

Analyse Numerique

David Wiedemann

Table des matières

1	Representation de nombres en arithmetique finie	3
1.1	Representation des nombres dans les ordinateurs	3
1.2	Approximation de \mathbb{R} dans $\mathcal{F}(2, 53, -1021, 1024)$	3
1.3	Operations dans \mathcal{F}	4
1.4	Parentese sur le concept de stabilite	4
2	Integration Numerique	4
2.1	Formules d'integration de Newton-Cotes	5
2.2	Formules de quadrature d'ordre optimal	8
2.3	Noeuds d'integration optimaux : Formule de Gauss	9
2.4	Etude d'erreur des formules de quadrature	11
3	Interpolation de fonctions	12
3.1	Polynomes de Lagrange	12
3.2	Interpolation sur les points de Chebyshev	14
3.3	Approximation par des polynomes dans la norme L^2	16
3.4	Erreurs d'arrondissement	17
3.5	Interpolation par polynomes par parties	18
3.6	Approximation dans la norme L^2	19
4	Resolution de systemes lineaires	19
4.1	Quelques rappels sur les matrices	20
4.2	Methodes iteratives pour la resolution de systemes lineaires	20

List of Theorems

2	Proposition	3
1	Definition	4
2	Definition (Formule de Quadrature)	5
3	Definition	6
4	Theorème	6
7	Theorème (Thm. fondamental de la theorie de l'integration)	8

8	Lemme	9
4	Définition (Polynomes de Legendre)	9
9	Théorème (Forme des polynomes de Legendre)	9
10	Théorème	10
11	Lemme	10
5	Définition	10
12	Théorème (Erreurs dans les formules de quadrature)	11
13	Théorème (Theoreme de Weierstrass)	12
6	Définition	12
14	Proposition	13
16	Théorème (Representation de l'erreur)	13
17	Théorème	14
18	Théorème	14
19	Proposition	15
20	Théorème	15
7	Définition (Lebesgue constant)	17
23	Théorème	18
24	Théorème (Behaviour of lebesgue constant)	18
8	Définition	18
25	Théorème	18
26	Théorème	19
9	Définition (matrice creuse)	20
27	Proposition	20
10	Définition	20
28	Théorème	21

Lecture 1: Representation de nombres en arithmetique finie

Thu 03 Mar

1 Representation de nombres en arithmetique finie

Notons $\mathcal{F}(\beta, t, L, U)$ l'ensemble des nombres representables sous la forme $(-1)^s(0, \alpha_1 \dots \alpha_t)_\beta \beta^e$ ou e est l'exposant, $L \leq e \leq U, 0 \leq \alpha_i < \beta, \alpha_1, \dots, \alpha_t$ est la mantisse et s le signe.

Cette representation est la representation floating point.

1.1 Representation des nombres dans les ordinateurs

On appelle les nombres en double precision l'ensemble

$$\mathcal{F}(2, 53, -1021, 1024)$$

Bien que les valeurs maximales et minimales sont tres grandes ($2 \cdot 10^{-308}$ et $2 \cdot 10^{308}$), mais on en saute beaucoup.

Tous les nombres dans \mathcal{F} sont de la forme $\frac{p}{2^n}, p \in \mathbb{N}$.

On regarde la distance entre deux nombres consecutifs de \mathcal{F} .

Pour un exposant fixe, $[2^p, 2^{p+1}]$, le premier nombre apres 2^p est

$$(0.10 \dots 01)2^{p+1} = 2^p + 2^{p+1-t}$$

Donc dans ce cas, on a que le spacing est donne par 2^{p-52} .

Remarque

Si on a que des entiers dans un intervalle $[\beta^p, \beta^{p+1}]$, alors $\beta^{p+1-t} = 1$.

1.2 Approximation de \mathbb{R} dans $\mathcal{F}(2, 53, -1021, 1024)$

Soit $x \in \mathbb{R}$, on appelle $fl(x) \in \mathcal{F}(2, 53, -1021, 1024)$.

Notons $x = (-1)^s(0, \alpha_1 \dots \alpha_{t-1} \alpha_t \alpha_{t+1} \dots)_\beta \beta^e$, on definit alors

$$fl(x) = (-1)^s(0, \alpha_1 \dots \alpha_{t-1} \tilde{\alpha}_t)_\beta \beta^e$$

on fait l'hypothese ici que au moins un des α_i est non nul.

On veut borner $|x - fl(x)| \leq \frac{1}{2} \text{spacing} = \frac{1}{2} \beta^{e-t}$.

Bien que l'erreur absolue est, en principe, grande, l'erreur relative sera bornee, on a en effet

$$\frac{|x - fl(x)|}{|x|} \leq \frac{1}{2} \beta^{e-t} \frac{1}{|x|} \leq \frac{1}{2} \beta^{1-t} (\simeq 10^{-16} \text{ dans notre systeme})$$

On appelle cette erreur la "machine precision" et on la note u

Proposition 2

On peut egalement ecrire que

$$x \in \mathbb{R} \quad fl(x) = x(1 + \epsilon), |\epsilon| \leq u$$

1.3 Operations dans \mathcal{F}

Soit $x, y \in \mathbb{R}$, $x + y \mapsto fl[fl(x) + fl(y)]$, qu'elle est l'erreur relative commise ?

$$\frac{|fl[fl(x) + fl(y)] - (x + y)|}{|x + y|}$$

En utilisant la proposition ci-dessus, notons $fl(x) = x(1 + \epsilon_1)$, $fl(y) = y(1 + \epsilon_2)$, on a alors

$$\begin{aligned} |(x(1+\epsilon_1)+y(1+\epsilon_2))(1+\epsilon_3)-(x+y)| \cdot \frac{1}{|x+y|} &\leq \frac{x\epsilon_1 + y\epsilon_2 + \epsilon_3(x+y) - (x+y)}{|x+y|} + \text{petit} \\ &\leq \left(\frac{|x|}{|x+y|} + \frac{|y|}{|x+y|} + 1 \right) u \end{aligned}$$

On remarque que si $x > 0, y < 0$, il est possible de commettre une erreur tres grande.

On dit que la soustraction est une operation instable.

1.4 Parenthese sur le concept de stabilite

On veut resoudre $y = G(x)$.

Definition 1

La resolution de $y = G(x)$ est stable si une petite perturbation de x correspond a une petite perturbation de y , ie.

$$y + \delta y = G(x + \delta x)$$

On appelle alors le conditionnement absolu du probleme

$$\kappa_{abs} = \sup_{\delta x} \frac{\|\delta y\|}{\|\delta x\|}$$

Et on appelle perturbation relative du probleme

$$\kappa_{rel} = \sup_{\delta x} \frac{\|\delta y\| / \|y\|}{\|\delta x\| / \|x\|}$$

Lecture 2: Integration Numerique

Thu 10 Mar

2 Integration Numerique

On veut construire des algorithmes pour calculer de maniere approchee $\int_a^b f(x)dx$

2.1 Formules d'intégration de Newton-Cotes

On écrit

$$\int_a^b f(x)dx = \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} f(x)dx$$

Chacun des termes de la somme se réécrit comme

$$\int_{x_i}^{x_{i+1}} f(x)dx = \int_0^1 f(x_i + th_i)h_i dt$$

Et on trouve

$$\int_a^b f(x)dx = \sum_{i=1}^{N-1} h_i \int_0^1 f(x_i + th_i)dt$$

Ainsi, il suffit de trouver un algorithme pour calculer des intégrales de la forme $\int_0^1 g(t)dt$. La manière la plus naïve pour approximer cette intégrale serait de prendre $\int_0^1 g(t)dt \approx g(\frac{1}{2})$, et on note $Q_1^{nc}(g) = g(\frac{1}{2})$.

Une manière moins naïve de faire est d'approcher g par une fonction linéaire et de prendre l'approximation

$$\int_0^1 g(t)dt \approx \frac{1}{2} (g(0) + g(1)) = Q_2^{nc}(g) \text{ (formule de Newton-Cote a deux noeuds)}$$

ou encore

$$\int_0^1 \approx \frac{1}{6} (g(0) + 4g(\frac{1}{2}) + g(1)) = Q_3^{nc}(g) \text{ (formule de cote a trois noeuds ou formule de Simpson)}$$

De manière générale, on appelle formule de Newton-Cotes à S noeuds

$$\int_0^1 g(t)dt \approx \int_0^1 p(t)dt$$

où $p(t)$ est le polynôme de degré $s - 1$ passant par les points $(c_i, g(c_i))$, où $0 \leq c_1 \leq \dots \leq c_{s-1} < c_s \leq 1$.

Ainsi, de manière générale

$$Q_S^{nc}(g) = \sum_{i=1}^s b_i g(c_i)$$

où b_i sont les poids des formules de N.C.

On veut donc essayer de trouver des formules qui donnent les poids de l'intégration de Newton-Cotes.

Definition 2 (Formule de Quadrature)

Une formule de quadrature $Q_s(f)$ est donnée par n'importe quelle en-

semble de couples $(\{b_i\}_{i=1}^s, \{c_i\}_{i=1}^s) :$

$$Q_s(f) = \sum_{i=1}^N b_i f(c_i)$$

Definition 3

$Q_s(\cdot)$ est d'ordre s quand elle est exacte sur tout polynome de degre $\leq s-1$

Remarque

Par definition les formules Q_s^{nc} sont d'ordre s .

Theorème 4

Etant donne s noeuds distincts $\{c_i\}_{i=1}^N$, la formule donnee par $(\{b_i\}, \{c_i\})$ est d'ordre s si et seulement si les poids verifient

$$\sum_{i=1}^s c_i^{q-1} b_i = \frac{1}{q} \quad \forall q = 1, \dots, s$$

Preuve

Supposons que Q est d'ordre s , alors prenons

$$p(t) = t^q \quad q = 1 \dots s$$

On ecrit

$$\int_0^1 p(t) dt = \int_0^1 t^{q-1} dt = \frac{1}{q}$$

d'autre part

$$\sum_{i=1}^s b_i p(c_i) = \sum_{i=1}^s b_i p(c_i) = \sum_{i=1}^s b_i c_i^{q-1}$$

Dans l'autre sens, si $\sum_{i=1}^s c_i^{q-1} b_i = \frac{1}{q}$, alors la formule est exacte sur tout monome (par le raisonnement ci-dessus), par linearite, elle sera donc exacte sur n'importe quel polynome. \square

On montre maintenant qu'en fait les poids b_i sont uniques étant donné les c_i , en effet, étant donné le théorème ci-dessus, on a

$$\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ c_1 & c_2 & c_3 & \dots & c_s \\ c_1^2 & c_2^2 & c_3^2 & \dots & c_s^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_1^{s-1} & c_2^{s-1} & c_3^{s-1} & \dots & c_s^{s-1} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ \vdots \\ \vdots \\ \vdots \\ b_s \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1}{2} \\ \frac{1}{3} \\ \vdots \\ \frac{1}{s} \end{pmatrix}$$

Ainsi, soit la matrice A ci-dessus est inversible, alors il y a un seul choix de poids pour la formule de N.C.

Par un théorème d'algèbre linéaire, la matrice est inversible. En appliquant donc ceci à une fonction f générale, on trouve

$$\begin{aligned} \int_a^b f(x) dx &= \sum_{j=0}^{N-1} \int_{x_j}^{x_{j+1}} f(x) dx = \sum_{j=0}^{N-1} h_j \int_0^1 f(x_j + th_j) dt \\ &= \sum_{j=0}^{N-1} h_j Q_s^{nc}(f(x_j + th_j)) = \sum_{j=0}^{N-1} h_j \sum_{i=1}^s b_i f(x_j + c_i h_j) \end{aligned}$$

Remarque

Pour les noeuds c_i fixes, il existe un seul choix de poids qui garantit que Q_s est d'ordre s .

Quel est le choix optimal des noeuds ?

- **Choix 1** Choisir des noeuds équidistants.

Ce choix rend le calcul instable en arithmétique finie.

En effet, supposons qu'on veut intégrer $f(x) > 0$, on aura $\sum_{i=1}^s f(ih)b_i$.

Alors les poids oscillent fortement.

- **Choix 2** On cherche à comprendre où placer les noeuds pour maximiser l'ordre de la formule.

Exemple

On considère à nouveau la formule de Simpson

$$Q_3^{nc}(g) = \frac{1}{6} \left[g(0) + 4g\left(\frac{1}{2}\right) + g(1) \right]$$

Ainsi, pour $c_i = 0, \frac{1}{2}, 1$ on a les poids $b_i = \frac{1}{6}, \frac{2}{3}, \frac{1}{6}$. Est-ce que cette formule est d'ordre 4 ?

$$\int_0^1 t^3 dt = \frac{1}{4} = \sum_i b_i c_i^3 = \frac{1}{4} \text{ (en substituant les valeurs)}$$

Est-elle aussi d'ordre 5 ?

$$\int_0^1 t^4 dt = \frac{1}{5} = \sum_i b_i c_i^4 = \frac{2}{3} \frac{1}{16} + \frac{1}{6} \neq \frac{1}{5}$$

2.2 Formules de quadrature d'ordre optimal

On veut donc choisir des noeuds c_1, \dots, c_s pour maximiser l'ordre de la formule de quadrature

Theorème 7 (Thm. fondamental de la theorie de l'integration)

Soit $(\{b_i\}, \{c_i\})$ une formule de quadrature d'ordre s , $Q_s(\cdot)$.

Soit $M(t) = (t - c_1)(t - c_2) \dots (t - c_s)$, alors la formule $Q_s(\cdot)$ est d'ordre $p \geq s + m$ si et seulement si

$$\int_0^1 M(t)g(t)dt = 0$$

Preuve

Soit $f(t)$ un polynome de degre $s + m - 1$, prenons $r(t)$ un polynome de degre $s - 1$ passant par les points $(c_i, f(c_i))$.

Alors $f(t) - r(t)$ est un polynome de degre $s + m - 1$ est un polynome s'annulant sur tous les noeuds.

Ainsi

$$f(t) - r(t) = M(t)g_f(t) \text{ avec } \deg g_f \leq m - 1$$

\Leftarrow

Supposons que $\int_0^1 M(t)g(t)dt = 0 \forall$ polynome $g(t) : \deg g \leq m - 1$.

On demontre que la formule est d'ordre $s + m - 1$.

Soit f un polynome $\deg f \leq s + m - 1$, on peut donc ecrire

$$f(t) = r(t) + \underbrace{\int_0^1 M(t)g_f(t)dt}_{=0}$$

De meme, on a que

$$Q_s(f) = \sum_{i=1}^s b_i f(c_i) = \sum_{i=1}^s b_i \left[r(c_i) + \underbrace{M(c_i)g_f(c_i)}_{=0} \right] = \int_0^1 r(t)dt$$

Et donc la formule est exacte

\Rightarrow

Supposons que la formule est d'ordre $s + m$, demontrons que $\int_0^1 M(t)g(t)dt = 0 \forall g, \deg g \leq m - 1$, ainsi

$$\int_0^1 M(t)g(t)dt = \sum_{i=1}^s b_i M(c_i)g(c_i) = 0$$

□

Lecture 3: Integration Numerique

Thu 17 Mar

Lemme 8

Si une formule à s noeuds est d'ordre p , alors $p \leq 2s$

Preuve

Supposons que $p = 2s + 1$, si Q_s est d'ordre $2s + 1$, par le theoreme fondamental, ceci implique que

$$\int_0^1 M(t)g(t) dt = 0 \quad \forall g(t) : \deg g \leq s$$

Ainsi, en particulier pour $g(t) = M(t)$ on a

$$\int_0^1 M(t)^2 dt = 0$$

□

et donc $M(t) = 0$

On se demande maintenant si on peut trouver la valeur des noeuds de maniere facile ?

2.3 Noeuds d'integration optimaux : Formule de Gauss

Definition 4 (Polynomes de Legendre)

On considere la suite de polynomes $\{p_k\}_{k=0,\dots,n}$, avec $\deg p_k = k$ et $\int_{-1}^1 p_k(x) \cdot g(x) dx = 0 \quad \forall g(x) \deg g \leq k-1$

Theorème 9 (Forme des polynomes de Legendre)

Les polynomes de Legendre ont la forme

$$p_k(x) = \frac{1}{2^k k!} \frac{d^k}{dx^k} [(x^2 - 1)^k]$$

Preuve

On veut montrer que

$$\int_{-1}^1 p_k(x)g(x)dx = 0 \quad \forall g \quad \deg g \leq k-1$$

$$\int_{-1}^1 \frac{d^k}{dx^k} [(x^2 - 1)^k] g(x) dx = - \int_{-1}^1 \frac{d^k}{dx^{k-1}} [(x^2 - 1)^k] \frac{d}{dx} g(x) dx \left[\frac{d^{k-1}}{dx^{k-1}} [(x^2 - 1)^k] \cdot g \right]_{-1}^1$$

$$= (-1)^k \int_{-1}^1 (x^2 \ominus 1)^k \underbrace{\frac{d^k}{dx^k} g(x)}_{=0}$$

Theorème 10

Toutes les racines de P_k sont reelles, distinctes et dans l'intervalle $(-1, 1)$.

Preuve

Par l'absurde supposons qu'il y a τ_1, \dots, τ_r racines distinctes de $p_k(x)$ dans l'intervalle $(-1, 1)$, $r < k$. Ainsi $g(x) = (x - \tau_1) \dots (x - \tau_r)$ $\deg g \leq k - 1$

Par hypothese, on a donc

$$\int_{-1}^1 p_k(x)g(x) = \int_{-1}^1 qg^2$$

Or q ne change pas de signe, donc l'integrale ne peut pas etre nulle. \square

Lemme 11

Les polynomes de Legendre se calculent par

$$(k+1)P_{k+1} = (2k+1)xP_k - kP_{k-1}$$

On cherchait c_1, \dots, c_s tel que $\deg M = s$ et tel que

$$\int_0^1 M(t)g(t) = 0 \forall g : \deg g \leq s-1$$

Choisissons donc

$$M(t) = P_s(2t-1)$$

En effet

$$\int_0^1 M(t)g(t)dt = \int_{-1}^1 P_s(x)g\left(\frac{x}{2}+1\right)\frac{1}{2}dx$$

On a que $P_s(2t-1)$ a aussi s racines distinctes dans l'intervalle $(0, 1)$.

Ces racines sont les deux d'integration optimaux.

Definition 5

La formule de quadrature $(\{b_i\}, \{c_i\})$ avec c_i choisis comme racines de $P_s(2t-1)$ et b_i les poids corresponndants s'appelle formule de quadrature de Gauss.

2.4 Etude d'erreur des formules de quadrature

Theorème 12 (Erreurs dans les formules de quadrature)

Soit $f \in C^r([a, b])$, $r \geq p$.

Soit $Q_s(\cdot)$ une formule de quadrature d'ordre p .

$$I_n(f) = \sum_{j=0}^{n-1} h_j \sum_{i=1}^s b_i f(x_j + c_i h_j)$$

On a alors que

$$\left| \int_a^b f(x) dx - I_n(f) \right| \leq C \frac{h^p}{p!} \max_{x \in [a, b]} |f^{(p)}(x)|$$

ou $h = \max_j h_j$ et C ne depend ni de f , ni de p ni de h , mais depend de

$$\frac{\max h_i}{\min h_i}$$

Preuve

Dans cette demonstration, C indiquera une constante generique qui ne depend pas de h, f, p .

On definit

$$E_n(f) = \left| \sum_{j=0}^{h-1} \int_0^1 f(x_j + h_j t) dt - \sum_{i=1}^s b_i f(x_j + h_j c_i) \right|$$

Posons $g(t) = f(x_j + h_j t)$ et

$$E_h^j(f) = \left| \int_0^1 g(t) dt - \sum_{i=1}^s b_i g(c_i) \right| (= E(g))$$

Supposons d'abord que $g(x)$ est une fonction entiere, alors

$$g(t) = \sum_{r=0}^{p-1} \frac{g^{(r)}(0)}{r!} t^r + \sum_{r \geq p} \frac{g^{(r)}(0)}{r!} t^r$$

La formule de quadrature est exacte sur la premiere partie, ainsi

$$\begin{aligned} E(g) &= \left| \int_0^1 \sum_{r \geq p} \frac{g^{(r)}(0)}{r!} t^r - \sum_{r \geq p} \sum_{i=1}^s b_i \frac{g^{(r)}(0)}{r!} c_i^r \right| \\ &= \left| \sum_{r \geq p} \frac{g^{(r)}(0)}{r!} \underbrace{\left[\frac{1}{r+1} - \sum_{i=1}^s b_i c_i^r \right]}_{=C_r} \right| \end{aligned}$$

$$= c_p \frac{g^{(p)}(0)}{p!} + \text{"reste"}$$

On a

$$g^{(p)}(0) = (f(x_j + th_j)^{(p)})|_{t=0} = h_j^p \cdot f^{(p)}(x_j)$$

On peut aussi montrer que

$$c_p = \left| \frac{1}{p+1} - \sum b_i c_i^p \right| \leq 2$$

$$\text{Ainsi } E_n^j(f) \leq 2 \frac{1}{p!} h_j^p |f^{(p)}(x_j)|$$

□

Lecture 4: Interpolation de fonctions

Thu 24 Mar

3 Interpolation de fonctions

3.1 Polynomes de Lagrange

On considère le problème d'interpolation à l'aide de polynômes.

Theorème 13 (Theoreme de Weierstrass)

Soit $f \in C^0([a, b])$ alors il existe un polynome p_n de degré n qui vérifie

$$\lim_{n \rightarrow +\infty} \|f - p_n\| = 0$$

Pour la norme L^∞ .

Étant donné $f(x_0), \dots, f(x_n)$, on cherche un polynôme de degré n qui approxime $f(x)$.

Définition 6

Étant donné une partition de $[a, b]$ x_0, \dots, x_n .

On appelle $\{l_i(x)\}$ les polynômes de Lagrange, les polynômes $l_i(x)$ tels que

$$l_i(x_j) = \delta_{ij}, l_i \in \mathbb{P}_n$$

En général, on a

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

Ainsi, on peut considérer

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

comme polynome interpolant et on remarque que $p_n(x_j) = f(x_j)$.
On se demande donc maintenant pour $f \in C^k([a, b])$, $k > 0$ si on peut borner $\|f - p_n\|$ par une quantité dépendant de n .

Proposition 14

Etant donné une partition x_i .

Soit $d_n(x)$ une fonction de classe $C^n([a, b])$ tel que $d_n(x_i) = 0 \forall x_i$ de la partition.

Alors $\exists \xi \in (x_0, x_n)$ tel que $d_n^{(n)}(\xi) = 0$

Remarque

Si f est régulière, alors $f(x) - p_n(x)$ est régulière et $f(x_i) - p_n(x_i) = 0$

Preuve

On doit appliquer le théorème de Rolle n fois.

En effet, on a $d_n(x_0) = d_n(x_1) = 0$ et donc $\exists y_0$ tel que $d'(y_0) = 0$ et de manière générale, on a

$$d_n(x_i) = d_n(x_{i+1}) = 0 \implies \exists y_i \text{ tel que } d'(y_i) = 0$$

On réapplique le théorème de Rolle à y_1, \dots, y_n

□

Théorème 16 (Représentation de l'erreur)

Soit $f \in C^{n+1}([a, b])$ et soit p_n le polynôme d'interpolation de f sur la partition (x_0, \dots, x_n) alors $\forall x \in [a, b] \exists \xi \in (a, b)$:

$$f(x) - p_n(x) = f^{(n+1)}(\xi) \cdot \pi_n(x)$$

$$\text{ou } \pi_n(x) = \frac{1}{(n+1)!} (x - x_0) \dots (x - x_n)$$

Preuve

On va démontrer le résultat pour tout point $x \in [a, b]$.

Si $x = x_i$, alors $f(x_i) - p_n(x_i) = f^{(n+1)}(\xi) \cdot 0$ ce qui est toujours vrai.

Donc, si $x \neq x_i$, alors $\pi_n(\bar{x}) \neq 0$.

Donc $\exists \eta \in \mathbb{R} : f(x) - p_n(x) = \eta \pi_n(x)$.

On peut donc prendre $d_{n+1}(x) = f(x) - p_n(x) - \eta \pi_n(x)$, alors d_{n+1} s'annule sur les x_i et sur \bar{x} .

On peut donc appliquer la proposition d'avant à d_{n+1} ,

$$\exists \xi : d_{n+1}^{(n+1)}(\xi) = 0$$

Ainsi

$$d_{n+1}^{(n+1)} = f^{(n+1)}(x) - 0 - \eta \underbrace{\frac{d^{(n+1)}}{dx^{n+1}} \pi_n}_{=1}$$

□

Et donc il existe ξ tel que $f^{(n+1)}(\xi) - \eta = 0$

On va essayer d'utiliser la representation de l'erreur pour trouver une estimation de l'erreur

En effet

$$\|f(x) - p_n(x)\| = \max_{x \in [a,b]} \|f^{(n+1)}(\xi) \pi_n(x)\| \leq \|f^{(n+1)}(x)\| \|\pi_n\|$$

On a

$$\|\pi_n\| = \left\| \frac{1}{(n+1)!} (x - x_0) \dots (x - x_{n-1}) \right\| \leq \frac{1}{(n+1)!} (b - a)^{n+1}$$

Ainsi

$$\|f - p_n\| \leq \frac{1}{4} \frac{1}{(n+1)!} (b - a)^{n+1} \|f^{(n+1)}\|$$

Pour quelle classe de fonctions puis-je donc deduire que $\lim_{n \rightarrow +\infty} \|f - p_n\| = 0$?

Clairment $f(x) = \frac{1}{1+x^2}$ n'appartient pas a cette classe.

Theorème 17

Soit x_0, \dots, x_n une partition equidistante de l'intervalle $[a, b]$ et soit $f : [-\alpha, \alpha] \rightarrow \mathbb{R}$ une fonction analytique.

Si f admet un developpement en serie entiere en x_0 de rayong R avec $R > 3\alpha$, alors $\lim_{n \rightarrow +\infty} \|f - p_n\| = 0$

En effet si $R > 3\alpha \exists a \in \mathbb{R}^+, a < 1 \ \|f - p_n\| \leq C(R)a^{n+2}$

Lecture 5: qqchose

Thu 31 Mar

3.2 Interpolation sur les points de Chebyshev

En partant de la caracterisation de l'erreur d'interpolation

$$\|f - p^n\| \leq \max |f^{(n+1)}(\eta)| \max |\pi_n(x)|$$

Comment $\pi_n(x)$ depend des points choisis et quels sont les points minimisant $\|f - p^n\|$?

On se pose sur l'intervalle $[-1, 1]$, $p_n(x) = x^n + \sum_{i=1}^{n-1} a_i x^i$, quels sont les coefficients a_1, \dots, a_{n-1} tel que

$$\min_{p_n \in \mathbb{P}_n^1} \max_x |p_n(x)|$$

Theorème 18

Ces polynomes existent pour tout n et il sont de la forme $T_n(x) = \cos(n \arccos(x))$, $x \in [-1, 1]$

On procede par etapes

Proposition 19

Si $p_n \in \mathbb{P}_n^1$ minimise le problème ci-dessus, alors p_n prend la valeur $L = \max_x |p_n(x)|$ exactement $n + 1$ fois.

Preuve

On montre le cas $n = 3$.

Supposons que p_3 atteint le min seulement 3 fois, $p(x_1) = L, p(x_2) = -L, p(x_3) = L$.

Prenons q_2 tel que $q_2(x_1) > 0, q_2(x_2) < 0, q_2(x_3) > 0$.

Alors $p_3 - \epsilon q_2 \in \mathbb{P}_3^1$.

Alors $p_3 - \epsilon q_2$ a diminué sa valeur en x_1, x_2, x_3 mais donc le polynôme p_3 n'était pas minimisant. \square

Les polynômes de Chebyshev sont des polynômes

On vérifie juste quelques cas

$$T_0 = \cos 0 = 1$$

$$T_1 = \cos \arccos x = x$$

$$T_2 = \cos(2 \arccos x) = 2x^2 - 1$$

De plus, $T_n(x) \leq 1 \forall x$ et les racines de $T_n(x)$ sont $\cos(\frac{2k+1}{2n}\pi), k = 0, \dots, n-1$.

De plus, $T_n(x)$ atteint -1 et 1 exactement $n + 1$ fois.

Ainsi,

$$\min_{p_n \in \mathbb{P}_n^1} \max_{x \in [-1,1]} |p_n(x)| = \max_x |2^{-n} T_n(x)|$$

En revenant au problème d'interpolation

$$\|f - p^n\| \leq \max_{\eta \in [-1,1]} |f^{(n+1)}(\eta)| \max_x |\pi_n(x)|$$

Ainsi, en prenant $\pi_n(x) = \frac{1}{(n+1)!} 2^{-n} T_{n+1}(x)$.

Les points d'interpolation qui minimisent l'erreur sont donc les racines de $T_{n+1}(x)$

Théorème 20

Soit $f : [-1, 1] \rightarrow \mathbb{R}$, f Lipschitz, et si $p_n^c(x)$ est le polynôme interpolant de $f(x)$ sur les points de Chebyshev, alors

$$\lim_{n \rightarrow +\infty} \|f - p_n^c\| = 0$$

Remarque

On peut passer de $[a, b]$ vers $[-1, 1]$ à travers une transformation linéaire.

3.3 Approximation par des polynomes dans la norme L^2

Jusqu'ici, on a cherché à minimiser

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

On cherche maintenant un polynome p_n tel que $\|f - p_n\|_{L^2}$ est minimale.

On cherche donc p_n qui minimise $\int_a^b (f(x) - p_n(x))^2 dx$.

On écrit donc

$$p_n(x) = \sum_k \alpha_k p_k(x)$$

où les p_k sont des polynomes de Legendre.

On cherche donc p_n tel que

$$\int_a^b (f - p_n)^2 \leq \int_a^b (f - q_n)^2 \forall q_n$$

Ainsi,

$$\int_a^b (f - \sum_k \alpha_k p_k)^2 = \int_a^b f^2 - 2 \sum_k \alpha_k \int_a^b f p_k + \sum_k \sum_{k'} \alpha_k \alpha_{k'} \int_a^b p_k p_{k'}$$

Sauf que les polynomes de Legendre sont orthogonaux pour la norme L^2 et donc

$$\int_a^b (f - \sum_k \alpha_k p_k)^2 = \int_a^b f^2 - 2 \sum_k \alpha_k \int_a^b f p_k + \sum_k \alpha_k^2 \int_a^b p_k^2$$

On cherche donc les coefficients α_k tel que

$$\begin{aligned} \frac{d}{d\alpha_i} \left(\int_a^b (f - p_n)^2 \right) &= 0 \\ \iff -2 \int_a^b f p_i + 2\alpha_i \int_a^b p_i^2 &= 0 \end{aligned}$$

Ainsi, $\alpha_i = \frac{\int_a^b f p_i}{\int_a^b p_i^2}$.

Remarque

Pour calculer $p_n(x)$ on n'utilise pas une interpolation sur les points, mais on a besoin de connaître $\int f p_k$.

En général $\int_a^b f p_k$ ne peut pas être calculée exactement, donc on peut écrire

$$\int_a^b f p_k = Q(f p_k)$$

Soit x_0, \dots, x_n $n+1$ points de Gauss sur l'intervalle $[a, b]$, alors

$$\int_a^b f p_k \simeq \sum_i f(x_i) p_k(x_i) c_i$$

Il s'agit d'une integrale approche mais c'est le mieux qu'on puisse faire avec $n + 1$ approximations.

Ainsi, on obtient un polynome moins optimal

$$\tilde{p}_n = \sum_{k=0}^n \tilde{\alpha}_k p_k$$

ou \tilde{p}_n est calculee grace a $n + 1$ evaluations de f .

Quel est alors le comportement asymptotique de $\|f - \tilde{p}_n\|_{L^2}$ par rapport a n

Lecture 6: Effets des erreurs d'arrondissement

Thu 07 Apr

3.4 Erreurs d'arrondissement

En realite, lorsqu'on interpole en pratique, a chaque etape de calcul, on com-
met une erreur d'arrondissement.

On remarque donc par exemple que, lorsqu'on interpole sur des noeuds equi-
distants une fonction telle que $\sin x$ des grandes erreurs aux bords, meme si l'on
s'attendrait a obtenir une convergence uniforme.

Donc on pose $\hat{f}(x_i) = f(x_i)(1 + \epsilon)$ avec ϵ une certaine erreur machine et on veut
etudier l'erreur due au "round off".

En substituant cette valeur dans les valeurs de p_n on obtient

$$\hat{p}_n := \sum_i \hat{f}(x_i) l_i(x)$$

Ou les l_i sont les polynomes interpolant.

On peut donc calculer la difference entre p_n et \hat{p}_n

$$|p_n - \hat{p}_n| \leq \sum_i |\epsilon f(x_i) l_i(x)| \leq \epsilon \|f\|_\infty \sum_i |l_i(x)|$$

Ceci motive la definition suivante

Definition 7 (Lebesgue constant)

$$\Lambda_n = \max_x \sum_i |l_i(x)|$$

et clairement Λ_n va dependre du choix des x_i .

Le calcul ci-dessus montre que

Theorème 23

Soit $n \in \mathbb{N}$, $f \in C^0([a, b])$ et p_n les polynomes d'interpolations de Lagrange, alors

$$\|p_n - \hat{p} - n\|_\infty \leq \epsilon \Lambda_n \|f\|_\infty$$

Donc pour contrôler l'erreur, il nous faut contrôler Λ_n , en fait

Theorème 24 (Behaviour of lebesgue constant)

— Si les noeuds sont equidistants, alors

$$\Lambda_n \approx \frac{2^{n+1}}{\epsilon n \log n} \text{ quand } n \rightarrow \infty$$

— Pour les points de chebychev, on a

$$\lambda_n \approx \frac{2}{\pi} \log n \text{ quand } n \rightarrow \infty$$

Mais meme dans le cas des noeuds optimaux, on voit que l'erreur va tout de meme tendre vers l'infini.

On essaie donc d'approximer les fonctions par des fonctions lineaires.

3.5 Interpolation par polynomes par parties**Definition 8**

Pour un $N \in \mathbb{N}$ fixe et $s \in \mathbb{N}$. On considere $f \in C^0$ et une partition a_i d'un intervalle $[a, b]$. Pour chaque i , on construit $p^{(i)}$ le polynome d'interpolation de lagrange locale pour s points choisis dans $[a_i, a_{i+1})$. On recolle alors les $p^{(i)}$ en une fonction \tilde{p}_s

Et on a un theoreme qui nous borne l'erreur :

Theorème 25

Soit $N \in \mathbb{N}$, $N \geq 1$ et $s \in \mathbb{N}$, $f \in C^{s+1}([a, b])$ et \tilde{p}_s le polynome d'interpolation par parties sur une partition generale, alors

$$\|f - \tilde{p}_s\|_\infty \leq \frac{H^{s+1}}{4(s+1)!} \|f^{(s+1)}\|_\infty$$

ou $H := \max |a_{i+1} - a_i|$

Preuve

On a

$$\|f - \tilde{p}_s\|_\infty = \max_i \left\| f - p_s^{(i)} \right\|_{\infty, [a_i, a_{i+1})} \leq \frac{1}{4(s+1)!} H^{s+1} \left\| f^{(s+1)} \right\|_\infty \quad \square$$

3.6 Approximation dans la norme L^2

Etant donne f , on veut trouver le meilleur polynome p^* qui minimisera la distance dans la norme L^2 , ie.

$$p^* = \operatorname{argmin}_{q_n \in \mathbb{P}_n} \|f - q_n\|_2^2$$

Theorème 26

La solution optimale est donnee par

$$p^* = \sum_k \alpha_k p_k \text{ avec } \alpha_k = \frac{\int_a^b f p_k dt}{\int_a^b |p_k|^2 dt}$$

Lecture 7: Resolution de systemes lineaires

Fri 08 Apr

4 Resolution de systemes lineaires

On cherche a resoudre $Ax = b$, $A \in \mathbb{R}^{n \times n}$ inversible, $b \in \mathbb{R}^n$.

Pour n grand, la regle de Cramer donne des erreurs d'arrondi enorme.

On souhaite donc utiliser la regle de Gauss.

On cherche a ecrire $A = LU$, L triangulaire inferieure et U triangulaire superieure.

Si on obtient cette decomposition, on a

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

Ainsi, $Ax = b$ devient la resolution de deux systemes triangulaires.

Combien d'operations faut-il pour calculer $y = L^{-1}B$?

Par elimination directe, il faut $O(n^2)$ operations.

Ainsi, une fois que la decomposition $A = LU$ a ete faite, on peut toujours resoudre le systeme.

On se demande donc combien d'operations il faut pour calculer L et U ?

Cette decomposition n'est pas toujours possible, cependant, si A est inversible

$\exists P$ un pivoting tel que PA admet une decomposition LU .

Une matrice pivoting est de la forme $(e_{\sigma(1)} \dots e_{\sigma(n)})$.

Ainsi $PAx = Pb$ peut se resoudre par elimination de Gauss.

L'élimination de Gauss est un algorithme en $O(n^3)$ (donc tres couteux).

Donc on va essayer de trouver un algorithme plus efficace.

Il y a quelques autres problemes lies a LU tel que la stabilite.

Definition 9 (matrice creuse)

A est creuse si $\#nz(A) = O(n)$ ou $\#nz$ est le nombre de valeurs non nulles.

Si A est creuse, sa decomposition LU n'est pas creuse, donc cette methode n'est pas viable en general.

4.1 Quelques rappels sur les matrices

On definit

$$\|A\|_p = \max_x \frac{\|Ax\|}{\|x\|}$$

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}$$

et

$$\rho(A) = \max_i |\lambda_i|$$

Proposition 27

$\forall p$, on a

$$\|A\|_p \geq \rho(A)$$

et

$$\|A\|_2 = \rho(A)$$

Definition 10

Le conditionnement de A est defini par

$$K_p(A) = \|A\|_p \|A^{-1}\|_p$$

si A n'est pas inversible, le conditionnement de A est $+\infty$.

On definit de plus

$$K_2(A) = K(A) = \rho(A)\rho(A^{-1})$$

4.2 Methodes iteratives pour la resolution de systemes lineaires

On se contente de calculer une suite x^k une suite telle que $\lim_{k \rightarrow +\infty} x^k = x_*$ la solution.

On definit $A = P - (P - A)$ avec P une matrice a definir, alors

$$Px_* - (P - A)x_* = b$$

Alors

$$Px^{k+1} - (P - A)x^k = b \implies Px^{k+1} = (P - A)x^k + b$$

Qui est un systeme verifie pour x_* .

On pose $r^k = b - Ax^k$ et on a donc

$$\begin{cases} Px^{k+1} = Px^k + r^k \\ x^0 = x_0 \text{ a choisir} \end{cases}$$

Pour calculer x^{k+1} , on doit pouvoir inverser P et donc P doit etre simple a inverser.

La convergence $x^k \rightarrow x_*$ depend donc du choix de P .

Ecrivons $A = A_D + A_L + A_U$.

On peut alors choisir $P = A_D + A_L$ ou $P = A_D$, dans les deux cas P est simple a inverser.

Methode de Richarson statique

La methode statique s'ecrit $Px^{k+1} = Px^k + \alpha r^k$ pour α une constante et la methode dynamique $Px^{k+1} = Px^k + \alpha_k r^k$ ou les α_k sont a choisir.

Lecture 8: Methodes Iteratives

Thu 21 Apr

Theorème 28

Une methode iterative si et seulement si pour toute estimation initiale x_0 , le rayon spectral de $B := \text{Id} - P^{-1}A$ est inferieur a 1.