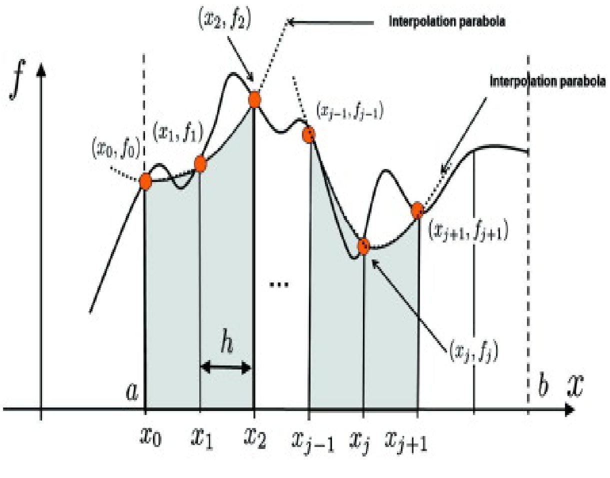
Faculté des Sciences et Techniques Errachidia

Module : Méthodes numériques (M148)

S4, Parcours : MIP

Notes de cours



Responsable du module

Pr. Zakia HAMMOUCH

A.U. 2018/2019

Table des matières

Avant-propos

Ce polycopié se trouve être un support pédagogique de cours, destiné aux étudiants de 2ème Année de Licence parcours MIP. Ce document regroupe un certain nombre de méthodes étudiées dans les différents chapitres du cours de méthodes numériques.

Chapitre 1

Interpolation polynômiale

1.1 Introduction

Dans ce chapitre, nous nous donnons une fonction f supposée continue, définie sur un intervalle de $\mathbb{R}$, la plupart du temps nous noterons $[a, b]$ cet intervalle. Nous cherchons à approcher, dans un sens a préciser, cette fonction par un polynôme. Par interpolation polynomiale on entend chercher un polynôme dont le graphe passe par tous les points donnés (ou connus) de f.

1.1.1 Rappels sur les polynômes

Soit $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. Nous notons $\mathcal{P}(\mathbb{K}))$ l'ensemble des fonctions polynômes. Autrement dit, pour tout $P \in \mathcal{P}(\mathrm{K})$, il existe un entier $n \in \mathbb{N}$ et des coefficients $a_i, i = 0, \ldots, n$ dans $\mathbb{K}$ tels que

$$P(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x^1 + a_2 x^2 + \ldots + a_n x^n$$

pour tout $x \in \mathbb{K}$

Définition 1.1.1.
- *Le degré d'un polynôme non nul est l'indice maximum d'un coefficient non nul. Par convention, le degré d'un polynôme nul (dont tous les coefficients sont nuls) est* $-\infty$.
- *Un monôme est un polynôme de la forme* $a_n x^n$, *pour* $n \in \mathbb{N}, a_n$ *et* $x \in \mathbb{K}$.

Remarque 1.1.1.
- Notons que l'ensemble $\mathcal{P}(\mathrm{K})$ est un K-espace vectoriel.
- Pour tout $n \in \mathbb{N}^*$, nous noterons $\mathcal{P}_n(\mathbb{K})$ le sous-espace vectoriel de $\mathcal{P}(\mathrm{K})$ des polynômes de degré inférieur ou égal à n dont $\{1, x, \ldots, x^n\}$ est une base appelée base canonique.
- La dimension de $\mathcal{P}_n(\mathbb{K})$ est $n + 1$

Nous avons alors la proposition suivante.

Proposition 1.1.1. *Toute famille de polynômes de degré deux à deux distincts est libre. Par conséquent, si* $(P_i)_{i \geq 0}$ *est une suite de polynômes telle que le degré de* P_i *soit égal à* i *alors* $\{P_0, P_1, \ldots, P_n\}$ *est une base de* $\mathcal{P}_n(\mathbb{K})$

Définition 1.1.2. *Soient* $P \in \mathcal{P}(\mathrm{K})$ *et* $x_0 \in \mathbb{K}$.
- x_0 *est une racine (on dit aussi zéro) de* P *lorsque* $P(x_0) = 0$.

- *x_0 est une racine d'ordre k de P si $(x-x_0)^k$ est un diviseur de P tandis que $(x-x_0)^{k+1}$ ne l'est pas,*
- *un polynôme de degré n est scindé lorsqu'il est le produit de n polynômes de degré 1. Autrement dit, il admet n racines distinctes,*
- *un polynôme est dit unitaire (ou normalisé) quand son coefficient dominant est 1.*

Proposition 1.1.2. *Soit $P \in \mathcal{P}(\mathrm{K})$*

- *Si le polynôme P est non nul, alors il possède un nombre fini de racines distinctes inférieur ou égal à son degré.*
- *Si le polynôme P s'annule en un nombre infini de points, alors il est nul.*
- *Si le polynôme P est de degré n et s'annule en $n+1$ points alors il est nul.*
- *Soient P et Q deux polynômes, alors $d^0(P+Q) \leq \max(d^\circ P, d^\circ Q)$*
- *Soient P et Q deux polynômes, alors $d'(PQ) = d^\circ P + d^\circ Q$*

Définition 1.1.3. *Le polynôme dérivé du polynôme $P(x) = \sum_{k=0}^{n} a_k x^k$ est le polynôme $P'(x) = \sum_{k=0}^{n} k a_k x^{k-1}$.*

Proposition 1.1.3. *Soit $P \in \mathcal{P}(\mathrm{K})$, qui s'écrit $P(x) = \sum_{k=0}^{n} a_k x^k$ alors*

- *$P(x) = \sum_{k=0}^{n} \frac{P^{(k)}(x_0)}{k!}(x-x_0)^k$, où $P^{(k)}$ désigne la dérivée k-ème de P, et $x_0 \in \mathbb{K}$. De cette façon un polynôme de degré n est entièrement déterminé par la connaissance de $P(x_0), \ldots, P^{(n)}(x_0)$.*
- *x_0 est une racine d'ordre k de P si $P(x_0) = \ldots = P^{(k-1)}(x_0) = 0$ mais $P^{(k)}(x_0) \neq 0$.*

Intéressons nous maintenant à l'interpolation polynomiale à proprement parlé. II existe plusieurs méthodes. Nous allons décrire ici les plus "incontournables" ou "classiques" selon les points de vue. Commençons par la plus célèbre ; l'interpolation de Lagrange.

1.2 Interpolation de Lagrange

Soit $f : [a,b] \to \mathbb{R}$ continue. Nous nous donnons $n+1$ points distincts dans $[a,b]$. $a \leq x_0 < x_1 < \ldots < x_n \leq b$ appelés nœuds. Pour plus de simplicité, quand ce n'est pas précisé nous noterons $\mathcal{P}_n$ au lieu de $\mathcal{P}_n(\mathbb{R})$ l'espace des polynômes de degré inférieur ou égal à n à coefficients réels. Rappelons que cet espace est de dimension $n+1$.

1.2.1 Polynômes de Lagrange

Théorème 1.2.1. *Pour tout choix de nœuds $x_0, x_1, \ldots, x_n$ dans $[a,b]$, il existe un unique polynôme P_n de degré inférieur ou égal à n qui coïncide avec f aux points $x_0, x_1, \ldots, x_n$ (i. e. $P(x_j) = f(x_j)$ pour tout $j = 0, \ldots, n$). Ce polynôme s'écrit*

$$P_n(x) = \sum_{j=0}^{n} f(x_j) L_j(x), \tag{1.1}$$

où

$$L_j(x) = \prod_{\substack{k=0 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k}$$

pour tout $j = 0, \ldots, n$.

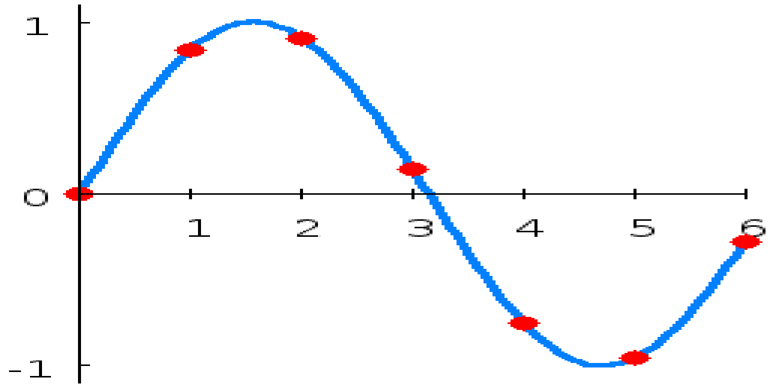


FIGURE 1.1 – Les points correspondent aux noeuds (x_k, y_k), et la courbe continue représente le polynôme d'interpolation.

Remarque 1.2.1.

— Les polynômes de Lagrange sont tels que

$$L_j(x_k) = \delta_{jk} = \begin{cases} 1, & \text{si } j = k \\ 0, & \text{sinon.} \end{cases}$$

on rappelle que δ_{jk} est appelé symbole de Kronecker.

— L'écriture (1.1) n'est pas utilisée en pratique. On ne peut pas calculer facilement le polynôme d'interpolation de f aux point $x_0, x_1, \ldots, x_n$ à partir du polynôme d'interpolation aux nœuds $x_0, x_1, \ldots, x_n$ étant donné que chacun des L_j dépend de tous les nœuds.

Il existe une autre forme, plus pratique à utiliser : la forme de Newton.

1.2.2 Forme de Newton

Construisons la méthode de Newton pas après pas. En commençant avec 1 nœud, puis 2 nœuds, puis n nœuds.

1. Avec un seul nœud x_0 : P_0 de degré 0 et $P_0(x_0) = f(x_0)$, donc

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

2. Avec 2 nœuds x_0 et x_1 : P_1 de degré inférieur où égal à 1, nous avons

$$P_1(x_0) = f(x_0), P_1(x_1) = f(x_1).$$

Nous écrivons alors

$$P_1(x) = P_0(x) + R_1(x), \text{ où } \deg(R_1) \leq 1$$
$$\text{Or } P_1(x_0) = f(x_0) = P_0(x_0) \text{ donc } R_1(x_0) = 0$$

D'où $R_1(x) = a_1(x - x_0)$. De plus $P_1(x_1) = f(x_1) = f(x_0) + a_1(x_1 - x_0)$, et donc

$$a_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

Nous notons alors $f[x_0, x_1] := \frac{f(x_1) - f(x_0)}{x_1 - x_0}$ le coefficient a_1. Et donc,

$$P_1(x) = f(x_0) + f[x_0, x_1](x - x_0).$$

3. Avec 3 nœuds x_0, x_1 et x_2 : P_2 de degré inférieur ou égal à 2, nous avons

$$P_2(x_0) = f(x_0), P_2(x_1) = f(x_1) \text{ et } P_2(x_2) = f(x_2)$$

Nous écrivons alors

$$P_2(x) = P_1(x) + R_2(x), \text{ où } \deg(R_2) \leq 2$$

Or

$$\begin{array}{lll} P_2(x_0) = f(x_0) = P_1(x_0) & \text{donc} & R_1(x_0) = 0, \text{ et} \\ P_2(x_1) = f(x_1) = P_1(x_1) & \text{donc} & R_2(x_1) = 0 \end{array}$$

D'où $R_2(x) = a_2(x - x_0)(x - x_1)$ De plus $P_2(x_2) = f(x_2) = f(x_0) + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1)$

En soustrayant I'expression de $P_2(x_2)$ à celle de $P_2(x_1)$, nous obtenons

$$a_2 = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$$

où $f[x_1, x_2] = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$ et $f[x_0, x_1]$ est défini précédemment comme le coefficient a_1. Nous notons alors $f[x_0, x_1, x_2] := \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$ le coefficient a_2. Et donc,

$$P_2(x) = f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1).$$

De façon générale nous pouvons construire le cas à n nœuds. Pour le calcul des coefficients : nous introduisons la notation des différences divisées $f[x_0] = f(x_0)$ pour tout réel x_0 et pour tout $n \in \mathbb{N}^*$, pour tous nœuds réel $x_0, x_1, \ldots, x_n$

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

Théorème 1.2.2. *Pour tout $n \in \mathbb{N}^*$, pour tous nœuds $x_0, x_1, \ldots, x_n$ dans $[a, b]$, il existe un unique polynôme P_n de degré inférieur ou égal à n qui coïncide avec f aux points $x_0, x_1, \ldots, x_n$ (i.e. $P(x_j) = f(x_j)$[1] pour tout $j = 1, \ldots, n$). Ce polynôme s'écrit*

$$P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + \ldots + f[x_0, \ldots, x_n] \prod_{k=0}^{n-1} (x - x_k)$$

Pour construire les coefficients de Newton nous procédons de la façon suivante :

$$\begin{aligned} a_0 &= f(x_0) \\ a_1 &= f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \\ a_2 &= f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} \\ &\vdots \\ a_n &= f[x_0, x_1, \cdots, x_n] = \frac{f[x_1, x_2, \cdots, x_n] - f[x_0, x_1, x_2, \cdots, x_{n-1}]}{x_n - x_0}. \end{aligned}$$

Remarque 1.2.2. Cette fois-ci, la forme de Newton est bien adaptée à un algorithme, il suffit de calculer les coefficients à l'aide de la définition des différences divisées par récurrence.

1.2.3 Erreur d'interpolation

Théorème 1.2.3. *Soient* $\{x_0, \ldots, x_n\}$, $(n+1)$ *nœuds distincts et* x *un point appartenant au domaine de définition de* f. *Si* $f \in C^{n+1}(I_x)$, *où* I_x *est le plus petit intervalle contenant les nœuds* $x_0, \ldots, x_n$ *et le point* x, *alors, l'erreur d'interpolation au point* x *est donnée par :* $\exists \xi_x \in I_x$ *tel que*

$$E_n(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!} \Pi_{k=0}^{n}(x - x_k).$$

D'où la majoration suivante :

$$|E_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |\Pi_{k=0}^{n}(x - x_k)|.$$

Preuve. *Voir TD.*

Remarque 1.2.3. Notons que M_{n+1} est fini car $f^{(n+1)}$ est continue sur le compact I_x.

Chapitre 2

Intégration numérique

2.1 Principe

Pour les méthodes élémentaires/simples (à opposer aux méthodes composites, voir dernière section du chapitre) de type interpolatrice, on approche $I_{[a,b]}(f) = \int_a^b f(x)\mathrm{d}x$ par l'intégrale sur $[a, b]$ du polynôme d'interpolation de Lagrange de f aux points $(x_i)_{i=0}^n$. On note p le polynôme d'interpolation, dans la base des polynômes de Lagrange p s'écrit

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

En intégrant cette expression sur l'intervalle $[a, b]$ on obtient l'expression générale des formules de quadrature élémentaires de type interpolatrice

$$I_{[a,b]}(f) = \int_a^b f(x)\mathrm{d}x \simeq \tilde{I}_{[a,b]}(f) = \int_a^b p(x)\mathrm{d}x = \sum_{i=0}^{n} \alpha_i f(x_i) \qquad \text{avec} \qquad \alpha_i = \int_a^b l_i(x)\mathrm{d}x$$

2.1.1 Intervalle de référence

Il est souvent plus facile de faire les calculs sur l'intervalle de référence $[-1, 1]$. Une fois ces calculs réalisés, on revient à l'intervalle $[a, b]$ grâce au changement de variable φ.

$$\begin{aligned} \varphi : [-1, 1] &\to [a, b] \\ t &\mapsto \frac{a+b}{2} + t\frac{b-a}{2}. \end{aligned}$$

Pour f continue sur $[a, b]$, on pose $\tilde{f}(t) = f(\varphi(t)) = f(x)$. Alors $\tilde{f}$ est continue sur $[-1, 1]$ et on obtient

$$\int_a^b f(x)\mathrm{d}x = \frac{b-a}{2}\int_{-1}^{1} \tilde{f}(t)\mathrm{d}t.$$

Remarque 2.1.1. Dans le cas où l'intervalle de référence est $[0, 1]$, on utilise le changement de variable suivant : $\psi(t) = (b-a)t + a$, (voir TD).

2.2 Exemples de formules de quadrature élémentaires

Une catégorie remarquable de formules de quadrature par interpolation est celle des formules de Newton-Cotes. Ce sont des formules obtenues avec des points d'interpolation équidistants

$$x_i = a + i\frac{b-a}{n}, \quad n \in \mathbb{N}^*.$$

• **Interpolation à un point : Formule du rectangle à gauche**
Sur l'intervalle de référence $[-1,1]$: $\tilde{I}_{[-1,1]}(\tilde{f}) = 2\tilde{f}(-1)$.
Sur l'intervalle $[a,b]$: $\tilde{I}_{[a,b]}(f) = (b-a)f(a)$
• **Formule du rectangle à droite :**
Sur l'intervalle de référence $[-1,1]$: $\tilde{I}_{[-1,1]}(\tilde{f}) = 2\tilde{f}(1)$.
Sur l'intervalle $[a,b]$: $\tilde{I}_{[a,b]}(f) = (b-a)f(b)$.

• **Formule du point milieu :**
sur l'intervalle de référence $[-1,1]$: $\tilde{I}_{[-1,1]}(\tilde{f}) = 2\tilde{f}(0)$.
Sur l'intervalle $[a,b]$: $\tilde{I}_{[a,b]}(f) = (b-a)f\left(\frac{a+b}{2}\right)$.

• **Interpolation à deux points : formule de trapèze**
Sur l'intervalle de référence $[-1,1]$: $\tilde{I}_{[-1,1]}(\tilde{f}) = \tilde{f}(-1) + \tilde{f}(1)$.
Sur l'intervalle $[a,b]$: $\tilde{I}_{[a,b]}(f) = (b-a)\frac{f(a)+f(b)}{2}$.
• **Interpolation à trois points : formule de Simpson**
Sur l'intervalle de référence $[-1,1]$: $\tilde{I}_{[-1,1]}(\tilde{f}) = \frac{\tilde{f}(-1)+4\tilde{f}(0)+\tilde{f}(1)}{3}$.
Sur l'intervalle $[a,b]$ $\tilde{I}_{[a,b]}(f) = \frac{b-a}{6}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right)$.

2.3 Étude de l'erreur pour les formules de quadrature élémentaires

2.3.1 Degré de précision d'une méthode d'intégration numérique

Définition 2.3.1. *On définit le degré de précision (ou degré d'exactitude) d'une formule d'intégration numérique comme le plus grand entier $r \geq 0$ pour lequel on assure*

$$\forall q \in \mathbb{R}_r[X], \quad I_{[a,b]}(q) = \tilde{I}_{[a,b]}(q).$$

Autrement dit, une formule de quadrature est dite d'ordre r si elle est exacte sur $\mathbb{R}_r[X]$ et inexacte pour au moins un polynôme de degré $r+1$.

Remarque 2.3.1.
- Pour vérifier qu'une formule est au moins d'ordre r, il suffit de vérifier qu'elle est exacte sur la base canonique $(1, X, \ldots, X^r)$.
- Pour les formule de quadrature considérées dans ce cours, on a

$$I_{[a,b]}(f) = \int_a^b f(x)\mathrm{d}x \simeq \int_a^b p(x)\mathrm{d}x,$$

  où p est le polynôme d'interpolation de Lagrange de f aux points $(x_i)_{i=0}^n$. On en déduit que le degré de précision de la méthode de quadrature est au moins n, le degré du polynôme d'interpolation. En effet si $f \in \mathbb{R}_n[x]$ alors p coïncide avec son polynôme d'interpolation sur tout $[a,b]$ et donc la formule de quadrature est exacte.

• **Méthode du rectangle à gauche :**
Cette méthode est au moins de degré 0 (exacte pour les constantes). Pour savoir si elle est au moins de degré 1, on teste la méthode avec $f(x) = x$, on a

$$\int_a^b f(x)\mathrm{d}x = \int_a^b x\mathrm{d}x = \frac{b^2 - a^2}{2} \neq (b-a)a = (b-a)f(a).$$

On en déduit que la méthode du rectangle à gauche est de degré de précision 0.
• **Méthode du rectangle à droite :**
Cette méthode est au moins de degré 0 (exacte pour les constantes). Pour savoir si elle est au moins de degré 1, on teste la méthode avec $f(x) = x$, on a

$$\int_a^b f(x)\mathrm{d}x = \int_a^b x\mathrm{d}x = \frac{b^2 - a^2}{2} \neq (b-a)b = (b-a)f(b).$$

On en déduit que la méthode du rectangle à droite est de degré de précision 0.
• **Méthode du point milieu :**
Cette méthode est au moins de degré 0 (exacte pour les constantes). Pour savoir si elle est au moins de degré 1, on teste la méthode avec $f(x) = x$, on a

$$\int_a^b f(x)\mathrm{d}x = \int_a^b x\mathrm{d}x = \frac{b^2 - a^2}{2} = (b-a)\frac{a+b}{2} = (b-a)f\left(\frac{a+b}{2}\right).$$

Donc la méthode est au moins d'ordre 1. On teste alors $f(x) = x^2$

$$\int_a^b f(x)\mathrm{d}x = \int_a^b x^2\mathrm{d}x = \frac{b^3 - a^3}{3} \neq (b-a)\left(\frac{a+b}{2}\right)^2 = (b-a)f\left(\frac{a+b}{2}\right).$$

On en déduit que la méthode du point milieu est de degré de précision 1.

• **Méthode du trapèze :**

Cette méthode est moins de degré 1 car le polynôme d'interpolation considéré est d'ordre 1. On teste

$$f(x) = x^2 \int_a^b f(x)\mathrm{d}x = \int_a^b x^2\mathrm{d}x = \frac{b^3 - a^3}{3} \neq (b-a)\frac{a^2+b^2}{2} = (b-a)\frac{f(a)+f(b)}{2}.$$

On en déduit que la méthode du trapèze est de degré de précision 1.
• **Méthode du Simpson :**

Cette méthode est moins de degré 2 car le polynôme d'interpolation considéré est d'ordre 2. Pour simplifier, on se place sur $[-1, 1]$, si la formule est exacte sur $[-1, 1]$ alors par le changement de variable qui permet de se ramener à $[a, b]$, la formule sera aussi exacte sur $[a, b]$. On teste $\tilde{f}(t) = t^3$. On rappelle que la formule de quadrature de Simpson sur $[-1, 1]$ est donnée par

$$\tilde{I}_{[-1,1]}(\tilde{f}) = \frac{\tilde{f}(-1) + 4\tilde{f}(0) + \tilde{f}(1)}{3}$$

On vérifie alors que

$$\int_{-1}^1 \tilde{f}(t)\mathrm{d}t = \int_{-1}^1 t^3\mathrm{d}t = 0 = \frac{(-1)^3 + 4 \times 0^3 + 1^3}{3}$$

mais que

$$\int_{-1}^{1} t^4 \mathrm{d}t = \frac{2}{5} \neq \frac{2}{3} = \frac{(-1)^4 + 4 \times 0^4 + 1^4}{3}$$

ce qui montre que la méthode de Simpson est de degré de précision 3, (Voir TD).

2.3.2 Majorations d'erreur pour les méthodes de quadrature élémentaires

Maintenant qu'on a une première idée de la précision des méthodes simples grâce au degré de précision, nous allons essayer de caractériser un peu mieux l'erreur commise quand on utilise une formule de quadrature élémentaire. On s'intéresse donc à l'erreur $E(f)$ et notre but est de trouver pour chacune des méthodes introduites précédemment $M(f)$ (qui dépendra de l'intervalle $[a,b]$, de f et de ses dérivées), le plus petit possible, tel que

$$|E(f)| = \left| I_{[a,b]}(f) - \tilde{I}_{[a,b]}(f) \right| \leq M(f)$$

2.3.3 Quelques outils/rappels

Théorème 2.3.1. *(Théorème des accroissements finis TAF) Soit $f : [a,b] \to \mathbb{R}$ une fonction continue sur $\mathbb{R}$ dérivable sur $]a,b[$. Il existe $c \in]a,b[$ tel que*

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

Théorème 2.3.2. *(Formule de Taylor-Lagrange) Soient $k \in \mathbb{N}, f : [a,b] \to \mathbb{R}$ de classe $\mathcal{C}^k$ sur $[a,b]$ dérivable $k+1$, sur $]a,b[$. Il existe $c \in]a,b[$ tel que*

$$f(b) = f(a) + (b-a)f'(a) + \frac{(b-a)^2}{2!} f''(a) + \ldots + \frac{(b-a)^k}{k!} f^{(k)}(a) + \frac{(b-a)^{k+1}}{(k+1)!} f^{(k+1)}(c).$$

Théorème 2.3.3. *(Erreur interpolation de Lagrange) Soient $n \in \mathbb{N}^*, f : [a,b] \to \mathbb{R}$ une application continue sur $[a,b]$ de classe $\mathcal{C}^{n+1}$ sur $]a,b[$.*
On prend $x_0 < \ldots < x_n$, $(n+1)$ réels distincts dans $[a,b]$, tels que :

$$x_i = a + ih \qquad \text{avec } h = \frac{b-a}{n}.$$

Soient enfin p le polynôme d'interpolation de Lagrange de f aux points x_i et e l'erreur qu'on définit par :

$$e(x) = f(x) - p(x).$$

Alors

$$\forall x \in [a,b], \quad |e(x)| \leq \tfrac{M_{n+1}}{(n+1)!} |\pi(x)|.$$

Avec $M_{n+1} = \max_{x \in [a,b]} \left| f^{(n+1)}(\xi_x) \right|$ et $\pi(x) = \Pi_{i=0}^{n} (x - x_i)$.

•Méthode du rectangle à gauche :

Soit f de classe $\mathcal{C}^1$ sur $[a,b]$, on a

$$\begin{aligned}
|E(f)| = \left|I_{[a,b]}(f) - \tilde{I}_{[a,b]}(f)\right| &= \left|\int_a^b f(x)dx - (b-a)f(a)\right| \\
&= \left|\int_a^b (f(x) - f(a))dx\right| \\
&\leq \int_a^b |f(x) - f(a)|dx \quad \text{(Par inégalité triangulaire)} \\
&\leq \int_a^b \max_{c\in[a,b]} |f'(c)|\, |x-a|dx \quad \text{(Par TAF)} \\
&= M_1 \int_a^b (x-a)dx \\
&= \frac{(b-a)^2}{2} M_1.
\end{aligned}$$

Est-ce que $M(f) = \frac{(b-a)^2}{2} M_1$ est le plus petit majorant possible ?

La réponse et oui ; si on prend la fonction $f(x) = x$, on a $M_1(f) = \max_{[a,b]} |f'(x)| = 1$ et donc

$$\begin{aligned}
E(f) &= \int_a^b f(x)\mathrm{d}x - (b-a)f(a) \\
&= \frac{b^2 - a^2}{2} - (b-a)a \\
&= \frac{(b-a)^2}{2} = \frac{(b-a)^2}{2} M_1.
\end{aligned}$$

On en déduit que cette fonction f particulière réalise l'égalité dans l'estimation d'erreur précédente et donc que la majoration trouvée est optimale (on ne peut pas trouver de plus petit majorant pour cette méthode).

Proposition 2.3.1. *(Erreur pour la méthode du rectangle à gauche)*

$$|E(f)| \leq \frac{M_1}{2}(b-a)^2.$$

• Méthode du rectangle à droite :
Le même raisonnement (en remplaçant $f(a)$ par $f(b)$) conduit à la même estimation. L'erreur pour la méthode du rectangle à droite est donnée par :

$$|E(f)| \leq \frac{M_1}{2}(b-a)^2.$$

• Méthode du point milieu :

Soit f une fonction de classe $\mathcal{C}^2$ sur $[a,b]$, on a

$$\begin{aligned}
|E(f)| = \left|I_{[a,b]}(f) - \tilde{I}_{[a,b]}(f)\right| &= \left|\int_a^b f(x)\mathrm{d}x - (b-a)f\left(\frac{a+b}{2}\right)\right| \\
&= \left|\int_a^b \left(f(x) - f\left(\frac{a+b}{2}\right)\right)\mathrm{d}x\right| \\
&\leq \int_a^b \left|f(x) - f\left(\frac{a+b}{2}\right)\right|\mathrm{d}x \\
&\leq \int_a^b \max_{c\in[a,b]} |f'(c)| \left|x - \frac{a+b}{2}\right| dx \\
&= \frac{(b-a)^2}{4} M_1.
\end{aligned}$$

Est-ce que $M(f) = \frac{(b-a)^2}{2} M_1$ est le plus petit majorant possible ? Non, car on a vu précédemment que la méthode était exacte pour les polynômes de degré 1, donc que l'erreur pour des polynômes de degré 1 était nulle. Cette information pour l'instant n'apparaît pas dans notre majoration d'erreur.
Comment obtenir une majoration optimale pour la méthode du point milieu ?
Pour simplifier les calculs, on se place sur l'intervalle de référence $[-1,1]$ et on s'intéresse à l'erreur

$$|E(\tilde{f})| = \left|\int_{-1}^1 \tilde{f}(t)\mathrm{d}t - \tilde{I}_{[-1,1](\tilde{f})}\right|,$$

pour revenir à l'erreur $|E(f)|$ sur $[a,b]$, on appliquera le changement de variable φ. Sur l'intervalle $[-1,1]$ l'erreur pour la méthode du point milieu s'écrit :

$$|E(\tilde{f})| = \left|\int_{-1}^1 \tilde{f}(t)\mathrm{d}t - 2\tilde{f}(0)\right|.$$

D'après la formule de Taylor-Lagrange à l'ordre 2 , il existe $c \in]a,b[$ tel que

$$\tilde{f}(t) = \tilde{f}(0) + t\tilde{f}'(0) + \frac{t^2}{2}\tilde{f}''(c).$$

et on pose $\tilde{q}(t) = \tilde{f}(0) + t\tilde{f}'(0)$. Comme $\tilde{q}$ est un polynôme de degré 1, la formule du point milieu est exacte pour $\tilde{q}$

$$\int_{-1}^1 \tilde{q}(t)\mathrm{d}t = \tilde{I}_{[-1,1]}(\tilde{q}) = 2\tilde{q}(0) = 2\tilde{f}(0)$$

Ainsi

$$\begin{aligned}
|E(\tilde{f})| &\leq \left|\int_{-1}^1 \tilde{f}(t)\mathrm{d}t - \int_{-1}^1 \tilde{q}(t)\mathrm{d}t\right| \\
&\leq \int_{-1}^1 |\tilde{f}(t) - \tilde{q}(t)|\mathrm{d}t \\
&= \int_{-1}^1 \left|\tilde{f}(t) - \tilde{f}(0) - \tilde{f}'(0)t\right|\mathrm{d}t \\
&\leq \int_{-1}^1 \max_{[-1,1]} \left|\tilde{f}''(c)\right| \frac{t^2}{2}\mathrm{d}t \quad \text{Par Taylor Lagrange} \\
&= \frac{\tilde{M}_2}{2} \times \frac{2}{3} = \frac{\tilde{M}_2}{3}.
\end{aligned}$$

où on a noté $\tilde{M}_k = \max_{c\in[-1,1]} \left|\tilde{f}^{(k)}(c)\right|$. On vérifie que cette majoration est optimale en testant $\tilde{f}(t) = t^2$, on

$$\int_{-1}^{1} \tilde{f}(t)\mathrm{d}t - 2\tilde{f}(0) = \int_{-1}^{1} t^2\mathrm{d}t - 2\times 0 = \frac{2}{3} = \frac{\tilde{M}_2}{3}$$

Si on revient à l'intervalle de départ $[\,a,b]$ en appliquant φ, on a

$$|E(f)| = \left|\int_a^b f(x)\mathrm{d}x - \tilde{I}_{[a,b]}(f)\right| = \frac{b-a}{2}\left|\int_{-1}^{1} \tilde{f}(t)\mathrm{d}t - \tilde{I}_{[-1,1]}(\tilde{f})\right| = \frac{b-a}{2}|E(\tilde{f})|$$

et donc

$$|E(f)| \leq \frac{b-a}{2} \times \frac{\tilde{M}_2}{3}$$

On conclut en revenant aux dérivées de f. Pour cela, on rappelle que

$$\tilde{f}(t) = f(\varphi(t))$$

et donc par différentiation d'une fonction composée

$$\frac{\mathrm{d}\tilde{f}}{\mathrm{d}t}(t) = \frac{\mathrm{d}f}{\mathrm{d}\varphi(t)}(\varphi(t)) \times \frac{\mathrm{d}\varphi}{\mathrm{d}t}(t) = \frac{\mathrm{d}f}{\mathrm{d}x}(x) \times \frac{(b-a)}{2}$$

Par récurrence sur k, l'ordre de dérivation, on montre que

$$\tilde{f}^{(k)}(t) = \left(\frac{b-a}{2}\right)^k f^{(k)}(x)$$

ce qui, avec $k = 2$ et en passant au max dans l'égalité, donne

$$\tilde{M}_2 = \left(\frac{b-a}{2}\right)^2 M_2 = \frac{(b-a)^2}{4} M_2.$$

Proposition 2.3.2. *(Erreur pour la méthode du point milieu)*

$$|E(f)| \leq \frac{M_2}{24}(b-a)^3.$$

- **Méthode de trapèze :**

Comme pour la méthode du point milieu, on prend f de classe $\mathcal{C}^2$ et on se ramène à la fonction $\tilde{f}$ définie sur $[-1,1]$. Soit $\tilde{p}$ le polynôme d'interpolation de Lagrange de $\tilde{f}$ aux points $-1, 1$. On sait que l'erreur d'interpolation est majorée par

$$|e(t)| = |\tilde{f}(t) - \tilde{p}(t)| \leq \frac{|t^2-1|}{2}\tilde{M}_2.$$

On en déduit que

$$\begin{aligned}
|E(\tilde{f})| &= \left|I_{[-1,1]}(\tilde{f}) - I_{[-1,1]}(\tilde{p})\right| \\
&= |\int_{-1}^{1} \tilde{f}(t) - \tilde{p}(t)\Big)\,\mathrm{d}t| \\
&\leq \int_{-1}^{1} |\tilde{f}(t) - \tilde{p}(t)|\mathrm{d}t \\
&\leq \frac{\tilde{M}_2}{2}\int_{-1}^{1} \left(1-t^2\right)\mathrm{d}t = \frac{2\tilde{M}_2}{3}.
\end{aligned}$$

Cette majoration est optimale car on atteint l'égalité (en valeur absolue) avec la fonction $\tilde{f}(t) = t^2$

$$\int_{-1}^{1} \tilde{f}(t)\mathrm{d}t - [f(-1) + f(1)] = \int_{-1}^{1} t^2\mathrm{d}t - 2 = -\frac{4}{3} = -\frac{2\tilde{M}_2}{3}$$

On revient à l'intervalle $[a, b]$ de la même manière que pour la méthode du point milieu.

Proposition 2.3.3. *(Erreur pour la méthode du trapèze)*

$$|E(f)| \leq \frac{M_2}{12}(b-a)^3.$$

Remarque 2.3.2. La méthode des trapèzes, bien que construite à partir d'un polynôme d'interpolation de degré 1 est moins précise que celle du point milieu construite elle à partir d'un polynôme de degré 0.

- **Méthode de Simpson :**

Soit $\tilde{f}$ de classe $\mathcal{C}^4$ sur $[a, b]$ et $\tilde{f}(t) = f(\varphi(t))$ définie sur $[-1, 1]$. Si, comme pour la méthode du trapèze, on considère $\tilde{p}$ le polynôme d'interpolation de $\tilde{f}$, on obtient une majoration d'erreur dépendant de $\tilde{M}_3$ et donc non optimale car on sait que la méthode de Simpson est d'ordre de précision 3 (*cf.* première majoration faite pour la méthode du point milieu). Essayons de dériver directement la majoration optimale en généralisant l'idée utilisée pour la méthode du point milieu. Soit $\tilde{q}$ l'unique polynôme de degré 3 tel que

$$\tilde{q}(0) = \tilde{f}(0), \quad \tilde{q}'(0) = \tilde{f}'(0), \quad \tilde{q}(-1) = \tilde{f}(-1), \quad \tilde{q}(1) = \tilde{f}(1).$$

Lemme 2.3.1. *Lorsqu'on interpole $\tilde{f}$ par $\tilde{q}$ on commet une erreur $\tilde{e}$ majorée par*

$$|\tilde{e}(t)| = |\tilde{f}(t) - \tilde{q}(t)| \leq \frac{\tilde{M}_4}{24} t^2 \left(1 - t^2\right) \quad \forall t \in [-1, 1].$$

Preuve. *On pose $\tilde{\pi}(t) = (t-0)^2(t-(-1))(t-1) = t^2\left(t^2-1\right)$. On va montrer qu'il existe $c \in]-1, 1[$ tel que*

$$\tilde{e}(t) = \frac{\tilde{f}^{(4)}(c)}{24}\tilde{\pi}(t).$$

Si $t = -1, 0$ ou 1 (points d'interpolation), alors

$$\begin{cases} \tilde{e}(t) = 0 \\ \tilde{\pi}(t) = 0 \end{cases}$$

et l'égalité est trivialement vérifiée pour tout $c \in]-1, 1[$.
Supposons $t \in [-1, 1] \backslash \{-1, 0, 1\}$, comme pour l'interpolation de Lagrange, on introduit la fonction auxiliaire

$$F_t(u) = \tilde{e}(u) - \frac{\tilde{e}(t)}{\tilde{\pi}(t)}\tilde{\pi}(u)$$

F_t est de classe $\mathcal{C}^4$ sur $[-1, 1]$ et

$$\begin{aligned} F_t^{(4)}(u) &= \tilde{e}^{(4)}(u) - \frac{\tilde{e}(t)}{\tilde{\pi}(t)}\tilde{\pi}^{(4)}(u) \\ &= \tilde{f}^{(4)}(u) - \underbrace{\tilde{q}^{(4)}(u)}_{=0} - 4!\frac{\tilde{e}(t)}{\tilde{\pi}(t)} \end{aligned}$$

Car le polynôme $\tilde{q}$ est de degré 3.
S'il existe un zéro (racine) noté c de $F_t^{(4)}$ dans $]-1,1[$ alors on a l'égalité annoncée car

$$F_t^{(4)}(c) = 0 \Longleftrightarrow f^{(4)}(c) - 24\frac{\tilde{e}(t)}{\tilde{\pi}(t)} = 0.$$

On est donc ramené à montrer qu'il existe une racine de $F_t^{(4)}$ sur $]-1,1[$.
Commençons par étudier les racines de F_t. La fonction F_t s'annule en quatre points distincts $-1,0,1$ et $t \in [-1,1]\backslash\{-1,0,1\}$. On classe ces points par ordre croissant

$$-1 = c_0^{(0)} < c_1^{(0)} < c_2^{(0)} < c_3^{(0)} = 1 \quad \text{avec} \quad c_1^{(0)} = \min\{0,t\}, c_2^{(0)} = \max\{0,t\}.$$

D'après le théorème de Rolle , il existe

$$\left\{ \begin{array}{l} c_0^{(1)} \in \left] c_0^{(0)}, c_1^{(0)} \right[\\ c_1^{(1)} \in \left] c_1^{(0)}, c_2^{(0)} \right[\\ c_2^{(1)} \in \left] c_2^{(0)}, c_3^{(0)} \right[\end{array} \right. \quad \text{tels que} \quad F_t'\left(c_0^{(1)}\right) = F_t'\left(c_1^{(1)}\right) = F_t'\left(c_2^{(1)}\right) = 0$$

La dérivée première F_t' s'annule donc en au moins trois points. De plus

$$\left\{ \begin{array}{l} \tilde{e}'(0) = \tilde{f}'(0) - \tilde{q}'(0) = 0 \quad \Rightarrow F_t'(0) = 0 \\ \tilde{\pi}'(0) = 0. \end{array} \right.$$

Au final F_t' s'annule en au moins quatre points distincts de $]-1,1[$. Grâce au théorème de Rolle, la dérivée seconde F_t'' s'annule en au moins trois points, la dérivée troisième en au moins deux points et enfin la dérivée quatrième $F_t^{(4)}$ en au moins un point ce qui achève la preuve.
Revenons à la majoration d'erreur pour la méthode de Simpson. Sur l'intervalle $[-1,1]$ la formule de quadrature est donnée par

$$\tilde{I}_{[-1,1]}(\tilde{f}) = \frac{\tilde{f}(-1) + 4\tilde{f}(0) + \tilde{f}(1)}{3} = \frac{\tilde{q}(-1) + 4\tilde{q}(0) + \tilde{q}(1)}{3}.$$

Par ailleurs, on sait que la méthode de Simpson est d'ordre de précision 3 donc en particulier exacte pour $\tilde{q}$ polynôme de degré 3

$$I_{[-1,1]}(\tilde{q}) = \int_{-1}^{1} \tilde{q}(t)\mathrm{d}t = \frac{\tilde{q}(-1) + 4\tilde{q}(0) + \tilde{q}(1)}{3},$$

ce qui entraîne la majoration d'erreur

$$\begin{aligned} |E(\tilde{f})| &= \left| \int_{-1}^{1} \tilde{f}(t)\mathrm{d}t - \frac{\tilde{f}(-1) + 4\tilde{f}(0) + \tilde{f}(1)}{3} \right| \\ &= \left| \int_{-1}^{1} \tilde{f}(t)\mathrm{d}t - \int_{-1}^{1} \tilde{q}(t)\mathrm{d}t \right| \\ &\leq \int_{-1}^{1} |\tilde{f}(t) - \tilde{q}(t)|\mathrm{d}t \\ &\leq \frac{\tilde{M}_4}{24} \int_{-1}^{1} t^2\left(1-t^2\right)\mathrm{d}t = \frac{\tilde{M}_4}{24}\left[\frac{t^3}{3} - \frac{t^5}{5}\right]_{-1}^{1} = \frac{\tilde{M}_4}{90}. \end{aligned}$$

On conclut en revenant à l'intervalle $[a,b]$

$$|E(f)| \leq \frac{b-a}{2}|E(\tilde{f})| \quad \text{et} \quad \tilde{M}_4 = \left(\frac{b-a}{2}\right)^4 M_4.$$

Proposition 2.3.4. *(Erreur pour la méthode de Simpson)*

$$|E(f)| \leq \frac{M_4}{2880}(b-a)^5.$$

2.4 Méthodes de quadrature composites

2.4.1 Principe

Pour approcher $I_{[a,b]}(f)$, les méthodes dites composites consistent à partitionner l'intervalle $[a,b]$ en petits sous-intervalles $[a_{j-1},a_j]$ sur lesquels on applique une méthode élémentaire du type de celles présentées en section précédente. On ne considérera ici que le cas où les sous-intervalles sont de même taille (subdivision uniforme).

Plus précisément, soient $J \in \mathbb{N}^*$ et $h = \frac{b-a}{J}$, on pose $a_j = a + jh$ pour tout $j = 0, \ldots, J$. Par la relation de Chasles

$$\int_a^b f(x)\mathrm{d}x = \sum_{j=1}^{J} \int_{a_{j-1}}^{a_j} f(x)\mathrm{d}x.$$

Puis par une formule de quadrature élémentaire, sur $[a_{j-1},a_j]$

$$\int_{a_{j-1}}^{a_j} f(x)\mathrm{d}x \simeq \tilde{I}_j(f)$$

Ce qui nous donne la formule de quadrature composite

$$I_{[a,b]}(f) = \sum_{i=1}^{J} \tilde{I}_j(f).$$

Remarque 2.4.1.

— Dans la suite on va supposer qu'on applique la même formule de quadrature élémentaire sur tous les sous-intervalles.

— Pour obtenir une majoration de l'erreur totale commise, on écrit

$$\begin{aligned} E_J(f) &= \left| \int_a^b f(x)\mathrm{d}x - \sum_{j=1}^{J} \tilde{I}_j(f) \right| \\ &= \left| \sum_{j=1}^{J} \int_{a_{j-1}}^{a_j} f(x)\mathrm{d}x - \tilde{I}_j(f) \right| \\ &\leq \sum_{j=1}^{J} \left| \int_{a_{j-1}}^{a_j} f(x)\mathrm{d}x - \tilde{I}_j(f) \right| = \sum_{j=1}^{J} E_j(f). \end{aligned}$$

2.4.2 Méthodes composites couramment utilisées

- **Méthode des rectangles à gauche :**

$$I_{[a,b]}(f) = \int_a^b f(x)\mathrm{d}x \simeq h \sum_{j=1}^{J} f(a_{j-1})$$
$$E_J(f) = \left| \int_a^b f(x)\mathrm{d}x - h \sum_{j=1}^{J} f(a_{j-1}) \right| \leq \frac{Jh^2}{2} M_1 \leq \frac{(b-a)^2}{2J} M_1.$$

- **Méthode des rectangles à droite :**

$$I_{[a,b](f)} = \int_a^b f(x)\mathrm{d}x \simeq h\sum_{j=1}^{J} f\left(a_j\right).$$

$$E_J(f) = \left|\int_a^b f(x)\mathrm{d}x - h\sum_{j=1}^{J} f\left(a_j\right)\right| \leq \frac{Jh^2}{2}M_1 \leq \frac{(b-a)^2}{2J}M_1.$$

• Méthodes des points milieux :

$$I_{[a,b]}(f) = \int_a^b f(x)\mathrm{d}x \simeq h\sum_{j=1}^{J} f\left(\frac{a_{j-1}+a_j}{2}\right).$$

$$E_J(f) = \left|\int_a^b f(x)\mathrm{d}x - h\sum_{j=1}^{J} f\left(\frac{a_{j-1}+a_j}{2}\right)\right| \leq \frac{(b-a)^3}{24J^2}M_2.$$

• Méthode des trapèzes :

$$I_{[a,b]}(f) = \int_a^b f(x)\mathrm{d}x \simeq h\sum_{j=1}^{J} \frac{f(a_{j-1})+f(a_j)}{2}$$
$$E_J(f) = \left|\int_a^b f(x)\mathrm{d}x - h\sum_{j=1}^{J} \frac{f(a_{j-1})+f(a_j)}{2}\right| \leq \frac{(b-a)^3}{12J^2}M_2.$$

Remarque 2.4.2. En pratique on n'implémente pas directement cette formule car on voit qu'elle nécessite de calculer deux fois chaque $f\left(a_j\right), j = 1, \ldots, J-1$. On réécrit la méthode sous la forme

$$\begin{aligned} I_{[a,b]}(f) \simeq & h\sum_{j=1}^{J} \frac{f\left(a_{j-1}\right)+f\left(a_j\right)}{2} \\ &= \frac{h}{2}\left(\sum_{j=1}^{J} f\left(a_{j-1}\right) + \sum_{j=1}^{J} f\left(a_j\right)\right) \\ &= \frac{h}{2}\left(\sum_{j=0}^{J-1} f\left(a_j\right) + \sum_{j=1}^{J} f\left(a_j\right)\right) \\ &= h\frac{f(a)+f(b)}{2} + h\sum_{j=1}^{J-1} f\left(a_j\right). \end{aligned}$$

- **Méthode de Simpson composite :**

$$I_{[a,b]}(f) = \int_a^b f(x)\mathrm{d}x \simeq h\sum_{j=1}^{J} \frac{f\left(a_{j-1}\right) + 4f\left(a_{j-1/2}\right) + f\left(a_j\right)}{6}$$

avec

$$a_{j-1/2} = \frac{a_{j-1}+a_j}{2} = a + (j-1/2)h$$

ou encore (forme plus économe en calculs)

$$I_{[a,b]}(f) = \frac{h}{3}\left(\frac{f(a)+f(b)}{2} + \sum_{j=1}^{J-1} f\left(a_j\right) + 2\sum_{j=1}^{J} f\left(a_{j-1/2}\right)\right).E_J(f) = \left|\int_a^b f(x)\mathrm{d}x - h\sum_{j=1}^{J} \frac{f(a_{j-1})+4f\left(a_{j-1/2}\right.}{6}\right.$$

Remarque 2.4.3. La méthode de Simpson est très utilisée, elle constitue un compromis entre précision (erreur en h^4) et quantité de calculs ($2J + 1$ évaluations) considéré généralement comme satisfaisant.

Chapitre 3

Dérivation numérique

3.1 Principe

La dérivation numérique consiste à dériver de façon approchée une fonction sur un intervalle borné $[a, b]$, c'est-à-dire calculer la pente de la courbe représentant la fonction, à partir d'un calcul ou d'une mesure en un nombre fini de points.

— La répartition des points en abscisse est généralement uniforme (pas d'échantillonnage constant h) mais il existe des méthodes à pas variable, ou encore à pas adaptatif.

— La dérivation des polynômes étant très simple, l'opération consiste généralement à construire une interpolation polynomiale par morceaux (de degré plus ou moins élevé) puis de dériver le polynôme sur chaque morceau. On obtient alors des formules dites de différences finies.

— La précision de la dérivation numérique peut alors s'améliorer en augmentant le nombre de points n (en diminuant le pas d'échantillonnage h) ou en augmentant le degré de l'interpolation polynomiale (sous réserve de bonnes propriétés de continuité de la courbe).

3.2 Dérivée première

Soit f une fonction connue seulement par sa valeur en $(n+1)$ points donnés $x_i \quad i = 0; 1; \cdots; n$ distincts. Les formules de différence les plus simples basées sur l'utilisation de la ligne droite pour interpoler les données utilisent deux points pour estimer la dérivée. On suppose connue la valeur de la fonction en x_{i-1}, x_i et x_{i+1}; on pose $f(x_{i-1}) = y_{i-1}, f(x_i) = y_i$, et $f(x_{i+1}) = y_{i+1}$ Si on suppose que l'espace entre deux points successifs est constant, donc on pose $h = x_i - x_{i-1} = x_{i+1} - x_i$ Alors les formules standards en deux points sont :

— Formule de différence progressive :

$$f'(x_i) \simeq \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}.$$

— Formule de différence régressive :

$$f'(x_i) \simeq \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}.$$

— Formule de différence centrale :

$$f'(x_i) \simeq \frac{f(x_{i+1}) - f(x_{i-1})}{x_{i+1} - x_{i-1}} = \frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}}.$$

Remarque 3.2.1.

Les formules de différences classiques peuvent être trouvées en utilisant la formule de Taylor.

$$f(x+h) = f(x) + hf'(x) + \tfrac{h^2}{2} f''(\eta)$$
$$x \leq \eta \leq x+h$$

-Formule progressive :

$$f'(x_i) = \tfrac{f(x_{i+1}) - f(x_i)}{h} - \tfrac{h}{2} f''(\eta)$$
$$x_i \leq \eta \leq x_i + h$$

l'erreur est $\frac{h}{2} f''(\eta)$ donc en $O(h)$: Cette formule peut être trouvée aussi en utilisant le polynôme d'interpolation de Lagrange pour les points $(x_i, f(x_i))$ et $(x_{i+1}, f(x_{i+1}))$.
- Formule régressive

$$h = x_i - x_{i-1}$$
$$f'(x_i) = \tfrac{f(x_i) - f(x_{i-1})}{h} + \tfrac{h}{2} f''(\eta)$$
$$x_{i-1} \leq \eta \leq x_i$$

La formule de différence centrale de la dérivée en x_i peut être trouvée en utilisant la formule de Taylor d'ordre 3 avec $h = x_{i+1} - x_i = x_i - x_{i-1}$
- La formule de différence centrale de la dérivée en x_i peut être trouvée en utilisant la formule de Taylor d'ordre 3 avec $h = x_{i+1} - x_i = x_i - x_{i-1}$

$$f(x_{i+1}) = f(x_i) + hf'(x_i) + \tfrac{h^2}{2} f''(x_i) + \tfrac{h^3}{3!} f^{(3)}(\eta_1)$$
$$f(x_{i-1}) = f(x_i) - hf'(x_i) + \tfrac{h^2}{2} f''(x_i) - \tfrac{h^3}{3!} f^{(3)}(\eta_2)$$
$$x_i \leq \eta_1 \leq x_{i+1}, x_{i-1} \leq \eta_2 \leq x_i$$

si on suppose que $f^{(3)}$ est continue sur $[x_{i-1}, x_{i+1}]$ on peut écrire la formule suivante :

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h} + \frac{h^2}{6} f^{(3)}(\eta)$$

l'erreur est $\frac{h^2}{6} f^{(3)}(\eta)$ donc en $O(h^2)$.
-La formule de différence centrale peut aussi être trouvée à partir du polynôme d'interpolation de Lagrange en 3 points. On peut interpoler les données par un polynôme au lieu d'utiliser la droite, nous obtenons alors les formules de différence qui utilisent plus de deux points. On suppose que le pas h est constant.
- Formule de différence progressive utilisant trois points :

$$f'(x_i) \simeq \frac{-f(x_{i+2}) + 4f(x_{i+1}) - 3f(x_i)}{x_{i+2} - x_i}$$

Formule de différence régressive utilisant trois points :

$$f'(x_i) \simeq \frac{3f(x_i) - 4f(x_{i-1}) + f(x_{i-2})}{x_i - x_{i-2}}$$

3.3 Formule générale en trois points

La formule d'approximation en 3 points de la dérivée première, basée sur le polynôme d'interpolation de Lagrange, n'utilise pas des points équidistants. Étant donné trois points $(x_1, y_1), (x_2, y_2)$ et (x_3, y_3) avec $x_1 < x_2 \leq x_3$, la formule suivante permet d'approcher la dérivée en un point $x \in [x_1, x_3]$. Les dérivées aux points xi sont les suivantes :

$$\begin{array}{l} f'(x_1) = \frac{2x_1 - x_2 - x_3}{(x_1 - x_2)(x_1 - x_3)} y_1 + \frac{x_1 - x_3}{(x_2 - x_1)(x_2 - x_3)} y_2 + \frac{x_1 - x_2}{(x_3 - x_1)(x_3 - x_2)} y_3 \\ f'(x_2) = \frac{x_2 - x_3}{(x_1 - x_2)(x_1 - x_3)} y_1 + \frac{2x_2 - x_1 - x_3}{(x_2 - x_1)(x_2 - x_3)} y_2 + \frac{x_2 - x_1}{(x_3 - x_1)(x_3 - x_2)} y_3 \\ f(x_3) = \frac{x_3 - x_2}{(x_1 - x_2)(x_1 - x_3)} y_1 + \frac{x_3 - x_1}{(x_2 - x_1)(x_2 - x_3)} y_2 + \frac{2x_3 - x_2 - x_1}{(x_3 - x_2)(x_3 - x_1)} y_3 \end{array}$$

Le polynôme de Lagrange est donnée par

$$P(x) = L_1(x)y_1 + L_2(x)y_2 + L_3(x)y_3$$

où

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

$$\begin{array}{l} L_2(x) = \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} \\ L_3(x) = \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)} \end{array}$$

L'approximation de la dérivée première est donnée par $f'(x) \simeq P'(x)$, qui peut s'écrire

$$P'(x) = L_1'(x)y_1 + L_2'(x)y_2 + L_3'(x)y_3$$

où

$$\begin{array}{l} L_1'(x) = \frac{2x - x_2 - x_3}{(x_1 - x_2)(x_1 - x_3)} \\ L_2'(x) = \frac{2x - x_1 - x_3}{(x_2 - x_1)(x_2 - x_3)} \\ L_3'(x) = \frac{2x - x_2 - x_1}{(x_3 - x_2)(x_3 - x_1)} \end{array}$$

Donc

$$f'(x) \simeq \frac{2x - x_2 - x_3}{(x_1 - x_2)(x_1 - x_3)} y_1 + \frac{2x - x_1 - x_3}{(x_2 - x_1)(x_2 - x_3)} y_2 + \frac{2x - x_2 - x_1}{(x_3 - x_2)(x_3 - x_1)} y_3$$

3.4 Dérivées d'ordre supérieur

Les formules de dérivées d'ordre supérieur, peuvent être trouvées à partir des dérivées du polynôme de Lagrange ou en utilisant les formules de Taylor. Par exemple, étant donné 3 points $x_{i-1}; x_i; x_{i+1}$ équidistants, la formule de la dérivée seconde est donnée par :

$$f''(x_i) \simeq \frac{1}{h^2} [f(x_{i+1}) - 2f(x_i) + f(x_{i-1})]$$

l'erreur est en $O(h^2)$:
- Dérivée seconde à partir du polynôme de Taylor

$$\begin{array}{c} f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \frac{h^3}{3!} f^{(3)}(x) + \frac{h^4}{4!} f^{(4)}(\eta_1) \\ f(x-h) = f(x) - hf'(x) + \frac{h^2}{2!} f''(x) - \frac{h^3}{3!} f^{(3)}(x) + \frac{h^4}{4!} f^{(4)}(\eta_2) \\ x \leq \eta_1 \leq x + h \text{ et } x - h \leq \eta_2 \leq x \\ f''(x) \simeq \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \end{array}$$

l'erreur est en $O(h^2)$: Pour obtenir les formules de la troisième et la quatrième dérivée, on prend une combinaison linéaire des développement de Taylor, pour $f(x+2h)$, $f(x+h), f(x-h)$ et $f(x-2h)$.
La table suivante donne différentes formules centrales toutes en $O(h^2)$:

$$f'(x_i) \simeq \frac{1}{2h}\left[f(x_{i+1}) - f(x_{i-1})\right].$$

$$\begin{gathered} f''(x_i) \simeq \tfrac{1}{h^2}\left[f(x_{i+1}) - 2f(x_i) + f(x_{i-1})\right] \\ f^{(3)}(x_i) \simeq \tfrac{1}{2h^3}\left[f(x_{i+2}) - 2f(x_{i+1}) + 2f(x_{i-1}) - f(x_{i-2})\right] \\ f^{(4)}(x_i) \simeq \tfrac{1}{h^4}\left[f(x_{i+2}) - 4f(x_{i+1}) + 6f(x_i) - 4f(x_{i-1}) + f(x_{i-2})\right] \end{gathered}$$

En utilisant les polynômes d'interpolation de Lagrange les dérivées d'ordre p sont calculées par :

$$f^{(p)}(\alpha) \sim \sum_{i=0}^{n} A_i(\alpha) f(x_i)$$

où $A_i(\alpha) = L_i^{(p)}(\alpha) \quad p \leq n.$

Remarque 3.4.1. *La formule est exacte pour les polynômes de degrés n : Le système linéaire donnant les $A_i(\alpha)$ a un déterminant de type Vandermonde différent de zéro si les x_i sont distincts. Les $A_i(\alpha)$ sont indépendants de f et peuvent être calculés une fois pour toutes.*

La formule est exacte pour les polynômes de degrés n : Le système linéaire donnant les $A_i(\alpha)$ a un déterminant de type Vandermonde différent de zéro si les x_i sont distincts. Les $A_i(\alpha)$ sont indépendants de f et peuvent être calculés une fois pour toutes.

3.5 Étude de l'erreur commise

D'après le chapitre *interpolation polynômiale*, si f est connue en $(n+1)$ points $x_i; i = 0; \ldots; n$ alors $f(x) = P_n(x) + e(x)$; où $e(x)$ est l'erreur d'interpolation. En dérivant on obtient

$$\begin{aligned} f'(x) = P_n'(x) + e'(x) = \textstyle\sum_{i=0}^{n} A_i(x) f(x_i) + e'(x) Avec : e'(x) = \tfrac{d}{dx}\left[\tfrac{1}{(n+1)!} L(x) f^{(n+1)}(\xi_x)\right] \\ = \tfrac{1}{(n+1)!} L'(x) f^{(n+1)}(\xi_x) + \tfrac{1}{(n+2)!} L(x) \tfrac{d}{dx}\left[f^{(n+2)}(\xi_x)\right] \end{aligned}$$

On remarque tout de suite que l'erreur de dérivation est nulle si f est un polynôme de degré inférieur ou égale à n. Si on prend pour x un point x_i; le second terme de la dernière somme s'annule, sinon il faut connaitre $\frac{d}{dx}\left[f^{(n+1)}(\xi_x)\right]$; ce qui est difficile car la fonction $x \to \xi_x$ étant inconnue. On constate qu'on devra se contenter d'une estimation :

$$|e'(x)| \leq \frac{1}{(n+1)!} L'(x) M_{n+1} + \frac{1}{(n+2)!} L(x) M_{n+2},$$

avec $M_{n+1} = \max_{[a,b]} |f^{(n+1)(x)}|$ et $M_{n+2} = \max_{[a,b]} |f^{(n+2)(x)}|$.

Chapitre 4

Résolution numérique des équations non linéaires $f(x) = 0$

4.1 Introduction

Ce chapitre a pour but de rechercher des solutions de l'équation non linéaire $f(x) = 0$ où f est une fonction donnée. Les méthodes numériques pour approcher une solution consistent à localiser grossièrement un zéro de f en procédant le plus souvent par des considérations graphiques ; la solution grossière est noté x_0, et aussi à construire à partir de x_0 une suite $x_1, x_2, x_3, \ldots$ telle que

$$\lim_{n \to \infty} x_n = \bar{x} \quad \text{où } \bar{x} \text{ vérifie } f(\bar{x}) = 0.$$

4.2 Existence de solutions et localisation des solutions

On se donne une application f continue d'un ouvert I de $\mathbb{R}$ dans $\mathbb{R}$ et on cherche à approcher numériquement une solution de l'équation $f(x) = 0$. Est ce qu'un tel x existe ?

Définition 4.2.1. *Soit $f : I \longrightarrow \mathbb{R}$, on appelle zéro de f tout $\bar{x} \in I$ qui satisfait*

$$f(\bar{x}) = 0.$$

On dit aussi que $\bar{x}$ est une racine de f.

Définition 4.2.2. *On appelle point fixe de f tout $\bar{x}$ qui satisfait*

$$f(\bar{x}) = \bar{x}.$$

Exercice 4.2.1.

- *Déterminer les zéros de l'application $f(x) = x^2 - 5x + 6$. Faire un dessin.*
- *Déterminer les points fixes de l'application $g(x) = x^2 - 4x + 6$. Faire un dessin. En déduire une interprétation graphique d'un zéro et d'un point fixe d'une fonction.*

Remarque 4.2.1. Si $\bar{x}$ est un zéro de f alors $\bar{x}$ est un point fixe de $g : x \mapsto f(x) + x$.

4.3 Premiers résultats théoriques

Théorème 4.3.1. *(Théorème des valeurs intermédiaires) Soit f une fonction continue sur $I = [a, b]$. Alors f atteint toutes les valeurs entre $f(a)$ et $f(b)$*

$$\forall d \in [f(a), f(b)] \text{ il existe } c \in I \text{ tel que } f(c) = d.$$

Corollaire 4.3.1. *Soit $f : I = [a, b] \to \mathbb{R}$ une application continue telle que $f(a)f(b) < 0$, c'est-à-dire que $f(a)$ et $f(b)$ sont non nuls et de signes opposés. Alors il existe $\bar{x} \in]a, b[$ tel que $f(\bar{x}) = 0$. Si de plus f est strictement monotone, alors $\bar{x}$ est unique.*

Exemple 4.3.1.

— *Une fonction polynomiale à coefficients réels de degré impair admet au moins un zéro sur $\mathbb{R}$.*

— *L'équation $x(1 + 2x) = e^x$ admet une unique solution dans l'intervalle $]0, 1[$.*

Corollaire 4.3.2. *(Théorème de point fixe) Soit $g : [a, b] \to [a, b]$ continue sur $[a, b]$. Alors g admet un point fixe $\bar{x}$ dans l'intervalle $[a, b]$*

Preuve. *Supposons par l'absurde que g n'admet pas de point fixe sur $[a, b]$. Alors en particulier*

$$\begin{cases} g(a) > a \\ g(b) < b \end{cases}$$

Posons $f(x) = g(x) - x$, f est continue puisque g l'est. De plus

$$\begin{cases} f(a) > 0 \\ f(b) < 0 \end{cases}$$

Le TVI nous donne alors l'existence d'un zéro de f dans $[a, b]$, $\bar{x}$, qui est donc par définition de f un point fixe de g.

4.4 Construction de solutions approchées

Définition 4.4.1. *On appelle méthode itérative un procédé de calcul de la forme*

$$x_{n+1} = F(x_n), \quad n = 0, 1, 2, \ldots$$

qui part d'une valeur donnée x_0 pour calculer x_1, puis à l'aide de x_1 on calcule x_2, etc. La formule qui donne x_{n+1} est dite formule de récurrence. Le procédé est dit convergent si x_n tend vers un nombre fini lorsque n tend vers l'infini.

Définition 4.4.2. *Soit p un entier positif. On dit qu'une méthode convergente est d'ordre p s'il existe une constante C telle que si $\bar{x} = \lim_n x_n$ alors*

$$|\bar{x} - x_{n+1}| \leq C |\bar{x} - x_n|^p$$

ou encore

$$\lim_{n \to \infty} \frac{|\bar{x} - x_{n+1}|}{|\bar{x} - x_n|^p} = C.$$

Si $p = 1$ on parle de convergence linéaire, si $p = 2$ la convergence est dite quadratique.

Remarque 4.4.1. Dans le cas où $p = 1$, il est nécessaire que $C < 1$ pour que (x_n) converge vers $\bar{x}$.

4.4.1 Méthodes de la dichotomie et de Lagrange

Principe des deux méthodes

Ces méthodes s'appuient sur le théorème des valeurs intermédiaires. On considère la fonction f continue sur $[a, b]$ à valeurs dans $\mathbb{R}$ avec $f(a)f(b) < 0$. On sait donc qu'il existe un zéro de f dans $I_0 =]a, b[$ qu'on note $\bar{x}$. Pour localiser $\bar{x}$ on va calculer à chaque itération un sous-intervalle $I_n = [a_n, b_n]$ de I_{n-1} dans lequel $\bar{x}$ est localisé.

• Algorithme de dichotomie :

La méthode de dichotomie consiste à découper l'intervalle I_n en deux intervalles de même longueur. Concrètement, supposons par exemple que $f(a_n) < 0, f(b_n) > 0$ et notons $x_n = \frac{a_n+b_n}{2}$. On étudie le signe de $f(x_n)$.

— Si $f(x_n) = 0$ alors $\bar{x} = x_n$.

— Si $f(x_n) < 0$, d'après de TVI, il existe un zéro de f sur l'intervalle $]x_n, b_n[$ On pose donc
$$a_{n+1} = x_n \text{ et } b_{n+1} = b_n.$$

— Si $f(x_n) > 0$ d'après de TVI, il existe un zéro de f sur l'intervalle $]a_n, x_n[$. On pose donc
$$a_{n+1} = a_n \text{ et } b_{n+1} = x_n.$$

On poursuit alors la construction jusqu'à obtenir la précision souhaitée.
La méthode de dichotomie consiste donc à construire deux suites adjacentes (a_n) et (b_n) (qui sont les extrémités des intervalles successifs dans lesquels $\bar{x}$ est localisé) convergeant vers $\bar{x}$.

Théorème 4.4.1. *Soient a et b dans $\mathbb{R}$ tels que $a < b$, et $f : [a, b] \mapsto \mathbb{R}$ une application continue possédant un unique zéro noté $\bar{x} \in]a, b[$. On suppose de plus que $f(a)f(b) < 0$. Alors les deux suites (a_n) et (b_n) convergent vers $\bar{x}$ et on a les majorations d'erreur suivantes :*
$$\forall n \geq 0, \quad 0 \leq \bar{x} - a_n \leq \frac{b-a}{2^n}, \quad 0 \leq b_n - \bar{x} \leq \frac{b-a}{2^n}.$$

Preuve. *On va montrer que les deux suites (a_n) et (b_n) sont adjacentes. Plus précisément, on va montrer par récurrence sur n que*
$$\forall n \geq 1, \quad a_{n-1} \leq a_n, \quad b_n \leq b_{n-1}, \quad b_n - a_n = \frac{b-a}{2^n} \quad (\text{ ou bien } a_n = b_n = \bar{x})\,(1)$$

On rappelle que : (a_n) et (b_n) sont dites adjacentes si l'une est croissante, l'autre décroissante et $|a_n - b_n| \to 0$ quand $n \to \infty$.
D'autre part ont sait que : si (a_n) et (b_n) sont adjacentes alors ces deux suites sont convergentes et ont la même limite $l \in \mathbb{R}, \forall n, a_n \leq l \leq b_n$.
- Initialisation. $n = 1$. Si $f(x_0)$ est du signe de $f(a)$ alors $a_1 = x_0 = \frac{a_0+b_0}{2}$ et $b_1 = b_0 = b$. Ainsi on vérifie que :
$$a_0 \leq \tfrac{a+b}{2} = a_1 \leq b_1 = b_0$$
$$b_1 - a_1 = b - \tfrac{a+b}{2} = \tfrac{b-a}{2}$$

On fait le raisonnement si $f(x_0)f(a) < 0$.
-Hérédité : On suppose la propriété (1) est vraie au rang n. On étudie le signe de $f(x_n) =$

$f\left(\frac{a_n+b_n}{2}\right)$. *Si* $f(x_n)\, f(a) > 0$ *(l'opposé se traitant de manière complètement analogue. Alors*

$$a_{n+1} = x_n \quad \text{et} \qquad a_n \leq \tfrac{a_n+b_n}{2} = a_{n+1},$$

et par hypothèse de récurrence on obtient

$$b_{n+1} - a_{n+1} = b_n - \frac{a_n+b_n}{2} = \frac{b_n - a_n}{2} = \frac{1}{2}\frac{b-a}{2^n} = \frac{b-a}{2^{n+1}},$$

ce qui achève la démonstration de (1).
On a ainsi montré que (a_n) *et* (b_n) *sont adjacentes, on note* l *la limite commune. Afin de conclure la démonstration du théorème, il reste à montrer que* $l = \bar{x}$. *Pour cela on va prouver que* $f(l) = 0$. *Pour tout* n *on a*

$$\begin{cases} f(a_n)\, f(a) \geq 0, \\ f(b_n)\, f(b) \geq 0, \end{cases}$$

ce qui donne en passant à la limite par continuité de f *:*

$$\begin{cases} f(l) f(a) \geq 0, \\ f(l) f(b) \geq 0. \end{cases}$$

Comme par ailleurs on a supposé que a *et* b *étaient non nuls et de signes opposés on a nécessairement*

$$\begin{cases} f(l) \leq 0 \\ f(l) \geq 0 \end{cases} \quad \Longrightarrow \quad f(l) = 0$$

On conclut enfin en disant que

$$\begin{array}{c} 0 \leq \bar{x} - a_n \leq b_n - a_n \leq \frac{b-a}{2^n} \\ 0 \leq b_n - x_n \leq b_n - a_n \leq \frac{b-a}{2^n}. \end{array}$$

Remarque 4.4.2. Comme on l'a dit, les itérations s'achèvent à la m-ième étape quand

$$|\bar{x} - x_m| \leq |I_m| < \varepsilon.$$

où ε est une tolérance fixée. On a $|I_m| = \frac{b-a}{2^m}$ donc pour avoir une erreur inférieure à ε, on doit prendre le plus petit m tel que

$$m \geq \frac{\log\left(\frac{b-a}{\varepsilon}\right)}{\log(2)} = \log_2\left(\frac{b-a}{\varepsilon}\right).$$

La méthode de dichotomie ne garantit pas la réduction monotone de l'erreur absolue d'une itération à l'autre, c'est-à-dire qu'on n'a pas

$$|\bar{x} - x_{n+1}| \leq C_n\, |\bar{x} - x_n| \qquad \text{pour tout } n \geq 0,$$

avec $C_n < 1$. La méthode de dichotomie n'est pas une méthode d'ordre 1.
Son avantage est qu'elle est facile à implémenter, une fois un zéro isolé on a convergence à coup sûr. Son inconvénients : Convergence lente, méthode pas généralisable en dimension supérieure, ne s'applique pas par exemple pour chercher les extremas, par ex. pour $x \mapsto x^2$.

• **Algorithme de Lagrange :**
Plutôt que de couper l'intervalle en deux intervalles de même longueur, on découpe $I_n = [a_n, b_n]$ en $[a_n, x_n]$ et $[x_n, b_n]$ où x_n est le point d'intersection de la droite passant par $(a_n, f(a_n))$ et $(b_n, f(b_n))$ et l'axe des abscisses. Autrement dit x_n satisfait les équations suivantes

$$\frac{f(b_n) - f(a_n)}{b_n - a_n}(x_n - b_n) + f(b_n) = 0$$

$$x_n = \frac{a_n f(b_n) - b_n f(a_n)}{f(b_n) - f(a_n)}$$

On est alors ramené comme précédemment à étudier le signe de $f(x_n)$.

Proposition 4.4.1. *Si f (régulière sur $[a,b]$) avec f'' de signe constant sur $[a,b]$ (càd f convexe ou concave sur $[a,b]$ alors soit il existe un n tel que $f(x_n) = 0$, soit x_n est bien défini pour tout n et x_n converge à l'ordre 1 vers $\bar{x}$ où $\bar{x}$ est l'unique zéro de f dans $[a,b]$.*

Preuve. *Supposons que $f(a) < 0$ et $f(b) > 0$ avec f convexe ($f'' \geq 0$).*
-f admet un unique zéro sur $[a,b]$.
- Dans la suite on suppose qu'il n'existe pas d'indice pour lequel $x_n = \bar{x}$. Montrons que pour tout n on a $a_{n+1} = x_n, b_{n+1} = b$ avec (x_n) strictement croissante. On raisonne par récurrence, on sait que $x_n > a_n = x_{n-1}$ car $f(a_n) < 0$. Or, x_n appartient au segment $[a_n, b_n]$ donc s'écrit comme barycentre des points a_n et b_n :

$$x_n = \lambda_n a_n + (1-\lambda_n) b_n \quad \textit{où } \lambda_n = \frac{x_n - b_n}{a_n - b_n},$$

l'inégalité de convexité nous donne alors

$$f(x_n) = f(\lambda_n a_n + (1-\lambda_n) b_n) \leq \lambda_n f(a_n) + (1-\lambda_n) f(b_n) = 0.$$

Ceci montre bien qu'on pose au rang suivant $a_{n+1} = x_n$.

-$x_n \to \bar{x}$: d'après ce qui précédé, la suite (x_n) est croissante, majorée par b. C'est donc une suite convergente, on note sa limite l. Reste à montrer donc que $l = \bar{x}$. Pour cela, écrivons la formule de récurrence donnant la suite (x_n) :

$$x_{n+1} = x_n - \frac{b - x_n}{f(b) - f(x_n)} f(x_n) \, (2)$$

En multipliant les deux membres de l'équation par $f(b) - f(x_n)$ et en passant à la limite $n \to \infty$ on obtient

$$(b - l) f(l) = 0(3)$$

Par ailleurs l'inégalité $x_n < \bar{x} < b$ implique qu'à la limite : $l \leq \bar{x} < b$. En revenant à l'équation (3) on en déduit que $f(x_l) = 0$ donc que $l = \bar{x}$.
la convergence est linéaire : en utilisant la formule de récurrence (2) on écrit

$$\bar{x} - x_{n+1} = \bar{x} - x_n + \frac{b - x_n}{f(b) - f(x_n)} f(x_n),$$

puis

$$\begin{aligned}\bar{x} - x_{n+1} &= (\bar{x} - x_n)\left(1 + \frac{b - x_n}{f(b) - f(x_n)} \frac{f(x_n)}{\bar{x} - x_n}\right) \\ &= (\bar{x} - x_n)\left(1 - \frac{b - x_n}{f(b) - f(x_n)} \frac{f(\bar{x}) - f(x_n)}{\bar{x} - x_n}\right).\end{aligned}$$

Passons à la limite $n \to \infty$ dans cette dernière équation pour aboutir à

$$\lim_{n\to\infty} \frac{\bar{x} - x_{n+1}}{\bar{x} - x_n} = 1 - \frac{b - \bar{x}}{f(b) - f(\bar{x})} f'(\bar{x}) = C.$$

Il nous faut donc montrer que $0 \leq C < 1$. C'est la convexité de f qui permet de conclure. En effet elle nous permet d'écrire les inégalités suivantes

$$\begin{cases} f(b) \leq f(\bar{x}) + f'(\bar{x})(b - \bar{x}), \\ f(a) \leq f(\bar{x}) + f'(\bar{x})(b - \bar{x}), \end{cases}$$

ou encore en utilisant que $a < \bar{x} < b$ et $f(\bar{x}) = 0$.

$$\begin{cases} f'(\bar{x}) \leq \frac{f(b)}{b-\bar{x}} & \Longrightarrow C \geq 0, \\ f'(\bar{x}) \geq \frac{f(a)}{a-x} & \Longrightarrow C < 1. \end{cases}$$

Exercice 4.4.1. *Décrire les méthodes de dichotomie et de Lagrange et les utiliser pour calculer le zéro de la fonction : $f(x) = 2x^3 - x - 5$ dans l'intervalle $[1, 2]$ avec une précision de 10^{-2}.*

- **Méthodes de points fixe :**

Théorème 4.4.2. *(Théorème de point fixe contractant) Soient I un intervalle fermé non vide de $\mathbb{R}$ et $g : I \to I$ une application strictement contractante càd qu il existe une constante $0 < k < 1$ telle que*

$$\forall (x, y) \in I^2 \quad |g(x) - g(y)| < k|x - y|.$$

Alors il existe un unique $\bar{x} \in I$ tel que $g(\bar{x}) = \bar{x}$ la suite définie par

$$\begin{cases} x_0 \in I, \\ \forall n \in \mathbb{N} \quad x_{n+1} = g(x_n), \end{cases}$$

converge vers $\bar{x}$. De plus, on a la majoration d'erreur

$$|x_n - \bar{x}| \leq k^n |x_0 - \bar{x}| \quad \forall n \in \mathbb{N}.$$

Preuve. *On se place dans le cas plus simple où $I = [a, b]$.*
- Existence d'un point fixe : facile via le corollaire 2 (on remarquera pour cela que la propriété de k-contractivité entraîne la continuité de g).
- Unicité du point fixe : soient $\bar{x}$ et $\tilde{x}$ deux points fixes de g sur I. Comme g est strictement contractante on a :

$$|\bar{x} - \tilde{x}| = |g(\bar{x}) - g(\tilde{x})| \leq k|\bar{x} - \tilde{x}|.$$

La condition $k < 1$ impose donc $|\bar{x} - \tilde{x}| = 0$ càd $\bar{x} = \tilde{x}$.
- Convergence et majoration d'erreur : On va la démontrer par récurrence sur n que

$$|x_n - \bar{x}| \leq k^n |x_0 - \bar{x}|.$$

— *Initialisation : on a bien $|x_0 - \bar{x}| \leq k^0 |x_0 - \bar{x}|$.*

— *Hérédité : on suppose l'inégalité vérifiée au rang n. Au rang $n+1$ on a*

$$|x_{n+1} - \bar{x}| = |g(x_{n+1}) - g(\bar{x})| \leq k\,|x_n - \bar{x}| \leq k^{n+1}\,|x_0 - \bar{x}|\,.$$

L'inégalité est bien satisfaite au rang $n+1$ ce qui conclut la preuve par récurrence. Comme $k < 1$, on a $k^n \to 0$ et cette inégalité permet ainsi de montrer la convergence de x_n vers $\bar{x}$.
Dans le cas plus général où I est un fermé de $\mathbb{R}$, l'existence de $\bar{x}$ et la convergence de la suite s'obtiennent simultanément en montrant que (x_n) est une suite de Cauchy. En effet comme g est k-contractante on peut écrire pour $n \geq 1$:

$$\begin{aligned}|x_{n+1} - x_n| &= |g(x_n) - g(x_{n-1})| \\ &\leq k\,|x_n - x_{n-1}| \\ &\leq k^n\,|x_1 - x_0|\,.\end{aligned}$$

Soit maintenant $p > n \geq 0$, on a

$$\sum_{j=n+1}^{p} (x_j - x_{j-1}) = x_p - x_n,$$

et l'inégalité précédente donne :

$$\begin{aligned}|x_p - x_n| &\leq \sum_{j=n+1}^{p} k^{j-1}\,|x_1 - x_0| \\ &= \sum_{j=0}^{p-(n+1)} k^{n+j}\,|x_1 - x_0| \\ &= k^n\,|x_1 - x_0| \sum_{j=0}^{p-(n+1)} k^j \\ &\leq k^n\,|x_1 - x_0| \sum_{j=0}^{+\infty} k^j \\ &= \frac{k^n\,|x_1 - x_0|}{1-k} \to 0 \qquad \textit{quand} \quad n \to \infty.\end{aligned}$$

La suite (x_n) est donc une suite de Cauchy dans I qui est un fermé de $\mathbb{R}$. Elle converge donc dans I, on note l sa limite ; par continuité de g, en passant à la limite dans l'égalité $x_{n+1} = g(x_n)$, on obtient $l = g(l)$.
L'unicité se démontre de la même manière que pour le cas d'un segment.

Remarque 4.4.3. I peut être de la forme $I = \mathbb{R}$, $I =]-\infty, a[$, $[a, +\infty[$ ou $I = [a, b]$. .

Corollaire 4.4.1. *Supposons $g : \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^1$ et soit $\bar{x}$ un point fixe de g. Si $|g'(\bar{x})| < 1$ alors il existe $\varepsilon > 0$ tel que si x_0 satisfait $|\bar{x} - x_0| \leq \varepsilon$, alors la suite donnée par $x_{n+1} = g(x_n)$. converge vers $\bar{x}$ lorsque n tend vers l'infini.*

Définition 4.4.3. *Soit $\bar{x}$ un point fixe d'une application g, on dit que $\bar{x}$ est un point fixe attractif si $|g'(x)| < 1$, répulsif si $|g'(\bar{x})| > 1$.*

Le corollaire précédent nous donne donc la convergence locale autour des point fixes attractifs. A contrario le résultat suivant donne la non-convergence pour les points fixes répulsifs.

Proposition 4.4.2. *Soient $g : I \to \mathbb{R}$ une application de classe $\mathcal{C}^1$ et $\bar{x}$ un point fixe répulsif de g. On pose (x_n) la suite définie par l'approximation de point fixe $x_{n+1} = g(x_n)$. Alors soit la suite (x_n) est stationnaire, égale à $\bar{x}$ à partir d'un certain rang, soit (x_n) ne converge pas vers $\bar{x}$.*

Preuve. *Comme g' est continue sur I, il existe $\delta > 0$ tel que*

$$\forall x \in I, \quad |\bar{x} - x| < \delta \Longrightarrow |g'(x)| > 1.$$

Soit $x_0 \in I$. Supposons par l'absurde que la suite (x_n) converge vers $\bar{x}$. Alors il existe un rang N_0 (qui dépend de δ) tel que

$$\forall n \geq N_0, \quad |x_n - \bar{x}| \leq \delta,$$

et donc

$$\forall n \geq N_0, \quad |g'(x_n)| > 1.$$

D'après le théorème des accroissements finis il existe c_n entre x_n et $\bar{x}$ tel que :

$$g(x_n) - g(\bar{x}) = g(c_n)(x_n - \bar{x}).$$

Donc pour tout $n \geq N_0$, on a :

$$|x_{n+1} - \bar{x}| = |g(x_n) - g(\bar{x})| = |g'(c_n)| |x_n - \bar{x}| > |x_n - \bar{x}|.$$

Si pour tout $n \leq N_0$ on a $x_n \neq \bar{x}$, l'inégalité précédente entre en contradiction avec la convergence des (x_n) vers $\bar{x}$.

• La méthode de Newton :

Soit f une fonction définie sur un intervalle $[a,b]$, continument dérivable sur $[a,b]$ (i.e. de classe $\mathcal{C}^1$ sur $[a,b]$). Soit $\bar{x}$ un zéro de f que $[a,b]$ tel que $f'(\bar{x}) \neq 0$. Par relaxation on peut écrire :

$$\{f(\bar{x}) = 0\} \Longleftrightarrow \{\forall \lambda \neq 0, \quad \bar{x} = g_\lambda(\bar{x})\} \quad g_\lambda(x) = x + \lambda f(x)$$

Alors le meilleur choix de constante λ est celui pour lequel la méthode est d'ordre 2, soit celui pour lequel $g'_\lambda(\bar{x}) = 0$ donc :

$$\lambda = -\frac{1}{f'(\bar{x})}$$

En pratique il n'est pas possible (sauf cas particulier) de calculer $f'(\bar{x})$ puisqu'on ne connait pas $\bar{x}$. Une solution consiste alors à approximer à chaque itération $f'(\bar{x})$ par $f'(x_n)$. C'est la méthode de Newton.

Théorème 4.4.3. *(Théorème de convergence globale) On reprend les hypothèses précédentes. et on suppose de plus que f est de classe $\mathcal{C}^2$ sur I et que f' et f'' ne s'annulent pas sur I. Soit $x_0 \in I$ tel que $f(x_0)$ soit du même signe que f'' (on suppose qu'il existe au moins un tel x_0). Alors la suite définie par la méthode de Newton*

$$\begin{cases} x_0 \text{ donné} \\ x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = g(x_n) \end{cases}$$

est bien définie et converge de manière monotone vers $\bar{x}$ unique zéro de f (également point fixe de g) sur I. De plus $g'(\bar{x}) = 0$.

Preuve. *Comme f' est de signe constant sur I, f est strictement monotone sur I et l'équation $f(x)=0$ admet au plus une solution. Le point $\bar{x}$ est donc l'unique zéro de f sur I. L'égalité $g'(\bar{x})=0$ découle directement de la définition de la méthode de Newton (on a choisit λ dans la méthode de relaxation pour avoir cette égalité). Reste à prouver la convergence de la suite (x_n) vers $\bar{x}$. On remarque que quitte à changer f en $-f$, on peut supposer sans perte de généralité que $f''>0$ (le cas $f''<0$ se traitant de manière identique). On doit maintenant distinguer deux cas $f'>0$ et $f'<0$.*
•$f'>0$: f est donc strictement croissante sur I et comme $f(x_0)>0=f(\bar{x})$ ceci implique que $x_0>\bar{x}$. Faire le tableau de signe de g', g : on observe que g est décroissante pour $x<\bar{x}$ et croissante pour $x>\bar{x}$. Comme $x_1=g(x_0)>g(\bar{x})$, on a $x_1>\bar{x}$. D'autre part

$$f(x_0)>0, f'(x_0)>0 \Longrightarrow x_1=x_0-\frac{f(x_0)}{f'(x_0)}<x_0$$

donc on a $\bar{x}<x_1<x_0$. Par récurrence on montre que la suite (x_n) décroit. Comme elle est minorée par $\bar{x}$ elle converge et sa limite est l'unique point fixe de g sur I : $\bar{x}$.
• $f'<0$: les variations sont échangées, on montre que $x_0<x_1<\bar{x}$ et on obtient une suite (x_n) croissante majorée qui doit converger vers $\bar{x}$.

Remarque 4.4.4. Si on enlève l'hypothèse $f(x_0)$ de même signe que f'' : si $x_1 \in I$ alors on peut montrer que $f(x_1)$ est de même signe que f'' et donc on commence l'itération à partir de x_1.

- Interprétation géométrique : La définition de x_{n+1} peut se réécrire sous la forme

$$0=f'(x_n)(x_{n+1}-x_n)+f(x_n).$$

Ceci signifie que $(x_{n+1},0)$ est le point d'intersection de la tangente à f en x_n avec l'axe des abscisses.

Théorème 4.4.4. *(Convergence quadratique locale) Soit $I=[a,b]$, on considère la fonction f définie sur I de classe $\mathcal{C}^2$ et $\bar{x} \in I$ tel que $f(\bar{x})=0$ mais $f'(\bar{x}) \neq 0$. Soient m et $M>0$ tels que*

$$|f'(x)| \geq m, \quad |f''(x)| \leq M \quad \forall x \in I.$$

On pose $c=\frac{M}{2m}(b-a)$, Soit $x_0 \in I$. On suppose que la suite définie par la méthode de Newton à partir de x_0 est bien définie, est à valeurs dans I et converge vers $\bar{x}$. Alors on a :

- $|x_{n+1}-\bar{x}| \leq \frac{M}{m}\frac{(\bar{x}-x_n)^2}{2}$ *pour tout* $n \geq 0$,
- $|x_n-\bar{x}| \leq \frac{2m}{M}c^{2^n}$ *pour tout* $n \geq 0$.

- Si de plus $c<1$ alors

$$\forall \varepsilon>0, \quad n \geq \ln\left(\ln\left(\frac{M\varepsilon}{2m}\right)/\ln c\right)/\ln 2 \Rightarrow |x_n-\bar{x}| \leq \varepsilon.$$

Preuve. *La formule de Taylor-Lagrange à l'ordre 2 appliquée à f au point x_n s'écrit*

$$f(\bar{x})=f(x_n)+(\bar{x}-x_n)f'(x_n)+\frac{(\bar{x}-x_n)^2}{2}f''(c_n),$$

avec c_n dans l'intervalle entre x_n et $\bar{x}$. Comme $f(\bar{x})=0$, on obtient en divisant par $f'(x_n)$:

$$-\frac{f(x_n)}{f'(x_n)}-\bar{x}+x_n=\frac{(\bar{x}-x_n)^2}{2}\frac{f''(c_n)}{f'(x_n)},$$

c'est-à-dire

$$x_{n+1} - \bar{x} = \frac{(\bar{x} - x_n)^2}{2} \frac{f''(c_n)}{f'(x_n)},$$

Il en découle l'inégalité

$$|x_{n+1} - \bar{x}| \leq \frac{(\bar{x} - x_n)^2}{2} \frac{M}{m}.$$

Le deuxième point s'en déduit alors par simple récurrence :
-Initialisation : pour $n = 0$ on a bien l'inégalité suivante

$$|x_0 - \bar{x}| \leq (b - a) = \frac{2m}{M} c.$$

- Hérédité : on suppose qu'au rang n l'inégalité est vérifiée i.e.

$$|x_n - \bar{x}| \leq \frac{2m}{M} c^{2^n}.$$

D'après ce qui précède,

$$\begin{aligned} |x_{n+1} - \bar{x}| &\leq (\bar{x} - x_n)^2 \frac{M}{m} \\ &\leq \left(\frac{2m}{M} c^{2^n}\right)^2 \frac{M}{m} \\ &\leq \frac{2m}{M} c^{2^{n+1}}. \end{aligned}$$

Ceci permet de conclure la démonstration du second point du théorème. Supposons que $c < 1$ et soit $\varepsilon > 0$. D'après les estimations précédentes on a $|\bar{x} - x_n| < \varepsilon$ dès que $\frac{2m}{M} c^{2^n} < \varepsilon$ càd

$$n \ln 2 \geq \ln\left(\left(\frac{M\varepsilon}{2m}\right) / \ln c\right).$$

• **Méthode de la corde :** Cette méthode permet d'éviter qu'à chaque itération on ait à évaluer $f'(x_n)$. La méthode de la corde consiste à remplacer $f'(x_n)$ par $f'(x_0)$ ce qui donne

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)}.$$

- Interprétation géométrique : Le calcul des itérés s'effectue en prenant toujours la même pente $f'(x_0)$.

Théorème 4.4.5. *Supposons f continûment dérivable ($\mathcal{C}^1$) qui admet un zéro $\bar{x}$ tel que $f'(\bar{x}) \neq 0$ Alors il existe $\varepsilon > 0$ tel que si x_0 satisfait $|\bar{x} - x_0| \leq \varepsilon$, la suite (x_n) donné par la méthode de la corde converge vers $\bar{x}$. La convergence est linéaire.*

• **Méthode de la sécante :**

Toujours dans la même idée d'éviter le calcul de la dérivée de f, on peut faire l'approximation

$$f'(x_n) \simeq \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}},$$

ce qui nous définit la méthode de la sécante

$$\begin{cases} x_0, x_1 \text{ donnés dans } I, x_0 \neq x_1, \\ \forall n > 0, \quad x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} f(x_n). \end{cases}$$

Théorème 4.4.6. *Soit f de classe $\mathcal{C}^2$ qui admet un zéro $\bar{x}$ tel que $f'(\bar{x}) \neq 0$. Alors il existe $\varepsilon > 0$ tel que si x_0 satisfait $|\bar{x} - x_0| \leq \varepsilon$, la suite (x_n) définie par la méthode de la sécante converge vers $\bar{x}$ et on*

$$|\bar{x} - x_{n+1}| \leq C\, |\bar{x} - x_n|^{\varphi}\,,$$

où φ est le nombre d'or $\left(\varphi = \frac{1+\sqrt{5}}{2} \sim 1,62\right)$.

Références

1. Hairer, Cours d'analyse numérique, université de Genève
2. A. Magnus, Analyse numérique, cours UCL, Louvain la Neuve.
3. A. Quarteroni, R. Sacco, F. Saleri, Méthodes numériques pour le calcul scientifique, Springer Paris, 2000.
4. M. Schatzman, Analyse numérique pour la licence, Inter Édition, 1993.
5. J. Fabbri, Notes de cours analyse numérique : Université de Tours France.
6. j.P. Chehab, Interpolation polynomiale Université de Picardie Jules Vernes LAMFA, version du 2 février 2014.
7. T. Mekkaoui, Polycopié de cours : Méthodes numeriques, 2014-2015.
8. J. Stoer, R. Burlisch,Introduction to Numerical Analysis, text in Applied Mathematics, 12, Springer, 3rd Edition 2002.