ANALYSE NUMERIQUE

Deuxième année M.I.

Prof. HAMRI NASR-EDDINE

Table des matières

ANALYSE NUMERIQUE 1

1. NOTIONS D'ERREURS

1.1 PRÉLIMINAIRES

L'outil fondamental en analyse numérique (qui nous fournit des méthodes de calcul pour l'étude et la solution approchée de problèmes mathématiques dont la résolution est généralement impossible ou impraticable), demeure la formule de Taylor.

Ces solutions approchées sont le plus souvent calculées sur ordinateur au moyen d'algorithmes convenables. Dans ce qui suit nous rappelons quelques théorèmes dont la connaissance est impérative pour une meilleure compréhension de la suite.

Théoreme 1.1.1 — de la valeur intermédiaire. Soit f une fonction définie sur un intervalle $[a,b]$, on définit $m = \inf_{x\in[a,b]} f(x)$ et $M = \sup_{x\in[a,b]} f(x)$. Alors pour tout y dans $[m,M]$, il existe au moins un point x dans $[a,b]$ pour lequel

$$f(x) = y.$$

Théoreme 1.1.2 — des accroissements finis. Soit f une fonction définie et continue sur un intervalle $[a,b]$, diff érentiable sur $]a,b[$. Alors il existe au moins un point c dans $[a,b]$ pour lequel

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

Théoreme 1.1.3 — de la moyenne. Soit $w(x)$ une fonction non négative définie et intégrable sur un intervalle $[a,b]$, et soit $f(x)$ une fonction continue sur $]a,b[$. Alors

$$\int_a^b w(x)f(x)dx = f(\xi)\int_a^b w(x)dx$$

pour $\xi \in [a,b]$.

Théoreme 1.1.4 — de Taylor. Soit f une fonction définie et continue sur un intervalle $[a,b]$, $(n+1)$ fois dérivable sur $[a,b]$ pour $n \geq 0$, et soit $x, x_0 \in [a,b]$. Alors

$$f(x) = p_n(x) + R_{n+1}(x). \tag{1.1}$$

Où

$$\begin{aligned} p_n(x) &= f(x_0) + \frac{(x-x_0)}{1!} f'(x_0) \\ &\quad + \ldots + \frac{(x-x_0)^n}{n!} f^n(x_0) \end{aligned}$$

Et

$$\begin{aligned} R_{n+1}(x) &= \frac{1}{n!} \int_{x_0}^{x} (x-t)^n f^{(n+1)}(t) dt \\ &= \frac{(x-x_0)^{n+1}}{(n+1)!} f^{(n+1)}(\xi) \end{aligned} \tag{1.2}$$

pour $\xi \in]x_0, x[$

En utilisant la formule de Taylor on obtient par exemple les formules suivantes :

$$e^x = 1 + x + \frac{x^2}{2} + \ldots + \frac{x^n}{n!} + \frac{x^{n+1}}{(n+1)!} e^{\xi_x} \tag{1.3}$$

$$\begin{aligned} \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots + (-1)^n \frac{x^{2n}}{2n!} + \\ &\quad + (-1)^{n+1} \frac{x^{2n+2}}{(2n+2)!} \cos(\xi_x) \end{aligned} \tag{1.4}$$

$$(1-x)^{-1} = 1 + x + x^2 + \ldots + x^n + \frac{x^{n+1}}{1-x}, \quad x \neq 1 \tag{1.5}$$

De cette dernière formule nous pouvons déduire :

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \quad |x| < 1 \tag{1.6}$$

On peut calculer les séries de Taylor de n'importe quelle fonction suffisamment dérivable avec autant de termes que l'on veut. Cependant à cause de la complexité de la différentiation de plusieurs fonctions, il est souvent préférable d'obtenir indirectement leur polynôme d'approximation de Taylor $p_n(x)$ ou leur séries de Taylor, en utilisant l'un des développement limités connus. Les trois exemples qui suivent montrent que les erreurs sont plus simples que lorsque l'on utilise la formule de l'erreur (1.2).

1.1.1 Exemples

1. $f(x) = e^{-x^2}$, En remplaçant x par $-x^2$ dans (1.3), on obtient :

$$e^{-x^2} = 1 - x^2 + \frac{x^4}{2} - \ldots + (-1)^n \frac{x^2 n}{n!} + (-1)^{n+1} \frac{x^{2n+2}}{(n+1)!} e^{\xi_x}$$

   avec $\xi_x \in [-x^2, 0]$.
2. $f(x) = \frac{1}{\tan(x)}$, En posant $x = -u^2$ dans le développement de $\frac{1}{1-x}$ on a :

$$\frac{1}{1+u^2} = 1 - u^2 + u^4 - \ldots + (-1)^n u^{2n} + (-1)^{n+1} \frac{u^{2n+2}}{1+u^2}$$

en intégrant sur $[0,x]$ on aboutit à :

$$\frac{1}{\tan(x)} = x - \frac{x^3}{3} + \frac{x^5}{5} - \ldots + (-1)^n \frac{x^{2n+1}}{2n+1} + (-1)^n \int_0^x \frac{u^{2n+2}}{1+u^2} du$$

En appliquant le théorème de la moyenne, on obtient :

$$\int_0^x \frac{u^{2n+2}}{1+u^2} du = \frac{x^{2n+3}}{2n+3} \cdot \frac{1}{1+\xi_x^2}$$

avec $\xi_x \in [0,x]$.

3. $f(x) = \int_0^1 \sin(xt)dt$, Utilisant le développement de $\sin x$, et intégrant, on écrit :

$$f(x) = \sum_{j=1}^{n} (-1)^{j-1} \frac{x^{2j-1}}{2j!} + (-1)^n \frac{x^{2n+1}}{(2n+1)!} \int_0^1 t^{2n+1} \cos(\xi_{xt}) dt$$

avec $\xi_{xt} \in [0,xt]$. L'intégrale dont le reste est bornée par $\frac{1}{2n+2}$, mais on peut aussi la mettre sous une forme simplifiée, et en appliquant le théorème de la moyenne on a :

$$\int_0^1 \sin(xt)dt = \sum_{j=1}^{n} (-1)^{j-1} \frac{x^{2j-1}}{2j!} + (-1)^n \frac{x^{2n+1}}{(2n+1)!} \cos(\xi_x)$$

avec $\xi_x \in [0,x]$.

1.2 ERREURS ABSOLUES et ERREURS RELATIVES

Un nombre approché x est légèrement différent du nombre exact X, et qui dans les calculs remplace X.

- Si $x < X$, x est dit valeur par *défaut* .
- Si $x > X$, x est dit valeur par *excès* .

On note généralement $x \approx X$.

▪ Exemple 1.1

$$1,41 < \sqrt{2} < 1,42$$

▪

Définition 1.2.1 On appelle erreur Δx d'un nombre approché, la valeur :

$$\Delta x = X - x,$$

C'est-à-dire

$$X = x + \Delta x$$

Définition 1.2.2 On appelle erreur absolue Δ d'un nombre x la valeur

$$\Delta = | X - x | \tag{1.7}$$

R

1. Si X est connu, l'erreur absolue est déterminée par (1.7).
2. Si X est inconnu, l'erreur absolue Δ est impossible à déterminer. Dans ce cas on introduit la limite supérieure de l'erreur absolue.

Définition 1.2.3 On appelle limite supérieure ou borne supérieure de l'erreur absolue tout nombre supérieur ou égal à l'erreur absolue de ce nombre. C'est-à-dire :

$$\Delta = \mid X - x \mid \leq \Delta_x$$

si Δ_x désigne la borne supérieure, donc

$$x - \Delta_x \leq X \leq x + \Delta_x$$

On note

$$X = x \pm \Delta_x.$$

■ **Exemple 1.2** Trouver la borne supérieure d'erreur absolue de $\pi = 3,14$.

$$3,14 \leq \pi \leq 3,15$$

Dans ce cas $\mid x - \pi \mid \leq 0,01$, on peut poser $\Delta_x = 0,01$, comme $3,140 \leq \pi \leq 3,142$, une meilleure estimation de la borne d'erreur absolue est $\Delta = 0,002$. ■

Définition 1.2.4 On appelle erreur relative notée δ d'un nombre x, le rapport suivant :

$$\delta = \frac{\Delta}{\mid X \mid}, \quad (X \neq 0)$$

C'est-à-dire

$$\Delta = \mid X \mid \delta.$$

Définition 1.2.5 La borne supérieure de l'erreur relative δ_x est un nombre supérieur ou égal à l'erreur relative de ce nombre. C'est-à-dire :

$$\delta \leq \delta_x$$

c'est-à-dire

$$\frac{\Delta}{\mid X \mid} \leq \delta_x,$$

donc

$$\Delta \leq \mid X \mid \delta_x$$

On peut aussi utiliser

$$\Delta_x = \mid x \mid \delta_x$$

car $X \approx x$.

R Si l'on connait une borne d'erreur relative δ_x on a :

$$X = x(1 \pm \delta_x)$$

C'est-à-dire

$$\delta_x = \frac{\Delta_x}{x - \Delta_x}.$$

De même on obtient :

$$\Delta_x = \frac{x\delta_x}{1-\delta_x}.$$

■ **Exemple 1.3** En cherchant la constante de gaz, on a obtenu $R = 29,25$, l'erreur relative étant $1^0/_{00}$ trouver un encadrement de R. On a $\delta_x = 0,001$, donc $\Delta_x = R\delta_x = 0,03$, c'est-à-dire :

$$29,22 \leq R \leq 29,28.$$

■

1.3 PRINCIPALES SOURCES D'ERREURS

Les erreurs commises dans les problèmes peuvent être des :

Erreurs inhérentes au problème : Erreurs dûes à la position même du problème. Le modèle théorique est trés rarement fidèle au modèle réel. Lors de l'étude d'un phénomène de la nature on est souvent contraint d'admettre certaines conditions.

Erreurs de la méthode : Il arrive qu'il soit difficile ou même impossible de résoudre un problème enoncé en termes exacts. On le remplace par un problème approché.

Erreurs de troncature : Associées aux processus infinis. Les fonctions données dans les formules le sont sous forme de suites infinies ou de séries, on est donc obligé de mettre fin à un certain terme de la suite. Par exemple, l'approximation d'une somme infinie par une somme finie, l'approximation de la limite d'une suite par un terme de "grand indice" ou encore l'approximation d'une intégrale par une somme finie.

Erreurs initiales : Dûes à la présence dans les formules de paramètres dont les valeurs sont approchées.

Erreurs d'arrondi : Dûes au système de numérisation.

Erreurs propagées : Les erreurs des données de départ se repercutent sur le résultat des calculs.

1.4 SERIE D'EXERCICES

Exercice 1.1 Trouver une borne de l'erreur absolue du nombre $x = 3.14$ qui remplace $\pi(\pi = 3.1415926...)$ dans les deux cas suivants :

1. $3.14 < \pi < 3.142$.
2. $3.14 < \pi < 3.15$.

■

Exercice 1.2 Supposons que $x_1, x_2, x_3, ...x_n$ approchent respectivement $X_1, X_2, X_3, ..., X_n$ et que dans chaque cas la borne supérieure de l'erreur absolue est ε. Montrer que la borne supérieure de l'erreur de la somme des $x_i (i = 1, 2, ...n)$ est égale à $n\varepsilon$.

■

Exercice 1.3 Donner les bornes des erreurs relatives de $a = 1.414$ et $b = 1.41$ qui approchent $\sqrt{2} = 1.414214.....$

■

Exercice 1.4 En recherchant la constante des gaz de l'air on a obtenu $R \simeq 29.25$. La borne de l'erreur relative de cette valeur étant $1^0/_{00}$, trouver les limites entre lesquelles est comprise R. ■

2. APPROXIMATION

2.1 GÉNÉRALITÉS

Soit $X = \{x_i,\ \ i = 1,2,\ldots n\}$ un ensemble de n points appartenant à l'intervalle $[a,b] \subset \mathbb{R}$.

Supposons qu'à chaque x_i de X, on sache associer un $y_i \in \mathbb{R}$, résultat d'une expérience ou valeur donnée par une table, mais que cette opération soit impossible avec $x \in [a,b] \setminus X$.

Notons

$$\mathscr{D} = \{(x_i, y_i),\ \text{ pour }\ x_i \in X\}.$$

On veut déterminer une fonction, *facilement calculable*, qui permette d'obtenir une estimation *"raisonnable"* de la réponse du phénomène pour la valeur x.

Notons g cette fonction et $\mathscr{G} = \{(x_i, g(x_i)),$ pour $x_i \in X\}$. Si on définit une *"distance"* entre $\mathscr{D}$ et $\mathscr{G}$, nous pouvons chercher une fonction g, d'un type préalablement choisi (un polynôme par exemple), telle que cette distance soit minimale.

C'est ce qu'on appelle une **approximation**.

Un procédé général consiste à limiter la représentation de g aux combinaisons linéaires des $n+1$ fonctions d'une certaine base, choisies à priori.

Si on désigne par

$$u_1(x), u_2(x), u_3(x), \ldots, u_{n+1}(x)$$

les fonctions de la base, les représentations utilisées seront les combinaisons

$$a_1u_1(x) + a_2u_2(x) + a_3u_3(x) + \ldots + a_{n+1}u_{n+1}(x) \tag{2.1}$$

qui dépendent des $n+1$ coefficients a_i que nous devrions calculer pour définir chaque approximation.

R On utilise un procédé analogue lorsque on étudie la représentation de fonctions par des séries de fonctions.
On prend une base infinie $u_1(x), u_2(x), \ldots, u_{n+1}(x), \ldots$ et on considère les séries

$$a_1u_1(x) + a_2u_2(x) + a_3u_3(x) + \ldots + a_{n+1}u_{n+1}(x) + \ldots \tag{2.2}$$

L'equation (2.1) apparaît alors comme la somme partielle des $n+1$ premiers termes de (3.2).

2.2 APPROXIMATION

2.2.1 Meilleure approximation

Soit $(\mathbf{E},d)$ un espace métrique et $\mathbf{F}\subset\mathbf{E}$. Chercher à approcher un élément f de $\mathbf{E}$ par un élément de $\mathbf{F}$, c'est donc déterminer g de $\mathbf{F}$ tel que :

$$d(f,g)=\min_{h\in\mathbf{F}} d(f,h) \tag{2.3}$$

S'il existe, cet élément sera appelé *meilleure approximation* de f dans $\mathbf{F}$ au sens de la distance d.

Théoreme 2.2.1 — d'existence. Si $\mathbf{F}$ est une partie compacte de $\mathbf{E}$, alors il existe au moins un élément g de $\mathbf{F}$ tel que :

$$d(f,g)=\min_{h\in\mathbf{F}} d(f,h).$$

Corollaire 2.2.2 Soit $\mathbf{E}$ un espace normé. Si $\mathbf{F}$ est un sous espace vectoriel de $\mathbf{E}$ de dimension finie, alors il existe au moins un élément g de $\mathbf{F}$ tel que :

$$\| f-g \| = \min_{h\in\mathbf{F}} \| f-h \| .$$

R

1. Nous supposerons dorénavant que $\mathbf{E}$ est l'ensemble des fonctions continues sur un intervalle $[a,b]$ de $\mathbb{R}$.
2. g est le plus souvent cherchée sous la forme suivante :

$$g(x)=a_1u_1(x)+a_2u_2(x)+a_3u_3(x)+\ldots+a_{n+1}u_{n+1}(x)$$

où les $u_i(x)$ sont des fonctions choisies dans :

— la classe des monômes :

$$1,x,x^2,\ldots,x^n.$$

— la classe des fonctions trigonométriques :

$$1,\sin x,\sin 2x,\ldots,\sin nx,\cos x,\cos 2x,\ldots,\cos nx.$$

— la classe des fonctions exponentielles :

$$1,\exp x,\exp 2x,\ldots,\exp nx$$

C'est-à-dire que :

—

$$g(x)=a_1x^n+a_2x^{n-1}+a_3x^{n-2}+\ldots+a_nx+a_{n+1},$$

ou ;

$$\begin{aligned} g(x) \quad = \quad & c+a_1\cos x+a_2\cos 2x+a_3\cos 3x+\ldots+a_n\cos nx+\\ & +b_1\sin x+b_2\sin 2x+b_3\sin 3x+\ldots+b_n\sin nx, \end{aligned}$$

ou ;

$$g(x)=a_1\exp b_1x+a_2\exp b_2x+a_3\exp b_3x+\ldots+a_n\exp b_nx,$$

ou aussi ;

$$g(x)=\frac{a_1x^n+a_2x^{n-1}+a_3x^{n-2}+\ldots+a_nx+a_{n+1}}{b_1x^n+b_2x^{n-1}+b_3x^{n-2}+\ldots+b_nx+b_{n+1}},$$

3. L'approximation de f par des polynômes, dite **approximation polynomiale**, est la plus fréquente.

Théoreme 2.2.3 Si f est une fonction continue sur $[a,b]$, pour tout $\varepsilon > 0$, il existe un polynôme P_n de degré inférieur ou égal à n tel que :

$$\max_{x\in[a,b]} \mid f(x) - P_n(x) \mid < \varepsilon.$$

2.3 APPROXIMATION AU SENS DES MOINDRES CARRÉS

Définition 2.3.1 Soit $\mathbf{E}$ un espace vectoriel. Pour tout couple (f,g) de $\mathbf{E}\times\mathbf{E}$, on définit le produit scalaire noté $\langle f,g\rangle$ et la norme associée par :

$$|| f ||= \sqrt{\langle f,f\rangle}.$$

Pour tout $f \in \mathbf{E}$, il existe $g \in \mathbf{F}$ ($\mathbf{F}$ est un sous espace vectoriel de $\mathbf{E}$, de dimension finie) tel que :

$$|| f-g ||= \min_{h\in\mathbf{F}} || f-h || .$$

Théoreme 2.3.1 Une condition nécessaire et suffisante pour que g soit une meilleure approximation de f est que :

$$\langle f-g,h\rangle = 0 \qquad \text{pour tout} \quad h \in \mathbf{F} \tag{2.4}$$

g est appelée meilleure approximation de f au sens des moindres carrés.

Démonstration. La Condition est nécessaire : Soit g la meilleure approximation de f. Faisons un raisonnement par l'absurde. Supposons qu'il existe $h_1 \in \mathbf{F}$ tel que :

$$\langle f-g,h_1\rangle = c \neq 0.$$

Soit $h_2 \in \mathbf{F}$ défini par :

$$h_2 = g + \frac{c}{|| h_1 ||^2} h_1.$$

on a

$$\begin{aligned} || f-h_2 || &= \sqrt{\langle f-g-\frac{c}{|| h_1 ||^2}h_1, f-g-\frac{c}{|| h_1 ||^2}h_1\rangle} \\ &= \sqrt{|| f-g ||^2 - \frac{c^2}{|| h_1 ||^2}} < || f-g || \end{aligned}$$

Ce qui est absurde.

La Condition est suffisante : Soit $g_1 \in \mathbf{F}$ tel que : $\langle f-g_1,h\rangle = 0$ pour tout $h \in \mathbf{F}$ on a alors :

$$\begin{aligned} || f-h || &= \sqrt{\langle f-h, f-h\rangle} \\ &= \sqrt{\langle f-g_1-h+g_1, f-g_1-h+g_1\rangle} \\ &= \sqrt{|| f-g_1 ||^2 + || h-g_1 ||^2}. \end{aligned}$$

D'où $|| f-g_1 || < || f-h ||$ pour tout $h \in \mathbf{F}$ et donc $g_1 = g$. ■

Théoreme 2.3.2 La meilleure approximation au sens des moindres carrés est unique.

Démonstration. Soient g_1 et g_2 deux meilleures approximations de f on a donc :

$$\langle f-g_1,h\rangle = 0 = \langle f-g_2,h\rangle \quad \text{pour tout} \quad h \in \mathbf{F}$$

en particulier pour $h = g_1 - g_2$, d'où

$$\begin{array}{rl} | & |\, g_1-g_2\,| = \sqrt{\langle g_1-g_2+f-f, g_1-g_2\rangle} = \\ = & \sqrt{\langle f-g_1, g_1-g_2\rangle + \langle f-g_2, g_1-g_2\rangle} = 0 \end{array}$$

donc $g_1 = g_2$. ■

2.4 CARACTÉRISATION

Il s'agit maintenant de construire le polynôme d'approximation qu'on note $Q_n(x)$ défini par :

$$Q_n(x) = a_1 + a_2x + a_3x^2 + \ldots + a_{n+1}x^n.$$

tel que la différence

$$\varepsilon(x) = f(x) - Q_n(x)$$

ait une norme aussi petite que possible.

Soit $\mathbf{P}_n$ l'ensemble des polynômes de degré inférieur ou égal à n. Et soit f une fonction continue sur un intervalle $[a,b]$. La meilleure approximation de f par un polynôme Q_n de $\mathbf{P}_n$ s'écrit :

$$Q_n(x) = b_1u_n(x) + b_2u_{n-1}(x) + \ldots + b_{n+1}u_0(x) = \sum_{i=1}^{n+1} b_iu_{n+1-i}(x).$$

(Où les u_i forment une base de $\mathbf{P}_n$, c'est-à-dire : $u_i(x) = x^i \quad i = 0,1,\ldots,n$), et un polynôme quelconque P_n de $\mathbf{P}_n$ s'écrit :

$$P_n(x) = a_1u_n(x) + a_2u_{n-1}(x) + \ldots + a_{n+1}u_0(x) = \sum_{i=1}^{n+1} a_iu_{n+1-i}(x).$$

La condition nécessaire et suffisante du théorème s'écrit :

$$\langle f - \sum_{i=1}^{n+1} b_iu_{n+1-i}, \sum_{i=1}^{n+1} a_iu_{n+1-i}\rangle = 0$$

pour tout élément $(a_1,a_2,\ldots,a_n+1)$ de $\mathbb{R}^{n+1}$. Ce qui donne le système d'équations :

$$\sum_{i=1}^{n+1} b_i\langle u_{n+1-i}, u_{n+1-j}\rangle = \langle f, u_{n+1-j}\rangle \quad j = 1,2,\ldots,n+1 \tag{2.5}$$

qui admet une solution unique.

2.4.1 Norme

R Sur l'ensemble des fonctions continues $\mathscr{C}([a,b])$, on utilise le produit scalaire suivant : Soit ω une fonction positive n'ayant qu'un nombre fini de racines sur $[a,b]$ et telle que : $\int_a^b \omega(x)h(x)dx$ existe pour tout $h \in \mathscr{C}([a,b])$. On suppose ω continue par morceaux et on pose :

$$\langle f,g\rangle = \int_a^b \omega(x)f(x)g(x)dx.$$

R Le plus souvent dans le cas de l'approximation polynomiale on prend $\omega(x) = 1$.

La meilleure approximation de $f \in \mathscr{C}([a,b])$ par un polynôme de $\mathbf{P}_n$est donc le polynôme Q_n défini par :

$$\int_a^b \omega(x)[f(x)-Q_n(x)]^2dx = \min_{P_n\in\mathbf{P}_n}\int_a^b \omega(x)[f(x)-P_n(x)]^2dx.$$

Ce polynôme existe et vérifie :

$$\int_a^b \omega(x)[f(x)-Q_n(x)]P_n(x)dx = O \quad \text{pour tout} \quad P_n \in \mathbf{P}_n$$

Ses coefficients b_i sont donnés par le système d'équations linéaires :

$$\sum_{i=1}^{n+1} b_i \int_a^b \omega(x)x^{2n+2-i-j}dx = \int_a^b \omega(x)f(x)x^{n+1-j}dx \quad j = 1,\ldots,n+1 \tag{2.6}$$

■ **Exemple 2.1** Le polynôme Q_2 qui réalise la meilleure approximation au sens des moindres carrés de la fonction $f(x) = x^3 - x^2 - \frac{1}{4}x + \frac{1}{4}$ sur l'intervalle $[-1,1]$ avec $\omega(x) = 1$ est donné par la condition :

$$\sum_{i=1}^{3} b_i \int_{a=-1}^{b=1} 1.x^{2.2+2-i-j}dx = \int_{a=-1}^{b=1} 1.(x^3 - x^2 - \frac{1}{4}x + \frac{1}{4})x^{2+1-j}dx \quad j = 1,2,3$$

c'est-à-dire le système :

$$\left\{\begin{array}{rcccc} \frac{2}{5}b_1 + \frac{2}{3}b_3 & = & -\frac{2}{5} + \frac{1}{6} & = & -\frac{7}{30} \\ \frac{2}{3}b_2 & = & \frac{2}{5} - \frac{1}{6} & = & \frac{7}{30} \\ \frac{2}{3}b_1 + 2b_3 & = & -\frac{2}{3} + \frac{1}{2} & = & -\frac{1}{6} \end{array}\right.$$

qui donne comme solution :

$$b_1 = -1; b_2 = \frac{7}{20}; b_3 = \frac{1}{4}.$$

le polynôme cherché est donc :

$$Q_2(x) = -x^2 + \frac{7}{20}x + \frac{1}{4}.$$

L'erreur commise pour cette approximation s'évalue comme suit :

$$\varepsilon_2 = f(x) - Q_2(x) = (x^3 - x^2 - \frac{1}{4}x + \frac{1}{4}) - \left(-x^2 + \frac{7}{20}x + \frac{1}{4}\right) = x^3 - \frac{3}{5}x.$$

■

■ **Exemple 2.2** Le polynôme Q_2 qui réalise la meilleure approximation au sens des moindres carrés de la fonction $f(x) = |x|$ sur l'intervalle $[-1,1]$ avec $\omega(x) = 1$ est donné par la condition :

$$\sum_{i=1}^{3} b_i \int_{-1}^{1} x^{6-i-j} dx = \int_{-1}^{1} |x| x^{3-j} dx \quad j = 1,2,3$$

c'est-à-dire le système :

$$\begin{cases} \frac{2}{5}b_1 + \frac{2}{3}b_3 &= \frac{1}{2} \\ \frac{2}{3}b_2 &= 0 \\ \frac{2}{3}b_1 + 2b_3 &= 1 \end{cases}$$

qui donne comme solution :

$$b_1 = \frac{15}{16}; b_2 = 0; b_3 = \frac{3}{16}.$$

le polynôme cherché est donc :

$$Q_2(x) = \frac{15}{16}x^2 + \frac{3}{16}.$$

L'erreur commise pour cette approximation s'évalue comme suit :

$$\varepsilon_2 = f(x) - Q_2(x) = |x| - \left(\frac{15}{16}x^2 + \frac{3}{16}\right) = \begin{cases} \frac{15}{16}x^2 - x - \frac{3}{16} & \text{si} \quad x < 0 \\ -\frac{15}{16}x^2 + x - \frac{3}{16} & \text{si} \quad x > 0 \end{cases}$$

■

2.5 SERIE D'EXERCICES

Exercice 2.1 Trouver le polynôme P_2 qui réalise la meilleure approximation au sens des moindres carrés de la fonction $f(x) = |x-1|$ sur l'intervalle $[0,2]$ en prenant $\omega(x) = 1$. ■

Exercice 2.2 Trouver le polynôme P_2 qui réalise la meilleure approximation au sens des moindres carrés de la fonction $f(x) = |3x-5|$ sur l'intervalle $[-1,2]$ en prenant $\omega(x) = 1$. ■

Exercice 2.3 Trouver le polynôme P_2 qui réalise la meilleure approximation au sens des moindres carrés de la fonction donnée par le tableau suivant :

x	-1	$-0,5$	0	$0,5$	1
$f(x)$	$-0,75$	0	$0,25$	0	0

■

3. INTERPOLATION POLYNOMIALE

3.1 GÉNERALITÉS

Le problème de l'interpolation peut se poser dans plusieurs situations. Par exemple :

1- Une fonction f n'est pas complètement définie. On connait simplement un nombre fini de ses valeurs sur un intervalle $[a,b]$, (par exemple des données expérimentales) $x_1, x_2, \ldots, x_{n+1}$ supposés tels que :

$$a < x_1 \leq x_2 \leq \ldots \leq x_n \leq x_{n+1} < b$$

et les valeurs de f en ces points

$$f(x_1) = y_1, f(x_2) = y_2, \ldots, f(x_n) = y_n, f(x_{n+1}) = y_{n+1}$$

On peut alors vouloir déterminer une fonction φ, définie sur tout l'intervalle $[a,b]$ et constituant une certaine *"approximation de f"* (dans un sens à préciser). Il est alors souhaitable que $\varphi(x)$ soit numériquement *facile à évaluer*

2- $f(x)$ est connu pour tout x, mais son évaluation numérique est complexe. On peut alors vouloir déterminer une fois pour toutes un nombre fini de valeurs

$$f(x_1), f(x_2), \ldots, f(x_n), f(x_{n+1})$$

et pour tout

$$x \neq x_1, x_2, \ldots, x_n, x_{n+1}$$

approcher $f(x)$ en fonction de ces valeurs.

Cadre général de l'interpolation : *Chercher une fonction φ(appelé l'interpolant), d'un type préalablement choisi qui interpole f sur $[a,b]$, c'est déterminer cette fonction φ telle que*

$$\varphi(x_i) = f(x_i), \quad i = 1, 2, \ldots, n+1 \tag{3.1}$$

Cela signifie d'un point de vue géométrique, qu'il faut trouver une courbe d'équation $y = \varphi(x)$ et d'un type donné passant par le système de points

$$M_i = (x_i, y_i) \quad i = 1, 2, \ldots, n+1.$$

Le problème ainsi posé peut avoir une infinité de solutions ou ne pas avoir du tout de solution. Cependant, il admet une solution et une seule si l'on cherche non pas une fonction quelconque $\varphi(x)$ mais un polynôme de degré inférieur ou égal à n qui vérifie (3.1) et est tel que :

$$P_n(x_1) = y_1, \quad P_n(x_2) = y_2, \ldots\ldots, P_n(x_{n+1}) = y_{n+1}.$$

Nous obtenons la formule d'interpolation :

$$y = \varphi(x).$$

Que nous utiliserons donc pour le calcul approché des valeurs de la fonction donnée $f(x)$ pour des valeurs de x différentes des points d'interpolation x_i, $i = 1, \ldots n+1$.

R Si $x \notin [x_1; x_{n+1}]$ on parle d'une extrapolation.

Proposition 3.1.1 Si $x_1, x_2, \ldots, x_{n+1}$ sont $(n+1)$ points distincts de $\mathbb{R}$ et si $y_1, y_2, \ldots, y_{n+1}$ sont $(n+1)$ réels, alors il existe un et un seul polynôme réel P de degré inférieur ou égal à n tel que

$$P_n(x_1) = y_1, \quad P_n(x_2) = y_2, \ldots\ldots, P_n(x_{n+1}) = y_{n+1}.$$

Démonstration. Tout polynôme réel de degré inférieur ou égal à n s'écrit de manière unique sous la forme

$$P_n(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_n$$

l'ensemble $\mathbb{R}[X]$ de ces polynômes est donc un espace vectoriel réel de dimension $(n+1)$. L'application

$$\begin{array}{ccc} \mathbb{R}[X] & \to & \mathbb{R}^{n+1} \\ P & \mapsto & \begin{pmatrix} P(x_1) \\ \vdots \\ P(x_{n+1}) \end{pmatrix} \end{array}$$

est linéaire et injective car un polynôme de degré inférieur ou égal à n qui a $(n+1)$ zéros distincts est nul. Il s'ensuit que cette application est aussi surjective, d'où la conclusion. ■

R Pour trouver les coéfficients du polynôme $P_n(x)$ c'est-à -dire déterminer $a_0, a_1, \ldots, a_n$, il suffit de résoudre le système linéaire (de Cramer) suivant :

$$\begin{pmatrix} 1 & x_1 & x_1^n \\ \vdots & \vdots & \vdots \\ 1 & x_{n+1} & x_{n+1}^n \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_{n+1} \end{pmatrix}$$

La matrice de ce système est une matrice de Vandermonde. Cependant, cette méthode de détermination des coefficients est peu efficace. On préfère des méthodes basées sur des formules explicites pour le polynôme $P_n(x)$ telles que les formules dites de Lagrange et de Newton établies ci-dessous.

3.2 POLYNOME DE LAGRANGE

Soit $\{x_1, x_2, ...x_{n+1}\}$, $(n+1)$ valeurs distinctes de $[a,b]$ données. On suppose que que l'on connaisse les valeurs correspondantes de $y = f(x)$ c'est-à-dire on a :

$$f(x_1) = y_1, \;\; f(x_2) = y_2, ...f(x_{n+1}) = y_{n+1}.$$

On veut construire un polynôme $L_n(x) = \sum_{i=1}^{n+1} p_i(x) f(x_i)$, de degré inférieur ou égal à n, vérifiant :

$$L_n(x_i) = y_i, \;\; i = 1, 2, ...n+1.$$

où les fonctions $p_i(x)$ sont telles que :

$$p_i(x_j) = \delta_{ij} = \begin{cases} 1 & \text{si} \quad j = i \\ 0 & \text{si} \quad j \neq i \end{cases}$$

Où δ_{ij} est le symbole de Kronecker. **Résolution du problème :** Le polynôme à obtenir s'annulant en (n) points $x_1, x_2 ..., x_{i-1}, x_{i+1}, ..., x_{n+1}$, il s'ecrit donc :

$$p_i(x) = C_i(x - x_1)(x - x_2)...(x - x_{i-1})(x - x_{i+1})...(x - x_n), \tag{3.2}$$

où C_i est une constante. Posant dans (3.2) $x = x_i$ et comme $p_i(x_i) = 1$, on a donc :

$$C_i(x_i - x_1)(x_i - x_2)...(x_i - x_{i-1})(x_i - x_{i+1})...(x_i - x_n) = 1$$

c'est-à-dire :

$$C_i = \frac{1}{(x_i - x_1)(x_i - x_2)...(x_i - x_{i-1})(x_i - x_{i+1})...(x_i - x_n)}$$

en portant cette valeur dans (3.2) on obtient :

$$p_i(x) = \frac{(x - x_1)(x - x_2)...(x - x_{i-1})(x - x_{i+1})...(x - x_n)}{(x_i - x_1)(x_i - x_2)...(x_i - x_{i-1})(x_i - x_{i+1})...(x_i - x_n)} = \prod_{\substack{j=1 \\ j \neq i}}^{n+1} \frac{(x - x_j)}{(x_i - x_j)}. \tag{3.3}$$

Le polynôme $L_n(x)$ qui vérifie les conditions $L_n(x_i) = y_i$, est alors de la forme :

$$L_n(x) = \sum_{i=1}^{n+1} p_i(x) y_i \tag{3.4}$$

En effet, d'une part le polynôme $L_n(x)$ ainsi construit est un polyn ôme de degré inférieur ou égal à n. Et d'autre part comme $p_i(x_j) = 1$ si $j = i$ et 0 sinon, on a :

$$L_n(x_j) = \sum_{i=1}^{n+1} p_i(x_j) y_i = p_j(x_j) y_i = y_i \qquad j = 1, 2, ..., n+1$$

En portant la valeur de $p_i(x)$ dans (3.4) tirée de (3.3) on obtient :

$$\begin{aligned} L_n(x) &= \sum_{i=1}^{n+1} \frac{(x - x_1)(x - x_2)...(x - x_{i-1})(x - x_{i+1})...(x - x_n)}{(x_i - x_1)(x_i - x_2)...(x_i - x_{i-1})(x_i - x_{i+1})...(x_i - x_n)} y_i = \\ &= \sum_{i=1}^{n+1} \left(\prod_{\substack{j=1 \\ j \neq i}}^{n+1} \frac{(x - x_j)}{(x_i - x_j)} \right) y_i. \end{aligned} \tag{3.5}$$

Le polynôme donné par (3.5) est appelé le *polynôme d'interpolation de Lagrange*. et

$$p_i(x) = \left(\prod_{\substack{j=1 \\ j \neq i}}^{n+1} \frac{(x - x_j)}{(x_i - x_j)} \right) \tag{3.6}$$

sont appelés les *coeficients de Lagrange*.

Proposition 3.2.1 Le polynôme d'interpolation de Lagrange est unique.

Démonstration. Faisons un raisonnement par l'absurde. Soit $\check{L}_n(x)$ un polynôme de degré inférieur ou é gal à n, distinct de $L_n(x)$ et est tel que :

$$\check{L}_n(x) = y_i \qquad i = 1, 2, ..., n+1$$

Le polynôme

$$Q_n(x) = \check{L}_n(x) - L_n(x)$$

dont le degré est aussi inférieur ou égal à n, s'annule en $n+1$ points $\{x_1, x_2 ...., x_{n+1}\}$, c'est-à-dire :

$$Q_n(x) = 0$$

donc

$$\check{L}_n(x) = L_n(x)$$

■

■ **Exemple 3.1** Construire les coefficients de Lagrange pour $n+1 = 2$ et $n+1 = 3$. ■

- Cas $n+1 = 2$: En appliquant la formule (3.6) on obtient

$$p_1(x) = \frac{x - x_2}{x_1 - x_2}; \quad p_2(x) = \frac{x - x_1}{x_2 - x_1}$$

et alors

$$L(x) = \frac{x - x_2}{x_1 - x_2} f(x_1) + \frac{x - x_1}{x_2 - x_1} f(x_2)$$

- Cas $n+1 = 3$: En appliquant la formule (3.6) on obtient

$$p_1(x) = \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)}; \quad p_2(x) = \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)}; \quad p_3(x) = \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)};$$

et alors

$$L_2(x) = \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} f(x_1) + \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} f(x_2) + \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)} f(x_3).$$

■ **Exemple 3.2** Construire le polynôme de Lagrange de la fonction donnée par $y = f(x) = \sin \pi x$ pour les points $x_1 = 0$, $x_2 = \frac{1}{6}$, $x_3 = \frac{1}{2}$. ■

On calcule d'abord les valeurs correspondantes aux points d'interpolation de la fonction $y = f(x)$:

$$y_1 = 0, \quad y_2 = \sin\frac{\pi}{6} = \frac{1}{2}, \quad y_3 = \sin\frac{\pi}{2} = 1$$

On applique ensuite les formules (3.5), on obtient le polynô me de Lagrange de degré inférieur ou égal à 2 suivant :

$$L_2(x) = \frac{(x-\frac{1}{6})(x-\frac{1}{2})}{(0-\frac{1}{6})(0-\frac{1}{2})}.0 + \frac{(x)(x-\frac{1}{2})}{\frac{1}{6}(\frac{1}{6}-\frac{1}{2})}.\frac{1}{2} + \frac{(x)(x-\frac{1}{6})}{\frac{1}{2}(\frac{1}{2}-\frac{1}{6})}.1$$

Ou

$$L_2(x) = -3x^2 + \frac{7}{2}x$$

■ **Exemple 3.3** Soit la fonction $y = f(x)$ donnée par le tableau suivant :

x	y
321,0	2,50651
322,8	2,50893
324,2	2,51081
325,0	2,51188

On demande le calcul de la valeur de f en 323,5. ■

On pose $x = 323,5$; le polynôme de Lagrange sera un polynôme de degr é inférieur ou égal à $n = 3$. D'après les formules (3.5), on a :

$$\begin{aligned} f(323,5) &= \frac{(323,5-322,8)(323,5-324,2)(323,5-325,0)}{(321,0-322,8)(321,0-324,2)(321,0-325,0)}.2,50651 + \\ &+ \frac{(323,5-321,0)(323,5-324,2)(323,5-325,0)}{(322,8-321,0)(322,8-324,2)(322,8-325,0)}.2,50893 + \\ &+ \frac{(323,5-321,0)(323,5-322,8)(323,5-325,0)}{(324,2-321,0)(324,2-322,8)(324,2-325,0)}.2,51081 + \\ &\frac{(323,5-321,0)(323,5-322,8)(323,5-324,2)}{(325,0-321,0)(325,0-322,8)(325,0-324,2)}.2,51188 \\ &= -0,07996 + 1,18794 + 1,83897 - 0,43708 = 2,50987 \end{aligned}$$

3.2.1 Cas où les points sont equidistants

Soit $\{x_1, x_2 ...., x_{n+1}\}$, $(n+1)$ points d'interpolation de $[a,b]$, on suppose que :

$$x_2 = x_1 + h, \;\; x_3 = x_2 + h = x_1 + 2h, \;\; ...x_j = x_{j-1} + h = x_1 + (j-1)h, \;\; ...x_{n+1} = x_n + h = x_1 + nh$$

Où $h = \frac{b-a}{n}$ représente le pas de la subdivision. Les points d'interpolation x_i sont alors dits *équidistants*. Dans ce cas, les coefficients de Lagrange peuvent être simplifiés, en posant :

$$x = x_1 + th$$

on aura :

$$t_1 = 0, \;\; t_2 = 1, \;\; ...t_n = n.$$

D'où

$$l_n(t) = \sum_{i=1}^{n+1}\left(\prod_{\substack{j=1\\ j\neq i}}^{n+1}\frac{(x_1+th-x_1-(j-1)h)}{((x_1+(i-1)h)-x_1+(j-1)h)}\right)y_i \qquad (3.7)$$

$$= \sum_{i=1}^{n+1}\left(\prod_{\substack{j=1\\ j\neq i}}^{n+1}\frac{(t-j+1)}{(i-j)}\right)y_i$$

Les coefficients de $l_n(t)$ sont donc indépendants du pas h et des points x_i ; ils peuvent être représentés dans une table.

■ **Exemple 3.4** Soit la fonction $y = \cos x$ donnée par le tableau suivant :

x	5,0	5,1	5,2	5,3	5,4	5,5	5,6	5,7
y	0,283662	0,377977	0,468516	0,554374	0,634692	0,708669	0,775565	0,834712
t	1	2	3	4	5	6	7	8

Calculer $\cos 5,34$. ■

Posons

$$x = 0,1t + 5$$

Les valeurs de la nouvelle variable t associées aux points d'interpolation seront alors :

$$t = 0,1,2,3,4,5,6,7$$

Il faut donc trouver la valeur de y pour $x = 5,34$, c'est-à-dire pour $t = 3,4$, les points étant équidistants, on a donc :

$$l_n(3,4) = \sum_{i=1}^{8}\left(\prod_{\substack{j=1\\ j\neq i}}^{8}\frac{(3,47-j+1)}{(i-j)}\right)y_i$$

$$= 0,592864$$

Donc $\cos 5,34 = 0,592864$.

3.3 ESTIMATION DE L'ERREUR DANS L'INTERPOLATION DE LAGRANGE

Soit $L_n(x)$ le polynôme de Lagrange qui interpole la fonction $y = f(x)$ aux points $\{x_1, x_2 ...., x_{n+1}\}$ c'est-à -dire qui vérifie :

$$L_n(x_1) = y_1, \;\; L_n(x_2) = y_2, \;\; ..., L_n(x_{n+1}) = y_{n+1}.$$

La question que l'on se pose maintenant est : quelle est l'approximation du polynôme construit par rapport à la fonction $f(x)$? ou en d'autres termes quelle est la grandeur du reste :

$$R_n(x) = f(x) - L_n(x)$$

Pour évaluer cette erreur commise dans l'interpolation par le polynô me de Lagrange, nous supposons que la fonction f est continue, dé rivable et à dérivées continues jusqu'à l'ordre $(n+1)$ dans le domaine $[a,b]$ contenant x et les points d'interpolation x_i. Soit

$$g(x) = f(x) - L_n(x) - k(x-x_1)(x-x_2).....(x-x_{n+1})$$

Où k est une constante. La fonction g possède $(n+1)$ racines aux points

$$x_1, x_2, ...x_{n+1}$$

Choisissons k de sorte que $g(x)$ ait une $(n+2)^{ième}$ racine en un point quelconque fixé $\tilde{x}$ de $[a,b]$, autre que les points d'interpolation. Il suffit pour cela de poser

$$f(\tilde{x}) - L_n(\tilde{x}) - k(\tilde{x}-x_1)(\tilde{x}-x_2).....(\tilde{x}-x_n)$$

D'où

$$k = \frac{f(\tilde{x}) - L_n(\tilde{x})}{(\tilde{x}-x_1)(\tilde{x}-x_2).....(\tilde{x}-x_n)} \tag{3.8}$$

car $\tilde{x}$ n'est pas un point d'interpolation. Pour cette valeur de k, la fonction $g(x)$ admet $(n+2)$ racines sur $[a,b]$ et elle s'annule aux extrémités de chacun des intervalles suivants :

$$[x_1,x_2]\,,\quad [x_2,x_3]\,,\quad ...\quad [x_i,\tilde{x}]\,,\quad [\tilde{x},x_{i+1}]\,,\quad ...[x_n,x_{n+1}]\,,$$

En appliquant le théorème de Rolle à chacun de ces intervalles, on voit que la dérivée $g'(x)$ admet au moins $(n+1)$ racines sur $[a,b]$. De même, la dérivée seconde $g''(x)$ s'annule au moins n fois sur $[a,b]$. En opérant de même pour les dérivées successives de la fonction g, on aboutit à la conclusion que dans $[a,b]$, la dérivée $g^{(n+1)}(x)$ possède au moins une racine. Notons par ξ cette racine : on a donc $g^{(n+1)}(\xi) = 0$. Comme

$$g^{(n+1)}(x) = f^{(n+1)}(x) - k(n+1)!$$

Pour $x = \xi$ on obtient :

$$0 = f^{(n+1)}(\xi) - k(n+1)!$$

Donc

$$k = \frac{f^{(n+1)}(\xi)}{(n+1)!} \tag{3.9}$$

En identifiant les équations (3.8) et (3.9) on obtient :

$$\frac{f(\tilde{x}) - L_n(\tilde{x})}{(\tilde{x}-x_1)(\tilde{x}-x_2).....(\tilde{x}-x_n)} = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

C'est-à-dire

$$f(\tilde{x}) - L_n(\tilde{x}) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(\tilde{x}-x_1)(\tilde{x}-x_2).....(\tilde{x}-x_{n+1}) \tag{3.10}$$

Comme $\tilde{x}$ est complètement arbitraire, l'équation (3.10) s'écrit :

$$R_n(x) = f(x) - L_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_1)(x-x_2).....(x-x_{n+1}) \tag{3.11}$$

Où $\xi \in [a,b]$ dépend de x.

R L'équation (3.11) est vraie pour tout point de $[a,b]$, y compris les points d'interpolation. En posant

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

Nous obtenons l'estimation de l'erreur absolue dans l'interpolation par le polynôme de Lagrange sous la forme :

$$|R_n(x)| = |f(x) - L_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |(x-x_1)(x-x_2).....(x-x_{n+1})|$$

■ **Exemple 3.5** Avec quelle précision peut-on calculer $\sqrt{115}$ à l'aide d'une interpolation par le polynôme de Lagrange de la fonction $y = \sqrt{x}$ si l'on prend les points d'interpolation

$$x_1 = 100, \quad x_2 = 121, \quad x_3 = 144.$$

■

Comme on a

$$y' = \frac{1}{2}x^{-\frac{1}{2}}, \quad y'' = -\frac{1}{4}x^{-\frac{3}{2}}, \quad y''' = \frac{3}{8}x^{-\frac{5}{2}}.$$

Il s'ensuit

$$M_3 = \max \left| y''' \right| = \frac{3}{8}(100)^{-\frac{5}{2}} = \frac{3}{8}10^{-5} \quad \text{pour} \quad 100 \leq x \leq 144$$

Donc

$$\begin{aligned} |R_2(x)| &\leq \frac{3}{8}10^{-5}\frac{1}{3!}|(115-100)(115-121)(115-144)| = \\ &= \frac{1}{16}10^{-5}.15.6.29 \approx 1,6.10^{-3} \end{aligned}$$

3.4 POLYNOME DE NEWTON

3.4.1 Différences finies

Définition 3.4.1 Soit $y = f(x)$ une fonction donnée. On pose $\Delta x = x_{i+1} - x_i = h$ une valeur fixée de l'accroissement de x. On appelle *différence d'ordre un* de la fonction y l'expression :

$$\Delta y = \Delta f(x) = f(x+\Delta x) - f(x)$$

On définit de façon analogue les différences d'ordres supé rieurs

$$\Delta^n y = \Delta(\Delta^{n-1} y) \qquad (n = 2,3,.....)$$

■ **Exemple 3.6** $\Delta^2 y = \Delta[f(x+\Delta x) - f(x)] = [f(x+2\Delta x) - f(x+\Delta x)] - [f(x+\Delta x) - f(x)] = f(x+2\Delta x) - 2f(x+\Delta x) + f(x)$. ■

■ **Exemple 3.7** Construire les différences de $f(x) = x^3$, en prenant le pas $\Delta x = 1$. On a $\Delta f(x) = (x+1)^3 - x^3 = 3x^2 + 3x + 1$, $\Delta^2 f(x) = \left[3(x+1)^2 + 3(x+1) + 1\right] - (3x^2+3x+1) = 6x+6$, $\Delta^3 f(x) = [6(x+1)+6] - (6x+6) = 6$, $\Delta^n f(x) = 0$, pour $n > 3$. ■

Proposition 3.4.1 Si $f(x) = P_n(x) = a_0x^n + a_1x^{n-1} + ... + a_n$ est un polynôme de degré n, alors $\Delta^n f(x) = \Delta^n P_n(x) = n!a_0h^n = C$ où $\Delta x = h$ et C est une constante.

Démonstration. En effet, on a :

$$\begin{aligned}\Delta f(x) &= \Delta P_n(x) = P_n(x+h) - P_n(x) = \\ &= a_0\left[(x+h)^n - x^n\right] + a_1\left[(x+h)^{n-1} - x^{n-1}\right] + ... \\ &\quad ... + a_{n-1}\left[(x+h) - h\right]\end{aligned}$$

En utilisant la formule du binôme de Newton, on voit que $\Delta f(x) = \Delta P_n(x)$ est un polynôme de degré $(n-1)$:

$$\Delta f(x) = \Delta P_n(x) = b_0 x^{n-1} + b_1 x^{n-2} + \cdots + b_{n-1}$$

Où

$$b_0 = nha_0$$

Suivant le même raisonnement la différence seconde $\Delta^2 f(x) = \Delta^2 P_n(x)$ est un polynôme de degré $(n-2)$:

$$\Delta^2 f(x) = \Delta^2 P_n(x) = c_0 x^{n-2} + c_1 x^{n-3} + \cdots + c_{n-2}$$

et

$$c_0 = (n-1)hb_0 = n(n-1)h^2 a_0$$

En raisonnant ainsi on établit de proche en proche que

$$\Delta^n f(x) = \Delta^n P_n(x) = n!h^n a_0$$

D'où l'on conclut

$$\Delta^m f(x) = \Delta^m P_n(x) = 0 \quad \text{pour} \quad m > n$$

■

R Le symbole Δ (delta) peut être considéré comme un *opérateur* qui associe à la fonction $y = f(x)$ la fonction $\Delta y = f(x+\Delta x) - f(x)$ (Δx étant une constante).

R L'opérateur Δ a les propriétés suivantes :

1. $\Delta(u+v) = \Delta u + \Delta v$;
2. $\Delta(Ku) = K\Delta u$ (K est une constante) ;
3. $\Delta^m(\Delta^n u) = \Delta^{m+n} u$ (m et n entiers non négatifs) ;
4. On pose par définition $\Delta^0 u = u$.

3.4.2 Différences divisées

Définition 3.4.2 Soit $\{x_1, x_2 ...., x_{n+1}\}$, $(n+1)$ points d'interpolation de $[a,b]$, et :

$$f(x_1) = y_1, \quad f(x_2) = y_2, \quad ..., f(x_{n+1}) = y_{n+1}.$$

on pose :

$$\Delta x_i = x_{i+1} - x_i \neq 0 \qquad (i = 1, 2, ...)$$

On appelle *différences divisées d'ordre* 1, les relations données par :

$$f[x_i, x_{i+1}] = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}, \quad (i = 1, 2, ...)$$

■ **Exemple 3.8** $f[x_1, x_2] = \frac{y_2 - y_1}{x_2 - x_1}$, $f[x_2, x_3] = \frac{y_3 - y_2}{x_3 - x_2}$, etc... ■

D'une façon analogue on définit les *différences divisées d'ordre 2*

$$f[x_i, x_{i+1}, x_{i+2}] = \frac{f[x_{i+1}, x_{i+2}] - f[x_i, x_{i+1}]}{x_{i+2} - x_i}, \quad (i = 1, 2, ...)$$

■ **Exemple 3.9** $f[x_1, x_2, x_3] = \frac{f[x_2, x_3] - f[x_1, x_2]}{f[x_3, x_1]}$. ■

D'une façon générale, les *différences divisées d'ordre n* s'obtiennent à partir des différences divisées d'ordre $(n-1)$ à l'aide de la relation de récurrence suivante :

$$f[x_i, x_{i+1}, ...x_{i+n}] = \frac{f[x_{i+1}, ...x_{i+n}] - f[x_i, x_{i+n-1}]}{x_{i+n} - x_i}, \quad (n = 1, 2, ...; i = 1, 2, ...)$$

R Les différences divisées sont symétriques de leurs arguments, c'est-à-dire que les différences divisées ne changent pas avec la permutation des éléments. Plus précisément, pour toute permutation σ de $\{1, 2...., k\}$ on a

$$f[x_{\sigma(1)}, x_{\sigma(2)}, ...x_{\sigma(k)}] = f[x_1, x_2, ...x_k]$$

■ **Exemple 3.10** $f[x_1, x_2] = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y_1 - y_2}{x_1 - x_2} = f[x_2, x_1]$ ■

Théoreme 3.4.2 Soit p_n le polynôme d'interpolation de f aux points $\{x_1, x_2...., x_{n+1}\}$, alors pour tout k de $\{1, 2...., n\}$ on a :

$$f[x_1, x_2, ...x_k] = \frac{f[x_2, ...x_{k+1}] - f[x_1, x_k]}{x_{k+1} - x_1},$$

Démonstration. L'idée est de construire une fonction polynôme qui interpole f sur $\{x_1, x_2...., x_{k+1}\}$; puis on identifie les deux expressions disponibles de son coefficient dominant. Soit p défini par

$$p(x) = \frac{x - x_1}{x_{k+1} - x_1} q_k(x) + \frac{x_{k+1} - x}{x_{k+1} - x_1} p_k(x)$$

Où q_k (respectivement p_k) désigne le polynôme d'interpolation de f sur $\{x_2, x_3...., x_{k+1}\}$ (respectivement $\{x_1, x_2...., x_k\}$). Par définition, p est un polynôme de degré inférieur ou égal à k. Montrons que p interpole f sur $\{x_1, x_2...., x_k\}$. On evalue d'abord $p(x_1)$: la contribution du premier terme de la somme est nulle ; il vient $p(x_1) = p_k(x_1)$ soit $p(x_1) = f(x_1)$ par d éfinition de p_k. On évalue alors $p(x_k)$ par le même type de raisonnement, on montre que $p(x_k) = f(x_k)$. Puis on considè re x_j pour tout j dans $\{2, 3...., k\}$. On sait que $p_k(x_j) = q_k(x_j) = f(x_j)$ grâce aux définitions de p_k et de q_k. Un calcul simple montre que : $p(x_j) = f(x_j)$; en effet

$$p(x_j) = \frac{1}{x_{k+1} - x_1} [(x_j - x_1) f(x_j) + (x_{k+1} - x_j) f(x_j)] \tag{3.12}$$

Soit

$$p(x_j) = f(x_j)$$

p est le polynôme d'interpolation de f sur $\{x_1, x_2 ...., x_{k+1}\}$. Soit α le coefficient dominant de p; d'après ce qui préc ède, $\alpha = f[x_1, ..., x_{k+1}]$. Sur l'expression (3.12), on voit que le coefficient dominant de p est donné par

$$\alpha = \frac{1}{x_{k+1} - x_1} [\text{coef dominant } (q_k) - \text{coef dominant } (p_k)]$$

comme (q_k) et (p_k) sont des polynômes d'interpolation sur $\{x_2, x_3 ...., x_{k+1}\}$ et $\{x_1, x_2 ...., x_k\}$ respectivement, l'égalité cherché e en découle. ■

R Les différences divisées forment généralement un tableau du type suivant :

x	$f(x)$	Différences divisées			
		Ordre 1	Ordre 2	Ordre 3	Ordre 4
x_1	$f[x_1]$	$f[x_1,x_2]$	$f[x_1,x_2,x_3]$	$f[x_1,x_2,x_3,x_4]$	$f[x_1,x_2,x_3,x_4,x_5]$
x_2	$f[x_2]$	$f[x_2,x_3]$	$f[x_2,x_3,x_4]$	$f[x_2,x_3,x_4,x_5]$	
x_3	$f[x_3]$	$f[x_3,x_4]$	$f[x_3,x_4,x_5]$		
x_4	$f[x_4]$	$f[x_4,x_5]$			
x_5	$f[x_5]$				

■ **Exemple 3.11** Donner les différences divisées de la fonction donnée par le tableau suivant :

x	0	0,2	0,3	0,4	0,7	0,9
y	132,651	148,877	157,464	166,375	195,112	216,000

■

Les résultats sont portés sur le tableau suivant :

x	$f(x)$	Différences divisées			
		Ordre 1	Ordre 2	Ordre 3	Ordre 4
0	132,651	81,13			
0,2	148,877	85,87	15,8		
0,3	157,464	89,11	16,2	1	0
0,4	166,375	95,79	16,7	1	0
0,7	195,112	104,44	17,3	1	
0,9	216,000				

3.4.3 Polynôme d'interpolation de Newton :

On a par définition

$$f[x, x_1] = \frac{y_1 - y}{x_1 - x} = \frac{f(x) - f(x_1)}{x - x_1},$$

donc

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

et

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

donc

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

en réitérant le procédé, on obtient :

$$\begin{aligned} f(x) \quad = \quad & f(x_1) + (x - x_1) f[x_1, x_2] + (x - x_1)(x - x_2) f[x_1, x_2, x_3] + \cdots + \\ & + (x - x_1)(x - x_2) \cdots\cdots (x - x_n) f[x_1, x_2, ..., x_{n+1}] \\ & + (x - x_1)(x - x_2) \cdots\cdots (x - x_n)(x - x_{n+1}) f[x, x_1, x_2, ..., x_{n+1}] \end{aligned} \tag{3.13}$$

R Si f est un polynôme de degré inférieur ou égal à n on a :

$$f[x, x_1, x_2, ..., x_{n+1}] = 0$$

R Le polynôme défini par

$$\begin{aligned} P_n(x) \quad = \quad & f(x_1) + (x - x_1) f[x_1, x_2] + (x - x_1)(x - x_2) f[x_1, x_2, x_3] + \cdots + \\ & + (x - x_1)(x - x_2) \cdots\cdots (x - x_n) f[x_1, x_2, ..., x_{n+1}] \end{aligned}$$

est le polynôme d'interpolation de degré inférieur ou égal à n de f pour les points $\{x_1, x_2 ...., x_{n+1}\}$ appelé *polynôme d'interpolation de Newton*. C'est-à-dire qu'il vérifie :

$$P_n(x_i) = f(x_i) \quad i = 1, 2, ...n + 1.$$

3.4.4 Erreur d'interpolation

En comparant $f(x)$ et $P_n(x)$, nous obtenons la formule générale de l'erreur :

$$\varepsilon(x) = f(x) - P_n(x) = (x - x_1)(x - x_2) \cdots\cdots (x - x_{n+1}) f[x, x_1, x_2, ..., x_{n+1}]$$

que l'on peut ecrire sous la forme :

$$\varepsilon(x) = \left(\prod_{i=1}^{n+1} (x - x_i) \right) f[x, x_1, x_2, ..., x_{n+1}]$$

3.4.5 Autre écriture du polynôme d'interpolation de Newton

Cas des points équidistants

En posant le pas du tableau $h = \Delta x$ ($\Delta x_i = x_{i+1} - x_i$, $i = 1, 2, ..., n + 1$) et $y_1 = f(x_1), y_2 = f(x_2), \ldots, y_n = f(x_n), y_{n+1} = f(x_{n+1})$; alors nous obtenons :

$$f(x) = y_1 + k\Delta y_1 + \frac{k(k-1)}{2!}\Delta^2 y_1 + \cdots + \frac{k(k-1)...(k-n+1)}{n!}\Delta^n y_1 + R_n(x)$$

Où

$$k = \frac{x - x_1}{h} \quad \text{et} \quad \Delta y_1 = y_2 - y_1; \; \Delta^2 y_1 = \Delta y_2 - \Delta y_1,$$

et

$$R_n(x) = \frac{h^{n+1} k(k-1)...(k-n)}{(n+1)!} f^{(n+1)}(\xi_x) \simeq \frac{k(k-1)...(k-n)}{(n+1)!}\Delta^{n+1} y_1; \qquad \xi_x \in [x, x_i]$$

■ **Exemple 3.12** Sachant que $\sin 26^\circ = 0,43837$; $\sin 27^\circ = 0,45399$ et $\sin 28^\circ = 0,46947$, calculer $\sin 26^\circ 15'$. ■

Le tableau des valeurs se présente comme suit :

i	x_i	y_i	Δy_i	$\Delta^2 y_i$
1	26	0,43837		
			0,01562	
2	27	0,45399		−0,00014
			0,01548	
3	28	0,46947		

comme $h = 60'$ et $k = \frac{1575'-1560'}{60} = \frac{1}{4}$ donc

$$\sin 26^\circ 15' = 0,43837 + \frac{1}{4}.0,01562 + \frac{\frac{1}{4}(\frac{1}{4}-1)}{2}(-0,00014) = 0,44229.$$

L'erreur

$$|R_2(x)| \leq \frac{\frac{1}{4}(\frac{1}{4}-1)(\frac{1}{4}-2)}{3!}(\frac{\pi}{180}) \simeq 0,2510^{-6}$$

car comme $y = \sin x$ alors $\left|y^{(n)}\right| \leq 1$.

R Si l'on fixe n et que l'on pose

$$h = \max_{h\in\{1,\ldots,n\}} |x_{k+1} - x_k|$$

alors pour tout $x \in [x_1, x_2]$, on a

$$|(x-x_1)\cdots(x_1-x_{n+1})| \leq h(h)(2h)\cdots(nh) \leq h^{n+1}n!.$$

Si f est de classe C^{n+1} sur $[a,b]$ et si on a $x_1 = a$ et $h < \frac{b-a}{n}$, alors

$$\sup_{x\in[x_1,x_2]} |f(x) - P_n(x)| \leq \frac{\sup_{x\in[a,b]} \left|f^{(n+1)}(x)\right|}{n+1} h^{n+1}.$$

Comme le second membre tend vers zéro avec h, on peut donc rendre

$$\sup_{x\in[x_1,x_2]} |f(x) - P_n(x)|$$

aussi petit que l'on veut (avec h suffisamment petit). Ceci montre que l'interpolation polynomiale peut être utilisée pour approcher les valeurs de $f(x)$.

R La remarque précédente n'est en général pas vraie, même si f est très régulière sur $[a,b]$. Dans le cas o ù $x_{k+1} - x_k$ ne dépend pas de k, c'est là ce qu'on appelle le phénomène de Runge. Par exmple si

$$f(x) = \frac{1}{1+25x^2}$$

et si $[a,b] = [-1,1]$, on peut montrer que le polyn ôme $P_n(x)$ interpolant $f(x)$ aux points

$$x_k = -1 + \frac{2k}{n}, \quad (k = 1,\ldots,n+1),$$

ne tend pas vers f lorsque $n \to \infty$.

3.5 SERIE D'EXERCICES

Exercice 3.1 On considère la fonction $f(x)$ donnée par le tableau suivant :

x	1	4	6
$f(x)$	1.5709	1.5727	1.5751

Trouver une approximation de $f(3.5)$ en utilisant le polynôme d'interpolation de Lagrange du second degré . ■

Exercice 3.2 Ecrire les différences divisées de $f(x) = x^2$ et de $g(x) = x^3$. ■

Exercice 3.3 1. Construire le polynôme d'interpolation de Newton de la fonction $y = f(x)$ donnée par le tableau suivant :

x	0	2.5069	5.0154	7.52270
$f(x)$	0.3989423	0.3988169	0.3984408	0.3978138

Trouver à l'aide de ce polynôme $f(3.7608)$. ■

Exercice 3.4 Soit le tableau des valeurs $y = \log x$, trouver $\log 1005$.

x	1000	1010	1020	1030	1040	1050
y	3.0000000	3.0043214	3.0086002	3.0128372	3.0170333	3.0211893

■

Exercice 3.5 Trouver $\sin 14°$ à partir du tableau des valeurs données ci-dessous de la fonction $y = \sin x$, le pas étant $h = 5°$.

x	15°	20°	25°	30°	35°	40°	45°	50°	55°
y	0.2588	0.3420	0.4226	0.5000	0.5736	0.6428	0.7071	0.7660	0.8192

■

Exercice 3.6 1. En interpolant par un polynôme de degré 3 et en utilisant la formule appropriée calculer pour $x = 1.05$, à l'aide de la table donnée ci-dessous, les valeurs de la fonction $y = \sin x$.

x	1.0	1.1	1.2	1.3
y	0.841471	0.891207	0.932039	0.963558

2. Donner une majoration de l'erreur.

■

4. INTEGRATION ET DÉRIVATION NUMÉRIQUES

4.1 INTÉGRATION NUMÉRIQUE

4.1.1 Méthode Générale

Lorsque l'intégrale définie d'une fonction continue sur un intervalle $[a,b]$ ne peut pas être évaluée analytiquement ou lorsque l'intégrale n'est pas donnée sous forme analytique mais numériquement en un certain nombre de valeurs discrètes, l'intégration numérique peut être utilisée.

Il existe plusieurs méthodes permettant d'évaluer les intégrales de fonctions bornées sur un intervalle $[a,b]$. La présence de singularité dans les fonctions (ou dans certaines fonctions) rend les calculs parfois difficiles.

Le problème de l'intégration numérique d'une fonction consiste à chercher la valeur de l'intégrale définie à partir de plusieurs valeurs de la fonction sous le signe somme. L'intégrale à évaluer étant :

$$I = \int_a^b f(x)dx. \tag{4.1}$$

L'intégration numérique consiste à remplacer l'intégrale (4.1) par une somme discrète sur un nombre fini de points :

$$I_N = \sum_{i=1}^{N} A_i f(x_i)$$

où A_i et x_i sont des variables à préciser.

Pour que l'évaluation numérique soit correcte, il est nécessaire d'imposer que toute méthode d'intégration vérifie :

$$\lim_{N\to\infty} I_N = I \tag{4.2}$$

Au delà de la vérification de ce critère, la qualité d'une méthode sera évaluée par la manière dont la convergence vers le résultat exact s'effectue.

Nous supposerons dorénavant que les fonctions sont de classe C^1(continues et à dérivées continues) sur $[a,b]$, mais aussi pour toute fonction f :

$$f'(x) < K \quad \forall x \in [a,b],$$

où K est une constante finie.

Ce qui veut dire que la dérivée de f n'est pas singulière.sur $[a,b]$.

On se propose alors de chercher une approximation de I. On remplace f sur $[a,b]$ par une fonction d'interpolation φ (un polynôme par exemple) pour considérer :

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

4.1.2 Approximation d'une intégrale

Soit ω une fonction positive définie sur $[c,d]$, on veut approcher

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

on suppose que f est connue en $(n+1)$ points $x_1, x_2, ..., x_{n+1}$. On écrit

$$I = \sum_{i=1}^{n+1} \alpha_i f(x_i) + R$$

où α_i seront choisi de telle sorte que R soit nul lorsque f est d'un type déterminé.

Lorsque f est quelconque, on suppose que

$$A = \sum_{i=1}^{n+1} \alpha_i f(x_i)$$

est une valeur approchée de I et R est suffisamment petit.

Soient $v_1, v_2, ..., v_{n+1}$ $(n+1)$ fonctions linéairement indépendantes continues sur $[a,b]$. On note $\mathscr{F}_n$ le sous espace engendré par ces fonctions. Pour que R soit nul, si $f \in \mathscr{F}_n$ on obtient :

$$I_k = \int_a^b \omega(x)v_k(x)dx = \sum_{i=1}^{n+1} \alpha_i v_k(x_i); \quad k = 1, ..., n+1.$$

On suppose que les fonctions [1] $v_1, v_2, ..., v_{n+1}$ sont telles que le système admette une soultion unique.

Soit A_n la fonction d'interpolation de f dans $\mathscr{F}_n$ vérifiant :

$$A_n(x) = \sum_{k=1}^{n+1} a_k v_k(x)$$

et

$$A_n(x_i) = f(x_i) \quad \text{pour} \quad i = 1, ..., n+1.$$

nous avons

$$\int_a^b \omega(x)A_n(x)dx = \sum_{i=1}^{n+1} \alpha_i A_n(x_i)$$

1. vérifient les conditions de Haar.

si

$$\varepsilon(x) = f(x) - A_n(x)$$

nous obtenons :

$$\int_c^d \omega(x) f(x) dx = \sum_{i=1}^{n+1} \alpha_i f(x_i) + \int_c^d \omega(x) \varepsilon(x) dx$$

Si la fonction d'interpolation $A_n(x)$ est le polynôme d'interpolation de f, c'est-à-dire que :

$$\{v_1(x), v_2(x), ..., v_{n+1}(x)\} = \{1, x, ..., x^n\}$$

On choisira $\{\alpha_1, \alpha_2, ..., \alpha_{n+1}\}$ de telle sorte que R soit nul quand f est un polynôme quelconque de degré inférieur ou égal à n.

4.1.3 Utilisation de l'interpolation polynomiale

Soit

$$A_n(x) = P_n(x) = \sum_{i=1}^{n+1} L_i(x) f(x_i)$$

avec

$$L_i(x) = \prod_{\substack{j=1 \\ j \neq i}}^{n+1} \frac{x - x_j}{x_i - x_j}$$

d'où

$$P = \int_c^d \omega(x) P_n(x) dx = \sum_{i=1}^{n+1} \left(\int_c^d \omega(x) L_i(x) dx \right) f(x_i)$$

R Sous certaines conditions sur ω, $\{\alpha_i\}$ et $\{x_i\}$ P est une approximation de I.

R Les α_i sont indépendants de f, ils peuvent être calculés une bonne fois pour toute.

4.1.4 Etude de l'erreur d'intégration

Si f est $(n+1)$ fois continument dérivable sur $[a,b]$, on sait qu'il existe $\xi_x \in [a,b]$ tel que :

$$\varepsilon(x) = f(x) - A_n(x) = L(x) \frac{f^{(n+1)}(\xi_x)}{(n+1)!}$$

avec

$$L(x) = \prod_{i=1}^{n+1} (x - x_i)$$

en intégrant on obtient :

$$R = I - P = \int_c^d \omega(x) L(x) \frac{f^{(n+1)}(\xi_x)}{(n+1)!} dx$$

si $\omega(x)L(x)$ est de signe constant dans $[c,d]$ (en particulier si $[c,d]$ ne contient aucun des points d'interpolation, avec ω de signe constant sur $[c,d]$) le théorème de la moyenne donne :

$$R = \frac{f^{(n+1)}(\eta)}{(n+1)!}\int_c^d \omega(x)L(x)dx, \qquad \eta \in [c,d]$$

si on connait une borne supérieure de $f^{(n+1)}$

$$M_{n+1} = \sup_{x\in[a,b]} \left|f^{(n+1)}(x)\right|$$

on a

$$|R| \leq \frac{M_{n+1}}{(n+1)!}\int_c^d |\omega(x)L(x)|\,dx$$

4.1.5 Convergence des méthodes d'intégration

Soit f une fonction continue sur $[a,b]$ et ω une fonction (poids) définie sur $[a,b]$ telle que

$$\int_a^b |\omega(x)|\,dx \leq M$$

on approche

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

par

$$A = \sum_{i=1}^{n+1} \alpha_i f(x_i)$$

où α_i et x_i sont à déterminer.

Problem 4.1 Savoir si en augmentant le nombre de points x_i, on obtiendrait une valeur de A de plus en plus proche de I. Plus précisemment a-t-on :

$$\lim_{n\to\infty} \sum_{i=1}^{n+1} \alpha_i f(x_i) = \int_a^b \omega(x)f(x)dx?$$

La réponse n'est positive que sous certaines conditions sur $\{\alpha_i\}$ et $\{x_i\}$.

Théoreme 4.1.1 Lorsque f est une fonction continue sur $[a,b]$ nous avons

$$\lim_{n\to\infty} \sum_{i=1}^{n+1} \alpha_i f(x_i) = \int_a^b \omega(x)f(x)dx \tag{4.3}$$

si les conditions suivantes sont vérifiées :

1. l'équation (4.3) est vrai pour un polynôme P quelconque
2. $\sum_{i=1}^{n+1} |\alpha_i|$ est borné pour tout n.

Démonstration. Soit P un polynôme quelconque. Posons

$$\varepsilon_f(x) = \int_a^b \omega(x)f(x)dx - \sum_{i=1}^{n+1} \alpha_i f(x_i)$$

et

$$\varepsilon_P(x) = \int_a^b \omega(x)P(x)dx - \sum_{i=1}^{n+1} \alpha_i P(x_i)$$

on a donc

$$\varepsilon_f(x) = \int_a^b \omega(x)P(x)dx + \int_c^d \omega(x)(f(x)-P(x))dx - \sum_{i=1}^{n+1} \alpha_i P(x_i) + \sum_{i=1}^{n+1} \alpha_i (f(x_i)-P(x_i))$$

ou

$$\varepsilon_f(x) = \varepsilon_P(x) + \int_a^b \omega(x)(f(x)-P(x))dx - \sum_{i=1}^{n+1} \alpha_i (f(x_i)-P(x_i))$$

soit $\varepsilon > 0$ un nombre déstiné à tendre vers 0. D'après le théorème de Weirstrass [2] nous pouvons prendre un polynôme P tel que

$$\max_{x\in[a,b]} |f(x)-P(x)| \leq \varepsilon$$

nous avons alors

$$\left|\varepsilon_f(x)\right| \leq |\varepsilon_P(x)| + \varepsilon \int_a^b |\omega(x)|\,dx - \varepsilon \sum_{i=1}^{n+1} |\alpha_i|$$

c'est-à-dire

$$\left|\varepsilon_f(x)\right| \leq |\varepsilon_P(x)| + \varepsilon(M - \sum_{i=1}^{n+1} |\alpha_i|)$$

si $\lim_{n\to\infty} |\varepsilon_P(x)| = 0$ pour tout polynôme P et si $\sum_{i=1}^{n+1} |\alpha_i| \leq N$ pour tout n on a

$$\lim_{n\to\infty} \left|\varepsilon_f(x)\right| \leq \varepsilon(M+N)$$

■

4.1.6 Formules de Newton Cotes

On suppose que f est connue en $(n+1)$ points $x_1, x_2, ..., x_{n+1}$ équidistants, et tels que :

$$\begin{aligned} x_1 &= a,\ x_2 = x_1 + h, \\ &\cdots\cdots\cdots\cdots\cdots \\ x_i &= x_{i-1} + h = a + (i-1)h \\ &\cdots\cdots\cdots\cdots\cdots \\ x_{n+1} &= x_n + h = a + nh \end{aligned}$$

on prend $\omega(x) = 1, \forall x \in [a,b]$.

On pose

$$\int_{x_{1-k}}^{x_{n+1+k}} f(x)dx = \sum_{i=1}^{n+1} \alpha_i f(x_i) + R \tag{4.4}$$

2. Si f est continue sur [a,b], il existe un polynôme de degré inférieur ou égal à n qui approche f.

— Si $k=0$, on dit que la formule (4.4) est de type fermé.
— Si $k=1$, on dit que la formule (4.4) est de type ouvert.

Les coefficients α_i sont donnés par

$$\alpha_i = \int_{x_{1-k}}^{x_{n+1+k}} L_i(x)dx$$

comme les x_i sont équidistants on peut poser $x = a+th$ et on a :

$$\alpha_i = \int_{x_{1-k}}^{x_{n+1+k}} l_i(t)dt$$

avec

$$l_i(t) = \prod_{\substack{j=1 \\ j\neq i}}^{n+1} \frac{t-j+1}{i-j}$$

R

$$\alpha_1 = \alpha_{n+1;}\alpha_2 = \alpha_{n;}\alpha_3 = \alpha_{n-1;}.....\text{et} \quad \sum_{i=1}^{n+1} \alpha_i = b-a+2kn \tag{4.5}$$

R Les méthodes de Newton Cotes sont convergentes pour les polynômes mais ne le sont pas pour une fonction continue quelconque.
En effet si d'après l'équation (4.5) $\sum_{i=1}^{n+1} \alpha_i$ est bien bornée, il n'en est pas de même de $\sum_{i=1}^{n+1} |\alpha_i|$ car les α_i ne sont pas toujours de même.

En intégrant les formules d'interpolation on a :

4.1.7 Formule de type fermé : des trapèzes et de Simpson

1. $\int_{x_1}^{x_2} f(x)dx = \frac{h}{2}(f(x_1)+f(x_2)) - \frac{h^3}{12}f''(\xi), \quad \xi \in [x_1,x_2]$ (*formule des trapèzes*)
2. $\int_{x_1}^{x_3} f(x)dx = \frac{h}{3}(f(x_1)+4f(x_2)+f(x_3)) - \frac{h^5}{90}f''(\xi), \quad \xi \in [x_1,x_3]$ (*formule de Simpson*)
3. $\int_{x_1}^{x_4} f(x)dx = \frac{3h}{8}(f(x_1)+3f(x_2)+3f(x_3)+f(x_4)) - \frac{3h^5}{80}f''(\xi), \quad \xi \in [x_1,x_4]$ (*formule de Newton*)

4.1.8 Formule de type ouvert :

1. $\int_{x_1}^{x_3} f(x)dx = 2hf(x_2) + \frac{h^3}{3}f''(\xi), \quad \xi \in [x_1,x_3]$ (*formule de Poncelet*)
2. $\int_{x_1}^{x_4} f(x)dx = \frac{3h}{2}(f(x_2)+f(x_3)) + \frac{3h^3}{4}f''(\xi), \quad \xi \in [x_1,x_4]$
3. $\int_{x_1}^{x_5} f(x)dx = \frac{4h}{3}(2f(x_2)-f(x_3)+2f(x_4)) + \frac{14h^5}{45}f^{(4)}(\xi), \quad \xi \in [x_1,x_5]$
4. $\int_{x_1}^{x_6} f(x)dx = \frac{5h}{24}(11f(x_2)+f(x_3)+f(x_4)+11f(x_5)) + \frac{95h^5}{144}f^{(4)}(\xi), \quad \xi \in [x_1,x_6]$

R Les formules d'intégration données entre x_1 et x_2, x_3 ....peuvent être modifiées pour d'autres points d'interpolation.Exemple : $\int_{x_2}^{x_3} f(x)dx = \frac{h}{2}(f(x_2)+f(x_3)) - \frac{h^3}{12}f''(\xi), \quad \xi \in [x_2,x_3]$
$\int_{x_2}^{x_4} f(x)dx = 2hf(x_3) + \frac{h^3}{3}f''(\xi), \quad \xi \in [x_2,x_4]$

4.1.9 Intégration par la méthode de Gauss

Polynôme de Legendre

Définition 4.1.1 On appelle polynômes de Legendre, les polynômes qui s'écrivent sous la forme :

$$P_n(x) = \frac{1}{2^n n!} \cdot \frac{\partial^n}{\partial x^n} \left[(x^2-1)^n \right], \quad (n = 0, 1, ...)$$

Propriétés : Les polynômes de Legendre vérifient les propriétés suivantes :

1. $P_n(1) = 1$ et $P_n(-1) = (-1)^n \quad (n = 0, 1, 2, ...)$.
2. $\int_{-1}^{1} P_n(x) Q_k(x) dx = 0, \quad (k < n)$, où $Q_k(x)$ est un polynôme de degré $k < n$.
3. Le polynôme de Legendre $P_n(x)$ possède n racines réelles distinctes dans $[-1, 1]$.

■ **Exemple 4.1**

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2 - 1) \\ P_3(x) &= \frac{1}{2}(5x^3 - 3x) \\ P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3) \end{aligned}$$

■

Formule de quadrature de Gauss

Soit $y = f(t)$ une fonction définie sur $[-1, 1]$.

Problem 4.2 Comment choisir $t_1, t_2, ..., t_n$ et $A_1, A_2, ...A_n$ pour que la formule de quadrature

$$\int_{-1}^{1} f(t) dt = \sum_{i=1}^{n} A_i f(t_i) \tag{4.6}$$

soit exacte pour tout polynôme $f(t)$ de degré N le plus grand.

Solution : Comme on a $2n$ constantes t_i et A_i $(i = 1, 2, ..., n)$ alors que le polynôme de degré $2n-1$ est défini par $2n$ coefficients, ce degré maximal dans le cas général est $N = 2n-1$.

Posant $\int_{-1}^{1} t^k dt = \sum_{i=1}^{n} A_i t_i^k \quad (k = 0, 1, ..., 2n-1)$ et $f(t) = \sum_{k=0}^{2n-1} C_k t^k$ alors

$$\int_{-1}^{1} f(t) dt = \sum_{k=0}^{2n-1} C_k \int_{-1}^{1} t^k dt = \sum_{k=0}^{2n-1} C_k \sum_{i=1}^{n} A_i t_i^k = \sum_{i=1}^{n} A_i f(t_i)$$

comme

$$\int_{-1}^{1} t^k dt = \frac{1 - (-1)^{k+1}}{k+1} = \begin{cases} \frac{2}{k+1} & \text{si } k \text{ est pair} \\ 0 & \text{si } k \text{ est impair} \end{cases}$$

pour résoudre le problème il suffit de déterminer t_i et A_i à partir du systéme non linéaire de $2n$ équations

$$\begin{cases} \sum_{i=1}^{n} A_i & = & 2 \\ \sum_{i=1}^{n} A_i t_i & = & 0 \\ \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} A_i t_i^{2n-2} & = & \frac{2}{2n-1} \\ \sum_{i=1}^{n} A_i t_i^{2n-1} & = & 0 \end{cases} \tag{4.7}$$

Pour résoudre le système (4.7), on considère

$$f(t) = t^k P_n(t) \quad (k = 0, 1, ...n-1)$$

où $P_n(t)$ est le polynôme de Legendre.

Les degrés de ce polynôme ne dépassant pas $2n-1$, ces polynômes doivent vérifier :

$$\int_{-1}^{1} f(t)dt = \sum_{i=1}^{n} A_i f(t_i)$$

et

$$\int_{-1}^{1} t^k P_n(t)dt = \sum_{i=1}^{n} A_i t_i^k P_n(t_i) \quad (k = 0, 1, ...n-1)$$

comme

$$\int_{-1}^{1} P_n(x)Q_k(x)dx = 0, \text{ pour } (k < n)$$

on

$$\int_{-1}^{1} t^k P_n(x)dx = 0, \text{ pour } (k < n)$$

et donc

$$\sum_{i=1}^{n} A_i t_i^k P_n(t_i) = 0 \quad (k = 0, 1, ...n-1) \tag{4.8}$$

si l'on pose $P_n(t_i) = 0$ pour $(i = 1, 2, ..., n)$ les égalités (4.8) sont vérifiées pour tout A_i.Si on connait t_i, on trouve à partir du système linéaire des n premières équations du système, les constantes A_i pour $(i = 1, 2, ..., n)$.

R La formule (4.8) où les t_i sont les racines du polynôme de Legendre $P_n(t)$ et où les A_i pour $(i = 1, 2, ..., n)$ sont définis à partir du système, s'appelle *formule de quadrature de Gauss*.

■ **Exemple 4.2** Trouver la formule de quadrature de Gauss, dans le cas de trois ordonnées $(n = 3)$.

■

Solution : Comme le polynôme de Legendre de degré 3 est le polynôme :

$$P_3(t) = \frac{1}{2}(5t^3 - 3t)$$

en l'annulant on obtient ses racines qui sont données par :

$$\begin{aligned} t_1 &= -\sqrt{\frac{3}{5}} \simeq -0,7745 \\ t_2 &= 0 \\ t_3 &= \sqrt{\frac{3}{5}} \simeq 0,7745 \end{aligned}$$

pour la détermination des coefficients A_i pour $(i = 1, 2, 3)$, on obtient le système :

$$\begin{cases} \sum_{i=1}^{3} A_i &= 2 \\ \sum_{i=1}^{3} A_i t_i &= 0 \\ \sum_{i=1}^{3} A_i t_i^2 &= \frac{2}{3} \end{cases}$$

c'est-à-dire

$$\begin{cases} A_1+A_2+A_3 &= 2 \\ -\sqrt{\frac{3}{5}}A_1+\sqrt{\frac{3}{5}}A_3 &= 0 \\ \frac{3}{5}A_1+\frac{3}{5}A_3 &= \frac{2}{3} \end{cases}$$

dont la solution est : $A_1 = A_3 = \frac{5}{9}$, $A_2 = \frac{8}{9}$.D'où :

$$\int_{-1}^{1} f(t)dt = \sum_{i=1}^{3} A_i f(t_i) = \frac{1}{9}\left[5f(-\sqrt{\frac{3}{5}})+8f(0)+5f(\sqrt{\frac{3}{5}})\right]$$

4.1.10 Calcul de $\int_a^b f(x)dx$

Pour calculer $\int_a^b f(x)dx$, on fait le changement de variable suivant :

$$x = \frac{b+a}{2}+\frac{b-a}{2}t$$

on obtient :

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

en appliquant la formule de quadrature de Gauss on a :

$$\int_a^b f(x)dx = \frac{b-a}{2}\sum_{i=1}^{n} A_i f(x_i)$$

où

$$x_i = \frac{b+a}{2}+\frac{b-a}{2}t_i \quad (i = 1,2,...,n)$$

et où t_i sont les racines du polynôme de Legendre $P_n(t)$, c'est-à-dire $P_n(t_i) = 0$.

4.1.11 Erreur de l'intégration par la méthode de Gauss

Le reste de la formule de Lagrange à n points est donné par

$$R_n = \frac{(b-a)^{2n-1}(n!)^4 f^{(4)}(\xi)}{(2n!)^3(2n+1)}$$

d'où l'on tire

$$\begin{aligned} R_2 &= \frac{1}{135}\left(\frac{b-a}{2}\right)^5 f^{(4)}(\xi) \\ R_3 &= \frac{1}{15750}\left(\frac{b-a}{2}\right)^7 f^{(6)}(\xi) \end{aligned}$$

■ **Exemple 4.3** Calculer par la méthode de quadrature de Gauss à trois ordonnées, l'intégrale suivante :

$$\int_0^1 \sqrt{1+2x}dx$$

■

Solution : Comme $a = 0$ et $b = 1$, d'après le changement de variable :

$$x_i = \frac{b+a}{2} + \frac{b-a}{2} t_i \quad (i = 1,2,3)$$

on obtient :

$$\begin{aligned} x_1 &= \frac{1}{2} + \frac{1}{2} t_1 = \frac{1}{2} + \frac{1}{2}.(-\sqrt{\frac{3}{5}}) \simeq 0,112 \\ x_2 &= \frac{1}{2} + \frac{1}{2} t_2 = \frac{1}{2} + \frac{1}{2}.0 \simeq 0,500 \\ x_3 &= \frac{1}{2} + \frac{1}{2} t_3 = \frac{1}{2} + \frac{1}{2}.(\sqrt{\frac{3}{5}}) \simeq 0,887 \end{aligned}$$

donc les coefficients C_i sont :

$$\begin{aligned} C_1 &= \frac{b-a}{2} A_1 = \frac{1}{2}.\frac{5}{9} \simeq 0,277 \\ C_2 &= \frac{b-a}{2} A_2 = \frac{1}{2}.\frac{8}{9} \simeq 0,444 \\ C_3 &= \frac{b-a}{2} A_3 = \frac{1}{2}.\frac{5}{9} \simeq 0,277 \end{aligned}$$

donc

$$\int_0^1 f(x)dx = \int_0^1 \sqrt{1+2x}dx = \sum_{i=1}^{3} C_i f(x_i) = 1,398$$

l'erreur commise dans le calcul de cette intégrale s'évalue de la manière suivante :

$$R_3 = \frac{1}{15750} \left(\frac{b-a}{2} \right)^7 f^{(6)}(\xi) \text{ où } \xi \in [0,1]$$

comme $f^{(6)}(x) = -945(1+2x)^{\frac{-11}{2}}$ alors $\max_{x\in[0,1]} \left| f^{(6)}(x) \right| = 945$ donc

$$R_3 \leq \frac{945}{15750} (\frac{1}{2})^7 = 0,5.10^{-3}$$

4.2 SERIE D'EXERCICES

Exercice 4.1 1. Soit f une fonction possédant 4 dérivées continues dans l'intervalle $[0,5]$, donner une évaluation de $\int_0^5 f(x)dx$ sachant que $x_1 = 1; x_2 = 2$ et $x_3 = 4$.

2. Soit g une fonction possédant 2 dérivées continues dans l'intervalle $[0,2]$, donner une évaluation de $\int_0^2 (x-1)g(x)dx$ sachant que $x_1 = 0; x_2 = 2$.

■

Exercice 4.2 1. Dans l'intervalle $[a,b]$ on prend $x_1 = a, x_2 = a+h = b$, $f \in C^2[a,b]$, évaluer la formule des trapèzes $\int_a^b f(x)dx$.

2. On prend $x_1 = a, x_2 = a+h, x_3 = a+2h, ...x_{n+1} = a+nh = b$, retrouver la formule des trapèzes généralisée.
3. Déduire la valeur approximative de l'intégrale

$$\int_0^1 \frac{dx}{1+x}$$

pour $n = 6$.
4. Calculer la valeur exacte de cette intégrale et déterminer les erreurs absolues et relatives.

■

Exercice 4.3 On suppose que f est une fonction 3 fois continument dérivable dans $[-h,h]$ et $f^{(4)}(x)$ continue dans cet intervalle, f est donnée aux points : $x_1 = -h, x_2 = 0, x_3 = h$.

1. Etablir la formule suivante :

$$\int_{-h}^{x} f(t)dt = \frac{2x^3 - 3hx^2 + 5h^3}{12h^2} f(-h) - \frac{x^3 - 3h^2x - 2h^3}{3h^2} f(0) + \frac{2x^3 + 3hx^2 - h^3}{12h^2} f(h) + \varepsilon(x)$$

On donnera une expression de $\varepsilon(x)$.
2. On suppose que $x \in [-h,x]$:

a) Montrer que $\varepsilon(x)$ peut s'écrire sous la forme :

$$\varepsilon(x) = \frac{1}{24}(x^2 - h^2)^2 f^{(3)}(\zeta),$$

avec $\zeta \in [-h,h]$.

b) Utiliser le résultat précédent pour donner une valeur approchée de :

$$\int_{-\frac{1}{4}}^{0} \frac{dt}{1+t}$$

Quelle est la précision obtenue ?

Comparer ce résultat à celui que donnerait la formule des trapèzes utilisant les points $x_1 = -\frac{1}{4}$ et $x_2 = 0$.

■

Exercice 4.4 Déduire la formule de Gauss de la fonction f sur l'intervalle $[-1,1]$ pour le cas de trois ordonnées, on prendra pour la fonction poids $\omega(x)=1$.

1. En utilisant la formule de Gauss à trois ordonnées, calculer l'intégrale :

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

2. En déduire $\int_0^1 \sqrt{1+2x}\,dx$.

■

4.3 DÉRIVATION NUMÉRIQUE

4.3.1 Généralités :

Pour résoudre un certain nombre de problèmes pratiques (étudier la vitesse d'un changement à l'intérieur d'un système par exemple), il est nécessaire parfois de calculer les dérivées d'une fonction $y=f(x)$ supposée dérivable mais connue de façon discrète sur un intervalle $[a,b]$, ou que l'expression analytique compliquée de cette fonction rende difficile sa dérivation.

Comment fournir une valeur approchée de la dérivée, d'ordre un ou supérieur, de $f(x)$ en un point de $[a,b]$.

Le principe est d'approcher la fonction à dériver par un polynôme d'interpolation $P_n(x)$ dont on calcule la dérivée ensuite, c'est-à-dire en posant :

$$f'(x)=P'_n(x)$$

Les dérivées d'ordre supérieur de $f(x)$ s'obtiennent de la même façon.

Si l'on connait l'erreur d'interpolation :

$$\varepsilon(x)=f(x)-P_n(x)$$

l'erreur de la dérivée $P'_n(x)$ est donnée par :

$$r(x)=f'(x)-P'_n(x)=\varepsilon'(x)$$

Soit f une fonction numérique $y=f(x)$ dérivable (respectivement p fois dérivable), donnée aux points équidistants $\{x_1,x_2,....,x_{n+1}\}\subset[a,b]$ par

$$y_i=f(x_i)$$

Pour chercher sur $[a,b]$ la dérivée $y'=f'(x)$, (respectivement la dérivée d'ordre p c'est-à-dire $y^{(p)}=f^{(p)}(x)$)

Nous allons chercher la valeur de la dérivée sous la forme :

$$y^{(p)}(x)=f^{(p)}(x)=\sum_{i=1}^{n+1}\alpha_i(x)f(x_i)+r(x)$$

les $\alpha_i(x)$ seront choisis de telle sorte que la fonction reste $r(x)$ soit nulle si f est d'un type déterminé (un polymôme).

Si f est quelconque et $r(x)$ suffisamment petit nous considérons que :

$$D(x) = \sum_{i=1}^{n+1} \alpha_i(x) f(x_i)$$

est une approximation de $f^{(p)}(x)$.

Soient $v_1, v_2, ..., v_{n+1}$ $(n+1)$ fonctions linéairement indépendantes p fois continument dérivables sur $[a,b]$. On note $\mathscr{F}_n$ le sous espace engendré par ces fonctions. Pour que $r(x)$ soit nul, il faut que les fonctions $v(x)$ vérifient

$$v^{(p)}(x) = \sum_{i=1}^{n+1} \alpha_i(x) v_k(x_i); \quad k = 1, ..., n+1.$$

On suppose que les fonctions[3] $v_1, v_2, ..., v_{n+1}$ sont telles que le système admette une soultion unique.

Soit P_n la fonction d'interpolation de f dans $\mathscr{F}_n$ vérifiant :

$$P_n(x) = \sum_{k=1}^{n+1} a_k v_k(x)$$

et

$$P_n(x_i) = f(x_i) \quad \text{pour} \quad i = 1, ..., n+1.$$

nous avons

$$P_n^{(p)}(x) = \sum_{i=1}^{n+1} \alpha_i(x) P_n(x_i)$$

si

$$\varepsilon(x) = f(x) - P_n(x)$$

nous obtenons :

$$f_n^{(p)}(x) = \sum_{i=1}^{n+1} \alpha_i(x) f_n(x_i) + \varepsilon^{(p)}(x)$$

R Si la fonction d'interpolation $P_n(x)$ est le polynôme d'interpolation de f, c'est-à-dire que :

$$\{v_1(x), v_2(x), ..., v_{n+1}(x)\} = \{1, x, ..., x^n\}$$

On choisira $\{\alpha_1(x), \alpha_2(x), ..., \alpha_{n+1}(x)\}$ de telle sorte que $r(x)$ soit nul quand f est un polynôme quelconque de degré inférieur ou égal à n.

3. vérifient les conditions de Haar.

4.3.2 Utilisation de l'interpolation polynomiale

Nous pouvons remplacer la fonction f par son polynôme d'interpolation de Newton (par exemple)

$$\begin{aligned} y &= f(x) = y_1 + k\Delta y_1 + \frac{k(k-1)}{2!}\Delta^2 y_1 + \frac{k(k-1)(k-2)}{3!}\Delta^3 y_1 + \cdots \\ &= y_1 + k\Delta y_1 + \frac{k^2-k}{2}\Delta^2 y_1 + \frac{k^3-3k^2+2k}{6}\Delta^3 y_1 + \cdots \end{aligned}$$

Où

$$k = \frac{x - x_1}{h} \quad \text{et} \quad h = x_{i+1} - x_i \quad (i = 1, 2, ...)$$

Comme

$$y'(x) = \frac{dy}{dx} = \frac{dy}{dk}\frac{dk}{dx} = \frac{1}{h}\frac{dy}{dk}$$

On obtient

$$y'(x) = \frac{1}{h}\left[\Delta y_1 + \frac{2k-1}{2}\Delta^2 y_1 + \frac{3k^2-6k+2}{6}\Delta^3 y_1 + \cdots\right]$$

D'une façon analogue, comme

$$y''(x) = \frac{d(y')}{dx} = \frac{d(y')}{dk}\frac{dk}{dx}$$

on a

$$y''(x) = \frac{1}{h^2}\left[\Delta^2 y_1 + (k-1)\Delta^3 y_1 + \cdots\right]$$

R On procède de la même manière pour chercher les dérivées d'un ordre quelconque.

R Pour chercher les dérivées $y'(x), y''(x)$ en un point fixé x, il faut prendre comme x_1 la valeur tabulée de l'argument la plus proche de ce point x.

R Lorsqu'on cherche la valeur de la dérivée ou des dérivées de y aux points d'interpolation x_i, on peut considérer toute valeur tabulée comme étant une valeur initiale ; on a alors

$$y'(x_1) = \frac{1}{h}\left[\Delta y_1 - \frac{\Delta^2 y_1}{2} + \frac{\Delta^3 y_1}{3} - \frac{\Delta^4 y_1}{4} + \frac{\Delta^5 y_1}{5} - \cdots\right]$$

et

$$y''(x_1) = \frac{1}{h^2}\left[\Delta^2 y_1 - \frac{\Delta^3 y_1}{3} + \frac{11}{12}\Delta^4 y_1 - \frac{5}{6}\Delta^5 y_1 + \cdots\right]$$

4.3.3 Erreur de dérivation

cas de $\varepsilon' = f' - P_n'$

Soit $P_n(x)$ le polynôme d'interpolation de f. Nous avons

$$P_n(x) = \sum_{i=1}^{n+1} a_i v_i(x) = \sum_{i=1}^{n+1} L_i(x) f(x_i)$$

On sait que pour tout $x \in [a,b]$, il existe $\xi_x \in [a,b]$ tel que

$$\varepsilon(x) = f(x) - P_n(x) = \frac{1}{(n+1)!} \left(\prod_{i=1}^{n+1} (x - x_i) \right) f^{(n+1)}(\xi_x).$$

Par ailleurs $\varepsilon(x) = L(x)g(x)$ avec $L(x) = \prod_{i=1}^{n+1}(x-x_i)$ et $g(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi_x)$, en dérivant nous obtenons

$$\varepsilon'(x) = f'(x) - P_n'(x) = L'(x)g(x) + L(x)g'(x). \tag{4.9}$$

Le point x pour lequel nous cherchons une approximation peut être :

- soit l'un des points d'interpolation x_i.
- soit un point de $[a,b]$ différent de x_i pour tout i.

L'erreur commise ne sera pas la même dans les deux cas.

a) si $x = x_i$ dans ce cas $L(x_i) = 0$ et

$$L'(x_i) = \lim_{x \to x_i} \frac{\prod_{j=1}^{n+1}(x - x_j)}{x - x_i} = \lim_{x \to x_i} \prod_{\substack{j=1 \\ j \neq i}}^{n+1} (x - x_j) = \prod_{\substack{j=1 \\ j \neq i}}^{n+1} (x_i - x_j)$$

d'où

$$\varepsilon'(x) = \frac{1}{(n+1)!} \left(\prod_{\substack{j=1 \\ j \neq i}}^{n+1} (x_i - x_j) \right) f^{(n+1)}(\xi_{x_i})$$

b) si $x \neq x_i$ pour tout i, dans ce cas nous devons connaitre une estimation de $g\prime(x)$. Si f est $(n+2)$ fois continument dérivable, pour tout $x \in [a,b]$, il existe un élément η_x tel que

$$g\prime(x) = \frac{1}{(n+2)!} f^{(n+2)}(\eta_x)$$

dans ce cas on a :

$$\varepsilon'(x) = \frac{1}{(n+1)!} L'(x) f^{(n+1)}(\xi_{x_i}) + \frac{1}{(n+2)!} L(x) f^{(n+2)}(\eta_x)$$

Si on connait des bornes de $f^{(n+1)}$ et $f^{(n+2)}$, c'est-à-dire si on note

$$M_p = \max_{x \in [a,b]} \left| f^{(p)}(x) \right|$$

nous avons alors

$$\left| \varepsilon'(x_i) \right| \leq \frac{M_{n+1}}{(n+1)!} \prod_{\substack{j=1 \\ j \neq i}}^{n+1} (x_i - x_j)$$

et

$$\left| \varepsilon'(x) \right| \leq \frac{M_{n+1}}{(n+1)!} \left| L'(x) \right| + \frac{M_{n+2}}{(n+2)!} \left| L(x) \right| \quad \text{pour } x \in [a,b]; x \neq x_i; i = 1,2,...,n$$

Si $P_m(x)$ est un polynôme de Newton contenant les différences $\Delta y_1, \Delta^2 y_1, ..., \Delta^m y_1$ et si l'erreur correspondante est donnée par :

$$\varepsilon_m(x) = f(x) - P_m(x)$$

l'erreur de la dérivée s'écrit :

$$r(x) = \varepsilon'_m(x) = f'(x) - P'_m(x)$$

comme

$$\varepsilon_m(x) = \frac{(x-x_1)(x-x_2)\cdots(x-x_m)}{(m+1)!} f^{(m+1)}(\xi) = h^{m+1}\frac{k(k-1)\cdots(k-m)}{(m+1)!} f^{(m+1)}(\xi)$$

où $\xi \in [x, x_m]$. En supposant que $f \in C^{(m+2)}$ on obtient :

$$r(x) = \frac{d\varepsilon_m(x)}{dk}\frac{dk}{dx} = \frac{h^m}{(m+1)!}\left[f^{(m+1)}(\xi)\frac{d}{dk}\left[k(k-1)\cdots(k-m)\right] + k(k-1)\cdots(k-m)\frac{d}{dk}f^{(m+1)}(\xi)\right] \tag{4.10}$$

en supposant $\frac{d}{dk}f^{(m+1)}(\xi)$ bornée et tenant compte du fait que $\frac{d}{dk}\left[k(k-1)\cdots(k-m)\right]_{k=0} = (-1)^m m!$, on en tire avec $x = x_1$ et par suite avec $k = 0$,

$$r(x) = (-1)^m \frac{h^m}{(m+1)} f^{(m+1)}(\xi) \tag{4.11}$$

■ **Exemple 4.4** Calculer $y\prime(50)$ pour la fonction $y = f(x)$ donnée par le tableau suivant :

x	50	55	60	65
y	1,6990	1,7404	1,7782	1,8129

■

Solution : Formons le tableau des différences finies comme suit :

x	y	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
50	1,6990			
		0,0414		
55	1,7404		−0,0036	
		0,0378		0,0005
60	1,7782		−0,0031	
		0,0347		
65	1,8129			

cas de $\varepsilon^{(p)} = f^{(p)} - P_n^{(p)}$

En généralisant l'équation (4.9) on obtient :

$$\varepsilon^{(p)}(x) = \sum_{j=0}^{p} \mathbb{C}_p^j L^{(p-j)}(x) g^{(j)}(x) \tag{4.12}$$

en supposant que f soit $(n+1+p)$ continument dérivable, cette equation (4.12) s'écrit

$$\varepsilon^{(p)}(x) = \sum_{j=0}^{p} \mathbb{C}_p^j L^{(p-j)}(x) \frac{f^{(n+1+j)}(\xi_j)}{(n+1+j)!}, \quad \xi_j \in [a,b]$$

en réecrivant autrement l'équation (4.10) on obtient :

$$\frac{\partial^p}{\partial x^p} f[x, x_1, ..., x_{n+1}] = f\left[\underbrace{x, x, ...x}_{p+1\ fois}, x_1, ..., x_{n+1}\right]$$

on a :

$$\varepsilon^{(p)}(x) = \sum_{j=0}^{p} \mathbb{C}_p^j L^{(p-j)}(x) f\left[\underbrace{x, x, ...x}_{j+1\ fois}, x_1, ..., x_{n+1}\right]$$

lorsque les points sont équidisants c'est-à-dire si : $x_i = x_{i-1} + h = x_1 + (i-1)h$, $i \geq 1$ et si nous posons : $x = x_1 + th$, nous avons :

$$L_i(x) = l_i(t) = \prod_{\substack{j=1 \\ j \neq i}}^{n+1} \frac{(t-j+1)}{i-j}$$

et

$$P_n(x) = p_n(t) = \sum_{i=1}^{n+1} l_i(t) f(x_i)$$

d'où

$$P_n'(x) = p_n'(t) = \sum_{i=1}^{n+1} l_i'(t) f(x_i)$$

et

$$P_n^{(p)}(x) = p_n^{(p)}(t) = \sum_{i=1}^{n+1} l_i^{(p)}(t) f(x_i)$$

R Les coefficients $l_i(t)$, $l_i'(t)$, $l_i^{(p)}(t)$ ne dépendant ni de h ni de x_1, peuvent être tabulés.

■ **Exemple 4.5** Trouver l'extremum de la fonction donnée par le tableau suivant :

x	1,80	1,82	1,84	1,86	1,88	1,90
y	0,5815170	0,5817731	0,5818649	0,5817926	0,5815566	0,5811571

■

Solution : Le tableau des différences est donnée par le tableau suivant :

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
1,80	0,5815170			
		0,002561		
1,82	0,5817731		−0,001643	
		0,000918		0,000002
1,84	0,5818649		−0,001641	
		−0,000723		0,000004
1,86	0,5817926		−0,001637	
		−0,002360		0,000002
1,88	0,5815566		−0,001635	
		−0,003995		
1,90	0,5811571			

l'extremum est atteint pour $f\prime(x)=0$ c'est-à-dire :

$$0=\frac{0,000918-0,000723}{2}+k(-0,001641)+\frac{3k^2-1}{6}\frac{0,000002+0,000004}{2}$$

ou

$$0=\frac{3}{2}k^2-1641k+97$$

ou aussi

$$k=\frac{97}{1641}+\frac{1}{1094}k^2$$

ce qui donne $k=0,05911$ d'où $x=x_1+kh=1,84+0,05911.0,02=1,8411822$.

4.3.4 Algorithmes de dérivation

Les formules de dérivation numériques déduites au paragraphe précédent pour la fonction $y=f(x)$ au point $x=x_1$ ont l'inconvenient de n'utiliser que des valeurs de la fonction pour $x>x_1$.

Les formules de dérivation qui tiennent compte des valeurs de $y=f(x)$ aussi bien pour $x>x_1$ que pour $x<x_1$ sont relativement plus exactes. Ces formules s'appellent *formules de dérivation par différences centrales*.

Soient $...,x_{-3},x_{-2},x_{-1},x_0,x_1,x_2,x_3,...$un système de points équidistants et numérotés symétriquement par rapport à x_0, à pas $x_{i+1}-x_i=h$ et $y_i=f(x_i)$ les valeurs correspondantes de $y=f(x)$. Si on pose

$$k=\frac{x-x_1}{h}$$

alors si le le polynôme d'interpolation est le polynôme de Stirling, on aura :

$$\begin{aligned} y &= f(x)=y_0+k\Delta y_{-\frac{1}{2}}+\frac{k^2}{2}\Delta^2 y_{-1}+\frac{k^2(k^2-1)}{3!}\Delta^3 y_{-\frac{3}{2}}+\frac{k^2(k^2-1)}{4!}\Delta^4 y_{-2}+ \\ &\quad +\frac{k^2(k^2-1)(k^2-2^2)}{5!}\Delta^5 y_{-\frac{5}{2}}+\frac{k^2(k^2-1)(k^2-2^2)}{6!}\Delta^6 y_{-3}+\cdots \end{aligned} \tag{4.13}$$

où $\Delta y_i=y_{i+1}-y_i$; $\Delta y_{-\frac{1}{2}}=\frac{\Delta y_{-1}+\Delta y_0}{2}$, $\Delta^3 y_{-\frac{3}{2}}=\frac{\Delta^3 y_{-2}+\Delta^3 y_{-1}}{2}$, $\Delta^5 y_{-\frac{5}{2}}=\frac{\Delta^5 y_{-3}+\Delta y_{-2}}{2}$, etc...

En tenant compte de :

$$\frac{dk}{dx}=\frac{1}{h}$$

on obtient de la formule (4.13) :

$$\begin{aligned} y' &= f'(x)=\frac{1}{h}(\Delta y_{-\frac{1}{2}}+k\Delta^2 y_{-1}+\frac{3k^2-1}{6}\Delta^3 y_{-\frac{3}{2}}+\frac{2k^2-k}{12}\Delta^4 y_{-2}+ \\ &\quad +\frac{5k^4-15k^2+4}{120}\Delta^5 y_{-\frac{5}{2}}+\frac{3k^5-10k^3+4k}{360}\Delta^6 y_{-3}+\cdots) \end{aligned}$$

et

$$\begin{aligned} y'' &= f''(x)=\frac{1}{h^2}(\Delta^2 y_{-1}+k\Delta^3 y_{-\frac{3}{2}}+\frac{6k^2-1}{12}\Delta^4 y_{-2}+ \\ &\quad +\frac{2k^3-3k}{12}\Delta^5 y_{-\frac{5}{2}}+\frac{15k^4-30k^2+4}{360}\Delta^6 y_{-3}+\cdots) \end{aligned}$$

en particulier si k=0, on a :

$$y'(x_0) = \frac{1}{h}(\Delta y_{-\frac{1}{2}} - \frac{1}{6}\Delta^3 y_{-\frac{3}{2}} + \frac{1}{30}\Delta^5 y_{-\frac{5}{2}} + \cdots) \tag{4.14}$$

et

$$y''(x_0) = \frac{1}{h^2}(\Delta^2 y_{-1} - \frac{1}{12}\Delta^4 y_{-2} + \frac{1}{90}\Delta^6 y_{-3} + \cdots) \tag{4.15}$$

■ **Exemple 4.6** Calculer la dérivée $y'(1)$ et la dérivée seconde $y''(1)$ de la fonction donnée par le tableau suivant :

x	0,96	0,98	1,00	1,02	1,04
y	0,7825361	0,7739332	0,7651977	0,7563321	0,7473390

■

Solution : En composant les différences de la fonction $y = f(x)$ on obtient :

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
0,96	0,7825361				
		−0,086029			
0,98	0,7739332		−0,001326		
		−0,87355		0,000025	
1,00	0,7651977		−0,001301		0,000001
		−0,88656		0,000026	
1,02	0,7563321		−0,001275		
		−0,89931			
1,04	0,7473390				

et en appliquant (4.14) on a :

$$\begin{aligned} y'(1) &= \frac{1}{0,02}(-\frac{87355+88656}{2}.10^{-7} - \frac{1}{6}.\frac{25+26}{2}.10^{-7} + \frac{1}{30}.1.10^{-7}) = \\ &= -50(88005,5+4,2+0).10^{-7} = -0,4400485. \end{aligned}$$

et

$$\begin{aligned} y''(1) &= \frac{1}{0,02^2}(-1301.10^{-7} - \frac{1}{12}.1.10^{-7}) = \\ &= -2500.1301.10^{-7} = -0,325250. \end{aligned}$$

R Quand les points d'interpolation sont équidistants, les différences divisées sont remplacées par différences finies.
- On appelle *différences finies* d'ordre 1 *progressives*, notées $\nabla_h f$, la fonction définie par :

$$\nabla_h f(x) = \frac{1}{h}(f(x+h) - f(x))$$

-On appelle *différences finies* d'ordre 1 *régressives*, notées $\bar{\nabla}_h f$, la fonction définie par :

$$\bar{\nabla}_h f(x) = \frac{1}{h}(f(x) - f(x-h))$$

-On appelle *différences finies* d'ordre 1 *centrales*, notées $\delta_h f$, la fonction définie par :

$$\delta_h f(x) = \frac{1}{h}(f(x+\frac{h}{2}) - f(x-\frac{h}{2}))$$

- On définit les *différences finies* d'ordre k *progressives*, notées $\nabla_h^k f$, la fonction définie par :

$$\nabla_h^k f(x) = \nabla_h(\nabla_h^{k-1} f(x))$$

-On appelle *différences finies* d'ordre 1 *régressives*, notées $\bar{\nabla}_h^k f$, la fonction définie par :

$$\bar{\nabla}_h^k f(x) = \bar{\nabla}_h(\bar{\nabla}_h^{k-1} f(x))$$

-On appelle *différences finies* d'ordre 1 *centrales*, notées $\delta_h^k f$, la fonction définie par :

$$\delta_h^k f(x) = \delta_h(\delta_h^{k-1} f(x))$$

R On appelle différences non divisées le produit

$$\nabla^k = \nabla_h^k h^k$$

pour les différences non divisées progressives,

$$\bar{\nabla}^k = \bar{\nabla}_h^k h^k$$

pour les différences non divisées régressives, et

$$\delta^k = \delta_h^k h^k$$

pour les différences non divisées centrales.

4.3.5 Formules centrales de dérivation

$$f'(x_0) = \frac{1}{2h}(f(x_1) - f(x_{-1})) - \frac{h^2}{6} f^{(3)}(\xi), \quad \xi \in [x_{-1}, x_1]$$

$$f'(x_0) = \frac{1}{12h}(f(x_{-2}) - 8f(x_{-1}) + 8f(x_1) - f(x_2)) + \frac{h^4}{30} f^{(5)}(\xi), \quad \xi \in [x_{-2}, x_2]$$

$$f''(x_0) = \frac{1}{h^2}(f(x_1) - 2f(x_0) + f(x_{-1})) + \frac{h^2}{12} f^{(4)}(\xi), \quad \xi \in [x_{-1}, x_1]$$

$$f''(x_0) = \frac{1}{24h^2}(-2f(x_{-2}) + 32f(x_{-1}) - 60f(x_0) + 32f(x_1) - 2f(x_2)) + \frac{h^4}{90} f^{(6)}(\xi), \quad \xi \in [x_{-2}, x_2]$$

4.3.6 Formules non centrales de dérivation

$$f'(x_{-1}) = \frac{1}{2h}(-3f(x_{-1}) + 4f(x_0) - f(x_1)) - \frac{h^2}{3} f^{(3)}(\xi), \quad \xi \in [x_{-1}, x_1]$$

$$f'(x_1) = \frac{1}{2h}(f(x_{-1}) - 4f(x_0) + 3f(x_1)) + \frac{h^2}{3} f^{(3)}(\xi), \quad \xi \in [x_{-1}, x_1]$$

$$f'(x_{-1}) = \frac{1}{12h}(-25f(x_2) + 48f(x_{-1}) - 36f(x_0) + 16f(x_1) - 3f(x_2)) + \frac{h^4}{5} f^{(5)}(\xi), \quad \xi \in [x_{-2}, x_2]$$

$$f'(x_1) = \frac{1}{12h}(-f(x_{-2}) + 6f(x_{-1}) - 18f(x_0) + 10f(x_1) + 3f(x_2)) - \frac{h^4}{20} f^{(5)}(\xi), \quad \xi \in [x_{-2}, x_2]$$

$$f'(x_2) = \frac{1}{12h}(3f(x_{-2}) - 16f(x_{-1}) + 36f(x_0) - 48f(x_1) + 25f(x_2)) - \frac{h^4}{5} f^{(5)}(\xi), \quad \xi \in [x_{-2}, x_2]$$

4.4 SERIE D'EXERCICES

Exercice 4.5 Soit f une fonction possédant $(n+2)$ dérivées continues dans l'intervalle $[a,b]$, l'erreur d'interpolation polynomiale en $(n+1)$ points est donnée par :

$$\varepsilon_n(x) = f(x) - p_n(x) = \frac{(x-x_1)(x-x_2)....(x-x_{n+1})f^{(n+1)}(\xi_x)}{(n+1)!}$$

1. Calculer $\varepsilon_n'(x)$;
2. Donner une évaluation de $\varepsilon_n'(x)$ en un point d'interpolation x_k;
3. Pour $n=1, x_1=a, x_2=a+h$, calculer $p_1'(a)$ et $\varepsilon_1'(a)$;
4. Pour $n=2, x_1=a, x_2=a+h, x_3=a+2h$, calculer $p_2'(a)$ et $\varepsilon_2'(a)$.

■

Exercice 4.6 Calculer $y'(0,97)$ de la fonction $y=f(x)$ donnée par le tableau suivant :

x	0,96	0,98	1,00	1,02	1,04
y	0,7825361	0,7739332	0,7651977	0,7563321	0,7473390

■

Exercice 4.7 Calculer $y'(50)$ et $y''(50)$ de la fonction $y=\log(x)$ donnée par le tableau suivant :

x	50	55	60	65
y	1,6990	1,7404	1,7782	1,8129

■

Exercice 4.8

1. Soit $f(x)=e^x$, on donne le tableau suivant :

x	0,4	0,6	0,7	1,0
y	1,491825	1,822119	2,013753	2,718282

2. Calculer $f'(0,8)$ et donner une majoration de l'erreur.
3. Calculer $f''(0,8)$ et donner une majoration de l'erreur.

■

5. RESOLUTION D'UN SYSTEME LINEAIRE

5.1 METHODES DIRECTES

5.1.1 Rappel

Rang d'une matrice

Définition 5.1.1 Soit $A \in M_{m,n}(\mathbb{R})$ une matrice de m lignes et de n colonnes. Le rang de A est égal au nombre maximum de colonnes de A linéairement indépendant et on le note :

$$rang\ A$$

Le rang de A est donc aussi égal à la dimension de l'image de toute application linéaire représenté par A.

Proposition 5.1.1 - $rang\ A = rang\ A^t$ - si $A \in M_{m,n}(\mathbb{R})$ alors $rang\ A \leq \inf(m,n)$ - si $A \in M_n(\mathbb{R})$ alors A inversible $\Leftrightarrow$ $rang\ A = n$.

Définition 5.1.2 Soit $A \in M_{m,n}(\mathbb{R})$, on appelle matrice extraite de A une matrice obtenue par sélection de lignes et de colonnes. Par exemple si I (respectivement J) est un sous ensemble de $\{1,2,...m\}$ (respectivement $\{1,2,...n\}$) on définit une matrice extraite B de A par :

$$B = (a_{ij})_{i \in I, j \in J}$$

Théoreme 5.1.2 Soit $A \in M_{m,n}(\mathbb{R})$ une matrice de *rang* r alors :
- le rang de toute matrice extraite de A est inférieur ou égal à r,
- toute matrice carrée inversible extraite de A est d'ordre inf érieur ou égal à r,
- il existe une matrice carrée extraite de A d'ordre r qui est inversible.

5.1.2 Systèmes linéaires

Définition 5.1.3 On appelle système de m équations à n inconnues $x_1, x_2, ..., x_n$ la famille

d'équations :

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n & = & b_2 \\ ..... & ... & ... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n & = & b_m \end{cases} \tag{5.1}$$

que l'on peut mettre sous la forme matricielle suivante :

$$Ax = b$$

où $A \in M_{m,n}(\mathbb{R})$ est une matrice donnée par ses éléments (a_{ij}), $1 \leq i \leq m, 1 \leq j \leq n$, $b \in \mathbb{R}^m$ de composantes $(b_1, b_2, ..., b_m)$ et x le vecteur inconnu de composantes $(x_1, x_2, ..., x_n)$.

R Si $b = 0$, on dit que le système (5.1) est homogène. Dans le cas où $m < n$, on dit que le système est sous determiné et dans le cas où $n > m$ on parle d'un système sur determiné.

Théoreme 5.1.3 Une condition suffisante pour que le système (5.1) admette au moins une solution est que $rang\ A = m$.

Corollaire 5.1.4 Soit $A \in M_n(\mathbb{R})$ une matrice carrée, une condition nécessaire et suffisante pour que le système $Ax = b$ admette une solution unique est que A soit inversible. Autrement dit $\det A \neq 0$ (ou $rang A = n$). Le système (5.1) est alors dit système de Cramer.

5.1.3 Résolution d'un système triangulaire supérieur

Définition 5.1.4 Soit $A \in M_n(\mathbb{R})$ une matrice carrée, A est dite triangulaire supérieure si :

$$a_{ij} = 0, \quad \forall i > j$$

Dans ce cas le système $Ax = b$ est dit triangulaire supérieur.

R On ne traitera pas le cas des systèmes triangulaires inférieurs car la technique de résolution est identique.

Le système d'équations $Ax = b$ triangulaire supérieur a la forme suivante :

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n & = & b_1 \\ a_{22}x_2 + ... + a_{2n}x_n & = & b_2 \\ ..... & ... & ... \\ a_{nn}x_n & = & b_n \end{cases}$$

Théoreme 5.1.5 Soit le système triangulaire supérieur $Ax = b$ où $A \in M_n(\mathbb{R})$ est une matrice carrée et $b \in \mathbb{R}^n$, si :

$$a_{kk} \neq 0, \quad \forall k \in [1, n]$$

alors le système admet une solution unique et cette solution x^* est telle que :

$$x_k^* = \frac{b_k - \sum_{j=k+1}^{n} a_{kj} x_j^*}{a_{kk}}, \quad k = \{n, n-1, ..., 1\} \tag{5.2}$$

Nous allons étudier les méthodes de résolution du système de n équations linéaires à n inconnues $Ax = b$, par les mé thodes directes. Nous entendons par méthodes directes des méthodes qui mènent à la solution en un nombre fini d'opérations élémentaires. Ces méthodes sont utilisées seulement si le nombre d'équations du système n'est pas trop élevé (généralement $n \leq 100$). La méthode de Cramer en est une, mais elle est numériquement inacceptable. Car sa mise en oeuvre demande le calcul de $n+1$ dé terminants et n divisions. Pour calculer chaque déterminant, nous devons effectuer $n!n$ multiplications et $n!-1$ additions soit un total de $(n+1)^2 n!-1$ opé rations élémentaires Par exemple, pour $n = 5$ on obtient 4319 opérations élé mentaires. Pour $n = 10$ on obtient à peu près 4.10^8 opérations él émentaires Or, dans la pratique, nous aurons à résoudre des systèmes d'ordres $n = 100$, $n = 1000$ voire même plus. Il est donc impossible de r ésoudre de tels systèmes par la méthode de Cramer. Dans ce chapitre, nous présentons essentiellement la méthode d'é liminations successives de Gauss et son interprétation matricielle, laquelle débouche sur la méthode de Cholesky pour un système à matrice définie positive. Si la matrice A n'est plus triangulaire, nous sommes amenés à chercher une matrice M inversible telle que la matrice produit MA soit triangulaire. On résoudra alors le système :

$$MAx = Mb$$

par l'algorithme (5.2). Nous nous limitons bien entendu à des systèmes $Ax = b$ avec $detA \neq 0$.

5.2 Méthode de Gauss

Soit $A \in M_n(\mathbb{R})$ une matrice carrée donnée et $b \in \mathbb{R}^n$. On cherche x^* solution du système linéaire :

$$Ax = b$$

La méthode de Gauss consiste à construire un système é quivalent plus facile à résoudre (à matrice triangulaire supé rieure par exemple). Deux systèmes linéaires définis par deux matrices $A \in M_n(\mathbb{R})$ et $U \in M_n(\mathbb{R})$ sont dits équivalents si leurs solutions sont identiques.

R - Les transformations élémentaires suivantes appliquées à un système linéaire engendre un système linéaire équivalent : - Une équation peut être remplacée par cette même é quation à laquelle on ajoute ou on retranche un certain nombre de fois une autre ligne. - La multiplication d'une équation par une constante non nulle. - La permutation de deux lignes ou de deux colonnes.

La représentation d'un système linéaire peut se faire à travers une matrice de dimension $n.(n+1)$ appelé matrice augmentée. La matrice est noté $\tilde{A} = [A|b]$ et a pour forme gé nérale :

$$\tilde{A} = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \cdots & \cdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array}\right)$$

La résolution du système linéaire ayant pour matrice augmenté e $\tilde{A}$ peut se faire en appliquant des transformations élé mentaires permettant d'obtenir un système équivalent. L'objectif de l'algorithme de Gauss est la construction d'un système triangulaire supérieur équivalent, en annulant au fur et à mesure les termes en dessous de la diagonale.

Définition 5.2.1 On appelle pivot de la transformation, l'élément a_{kk} de la matrice utilisée pour annuler les termes a_{jk}, $j > k$. La ligne k est alors appelée ligne pivot.

Théoreme 5.2.1 Soit un système linéaire défini par une matrice A d'ordre n et $b \in \mathbb{R}^n$. Si A est non singulière alors il existe une matrice U d'ordre n triangulaire supérieure et $y \in \mathbb{R}^n$ tels que $Ux = y$ soit équivalent à $Ax = b$. La résolution du système $Ax = b$ se fait ensuite par résolution du système triangulaire supérieur.

Démonstration. Construisons la matrice augmentée $\tilde{A}^{(1)} = \left[A^{(1)}|b^{(1)}\right]$

$$\tilde{A}^{(1)} = \left(\begin{array}{cccc|c} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} & b_1^{(1)} \\ a_{21}^{(1)} & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} & b_2^{(1)} \\ \cdots & \cdots & \ddots & \vdots & \vdots \\ a_{n1}^{(1)} & a_{n2}^{(1)} & \cdots & a_{nn}^{(1)} & b_n^{(1)} \end{array}\right)$$

l'exposant indiquant le nombre de fois qu'une valeur a été stocké e à la location i, j donnée. La première étape de l'algorithme de Gauss est d'annuler l'ensemble des coefficients de la premi ère colonne en dessous de la diagonale. Cela s'obtient si $a_{11} \neq 0$ en réalisant la transformation suivante sur la ligne $i > 1$:

$$a_{ij}^{(2)} = a_{ij}^{(1)} - g_{i1} a_{1j}^{(1)}, \quad j \in [1, n+1]$$

où $g_{i1} = \frac{a_{i1}^{(1)}}{a_{11}^{(1)}}$, on obtient le système équivalent à l'étape 2 suivant donné par sa matrice augment ée :

$$\tilde{A}^{(2)} = \left(\begin{array}{cccc|c} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} & b_1^{(1)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} & b_2^{(2)} \\ \cdots & \cdots & \ddots & \vdots & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} & b_n^{(2)} \end{array}\right)$$

les étapes suivantes consistent à refaire le même procédé pour les colonnes suivantes. Ainsi l'étape k consiste à éliminer l'inconnu x_k dans les équations $k+1, ..., n$. Ce qui donne les formules suivantes définies pour les lignes $i = k+1, ..., n$ en supposant que le $k^{ième}$ pivot $a_{kk}^{(k)} \neq 0$:

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - g_{ik} a_{kj}^{(k)}, \quad j \in [k, n+1] \tag{5.3}$$

avec $g_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}$. A la dernière étape c'est-à-dire à $k = n$, on obtient le syst ème équivalent suivant :

$$\tilde{A}^{(n)} = \left(\begin{array}{ccccc|c} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & \cdots & a_{1n}^{(1)} & b_1^{(1)} \\ 0 & a_{22}^{(2)} & \cdots & \cdots & a_{2n}^{(2)} & b_2^{(2)} \\ \cdots & \cdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1n-1}^{(n-1)} & a_{n-1n}^{(n-1)} & b_{n-1}^{(n-1)} \\ 0 & \cdots & \cdots & 0 & a_{nn}^{(n)} & b_n^{(n)} \end{array}\right)$$

la matrice U est donc définie comme étant la matrice $\tilde{A}^{(n)}$ et y le vecteur $b^{(n)}$. ■

R - La ligne i de la matrice $\tilde{A}^{(k)}$ n'est plus modifiée par l'algorithme dès lors que $i \leq k$. - A l'étape k, on pratique l'élimination sur une matrice de taille $n-k+1$ lignes et $n-k+2$ colonnes.

R Si lors de l'élimination l'élément $a_{kk}^{(k)}$ à l'é tape k est nul alors la ligne k ne peut pas être utilisée comme ligne pivot. Dans ce cas, on cherche une ligne $j > k$ telle $a_{jk}^{(k)} \neq 0$. Si une telle ligne existe, alors on permute la ligne j et la ligne k sinon le système n'admet pas de solution.

R Pour minimiser les erreurs d'arrondi, on choisit la valeur du pivot la plus grande en valeur absolue. Pour ce faire deux stratégie sont possibles :

1. La méthode dite à *pivot partiel* : Au $k^{ième}$ pas de l'élimination, on choisit comme ligne de pivot celle qui, parmi les $n-k+1$ restantes, a l'élément de module maximum en colonne et on permute dans $\tilde{A}^{(k)}$ la $k^{ième}$ ligne naturelle et celle qui réalise ce maximum.
2. La méthode dite à *pivot total* : Au $k^{ième}$ pas de l'élimination, on choisit comme pivot l'élément de plus grand module dans la matrice d'ordre $n-k+1$ restante. On permute donc dans $\tilde{A}^{(k)}$ la $k^{ième}$ colonne naturelle et celle du pivot, ce qui modifiera l'ordre des composantes du résultat. A la fin du processus, il ne faudra pas oublier de remettre dans l'ordre initial les composantes de la solution x.

5.2.1 Interprétation matricielle de la méthode de Gauss

Supposons que l'on puisse effectuer l'élimination sans permutation des lignes et des colonnes. Considérons alors les matrices

$$G^{(k)} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & & & \vdots \\ \vdots & \vdots & 1 & & & \vdots \\ 0 & \cdots & -g_{k+1k} & 1 & & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -g_{nk} & 0 & \cdots & 1 \end{pmatrix}, \quad k = 1,2,...,n-1$$

avec

$$g_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, \quad \text{pour} \quad i = k+1,...,n.$$

le système (5.3) peut se mettre sous la forme

$$\tilde{A}^{(k+1)} = G^{(k)}\tilde{A}^{(k)}$$

ce qui donne

$$\tilde{A}^{(n)} = G^{(n-1)}.G^{(n-2)}.G^{(n-3)}.....G^{(1)}.\tilde{A}^{(1)}$$

avec

$$\tilde{A}^{(n)} = \left[A^{(n)}|b^{(n)}\right]$$

Posons

$$\begin{aligned} U &= A^{(n)} \\ L &= \left(G^{(n-1)}.G^{(n-2)}.G^{(n-3)}.....G^{(1)}\right)^{-1} \end{aligned}$$

U (pour Upper) est une matrice triangulaire supérieure et L (pour Lower) est une matrice triangulaire inférieure à diagonale unité . Donc nous avons écrit A sous la forme : $A = LU$ où

$$L = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ g_{21} & 1 & \ddots & \cdots & \cdots & \vdots \\ g_{31} & g_{32} & 1 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ g_{n1} & \cdots & \cdots & \cdots & \cdots & 1 \end{pmatrix}, \text{ et } U = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & \cdots & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & a_{n-1n-1}^{(n-1)} & a_{n-1n}^{(n-1)} \\ 0 & \cdots & \cdots & \cdots & 0 & a_{nn}^{(n)} \end{pmatrix}$$

nous sommes donc amenés à résoudre successivement les deux syst èmes triangulaires :

$$\begin{cases} Ly = b \\ Ux = y \end{cases} \quad \text{où } y = b^{(n)}$$

5.3 Méthodes LU

La première phase de la méthode de Gauss consistait à transformer le système $Ax = b$ en un système triangulaire $Ux = y$ avec U une matrice triangulaire supérieure. Supposons qu'aucune permutation n'ait été effectuée, on peut alors montrer que U et y ont été obtenus à partir de A et b en les multipliant par une même matrice R triangulaire et inversible, c'est-à-dire

$$U = RA \quad \text{et} \quad y = Rb$$

on a donc $A = R^{-1}U$. Et si on pose $L = R^{-1}$et $U = R$, on peut donc dé composer A en un produit de matrice triangulaire inférieure L et une matrice triangulaire supérieure U. La méthode de Gauss appartient donc à la classe des méthodes dites méthodes LU. Elles consistent à obtenir une décomposition de la matrice A du type LU et à résoudre le système triangulaire $Ly = b$ puis ensuite le système triangulaire $Ux = y$ (L et U étant supposé es inversibles).

$$Ax = b \Longleftrightarrow LUx = b \Leftrightarrow \begin{cases} Ly = b \\ Ux = y \end{cases}$$

5.3.1 Décomposition LU

Définition 5.3.1 Une matrice A non singulière, admet une factorisation triangulaire si il existe une matrice L triangulaire inférieure et une matrice U triangulaire supérieure telles que :

$$A = LU$$

Théoreme 5.3.1 Soit le système linéaire $Ax = b$, si au cours de l'élimination de Gauss de la matrice A, aucun pivot n'est nul alors il existe une matrice L triangulaire inférieure et une matrice U triangulaire supé rieure telles que :

$$A = LU$$

si de plus on impose $l_{kk} = 1$ alors la factorisation est unique.

La matrice U s'obtient en appliquant la méthode de Gauss tandis que la matrice L s'écrit de la manière suivante :

$$L = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ l_{21} & 1 & \cdots & 0 & 0 \\ l_{31} & l_{32} & \ddots & 0 & 0 \\ \vdots & \cdots & \ddots & 1 & 0 \\ l_{n1} & l_{n2} & \cdots & l_{nn-1} & 1 \end{pmatrix}$$

où pour $i > 1$ on a $l_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}$. Ainsi la matrice L est composée des facteurs multiplicatifs permettant d'annuler les éléments sous le pivot. Comme il existe des problèmes simples pour lesquelles un des pivots est nul, le théorème suivant permet d'étendre la factorisation LU à un cadre plus général.

Théoreme 5.3.2 Une condition nécessaire et suffisante pour qu'une matrice A inversible puisse se factoriser sous la forme $A = LU$ est que $\det(A_k) \neq 0$, $\forall k = 1, 2, ...n-1$. Où $A_k = (a_{ij})_{\substack{i=1,2,..,k \\ j=1,2,.....,k}}$.

Démonstration. 1) Si $A = LU$, $A_k = L_k U_k$ et si A est inversible, $\det U = \prod_{i=1}^{n} u_{ii} = \prod_{i=1}^{n} a_{ii}^{(i)} \neq 0$. Donc $\det(A_k) = \det(L_k).\det(U_k) = \det(U_k) = \prod_{i=1}^{k} a_{ii}^{(i)} \neq 0$. 2) Supposons que $\det(A_k) \neq 0 \ \ \forall k = 1, 2, ..., n-1$. Celà est vrai en particulier pour $k = 1$, donc $a_{11}^{(1)} = \det(A_1)$ et la première étape de l'élimination de Gauss est possible. Par récurrence, si on a obtenu $A^{(k)}$ pour $k \leq n-1$

$$A^{(k)} = G^{(k-1)}.G^{(k-2)}.G^{(k-3)}.....G^{(1)}.A^{(1)}$$

alors $\det(A_k^{(k)}) = \det(G_k^{(k-1)})...\det(G_k^{(1)})\det(A^{(1)}) = \det(A^{(1)}) \neq 0$ $A_k^{(k)}$ étant triangulaire on a $\prod_{i=1}^{k} a_{ii}^{(i)} \neq 0$ donc $a_{kk}^{(k)} \neq 0$, donc la $k^{ième}$ étape de l'élimination est possible. On obtiendra finalement A=LU. ■

Théoreme 5.3.3 — méthode à pivot partiel. Soit A une matrice carrée d'ordre n inversible, alors il existe une matrice de permutation P telle que les pivots de PA soient non nuls. Ainsi il existe deux matrices L et U telles que $PA = LU$.

R le système linéaire $Ax = b$ est équivalent au système $PAx = Pb$ et la résolution du système se fait selon les étapes suivantes :

1. Construire U, L et P,
1. Calculer Pb,
2. Résoudre $Ly = Pb$ (système triangulaire inférieur),
3. Résoudre $Ux = y$ (système triangulaire supérieur).

5.4 Méthode de Cholesky

Certains systèmes présentent des propriétés particuliè res. Les matrices associées à ces systèmes peuvent être sym ètriques, à bande, etc...La méthode de cholesky a pour but la r ésolution de systèmes linéaires pour lesquels la matrice associ ée est symétrique définie positive.

Définition 5.4.1 — Matrice symétrique. Soit A une matrice carrée d'ordre n. On dit que A est symé trique si on a

$$A = A^t.$$

Définition 5.4.2 — Matrice définie positive. Soit A une matrice carrée d'ordre n. On dit que A est définie positive si elle vérifie la condition suivante :

$$\forall x \in \mathbb{R}^n \text{ et } x \neq 0, \quad \langle Ax, x\rangle > 0$$

où $\langle .,. \rangle$ désigne le produit scalaire dans $\mathbb{R}^n$. C'est-à-dire :

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i, \quad \forall x, y \in \mathbb{R}^n$$

On définit la norme induite par :

$$\|x\| = \langle x, x\rangle^{\frac{1}{2}}$$

Proposition 5.4.1 Si A est une matrice symétrique définie positive alors :

1. $a_{ii} > 0$,
2. $a_{ij} < a_{ii} a_{jj} \quad \forall i \neq j$,
3. $\max_{j,k} |a_{jk}| < \max_i |a_{ii}|$.

Théoreme 5.4.2 Si la matrice A est une matrice carrée définie positive alors elle est inversible.

Corollaire 5.4.3 Si la matrice A est une matrice carrée définie positive alors le système linéaire $Ax = b$ où $x, b \in \mathbb{R}^n$ admet une solution et une seule.

Théoreme 5.4.4 Soit M une matrice carrée telle et non singulière alors la matrice $A = MM^t$ est symétrique définie positive.

■ **Exemple 5.1** Soit

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

alors

$$A = MM^t = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$$

est définie positive. ■

5.4.1 Factorisation de Cholesky

Théoreme 5.4.5 Soit A une matrice carrée d'ordre n symétrique définie positive, A peut alors se décomposer et de manière unique en

$$A = LL^t$$

où L est une matrice triangulaire inférieure avec des élé ments diagonaux positifs.

Ainsi ce théorème permet de déduire que la méthode de construction des matrices définies positives engendre en fait l'ensemble des matrices symétriques définies positives. Si A est une matrice symétrique définie positive alors le systè me $Ax = b$ peut être décomposé en $LL^t x = b$ et ce système peut se résoudre en résolvant les systèmes triangulaires :

$$\begin{cases} Ly &= b \\ L^t x &= y \end{cases}$$

5.4.2 Algorithme de décomposition de Cholesky

Soit A une matrice symétrique définie positive alors on a $A = LL^t$. Pour résoudre le système

$$Ax = b \tag{5.4}$$

le théorème précédent nous permet d'écrire (5.4) sous la forme $LL^t x = b$ avec L une matrice triangulaire inférieure inversible. On est donc amené à résoudre

$$\begin{cases} Ly &= b \\ L^t x &= y \end{cases}$$

Le problème consiste donc à construire explicitement la matrice $L = (l_{ij})$ triangulaire inférieure telle que

$$A = LL^t \quad \text{où} \quad A = (a_{ij})$$

ce qui équivaut à

$$a_{ij} = \sum_{k=1}^{j} l_{ik} l_{jk}, \quad j \leq i.$$

Soit :

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1n} \\ 0 & l_{22} & \cdots & l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & l_{nn} \end{pmatrix}$$

en remarquant que a_{ij} est le produit de la ligne i de L et la colonne j de L^t alors on a :

$$a_{i1} = \sum_{k=1}^{n} l_{ik} l_{1k} = l_{i1} l_{11} + l_{i2} l_{12} + \cdots + l_{in} l_{1n} = l_{i1} l_{11}$$

en particulier pour $i = 1$, on a $l_{11} = \sqrt{a_{11}}$ (l_{11} est bien positif). la connaissance de l_{11} permet de construire la première colonne de la matrice L car :

$$l_{i1} = \frac{a_{i1}}{l_{11}}$$

En raisonnant de la même manière pour la deuxième colonne de L , on a :

$$a_{i2} = \sum_{k=1}^{n} l_{ik} l_{2k} = l_{i1} l_{21} + l_{i2} l_{22}$$

en prenant $i = 2$ alors $a_{22} = l_{21}^2 + l_{22}^2$. D'où l'on tire

$$l_{22} = \sqrt{a_{22} - l_{21}^2}$$

ensuite on a :

$$l_{i2} = \frac{a_{i2} - l_{i1} l_{21}}{l_{22}} \qquad i = 3,4,...,n$$

On peut généraliser la procédure au calcul de la colonne j en supposant que les $(j-1)$ colonnes ont déjà été calculé e. Ainsi :

$$a_{ij} = \sum_{k=1}^{n} l_{ij} l_{1j} = l_{ij} l_{j1} + l_{ij} l_{j2} + \cdots + l_{ik} l_{jk} + \cdots + l_{in} l_{jj}$$

et seul l_{ij} et l_{jj} ne sont pas connus. Si on pose $i = j$, on obtient :

$$l_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2}$$

et par conséquent

$$l_{ij} = \frac{a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk}}{l_{jj}} \qquad i > j \qquad \begin{array}{rcl} j & = & 2,.....,n \\ i & = & j+1,...,n \end{array}$$

R La décomposition de A symétrique définie positive, sous la forme $A = LL^t$ est unique à une matrice diagonale unité près. C'est-à-dire si $A = LL^t = MM^t$, alors $M = DL$ avec D matrice diagonale telle que $d_{ii} = \pm 1$.

R La méthode de Cholesky permet de calculer $\det A$ par

$$\det A = \prod_{i=1}^{n} l_{ii}^2$$

5.5 SERIE D'EXERCICES

Exercice 5.1 Résoudre le système d'équations linéaires suivant :

$$\begin{cases} -3x_1 - x_2 & = & 5 \\ -2x_1 + x_2 + x_3 & = & 0 \\ 2x_1 - x_2 + 4x_3 & = & 15 \end{cases}$$

1. En appliquant les formules de Cramer.
2. En triangularisant la matrice du système associée par la méthode de Gauss.

■

Exercice 5.2 Résoudre le système d'équations linéaires suivant :

$$\begin{cases} 2x_1 + x_2 - 5x_3 + x_4 & = & 8 \\ x_1 - 3x_2 - 6x_4 & = & 9 \\ 2x_2 - x_3 + 2x_4 & = & -5 \\ x_1 + 4x_2 - 7x_3 + 6x_4 & = & 0 \end{cases}$$

En appliquant le principe de triangularisation de Gauss. ■

Exercice 5.3 Soit le système d'équations linéaires suivant :

$$\begin{cases} x_1 + 2x_3 & = & 2 \\ 5x_2 + 4x_3 & = & 0 \\ 2x_1 + 4x_2 + 14x_3 & = & 5 \end{cases}$$

1. Montrer que la matrice associée à ce système est définie positive.
2. Résoudre ce système en utilisant la méthode de Choleski.

■

Exercice 5.4 Soit les systèmes d'équations linéaires suivants :

$$\begin{cases} x_1 + x_2 + 2x_3 & = & 1 \\ 5x_1 + 5x_2 & = & 3 \\ 3x_1 + x_2 + x_3 & = & -2 \end{cases} \quad \text{et} \quad \begin{cases} x_1 + 4x_2 + x_3 + 3x_4 & = & 2 \\ -x_2 + 3x_3 - x_4 & = & 0 \\ 3x_1 + x_2 + 2x_4 & = & 1 \\ x_1 - 2x_2 + 5x_3 + x_4 & = & -2 \end{cases}$$

Effectuer la résolution en mettant, si cela est possible, la matrice associée à chacun de ces deux systèmes, sous forme d'un produit de deux matrices triangulaires de structures différentes.

■

5.6 METHODES INDIRECTES

Introduction

Les méthodes directes de résolution de systèmes linéaires fournissent une solution x au problème $Ax = b$ en un nombre fini d'op érations. Si l'ordre n de la matrice A est élevé, le nombre d'opérations est aussi élevé et de plus, le résultat obtenu n'est pas rigoureusement exact. Par ailleurs, il existe des cas où les structures du système liné aire ne sont pas tirés à profit par les méthodes directes. C'est par exemple le cas des systèmes où la matrice A est très creuse. C'est la raison pour laquelle, dans ce cas , on préfère utiliser des méthodes itératives. L'objectif est de construire une suite de

vecteurs $\{x^{(k)}\}_{k=1,2...,n}$ qui tend vers un vecteur $\bar{x}$, solution exacte du problème $Ax = b$. Souvent, on part d'une approximation $\{x^{(0)}\}$ de $\bar{x}$ obtenue en géné ral par une méthode directe.

5.6.1 Les méthodes itératives

L'objectif est de résoudre un système du type $Ax = b$. Pour cela, nous allons décomposer la matrice A en

$$A = M - N$$

de sorte que M soit inversible. Ainsi, le système devient :

$$Mx = Nx + b$$

et nous chercherons par récurrence une suite de vecteurs $x^{(i)}$ obtenu à partir d'un vecteur $x^{(0)}$ et de la relation

$$Mx^{(k+1)} = Nx^{(k)} + b$$

c'est-à-dire

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b$$

Cette relation est une relation de récurrence du premier ordre. Nous pouvons en déduire une relation reliant l'erreur $e^{(k)} = x^{(k)} - \bar{x}$ à $e^{(k-1)} = x^{(k-1)} - \bar{x}$:

$$M(x^{(k)} - \bar{x}) = N(x^{(k-1)} - \bar{x})$$

puisque $M\bar{x} = N\bar{x} + b$ et donc $e^{(k)} = M^{-1}Ne^{(k-1)}$ pour $k = 1, 2, ...$ Si on pose $B = M^{-1}N$, nous avons alors

$$e^{(k)} = Be^{(0)}$$

La convergence de la suite $x^{(k)}$ vers la solution $\bar{x}$ est donné par le proposition suivant :

Proposition 5.6.1 Le choix de la décomposition de A devra obeir aux règles suivantes :

R

Proposition 5.6.2
1. Le rayon spectral $\rho(M^{-1}N)$ doit être strictement infé rieur à 1.
2. La résiolution de $Mx^{(k)} = Nx^{(k-1)} + b$ doit être simple et n écessiter le moins d'opérations possibles
3. Pour obtenir la meilleure convergence, $\rho(M^{-1}N)$ doit être le plus petit possible.

On voit que la convergence dépend de la décomposition.

5.6.2 Différentes décomposition de A

On écrit la matrice A sous la forme

$$A = D + E + F$$

avec D la matrice diagonale suivante :

$$D = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

E la matrice triangulaire inférieure suivante

$$E = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ a_{21} & 0 & 0 & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn-1} & 0 \end{pmatrix}$$

et F la matrice triangulaire supérieure

$$F = \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \vdots & 0 & \ddots & \vdots \\ 0 & & \ddots & a_{n-1n} \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

Nous obtiendrons donc la décomposition $A = M - N$ à partir de diffé rents types de regroupement de ces matrices D, E et F.

5.6.3 Méthode de Jacobi

On pose

$$M = D \quad \text{et} \quad N = -(E+F)$$

ainsi, $B = M^{-1}N = D^{-1}(-E-F)$, ce qui implique :

$$x^{(k+1)} = D^{-1}(-E-F)x^{(k)} + D^{-1}b$$

si on exprime cette relation en fonction des éléments de la matrice A nous avons :

$$x_i^{(k+1)} = -\sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}}, \qquad i = 1,2,...,n$$

5.6.4 Méthode de Gauss-Seidel

Cette méthode utilise

$$M = D+E \quad \text{et} \quad N = -F$$

D'où

$$B = -(D+E)^{-1}F,$$

et alors on a :

$$x^{(k+1)} = -(D+E)^{-1}Fx^{(k)} + (D+E)^{-1}b$$

le calcul de l'inverse de $(D+E)$ peut être évité. Si on écrit $(D+E)x^{(k+1)} = -Fx^k + b$, on obtient

$$\sum_{j=1}^{n} a_{ij}x_j^{(k+1)} = -\sum_{j=i+1}^{n} a_{ij}x_j^{(k)} b_i,$$

d'où

$$x_i^{(k+1)} = -\frac{1}{a_{ii}}\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \frac{1}{a_{ii}}\sum_{j=i+1}^{i-1} a_{ij}x_j^{(k)} + \frac{b_i}{a_{ii}}, \qquad i = 1,2,...,n.$$

5.6.5 Méthode de relaxation

On donne un paramètre $\omega \in]0,2[$, appelé facteur de relaxation, et on pose

$$M = \frac{D}{\omega} + E \quad \text{et} \quad N = \left(\frac{1-\omega}{\omega}\right) D - F$$

et par conséquent

$$\left(\frac{D}{\omega} + E\right) x^{(k+1)} = \left(\left(\frac{1-\omega}{\omega}\right) D - F\right) x^{(k)} + b$$

d'où

$$x_i^{(k+1)} = (1-\omega)\, x_i^{(k)} + \frac{\omega}{a_{ii}} \left(- \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k)} + b_i \right) \qquad i = 1,2,...,n.$$

Comme on peut le constater, la méthode de Gauss-Seidel correspond à la méthode de relaxation pour $\omega = 1$.

5.7 Convergence des méthodes itératives

La convergence des méthodes itératives dépend fortement du rayon spectral de A, Nous étudions d'abord les propriétés de certaines matrices et la localisation de leurs valeurs propres.

Définition 5.7.1 Soit $A \in \mathscr{M}_{m,n}(\mathbb{R})$ une matrice. On définit la norme matricielle induite à partir de la norme vectorielle sur $\mathbb{R}^n$ par

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

Proposition 5.7.1 Soit A et B deux matrices telles que leur multiplication soit compatible alors on a :

$$\|AB\| \leq \|A\| \, \|B\|$$

pour toute norme induite.

Théoreme 5.7.2 — Gerschgorin-Hadamard. Les valeurs propres de la matrice A appartiennent à la reunion des n disques D_k pour $k = 1,2,...,n$ du plan complexe ($\lambda \in \cup_{k=1}^{n} D_k$ où D_k, appelé disque de Gerschgorin, est dé fini par :

$$|z - a_{kk}| \leq \sum_{\substack{j=1 \\ j \neq i}} |a_{kj}|$$

5.7.1 Cas général

On considère une méthode itérative définie comme :

$$\begin{cases} x^{(0)} & \text{donné} \\ x^{(k+1)} & = \quad Cx^{(k)} + D \end{cases}$$

Théoreme 5.7.3 Soit A une matrice carré d'ordre n, pour que $\lim_{k\to\infty} A^k = 0$, il faut et il suffit que $\rho(A) < 1$.

Théoreme 5.7.4 Si il existe une norme induite telle que $\|C\| < 1$ alors la méthode itérative décrite ci-dessus est convergente quelque soit $x^{(0)}$ et elle converge vers la solution de :

$$(I_d - C)x = D$$

Théoreme 5.7.5 Une condition nécessaire et suffisante de convergence de la méthode ci-dessus est que :

$$\rho(C) < 1$$

R la condition de convergence donnée par le rayon spectral n'est pas dé pendante de la norme induite, cependant elle peut être utile car le calcul du rayon spectral peut être difficile.

Cas des matrices à diagonale dominante

Définition 5.7.2 Une matrice est dite à diagonale dominante si :

$$\forall i, 1 \leq i \leq n, \quad |a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|$$

Théoreme 5.7.6 Si A est une matrice à diagonale strictement dominante, alors A est inversible et en outre, les méthodes de Jacobi et de Gauss-Seidel convergent.

Démonstration. si A est une matrice à diagonale strictement dominante, on montre que A est inversible en démontrant que 0 n'est pas une valeur propre (c'est-à-dire $KerA = 0$). Posons $B = M^{-1}N$ est soit λ et v tels que $Bv = \lambda v$ avec $v \neq 0$. Puisque l'on s'interesse à $\rho(B) < 1$, on s'interesse en fait à la plus grande valeur propre de plus grand module de B. Ainsi, on peut supposer que $\lambda \neq 0$. L'équation $Bv = \lambda v$ devient :

$$\left(M - \frac{1}{\lambda}N\right) v = 0$$

- Pour Jacobi ; l'équation devient :

$$\left(D + \frac{1}{\lambda}E + \frac{1}{\lambda}F\right) v = 0$$

soit $C = D + \frac{1}{\lambda}E + \frac{1}{\lambda}F$. si $|\lambda| \geq 1$, on aurait :

$$|c_{ii}| = |a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \geq \sum_{\substack{j=1 \\ j\neq i}}^{n} \left|\frac{a_{ij}}{\lambda}\right| = \sum_{\substack{j=1 \\ j\neq i}}^{n} |c_{ij}|$$

donc C serait à diagonale strictement dominante et par conséquent inversible. C inversible implique que $Cv = 0$ donc $v = 0$. Or $v \neq 0$, d'o ù la contradiction et donc on a bien $|\lambda| < 1$. - Pour Gauss-Seidel ; l'équation devient :

$$\left(D + E + \frac{1}{\lambda}F\right) v = 0$$

en posant encore $C = D + E + \frac{1}{\lambda}F$. et en supposant $|\lambda| \geq 1$, on aurait :

$$|c_{ii}| = |a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \geq \sum_{j<1}^{n} \left|\frac{a_{ij}}{\lambda}\right| + \sum_{j>1}^{n} \left|\frac{a_{ij}}{\lambda}\right| = \sum_{\substack{j=1 \\ j\neq i}}^{n} |c_{ij}|$$

et on obtient le même type de contradiction. ■

Cas des matrices symétriques définies positives

Théoreme 5.7.7 Si A est une matrice symétrique définie positive, alors les mé thodes de Gauss-Seidel et de relaxation pour ($\omega \in]0,2[$) convergent.

La convergence de la méthode est d'autant plus rapide que $\rho(M^{-1}N)$ est petit. Or cette matrice $B = M^{-1}N$ dépend de ω. Une é tude théorique des valeurs propres de B montre que l'allure de la courbe $\rho(B)$ en fonction de B est décroissante entre 0 et ω_{opt} et croissante entre ω_{opt} et 2. Par ailleurs, on a toujours $1 < \omega_{opt} < 2$. On a donc intérêt à choisir ω le plus proche possible de ω_{opt}.

La méthode de correction

Soit le vecteur reste en x défini comme :

$$r(x) = b - Ax$$

et $\left\{r^{(k)}\right\}$ le reste en $\left\{x^{(k)}\right\}$. On appelle également l'erreur en k le vecteur

$$e^{(k)} = x^{(k)} - \bar{x}$$

où $\bar{x}$ est la solution. si on a une approximation $\left\{x^{(0)}\right\}$ de x, la relation suivante est vérifiée :

$$Ae^{(0)} = A(x^{(0)} - \bar{x}) = A(x^{(0)}) - b = -r^{(0)}$$

ce qui signifie que $e^{(0)}$ est la solution du système $Ax = -r^{(0)}$ et théoriquement, on a $\bar{x} = x^{(0)} - e^{(0)}$. Pratiquement, en appliquant au système $Ax = -r^{(0)}$ la méthode directe qui nous a fourni $x^{(0)}$, on n'obtient pas directement $e^{(0)}$, mais une approximation $y^{(0)}$ de $e^{(0)}$. Si on pose $x^{(1)} = x^{(0)} - y^{(0)}$, $x^{(1)}$ est une nouvelle approximation de $\bar{x}$, en itérant les calculs précédents, on obtient :

$$Ae^{(1)} = A(x^{(1)} - \bar{x}) = A(x^{(1)}) - b = -r^{(1)}$$

la résolution du système $Ax = -r^{(1)}$ donnera une approximation $y^{(1)}$ de $e^{(1)}$, et une nouvelle approximation $x^{(2)}$ de $\bar{x}$:

$$x^{(2)} = x^{(1)} - y^{(1)} = x^{(0)} - y^{(0)} - y^{(1)}$$

Ces calculs peuvent être itérer autant de fois que nécessaire, pour s'arrêter lorsque le reste est suffisament petit. A la $k^{ième}$ itération, les relations suivantes sont vérifiées pour $y^{(k-1)}$ approximation de $e^{(k-1)}$:

$$x^{(k)} = x^{(k-1)} - y^{(k-1)} = x^{(0)} - \sum_{i=0}^{k-1} y^{(i)}$$

avec $y^{(i)}$ une approximation de $e^{(i)}$, solution de $Ax = -r^{(i)}$ et $i = 0,1,2,...,k-1$. Si nous nous arrêtons lorsque $k = N$, il est nécessaire de ré soudre $N+1$ systèmes linéaires : d'abord $Ax = b$, pour obtenir $x^{(0)}$ puis $Ax = -r^{(i)}$ et $i = 0,1,2,...,N-1$ afin d'obtenir $y^{(i)}$. Une fois la matrice A décomposée (en LU ou Cholesky), il s'agit donc de résoudre les systèmes $LUx = -r^{(i)}$ où $-r^{(i)}$ a été calculé par la relation $r^{(i)} = b - Ax^{(i)}$.

5.8 SERIE D'EXERCICES

Exercice 5.5 Résoudre le système d'équations linéaires suivant :

$$\begin{cases} 10x_1 - 2x_2 - 2x_3 &= 6 \\ -x_1 + 10x_2 - 2x_3 &= 7 \\ -x_1 - x_2 + 10x_3 &= 8 \end{cases}$$

Par la méthode des approximations successives. Arrêter les calculs dès que :

$$\left| x_i^{(k+1)} - x_i^{(k)} \right| < 10^{-2}$$

■

Exercice 5.6 Résoudre le système d'équations linéaires suivant :

$$\begin{cases} 10x_1 - 2x_2 - 2x_3 &= 6 \\ -x_1 + 10x_2 - 2x_3 &= 7 \\ -x_1 - x_2 + 10x_3 &= 8 \end{cases}$$

Par la méthode de Seidel. Arrêter les calculs dès que :

$$\left| x_i^{(k+1)} - x_i^{(k)} \right| < 10^{-2}$$

■

Exercice 5.7 Résoudre le système d'équations linéaires suivant :

$$\begin{cases} 10x_1 - 2x_2 - 2x_3 &= 6 \\ -x_1 + 10x_2 - 2x_3 &= 7 \\ -x_1 - x_2 + 10x_3 &= 8 \end{cases}$$

Par la méthode de relaxation. Faire les calculs avec deux décimales. ■

Exercice 5.8 Résoudre le système d'équations linéaires suivant :

$$\begin{cases} 10x_1 + x_2 + x_3 &= 12 \\ 2x_1 + 10x_2 + x_3 &= 13 \\ 2x_1 + 2x_2 + 10x_3 &= 14 \end{cases}$$

Par la méthode de relaxation. Faire les calculs avec quatre décimales. ■