

ANALYSE NUMÉRIQUE

Prof. HAMRI Nasr-eddine
Département de Mathématiques
Université Abdelhafid Boussouf - Mila



TABLE DES MATIÈRES

1	NOTIONS D'ERREURS	7
1	PRÉLIMINAIRES	8
1.1	Exemples	9
2	Erreurs absolues et Erreurs relatives	9
2.1	Exemple	10
2.2	Exemples	10
3	PRINCIPALES SOURCES D'ERREURS	11
4	PRECISION, CHIFFRES SIGNIFICATIFS	11
4.1	Chiffres significatifs	11
5	Cumulation des erreurs d'arrondi	12
5.1	Erreurs d'arrondi sur une somme	12
5.2	Erreurs d'arrondi sur un produit	13
6	Représentation approchée des nombres réels	13
6.1	Nombres en virgule flottante	14
6.2	Non-associativité des opérations arithmétiques.	14
6.3	Phénomènes de compensation.	15
7	SERIE D'EXERCICES	15
2	APPROXIMATION	17
1	GÉNÉRALITÉS	17
2	APPROXIMATION	17
2.1	Meilleure approximation	17
3	APPROXIMATION AU SENS DES MOINDRES CARRÉS	19
4	CARACTÉRISATION	20
4.1	Norme	20
5	SERIE D'EXERCICES	22
3	INTERPOLATION POLYNOMIALE	23
1	GÉNÉRALITÉS	24
2	POLYNOME DE LAGRANGE	25
2.1	Cas où les points sont equidistants	27
3	Estimation de l'erreur dans l'interpolation de Lagrange	28
4	POLYNOME DE NEWTON	30
4.1	Différences finies	30
4.2	Différences divisées	31
4.3	Polynôme d'interpolation de Newton :	33
4.4	Erreur d'interpolation	33
4.5	Autre écriture du polynôme d'interpolation de Newton	34
5	INTERPOLATION CUBIQUE DE HERMITE	35
6	SERIE D'EXERCICES	35

4	INTEGRATION ET DÉRIVATION NUMÉRIQUE	37
1	INTÉGRATION NUMÉRIQUE	38
1.1	Méthode Générale	38
1.2	Approximation d'une intégrale	38
1.3	Utilisation de l'interpolation polynomiale	39
1.4	Etude de l'erreur d'intégration	40
1.5	Convergence des méthodes d'intégration	40
1.6	Formules de Newton Cotes	42
1.7	Formule de type fermé : des trapèzes et de Simpson	42
1.8	Formule de type ouvert :	43
1.9	Intégration par la méthode de Gauss	43
1.10	Calcul de $\int_a^b f(x)dx$	45
1.11	Erreur de l'intégration par la méthode de Gauss	45
2	SERIE D'EXERCICES	46
3	DÉRIVATION NUMÉRIQUE	48
3.1	Généralités :	48
3.2	Utilisation de l'interpolation polynomiale	49
3.3	Erreur de dérivation	50
3.4	Algorithmes de dérivation	53
3.5	Formules centrales de dérivation	55
3.6	Formules non centrales de dérivation	55
4	SERIE D'EXERCICES	56
5	RESOLUTION D'UN SYSTEME LINEAIRE	57
1	METHODES DIRECTES	57
1.1	Rappel	57
1.2	Systèmes linéaires	57
1.3	Résolution d'un système triangulaire supérieur	58
2	Méthode de Gauss	59
2.1	Interprétation matricielle de la méthode de Gauss	60
3	Méthodes LU	61
3.1	Décomposition LU	61
4	Méthode de Cholesky	62
4.1	Factorisation de Cholesky	63
4.2	Algorithme de décomposition de Cholesky	64
5	SERIE D'EXERCICES	65
6	METHODES INDIRECTES	66
6.1	Les méthodes itératives	66
6.2	Différentes décomposition de A	67
6.3	Méthode de Jacobi	67
6.4	Méthode de Gauss-Seidel	67
6.5	Méthode de relaxation	68
7	Convergence des méthodes itératives	68
7.1	Cas général	68
8	SERIE D'EXERCICES	70

6	CALCUL DES VALEURS PROPRES ET VECTEURS PROPRES	73
1	Introduction	74
2	RAPPELS	74
3	Calcul direct de $\det(A - \lambda I)$	74
4	Méthode de Krylov	74
5	MÉTHODE DE LEVERRIER	76
6	Valeurs et Vecteurs Propres	77
7	La condition du calcul des valeurs propres	77
	7.1 Condition du calcul des vecteurs propres	79
8	La méthode de la puissance	80
9	Méthode de la puissance inverse de Wielandt	81
10	VALEURS PROPRES ET VECTEURS PROPRES	82
11	LA CONDITION DU CALCUL DES VALEURS PROPRES	83
	11.1 Condition du calcul des vecteurs propres	85
12	LA METHODE DE LA PUISSANCE	86
13	METHODE DE LA PUISSANCE INVERSE DE WIELANDT	87
14	Transformation sous forme tridiagonale (ou de Hessenberg)	89
	14.1 a) A l'aide des transformations élémentaires	89
	14.2 b) A l'aide des transformations orthogonales	90
	14.3 Méthode de bisection pour des matrices tridiagonales	90
	14.4 Méthode de bisection.	92
15	L'itération orthogonale	92
	15.1 Généralisation de la méthode de la puissance (pour calculer les deux va- leurs propres dominantes).	93
	15.2 Méthode de la puissance (pour le calcul de toutes les valeurs propres)	94
	15.3 L'algorithm QR	95
	15.4 Accélération de la convergence	96
	15.5 Critère pour arrêter l'itération.	96
	15.6 Le "double shift" de Francis	97
	15.7 Etude de la convergence	98
16	Exercices	98
17	TRANSFORMATION SOUS FORME TRIDIAGONALE (ou de HESSENBERG)	101
	17.1 a) A l'aide des transformations élémentaires	102
	17.2 b) A l'aide des transformations orthogonales	102
	17.3 Méthode de bisection pour des matrices tridiagonales	103
	17.4 Méthode de bisection.	104
18	L'ITERATION ORTHOGONALE	105
	18.1 Généralisation de la méthode de la puissance (pour calculer les deux va- leurs propres dominantes).	105
	18.2 Méthode de la puissance (pour le calcul de toutes les valeurs propres)	107
	18.3 L'algorithm QR	107
	18.4 Accélération de la convergence	108
	18.5 Critère pour arrêter l'itération.	109
	18.6 Le "double shift" de Francis	110
	18.7 Etude de la convergence	111
19	Exercices	111

6 METHODES INDIRECTES

6.0.1 Introduction

Les méthodes directes de résolution de systèmes linéaires fournissent une solution x au problème $Ax = b$ en un nombre fini d'opérations. 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 linéaire 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 $\{x^{(k)}\}_{k=1,2,\dots,n}$ 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 général par une méthode directe.

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 $k = 1, 2, \dots$. 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 154. Le choix de la décomposition de A devra obéir aux règles suivantes :

Remarque 155.

Proposition 156. 1. Le rayon spectral $\rho(M^{-1}N)$ doit être strictement inférieur à 1.

2. La résolution de $Mx^{(k)} = Nx^{(k-1)} + b$ doit être simple et nécessiter 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.

6.2 Différentes décomposition de A

On écrit la matrice A sous la forme

$$A = D + E + F$$

avec D la matrice diagonale suivante :

$$D = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

E la matrice triangulaire inférieure suivante

$$E = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ a_{21} & 0 & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn-1} & 0 \end{pmatrix}$$

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 différents types de regroupement de ces matrices D, E et F .

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 :

$$x_i^{(k+1)} = - \sum_{\substack{j=1 \\ j \neq i}}^n \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}}, \quad i = 1, 2, \dots, n$$

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

$$x_i^{(k+1)} = -\frac{1}{a_{ii}} \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \frac{1}{a_{ii}} \sum_{j=i+1}^{i-1} a_{ij}x_j^{(k)} + \frac{b_i}{a_{ii}}, \quad i = 1, 2, \dots, n.$$

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

$$x_i^{(k+1)} = (1-\omega)x_i^{(k)} + \frac{\omega}{a_{ii}} \left(-\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)} + b_i \right) \quad i = 1, 2, \dots, n.$$

Comme on peut le constater, la méthode de Gauss-Seidel correspond à la méthode de relaxation pour $\omega = 1$.

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 157. Soit $A \in \mathcal{M}_{m,n}(\mathbb{R})$ 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 158. Soit A et B deux matrices telles que leur multiplication soit compatible alors on a :

$$\|AB\| \leq \|A\| \|B\|$$

pour toute norme induite.

Théorème 159 (Gerschgorin-Hadamard). Les valeurs propres de la matrice A appartiennent à la réunion des n disques D_k pour $k = 1, 2, \dots, n$ du plan complexe ($\lambda \in \cup_{k=1}^n D_k$ où D_k , appelé disque de Gerschgorin, est défini par :

$$|z - a_{kk}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{kj}|$$

7.1 Cas général

On considère une méthode itérative définie comme :

$$\begin{cases} x^{(0)} & \text{donné} \\ x^{(k+1)} & = Cx^{(k)} + D \end{cases}$$

Théorème 160. Soit A une matrice carré d'ordre n , pour que $\lim_{k \rightarrow \infty} A^k = 0$, il faut et il suffit que $\rho(A) < 1$.

Théorème 161. 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 :

$$(I_d - C)x = D$$

Théorème 162. Une condition nécessaire et suffisante de convergence de la méthode ci-dessus est que :

$$\rho(C) < 1$$

Remarque 163. la condition de convergence donnée par le rayon spectral n'est pas dépendante de la norme induite, cependant elle peut être utile car le calcul du rayon spectral peut être difficile.

7.1.1 Cas des matrices à diagonale dominante

Définition 164. Une matrice est dite à diagonale dominante si :

$$\forall i, 1 \leq i \leq n, \quad |a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$$

Théorème 165. 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 $\text{Ker} A = 0$). Posons $B = M^{-1}N$ est soit λ et v tels que $Bv = \lambda v$ avec $v \neq 0$. Puisque l'on s'intéresse à $\rho(B) < 1$, on s'intéresse 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'où 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 :

$$|c_{ii}| = |a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \geq \sum_{j < i}^n \left| \frac{a_{ij}}{\lambda} \right| + \sum_{j > i}^n \left| \frac{a_{ij}}{\lambda} \right| = \sum_{\substack{j=1 \\ j \neq i}}^n |c_{ij}|$$

et on obtient le même type de contradiction.

cqfd

7.1.2 Cas des matrices symétriques définies positives

Théorème 166. Si A est une matrice symétrique définie positive, alors les méthodes 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 étude théorique des valeurs propres de B montre que l'allure de la courbe $\rho(B)$ en fonction de ω 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} .

7.1.3 La méthode de correction

Soit le vecteur reste en x défini comme :

$$r(x) = b - Ax$$

et $\{r^{(k)}\}$ le reste en $\{x^{(k)}\}$. 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 $\{x^{(0)}\}$ 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 itérés autant de fois que nécessaire, pour s'arrêter lorsque le reste est suffisamment petit. A la $k^{\text{ième}}$ 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 $i = 0, 1, 2, \dots, k-1$. Si nous nous arrêtons lorsque $k = N$, il est nécessaire de résoudre $N+1$ systèmes linéaires : d'abord $Ax = b$, pour obtenir $x^{(0)}$ puis $Ax = -r^{(i)}$ et $i = 0, 1, 2, \dots, N-1$ 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 $r^{(i)} = b - Ax^{(i)}$.

8 SERIE D'EXERCICES

Exercice 167. 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 168. 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 169. 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 170. 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.