v^{(n)}(x)] + \frac{d}{dx}\left(\frac{(1-x)^{n+1}}{(n+1)!}v^{(n+1)}(x)\right) \\
&= \frac{(1-x)^n}{n!}v^{(n+1)}(x) - (n+1)\frac{(1-x)^n}{(n+1)!}v^{(n+1)}(x) + \frac{(1-x)^{n+1}}{(n+1)!}v^{(n+2)}(x)
\end{aligned}$$

d'où le résultat.

De l'égalité (2.7), intégrée entre 0 et 1, on déduit immédiatement la proposition

Proposition 2.11 *Soit v une fonction de classe C^{n+1} sur $[0,1]$. On a l'égalité*

$$v(1) - v(0) - v'(0) - \cdots - \frac{1}{n!}v^n(0) = \int_0^1 \frac{(1-t)^n}{n!}v^{(n+1)}(t)dt. \tag{2.8}$$

Démonstration du théorème 2.10 On pose $v(t) = f(a + t(x-a))$. On a alors pour tout $t \in I$

$$v^{(n)}(t) = f^{(n)}(a + t(x-a))(x-a)^n.$$

En réecrivant l'égalité (2.29), on déduit (2.10).

3. Formule de Taylor avec reste de Lagrange.
Le théorème suivant est dû au mathématicien Lagrange.

Théorème 2.12 *Soit $n \in \mathbb{N}$. Soient f une fonction définie sur I, $n+1$ fois dérivable sur I et $a \in I$. Alors on a :*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x), \tag{2.9}$$

où

$$R_n(x) = \frac{f^{(n+1)}(a+\theta_x(x-a))}{(n+1)!}(x-a)^{n+1}, \quad \theta_x \in]0,1[.$$

Remarque 2.13 *Le théorème précédent est une généralisation de la formule des accroissements finis. En effet, dans le cas où $n = 0$, on retrouve le théorème des accroissements finis. Rappelons que dans la précédente section, on a exprimé le théorème des accroissements finis sous la forme suivante : étant donné $x > a$*

$$f(x) - f(a) = f'(c)(x-a), \quad c \in]a,x[.$$

Dire que $c \in]a,x[$, c'est dire qu'il existe $\theta \in]0,1[$ tel que $c = \theta x + (1-\theta)a = a + \theta(x-a)$. On retrouve ainsi la formule (2.9) dans le cas $n = 0$.

2.2 Interpolation de Lagrange

2.2.1 Existence et unicité du polynôme de Lagrange

On considère une fonction f définie sur $[a,b]$ à valeurs réelles et (x_i) $n+1$ points de $[a,b]$ tels que $a \leq x_0 < x_1 < \cdots < x_{n-1} < x_n \leq b$. On cherche un polyôme de degré minimal satisfaisant les conditions

$$P(x_i) = f(x_i), \quad \forall\, i = 0, \cdots, n. \tag{2.10}$$

Théorème 2.14 *Il existe un unique polynôme de degré inférieur ou égal à n satisfaisant les conditions (2.10).*

Preuve a. Existence du polynôme d'interpolation.
Contruisons des polynômes l_i $i = 0, \cdots, n$ tels que

$$l_i(x_j) = \delta_{i,j}.$$

On appelle l_i le ième polynôme élémentaire de Lagrange. Pour tout $j \neq i$, x_j est racine de l_i, donc d'après le théorème 2.6, on a

$$l_i(x) = C. \prod_{j=0, j\neq i}^{n} (x - x_j).$$

La constante C est déterminée par la condition $l(x_i) = 1$, et on obtient immédiatement $C = \dfrac{1}{\prod_{j=0,j\neq i}^{n}(x_i - x_j)}$. Par construction, le polynôme l_i recherché est donné par

$$l_i(x) = \prod_{j=0, j\neq i} \frac{x - x_j}{x_i - x_j}.$$

Le polynôme défini par

$$P(x) = \sum_{i=0}^{n} f(x_i) l_i(x)$$

satisfait les conditions (2.10).

b. Unicité du polynôme d'interpolation
Supposons qu'il existe P et Q satisfaisant 2.10. Alors le polynôme $P - Q$ admet $n + 1$ racines et son degré est inférieur ou égal à n. On déduit de la remarque 2.7 qu'il est identiquement nul. Donc $P = Q$.

Une autre approche possible pour déterminer le polynôme interpolant f est de chercher P sous la forme $P(x) = a_0 + a_1 x + \cdots + a_n x^n$, a_i à déterminer de telle sorte que (2.10) soit vérifiée. Le système linéaire obtenue est de la forme

$$A_n X = b_n,$$

où $b_n = (f(x_0), \cdots, f(x_n))^t$, $X = (a_i)^t_{i=0,\cdots,n}$ et $(A_n)_{i,j} = x_{i-1}^{j-1}$, $1 \leq i, j \leq n + 1$.

Proposition 2.15 *On a l'égalité*

$$\det A_n = \prod_{0\leq j<i\leq n} (x_i - x_j).$$

A_n est inversible et la solution du système $A_n X = b_n$ existe et est unique.

Preuve
Effectuons un raisonnement par réccurence sur n. Le résultat est vrai pour $n = 1$. En effet, dans ce cas,

$$\det A_1 = x_1 - x_0 \quad \text{et} \quad \prod_{0 \leq j < i \leq 1} (x_i - x_j) = x_1 - x_0.$$

Supposons le résultat vrai au rang $n-1$ ($n \geq 2$) et montrons qu'alors il est vrai au rang n. Remplaçons x_n par x dans l'expression de A_n (on notera par $A_n(x)$ la matrice ainsi obtenue) et considérons l'application $x \mapsto detA_n(x)$ notée ψ. Remarquons que ψ est un polynôme de degré inférieur ou égal à n en l'indéterminée x et que, d'après les propriétés du déterminant, ψ s'annule en x_0, ...,x_{n-1}. On a donc

$$\psi(x) = C \prod_{j=0}^{n-1} (x - x_j).$$

Déterminons C, le coefficient du monôme de plus haut degré de ψ. La constante C est obtenue en développant le déterminant de A_n par rapport à la dernière ligne et précisément, on a

$$C = \det A_{n-1}$$

Mais par hypothèse de récurrence, on a

$$C = \prod_{0 \leq j < i \leq n-1} (x_i - x_j).$$

Finalement, on obtient

$$\det A_n = \prod_{0 \leq j < i \leq n-1} (x_i - x_j). \prod_{j=0}^{n-1} (x_n - x_j) = \prod_{0 \leq j < i \leq n} (x_i - x_j).$$

Proposition 2.16 *Le système* (l_i) $i = 0, \cdots, n$ *constitue une base de* $\mathbb{R}_n[X]$.

Preuve La dimension de $\mathbb{R}_n[X]$ est égale à $n+1$. Pour montrer que le système constitue une base, il suffit de montrer qu'il est libre. Soit un $n+1$-uplet $(a_0, a_1, \cdots, a_n)$ tel que

$$\sum_{i=0}^{n} a_i l_i(x) = 0. \tag{2.11}$$

Posons $x = x_k$, $k \in \{0, \cdots, n\}$ dans (2.11). On obtient alors $a_k = 0$. Le système est donc libre et il constitue une base de $\mathbb{R}_n[X]$.

Exemple de calcul du polynôme d'interpolation
Considérons la fonction

$$f(x) = \frac{2^x}{x+1}.$$

On veut interpoler cette fonction aux points $x_0 = 0$, $x_1 = 1$ et $x_2 = 2$. On a $f(x_0) = 1$, $f(x_1) = 1$ et $f(2) = \frac{4}{3}$. Les polynômes élémentaires de Lagrange sont donnés par

$$l_0(x) = \frac{(x-1)(x-2)}{2},$$

$l_1(x) = -x(x-2)$ et $l_2(x) = \frac{x(x-1)}{2}$. D'après le théorème 2.14, le polynôme P interpolant f est donné par

$$P(x) = l_0(x) + l_1(x) + \frac{4}{3}l_2(x).$$

2.2.2 Estimation de l'erreur dans le cas où f est de classe C^{n+1}

On fait ici l'hypothèse supplémentaire que f est de classe C^{n+1} et on se propose de déterminer une expression de l'erreur $f(x) - P_n(x)$. On utilisera une partie des résultats obtenus dans le lemme suivant, dont la démonstration repose sur le théorème de Rolle :

Lemme 2.17 *Soit $n \in \mathbb{N}^*$. Soit f une fonction de classe C^n sur $[a,b]$ admettant $n+1$ racines distinctes dans $[a,b]$. Alors il existe $\rho \in]a,b[$ tel que*

$$f^{(n)}(\rho) = 0.$$

Preuve Le résultat est vrai pour $n = 1$ d'après le théorème de Rolle.
Supposons le résultat vrai pour $n \geq 1$ et montrons qu'il est alors vrai au rang $n+1$. Soient x_1, ..., x_{n+2} les $n+2$ racines simples de f. Appliquons le théorème de Rolle sur chaque intervalle de la forme $[x_i, x_{i+1}]$, pour $i = 1, \cdots, n+1$. Il existe $\rho_i \in]x_i, x_{i+1}[$ tel que

$$f'(\rho_i) = 0.$$

La fonction f' admet donc $n+1$ racines distinctes et par hypothèse de récurrence, il existe $\rho \in]a,b[$ tel que

$$(f')^{(n)}(\rho) = 0.$$

Le résultat est donc établi.

Définition 2.18 *On dit qu'une fonction f de classe C^n sur $\mathbb{R}$ admet x_0 comme racine de multiplicité n si*

$$f(x_0) = f'(x_0) = \cdots = f^{(n-1)}(x_0) = 0.$$

En particulier, on dit que x_0 est une racine double de f si

$$f(x_0) = f'(x_0) = 0.$$

On peut déduire du lemme 2.17 le lemme suivant :

Lemme 2.19 *Soit f une fonction de classe C^{2n+2} sur $[a,b]$ admettant $n+1$ racines doubles distinctes dans $[a,b]$ et une racine simple. Alors il existe $\rho \in]a,b[$ tel que*

$$f^{(2n+2)}(\rho) = 0.$$

Preuve
Soient x_1, ..., x_{n+1} les $n+1$ racines doubles de f et y l'unique racine simple. On peut supposer sans perdre en généralités que $x_1 < x_2 < \cdots < x_{n+1} < y$. La fonction f' admet pour racines simples $x_1 < x_2 < \cdots < x_{n+1}$. D'autre part, en appliquant le théorème de Rolle sur les intervalles $[x_i, x_{i+1}]$ pour $i = 1, \cdots, n$ et sur $[x_{n+1}, y]$, on obtient l'existence de $n+1$ racines simple de f', distinctes de x_1, ..., x_{n+1}. La fonction f' admet donc $2n+2$ racines simples, et d'après le lemme 2.17, on déduit qu'il existe $\rho \in]a,b[$ tel que

$$f'^{(2n+1)}(\rho) = 0,$$

ce qui achève la preuve du lemme 2.19.

Théorème 2.20 *Soit $f \in C^{n+1}([a,b])$. Pour tout $x \in [a,b]$, il existe $\rho_x \in]a,b[$ tel que*

$$f(x) - P_n(x) = \frac{f^{(n+1)}(\rho_x)}{(n+1)!} \prod_{i=0}^{n} (x - x_i). \tag{2.12}$$

Preuve
Remarquons que si $x = x_0, \cdots, x_n$, la conclusion du théorème 2.20 est vraie. Pour x distinct de $x_0, \cdots, x_n$, on pose

$$K_x = \frac{f(x) - P_n(x)}{\prod_{i=0}^{n}(x - x_i)}$$

On considère la fonction auxiliaire, pour $t \in [a,b]$

$$\phi_x := f(t) - P_n(t) - K_x \prod_{i=0}^{n} (t - x_i).$$

Remarquons que ϕ_x admet pour racines $x_0, \cdots, x_n$ et que compte tenu de la définition de K_x, on a $\phi_x(x) = 0$. La fonction ϕ_x admet donc $n+2$ racines simples distinctes, et de plus, elle est de classe C^{n+1}. D'après le lemme 2.17, on déduit qu'il existe $\rho_x \in]a, b[$ tel que

$$\phi_x^{(n+1)}(\rho_x) = 0.$$

Or, comme P_n est de degré inférieur ou égal à n, on a $P_n^{(n+1)} = 0$. Par ailleurs, $\prod_{i=0}^{n}(t - x_i)$ est un polynôme de degré $n+1$ dont la dérivée $n+1$ ième est égale à $(n+1)!$. On obtient finalement

$$\phi_x^{(n+1)}(\rho_x) = f^{(n+1)}(\rho_x) - K_x(n+1)! = 0,$$

soit

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

On en déduit (2.12).

Du théorème 2.20, on déduit immédiatement le corollaire

Corollaire 2.21 *Soient $f \in C^{n+1}([a, b])$ et P_n le polynôme qui interpole f aux points $x_0, \cdots, x_n$. On a l'estimation*

$$|f(x) - P_n(x)| \leq \frac{\max_{x\in[a,b]} |f^{(n+1)}(x)|}{(n+1)!} \prod_{i=0}^{n} |x - x_i|, \quad \forall\, x \in [a, b]. \qquad (2.13)$$

Preuve
On a

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

L'inégalité (2.13) découle alors de l'inégalité précédente et de (2.12).

3. Applications : calcul d'une valeur approchée de $\ln 9.2$ connaissant une valeur approchée de $\ln 9$ et $\ln 9.5$. On donne $\ln 9 = 2.19722$ et $\ln 9.5 = 2.25129$. Une valeur approchée de $\ln 9.2$ est donnée par $P_1(9.2)$ où P_1 est le polynôme qui interpole f définie par $f(x) = \ln x$ aux points $x_0 = 9$ et $x_1 = 9.5$.
Le polynôme P_1 est donné par

$$P_1(x) = 0.10814(x - 9) + 2.19722.$$

et une valeur approchée de $\ln 9.2$ est donnée par 2.21884..
Déterminons une majoration de l'erreur $\ln 9.2 - P_1(9.2)$. D'après le corollaire 2.21, on déduit que

$$|\ln 9.2 - P_1(9.2)| \leq \frac{\max_{x\in[9;9.5]} |f''(x)|}{2!}(9.2-9)(9.5-9.2)$$

Or, $\max_{x\in[9;9.5]} |f''(x)| = \max_{x\in[9;9.5]} \frac{1}{x^2} = \frac{1}{81}$, d'où on obtient

$$|\ln 9.2 - P_1(9.2)| \leq \frac{\max_{x\in[9;9.5]} |f''(x)|}{2!}(9.2-9)(9.5-9.2) = 3.7037.10^{-4}.$$

2.3 Différences divisées

2.3.1 Polynôme de Newton

La méthode de Lagrange comporte divers inconvénients. Par exemple, si on introduit un point d'interpolation supplémentaire, il est nécessaire de recalculer tous les polynômes élémentaires de Lagrange afin de déterminer le polynôme d'interpolation de f.
L'objectif dans cette partie est de déterminer les polynômes P_n par *récurrence* sur n. Soit $n \geq 1$. Supposons que P_{n-1}, le polynôme qui interpole f aux points $x_0, x_1, \cdots, x_{n-1}$ soit déterminé. On cherche donc P_n sous la forme

$$P_n(x) = P_{n-1}(x) + g_n(x),$$

g_n polynôme à déterminer. Puisque $P_n(x_i) = P_{n-1}(x_i)$ pour $i = 0, \cdots, n-1$, on a $g_n(x_i) = 0$ pour tout $i = 0, \cdots, n-1$. Donc on a

$$g_n(x) = a_n \prod_{i=0}^{n-1}(x - x_i).$$

Le coefficient a_n se note $f[x_0, \cdots, x_n]$: c'est la n-ième différence divisée de f aux points $x_0, \cdots, x_n$. C'est le coefficient du monôme de plus haut degré de P_n. Observons que

$$a_n = \frac{P_n(x_n) - P_{n-1}(x_n)}{\prod_{i=0}^{n-1}(x_n - x_i)}.$$

L'objectif est de calculer a_n en effectuant une récurrence sur n. Par définition, on pose

$$f[x_0] := f(x_0).$$

Calculons g_n dans le cas $n = 1$. On a

$$P_1(x) = P_0(x) + g_0(x),$$

avec $P_0(x) = f(x_0)$ et

$$P_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0).$$

On en déduit que $g_0(x) = \frac{f(x_1)-f(x_0)}{x_1-x_0}(x - x_0)$, et on pose

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \tag{2.14}$$

Afin de déterminer $f[x_0, \cdots, x_n]$ pour $n \geq 2$ quelconque, établissons le lemme d'Aitken.

Lemme 2.22 *Soit P le polynôme qui interpole f aux points $x_0, \cdots, x_n$ et Q le polynôme qui interpole f aux points $x_1, \cdots, x_{n+1}$. Alors le polynôme qui interpole f aux points $x_0, \cdots, x_{n+1}$ est donné par*

$$R(x) = \frac{(x_{n+1} - x)P(x) - (x_0 - x)Q(x)}{x_{n+1} - x_0}. \tag{2.15}$$

Preuve En effet, on a

$$R(x_0) = P(x_0) = f(x_0),$$

et $R(x_{n+1}) = Q(x_{n+1}) = f(x_{n+1})$. Pour $i \neq 0, n+1$, on a

$$R(x_i) = \frac{(x_{n+1} - x_i)P(x_i) - (x_0 - x_i)Q(x_i)}{x_{n+1} - x_0} = \frac{((x_{n+1} - x_i) - (x_0 - x_i))f(x_i)}{x_{n+1} - x_0} = f(x_i).$$

On déduit du lemme 2.22 la proposition suivante :

Proposition 2.23 *On a $f[x_0] = f(x_0)$ et pour $n \geq 1$*

$$f[x_0, \cdots, x_n] = \frac{f[x_1, \cdots, x_n] - f[x_0, \cdots, x_{n-1}]}{x_n - x_0} \tag{2.16}$$

Preuve Soit $n \geq 1$. Appliquons le lemme d'Aitken en considérant les polynômes P et Q qui interpolent f respectivement aux points $x_0, \cdots, x_{n-1}$ et $x_1, \cdots, x_n$. Le coefficient du monôme de plus haut degré dans R est donné par $f[x_0, \cdots, x_n]$, celui de P par $f[x_0, x_1, \cdots, x_{n-1}]$ et celui de Q par $f[x_1, x_2, \cdots, x_n]$. D'après (2.15), on déduit que le coefficient du monôme de plus haut degré dans R est égal à

$$\frac{f[x_1, \cdots, x_n] - f[x_0, \cdots, x_{n-1}]}{x_n - x_0}.$$

Ceci achève la preuve de la proposition 2.23.

On déduit immédiatement de la proposition 2.23 le théorème :

Théorème 2.24 *Le polynome P_n qui interpole f aux points (x_i) est donné par*

$$P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + \cdots + f[x_0, \cdots, x_n] \prod_{i=0}^{n-1} (x - x_i).$$

où $f[x_0, \cdots, x_n]$ est donnée par (2.16).

À titre de comparaison avec la méthode de Lagrange, reprenons l'exemple de calcul du polynôme d'interpolation donné à la sous-section précédente. On a $x_0 = 0$, $x_1 = 1$ et $x_2 = 2$ et $f(x_0) = 1$, $f(x_1) = 1$ et $f(2) = \frac{4}{3}$. Appliquant la proposition 2.23, on obtient

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = 0$$

et

$$f[x_1, x_2] = \frac{f(x_2) - f(x_1)}{x_2 - x_1} = \frac{1}{3}$$

puis

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = \frac{1}{6}.$$

D'après le théorème 2.24, on en déduit l'expression suivante de P_2 :

$$P_2(x) = f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) = 1 + \frac{x(x-1)}{6}.$$

2.3.2 Propriétés des différences divisées

Proposition 2.25 *Soit σ une permutation de $\{0, \cdots, n\}$. Alors on a*

$$f[x_{\sigma(0)}, \cdots, x_{\sigma(n)}] = f[x_0, \cdots, x_n]. \tag{2.17}$$

Preuve
En effet, le polynôme P_n qui interpole f aux points $x_0, x_1, \cdots, x_n$ est égal au polynôme Q qui interpole f aux points $x_{\sigma(0)}, \cdots, x_{\sigma(n)}$. Or, le coefficient du monôme de plus haut degré de P_n vaut $f[x_0, \cdots, x_n]$ et celui de Q est égal à $f[x_{\sigma(0)}, \cdots, x_{\sigma(n)}]$ d'où (2.17).

Etablissons la proposition

Proposition 2.26 *Soient $p \in I\!R_n[X]$ et $(x_i)_{i \in \{0, \cdots, n1\}}$ $n+2$ points distincts de $[a, b]$ tels que $a \leq x_0 < x_1 < \cdots < x_n < x_{n+1} \leq b$. Alors $p[x_0, \cdots, x_n]$ est indépendant du choix des points d'interpolation $x_0, \cdots, x_n$.*
De plus, on a

$$p[x_0, \cdots, x_{n+1}] = 0, \quad \forall\, p \in I\!R_n[X].$$

Preuve
En effet, soit a_n le coefficient du monôme de plus haut degré de p. Alors d'après le théorème 2.24, quelque soit $(x_0, \cdots, x_n) \in I\!R^{n+1}$, $n+1$ points distincts, on a

$$p[x_0, \cdots, x_n] = a_n.$$

D'autre part, d'après la proposition 2.23 et ce qui précède, on a

$$p[x_0, \cdots, x_{n+1}] = \frac{p[x_1, \cdots, x_{n+1}] - p[x_0, \cdots, x_n]}{x_{n+1} - x_0} = \frac{a_n - a_n}{x_{n+1} - x_0} = 0,$$

ce qui achève la preuve de la proposition 2.26.

On pose

$$\prod_i (x_0, \cdots, x_n) = \prod_{j=0, j\neq i}^{n} (x_i - x_j), \quad 0 \leq i \leq n.$$

On peut montrer par récurrence la proposition suivante :

Proposition 2.27 *Soit $n \in I\!N^*$. On a l'égalité*

$$f[x_0, \cdots, x_n] = \sum_{i=0}^{n} \frac{f(x_i)}{\prod_i (x_0, \cdots, x_n)}.$$

Preuve On raisonne par récurrence sur n.
Si $n = 1$, on a $f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$ et

$$\sum_{i=0}^{1} \frac{f(x_i)}{\prod_i (x_0, x_1)} = \frac{f(x_0)}{(x_0 - x_1)} + \frac{f(x_1)}{(x_1 - x_0)} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

Le résultat est donc vrai dans ce cas. Supposons le résultat vrai à l'ordre n et montrons qu'il est vrai à l'ordre $n+1$.
On a

$$f[x_0, \cdots, x_n, x_{n+1}] = \frac{f[x_1, \cdots, x_n, x_{n+1}] - f[x_0, \cdots, x_n]}{x_{n+1} - x_0}.$$

Par hypothèse de récurrence, on a

$$f[x_0, \cdots, x_n, x_{n+1}] = \frac{\displaystyle\sum_{i=1}^{n+1} \frac{f(x_i)}{\prod_i (x_1, \cdots, x_{n+1})} - \sum_{i=0}^{n} \frac{f(x_i)}{\prod_i (x_0, \cdots, x_n)}}{x_{n+1} - x_0}.$$

Par définition, on a

$$(x_i - x_0) \prod_i (x_1, \cdots, x_{n+1}) = \prod_i (x_0, \cdots, x_{n+1}), \quad \forall\, i \neq 0$$

et

$$(x_i - x_{n+1})\prod_i(x_0,\cdots,x_n) = \prod_i(x_0,\cdots,x_{n+1}),\ \ \forall\, i \neq n+1.$$

Or,

$$\frac{\sum_{i=1}^{n+1}\frac{f(x_i)}{\prod_i(x_1,\cdots,x_{n+1})} - \sum_{i=0}^{n}\frac{f(x_i)}{\prod_i(x_0,\cdots,x_n)}}{x_{n+1}-x_0}$$
$$= \frac{f(x_{n+1})}{\prod_{n+1}(x_1,\cdots,x_{n+1})} + \sum_{i=1}^{n}\left(\frac{f(x_i)}{\prod_i(x_1,\cdots,x_{n+1})} - \frac{f(x_i)}{\prod_i(x_0,\cdots,x_n)}\right) - \frac{f(x_0)}{\prod_0(x_1,\cdots,x_{n+1})}.$$

De plus,

$$\sum_{i=1}^{n}\left(\frac{f(x_i)}{\prod_i(x_1,\cdots,x_{n+1})}) - \frac{f(x_i)}{\prod_i(x_0,\cdots,x_n)}\right) = \sum_{i=1}^{n}\frac{f(x_i)(x_i - x_0 - x_i + x_{n+1})}{\prod_i(x_0,\cdots,x_{n+1})}$$
$$= \sum_{i=1}^{n}\frac{f(x_i)(x_{n+1}-x_0)}{\prod_i(x_0,\cdots,x_{n+1})}.$$

On en déduit que le résultat est vrai au rang $n+1$ ce qui achève la preuve de la proposition 2.27.

On pose $w(x_0,\cdots,x_{n+1}) = \sum_{i=0}^{n+1}\frac{1}{|\prod_i(x_0,\cdots,x_{n+1})|}$. De la proposition 2.27, on peut déduire la proposition

Proposition 2.28 *Soit $f \in C^0([a,b])$. Quelque soit $p \in I\!R_n[X]$, on a*

$$\|f-p\|_\infty \geq \frac{|f[x_0,\cdots,x_{n+1}]|}{w(x_0,\cdots,x_{n+1})},$$

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

Preuve D'après les propositions 2.26 et 2.27, on déduit l'égalité

$$f[x_0,x_1,\cdots,x_n,x_{n+1}]$$
$$= f[x_0,x_1,\cdots,x_n,x_{n+1}] - p[x_0,x_1,\cdots,x_n,x_{n+1}] = \sum_{i=0}^{n+1}\frac{f(x_i)-p(x_i)}{\prod_i(x_0,\cdots,x_{n+1})}.$$

Par inégalité triangulaire, on en déduit aussitôt l'inégalité

$$|f[x_0,x_1,\cdots,x_n,x_{n+1}]| \leq \sum_{i=0}^{n+1}|\frac{f(x_i)-p(x_i)}{\prod_i(x_0,\cdots,x_{n+1})}| \leq \|f-p\|_\infty\sum_{i=0}^{n+1}|\frac{1}{\prod_i(x_0,\cdots,x_{n+1})}|,$$

ce qui achève la preuve de la proposition 2.28.

2.3.3 Détermination de l'erreur

On pose $e_n(x) = f(x) - P_n(x)$.

Proposition 2.29 *On a pour tout $x \in [a, b]$,*

$$e_n(x) = f[x_0, x_1, \cdots, x_n, x] \prod_{i=0}^{n} (x - x_i).$$

Preuve
Soit $\bar{x}$ différent de $x_0, x_1, \cdots, x_n$. Soit P_n le polynôme qui interpole f aux points $x_0, x_1, \cdots, x_n$. Le polynôme $\bar{P}$ qui interpole f aux points $x_0, x_1, \cdots, x_n, \bar{x}$ est donné par

$$\bar{P}(x) = P_n(x) + f[x_0, x_1, \cdots, x_n, \bar{x}] \prod_{i=0}^{n} (x - x_i).$$

Au point $x = \bar{x}$, on a

$$\bar{P}(\bar{x}) = f(\bar{x}) = P_n(\bar{x}) + f[x_0, x_1, \cdots, x_n, \bar{x}] \prod_{i=0}^{n} (\bar{x} - x_i).$$

On en déduit que

$$e_n(\bar{x}) := f(\bar{x}) - P_n(\bar{x}) = f[x_0, x_1, \cdots, x_n, \bar{x}] \prod_{i=0}^{n} (\bar{x} - x_i),$$

ce qui achève la preuve de la proposition 2.29.

Du théorème 2.20 et de la proposition 2.29, on déduit immédiatement le théorème

Théorème 2.30 *Soient $f \in C^{n+1}([a, b])$ et $a \leq x_0 < x_1 < \cdots < x_n \leq b$. Pour tout $x \in [a, b]$, il existe $\rho_x \in]a, b[$ tel que*

$$f[x_0, x_1, \cdots, x_n, x] = \frac{f^{(n+1)}(\rho_x)}{(n+1)!}. \tag{2.18}$$

Preuve On a établi deux expressions de l'erreur $|f(x) - P_n(x)|$ (voir théorème 2.20 et proposition 2.29). En comparant ces deux expressions, on déduit (2.18).

2.4 Interpolation de Hermite

2.4.1 Existence et unicité du polynôme de Hermite

Soient f une fonction dérivable et $x_i \in [a,b]$, $i \in \{0,\cdots,n\}$ $n+1$ points distincts. Pour $i \in \{0,\cdots,n\}$, on pose $y_i = f(x_i)$ et $z_i = f'(x_i)$ et on cherche H_n un polynôme de degré minimal défini par les relations

$$\begin{cases} H_n(x_i) = f(x_i), & i \in \{0,\cdots,n\} \\ H_n'(x_i) = f'(x_i) & i \in \{0,\cdots,n\}. \end{cases} \tag{2.19}$$

Théorème 2.31 *Il existe un unique polynôme de degré inférieur ou égal à* $2n+1$ *satisfaisant (2.19). Il est donné par*

$$H_n(x) = \sum_{i=0}^{n} A_i(x) y_i + \sum_{i=0}^{n} B_i(x) z_i, \tag{2.20}$$

où

$$A_i(x) = l_i^2(x)\,(1 - 2l_i'(x_i)(x - x_i)) \quad \text{et} \quad B_i(x) = l_i^2(x)(x - x_i).$$

Preuve
Existence de H_n. On va chercher H_n sous la forme

$$H_n(x) = \sum_{i=0}^{n} A_i(x) y_i + \sum_{i=0}^{n} B_i(x) z_i,$$

A_i et B_i à déterminer. Déterminons les A_i. Si on pose

$$\begin{cases} A_i(x_i) = 1, \;\; A_i(x_j) = 0, \;\; j \neq i, \\ A_i'(x_j) = 0 \;\; \forall\, j \in \{0,\cdots,n\}, \end{cases}$$

et

$$\begin{cases} B_i(x_j) = 0, \;\; \forall\, j \in \{0,\cdots,n\}, \\ B_i'(x_j) = 0 \;\; i \neq j, \quad B_i'(x_i) = 1, \end{cases}$$

alors on a

$$H_n(x_i) = y_i, \quad \text{et} \quad H_n'(x_i) = z_i \;\; \forall\, i \in \{0,\cdots,n\}.$$

Par conséquent, la fonction A_i admet $n-1$ racines doubles $(x_j)_{j\neq i}$ et satisfait les deux conditions $A_i(x_i) = 1$ et $A_i'(x_i) = 0$. Par conséquent, on cherche A_i sous la forme $A_i(x) = \prod_{j\neq i}(x - x_j)^2(ax + b)$, le terme $ax + b$ étant à déterminer de telle sorte que

$$A_i(x_i) = 1 \quad \text{et} \quad A_i'(x_i) = 0. \tag{2.21}$$

Remarquons que A_i est de degré $2n+1$ et que l'on peut exprimer cette fonction à l'aide des polynômes élémentaires de Lagrange. Finalement, on cherchera A_i sous la forme

$$A_i(x) := l_i^2(x)(ax+b).$$

Les deux conditions (2.21) sont satisfaites si et seulement si a et b satisfont le système linéaire

$$\begin{cases} ax_i + b = 1, \\ 2l_i'(x_i)(ax_i+b) + a = 0. \end{cases}$$

On obtient après résolution du système $a = -2l_i'(x_i)$ et $b = 1 + 2l_i'(x_i)x_i$, d'où

$$A_i(x) = l_i^2(x)(1 - 2l_i'(x_i)(x - x_i)).$$

On procède de même avec B_i.
On obtient

$$\begin{cases} B_i(x_j) = 0, \quad \forall\, j, \\ B_i'(x_j) = 0 \quad i \neq j, \quad B_i'(x_i) = 1. \end{cases}$$

B_i admet x_j $j \neq i$ comme racines doubles et x_i comme racine simple. On en déduit que

$$B_i(x) = C. \prod_{j=0, j\neq i}^{n} (x - x_j)^2 (x - x_i),$$

que l'on peut aussi écrire sous la forme

$$B_i(x) = \tilde{C}.l_i(x)^2(x - x_i).$$

On a $B_i'(x_i) = \tilde{C}.1 = 1$. Le polynôme B_i est également de degré $2n+1$, et compte tenu de (2.20), on en déduit que H_n est de degré inférieur ou égal à $2n+1$. On a donc établi l'existence de H_n.

Unicité de H_n. On suppose qu'il existe deux polynômes H_n et G_n de degré inférieur ou égal à $2n+1$ satisfaisant (2.19). Alors, la différence $H_n - G_n$ admet $n+1$ racines doubles donc si $H_n - G_n$ est non nul, son degré est de $2n+2$. Contradiction. Donc $H_n = G_n$. La preuve du théorème 2.31 est achevée.

2.4.2 Estimation de l'erreur

Pour $x \in [a, b]$, on pose $E(x) := f(x) - H_n(x)$.

Théorème 2.32 *Soit f une fonction définie sur $[a,b]$ de classe C^{2n+2}. Pour tout $x \in [a,b]$, il existe $\rho_x \in [a,b]$ tel que*

$$E(x) = \frac{f^{(2n+2)}(\rho_x)}{(2n+2)!} \prod_{i=0}^{n} (x - x_i)^2.$$

Preuve
On introduit pour $x \in [a,b]$ fixé, la fonction

$$\phi_x(y) = f(y) - H_n(y) - (f(x) - H_n(x)) \prod_{i=0}^{n} \frac{(y - x_i)^2}{(x - x_i)^2}.$$

Remarquons que la fonction ϕ_x admet $n+1$ racines doubles $x_0, x_1, \cdots, x_n$ et une racine simple, x. On applique alors le lemme 2.19. On déduit qu'il existe $\rho_x \in]a,b[$ tel que

$$\phi_x^{(2n+2)}(\rho_x) = 0.$$

Or, comme $H_n^{(2n+2)}(x) = 0$ pour tout x et la dérivée $2n+2$ ième de $y \mapsto (y - x_i)^2$ vaut $(2n+2)!$, on obtient

$$\phi_x^{(2n+2)}(y) = f^{(2n+2)}(y) - (f(x) - H_n(x)) \frac{(2n+2)!}{\prod_{i=0}^{n} (x - x_i)^2}$$

On en déduit immédiatement le résultat cherché.

2.5 Minimisation de l'erreur

Dans la suite, on note par E_n l'ensemble des polynômes unitaires de degré n. On suppose que $f \in C^{n+1}([a,b])$. D'après le théorème 2.20, l'erreur dépend de deux termes :

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

et

$$\max_{x \in [a,b]} |\prod_{i=0}^{n} (x - x_i)|.$$

La question est de déterminer comment choisir les (x_i) de telle sorte que $\max_{x \in [a,b]} |\prod_{i=0}^{n} (x - x_i)|$ soit minimal ?
Nous allons montrer dans cette sous-section qu'il existe un polynôme unitaire q scindé de degré $n+1$ tel que :

$$\max_{x \in [a,b]} |q(x)| \leq \max_{x \in [a,b]} |v(x)| \quad \forall\, v \in E_{n+1}$$

2.5.1 Polynôme de Tchebychev

Pour $x \in [-1,1]$ et $n \in \mathbb{N}$, on pose

$$T_n(x) = \cos(n\text{arcos}(x)).$$

Proposition 2.33 *La fonction T_n est un polynôme de degré n. De plus, pour $n \geq 1$, le coefficient du monÃ´me de plus haut degré de T_n est égal à 2^{n-1}.*
Pour tout $x \in [-1,1]$ et $n \in \mathbb{N}^*$, on a la relation

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \tag{2.22}$$

Preuve
On a pour tout $\theta \in \mathbb{R}$

$$\cos((n+1)\theta) = \cos n\theta \cos\theta - \sin n\theta \sin\theta$$

et

$$\cos((n-1)\theta) = \cos n\theta \cos\theta + \sin n\theta \sin\theta$$

donc

$$\cos((n+1)\theta) + \cos((n-1)\theta) = 2\cos n\theta \cos\theta$$

et en faisant le choix $\theta = \text{arcos}x$, on obtient (2.22).
Remarquons que $T_0(x) = 1$ et $T_1(x) = x$. En effectuant un raisonnement par récurrence, et en utilisant (2.22), on déduit le résultat demandé. En effet, le résultat est vrai pour $n = 0$ et $n = 1$. Supposons le résultat vrai pour $k \in \{0, \cdots, n\}$. Alors d'après (2.22), T_{n+1} est un polynôme de degré $n+1$ et le coefficient du monôme de plus haut degré est égal à $2.2^{n-1} = 2^n$.

Proposition 2.34 *Pour $n \geq 1$, le polynôme T_n admet n racines simples*

$$x_k = \cos(\frac{2k-1}{2n}\pi), \quad k = 1, \cdots, n.$$

De plus, T_n atteint ses extremums dans l'intervalle $]-1,1[$ aux $n-1$ points $x'_k = \cos(\frac{k}{n}\pi)$ $k = 1, \cdots, n-1$. En ces points, on a $T_n(x'_k) = (-1)^k$. De plus, en $x'_0 := -1$ et $x'_n = 1$, on a

$$T_n(x'_0) = (-1)^n \qquad et \quad T_n(x'_n) = 1.$$

Preuve On a $T_n(x) = 0$ si et seulement si $\cos(n\text{arcos}(x)) = 0$, soit $n\text{arcos}(x) = \frac{\pi}{2} + k\pi$ ou encore

$$x = \cos(\frac{\pi}{2n} + \frac{k\pi}{n}), \quad k \in Z.$$

T_n est degré n, il admet au plus n racines notées x_k. On en déduit que

$$x_k = \cos(\frac{(2k-1)\pi}{2n}), \quad k = 1, \cdots, n.$$

On a pour tout $x \in]-1, 1[$

$$T_n'(x) = \frac{n}{\sqrt{1-x^2}} \sin(n\text{arcos}(x)).$$

Donc $T_n'(x) = 0$ si et seulement si

$$n\text{arcos}(x) = k\pi, \;\; k \in Z.$$

La racines de T_n' sont données par

$$x_k' = \cos(\frac{k\pi}{n}), \quad k = 1, \cdots, n-1.$$

Comme la fonction T_n' change de signe au voisinage de x_k', on en déduit que $x_k' \quad k = 1, \cdots, n-1$ sont des extremums de T_n. En ces points, on a

$$T_n(x_k') = \cos(k\pi) = (-1)^k.$$

De plus, $T_n(-1) = \cos(n\pi) = (-1)^n$ et $T_n(1) = \cos(0) = 1$.

2.5.2 Minimisation de $\max_{x\in[a,b]} \prod_{i=0}^{n} |x - x_i|$

L'objectif est de minimiser $\max_{x\in[a,b]} \prod_{i=0}^{n} |x - x_i|$ en choisissant les x_i au mieux dans $[a, b]$.

Dans la suite, on pose $\bar{T}_n = \dfrac{T_n}{2^{n-1}}$.

Théorème 2.35 *Soit $p \in E_n$.*
On a l'inégalité

$$\frac{1}{2^{n-1}} = \max_{-1\le x\le 1} |\bar{T}_n(x)| \le \max_{-1\le x\le 1} |p(x)|.$$

Preuve
On suppose qu'il existe $P \in E_n$ tel que

$$\max_{-1\le x\le 1} |P(x)| < \max_{-1\le x\le 1} |\bar{T}_n(x)| = \frac{1}{2^{n-1}}. \tag{2.23}$$

Posons $r = \bar{T}_n - P$. Le degré de r est inférieur ou égal à $n-1$ et $r \neq 0$. D'autre part, d'après la proposition 2.34, on a

$$r(x_k') = \bar{T}_n(x_k') - P(x_k') = \frac{(-1)^k}{2^{n-1}} - P(x_k'), \;\; k = 0, \cdots, n.$$

Comme $\max_{-1\leq x\leq 1}|P(x)| < \dfrac{1}{2^{n-1}}$, le signe de $r(x'_k)$ dépend du signe de $\dfrac{(-1)^k}{2^{n-1}}$, il est donc positif si k est pair et négatif si k est impair. Comme r est continue, on déduit du théorème des valeurs intermédiaires appliquée entre x'_k et x'_{k+1} $(k = 0, \cdots, n)$ que r admet au moins n zéros. Or, comme le degré de r est inférieur ou égal à $n-1$, on a $r = 0$. Donc $\bar{T}_n - P = 0$ ce qui contredit (2.23).

On déduit de cette étude que pour minimiser $\max_{x\in[a,b]}|\prod_{i=0}^{n}(x - x_i)|$, il faut choisir pour x_i les racines de T_{n+1}. Ainsi, on a établi le théorème

Théorème 2.36 *On suppose que $a = -1$, $b = 1$ et que les points d'interpolation x_i sont les racines de T_{n+1}. Alors, pour tout $x \in [-1, 1]$, on a l'estimation suivante :*

$$|f(x) - P_n(x)| \leq \frac{1}{2^n}\max_{x\in[-1,1]}|\frac{f^{n+1}(x)}{(n+1)!}|.$$

De plus, ce choix des points d'interpolation est le meilleur possible au sens où pour tout $(y_i)_{(i\in\{0,\cdots,n\})}$ $y_i \in [-1, 1]$ (y_i distincts deux-à-deux), on a

$$\frac{1}{2^n} \leq \max_{y\in[-1,1]}|\prod_{i=0}^{n}(y - y_i)|.$$

Il faut à présent établir un résultat analogue pour un intervalle quelconque $[a, b]$. Soit ϕ la bijection affine définie sur $[-1, 1]$ dont l'image est $[a, b]$ avec $\phi(-1) = a$ et $\phi(1) = b$. On a

$$\phi(x) = \frac{b-a}{2}x + \frac{b+a}{2}. \tag{2.24}$$

On pose

$$u_i := \phi(x_i) \tag{2.25}$$

pour $i = 0, \cdots, n$ (x_i définis dans la proposition 2.34) et pour $x \in [-1, 1]$, $u = \phi(x)$. Le théorème suivant généralise le théorème 2.36 au cas d'un intervalle quelconque :

Théorème 2.37 *Soient $x \in [a, b]$ et P_n le polynôme qui interpole f aux points (u_i), pour $i = 0, \cdots, n$. On a alors pour $x \in [a, b]$*

$$|f(x) - P_n(x)| \leq \frac{(b-a)^{n+1}}{(n+1)!2^{2n+1}}\max_{x\in[a,b]}|f^{(n+1)}(x)|.$$

Ce choix est le meilleur possible au sens pour tout $(y_i)_{(i\in\{0,\cdots,n\})}$ $y_i \in [a, b]$, on a

$$\frac{(b-a)^{n+1}}{2^{2n+1}} \leq \max_{y\in[a,b]}|\prod_{i=0}^{n}(y - y_i)|.$$

Afin de prouver le théorème 2.37, établissons le théorème

Théorème 2.38 *Soit p un polynôme unitaire de degré n, scindé sur $[a,b]$. Soit $\tilde{P}$ défini par $\tilde{P}(u) = \prod_{i=0}^{n}(u-u_i)$, $u \in [a,b]$ et (u_i) définies dans (2.25). On a l'inégalité*

$$\max_{a\leq x\leq b} |\tilde{P}(x)| = \frac{(b-a)^{n+1}}{2^{2n+1}} \leq \max_{a\leq x\leq b} |p(x)|.$$

Preuve Soient (z_i) les racines simples de p dans $[a,b]$ et (y_i) définie par $y_i = \phi^{-1}(z_i)$, $i = 1, \cdots, n$. Pour tout $u \in [a,b]$, il existe un unique $x \in [-1,1]$ tel que

$$|\tilde{P}(u)| = |\prod_{i=0}^{n}(u-u_i)| = \prod_{i=0}^{n}|\phi(x)-\phi(x_i)| = \frac{(b-a)^{n+1}}{2^{n+1}}|\prod_{i=0}^{n}(x-x_i)|.$$

On a alors

$$\max_{u\in[a,b]} |\tilde{P}(u)| = \frac{(b-a)^{n+1}}{2^{n+1}} \max_{x\in[-1,1]} |\prod_{i=0}^{n}(x-x_i)| = \frac{(b-a)^{n+1}}{2^{n+1}}.\frac{1}{2^n} = \frac{(b-a)^{n+1}}{2^{2n+1}}, \tag{2.26}$$

et

$$\max_{u\in[a,b]} |p(u)| = \frac{(b-a)^{n+1}}{2^{n+1}} \max_{x\in[-1,1]} |\prod_{i=0}^{n}(x-y_i)|.$$

On déduit alors du théorème 2.35 l'inégalité

$$\max_{u\in[a,b]} |\tilde{P}(u)| = \frac{(b-a)^{n+1}}{2^{2n+1}} \leq \max_{u\in[a,b]} |p(u)|. \tag{2.27}$$

Preuve du théorème 2.37
La preuve du théorème 2.37 découle immédiatement du théorème 2.38 (voir (2.26) et (2.27)).

Remarque 2.39 *Considérons la fonction $f(x) = \dfrac{1}{1+x^6}$ sur $[-4;4]$. Posons $x_i = -4 + ih$, avec $h = \frac{8}{N}$, $i = 0, \cdots, N$. Le polynôme qui interpole la fonction f aux points équidistants (x_i) approche f de manière très mauvaise au voisinage de -4 et 4. Ce phénomène est appelé le phénomène de Runge. Un moyen d'y remédier est d'utiliser pour points d'interpolation les images par $\phi(t) := 4t$ des zéros du polynôme de Tchebychev T_{N+1}.*

2.6 Introduction à l'approximation uniforme par des polynômes

On note par $\mathbb{R}_n[X]$ l'anneau des polynômes de degré inférieur ou égal à n. Soit $f : [a, b] \to \mathbb{R}$ une fonction continue sur $[a, b]$. On pose

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

Dans cette section, on étudie le problème de l'approximation uniforme de f. On cherche un polynôme $P_0 \in \mathbb{R}_n[X]$ tel que

$$\|f - P_0\|_\infty = \inf_{p\in\mathbb{R}_n[X]} \|f - p\|_\infty. \tag{2.28}$$

On va montrer qu'un tel polynôme existe et qu'il est unique.

2.6.1 Existence et unicité du polynôme de meilleure approximation

Le théorème suivant donne la réponse à la question posée précédemment.

Théorème 2.40 *Etant donné $f \in C^0([a, b])$, il existe un unique polynôme P_0 satisfaisant (2.28).*

Preuve
Existence Observons que si $f \in \mathbb{R}_n[X]$, on a $P_0 = f$. Supposons que $f \notin \mathbb{R}_n[X]$. Considérons l'application ϕ définie sur $\mathbb{R}_n[X]$ à valeurs dans $\mathbb{R}^+$ définie par $\psi(p) = \|p - f\|_\infty$. La fonction ϕ est continue sur $\mathbb{R}_n[X]$. En effet, soit $p_0 \in \mathbb{R}_n[X]$. On a par inégalité triangulaire

$$|\phi(p) - \phi(p_0)|_\infty \leq \|p - p_0\|_\infty.$$

Donc pour tout $\epsilon > 0$, il existe $\eta = \epsilon$, tel que $\|p - p_0\|_\infty < \eta$ implique $|\phi(p) - \phi(p_0)|_\infty < \epsilon$.
D'autre part, $\phi(p)$ tend vers $+\infty$ quand $\|p\|_\infty$ tend vers $+\infty$ puisque, par inégalité triangulaire

$$\phi(p) \geq \|p\|_\infty - \|f\|_\infty.$$

(On dit alors que ϕ est coercive.)
Donc, il existe $R > 0$ tel que

$$\phi(p) \geq \phi(p_0) \quad \forall\, p \text{ tel que } \|p\| \geq R.$$

On a donc

$$\inf_{p\in\mathbb{R}_n[X]} \|f - p\|_\infty = \inf_{p\in B(0,R)} \|f - p\|_\infty,$$

où $B(0,R)$ représente la boule fermée de centre O et de rayon R incluse dans $(\mathbb{R}_n[X], \|.\|_\infty)$.
Or, la boule $B(0,R)$ est un fermé borné de $\mathbb{R}_n[X]$, donc c'est un sous-ensemble compact de $\mathbb{R}_n[X]$. De plus, il s'agit d'un sous-ensemble connexe de $\mathbb{R}_n[X]$. En effet, la boule est connexe par arcs (on peut relier deux points quelconques de la boule par un segment), donc connexe.
L'image d'un sous-ensemble connexe et compact de $\mathbb{R}_n[X]$ par ϕ, application continue, est un sous-ensemble compact et connexe de $\mathbb{R}^+$. Les compacts connexes de $\mathbb{R}$ sont exactement les intervalles fermés bornés. On a donc

$$\phi(B(0,R)) = [\alpha, \beta] \subset \mathbb{R}^+.$$

Il existe $P_0 \in B(0,R)$ tel que $\phi(P_0) = \alpha$ et donc

$$\inf_{p \in \mathbb{R}_n[X]} \|f - p\|_\infty = \inf_{p \in B(0,R)} \|f - p\|_\infty = \phi(P_0).$$

Unicité
Etape 1 Montrons dans un premier temps qu'il existe $n+2$ points (x_i) en lesquels

$$|f(x_i) - P_0(x_i)| = \|f - P_0\|_\infty.$$

Supposons que ce ne soit pas le cas, que l'égalité soit satisfaite en seulement k points $1 \leq k < n+2$. Soit q l'unique polynôme qui interpole f aux points x_i $i = 1, \cdots k$. q est donc de degré inférieur ou égal à n. Par continuité de $x \mapsto f(x) - q(x)$ aux points (x_i), on déduit qu'il existe V_ϵ un ouvert tel que $x_i \in V_\epsilon$ pour tout i et

$$|f(x) - q(x)| \leq \epsilon, \quad \forall\, x \in V_\epsilon. \tag{2.29}$$

Pour $t \in]0,1[$, on pose $P_t = (1-t)P_0 + tq$. On a

$$P_t - f = (1-t)P_0 + tq - (1-t)f - tf = (1-t)(P_0 - f) + t(q - f).$$

Pour $x \in V_\epsilon$, d'après (2.29), on a

$$|(f - P_t)(x)| \leq (1-t)\|f - P_0\|_\infty + t\epsilon.$$

Pour $x \notin V_\epsilon$, on a

$$|(f - P_t)(x)| \leq \sup_{y \notin V_\epsilon} |(f - P_0)(y)| + tA,$$

oÃ¹A=$\|f - q\|$. Observons que

$$\sup_{y \notin V_\epsilon} |(f - P_0)(y)| < \|f - P_0\|_\infty.$$

L'inégalité précédente est stricte car $x_i \notin V_\epsilon$ et par définition des (x_i), $y \mapsto |f(x) - P_0(x)|$ atteint son maximum en ces points. Choisissons t assez petit de telle sorte que

$$\sup_{y \notin V_\epsilon} |(f - P_0)(y)| + tA < \|f - P_0\|_\infty. \tag{2.30}$$

D'autre part, posant $\epsilon = \frac{\|f-P_0\|_\infty}{2}$, on obtient pour tout $t \in [0,1]$

$$\sup_{x \in V_\epsilon} |(f - P_t)(x)| \leq (1 - t/2)\|f - P_0\|_\infty < \|f - P_0\|_\infty. \tag{2.31}$$

En conslusion, compte-tenu de (2.30) et (2.31), on a

$$\|f - P_t\|_\infty < \|f - P_0\|_\infty.$$

On obtient une contradiction puique P_0 est le polynôme de meilleur approximation (P_t est un polynôme de meilleure approximation de f que P_0 !).

Etape 2

Considérons à présent le polynôme $P = \frac{P_1+P_2}{2}$ où P_1 et P_2 sont deux polynôme de meilleure approximation. P est également un polynôme de meilleure approximation. En effet, par inégalité triangulaire

$$\|\frac{P_1 + P_2}{2} - f\|_\infty \leq \frac{1}{2}(\|f - P_1\|_\infty + \|f - P_2\|_\infty) = \|f - P_1\|_\infty.$$

Donc $\|f - P_1\| = \|f - P_2\| = \|\frac{P_1+P_2}{2} - f\|$. Soient les $n+2$ points (x_i) définis par

$$\|f - \frac{P_1 + P_2}{2}\| = |f(x_i) - \frac{P_1 + P_2}{2}(x_i)|.$$

De tels points existent d'après l'étape 1. On a

$$\|\frac{P_1 + P_2}{2} - f\|_\infty = |(\frac{P_1 + P_2}{2} - f)(x_i)| \leq \frac{1}{2}|(P_1 - f)(x_i)| + \frac{1}{2}|(P_2 - f)(x_i)| \leq \|f - P_1\|_\infty.$$

Comme $\|\frac{P_1+P_2}{2} - f\|_\infty = \|f - P_1\|_\infty$, les inégalités ci-dessus sont en réalités des égalités. Donc

$$\frac{1}{2}|(P_1 - f)(x_i)| + \frac{1}{2}|(P_2 - f)(x_i)| = \|f - P_1\|_\infty = \|f - P_2\|_\infty.$$

Il en résulte que l'on doit avoir $\frac{1}{2}|(P_2 - f)(x_i)| \geq \frac{1}{2}\|f - P_1\|_\infty = \frac{1}{2}\|f - P_2\|_\infty$. Donc on a

$$|(P_1 - f)(x_i)| = |(P_2 - f)(x_i)| = \|f - P_1\|_\infty.$$

On obtient finalement

$$|(P_1 - f)(x_i)| = |(P_2 - f)(x_i)|.$$

Or $(P_1 - f)(x_i)$ et $(P_2 - f)(x_i)$ possèdent le même signe, puisque si

$$(P_1 - f)(x_i) = -(P_2 - f)(x_i),$$

alors $(\frac{P_1+P_2}{2} - f)(x_i) = 0$ ce qui implique que $f \in \mathbb{R}[X]$. On en déduit que $P_1(x_i) = P_2(x_i)$ pour tout i. Donc le polynôme $P_1 - P_2$ admet $n+2$ racines, il est nul et $P_1 = P_2$.

Remarque 2.41 *Le schéma de la preuve suivi pour établir l'existence dans le théorème 2.40 s'adapte sans difficulté au cas où on remplace $\mathbb{R}_n[X]$ par un espace vectoriel normé de dimension fini.*
Soit une fonction ϕ définie sur un espace vectoriel normé E de dimension finie à valeurs réelles, coercive (c'est-à-dire que $\phi(x)$ tend vers $+\infty$ quand $\|x\|_E$ tend vers $+\infty$) et continue sur E. Alors il existe un élément $u \in E$ (non nécessairement unique) tel que

$$\phi(u) = \inf_{v \in E} \phi(v).$$

Remarque 2.42 *On a montré (voir théorème 2.35) que le polynôme de Tchebychev de degré n est la meilleure approximation de 0 par des polynômes unitaires de degré n sur l'intervalle $[-1, 1]$.*

Une caractérisation du polynôme de meilleure approximation est donnée dans théorème suivant, dû à Tchebychev, que l'on admettra :

Théorème 2.43 *Soit $f \in C^0([a, b])$. Le polynôme $p \in \mathbb{R}_n[X]$ est la meilleure approximation uniforme de f sur $[a, b]$ si et seulement si il existe $n+2$ points $a \leq x_0 < x_1 < \cdots < x_{n+1} \leq b$ tel que*

$$(-1)^i (f(x_i) - p(x_i)) = \epsilon_0 \|f - p\|_\infty, \quad i = 0, \cdots, n+1, \tag{2.32}$$

où $\epsilon_0 = sng(f(x_0) - p(x_0))$.

Remarque 2.44 *Nous avons établi au cours de la preuve du théorème 2.40 un résultat plus faible que la condition (2.32) : si P_0 est le polynôme de meilleure approximation de f, il existe $n+2$ points (x_i) en lesquels*

$$|f(x_i) - P_0(x_i)| = \|f - P_0\|_\infty.$$

Exemple Meilleure approximation de x^{n+1} sur $[-1,1]$. On cherche $p_n \in \mathbb{R}_n[X]$ tel que

$$\|x^{n+1} - p_n\|_\infty = \inf_{p \in \mathbb{R}_n[X]} \|x^{n+1} - p\|_\infty.$$

Soit T_{n+1} le $n+1$ ième polynôme de Tchebychev. Le coefficient du monôme de plus haut degré de T_{n+1} est 2^n. On pose

$$p_n := x^{n+1} - 2^{-n} T_{n+1}(x).$$

D'après la proposition 2.33, p_n est un polynôme de degré n. On va montrer que d'après le théorème de Tchebychev et la proposition 2.34, p_n est le polynôme de meilleure approximation de x^{n+1} et

$$\text{dist}(x^{n+1}, \mathbb{R}_n[X]) = \frac{1}{2^n}.$$

D'après la proposition 2.34, on a

$$\max_{[-1,1]} |x^{n+1} - p_n(x)| = \frac{1}{2^n} \max_{x \in [-1,1]} |T_{n+1}(x)| = \frac{1}{2^n}.$$

De plus,

$$2^{-n} T_{n+1}(x'_k) = \frac{(-1)^k}{2^n}, \quad k = 0, \cdots, n+1$$

donc $x \mapsto x^{n+1} - p_n(x)$ atteint $\max_{[-1,1]} |x^{n+1} - p_n(x)|$ en changeant de signe $n+2$ fois : d'après le théorème de Tchebychev, p_n est bien le polynôme recherché.

Remarque 2.45 *En général, la détermination du polynôme de meilleure approximation conduit à résoudre un système d'équations non linéaires, comportant de nombreuses inconnues. Un algorithme de résolution de ce système est l'algorithme de Remez, non abordée dans le cadre de ce cours. En pratique, on préférera utiliser le polynôme d'interpolation de Lagrange plutôt que le polynôme de meilleure approximation. La résolution numérique du système non linéaire est beaucoup trop coûteuse pour être "rentable".*

2.6.2 Polynôme d'interpolation de Lagrange et polynôme de meilleure approximation

On a le théorème suivant.

Théorème 2.46 *Soit $f \in C^0([a,b])$ (x_i) $n+1$ points distincts de $[a,b]$. Soit P_n le polynôme de Lagrange interpolant f aux points (x_i) construit au théorème 2.14. Alors on a*

$$dist(f, \mathbb{R}_n[X]) \leq \|f - P_n\|_\infty \leq (1 + \|\Lambda_n\|_\infty) dist(f, \mathbb{R}_n[X]),$$

où

$$\Lambda_n(x) = \sum_{i=0}^{n} |l_i(x)|,$$

l_i ième polynôme élémentaire de Lagrange.

Preuve Notons par P le polynôme de meilleure approximation défini dans le théorème 2.40. On a par inégalité triangulaire

$$\|f - P_n\|_\infty \leq \|f - P\|_\infty + \|P - P_n\|_\infty. \tag{2.33}$$

Le polynôme qui interpole P aux points (x_i) est lui-même ($P(x) = \sum_{i=0}^{n} P(x_i) l_i(x)$), donc pour $x \in [a, b]$, on a

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

Par inégalité triangulaire, on en déduit que

$$|P(x) - P_n(x)| \leq \sum_{i=0}^{n} |P(x_i) - f(x_i)||l_i(x)| \leq \|f - P\|_\infty \Lambda_n(x).$$

Il en résulte que

$$\|P - P_n\|_\infty \leq \|f - P\|_\infty \|\Lambda_n\|_\infty.$$

De (2.33) et de l'inégalité précédente, on déduit le résultat recherché.

2.7 Compléments sur l'interpolation

2.7.1 Fonctions splines

On considère dans la suite $f : [a, b] \to \mathbb{R}$ et $x_0 = a < x_1 < \cdots < x_n = b$, $n + 1$ points distincts de $[a, b]$.

Définition 2.47 *Soit $n \in \mathbb{N}^*$. On appelle fonction spline, une fonction s de classe C^2 sur $[a, b]$ interpolant f aux points $(x_i)_{0 \leq i \leq n}$, telle que la restriction de s à l'intervalle $[x_i, x_{i+1}]$ $i = 0, \cdots, n-1$, notée s_i, est un polynôme de degré inférieur ou égal à 3.*

Comme s interpole f aux points (x_i), on a les égalités :

$$s_i(x_i) = f_i := f(x_i),\ i = 0, \cdots, n-1 \quad \text{et} \quad s_{i-1}(x_i) = f_i,\ \ i = 1, \cdots, n \tag{2.34}$$

Comme la fonction s est de classe $C^2([a,b])$, on a les relations suivantes qui expriment la continuité de s et de ses dérivées premières et secondes aux points (x_i) : pour $i = 1, \cdots, n-1$, on a

$$s'_{i-1}(x_i) = s'_i(x_i), \quad s''_{i-1}(x_i) = s''_i(x_i). \tag{2.35}$$

La fonction s_i est un polynôme de degré 3 pour tout $i = 0, \cdots, n-1$, donc déterminer s revient à déterminer la valeurs de $4n$ inconnues. Les conditions (2.34) conduisent à écrire un système linéaire à $2n$ équations, quant aux conditions (2.35), elles se traduisent par un système linéaire à $n-1+n-1$ équations. On obtient donc un système à $4n-2$ équations à $4n$ inconnues. Il est donc nécessaire d'ajouter deux autres conditions afin d'assurer l'unicité de la fonction spline. Plusieurs choix sont possibles, comme par exemple poser $s''(a) = s''(b) = 0$.

Théorème 2.48 *Il existe une unique fonction spline au sens de la définition 2.47, satisfaisant la condition*

$$s''(a) = s''(b) = 0. \tag{2.36}$$

Afin d'établir le théorème 2.48, on aura recours à la proposition suivante

Proposition 2.49 *Soit $A \in M_n(I\!R)$ une matrice à diagonale strictement dominante, c'est-$\tilde{A}$ -dire telle que pour tout $i \in \{1, \cdots, n\}$, on a*

$$\sum_{j=1, j\neq i}^{n} |a_{ij}| < |a_{ii}|. \tag{2.37}$$

Alors, la matrice A est inversible.

Preuve Montrons que l'ensemble des vecteurs X tels que $AX = 0$ est réduit à $\{0\}$. Supposons que cela ne soit pas le cas. Soit X un vecteur non nul tel que $AX = 0$. Soit i_0 tel que

$$|X_{i_0}| = \max_{i\in\{1,\cdots,n\}} |X_i|.$$

Puisque $X \neq 0$, on a $X_{i_0} \neq 0$. On a

$$(AX)_{i_0} = \sum_{j=1}^{n} a_{i_0 j} X_j = 0,$$

donc par inégalité triangulaire

$$|a_{i_0 i_0}| \leq \sum_{j=1, j\neq i_0}^{n} |a_{i_0 j}| |\frac{X_j}{X_{i_0}}| \leq \sum_{j=1, j\neq i_0}^{n} |a_{i_0 j}|.$$

Cette inégalité contredit (2.37). Il est résulte que le noyau de A est vide et donc A est inversible.

Preuve du théorème 2.48
Etape 1. Détermination de s en fonction de s_i''.
On pose $h_i = x_{i+1} - x_i$, $i = 0, \cdots, n-1$. Pour $i = 1, \cdots, n-1$, on note par s_i'' la valeur de $s_i''(x_i)$. Comme $s''(a) = s''(b) = 0$, on pose $s_0'' = s_n'' = 0$.
Par interpolation aux points x_i et x_{i+1} et compte-tenu de (2.35), on obtient pour $i = 0, \cdots, n-1$

$$s_i''(x) = s_i'' \frac{x_{i+1} - x}{h_i} + s_{i+1}'' \frac{x - x_i}{h_i}.$$

En intégrant deux fois l'égalité précédente, on obtient

$$s_i(x) = s_i'' \frac{(x_{i+1} - x)^3}{6h_i} + s_{i+1}'' \frac{(x - x_i)^3}{6h_i} + a_i(x_{i+1} - x) + b_i(x - x_i),$$

où a_i et b_i sont des constantes à déterminer. Les conditions $s_i(x_i) = f_i$ et $s_i(x_{i+1}) = f_{i+1}$ se traduisent par

$$s_i'' \frac{h_i^2}{6} + a_i h_i = f_i, \quad s_{i+1}'' \frac{h_i^2}{6} + b_i h_i = f_{i+1}.$$

Remplaçant alors a_i et b_i par leur valeur respective, on obtient pour $i = 0, \cdots, n-1$:

$$s_i(x) = s_i'' \left(\frac{(x_{i+1} - x)^3}{6h_i} - \frac{h_i}{6}(x_{i+1} - x) \right) + s_{i+1}'' \left(\frac{(x - x_i)^3}{6h_i} - \frac{h_i}{6}(x - x_i) \right)$$
$$+ \frac{f_i}{h_i}(x_{i+1} - x) + \frac{f_{i+1}}{h_i}(x - x_i).$$

Etape 2. Détermination des valeurs de s_i''
Afin de déterminer s_i pour tout $i \in \{0, \cdots, n-1\}$, il faut et il suffit de déterminer la valeur de s_i'' pour tout $i \in \{1, \cdots, n-1\}$.
D'après l'étape 1, on a

$$s_i'(x) = s_i'' \left(-\frac{(x_{i+1} - x)^2}{2h_i} + \frac{h_i}{6} \right) + s_{i+1}'' \left(\frac{(x - x_i)^2}{2h_i} - \frac{h_i}{6} \right) - \frac{f_i}{h_i} + \frac{f_{i+1}}{h_i}.$$

La relation $s_i'(x_i) = s_{i-1}'(x_i)$ équivaut à

$$-s_i'' \frac{h_i}{3} - s_{i+1}'' \frac{h_i}{6} - \frac{f_i - f_{i+1}}{h_i} = s_{i-1}'' \frac{h_{i-1}}{6} + s_i'' \frac{h_{i-1}}{3} - \frac{f_{i-1} - f_i}{h_{i-1}}.$$

ou encore

$$h_i s''_{i+1} + 2(h_i + h_{i-1})s''_i + h_{i-1}s''_{i-1} = 6\left(\frac{f_{i+1} - f_i}{h_i} - \frac{f_i - f_{i-1}}{h_{i-1}}\right).$$

pour $i = 1, 2, \cdots, n-1$. Nous avons obtenu un système linéaire de $n-1$ équations à $n-1$ inconnues (à noter que d'après (2.36), s''_0 et s''_n sont connus). La matrice A du système linéaire obtenu précédemment est tridiagonale, à diagonale *strictement dominante*. D'après la proposition 2.49, elle est donc inversible et la solution est unique. On a donc établi l'existence et l'unicité de la fonction spline.

On désigne par $\mathcal{G}$ l'ensemble des fonctions de classe $C^2([a,b])$ interpolant f aux points x_i et satisfaisant l'une des deux conditions aux limites suivantes :

$$s'(a) = f'(a), \quad s'(b) = f'(b), \quad ou \quad s''(a) = s''(b) = 0. \tag{2.38}$$

La fonction spline possède la propriété remarquable suivante :

Théorème 2.50 *La fonction spline est l'unique fonction qui minimise l'énergie de flexion, autrement dit,*

$$\min_{g \in \mathcal{G}} \int_a^b g''(x)^2 dx = \int_a^b s''(x)^2 dx. \tag{2.39}$$

Preuve Etape 1. La fonction spline minimise l'énergie de flexion. Montrons que

$$\int_a^b s''(x)e''(x)dx = 0, \tag{2.40}$$

où $e(x) := f(x) - s(x)$. Effectuons deux intégrations par parties successives de $\int_a^b s''(x)e''(x)dx$. On obtient

$$\int_a^b s''(x)e''(x)dx = [s''e']_a^b - \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} s^{(3)}(x)e'(x)dx$$

puis

$$\int_a^b s''(x)e''(x)dx = s''(b)e'(b) - s''(a)e'(a) - \sum_{i=0}^{n-1}\left(-\int_{x_i}^{x_{i+1}} s^{(4)}(x)e(x)dx + [s^{(3)}(x)e(x)]_{x_i}^{x_{i+1}}\right).$$

Or, d'une part $s^{(4)} = 0$ puisque s est une fonction polynômiale de degré inférieur ou égal à 3 par morceaux et $e(x_i) = 0$ pour tout i. Il en résulte que

$$\int_a^b s''(x)e''(x)dx = s''(b)e'(b) - s''(a)e'(a),$$

et d'après (2.38), on a $s''(b)e'(b) - s''(a)e'(a) = 0$. On en déduit donc (2.40). Il en résulte que

$$\int_a^b f''(x)^2 dx = \int_a^b e''(x)^2 dx + \int_a^b s''(x)^2 dx, \tag{2.41}$$

puisque $e + s = f$ et

$$\int_a^b ((e+s)'')(x)^2 dx = \int_a^b e''(x)^2 dx + \int_a^b s''(x)^2 dx + 2\int_a^b e''(x)s''(x)dx.$$

La relation (2.41) est vraie pour toute fonction $g \in \mathcal{G}$, on en déduit immédiatement que

$$\int_a^b g''(x)^2 dx \geq \int_a^b s''(x)^2 dx, \quad \forall\, g \in \mathcal{G}$$

puis que

$$\min_{g\in\mathcal{G}} \int_a^b g''(x)^2 dx = \int_a^b s''(x)^2 dx.$$

Etape 2. Unicité du minimiseur
Soient $(s_1, s_2) \in \mathcal{G}^2$, deux minimiseurs de l'énergie. Alors d'après la relation (2.41), on a

$$\int_a^b s_1''(x)^2 dx = \int_a^b (s_1''(x) - s_2''(x))^2 dx + \int_a^b s_2''(x)^2 dx.$$

Comme $\int_a^b s_1''(x)^2 dx = \int_a^b s_2''(x)^2 dx$, il en résulte que

$$\int_a^b (s_1''(x) - s_2''(x))^2 dx = 0.$$

Donc les fonctions s_1 et s_2 diffèrent d'une fonction affine. Les conditions $s_1(x_j) = s_2(x_j) = f(x_j)$, $j = 0, \cdots, n$ impliquent que les fonctions s_1 et s_2 sont égales, ce qui achève la preuve du théorème 2.50.

3 Intégration numérique

Dans toute la suite, on considère une fonction f définie sur $[a, b]$ à valeurs réelles, intégrables sur $[a, b]$ et $a \leq x_0 < x_1 < \cdots < x_n \leq b$, $n + 1$ points distincts de $[a, b]$. On pose

$$I(f) = \int_a^b f(t)dt$$

et on se propose de déterminer une valeur approchée de $I(f)$. Comme annoncé dans l'introduction, la première méthode consistera à approcher $I(f)$ par $I(p_n)$ où p_n est le polynôme qui interpole f aux points $(x_i)_{0\leq i\leq n}$.

Dans ce chapitre consacré à l'intégration numérique, on utilisera régulièrement la deuxième formule de la moyenne :

Théorème 3.1 *Soit f une fonction continue sur $[a,b]$ et g une fonction positive, intégrable sur $[a,b]$. Alors il existe $\rho \in [a,b]$ tel que*

$$\int_a^b f(t)g(t)dt = f(\rho)\int_a^b g(t)dt. \tag{3.1}$$

Remarque 3.2 *La version "discrète" de la formule (3.1) a été établi dans la proposition 2.5. La preuve de (3.1) est analogue à celle donnée pour établir (2.1).*

Preuve On pose $\psi(x) := \displaystyle\int_a^b f(t)g(t)dt - f(x)\int_a^b g(t)dt$. La fonction g étant positive, on en déduit que $\int_a^b g(t)dt \geq 0$. Comme la fonction f est continue sur $[a,b]$, elle admet un minimum et un maximum atteint respectivement en $\bar{x}$ et $\hat{x}$. On a alors, comme g est positive sur $[a,b]$

$$\int_a^b f(t)g(t)dt \geq \int_a^b f(\bar{x})g(t)dt$$

et donc

$$\psi(\bar{x}) \geq \int_a^b f(\bar{x})g(t)dt - f(\bar{x})\int_a^b g(t)dt = 0.$$

On montre de même que

$$\psi(\hat{x}) \leq 0.$$

On déduit du théorème des valeurs intermédiaires qu'il existe $\zeta \in [a,b]$ tel que $\psi(\zeta) = 0$, ce qui achève la preuve du théorème.

3.1 Formules de quadratures

On se propose d'approcher $I(f)$ par une expression de la forme suivante, dite formule de quadrature à $n+1$ points

$$I_n(f) := \sum_{i=0}^{n} \alpha_i f(x_i), \tag{3.2}$$

$(x_i)_{\{i=0,\cdots,n\}} \in [a,b]$ distincts deux-à-deux et $(\alpha_i)_{\{i=0,\cdots,n\}}$ réels.
On pose

$$E(f) = I(f) - I_n(f).$$

On désignera par V un sous-espace vectoriel de fonctions définies sur $[a,b]$ à valeurs réelles, intégrables sur $[a,b]$.

Définition 3.3 *On dit qu'une formule de quadrature est exacte sur l'ensemble V si*

$$I(f) - I_n(f) = 0, \quad \forall\, f \in V.$$

Définition 3.4 *Nous dirons qu'une formule de quadrature à $n+1$ points est d'ordre n si elle est exacte pour tout polynôme de degré inférieur ou égal à n. Autrement dit*

$$I(p) - I_n(p) = 0, \quad \forall\, p \in V := I\!R_n[X].$$

Proposition 3.5 *Une formule de quadrature à $n+1$ points est exacte sur $I\!R_n[X]$ si et seulement si elle est de type interpolation à $n+1$ points, c'est-à-dire si*

$$\alpha_k := \int_a^b l_k(t)dt, \quad \forall\, k \in \{0,\cdots,n\}$$

où l_k représente le k-ième polynôme élémentaire de Lagrange.

Preuve Supposons la formule exacte sur $I\!R_n[X]$. Alors pour tout polynôme élémentaire de Lagrange l_i, on a :

$$\int_a^b l_k(t)dt = \sum_{i=0}^{n} \alpha_i l_k(x_i) = \alpha_k, \quad k = 1,\cdots,n.$$

Réciproquement, supposons $\alpha_k := \displaystyle\int_a^b l_k(t)dt$ pour tout k. $I\!R_n[X]$. Soit $P \in I\!R_n[X]$. Le polynôme qui interpole P aux points (x_i) n'est autre que P. On a donc

$$\int_a^b P(t)dt = \int_a^b \sum_{i=0}^{n} P(x_i) l_i(t)dt = \sum_{i=0}^{n} \alpha_i P(x_i).$$

On a montré que si la formule est de type interpolation, elle est exacte sur $I\!R_n[X]$.

Proposition 3.6 *Soit $m \in I\!N^*$. Une formule de quadrature à $n+1$ points est exacte sur $I\!R_m[X]$ si et seulement si*

$$E(x^i) = 0, \quad \forall\, i \in \{0,\cdots,m\}. \tag{3.3}$$

Preuve Observons que $f \mapsto E(f)$ est linéaire (en effet, les applications $f \mapsto \int_a^b f(t)dt$ et $f \mapsto \sum_{i=0}^n f(x_i)\lambda_i$ sont linéaires). Si les égalités (3.3) sont satisfaites, par linéarité, on obtient pour tout $(\beta_0, \beta_1, \cdots, \beta_m) \in I\!R^m$

$$\sum_{i=0}^{m} \beta_i E(x^i) = E(\sum_{i=0}^{m} \beta_i x^i) = 0,$$

ce qui achève la preuve de la proposition 3.6.

3.2 Formules de Newton-Cotes

3.2.1 Formule des rectangles

La méthode des rectangles consiste à approcher $I(f)$ par $f(a)(b-a)$ (méthode des rectangles à gauche), ou $f(b)(b-a)$ (méthode des rectangles à droite). La formule de quadrature (3.2) se réduit à

$$I_0(f) := \sum_{i=0}^{0} \alpha_i f(x_i) = (b-a)f(a).$$

Ici, $x_0 = a$ et $\alpha_0 = b - a$.
L'erreur est donné dans la proposition suivante :

Proposition 3.7 *Supposons f de classe C^1 sur $[a, b]$. L'erreur dans la méthode des rectangles est donnée par :*

$$f'(\eta)\frac{(b-a)^2}{2}, \quad \eta \in]a, b[. \tag{3.4}$$

De plus, la méthode des rectangles est une méthode exactement d'ordre 0.

Preuve Pour tout $x \in]a, b[$, il existe $c_x \in]a, x[$

$$f(x) - f(a) = f'(c_x)(x-a).$$

Intégrant cette égalité entre a et b et utilisant la deuxième formule de la moyenne en posant $g(x) = x - a$, on déduit qu'il existe $\eta \in]a, b[$ tel que

$$\int_a^b f(t)dt = f(a)(b-a) + f'(\eta)\frac{(b-a)^2}{2}.$$

Par définition, la méthode est d'ordre 0. Montrons qu'elle n'est pas d'ordre 1. En effet, d'une part on a

$$\int_a^b x dx = \frac{b^2}{2} - \frac{a^2}{2},$$

d'autre part, on a

$$f(a)(b-a) = a(b-a).$$

Or

$$\frac{b^2}{2} - \frac{a^2}{2} = a(b-a)$$

équivaut à $\frac{1}{2}(b-a)^2 = 0$, donc $b = a$.

3.2.2 Formule des trapèzes

On approche f par son polynôme d'interpolation de degré 1. Une approximation de $\int_a^b f(t)dt$ est donnée par $\int_a^b P_1(t)dt$, où P_1 est donné par

$$P_1(t) = f(a) + f[a,b](t-a).$$

On a

$$\int_a^b P_1(t)dt = (f(a)+f(b))\frac{b-a}{2}.$$

L'erreur est donné dans la proposition suivante :

Proposition 3.8 *Supposons f de classe C^2 sur $[a,b]$. L'erreur dans la méthode des trapèzes est donnée par :*

$$E(f) = \frac{f''(\eta)}{2}\int_a^b (x-a)(x-b)dx = -f''(\eta)\frac{(b-a)^3}{12}, \quad \eta \in [a,b]. \tag{3.5}$$

De plus, la méthode des trapèzes est une méthode d'ordre 1.

Preuve
Etape 1. Estimation de l'erreur D'après le théorème 2.20, on déduit que

$$\int_a^b f(t)dt - \int_a^b P_1(t)dt = \int_a^b \frac{f''(\eta_x)}{2}(x-a)(x-b)dx.$$

La fonction $g(x) = (x-a)(x-b)$ est de signe constant sur $[a,b]$ (g est négative), donc d'après la deuxième formule de la moyenne, on obtient l'estimation

$$E(f) = \frac{f''(\eta)}{2}\int_a^b (x-a)(x-b)dx = -f''(\eta)\frac{(b-a)^3}{12}, \quad \eta \in [a,b].$$

Étape 2. Détermination de l'ordre de la méthode.
Par construction, la méthode est au moins d'ordre 1. Montrons qu'elle est exactement d'ordre 1.

Étudions d'abord le cas particulier suivant. Posons $f(x) = x^2$, $a = -1$ et $b = 1$. On a

$$\int_a^b x^2 dx = \frac{2}{3}.$$

D'autre part

$$(f(a) + f(b))\frac{b-a}{2} = 2.$$

La méthode n'est donc pas d'ordre 1 dans ce cas.

Cas où a et b sont quelconques.

Posons $t = \phi(x)$, $x \in [-1, 1]$ où ϕ est définie en (2.24). On a

$$\int_{-1}^1 f(x)dx = \frac{2}{b-a}\int_a^b fo\phi^{-1}(t)dt$$

et alors d'après l'étape 1

$$\int_{-1}^1 f(x)dx \neq (f(-1) + f(1))\frac{1-(-1)}{2}$$

et comme $(f(-1) + f(1))\frac{1-(-1)}{2} = (fo\phi^{-1}(a) + fo\phi^{-1}(b))$ on obtient

$$\int_a^b fo\phi^{-1}(t)dt \neq (fo\phi^{-1}(a) + fo\phi^{-1}(b))\frac{b-a}{2}.$$

Par conséquent, la formule n'est pas exacte si on choisit $g(t) = fo\phi^{-1}(t)$, ce qui démontre que la formule n'est pas d'ordre 2. Nous venons donc d'établir que la méthode est exactement d'ordre 1, ce qui achève la preuve de la proposition 3.8.

3.2.3 Méthode de Simpson

On approche f par son polynôme d'interpolation de degré 2. Une approximation de $\int_a^b f(t)dt$ est donnée par $\int_a^b P_2(t)dt$, où P_2 est le polynôme qui interpole f en $x_0 = a$, $x_1 = \frac{a+b}{2}$ et $x_2 = b$. P_2 est donné par

$$P_2(t) = f(a) + f[a, b](x - a) + f[a, \frac{a+b}{2}, b](x - a)(x - b).$$

Après intégration, on trouve

$$\int_a^b P_2(t)dt = \frac{b-a}{6}\left(f(a) + 4f(\frac{a+b}{2}) + f(b)\right).$$

La méthode de Simpson est d'ordre 2 au moins. Montrons qu'elle est exactement d'ordre 3.

Proposition 3.9 *La méthode de Simpson est une méthode exactement d'ordre 3.*

Preuve Par construction, la méthode est une méthode d'ordre 2 au moins. Montrons qu'elle est d'ordre 3.
Posons $f(x) = x^3$. On obtient d'une part

$$\int_a^b f(t)dt = \frac{b^4}{4} - \frac{a^4}{4}$$

et d'autre part

$$\frac{b-a}{6}\left(a^3 + 4(\frac{a+b}{2})^3 + b^3\right) = \frac{b-a}{6}(a^3 + \frac{1}{2}(a+b)^3 + b^3)$$
$$= \frac{b-a}{6}(\frac{3}{2}a^3 + \frac{3}{2}b^3 + \frac{3}{2}a^2b + \frac{3}{2}ab^2) = \frac{b^4}{4} - \frac{a^4}{4}.$$

D'après la proposition 3.6, la méthode est donc au moins d'ordre 3. Montrons qu'elle n'est pas d'ordre 4.
Etape 1. Cas où $a = -1$ et $b = 1$.
Posons $f(x) = x^4$, $a = -1$ et $b = 1$. On a alors

$$\int_a^b f(t)dt = \frac{b^5}{5} - \frac{a^5}{5} = \frac{2}{5}.$$

D'autre part,

$$\frac{b-a}{6}(a^4 + \frac{1}{2}(a+b)^4 + b^4). = \frac{2}{3}.$$

On en déduit que la méthode n'est pas d'ordre 4 dans ce cas.

Etape 2 Cas où a et b sont quelconques.
Posons $t = \phi(x)$, $x \in [-1, 1]$ où ϕ est définie en (2.24). On a alors d'après l'étape 1

$$\int_{-1}^1 f(x)dx = \frac{2}{b-a}\int_a^b fo\phi^{-1}(t)dt \neq \frac{1}{3}\left(fo\phi^{-1}(a) + 4fo\phi^{-1}(\frac{a+b}{2}) + fo\phi^{-1}(b)\right).$$

Par conséquent, la formule n'est pas exacte si on choisit $g(t) = fo\phi^{-1}(t)$, ce qui démontre que la formule n'est pas d'ordre 4.

Remarque 3.10 *On montrera un peu plus loin que pour $f \in C^4([a,b])$, l'erreur dans cette méthode est donnée par*

$$E(f) = -f^{(4)}(\eta)\frac{(b-a)^5}{2880}, \quad \eta \in]a,b[.$$

La méthode de simpson, en raison de la simplicité de sa mise en oeuvre et de sa précision est la plus utilisée par les calculatrices pour tous calculs approchés d'intégrales de fonctions explicites.

3.3 Méthodes composites

3.3.1 Méthode composite des rectangles

Soit $N \in \mathbb{N}^*$. On pose $h = \dfrac{b-a}{N}$. On pose $x_i = a + ih$ $i = 0, \cdots, N$. On va appliquer la méthode des rectangles exposée précédemment sur chaque intervalle $[x_i, x_{i+1}]$. D'après la proposition 3.7, on obtient

$$\int_{x_{i-1}}^{x_i} f(x)dx = (x_i - x_{i-1})f(x_{i-1}) + f'(\eta_i)\frac{(x_i - x_{i-1})^2}{2}, \quad \eta_i \in]x_{i-1}, x_i[$$

donc d'après la relation de Chasles

$$\int_a^b f(x)dx = \sum_{i=1}^{N} hf(x_{i-1}) + \sum_{i=1}^{N} f'(\eta_i)\frac{(x_i - x_{i-1})^2}{2}.$$

Proposition 3.11 *L'erreur dans la méthode composite des rectangles est donnée par*

$$\frac{f'(\eta)(b-a)h}{2}, \quad \eta \in [a, b]. \tag{3.6}$$

Preuve D'après la proposition 2.5 appliquée en posant $g_i = \frac{(x_i - x_{i-1})^2}{2}$, on obtient

$$\sum_{i=1}^{N} f'(\eta_i)\frac{(x_i - x_{i-1})^2}{2} = \frac{h^2}{2}f'(\eta).N = \frac{f'(\eta)(b-a)h}{2}$$

d'où (3.6).

3.3.2 Méthode composite des trapèzes

On procéde de même qu'avec la méthode des rectangles. On va appliquer la méthode des trapèzes sur chaque intervalle $[x_i, x_{i+1}]$. On a alors

$$\int_a^b f(t)dt = \sum_{i=0}^{N-1} \frac{f(x_i) + f(x_{i+1})}{2}h + \sum_{i=0}^{N-1} -f''(\eta_i)\frac{(x_{i+1} - x_i)^3}{12}.$$

soit

$$\int_a^b f(t)dt = h\frac{f(a) + f(b)}{2} + h\sum_{i=1}^{N-1} f(x_i) + E(f)$$

avec

$$E(f) = -\sum_{i=0}^{N-1} f''(\eta_i)\frac{(x_{i+1} - x_i)^3}{12}.$$

On

Proposition 3.12 *L'erreur dans la méthode composite des trapèzes est donnée par*

$$-\frac{f''(\eta)h^2(b-a)}{12}, \quad \eta \in [a,b]. \tag{3.7}$$

Preuve Appliquons la proposition 3.5 à l'expression $-\sum_{i=0}^{N-1} f''(\eta_i)\frac{(x_{i+1}-x_i)^3}{12}$ en posant $g_i := \frac{(x_{i+1}-x_i)^3}{12}$ ($g_i \geq 0$ pour tout i). On obtient alors que l'erreur est donnée par (3.7).

3.3.3 Méthode composite de Simpson

On pose $f_{i-\frac{1}{2}} = f(\frac{x_{i-1}+x_i}{2})$ et $f_i = f(x_i)$. On obtient alors d'après la remarque (3.10)

$$\int_{x_{i-1}}^{x_i} f(x)dx = \frac{h}{6}[f_{i-1} + 4f_{i-\frac{1}{2}} + f_i] - \frac{f^{(4)}(\eta_i)(\frac{h}{2})^5}{90}$$

et en raisonnant comme précédemment, on obtient

$$I(f) = \frac{h}{6}\sum_{i=1}^{N}(f_{i-1} + 4f_{i-\frac{1}{2}} + f_i) + E(f)$$

avec

$$E(f) = -\frac{f^{(4)}(\zeta)(\frac{h}{2})^4(b-a)}{180}, \quad \zeta \in]a,b[\tag{3.8}$$

3.4 Applications

On considère

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

On veut obtenir une valeur approchée de I avec une erreur inférieure ou égale à 10^{-6}. Combien faut-il prendre de points d'intégrations pour obtenir une telle erreur lorsqu'on utilise la méthode des trapèzes ?
Sur cet exemple, on a $f(x) = e^{-x^2}$, $a = 0$, $b = 1$ et $h = \frac{1}{N}$. D'après (3.5), l'erreur est donnée par

$$-\frac{f''(\eta)}{12N^2}, \quad \eta \in [0,1].$$

Majorons l'expression $\max_{\eta\in[0,1]} |f''(\eta)|$. On a pour tout $x \in I\!R$,

$$f''(x) = e^{-x^2}(4x^2 - 2) \text{ et } f^{(3)}(x) = e^{-x^2}4x(3 - 2x^2).$$

On a $f^{(3)}(x) \geq 0$ sur $[0,1]$, donc f'' croit sur $[0,1]$. On en déduit que

$$\max_{\eta\in[0,1]} |f''(\eta)| = \max(|f''(0)|, |f''(1)|) = \max(2, 2e^{-1}) = 2.$$

Il faut donc choisir N tel que

$$\frac{2}{12N^2} \leq 10^{-6},$$

soit

$$N \geq \frac{1000}{\sqrt{6}} = 408.24.$$

L'entier $N = 409$ convient.

Avec la méthode de Simpson, compte tenu de (3.8), l'erreur est de la forme $f^{(4)}(\eta)\frac{1}{(2N)^4 180}$. On a pour tout $x \in I\!R$,

$$f^{(4)}(x) = 4e^{-x^2}(3 - 12x^2 + 4x^4) \quad \text{et} \quad f^{(5)}(x) = 8xe^{-x^2}(-4x^4 + 20x^2 - 15).$$

Etudions le signe de $f^{(5)}$. Cette fonction est décroissante sur $[0, \sqrt{x_1}]$ et croissante sur $[\sqrt{x_1}, 1]$ où $x_1 = \frac{5-\sqrt{10}}{2}$. On en déduit que

$$\max_{\eta \in [0,1]} |f^{(4)}(x)| = \max(|f^{(4)}(0)|, |f^{(4)}(\sqrt{x_1})|, |f^{(4)}(1)|) = 12.$$

L'erreur est inférieure à 10^{-6} si

$$12\frac{1}{(2N)^4.180} \leq 10^{-6},$$

soit

$$N \geq 8.035.$$

L'entier $N = 9$ convient.

3.5 Méthode de Péano pour le calcul de l'erreur

3.5.1 Noyau de Péano

On considère la formule de quadrature $\sum_{j=0}^{m} \lambda_j f(x_j)$. On pose $t_+ = \max(t, 0)$.

On pose

$$R(f) = \int_a^b f(t)dt - \sum_{j=0}^{m} \lambda_j f(x_j).$$

La preuve du théorème suivant repose sur la formule de Taylor avec reste intégral (2.10).

Théorème 3.13 *Soit $n \in \mathbb{N}$. Si la formule de quadrature $\sum_{j=0}^{m} \lambda_j f(x_j)$ est d'ordre n (exacte pour les polynômes de degré inférieur ou égal à n), alors pour tout $f \in C^{n+1}([a,b])$, on a*

$$R(f) = \int_a^b f(t)dt - \sum_{j=0}^{m} \lambda_j f(x_j) = \int_a^b K(t) f^{(n+1)}(t)dt$$

où

$$K(t) := \frac{1}{n!} R(x \mapsto (x-t)_+^n), \quad (x-t)_+^n = \begin{cases} (x-t)^n & si\ x \geq t \\ 0 & sinon. \end{cases}$$

Preuve
On pose

$$P(x) := f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

D'après la formule de Taylor avec reste intégral (voir théorème 2.10), on obtient

$$R(f) = R(P) + R(\int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)dt).$$

La formule étant exacte à l'ordre n, on obtient $R(P) = 0$. On a donc

$$R(f) = R(\int_a^b \frac{(x-t)_+^n}{n!} f^{(n+1)}(t)dt).$$

Comme la fonction

$$f^{(n+1)}(t)(x-t)_+^n$$

est intégrable sur $[a,b] \times [a,b]$, on a d'après le théorème de Fubini

$$\begin{aligned}
&\tfrac{1}{n!} R\left(\int_a^b f^{(n+1)}(t)(x-t)_+^n dt\right) \\
&= \tfrac{1}{n!}\left(\int_a^b \int_a^b f^{(n+1)}(t)(x-t)_+^n dt dx - \sum_{j=0}^{m} \lambda_j \int_a^b f^{(n+1)}(t)(x_j-t)_+^n dt\right) \\
&= \frac{1}{n!}\left(\int_a^b f^{(n+1)}(t)(\int_a^b (x-t)_+^n dx - \sum_{j=0}^{m} \lambda_j (x_j-t)_+^n)dt\right) = \int_a^b K(t) f^{(n+1)}(t)dt.
\end{aligned}$$

Définition 3.14 *La fonction K définie dans le théorème précédent est le noyau de Péano d'ordre n de la formule d'intégration approchée considérée.*

On déduit du théorème 3.13, la proposition suivante :

Proposition 3.15 *Si le noyau de Péano d'ordre n associé à une formule de quadrature est de signe constant, alors il existe* $\zeta \in [a,b]$ *tel que*

$$R(f) = \int_a^b K(t)dt.f^{(n+1)}(\zeta). \tag{3.9}$$

De plus,

$$\int_a^b K(t)dt = \frac{1}{(n+1)!}R(x \mapsto x^{n+1}).$$

Preuve
D'après le théorème 3.13, on a $R(f) = \int_a^b K(t)f^{(n+1)}(t)dt$. Comme K est de signe constant sur $[a,b]$, d'après la deuxième formule de la moyenne (3.1), on obtient qu'il existe $\zeta \in [a,b]$ tel que

$$R(f) = \int_a^b K(t)dt.f^{(n+1)}(\zeta).$$

Appliquant (3.9) à $f(x) = x^{n+1}$, on obtient

$$\int_a^b K(t)dt = \frac{1}{(n+1)!}R(x^{n+1}),$$

ce qui achève la preuve de la proposition 3.15.

3.5.2 Exemple de calcul de noyau de Péano et estimation d'erreur

On considère la méthode des trapèzes sur $[-1,1]$. Dans la méthode des trapèzes appliquée sur $[-1,1]$, on approche $\int_{-1}^{1} f(t)dt$ par $f(-1)+f(1)$.

Proposition 3.16 *Le noyau de Péano associé à la méthode des trapèzes sur* $[-1,1]$ *est donné par*

$$K(t) = \frac{1}{1!}R[x \mapsto (x-t)_+] = \frac{t^2-1}{2}, \quad t \in [-1,1].$$

L'erreur dans la méthode des trapèzes sur $[-1,1]$ *est donnée par*

$$-\frac{2}{3}f^{(2)}(\zeta).$$

Preuve Par définition, on a

$$K(t) = \frac{1}{1!}R[x \mapsto (x-t)_+] = \int_{-1}^{1}(x-t)_+dx - (1-t)_+ - (-1-t)_+.$$

Par définition, on a $(1-t)_+ = 1-t$ et $(-1-t)_+ = 0$. De plus,

$$\int_{-1}^{1} (x-t)_+ dx = \int_t^1 (x-t)dx = \frac{(1-t)^2}{2}.$$

On a donc

$$K(t) = \frac{1}{1!} R[x \mapsto (x-t)_+] = \frac{t^2-1}{2}.$$

Remarquons que le noyau de Péano K est de signe constant (négatif) sur $[-1,1]$. D'après la proposition 3.15, on en déduit l'expression de l'erreur suivante :

$$R(f) = \int_a^b f^{(2)}(t)K(t)dt = f^{(2)}(\zeta).\frac{1}{2!}R(x^2)$$

On a

$$R(x^2) = \int_{-1}^{1} t^2 dt - 2 = \frac{2}{3} - 2 = -\frac{4}{3}.$$

L'erreur dans la méthode des trapèzes sur $[-1,1]$ est donnée par

$$-\frac{2}{3} f^{(2)}(\zeta).$$

On retrouve bien la formule donnée en (3.5) en posant $a=-1$ et $b=1$.

Proposition 3.17 *Le noyau de Péano de la méthode de Simpson localisée sur* $[-1,1]$ *est donné par*

$$K(t) = -\frac{(1-|t|)^3(1+3|t|)}{72}.$$

L'erreur dans la méthode de Simpson sur $[a,b]$ *est donnée par :*

$$E(f) = -\frac{(b-a)^5}{2^5.90} f^{(4)}(\zeta).$$

Preuve Établissons ce résultat pour $t \geq 0$ (le cas $t \leq 0$ est laissé au lecteur en exercice). Par définition du noyau, on a pour $t \in [-1,1]$

$$K(t) = \frac{1}{6}\left(\int_{-1}^{1} (x-t)_+^3 dx - \frac{1}{3}(-1-t)_+^3 + \frac{4}{3}(-t)_+^3 + \frac{1}{3}(1-t)_+^3 \right).$$

Pour $t \geq 0$, on a $(-1-t)_+ = 0$, $(-t)_+^3 = 0$ et $(1-t)_+^3 = 1-t^3$. Par ailleurs, on a

$$\int_{-1}^{t} (x-t)_+^3 dx = 0 \quad \text{et} \quad \int_t^1 (x-t)_+^3 dx = \int_t^1 (x-t)^3 dx = \frac{(1-t)^4}{4}.$$

On en déduit que

$$K(t) = \frac{1}{6}\left(\frac{(1-t)^4}{4} - \frac{1}{3}(1-t)^3\right) = -\frac{(1-t)^3(1+3t)}{72}.$$

On observe que K est de signe constant sur $[-1,1]$. D'après la proposition 3.15, on a

$$R(f) = \int_{-1}^{1} K(t)dt.f^{(4)}(\zeta), \quad \zeta \in [-1,1].$$

De plus,

$$\int_{-1}^{1} K(t)dt = \frac{1}{4!}R(x^4).$$

Comme

$$\int_{-1}^{1} K(t)dt = \frac{1}{4!}R(x^4) = \frac{1}{24}(\int_{-1}^{1} x^4dx - \frac{4}{6}) = -\frac{1}{90},$$

on obtient

$$R(f) = -\frac{1}{90}f^{(4)}(\zeta). \tag{3.10}$$

Traitons le cas général à partir de ce cas particulier. Soit $f \in C^4([a,b])$ et effectuons le changement de variable $t = \phi(x) := \frac{b-a}{2}x + \frac{b+a}{2}$ dans $\int_a^b f(t)dt$. On a $dt = \frac{b-a}{2}dx$. Posons $g = fo\phi$. On obtient alors

$$\begin{aligned} E(f) &:= \int_{-1}^{1} fo\phi(x)\ \frac{b-a}{2}dx - \frac{b-a}{6}(fo\phi(-1) + 4fo\phi(0) + fo\phi(-1)) \\ &= \frac{b-a}{2}(\int_{-1}^{1} g(x)dx - (\frac{1}{3}g(-1) + \frac{4}{3}g(0) + \frac{1}{3}g(-1)), \end{aligned}$$

donc d'après l'estimation d'erreur obtenue en (3.10), on déduit que

$$E(f) = -\frac{b-a}{2}\frac{1}{90}g^{(4)}(\zeta).$$

Or, $g^{(4)}(x) = \frac{(b-a)^4}{2^4}f^{(4)}o\phi(x)$. On en déduit donc que

$$E(f) = -\frac{(b-a)^5}{2^5.90}f^{(4)}(\phi(\zeta)).$$

On a bien retrouvé l'estimation annoncée dans la remarque 3.10.

3.6 Formules de Newton-Côtes

3.6.1 Formules de Newton-Côtes de type *fermé*

Soit $n \in \mathbb{N}^*$. On souhaite généraliser le travail effectué précédemment dans le cas $n = 0$, $n = 1$ et $n = 2$. Plus généralement, on approche $\int_a^b f(t)dt$ par $\int_a^b P_n(t)dt$ où P_n interpole f aux points équidistants $x_i = a + ih$, $h = \frac{b-a}{n}$, $i = 0, \cdots, n$. La formule de quadrature est donnée par

$$\sum_{j=0}^{n} \alpha_j^n f(a + jh), \tag{3.11}$$

avec $\alpha_j^n = \int_a^b l_j(t)dt$, $j = 0, \cdots, n$. On a donc

$$\int_a^b f(t)dt = \sum_{j=0}^{n} \alpha_j^n f(a + jh) + E(f).$$

On pose

$$B_j^n = \frac{1}{b-a} \int_a^b l_j(t)dt.$$

Proposition 3.18 *On a*

$$B_j^n = \frac{(-1)^{n-j}}{j!(n-j)!n} \int_0^n \prod_{k=0, k\neq j}^{n} (y-k)dy. \tag{3.12}$$

De plus,

$$B_j^n = B_{n-j}^n, \quad j = 0, \cdots, n.$$

Preuve On effectue le changement de variable $y = \frac{x-a}{h}$. On obtient puisque $x = a + hy$ et $x_k = a + kh$

$$\prod_{k=0, k\neq j}^{n} (x - x_k) = h^n \prod_{k=0, k\neq j}^{n} (y - k)$$

et

$$\prod_{k=0, k\neq j}^{n} (x_j - x_k) = h^n \prod_{k=0, k\neq j}^{n} (j - k) = h^n (-1)^{n-j} j!(n-j)!.$$

On en déduit que

$$B_j^n = \frac{(-1)^{n-j}}{j!(n-j)!(b-a)} \int_0^n \prod_{k=0, k\neq j}^{n} (y-k)hdy$$

et comme $b-a=hn$, on déduit (3.12).
Pour obtenir l'égalité $B_j^n = B_{n-j}^n$, on calcule B_{n-j}^n. On a

$$B_{n-j}^n = \frac{(-1)^{n-(n-j)}}{(n-j)!j!n} \int_0^n \prod_{k=0,k\neq n-j}^{n} (y-k)dy,$$

donc comme $\prod_{k=0,k\neq n-j}^{n}(y-k) = \prod_{k=0,k\neq j}^{n}(y-n+k)$ (effectuer le changement d'indice $k = n-k'$),

$$B_{n-j}^n = \frac{(-1)^{j}}{(n-j)!j!n} \int_0^n \prod_{k=0,k\neq j}^{n} (y-n+k)dy.$$

On effectue le changement de variable $u = n-y$ dans la dernière intégrale. On obtient

$$B_{n-j}^n = \frac{(-1)^{j}}{(n-j)!j!n} \int_n^0 \prod_{k=0,k\neq j}^{n} -(u-k)(-du),$$

soit

$$B_{n-j}^n = \frac{(-1)^{j}(-1)^n}{(n-j)!j!n} \int_0^n \prod_{k=0,k\neq j}^{n} (u-k)du.$$

Comme $(-1)^{n+j} = (-1)^{n-j}$, on obtient le résultat désiré, ce qui achève la preuve de la proposition 3.18.

Les formules précédentes sont les formules fermés de Newton-Cotes de degré n.
Concernant l'étude de l'ordre de la méthode, on peut généraliser le travail effectué avec $n=1$ et $n=2$ et établir le théorème suivant :

Théorème 3.19 *La formule de quadrature (3.11) est d'ordre n si n est impair, et d'ordre $n+1$ si n est pair.*

Preuve On se bornera à montrer le résultat dans le cas où $a=-1$ et $b=1$. D'après la proposition 3.5, la formule est au moins d'ordre n puisqu'elle est de type interpolation à $n+1$ points.
Supposons n pair, soit $n=2p$. On a

$$E(x^{2p+1}) = \int_{-1}^{1} x^{2p+1}dx - 2\sum_{k=0}^{2p} B_k^n x_k^{2p+1}$$

Comme $x \mapsto x^{2p+1}$ est impaire, on a $\displaystyle\int_{-1}^{1} x^{2p+1}dx = 0$. Rappelons que compte tenu de la proposition 3.18, on a $B_k^n = B_{n-k}^n$ et que d'autre part, par symétrie de $[-1,1]$ par rapport à 0, on a

$$x_{p-k} = -x_{p+k}, \quad k = 0, \cdots, p,$$

donc $x_k = -x_{n-k}$ $k = 0, \cdots, 2p$. On obtient (car $x_p = 0$)

$$\sum_{k=0}^{2p} B_k^n x_k^{2p+1} = \sum_{k=0}^{p-1} B_k^n x_k^{2p+1} - \sum_{k=p+1}^{2p} B_{n-k}^n x_{n-k}^{2p+1}.$$

Effectuant le changement d'indice $k_1 = n-k$ dans $\displaystyle\sum_{k=p+1}^{2p} B_{n-k}^n x_{n-k}^{2p+1}$, on obtient

$$\sum_{k=p+1}^{2p} B_{n-k}^n x_{n-k}^{2p+1} = \sum_{k_1=p-1}^{0} B_{k_1}^n x_{k_1}^{2p+1}.$$

Il en résulte que $\displaystyle\sum_{k=0}^{2p} B_k^n x_k^{2p+1} = 0$. Donc la formule est d'ordre $n+1$ si n est pair. On peut montrer qu'elle n'est pas d'ordre strictement supérieur à $n+1$.
On peut montrer que $E(x^{2p+2}) \neq 0$ et $E(x^{2p}) \neq 0$ dans le cas où n est impair $(n = 2p-1)$.

Concernant l'étude de la convergence, on peut généraliser le travail effectué pour $n = 1$ et $n = 2$ et établir le théorème suivant :

Théorème 3.20 *L'erreur dans les formules de Newton-Cotes est en $\mathcal{O}(h^{n+1})$ si la formule est d'ordre n avec n impair et d'ordre $\mathcal{O}(h^{n+2})$ si n est pair.*

3.6.2 Méthode de Newton-Cotes de type ouvert

On peut construire aussi des formules de Newton-Cotes en ne prenant pas les extrémités de l'intervalle d'intégration comme abscisses d'interpolation, ce sont les formules de type ouvert.
Une exemple est donné par la méthode du point milieu. On approche $\int_a^b f(t)dt$ par $(b-a)f(\frac{a+b}{2})$. La méthode composite du point milieu s'écrit sous la forme

$$S_n := \sum_{i=0}^{n-1}(x_{i+1} - x_i)f(\frac{x_i + x_{i+1}}{2}) = h\sum_{i=0}^{n-1} f(\frac{x_i + x_{i+1}}{2}).$$

On peut montrer la proposition

Proposition 3.21 *Soient $f \in C^2([a,b])$ et $M := \max_{x\in[a,b]} |f''(x)|$. On a l'inégalité*

$$|\int_a^b f(t)dt - S_n| \leq \frac{Mh^2(b-a)}{24}.$$

Preuve
Etape 1 Posons $n = 1$. On note $c = \frac{a+b}{2}$. Comme $\int_a^b (x-c)f'(c))dx = 0$, on a que

$$(b-a)f(c) = \int_a^b (f(c) + (x-c)f'(c))dx.$$

On a alors

$$\int_a^b f(x)dx - (b-a)f(c) = \int_a^b (f(x) - f(c) - (x-c)f'(c))dx.$$

D'après la formule de Taylor-Lagrange à l'ordre 2 (2.9) appliquée au point $x = c$, on déduit qu'il existe $\theta \in]0,1[$ tel que

$$f(x) - f(c) - (x-c)f'(c) = \frac{f^{(2)}(\theta c + (1-\theta)x}{2})(x-c)^2.$$

Comme $|f''(x)| \leq M$ pour tout $x \in [a,b]$, on obtient la majoration pour tout $x \in [a,b]$

$$|f(x) - f(c) - (x-c)f'(c)| \leq \frac{M}{2}(x-c)^2.$$

Intégrant l'inégalité entre $[a,b]$, il vient

$$|\int_a^b f(x)dx - (b-a)f(c)| \leq \frac{M}{2}\int_a^b (x-c)^2 dx = \frac{M(b-a)^3}{24}.$$

Etape 2 Cas général.
D'après la première étape appliquée en posant $a = x_i$ et $b = x_{i+1}$, on obtient

$$|\int_{x_i}^{x_{i+1}} f(x)dx - hf(\frac{x_{i+1}+x_i}{2})| \leq \frac{Mh^3}{24}.$$

Appliquant alors la relation de chasles entre a et b, on obtient

$$|\int_a^b f(t)dt - S_n| = |\sum_{i=0}^{n-1}(\int_{x_i}^{x_{i+1}} f(t)dt - f(\frac{x_i+x_{i+1}}{2})h)|$$

puis par inégalité triangulaire

$$|\int_a^b f(t)dt - S_n| \leq \sum_{i=0}^{n-1} \frac{Mh^3}{24} = \frac{M(b-a)h^2}{24}.$$

3.7 Convergence et stabilité

Les formules de quadrature s'expriment en fonction du paramètre n. Il est raisonnable de penser que si n augmente, on obtienne un résultat plus précis, et à la limite, on obtienne la valeur exacte de l'intégrale. On considère dans la suite que les points x_i ($i \in \{0, \cdots, n\}$) dépendent également de n et on note les points d'intégration par x_i^n. On supposera que $x_i^n \in [a, b]$ ($(a, b) \in I\!R^2$) pour tout i et pour tout $n \in I\!N$.

Définition 3.22 *On dit qu'une formule de quadrature* $L_n(f) = \sum_{k=0}^{n} A_k^n f(x_k^n)$ *converge sur un ensemble* V *si quel que soit* $f \in V$ *on a :*

$$\lim_{n \to +\infty} \sum_{k=0}^{n} A_k^n f(x_k^n) = \int_a^b f(x)dx.$$

On pose pour $n \in I\!N$

$$E_n(f) = \int_a^b f(x)dx - \sum_{k=0}^{n} A_k^n f(x_k^n).$$

Remarque 3.23 *Dans le cas d'une approximation de* $\int_a^b f(t)dt$ *par* $\int_a^b P_n(t)dt$ *où* P_n *désigne le polynôme de Lagrange qui interpole* f *aux points* (x_i), $i = 0, \cdots, n$, *on a*

$$A_i^n = \int_a^b L_i^n(t)dt, \quad \forall\, i = 0, \cdots, n$$

où L_i^n *désigne le i-ième polynôme élémentaire de Lagrange aux points* (x_i). *Les points* x_i *dépendent du paramètre n puisque* $x_i = x_0 + i\frac{b-a}{n}$ *(on les notera donc* x_i^n*).*

Nous allons par la suite donner une condition nécessaire et suffisante pour qu'une formule de quadrature de type interpolation converge. Ceci nous conduit à introduire la notion de stabilité.

3.7.1 Stabilité

Pour qu'une méthode soit jugée bonne, il est nécessaire qu'elle soit peu sensible aux erreurs de calcul. Dans une formule de la forme $\sum_{k=0}^{n} A_k^n f(x_k^n)$, les

erreurs que l'on peut commettre portent sur les $f(x_k^n)$. Il faut donc évaluer la différence entre un calcul effectué avec $f(x_k^n)$ et un calcul effectué avec $f(x_k^n) + \epsilon_k$, c'est-à-dire évaluer :

$$\sum_{k=0}^{n} A_k^n (f(x_k^n) + \epsilon_k) - \sum_{k=0}^{n} A_k^n f(x_k^n) = \sum_{k=0}^{n} A_k^n \epsilon_k.$$

Définition 3.24 *On dit qu'une formule de quadrature de la forme* $L_n(f) = \sum_{k=0}^{n} A_k^n f(x_k^n)$ *est stable si il existe une constante* $M > 0$ *indépendante de* n*, telle que pour tout* $(\epsilon_0, \epsilon_1, \cdots, \epsilon_n)$*, on a*

$$|\sum_{k=0}^{n} A_k^n \epsilon_k| \leq M \max_{k=0,\cdots,n} |\epsilon_k| \quad \forall n \in I\!N. \tag{3.13}$$

On peut alors établir le théorème

Théorème 3.25 *La formule de quadrature* $\sum_{k=0}^{n} A_k^n f(x_k^n)$ *est stable si et seulement si il existe une constante* $M > 0$ *telle*

$$\sum_{k=0}^{n} |A_k^n| \leq M, \quad \forall\, n \in I\!N. \tag{3.14}$$

Preuve La condition est suffisante puisque

$$\sum_{k=0}^{n} |A_k^n||\epsilon_k| \leq \max_k |\epsilon_k| \sum_{k=0}^{n} |A_k^n| \leq M \max_k |\epsilon_k|.$$

Montrons qu'elle est nécessaire. Raisonnons par contraposée. S'il n'existe pas de constante satisfaisant (3.14), alors pour tout $M > 0$, il existe $n(M) \in I\!N$ tel que

$$\sum_{k=0}^{n(M)} |A_k^{n(M)}| > M.$$

Il existe donc une fonction ϕ croissante définie sur $I\!N$ et à valeurs dans $I\!N$ telle que

$$\lim_{n \to +\infty} \sum_{k=0}^{\phi(n)} |A_k^{\phi(n)}| = +\infty.$$

Pour construire ϕ, il suffit de faire varier M en posant successivement $M = 1, 2, 3, \cdots$. On pose alors $\phi(n) := n(M)$.
Pour $k \in \{0, \cdots, \phi(n)\}$ tel que $A_k^{\phi(n)} \neq 0$, on pose

$$\epsilon_k = \frac{A_k^{\phi(n)}}{|A_k^{\phi(n)}|},$$

et $\epsilon_k = 0$ sinon. Donc $\max_{k\in\{0,\cdots,\phi(n)\}} |\epsilon_k| = 1$. On a

$$|\sum_{k=0}^{\phi(n)} A_k^{\phi(n)} \epsilon_k| = \sum_{k=0}^{\phi(n)} |A_k^{\phi(n)}| \to +\infty$$

quand n tend vers $+\infty$, ce qui contredit (3.13).

3.7.2 Convergence

Nous donnons par la suite une condition nécessaire et suffisante pour que la formule de quadrature introduite à la définition 3.22 soit convergente. La preuve nécessite d'avoir recours au théorème de Banach-Steinhauss, que nous admettrons. Son énoncé est le suivant :

Théorème 3.26 *Soient E et F deux espaces vectoriels normés complets. Soit (f_n) une suite d'applications linéaires continues définies sur E à valeurs dans F telles que, pour tout $x \in E$, on a*

$$\sup_{n\in I\!N} \|f_n(x)\|_F < +\infty.$$

Alors

$$\sup_{n\in I\!N} \|f_n\| < +\infty \quad (\|f_n\| := \sup_{x\neq 0} \frac{\|f_n(x)\|_F}{\|x\|_E}).$$

On rappelle également que l'espace vectoriel des polynômes est dense dans $(C^0([a,b]), \|.\|_\infty)$.

Théorème 3.27 *Soit $f \in C^0([a,b])$. Pour tout $\epsilon > 0$, il existe $P \in I\!R[X]$ tel que*

$$\|f - P\|_\infty < \epsilon.$$

On rappelle enfin qu'une application linéaire définie sur un espace vectoriel normé E à valeurs dans un espace vectoriel normé F est continue sur E si et seulement si il existe $k > 0$ telle que

$$\|f(x)\|_F \leq k\|x\|_E, \ \ \forall\, x \in E.$$

Le théorème suivant fournit une condition nécessaire et suffisante pour que la méthode de quadrature soit convergente.

Théorème 3.28 *Une condition nécessaire et suffisante pour que la formule de quadrature $\sum_{i=0}^n A_i^n f(x_i^n)$ soit convergente sur $C^0([a,b])$ est que*

- *i.* $\exists M > 0,\ \sum_{i=0}^{n} |A_i^n| < M, \quad \forall\, n \in I\!N.$
- *ii.* $\forall\, N \in I\!N,\ \lim_{n\to+\infty} E_n(x^N) = 0.$

Preuve
Montrons que la condition est suffisante. Soient $\epsilon > 0$ et $f \in C^0([a,b])$. D'après le théorème 3.27, il existe $P \in I\!R[X]$ tel que

$$\|f - P\|_\infty < \frac{\epsilon}{2(M+b-a)}.$$

D'autre part, par linéarité de $f \mapsto E_n(f)$, on a

$$E_n(f) = E_n(f-P) + E_n(P).$$

Il existe $m+1$ réels β_i tels que

$$P = \sum_{i=0}^{m} \beta_i x^i.$$

Par linéarité de E_n, on déduit que

$$E_n(P) = \sum_{i=0}^{m} \beta_i E_n(x^i).$$

D'après ii., pour tout $i \in \{0, \cdots, m\}$, $E_n(x^i)$ tend vers 0 quand n tend vers $+\infty$. On en déduit qu'il existe n_0 tel que, pour tout $n \geq n_0$ on a

$$|E_n(P)| < \frac{\epsilon}{2}. \tag{3.15}$$

Par ailleurs, on a par inégalité triangulaire

$$\begin{aligned} &|E_n(f-P)| \\ &\leq |\textstyle\sum_{i=0}^{n} A_i^n (f(x_i^n) - P(x_i^n))| + \int_a^b |f(t) - P(t)|dt \leq \|f-P\|_\infty \left(\textstyle\sum_{i=0}^{n} |A_i^n| + (b-a)\right), \end{aligned}$$

et d'après ii., on obtient alors

$$|E_n(f-P)| \leq (M+b-a)\|f-P\|_\infty < \frac{\epsilon}{2}. \tag{3.16}$$

D'après (3.15) et (3.16), on déduit que pour tout $n \geq n_0$, on a

$$|E_n(f)| \leq \epsilon.$$

La convergence de la méthode est donc établie.
Établissons à présent la réciproque. ii. est alors vrai puisque $x \mapsto x^N$ est une

fonction continue sur $\mathbb{R}$. Prouvons i.
Pour tout $f \in C^0([a,b])$, la suite de réels $(E_n(f))$ converge, donc elle est bornée. D'autre part, $f \mapsto E_n(f)$ est linéaire continue sur $C^0([a,b])$. En effet, pour tout $f \in C^0([a,b])$, on a

$$|\int_a^b f(t)dt| \leq (b-a)\|f\|_\infty$$

et

$$|\sum_{i=0}^{n} A_i^n f(x_i^n)| \leq \sum_{i=0}^{n} |A_i^n|.\|f\|_\infty$$

donc, on en déduit que

$$|E_n(f)| \leq ((b-a) + \sum_{i=0}^{n} |A_i^n|)\|f\|_\infty.$$

Les espaces $E = (C^0([a,b]), \|.\|_\infty)$ et $F = \mathbb{R}$ muni de la norme $|.|$ sont des espaces complets. D'autre part, comme la suite $(E_n(f))$ converge vers 0 pour tout f, elle est bornée. On peut donc appliquer le théorème de Banach-Steinhauss (voir 3.26) avec $f_n := E_n$, $E = (C^0([a,b]), \|.\|_\infty)$ et $F = \mathbb{R}$ muni de la norme $|.|$. On en déduit qu'il existe $C > 0$ tel que

$$\|E_n\| \leq C, \quad \forall\, n \in \mathbb{N}. \tag{3.17}$$

Considérons alors une suite de fonctions (f_n) ($f_n \in C^0([a,b])$ pour tout n) telle que $\|f_n\|_\infty = 1$ pour tout $n \in \mathbb{N}$ et $f_n(x_i^n) = 1$ si $A_i^n > 0$ et $f_n(x_i^n) = -1$ si $A_i^n < 0$. On a alors

$$\sum_{i=0}^{n} |A_i^n| = \sum_{i=0}^{n} A_i^n.f_n(x_i) = -E_n(f_n) + \int_a^b f_n(x)dx.$$

D'après (3.17), on en déduit que

$$\sum_{i=0}^{n} |A_i^n| \leq C + b - a.$$

Ainsi, i. est réalisé avec $M = C + b - a$, ce qui achève la preuve du théorème 3.28.

3.8 Formules de Gauss

3.8.1 Polynôme orthogonaux

Dans cette section, une fonction poids est une fonction définie sur un ouvert $]a,b[$ de $\mathbb{R}$ à valeurs réelles, positive et intégrable sur $]a,b[$.
Dans la suite, on considère le produit scalaire

$$(f,g) = \int_a^b f(x)g(x)w(x)dx, \tag{3.18}$$

où w est une fonction poids.

Définition 3.29 *Une suite de polynômes (P_n) est une suite de polynômes orthogonaux si*

- $deg\, P_i = i, \quad \forall i \in \mathbb{N}$,
- $(P_i, P_j) = 0 \ \forall\, (i,j) \in \mathbb{N}^2,\ i \neq j$.

Proposition 3.30 *Une suite de polynômes $(P_n)_{n\in\mathbb{N}}$ telle que $deg\, P_n = n$ pour tout $n \in \mathbb{N}$ constitue une base de $\mathbb{R}[X]$.*

Preuve Montrons que le système $\{P_0, P_1, \cdots, P_n\}$ constitue une base de $\mathbb{R}_n[X]$ pour tout $n \in \mathbb{N}$. Le résultat est vrai pour $n = 0$. Supposons le résultat vrai pour l'entier $n-1$, $n \geq 1$, n quelconque. Considérons l'égalité

$$\sum_{i=0}^{n} \alpha_i P_i = 0.$$

On a

$$\alpha_n P_n = -\sum_{i=0}^{n-1} \alpha_i P_i,$$

et par hypothèse ($\deg P_i = i$) le degré de $-\sum_{i=0}^{n-1} \alpha_i P_i$ est inférieur ou égal à $n-1$. Par conséquent, l'égalité précédente ne peut être satisfaite que si $\alpha_n = 0$. Par hypothèse de récurrence, on en déduit que $\alpha_0 = \alpha_1 = \cdots = \alpha_{n-1} = 0$. Donc le système $P_0, P_1, \cdots, P_n$ constitue une base de $\mathbb{R}_n[X]$ parce qu'il compte $n+1$ vecteurs constituant un système libre dans un espace vectoriel de dimension $n+1$.

Le premier objectif est de construire une suite de polynômes orthogonaux pour le produit scalaire (3.18). Cette construction repose sur le procédé d'orthogonalisation de Gram-Schmidt.
On a la proposition fondamentale suivante :

Proposition 3.31 *Quel que soit le poids w intégrable sur $[a,b]$, il existe une suite de polynômes orthogonaux (P_i) au sens de la définition 3.29. La suite constituée des polynômes $P_0(x)=1$ et pour $n \geq 1$*

$$P_n(x) = x^n - \sum_{i=0}^{n-1} c_{in} P_i$$

avec pour $i = 0, \cdots, n-1$

$$c_{in} = \frac{(x^n, P_i)}{(P_i, P_i)} \tag{3.19}$$

satisfait les deux conditions de la définition 3.29.

Preuve On construit cette suite de polynômes à partir des vecteurs de la base canonique de $\mathbb{R}[X]$ par le procédé d'orthogonalisation de Gram-Schmidt. On pose $P_0(x) = 1$. On construit P_1 en déterminant le projeté de x sur la droite engendrée par le vecteur 1. Le projeté orthogonal de x noté $P(1)$ appartient à la droite vectorielle engendrée par 1 (donc il s'écrit sous la forme $\alpha 1$) et satisfait

$$(x - P(x), 1) = 0,$$

donc

$$\alpha = c_{01} = \frac{(x,1)}{(1,1)}.$$

Le polynôme P_1 recherché est donc défini par

$$P_1(x) = x - c_{0,1}.$$

Supposons avoir déterminé les vecteurs P_0, P_1, $\cdots$, P_n ($n \geq 1$). On projette le vecteur x^{n+1} sur l'espace vectoriel vect$(P_0, P_1, \cdots, P_n\}$. Notons $P(x^{n+1})$ le projeté orthogonal de x^{n+1}. Ce vecteur s'écrit sous la forme $P(x^{n+1}) = \sum_{i=0}^{n} \alpha_i P_i$, et il satisfait

$$(x^{n+1} - P(x^{n+1}), v) = 0 \quad \forall\, v \in \text{vect}(P_0, P_1, \cdots, P_n),$$

et en particulier en prenant $v = P_j$ (j$\in \{0, \cdots, n\}$), on obtient :

$$(x^{n+1} - P(x^{n+1}), P_j) = 0.$$

On en déduit que

$$(x^{n+1} - \alpha_j P_j, P_j) = 0,$$

donc

$$\alpha_j = \frac{(x^{n+1}, P_j)}{(P_j, P_j)}.$$

Le vecteur P_{n+1} défini par $x^{n+1} - P(x^{n+1})$ est orthogonal à P_i pour tout $i = 1, \cdots, n$. On en déduit (3.19).

Dans la suite, on note par a_n le coefficient du monôme de plus haut degré de P_n.

Remarque 3.32 *Si on pose $a = -1$, $b = 1$ et $w = 1$ dans (3.18), on obtient les polynômes orthogonaux de Legendre.*
Si on prend $a = -\infty$ et $b = +\infty$, et $w(x) = e^{-x^2}$, on obtient les polynômes orthogonaux de Hermite.
Enfin, avec le choix $a = -1$, $b = 1$ et $w = \frac{1}{\sqrt{1-x^2}}$, on obtient les polynômes de Tchebychev.

Proposition 3.33 *Soit $k, n \in \mathbb{N}^*$, $k < n$. Si $P \in \mathbb{R}_k[X]$, alors on a*

$$(P_n, P) = 0.$$

D'autre part, si $P \in \mathbb{R}[X]$ est de degré n et si $P \in \mathbb{R}_{n-1}[X]^\perp$, alors il existe $C \in \mathbb{R}^*$ tel que $P_n = CP$.

Preuve Il suffit d'écrire P dans une base de $\mathbb{R}_k[X]$ constituée de polynômes orthogonaux. On a alors

$$(P_n, P) = (P_n, \sum_{i=0}^{k} \alpha_k P_k) = \sum_{i=0}^{k} \alpha_k (P_k, P_n) = 0.$$

D'autre part, si $P \in (\mathbb{R}_{n-1}[X])^\perp$ et deg $P = n$, on a $P = \sum_{i=0}^{n} \alpha_i P_i$ et $(P, P_j) = \alpha_j = 0$ pour tout $j = 1, \cdots, n-1$ d'où le résultat.

On peut alors montrer que le polynôme P_n admet n racines simples dans $]a, b[$.

Proposition 3.34 *Soit $n \in \mathbb{N}^*$. Le polynôme P_n admet n racines simples dans $]a, b[$.*

Preuve
Soient $x_1, x_2, \cdots, x_j$ les racines distinctes de P_n se trouvant dans $]a, b[$. On a $j \leq n$. Supposons $j < n$. Le polynôme P_n va changer de signe en toute racine de multiplicité impaire. Posons pour $j \geq 1$

$$Q(x) = \prod_{k=1}^{j} (x - x_k)^{\epsilon(k)} \quad \text{où } \epsilon(k) = 1 \text{ si } x_k \text{ est de multiplicité impaire, } 0 \text{ sinon}$$

et si $j = 0$, $Q(x) = 1$.
On remarque que le produit $P_n Q$ ne change pas de signe dans $]a, b[$ et que $\deg Q \leq n - 1$. On a donc d'après la proposition 3.33

$$(P_n, Q) = 0,$$

ce qui est impossible donc $j = n$. Conclusion : Toutes les racines de P_n sont dans $]a, b[$ et sont simples ce qui achève la preuve de la proposition.

On peut également établir la proposition suivante :

Proposition 3.35 *Les polynômes orthogonaux vérifient une relation de récurrence à trois termes*

$$P_{i+1} = A_i(x - B_i)P_i(x) - C_i P_{i-1}(x), \quad i \in I\!N \tag{3.20}$$

où

$$A_i = \frac{a_{i+1}}{a_i}, \quad B_i = \frac{(xP_i, P_i)}{(P_i, P_i)}, \quad C_i = \frac{A_i(P_i, P_i)}{A_{i-1}(P_{i-1}, P_{i-1})}$$

et

$$P_{-1}(x) = 0.$$

Preuve
On considère le polynôme $Q_n = P_{n+1} - A_n x P_n$ et on pose

$$A_n = \frac{a_{n+1}}{a_n}.$$

Avec ce choix de A_n, Q_n est de degré n. Ecrivons ce polynôme dans la base P_0, P_1, $\cdots$, P_n

$$Q_n = \sum_{i=0}^{n} \alpha_i Pi.$$

Nous avons pour $j = 0, \cdots, n$

$$\alpha_j = (Q_n, P_j) = (P_{n+1}, P_j) - A_n(xP_n, P_j) = -A_n(P_n, xP_j).$$

Mais $(P_n, xP_j) = 0$ pour tout $j = 0, \cdots, n-2$, donc

$$Q_n = \alpha_n P_n + \alpha_{n-1} P_{n-1}.$$

Déterminons α_n et α_{n-1}. Nous pouvons écrire

$$xP_{n-1} = \frac{a_{n-1}}{a_n} P_n + q_{n-1},$$

où le degré de q_{n-1} est inférieur ou égal à $n-1$. On a

$$(Q_n, P_{n-1}) = (P_{n+1} - A_n x P_n, P_{n-1}) = \alpha_{n-1}$$

ou encore

$$\alpha_{n-1}(P_{n-1}, P_{n-1}) = -A_n(P_n, xP_{n-1}) = -A_n(\frac{a_{n-1}}{a_n}P_n + q_{n-1}, P_n) = -A_n\frac{a_{n-1}}{a_n}(P_n, P_n)$$

Donc

$$\alpha_{n-1} = -\frac{A_n(P_n, P_n)}{A_{n-1}(P_{n-1}, P_{n-1})}.$$

Déterminons à présent α_n. On a

$$(Q_n, P_n) = (P_{n+1} - A_n x P_n, P_n) = -A_n(xP_n, P_n) = \alpha_n(P_n, P_n),$$

donc

$$\alpha_n = -\frac{A_n.(xP_n, P_n)}{(P_n, P_n)}.$$

On en déduit (3.20).

3.8.2 Formules de quadrature d'ordre maximal

On considère une formule de quadrature générale $\sum_{i=1}^{n} \lambda_i f(x_i)$. L'objectif est de choisr les (λ_i) et les (x_i) de telle sorte que la formule de quadrature soit d'ordre le plus élevé possible. Nous savons d'après la proposition 3.5 qu'une condition nécessaire et suffisante pour que la formule soit d'ordre $n-1$ (attention, ici, nous travaillons avec n points au lieu de $n+1$ points) est qu'elle soit de type interpolation. Il a été établi dans cette proposition que $\lambda_i = \int_a^b l_i(t)dt$ où l_i est le ième polynôme élémentaire de Lagrange. L'objectif est de choisir les (x_i) au mieux de telle sorte à rendre l'ordre de la formule le plus élevé possible. La réponse à cette question est donnée dans le théorème suivant :

Théorème 3.36 *L'unique formule de quadrature à n points d'ordre maximal est la formule par interpolation construite en prenant pour noeuds les zéros du n-ième polynôme orthogonal construit dans la proposition 3.31 par rapport au poids w. La formule ainsi déterminée est d'ordre $2n-1$. Elle est dite formule Gaussienne.*

Preuve
Soit P_n le n-ième polynôme orthogonal défini dans la proposition 3.31 et $P \in I\!R_{2n-1}[X]$. On a

$$P = P_n q + r,$$

r polynôme de degré inférieur ou égal à $n-1$. Soient x_j les zéros de P_n. Montrons que la formule est bien d'ordre au moins $2n-1$. Compte-tenu des degrés respectifs de P_n et q, on a d'après la proposition 3.33

$$\int_a^b P_n(x)q(x)w(x)dx = 0.$$

D'autre part, comme $P_n(x_j) = 0$ pour tout $j = 1, \cdots, n$ et que

$$\int_a^b r(x)w(x)dx = \sum_{j=1}^{n} \lambda_j r(x_j)$$

puisque la formule est de type interpolation, on a

$$\int_a^b P(x)w(x)dx = \int_a^b P_n(x)q(x)w(x)dx + \int_a^b r(x)w(x)dx$$
$$= \sum_{j=1}^{n} \lambda_j P_n(x_j)q(x_j) + \sum_{j=1}^{n} \lambda_j r(x_j) = \sum_{j=1}^{n} \lambda_j P(x_j).$$

La formule ainsi définie est donc au moins d'ordre $2n-1$.
Elle n'est pas de degré $2n$ puisque

$$\int_a^b P_n(x)^2 w(x)dx - \sum_{j=1}^{n} \lambda_j P_n(x_j)^2 = \int_a^b P_n(x)^2 dx \neq 0.$$

Réciproquement, considérons une formule de quadrature exacte d'ordre $k \geq 2n-1$, notée $\sum_{j=1}^{n} \mu_j f(y_j)$. Comme $k > n$, on a vu que l'on a nécessairement

$$\mu_j = \int_a^b l_j(t)dt, \quad \forall\, j \in \{1, \cdots, n\}.$$

Montrons à présent que $y_j = x_j$ pour tout $j = 1, \cdots, n$.
Considérons le polynôme $\tilde{p}(x) = \prod_{i=1}^{n}(x - y_j)$. Pour tout $P \in I\!R_{2n-1}[X]$, $(\deg P \geq n)$, on a

$$P = \tilde{p}Q + r$$

avec $\deg r \leq n-1$. On obtient alors

$$\int_a^b P(x)w(x)dx = \int_a^b \tilde{p}(x)Q(x)w(x)dx + \int_a^b r(x)w(x)dx = \sum_{i=1}^{n} P(y_j)\mu_j = \sum_{i=1}^{n} r(y_j)\mu_j.$$

On a donc nécessairement

$$\int_a^b \tilde{p}(x)Q(x)w(x)dx = (\tilde{p}, Q) = 0.$$

Le polynôme P étant quelconque, on a montré que

$$\int_a^b \tilde{p}(x)Q(x)w(x)dx = (\tilde{p}, Q) = 0. \quad \forall\, Q \in I\!R_{n-1}[X].$$

Il résulte de la proposition 3.33 que $\tilde{p} = kP_n$ $(k \neq 0)$ et par conséquent, les racines de P_n sont égales à celles de $\tilde{p}$, autrement dit, $y_j = x_j$ pour tout $j = 1, \cdots, n$.

On peut alors obtenir l'estimation d'erreur suivante :

Théorème 3.37 *Pour $f \in C^{2n}(a, b])$, l'erreur de quadrature dans la formule de Gauss est donnée par*

$$\int_a^b f(x)dx - \sum_{j=1}^{n} \lambda_j f(x_j) = \frac{f^{(2n)}(\alpha)}{(2n)!} \int_a^b \prod_{i=1}^{n} (x - x_i)^2 dx, \quad \alpha \in]a, b[. \quad (3.21)$$

Preuve
Considérons le polynôme de Hermite H_{2n-1} interpolant f en $x_1, x_2, \cdots, x_n$. On a montré que l'erreur $f(x) - H_{2n-1}(x)$ est donnée par

$$f(x) - H_{2n-1}(x) = \frac{f^{(2n)}(\zeta_x)}{(2n)!} \prod_{i=1}^{n} (x - x_i)^2.$$

Intégrons l'égalité précédente entre a et b. D'après la deuxième formule de la moyenne, on obtient

$$\int_a^b \frac{f^{(2n)}(\zeta_x)}{(2n)!} \prod_{i=1}^{n} (x - x_i)^2 dx = \frac{f^{(2n)}(\alpha)}{(2n)!} \int_a^b \prod_{i=1}^{n} (x - x_i)^2 dx, \quad \alpha \in]a, b[.$$

D'autre part, comme la formule est d'ordre $2n - 1$ et que le degré de H_{2n-1} est égal à $2n - 1$, on obtient

$$\int_a^b H_{2n-1}(x)\omega(x)dx = \sum_{i=1}^{n} H_{2n-1}(x_i)\lambda_j = \sum_{i=1}^{n} f(x_i)\lambda_j,$$

ce qui achève la preuve du théorème 3.37.

Exemple
Si $w(x) = 1$ $a = -1$ et $b = 1$, les polynômes orthogonaux de Legendre sont donnés par $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = x^2 - \frac{1}{3}$, $P_3(x) = x^3 - \frac{3}{5}x$, ... Les racines du polynôme P_2 sont données par $-\frac{1}{\sqrt{3}}$ et $\frac{1}{\sqrt{3}}$. On pose $x_1 = -\frac{1}{\sqrt{3}}$ et

$x_2 = \frac{1}{\sqrt{3}}$. Les nombres λ_1 et λ_2 sont égaux respectivement à $\int_{-1}^{1} l_0(t)dt$ et $\int_{-1}^{1} l_1(t)dt$. On a $l_0(x) = \frac{\sqrt{3}}{2}(x + \frac{1}{\sqrt{3}})$, d'où on déduit que

$$\int_{-1}^{1} l_0(t)dt = 1.$$

On montre de même que $\lambda_1 = 1$. Il en résulte que la formule de quadrature à deux points (avec le choix $a = -1$ et $b = 1$) s'écrit :

$$\int_{-1}^{1} f(t)dt \sim f(-\frac{1}{\sqrt{3}}) + f(\frac{1}{\sqrt{3}}).$$

D'après le théorème 3.37, on déduit qu'il existe $\zeta \in]-1, 1[$ tel que

$$E(f) = \frac{f^{(4)}(\zeta)}{4!} \int_{-1}^{1} (x + \frac{1}{\sqrt{3}})^2 (x - \frac{1}{\sqrt{3}})^2 dx = \frac{1}{135} f^{(4)}(\zeta).$$

3.9 Méthode de Romberg

Dans cette partie, l'objectif est de déterminer une méthode permettant d'accélérer la vitesse de convergence de la méthode des trapèzes. La méthode présentée ici est due à Romberg.

3.9.1 Polynômes de Bernouilli

Proposition 3.38 *Il existe une unique suite de polynômes (B_n) tels que $B_0(x) = 1$ et*

$$B_n'(x) = nB_{n-1}(x), \quad \forall\, n \in I\!N^* \tag{3.22}$$

et

$$\int_0^1 B_n(x)dx = 0. \tag{3.23}$$

Les nombres $b_n = B_n(0)$ sont appelés les nombres de Bernouilli.

Preuve On construit les polynômes par récurrence. B_0 est défini. Supposons B_{n-1} construit pour $n \geq 1$. Alors

$$B_n(x) = n \int_0^x B_{n-1}(t)dtdx + k$$

où k est à choisir de telle sorte que (3.23) soit vérifié. k est donc déterminée de façon unique en posant

$$k = -n \int_0^1 \int_0^x B_{n-1}(t)dtdx.$$

On peut alors établir la proposition suivante :

Proposition 3.39 *Pour tout $n \in \mathbb{N}$, on a*

$$(-1)^n B_n(1-x) = B_n(x). \tag{3.24}$$

De plus, on a

$$b_{2n+1} = 0 \ \ \forall\, n \geq 1. \tag{3.25}$$

Preuve Observons que $B_0(1) = B_0(0)$ et montrons que pour tout $n \neq 1$, on a

$$B_n(0) = B_n(1). \tag{3.26}$$

En effet, pour $n \geq 2$, on a d'après (3.22)

$$B_n(1) - B_n(0) = \int_0^1 B_n'(t)dt = n \int_0^1 B_{n-1}(t)dt = 0.$$

On pose $c_n(x) = (-1)^n B_n(1-x)$. On a $c_0(x) = 1$ et

$$c_n'(x) = (-1)^{n+1} B_n'(1-x) = (-1)^{n+1} n B_{n-1}(1-x) = n c_{n-1}(x).$$

Enfin,

$$\int_0^1 c_n(x)dx = \int_0^1 (-1)^n B_n(1-x)dx = 0.$$

Par unicité de la suite des polynômes de Bernouilli définis dans la proposition 3.38, on déduit l'égalité (3.24). D'autre part, pour $n \neq 1$, on a d'après (3.24) et (3.26)

$$b_n = (-1)^n b_n,$$

d'où (3.25).

3.9.2 Formule sommatoire d'Euler Mac Laurin

En premier lieu, établissons la formule sommatoire d'Euler Mac Laurin.

Théorème 3.40 *(Formule D'Euler-Mac Laurin) Soient m, n deux entiers tels que $m < n$, soit $r \in \mathbb{N}^*$ et $f \in C^r([m,n])$. On a*

$$\sum_{k=m}^{n} f(k) = \int_m^n f(t)dt + \frac{1}{2}(f(m)+f(n)) + \sum_{p=1}^{E(\frac{r}{2})} \frac{b_{2p}}{(2p)!}(f^{(2p-1)}(n) - f^{(2p-1)}(m))$$
$$+\frac{(-1)^{r+1}}{r!}\int_m^n \tilde{B}_r(t)f^{(r)}(t)dt, \tag{3.27}$$

où $\tilde{B}_r(t)$ désigne la fonction 1-périodique qui coÃ¯ncide avec B_r sur $[0,1[$ et $E(x)$ désigne la partie entière de x.

Preuve On procède par récurrence sur r. Supposons $r = 1$. On a

$$B_1(x) = x - \frac{1}{2}.$$

Pour $k \in \{m, \cdots, n-1\}$, considérons la fonction $\tilde{B}_1$ sur $[k, k+1[$. Prolongeons la fonction $\tilde{B}_1$ à gauche de $k+1$ par continuité en posant $\tilde{B}_1(k+1) = \frac{1}{2}$ (la fonction ainsi obtenue est de classe C^1 sur $[k, k+1]$, on la note $\tilde{B}_1$ par commodité).
Pour tout $k \in \{m, \cdots, n-1\}$, une intégration par parties sur $[k, k+1]$ appliquée aux fonctions de classe C^1 f et $\tilde{B}_1$ conduit à

$$\int_k^{k+1} f(t)dt = \int_k^{k+1} \tilde{B}_1'(t)f(t)dt = \frac{1}{2}\left(f(k+1)+f(k)\right) - \int_k^{k+1} \tilde{B}_1(t)f'(t)dt.$$

Donc, en sommant sur k, on obtient

$$\int_m^n f(t)dt = \frac{1}{2}(f(m)+f(n)) + \sum_{k=m}^{n} f(k) - \int_m^n f'(t)\tilde{B}_1(t)dt,$$

ce qui établit la formule dans le cas $r = 1$.
On suppose la formule démontrée à un rang $r \geq 1$. Soit $f \in C^{r+1}([m,n])$. D'après (3.24) et (3.26), la formule d'intégration par parties donne

$$\int_m^n \tilde{B}_r(t)f^{(r)}(t)dt = [\frac{\tilde{B}_{r+1}(t)}{r+1}f^{(r)}(t)]_m^n - \int_m^n \frac{\tilde{B}_{r+1}(t)}{r+1}f^{(r+1)}(t)dt$$
$$= \frac{b_{r+1}}{r+1}(f^{(r)}(n) - f^{(r)}(m)) - \int_m^n \frac{\tilde{B}_{r+1}(t)}{r+1}f^{(r+1)}(t)dt.$$

On a utilisé le fait que $\tilde{B}_{r+1}(m) = B_{r+1}(0)$ pour tout $m \in \mathbb{N}$, $m \geq 2$.
Par hypothèse de récurrence, on en déduit que

$$\sum_{k=m}^{n} f(k) = \int_m^n f(t)dt + \frac{1}{2}(f(m)+f(n)) + \sum_{p=1}^{E(\frac{r}{2})} \frac{b_{2p}}{(2p)!}(f^{(2p-1)}(n) - f^{(2p-1)}(m))$$
$$+\frac{(-1)^{r+1}}{r!}[\frac{b_{r+1}}{r+1}(f^{(r)}(n) - f^{(r)}(m)) - \int_m^n \frac{\tilde{B}_{r+1}}{r+1} f^{(r+1)}(t)dt]. \tag{3.28}$$

En distinguant le cas où r est pair du cas où r est impair, et en utilisant le fait que $b_{2p+1} = 0$ pour tout $p \in \mathbb{N}^*$, on obtient la formule (3.27). En effet, si r est pair, alors $r+1$ est impair et $b_{r+1} = 0$. L'égalité (3.28) devient

$$\sum_{k=m}^{n} f(k) = \int_m^n f(t)dt + \frac{1}{2}(f(m)+f(n)) + \sum_{p=1}^{E(\frac{r}{2})} \frac{b_{2p}}{(2p)!}(f^{(2p-1)}(n) - f^{(2p-1)}(m))$$
$$+\frac{(-1)^{r+2}}{(r+1)!} \int_m^n \tilde{B}_{r+1}(t) f^{(r+1)}(t)dt$$

ce qui établi (3.27) dans ce cas. Si r est impair, $r = 2t+1$ alors $r+1 = 2t+2$ et (3.28) devient alors

$$\sum_{k=m}^{n} f(k) = \int_m^n f(t)dt + \frac{1}{2}(f(m)+f(n)) + \sum_{p=1}^{t} \frac{b_{2p}}{(2p)!}(f^{(2p-1)}(n) - f^{(2p-1)}(m))$$
$$+\frac{b_{2t+2}}{(2t+2)!}(f^{(2t+1)}(n) - f^{(2t+1)}(m)) + \frac{(-1)^{r+2}}{(r+1)!} \int_m^n \tilde{B}_{r+1}(t) f^{(r+1)}(t)dt],$$

ce qui est le résultat attendu.

La formule d'Euler Mac-Laurin a de très nombreuses applications. Elle permet notamment d'obtenir un développement asymptotique de certaines suites. Un exemple d'application est donné par la détermination d'un développement asymptotique de la suite (u_n) définie par

$$u_n = \sum_{i=1}^{n} \frac{1}{k} - \ln n,$$

qui est une suite convergente (elle converge vers la constante d'Euler γ). On a la proposition

Proposition 3.41 *Il existe $\gamma > 0$ tel que pour tout $r \in \mathbb{N}^*$, on a quand $n \to +\infty$*

$$u_n = \gamma + \frac{1}{2n} - \sum_{p=1}^{r} \frac{b_{2p}}{2p} \frac{1}{n^{2p}} + \mathcal{O}(\frac{1}{n^{2r+1}}). \tag{3.29}$$

Preuve La formule d'Euler-Mac Laurin appliquée à la fonction $t \mapsto \dfrac{1}{t}$ au rang $2r+1$ avec $m=1$ s'écrit

$$\sum_{k=1}^{n} \frac{1}{k} = \int_1^n \frac{dt}{t} + \frac{1}{2} + \frac{1}{2n} - \sum_{p=1}^{r} \frac{b_{2p}}{2p} \left(\frac{1}{n^{2p}} - 1 \right) - \int_1^n \frac{\tilde{B}_{2r+1}(t)}{t^{2r+2}} dt.$$

Comme la fonction $\tilde{B}_{2r+1}$ est bornée sur $\mathbb{R}$, la fonction $t \mapsto \frac{\tilde{B}_{2r+1}(t)}{t^{2r+2}}$ est intégrable sur $[1, +\infty[$. En notant

$$\gamma_r = \frac{1}{2} + \sum_{p=1}^{r} \frac{b_{2p}}{2p} - \int_1^{+\infty} \frac{\tilde{B}_{2r+1}(t)}{t^{2r+2}} dt,$$

on obtient

$$u_n = \gamma_r + \frac{1}{2n} - \sum_{p=1}^{r} \frac{b_{2p}}{2p} \frac{1}{n^{2p}} + \int_n^{+\infty} \frac{\tilde{B}_{2r+1}(t)}{t^{2r+2}} dt.$$

Comme $\frac{1}{2n} - \sum_{p=1}^{r} \frac{b_{2p}}{2p} \frac{1}{n^{2p}} \to 0$ quand n tend vers $+\infty$ et

$$\lim_{n \to +\infty} \int_n^{+\infty} \frac{\tilde{B}_{2r+1}(t)}{t^{2r+2}} dt = 0,$$

on en déduit que la suite (u_n) converge vers γ_r, et par unicité de la limite, γ_r est indépendant de r. Ainsi, on a établi (3.29).

Une autre application de la formule sommatoire d'Euler Mac Laurin et qui nous conduira vers la méthode de Romberg est la suivante. On considère un entier $N \in \mathbb{N}^*$, $(a,b) \in \mathbb{R}^2$ $(a<b)$ et on pose $h = \frac{b-a}{N}$. Soit $f \in C^\infty([a,b])$. On considère la méthode des trapèzes composites pour approcher l'intégrale de f entre a et b. On pose

$$T_f(h) = \frac{h}{2} \left(f(a) + f(b) \right) + h \sum_{k=1}^{N-1} f(a+kh). \tag{3.30}$$

Théorème 3.42 *Soit $r \in \mathbb{N}^*$. Il existe des réels $(a_i)_{1 \le i \le E(\frac{r}{2})}$ tels que*

$$\int_a^b f(x)dx - T_f(h) = \sum_{i=1}^{E(\frac{r}{2})} a_i h^{2i} + \mathcal{O}(h^r). \tag{3.31}$$

L'erreur admet donc un développement en puissances de h^2.

Preuve Appliquons la formule d'Euler Mac Laurin à la fonction g définie sur $\mathbb{R}$ par $g(u) = f(a+uh)$ entre $m = 0$ et $n = N$. D'après (3.27), on obtient

$$\sum_{k=0}^{N} f(a+kh) = \int_0^N f(a+uh)du + \frac{1}{2}(g(N)+g(0))$$
$$+ \sum_{p=1}^{E(\frac{r}{2})} \frac{b_{2p}}{(2p)!}(g^{(2p-1)}(N) - g^{(2p-1)}(0)) + \frac{(-1)^{r+1}}{r!}\int_0^N \tilde{B}_r(t)g^{(r)}(t)dt,$$

ou encore comme $\int_0^N f(a+uh)du = \frac{1}{h}\int_a^b f(x)dx$ (effectuer le changement de variable $x = a+uh$) et $g^{(n)}(u) = f^{(n)}(a+uh).h^n$

$$\sum_{k=0}^{N} f(a+kh) = \frac{1}{h}\int_a^b f(x)dx + \frac{1}{2}(f(b)+f(a))$$
$$+ \sum_{p=1}^{E(\frac{r}{2})} \frac{b_{2p}}{(2p)!}(f^{(2p-1)}(b) - f^{(2p-1)}(a))h^{2p-1} \qquad (3.32)$$
$$+ \frac{(-1)^{r+1}}{r!}\int_0^N \tilde{B}_r(t)f^{(r)}(a+th)h^r(t)dt.$$

Comme la fonction $\tilde{B}_r$ est bornée sur $[0,N]$ (car 1-périodique) et comme $f^{(r)}$ est bornée sur $[a,b]$ (car continue sur cette intervalle), on obtient l'inégalité :

$$|\frac{(-1)^{r+1}}{r!}\int_0^N \tilde{B}_r(t)f^{(r)}(a+th)h^r dt| \leq \max_{x\in[a,b]}|f^{(r)}(x)| \max_{x\in[a,b]}|\tilde{B}_r(t)||\frac{h^{r-1}}{r!}.$$

donc on a

$$|\frac{(-1)^{r+1}}{r!}\int_0^N \tilde{B}_r(t)f^{(r)}(a+th)h^r dt| = \mathcal{O}(h^{r-1}).$$

On en déduit en multipliant les deux membres de l'égalité (3.32) par h que

$$T_f(h) = \int_a^b f(x)dx + \sum_{p=1}^{E(\frac{r}{2})} a_p h^{2p} + \mathcal{O}(h^r), \qquad (3.33)$$

avec

$$a_p = \frac{b_{2p}}{(2p)!}(f^{(2p-1)}(b) - f^{(2p-1)}(a)),$$

ce qui établit que le reste admet un développement en puissances de h^2.

3.9.3 Description de la méthode de Romberg

D'après l'étude menée sur la méthode composites des trapèzes, et précisément l'estimation de l'erreur obtenue dans (3.7) (en supposant que $b-a$ est de l'ordre de quelques unités), pour assurer une précision de 10^{-6}, il suffit que le pas h soit de l'ordre de 10^{-3}. Un tel calcul nécessitera une boucle comportant au moins plusieurs milliers d'itérations et le calcul sera propice à des propagations d'erreur d'arrondis. La méthode de Romberg permet d'accélérer la vitesse de convergence de de la méthode des trapèzes. La méthode de Romberg repose sur le procédé d'extrapolation de Richardson. L'idée du procédé est de combiner plusieurs développements de Taylor d'une fonction v au voisinage de 0 pour déterminer $v(0)$avec l'erreur la plus faible possible. Exemple : si $v(h) = v(0) + c_1 h + \mathcal{O}(h^2)$, on a aussi pour un réel $r \in]0,1[$, fixé (souvent $r = \frac{1}{2}$) $v(rh) = v(0) + c_1 rh + \mathcal{O}(h^2)$. On a alors

$$\frac{v(rh) - rv(h)}{1-r} = v(0) + \mathcal{O}(h^2).$$

On obtient ainsi une approximation de $v(0)$ à un $\mathcal{O}(h^2)$ près au lieu d'un $\mathcal{O}(h)$ près. Cette combinaison linéaire permet donc d'obtenir une meilleure valeur approchée de $v(0)$.
Le théorème 3.42 permet d'écrire en remplaçant r par $2n+1$ ($n \in I\!N^*$) :

$$T_f(h) = \int_a^b f(t)dt - c_2 h^2 - c_4 h^4 + \cdots - c_{2n} h^{2n} + \mathcal{O}(h^{2n+1}). \tag{3.34}$$

La méthode de Romberg débute par le calcul des approximations intégrales de f pour les pas $\frac{h}{2}$, $\frac{h}{4}$, $\frac{h}{8}$ $\cdots$ que l'on dispose dans une colonne. On observe que l'on passe facilement de $T_f(h)$ à $T_f(\frac{h}{2})$ (où $Nh = b-a$) en ajoutant les images des abscisses intermédiaires situées au milieu des intervalles de subdivisions : $T_f(\frac{h}{2}) = \frac{1}{2}\left(T_f(h) + M_h\right)$ où
$M_h = h\left(f(a+\frac{h}{2}) + f(a+\frac{3h}{2}) + \cdots + f(a+\frac{(2N-1)h}{2})\right)$. Le tableau de la méthode de Romberg se construit à partir de la première colonne dont les éléments sont notés $T_{00} = T_f(h)$, $T_{10} = T_f(\frac{h}{2})$, $\cdots$, $T_{m0} = T_f(\frac{h}{2^m})$ $m \in I\!N^*$. On pose alors pour $n = 1, \cdots, m$

$$T_{n,1} = \frac{4T_{n,0} - T_{n-1,0}}{4-1}.$$

Appliquant cette formule, d'après (3.34), on obtient par exemples

$$T_{11} = \int_a^b f(t)dt + \frac{c_4}{4}h^4 + \frac{5}{16}c_6 h^6 \cdots + \mathcal{O}(h^{2n+1}),$$

$$T_{21} = \int_a^b f(t)dt + \frac{c_4}{4}(\frac{h}{2})^4 + \frac{5}{16}c_6(\frac{h}{2})^6 + \cdots + \mathcal{O}(h^{2n+1}),$$

Ainsi, on obtient une valeur approchée de $\int_a^b f(t)dt$ à un $\mathcal{O}(h^4)$ près dans le premier cas et à un $\mathcal{O}((\frac{h}{2})^4)$ près dans le second cas. On construit de la sorte la deuxième colonne du tableau : elle est constituée des éléments $T_{11}, T_{21}, \cdots, T_{m1}$. Afin d'améliorer la précision de l'approximation, on peut évaluer

$$T_{22} := \frac{2^4 T_{21} - T_{11}}{2^4 - 1} = \int_a^b f(t)dt + \mathcal{O}(h^6).$$

On peut généraliser ce qui précède en introduisant la formule de récurrence pour $k = 0, \cdots, m-1$, $n = k+1, \cdots, m$

$$T_{n,k+1} = \frac{2^{2k+2} T_{n,k} - T_{n-1,k}}{4^{k+1} - 1},$$

qui permet de déterminer la colonne $k+1$ du tableau. La dernière valeur $T_{m,m}$ fournit une valeur approchée de l'intégrale à $\mathcal{O}(h^{2m+2})$ près. En pratique, il est inutile de calculer $T_{m,m}$.

4 Résolution de l'équation $f(x) = 0$

4.1 Introduction

Dans toute la première partie, on considére une fonction f définie sur $[a, b]$ à valeurs réelles, continue sur $[a, b]$ telle que $f(a).f(b) \leq 0$. D'après le théorème des valeurs intermédiaires, on sait que f admet au moins une racine dans $[a, b]$, notée l.
Mis à part quelque cas simple, par l'exemple les équations $ax^2 + bx + c = 0$ et $ax^3 + bx^2 + cx + d = 0$, on ne peut pas résoudre algébriquement l'équation $f(x) = 0$.
En pratique, on cherche donc une solution approchée de la solution l en construisant une suite numérique (u_n) qui converge vers l. On se propose ici de donner plusieurs méthodes de résolution de l'équation $f(x) = 0$. Une méthode déjé abordée en L1 est la méthode de dichotomie, mais cette méthode se révéle peu efficace, car assez lente (la convergence est "géométrique" de raison $\frac{1}{2}$).
Nous présentons ici diverses méthodes de type point fixe, dont la méthode des approximations successives. Cette dernière repose sur le théorème du point fixe. En pratique, on remplace l'équation $f(x) = 0$ par une équation équivalente, par exemple $x = x - f(x)$ et on cherche les points fixes de l'équation

$g(x) = x$ avec $g(x) = x - f(x)$. D'autres méthodes seront présentées, par exemples la méthode de la corde ou encore la méthode de Newton. Nous montrerons que la méthode de Newton se révéle la plus efficace lorsqu'elle converge, la convergence étant quadratique.
Les problémes posés par l'introduction de telles suites numériques sont les suivants :
1. La suite (x_n) converge-t-elle ?
2. Si la suite converge, sa limite est-elle l ?
Si la réponse à l'une de ces questions est non, alors la méthode considérée n'est pas satisfaisante.
Un autre probléme se pose : si on veut calculer la solution à ϵ prés, combien faut-il d'itérations pour y parvenir, et comment arrêter les itérations dés que cette condition est remplie ?

Dans une seconde partie, on envisage d'étudier le cas ou f est définie sur un ouvert d'un espace vectoriel normé complet (éventuellement de dimension infinie) à valeurs dans un espace vectoriel normé Y. On généralisera la méthode de Newton étudié dans le cas de la dimension un.

4.2 La méthode de dichotomie

On considère une fonction f définie sur $[a, b]$ à valeurs réelles, continue sur $[a, b]$. Soit (u_n) une suite de $[a, b]$ convergeant vers l. On rappelle (voir cours de topologie) que d'une part, $l \in [a, b]$ et que d'autre part, la continuité de f entraîne que

$$\lim_{n\to+\infty} f(u_n) = f(\lim_{n\to+\infty} u_n).$$

On suppose dans cette sous-section ainsi que dans la suivante que la fonction f posséde une **unique** racine notée l dans l'intervalle $[a, b]$. La méthode de dichotomie consiste à introduire à chaque étape le milieu du segment $[a, b]$, $c = \dfrac{a+b}{2}$, puis à déterminer l'intervalle contenant la racine de f en ayant recours au théorème des valeurs intermédiaires. L'algorithme est donc le suivant : on pose

$$a_0 = a,\ b_0 = b \quad \text{et} \quad c_0 = \frac{a+b}{2}.$$

Si $f(a_0).f(c_0) \leq 0$, alors d'après le théorème des valeurs intermédiaires $l \in [a_0, c_0]$, et on pose $a_1 = a_0$ et $b_1 = c_0$, sinon, $l \in [c_0, b_0]$ et on pose $a_1 = c_0$ et $b_1 = b_0$.
Pour $n \geq 1$, supposons déterminé les réels $a_0, \cdots, a_{n-1}$, $b_0, \cdots, b_{n-1}$.

À l'étape n, on pose

$$c_{n-1} = \frac{a_{n-1} + b_{n-1}}{2}.$$

Si $f(a_{n-1}).f(c_{n-1}) \leq 0$, alors $l \in [a_{n-1}, c_{n-1}]$, et on pose $a_n = a_{n-1}$ et $b_n = c_{n-1}$, sinon, $l \in [c_{n-1}, b_{n-1}]$ et on pose $a_n = c_{n-1}$ et $b_n = b_{n-1}$.

Montrons que la méthode de dichotomie converge.

Théorème 4.1 *Les suites de réels (a_n) et (b_n) sont adjacentes, elles convergent vers la solution de l'équation $f(x) = 0$. De plus, la convergence est "géométrique". Précisément, on a*

$$|a_n - l| \leq \frac{|a-b|}{2^n} \quad \forall\, n \in I\!N \quad et \quad |b_n - l| \leq \frac{|a-b|}{2^n} \quad \forall\, n \in I\!N.$$

Preuve
Supposons que f est strictement négative sur $[a, l]$ et strictement positive sur $[l, b]$. Par récurrence, on montre que

$$a_n \leq b_n \quad \forall\, n \in I\!N. \tag{4.1}$$

En effet, on a $a_0 < b_0$, et si on suppose $a_{n-1} \leq b_{n-1}$ pour $n \geq 1$. On obtient (si $l \in [a_{n-1}, c_{n-1}]$))

$$a_n - b_n = a_{n-1} - \frac{a_{n-1} + b_{n-1}}{2} = \frac{a_{n-1} - b_{n-1}}{2} < 0,$$

ou (si $l \in [c_{n-1}, b_{n-1}]$)

$$a_n - b_n = \frac{a_{n-1} + b_{n-1}}{2} - b_{n-1} = \frac{a_{n-1} - b_{n-1}}{2} < 0.$$

Donc $a_n \leq b_n$ pour tout $n \in I\!N$. Par ailleurs, on a établi que

$$a_n - b_n = \frac{a_{n-1} - b_{n-1}}{2}, \quad \forall\, n \in I\!N.$$

Par récurrence, on en déduit que

$$a_n - b_n = \frac{a-b}{2^n} \quad \forall\, n \in I\!N. \tag{4.2}$$

De plus, d'après (4.1) et (4.2), on en déduit que les suites (a_n) et (b_n) convergent vers la même limite notée $\hat{l}$. En effet, la suite (a_n) est croissante (et (b_n) est décroissante) puisque par définition de (a_n), soit

$$a_{n+1} - a_n = 0,$$

soit

$$a_{n+1} - a_n = \frac{a_n + b_n}{2} - a_n = \frac{b_n - a_n}{2}.$$

De (4.1), on déduit que (a_n) est croissante, majorée par b_0 (respectivement, (b_n) est décroissante et minorée par a_0). Par conséquent, elles convergent toutes deux, et compte tenu de (4.2), elles convergent vers la même limite.
De plus, comme $f(a_n) \leq 0$ pour tout n, on obtient par passage à la limite $\lim_{n\to+\infty} f(u_n) = f(\hat{l}) \leq 0$. De même, comme $f(b_n) \geq 0$, par passage à la limite, on obtient $f(\hat{l}) \geq 0$. Donc $f(\hat{l}) = 0$, et comme f admet une unique racine dans $[a, b]$, on a $l = \hat{l}$. D'autre part, comme $l \in [a_n, b_n]$, d'après (4.2), on déduit que

$$|a_n - b_n| = |a_n - l| + |b_n - l| \leq \frac{|a-b|}{2^n} \quad \forall\, n \in I\!N.$$

Donc on a $|a_n - l| \leq \frac{|a-b|}{2^n} \quad \forall\ n \in I\!N$ et $|b_n - l| \leq \frac{|a-b|}{2^n} \quad \forall\ n \in I\!N$. La convergence de (a_n) et (b_n) vers l est donc "géométrique" (la convergence est de l'ordre de $\frac{1}{2^n}$). La preuve du théorème 4.1 est achevée.

4.3 La méthode des approximations successives

Dans cette partie ainsi que dans ce chapitre, on sera amené à utiliser les lemmes suivants

Lemme 4.2 *On considère une suite de nombre réels (u_n) satisfaisant la condition suivante : il existe $k \in]0, 1[$ tel que*

$$|u_{n+1}| \leq k|u_n|, \quad \forall\, n \in I\!N. \tag{4.3}$$

Alors la suite (u_n) converge vers 0 et la vitesse de convergence est géométrique.

Preuve Montrons par récurrence que

$$|u_n| \leq k^n|u_0|, \quad \forall\, n \in I\!N.$$

Le résultat est vrai pour $n = 0$, puisque $|u_0| \leq k^0|u_0|$.
Supposons le résultat est vrai au rang n. On a alors d'après (4.3) et par hypothèse de récurrence :

$$|u_{n+1}| \leq k|u_n| \leq k.k^n|u_0|.$$

Le résultat est donc vrai pour tout $n \in I\!N$. Comme $k \in]0, 1[$, on en déduit que k^n tend vers 0 quand n tend vers l'infini, et il en résulte que (u_n) tend vers 0 et la vitesse de convergence est géométrique.

Lemme 4.3 *On considère une suite de nombre réels (u_n) positifs satisfaisant la condition suivante : il existe $C > 0$ tel que*

$$u_{n+1} \leq C|u_n|^2, \quad \forall\, n \in I\!N. \tag{4.4}$$

Alors on a

$$u_n \leq C^{2^k-1} u_0^{2^k}.$$

Preuve Effectuons un raisonnement par récurrence sur n. Le résultat est vrai pour $n = 0$. Supposons le vrai au rang n ($n \geq 0$). D'après (4.4) et l'hypothèse de récurrence, on obtient

$$u_{n+1} \leq C u_n^2 \leq C \left(C^{2^n-1} u_0^{2^n}\right)^2 = C^{2^{n+1}-1} u_0^{2^{n+1}}.$$

Le résultat est donc vrai pour tout $n \in I\!N$.

4.3.1 Le théorème du point fixe

Définition 4.4 *Soit f une fonction définie sur $[a, b]$ à valeurs réelles. On dit que $\alpha \in [a, b]$ est un point fixe de f si $f(\alpha) = \alpha$.*

Définition 4.5 *Soit f une fonction définie sur $[a, b]$ à valeurs réelles. On dit que f est une fonction contractante si il existe un réel $k \in]0, 1[$ tel que*

$$|f(x) - f(y)| \leq k|x - y| \quad \forall\, (x, y) \in [a, b]^2. \tag{4.5}$$

On rappelle le résultat suivant établi dans le cours de topologie.
Soit (u_n) une suite d'un espace vectoriel normé E, et F une partie fermée de E. Si $u_n \in F$ pour tout $n \in I\!N$ et si (u_n) converge vers $l \in E$, alors $l \in F$.

Le théorème du point fixe joue une très grand rôle en analyse. La méthode des approximations successives présentées ici repose sur ce théorème dont on donne l'énoncé et la démonstration.

Théorème 4.6 *(théorème du point fixe)*
Soit f une fonction définie sur $[a, b]$ à valeurs réelles satisfaisant les deux conditions suivantes :

- *$f([a, b]) \subset [a, b]$ (on dit que $[a, b]$ est stable par f).*
- *f est une fonction contractante sur $[a, b]$ au sens de la définition 4.5.*

Alors il existe un unique $\alpha \in [a, b]$ tel que $f(\alpha) = \alpha$. De plus, la suite (u_n) définie par $u_{n+1} = f(u_n)$, $u_0 \in [a, b]$ converge vers α.

Preuve
S'il existe un point fixe de f alors il est unique. En effet, soient x_1 et x_2 deux points fixes de f tels que $x_1 \neq x_2$. Alors, comme f est contractante, on a

$$|f(x_1) - f(x_2)| \leq k|x_1 - x_2|,$$

donc

$$k \geq \frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} = 1.$$

Contradiction ($k \in]0,1[$).
Pour montrer l'existence du point fixe, nous allons montrer que la suite (u_n) définie par $u_{n+1} = f(u_n)$ est une suite de Cauchy. Remarquons que compte-tenu de (4.5)

$$|u_{n+1} - u_n| = |f(u_n) - f(u_{n-1})| \leq k|u_n - u_{n-1}| \ \ \forall\, n \geq 1.$$

D'après le lemme 4.2 appliqué à la suite $(|u_{n+1} - u_n|)$, on obtient

$$|u_{n+1} - u_n| \leq k^n |u_1 - u_0|, \quad \forall\, n \in I\!N.$$

Par inégalité triangulaire, on en déduit que pour $p \in I\!N$ et $n \in I\!N$, on a

$$|u_{n+p} - u_n| \leq \sum_{i=n}^{n+p-1} |u_{i+1} - u_i| \leq \sum_{i=n}^{n+p-1} k^i |u_1 - u_0|.$$

Comme la série $\sum k^n$ converge (car $0 < k < 1$), on en déduit que (u_n) est une suite de Cauchy dans $I\!R$, espace complet. La suite (u_n) converge vers α, et comme $[a,b]$ est fermé, on a $\alpha \in [a,b]$.
Comme f est continue, on a

$$\lim_{n\to\infty} u_{n+1} = \alpha = \lim_{n\to\infty} f(u_n) = f(\lim_{n\to\infty} u_n) = f(\alpha).$$

On obtient que α est un point fixe de f. Ceci achève la preuve du théorème 4.6.

Remarque 4.7 *On peut établir un résultat analogue dans un cadre beaucoup plus général que celui donné dans le théorème 4.6, par exemple dans le cadre des espaces métriques complets.*
Ici, soit E un espace vectoriel normé complet et ϕ une application définie sur E à valeurs dans E (E est un espace métrique pour la distance $d(x,y) = \|x-y\|$). On suppose que ϕ est contractante, c'est-à-dire qu'il existe $k \in]0,1[$ tel que

$$\|\phi(x) - \phi(y)\|_E \leq k\|x - y\|_E \quad \forall\, (x,y) \in E^2.$$

Alors il existe un unique $\alpha \in E$ tel que $\phi(\alpha) = \alpha$. De plus, la suite (u_n) définie par $u_{n+1} = \phi(u_n)$, $u_0 \in E$ converge vers α.
La démonstration de ce résultat très important est identique à celle donnée dans le théorème 4.6. Pour l'obtenir, il suffit de remplacer la norme sur $\mathbb{R}$ par celle sur E.

Etudions le cas particuliers où f est de classe C^1 sur $[a, b]$.

Proposition 4.8 *Soit $f \in C^1([a, b])$. On suppose que*

$$k := \max_{x\in[a,b]} |f'(x)| < 1.$$

Alors f est contractante sur $[a, b]$.

Preuve D'après le théorème des accroissements finis appliqué entre x et y $(x, y) \in [a, b]^2$, on a

$$|f(y) - f(x)| = |f'(c)|.|y - x|, \quad c \in]a, b[.$$

Compte-tenu de l'hypothése, on en déduit que

$$|f(y) - f(x)| \leq k|y - x| \ \ \forall\, (x, y) \in [a, b]^2.$$

4.3.2 Résolution de l'équation $f(x) = 0$ par la méthode du point fixe

Exemple Considérons l'équation $x^3 - x^2 - 1 = 0$. Une étude de la fonction $f(x) = x^3 - x^2 - 1$ permet de montrer que cette équation admet une unique racine dans $[1, 2]$. En effet, la dérivée de f est donnée par $f'(x) = 3x^2 - 2x$ et f' est de signe strictement positif sur $[1, 2]$, donc la fonction f croît strictement sur $[1, 2]$. Comme $f(1).f(2) = -3 < 0$, on déduit du théorème des valeurs intermédiaires qu'il existe un unique élément $l \in [1, 2]$ tel $f(l) = 0$. Transformons l'équation de telle sorte à l'écrire comme un problème de point fixe. Plusieurs transformations sont possibles. On peut par exemple écrire l'équation sous la forme

$$x = x^3 - x^2 + x - 1$$

ou encore

$$x = (x^2 + 1)^{\frac{1}{3}}.$$

Étudions les solutions de l'équation $x = (x^2 + 1)^{\frac{1}{3}}$. Posons $g(x) = (x^2 + 1)^{\frac{1}{3}}$ et montrons que le théorème du point fixe 4.6 s'applique sur $[1, 2]$ à g. La fonction g est croissante sur $[1, 2]$, on a donc

$$g([1, 2]) = [g(1), g(2)].$$

Or, $g(1) = 2^{\frac{1}{3}} \in [1,2]$ et $g(2) = 5^{\frac{1}{3}} \in [1,2]$, donc la premiére hypothése du théorème 4.6 est bien satisfaite.
D'autre part, on a

$$g'(x) = \frac{2}{3}x(x^2+1)^{-\frac{2}{3}}.$$

Pour tout $x \in [1,2]$, on a

$$|g'(x)| \leq \frac{2}{3}22^{-\frac{2}{3}} < 1 \quad \forall\, x \in [1,2].$$

D'après le théorème 4.6 et la proposition 4.8, on déduit que la suite (x_n) définie par $x_0 \in [1,2]$ et $x_{n+1} = g(x_n)$ converge vers l'unique point fixe de g dans $[1,2]$. Ce point fixe n'est autre que la solution de l'équation $f(x) = 0$.

Remarque 4.9 *Un intérêt du théorème 4.6 est que la convergence est assurée pour un choix quelconque de x_0 dans $[a,b]$. Il n'en va pas de même dans la méthode de Newton que nous aborderons ultérieurement, qui peut s'avérer divergente si x_0 est choisi trop loin du point fixe.*

Soit $\epsilon > 0$. On peut se demander combien d'itérations sont nécessaires pour obtenir une estimation de l'erreur $e_n := |x_n - l|$ plus petite que ϵ. C'est l'objet de la proposition :

Proposition 4.10 *Soit la suite (u_n) définie dans le théorème 4.6. Alors le nombre d'itérations nécessaires pour que $|u_n - l| \leq \epsilon$ est donné par*

$$n \geq \frac{\ln(\epsilon) - \ln|u_0 - l|}{\ln k}.$$

Preuve
On a

$$|u_n - l| = |f(u_{n-1}) - f(l)| \leq k|u_{n-1} - l|.$$

On en déduit par récurrence que

$$|u_n - l| \leq k^n|u_0 - l|, \quad \forall\, n \in I\!N.$$

Donc une condition suffisante pour obtenir que $|u_n - l| \leq \epsilon$ est que

$$k^n|u_0 - l| \leq \epsilon.$$

Cette condition équivaut à

$$n\ln k + \ln|u_0 - l| \leq \ln(\epsilon),$$

soit

$$n \geq \frac{\ln(\epsilon) - \ln|u_0 - l|}{\ln k}.$$

Définition 4.11 *Soit f une fonction de classe $C^1([a,b])$ et α un point fixe de f. On dit que α est un point attractif si*

$$|f'(\alpha)| < 1.$$

On dit que α est répulsif si

$$|f'(\alpha)| > 1.$$

L'une des difficultés pour appliquer le théorème du point fixe réside dans la détermination d'un intervalle stable par f. La proposition suivante donne une réponse à cette interrogation :

Proposition 4.12 *Soit $g : [a,b] \to I\!R$ une fonction de classe C^1 sur $[a,b]$. Soit $l \in [a,b]$ une point fixe de g. On suppose que*

$$|g'(l)| < 1.$$

Alors il existe un intervalle $[\alpha,\beta] \subset [a,b]$ contenant l pour lequel la suite définie par $x_0 \in [a,b]$ et $x_{n+1} = g(x_n)$ converge vers l.

Preuve
On suppose que $0 < g'(l) < 1$. Comme g' est continue au point l, il existe un intervalle $[\alpha,\beta]$ contenant l tel que

$$0 < g'(x) < 1 \quad \forall\, x \in [\alpha,\beta].$$

En effet, par continuité de g' au point $x = l$, on a :

$$\forall\, \epsilon > 0,\ \ \exists\, \eta > 0\ |x - l| \leq \eta,\ \ |g'(x) - g'(l)| < \epsilon.$$

On choisit alors $\epsilon > 0$ assez petit pour que

$$0 < g'(l) - \epsilon < g'(x) < g'(l) + \epsilon < 1,\ \ \forall\, x \in [l - \eta, l + \eta].$$

On pose alors $\alpha = l - \eta$ et $\beta = l + \eta$. Reste à montrer que $[\alpha,\beta]$ est stable par g. On a d'après le théorème des accroissements finis appliqué entre α et l

$$l - g(\alpha) = g(l) - g(\alpha) = g'(\gamma)(l - \alpha), \quad \gamma \in]\alpha, l[.$$

Comme $g'(\gamma) < 1$, on a

$$l - g(\alpha) \leq (l - \alpha),$$

soit $\alpha \leq g(\alpha)$. On montre de même que $\beta \geq g(\beta)$. Comme g est croissante sur $[\alpha,\beta]$, on a bien $g([\alpha,\beta]) \subset [\alpha,\beta]$. D'autre part, pour tout $x \in [\alpha,\beta]$, on a $|g'(x)| < 1$. D'après la proposition 4.8, la fonction g est donc contractante

sur cet intervalle. On peut alors appliquer le théorème du point fixe à g restreint à $[\alpha, \beta]$ et obtenir ainsi la conclusion désirée.
Le cas $-1 < g'(l) < 0$ se traite de maniére analogue.

Etudions à présent le cas $|g'(l)| > 1$. On a la proposition :

Proposition 4.13 *Soit l une solution de l'équation $g(x) = x$. Si g' est continue au voisinage de l et si $|g'(l)| > 1$, alors la suite définie par $x_{n+1} = g(x_n)$, $x_0 \neq l$ ne converge pas vers l.*

Preuve Supposons $x_0 \neq l$. Soit $[\alpha, \beta]$ un intervalle contenant l et tel que

$$k := \min_{x \in [\alpha, \beta]} |g'(x)| > 1.$$

Un tel intervalle existe puisque g' est continue au point l et $|g'(l)| > 1$. Soit $n \in \mathbb{N}$. Alors, seules deux éventualités sont possibles :
ou bien $x_n \notin [\alpha, \beta]$ ou $x_n \in [\alpha, \beta]$ et alors d'après le théorème des accroissements finis appliqué à g entre x_n et l, on obtient

$$|g(x_n) - g(l)| = |g'(\eta)(x_n - l)|, \quad \eta \in]x_n, l[\cup]l, x_n[.$$

soit

$$|x_{n+1} - l| \geq k|x_n - l|.$$

On en déduit que l'erreur $e_n = |x_n - l|$ ne peut pas tendre vers 0. En effet, ou bien il existe une infinité d'entiers n tels que $x_n \notin [\alpha, \beta]$ et par conséquent, (x_n) ne converge pas vers l, ou alors il existe un entier n_0 tel que pour tout $n \geq n_0$, on a

$$|x_{n+1} - l| \geq k|x_n - l|,$$

et dans ce cas, on déduit par récurrence que

$$|x_n - l| \geq k^{n-n_0}|x_{n_0} - l|,$$

et par conséquent, la suite (x_n) ne converge pas vers l puisque k^{n-n_0} tend vers $+\infty$ quand n tend vers l'infini.

Récapitulons la marche à suivre afin d'étudier des suites de la forme $u_{n+1} = g(u_n)$ (g dérivable) en utilisant la méthode du point fixe. Soit l un point fixe de g.

- Si $|g'(l)| > 1$, ou on élimine la méthode ou on peut travailler avec g^{-1} puisque
$$(g^{-1})'(l) = \frac{1}{g'(g^{-1}(l))} = \frac{1}{g'(l)} < 1.$$
- Si $|g'(l)| < 1$, il faut trouver un intervalle $[a, b]$ stable par la fonction g.
- Si $|g'(l)| = 1$, on peut avoir convergence ou divergence.

4.4 La méthode de la corde

4.4.1 Fonctions convexes

On rappelle quelques résultats concernant les fonctions convexes.

Définition 4.14 *Soit $f : I \to \mathbb{R}$. On dit que f est convexe sur I si pour tout $(x, y) \in I^2$, pour tout $t \in [0, 1]$, on a*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y). \tag{4.6}$$

On dit que f est strictement convexe sur I si pour tout $(x, y) \in I^2$, pour tout $t \in]0, 1[$, on a

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y).$$

Lorsqu'on inverse le sens des inégalités précédentes, on dit que f est concave sur I.

De la définition 4.14, on déduit la proposition :

Proposition 4.15 *Soit f une fonction convexe définie sur $[a, b]$, et $s, t, u \in [a, b]$ tels que $s < t < u$. Alors on a :*

$$\frac{f(t) - f(s)}{t - s} \leq \frac{f(u) - f(t)}{u - t}. \tag{4.7}$$

Si on suppose que f est dérivable sur I, on peut caractériser la convexité gràce à la dérivée première.

Théorème 4.16 *Soit $f : I \to \mathbb{R}$ une fonction dérivable sur I. La fonction f est convexe si et seulement si*

$$f(y) \geq f(x) + f'(x)(y - x), \quad \forall\, (x, y) \in I^2. \tag{4.8}$$

La fonction f est strictement convexe si et seulement si

$$f(y) > f(x) + f'(x)(y - x), \quad \forall\, (x, y) \in I^2. \tag{4.9}$$

La fonction f est concave si et seulement si les inégalités dans (4.8) et (4.9) sont inversées.

Démonstration Si f est convexe, on peut écrire

$$f(x + t(y - x)) \leq (1-t)f(x) + tf(y),$$

soit, pour $t \neq 0$

$$\frac{f(x+t(y-x))-f(x)}{t} \leq f(y)-f(x),$$

inégalité que l'on peut réecrire sous la forme

$$\frac{f(x+t(y-x))-f(x)}{t(y-x)}(y-x) \leq f(y)-f(x). \tag{4.10}$$

Or, comme f est dérivable au point x, on a

$$\lim_{t\to 0} \frac{f(x+t(y-x))-f(x)}{t(y-x)} = f'(x).$$

Faisant tendre t vers 0 dans l'inégalité (4.10), on obtient

$$\lim_{t\to 0} \frac{f(x+t(y-x))-f(x)}{t(y-x)}(y-x) \leq f(y)-f(x),$$

d'où

$$f'(x)(y-x) \leq f(y)-f(x).$$

Réciproquement, supposons (4.8) satisfaite. Alors, remplaéant x par $y + t(x-y)$ dans (4.8), on obtient pour $t \in]0,1[$

$$f(y) \geq f(y+t(x-y)) - tf'(y+t(x-y))(x-y).$$

De même, remplaéant y par x et x par $y+t(x-y)$ dans (4.8), on obtient

$$f(x) \geq f(y+t(x-y)) + (1-t)f'(y+t(x-y))(x-y),$$

et il suffit d'additionner les deux inégalités ci-dessus, multipliées respectivement par $(1-t)$ et t pour obtenir (4.18).
On admettra (4.9).

Remarque 4.17 *Le théorème 4.16 exprime que la courbe représentative de f est au-dessus de la tangente au point d'abscisse x_0, x_0 quelconque appartenant à I.*

Exemple d'application Considérons la fonction $f(x) = \ln x$ pour $x > 0$. D'après le théorème 4.16, pour établir la concavité de f sur $]0,+\infty[$, il suffit de montrer que

$$\frac{1}{x}(y-x) \geq \ln y - \ln x, \quad \forall\, (x,y) \in]0,+\infty[^2,$$

soit encore

$$\ln(\frac{y}{x}) \leq \frac{y}{x} - 1 \ \ \forall\, (x, y) \in]0, +\infty[^2.$$

On considère la fonction auxiliaire g définie par

$$g(u) = \ln(\frac{u}{x}) - \frac{u}{x} + 1, \quad \forall\, u > 0.$$

On a $g'(u) = \frac{1}{u} - \frac{1}{x}$. g admet un unique maximum atteint en $u = x$, et $g(u) = 0$. Donc $g(u) \leq 0$ pour tout $u > 0$. Il en résulte que f est concave.

On admettra le théorème suivant qui sera établi dans le chapitre consacré aux développements limités.

Théorème 4.18 *Soit f une fonction deux fois dérivable sur I. Alors f est convexe sur I si et seulement si*

$$f''(x) \geq 0, \quad \forall\, x \in I.$$

Si

$$f''(x) > 0, \quad \forall\, x \in I$$

alors f est strictement convexe sur I.

Exemples 1. La fonction exponentielle est dérivable sur $\mathbb{R}$, de dérivée seconde égale à e^x. Cette fonction est donc strictement convexe sur $\mathbb{R}$.
2. Les fonctions de la forme ax^2+bx+c avec $a > 0$ sont strictement convexes sur $\mathbb{R}$. En effet, on a $(ax^2 + bx + c)^{(2)} = 2a > 0$.
3. On considère la fonction définie sur $\mathbb{R}$ par $f(x) = x^2 + 2\sin x$. Montrons que f est convexe sur $\mathbb{R}$. La fonction f est deux fois dérivable sur $\mathbb{R}$ et on a pour tout $x \in \mathbb{R}$ l'égalité $f''(x) = 2 - 2\cos x$. Comme $f''(x) \geq 0$ sur $\mathbb{R}$, on déduit du théorème 4.18 que f est convexe sur $\mathbb{R}$.

De la convexité de f sur $\mathbb{R}$, on peut déduire l'inégalité

$$\sin x \geq x - \frac{x^2}{2}, \quad \forall\, x \in \mathbb{R}. \tag{4.11}$$

En effet, la tangente (T) à la courbe représentative de f au point $x = 0$ est donnée par $y = 2x$. Or, d'après le théorème 4.16, la courbe représentative de f est au-dessus de la tangente (T). On en déduit que

$$x^2 + 2\sin x \geq 2x,$$

soit (4.11).

4.4.2 Convergence de la méthode de la corde

Soit f une fonction de classe C^2 sur $[a,b]$, convexe et strictement croissante sur $[a,b]$ tel que $f(a).f(b) < 0$. On considère le segment dont les extrémités sont les points $A(a, f(a))$ et $B(b, f(b))$. Ce segment coupe l'axe des abscisses au point de coordonnée $(x_1, 0)$. Calculons x_1. L'équation de la droite (AB) est donnée par

$$y = \tau(x - a) + f(a),$$

avec $\tau = \dfrac{f(b) - f(a)}{b - a}$. x_1 satisfait

$$0 = \tau(x_1 - a) + f(a),$$

donc $x_1 = a - \dfrac{f(a)}{\tau}$. On peut alors construire une suite par récurrence en procédant de la faéon suivante : étant donné x_n, le terme x_{n+1} est l'intersection de la droite passant par $A_n(x_n, f(x_n))$ et le point $B(b, f(b))$. La suite récurrente ainsi définie est

$$\begin{cases} x_0 = a \\ x_{n+1} = x_n - \dfrac{f(x_n)}{\tau_n} \end{cases} \tag{4.12}$$

où

$$\tau_n = \frac{f(b) - f(x_n)}{b - x_n}.$$

On peut montrer le théorème

Théorème 4.19 *Soit $f \in C^2([a,b])$, convexe et strictement croissante sur $[a,b]$ tel que $f(a).f(b) < 0$. Alors la suite (x_n) converge vers α, unique zéro de f dans $[a,b]$, et la convergence est géométrique.*

Preuve
Etape 1 La suite (x_n) converge vers α.
Montrons par récurrence sur k que $x_k \leq \alpha$ pour tout k.
Le résultat est vrai pour $k = 0$ puisque $x_0 = a$. Supposons $x_k \leq \alpha$. On a

$$x_{k+1} - \alpha = x_k - \alpha - \frac{f(x_k)(b - x_k)}{f(b) - f(x_k)} = \frac{f(b)(x_k - \alpha) + f(x_k)(\alpha - b)}{f(b) - f(x_k)}. \tag{4.13}$$

Comme f est convexe et $x_k \leq \alpha < b$, on a d'après la proposition 4.15

$$\frac{f(\alpha) - f(x_k)}{\alpha - x_k} \leq \frac{f(b) - f(\alpha)}{b - \alpha}.$$

De cette inégalité, et comme $f(\alpha) = 0$, on déduit que

$$f(b)(x_k - \alpha) + f(x_k)(\alpha - b) \leq 0.$$

Donc d'après (4.13), on déduit que $x_{k+1} \leq \alpha$. Par récurrence, on a donc $x_k \leq \alpha$ pour tout k. Comme f est croissante et α est un zéro de f, on en déduit que

$$f(x_k) \leq 0, \quad \forall\, k \in I\!N.$$

Il résulte de (4.12) que (x_n) est croissante et majorée par α, donc elle converge vers l. On en déduit que $\lim_{n\to+\infty} \tau_k$ existe et vaut $\frac{f(b)-f(l)}{b-l} \neq 0$.
Passons à la limite dans (4.12). On obtient

$$l = l - \frac{f(l)}{\lim_{n\to+\infty} \tau_k},$$

donc $f(l) = 0$, et comme f admet pour unique racine α, on a $l = \alpha$.
Etape 2 Estimation de l'erreur $\epsilon_k := \alpha - x_k$.

Compte-tenu de l'étape 1, on a $\epsilon_k \geq 0$ pour tout $k \geq 0$. Par ailleurs, on a d'après (4.12)

$$\epsilon_{k+1} = \epsilon_k + \frac{f(x_k)}{\tau_k}. \tag{4.14}$$

D'autre part, la suite (τ_k) est bornée puisqu'elle converge et elle est composée de termes positifs. Il existe $M > 0$ tel que

$$0 \leq \tau_k \leq M, \quad \forall\, k \geq 0. \tag{4.15}$$

D'après le théorème des accroissements finis entre x_k et α, on a

$$f(x_k) - f(\alpha) = (x_k - \alpha) f'(\theta_k), \quad \theta_k \in]x_k, \alpha[.$$

Puisque f' est croissante (car f est convexe et deux fois dérivable)), il en résulte que

$$-\frac{f(x_k)}{\epsilon_k} = f'(\theta_k) \in [f'(a), f'(\alpha)] \tag{4.16}$$

De (4.14) et (4.16), on déduit que

$$\epsilon_{k+1} = \epsilon_k + \frac{f(x_k)}{\tau_k \epsilon_k}\epsilon_k = \epsilon_k(1 - \frac{f'(\theta_k)}{\tau_k}).$$

Minorons $\dfrac{f'(\theta_k)}{\tau_k}$. Comme f est convexe sur $[a, b]$ et de classe C^2, la fonction f' est croissante. D'après (4.16) et (4.15), on obtient :

$$\frac{f'(\theta)}{\tau_k} \geq \frac{f'(a)}{M}, \quad \forall\, k \geq 0.$$

On déduit que pour tout $k \geq 0$, on a

$$0 \leq \epsilon_{k+1} \leq (1 - \frac{f'(a)}{M})\epsilon_k.$$

D'après le lemme 4.2, on déduit que

$$0 \leq \epsilon_k \leq (1 - \frac{f'(a)}{M})^k \epsilon_0, \quad \forall\, k \in I\!N. \tag{4.17}$$

Donc la convergence de (x_k) vers α est du même ordre que celle de $(1 - \frac{f'(a)}{M})^k$. La convergence est géométrique, et la preuve du théorème est achevée.

4.5 La méthode de Newton

4.5.1 Description et convergence de la méthode

Soit f une fonction définie sur $[a, b]$ à valeurs réels, de classe C^1 et convexe sur $[a, b]$ admettant une racine l. Soit $x_0 \in [a, b]$. On considère la tangente (T) à la courbe représentative de f au point d'abscisse x_0. Son équation est donnée par

$$y = f(x_0) + f'(x_0)(x - x_0).$$

Elle coupe l'axe des abscisses au point x_1. Le point de coordonnée $(x_1, 0)$ appartient à (T), on en déduit que

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

On peut alors considèrer la tangente à la courbe représentative de f au point x_1 et raisonner comme précédement. Réitirant ce procédé, on construit une suite numérique (x_n) dont on peut penser qu'elle converge vers l.
La suite générée ici est donnée par

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{4.18}$$

Si (x_n) converge vers l et si $f'(l) \neq 0$, on alors par passage à la limite

$$l = l - \frac{f(l)}{f'(l)},$$

donc $f(l) = 0$.
La méthode de Newton est convergente si x_0 est choisi assez proche de l. C'est ce qu'exprime le théorème suivant :

Théorème 4.20 *Soit f de classe C^2 sur $[a,b]$. On suppose qu'il existe $l \in]a,b[$ tel que $f(l)=0$ et $f'(l) \neq 0$.*
Alors si $|x_0 - l|$ est assez petit, la suite (x_n) est bien définie et converge vers l. De plus, il existe une constante $C>0$ telle que pour tout n

$$|x_n - l| \leq \frac{1}{C}(C|x_0 - l|)^{2^n}. \tag{4.19}$$

Preuve
Comme $f'(l) \neq 0$, par continuité de f' au point l, on déduit qu'il existe $\eta > 0$ tel que $f'(x) \neq 0$ sur $J :=]-\eta + l, l + \eta[$. En effet,

$$\forall\, \epsilon > 0, \quad \exists \eta > 0, \quad |x - l| \leq \eta, \quad |f'(x) - f'(l)| < \epsilon.$$

Il en résulte que pour tout $x \in J :=]-\eta + l, l + \eta[$, on a

$$-\epsilon + f'(l) < f'(x) < \epsilon + f'(l).$$

Supposons $f'(l) > 0$. Il suffit alors de choisir ϵ assez petit de telle sorte que $-\epsilon + f'(l) > 0$ pour obtenir la stricte positivité de f' sur J. On raisonne de manière analogue pour traiter le cas $f'(l) < 0$.
Quitte à travailler avec $-f$, on peut supposer que $f'(x) > 0$ sur J. Posons

$$\phi(x) = x - \frac{f(x)}{f'(x)}.$$

On a l'égalité

$$\phi(x) - l = x - l - \frac{f(x) - f(l)}{f'(x)}.$$

D'après la formule de Taylor-Lagrange appliquée à f entre x et l, on obtient

$$f(l) - f(x) = f'(x)(l - x) + \frac{f''(\eta_x)(l - x)^2}{2}, \quad \eta_x \in]l, x[\cup]x, l[.$$

Il en résulte que

$$|\phi(x) - l| = \frac{|f''(\eta_x)|}{2|f'(x)|}|x - l|^2$$

puis que

$$|\phi(x) - l| \leq \frac{\max_{x\in[a,b]} |f''(x)|}{2|f'(x)|}|x - l|^2 \leq C|x - l|^2 \tag{4.20}$$

avec $C = \max\limits_{x\in[a,b]} |f''(x)|\}2 \min\limits_{x\in J} f'(x)$. Quitte à réduire η, on peut supposer $\eta < \frac{1}{C}$. Alors si $x \in J$, on a $\phi(x) \in J$ puisque $|\phi(x) - l| \leq C\eta^2 < \eta$. On a donc établi que pour $x_0 \in J$, la suite $x_{n+1} = \phi(x_n)$ est bien définie (et $x_n \in J$

pour tout n).
Posons $\epsilon_k = |x_k - l|$. D'après (4.20), il existe une constante $C > 0$ telle que

$$\epsilon_{k+1} \leq C\epsilon_k^2 \quad \forall\, k \in I\!N. \tag{4.21}$$

D'après le lemme 4.3, on déduit que

$$\epsilon_k \leq C^{2^k-1}\epsilon_0^{2^k} \quad \forall\, k \geq 0.$$

On en déduit que si

$$C|x_0 - l| < 1,$$

alors (x_n) converge vers l (puisqu'alors $(C|x_0-l|)^{2^n}$ tend vers 0 quand n tend vers l'infini), et de plus, l'inégalité (4.19) est établie. Ainsi, pour assurer la convergence de la méthode, il est nécessaire que la donnée initiale soit assez proche de l.

Remarque 4.21 *La méthode de Newton ne converge pas nécessairement vers la solution de l'équation $f(x) = 0$, comme l'indique le théorème 4.20. En effet, si x_0 est choisi trop loin de la solution de l'équation, la méthode peut diverger. En pratique, on peut appliquer la méthode de dichotomie afin de s'approcher de la solution de l'équation, puis mettre en oeuvre la méthode de Newton qui converge beaucoup plus vite vers la solution que la méthode de dichotomie.*

Nous allons donner à présent des conditions suffisantes sur f permettant d'assurer la convergence de la suite définie en (4.18) pour certaines valeurs de x_0.

Théorème 4.22 *Soit $f \in C^2([a,b])$. On suppose que*

- *(1) $f(a).f(b) < 0$.*
- *(2) $f'(x) \neq 0$ pour tout $x \in [a,b]$ (f est strictement monotone).*
- *(3) $f''(x) \neq 0$ pour tout $x \in [a,b]$ (f ne change pas de concavité).*

Alors pour tout $x_0 \in [a,b]$ tels que $f(x_0).f''(x_0) > 0$, la suite définie en (4.18) converge vers l'unique solution de l'équation $f(x) = 0$.

Preuve Dans la suite on suppose que $f''(x) > 0$ sur $[a,b]$ (et donc $f(x_0) > 0$ compte tenu de l'hypothèse $f(x_0).f''(x_0) > 0$) et que $f'(x) > 0$ sur $[a,b]$. Les trois autres cas se traitent de façon analogue à celui-ci et leurs démonstrations sont laissées au lecteur.
Les conditions (1) et (2) assurent l'existence et l'unicité d'une racine simple

$l \in [a, b]$ de l'équation $f(x) = 0$.
Étape 1. La suite (x_n) est minorée par l
De (4.18) et de la formule de Taylor-Lagrange appliquée à f entre x_n et l,

$$f(l) - f(x_n) = f'(x_n)(l - x_n) + \frac{f''(\epsilon_n)}{2}\frac{(l - x_n)^2}{2},\ \epsilon_n \in]l, x_n[$$

on déduit les égalités

$$\begin{aligned} x_{n+1} - l &= x_n - l + \tfrac{f(l)-f(x_n)}{f'(x_n)} \\ &= x_n - l + \frac{(l - x_n)f'(x_n) + \frac{(l-x_n)^2}{2}f''(\epsilon_n)}{f'(x_n)} \\ &= \tfrac{(x_n-l)^2}{2}\tfrac{f''(\epsilon_n)}{f'(x_n)}. \end{aligned}$$

Il en résulte que si $f''(x)$ et $f'(x)$ sont de même signe sur $[a, b]$ alors pour $n \geq 0$, $x_{n+1} - l > 0$. La suite est minorée par l à partir du rang $n \geq 1$.
Étape 2. La suite (x_n) est décroissante.
la fonction f est strictement croissante sur $[a, b]$ et on a donc $f(x) \geq 0$ pour $x \geq l$.
On a alors puisque $f(x_0) > 0$

$$l < x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} < x_0.$$

Puisqu'on a supposé que $f'(x) > 0$ sur $[a, b]$, la fonction f est croissante sur $[a, b]$ et comme $x_n \geq l$ pour tout $n \geq 1$, on déduit que $f(x_n) \geq f(l) = 0$. On a pour $n \geq 1$

$$l < x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} < x_n.$$

Donc la suite (x_n) est décroissante. Comme elle est minorée par l, elle converge, et on a vu qu'elle converge vers l.

4.5.2 Méthode de Régula Falsi

La méthode de Newton comporte un autre inconvénient que celui de ne pas converger pour n'importe quelles valeurs de x_0. En pratique, on ne connait pas nécessairement l'expression de f' en tous points. On peut alors approcher la valeur $f'(x_n)$ par le quotient aux différences finies

$$\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}.$$

On obtient alors la méthode de régula-falsi :

$$x_{n+1} = x_n - \frac{f(x_n)(x_n - x_{n-1})}{f(x_n) - f(x_{n-1})}. \tag{4.22}$$

On peut alors montrer le théorème :

Théorème 4.23 *Soit $f \in C^2([a,b])$. On suppose que f admet une unique racine $l \in [a,b]$ et que $f'(l) \neq 0$ et $f''(l) \neq 0$. Alors il existe $\eta > 0$ tel que si $x_0, x_1 \in]-\eta + l, l + \eta[$, la suite (x_n) converge vers l.*

4.5.3 Ordre d'une méthode

Définition 4.24 *Soient $g \in C^0([a,b]$ à valeurs dans $[a,b]$ et $x_0 \in [a,b]$ On considère la suite $x_{n+1} = g(x_n)$ et on suppose que (x_n) converge vers l.*
Une méthode définie par $x_{n+1} = g(x_n)$ est dite d'ordre p s'il existe $C > 0$ telle que

$$|x_{n+1} - l| \leq C|x_n - l|^p \quad \forall\, n \in I\!N. \tag{4.23}$$

Une méthode d'ordre 1 est dite linéaire, une méthode d'ordre 2 est dite quadratique.

Remarque 4.25 *D'après la définition 4.24, si une méthode est d'ordre p, elle est aussi d'ordre $m < p$ puisque à partir d'un certain rang, on a l'inégalité*

$$|x_n - l|^p \leq |x_n - l|^m.$$

Précisons l'ordre des méthodes de résolution de l'équation $f(x) = 0$ rencontrées dans ce cours.

Proposition 4.26 *La méthode des approximations successives ainsi que la méthode de la corde sont des méthodes d'ordre 1 au moins. La méthode de Newton est une méthode d'ordre 2 au moins.*

Preuve On a montré que la méthode du point fixe est au moins une méthode d'ordre 1 puisque, sous les hypothéses du théorème 4.6, on a

$$|x_{n+1} - l| = |f(x_n) - f(l)| \leq k|x_n - l|, \quad \forall\, n \in I\!N.$$

Il en va de même dans la méthode de la corde, compte-tenu de ce qui a été établi en (4.17).
D'après (4.21), il résulte que la méthode de Newton est d'ordre 2 au moins.
Enfin, on peut montrer la proposition suivante :

Proposition 4.27 *La méthode de Régula Falsi est d'ordre* $p = \frac{1+\sqrt{5}}{2}$.

Explicitons à présent la définition 4.24 dans le cas où g est trés "réguliére", par exemple de classe C^p.

Proposition 4.28 *Soit g une fonction de classe C^p sur $[a,b]$ et l un point fixe de g. On suppose que $[a,b]$ est stable par g et on considère la suite $x_{n+1} = g(x_n)$.*
La méthode est d'ordre p si et seulement si

$$g'(l) = g''(l) = \cdots = g^{(p-1)}(l) = 0, \quad et \quad g^{(p)}(l) \neq 0.$$

Preuve La fonction g étant de classe C^p au point $x = l$, elle admet un développement limité à l'ordre p en ce point. On a donc

$$e_{n+1} = x_{n+1} - l = g(x_n) - g(l) = \sum_{k=1}^{p} \frac{g^k(l)}{k!}(x_n - l)^k + o((x_n - l)^p).$$

Supposons que

$$g'(l) = g''(l) = \cdots = g^{(p-1)}(l) = 0, \quad \text{et} \quad g^{(p)}(l) \neq 0.$$

On a alors

$$|x_{n+1} - l| = |g^{(p)}(l)||x_n - l|^p + o((x_n - l)^p),$$

Donc, on a bien (4.23).
Supposons qu'il existe un entier m tel que $m < p$ et $g^{(m)}(l) \neq 0$ (et supposons que m soit le plus petit entier satisfaisant cette propriété). Alors

$$\frac{|x_{n+1} - l|}{|x_n - l|^p} = \frac{|g^{(m)}(l)|}{|x_n - l|^{p-m}}(1 + o(x_n - l)).$$

Comme $\frac{|g^{(m)}(l)|}{|x_n - l|^{p-m}}$ tend vers $+\infty$ quand n tend vers l'infini ($p - m > 0$), on $\frac{|x_{n+1} - l|}{|x_n - l|^p}$ tend vers $+\infty$ quand n tend vers $+\infty$: la méthode ne peut donc pas être d'ordre p. Ceci achève la preuve de la proposition 4.26.

4.6 Accélération de la convergence

Théorème 4.29 *Si la méthode définie par $x_{n+1} = g(x_n)$ converge vers l et si $\frac{x_{n+1} - l}{x_n - l} \to A \in \mathbb{R}$ alors la suite (x'_n) définie par*

$$x'_n = x_n - \frac{(x_{n+1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n} \tag{4.24}$$

converge vers l plus rapidement, c'est-à-dire que

$$\lim_{n\to+\infty} \frac{x'_n - l}{x_n - l} = 0.$$

Preuve Posons $e_n = x_n - l$. Par hypothèse, il existe A et (ϵ_n) tels que

$$e_{n+1} = (A + \epsilon_n)e_n, \tag{4.25}$$

où ϵ_n tend vers 0 quand n tend vers $+\infty$. En effet, il suffit de poser pour tout $n \in \mathbb{N}$

$$\epsilon_n = \frac{e_{n+1}}{e_n} - A.$$

On a

$$e_{n+2} = (A + \epsilon_{n+1})e_{n+1} = (A + \epsilon_{n+1})(A + \epsilon_n)e_n.$$

D'autre part,

$$x_{n+2} - 2x_{n+1} + x_n = x_{n+2} - l - 2(x_{n+1} - l) + x_n - l = e_{n+2} - 2e_{n+1} + e_n.$$

Il en résulte que

$$\begin{array}{l} x_{n+2} - 2x_{n+1} + x_n = ((A + \epsilon_{n+1})(A + \epsilon_n) - 2(A + \epsilon_n) + 1)e_n \\ = ((A-1)^2 + \theta_n)e_n \end{array} \tag{4.26}$$

avec

$$\theta_n = (\epsilon_{n+1} + \epsilon_n)A - 2\epsilon_n + \epsilon_{n+1}\epsilon_n.$$

De plus, d'après (4.25)

$$x_{n+1} - x_n = (A - 1 + \epsilon_n)e_n.$$

Donc,

$$x'_n - l = x_n - l - \frac{(x_{n+1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n} = e_n - \frac{(A - 1 + \epsilon_n)^2 e_n^2}{((A-1)^2 + \theta_n)e_n}$$

ou encore

$$\frac{x'_n - l}{x_n - l} = \frac{\theta_n - 2\epsilon_n(A-1) - \epsilon_n^2}{(A-1)^2 + \theta_n}. \tag{4.27}$$

Remarquons que θ_n tend vers 0 quand n tend vers l'infini (car ϵ_n tend vers 0 quand n tend vers $+\infty$). Donc $\dfrac{\theta_n - 2\epsilon_n(A-1) - \epsilon_n^2}{(A-1)^2 + \epsilon_n}$ tend vers 0 quand n tend vers $+\infty$. Compte-tenu de (4.27), on en déduit que $\dfrac{x'_n - l}{x_n - l}$ tend vers 0 quand n tend vers $+\infty$.