Analyse numérique non linéaire

Fabien Priziac

Licence 3 Spécialité Mathématiques, année universitaire 2021-2022

Table des matières

Chapitre 1

Equations non linéaires

1.1 Introduction

Soient I un intervalle de $\mathbb{R}$ et $f : I \to \mathbb{R}$ sur fonction sur I. On souhaite résoudre *numériquement* l'équation

$$\text{(E)} \quad f(x) = 0, \ x \in I,$$

i.e. déterminer une approximation suffisamment proche pour une (chaque si possible) solution de (E).

Plus précisément, pour une erreur d'approximation tolérée de $\epsilon \in]0, +\infty[$, si $x \in I$ est une solution de (E), on souhaite être capable de construire un nombre $x_0 \in I$ tel que

- $|x - x_0| \leqslant \epsilon$,
- $|f(x_0)| \leqslant \widetilde{\epsilon}$,

où $\widetilde{\epsilon}$ est un autre paramètre d'erreur : on souhaite que l'approximation considérée de la solution x soit "presque une solution" de (E).

Une démarche classique consiste à construire à l'aide d'un algorithme une suite $(x_n)_{n\in\mathbb{N}}$ de nombres de I convergeant vers une solution x de (E). On représente alors la "vitesse" de convergence de la suite $(x_n)_{n\in\mathbb{N}}$ vers x par la notion d'*ordre de convergence* :

Définition 1.1.1. *Soit $k \in]1, +\infty[$. On dit que la convergence de la suite $(x_n)_{n\in\mathbb{N}}$ vers x est*

- *<u>linéaire</u> (ou <u>d'ordre 1</u>) s'il existe $C \in [0; 1[$, $N \in \mathbb{N}\backslash\{0\}$ tels que, pour tout entier naturel n au moins égal à N, $|x_{n+1} - x| \leqslant C\,|x_n - x|$,*
- *<u>d'ordre k</u> s'il existe $C \in [0; +\infty[$, $N \in \mathbb{N}\backslash\{0\}$ tels que, pour tout entier naturel n au moins égal à N, $|x_{n+1} - x| \leqslant C\,|x_n - x|^k$.*

Remarque 1.1.2.

- Une convergence d'ordre 2 est également appelée <u>convergence quadratique</u>.
- Si la convergence de la suite $(x_n)_{n\in\mathbb{N}}$ vers x est d'ordre $k \in]1, +\infty[$, alors, avec les notations ci-dessus, si $N \in \mathbb{N}$, on a, pour $p \in \mathbb{N}\backslash\{0\}$,

$$|x_{N+p} - x| \leqslant C^{1+k+\cdots+k^{p-1}} |x_N - x|^{k^p}$$

(on le montre par récurrence sur p). Or

$$C^{1+k+\cdots+k^{p-1}} = C^{\frac{1-k^p}{1-k}} = C^{\frac{k^p-1}{k-1}} = \left(C^{\frac{1}{k-1}}\right)^{k^p-1}$$

donc, si l'on note $\widetilde{C} := C^{\frac{1}{k-1}}$, on a

$$|x_{N+p} - x| \leqslant \frac{1}{\widetilde{C}} \left(\widetilde{C}|x_N - x|\right)^{k^p}.$$

Comme $x_n \underset{n\to+\infty}{\longrightarrow} x$, on sait qu'il existe un rang $N \in \mathbb{N}$ tel que, pour tout entier n au moins égal à N, $|x_N - x| < \frac{1}{\widetilde{C}}$ et donc $\widetilde{C}|x_N - x| < 1$: on a alors $|x_{N+p} - x| \leqslant \frac{1}{\widetilde{C}} \left(\widetilde{C}|x_N - x|\right)^{k^p} \underset{p\to+\infty}{\longrightarrow} 0$ et on n'a donc pas besoin de "limiter l'amplitude" de C, contrairement au cas de la convergence linéaire.

Tous les algorithmes mis en place pour résoudre *numériquement* l'équation (E), i.e. pour déterminer une solution approchée de (E) à des erreurs prescrites près, nécessitent des *conditions d'arrêt* qui assurent que la valeur fournie par l'algorithme est bien une solution *numérique* de (E) et/ou qui évitent les temps de calculs trop longs.

Dans ce chapitre, nous allons décrire de tels algorithmes et nous discuterons de leurs conditions de convergence, de leurs ordres de convergence et de leurs dépendances aux *conditions initiales*, i.e. à l'initialisation de la suite récurrente construite.

1.2 Méthode de dichotomie

Soient $a, b \in \mathbb{R}$ avec $a < b$ et soit $f : [a, b] \to \mathbb{R}$ sur fonction continue sur le segment $[a, b]$. Supposons que $f(a)f(b) < 0$: comme les évaluations $f(a)$ et $f(b)$ sont de signes strictement différents, le théorème des valeurs intermédiaires assure qu'il existe (au moins) un réel $c \in [a, b]$ tel que $f(c) = 0$, autrement dit l'équation

$$\text{(E)} \quad f(x) = 0, \ x \in [a, b]$$

possède (au moins) une solution.

La méthode dite *de dichotomie* consiste alors à construire une suite d'intervalles fermés $I_n = [u_n, v_n]$, $n \in \mathbb{N}$, inclus dans $[a, b]$ (autrement dit, pour tout $n \in \mathbb{N}$, $a \leqslant u_n \leqslant v_n \leqslant b$), telle que les suites $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ soient adjacentes et convergent vers une solution de (E).

L'algorithme constructif est le suivant :

- On pose $u_0 := a$, $v_0 := b$ et $I_0 := [u_0, v_0]$.
- Supposons que, pour $n \in \mathbb{N}$, le segment $I_n = [u_n, v_n]$ ait été construit, avec $f(u_n)f(v_n) < 0$. On calcule alors $\alpha_n := \frac{u_n+v_n}{2}$, et :
  - Si $f(\alpha_n) = 0$, on retourne la valeur α_n,
  - Sinon,

∗ Si $f(u_n)f(\alpha_n) < 0$, on pose $u_{n+1} := u_n$ et $v_{n+1} := \alpha_n$ et $I_{n+1} := [u_{n+1}, v_{n+1}]$
∗ Sinon, on pose $u_{n+1} := \alpha_n$ et $v_{n+1} := v_n$ et $I_{n+1} := [u_{n+1}, v_{n+1}]$

Le résultat suivant énonce que la méthode de dichotomie aboutit toujours à une solution au moins approchée de l'équation (E) :

Proposition 1.2.1. *On suppose que pour tout $n \in \mathbb{N}$, $f(\alpha_n) \neq 0$. Il existe alors $x_0 \in [a, b]$ tel que*

- $f(x_0) = 0$,
- *les suites $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ sont adjacentes et convergent vers x_0.*

Soient $\epsilon, \widetilde{\epsilon} \in]0, +\infty[$. Il existe $N \in \mathbb{N}$ tel que, pour tout entier naturel n au moins égal à N,

- $v_n - u_n \leqslant \epsilon$ *et donc* $|u_n - x_0| \leqslant \epsilon$ *et* $|v_n - x_0| \leqslant \epsilon$,
- $|f(u_n)| \leqslant \widetilde{\epsilon}$ *et* $|f(v_n)| \leqslant \widetilde{\epsilon}$.

Démonstration. Montrons tout d'abord que les suites $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ sont adjacentes : si $n \in \mathbb{N}$, on a $u_{n+1} \geqslant u_n$ et $v_{n+1} \leqslant v_n$, ainsi que $v_n \geqslant u_n$ et, si $n \in \mathbb{N}\backslash\{0\}$,

$$v_n - u_n = \frac{v_{n-1} - u_{n-1}}{2} = \cdots = \frac{v_0 - u_0}{2^n} \underset{n\to+\infty}{\longrightarrow} 0.$$

Notons $x_0 \in \mathbb{R}$ la limite commune des suites adjacentes $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$. On a $x_0 \in [a, b]$ car, pour tout $n \in \mathbb{N}$, $a \leqslant u_n \leqslant v_n \leqslant b$. De plus, pour tout $n \in \mathbb{N}$, $f(u_n)f(v_n) < 0$ donc, comme f est continue, $u_n \underset{n\to+\infty}{\longrightarrow} x_0$ et $u_n \underset{n\to+\infty}{\longrightarrow} x_0$, $f(x_0)f(x_0) \leqslant 0$ i.e. $(f(x_0))^2 \leqslant 0$. Ainsi, nécessairement, $(f(x_0))^2 = 0$ i.e. $f(x_0) = 0$.

Enfin, soit $N_1 \in \mathbb{N}$ tel que, pour tout n au moins égal à N_1, $v_n - u_n \leqslant \epsilon$ alors

$$|u_n - x_0| = x_0 - u_n \leqslant v_n - u_n \leqslant \epsilon$$

et

$$|v_n - x_0| = v_n - x_0 \leqslant v_n - u_n \leqslant \epsilon.$$

Soit maintenant $N_2 \in \mathbb{N}$ tel que, pour tout n au moins égal à N_2, $|f(u_n) - f(x_0)| \leqslant \widetilde{\epsilon}$ i.e. $|f(u_n)| \leqslant \widetilde{\epsilon}$ (f est continue donc $f(u_n) \underset{n\to+\infty}{\longrightarrow} f(x_0) = 0$).

Soit enfin $N_3 \in \mathbb{N}$ tel que, pour tout n au moins égal à N_3, $|f(v_n) - f(x_0)| \leqslant \widetilde{\epsilon}$ i.e. $|f(v_n)| \leqslant \widetilde{\epsilon}$ (f est continue donc $f(v_n) \underset{n\to+\infty}{\longrightarrow} f(x_0) = 0$).

Si on note alors $N := \max(N_1, N_2, N_3)$, on obtient alors le résultat voulu avec les notations de l'énoncé. □

Une proposition d'algorithme d'approximation numérique concret mettant en jeu la méthode de dichotomie appliquée à la fonction continue $f : [a, b] \to \mathbb{R}$ (sous l'hypothèse que $f(a)f(b) < 0$) est la suivante (on suppose également fixées les erreurs d'approximations ϵ et $\widetilde{\epsilon}$) :

$u \leftarrow a$
$v \leftarrow b$
$\alpha \leftarrow u$
while $v - u > \epsilon$ ou $|f(u)| > \widetilde{\epsilon}$ ou $|f(v)| > \widetilde{\epsilon}$ **do**
  $\alpha \leftarrow \frac{u+v}{2}$
  if $f(\alpha) = 0$ **then**
    Return α
    Stop
  else
    if $f(u)f(\alpha) < 0$ **then**
      $v \leftarrow \alpha$
    else
      $u \leftarrow \alpha$
    end if
  end if
end while
Return u (ou v)

La vitesse de convergence de la méthode de dichotomie est donnée par l'ordre de convergence de la suite $(v_n - u_n)_{n\in\mathbb{N}}$ vers 0 : c'est cette convergence qui conditionne l'arrêt de l'algorithme.

Proposition 1.2.2. *La convergence de la suite $(v_n - u_n)_{n\in\mathbb{N}}$ vers 0 est linéaire.*

Démonstration. Soit $n \in \mathbb{N}$, on a $v_{n+1} - u_{n_1} = \alpha_n - u_n = v_n - \alpha_n = \frac{1}{2}(v_n - u_n)$ □

De plus, l'égalité ci-dessus nous permet d'affirmer que, pour tout $n \in \mathbb{N}$ $v_n - u_n = \frac{b-a}{2^n}$: à chaque étape $n \in \mathbb{N}$ de la méthode de dichotomie, on a donc un écart de v_n et u_n à la solution x_0 majoré par $\frac{b-a}{2^n}$. Si ϵ est l'erreur tolérée pour x_0, on sait donc que l'on obtiendra une approximation de x_0 à ϵ près au bout de n étapes avec $\frac{b-a}{2^n} \leqslant \epsilon$.

La convergence de la méthode de dichotomie vers une solution de (E), bien que certaine, reste cependant lente, comparativement aux méthodes que nous allons présenter dans la suite.

1.3 Méthode du point fixe

On suppose ici que la fonction $f : [a, b] \to \mathbb{R}$ est de classe $\mathcal{C}^1$ sur le segment $[a, b]$. Soit $g : [a, b] \to \mathbb{R}$ une fonction telle que, pour tout $x \in [a, b]$, $f(x) = 0$ ssi $g(x) = x$: dans cette section, pour déterminer numériquement une solution de (E), nous allons appliquer une méthode dite *de point fixe* à la fonction g.

Une telle fonction g existe toujours : on peut considérer par exemple les fonctions

- $g : [a, b] \to \mathbb{R}$; $x \mapsto f(x) + x$,
- $g : [a, b] \to \mathbb{R}$; $x \mapsto x - f(x)$,
- $g : [a, b] \to \mathbb{R}$; $x \mapsto \lambda f(x) + x$, avec $\lambda \in \mathbb{R}\backslash\{0\}$,

d'autres fonctions pouvant également être considérées suivant l'expression et les propriétés de f, tant qu'elles vérifient les hypothèses de la méthode de point fixe suivante :

Théorème 1.3.1. *Soit $h : [a,b] \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$ sur $[a,b]$ telle que*

- $h([a,b]) \subset [a,b]$,
- $K := \sup_{x\in[a,b]} |h'(x)| = \max_{x\in[a,b]} |h'(x)| < 1$.

Alors h possède un unique point fixe dans $[a,b]$ i.e. un unique point $y \in [a,b]$ tel que $h(y) = y$.

De plus, si $y_0 \in [a,b]$ et si on note $(x_n)_{n\in\mathbb{N}}$ la suite récurrente définie par

$$\begin{cases} x_0 := y_0, \\ \text{pour tout } n \in \mathbb{N},\ x_{n+1} := h(x_n), \end{cases}$$

alors

- *la suite $(x_n)_{n\in\mathbb{N}}$ converge vers y,*
- *pour tout $n \in \mathbb{N}$, $|x_{n+1} - y| \leqslant K|x_n - y|$.*

Démonstration. Montrons tout d'abord que h possède un point fixe : si on note $\widetilde{h}$ la fonction $[a,b] \to \mathbb{R}$; $x \mapsto h(x) - x$, on a $\widetilde{h}(a) = h(a) - a \geqslant 0$ (car $h(a) \in [a,b]$) et $\widetilde{h}(b) = h(b) - b \leqslant 0$ (car $h(b) \in [a,b]$), donc, d'après le théorème des valeurs intermédiaires (h est continue sur $[a,b]$ donc $\widetilde{h}$ est continue sur $[a,b]$), il existe $y \in [a,b]$ tel que $\widetilde{h}(y) = 0$ i.e. $h(y) = y$.

Montrons ensuite que ce point fixe est unique : soient y_1, y_2 deux points fixes de h et supposons par l'absurde que $y_1 \neq y_2$. Comme h est de classe $\mathcal{C}^1$ sur $[a,b]$, d'après le théorème des accroissements finis, il existe $\xi \in]a,b[$ tel que

$$y_1 - y_2 = h(y_1) - h(y_2) = h'(\xi)(y_1 - y_2),$$

et donc $h'(\xi) = 1$. Mais, par hypothèse, $|h'(\xi)| < 1$, d'où une contradiction et donc $y_1 = y_2$.

Soit enfin $n \in \mathbb{N}$ et soit $\xi_n \in]a,b[$ tel que $h(x_n) - h(y) = h'(\xi_n)(x_n - y)$. On a

$$|x_{n+1} - y| = |h(x_n) - h(y)| = \left|h'(\xi_n)(x_n - y)\right| \leqslant K|x_n - y|,$$

et donc

$$|x_{n+1} - y| \leqslant K^{n+1}|x_0 - y| \underset{n\to+\infty}{\longrightarrow} 0$$

(car $0 \leqslant K < 1$). En particulier, $x_n \xrightarrow[n\to+\infty]{} y$. □

Ainsi, si la fonction g vérifie les hypothèses du théorème de point fixe 1.3.1 (c'est-à-dire si g est de classe $\mathcal{C}^1$ sur $[a,b]$, $g([a,b]) \subset [a,b]$ et $\max_{x\in[a,b]} g'(x) < 1$), l'équation

$$\text{(E)} \quad f(x) = 0,\ x \in [a,b] \Leftrightarrow g(x) = x,\ x \in [a,b]$$

possède une unique solution y et on peut approcher y en choisissant un point y_0 de $[a, b]$ et en calculant les termes de la suite

$$\begin{cases} x_0 := y_0, \\ \text{pour tout } n \in \mathbb{N},\ x_{n+1} := g(x_n), \end{cases}$$

la convergence de $(x_n)_{n\in\mathbb{N}}$ vers y étant linéaire.

Une question importante est maintenant celle de la condition d'arrêt de cet algorithme. Considérons la propriété suivante :

Proposition 1.3.2. *Soit $n \in \mathbb{N}$. On a*

$$|x_n - y| \leqslant \frac{1}{1-K}|x_{n+1} - x_n|.$$

Démonstration. D'après le théorème des accroissements finis, il existe $\eta_n \in]a, b[$ tel que

$$x_{n+1} - y = g(x_n) - g(y) = g'(\eta_n)(x_n - y)$$

On a donc

$$x_{n+1} - x_n = x_{n+1} - y + y - x_n = g'(\eta_n)(x_n - y) - (x_n - y) = \big(g'(\eta_n) - 1\big)(x_n - y)$$

et ainsi

$$|x_n - y| = \frac{1}{1 - g'(\eta_n)}|x_{n+1} - x_n| \leqslant \frac{1}{1-K}|x_{n+1} - x_n|$$

(on a $|g'(\eta_n)| \leqslant K < 1$). □

Une condition d'arrêt raisonnable est ainsi de tester à chaque étape $n \in \mathbb{N}$ de l'algorithme si l'écart entre x_{n+1} et x_n est suffisamment petit par rapport à une erreur tolérée fixée préalablement. Cependant, l'écart entre x_n et le point approché x_0 peut, si K est très proche de 1, demeurer très grand d'après la propriété précédente. Il faut donc en tenir compte lorsque l'on définit ce seuil d'erreur, afin de compenser le facteur $\frac{1}{1-K}$.

Si l'on fixe ainsi une erreur d'arrêt convenable $\epsilon \in]0, +\infty[$ pour l'écart $|x_{n+1} - x_n|$ ainsi qu'une erreur tolérée $\widetilde{\epsilon} \in]0, +\infty[$ pour l'écart $|f(x_n) - 0|$, voici une proposition d'algorithme mettant en œuvre la méthode de point fixe décrite ci-dessus. On choisit au préalable un point de départ y_0 de $[a, b]$:

$x \leftarrow y_0$
$\widetilde{x} \leftarrow g(x)$
while $|\widetilde{x} - x| > \epsilon$ ou $|f(x)| > \widetilde{\epsilon}$ **do**
  $x \leftarrow \widetilde{x}$
  $\widetilde{x} \leftarrow g(x)$
end while
Return x

Remarque 1.3.3. Le "temps de convergence" peut dépendre fortement du point initial y_0 choisi dans $[a, b]$.

1.4 Méthode de Newton

Dans cette section, on suppose que la fonction $f : [a, b] \to \mathbb{R}$ est de classe $\mathcal{C}^2$, et on suppose (que l'on a montré au préalable) qu'il existe un point y de $[a, b]$ tel que $f(y) = 0$ et $f'(y) \neq 0$.

On commence par se restreindre à un segment $[\alpha, \beta] \subset [a, b]$ sur lequel (on a montré au préalable que) f' ne s'annule pas et f s'annule en (au moins) un point. On suppose également que f ne s'annule pas en α et β. On considère ensuite la fonction

$$g : \begin{array}{ccc} [\alpha, \beta] & \to & \mathbb{R} \\ x & \mapsto & x - \frac{f(x)}{f'(x)} \end{array},$$

bien définie car f' ne s'annule pas sur $[\alpha, \beta]$. Remarquons alors que, pour $x \in [\alpha, \beta]$, $f(x) = 0$ ssi $g(x) = x$.

Nous avons ainsi ramené notre problème initial à une problème de point fixe pour g, et nous allons nous placer dans des conditions propices à l'application du théorème 1.3.1.

Commençons par remarquer que, comme f est de classe $\mathcal{C}^2$ sur $[a, b]$, g est de classe $\mathcal{C}^1$ sur $[\alpha, \beta]$. Ensuite, soit $x \in [\alpha, \beta]$, on a

$$g'(x) = 1 - \frac{\big(f'(x)\big)^2 - f(x)f''(x)}{\big(f'(x)\big)^2} = \frac{f(x)f''(x)}{\big(f'(x)\big)^2}.$$

En particulier, si y est un point de $[\alpha, \beta]$ tel que $f(y) = 0$, on a $g'(y) = 0$: soit $L \in [0, 1[$, comme g' est continue sur $[\alpha, \beta]$, il existe $\delta \in]0, +\infty[$ tel que $[y-\delta, y+\delta] \subset [\alpha, \beta]$ et, si $x \in [y-\delta, y+\delta]$, $|g'(x)| \leqslant L$. De plus, $g([y - \delta, y + \delta]) \subset [y - \delta, y + \delta]$ car, si $x \in [y - \delta, y + \delta]$ i.e. $|x - y| \leqslant \delta$, il existe, par le théorème des accroissements finis, $\xi \in]y-\delta, y+\delta[$ tel que $g(x)-g(y) = g'(\xi)(x-y)$ donc

$$|g(x) - y| = |g(x) - g(y)| = |g'(\xi)||x - y| \leqslant |x - y| \leqslant \delta$$

($\xi \in [y - \delta, y + \delta]$ donc $|g'(\xi)| \leqslant L < 1$). Comme de plus, $\max\limits_{x\in[y-\delta,y+\delta]} g'(x) \leqslant L < 1$, on peut appliquer le théorème 1.3.1 à la fonction $g : [y - \delta, y + \delta] \to \mathbb{R}$.

La méthode de Newton consiste alors à calculer les termes de la suite récurrente $(x_n)_{n\in\mathbb{N}}$ définie par

$$\begin{cases} x_0 := y_0 \\ \text{pour tout } n \in \mathbb{N},\ x_{n+1} := g(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}, \end{cases}$$

où le premier terme y_0 est choisi dans l'intervalle $[y - \delta, y + \delta]$. La suite $(x_n)_{n\in\mathbb{N}}$ converge alors vers le point y.

Remarque 1.4.1. La difficulté de la méthode de Newton est de choisir un point y_0 qui soit "suffisamment proche" de y.

Le fait que la méthode de Newton soit *locale* est compensé par la convergence quadratique de la suite $(x_n)_{n\in\mathbb{N}}$ vers y :

Théorème 1.4.2. *Il existe $C \in [0, +\infty[$ et $N \in \mathbb{N}$ tels que, pour tout entier naturel n au moins égal à N, $|x_{n+1} - y| \leqslant C\,|x_n - y|^2$.*

Démonstration. Soit $n \in \mathbb{N}$. D'après le théorème de Taylor-Lagrange, il existe $\xi_n \in]y-\delta, y+\delta[$ tel que

$$0 = f(y) = f(x_n) + f'(x_n)(y - x_n) + \frac{f''(\xi_n)}{2}(y - x_n)^2$$

et donc

$$y = x_n - \frac{f(x_n)}{f'(x_n)} - \frac{f''(\xi_n)}{2f'(x_n)}(y - x_n)^2 = g(x_n) - \frac{f''(\xi_n)}{2f'(x_n)}(y - x_n)^2 = x_{n+1} - \frac{f''(\xi_n)}{2f'(x_n)}(y - x_n)^2,$$

ce que l'on écrit sous la forme

$$x_{n+1} - y = \frac{f''(\xi_n)}{2f'(x_n)}(x_n - y)^2.$$

Si on note alors $M := \max\limits_{x\in[y-\delta,y+\delta]} |f''(x)|$ et $m := \min\limits_{x\in[y-\delta,y+\delta]} |f'(x)|$ (f' et f'' sont continues sur le segment $[y-\delta, y+\delta]$ donc sont bornées et atteignent leurs bornes sur ce segment, et f' ne s'annule pas sur ce segment par hypothèse), on a

$$|x_{n+1} - y| \leqslant \frac{M}{2m}|x_n - y|^2.$$

□

Remarque 1.4.3.

- L'interprétation géométrique de la méthode de Newton est la suivante. Pour $n \in \mathbb{N}$, le point x_{n+1} est l'abscisse du point d'intersection de la tangente à la courbe représentative de f au point $\big(x_n, f(x_n)\big)$ avec l'axe des abscisses. En effet, cette tangente a pour équation $z = (x - x_n)f'(x_n) + f(x_n)$ et le point d'intersection $(x_{n+1}, 0)$ de cette droite avec l'axe des abscisses vérifie

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

  i.e.

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

- Un inconvénient de la méthode de Newton est qu'elle nécessite un calcul d'évaluation de la fonction dérivée f' en les points x_n, $n \in \mathbb{N}$, ce qui peut s'avérer difficile. On peut éventuellement approcher cette évaluation à l'aide du taux d'accroissement mais il faudra alors en tenir compte dans l'algorithme de point fixe associé. Précisément, si η est l'erreur tolérée sur le calcul de tous les points x_n, $n \in \mathbb{N}$, et si, pour tout $n \in \mathbb{N}$, on note $\widetilde{x}_n$ l'approximation de x_n, on a

  $$|\widetilde{x}_n - y| \leqslant |x_n - y| + |\widetilde{x}_n - x_n| \leqslant L^n|x_0 - y| + \eta.$$

- Il existe une méthode de recherche d'un zéro de f qui ne passe pas par le calcul de la dérivée de f mais par le calcul de taux d'accroissements. Cette méthode, appelée méthode de la sécante consiste à calculer les termes de la suite $(x_n)_{n\in\mathbb{N}}$ donnée par

  $$\begin{cases} x_0 \in [a, b], \\ x_1 \in [a, b], \\ \text{pour tout } n \in \mathbb{N}\backslash\{0\},\ x_{n+1} := x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}, \end{cases}$$

(pour $n \in \mathbb{N}\backslash\{0\}$, x_{n+1} est l'abscisse du point d'intersection de l'axe des abscisses avec la droite d'équation $z = \frac{f(x_n)-f(x_{n-1})}{x_n-x_{n-1}}(x - x_n) + f(x_n)$ i.e. l'unique droite passant par les points $(x_{n-1}, f(x_{n-1}))$ et $(x_n, f(x_n)))$.

Toujours sous l'hypothèse que f est de classe $\mathcal{C}^2$ sur $[a, b]$, si y est un point de $[a, b]$ tel que $f(y) = 0$ et $f'(y) \neq 0$, on peut montrer qu'il existe $\delta \in]0, +\infty[$ tel que si $x_0, x_1 \in [y - \delta, y + \delta]$, alors la suite $(x_n)_{n\in\mathbb{N}}$ est bien définie et converge vers y, et l'ordre de cette convergence est $\frac{1+\sqrt{5}}{2}$.

Chapitre 2

Interpolation polynomiale

2.1 Introduction

Interpoler une fonction $f : I \to \mathbb{R}$ définie sur un intervalle I de $\mathbb{R}$ par une fonction polynomiale consiste à considérer une fonction polynomiale prenant les mêmes valeurs que f en un nombre fini de points de I. L'objectif est d'utiliser l'interpolation polynomiale pour approcher "le mieux possible" la fonction f par une fonction polynomiale, plus simple à évaluer.

2.2 Méthode de Horner

On commence ce chapitre par une section consacrée à une méthode efficace d'évaluation d'un polynôme de $\mathbb{R}[X]$ en un nombre réel : tant qu'à approcher une fonction réelle par une fonction polynomiale, autant savoir évaluer celle-ci efficacement.

Soit donc $P = a_0 + a_1X + \cdots + a_nX^n$ un polynôme de $\mathbb{R}[X]$ et soit $x \in \mathbb{R}$. Une méthode basique d'évaluation de P en x consiste en l'algorithme suivant :

$r \leftarrow 0$
for $i = 0$ à n **do**
  $r := r + a_i x^i$
end for
Return r

A chaque étape i de cet algorithme, $i \in \{0, \ldots, n\}$, i produits et une somme sont effectués, soit au total $\sum_{i=0}^{n} i = \frac{n(n+1)}{2}$ produits et n sommes : la complexité de cet algorithme est ainsi en $O(n^2)$.

L'algorithme suivant, dit méthode de Horner, repose quant à lui sur l'écrire suivante de $P(x)$: on a

$$P(x) = a_0 + a_1x + \cdots + a_nx^n = a_0 + x\big(a_1 + x\big(\cdots + x\big(a_{n-1} + xa_n\big)\cdots\big)\big).$$

La méthode de Horner consiste alors en l'algorithme suivant :

$r \leftarrow a_n$
for $i = n-1$ à 0 **do**
  $r \leftarrow a_i + xr$
end for

A chaque étape de l'algorithme, un seul produit et une seule somme sont effectués, soit au total n produits et n sommes : la complexité de la méthode de Horner est ainsi en $O(n)$.

2.3 Interpolation polynomiale et méthode de Lagrange

Interpoler une fonction réelle par une fonction polynomiale consiste à considérer une fonction polynomiale qui *interpole* un nombre fini de points du graphe de la fonction considérée, i.e. une fonction polynomiale dont le graphe passe par ces points.

Soit $n \in \mathbb{N}$ et soient $c_0, \dots, c_n \in \mathbb{R}$ deux à deux distincts et $y_0, \dots, y_n \in \mathbb{R}$. La méthode d'interpolation polynomiale dite de Lagrange consiste justement à construire un polynôme $P \in \mathbb{R}[X]$ de degré au plus n tel que, pour tout $i \in \{0, \dots, n\}$, $P(c_i) = y_i$.

Le théorème suivant énonce l'existence et l'unicité d'un tel polynôme, et donne un moyen de le construire : ce polynôme est donné par les polynômes de Lagrange associés aux centres $c_0, \dots, c_n$ définis par, pour $i \in \{0, \dots, n\}$,

$$L_i := \prod_{0 \leqslant k \leqslant n,\ k \neq i} \frac{X - c_k}{c_i - c_k} \in \mathbb{R}_n[X],$$

où $\mathbb{R}_n[X]$ désigne le $\mathbb{R}$-espace vectoriel des polynômes de $\mathbb{R}[X]$ de degré au plus n.

Théorème 2.3.1. *1. Pour tous $i, j \in \{0, \dots, n\}$, on a $L_i(c_j) = \delta_{i,j}$.*

2. Si on note $L := \sum_{i=0}^{n} y_i L_i \in \mathbb{R}_n[X]$, on a, pour tout $j \in \{0, \dots, n\}$, $L(c_j) = y_j$.

3. Si $P \in \mathbb{R}_n[X]$ vérifie, pour tout $j \in \{0, \dots, n\}$, $P(c_j) = y_j$, alors $P = L$.

Démonstration. Soient $i, j \in \{0, \dots, n\}$, on a

$$L_i(c_j) = \prod_{0 \leqslant k \leqslant n,\ k \neq i} \frac{c_j - c_k}{c_i - c_k} = \begin{cases} 0 & \text{si } j \neq i, \\ \displaystyle\prod_{0 \leqslant k \leqslant n,\ k \neq i} \frac{c_i - c_k}{c_i - c_k} = 1 & \text{si } j = i, \end{cases}$$

et donc

$$L(c_j) = \sum_{i=0}^{n} y_i L_i(c_j) = y_j L_j(c_j) = y_j.$$

Soit maintenant $P \in \mathbb{R}_n[X]$ tel que pour tout $j \in \{0, \dots, n\}$, $P(c_j) = y_j$, alors, pour tout $j \in \{0, \dots, n\}$, $(P - L)(c_j) = P(c_j) - L(c_j) = y_j - y_j = 0$: le polynôme $P - L$, de degré au plus n, possède $n + 1$ racines distinctes, et il s'agit donc du polynôme nul i.e. $P = L$. □

Remarquons également que la famille $\{L_0, \dots, L_n\}$ est une base du $\mathbb{R}$-espace vectoriel $\mathbb{R}_n[X]$. Il s'agit en effet d'une famille libre de $n+1$ éléments de $\mathbb{R}_n[X]$, qui est de dimension $n+1$ (la base dite *canonique* de $\mathbb{R}_n[X]$ est la famille $\{1, X, \dots, X^n\}$) : soient $\lambda_0, \dots, \lambda_n \in \mathbb{R}$ tels que $\sum_{i=0}^{n} \lambda_i L_i = 0$, alors, pour tout $j \in \{1, \dots, n\}$, $0 = \sum_{i=0}^{n} \lambda_i L_i(c_j) = \lambda_j$.

C'est pourquoi la famille $\{L_0, \dots, L_n\}$ des polynômes de Lagrange associés aux centres $c_0, \dots, c_n$ est également appelée base de Lagrange associée aux centres $c_0, \dots, c_n$.

Remarque 2.3.2. Si on note, pour tout $j \in \{0, \dots, n\}$,

$$\varphi_j : \begin{array}{ccc} \mathbb{R}_n[X] & \to & \mathbb{R} \\ P & \mapsto & P(c_j) \end{array},$$

l'application d'évaluation en c_j, alors la famille $\{\varphi_0, \dots, \varphi_n\}$ de $\mathcal{L}(\mathbb{R}_n[X], \mathbb{R}) = (\mathbb{R}_n[X])^*$ est la base duale de la base $\{L_0, \dots, L_n\}$ de $\mathbb{R}_n[X]$, autrement dit la famille $\{L_0, \dots, L_n\}$ de $\mathbb{R}_n[X]$ est la base antéduale de la base $\{\varphi_0, \dots, \varphi_n\}$ de $(\mathbb{R}_n[X])^*$.

Soit maintenant I un intervalle non réduit à un point de $\mathbb{R}$ et soit $f : I \to \mathbb{R}$ une fonction. Supposons également que les réels $c_0, \dots, c_n$ sont tous dans I. On cherche à construire un polynôme P de degré au plus n tel que pour tout $i \in \{1, \dots, n\}$, $P(c_i) = f(c_i)$: le théorème 2.3.1 montre l'existence et l'unicité d'un tel polynôme, et nous donne également un moyen de le construire de la base de Lagrange associée aux centres $c_0, \dots, c_n$

Définition 2.3.3. *On note $P_{f,c_0,\dots,c_n}$ l'unique polynôme de $\mathbb{R}_n[X]$ tel que pour tout $i \in \{1, \dots, n\}$, $P(c_i) = f(c_i)$: $P_{f,c_0,\dots,c_n}$ est appelé polynôme d'interpolation de f aux centres $c_0, \dots, c_n$.*

Proposition 2.3.4. *On a*

$$P_{f,c_0,\dots,c_n} = \sum_{i=0}^{n} f(c_i) L_i.$$

Remarque 2.3.5. Pour toute permutation σ sur l'ensemble $\{0, \dots, n\}$, $P_{f,c_0,\dots,c_n} = P_{f,c_{\sigma(0)},\dots,c_{\sigma(n)}}$.

Exemple 2.3.6. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction telle que $f(2) = 3$ et $f(5) = -6$. La base de Lagrange associée aux centres 2 et 5 est formée des polynômes $L_0 := \frac{X-5}{2-5} = \frac{1}{3}(5 - X)$ et $L_1 := \frac{X-2}{5-2} = \frac{1}{3}(X - 2)$, et le polynôme d'interpolation de Lagrange de f aux centres 2 et 5 est donc

$$3L_0 - 6L_1 = (5 - X) - 2(X - 2) = -3X + 9.$$

La méthode dite d'interpolation de Lagrange consiste ainsi à calculer le polynôme d'interpolation $P_{f,c_0,\dots,c_n}$ de f aux centres $c_0, \dots, c_n$ à l'aide de la base de Lagrange $\{L_0, \dots, L_n\}$ associée aux centres $c_0, \dots, c_n$. Un défaut de cette méthode est cependant que l'ajout d'un centre d'interpolation supplémentaire c_{n+1} change complètement la base de Lagrange associée au nouvel $n+2$-uplet $c_0, \dots, c_n, c_{n+1}$: plus précisément, les $n+1$ premiers polynômes de Lagrange associés aux centres $c_0, \dots, c_n, c_{n+1}$ sont différents des polynômes $L_0, \dots, L_n$ ci-dessus.

Nous allons introduire ci-dessous une autre base de $\mathbb{R}_n[X]$ qui permet de pallier ce défaut.

2.4 Méthode de Newton

On reprend les notations de la section précédente et on note $H_0 := 1 \in \mathbb{R}[X]$ et, pour $i \in \{1, \dots, n\}$,

$$H_i := \prod_{j=0}^{i-1}(X - c_j) = (X - c_0) \cdots (X - c_{i-1}).$$

Comme, pour $i \in \{0, \dots, n\}$, $\deg(H_i) = i$, la famille $\{H_0, \dots, H_n\}$ est une base de $\mathbb{R}_n[X]$, appelée base de Newton associée aux centres $c_0, \dots, c_n$: il existe donc $\alpha_0, \dots, \alpha_n \in \mathbb{R}$ uniques tels que

$$P_{f,c_0,\dots,c_n} = \sum_{i=0}^{n} \alpha_i H_i.$$

Soit $k \in \{0, \dots, n\}$. Nous allons montrer que le polynôme $\sum_{i=0}^{k} \alpha_i H_i$ est le polynôme d'interpolation de f aux centres $c_0, \dots, c_k$. En particulier, si $n \in \mathbb{N}\backslash\{0\}$, $P_{f,c_0,\dots,c_n}$ peut être calculé à partir de $P_{f,c_0,\dots,c_{n-1}}$ en y ajoutant $\alpha_n H_n$ (nous verrons par la suite une méthode algorithmique de calcul des coefficients $\alpha_0, \dots, \alpha_n$).

Proposition 2.4.1. *On a* $P_{f,c_0,\dots,c_k} = \sum_{i=0}^{k} \alpha_i H_i$.

Démonstration. Le polynôme $\sum_{i=0}^{k} \alpha_i H_i$ est de degré au plus k et, si $j \in \{0, \dots, k\}$,

$$\left(\sum_{i=0}^{k} \alpha_i H_i\right)(c_j) = \sum_{i=0}^{k} \alpha_i H_i(c_j) = \sum_{i=0}^{n} \alpha_i H_i(c_j) = \left(\sum_{i=0}^{n} \alpha_i H_i\right)(c_j) = P_{f,c_0,\dots,c_n}(c_j) = f(c_j)$$

(pour tout $i \in \{k+1, \dots, n\}$, $X - c_j$ divise H_i, car $0 \leqslant j \leqslant k$, donc, pour tout $i \in \{k+1, \dots, n\}$, $H_i(c_j) = 0$).

Par unicité du polynôme d'interpolation de f aux centres $c_0, \dots, c_k$, on a donc bien

$$\sum_{i=0}^{k} \alpha_i H_i = P_{f,c_0,\dots,c_k}.$$

□

Remarquons en particulier que, pour tout $i \in \{0, \dots, n\}$, α_i ne dépend donc que de f et $c_0, \dots, c_i$ (pas de $c_{i+1}, \dots, c_n$).

Définition 2.4.2. *Pour tout $i \in \{0, \dots, n\}$, on note $f[c_0, \dots, c_i] := \alpha_i$ et on appelle ce nombre réel la $i^{ème}$ différence divisée de f en les centres $c_0, \dots, c_i$.*

La méthode de Newton consiste à calculer le polynôme d'interpolation de f aux centres $c_0, \dots, c_n$ comme la somme

$$\sum_{i=0}^{n} f[c_0, \dots, c_i] H_i.$$

Les relations suivantes vont nous permettre de calculer par récurrence les différences divisées de f aux centres $c_0, \dots, c_i$ pour tout $i \in \{0, \dots, n\}$:

Proposition 2.4.3. *On a*

$$f[c_0] = f(c_0)$$

et si $i \in \{1, \dots, n\}$,

$$f[c_0, \dots, c_i] = \frac{f[c_1, \dots, c_i] - f[c_0, \dots, c_{i-1}]}{c_i - c_0}$$

Démonstration. Le polynôme de Lagrange associé au centre c_0 est 1 donc $P_{f,c_0} = f(c_0) \times 1 = f(c_0)$. D'autre part, $P_{f,c_0} = f[c_0]H_0 = f[c_0]$ et donc $f[c_0] = f(c_0)$.

A présent, soit $k \in \{1, \dots, n\}$ et considérons la base de Newton $\{Q_0, \dots, Q_k\}$ associée aux centres $c_k, c_{k-1}, \dots, c_0$: on a $Q_0 = 1$ et, pour $i \in \{1, \dots, k\}$,

$$Q_i = \prod_{j=0}^{i-1} (X - c_{k-j}) = (X - c_k) \cdots (X - c_{k-i+1}).$$

Alors

$$\begin{aligned} P_{f,c_0,\dots,c_k} = P_{f,c_k,\dots,c_0} &= \sum_{i=0}^{k} f[c_k, \dots, c_{k-i}] Q_i = \sum_{i=0}^{k} f[c_k, \dots, c_{k-i}] \prod_{j=0}^{i-1} (X - c_{k-j}) \\ &= \sum_{i=0}^{k} f[c_0, \dots, c_i] H_i = \sum_{i=0}^{k} f[c_0, \dots, c_i] \prod_{j=0}^{i-1} (X - c_j) \end{aligned}$$

En identifiant les coefficients de degré k des polynômes $\sum_{i=0}^{k} f[c_k, \dots, c_{k-i}] \prod_{j=0}^{i-1} (X - c_{k-j})$ et $\sum_{i=0}^{k} f[c_0, \dots, c_i] \prod_{j=0}^{i-1} (X - c_j)$, on obtient que $f[c_k, \dots, c_0] = f[c_0, \dots, c_k]$. En identifiant ensuite les coefficients de degré $k - 1$, on obtient que

$$f[c_k, \dots, c_1] + f[c_k, \dots, c_0] \left(- \sum_{j=0}^{k-1} c_{k-j} \right) = f[c_0, \dots, c_{k-1}] + f[c_0, \dots, c_k] \left(- \sum_{j=0}^{k-1} c_j \right)$$

et donc

$$\begin{aligned} f[c_k, \dots, c_1] - f[c_0, \dots, c_{k-1}] &= f[c_0, \dots, c_k] \left(- \sum_{j=0}^{k-1} c_j + \sum_{j=0}^{k-1} c_{k-j} \right) \\ &= f[c_0, \dots, c_k](c_k - c_0) \end{aligned}$$

Ainsi,

$$f[c_0,\dots,c_k] = \frac{f[c_k,\dots,c_1] - f[c_0,\dots,c_{k-1}]}{c_k - c_0} = \frac{f[c_1,\dots,c_k] - f[c_0,\dots,c_{k-1}]}{c_k - c_0}.$$

□

Remarque 2.4.4. 1. Plus généralement, si $k \in \{0,\dots,n\}$ et si σ est n'importe quelle permutation sur l'ensemble $\{0,\dots,k\}$, on a

$$f[c_{\sigma(0)},\dots,c_{\sigma(k)}] = f[c_0,\dots,c_k].$$

En effet, si l'on note alors, pour $j \in \{0,\dots,k\}$, $x_j := c_{\sigma(j)}$ et si $\{M_0,\dots,M_k\}$ désigne la base de Newton associée aux centres $x_0,\dots,x_k$, on a

$$\begin{aligned} P_{f,c_0,\dots,c_k} = P_{f,x_0,\dots,x_k} &= \sum_{i=0}^{k} f[x_0,\dots,x_i]M_i = \sum_{i=0}^{k} f[x_0,\dots,x_i]\prod_{j=0}^{i-1}(X-x_j) \\ &= \sum_{i=0}^{k} f[c_0,\dots,c_i]H_i = \sum_{i=0}^{k} f[c_0,\dots,c_i]\prod_{j=0}^{i-1}(X-c_j) \end{aligned}$$

et donc, en identifiant les coefficients de degré k des polynômes $\sum_{i=0}^{k} f[x_0,\dots,x_i]\prod_{j=0}^{i-1}(X-x_j)$ et $\sum_{i=0}^{k} f[c_0,\dots,c_i]\prod_{j=0}^{i-1}(X-c_j)$, on obtient que $f[x_0,\dots,x_k] = f[c_0,\dots,c_k]$ i.e. $f[c_{\sigma(0)},\dots,c_{\sigma(k)}] = f[c_0,\dots,c_k]$.

2. Pour $k \in \{0,\dots,n\}$, on a également l'expression

$$f[c_0,\dots,c_k] = \sum_{i=0}^{k} \frac{f(c_i)}{\prod\limits_{j=0,j\neq i}^{k}(c_i - c_j)}.$$

Montrons cette égalité à l'aide d'une démonstration par récurrence sur $k \in \{0,\dots,n\}$: on a $f[c_0] = f(c_0)$ et, si l'on suppose la propriété vérifiée au rang $k-1$ pour $k \in \{1,\dots,n\}$ fixé

et tout k-uplet de points distincts deux à deux de I, alors, par hypothèse de récurrence,

$$\begin{aligned} f[c_1,\dots,c_k]-f[c_0,\dots,c_{k-1}] &= \sum_{i=1}^{k}\frac{f(c_i)}{\prod_{j=1,j\neq i}^{k}(c_i-c_j)}-\sum_{i=0}^{k-1}\frac{f(c_i)}{\prod_{j=0,j\neq i}^{k-1}(c_i-c_j)} \\ &= \sum_{i=1}^{k}\frac{f(c_i)(c_i-c_0)}{\prod_{j=0,j\neq i}^{k}(c_i-c_j)}-\sum_{i=0}^{k-1}\frac{f(c_i)(c_i-c_k)}{\prod_{j=0,j\neq i}^{k}(c_i-c_j)} \\ &= \sum_{i=0}^{k}\frac{f(c_i)(c_i-c_0)}{\prod_{j=0,j\neq i}^{k}(c_i-c_j)}-\sum_{i=0}^{k}\frac{f(c_i)(c_i-c_k)}{\prod_{j=0,j\neq i}^{k}(c_i-c_j)} \\ &= \sum_{i=0}^{k}\frac{f(c_i)(c_k-c_0)}{\prod_{j=0,j\neq i}^{k}(c_i-c_j)}, \end{aligned}$$

et donc

$$f[c_0,\dots,c_k]=\frac{f[c_1,\dots,c_k]-f[c_0,\dots,c_{k-1}]}{c_k-c_0}=\sum_{i=0}^{k}\frac{f(c_i)}{\prod_{j=0,j\neq i}^{k}(c_i-c_j)}.$$

On n'utilisera cependant pas cette expression pour le calcul des différences divisées de f : on privilégiera la relation de récurrence de la proposition 2.4.3.

Un algorithme basé sur la relation de récurrence de la proposition 2.4.3 pour, si $k \in \{0,\dots,n\}$ calculer le réel $f[c_0,\dots,c_k]$ est, étant données les valeurs respectives $y_0,\dots,y_k$ de f en les points $c_0,\dots,c_k$,

$a_l \leftarrow y_l$, $l \in \{0,\dots,k\}$
for $i = 1$ à k **do**
  for $j = 0$ à $k-i$ **do**
    $a_j \leftarrow \frac{a_{j+1}-a_j}{c_{j+i}-c_j}$
  end for
end for
Return a_0

Exemple 2.4.5. On reprend l'exemple de l'exemple 2.3.6. La base de Newton associée aux centres 2 et 5 est formée des polynômes $H_0 := 1$ et $H_1 := X-2$, et on a $f[2] = f(2) = 3$ et

$$f[2,5]=\frac{f[5]-f[2]}{5-2}=\frac{f(5)-f(2)}{5-2}=\frac{-6-3}{3}=-3.$$

On a donc

$$P_{f,2,5} = f[2]H_0 + f[2,5]H_1 = 3 - 3(X-2) = -3X + 9,$$

ce qui est bien cohérent avec l'exemple 2.3.6.

Remarque 2.4.6. Une fois obtenue la décomposition

$$\begin{aligned} P_{f,c_0,\dots,c_n} &= \sum_{i=0}^{n} \alpha_i H_i \\ &= \sum_{i=0}^{n} \alpha_i \prod_{j=0}^{i-1} (X - c_j) \\ &= \alpha_0 + (X-c_0)\big(\alpha_1 + (X-c_1)\big(\cdots + (X-c_{n-2})(\alpha_{n-1} + (X-c_{n-1})\alpha_n\big)\cdots\big)\big) \end{aligned}$$

de $P_{f,c_0,\dots,c_n}$ dans la base de Newton associée aux centres $c_0, \dots, c_n$, on peut calculer l'évaluation de $P_{f,c_0,\dots,c_n}$ en un point $x \in I$ à l'aide d'un algorithme inspiré de la méthode de Horner :

$r \leftarrow \alpha_n$
for $i = n-1$ à 0 **do**
  $r \leftarrow \alpha_i + (x - c_i)r$
end for
Return r

2.5 Erreur d'interpolation

Reprenons les notations des sections précédentes : $f : I \to \mathbb{R}$ est une fonction sur un intervalle I de $\mathbb{R}$ et $c_0, \dots, c_n$ sont des nombres de I deux à deux distincts. On souhaite étudier l'erreur d'interpolation de f par le polynôme $P_{f,c_0,\dots,c_n}$. Précisément, on considère la fonction suivante :

Définition 2.5.1. *L'erreur d'interpolation de f aux centres $c_0, \dots, c_n$ est la fonction*

$$e : \begin{array}{ccc} I & \to & \mathbb{R} \\ x & \mapsto & f(x) - P_{f,c_0,\dots,c_n}(x) \end{array}.$$

Un premier résultat est le suivant :

Théorème 2.5.2. *On suppose que f est $n+1$ fois dérivable sur I. Alors, pour tout $x \in I$, il existe $\xi \in I$ tel que*

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

Pour démontrer ce théorème, nous utiliserons plusieurs propriétés intermédiaires. La première exprime l'erreur d'interpolation e à l'aide d'une différence divisée de f :

Proposition 2.5.3. *Soit $x \in I \backslash \{c_0, \dots, c_n\}$. Alors*

$$e(x) = f[c_0, \dots, c_n, x] \prod_{i=0}^{n} (x - c_i).$$

Démonstration. On a

$$P_{f,c_0,\dots,c_n,x} = P_{f,c_0,\dots,c_n} + f[c_0,\dots,c_n,x]\prod_{i=0}^{n}(X-c_i)$$

donc

$$f(x) = P_{f,c_0,\dots,c_n,x}(x) = P_{f,c_0,\dots,c_n}(x) + f[c_0,\dots,c_n,x]\prod_{i=0}^{n}(x-c_i),$$

et donc

$$e(x) = f(x) - P_{f,c_0,\dots,c_n}(x) = f[c_0,\dots,c_n,x]\prod_{i=0}^{n}(x-c_i).$$

□

La proposition suivante énonce que, si f est n-fois dérivable sur I, alors la différence divisée de f en les centres $c_0,\dots,c_n$ peut s'exprimer à l'aide de la dérivée $n^{\text{ème}}$ de f :

Proposition 2.5.4. *On suppose que f est n fois dérivable sur I. Alors il existe $\xi \in I$ tel que*

$$f[c_0,\dots,c_n] = \frac{f^{(n)}(\xi)}{n!}.$$

La preuve de la proposition 2.5.4 utilise la conséquence suivante du théorème de Rolle, qui en est également une généralisation :

Lemme 2.5.5. *Soit $m \in \mathbb{N}\backslash\{0\}$, soit J un intervalle de $\mathbb{R}$ et soit $h : J \to \mathbb{R}$ une fonction dérivable m fois s'annulant en $m+1$ points deux à deux distincts $a_0,\dots,a_m$ de J. Alors il existe $\xi \in J$ tel que $h^{(m)}(\xi) = 0$.*

Démonstration. Montrons cette propriété par récurrence sur $m \in \mathbb{N}\backslash\{0\}$.

Si $m = 1$ et si h est une fonction dérivable sur J s'annulant en deux points a_0 et a_1 de J avec $a_0 < a_1$, alors h est continue sur $[a_0, a_1]$ et dérivable sur $]a_0, a_1[$ donc, d'après le théorème de Rolle, il existe $\xi \in]a_0, a_1[$ tel que $h'(\xi) = 0$.

Supposons maintenant la propriété vérifiée au rang $m-1$ pour $m \in \mathbb{N}\backslash\{0;1\}$ fixé i.e. pour toute fonction dérivable $m-1$ fois sur J s'annulant en m points, il existe un point de J en lequel s'annule sa dérivée $(m-1)^{\text{ème}}$ s'annule, et considérons notre fonction h. Supposons sans perdre de généralité que $a_0 < a_1 < \ldots < a_m$. Comme h est de classe $\mathcal{C}^m$ sur J, pour tout $i \in \{0,\dots,m-1\}$, h est continue sur $[a_i, a_{i+1}]$ et dérivable sur $]a_i, a_{i+1}[$ donc, d'après le théorème de Rolle, il existe $b_i \in]a_i, a_{i+1}[$ tel que $h'(b_i) = 0$.

La dérivée h' de h, dérivable $m-1$ fois sur J, s'annule en les m points deux à deux distincts $b_0,\dots,b_{m-1}$: d'après l'hypothèse de récurrence, il existe donc $\xi \in J$ tel que $h^{(m)}(\xi) = (h')^{(m-1)}(\xi) = 0$. □

Remarque 2.5.6. Avec les notations ci-dessus, si $J = [a,b]$ avec $a, b \in \mathbb{R}$ tels que $a < b$, alors $\xi \in]a, b[$.

Démonstration de la proposition 2.5.4. La fonction f étant n fois dérivable sur I et $P_{f,c_0,\dots,c_n}$ étant un polynôme, la fonction e est n fois dérivable sur I. Elle s'annule de plus en les $n+1$ nombres deux à deux distincts $c_0,\dots,c_n$ de I. D'après le lemme précédent, il existe donc $\xi \in I$ tel que $e^{(n)}(\xi) = 0$. Or, pour tout $x \in I$,

$$e^{(n)}(x) = f^{(n)}(x) - n! f[c_0,\dots,c_n].$$

(la dérivée $n^{\text{ème}}$ d'un polynôme de la forme $\sum_{i=0}^{n} a_i X^i$, $a_0,\dots,a_n \in \mathbb{R}$, est $n!\,a_n$). Ainsi,

$$0 = e^{(n)}(\xi) = f^{(n)}(\xi) - n! f[c_0,\dots,c_n]$$

i.e.

$$f[c_0,\dots,c_n] = \frac{f^{(n)}(\xi)}{n!}.$$

□

Démonstration du théorème 2.5.2. Soit $x \in I$. On suppose tout d'abord que $x \notin \{c_0,\dots,c_n\}$. Alors, d'après la proposition 2.5.3,

$$e(x) = f[c_0,\dots,c_n,x] \prod_{i=0}^{n} (x - c_i).$$

Mais, d'après la proposition 2.5.4, comme f est dérivable $n+1$ fois sur J, il existe $\xi \in J$ tel que $f[c_0,\dots,c_n,x] = \frac{f^{(n+1)}(\xi)}{(n+1)!}$. On a alors

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

Si maintenant $x \in \{c_0,\dots,c_n\}$, alors $e(x) = 0$ et $\prod_{i=0}^{n}(x - c_i) = 0$, donc $e(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^{n}(x - c_i)$ pour tout $\xi \in J$. □

Supposons maintenant que $I = [a,b]$ avec $a, b \in \mathbb{R}$ tels que $a < b$ et, pour toute fonction $h : [a,b] \to \mathbb{R}$ continue, notons

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

Supposons également que f soit de classe $\mathcal{C}^{n+1}$ sur $[a,b]$. Alors, d'après le théorème 2.5.2, on a

$$\|e\|_{\infty,[a,b]} = \|f - P_{f,c_0,\dots,c_n}\|_{\infty,[a,b]} \leqslant \frac{\|f^{(n+1)}\|_{\infty,[a,b]}}{(n+1)!} \max_{x\in[a,b]} \left| \prod_{i=0}^{n} (x - c_i) \right|.$$

Une remarque importante est maintenant la suivante. Même si l'on augmente le nombre de points d'interpolations, la suite de fonctions donnée par les erreurs d'interpolations correspondantes ne converge pas nécessairement vers la fonction nulle, comme illustré par le phénomène

dit de Runge : la suite des interpolations polynomiales de la fonction $g : \begin{array}{ccc}[-5;5] & \to & \mathbb{R} \\ x & \mapsto & \frac{1}{1+x^2}\end{array}$ en des centres équirépartis sur l'intervalle $[-5;5]$ ne converge pas vers g.

Pour tout $n \in \mathbb{N}$, il existe cependant des centres d'interpolation $c_{n,0}, \dots, c_{n,n} \in [a,b]$ tels que, sous une condition suffisante de régularité sur f, la suite $\big(f - P_{f,c_{n,0},\dots,c_{n,n}}\big)_{n\in\mathbb{N}}$ des erreurs d'interpolation de la fonction f converge uniformément vers la fonction nulle i.e. $\big\|f - P_{f,c_{n,0},\dots,c_{n,n}}\big\|_{\infty,[a,b]} \xrightarrow[n\to+\infty]{}$, et tels que, pour tout $n \in \mathbb{N}$, la quantité $\max\limits_{x\in[a,b]} \left|\prod\limits_{i=0}^{n}(x - c_{n,i})\right|$ soit minimale, et donc telle que (la majoration ci-dessus de) l'erreur e soit minimale.

Nous allons construire cette suite de points $(c_i)_{i\in\mathbb{N}}$ à l'aide des racines des polynômes dits de Tchebychev :

Définition 2.5.7. *On pose $T_0 := 1$, $T_1 := X$ et, pour $n \in \mathbb{N}$, $T_{n+2} := 2XT_{n+1} - T_n$. Si $n \in \mathbb{N}$, le polynôme $T_n \in \mathbb{R}[X]$ est appelé $n^{ème}$ polynôme de Tchebychev.*

Remarque 2.5.8. Pour tout $n \in \mathbb{N}$, T_n est de degré n et, si $n \in \mathbb{N}\backslash\{0\}$, le coefficient dominant de T_n est 2^{n-1}.

Nous allons considérer la propriété suivante des polynômes de Tchebychev, qui nous permettra de déterminer leurs racines :

Proposition 2.5.9. *Soit $n \in \mathbb{N}$. Pour tout $\theta \in \mathbb{R}$, on a*

$$T_n\big(\cos(\theta)\big) = \cos(n\theta).$$

Démonstration. Montrons ce résultat pour récurrence sur $n \in \mathbb{N}$. Soit $\theta \in \mathbb{R}$. On a

$$T_0\big(\cos(\theta)\big) = 1 = \cos(0 \times \theta)$$

et

$$T_1\big(\cos(\theta)\big) = \cos(\theta) = \cos(1 \times \theta).$$

Supposons maintenant la propriété vérifiée jusqu'au rang $n+1$ pour $n \in \mathbb{N}$ fixé. On a alors

$$\begin{aligned} T_{n+2}\big(\cos(\theta)\big) &= 2\cos(\theta)T_{n+1}\big(\cos(\theta)\big) - T_n\big(\cos(\theta)\big) \\ &= 2\cos(\theta)\cos\big((n+1)\theta\big) - \cos(n\theta) \\ &= \cos\big(\theta + (n+1)\theta\big) + \cos\big(\theta - (n+1)\theta\big) - \cos(n\theta) \\ &= \cos\big((n+2)\theta\big) + \cos(-n\theta) - \cos(n\theta) \\ &= \cos\big((n+2)\theta\big) \end{aligned}$$

□

Corollaire 2.5.10. *Soit $n \in \mathbb{N}\backslash\{0\}$. Les racines de T_n sont les nombres (deux à deux distincts)*

$$\cos\left(\frac{(2k+1)\pi}{2n}\right),\ k \in \{0,\dots,n-1\}.$$

Démonstration. Soit $x \in \mathbb{R}$ et supposons que $x \in [-1;1]$: il existe donc un unique angle $\theta \in [0;\pi]$ tel que $x = \cos(\theta)$ et

$$\begin{aligned} T_n(x) = 0 \quad &\text{ssi} \quad T_n\big(\cos(\theta)\big) = \cos(n\theta) = 0 \\ &\text{ssi} \quad n\theta = \frac{\pi}{2} + k\pi \text{ avec } k \in \mathbb{Z} \\ &\text{ssi} \quad \theta = \frac{(2k+1)\pi}{2n} \text{ avec } k \in \mathbb{Z} \\ &\text{ssi} \quad \theta = \frac{(2k+1)\pi}{2n} \text{ avec } k \in \{0,\dots,n-1\} \\ &\text{ssi} \quad x = \cos\left(\frac{(2k+1)\pi}{2n}\right) \text{ avec } k \in \{0,\dots,n-1\}. \end{aligned}$$

Enfin, les nombres $\cos\left(\frac{(2k+1)\pi}{2n}\right)$, $k \in \{0,\dots,n-1\}$, sont deux à deux distincts et T_n est de degré n : le polynôme T_n ne possède donc pas d'autre racine. □

Remarque 2.5.11. En particulier, si $n \in \mathbb{N}\backslash\{0\}$, $T_n \in \mathbb{R}[X]$ est un polynôme scindé à racines simples.

Soit $n \in \mathbb{N}\backslash\{0\}$ et divisons à présent T_n par son coefficient dominant 2^{n-1} : le polynôme $\frac{1}{2^{n-1}}T_n$ est donc un polynôme unitaire de degré n de $\mathbb{R}[X]$. Il s'agit en fait du polynôme minimisant la norme $\|\cdot\|_{\infty,[-1;1]}$ parmi tous les polynômes unitaires de degré n :

Théorème 2.5.12. *On a*

$$\begin{aligned} \left\|\frac{1}{2^{n-1}}T_n\right\|_{\infty,[-1,1]} &= \inf\{\|Q\|_{\infty,[-1,1]},\ Q \in \mathbb{R}[X] \text{ de degré } n \text{ unitaire}\} \\ &= \min\{\|Q\|_{\infty,[-1,1]},\ Q \in \mathbb{R}[X] \text{ de degré } n \text{ unitaire}\} \end{aligned}$$

et le polynôme $\frac{1}{2^{n-1}}T_n$ unitaire de degré n est unique avec cette propriété.

Démonstration. Commençons par remarquer que

$$\|T_n\|_{\infty,[-1,1]} = \max_{x\in[-1;1]} |T_n(x)| = \max_{\theta\in[0;\pi]} \big|T_n\big(\cos(\theta)\big)\big| = \max_{\theta\in[0;\pi]} |\cos(n\theta)| = 1$$

et donc

$$\left\|\frac{1}{2^{n-1}}T_n\right\|_{\infty,[-1,1]} = \frac{1}{2^{n-1}}\|T_n\|_{\infty,[-1,1]} = \frac{1}{2^{n-1}}.$$

Supposons ensuite par l'absurde qu'il existe un polynôme $Q \in \mathbb{R}[X]$ unitaire de degré n tel que $\|Q\|_{\infty,[-1,1]} < \left\|\frac{1}{2^{n-1}}T_n\right\|_{\infty,[-1,1]}$ i.e. tel que $\|Q\|_{\infty,[-1,1]} < \frac{1}{2^{n-1}}$, et considérons le polynôme

$$R := \frac{1}{2^{n-1}}T_n - Q$$

de $\mathbb{R}[X]$ qui est de degré au plus $n-1$ car les polynômes $\frac{1}{2^{n-1}}T_n$ et Q de degré n sont tous deux unitaires.

Soit $k \in \{0, \dots, n\}$, on a

$$\begin{aligned} R\left(\cos\left(\frac{k\pi}{n}\right)\right) &= \frac{1}{2^{n-1}} T_n\left(\cos\left(\frac{k\pi}{n}\right)\right) - Q\left(\cos\left(\frac{k\pi}{n}\right)\right) \\ &= \frac{1}{2^{n-1}} \cos(k\pi) - Q\left(\cos\left(\frac{k\pi}{n}\right)\right) \\ &= \frac{1}{2^{n-1}} (-1)^k - Q\left(\cos\left(\frac{k\pi}{n}\right)\right) \end{aligned}$$

et donc, comme $\|Q\|_{\infty,[-1,1]} < \frac{1}{2^{n-1}}$,

$$\frac{1}{2^{n-1}}\big((-1)^k - 1\big) < R\left(\cos\left(\frac{k\pi}{n}\right)\right) < \frac{1}{2^{n-1}}\big((-1)^k + 1\big).$$

Ainsi, si k est pair, $R\left(\cos\left(\frac{k\pi}{n}\right)\right) > 0$, et si k est impair, $R\left(\cos\left(\frac{k\pi}{n}\right)\right) < 0$.

Par le théorème des valeurs intermédiaires, il existe donc, pour tout $k \in \{1, \dots, n\}$, un réel $x_k \in \left]\cos\left(\frac{k\pi}{n}\right), \cos\left(\frac{(k-1)\pi}{n}\right)\right[$ tel que $R(x_k) = 0$. En particulier, le polynôme R de degré au plus $n-1$ possède n racines distinctes : il s'agit donc du polynôme nul i.e. $Q = \frac{1}{2^{n-1}} T_n$, et on a alors $\|Q\|_{\infty,[-1,1]} = \frac{1}{2^{n-1}}$, ce qui est en contradiction avec l'hypothèse de départ sur Q.

Ainsi, on a bien

$$\min\{\|Q\|_{\infty,[-1,1]},\ Q \in \mathbb{R}[X] \text{ de degré } n \text{ unitaire}\} = \left\|\frac{1}{2^{n-1}} T_n\right\|_{\infty,[-1,1]} = \frac{1}{2^{n-1}}.$$

Montrons à présent que $\frac{1}{2^{n-1}} T_n$ est le seul polynôme unitaire de degré n avec cette propriété : soit $Q \in \mathbb{R}[X]$ unitaire de degré n tel que $\|Q\|_{\infty,[-1,1]} = \frac{1}{2^{n-1}}$, et montrons que $Q = \frac{1}{2^{n-1}} T_n$. Notons encore une fois $R := \frac{1}{2^{n-1}} T_n - Q \in \mathbb{R}[X]$: P est un polynôme de degré au plus $n-1$.

Notons ensuite, pour tout $k \in \{0, \dots, n\}$, $x_k := \cos\left(\frac{k\pi}{n}\right)$ et considérons le polynôme d'interpolation $P_{R,x_0,\dots,x_n}$ de la fonction polynomiale associée à R en les $n+1$ centres (deux à deux distincts) $x_0, \dots, x_n$.

Tout d'abord, comme R et $P_{R,x_0,\dots,x_n}$ sont deux polynômes de degré au plus n coïncidant en $n+1$ nombres réels distincts, $R = P_{R,x_0,\dots,x_n}$ (car alors le polynôme $R - P_{R,x_0,\dots,x_n}$ de degré au plus n s'annule en $n+1$ points). Ainsi,

$$R = P_{R,x_0,\dots,x_n} = \sum_{k=0}^{n} R(x_k) \prod_{j=0, j\neq k}^{n} \frac{X - x_j}{x_k - x_j} = \sum_{k=0}^{n} \frac{R(x_k)}{\prod_{j=0, j\neq k}^{n} x_k - x_j} \prod_{j=0, j\neq k}^{n} (X - x_j).$$

De plus, comme R est de degré au plus $n-1$, le coefficient de degré n du polynôme de droite est égal à 0 i.e.

$$\sum_{k=0}^{n} \frac{R(x_k)}{\prod_{j=0, j\neq k}^{n} x_k - x_j} = 0.$$

Or, si $k \in \{0, \dots, n\}$, on a

$$R(x_k) = \frac{1}{2^{n-1}} T_n(x_k) - Q(x_k) = \frac{1}{2^{n-1}} (-1)^k - Q(x_k)$$

et donc, comme $\|Q\|_{\infty,[-1,1]} \leqslant \frac{1}{2^{n-1}}$, de façon analogue à ci-dessus, $R(x_k) \geqslant 0$ si k est pair, et $R(x_k) \leqslant 0$ si k est impair.

D'autre part, comme la fonction cosinus est décroissante sur l'intervalle $[0; \pi]$, on a, si $j \in \{0, \dots, n\} \backslash \{k\}$, $x_k - x_j < 0$ si $j < k$ et $x_k - x_j > 0$ si $j > k$: la quantité $\prod_{j=0, j\neq k}^{n} x_k - x_j$ est donc du signe de $(-1)^k$, i.e. $\prod_{j=0, j\neq k}^{n} x_k - x_j > 0$ si k est pair, et $\prod_{j=0, j\neq k}^{n} x_k - x_j < 0$ si k est impair.

Au total, pour tout $k \in \{0, \dots, n\}$, on a $\frac{R(x_k)}{\prod_{j=0, j\neq k}^{n} x_k - x_j} \geqslant 0$. Or, nous avions montré que

$$\sum_{k=0}^{n} \frac{R(x_k)}{\prod_{j=0, j\neq k}^{n} x_k - x_j} = 0$$

donc, nécessairement, pour tout $k \in \{0, \dots, n\}$, $\frac{R(x_k)}{\prod_{j=0, j\neq k}^{n} x_k - x_j} = 0$ (une somme de nombres réels positifs est nulle si et seulement si chacun des nombres est nul) i.e. pour tout $k \in \{0, \dots, n\}$, $R(x_k) = 0$.

Le polynôme R de degré au plus $n-1$ s'annulant en les n points distincts $x_0, \dots, x_n$, il s'agit du polynôme nul i.e. $Q = \frac{1}{2^{n-1}} T_n$. □

A partir de ce résultat sur $[-1; 1]$, nous allons déduire l'unique polynôme unitaire de degré n qui minimise la norme $\| \cdot \|_{\infty,[a,b]}$ parmi tous les polynômes unitaires de degré n.

Pour cela, considérons l'application affine bijective

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

de bijection réciproque l'application affine

$$\psi : \begin{array}{ccc} [a;b] & \to & [-1,1] \\ x & \mapsto & \frac{2}{b-a} x - \frac{a+b}{b-a} \end{array} .$$

Si $Q \in \mathbb{R}[X]$ est un polynôme unitaire de degré n, on a

$$\begin{aligned}
\|Q\|_{\infty,[a,b]} &= \max_{x\in[a,b]} |Q(x)| \\
&= \max_{t\in[-1;1]} |Q(\varphi(t))| \\
&= \left(\frac{b-a}{2}\right)^n \max_{t\in[-1;1]} \left|\left(\frac{2}{b-a}\right)^n Q(\varphi(t))\right| \\
&= \left(\frac{b-a}{2}\right)^n \left\|\left(\frac{2}{b-a}\right)^n Q\circ\varphi\right\|_{\infty,[-1,1]}
\end{aligned}$$

et, si l'on considère φ comme le polynôme $\frac{b-a}{2}X + \frac{a+b}{2}$ de degré 1,

$$\left(\frac{2}{b-a}\right)^n Q\circ\varphi = \left(\frac{2}{b-a}\right)^n Q\left(\frac{b-a}{2}X + \frac{a+b}{2}\right) \in \mathbb{R}[X]$$

est un polynôme unitaire de degré n. Précisément, l'application $Q \in \mathbb{R}[X] \mapsto \left(\frac{2}{b-a}\right)^n Q\left(\frac{b-a}{2}X + \frac{a+b}{2}\right) \in \mathbb{R}[X]$ réalise une bijection de l'ensemble des polynômes unitaires de degré n dans lui-même, d'inverse

$$P \in \mathbb{R}[X] \mapsto \left(\frac{b-a}{2}\right)^n P\circ\psi = \left(\frac{b-a}{2}\right)^n P\left(\frac{2}{b-a}X - \frac{a+b}{b-a}\right) \in \mathbb{R}[X],$$

et on a aussi, si $P \in \mathbb{R}[X]$ est un polynôme unitaire de degré n,

$$\begin{aligned}
\|P\|_{\infty,[-1;1]} &= \max_{t\in[-1;1]} |P(t)| \\
&= \max_{x\in[a,b]} |P(\psi(x))| \\
&= \left(\frac{2}{b-a}\right)^n \max_{x\in[a,b]} \left|\left(\frac{b-a}{2}\right)^n P(\psi(x))\right| \\
&= \left(\frac{2}{b-a}\right)^n \left\|\left(\frac{b-a}{2}\right)^n P\circ\psi\right\|_{\infty,[a,b]}.
\end{aligned}$$

On en déduit :

Corollaire 2.5.13. *On a*

$$\min\{\|Q\|_{\infty,[a,b]},\ Q \in \mathbb{R}[X] \text{ de degré } n \text{ unitaire}\} = \left\|\frac{1}{2^{n-1}}\left(\frac{b-a}{2}\right)^n T_n\left(\frac{2}{b-a}X - \frac{a+b}{b-a}\right)\right\|_{\infty,[a,b]}$$

et le polynôme $\frac{1}{2^{n-1}}\left(\frac{b-a}{2}\right)^n T_n\left(\frac{2}{b-a}X - \frac{a+b}{b-a}\right)$ *est unique avec cette propriété.*

Démonstration. D'après les considérations ci-dessus, si $Q \in \mathbb{R}[X]$ est un polynôme unitaire de degré n, on a

$$\|Q\|_{\infty,[a,b]} = \left(\frac{b-a}{2}\right)^n \left\|\left(\frac{2}{b-a}\right)^n Q\circ\varphi\right\|_{\infty,[-1,1]}$$

et $\left(\frac{b-a}{2}\right)^n Q\circ\varphi \in \mathbb{R}[X]$ est un polynôme unitaire de degré n de telle sorte que

$$\begin{aligned}\|Q\|_{\infty,[a,b]} &\geqslant \left(\frac{b-a}{2}\right)^n \left\|\frac{1}{2^{n-1}}T_n\right\|_{\infty,[-1,1]}\\ &= \left(\frac{b-a}{2}\right)^n \left(\frac{2}{b-a}\right)^n \left\|\left(\frac{b-a}{2}\right)^n \frac{1}{2^{n-1}}T_n\circ\psi\right\|_{\infty,[a,b]}\\ &= \left\|\frac{1}{2^{n-1}}\left(\frac{b-a}{2}\right)^n T_n\left(\frac{2}{b-a}X-\frac{a+b}{b-a}\right)\right\|_{\infty,[a,b]},\end{aligned}$$

et $\|Q\|_{\infty,[a,b]} = \left\|\frac{1}{2^{n-1}}\left(\frac{b-a}{2}\right)^n T_n\left(\frac{2}{b-a}X-\frac{a+b}{b-a}\right)\right\|_{\infty,[a,b]}$ ssi

$$\left(\frac{b-a}{2}\right)^n \left\|\left(\frac{2}{b-a}\right)^n Q\circ\varphi\right\|_{\infty,[-1,1]} = \left(\frac{b-a}{2}\right)^n \left\|\frac{1}{2^{n-1}}T_n\right\|_{\infty,[-1,1]}$$

ssi

$$\begin{aligned}\left\|\left(\frac{2}{b-a}\right)^n Q\circ\varphi\right\|_{\infty,[-1,1]} = \left\|\frac{1}{2^{n-1}}T_n\right\|_{\infty,[-1,1]} \quad &\text{ssi}\quad \left(\frac{2}{b-a}\right)^n Q\circ\varphi = \frac{1}{2^{n-1}}T_n\\ &\text{ssi}\quad Q\circ\varphi = \frac{1}{2^{n-1}}\left(\frac{b-a}{2}\right)^n T_n\\ &\text{ssi}\quad Q = \frac{1}{2^{n-1}}\left(\frac{b-a}{2}\right)^n T_n\circ\psi\\ &\text{ssi}\quad Q = \frac{1}{2^{n-1}}\left(\frac{b-a}{2}\right)^n T_n\left(\frac{2}{b-a}X-\frac{a+b}{b-a}\right).\end{aligned}$$

□

Remarque 2.5.14. On a

$$\begin{aligned}\left\|\frac{1}{2^{n-1}}\left(\frac{b-a}{2}\right)^n T_n\left(\frac{2}{b-a}X-\frac{a+b}{b-a}\right)\right\|_{\infty,[a,b]} &= \left\|\frac{1}{2^{n-1}}\left(\frac{b-a}{2}\right)^n T_n\circ\psi\right\|_{\infty,[a,b]}\\ &= \left\|\left(\frac{b-a}{2}\right)^n \frac{1}{2^{n-1}}T_n\circ\psi\right\|_{\infty,[a,b]}\\ &= \left(\frac{b-a}{2}\right)^n \left\|\frac{1}{2^{n-1}}T_n\right\|_{\infty,[-1,1]}\\ &= \left(\frac{b-a}{2}\right)^n \frac{1}{2^{n-1}}\end{aligned}$$

Revenons à notre problème de départ : nous souhaitions construire des centres d'interpolation $c_0,\dots,c_n$ telle que la quantité $\max\limits_{x\in[a,b]}\left|\prod\limits_{i=0}^n (x-c_i)\right| = \left\|\prod\limits_{i=0}^n (X-c_i)\right\|_{\infty,[a,b]}$ soit minimale. Nous venons de démontrer que cette quantité est minimale pour

$$\prod_{i=0}^n (X-c_i) = \frac{1}{2^n}\left(\frac{b-a}{2}\right)^{n+1} T_{n+1}\left(\frac{2}{b-a}X-\frac{a+b}{b-a}\right)$$

i.e. pour $c_0, \dots, c_n$ les racines du polynôme $\frac{1}{2^n}\left(\frac{b-a}{2}\right)^{n+1} T_{n+1}\left(\frac{2}{b-a}X - \frac{a+b}{b-a}\right)$.

Or, si $x \in [a,b]$,

$$\begin{aligned}\frac{1}{2^n}\left(\frac{b-a}{2}\right)^{n+1} T_{n+1}\left(\frac{2}{b-a}x - \frac{a+b}{b-a}\right) = 0 \quad &\text{ssi} \quad T_{n+1}\left(\frac{2}{b-a}x - \frac{a+b}{b-a}\right) = 0 \\ &\text{ssi} \quad \frac{2}{b-a}x - \frac{a+b}{b-a} = \cos\left(\frac{(2k+1)\pi}{2(n+1)}\right),\ k \in \{0,\dots,n\} \\ &\text{ssi} \quad x = \frac{a+b}{2} + \frac{b-a}{2}\cos\left(\frac{(2k+1)\pi}{2(n+1)}\right),\ k \in \{0,\dots,n\}.\end{aligned}$$

Définition 2.5.15. *On appelle les points*

$$c_{n,k}^T := \frac{a+b}{2} + \frac{b-a}{2}\cos\left(\frac{(2k+1)\pi}{2(n+1)}\right),\ k \in \{0,\dots,n\}$$

les centres de Tchebychev d'ordre n du segment $[a,b]$.

Pour tout $n \in \mathbb{N}$, notons e_n^T l'erreur d'interpolation de f aux centres de Tchebychev d'ordre n. On a alors en particulier :

Proposition 2.5.16. *Soit $n \in \mathbb{N}$ et supposons que la fonction $f : [a,b] \to \mathbb{R}$ soit de classe $\mathcal{C}^{n+1}$ sur $[a,b]$. Alors*

$$\left\|e_n^T\right\|_{\infty,[a,b]} \leqslant \frac{\left\|f^{(n+1)}\right\|_{\infty,[a,b]}}{2^n(n+1)!}\left(\frac{b-a}{2}\right)^{n+1}.$$

Démonstration. On a

$$\begin{aligned}\|e_n^T\|_{\infty,[a,b]} &\leqslant \frac{\left\|f^{(n+1)}\right\|_{\infty,[a,b]}}{(n+1)!}\left\|\prod_{i=0}^{n}(X - c_{n,i}^T)\right\|_{\infty,[a,b]} \\ &= \frac{\left\|f^{(n+1)}\right\|_{\infty,[a,b]}}{(n+1)!}\left\|\frac{1}{2^{n-1}}\left(\frac{b-a}{2}\right)^n T_n\left(\frac{2}{b-a}X - \frac{a+b}{b-a}\right)\right\|_{\infty,[a,b]} \\ &= \frac{\left\|f^{(n+1)}\right\|_{\infty,[a,b]}}{(n+1)!}\left(\frac{b-a}{2}\right)^{n+1}\frac{1}{2^n}\end{aligned}$$

(par la remarque 2.5.14).

□

L'inégalité de la proposition précédente nous permet de maîtriser l'erreur à chaque interpolation polynomiale de f aux centres de Tchebychev $c_{n,0}^T, \dots, c_{n,n}^T$, $n \in \mathbb{N}$.

En particulier, si f est de classe $\mathcal{C}^\infty$ sur $[a,b]$ et la suite $\left(\frac{\left\|f^{(n+1)}\right\|_{\infty,[a,b]}}{2^n(n+1)!}\left(\frac{b-a}{2}\right)^{n+1}\right)_{n\in\mathbb{N}}$ converge vers 0, alors la suite des erreurs d'interpolation $(e_n^T)_{n\in\mathbb{N}}$ converge uniformément vers la fonction nulle, i.e. $\left\|e_n^T\right\|_{\infty,[a,b]} \xrightarrow[n\to+\infty]{} 0$.

En fait, cette dernière propriété est vraie sous des hypothèses plus faibles :

Théorème 2.5.17. *Si f est de classe $\mathcal{C}^1$ sur $[a,b]$, alors la suite $(e_n^T)_{n\in\mathbb{N}}$ converge uniformément vers la fonction nulle.*

Remarque 2.5.18. Un défaut de l'interpolation polynomiale telle que nous venons de l'étudier est que plus le nombre de centres d'interpolation augmente, plus le degré du polynôme d'interpolation et plus la quantité de calculs associés augmentent. Une solution consiste, pour $n \in \mathbb{N}$, plutôt que d'interpoler la fonction $f : [a,b] \to \mathbb{R}$ par une application polynomiale sur $[a,b]$ en $n+1$ centres $c_0, \ldots, c_n$ donnés, à interpoler f par une fonction polynomiale par morceaux s telle que

- les centres $c_0, \ldots, c_n$ délimitent les "morceaux",
- sur chaque morceau, la restriction de s est polynomiale de degré "peu élevé" (par exemple un, deux ou trois),
- s est suffisamment régulière en les points de "recollement" $c_0, \ldots, c_n$.

Une telle fonction s est appelée une spline.

Chapitre 3

Méthode des moindres carrés

3.1 Introduction

Soit $N \in \mathbb{N}$ "grand", soient (x_0, y_0), ..., (x_N, y_N) des points du plan $\mathbb{R}^2$ tels que les réels $x_0, \ldots, x_N$ sont deux à deux distincts, et notons $\mathcal{N}$ l'ensemble de ces points (on dit que $\mathcal{N}$ est un nuage de points). On souhaiterait "approcher" le nuage de points $\mathcal{N}$ au sens des "moindres carrés" par une fonction polynomiale de degré "très" inférieur à N.

3.2 Approximation au sens des moindres carrés

Reprenons les notations de l'introduction et soit $m \in \mathbb{N}$ tel que $m \leqslant N$.

Définition 3.2.1. *Soit $S \in \mathbb{R}_m[X]$. On dit que S est une solution d'ordre m au problème des moindres carrés associé à $\mathcal{N}$, ou encore une approximation d'ordre m au sens des moindres carrés de $\mathcal{N}$ si*

$$\sum_{i=0}^{N} \big(y_i - S(x_i)\big)^2 = \min \left\{ \sum_{i=0}^{N} \big(y_i - R(x_i)\big)^2 \mid R \in \mathbb{R}_m[X] \right\}.$$

Nous allons montrer qu'une telle solution existe toujours, qu'elle est unique et que l'on peut la déterminer explicitement.

Si $P, Q \in \mathbb{R}_N[X]$, notons

$$\langle P, Q \rangle := \sum_{i=0}^{N} P(x_i)Q(x_i).$$

Alors :

Proposition 3.2.2. *L'application*

$$\langle \cdot, \cdot \rangle : \begin{array}{ccc} \mathbb{R}_N[X] \times \mathbb{R}_N[X] & \to & \mathbb{R} \\ (P, Q) & \mapsto & \langle P, Q \rangle \end{array}$$

est un produit scalaire sur $\mathbb{R}_N[X]$.

Démonstration. L'application $\langle \cdot, \cdot \rangle$ est bilinéaire symétrique et, si $P \in \mathbb{R}_N[X]$,

$$\langle P, P \rangle = \sum_{i=0}^{N} P(x_i)P(x_i) = \sum_{i=0}^{N} \big(P(x_i)\big)^2 \geqslant 0$$

De plus, si $\langle P, P \rangle = \sum_{i=0}^{N} P(x_i)^2 = 0$ alors, pour tout $i \in \{0, \dots, N\}$, $P(x_i) = 0$, et donc P possède $N+1$ racines deux à deux distinctes : P est donc le polynôme nul car P est de degré au plus N. □

Notons $\| \cdot \|$ la norme euclidienne associée au produit scalaire $\langle \cdot, \cdot \rangle$ sur $\mathbb{R}_N[X]$.

Notons également $P_{\mathcal{N}}$ le polynôme d'interpolation des points du nuage $\mathcal{N}$ i.e. l'unique polynôme de $\mathbb{R}_N[X]$ tel que pour tout $i \in \{0, \dots, N\}$, $P_{\mathcal{N}}(x_i) = y_i$: si $\{L_0, \dots, L_N\}$ désigne la base de Lagrange associée aux centres $x_0, \dots, x_N$,

$$P_{\mathcal{N}} = \sum_{i=1}^{N} y_i L_i$$

(cf. théorème 2.3.1).

Si $R \in \mathbb{R}_m[X] \subset \mathbb{R}_N[X]$, on a alors

$$\sum_{i=0}^{N} \big(y_i - R(x_i)\big)^2 = \sum_{i=0}^{N} \big(P_{\mathcal{N}}(x_i) - R(x_i)\big)^2 = \sum_{i=0}^{N} \big((P_{\mathcal{N}} - R)(x_i)\big)^2 = \|P_{\mathcal{N}} - R\|^2 = d(P_{\mathcal{N}}, R)^2$$

si d désigne la distance euclidienne associée au produit scalaire $\langle \cdot, \cdot \rangle$.

On peut ainsi reformuler le problème des moindres carrés associé à $\mathcal{N}$ de la manière suivante :

Proposition 3.2.3. *Soit $S \in \mathbb{R}_m[X]$. Alors S est une solution d'ordre m au problème des moindres carrés associé à $\mathcal{N}$ si et seulement si*

$$d(P_{\mathcal{N}}, S)^2 = \min \big\{ d(P_{\mathcal{N}}, R)^2 \ \mid \ R \in \mathbb{R}_m[X] \big\}$$

si et seulement si

$$d(P_{\mathcal{N}}, S) = \min \{ d(P_{\mathcal{N}}, R) \ \mid \ R \in \mathbb{R}_m[X] \}$$

si et seulement si

$$d(P_{\mathcal{N}}, S) = d(P_{\mathcal{N}}, \mathbb{R}_m[X]) = d\big(P_{\mathcal{N}}, p_{\mathbb{R}_m[X]}(P_{\mathcal{N}})\big)$$

(où $p_{\mathbb{R}_m[X]} : \mathbb{R}_N[X] \to \mathbb{R}_m[X]$ désigne la projection orthogonale sur $\mathbb{R}_m[X]$) si et seulement si

$$S = p_{\mathbb{R}_m[X]}(P_{\mathcal{N}}).$$

Démonstration. La première équivalence est obtenue à l'aide des réécritures précédentes.

La seconde équivalence est obtenue, pour le sens direct, par application de la racine carrée qui est une fonction continue de $[0; +\infty[$ dans $[0; +\infty[$ et, pour le sens réciproque, par application de la fonction carrée qui est continue de $[0; +\infty[$ dans $[0; +\infty[$.

Les deux dernières équivalences sont obtenues en utilisant les propriétés des espaces euclidiens (plus généralement des espaces préhilbertiens) relatives à la distance à un sous-espace vectoriel (de dimension finie) et à la projection orthogonale sur un tel sous-espace. □

Le problème des moindres carrés associé au nuage de points $\mathcal{N}$ possède donc une unique solution d'ordre m, qui est $p_{\mathbb{R}_m[X]}(P_{\mathcal{N}})$.

3.3 Calcul de la solution au problème des moindres carrés

Reprenant les notations de la section précédente, nous allons dans cette section donner une méthode pour calculer la solution d'ordre m au problème des moindres carrés associé à $\mathcal{N}$ i.e. $p_{\mathbb{R}_m[X]}(P_{\mathcal{N}})$.

Soit $S \in \mathbb{R}_m[X]$ et notons V le vecteur colonne des coordonnées de S dans la base $\{1, X, \ldots, X^m\}$ de $\mathbb{R}_m[X]$. On a alors le résultat suivant :

Proposition 3.3.1. *Le polynôme S est l'unique solution d'ordre m du problème des moindres carrés associé à $\mathcal{N}$ ssi*

$$ {}^tCCV = {}^tCY $$

où

$$ C := \begin{pmatrix} 1 & x_0 & \cdots & x_0^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^m \end{pmatrix} \qquad et \qquad Y := \begin{pmatrix} y_0 \\ \vdots \\ y_N \end{pmatrix}. $$

Démonstration. On a tout d'abord $S = p_{\mathbb{R}_m[X]}(P_{\mathcal{N}})$ ssi $P_{\mathcal{N}} - S \in (\mathbb{R}_m[X])^\perp$ ($\mathbb{R}_N[X] = \mathbb{R}_m[X] \oplus (\mathbb{R}_m[X])^\perp$), et

$$ \begin{aligned} P_{\mathcal{N}} - S \in (\mathbb{R}_m[X])^\perp \quad & \text{ssi} \quad \forall R \in \mathbb{R}_m[X],\ \langle P_{\mathcal{N}} - S, R \rangle = 0 \\ & \text{ssi} \quad \forall R \in \mathbb{R}_m[X],\ \sum_{i=0}^{N} \big(P_{\mathcal{N}}(x_i) - S(x_i)\big) R(x_i) = 0 \\ & \text{ssi} \quad \forall R \in \mathbb{R}_m[X],\ \sum_{i=0}^{N} \big(y_i - S(x_i)\big) R(x_i) = 0. \end{aligned} $$

Or, si $R \in \mathbb{R}_m[X]$ et si W désigne le vecteur colonne des coordonnées de R dans la base $\{1, X, \ldots, X^m\}$ de $\mathbb{R}_m[X]$, on a

$$ CW = \begin{pmatrix} R(x_0) \\ \vdots \\ R(x_M) \end{pmatrix} $$

et donc

$$ \sum_{i=0}^{N} \big(y_i - S(x_i)\big) R(x_i) = {}^t(Y - CV) CW. $$

Ainsi,

$$\begin{aligned} S = p_{\mathbb{R}_m[X]}(P_{\mathcal{N}}) \quad &\text{ssi} \quad \forall W \in \mathrm{M}_{m+1,1}(\mathbb{R}),\ {}^t(Y - CV)CW = 0 \\ &\text{ssi} \quad \forall W \in \mathrm{M}_{m+1,1}(\mathbb{R}),\ \left({}^tY - {}^tV{}^tC\right) CW = 0 \\ &\text{ssi} \quad \forall W \in \mathrm{M}_{m+1,1}(\mathbb{R}),\ \left({}^tYC - {}^tV{}^tCC\right) W = 0 \\ &\text{ssi} \quad {}^tYC - {}^tV{}^tCC = 0_{m+1,1} \\ &\text{ssi} \quad {}^tYC = {}^tV{}^tCC \\ &\text{ssi} \quad {}^tCY = {}^tCCV. \end{aligned}$$

□

Remarque 3.3.2. En particulier, avec les notations ci-dessus, comme la matrice symétrique ${}^tCC$ est carrée (de taille $m+1$) et comme le système ${}^tCCW = {}^tCY$, $W \in \mathrm{M}_{m+1,1}(\mathbb{R})$, possède une unique solution, la matrice ${}^tCC$ est inversible, et le vecteur colonne des coordonnées du polynôme $p_{\mathbb{R}_m[X]}(P_{\mathcal{N}})$ dans la base $\{1, X, \ldots, X^m\}$ de $\mathbb{R}_m[X]$ est donc

$$\left({}^tCC\right)^{-1} {}^tCY.$$

Exemple 3.3.3. On s'intéresse à l'unique approximation d'ordre 1 au sens des moindres carrés du nuage de points $\mathcal{N}$, aussi appelée droite de régression linéaire associée à $\mathcal{N}$ ou droite des moindres carrés associée à $\mathcal{N}$.

Si on note $S = a_0 + a_1 X$ ce polynôme, avec $a_0, a_1 \in \mathbb{R}$,

$$V := \begin{pmatrix} a_0 \\ a_1 \end{pmatrix},\ C := \begin{pmatrix} 1 & x_0 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} \quad \text{et} \quad Y := \begin{pmatrix} y_0 \\ \vdots \\ y_N \end{pmatrix},$$

on a alors

$${}^tCCV = {}^tCY \iff V = \left({}^tCC\right)^{-1} {}^tCY,$$

avec ${}^tCC = \begin{pmatrix} N+1 & \sum_{i=0}^{N} x_i \\ \sum_{i=0}^{N} x_i & \sum_{i=0}^{N} x_i^2 \end{pmatrix}$ et donc

$$\left({}^tCC\right)^{-1} = \frac{1}{\det\left({}^tCC\right)} \begin{pmatrix} \sum_{i=0}^{N} x_i^2 & -\sum_{i=0}^{N} x_i \\ -\sum_{i=0}^{N} x_i & N+1 \end{pmatrix}$$

où $\det\left({}^tCC\right) = (N+1)\sum_{i=0}^{N} x_i^2 - \left(\sum_{i=0}^{N} x_i\right)^2$.

Ainsi, comme ${}^tCY = \begin{pmatrix} \sum_{i=0}^{N} y_i \\ \sum_{i=0}^{N} x_i y_i \end{pmatrix}$,

$$a_0 = \frac{\left(\sum_{i=0}^{N} x_i^2\right)\left(\sum_{i=0}^{N} y_i\right) - \left(\sum_{i=0}^{N} x_i\right)\left(\sum_{i=0}^{N} x_i y_i\right)}{(N+1)\sum_{i=0}^{N} x_i^2 - \left(\sum_{i=0}^{N} x_i\right)^2}$$

et

$$a_1 = \frac{-\left(\sum_{i=0}^{N} x_i\right)\left(\sum_{i=0}^{N} y_i\right) + (N+1)\left(\sum_{i=0}^{N} x_i y_i\right)}{(N+1)\sum_{i=0}^{N} x_i^2 - \left(\sum_{i=0}^{N} x_i\right)^2}.$$

Chapitre 4

Intégration numérique

4.1 Introduction

Il est rarement possible de calculer expliciter l'intégrale d'une fonction continue sur un segment de $\mathbb{R}$. L'*intégration numérique* consiste à approcher une intégrale par une expression "facilement calculable", plus précisément par une *formule de quadrature*.

4.2 Formules de quadrature

Soient $a, b \in \mathbb{R}$ tels que $a < b$, et soit $f : [a, b] \to \mathbb{R}$ une fonction intégrable au sens de Riemann sur $[a, b]$. Une *formule de quadrature sur* $[a, b]$ est une expression $\mathbb{R}$-linéaire en f qui "approche" l'intégrale $\displaystyle\int_a^b f(t)\mathrm{d}t$ de f sur $[a, b]$. Plus précisément :

Définition 4.2.1. *Une formule de quadrature sur* $[a, b]$ *(ou d'intégration numérique sur* $[a, b]$*) est une forme linéaire*

$$J : \mathcal{RI}([a, b]) \to \mathbb{R},$$

où $\mathcal{RI}([a, b])$ *désigne le* $\mathbb{R}$*-espace vectoriel des fonctions intégrables au sens de Riemann sur* $[a, b]$*.*

Par exemple, si $n \in \mathbb{N}$, $\lambda_0, \ldots, \lambda_n \in \mathbb{R}$ et $x_0, \ldots, x_n \in [a, b]$, l'application

$$\begin{array}{ccc} \mathcal{RI}([a, b]) & \to & \mathbb{R} \\ f & \mapsto & \displaystyle\sum_{i=0}^{n} \lambda_i f(x_i) \end{array}$$

est une formule de quadrature sur $[a, b]$ (il s'agit d'une combinaison linéaire des évaluations en les points $x_0, \ldots, x_n$), appelée formule de quadrature de poids $\lambda_0, \ldots, \lambda_n$ en les points $x_0, \ldots, x_n$.

Les exemples de cette forme constituent une classe importante de formules de quadrature. Parmi ceux-ci, on trouve notamment les formules de quadrature "induites" par les polynômes d'interpolation en le sens suivant. Soient $c_0, \ldots, c_n \in [a, b]$ deux à deux distincts et notons $\{L_0, \ldots, L_n\}$ la base de Lagrange associée aux centres $c_0, \ldots, c_n$.

Proposition et Définition 4.2.2. *L'application*

$$J^L_{c_0,\dots,c_n} : \begin{array}{ccc} \mathcal{RI}([a,b]) & \to & \mathbb{R} \\ g & \mapsto & \displaystyle\int_a^b P_{g,c_0,\dots,c_n}(t)\mathrm{d}t \end{array}$$

(où, si $g \in \mathcal{RI}([a,b])$, $P_{g,c_0,\dots,c_n}$ est le polynôme d'interpolation de g en les centres $c_0,\dots,c_n$) est une formule de quadrature sur $[a,b]$, appelée formule de quadrature de Lagrange associée aux centres $c_0,\dots,c_n$: il s'agit de la formule de quadrature de poids $\displaystyle\int_a^b L_i(t)\mathrm{d}t$, $i \in \{0,\dots,n\}$, en les points $c_0,\dots,c_n$.

Démonstration. On a $P_{f,c_0,\dots,c_n} = \displaystyle\sum_{i=0}^n f(c_i)L_i$ et donc

$$J^L_{c_0,\dots,c_n}(f) = \int_a^b P_{f,c_0,\dots,c_n}(t)\mathrm{d}t = \int_a^b \left(\sum_{i=0}^n f(c_i)L_i(t)\right)\mathrm{d}t = \sum_{i=0}^n \left(\int_a^b L_i(t)\mathrm{d}t\right) f(c_i).$$

□

Remarque 4.2.3. On peut dès à présent établir une majoration de l'erreur d'approximation de $\displaystyle\int_a^b f(t)\mathrm{d}t$ par $J^L_{c_0,\dots,c_n}(f)$ dans le cas où f est de classe $\mathcal{C}^{n+1}$ sur $[a,b]$: si l'on suppose que $f \in \mathcal{C}^{n+1}([a,b])$, on a

$$\begin{aligned} \left|\int_a^b f(t)\mathrm{d}t - J^L_{c_0,\dots,c_n}(f)\right| &= \left|\int_a^b f(t)\mathrm{d}t - \int_a^b P_{f,c_0,\dots,c_n}(t)\mathrm{d}t\right| \\ &= \left|\int_a^b \big(f(t) - P_{f,c_0,\dots,c_n}(t)\big)\mathrm{d}t\right| \\ &\leqslant \int_a^b |f(t) - P_{f,c_0,\dots,c_n}(t)|\,\mathrm{d}t \\ &\leqslant \int_a^b \|f - P_{f,c_0,\dots,c_n}\|_{\infty,[a,b]}\,\mathrm{d}t \\ &= (b-a)\,\|f - P_{f,c_0,\dots,c_n}\|_{\infty,[a,b]} \\ &\leqslant (b-a)\frac{\left\|f^{(n+1)}\right\|_{\infty,[a,b]}}{(n+1)!}\max_{x\in[a,b]}\left|\prod_{i=0}^n (x-c_i)\right|. \end{aligned}$$

Comme pour tout $x \in [a,b]$ et tout $i \in \{0,\dots,n\}$, $|x - c_i| \leqslant b - a$, on peut également considération la majoration (moins fine)

$$\left|\int_a^b f(t)\mathrm{d}t - J^L_{c_0,\dots,c_n}(f)\right| \leqslant \frac{(b-a)^{n+2}}{(n+1)!}\left\|f^{(n+1)}\right\|_{\infty,[a,b]}.$$

Cependant, cette majoration n'assure pas la convergence de la formule de quadrature de Lagrange appliquée à f vers $\displaystyle\int_a^b f(t)\mathrm{d}t$ lorsque le nombre de centres d'interpolation augmente (il existe des *phénomènes de Runge*).

Etablissons également quelques résultats sur les poids de la formule de quadrature $J^L_{c_0,\dots,c_n}$ dans le cas où les centres $c_0,\dots,c_n$ sont équirépartis sur $[a,b]$:

Lemme 4.2.4. *On suppose que les centres $c_0,\dots,c_n$ sont les $n+1$ points équirépartis sur $[a,b]$ i.e. pour tout $i\in\{0,\dots,n\}$, $c_i = a + i\frac{b-a}{n}$. Alors, pour tout $i\in\{0,\dots,n\}$,*

$$\int_a^b L_{n-i}(t)\mathrm{d}t = \int_a^b L_i(t)\mathrm{d}t.$$

Démonstration. Soit $i\in\{0,\dots,n\}$. On a

$$c_{n-i} = a + (n-i)\frac{b-a}{n} = a + (b-a) - i\frac{b-a}{n} = a+b-\left(a+i\frac{b-a}{n}\right) = a+b-c_i$$

donc

$$\begin{aligned}
\int_a^b L_{n-i}(t)\mathrm{d}t &= \int_a^b \left(\prod_{0\leqslant k\leqslant n,\ k\neq n-i} \frac{t-c_k}{c_{n-i}-c_k}\right)\mathrm{d}t \\
&= \int_a^b \left(\prod_{0\leqslant k\leqslant n,\ k\neq n-i} \frac{t-c_k}{a+b-c_i-c_k}\right)\mathrm{d}t \\
&= \int_a^b \left(\prod_{0\leqslant k\leqslant n,\ k\neq n-i} \frac{t-(a+b)+a+b-c_k}{-c_i+a+b-c_k}\right)\mathrm{d}t \\
&= \int_a^b \left(\prod_{0\leqslant k\leqslant n,\ k\neq n-i} \frac{t-(a+b)+c_{n-k}}{-c_i+c_{n-k}}\right)\mathrm{d}t \\
&= \int_a^b \left(\prod_{0\leqslant k\leqslant n,\ k\neq n-i} \frac{a+b-t-c_{n-k}}{c_i-c_{n-k}}\right)\mathrm{d}t \\
&= \int_a^b \left(\prod_{0\leqslant k'\leqslant n,\ k'\neq i} \frac{a+b-t-c_{k'}}{c_i-c_{k'}}\right)\mathrm{d}t \text{ (en posant } k'=n-k) \\
&= -\int_b^a \left(\prod_{0\leqslant k'\leqslant n,\ k'\neq i} \frac{u-c_{k'}}{c_i-c_{k'}}\right)\mathrm{d}u \text{ (via le changement de variable } u=a+b-t) \\
&= \int_a^b \left(\prod_{0\leqslant k'\leqslant n,\ k'\neq i} \frac{u-c_{k'}}{c_i-c_{k'}}\right)\mathrm{d}u \\
&= \int_a^b L_i(t)\mathrm{d}t.
\end{aligned}$$

□

Remarquons également l'expression suivante des poids de la formule $J^L_{c_0,\dots,c_n}$ dans le cas de centres équirépartis :

Lemme 4.2.5. *On suppose que les centres $c_0, \dots, c_n$ sont les $n+1$ points équirépartis sur $[a, b]$. Alors, pour tout $i \in \{0, \dots, n\}$,*

$$\int_a^b L_i(t)\mathrm{d}t = \frac{b-a}{n}\frac{(-1)^{n-i}}{i!(n-i)!}\int_0^n \prod_{j=0,\ j\neq i}^{n}(u-j)\mathrm{d}u.$$

Démonstration. Pour simplifier les écritures, notons tout d'abord $h_n := \frac{b-a}{n}$. Soit ensuite $i \in \{0, \dots, n\}$. On a

$$\int_a^b L_i(t)\mathrm{d}t = \int_a^b \left(\prod_{j=0,\ j\neq i}^{n}\frac{t-c_j}{c_i-c_j}\right)\mathrm{d}t.$$

Comme les centres $c_0, \dots, c_n$ sont les $n+1$ centres équirépartis sur $[a, b]$, pour tout $j \in \{0, \dots, n\}$ tel que $j \neq i$, on a $c_i - c_j = (i-j)h_n$ et donc

$$\begin{aligned}
\prod_{j=0,\ j\neq i}^{n} c_i - c_j &= \left(\prod_{j=0}^{i-1}(i-j)h_n\right)\left(\prod_{j=i+1}^{n}(i-j)h_n\right)\\
&= \left(\prod_{j=1}^{i} jh_n\right)\left(\prod_{j=1}^{n-i}(-j)h_n\right)\\
&= \left(h_n^i i!\right)\left((-1)^{n-i}h_n^{n-i}(n-i)!\right)\\
&= (-1)^{n-i}h_n^n\, i!(n-i)!
\end{aligned}$$

Ainsi,

$$\int_a^b L_i(t)\mathrm{d}t = \frac{(-1)^{n-i}}{h_n^n\, i!(n-i)!}\int_a^b \left(\prod_{j=0,\ j\neq i}^{n}(t-c_j)\right)\mathrm{d}t.$$

Considérons maintenant le changement de variable $u = \frac{t-a}{h_n}$ et écrivons

$$\begin{aligned}
\int_a^b \left(\prod_{j=0,\ j\neq i}^{n}(t-c_j)\right)\mathrm{d}t &= h_n\int_0^n \left(\prod_{j=0,\ j\neq i}^{n}(a+h_nu-c_j)\right)\mathrm{d}u\\
&= h_n\int_0^n \left(\prod_{j=0,\ j\neq i}^{n}\big(a+h_nu-(a+jh_n)\big)\right)\mathrm{d}u\\
&= h_n\int_0^n \left(\prod_{j=0,\ j\neq i}^{n}h_n(u-j)\right)\mathrm{d}u\\
&= h_n^{n+1}\int_0^n \left(\prod_{j=0,\ j\neq i}^{n}(u-j)\right)\mathrm{d}u.
\end{aligned}$$

Au total, on a donc bien

$$\int_a^b L_i(t)\mathrm{d}t = \frac{(-1)^{n-i}h_n}{i!(n-i)!}\int_0^n \prod_{j=0,\ j\neq i}^{n}(u-j)\mathrm{d}u.$$

□

Terminons cette section par la notion d'ordre d'exactitude d'une formule de quadrature générale. Soit donc J une formule de quadrature sur $[a,b]$.

Définition 4.2.6. *On appelle ordre d'exactitude de J le plus grand entier naturel r tel que pour tout $P \in \mathbb{R}_r[X]$, $J(P) = \int_a^b P(t)\mathrm{d}t$ (où, dans l'écriture $J(P)$, P désigne, par abus de notation, la fonction polynomiale $[a,b] \to \mathbb{R}\ ;\ t \mapsto P(t)$ associée à P).*

Exemple 4.2.7. La formule de quadrature de Lagrange associée aux centres $c_0, \dots, c_n$ (non nécessairement équirépartis sur $[a,b]$) est d'ordre d'exactitude au moins n car, si $Q \in \mathbb{R}_n[X]$, $P_{Q,c_0,\dots,c_n} = Q$ (le polynôme $P_{Q,c_0,\dots,c_n} - Q$ de degré au plus n possède $n+1$ racines) et donc

$$J^L_{c_0,\dots,c_n}(Q) = \int_a^b P_{Q,c_0,\dots,c_n}(t)\mathrm{d}t = \int_a^b Q(t)\mathrm{d}t.$$

4.3 Méthodes composées

Reprenons les notations de la section précédente et considérons une *subdivision* du segment $[a,b]$: une subdivision de $[a,b]$ est un $(n+1)$-uplet $(x_0,\dots,x_n)$ de nombres de $[a,b]$ tels que $a = x_0 < \cdots < x_n = b$. Soit $(x_0,\dots,x_n)$ une telle subdivision de $[a,b]$, on peut alors *subdiviser* le segment $[a,b]$ en les segments $[x_i, x_{i+1}]$, $i \in \{0,\dots,n-1\}$, et on appelle pas de la subdivision $(x_0,\dots,x_n)$ la plus grande des longueurs des intervalles $[x_i,x_{i+1}]$, $i \in \{0,\dots,n-1\}$, i.e. le réel $\max\limits_{i\in\{0,\dots,n-1\}} x_{i+1} - x_i$.

Plutôt qu'une formule de quadrature sur $[a,b]$ donnée par "une seule expression sur $[a,b]$", nous allons considérer des formules de quadrature sur chaque segment $[x_i,x_{i+1}]$, $i \in \{0,\dots,n-1\}$, puis les *composer* en une formule de quadrature sur $[a,b]$:

Proposition et Définition 4.3.1. *Pour tout $i \in \{0,\dots,n-1\}$, soit $J_i : \mathcal{RI}([x_i,x_{i+1}]) \to \mathbb{R}$ une formule de quadrature sur $[x_i,x_{i+1}]$. L'application*

$$J : \begin{array}{ccc} \mathcal{RI}([a,b]) & \to & \mathbb{R} \\ g & \mapsto & \displaystyle\sum_{i=0}^{n-1} J_i\left(g_{|[x_i,x_{i+1}]}\right) \end{array}$$

est une formule de quadrature sur $[a,b]$, appelée formule de quadrature composée à partir de $J_0,\dots,J_{n-1}$.

Démonstration. Soit $i \in \{0,\dots,n-1\}$. Comme l'application $J_i : \mathcal{RI}([x_i,x_{i+1}]) \to \mathbb{R}$ est une forme linéaire sur $\mathcal{RI}([x_i,x_{i+1}])$, l'application $\mathcal{RI}([a,b]) \to \mathbb{R}\ ;\ f \mapsto J_i\left(f_{|[x_i,x_{i+1}]}\right)$ est une forme linéaire sur $\mathcal{RI}([a,b])$. La somme des applications $\mathcal{RI}([a,b]) \to \mathbb{R}\ ;\ f \mapsto J_i\left(f_{|[x_i,x_{i+1}]}\right)$, $i \in \{0,\dots,n-1\}$, est alors également une forme linéaire sur $\mathcal{RI}([a,b])$. □

Pour $i \in \{0, \dots, n-1\}$, soit $c_i \in [x_i, x_{i+1}]$. L'application

$$J_{c_0,\dots,c_{n-1}} : \begin{array}{ccc} \mathcal{RI}([a,b]) & \to & \mathbb{R} \\ f & \mapsto & \displaystyle\sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t = \sum_{i=0}^{n-1} f(c_i)(x_{i+1}-x_i) \end{array}$$

est un exemple de formule de quadrature composée, associée aux formules de quadrature sur $[x_i, x_{i+1}]$

$$\begin{array}{ccc} \mathcal{RI}([x_i,x_{i+1}]) & \to & \mathbb{R} \\ f & \mapsto & \displaystyle\int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t = f(c_i)(x_{i+1}-x_i) \end{array} ,$$

$i \in \{0, \dots, n-1\}$ ($J_{c_0,\dots,c_{n-1}}$ est la formule de quadrature de poids $x_1 - x_0, \dots, x_n - x_{n-1}$ en les points $c_0, \dots, c_{n-1}$).

Exemple 4.3.2.

- Si, pour tout $i \in \{0, \dots, n-1\}$, $c_i = x_i$, la formule de quadrature $J_{c_0,\dots,c_{n-1}} = J_{x_0,\dots,x_{n-1}}$ est appelée méthode des rectangles à gauche associée à la subdivision $(x_0, \dots, x_n)$ de $[a, b]$: on la note R_g.
- Si, pour tout $i \in \{0, \dots, n-1\}$, $c_i = x_{i+1}$, la formule de quadrature $J_{c_0,\dots,c_{n-1}} = J_{x_1,\dots,x_n}$ est appelée méthode des rectangles à droite associée à la subdivision $(x_0, \dots, x_n)$ de $[a, b]$: on la note R_d.
- Si, pour tout $i \in \{0, \dots, n-1\}$, $c_i = \frac{x_i + x_{i+1}}{2}$, la formule de quadrature $J_{c_0,\dots,c_{n-1}}$ est appelée méthode des points milieux associée à la subdivision $(x_0, \dots, x_n)$ de $[a, b]$: on la note R_m.

La formule de quadrature $J_{c_0,\dots,c_{n-1}}$ est d'ordre d'exactitude au moins 0 : si $\alpha \in \mathbb{R}$,

$$J_{c_0,\dots,c_{n-1}}(\alpha) = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} \alpha \,\mathrm{d}t = \int_a^b \alpha \,\mathrm{d}t.$$

Les méthodes des rectangles au gauche et à droite sont d'ordre d'exactitude 0. En effet, pour tout $i \in \{0, \dots, n-1\}$, pour tout $t \in]x_i, x_{i+1}[$, on a $x_i < t < x_{i+1}$ et donc

$$\int_{x_i}^{x_{i+1}} x_i \mathrm{d}t < \int_{x_i}^{x_{i+1}} t \,\mathrm{d}t < \int_{x_i}^{x_{i+1}} x_{i+1}\mathrm{d}t \quad \text{i.e.} \quad X(x_i)(x_{i+1}-x_i) < \int_{x_i}^{x_{i+1}} t \,\mathrm{d}t < X(x_{i+1})(x_{i+1}-x_i)$$

d'où

$$R_g(X) < \int_a^b t \,\mathrm{d}t < R_d(X).$$

Nous allons à présent majorer l'erreur d'approximation de l'intégrale d'une fonction f sur $[a, b]$ par la formule de quadrature $J_{c_0,\dots,c_{n-1}}$ dans le cas où f est une fonction de classe $\mathcal{C}^1$ sur $[a, b]$ et où la subdivision $(x_0, \dots, x_n)$ est *régulière* :

Définition 4.3.3. *On dit que la subdivision $(x_0, \dots, x_n)$ de $[a, b]$ est la subdivision régulière d'ordre n de $[a, b]$ si, pour tout $i \in \{0, \dots, n-1\}$, $x_{i+1} = x_i + \frac{b-a}{n}$ i.e. si, pour tout $i \in \{0, \dots, n\}$, $x_i = a + i\frac{b-a}{n}$.*

Supposons donc dans la suite que la subdivision $(x_0, \dots, x_n)$ est régulière et notons h_n son pas : on a $h_n = \frac{b-a}{n}$.

Proposition 4.3.4. *Si $f \in \mathcal{C}^1([a,b])$, on a*

$$\left| J_{c_0,\dots,c_{n-1}}(f) - \int_a^b f(t)\,\mathrm{d}t \right| \leqslant (b-a) h_n \left\| f' \right\|_{\infty,[a,b]}.$$

Démonstration. Soit $i \in \{0, \dots, n-1\}$. D'après le théorème des accroissements finis (f est de classe $\mathcal{C}^1$ sur $[a,b]$), pour tout $t \in [x_i, x_{i+1}]$, il existe $\xi_t \in]x_i, x_{i+1}[$ tel que

$$f(t) - f(c_i) = f'(\xi_t)(t - c_i)$$

et ainsi, pour tout $t \in [x_i, x_{i+1}]$,

$$|f(t) - f(c_i)| = \left|f'(\xi_t)\right| |t - c_i| \leqslant \left\| f' \right\|_{\infty,[a,b]} h_n$$

($t, c_i \in [x_i, x_{i+1}]$ donc $|t - c_i| \leqslant x_{i+1} - x_i = h_n$).

On a ainsi

$$\begin{aligned}
\left| \int_{x_i}^{x_{i+1}} f(t)\,\mathrm{d}t - \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t \right| &= \left| \int_{x_i}^{x_{i+1}} \big(f(t) - f(c_i)\big)\mathrm{d}t \right| \\
&\leqslant \int_{x_i}^{x_{i+1}} |f(t) - f(c_i)|\,\mathrm{d}t \\
&\leqslant \int_{x_i}^{x_{i+1}} \left\| f' \right\|_{\infty,[a,b]} h_n \mathrm{d}t \\
&\leqslant \left\| f' \right\|_{\infty,[a,b]} h_n (x_{i+1} - x_i) \\
&\leqslant \left\| f' \right\|_{\infty,[a,b]} h_n^2
\end{aligned}$$

et donc

$$\begin{aligned}
\left| J_{c_0,\dots,c_{n-1}}(f) - \int_a^b f(t)\,\mathrm{d}t \right| &= \left| \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t - \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(t)\mathrm{d}t \right| \\
&= \left| \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(t)\mathrm{d}t - \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t \right| \\
&= \left| \sum_{i=0}^{n-1} \left(\int_{x_i}^{x_{i+1}} f(t)\,\mathrm{d}t - \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t \right) \right| \\
&\leqslant \sum_{i=0}^{n-1} \left| \int_{x_i}^{x_{i+1}} f(t)\,\mathrm{d}t - \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t \right| \\
&\leqslant \sum_{i=0}^{n-1} \left\| f' \right\|_{\infty,[a,b]} h_n (x_{i+1} - x_i) \\
&= \left\| f' \right\|_{\infty,[a,b]} h_n (b-a).
\end{aligned}$$

□

Remarque 4.3.5.

- En particulier, cela montre que si pour tout $n \in \mathbb{N}$, pour tout $i \in \{0, \dots, n-1\}$, c_i^n est un choix de point de $[x_i, x_{i+1}]$, la suite réelle $\left(J_{c_0^n,\dots,c_{n-1}^n}\right)_{n\in\mathbb{N}}$ converge vers l'intégrale $\displaystyle\int_a^b f(t)\mathrm{d}t$ (car $h_n \xrightarrow[n\to+\infty]{} 0$), ce que l'on savait déjà car pour tout $n \in \mathbb{N}$, $J_{c_0^n,\dots,c_{n-1}^n}$ est une somme de Riemann et f a été supposée intégrable au sens de Riemann (l'hypothèse de continue dérivabilité sur f n'est pas nécessaire pour montrer la convergence de $\left(J_{c_0^n,\dots,c_{n-1}^n}\right)_{n\in\mathbb{N}}$ vers $\displaystyle\int_a^b f(t)\mathrm{d}t$). La borne de la proposition 4.3.4 permet cependant de maîtriser à chaque étape $n \in \mathbb{N}$ l'erreur de l'approximation de $\displaystyle\int_a^b f(t)\mathrm{d}t$ par $J_{c_0^n,\dots,c_{n-1}^n}$.

- On peut améliorer la borne de la proposition 4.3.4 si la formule de quadrature $J_{c_0,\dots,c_{n-1}}$ est la méthode des rectangles à gauche. En effet, si $f \in \mathcal{C}^1([a,b])$, on a

$$\left|R_g(f) - \int_a^b f(t)\,\mathrm{d}t\right| \leqslant \frac{b-a}{2} h_n \left\|f'\right\|_{\infty,[a,b]}.$$

Pour le montrer, reprenons la preuve de la proposition précédente et ses notations : si $i \in \{0, \dots, n-1\}$, on a, avec $c_i = x_i$, pour tout $t \in [x_i, x_{i+1}]$,

$$|f(t) - f(x_i)| = \left|f'(\xi_t)\right| |t - x_i| = \left|f'(\xi_t)\right| (t - x_i) \leqslant \left\|f'\right\|_{\infty,[a,b]} (t - x_i)$$

et

$$\begin{aligned}
\left|\int_{x_i}^{x_{i+1}} f(t)\,\mathrm{d}t - \int_{x_i}^{x_{i+1}} f(x_i)\mathrm{d}t\right| &\leqslant \int_{x_i}^{x_{i+1}} |f(t) - f(x_i)|\,\mathrm{d}t \\
&\leqslant \int_{x_i}^{x_{i+1}} \left\|f'\right\|_{\infty,[a,b]} (t - x_i)\mathrm{d}t \\
&= \int_0^{x_{i+1}-x_i} \left\|f'\right\|_{\infty,[a,b]} u\,\mathrm{d}u \\
&= \left\|f'\right\|_{\infty,[a,b]} \frac{(x_{i+1} - x_i)^2}{2} \\
&= \left\|f'\right\|_{\infty,[a,b]} \frac{h_n^2}{2}
\end{aligned}$$

d'où

$$\left|R_g(f) - \int_a^b f(t)\,\mathrm{d}t\right| = \left|J_{x_0,\dots,x_{n-1}}(f) - \int_a^b f(t)\,\mathrm{d}t\right| \leqslant \frac{b-a}{2} h_n \left\|f'\right\|_{\infty,[a,b]}.$$

On retrouve la même majoration pour la méthode des rectangles à droite : si $i \in \{0, \dots, n-1\}$, on a, avec $c_i = x_{i+1}$, pour tout $t \in [x_i, x_{i+1}]$,

$$|f(t) - f(x_{i+1})| = \left|f'(\xi_t)\right| |t - x_{i+1}| = \left|f'(\xi_t)\right| (x_{i+1} - t) \leqslant \left\|f'\right\|_{\infty,[a,b]} (x_{i+1} - t)$$

et

$$\begin{aligned}\left|\int_{x_i}^{x_{i+1}} f(t)\,\mathrm{d}t - \int_{x_i}^{x_{i+1}} f(x_{i+1})\mathrm{d}t\right| &\leqslant \int_{x_i}^{x_{i+1}} |f(t)-f(x_{i+1})|\,\mathrm{d}t\\ &\leqslant \int_{x_i}^{x_{i+1}} \left\|f'\right\|_{\infty,[a,b]}(x_{i+1}-t)\mathrm{d}t\\ &= -\int_{x_{i+1}-x_i}^{0} \left\|f'\right\|_{\infty,[a,b]} u\,\mathrm{d}u\\ &= \int_{0}^{x_{i+1}-x_i} \left\|f'\right\|_{\infty,[a,b]} u\,\mathrm{d}u\\ &= \left\|f'\right\|_{\infty,[a,b]} \frac{h_n^2}{2}\end{aligned}$$

d'où

$$\left|R_d(f) - \int_a^b f(t)\,\mathrm{d}t\right| = \left|J_{x_1,\ldots,x_n}(f) - \int_a^b f(t)\,\mathrm{d}t\right| \leqslant \frac{b-a}{2} h_n \left\|f'\right\|_{\infty,[a,b]}.$$

En ce qui concerne la méthode des points milieux, on peut déterminer une "meilleure" borne dans le cas où la fonction f est de classe $\mathcal{C}^2$, en ce sens que cette borne implique une convergence plus rapide (quadratique) de la méthode :

Proposition 4.3.6. *Si f est de classe $\mathcal{C}^2$ sur $[a,b]$, on a*

$$\left|R_m(f) - \int_a^b f(t)\mathrm{d}t\right| \leqslant \frac{b-a}{24} h_n^2 \left\|f''\right\|_{\infty,[a,b]}.$$

Démonstration. Soit $i \in \{0,\ldots,n-1\}$. D'après le théorème de Taylor-Lagrange, comme f est de classe $\mathcal{C}^2$ sur $[a,b]$, pour tout $t \in [x_i, x_{i+1}]$, il existe $\xi_t \in]x_i, x_{i+1}[$ tel que

$$f(t) = f(c_i) + (t-c_i)f'(c_i) + \frac{(t-c_i)^2}{2} f''(\xi_t),$$

où $c_i = \frac{x_i+x_{i+1}}{2}$. Ainsi,

$$\begin{aligned}\int_{x_i}^{x_{i+1}} f(t)\,\mathrm{d}t - \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t &= \int_{x_i}^{x_{i+1}} (t-c_i)f'(c_i)\mathrm{d}t + \int_{x_i}^{x_{i+1}} \frac{(t-c_i)^2}{2} f''(\xi_t)\mathrm{d}t\\ &= f'(c_i)\left[\frac{(x_{i+1}-c_i)^2}{2} - \frac{(x_i-c_i)^2}{2}\right] + \int_{x_i}^{x_{i+1}} \frac{(t-c_i)^2}{2} f''(\xi_t)\mathrm{d}t\\ &= \int_{x_i}^{x_{i+1}} \frac{(t-c_i)^2}{2} f''(\xi_t)\mathrm{d}t \quad (\text{car } |x_{i+1}-c_i| = |x_i-c_i| = \tfrac{h_n}{2}).\end{aligned}$$

Or

$$\begin{aligned}
\left|\int_{x_i}^{x_{i+1}} \frac{(t-c_i)^2}{2} f''(\xi_t)\mathrm{d}t\right| &\leqslant \int_{x_i}^{x_{i+1}} \frac{(t-c_i)^2}{2} \left|f''(\xi_t)\right| \mathrm{d}t \\
&\leqslant \int_{x_i}^{x_{i+1}} \frac{(t-c_i)^2}{2} \left\|f''\right\|_{\infty,[a,b]} \mathrm{d}t \\
&= \frac{\left\|f''\right\|_{\infty,[a,b]}}{2} \left[\frac{(x_{i+1}-c_i)^3}{3} - \frac{(x_i-c_i)^3}{3}\right] \\
&= \frac{\left\|f''\right\|_{\infty,[a,b]}}{6} \left[\left(\frac{h_n}{2}\right)^3 - \left(\frac{-h_n}{2}\right)^3\right] \\
&= \frac{\left\|f''\right\|_{\infty,[a,b]}}{48} \left[2h_n^3\right] \\
&= \frac{\left\|f''\right\|_{\infty,[a,b]}}{24} h_n^3 \\
&= \frac{\left\|f''\right\|_{\infty,[a,b]}}{24} h_n^2 (x_{i+1}-x_i).
\end{aligned}$$

Au total, on a donc

$$\begin{aligned}
\left|R_m(f) - \int_a^b f(t)\,\mathrm{d}t\right| &= \left|\sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t - \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(t)\mathrm{d}t\right| \\
&= \left|\sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(t)\mathrm{d}t - \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t\right| \\
&= \left|\sum_{i=0}^{n-1} \left(\int_{x_i}^{x_{i+1}} f(t)\,\mathrm{d}t - \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t\right)\right| \\
&\leqslant \sum_{i=0}^{n-1} \left|\int_{x_i}^{x_{i+1}} f(t)\,\mathrm{d}t - \int_{x_i}^{x_{i+1}} f(c_i)\mathrm{d}t\right| \\
&= \sum_{i=0}^{n-1} \left|\int_{x_i}^{x_{i+1}} \frac{(t-c_i)^2}{2} f''(\xi_t)\mathrm{d}t\right| \\
&\leqslant \sum_{i=0}^{n-1} \frac{\left\|f''\right\|_{\infty,[a,b]}}{24} h_n^2 (x_{i+1}-x_i) \\
&= \frac{\left\|f''\right\|_{\infty,[a,b]}}{24} h_n^2 (b-a).
\end{aligned}$$

□

4.4 Méthodes composées de Newton-Cotes

Reprenons les notations de la section précédente et supposons que la subdivision $(x_0, \dots, x_n)$ de $[a, b]$ est régulière.

Soit $d \in \mathbb{N}\backslash\{0\}$. On considère la formule de quadrature composée à partir des formules de quadratures de Lagrange associées aux $d+1$ centres équirépartis sur $[x_i, x_{i+1}]$, $i \in \{0, \dots, n-1\}$. Précisément, soit $i \in \{0, \dots, n-1\}$ et notons $c_{i,0}, \dots, c_{i,d}$ les $d+1$ centres équirépartis sur $[x_i, x_{i+1}]$ (i.e. pour tout $j \in \{0, \dots, d\}$, $c_{i,j} = x_i + j\frac{x_{i+1}-x_i}{d} = x_i + j\frac{h_n}{d}$) et $\{L_{i,0}, \dots, L_{i,d}\}$ la base de Lagrange associée.

On note ensuite $J_{d,i} := J^L_{c_{i,0},\dots,c_{i,d}}$ la formule de quadrature de Lagrange sur $\underline{[x_i, x_{i+1}]}$ associée aux centres $c_{i,0}, \dots, c_{i,d}$ (cf. Proposition et Définition 4.2.2).

Enfin, on note J_d la formule de quadrature composée à partir de $J_{d,0}, \dots, J_{d,n-1}$: on a

$$J_d(f) = \sum_{i=0}^{n-1} J_{d,i}\left(f_{|[x_i,x_{i+1}]}\right) = \sum_{i=0}^{n-1}\sum_{j=0}^{d} \omega_{i,j} f(c_{i,j})$$

où si $j \in \{0, \dots, d\}$, $\omega_{i,j}$ désigne le poids $\displaystyle\int_{x_i}^{x_{i+1}} L_{i,j}(t)\mathrm{d}t$ (J_d est la formule de quadrature de poids $\omega_{i,j}$, $i \in \{0, \dots, n-1\}$, $j \in \{0, \dots, d\}$, en les points $c_{i,j}$, $i \in \{0, \dots, n-1\}$, $j \in \{0, \dots, d\}$).

Définition 4.4.1. *La formule de quadrature J_d sur $[a, b]$ est appelée <u>méthode</u> (ou <u>formule</u>) <u>de Newton-Cotes d'ordre d</u> associée à la subdivision régulière $(x_0, \dots, x_n)$ de $[a, b]$.*

Nous allons dans la suite considérer différents exemples de méthodes de Newton-Cotes. Commençons par quelques propriétés générales :

Lemme 4.4.2. *Si $i \in \{0, \dots, n-1\}$ et $j \in \{0, \dots, d\}$, la quantité $\omega_{i,j}$ ne dépend pas de i.*

Démonstration. Soient $i \in \{0, \dots, n-1\}$ et $j \in \{0, \dots, d\}$, montrons que $\omega_{i,j} = \omega_{0,j}$. Pour tout $k \in \{0, \dots, d\}$, on a

$$c_{i,j} - c_{i,k} = x_i + j\frac{h_n}{d} - \left(x_i + k\frac{h_n}{d}\right) = j\frac{h_n}{d} - k\frac{h_n}{d} = x_0 + j\frac{h_n}{d} - \left(x_0 + k\frac{h_n}{d}\right) = c_{0,j} - c_{0,k}$$

et donc

$$\begin{aligned}
\omega_{i,j} &= \int_{x_i}^{x_{i+1}} L_{i,j}(t)\mathrm{d}t \\
&= \int_{x_i}^{x_{i+1}} \left(\prod_{0\leqslant k\leqslant d,\ k\neq j} \frac{t-c_{i,k}}{c_{i,j}-c_{i,k}} \right) \mathrm{d}t \\
&= \int_{x_i}^{x_{i+1}} \left(\prod_{0\leqslant k\leqslant d,\ k\neq j} \frac{t-c_{i,k}}{c_{0,j}-c_{0,k}} \right) \mathrm{d}t \\
&= \int_{0}^{x_{i+1}-x_i} \left(\prod_{0\leqslant k\leqslant d,\ k\neq j} \frac{u+x_i-c_{i,k}}{c_{0,j}-c_{0,k}} \right) \mathrm{d}t \text{ (via le changement de variable } u=t-x_i) \\
&= \int_{0}^{h_n} \left(\prod_{0\leqslant k\leqslant d,\ k\neq j} \frac{u+c_{i,0}-c_{i,k}}{c_{0,j}-c_{0,k}} \right) \mathrm{d}t \\
&= \int_{0}^{h_n} \left(\prod_{0\leqslant k\leqslant d,\ k\neq j} \frac{u+c_{0,0}-c_{0,k}}{c_{0,j}-c_{0,k}} \right) \mathrm{d}t \\
&= \int_{0}^{h_n} \left(\prod_{0\leqslant k\leqslant d,\ k\neq j} \frac{u+a-c_{0,k}}{c_{0,j}-c_{0,k}} \right) \mathrm{d}t \\
&= \int_{a}^{a+h_n} \left(\prod_{0\leqslant k\leqslant d,\ k\neq j} \frac{v-c_{0,k}}{c_{0,j}-c_{0,k}} \right) \mathrm{d}v \text{ (via le changement de variable } v=t+a) \\
&= \int_{x_0}^{x_1} \left(\prod_{0\leqslant k\leqslant d,\ k\neq j} \frac{v-c_{0,k}}{c_{0,j}-c_{0,k}} \right) \mathrm{d}v \\
&= \int_{x_0}^{x_1} L_{0,j}(t)\mathrm{d}t \\
&= \omega_{0,j}.
\end{aligned}$$

□

Remarque 4.4.3. Pour $i \in \{0,\dots,n-1\}$ et $j \in \{0,\dots,d\}$, on aurait également pu considérer l'expression

$$\omega_{i,j} = \int_{x_i}^{x_{i+1}} L_{i,j}(t)\mathrm{d}t = \frac{x_{i+1}-x_i}{d}\frac{(-1)^{d-j}}{j!(d-j)!}\int_0^d \prod_{k=0,\ k\neq j}^{d}(u-k)\mathrm{d}u = \frac{h_n}{d}\frac{(-1)^{d-j}}{j!(d-j)!}\int_0^d \prod_{k=0,\ k\neq j}^{d}(u-k)\mathrm{d}u,$$

donnée par le lemme 4.2.5, pour montrer que la quantité $\omega_{i,j}$ était indépendante de i.

En vertu du lemme précédent, si $j \in \{0,\dots,d\}$, nous noterons $\omega_j := \omega_{0,j}$ et on a donc

$$J_d(f) = \sum_{i=0}^{n-1}\sum_{j=0}^{d} \omega_j f(c_{i,j}).$$

Les nombres $\omega_0, \dots, \omega_d$ sont appelés les poids de la méthode de Newton-Costes d'ordre d. En vertu du lemme 4.2.4, pour tout $j \in \{0, \dots, d\}$, on a $\omega_{d-j} = \omega_j$.

Remarque 4.4.4. Considérant la remarque précédente 4.4.3, notons, pour tout $j \in \{0, \dots, d\}$,

$$\omega'_j := \frac{1}{d} \frac{(-1)^{d-j}}{j!(d-j)!} \int_0^d \prod_{k=0,\ k\neq j}^{d} (u-k) \mathrm{d}u.$$

Alors, pour tout $j \in \{0, \dots, d\}$, $\omega_j = h_n \omega'_j$ et ω'_j est indépendant de n.

Continuons par un premier résultat de majoration de l'erreur d'approximation de $\displaystyle\int_a^b f(t)\mathrm{d}t$ par $J_d(f)$, dans le cas où f est de classe $\mathcal{C}^{d+1}$ sur $[a, b]$:

Proposition 4.4.5. *Supposons que $f \in \mathcal{C}^{d+1}([a, b])$. Alors*

$$\left| \int_a^b f(t)\mathrm{d}t - J_d(f) \right| \leqslant h_n^{d+1} \frac{b-a}{(d+1)!} \left\| f^{(d+1)} \right\|_{\infty,[a,b]}.$$

Démonstration. On a

$$\begin{aligned}
\left| \int_a^b f(t)\mathrm{d}t - J_d(f) \right| &= \left| \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(t)\mathrm{d}t - \sum_{i=0}^{n-1} J_{d,i}\left(f_{|[x_i,x_{i+1}]}\right) \right| \\
&= \left| \sum_{i=0}^{n-1} \left(\int_{x_i}^{x_{i+1}} f(t)\mathrm{d}t - J_{d,i}\left(f_{|[x_i,x_{i+1}]}\right) \right) \right| \\
&= \left| \sum_{i=0}^{n-1} \left(\int_{x_i}^{x_{i+1}} f(t)\mathrm{d}t - J^L_{c_{i,0},\dots,c_{i,d}}\left(f_{|[x_i,x_{i+1}]}\right) \right) \right| \\
&\leqslant \sum_{i=0}^{n-1} \left| \left(\int_{x_i}^{x_{i+1}} f(t)\mathrm{d}t - J^L_{c_{i,0},\dots,c_{i,d}}\left(f_{|[x_i,x_{i+1}]}\right) \right) \right| \\
&\leqslant \sum_{i=0}^{n-1} \frac{(x_{i+1}-x_i)^{d+2}}{(d+1)!} \left\| f^{(d+1)} \right\|_{\infty,[x_i,x_{i+1}]} \text{ (par la remarque 4.2.3)} \\
&= \sum_{i=0}^{n-1} \frac{h_n^{d+2}}{(d+1)!} \left\| f^{(d+1)} \right\|_{\infty,[x_i,x_{i+1}]} \\
&\leqslant \sum_{i=0}^{n-1} \frac{h_n^{d+2}}{(d+1)!} \left\| f^{(d+1)} \right\|_{\infty,[a,b]} \\
&= n \frac{h_n^{d+2}}{(d+1)!} \left\| f^{(d+1)} \right\|_{\infty,[a,b]} \\
&= h_n^{d+1} \frac{b-a}{(d+1)!} \left\| f^{(d+1)} \right\|_{\infty,[a,b]}.
\end{aligned}$$

□

Remarquant que la formule de quadrature J_d sur $[a,b]$ a été construite à partir de la subdivision régulière $(x_0,\dots,x_n)$ de pas $h_n = \frac{b-a}{n}$ de $[a,b]$, que la quantité $J_d(f)$ dépend donc de $n \in \mathbb{N}$, et renotant $J_d^n(f) := J_d(f)$ la formule de quadrature d'ordre d construite à partir de la subdivision régulière $(x_0,\dots,x_n)$ de $[a,b]$ évaluée en f, on obtient le résultat de convergence suivant :

Corollaire 4.4.6. *Supposons que $f \in \mathcal{C}^{d+1}([a,b])$. Alors la suite $\left(J_d^n(f)\right)_{n\in\mathbb{N}}$ converge vers $\displaystyle\int_a^b f(t)\mathrm{d}t$.*

Par ailleurs :

Proposition 4.4.7. *La formule J_d est d'ordre d'exactitude au moins d.*

Démonstration. Si Q un polynôme de $\mathbb{R}_d[X]$, on a, pour tout $i \in \{0,\dots,n-1\}$, $P_{Q,c_{i,0},\cdots,c_{i,d}} = Q$ (car Q est un polynôme de degré au plus d) et donc

$$\begin{aligned} J_d(Q) &= \sum_{i=0}^{n-1} J_{d,i}\,(Q) \\ &= \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} P_{Q,c_{i,0},\cdots,c_{i,d}}(t)\mathrm{d}t \\ &= \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} Q(t)\mathrm{d}t \\ &= \int_a^b Q(t)\mathrm{d}t \end{aligned}$$

□

Remarque 4.4.8. Il est possible de montrer que

- si d est impair, l'ordre d'exactitude de J_d est exactement d,
- si d est pair, l'ordre d'exactitude de J_d est $d+1$.

On considère maintenant deux exemples de méthodes de Newton-Cotes : la formule de quadrature J_1, appelée méthode des trapèzes, et la formule J_2, appelée méthode de Simpson.

Proposition 4.4.9. *Si $d = 1$, on a $\omega_0 = \omega_1 = \frac{h_n}{2}$, et donc*

$$J_1(f) = \sum_{i=0}^{n-1} \big(\omega_0 f(c_{i,0}) + \omega_1 f(c_{i,1})\big) = \sum_{i=0}^{n-1} \big(f(x_i) + f(x_{i+1})\big)\frac{h_n}{2} = \sum_{i=0}^{n-1} \frac{f(x_{i+1}) + f(x_i)}{2}(x_{i+1} - x_i).$$

Démonstration. Supposons donc que $d = 1$ et soient $i \in \{0,\dots,n-1\}$. On a $c_{i,0} = x_i$ et $c_{i,1} = x_{i+1}$, et donc $L_{i,0} = \dfrac{X - c_1}{c_0 - c_1} = \dfrac{X - x_{i+1}}{x_i - x_{i+1}} = \dfrac{x_{i+1} - X}{h_n}$.

Ainsi,

$$\begin{aligned}\omega_0 &= \int_{x_i}^{x_{i+1}} L_{i,0}(t)\mathrm{d}t \\ &= \int_{x_i}^{x_{i+1}} \frac{x_{i+1}-t}{h_n}\mathrm{d}t \\ &= \frac{(x_{i+1}-x_i)^2}{2h_n} \\ &= \frac{h_n^2}{2h_n} \\ &= \frac{h_n}{2}\end{aligned}$$

et $\omega_1 = \omega_0 = \frac{h_n}{2}$. □

Remarque 4.4.10.

- Pour $i \in \{0,\dots,n-1\}$, la quantité $\frac{f(x_{i+1})+f(x_i)}{2}(x_{i+1}-x_i)$ est l'aire du trapèze de "bases" $f(x_i)$ et $f(x_{i+1})$ et de hauteur $x_{i+1}-x_i$, d'où l'appellation "méthode des trapèzes".
- On a

$$\begin{aligned}J_1(f) &= \sum_{i=0}^{n-1} \frac{f(x_{i+1})+f(x_i)}{2}(x_{i+1}-x_i) \\ &= \frac{1}{2}\left(\sum_{i=0}^{n-1} f(x_{i+1})(x_{i+1}-x_i) + \sum_{i=0}^{n-1} f(x_i)(x_{i+1}-x_i)\right) = \frac{1}{2}\big(R_d(f)+R_g(f)\big).\end{aligned}$$

- D'après la remarque 4.4.8, la méthode des trapèzes est d'ordre d'exactitude égal à 1.

Considérons maintenant la méthode de Simpson :

Proposition 4.4.11. *Si $d = 2$, on a $\omega_0 = \omega_2 = \frac{h_n}{6}$ et $\omega_1 = \frac{2}{3}h_n$.*

Démonstration. Supposons donc que $d = 2$ et soient $i \in \{0,\dots,n-1\}$. Par le lemme 4.2.5 (voir aussi la remarque 4.4.3), on a

$$\begin{aligned}\omega_0 &= \frac{h_n}{2}\frac{(-1)^{2-0}}{0!(2-0)!}\int_0^2 \prod_{k=0,\ k\neq 0}^{2}(u-k)\mathrm{d}u \\ &= \frac{h_n}{4}\int_0^2 (u-1)(u-2)\mathrm{d}u \\ &= \frac{h_n}{4}\int_0^2 (u^2-3u+2)\mathrm{d}u \\ &= \frac{h_n}{4}\left[\frac{u^3}{3}-\frac{3u^2}{2}+2u\right]_0^2 \\ &= \frac{h_n}{4}\times\frac{2}{3} \\ &= \frac{h_n}{6}\end{aligned}$$

et

$$\begin{aligned}
\omega_1 &= \frac{h_n}{2}\frac{(-1)^{2-1}}{1!(2-1)!}\int_0^2 \prod_{k=0,\ k\neq 1}^{2}(u-k)\mathrm{d}u \\
&= -\frac{h_n}{2}\int_0^2 u(u-2)\mathrm{d}u \\
&= -\frac{h_n}{2}\int_0^2 (u^2-2u)\mathrm{d}u \\
&= -\frac{h_n}{2}\left[\frac{u^3}{3}-u^2\right]_0^2 \\
&= -\frac{h_n}{2}\times\left(-\frac{4}{3}\right) \\
&= \frac{2h_n}{3}.
\end{aligned}$$

□

Remarque 4.4.12.

- Si $d=2$, on a, pour tout $i\in\{0,\dots,n-1\}$, $c_{i,0}=x_i$, $c_{i,1}=\frac{x_i+x_{i+1}}{2}$ et $c_{i,2}=x_{i+1}$ donc

$$\begin{aligned}
J_2(f) &= \sum_{i=0}^{n-1}\left(\omega_0 f(c_{i,0})+\omega_1 f(c_{i,1})+\omega_d f(c_{i,2})\right) \\
&= \sum_{i=0}^{n-1}\left(\frac{1}{6}f(x_i)+\frac{2}{3}f\left(\frac{x_i+x_{i+1}}{2}\right)+\frac{1}{6}f(x_{i+1})\right)h_n \\
&= \sum_{i=0}^{n-1}\left(\frac{1}{6}f(x_i)+\frac{2}{3}f\left(\frac{x_i+x_{i+1}}{2}\right)+\frac{1}{6}f(x_{i+1})\right)(x_{i+1}-x_i).
\end{aligned}$$

- D'après la remarque 4.4.8, la méthode de Simpson est d'ordre d'exactitude égal à 3.

4.5 Estimation de l'erreur d'intégration numérique et noyau de Peano

Soient $a,b\in\mathbb{R}$ tels que $a<b$, soit $N\in\mathbb{N}$, soient $\lambda_0,\dots,\lambda_N\in\mathbb{R}$, soient $y_0,\dots,y_N\in[a,b]$ et notons J la formule de quadrature de poids $\lambda_0,\dots,\lambda_N$ en les points $y_0,\dots,y_N$: si $f:[a,b]\to\mathbb{R}$ désigne à nouveau une fonction intégrable au sens de Riemann sur $[a,b]$, on a

$$J(f)=\sum_{i=0}^{N}\lambda_i f(y_i).$$

Notons ensuite

$$E(f):=\int_a^b f(t)\mathrm{d}t-J(f)$$

l'erreur d'approximation de l'intégrale de f sur $[a,b]$ par J ($E(f)$ est linéaire en f).

Nous allons donner une expression de $E(f)$ permettant de meilleures estimation et majoration de $E(f)$. Cette expression mettra en jeu le *noyau de Peano* :

Définition 4.5.1. *Soit $r \in \mathbb{N}$. Pour $t \in [a,b]$, soit $\psi_{r,t}$ la fonction*

$$\begin{array}{ccc} [a,b] & \to & \mathbb{R} \\ x & \mapsto & \begin{cases} (x-t)^r & \text{si } x-t \geqslant 0 \text{ i.e. si } t \leqslant x, \\ 0 & \text{sinon i.e. si } t > x. \end{cases} \end{array}$$

Pour tout $t \in [a,b]$, la fonction $\psi_{r,t}$ est intégrable au sens de Riemann sur $[a,b]$ et si on note

$$K_r : \begin{array}{ccc} [a,b] & \to & \mathbb{R} \\ t & \mapsto & E(\psi_{r,t}) = \displaystyle\int_a^b \psi_{r,t}(x)\mathrm{d}x - J(\psi_{r,t}) \end{array}$$

la fonction K_r est appelée noyau de Peano d'ordre r associé à J.

On a alors le résultat suivant :

Théorème 4.5.2. *Soit $r \in \mathbb{N}$ et supposons que $f \in \mathcal{C}^{r+1}([a,b])$ et que J est d'ordre d'exactitude au moins r. Alors*

$$E(f) = \frac{1}{r!}\int_a^b K_r(t) f^{(r+1)}(t)\mathrm{d}t.$$

Démonstration. Pour tout $x \in [a,b]$, on a, par le théorème de Taylor avec reste intégral,

$$f(x) = f(a) + (x-a)f'(a) + \cdots + \frac{(x-a)^r}{r!}f^{(r)}(a) + \int_a^x \frac{(x-t)^r}{r!}f^{(r+1)}(t)\mathrm{d}t.$$

Si l'on note alors P le polynôme $f(a) + (X-a)f'(a) + \cdots + \frac{(X-a)^r}{r!}f^{(r)}(a) \in \mathbb{R}_r[X]$ et $R : [a,b] \to \mathbb{R}$ la fonction définie par $R(x) := \displaystyle\int_a^x \frac{(x-t)^r}{r!}f^{(r+1)}(t)\mathrm{d}t$ si $x \in [a,b]$, on a

$$E(f) = E(P) + E(R) = E(R),$$

car la formule de quadrature J est d'ordre d'exactitude au moins r et P est de degré au plus r.

De plus, si $x \in [a,b]$,

$$R(x) = \int_a^x \frac{(x-t)^r}{r!}f^{(r+1)}(t)\mathrm{d}t = \int_a^b \frac{\psi_{r,t}(x)}{r!}f^{(r+1)}(t)\mathrm{d}t$$

d'où

$$\begin{aligned}
E(f) = E(R) &= \int_a^b R(x)\mathrm{d}x - J(R) \\
&= \int_a^b \left(\int_a^b \frac{\psi_{r,t}(x)}{r!} f^{(r+1)}(t)\mathrm{d}t \right) \mathrm{d}x - \sum_{k=0}^{N} \lambda_k \left(\int_a^b \frac{\psi_{r,t}(y_k)}{r!} f^{(r+1)}(t)\mathrm{d}t \right) \\
&= \int_a^b \left(\int_a^b \psi_{r,t}(x)\mathrm{d}x \right) \frac{f^{(r+1)}(t)}{r!}\mathrm{d}t - \int_a^b \left(\sum_{k=0}^{N} \lambda_k \psi_{r,t}(y_k) \right) \frac{f^{(r+1)}(t)}{r!}\mathrm{d}t \\
&= \int_a^b \left(\int_a^b \psi_{r,t}(x)\mathrm{d}x - \sum_{k=0}^{N} \lambda_k \psi_{r,t}(y_k) \right) \frac{f^{(r+1)}(t)}{r!}\mathrm{d}t \\
&= \int_a^b E(\psi_{r,t}) \frac{f^{(r+1)}(t)}{r!}\mathrm{d}t \\
&= \frac{1}{r!} \int_a^b K_r(t) f^{(r+1)}(t)\mathrm{d}t.
\end{aligned}$$

□

On obtient ainsi la majoration suivante :

Corollaire 4.5.3. *Avec les mêmes hypothèses que dans l'énoncé du théorème 4.5.2, on a*

$$|E(f)| \leqslant \frac{\left\| f^{(r+1)} \right\|_{\infty,[a,b]}}{r!} \int_a^b |K_r(t)|\,\mathrm{d}t.$$

Démonstration. On a

$$\begin{aligned}
|E(f)| &\leqslant \left| \frac{1}{r!} \int_a^b K_r(t) f^{(r+1)}(t)\mathrm{d}t \right| \\
&\leqslant \frac{1}{r!} \int_a^b |K_r(t)| \left| f^{(r+1)}(t) \right| \mathrm{d}t \\
&\leqslant \frac{1}{r!} \int_a^b |K_r(t)| \left\| f^{(r+1)} \right\|_{\infty,[a,b]} \mathrm{d}t \\
&= \frac{\left\| f^{(r+1)} \right\|_{\infty,[a,b]}}{r!} \int_a^b |K_r(t)|\,\mathrm{d}t.
\end{aligned}$$

□

Considérons à présent les méthodes de Newton-Cotes. Soit donc $d \in \mathbb{N}$ et supposons que $J = J_d$. Notons r l'ordre d'exactitude de la méthode J_d (si d est impair, $r = d$ et, si d est pair,

$r = d + 1$: cf. remarque 4.4.8). Soit $t \in [a, b]$, on a alors

$$\begin{aligned} K_r(t) &= E(\psi_{r,t}) \\ &= \int_a^b \psi_{r,t}(x)\mathrm{d}x - \sum_{i=0}^{n-1}\sum_{j=0}^{d} \omega_j \psi_{r,t}(c_{i,j}) \\ &= \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} \psi_{r,t}(x)\mathrm{d}x - \sum_{i=0}^{n-1}\sum_{j=0}^{d} \omega_j \psi_{r,t}(c_{i,j}) \\ &= \sum_{i=0}^{n-1} \left(\int_{x_i}^{x_{i+1}} \psi_{r,t}(x)\mathrm{d}x - \sum_{j=0}^{d} \omega_j \psi_{r,t}(c_{i,j}) \right). \end{aligned}$$

Considérons alors la propriété suivante :

Lemme 4.5.4. *Soit $i \in \{0, \ldots, n-1\}$. Si $t \notin [x_i, x_{i+1}]$, on a*

$$\int_{x_i}^{x_{i+1}} \psi_{r,t}(x)\mathrm{d}x - \sum_{j=0}^{d} \omega_j \psi_{r,t}(c_{i,j}) = 0.$$

Démonstration. Si $t < x_i$, pour tout $x \in [x_i, x_{i+1}]$, $t < x$ et donc $\psi_{r,t} : [x_i, x_{i+1}] \to \mathbb{R}$ est la fonction polynomiale de degré r associée au polynôme $(X-t)^r$: comme la méthode de Newton-Cotes d'ordre d associée à la subdivision (x_i, x_{i+1}) de $[x_i, x_{i+1}]$ est d'ordre d'exactitude r, on a

$$\int_{x_i}^{x_{i+1}} \psi_{r,t}(x)\mathrm{d}x - \sum_{j=0}^{d} \omega_j \psi_{r,t}(c_{i,j}) = 0.$$

Par ailleurs, si $t > x_{i+1}$, pour tout $x \in [x_i, x_{i+1}]$, $t > x$ et donc $\psi_{r,t} : [x_i, x_{i+1}] \to \mathbb{R}$ est la fonction nulle sur $[x_i, x_{i+1}]$. □

Ainsi

$$\begin{aligned} \int_a^b K_r(t)\mathrm{d}t &= \int_a^b \left(\sum_{i=0}^{n-1} \left(\int_{x_i}^{x_{i+1}} \psi_{r,t}(x)\mathrm{d}x - \sum_{j=0}^{d} \omega_j \psi_{r,t}(c_{i,j}) \right) \right) \mathrm{d}t \\ &= \sum_{i=0}^{n-1} \int_a^b \left(\int_{x_i}^{x_{i+1}} \psi_{r,t}(x)\mathrm{d}x - \sum_{j=0}^{d} \omega_j \psi_{r,t}(c_{i,j}) \right) \mathrm{d}t \\ &= \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} \left(\int_{x_i}^{x_{i+1}} \psi_{r,t}(x)\mathrm{d}x - \sum_{j=0}^{d} \omega_j \psi_{r,t}(c_{i,j}) \right) \mathrm{d}t \\ &= \sum_{i=0}^{n-1} \int_0^{x_{i+1}-x_i} \left(\int_{x_i}^{x_{i+1}} \psi_{r,u+x_i}(x)\mathrm{d}x - \sum_{j=0}^{d} \omega_j \psi_{r,u+x_i}(c_{i,j}) \right) \mathrm{d}u \\ &= \sum_{i=0}^{n-1} \int_0^{x_{i+1}-x_i} \left(\int_0^{x_{i+1}-x_i} \psi_{r,u+x_i}(y + x_i)\mathrm{d}y - \sum_{j=0}^{d} \omega_j \psi_{r,u+x_i}(c_{i,j}) \right) \mathrm{d}u \\ &= \sum_{i=0}^{n-1} \int_0^{h_n} \left(\int_0^{h_n} \psi_{r,u+x_i}(y + x_i)\mathrm{d}y - \sum_{j=0}^{d} \omega_j \psi_{r,u+x_i}(c_{i,j}) \right) \mathrm{d}u \end{aligned}$$

et, si $y \in [0, h_n]$,

$$\begin{aligned}\psi_{r,u+x_i}(y+x_i) &= \begin{cases}\big(y+x_i-(u+x_i)\big)^r = (y-u)^r & \text{si } y+x_i-(u+x_i) \geqslant 0 \text{ i.e. si } u \leqslant y,\\ 0 & \text{sinon i.e. si } u < y\end{cases}\\ &= \psi_{r,u}(y)\end{aligned}$$

donc

$$\begin{aligned}\int_a^b K_r(t)\mathrm{d}t &= \sum_{i=0}^{n-1}\int_0^{h_n}\left(\int_0^{h_n}\psi_{r,u}(y)\mathrm{d}y - \sum_{j=0}^{d}\omega_j\psi_{r,u}\left(j\frac{h_n}{d}\right)\right)\mathrm{d}u\\ &= n\int_0^{h_n}\left(\int_0^{h_n}\psi_{r,u}(y)\mathrm{d}y - \sum_{j=0}^{d}\omega_j\psi_{r,u}\left(j\frac{h_n}{d}\right)\right)\mathrm{d}u.\end{aligned}$$

Remarque 4.5.5. Il est possible de montrer que le noyau de Peano K_r associé à la méthode de Newton-Cotes J_d est de signe constant. Ainsi,

$$\int_a^b |K_r(t)|\,\mathrm{d}t = \left|\int_a^b K_r(t)\mathrm{d}t\right| = n\left|\int_0^{h_n}\left(\int_0^{h_n}\psi_{r,u}(y)\mathrm{d}y - \sum_{j=0}^{d}\omega_j\psi_{r,u}\left(j\frac{h_n}{d}\right)\right)\mathrm{d}u\right|.$$

Exemple 4.5.6. Pour la méthode des trapèzes J_1 d'ordre d'exactitude égal à 1, on a

$$\begin{aligned}\int_a^b K_1(t)\mathrm{d}t &= n\int_0^{h_n}\left(\int_0^{h_n}\psi_{1,u}(y)\mathrm{d}y - \frac{\psi_{1,u}(0)+\psi_{1,u}(h_n)}{2}h_n\right)\mathrm{d}u\\ &= n\int_0^{h_n}\left(\int_0^{h_n}\psi_{1,u}(y)\mathrm{d}y - \frac{\psi_{1,u}(0)+\psi_{1,u}(h_n)}{2}h_n\right)\mathrm{d}u\\ &= n\int_0^{h_n}\left(\int_u^{h_n}(y-u)\mathrm{d}y - \frac{h_n-u}{2}h_n\right)\mathrm{d}u\\ &= n\int_0^{h_n}\left(\frac{(h_n-u)^2}{2} - \frac{h_n-u}{2}h_n\right)\mathrm{d}u\\ &= \frac{n}{2}\int_0^{h_n}\big((h_n-u)^2-(h_n-u)h_n\big)\,\mathrm{d}u\\ &= -\frac{n}{2}\int_{h_n}^{0}\big(v^2-h_nv\big)\,\mathrm{d}v \text{ (via le changement de variable } v = h_n-u)\\ &= \frac{n}{2}\int_0^{h_n}\big(v^2-h_nv\big)\,\mathrm{d}v\\ &= \frac{n}{2}\left(\frac{h_n^3}{3} - h_n\frac{h_n^2}{2}\right)\\ &= -\frac{nh_n^3}{12}\\ &= -\frac{(b-a)^3}{12n^2}\end{aligned}$$

et

$$\int_a^b |K_r(t)|\,\mathrm{d}t = \frac{(b-a)^3}{12n^2}.$$

Ainsi, si f est de classe $\mathcal{C}^2$ sur $[a,b]$, d'après le corollaire 4.5.3,

$$\left|\int_a^b f(t)\mathrm{d}t - J_1(f)\right| \leqslant \frac{(b-a)^3}{12n^2}\left\|f^{(2)}\right\|_{\infty,[a,b]}.$$

En ce qui concerne la méthode de Simpson J_2 d'ordre d'exactitude égal à 3, on a

$$\begin{aligned}
\int_a^b K_3(t)\mathrm{d}t &= n\int_0^{h_n}\left(\int_0^{h_n}\psi_{3,u}(y)\mathrm{d}y - \left(\frac{1}{6}\psi_{3,u}(0)+\frac{2}{3}\psi_{3,u}\left(\frac{h_n}{2}\right)+\frac{1}{6}\psi_{3,u}(h_n)\right)h_n\right)\mathrm{d}u\\
&= n\int_0^{h_n}\left(\int_u^{h_n}(y-u)^3\mathrm{d}y - \left(\frac{2}{3}\psi_{3,u}\left(\frac{h_n}{2}\right)+\frac{1}{6}(h_n-u)^3\right)h_n\right)\mathrm{d}u\\
&= n\int_0^{h_n}\left(\frac{(h_n-u)^4}{4} - \left(\frac{2}{3}\psi_{3,u}\left(\frac{h_n}{2}\right)+\frac{1}{6}(h_n-u)^3\right)h_n\right)\mathrm{d}u\\
&= n\left(\int_0^{h_n}\frac{(h_n-u)^4}{4}\mathrm{d}u - \frac{2h_n}{3}\int_0^{h_n}\psi_{3,u}\left(\frac{h_n}{2}\right)\mathrm{d}u - \frac{h_n}{6}\int_0^{h_n}(h_n-u)^3\mathrm{d}u\right)\\
&= n\left(\frac{h_n^5}{20} - \frac{2h_n}{3}\int_0^{\frac{h_n}{2}}\left(\frac{h_n}{2}-u\right)^3\mathrm{d}u - \frac{h_n}{6}\times\frac{h_n^4}{4}\right)\\
&= n\left(\frac{h_n^5}{20} - \frac{2h_n}{3}\times\frac{\left(\frac{h_n}{2}\right)^4}{4} - \frac{h_n}{6}\times\frac{h_n^4}{4}\right)\\
&= n\left(\frac{h_n^5}{20} - \frac{h_n^5}{96} - \frac{h_n^5}{24}\right)\\
&= -\frac{nh_n^5}{480}\\
&= -\frac{(b-a)^5}{480n^4}
\end{aligned}$$

Ainsi, si f est de classe $\mathcal{C}^4$ sur $[a,b]$, d'après le corollaire 4.5.3,

$$\left|\int_a^b f(t)\mathrm{d}t - J_2(f)\right| \leqslant \frac{(b-a)^5}{2880n^4}\left\|f^{(4)}\right\|_{\infty,[a,b]}.$$

On a en fait la généralisation suivante des deux exemples précédents :

Théorème 4.5.7. *On a*

$$\int_a^b K_r(t)\mathrm{d}t = \frac{(b-a)^{r+2}}{n^{r+1}}\left(\frac{1}{(r+1)(r+2)} - \frac{1}{d^{r+1}(r+1)}\sum_{j=0}^{d}\omega'_j j^{r+1}\right)$$

et donc, si $f\in\mathcal{C}^{r+1}([a,b])$,

$$\left|\int_a^b f(t)\mathrm{d}t - J_d(f)\right| \leqslant \frac{(b-a)^{r+2}}{r!n^{r+1}}\left|\frac{1}{(r+1)(r+2)} - \frac{1}{d^{r+1}(r+1)}\sum_{j=0}^{d}\omega'_j j^{r+1}\right|\left\|f^{(r+1)}\right\|_{\infty,[a,b]}.$$

Démonstration. D'après les considérations précédentes, on a

$$\begin{aligned}\int_a^b K_r(t)\mathrm{d}t &= n\int_0^{h_n}\left(\int_0^{h_n}\psi_{r,u}(y)\mathrm{d}y - \sum_{j=0}^{d}\omega_j\psi_{r,u}\left(j\frac{h_n}{d}\right)\right)\mathrm{d}u\\ &= n\int_0^{h_n}\left(\int_0^{h_n}\psi_{r,u}(y)\mathrm{d}y - h_n\sum_{j=0}^{d}\omega'_j\psi_{r,u}\left(j\frac{h_n}{d}\right)\right)\mathrm{d}u\\ &= n\int_0^{h_n}\left(\int_0^{h_n}\psi_{r,u}(y)\mathrm{d}y\right)\mathrm{d}u - nh_n\sum_{j=0}^{d}\omega'_j\int_0^{h_n}\psi_{r,u}\left(j\frac{h_n}{d}\right)\mathrm{d}u\end{aligned}$$

Or

$$\begin{aligned}\int_0^{h_n}\left(\int_0^{h_n}\psi_{r,u}(y)\mathrm{d}y\right)\mathrm{d}u &= \int_u^{h_n}(y-u)^r\mathrm{d}y\\ &= \int_0^{h_n}\frac{(h_n-u)^{r+1}}{r+1}\mathrm{d}u\\ &= \frac{h_n^{r+2}}{(r+1)(r+2)}\end{aligned}$$

et, si $j\in\{0,\dots,d\}$,

$$\begin{aligned}\int_0^{h_n}\psi_{r,u}\left(j\frac{h_n}{d}\right)\mathrm{d}u &= \int_0^{j\frac{h_n}{d}}\left(j\frac{h_n}{d}-u\right)^r\mathrm{d}u\\ &= \frac{\left(j\frac{h_n}{d}\right)^{r+1}}{r+1}\\ &= \frac{j^{r+1}}{d^{r+1}(r+1)}h_n^{r+1}.\end{aligned}$$

Ainsi,

$$\begin{aligned}\int_a^b K_r(t)\mathrm{d}t &= \frac{nh_n^{r+2}}{(r+1)(r+2)} - nh_n\sum_{j=0}^{d}\omega'_j\frac{j^{r+1}}{d^{r+1}(r+1)}h_n^{r+1}\\ &= nh_n^{r+2}\left(\frac{1}{(r+1)(r+2)} - \frac{1}{d^{r+1}(r+1)}\sum_{j=0}^{d}\omega'_j j^{r+1}\right)\\ &= \frac{(b-a)^{r+2}}{n^{r+1}}\left(\frac{1}{(r+1)(r+2)} - \frac{1}{d^{r+1}(r+1)}\sum_{j=0}^{d}\omega'_j j^{r+1}\right).\end{aligned}$$

□

Remarque 4.5.8. Si d est pair, $r=d+1$ et la majoration donnée par le théorème 4.5.7 est donc meilleure que la majoration donnée par la proposition 4.4.5.

Chapitre 5

Résolution numérique des équations différentielles ordinaires d'ordre 1

5.1 Introduction

Soient

- I un intervalle de $\mathbb{R}$,
- $f : I \times \mathbb{R} \to \mathbb{R}$ une fonction,
- $(t_0, y_0) \in I \times \mathbb{R}$.

On considère l'équation différentielle

$$\text{(E)} \quad \begin{cases} \forall t \in I,\ y'(t) = f\big(t, y(t)\big) \\ y(t_0) = y_0 \end{cases} \quad , \ y : I \to \mathbb{R} \text{ dérivable.}$$

On cherche à résoudre *numériquement* l'équation différentielle (E) : l'idée est ici d'*approcher* les valeurs d'une solution $y : I \to \mathbb{R}$ de (E) en un nombre fini de points.

On commencer par rappeler une condition suffisante d'existence et d'unicité d'une solution pour l'équation différentielle (E) :

Théorème 5.1.1 (Théorème de Cauchy-Lipschitz)**.** *On suppose que*

- $f : I \times \mathbb{R} \to \mathbb{R}$ *est de classe* $\mathcal{C}^k$, $k \in \mathbb{N}\backslash\{0\}$,
- f *est lipschitzienne par rapport à sa deuxième variable i.e. il existe* $C \in]0; +\infty[$ *tel que pour tout* $t \in I$, *pour tous* $y_1, y_2 \in \mathbb{R}$,

$$|f(t, y_1) - f(t, y_2)| \leqslant C|y_1 - y_2|.$$

Alors l'équation différentielle (E) *possède une et une seule solution* $y : I \to \mathbb{R}$, *qui est de classe* $\mathcal{C}^{k+1}$.

Remarque 5.1.2. 1. Nous ne donnerons pas la démonstration du théorème de Cauchy-Lipschitz.

2. Dans le cas particulier où f ne dépend pas de la deuxième variable, on peut donner un énoncé plus fort : si $f : I \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^k$, $k \in \mathbb{N}$, alors l'équation différentielle
$$\begin{cases} \forall t \in I,\ y'(t) = f\big(t\big) \\ y(t_0) = y_0 \end{cases}, \ y : I \to \mathbb{R} \text{ dérivable}$$
possède une unique solution $y : I \to \mathbb{R}$ de classe $\mathcal{C}^{k+1}$, donnée par, si $t \in I$,
$$y(t) = y_0 + \int_{t_0}^{t} f(u)\mathrm{d}u.$$

3. En général, même si une solution y de (E) existe et est unique, il n'y a pas d'expression explicite pour y.

Décrivons à présent la méthode de résolution numérique de (E) que nous allons aborder. On suppose que I est un segment $[t_0, t_0 + T]$, avec $T \in]0; +\infty[$, et que (E) possède une unique solution y sur $I = [t_0, t_0 + T]$.

Soit ensuite $N \in \mathbb{N}\backslash\{0\}$ et considérons une subdivision $(t_0, \ldots, t_N)$ du segment $[t_0, t_0 + T]$. Nous allons approcher les valeurs $y(t_1), \ldots, y(t_N)$ par des réels $y_1, \ldots, y_N$ respectivement. Plus précisément, si $i \in \{1, \ldots, N\}$, le réel y_i sera calculé par récurrence à partir des termes précédents. Si $r \in \mathbb{N}\backslash\{0\}$ et, pour $i \in \{r, \ldots, N\}$, y_i est calculé à partir des termes $y_{i-r}, \ldots, y_{i-1}$, on parle de *méthode de résolution numérique à r pas*. Dans ce chapitre, nous ne traiterons que des méthodes de résolution numérique à un pas.

5.2 Méthode de résolution numérique à un pas

On reprend les notations et les hypothèses de l'introduction (on rappelle notamment que l'on a supposé que (E) possédait une unique solution notée y) et pour $i \in \{0, \ldots, N-1\}$, notons $h_i := t_{i+1} - t_i$.

Une méthode de résolution numérique de (E) à un pas consiste à considérer une fonction continue
$$\phi : [t_0, t_0 + T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$
et à calculer les réels $y_1, \ldots, y_N$ donnés par la relation de récurrence
$$y_{n+1} = y_n + h_n \phi(t_n, y_n, h_n),\ n \in \{0, \ldots, N-1\}.$$

Un premier exemple d'un telle méthode à un pas est la méthode d'Euler :

Définition 5.2.1. *La méthode d'Euler pour la résolution numérique de* (E) *consiste à calculer les réels $y_1, \ldots, y_N$ donnés par la relation de récurrence*
$$y_{n+1} = y_n + h_n f(t_n, y_n),\ n \in \{0, \ldots, N-1\}.$$

Remarque 5.2.2.

- La "fonction ϕ" associée à la méthode d'Euler est la fonction

$$\begin{array}{ccc} [t_0, t_0+T]\times\mathbb{R}\times\mathbb{R} & \to & \mathbb{R} \\ (t,\gamma,\kappa) & \mapsto & f(t,\gamma) \end{array}$$

- Soit $n \in \{0,\dots,N-1\}$. D'après le théorème de Taylor-Young, il existe une fonction $\epsilon : [t_n, t_{n+1}] \to \mathbb{R}$ telle que $\epsilon(t) \xrightarrow[t\to t_n]{} 0$ et, si $t \in [t_n, t_{n+1}]$,

$$y(t) = y(t_n) + (t-t_n)y'(t_n) + (t-t_n)\epsilon(t) = y(t_n) + (t-t_n)f\big(t_n, y(t_n)\big) + (t-t_n)\epsilon(t)$$

(car y est une solution de (E)) et donc, si $\big(y(t_n), y'(t_n)\big) \neq (0,0)$,

$$y(t) \underset{t\to t_n}{\sim} y(t_n) + (t-t_n)f\big(t_n, y(t_n)\big).$$

C'est cette relation qui motive la méthode d'Euler.

Considérons une méthode de résolution numérique de (E) à un pas donnée par une fonction continue $\phi : [t_0, t_0+T]\times\mathbb{R}\times\mathbb{R} \to \mathbb{R}$. Nous allons considérer une première notion d'erreur pour cette méthode :

Définition 5.2.3. *Soit $n \in \{0,\dots,N-1\}$. On note*

$$e_n := y(t_{n+1}) - \big(y(t_n) + h_n\phi(t_n, y(t_n), h_n)\big)$$

la quantité appelée erreur de consistance locale de la méthode considérée sur le segment $[t_n, t_{n+1}]$.

Si $n \in \{0,\dots,N-1\}$, la quantité e_n mesure l'écart entre la "véritable" valeur de y en t_{n+1} et le réel obtenu en "appliquant la méthode" à la valeur de y en t_n.

Nous allons dans l'exemple ci-dessous déterminer un "ordre de grandeur" de l'erreur de consistance locale pour la méthode d'Euler associée à (E) :

Exemple 5.2.4. Supposons que la fonction $f : I\times\mathbb{R} \to \mathbb{R}$ soit de classe $\mathcal{C}^1$ et que la subdivision $(t_0,\dots,t_n)$ de $[t_0, t_0+T]$ soit régulière (i.e. pour tout $n \in \{0,\dots,N-1\}$, $h_n = \frac{T}{N}$ et, pour tout $n \in \{0,\dots,N\}$, $t_n = t_0 + n\frac{T}{d}$) et notons $h := \frac{T}{N}$.

Soit ensuite $n \in \{0,\dots,N-1\}$. On a, avec les notations ci-dessus,

$$\begin{array}{rcl} e_n & = & y(t_{n+1}) - \big(y(t_n) + hf(t_n, y(t_n))\big) \\ & = & y(t_n+h) - \big(y(t_n) + hy'(t_n)\big) \end{array}$$

et, par le théorème de Taylor-Lagrange, il existe donc $\xi_n \in]t_n, t_n+h[\subset [t_0, t_0+T]$ tel que

$$e_n = \frac{h^2}{2}y''(\xi_n).$$

Remarquons alors que, comme pour tout $t \in I$, $y'(t) = f(t, y(t))$, on a, pour tout $t \in I$,

$$y''(t) = \frac{\partial f}{\partial t}(t, y(t)) + y'(t)\frac{\partial f}{\partial \gamma}(t, y(t)) = \frac{\partial f}{\partial t}(t, y(t)) + f(t, y(t))\frac{\partial f}{\partial \gamma}(t, y(t)).$$

Ainsi,

$$e_n = \frac{h^2}{2}\left(\frac{\partial f}{\partial t}(\xi_n, y(\xi_n)) + f(\xi_n, y(\xi_n))\frac{\partial f}{\partial \gamma}(\xi_n, y(\xi_n))\right).$$

Remarque 5.2.5. Dans l'exemple précédent, nous avons utilisé le fait que l'application $k : t \in I \mapsto f(t, y(t)) \in \mathbb{R}$ était la composition des applications $f : (t, \gamma) \in I \times \mathbb{R} \to f(t, \gamma) \in \mathbb{R}$ et $g : t \in I \mapsto (t, y(t)) \in I \times \mathbb{R}$, et que donc, si $t \in I$,

$$\mathrm{d}_t k = \mathrm{d}_t(f \circ g) = \mathrm{d}_{g(t)} f \circ \mathrm{d}_t g,$$

cette dernière application linéaire étant donnée par la matrice

$$\left(\frac{\partial f}{\partial t}(t, y(t)) \quad \frac{\partial f}{\partial \gamma}(t, y(t))\right) \begin{pmatrix} 1 \\ y'(t) \end{pmatrix} = \left(\frac{\partial f}{\partial t}(t, y(t)) + y'(t)\frac{\partial f}{\partial \gamma}(t, y(t))\right).$$

Ainsi, on a bien

$$y''(t) = k'(t) = \frac{\partial f}{\partial t}(t, y(t)) + y'(t)\frac{\partial f}{\partial \gamma}(t, y(t)).$$

Dans la suite, afin de simplifier les écritures, nous écrirons, si $g : I \times \mathbb{R} \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^1$,

$$g^{[1]} := \begin{array}{ccc} I \times \mathbb{R} & \to & \mathbb{R} \\ (t, \gamma) & \mapsto & \frac{\partial g}{\partial t}(t, \gamma) + g(t, \gamma)\frac{\partial g}{\partial \gamma}(t, \gamma) \end{array}$$

Par exemple, dans l'exemple 5.2.4, l'erreur de consistance locale e_n pour la méthode d'Euler s'écrit

$$e_n = \frac{h^2}{2} f^{[1]}(\xi_n, y(\xi_n)).$$

Si g est de classe $\mathcal{C}^{n+1}$, $n \in \mathbb{N}\backslash\{0\}$, on définit ensuite par récurrence

$$g^{[n+1]} := \begin{array}{ccc} I \times \mathbb{R} & \to & \mathbb{R} \\ (t, \gamma) & \mapsto & \frac{\partial g^{[n]}}{\partial t}(t, \gamma) + g(t, \gamma)\frac{\partial g^{[n]}}{\partial \gamma}(t, y) \end{array}$$

Remarquons alors que, si f est de classe $\mathcal{C}^p$, $p \in \mathbb{N}\backslash\{0\}$, sur $I \times \mathbb{R}$, l'unique solution y de (E) est de classe $\mathcal{C}^{p+1}$ sur I et on a, si $t \in I$,

$$y^{(p+1)}(t) = f^{[p]}(t, y(t)).$$

En effet, cette égalité est vraie pour $p = 1$ et, si on la suppose vraie au rang $p - 1$ pour $p \in \mathbb{N}\backslash\{0; 1\}$ fixé, on a

$$\begin{aligned} y^{(p+1)}(t) &= \left(y^{(p)}\right)'(t) \\ &= \frac{\partial f^{[p-1]}}{\partial t}(t, y(t)) + y'(t)\frac{\partial f^{[p-1]}}{\partial \gamma}(t, y(t)) \\ &= \frac{\partial f^{[p-1]}}{\partial t}(t, y(t)) + f(t, y(t))\frac{\partial f^{[p-1]}}{\partial \gamma}(t, y(t)) \\ &= f^{[p]}(t, y(t)). \end{aligned}$$

Cette dernière écriture motive la définition, ci-dessous, de méthodes de résolution numérique de (E) à un pas généralisant la méthode d'Euler.

Soit $p \in \mathbb{N}\backslash\{0\}$.

Définition 5.2.6. *On suppose que la fonction $f : I \times \mathbb{R} \to \mathbb{R}$ est de classe $\mathcal{C}^{p-1}$ sur $I \times \mathbb{R}$, que la subdivision $(t_0, \ldots, t_N)$ de $[t_0, t_0+T]$ est régulière et on note $h := \frac{T}{N}$. La méthode de Taylor d'ordre p pour la résolution numérique de* (E) *consiste à calculer les réels $y_1, \ldots, y_N$ donnés par la relation de récurrence*

$$y_{n+1} = y_n + \sum_{k=1}^{p} \frac{h^k}{k!} f^{[k-1]}(t_n, y_n)$$

(où $f^{[0]}$ désigne la fonction f).

Remarque 5.2.7. Conservons les hypothèses et les notations de la définition précédente.

- La méthode d'Euler est, si la subdivision considérée de $[t_0, t_0+T]$ est régulière, la méthode de Taylor d'ordre 1.

- La méthode de Taylor d'ordre p est la méthode à un pas associée à la "fonction ϕ" donnée par

$$\phi(t, \gamma, \kappa) = \sum_{k=1}^{p} \frac{\kappa^{k-1}}{k!} f^{[k-1]}(t, \gamma)$$

  si $(t, \gamma, \kappa) \in [t_0, t_0 + T] \times \mathbb{R} \times \mathbb{R}$,

- Supposons que f soit de classe $\mathcal{C}^p$ sur $I \times \mathbb{R}$ et soit $n \in \{0, \ldots, N-1\}$. D'après le théorème de Taylor-Lagrange, il existe un réel $\xi_n \in]t_n, t_n + h[$ tel que

$$\begin{aligned} y(t_n + h) &= y(t_n) + \sum_{k=1}^{p} \frac{h^k}{k!} y^{(k)}(t_n) + \frac{h^{p+1}}{(p+1)!} y^{(p+1)}(\xi_n) \\ &= y(t_n) + \sum_{k=1}^{p} \frac{h^k}{k!} f^{[k-1]}(t_n, y(t_n)) + \frac{h^{p+1}}{(p+1)!} f^{[p]}(\xi_n, y(\xi_n)). \end{aligned}$$

  L'erreur de consistance locale de la méthode de Taylor d'ordre p sur $[t_n, t_{n+1}]$ est donc donnée par, comme $t_{n+1} = t_n + h$,

$$\begin{aligned} e_n &= y(t_n + h) - \left(y(t_n) + \sum_{k=1}^{p} \frac{h^k}{k!} f^{[k-1]}(t_n, y(t_n)) \right) \\ &= \frac{h^{p+1}}{(p+1)!} f^{[p]}(\xi_n, y(\xi_n)). \end{aligned}$$

- Deux défauts d'application de la méthode de Taylor d'ordre p est que la fonction f considérée peut ne pas être de classe $\mathcal{C}^{p-1}$ et que le calcul des expressions $f^{[k-1]}(t_n, y_n))$, $k \in \{1, \ldots, p\}$, $n \in \{0, \ldots, N-1\}$, peut être coûteux.

5.3 Consistance, stabilité et convergence des méthodes à un pas

Reprenons notre équation différentielle

$$\text{(E)} \quad \begin{cases} \forall t \in [t_0, t_0+T],\ y'(t) = f\big(t, y(t)\big) \\ y(t_0) = y_0 \end{cases} \quad , \ y : I \to \mathbb{R} \text{ dérivable}$$

dont on suppose qu'elle possède une unique solution y, et considérons une subdivision quelconque $(t_0, \dots, t_N)$ de $[t_0, t_0+T]$. Notons

$$h := \max_{i\in\{0,\dots,n-1\}} h_i = \max_{i\in\{0,\dots,n-1\}} (t_{i+1} - t_i)$$

le pas de la subdivision $(t_0, \dots, t_N)$.

Considérons ensuite une méthode de résolution numérique de (E) à un pas donnée par une fonction $\phi : [t_0, t_0+T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. On se pose la question de la *convergence* de la méthode vers la solution y de (E).

Introduisons tout d'abord les notions de *consistance* et de *stabilité* :

Définition 5.3.1.

- *On appelle erreur de consistance associée à la subdivision $(t_0, \dots, t_N)$ la quantité*

$$e(t_0, \dots, t_N) := \sum_{n=0}^{N-1} |e_n|$$

et on dit que la méthode considérée est consistante si $e(t_0, \dots, t_N)$ tend vers 0 quand le pas h tend vers 0, i.e.

$$\forall \epsilon > 0, \exists \delta > 0, \forall (t_0, \dots, t_N) \text{ subdivision de } [t_0, t_0+T], h := \max_{i\in\{0,\dots,n-1\}} (t_{i+1} - t_i) < \delta \Rightarrow e(t_0, \dots, t_N) < \epsilon.$$

- *On dit que la méthode est stable s'il existe une constante $S \in]0; +\infty[$ telle que pour toute subdivision $(t_0, \dots, t_N)$ de $[t_0, t_0+T]$, pour tout N-uplet $(\epsilon_1, \dots, \epsilon_N)$ de nombres réels positifs ou nuls, si $y_1, \dots, y_N$ et $\widetilde{y}_0, \dots, \widetilde{y}_N$ sont les réels définis respectivement par la relation de récurrence*

$$\forall n \in \{0, \dots, N-1\},\ y_{n+1} = y_n + h_n \phi(t_n, y_n, h_n)$$

et

$$\begin{cases} \widetilde{y}_0 \in \mathbb{R} \\ \forall n \in \{0, \dots, N-1\},\ \widetilde{y}_{n+1} = \widetilde{y}_n + h_n \phi(t_n, \widetilde{y}_n, h_n) + \epsilon_{n+1} \end{cases}$$

on a

$$\max_{0\leqslant n\leqslant N} |\widetilde{y}_n - y_n| \leqslant S \sum_{i=0}^{N} |\epsilon_i|$$

où $\epsilon_0 := |\widetilde{y}_0 - y_0|$. Dans ce cas, on dit que la constante S est une constante de stabilité de la méthode.

Remarque 5.3.2. La condition de stabilité est une condition de maitrise de l'erreur "globale" engendrée par des erreurs produites à chaque étape de la méthode.

On en vient maintenant à la notion de *convergence* de la méthode considérée. Le but est d'approcher "le mieux possible" des images de points de $[t_0, t_0 + T]$ par l'unique solution y de (E). Plus précisément, on souhaite déterminer des subdivisions $(t_0, \dots, t_N)$, $N \in \mathbb{N}\backslash\{0\}$, de $[t_0, t_0+T]$ et des réels $y_1, \dots, y_N$, $N \in \mathbb{N}$, tels que pour tout $n \in \{1, \dots, N\}$, l'erreur $|y_n - y(t_n)|$ soit plus petite qu'une erreur ϵ prescrite.

Si les nombres $y_1, \dots, y_N$ sont les réels définis à l'aide de la relation de récurrence

$$y_{n+1} = y_n + h_n \phi(t_n, y_n, h_n),\ n \in \{0, \dots, N-1\},$$

on est ainsi amené à définir l'erreur globale

$$E(t_0, \dots, t_n) := \max_{1 \leqslant n \leqslant N} |y_n - y(t_n)|.$$

de la méthode considérée (on rappelle que $y(t_0) = y_0$ donc $y_0 - y(t_0) = 0$), ainsi que la notion de convergence suivante :

Définition 5.3.3. *On dit que la méthode est convergente si l'erreur globale $E(t_0, \dots, t_n)$ tend vers 0 quand le pas h tend vers 0, i.e. si*

$$\forall \epsilon > 0, \exists \delta > 0, \forall (t_0, \dots, t_N) \text{ subdivision de } [t_0, t_0 + T], h := \max_{i \in \{0,\dots,n-1\}} (t_{i+1} - t_i) < \delta \Rightarrow E(t_0, \dots, t_N) < \epsilon.$$

Une condition suffisante de convergence de la méthode considérée est que la méthode soit consistante et stable :

Proposition 5.3.4. *On suppose que la méthode considérée est à la fois consistante et stable. Alors la méthode est convergente.*

Démonstration. Pour $n \in \{0, \dots, N\}$, on pose $\widetilde{y}_n := y(t_n)$. On a alors, pour tout $n \in \{0, \dots, N-1\}$, par définition,

$$e_n = \widetilde{y}_{n+1} - \left(\widetilde{y}_n + h_n \phi(t_n, \widetilde{y}_n, h_n)\right)$$

i.e.

$$\widetilde{y}_{n+1} = \widetilde{y}_n + h_n \phi(t_n, \widetilde{y}_n, h_n) + e_n.$$

Comme la méthode a été supposée stable, si $S \in]0; +\infty[$ est une constante de stabilité de la méthode, on a

$$\max_{0 \leqslant n \leqslant N} |\widetilde{y}_n - y_n| \leqslant S \left(|\widetilde{y}_0 - y_0| + \sum_{i=0}^{N-1} |e_i| \right) = S \sum_{i=0}^{N-1} |e_i|$$

($\widetilde{y}_0 = y(t_0) = y_0$) et donc

$$E(t_0, \dots, t_n) = \max_{0 \leqslant n \leqslant N} |y_n - y(t_n)| = \max_{0 \leqslant n \leqslant N} |\widetilde{y}_n - y_n| \xrightarrow[h \to 0]{} 0$$

car $\sum_{i=0}^{N-1} |e_i| \xrightarrow[h \to 0]{} 0$, la méthode considérée ayant été supposée consistante. □

La preuve de la proposition précédente nous montre en particulier que, si la méthode est stable et consistante, la "vitesse de convergence" de l'erreur de consistance vers 0 détermine la vitesse de convergence de l'erreur globale vers 0. On est ainsi amené à définir une notion d'*ordre de consistance* de la méthode considérée :

Définition 5.3.5. *Soit $p \in \mathbb{N}\backslash\{0\}$. On dit que la méthode considérée est d'ordre de consistance au moins p s'il existe une constante positive $M \in [0; +\infty[$ telle que*

$$e(t_0, \ldots, t_N) \leqslant h^p M.$$

Remarquons qu'une méthode d'ordre de consistance au moins 1 est nécessairement consistante.

Si la méthode est d'ordre de consistance au moins $p \in \mathbb{N}\backslash\{0\}$, i.e. s'il existe $M \in [0; +\infty[$ telle que $e(t_0, \ldots, t_N) \leqslant h^p M$, et si la méthode est de plus stable, alors, d'après la preuve de la proposition 5.3.4, on a

$$E(t_0, \ldots, t_N) \leqslant h^p S M,$$

si S est une constante de stabilité de la méthode.

Exemple 5.3.6. Soit $p \in \mathbb{N}\backslash\{0\}$. Si la fonction $f : I \times \mathbb{R} \to \mathbb{R}$ est de classe $\mathcal{C}^p$ et si la subdivision $(t_0, \ldots, t_N)$ est régulière, la méthode de Taylor d'ordre p est d'ordre de consistance au moins p (en particulier la méthode d'Euler est d'ordre de consistance au moins 1) : on a, avec les notations de la remarque 5.2.7,

$$\begin{aligned}
e(t_0, \ldots, t_N) &= \sum_{n=0}^{N-1} |e_n| \\
&= \sum_{n=0}^{N-1} \left| \frac{h^{p+1}}{(p+1)!} f^{[p]}(\xi_n, y(\xi_n)) \right| \\
&\leqslant \sum_{n=0}^{N-1} \frac{h^{p+1}}{(p+1)!} \max_{t \in [t_0, t_0+T]} \left| f^{[p]}(t, y(t)) \right| \\
&= \frac{h^{p+1}}{(p+1)!} \left\| f^{[p]}(\cdot, y(\cdot)) \right\|_{\infty, [t_0, t_0+T]} \left(\sum_{n=0}^{N-1} 1 \right) \\
&= N \frac{T^{p+1}}{N^{p+1}} \frac{\left\| f^{[p]}(\cdot, y(\cdot)) \right\|_{\infty, [t_0, t_0+T]}}{(p+1)!} \\
&= h^p \frac{T \left\| f^{[p]}(\cdot, y(\cdot)) \right\|_{\infty, [t_0, t_0+T]}}{(p+1)!}.
\end{aligned}$$

Nous avons par ailleurs le résultat suivant :

Théorème 5.3.7. *Soit $p \in \mathbb{N}\backslash\{0\}$. Supposons que la fonction f soit de classe $\mathcal{C}^p$ sur $I \times \mathbb{R}$ et que ϕ soit de classe $\mathcal{C}^p$ par rapport à sa troisième variable κ. Supposons de plus que pour tout $k \in \{0, \ldots, p-1\}$, pour tout $t \in [t_0, t_0 + T]$, pour tout $\gamma \in \mathbb{R}$,*

$$\frac{\partial^k \phi}{\partial \kappa^k}(t, \gamma, 0) = \frac{1}{k+1} f^{[k]}(t, \gamma).$$

Supposons enfin que la subdivision $(t_0, \dots, t_N)$ de $[t_0, t_0 + T]$ est régulière.

Alors la méthode à un pas associée à ϕ est d'ordre de consistance au moins p.

Démonstration. Soit $n \in \{0, \dots, N-1\}$. D'après le théorème de Taylor-Lagrange, Il existe $\kappa_n \in]0, h[$ tel que

$$\phi(t_n, y(t_n), h) = \sum_{k=0}^{p-1} \frac{h^k}{k!} \frac{\partial^k \phi}{\partial \kappa^k}(t_n, y(t_n), 0) + \frac{h^p}{p!} \frac{\partial^p \phi}{\partial \kappa^p}(t_n, y(t_n), \kappa_n)$$

D'autre part, il existe $\xi_n \in]t_n, t_{n+1}[$ tel que

$$\begin{aligned} y(t_{n+1}) &= y(t_n) + \sum_{k=1}^{p} \frac{h^k}{k!} y^{(k)}(t_n) + \frac{h^{p+1}}{(p+1)!} y^{(p+1)}(\xi_n) \\ &= y(t_n) + \sum_{k=1}^{p} \frac{h^k}{k!} f^{[k-1]}(t_n, y(t_n)) + \frac{h^{p+1}}{(p+1)!} f^{[p]}(\xi_n, y(\xi_n)) \end{aligned}$$

Ainsi,

$$\begin{aligned} e_n &= y(t_{n+1}) - \big(y(t_n) + h\phi(t_n, y(t_n), h)\big) \\ &= \sum_{k=1}^{p} \frac{h^k}{k!} f^{[k-1]}(t_n, y(t_n)) + \frac{h^{p+1}}{(p+1)!} f^{[p]}(\xi_n, y(\xi_n)) - \left(\sum_{k=0}^{p-1} \frac{h^{k+1}}{k!} \frac{\partial^k \phi}{\partial \kappa^k}(t_n, y(t_n), 0) + \frac{h^{p+1}}{p!} \frac{\partial^p \phi}{\partial \kappa^p}(t_n, y(t_n), \kappa_n) \right) \\ &= \sum_{k=0}^{p-1} \frac{h^{k+1}}{(k+1)!} f^{[k]}(t_n, y(t_n)) + \frac{h^{p+1}}{(p+1)!} f^{[p]}(\xi_n, y(\xi_n)) - \left(\sum_{k=0}^{p-1} \frac{h^{k+1}}{k!} \frac{\partial^k \phi}{\partial \kappa^k}(t_n, y(t_n), 0) + \frac{h^{p+1}}{p!} \frac{\partial^p \phi}{\partial \kappa^p}(t_n, y(t_n), \kappa_n) \right) \\ &= \frac{h^{p+1}}{(p+1)!} f^{[p]}(\xi_n, y(\xi_n)) - \frac{h^{p+1}}{p!} \frac{\partial^p \phi}{\partial \kappa^p}(t_n, y(t_n), \kappa_n) \end{aligned}$$

car, par hypothèse, pour tout $k \in \{0, \dots, p-1\}$, $\dfrac{1}{k+1} f^{[k]}(t_n, y(t_n)) = \dfrac{\partial^k \phi}{\partial \kappa^k}(t_n, y(t_n), 0)$.

On a donc

$$\begin{aligned} |e_n| &\leqslant h^{p+1} \left(\frac{1}{(p+1)!} \left| f^{[p]}(\xi_n, y(\xi_n)) \right| + \frac{1}{p!} \left| \frac{\partial^p \phi}{\partial \kappa^p}(t_n, y(t_n), \kappa_n) \right| \right) \\ &\leqslant h^{p+1} \left(\frac{1}{(p+1)!} \left\| f^{[p]}(\cdot, y(\cdot)) \right\|_{\infty, [t_0, t_0+T]} + \frac{1}{p!} \left\| \frac{\partial^p \phi}{\partial \kappa^p}(\cdot, y(\cdot), \bullet) \right\|_{\infty, [t_0, t_0+T] \times [0,h]} \right) \end{aligned}$$

Notons M le facteur de h^{p+1} ci-dessus. On a alors

$$\begin{aligned} e(t_0, \dots, t_N) &= \sum_{n=0}^{N-1} |e_n| \\ &\leqslant \sum_{n=0}^{N-1} h^{p+1} M \\ &= N h^{p+1} M \\ &= h^p T M \end{aligned}$$

$(h = \frac{T}{N})$. □

Remarque 5.3.8. Supposons que la fonction f soit de classe $\mathcal{C}^p$ et que la subdivision $(t_0, \dots, t_N)$ soit régulière. Alors la méthode de Taylor d'ordre p vérifie l'hypothèse du théorème 5.3.7. En effet, la méthode de Taylor d'ordre p est la méthode à un pas associée à la fonction

$$\begin{array}{rccc} & [t_0, t_0+T] \times \mathbb{R} \times \mathbb{R} & \to & \mathbb{R} \\ \phi : & (t, \gamma, \kappa) & \mapsto & \displaystyle\sum_{k=0}^{p-1} \frac{\kappa^k}{(k+1)!} f^{[k]}(t, \gamma) \end{array}$$

de classe $\mathcal{C}^\infty$ par rapport à sa troisième variable, et on a, si $k \in \{0, \dots, p-1\}$ et $t \in [t_0, t_0+T]$,

$$\frac{\partial^k \phi}{\partial \kappa}(t, \gamma, 0) = \frac{1}{k+1} f^{[k]}(t, \gamma).$$

Continuons par une condition suffisante de stabilité pour la méthode considérée :

Théorème 5.3.9. *On suppose que la fonction ϕ est lipschitzienne par rapport à sa deuxième variable γ, i.e. qu'il existe une constante $C \in]0; +\infty[$ telle que pour tout $t \in I$, pour tout $\kappa \in \mathbb{R}$, pour tous $y_1, y_2 \in \mathbb{R}$,*

$$|\phi(t, y_1, \kappa) - \phi(t, y_2, \kappa)| \leqslant C|y_1 - y_2|.$$

Alors la méthode à un pas associée à ϕ est stable.

Démonstration. On reprend les notations de la définition 5.3.1. Soit alors $n \in \{0, \dots, N-1\}$, on a, en utilisant le fait que, pour tout $x \in \mathbb{R}$, $1 + x \leqslant e^x$,

$$\begin{aligned} |\widetilde{y}_{n+1} - y_{n+1}| &= |\widetilde{y}_n - y_n + h_n \left(\phi(t_n, \widetilde{y}_n, h_n) - \phi(t_n, y_n, h_n)\right) + \epsilon_{n+1}| \\ &\leqslant |\widetilde{y}_n - y_n| + h_n \left|\phi(t_n, \widetilde{y}_n, h_n) - \phi(t_n, y_n, h_n)\right| + |\epsilon_{n+1}| \\ &\leqslant |\widetilde{y}_n - y_n| + h_n C \left|\widetilde{y}_n - y_n\right| + |\epsilon_{n+1}| \\ &= (1 + h_n C) \left|\widetilde{y}_n - y_n\right| + |\epsilon_{n+1}| \\ &\leqslant e^{h_n C} \left|\widetilde{y}_n - y_n\right| + |\epsilon_{n+1}| \\ &\leqslant e^{h_n C} \left(e^{h_{n-1} C} \left|\widetilde{y}_{n-1} - y_{n-1}\right| + |\epsilon_n|\right) + |\epsilon_{n+1}| \\ &\dots \\ &\leqslant e^{h_n C + \dots + h_0 C} \left|\widetilde{y}_0 - y_0\right| + \left(\sum_{k=1}^{n} e^{h_n C + \dots + h_k C} |\epsilon_k|\right) + |\epsilon_{n+1}| \\ &\leqslant e^{(h_0 + \dots + h_{N-1}) C} \sum_{k=0}^{n} |\epsilon_k| + |\epsilon_{n+1}| \\ &\leqslant e^{TC} \sum_{k=0}^{n} |\epsilon_k| + |\epsilon_{n+1}| \\ &\leqslant S \sum_{k=0}^{n+1} |\epsilon_k| \\ &\leqslant S \sum_{k=0}^{N} |\epsilon_k| \end{aligned}$$

où S est le maximum de e^{TC} et 1. Enfin,

$$|\widetilde{y}_0 - y_0| = \epsilon_0 \leqslant \sum_{k=0}^{N} |\epsilon_k| \leqslant S \sum_{k=0}^{N} |\epsilon_k|\,,$$

et on a donc

$$\max_{0\leqslant n\leqslant N} |\widetilde{y}_n - y_n| \leqslant S \sum_{i=0}^{N} |\epsilon_i|.$$

□

Corollaire 5.3.10. *Supposons que la fonction f soit lipschitzienne par rapport à sa deuxième variable γ. Alors la méthode d'Euler associée est convergente.*

Démonstration. La "fonction ϕ" associée à la méthode d'Euler est donnée par $\phi(t,\gamma,\kappa) = f(t,\gamma)$: si f est lipschitzienne par rapport à sa seconde variable, alors ϕ est également lipschitzienne par rapport à sa deuxième variable, et la méthode d'Euler est donc stable par le théorème précédent 5.3.9. De plus, la méthode d'Euler est d'ordre de consistance au moins 1 (cf. exemple 5.3.6) et est donc également consistante : par la proposition 5.3.4, elle est donc convergente. □

5.4 Méthodes de Runge-Kutta

Dans cette section, nous allons considérer des méthodes de résolution numérique de (E) à un pas construites à l'aide de formules d'intégration numérique d'ordre d'exactitude au moins 0.

Le principe est le suivant. Soient $c_1,\dots,c_q \in [0;1]$, $q \in \mathbb{N}\backslash\{0\}$, tels que $0 = c_1 \leqslant \cdots \leqslant c_q \leqslant 1$. Soit maintenant $n \in \{0,\dots,N-1\}$ et, si $i \in \{1,\dots,q\}$, posons

$$t_{n,i} := t_n + c_i h_n.$$

Soient ensuite $b_1,\dots,b_q \in \mathbb{R}$ tels que $\displaystyle\sum_{i=1}^{q} b_i = 1$ et considérons la formule de quadrature sur $[t_n, t_{n+1}]$ donnée par

$$h_n \sum_{i=1}^{q} b_i g(t_{n,i})$$

si g est une fonction Riemann-intégrable sur $[t_n, t_{n+1}]$. Pour motiver ces considérations, remarquons que, comme pour tout $t \in I$, $y'(t) = f(t, y(t))$, on a

$$y(t_{n+1}) - y(t_n) = \int_{t_n}^{t_{n+1}} f(t, y(t))\mathrm{d}t,$$

et que l'on peut ensuite approcher cette dernière intégrale par

$$h_n \sum_{i=1}^{q} b_i f\big(t_{n,i}, y(t_{n,i})\big).$$

Lemme 5.4.1. *La formule de quadrature ci-dessus, définie par les réels $c_1, \dots, c_q$ et $b_1, \dots, b_q$, est d'ordre d'exactitude au moins* 0.

Démonstration. On a

$$h_n \sum_{i=1}^{q} b_i = h_n = t_{n+1} - t_n = \int_{t_n}^{t_{n+1}} \mathrm{d}t.$$

□

Supposons enfin que $q \geqslant 2$, soit $i \in \{2, \dots, q\}$ et soient $a_{i,1}, \dots, a_{i,i-1} \in \mathbb{R}$ tels que $\sum_{j=1}^{i-1} a_{i,j} = c_i$. Considérons alors la formule de quadrature sur $[t_n, t_{n,i}]$ donnée par

$$h_n \sum_{j=1}^{i-1} a_{i,j} g(t_{n,j})$$

si g est une fonction Riemann-intégrable sur $[t_n, t_{n,i}]$: on a

$$y(t_{n,i}) - y(t_n) = \int_{t_n}^{t_{n,i}} f(t, y(t)) \mathrm{d}t$$

et on peut approcher cette dernière intégrale par

$$h_n \sum_{j=1}^{i-1} a_{i,j} f\big(t_{n,j}, y(t_{n,j})\big).$$

Lemme 5.4.2. *La formule de quadrature ci-dessus, définie par les réels $c_i, \dots, c_i$ et $a_{i,1}, \dots, a_{i,i-1}$, est d'ordre d'exactitude au moins* 0.

Démonstration. On a

$$h_n \sum_{j=1}^{i-1} a_{i,j} = h_n c_i = t_{n,i} - t_n = \int_{t_n}^{t_{n,i}} \mathrm{d}t.$$

□

Les approximations successives

$$\begin{cases} y(t_{n,2}) \sim y(t_n) + h_n a_{2,1} f\big(t_n, y(t_n)\big) \\ y(t_{n,3}) \sim y(t_n) + h_n a_{3,1} f\big(t_n, y(t_n)\big) + h_n a_{3,2} f\big(t_{n,2}, y(t_{n,2})\big) \\ \vdots \\ y(t_{n,q}) \sim y(t_n) + h_n \displaystyle\sum_{j=1}^{q-1} a_{i,j} f\big(t_{n,j}, y(t_{n,j})\big) \\ \\ y(t_{n+1}) \sim y(t_n) + h_n \displaystyle\sum_{i=1}^{q} b_i f\big(t_{n,i}, y(t_{n,i})\big) \end{cases}$$

motivent la méthode de résolution numérique de (E) à un pas suivante, construite par récurrence : on considère les réels $y_{n,1}, \dots, y_{n,q}$ définis successivement par

$$\begin{cases} y_{n,1} & := y_n \\ y_{n,2} & := y_n + h_n a_{2,1} f\big(t_n, y_n\big) \\ y_{n,3} & := y_n + h_n a_{3,1} f\big(t_n, y_n\big) + h_n a_{3,2} f\big(t_{n,2}, y_{n,2}\big) \\ & \vdots \\ y_{n,q} & := y_n + h_n \displaystyle\sum_{j=1}^{q-1} a_{i,j} f\big(t_{n,j}, y_{n,j}\big) \end{cases}$$

puis le réel

$$y_{n+1} := y_n + h_n \sum_{i=1}^{q} b_i f\big(t_{n,i}, y_{n,i}\big).$$

Définition 5.4.3. *La méthode ci-dessus est appelée méthode de Runge-Kutta associée aux nœuds $c_1, \dots, c_q$, aux poids $b_1, \dots, b_q$ et aux coefficients de Runge-Kutta $a_{i,j}$, $i \in \{2, \dots, q\}$, $j \in \{1, \dots, i-1\}$.*

On réunit ces paramètres sous la forme du tableau

$$\begin{array}{c|ccccc} 0 & & & & & \\ c_2 & a_{2,1} & & & & \\ c_3 & a_{3,1} & a_{3,2} & & & \\ \vdots & & & \ddots & & \\ c_q & a_{q,1} & a_{q,2} & \cdots & a_{q,q-1} & \\ \hline & b_1 & b_2 & \cdots & b_{q-1} & b_q \end{array}$$

appelé tableau de Butcher.

Avant de donner quelques exemples de méthodes de Runge-Kutta, énonçons, sans démonstration, le théorème de convergence suivant :

Théorème 5.4.4. *Les méthodes de Runge-Kutta sont consistantes. Si f est lipschitzienne en sa seconde variable, elles sont également stables. Ainsi, si f est lipschitzienne en sa seconde variable, toute méthode de Runge-Kutta de résolution numérique de* (E) *est convergente.*

Exemple 5.4.5. 1. La méthode de Runge-Kutta associée au tableau de Butcher

$$\begin{array}{c|c} 0 & \\ \hline & 1 \end{array}$$

est la méthode à un pas donnée par l'expression, si $n \in \{0, \dots, N-1\}$,

$$y_{n+1} = y_n + h_n f(t_n, y_n),$$

i.e. la méthode d'Euler.

2. La méthode de Runge-Kutta associée au tableau de Butcher

$$\begin{array}{c|cc} 0 & & \\ \frac{1}{2} & \frac{1}{2} & \\ \hline & 0 & 1 \end{array}$$

est la méthode à un pas donnée par l'expression, si $n \in \{0, \dots, N-1\}$,

$$y_{n+1} = y_n + h_n f\left(t_n + \frac{h_n}{2}, y_n + \frac{h_n}{2} f(t_n, y_n))\right).$$

Cette méthode est appelée méthode de Runge.

3. La méthode de Runge-Kutta associée au tableau de Butcher

$$\begin{array}{c|ccc} 0 & & & \\ \frac{1}{3} & \frac{1}{3} & & \\ \frac{2}{3} & 0 & \frac{2}{3} & \\ \hline & \frac{1}{4} & 0 & \frac{3}{4} \end{array}$$

est la méthode à un pas donnée par l'expression, si $n \in \{0, \dots, N-1\}$,

$$y_{n+1} = y_n + h_n \left(\frac{1}{4} f(t_n, y_n) + \frac{3}{4} f\left(t_n + \frac{2}{3}h_n, y_n + \frac{2}{3}h_n f\left(t_n + \frac{h_n}{3}, y_n + \frac{h_n}{3} f(t_n, y_n)\right)\right)\right).$$

Cette méthode est appelée méthode de Heun.