Table des matières

Préface **5**

Introduction générale **6**

1 Méthodes directes de résolution des systèmes linéaires **10**
1.1 Rappels d'algèbre linéaire, calcul matriciel 10
1.1.1 Généralités 10
1.1.2 Matrices particulières 11
1.1.3 Opérations sur les matrices 12
1.1.4 Matrice inverse 13
1.1.5 Déterminant d'une matrice 13
1.1.6 Produit de matrices par blocs 14
1.2 Systèmes linéaires et les méthodes directes 14
1.2.1 Position du problème 14
1.3 Rappels sur les systèmes linéaires 15
1.3.1 Méthodes classiques pour résoudre (S) sans calculer A^{-1} . . . 16
1.4 Méthodes directes 17
1.4.1 Systèmes à matrice triangulaire inférieure 17
1.4.2 Systèmes à matrice triangulaire supérieure 18
1.5 Méthodes d'élimination de Gauss 20
1.5.1 Méthode d'élimination de Gauss sans échange 20
1.5.2 Méthode d'élimination de Gauss avec échange 22
1.5.3 Méthode de Gauss avec choix du pivot 24
1.5.4 Nombre d'opérations pour l'élimination de Gauss 27
1.6 Méthode d'élimination de Gauss-Jordan 28
1.6.1 Nombre d'opérations pour l'élimination de Gauss-Jordan . . . 29

1.6.2 Calcul de l'inverse d'une matrice par l'élimination de G-J . . 29
1.7 Méthode de factorisation LU de Doolittle 30
1.7.1 Algorithme de la factorisation LU de Doolittle 32
1.7.2 Résolution du système $AX = b$ par la factorisation LU 32
1.8 Factorisation de Cholesky 33
1.8.1 Algorithme de la factorisation de Cholesky 34
1.8.2 Coût de la méthode de Cholesky 35
1.9 Exercices corrigés sur le chapitre 37

2 Méthodes itératives de résolution des systèmes linéaires 43
2.1 Analyse matricielle, Normes 43
2.1.1 Valeurs et vecteurs propres 45
2.1.2 Norme matricielle, rayon spectral et convergence 47
2.2 Systèmes linéaires et les méthodes itératives 48
2.2.1 Position du problème 48
2.3 Convergence des méthodes itératives 49
2.4 Étude de quelques méthodes itératives 50
2.4.1 Méthode de Jacobi 50
2.4.2 Méthode de Gauss-Seidel 52
2.4.3 Convergence des méthodes de Jacobi et de Gauss-Seidel . . . 53
2.5 Nombre d'itérations nécessaires 55
2.5.1 Estimation de l'erreur d'un processus itératif 55
2.6 Comparaison de deux méthodes itératives 57
2.7 Exercices corrigés sur le chapitre 59

3 Résolution d'équations non linéaires f(x)=0 65
3.1 Rappels de quelques éléments d'analyse 65
3.2 Résolution d'équations non linéaires f(x)=0 68
3.2.1 Position du problème 68
3.3 Localisation et séparation des racines 68
3.3.1 Méthode graphique 68
3.3.2 Méthode algébrique de balayage 69
3.3.3 Évaluation de l'erreur d'une racine approchée 71
3.4 Méthode de dichotomie 72
3.4.1 Principe de la méthode de dichotomie 72

3.4.2 Convergence de la méthode de dichotomie 73
3.4.3 Critère d'arrêt de la méthode de dichotomie 73
3.5 Méthode du point fixe 74
3.5.1 Principe de la méthode du point fixe 74
3.5.2 Estimation de l'erreur de la méthode du point fixe 77
3.5.3 Critère d'arrêt de la méthode du point fixe 77
3.5.4 Interprétation géométrique de la méthode du point fixe 78
3.6 Méthode de Newton-Raphson 81
3.6.1 Principe de la méthode de Newton-Raphson 81
3.6.2 Estimation de l'erreur de la méthode de Newton 83
3.6.3 Critère d'arrêt de la méthode de Newton 84
3.6.4 Interprétation géométrique de la méthode de Newton 84
3.7 Méthode de la sécante 86
3.7.1 Principe de la méthode de la sécante 86
3.7.2 Critère d'arrêt de la méthode de la sécante 87
3.7.3 Interprétation géométrique de la méthode de la sécante 87
3.8 Exercices corrigés sur le chapitre 89

4 Interpolation polynomiale **95**
4.1 Position du problème 95
4.2 Interpolation polynomiale 96
4.2.1 Méthode directe pour la recherche du polynôme P_n 97
4.3 Polynôme d'interpolation de Lagrange 98
4.3.1 Détermination pratique des polynômes de Lagrange 99
4.3.2 Erreur d'interpolation de la formule de Lagrange 101
4.3.3 Cas particulier où les points d'interpolation sont équidistants 103
4.4 Polynôme d'interpolation de Newton 104
4.4.1 Différences divisées 104
4.4.2 Formule explicite des différences divisées 105
4.4.3 Erreur d'interpolation de la formule de Newton 106
4.4.4 Détermination pratique des différences divisées 107
4.5 Différences finies ascendantes 109
4.5.1 Tableau des différences finies ascendantes 110
4.5.2 Polynôme d'interpolation ascendant de Newton 111

4.5.3 Évaluation de l'erreur de la première formule de Newton . . . 112
4.6 Différences finies descendantes 114
4.6.1 Tableau des différences finies descendantes 114
4.6.2 Polynôme d'interpolation descendant de Newton 115
4.7 Polynôme d'interpolation d'Hermite 116
4.7.1 Principe de l'interpolation d'Hermite 116
4.7.2 Erreur d'interpolation d'Hermite 117
4.8 Exercices corrigés sur le chapitre 119

Bibliographie **122**

Préface

Ce cours est le fruit d'un enseignement, d'une dizaine d'années, donné aux étudiants de deuxième année ingénieur toutes filières confondues, de deuxième année licence de mathématiques et de deuxième année licence chimie. Ce qui correspond actuellement au niveau L2-S3 et L2-S4. Le module enseigné s'intitulait d'abord module "TM011" puis module "Analyse Numérique I et II" et plus tard module "Méthodes Numériques". Il se compose à la fois d'un cours magistral, d'une séance de travaux dirigés et d'une séance de travaux pratiques. Le contenu du cours est très classique et doit beaucoup aux ouvrages de B. Demidovitch et M. Lahrib que l'on trouvera dans la bibliographie.

Introduction générale

« Les solutions analytiques exactes, qui sont rares en physique, sont élégantes, mais n'ont pas plus de valeur intrinsèque que les solutions numériques. On ne doit pas sous-estimer la facilité et la puissance des méthodes de calcul numérique ».

M.A. Ruderman
Cours de physique de Berkeley

L'analyse numérique est une discipline des mathématiques. Elle s'intéresse tant aux fondements théoriques qu'à la mise en pratique des méthodes permettant de résoudre, par des calculs purement numériques, des problèmes d'analyse mathématique. Plus formellement, l'analyse numérique est l'étude des algorithmes permettant de résoudre les problèmes de mathématiques continues (distinguées des mathématiques discrètes). Cela signifie qu'elle s'occupe principalement de répondre numériquement à des questions à variable réelle ou complexe comme l'algèbre linéaire numérique sur les champs réels ou complexes, la recherche de solution numérique d'équations différentielles et d'autres problèmes liés survenant dans les sciences physiques et l'ingénierie.

Certains problèmes de mathématiques peuvent être résolus numériquement sur ordinateur de façon exacte par un algorithme en un nombre fini d'opérations. Ces algorithmes sont parfois appelés méthodes directes ou qualifiés de finis. Des exemples sont l'élimination de Gauss-Jordan pour la résolution d'un système d'équations linéaires et l'algorithme du simplexe en optimisation linéaire.
Cependant, aucune méthode directe n'est connue pour certains problèmes dits NP-complets où aucun algorithme de calcul direct en temps polynomial n'est connu

à ce jour. Dans de tels cas, il est parfois possible d'utiliser une méthode itérative pour tenter de déterminer une approximation de la solution. Une telle méthode démarre depuis une valeur devinée ou estimée grossièrement et trouve des approximations successives qui devraient converger vers la solution sous certaines conditions. Même quand une méthode directe existe, une méthode itérative peut être préférable car elle est souvent plus efficace et même souvent plus stable.

Par ailleurs, certains problèmes continus peuvent parfois être remplacés par un problème discret dont la solution est connue pour approcher celle du problème continu, ce procédé est appelé discrétisation. Par exemple la solution d'une équation différentielle est une fonction. Cette fonction peut être représentée de façon approchée par une quantité finie de données, par exemple par sa valeur en un nombre fini de points de son domaine de définition, même si ce domaine est continu.

L'utilisation de l'analyse numérique est grandement facilitée par les ordinateurs. L'accroissement de la disponibilité et de la puissance des ordinateurs depuis la seconde moitié du xxe siècle a permis l'application de l'analyse numérique dans de nombreux domaines scientifiques, techniques et économiques, avec souvent des effets révolutionnaires.

L'étude et la résolution d'un problème d'analyse numérique en vraie grandeur passe par diverses phases:

(1) Le problème provient tout d'abord en général, de la mécanique, de la physique, des sciences de l'ingénieur. Il y a ainsi un travail de modélisation, se traduisant par une mise en équations. Le plus souvent cette modélisation est non triviale. Elle est faite, ou du moins engagée par celui qui pose le problème (physicien, mécanicien, ingénieur).

(2) L'étape précédente conduit à un modèle. C'est alors le travail du mathématicien d'en étudier la cohérence. Il propose un cadre fonctionnel adéquat et s'efforce de montrer que le problème posé, ou du moins la traduction mathématique qui en a été construite, admet une solution unique. Malheureusement, dans de nombreuses situations réelles, cette étude ne peut être menée à son terme, soit du fait de la difficulté mathématique soit du fait du manque de temps ou de moyens.

(3) Le pas suivant consiste à définir une approximation du modèle, afin de permettre une résolution numérique du problème posé. Là commence le travail proprement dit de l'analyste numéricien. Il lui faut construire un algorithme de calcul et, lorsque faire ce peut, démontrer que l'algorithme définit bien une solution approchée du problème.

(4) Enfin, il reste à écrire un logiciel implantant l'algorithme et à valider les résultats fournis en bout de chaîne par la machine.

Le champ d'application de l'analyse numérique précède de nombreux siècles l'invention des calculatrices modernes. En fait, bon nombre de mathématiciens du passé étaient préoccupés par l'analyse numérique, comme en témoignent évidemment les noms des algorithmes les plus importants tels que la méthode de Newton, l'interpolation lagrangienne, l'élimination de Gauss-Jordan ou la méthode d'Euler.

Pour faciliter les calculs manuels, de volumineux livres ont été édités, contenant des formules et tables de données telles que les points d'interpolation et coefficients de fonctions. Les plus connus sont les tables de logarithmes et les tables trigonométriques. À l'aide de ces tables, on pouvait rechercher les valeurs à utiliser dans les formules données, et obtenir de très bonnes estimations de certaines fonctions. Un travail fondamental dans ce domaine est l'ouvrage Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables publié par le NIST en 1964, et édité par Milton Abramowitz et Irene Stegun. Ce livre comprend un très grand nombre de formules et fonctions usuelles, et leurs valeurs en de nombreux points. Il est moins destiné au calcul à la main (les tables ne disposant pas d'aide au calcul, comme les différences tabulaires) qu'à la mise au point de programmes sur ordinateur, en facilitant les tests et en fournissant de nombreuses formules.

La règle à calcul représentait aussi une application pratique de ces anciennes tables numériques pour l'approximation rapide (généralement limitée à 3 ou 4 chiffres significatifs) de certaines fonctions continues à variable réelle simples (comme les fonctions trigonométriques, logarithmiques et exponentielles, et l'approximation rapide de la multiplication). Elle fut longtemps en usage notamment dans l'ingénierie, jusqu'au début des années 1980, avant que les calculatrices dites scien-

tifiques ne soient largement répandues et accessibles au grand public à un prix modique.

La calculatrice mécanique a aussi été développée comme un outil pour le calcul manuel dès son invention, liée au développement de l'horlogerie, notamment dans les applications commerciales (comme la caisse enregistreuse largement diffusée depuis le XIX^e siècle). Ces calculatrices ont évolué en ordinateurs électroniques dans les années 1940 grâce à la découverte des propriétés discrètes de la jonction P-N des semi-conducteurs (et son application dans le transistor), et on a vite découvert que ces ordinateurs seraient également utiles à des fins administratives. Mais l'invention de l'ordinateur a aussi influencé et largement étendu le champ d'application de l'analyse numérique, puisque dorénavant des calculs bien plus longs et compliqués peuvent être réalisés.

Comme applications, les algorithmes d'analyse numérique sont appliqués de façon routinière pour résoudre de nombreux problèmes dans les sciences appliquées et l'ingénierie. Des exemples sont la conception de structures comme les ponts, systèmes aéronautiques ou automobiles ou de systèmes complexes et chaotiques (voir prévision numérique du temps, modèles climatiques en météorologie), l'analyse, la modélisation ou la conception d'objets chimiques (voir chimie numérique), la recherche pétrolière et la géodésie, l'astrophysique, et les arts graphiques et la modélisation 3D (effets spéciaux au cinéma, les dessins animés, les jeux vidéo), les statistiques appliquées (démographie, modèles économiques...), l'analyse financière ou boursière...etc.

En fait, pratiquement tous les superordinateurs ou supercalculateurs mettent en pratique continûment des algorithmes d'analyse numérique. Par conséquent, l'efficacité des algorithmes joue un rôle important, et une méthode heuristique peut être préférée à une méthode basée sur une solide fondation théorique, simplement parce qu'elle est plus efficace.

Chapitre 1

Méthodes directes de résolution des systèmes linéaires

1.1 Rappels d'algèbre linéaire, calcul matriciel

Dans cette section, nous tenons à rappeler quelques notions d'algèbre linéaire qui seront utilisées dans notre cours.
Dans toute la suite, on désigne par $\mathbb{K}$ un corps commutatif avec $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$.

1.1.1 Généralités

Une matrice de type (m,n) est un tableau rectangulaire à m lignes et n colonnes.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{1.1}$$

Les nombres a_{ij} pour ($i = 1,...,m$ et $j = 1,...,n$) sont les éléments ou termes de la matrice A. L'indice i désigne le numéro de la ligne et j désigne celui de la colonne.
Une matrice sous forme (1.1) peut se mettre sous la forme condensée
$A = (a_{ij})_{ij}$ ($i = 1,...,m$ et $j = 1,...,n$).
Une application f de $\mathbb{K}^n$ dans $\mathbb{K}^m$ est dite linéaire si elle satisfait
$f(X + Y) = f(X) + f(Y)$ et $f(\lambda X) = \lambda f(X)$ pour tout $X,Y \in \mathbb{K}^n$ et tout $\lambda \in \mathbb{K}$.

La matrice A est aussi la matrice d'une application linéaire f de $\mathbb{K}^n$ dans $\mathbb{K}^m$ avec $(e_1,e_2,...,e_n)$ une base de $\mathbb{K}^n$ et $(f_1,f_2,...,f_m)$ une base de $\mathbb{K}^m$ où f est définie par

$f(e_j) = \sum_{i=1}^{m} a_{ij} f_j$, ($j$ = 1,...,n). Le $j^{ème}$ vecteur colonne de A représente la décomposition de $f(e_i)$ dans la base (f_1,...,f_m).

Définition 1.1. Soit E et F deux $\mathbb{K}$-espaces vectoriels et f une application linéaire de E dans F. On appelle

- Noyau de f noté $Ker(f)$ l'ensemble $ker(f) = \{X \in E / f(X) = 0_F\}$,
- Image de f notée $Im(f)$ l'ensemble $Im(f) = \{Y \in F / \exists X \in E : Y = f(X)\}$,
- Rang de f noté $rg(f)$ la dimension de $Im(f)$.

Définition 1.2. L'application linéaire f est injective $\Leftrightarrow Ker(f) = \{0_E\}$.

Définition 1.3. L'application linéaire f est surjective $\Leftrightarrow Im(f) = F$.

Définition 1.4. L'application f est bijective si elle est à la fois injective et surjective.

Définition 1.5. Soit $A \in \mathcal{M}_{m,n}(\mathbb{K})$. On appelle

- Noyau de A noté $Ker(A)$ l'ensemble $ker(A) = \{X \in \mathbb{K}^n / A(X) = 0\}$,
- Image de A notée $Im(A)$ l'ensemble $Im(A) = \{Y \in \mathbb{K}^m / \exists X \in \mathbb{K}^n : Y = AX\}$,
- Rang de A noté $rg(A)$ la dimension de $Im(A)$.

1.1.2 Matrices particulières

Si $m = n$, la matrice A est dite matrice carrée d'ordre n. Une matrice carrée d'ordre n est dite diagonale si $a_{ij} = 0$, $\forall i \neq j$.
Si les $a_{ii} = 1$,$\forall$ i = 1,...,n et $a_{ij} = 0$, $\forall i \neq j$, la matrice est dite matrice unité ou identité d'ordre n et notée par Id_n.
Une matrice dont tous les éléments sont nuls est dite matrice nulle et se note $A = 0$.
Une matrice est triangulaire inférieure si $a_{ij} = 0$, $\forall j > i$.
Une matrice est triangulaire supérieure si $a_{ij} = 0$, $\forall j < i$.

Définition 1.6. Une matrice est dite "matrice creuse" si elle contient un certain nombre significatif de valeurs nulles (c'est à dire elle contient une majorité de zéro).

1.1.3 Opérations sur les matrices

1°) Somme de deux matrices:
La somme de deux matrices A et B de même type (m,n) est définie par $(A+B)_{ij} = (A)_{ij} + (B)_{ij}$. La différence de deux matrices se définit de façon analogue.

2°) Produit d'une matrice par un scalaire:
Pour $\lambda \in \mathbb{K}$, la matrice λA est définie par $(\lambda A)_{ij} = \lambda (A)_{ij}$.

3°) Produit de deux matrices:
Pour A de type (m,n) et B de type (n,p), le produit AB est de type (m,p) et défini par $(AB)_{ij} = \sum_{k=1}^{k=n} A_{ik} B_{kj}$. En général on a $AB \neq BA$.

4°) Transposée d'une matrices:
Si A est de type (m,n), sa transposée notée A^t est de type (n,m) et définie par $(A^t)_{ij} = A_{ji}$.

Propriétés de la tranposée

La matrice transposée vérifie les propriétés suivantes:

1. $(A^t)^t = A$,
2. $(A + B)^t = A^t + B^t$,
3. $(AB)^t = B^t A^t$.

Les matrices triangulaires sont importantes pour la résolution numérique des systèmes car elles ont les propriétés suivantes:

Propriétés des matrices triangulaires

1. Le produit de deux matrices triangulaires inférieures est triangulaire inférieure et le produit de deux matrices triangulaires supérieures est triangulaire supérieure.
2. La transposée d'une matrice triangulaire inférieure est triangulaire supérieure et réciproquement.

Définition 1.7. La matrice A est dite symétrique si elle est égale à sa transposée $A = A^t$, antisymétrique si $A = -A^t$.

1.1.4 Matrice inverse

Définition 1.8. Soit A une matrice carrée d'ordre n.
– A est dite inversible ou régulière s'il existe une matrice unique notée A^{-1} telle que $AA^{-1} = A^{-1}A = Id_n$. A^{-1} est la matrice inverse de A.
– Une matrice carrée est dite singulière si elle n'est pas inversible.

Propriétés de la matrice inverse

La matrice inverse vérifie les propriétés suivantes:

1. $(A^{-1})^{-1} = A$,
2. $(AB)^{-1} = B^{-1}A^{-1}$,
3. $(A^{-1})^t = (A^t)^{-1}$
4. Une matrice inverse (si elle existe) rend plus facile la résolution d'un système $AX = b$ (sa solution est $X = A^{-1}b$).

1.1.5 Déterminant d'une matrice

Le déterminant d'une matrice carrée A d'ordre n est une réel noté $det(A)$ ou

$$det(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{vmatrix}$$

Son calcul se fait par un développement par rapport à une ligne ou une colonne et possède les propriétés suivantes:

Propriétés des déterminants

1. $detId_n = 1$,
2. $det(A^t) = detA$,
3. $det(AB) = detA \times detB$,
4. Si A est inversible, on a $detA^{-1} = \dfrac{1}{detA}$,
5. $det(\lambda A) = \lambda^n detA$, $\forall \lambda \in \mathbb{K}$,
6. Le déterminant d'une matrice diagonale ou triangulaire est égal au produit des éléments diagonaux.

1.1.6 Produit de matrices par blocs

La matrice carrée A étant décomposée de la façon suivante:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \vdots & a_{1,j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \vdots & a_{i,j+1} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{i+1,1} & \cdots & a_{i+1,j} & \vdots & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \vdots & a_{n,j+1} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} A_{ij}^{11} & A_{i,n-j}^{12} \\ A_{n-i,j}^{21} & A_{n-i,n-j}^{22} \end{pmatrix}$$

où A^{11} est la matrice de type (i,j), A^{21} est la matrice de type $(n-i,j)$, A^{12} est la matrice de type $(i,n-j)$ et A^{22} est la matrice de type $(n-i,n-j)$.

De même, on considère la matrice B décomposée en $B = \begin{pmatrix} B_{jk}^{11} & B_{j,n-k}^{12} \\ B_{n-j,k}^{21} & B_{n-j,n-k}^{22} \end{pmatrix}$

Le produit des deux matrices partitionnées A et B est défini comme un produit de deux matrices d'ordre deux. Ainsi,

$$AB = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix} \times \begin{pmatrix} B^{11} & B^{12} \\ B^{21} & B^{22} \end{pmatrix} = \begin{pmatrix} A^{11}B^{11} + A^{12}B^{21} & A^{11}B^{12} + A^{12}B^{22} \\ A^{21}B^{11} + A^{22}B^{21} & A^{21}B^{12} + A^{22}B^{22} \end{pmatrix}$$

1.2 Systèmes linéaires et les méthodes directes

1.2.1 Position du problème

Le problème consiste à résoudre le système de n équations à n inconnues suivant:

$$(S) \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \quad \vdots \qquad\qquad\qquad\qquad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases} \tag{1.2}$$

où x_i sont les inconnues, a_{ij} et b_i sont des nombres réels $(i = 1,...,n)$ et $(j = 1,...,n)$.

La forme matricielle du système (S) s'écrit $A.X = b$ où

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}; \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}; \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

Les algorithmes de résolution des systèmes linéaires se classent en trois grandes catégories.

1. Les méthodes directes (méthodes de Gauss, Gauss-Jordan, Cholesky, décomposition LU...etc).
2. Les méthodes itératives ou indirectes (méthodes de Jacobi, Gauss seidel, relaxation...etc).
3. Les méthodes projectives (méthodes de la plus profonde descente, méthodes du gradient conjugué...etc). Ces méthodes ne seront pas traitées dans notre cours.

Remarque 1.1.
Le choix d'une méthode particulière pour la résolution d'un système linéaire dépendra de la taille du système et aussi du type et de la structure de la matrice A intervenant dans le problème. Elle peut être caractérisée par une structure (creuse, diagonale, triangulaire, symétrique, définie positive...etc).

1.3 Rappels sur les systèmes linéaires

Soit le système $A.X = b$ avec $X \in \mathbb{R}^n$, $r = rg(A)$ et $B = [A,b]$ avec $r' = rg(B)$.
Alors la solution de $A.X = b$

$$\begin{cases} \text{est possible si} \quad r = r' \\ \text{est unique si} \quad r = r' = n \\ \text{comporte une infinité de solutions si} \quad r = r' < n \\ \text{n'existe pas si} \quad r < r' \end{cases}$$

Remarque 1.2. Si la matrice A est inversible alors le système (S) admet une solution unique de la forme $X = A^{-1}b$.
Ainsi théoriquement, le problème revient à calculer la matrice inverse A^{-1} et en pratique ce calcul est difficile. En effet, le calcul de A^{-1} équivaut à résoudre n systèmes linéaires du type $Ax_i = e_i$, $(i = 1,...,n)$ où e_i désigne le $i^{ème}$ vecteur de la base canonique de $\mathbb{R}^n$, ce qui s'avère bien plus coûteux que la résolution d'un seul système.

1.3.1 Méthodes classiques pour résoudre (S) sans calculer A^{-1}

Méthode de Cramer et son principe

Si le déterminant de A est différent de zéro alors la solution du système (S) est donnée par les formules de Cramer suivantes: $\quad x_i = \dfrac{detA_i}{detA}, \ \forall\ i = 1,...,n$

où A_i est la matrice obtenue en remplaçant dans A la $i^{ème}$ colonne par le vecteur second membre b.

Remarque 1.3. Coût de la méthode de Cramer
Le nombre d'opérations élémentaires $(+, -, \times, \div)$ que nécessite la méthode de Cramer est de l'ordre de $(n+1)^2 n! - 1$. En effet, il faut d'abord calculer $(n+1)$ déterminants puis effectuer n divisions pour calculer les (x_i) $i = 1,...,n$.
Pour calculer chaque déterminant, nous devons effectuer $(n!n)$ multiplications et $(n! - 1)$ additions. Il en résulte que la résolution du système (S) nécessite le total de
$T_{Cramer} = (n+1)n!n + (n+1)(n! - 1) + n = (n+1)^2 n! - 1$ opérations élémentaires.
Par exemple $T_{Cramer} = 4319$ pour $n = 5$ et $T_{Cramer} = 4.10^8$ pour $n = 10$, ce qui semble énorme.

Exemple d'application

Résoudre le système suivant par la méthode de Cramer (S_1) $\begin{cases} x + 2y = 5 \\ 2x + y = 4 \end{cases}$

Solution

Les formules de Cramer donnent

$$x = \frac{detA_1}{detA} = \frac{\begin{vmatrix} 5 & 2 \\ 4 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix}} = \frac{-3}{-3} = 1 \qquad \text{et} \qquad y = \frac{detA_2}{detA} = \frac{\begin{vmatrix} 1 & 5 \\ 2 & 4 \end{vmatrix}}{\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix}} = \frac{-6}{-3} = 2$$

La solution de (S_1) est donc $X = \begin{pmatrix} x = 1 \\ y = 2 \end{pmatrix}$

Méthode de substitution (ou d'élimination) et son principe

Pour résoudre le système (S) par la méthode de substitution ou d'élimination, le plus simple est d'utiliser l'une des équations de (S) pour exprimer une inconnue

en fonction des autres. Dans les autres équations, on remplace (substitue) alors cette inconnue par l'expression obtenue, puis on simplifie. Apparaît ainsi un sous-système, plus facile à résoudre, qui comporte une inconnue de moins et ainsi de suite.

Exemple d'application

Résoudre le système (S_2) précédent par la méthode de substitution.

Solution

$$\begin{cases} x+2y=5 \\ 2x+y=4 \end{cases} \Rightarrow \begin{cases} x=-2y+5 \\ 2x+y=4 \end{cases} \Rightarrow \begin{cases} x=-2y+5 \\ 2(-2y+5)+y=4 \end{cases}$$

$$\begin{cases} x=-2y+5 \\ -3y=-6 \end{cases} \Rightarrow \begin{cases} x=-2y+5 \\ y=2 \end{cases} \Rightarrow \begin{cases} x=1 \\ y=2 \end{cases} \Rightarrow X=\begin{pmatrix} x=1 \\ y=2 \end{pmatrix}$$

1.4 Méthodes directes

Définition 1.9. Méthode directe
On appelle méthode directe de résolution du système (S) une méthode qui fournit, en l'absence d'erreurs d'arrondi, la solution exacte X (A et b étant connus) en un nombre fini d'opérations élémentaires du type $(+, -, \times, \div)$.

Remarque 1.4.
Nous verrons que ces méthodes consistent en la construction d'une matrice inversible S telle que SA soit une matrice triangulaire et le système équivalent obtenu $SAX = Sb$ étant plus facile à résoudre.

1.4.1 Systèmes à matrice triangulaire inférieure

Considérons un système dont la matrice A est triangulaire inférieure du type:

$$\begin{cases} a_{11}x_1 & = b_1 & (1) \\ a_{21}x_1 + a_{22}x_2 & = b_2 & (2) \\ & \ddots \quad \vdots & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n & = b_n & (n) \end{cases} \tag{1.3}$$

Si la matrice est régulière et triangulaire alors, comme det(A)=$\prod_{i=1}^{i=n} a_{ii}$, on en déduit que $a_{ii} \neq 0$, $\forall i = 1,...,n$. De l'équation (1), on tire $x_1 = \frac{b_1}{a_{11}}$ que l'on substitue ensuite dans l'équation (2) pour obtenir x_2, et ainsi de suite jusqu'à obtention de x_n.

Principe de la méthode de la descente

Si $a_{ii} \neq 0$ pour $i = 1,...,n$, la suite $\{x_i\}$ définie par

$$\begin{cases} x_1 = \dfrac{b_1}{a_{11}} \\ \\ x_i = \dfrac{1}{a_{ii}}(b_i - \sum_{j=1}^{i-1} a_{ij}.x_j) \ , \quad i = 2,...,n \end{cases} \tag{1.4}$$

est l'unique solution du système (1.3).

Coût de la méthode de la descente

Le coût ou le nombre d'opérations que nécessite la méthode de la descente est de n^2 opérations élémentaires. En effet, pour la ligne 1, le calcul de x_1 se fait en une opération (une division). Pour les lignes $i = 2,...,n$, le calcul de x_i s'effectue en $(i-1)$ multiplications, $(i-2)$ additions, une soustraction et une division soit $(2i-1)$ au total. Le calcul de la solution du système (1.3) par la descente s'effectue donc en un total de $T_{descente} = \sum_{i=1}^{i=n}(2i-1) = n^2$ opérations élémentaires.

1.4.2 Systèmes à matrice triangulaire supérieure

Lorsque A est une matrice triangulaire supérieure, la résolution du système est immédiate.

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ \qquad a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad\qquad \ddots \qquad\quad \vdots \\ \qquad\qquad\qquad a_{nn}x_n = b_n \end{cases} \tag{1.5}$$

On calcule successivement x_n à partir de la dernière équation, puis x_{n-1} à partir de l'avant-dernière et ainsi de suite.

Principe de la "méthode de la remontée"ou "retour arrière"

Si $a_{ii} \neq 0$ pour $i = 1,...,n$, la suite $\{x_i\}$ définie par

$$\begin{cases} x_n = \dfrac{b_n}{a_{nn}} \\ \\ x_i = \dfrac{1}{a_{ii}}(b_i - \sum\limits_{j=i+1}^{j=n} a_{ij}.x_j) \ , \quad i = n-1,...,1 \end{cases} \tag{1.6}$$

est l'unique solution du système (1.5).

Remarque 1.5. Le coût de la méthode de la remontée est égal au coût de la méthode de la descente (même principe de calcul).

Exemple d'application 1

Résoudre le système suivant par la méthode de la remontée $\begin{cases} 4x_1 + 4x_2 - x_3 = 2 \\ x_2 + 4x_3 = 1 \\ -7x_3 = 6 \end{cases}$

Solution

$$\begin{cases} 4x_1 + 4x_2 - x_3 = 2 \\ x_2 + 4x_3 = 1 \\ -7x_3 = 6 \end{cases} \Rightarrow \begin{cases} 4x_1 + 4x_2 - x_3 = 2 \\ x_2 + 4x_3 = 1 \\ x_3 = -\frac{6}{7} \end{cases} \Rightarrow \begin{cases} 4x_1 + 4x_2 - x_3 = 2 \\ x_2 = 1 - 4x_3 = \frac{31}{7} \\ x_3 = -\frac{6}{7} \end{cases}$$

$$\begin{cases} x_1 = \frac{1}{4}(2 - 4x_2 + x_3) = -\frac{29}{7} \\ x_2 = \frac{31}{7} \\ x_3 = -\frac{6}{7} \end{cases} \Rightarrow \text{ la solution du système est } X = \begin{cases} x_1 = -29/7 \\ x_2 = 31/7 \\ x_3 = -6/7 \end{cases}$$

Exemple d'application 2

Résoudre le système suivant par la méthode de la descente $\begin{cases} 4x_1 = 8 \\ 2x_1 + 3x_2 = 10 \\ 3x_1 - x_2 + 3x_3 = 13 \end{cases}$

Solution

$$\begin{cases} 4x_1 = 8 \\ 2x_1 + 3x_2 = 10 \\ 3x_1 - x_2 + 3x_3 = 13 \end{cases} \Rightarrow \begin{cases} x_1 = \frac{8}{4} = 2 \\ 2x_1 + 3x_2 = 10 \\ 3x_1 - x_2 + 3x_3 = 13 \end{cases} \Rightarrow \begin{cases} x_1 = 2 \\ x_2 = \frac{1}{3}(10 - 2x_1) = \frac{6}{3} = 2 \\ 3x_1 - x_2 + 3x_3 = 13 \end{cases}$$

$$\begin{cases} x_1 = 2 \\ x_2 = 2 \\ x_3 = \frac{1}{3}(13 - 3x_1 + x_2) = \frac{9}{3} = 3 \end{cases} \Rightarrow \text{ la solution du système est } X = \begin{cases} x_1 = 2 \\ x_2 = 2 \\ x_3 = 3 \end{cases}$$

1.5 Méthodes d'élimination de Gauss

Définition 1.10. :
On appelle opérations élémentaires ou transformations élémentaires sur une matrice les opérations suivantes:

1. Permuter deux lignes ou deux colonnes entre elles.
2. Multiplier tous les éléments d'une ligne ou une colonne de A par un même scalaire non nul.
3. Remplacer une ligne ou une colonne de A par la somme de cette même ligne ou même colonne avec une autre ligne ou colonne multipliée par un scalaire non nul

.

Définition 1.11.
Deux matrices sont dites équivalentes si l'une s'obtient de l'autre à la suite d'un nombre fini de transformations élémentaires

.

La "méthode d'élimination de Gauss "appelée encore "méthode de triangularisation de Gauss "consiste en premier lieu à transformer, par des opérations élémentaires simples sur la matrice A, le système $AX = b$ à résoudre en un système équivalent (c'est-à-dire ayant la ou les mêmes solutions), $S.A.X = S.b$ où $S.A$ est une matrice triangulaire supérieure. Si A est inversible, la solution du système peut ensuite être obtenue par une méthode de la remontée, mais le procédé d'élimination est en fait très général, la matrice pouvant être rectangulaire.

1.5.1 Méthode d'élimination de Gauss sans échange

Considérons le système linéaire initial (S) où A est une matrice inversible d'ordre n. Partons de la formule, $A^{(1)} = A$ et $b^{(1))} = b$.
Supposons de plus que le terme a_{11} de la matrice est non nul. L'inconnue x_1 peut être éliminée des lignes 2 à n du système en leur retranchant respectivement la première ligne multipliée par le coefficient $\frac{a_{i1}}{a_{11}}$ $(i = 2,...,n)$.
Notons par $A^{(2)}$ et $b^{(2)}$ la matrice et le vecteur second membre résultant des opérations précédentes, ainsi nous avons:
$a_{ij}^{(2)} = a_{ij} - \frac{a_{i1}}{a_{11}}.a_{1j}$ et $b_i^{(2)} = b_i - \frac{a_{i1}}{a_{11}}.b_1$ $(i = 2,...,n$ et $j = 2,...,n)$ et le système $A^{(2)}.X = b^{(2)}$ est équivalent au système initial (S).
De même, en supposant que le coefficient $a_{22}^{(2)}$ de $A^{(2)}$ est non nul, on peut procéder

à l'élimination de l'inconnue x_3 des lignes 3 à n de ce système et ainsi de suite.
On obtient, sous l'hypothèse $a_{kk}^{(k)} \neq 0$ $(k = 1,...,n-1)$ une suite finie de matrices $A^{(k)}$, $(k \geq 2)$ de la forme:

$$A^{(k)} = \begin{pmatrix} a_{11}^{(k)} & a_{12}^{(k)} & \cdots & \cdots & \cdots & a_{1n}^{(k)} \\ 0 & a_{22}^{(k)} & \cdots & \cdots & \cdots & a_{2n}^{(k)} \\ \vdots & \ddots & \ddots & \cdots & \cdots & \vdots \\ 0 & \cdots & 0 & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & a_{nk}^{(k)} & \cdots & a_{nn}^{(k)} \end{pmatrix}$$

Les quantités $a_{kk}^{(k)}$ $(k = 1,...,n-1)$ sont appelées pivots et sont supposées non nulles à chaque étape de l'algorithme. Ainsi, les formules résumant le passage du $k^{ème}$ système linéaire au $(k+1)^{ème}$ sont données par:

$$\begin{cases} a_{ij}^{(k+1)} = a_{ij}^{(k)} - \dfrac{a_{ik}^{(k)}}{a_{kk}^{(k)}}.a_{kj}^{(k)} \quad i = 2,...,n; \quad j = 2,...,n \\ \\ b_i^{(k+1)} = b_i^{(k)} - \dfrac{a_{ik}^{(k)}}{a_{kk}^{(k)}}.b_k^{(k)} \quad i = 2,...,n; \quad j = 2,...,n \end{cases} \tag{1.7}$$

Remarque 1.6.
En pratique, pour une résolution d'un système linéaire $A.X = b$ à la main par la méthode de Gauss, il est commode d'appliquer l'algorithme sur la matrice augmentée $[A\ b]$.

Exemple d'application

Résoudre le système linéaire suivant par la méthode d'élimination de Gauss sans échange.

$$\begin{cases} 2x_1 + 8x_2 + 4x_3 = 3 \\ 2x_1 + 10x_2 + 6x_3 = 1 \\ x_1 + 8x_2 + 5x_3 = 1 \end{cases} \tag{1.8}$$

Solution

A la première étape, le pivot vaut 2 et on soustrait de la deuxième (resp. troisième) équation la première équation multipliée par 1 (resp. par (1/2) pour obtenir

$$\begin{cases} 2x_1 + 8x_2 + 4x_3 = 3 \\ \quad 2x_2 + 2x_3 = -2 \\ \quad 4x_2 + 3x_3 = -1/2 \end{cases} \tag{1.9}$$

Le pivot vaut 2 à la deuxième étape. On retranche alors à la troisième équation la deuxième équation multipliée par 2, d'où le système

$$\begin{cases} 2x_1 + 8x_2 + 4x_3 = 3 \\ 2x_2 + 2x_3 = -2 \\ -x_3 = 7/2 \end{cases} \tag{1.10}$$

Le système (1.10) étant triangulaire supérieur, il est équivalent au système d'origine (1.8) et peut être résolu par la méthode de la remontée.

$$\begin{cases} 2x_1 + 8x_2 + 4x_3 = 3 \\ 2x_2 + 2x_3 = -2 \\ x_3 = -7/2 \end{cases} \Rightarrow \begin{cases} 2x_1 + 8x_2 + 4x_3 = 3 \\ x_2 = (1/2)(-2 - 2x_3) = (1/2)(-2 - 2(-7/2)) = 5/2 \\ x_3 = -7/2 \end{cases}$$

$$\begin{cases} x_1 = (1/2)(3 - 8x_2 - 4x_3) = -3/2 \\ x_2 = 5/2 \\ x_3 = -7/2 \end{cases} \Rightarrow \text{la solution de (1.8) est } X = \begin{cases} x_1 = -3/2 \\ x_2 = 5/2 \\ x_3 = -7/2 \end{cases}$$

Remarque 1.7.
La méthode d'élimination de Gauss sans échange, ne peut s'appliquer que si tous les pivots $a_{kk}^{(k)}$ $(k = 1,...,n-1)$ sont non nuls, ce qui élimine de fait l'application de cet algorithme aux matrices inversibles du type $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ par exemple.
De plus, le fait qu'une matrice soit inversible n'empêche pas l'apparition de pivot nul durant l'élimination de Gauss (voir exemple suivant).

Exemple d'application

Trianguläriser la matrice inversible $A = A^{(1)} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 7 & 8 & 9 \end{pmatrix} \Rightarrow A^{(2)} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & -1 \\ 0 & -6 & -12 \end{pmatrix}$

En appliquant l'élimination de Gauss sans échange, on remarque que le pivot $a_{22}^{(2)}$ est nul, d'où l'arrêt de l'algorithme à la seconde étape.

1.5.2 Méthode d'élimination de Gauss avec échange

Supposons qu'à la première étape le coefficient $a_{11}^{(1)}$ soit nul et qu'au moins l'un des coefficients de la première colonne de la matrice $A^{(1)} = A$ est non nul (sinon det(A)=0 et A serait non inversible).
On choisit un de ces éléments non nuls comme premier pivot d'élimination et l'on échange alors la première ligne du système avec celle du pivot non nul avant de

procéder à l'élimination de la première colonne de la matrice résultante.
On note $A^{(2)}$ et $b^{(2)}$ la matrice et le second membre du système obtenu et l'on réitère ce procédé. A l'étape k $(2 \leq k \leq n-1)$, la matrice $A^{(k)}$ est inversible et donc l'un au moins des éléments $a_{ik}^{(k)}$, $(k \leq i \leq n)$ est différent de zéro.
Après avoir choisi comme pivot l'un de ces éléments non nuls, on effectue l'échange de la ligne de ce pivot avec la $k^{ème}$ ligne de la matrice $A^{(k)}$, puis l'élimination conduisant à la matrice $A^{(k+1)}$.
Ainsi, on arrive après $(n-1)$ étapes à la matrice $A^{(n)}$ dont le coefficient $a_{nn}^{(n)}$ est $\neq 0$.

Exemple d'application

Considérons la résolution du système linéaire $AX = b$ où

$$\begin{cases} 2x_1 + 4x_2 - 4x_3 + x_4 = 0 \\ 3x_1 + 6x_2 + x_3 - 2x_4 = -7 \\ -x_1 + x_2 + 2x_3 + 3x_4 = 4 \\ x_1 + x_2 - 4x_3 + x_4 = 2 \end{cases} \tag{1.11}$$

Solution

Par application de la méthode d'élimination de Gauss avec échange sur la matrice augmentée $[A \;\; b] = [A^{(1)} \; b^{(1)}]$. On trouve successivement

$$\left(\begin{array}{cccc|c} & & A^{(1)} & & . & b^{(1)} \\ 2 & 4 & -4 & 1 & . & 0 \\ 3 & 6 & 1 & -2 & . & -7 \\ -1 & 1 & 2 & 3 & . & 4 \\ 1 & 1 & -4 & 1 & . & 2 \end{array}\right) \begin{array}{c} \text{opérations} \\ L_2 - (3/2)L_1 \\ L_3 + (1/2)L_1 \\ L_4 - (1/2)L_1 \end{array} \left(\begin{array}{cccc|c} & & A^{(2)} & & . & b^{(2)} \\ 2 & 4 & -4 & 1 & . & 0 \\ 0 & 0 & 7 & -7/2 & . & -7 \\ 0 & 3 & 0 & 7/2 & . & 4 \\ 0 & -1 & -2 & 1/2 & . & 2 \end{array}\right) \begin{array}{c} \text{opérations} \\ L_2 \longleftrightarrow L_3 \end{array}$$

$$\left(\begin{array}{cccc|c} 2 & 4 & -4 & 1 & . & 0 \\ 0 & 3 & 0 & 7/2 & . & 4 \\ 0 & 0 & 7 & -7/2 & . & -7 \\ 0 & -1 & -2 & 1/2 & . & 2 \end{array}\right) \begin{array}{c} \text{opérations} \\ L_4 + (1/3)L_2 \\ \longmapsto \end{array} \left(\begin{array}{cccc|c} & & A^{(3)} & & . & b^{(3)} \\ 2 & 4 & -4 & 1 & . & 0 \\ 0 & 3 & 0 & 7/2 & . & 4 \\ 0 & 0 & 7 & -7/2 & . & -7 \\ 0 & 0 & -2 & 5/3 & . & 10/3 \end{array}\right)$$

$$\begin{array}{c} \text{opérations} \\ L_4 + (2/7)L_3 \\ \longmapsto \end{array} \left(\begin{array}{cccc|c} & & A^{(4)} & & . & b^{(4)} \\ 2 & 4 & -4 & 1 & . & 0 \\ 0 & 3 & 0 & 7/2 & . & 4 \\ 0 & 0 & 7 & -7/2 & . & -7 \\ 0 & 0 & 0 & 2/3 & . & 4/3 \end{array}\right)$$

Le système $A^{(4)}X = b^{(4)}$ est équivalent au système initial $AX = b$, de plus il est triangulaire supérieur donc nous pouvons le résoudre par la méthode de la remontée. Ainsi la solution du système $AX = b$ est donnée par $X = (1, -1{,}0{,}2)^t$.

1.5.3 Méthode de Gauss avec choix du pivot

Exemple d'application

Sachant que la solution exacte du système suivant est $X = (1/3{,}2/3)$, nous allons essayer de la retrouver par la méthode d'élimination de Gauss.

$$\begin{cases} 0{,}0003x_1 + 3x_2 = 2{,}0001 \\ x_1 + x_2 = 1 \end{cases} \tag{1.12}$$

Solution

En choisissant le nombre 3.10^{-4} comme pivot à la première étape de l'élimination de Gauss, on obtient le système triangulaire suivant:

$$\begin{cases} 0{,}0003x_1 + 3x_2 = 2{,}0001 \\ -9999x_2 = -6666 \end{cases} \tag{1.13}$$

Sa solution est alors $x_1 = 0$ et $x_2 \approx 0{,}6667$, on conclut qu'elle est très éloignée de la solution exacte du système.
Par contre, si on échange d'abord les deux équations du système de départ pour utiliser le nombre 1 comme pivot, on trouve

$$\begin{cases} x_1 + x_2 = 1 \\ 2{,}9997x_2 = 1{,}9998 \end{cases} \tag{1.14}$$

Sa solution calculée vaut $x_1 \approx 1/3$ et $x_2 \approx 2/3$, on conclut que dans ce cas la solution trouvée est correcte.

Remarque 1.8.
Le choix d'un pivot très petit conduit à un mauvais résultat car l'algorithme conduit à des erreurs d'arrondi importantes. Pour éviter ce problème, nous allons utiliser l'une des deux méthodes suivantes.

Méthode de Gauss avec pivot partiel

A la $k^{ème}$ étape, on choisit comme pivot l'élément $a_{lk}^{(k)}$ tel que $a_{lk}^{(k)} = \max\limits_{k\leq i\leq n} \left|a_{ik}^{(k)}\right|$.

Méthode de Gauss avec pivot total

A la $k^{ème}$ étape, on choisit comme pivot l'élément $a_{lm}^{(k)}$ tel que $a_{lm}^{(k)} = \max\limits_{k\leq i,j\leq n} \left|a_{ij}^{(k)}\right|$.

Remarque 1.9.
Une permutation sur les colonnes entraîne une permutation sur les inconnues.

Exemple d'application 1

Triangulariser la matrice A en utilisant la méthode d'élimination de Gauss avec choix du pivot partiel.

$$\begin{pmatrix} A = A^{(1)} & & \\ 1 & 0 & 1 \\ -2 & 4 & 2 \\ 0 & -3 & -1 \end{pmatrix} \begin{array}{c}\text{opérations} \\ L_1 \longleftrightarrow L_2 \end{array} \begin{pmatrix} A^{(2)} & & \\ -2 & 4 & 2 \\ 1 & 0 & 1 \\ 0 & -3 & -1 \end{pmatrix} \begin{array}{c}\text{opérations} \\ L_2 + (1/2)L_1 \\ \longmapsto \end{array} \begin{pmatrix} A^{(3)} & & \\ -2 & 4 & 2 \\ 0 & 2 & 2 \\ 0 & -3 & -1 \end{pmatrix}$$

$$\begin{array}{c}\text{opérations} \\ L_2 \longleftrightarrow L_3 \end{array} \begin{pmatrix} A^{(4)} & & \\ -2 & 4 & 2 \\ 0 & -3 & -1 \\ 0 & 2 & 2 \end{pmatrix} \begin{array}{c}\text{opérations} \\ L_3 + (2/3)L_2 \\ \longmapsto \end{array} \begin{pmatrix} A^{(5)} & & \\ -2 & 4 & 2 \\ 0 & -3 & -1 \\ 0 & 0 & 4/3 \end{pmatrix}$$

Exemple d'application 2

Résoudre le système $AX = b$ suivant par la méthode de Gauss avec choix du pivot total.

$$\begin{cases} 0x_1 + x_2 + 3x_3 = 1 \\ 3x_1 + 3x_2 + x_3 = 0 \\ 4x_1 + 2x_2 - x_3 = 2 \end{cases} \tag{1.15}$$

Solution

Sur toute la matrice A, le nombre 4 étant le plus élevé donc il est choisi comme premier pivot d'où la permutation des lignes L_1 et L_3 du système pour obtenir

$$\left[\begin{array}{cccc|c} & A^{(1)} & & & b^{(1)} \\ 0 & 1 & 3 & . & 1 \\ 3 & 3 & 1 & . & 0 \\ 4 & 2 & -1 & . & 2 \end{array}\right] \begin{array}{c} \text{Opérations} \\ L_1 \leftrightarrow L_3 \\ \longmapsto \end{array} \left[\begin{array}{cccc|c} & A^{(2)} & & & b^{(2)} \\ 4 & 2 & -1 & . & 2 \\ 3 & 3 & 1 & . & 0 \\ 0 & 1 & 3 & . & 1 \end{array}\right]$$

On applique la méthode d'élimination de Gauss en soustrayant de la deuxième ligne L_2 la première ligne L_1 multipliée par le nombre $3/4$ pour obtenir

$$\begin{array}{c} \text{Opérations} \\ L_2 - 3/4L_1 \\ \longmapsto \end{array} \left[\begin{array}{cccc|c} & A^{(3)} & & & b^{(3)} \\ 4 & 2 & -1 & . & 2 \\ 0 & 3/2 & 7/4 & . & -3/2 \\ 0 & 1 & 3 & . & 1 \end{array}\right]$$

Le nombre 3 étant le plus élevé sur la sous matrice $\begin{pmatrix} 3/2 & 7/4 \\ 1 & 3 \end{pmatrix}$ donc il est choisi comme deuxième pivot en permutant d'abord les colonnes C_2 et C_3 de tout le système (avec aussi permutation des inconnues x_2 et x_3) pour obtenir

$$\begin{array}{c} \text{Opérations} \\ C_2 \leftrightarrow C_3 \\ \longmapsto \end{array} \left[\begin{array}{cccc|c} & A^{(4)} & & & b^{(4)} \\ 4 & -1 & 2 & . & 2 \\ 0 & 7/4 & 3/2 & . & -3/2 \\ 0 & 3 & 1 & . & 1 \end{array}\right]$$

On permute les lignes L_2 et L_3 du système pour obtenir

$$\begin{array}{c} \text{Opérations} \\ L_2 \leftrightarrow L_3 \\ \longmapsto \end{array} \left[\begin{array}{cccc|c} & A^{(5)} & & & b^{(5)} \\ 4 & -1 & 2 & . & 2 \\ 0 & 3 & 1 & . & 1 \\ 0 & 7/4 & 3/2 & . & -3/2 \end{array}\right]$$

On applique la méthode d'élimination de Gauss en soustrayant de la troisième ligne L_3 la deuxième ligne L_2 multipliée par le nombre $7/12$ pour obtenir

$$\begin{array}{c} \text{Opérations} \\ L_3 - 7/12L_2 \\ \longmapsto \end{array} \left[\begin{array}{cccc|c} & A^{(6)} & & & b^{(6)} \\ 4 & -1 & 2 & . & 2 \\ 0 & 3 & 1 & . & 1 \\ 0 & 0 & 11/12 & . & -25/12 \end{array}\right]$$

Le système $A^{(6)}X = b^{(6)}$ final étant triangulaire supérieur, il est résolu par la méthode de la remontée. $\left\{\begin{array}{c} 4x_1 - x_3 + 2x_2 = 2 \\ 3x_3 + x_2 = 1 \\ x_2 = -25/11 \end{array}\right. \Rightarrow \left\{\begin{array}{l} 4x_1 - x_3 + 2x_2 = 2 \\ x_3 = 1/3(1 - x_2) = 12/11 \\ x_2 = -25/11 \end{array}\right. \Rightarrow$

$$\begin{cases} x_1 = 1/4(2 + x_3 - 2x_2) = 21/11 \\ x_3 = 12/11 \\ x_2 = -25/11 \end{cases} \Rightarrow \text{la solution de (1.15) est } X = \begin{cases} x_1 = 21/11 \\ x_2 = -25/11 \\ x_3 = 12/11 \end{cases}$$

1.5.4 Nombre d'opérations pour l'élimination de Gauss

La résolution d'un système linéaire, de n équations à n inconnues, par la méthode d'élimination de Gauss nécessite un total de $T_{Gauss} = \frac{4n^3+9n^2-7n}{6}$. En effet,
a) Pour passer de la matrice $A^{(k)}$ à la matrice $A^{(k+1)}$ $(1 \leq k \leq n-1)$, on effectue $(n-k+1)(n-k)$ additions et autant de multiplications et $(n-k)$ divisions, ce qui correspond à un total de $\frac{n(n^2-1)}{3}$ additions plus $\frac{n(n^2-1)}{3}$ multiplications plus $\frac{(n-1)n}{2}$ divisions pour l'élimination complète.
b)A l'étape (k), pour la mise à jour du second membre, on effectue $(n-k)$ additions et multiplications. Soit un total de $\frac{(n-1)n}{2}$ additions et multiplications pour l'élimination complète.
c) Pour la résolution du système triangulaire par la méthode de la remontée, il faut $\frac{(n-1)n}{2}$ additions et multiplications et n divisions.
D'où $T_{Gauss} = 2\frac{n(n^2-1)}{3} + 3\frac{(n-1)n}{2} + n = \frac{4n^3+9n^2-7n}{6}$.
A titre d'exemple, pour $n = 5$ on obtient $T_{Gauss} = 115$ opérations et $T_{Cramer} = 4319$ opérations. Ce qui montre l'énorme amélioration que la méthode d'élimination de Gauss apporte par rapport à la méthode de Cramer.

Remarque 1.10.
La méthode d'élimination de Gauss permet de calculer aisément les déterminants de matrices grâce à la formule suivante:

$$det(A) = (-1)^p det(A^{(n)}) = \pm \prod_{k=1}^{k=n} a_{kk}^{(k)}$$

Où $A^{(n)}$ est la matrice triangulaire supérieure à l'étape (n); $a_{kk}^{(k)}$ est le pivot à l'étape (k) et p représente le nombre de permutations effectuées, sur les lignes et colonnes, durant l'élimination de Gauss.

Exemple

Reprenons l'exemple 1 et l'exemple 2 de la sous-section (1.5.3).
Exemple 1: $det(A) = (-1)^p$(produits des pivots) $= (-1)^2(-2)(-3)(\frac{4}{3}) = 8$.

Exemple 2: $det(A) = (-1)^p(det(A^{(5)}) = (-1)^3(4)(3)(\frac{11}{12}) = -11.$

1.6 Méthode d'élimination de Gauss-Jordan

La méthode d'élimination de Gauss-Jordan est une variante de la méthode d'élimination de Gauss. Elle permet non pas de trianguläriser la matrice A du système comme dans la méthode de Gauss mais à remplacer A par la matrice identité (Id).
A partir de la matrice augmentée $[A \;\; b]$, on effectue des tranformations élémentaires (comme normalisation puis réduction pour arriver au système final $[Id \;\; b']$. La solution $X = b'$ de ce dernier système est la solution du système de départ $AX = b$.
Chaque étape de la méthode doit subir au moins:

1. Une normalisation en divisant par le pivot non nul (pour avoir les 1 sur la diagonale principale).
2. Une réduction avec la méthode d'élimination de Gauss (pour avoir des zéros ailleurs sur toute la matrice).

Exemple d'application

Résoudre le système $AX = b$ suivant par la méthode de Gauss-Jordan.

$$\begin{cases} -x_1 + 3x_2 + 3x_3 = 1 \\ 2x_1 + 2x_2 \quad\quad = 0 \\ 3x_1 + 2x_2 + 6x_3 = 0 \end{cases} \tag{1.16}$$

Solution

On part de la matrice augmentée et l'on trouve successivement

$$\left[\begin{array}{cccc} & A^{(1)} & & b^{(1)} \\ -1 & 3 & 3 & 1 \\ 2 & 2 & 0 & 0 \\ 3 & 2 & 6 & 0 \end{array}\right] \begin{array}{c}\text{Normalisation} \\ (L_1/(-1)) \\ \longmapsto\end{array} \left[\begin{array}{cccc} & & & \\ 1 & -3 & -3 & -1 \\ 2 & 2 & 0 & 0 \\ 3 & 2 & 6 & 0 \end{array}\right] \begin{array}{c}\text{Réduction} \\ L_2 - 2L_1 \\ \longmapsto \\ L_3 - 3L_1\end{array}$$

$$\left[\begin{array}{cccc} & A^{(2)} & & b^{(2)} \\ 1 & -3 & -3 & -1 \\ 0 & 8 & 6 & 2 \\ 0 & 11 & 15 & 3 \end{array}\right] \begin{array}{c}\text{Normalisation} \\ (L_2/8) \\ \longmapsto\end{array} \left[\begin{array}{cccc} & & & \\ 1 & -3 & -3 & -1 \\ 0 & 1 & 3/4 & 1/4 \\ 0 & 11 & 15 & 3 \end{array}\right] \begin{array}{c}\text{Réduction} \\ L_1 + 3L_2 \\ \longmapsto \\ L_3 - 11L_2\end{array}$$

$$\left[\begin{array}{cccc} & A^{(3)} & & . \quad b^{(3)} \\ 1 & 0 & -3/4 & . \quad -1/4 \\ 0 & 1 & 3/4 & . \quad 1/4 \\ 0 & 0 & 27/4 & . \quad 1/4 \end{array}\right] \begin{array}{c} \text{Normalisation} \\ (\frac{L_3}{27/4}) \\ \longmapsto \end{array} \left[\begin{array}{cccc} & & & . \\ 1 & 0 & -3/4 & . \quad -1/4 \\ 0 & 1 & 3/4 & . \quad 1/4 \\ 0 & 0 & 1 & . \quad 1/27 \end{array}\right] \begin{array}{c} \text{Réduction} \\ L_1 + (3/4)L_3 \\ \longmapsto \\ L_2 - (3/4)L_3 \end{array}$$

$$\left[\begin{array}{cccc} & A^{(4)} = Id & & . \quad b^{(4)} = X \\ 1 & 0 & 0 & . \quad -2/9 \\ 0 & 1 & 0 & . \quad 2/9 \\ 0 & 0 & 1 & . \quad 1/27 \end{array}\right] \Rightarrow \text{la solution du système (1.16) est } X = \begin{cases} x_1 = -2/9 \\ x_2 = 2/9 \\ x_3 = 1/27 \end{cases}$$

1.6.1 Nombre d'opérations pour l'élimination de Gauss-Jordan

Si la matrice du système A est réelle, la méthode d'élimination de Gauss-Jordan nécessite $\frac{n(n^2-1)}{2}$ additions, $\frac{n(n^2-1)}{2}$ multiplications et $\frac{n(n+1)}{2}$ divisions donc soit un total de $T_{Gauss-Jordan} = \frac{2n^3+n^2-n}{2}$.
Pour $n = 5$ par exemple, on obtient $T_{Gauss-Jordan} = 135$ opérations élémentaires et $T_{Gauss} = 115$ opérations. On conclut que la méthode de Gauss-Jordan est légèrement moins efficace que la méthode d'élimination de Gauss pour la résolution des systèmes linéaires.

1.6.2 Calcul de l'inverse d'une matrice par l'élimination de G-J

La méthode d'élimination de Gauss-Jordan est très utile pour déterminer la matrice inverse d'une matrice carrée A inversible d'ordre n. Il suffit d'appliquer l'élimination de Gauss-Jordan à la matrice augmentée $[A\ Id]$ pour arriver au tableau $[Id\ A^{-1}]$.

Exemple d'application

Inverser la matrice $A = \begin{pmatrix} 2 & 8 & 4 \\ 2 & 10 & 6 \\ 1 & 8 & 2 \end{pmatrix}$ par la méthode de Gauss-Jordan.

Solution

On part de la matrice augmentée $[A \quad Id]$ et l'on trouve successivement

$$\left[\begin{array}{ccccccc} & A^{(1)} & & . & & Id & \\ 2 & 8 & 4 & . & 1 & 0 & 0 \\ 2 & 10 & 6 & . & 0 & 1 & 0 \\ 1 & 8 & 2 & . & 0 & 0 & 1 \end{array}\right] \begin{array}{c} \text{Normalisation} \\ (L_1/2) \\ \longmapsto \end{array} \left[\begin{array}{ccccccc} & & & . & & & \\ 1 & 4 & 2 & . & 1/2 & 0 & 0 \\ 2 & 10 & 6 & . & 0 & 1 & 0 \\ 1 & 8 & 2 & . & 0 & 0 & 1 \end{array}\right] \begin{array}{c} \text{Réduction} \\ L_2 - 2L_1 \\ \longmapsto \\ L_3 - L_1 \end{array}$$

$$\left[\begin{array}{ccccccc} & A^{(2)} & & . & & & \\ 1 & 4 & 2 & . & 1/2 & 0 & 0 \\ 0 & 2 & 2 & . & -1 & 1 & 0 \\ 0 & 4 & 0 & . & -1/2 & 0 & 1 \end{array}\right] \begin{array}{c}\text{Normalisation} \\ (L_2/2) \\ \longmapsto\end{array} \left[\begin{array}{ccccccc} & & & . & & & \\ 1 & 4 & 2 & . & 1/2 & 0 & 0 \\ 0 & 1 & 1 & . & -1/2 & 1/2 & 0 \\ 0 & 4 & 0 & . & -1/2 & 0 & 1 \end{array}\right] \begin{array}{c}\text{Réduction} \\ L_1 - 4L_2 \\ \longmapsto \\ L_3 - 4L_2\end{array}$$

$$\left[\begin{array}{ccccccc} & A^{(3)} & & . & & & \\ 1 & 0 & -2 & . & 5/2 & -2 & 0 \\ 0 & 1 & 1 & . & -1/2 & 1/2 & 0 \\ 0 & 0 & -4 & . & 3/2 & -2 & 1 \end{array}\right] \begin{array}{c}\text{Normalisation} \\ (L_3/(-4)) \\ \longmapsto\end{array}$$

$$\left[\begin{array}{ccccccc} & & & . & & & \\ 1 & 0 & -2 & . & 5/2 & -2 & 0 \\ 0 & 1 & 1 & . & -1/2 & 1/2 & 0 \\ 0 & 0 & 1 & . & -3/8 & 1/2 & -1/4 \end{array}\right] \begin{array}{c}\text{Réduction} \\ L_1 + 2L_3 \\ \longmapsto \\ L_2 - L_3\end{array} \left[\begin{array}{ccccccc} & A^{(4)} = Id & & . & A^{-1} & & \\ 1 & 0 & 0 & . & 7/4 & -1 & -1/2 \\ 0 & 1 & 0 & . & -1/8 & 0 & 1/4 \\ 0 & 0 & 1 & . & -3/8 & 1/2 & -1/4 \end{array}\right]$$

L'inverse de A est $A^{-1} = \begin{pmatrix} 7/4 & -1 & -1/2 \\ -1/8 & 0 & 1/4 \\ -3/8 & 1/2 & -1/4 \end{pmatrix}$

1.7 Méthode de factorisation LU de Doolittle

Soit $A = (a_{ij})$, $(1 \leq i,j \leq n)$ une matrice carrée d'ordre n et soit

$$A_k = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}$$

Théorème 1.1.
Si tous les mineurs principaux de A sont non nuls (c'est à dire $det(A_k) \neq 0$ $(\forall k = 1,...,n)$, alors il existe une unique matrice L triangulaire inférieure à diagonale unité et une unique matrice U triangulaire supérieure telles que $A = LU$.

Démonstration.
La démonstration se fait par récurrence sur n. La propriété est évidente à l'étape $n = 1$.
On suppose que la propriété est vraie à l'ordre $(n-1)$. On décompose la matrice A par

blocs sous la forme

$$A = \begin{pmatrix} A^{(n-1)} & c \\ d^t & a_{nn} \end{pmatrix} \tag{1.17}$$

où $A^{(n-1)}$ est la matrice d'ordre $(n-1) \times (n-1)$ des $(n-1)$ premières lignes et colonnes de A, c et d sont des vecteurs donnés par

$$c = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{n-1,n} \end{pmatrix} \qquad d = \begin{pmatrix} a_{n1} \\ \vdots \\ a_{n,n-1} \end{pmatrix}$$

La matrice $A^{(n-1)}$ a les mêmes mineurs principaux que A, on peut donc lui appliquer l'hypothèse de récurrence, il existe donc deux matrices $L^{(n-1)}$ triangulaire inférieure avec des 1 sur la diagonale et $U^{(n-1)}$ triangulaire supérieure telles que $A^{(n-1)} = L^{(n-1)}U^{(n-1)}$. Les matrices à chercher L et U peuvent aussi être décomposées par blocs sous la forme

$$L = \begin{pmatrix} L^{(n-1)} & 0 \\ l^t & 1 \end{pmatrix} \quad U = \begin{pmatrix} U^{(n-1)} & u \\ 0 & u_{nn} \end{pmatrix}$$

En effectuant le produit par blocs de L et U et en identifiant à la décomposition de A, on obtient le système d'équations suivantes

$$\begin{cases} A^{(n-1)} = L^{(n-1)}U^{(n-1)} \\ d^t = l^t U^{(n-1)} \quad \Rightarrow \quad l^t = d^t (U^{(n-1)})^{-1} \\ c = L^{(n-1)}u \quad \Rightarrow \quad u = (L^{(n-1)})^{-1}c \\ a_{nn} = l^t u + u_{nn} \quad \Rightarrow \quad u_{nn} = a_{nn} - d^t (U^{(n-1)})^{-1} (L^{(n-1))^{-1}} c = d^t (A^{(n-1)})^{-1} c \end{cases}$$

Pour l'unicité de la décomposition de A, supposons qu'il existe deux couples de matrices (L_1,U_1) et (L_2,U_2) tels que $A = L_1U_1 = L_2U_2$. Toutes ces matrices étant inversibles donc $(L_1U_1)^{-1} = (L_2U_2)^{-1}$ ce qui donne $U_1^{-1}L_1^{-1} = U_2^{-1}L_2^{-1}$ et aussi $U_1(U_2)^{-1} = (L_1)^{-1}L_2$.
Dans le membre de gauche, on a une matrice triangulaire supérieure et dans le membre de droite, on a une matrice triangulaire inférieure à diagonale unité.
Pour qu'elles soient égales, elles doivent coïncider avec l'identité, d'où $U_1(U_2)^{-1} = Id$ donc $U_1 = U_2$ (de même $L_1 = L_2$). □

1.7.1 Algorithme de la factorisation LU de Doolittle

L'algorithme de Doolittle est donné par les formules suivantes:

$$\begin{cases} u_{mj} = a_{mj} - \sum_{k=1}^{m-1} l_{mk}.u_{kj} & j = m,m+1,...,n \\ l_{mm} = 1 & 1 \leq m \leq n \\ l_{im} = \frac{1}{u_{mm}}(a_{im} - \sum_{k=1}^{m-1} l_{ik}.u_{km}) & i = m+1,m+2,...,n \end{cases} \tag{1.18}$$

Remarque 1.11.
Une méthode de factorisation LU dite de Crout est obtenue de manière similaire à celle de Doolittle mais en choisissant la matrice U à diagonale unité c'est à dire $u_{ii} = 1 \ \ \forall i = 1,...,n$.

1.7.2 Résolution du système $AX = b$ par la factorisation LU

Comme $A = LU$ d'où $AX = LUX = L(UX) = LY = b$ donc la résolution de $AX = b$ revient à la résolution de deux systèmes triangulaires suivants:

(1) LY=b système triangulaire inférieur à résoudre par la descente pour obtenir Y.

(2) UX=Y système triangulaire supérieur à résoudre par la remontée pour obtenir la solution X du système initial $AX = b$.

Remarque 1.12.
Grâce à la factorisation LU on peut calculer le déterminant d'une matrice carrée. En effet, $det(A) = det(LU) = (detL).(detU) = \prod_{i=1}^{i=n} u_{ii}$ car $detL = 1$.

Exemple d'application

Appliquer l'algorithme de factorisation LU pour résoudre le système suivant:

$$A = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{pmatrix} \quad X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Solution

Les trois sous matrices A_k pour $k = 1,...,3$ étant inversibles (car $det(A_k) \neq 0$) donc on peut décomposer A sous forme A=LU. En appliquant les formules (1.18) on a
Pour m=1 on a j=1,2,3 et i=2,3 alors $u_{11} = a_{11} = 1$; $u_{12} = a_{12} = 4$;
$u_{13} = a_{13} = 7$; $l_{21} = a_{21} = 2$ et $l_{31} = a_{31} = 3$.
Pour m=2 on a j=2,3 et i=3 alors $u_{22} = a_{22} - l_{21}u_{12} = 5 - 2(4) = -3$; $u_{23} = a_{23} - l_{21}u_{13} = 8 - 2(7) = -6$ et $l_{32} = \frac{1}{u_{22}}(a_{32} - l_{31}u_{12}) = 2$.
Pour m=3 on a j=3 alors $u_{33} = a_{33} - l_{31}u_{13} - l_{32}u_{23} = 1$.
Finalement, les matrices de factorisation sont:

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \quad U = \begin{pmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{pmatrix}$$

La résolution par la descente du système triangulaire inférieur $LY = b$ donne pour solution $Y = (1, -1,0)^t$.
La résolution par la remontée du système triangulaire supérieur $UX = Y$ donne pour solution $X = (\frac{-1}{3},\frac{1}{3},0)^t$ qui est la solution du système de départ $AX = b$.

1.8 Factorisation de Cholesky

Cette section traite le cas particulier des matrices symétriques définies positives. Pour de telles matrices, la méthode d'élimination de Gauss s'applique toujours puisque ces matrices sont aussi inversibles.

Théorème 1.2. *Factorisation de Cholesky*
Soit A une matrice symétrique définie positive alors il existe une matrice B triangulaire inférieure telle que $A = B.B^t$.

Démonstration.
Comme A est définie positive donc $det(A_k) > 0$ pour tout $k = 1,...,n$ alors on peut appliquer la factorisation LU de A où L est triangulaire inférieure à diagonale unité et U triangulaire supérieure.
De plus, les éléments diagonaux u_{kk} de la matrice U étant strictement positifs car
$det(A_k) = u_{11}u_{22}...u_{kk} > 0$ du fait que A est symétrique définie positive.
On intercale la matrice $\Lambda = diag(\sqrt{u_{ii}})$ diagonale d'éléments diagonaux $\sqrt{u_{ii}}$.

Soit $A = L.U = L(\Lambda.\Lambda^{-1})U = (L\Lambda).(\Lambda^{-1}U) = B.C = A$
Comme $A = A^t$ donc $B.C = (B.C)^t = C^t.B^t$ et $(C^t.B^t)^{-1} = (B.C)^{-1} = C^{-1}.B^{-1}$
$\Rightarrow (C^t.B^t)^{-1} = C^{-1}.B^{-1} \Rightarrow (B^t)^{-1}.(C^t)^{-1} = C^{-1}.B^{-1} \Rightarrow (B^t)^{-1}.(C^t)^{-1}.C^t = C^{-1}.B^{-1}.C^t$
$\Rightarrow (B^t)^{-1} = C^{-1}.B^{-1}.C^t \Rightarrow C.(B^t)^{-1} = B^{-1}.C^t$.
Comme $C.(B^t)^{-1}$ est triangulaire supérieure et $B^{-1}.C^t$ est triangulaire inférieure donc leur égalité prouve qu'elles sont diagonales et leurs éléments diagonaux sont égaux à 1, ceci prouve que $B^{-1}.C^t = Id$ c'est à dire $C^t = B$ ou que $C = B^t$ d'où $A = B.C = B.B^t$. □

Remarque 1.13.
Si on impose aux éléments diagonaux de B d'être strictement positifs alors la factorisation $A = B.B^t$ est unique.

1.8.1 Algorithme de la factorisation de Cholesky

La construction de la matrice de Cholesky est donnée par les formules suivantes:

$$\begin{cases} b_{11} = \sqrt{a_{11}} & \\ b_{i1} = \dfrac{a_{i1}}{b_{11}} & i = 2,3,...,n \\ b_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} b_{jk}^2} & j = 2,3,...,n \\ b_{ij} = (a_{ij} - \sum_{k=1}^{j-1} b_{ik}.b_{jk})/b_{jj} & i = j+1,j+2,...,n \end{cases} \tag{1.19}$$

Résolution du système $AX = b$ par Cholesky

Après décomposition de $A = B.B^t$, le système $AX = b$ se ramène alors à la résolution de deux systèmes triangulaires $B.Y = b$ et $B^t.X = Y$.

Remarque 1.14.
Grâce à la factorisation de Cholesky on peut calculer le déterminant d'une matrice carrée. En effet, $det(A) = det(B.B^t) = (detB).(detB^t) = (det(B))^2 = \prod_{i=1}^{i=n} (b_{ii})^2$.

Exemple d'application

Résoudre le système $AX = V$ suivant par la méthode de Cholesky.

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 8 & 2 \\ 1 & 2 & 10 \end{pmatrix} \quad X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad V = \begin{pmatrix} 10 \\ -5 \\ 1 \end{pmatrix}$$

Solution

1) Il faut d'abord vérifier que la matrice A est symétrique définie positive.
– On a $A = A^t \Rightarrow A$ est symétrique.
– Montrons que les déterminants mineurs $\Delta_k > 0$ pour k = 1,...,3. $\Delta_1 = 1 > 0$;
$\Delta_2 = \begin{vmatrix} 1 & 2 \\ 2 & 8 \end{vmatrix} = 4 > 0$ *et* $\Delta_3 = det(A) = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 8 & 2 \\ 1 & 2 & 10 \end{vmatrix} = 36 > 0.$
A étant symétrique définie positive donc on peut appliquer la factorisation de Cholesky.
2) Détermination de la matrice de Cholesky B en appliquant les formules (1.19).
On a $b_{11} = \sqrt{a_{11}} = 1$; pour i=2 et i=3, $b_{21} = \frac{a_{21}}{b_{11}} = 2$ et $b_{31} = \frac{a_{31}}{b_{11}} = 1$.
Pour j=2, $b_{22} = (a_{22} - b_{21}^2)^{\frac{1}{2}} = 2$; $b_{32} = (a_{32} - b_{31}.b_{21})/b_{22} = 0$.
et enfin $b_{33} = (a_{33} - (b_{31}^2 + b_{32}^2))^{\frac{1}{2}} = 3$.
d'où la matrice de cholesky est $B = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix}$ *et* $B^t = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$
3) La résolution par la descente du système triangulaire inférieur $BY = V$ donne pour solution $Y = (10, -25/2, -3)^t$.
La résolution par la remontée du système triangulaire supérieur $B^tX = Y$ donne pour solution $X = (-\frac{3}{2}, -25/4, -1)^t$ qui est la solution du système de départ $AX = V$.

1.8.2 Coût de la méthode de Cholesky

Le coût de calcul de la matrice de Cholesky B est de $C_L = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}$ opérations élémentaires auxquelles s'ajoutent le coût de la résolution des deux systèmes triangulaires $BY = b$ et $B^tX = Y$ à savoir n^2 opérations élémentaires chacun. Donc

Le total $T_{Cholesky} = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} + 2n^2 = \frac{2n^3 + 15n^2 + n}{6}$ opérations élémentaires.

Pour n=5, on obtient $T_{Cholesky} = 105$ alors que $T_{Gauss} = 115$ opérations élémentaires.
Pour n=10, on obtient $T_{Cholesky} = 585$ alors que $T_{Gauss} = 805$ opérations élémentaires.
Le gain de la méthode de Cholesky est plus intéressant pour de grands systèmes que l'élimination de Gauss.
De plus, la méthode de Cholesky est fortement conseillée (plutôt que Gauss) quand la matrice est symétrique définie positive.

1.9 Exercices corrigés sur le chapitre

Exercice 1 Soit une matrice A carrée symétrique d'ordre 4 telle que $\forall X \in \mathbb{R}^4 : X^t.A.X = x_1^2 + x_2^2 + (x_3 + x_4)^2 + x_4^2$ où $X^t = (x_1,x_2,x_3,x_4)$.

(1) Déterminer la matrice A ainsi définie.

(2) On pose $b = (1,1,2,3)^t$. Résoudre le système $AX = b$ par la méthode de Gauss.

(3) En déduire le déterminant de la matrice A.

Solution de l'exercice 1:

(1) La matrice A étant carrée symétrique d'ordre 4 d'où $A = \begin{pmatrix} a & b & c & d \\ b & y & z & t \\ c & z & e & f \\ d & t & f & h \end{pmatrix}$.

Par Hypothèses, on a $X^t.A.X = x_1^2 + x_2^2 + (x_3 + x_4)^2 + x_4^2 = x_1^2 + x_2^2 + x_3^2 + 2x_4^2 + 2x_3x_4$.
Le calcul de l'expression donne $X^t.A.X = ax_1^2 + bx_1x_2 + cx_1x_3 + dx_1x_4 + bx_1x_2 + yx_2^2 + zx_2x_3 + tx_2x_4 + cx_1x_3 + zx_2x_3 + ex_3^2 + fx_3x_4 + dx_1x_4 + tx_2x_4 + fx_3x_4 + hx_4^2$.
En identifiant les deux expressions de $X^t.A.X$, on trouve $a = 1$, $y = 1$, $e = 1$, $h = 2$, $f = 1$, $b = c = d = z = t = 0$. Ainsi, la matrice $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$.

(2) Résolution du système $AX = b$ par la méthode d'élimination de Gauss.

$$\left(\begin{array}{cccc|c} & & A^{(1)} & & b^{(1)} \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 & 3 \end{array}\right) \xrightarrow[L_4 - L_3 \longrightarrow]{\text{opérations}} \left(\begin{array}{cccc|c} & & A^{(2)} & & b^{(2)} \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right)$$

Le système final $A^{(2)}X = b^{(2)}$ étant triangulaire supérieur, il sera résolu par la méthode de la remontée. La solution obtenue $X = (1,1,1,1)^t$ est la solution du système de départ $AX = b$.

(3) $det(A) = (-1)^{p=0} det(A^{(2)}) = 1$.

Exercice 2 Soient les deux matrices suivantes:

$$A_\alpha = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & 0 & \alpha \end{pmatrix} \text{ pour } \alpha \in \mathbb{R}^* \text{ et } \quad B_\alpha = \frac{1}{4\alpha} \begin{pmatrix} -2\alpha & 2\alpha & 0 \\ 3\alpha - 2 & -\alpha + 2 & -2 \\ 4 & -4 & 4 \end{pmatrix}$$

(1) Montrer que A_α est inversible si et seulement si $\alpha \neq 0$.

(2) Pour $\alpha \neq 0$, en déduire que $\forall n \in \mathbb{N}^*$ et $\forall b \in \mathbb{R}^3$ le système $A_\alpha^n.X = b$ admet une solution unique.

(3) Pour $\alpha \neq 0$, calculer $A_\alpha.B_\alpha$ puis résoudre le système $A_\alpha.X = b$ où $b = (1,2,3)^t$.

(4) Résoudre le système $A_1.X = b$ et déduire la résolution du système $A_1^2.X = b$.

(5) Déduire les solutions du système $A_1^4.B_1^2.X = b$.

Solution de l'exercice 2:

(1) La matrice A_α est inversible ssi $det(A_\alpha) \neq 0$. Comme $det(A_\alpha) = -4\alpha \neq 0$ ssi $\alpha \neq 0$.

(2) Le système $A_\alpha^n.X = b$, $\forall n \in \mathbb{N}$ et $\forall b \in \mathbb{R}$ admet une solution unique ssi $det(A_\alpha^n) \neq 0$. Comme $det(A_\alpha^n) = (detA_\alpha)^n \neq 0$ car d'après la question (1) $det(A_\alpha) \neq 0$ pour $\alpha \neq 0$.

(3) On trouve $A_\alpha.B_\alpha = B_\alpha.A_\alpha = Id_3$ ce qui implique que $B_\alpha = A_\alpha^{-1}$.
La résolution du système $A_\alpha.X = b$ donne pour solution $X = A_\alpha^{-1}.b = B_\alpha.b = \left(\frac{1}{2}, \frac{\alpha - 4}{4\alpha}, \frac{2}{\alpha}\right)^t$ avec $b = (1,2,3)^t$.

(4) Pour $\alpha = 1$ dans la solution précédente, la solution du système $A_1.X = b$ est $X = \left(\frac{1}{2}, \frac{-3}{4}, 2\right)^t$. La résolution du système $A_1^2.X = A_1.(A_1.X) = b$ est équivalente à la résolution de deux systèmes $A_1.Y = b$ et $A_1.X = Y$. La solution de $A_1.Y = b$ est $Y = \left(\frac{1}{2}, \frac{-3}{4}, 2\right)^t$ et la solution de $A_1.X = Y$ est $X = A_1^{-1}.Y = B_1.Y = \left(\frac{-5}{8}, \frac{-17}{16}, \frac{13}{4}\right)^t$.
On conclut que la solution du système $A_1^2.X = b$ est $X = \left(\frac{-5}{8}, \frac{-17}{16}, \frac{13}{4}\right)^t$.

(5) On a $A_1^4.B_1^2.X = (A_1^2.B_1)^2.X = (A_1.(A_1.B_1))^2.X = (A_1.Id_3)^2.X = b$ car
$A_1.B_1 = A_1.A_1^{-1} = Id_3$ donc on conclut que les systèmes $A_1^4.B_1^2.X = b$ et $A_1^2.X = b$ sont équivalents donc ont la même solution $X = \left(\frac{-5}{8}, \frac{-17}{16}, \frac{13}{4}\right)^t$.

Exercice 3 Soit la matrice A et le vecteur b définis par:

$$A = \begin{pmatrix} m & m & m \\ 1 & m & m \\ 1 & m & m \end{pmatrix}, \quad \text{et} \quad b = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \quad \text{où} \quad m \in \mathbb{R}$$

1°) Le système $AX = b$ admet-il une solution unique?

2°) On considère la matrice A_α définie par $A_\alpha = \begin{pmatrix} m & m & m \\ 1 & m & m+\alpha \\ 1 & m & m \end{pmatrix}$

Déterminer les valeurs de m pour lesquelles le système $A_\alpha X = b$ admet une solution unique.

3°) Pour ces valeurs de m,

(1) Utiliser la méthode de Gauss pour résoudre le système $A_\alpha X = b$.

(2) Que peut-on dire de la solution lorsque $\alpha \mapsto 0$?

4°) Considérons le système $A_\alpha X = b_\varepsilon$ où $b_\varepsilon = (1,\varepsilon,1)^t$.

(1) Déterminer par la méthode de Gauss la solution de ce système.

(2) Discuter l'existence de la solution lorsque $\alpha \mapsto 0$.

5°) Considérons la matrice C définie par $C = \begin{pmatrix} m & 1 & m \\ 1 & 1 & 1 \\ m & 1 & m^2 \end{pmatrix}$

(1) Pour quelle valeur de m cette matrice est-elle symétrique définie positive?

(2) En utilisant la méthode de Cholesky, résoudre le système $CX = d$ pour $m = 2$ et $d = (1,1,1)^t$.

Solution de l'exercice 3:

(1) La matrice A admet deux colonnes (lignes) identiques donc son déterminant est nul et le système $AX = b$ n'admet pas de solutions pour tout $m \in \mathbb{R}$.

(2) Le système $A_\alpha X = b$ admet une solution unique ssi $det(A_\alpha) \neq 0$. Le calcul du déterminant de A_α donne $det(A_\alpha) = m\alpha(1-m)$. On conclut que $A_\alpha X = b$ admet une solution unique ssi $m \neq 0$ ou $m \neq 1$.

$(3)_{(1)}$ Pour $m = 0$, résolution de $A_\alpha X = b$ par la méthode d'élimination de Gauss

$$\begin{pmatrix} & A_\alpha^{(1)} & & . & b^{(1)} \\ 0 & 0 & 0 & . & 1 \\ 1 & 0 & \alpha & . & 2 \\ 1 & 0 & 0 & . & 1 \end{pmatrix} \begin{array}{c} \text{opérations} \\ L_1 \leftrightarrow L_3 \\ \longrightarrow \end{array} \begin{pmatrix} & A_\alpha^{(2)} & & . & b^{(2)} \\ 1 & 0 & 0 & . & 1 \\ 1 & 0 & \alpha & . & 2 \\ 0 & 0 & 0 & . & 1 \end{pmatrix}$$

$$\begin{array}{c}\text{opérations}\\ L_2-L_1\\ \longrightarrow\end{array}\begin{pmatrix} & A_\alpha^{(3)} & & . & b^{(3)}\\ 1 & 0 & 0 & . & 1\\ 0 & 0 & \alpha & . & 1\\ 0 & 0 & 0 & . & 1\end{pmatrix}$$

Le système $A_\alpha^{(3)}X = b^{(3)}$ étant impossible donc le système $A_\alpha X = b$ n'admet pas de solutions pour $m = 0$.

Pour $m = 1$, résolution de $A_\alpha X = b$ par la méthode d'élimination de Gauss:

$$\begin{pmatrix} & A_\alpha^{(1)} & & . & b^{(1)}\\ 1 & 1 & 1 & . & 1\\ 1 & 1 & 1+\alpha & . & 2\\ 1 & 1 & 1 & . & 1\end{pmatrix}\begin{array}{c}\text{opérations}\\ L_2-L_1\\ \longrightarrow\\ L_3-L_1\end{array}\begin{pmatrix} & A_\alpha^{(2)} & & . & b^{(2)}\\ 1 & 1 & 1 & . & 1\\ 0 & 0 & \alpha & . & 1\\ 0 & 0 & 0 & . & 0\end{pmatrix}$$

Le système $A_\alpha^{(2)}X = b^{(2)}$ équivalent au système $A_\alpha X = b$ admet une infinité de solutions pour $m = 1$.

$(3)_{(2)}$ Si $\alpha \mapsto 0$, le système $A_\alpha X = A_0 X = AX = b$ n'admet pas de solutions d'après la question 1°).

$(4)_{(1)}$ Résolution du système $A_\alpha X = b_\varepsilon$ où $b_\varepsilon = (1,\varepsilon,1)^t$.

$$\begin{pmatrix} & A_\alpha^{(1)} & & . & b_\varepsilon^{(1)}\\ m\neq 0 & m & m & . & 1\\ 1 & m & m+\alpha & . & \varepsilon\\ 1 & m & m & . & 1\end{pmatrix}\begin{array}{c}\text{opérations}\\ L_2-L_1/m\\ \longrightarrow\\ L_3-L_1/m\end{array}\begin{pmatrix} & A_\alpha^{(2)} & & . & b_\varepsilon^{(2)}\\ m & m & m & . & 1\\ 0 & m-1\neq 0 & m+\alpha-1 & . & \varepsilon-1/m\\ 0 & m-1 & m-1 & . & 1-1/m\end{pmatrix}$$

$$\begin{array}{c}\text{opérations}\\ L_3-L_2\\ \longrightarrow\end{array}\begin{pmatrix} & A_\alpha^{(3)} & & . & b_\varepsilon^{(3)}\\ m & m & m & . & 1\\ 0 & m-1 & m+\alpha-1 & . & \varepsilon-1/m\\ 0 & 0 & -\alpha\neq 0 & . & 1-\varepsilon\end{pmatrix}$$

Le système $A_\alpha^{(3)}X = b_\varepsilon^{(3)}$ est triangulaire supérieur, il est résolu par la méthode retour arrière et sa solution $X = \left(0, -\dfrac{\varepsilon}{\alpha}+\dfrac{1}{\alpha}+\dfrac{1}{m}, \dfrac{\varepsilon-1}{\alpha}\right)^t$ pour $m \neq 0$, $m \neq 1$ et $\alpha \neq 0$ est la solution du système $A_\alpha X = b_\varepsilon$ où $b_\varepsilon = (1,\varepsilon,1)^t$.

$(4)_{(2)}$ Si $\alpha = 0$ et $\varepsilon = 1$, le système admet une infinité de solutions et si $\alpha = 0$ et $\varepsilon \neq 1$, le système n'admet pas de solutions.

$(5)_{(1)}$ On a $C = C^t$ pour $\forall m \in \mathbb{R}$.

La matrice C est définie positive si les mineurs principaux $\Delta_k > 0$, $\forall k = 1,...,3$.
$\Delta_1 = m > 0$ et $\Delta_2 = \begin{vmatrix} m & 1 \\ 1 & 1 \end{vmatrix} = m - 1 > 0$ si $m > 1$ et
$\Delta_3 = det(C) = m(m-1)^2 > 0$ si $m > 0$.

On conclut que la matrice C est symétrique définie positive si $m > 1$ donc on peut appliquer la méthode de Cholesky pour $m > 1$.

$(5)_{(2)}$ Résolution du système $CX = d$ où $d = (1,1,1)^t$ avec $m = 2$ par la méthode de Cholesky (décomposition de $C = B.B^t$ où B triangulaire inférieure).

Algorithme de Cholesky:
$b_{11} = \sqrt{c_{11}} = \sqrt{2}$ et $b_{i1} = \dfrac{c_{i1}}{b_{11}}$ pour $i = 2,3$ d'où $b_{21} = \dfrac{1}{\sqrt{2}}$ et $b_{31} = \dfrac{2}{\sqrt{2}}$.
$b_{jj} = (c_{jj} - \sum_{k=1}^{j-1} b_{jk}^2)^{1/2}$ pour $j = 2,3$ d'où $b_{22} = \dfrac{1}{\sqrt{2}}$.
$b_{ij} = \left(c_{ij} - \sum_{k=1}^{j-1} b_{ik}.b_{jk} \right) / b_{jj}$ pour $j = j+1,...,n$.
$b_{22} = \dfrac{1}{\sqrt{2}}$, $b_{32} = (c_{32} - b_{31}b_{21})/b_{22} = 0$ et $b_{33} = (4 - b_{31}^2 - b_{32}^2)^{1/2} = \sqrt{2}$.

La matrice de Cholesky obtenue est $B = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 2/\sqrt{2} & 0 & \sqrt{2} \end{pmatrix}$

Résoudre le système $CX = d \Leftrightarrow BB^tX = d$ par la méthode de Cholesky revient à résoudre le système triangulaire inférieure $BY = d$ et le système triangulaire supérieure $B^tX = Y$.

1°) Le système $BY = d \Leftrightarrow \begin{cases} \sqrt{2}y_1 = 1 \\ (1/\sqrt{2})y_1 + (1/\sqrt{2})y_2 = 1 \\ (2/\sqrt{2})y_1 + \sqrt{2}y_3 = 1 \end{cases}$

La solution du système $BY = d$ est $Y = (y_1,y_2,y_3)^t = \left(1/\sqrt{2},1/\sqrt{2},0\right)^t$.

2°) Le système $B^tX = Y \Leftrightarrow \begin{cases} \sqrt{2}x_1 + (1/\sqrt{2})x_2 + (2/\sqrt{2})x_3 = 1/\sqrt{2} \\ (1/\sqrt{2})x_2 = 1/\sqrt{2} \\ \sqrt{2}x_3 = 0 \end{cases}$

La solution du système $B^tX = Y$ est $X = (x_1,x_2,x_3)^t = (0,1,0)^t$ qui représente la solution du système $CX = d$ pour $m = 2$.

Chapitre 2

Méthodes itératives de résolution des systèmes linéaires

2.1 Analyse matricielle, Normes

Définition 2.1. Une norme sur un $\mathbb{K}$-espace vectoriel E est une application notée $\|.\| : E \mapsto \mathbb{R}^+$ qui vérifie les propriétés suivantes:

1. $\|X\| = 0 \Leftrightarrow X = 0$,
2. $\|\alpha X\| = |\alpha|\|X\|$, $\forall\ \alpha \in \mathbb{K}$, $\forall X \in E$,
3. $\|X + Y\| \leq \|X\| + \|Y\|$, $\forall\ X,Y \in E$ (inégalité triangulaire).

Remarque 2.1. Une norme sur un espace vectoriel E est appelée norme vectorielle. De plus, un espace vectoriel est dit normé s'il est muni d'une norme.

Exemples de normes vectorielles

Les trois normes vectorielles les plus fréquemment utilisées sur $\mathbb{C}^n$ sont:

– La norme l^1 définie par $\|X\|_1 = \sum_{i=1}^{n} |x_i|$ où $X = (x_1,x_2,...,x_n)$,

– La norme l^2 définie par $\|X\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$ appelée norme euclidienne sur $\mathbb{C}^n$,

– La norme l^∞ définie par $\|X\|_\infty = \max_{1\leq i\leq n} |x_i|$.

Définition 2.2. Deux normes $\|.\|_a$ et $\|.\|_b$, définies sur un même espace vectoriel E, sont dites équivalentes s'il existe deux constantes $0 < \alpha \leq \beta$ telles que $\alpha\|X\|_a \leq \|X\|_b \leq \beta\|X\|_a$.

Conséquence

Dans un espace vectoriel E de dimension finie, toutes les normes sur E sont équivalentes.

Soit $\mathcal{M}_{m,n}(\mathbb{K})$ l'espace vectoriel, de matrices de type (m,n), de dimension finie $m.n$.

Définition 2.3. Une norme matricielle sur $\mathcal{M}_{m,n}(\mathbb{K})$ est une application notée $\|.\| : \mathcal{M}_{m,n}(\mathbb{K}) \mapsto \mathbb{R}^+$ qui vérifie les propriétés suivantes:

1. $\|A\| = 0 \Leftrightarrow A = 0$,
2. $\|\alpha A\| = |\alpha|\|A\|$, $\forall\ \alpha \in \mathbb{K}$, $\forall A \in \mathcal{M}_{m,n}(\mathbb{K})$,
3. $\|A + B\| \leq \|A\| + \|B\|$, $\forall\ A,B \in \mathcal{M}_{m,n}(\mathbb{K})$,
4. $\|A.B\| \leq \|A\|.\|B\|$, $\forall\ A,B \in \mathcal{M}_{m,n}(\mathbb{K})$.

Définition 2.4. Soit une norme vectorielle $\|.\|$ sur $\mathbb{K}^n$, on lui associe une norme matricielle dite norme matricielle subordonnée à la norme vectorielle donnée et définie par:

$$\|A\| = \sup_{X \in \mathbb{K}^n, X \neq 0} \frac{\|AX\|}{\|X\|} = \sup_{X \in \mathbb{K}^n, \|X\| \leq 1} \|AX\| = \sup_{X \in \mathbb{K}^n, \|X\| = 1} \|AX\| \tag{2.1}$$

Exemples de normes matricielles subordonnées

Les normes matricielles subordonnées, associées aux normes vectorielles usuelles l^1, l^2 et l^∞ de $\mathbb{R}^n$, sont respectivement données par:

- $\|A\|_1 = \max\limits_{1 \leq j \leq n} \left(\sum\limits_{i=1}^{m} |a_{ij}| \right)$ où $A = (a_{ij})$, $i = 1,...,m$ et $j = 1,...,n$,
- $\|A\|_2 = \left(\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}}$,
- $\|A\|_\infty = \max\limits_{1 \leq i \leq m} \left(\sum\limits_{j=1}^{n} |a_{ij}| \right)$.

Exemple d'application

Trouver les normes matricielles 1,2 et ∞ de la matrice $A = \begin{pmatrix} 2 & 1 & 1 & -1 \\ 2 & 3 & 2 & 0 \\ 1 & 1 & 2 & 3 \end{pmatrix}_{3 \times 4}$

Solution

$$\|A\|_1 = \max_{1\leq j\leq 4}\left(\sum_{i=1}^{3}|a_{ij}|\right) = \max_{1\leq j\leq 4}\left(|a_{1j}|+|a_{2j}|+|a_{3j}|\right) = \max(5,5,5,4) = 5.$$

$$\|A\|_2 = \left(\sum_{i=1}^{3}\sum_{j=1}^{4}|a_{ij}|^2\right)^{\frac{1}{2}} = \left[\sum_{i=1}^{3}\left(|a_{i1}|^2+|a_{i2}|^2+|a_{i3}|^2+|a_{i4}|^2\right)\right]^{\frac{1}{2}} = \sqrt{39}.$$

$$\|A\|_\infty = \max_{1\leq i\leq 3}\left(\sum_{j=1}^{4}|a_{ij}|\right) = \max_{1\leq i\leq 3}\left(|a_{i_1}|+|a_{i2}|+|a_{i3}|+|a_{i4}|\right) = \max(5,7,7) = 7.$$

2.1.1 Valeurs et vecteurs propres

Soit A une matrice d'ordre n, les valeurs propres de A sont les n zéros λ_i, $(i = 1,...,n)$ réelles ou complexes distinctes ou confondues, du polynôme caractéristique $P(\lambda) = det(A - \lambda Id_n)$ associé à A.
Le spectre de A est l'ensemble de ses valeurs propres. A toute valeur propre λ correspond au moins un vecteur non nul X tel que $AX = \lambda X$ appelé vecteur propre associé à la valeur propre λ.

Exemple d'application

Trouver les valeurs propres de la matrice $A = \begin{pmatrix} 1 & 3 \\ 0 & -2 \end{pmatrix}_{2\times 2}$ et donner les vecteurs propres associés à ces valeurs.

Solution

Déterminons d'abord le polynôme caractéristique $P(\lambda) = det(A - \lambda Id_2)$ associé à A,

$$P(\lambda) = det(A-\lambda Id_2) = \left|\begin{pmatrix} 1 & 3 \\ 0 & -2 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right| = \begin{vmatrix} 1-\lambda & 3 \\ 0 & -2-\lambda \end{vmatrix} = (1-\lambda)(-2-\lambda)$$

$P(\lambda) = -(1-\lambda)(2+\lambda) = 0$ si $\lambda_1 = 1$ ou $\lambda_2 = -2$, donc les valeurs propres associées à la matrice A sont $\lambda_1 = 1$ et $\lambda_2 = -2$.
Soit $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ avec $X \neq 0_{\mathbb{R}^2}$.

Pour déterminer le vecteur propre associé à la valeur propre $\lambda = 1$, il suffit de résoudre le système $AX = \lambda X = X \Leftrightarrow \begin{pmatrix} 1 & 3 \\ 0 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Leftrightarrow \begin{cases} x_1 + 3x_2 = x_1 \\ -2x_2 = x_2 \end{cases}$

Les solutions de ce système sont tous les vecteurs $X = \begin{pmatrix} x_1 \in \mathbb{R} \\ x_2 = 0 \end{pmatrix}$ et on conclut qu'il existe une infinité de vecteurs propres associés à la valeur propre $\lambda = 1$.

Définition 2.5.
Le rayon spectral de la matrice A d'ordre n est défini par $\rho(A) = \max\limits_{1 \leq i \leq n} |\lambda_i|$.

Remarque 2.2. Il est à noter que si les valeurs propres sont complexes, $|\lambda_i|$ représente la norme d'un nombre complexe qui est un réel positif.

Proposition 2.1.
Soit A une matrice carrée d'ordre n alors $\rho(A) \leq \|A\|$.

Démonstration.
Soit λ valeur propre de A et X le vecteur propre associé à λ donc $AX = \lambda X$ avec $X \neq 0_{\mathbb{R}^n}$ alors $\|AX\| = \|\lambda X\| = |\lambda|\|X\| \leq \|A\|\|X\|$ et $|\lambda|\|X\| \leq \|A\|\|X\|$ d'où $|\lambda| \leq \|A\|$ pour tout λ donc $\rho(A) \leq \|A\|$ car $\rho(A)$ est la plus grande valeur propre de A en module.

□

Définition 2.6.
La matrice A est dite définie positive si $X^t AX > 0$, $\forall X \in \mathbb{R}^n$, $X \neq 0$.

Soit $A = (a_{ij})$, $(1 \leq i,j \leq n)$ une matrice carrée d'ordre n et soit

$$A_k = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}$$

Critère de Sylvester

Pour que la matrice carrée A réelle symétrique soit définie positive il faut et il suffit que tous les mineurs principaux dominants soient strictement positifs c'est à dire $det(A_k) > 0$ $(\forall k = 1,...,n)$.

Définition 2.7. On dit que la suite de vecteurs $X^{(k)}$ converge vers le vecteur X si et seulement s'il existe une norme vectorielle telle que $\lim\limits_{k \to \infty} \|X^{(k)} - X\| = 0$.

Définition 2.8. La suite de matrices (A^k), $k \in \mathbb{N}$ converge vers la matrice A si et seulement s'il existe une norme matricielle telle que $\lim_{k\to\infty} \|A^k - A\| = 0$ (où $A^k = \underbrace{AA...A}_{k fois}$).

2.1.2 Norme matricielle, rayon spectral et convergence

Proposition 2.2.

Pour toute matrice A d'ordre n et tout $\epsilon > 0$, il existe au moins une norme matricielle $\|.\|$ telle que $\|A\| < \rho(A) + \epsilon$.

Proposition 2.3.

Les assertions suivantes sont équivalentes.

(1) $\lim_{k\to\infty} A^k = 0_{\mathcal{M}_n(\mathbb{K})} \Leftrightarrow \lim_{k\to\infty} \|A^k\| = 0$.

(2) $\lim_{k\to\infty} A^k X = 0_{\mathbb{R}^n} \Leftrightarrow \lim_{k\to\infty} \|A^k X\| = 0$ *pour tout* X.

(3) $0 \leq \rho(A) < 1$ *où* $\rho(A)$ *désigne le rayon spectral de la matrice* A.

(4) *Il existe une norme matricielle subordonnée* $\|.\|$ *telle que* $\|A\| < 1$.

Démonstration.

(1) $\Rightarrow$ (2) On a $0 \leq \|A^k X - 0_{\mathbb{R}^n}\| = \|A^k X\| \leq \|A^k\|.\|X\|$, comme $\lim_{k\to\infty} \|A^k\| = 0$. De cette double inégalité et d'après (1) on déduit que $\lim_{k\to\infty} A^k X = 0_{\mathbb{R}^n}$.

(2) $\Rightarrow$ (3) Prouvons la contraposée, non (3) est équivalent à $\rho(A) \geq 1$ donc il existe au moins une valeur propre λ telle que $|\lambda| \geq 1 \Rightarrow AX = \lambda X \Rightarrow \|AX\| = |\lambda|\|X\|$ or $AX = \lambda X \Rightarrow AAX = A(\lambda X) = \lambda(AX) = \lambda^2 X$ d'où $A^2X = \lambda^2 X$ et $A^k X = \lambda^k X$ d'où $\|A^k X\| = \|\lambda^k X\| = |\lambda|^k \|X\|$ et $\lim_{k\to\infty} \|A^k X\| = \lim_{k\to\infty} |\lambda|^k \|X\| = \begin{cases} +\infty & si\ |\lambda| > 1 \\ \|X\| & si\ |\lambda| = 1 \end{cases}$
X étant un vecteur propre donc $\|X\| \neq 0$ d'où $\lim_{k\to\infty} \|A^k X\| \neq 0$ ainsi $\lim_{k\to\infty} A^k X \neq 0_{\mathbb{R}^n}$.

(3) $\Rightarrow$ (4) D'après la proposition (2.2), $\forall \epsilon > 0$, il existe une norme $\|.\|$ telle que $\|A\| < \rho(A) + \epsilon$. On peut choisir un $\epsilon > 0$ tel que $1 - \rho(A) > \epsilon > 0$ c'est à dire $\|A\| < 1$.

(4) $\Rightarrow$ (1) Il existe une norme $\|.\| < 1$ telle que $\|A\| < 1$ d'où $\lim_{k\to\infty} \|A\|^k = 0$.
Comme $0 \leq \|A^k\| = \|\underbrace{AA...A}_{k fois}\| \leq \underbrace{\|A\|.\|A\|...\|A\|}_{k fois} = \|A\|^k$ donc $\lim_{k\to\infty} \|A^k\| = 0$.

□

Définition 2.9. Une matrice carrée d'ordre n est à diagonale strictement dominante par lignes (resp. par colonnes) si

$$|a_{ii}| > \sum_{j=1,j\neq i}^{n} |a_{ij}| \quad \left(\text{resp. } |a_{ii}| > \sum_{j=1,j\neq i}^{n} |a_{ji}|\right), \ i = 1,...,n \tag{2.2}$$

Théorème 2.1.
Soit A une matrice d'ordre n à diagonale strictement dominante (par lignes ou par colonnes). Alors, A est inversible.

Démonstration.
Faisons un raisonnement par absurde, supposons que A est à diagonale strictement dominante par lignes avec A non inversible alors il existe un vecteur non nul $X \in \mathbb{R}^n$ tel que $AX = 0 \ \Rightarrow \ \sum_{j=1}^{n} a_{ij}x_j = 0$, $(i = 1,...,n)$. Comme X est non nul donc il existe un indice $k \in \{1,...,n\}$ tel que $0 \neq |x_k| = \max_{1\leq i\leq n} |x_i|$ et on a alors $-a_{kk}x_k = \sum_{j=1,j\neq k}^{n} a_{kj}x_j \Rightarrow$
$|a_{kk}| \leq \sum_{j=1,j\neq k}^{n} |a_{kj}|.\frac{|x_j|}{|x_k|} \leq \sum_{j=1,j\neq k}^{n} |a_{kj}|$ ce qui contredit le fait que A est à diagonale strictement dominante par lignes. □

2.2 Systèmes linéaires et les méthodes itératives

2.2.1 Position du problème

Les méthodes directes étudiées au chapitre précédent sont très efficaces. Elles permettent d'obtenir la solution exacte du système linéaire considéré aux erreurs d'arrondi près. Elles ont l'inconvénient d'être très coûteuses en termes de nombre d'opérations donc de temps de calcul lorsque la taille du système est assez élevé. Dans ces conditions, il devient beaucoup plus commode de trouver les racines du système par des méthodes numériques dites méthodes itératives.
Dans ce chapitre, on va présenter des méthodes itératives parmi les plus simples à mettre en oeuvre, à savoir les méthodes de Jacobi et de Gauss-Seidel.

On cherche à résoudre le système linéaire $AX = b$ où A est une matrice carrée d'ordre n à coefficients dans $K = \mathbb{R}$ *ou* $\mathbb{C}$ et b un vecteur d'ordre n dans $\mathbb{K}$.

Définition 2.10. : Méthode itérative ou indirecte
On appelle méthode itérative de résolution du système linéaire $AX = b$ une méthode qui consiste à construire une suite de vecteurs $(X^{(k)})_{k\in\mathbb{N}}$ qui converge vers la solution X du système quand $k \mapsto \infty$.

2.3 Convergence des méthodes itératives

Définition 2.11.
Une méthode itérative est dite convergente si pour tout choix du vecteur initial $X^{(0)} \in \mathbb{R}^n$, on a $\lim_{k\to\infty} X^{(k)} = X$.

Les méthodes itératives, que nous allons étudier dans ce chapitre, consiste à transformer le système linéaire $AX = b$ en un système équivalent de la forme $CX + l = X$ où C est une matrice d'ordre n à coefficients dans $\mathbb{K}$ appelée "matrice d'itération" de la méthode et l un vecteur d'ordre n dans $\mathbb{K}$. Puis on construit une suite de vecteurs $X^{(k+1)}$ définis par:

$$\begin{cases} X^{(k+1)} = CX^{(k)} + l \\ X^{(0)} \text{ donné} \end{cases} \tag{2.3}$$

qui converge vers la solution exacte du système $CX + l = X$.

Question: Comment déterminer la matrice C et le vecteur l pour que la suite $(X^{(k)})_{k\in\mathbb{N}}$ converge vers X?

Définition 2.12.
La suite itérative $(X^{(k)})_{k\in\mathbb{K}}$ définie par $\begin{cases} X^{(k+1)} = CX^{(k)} + l \\ X^{(0)} \text{ donné} \end{cases}$ converge vers le vecteur X si $\lim_{k\to\infty}(X^{(k)} - X) = 0_{\mathbb{R}^n}$.

Définition 2.13.
On appelle erreur à l'étape (k) $(k \in \mathbb{N})$, le vecteur $e^{(k)}$ défini par $e^{(k)} = X^{(k)} - X$.

Conséquence

On déduit de cette définition qu'une méthode itérative de la forme (2.3) converge si et seulement si $\lim_{k\to\infty} X^{(k)} = X \Leftrightarrow \lim_{k\to\infty} e^{(k)} = 0 \Leftrightarrow \lim_{k\to\infty} C^{(k)} = 0_{\mathcal{M}_n(\mathbb{K})}$, ceci est dû

au fait que ($e^{(k)} = X^{(k)} - X = C^k.e^{(0)}$). Ce résultat est à montrer par récurrence sur k.

Théorème 2.2. *CNS de convergence d'une méthode itérative*
La méthode itérative de la forme (2.3) *est convergente si et seulement si* $\rho(C) < 1$.

Démonstration.
($\Rightarrow$) Par contraposée, supposons que $\rho(C) \geq 1$ et montrons que la méthode itérative (2.3) ne converge pas. Si $\rho(C) \geq 1$ alors il existe un vecteur $Y \in \mathbb{R}^n$ tel que C^kY ne converge pas vers 0 d'après la proposition (2.3). En choisissant $X^{(0)} = X + Y = A^{-1}b + Y$. L'égalité $X^{(k)} - X = C^k(X^{(0)} - X)$ devient $X^{(k)} - X = C^kY$ car $Y = X^{(0)} - X$ donc $X^{(k)} - X$ ne converge pas vers 0 donc $X^{(k)}$ ne converge pas vers X quand k tend vers ∞ et la méthode ne converge pas.
($\Leftarrow$) Supposons que $\rho(C) < 1$ alors l'égalité $X^{(k)} - X = C^k(X^{(0)} - X)$ tend vers 0 d'après la proposition (2.3) donc $X^{(k)}$ tend vers $X = A^{-1}b$ quand k tend vers ∞ et la méthode converge. □

2.4 Étude de quelques méthodes itératives

Soit le système linéaire $AX = b$ à résoudre. Décomposons la matrice A inversible sous la forme $A = D - E - F = \begin{pmatrix} \ddots & & -F \\ & D & \\ -E & & \ddots \end{pmatrix}$ où

$$d_{ij} = \begin{cases} a_{ii} & \text{si } i = j \\ 0 & \text{sinon} \end{cases}, \quad e_{ij} = \begin{cases} a_{ij} & \text{si } i > j \\ 0 & \text{sinon} \end{cases}, \quad f_{ij} = \begin{cases} a_{ij} & \text{si } i < j \\ 0 & \text{sinon} \end{cases}$$

2.4.1 Méthode de Jacobi

En supposant que les $a_{ii} \neq 0$, $\forall i = 1,...,n$, le système $AX = b$ est équivalent au système suivant:

$$x_i = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j\neq i}^{j=n} a_{ij}.x_j \right), \qquad i = 1,...,n \tag{2.4}$$

Pour un vecteur initial $X^{(0)}$ donné, le calcul de $X^{(k+1)}$ est basé sur:

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j\neq i}^{j=n} a_{ij}.x_j^{(k)} \right), \qquad i = 1,...,n \text{ et } k = 1,2,... \tag{2.5}$$

La décomposition de A dans la méthode de Jacobi est $A = D - E - F$ d'où
$AX = b \Leftrightarrow (D - E - F)X = b \Leftrightarrow DX - (E + F) = b \Leftrightarrow DX = (E + F)X + b.$
Comme D est inversible (car $a_{ii} \neq 0$) alors $X = D^{-1}(E + F)X + D^{-1}b$ et comme $A = D - (E + F)$ d'où $E + F = D - A$. Par conséquent,
$X = D^{-1}(E + F)X + D^{-1}b = D^{-1}(D - A)X + D^{-1}b = (Id_n - D^{-1}A)X + D^{-1}b.$

La matrice $J = Id_n - D^{-1}A$ est appelée matrice de Jacobi associée à la matrice A. Ainsi, l'algorithme ou méthode de Jacobi est défini par :

$$\begin{cases} X^{(k+1)} = JX^{(k)} + D^{-1}b \\ \\ X^{(0)} \text{ donné} \end{cases} \tag{2.6}$$

Les éléments de la matrice J sont donnés par

$$J = (j_{ij})_{1 \leq i,j \leq n} = \begin{cases} -\dfrac{a_{ij}}{a_{ii}} & \text{si} \quad i \neq j \\ \\ 0 & \text{si} \quad i = j \end{cases} \tag{2.7}$$

Exemple d'application sur la méthode de Jacobi

On considère la matrice A et le vecteur b

$$A = \begin{pmatrix} 4 & 0 & 0 & -1 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & -2 \\ -1 & 0 & -2 & 5 \end{pmatrix} \qquad \text{et} \qquad b = \begin{pmatrix} 2 \\ 4 \\ 0 \\ 7 \end{pmatrix}$$

1°) Déterminer la matrice de Jacobi J_A associée à la matrice A.
2°) Calculer le rayon spectral $\rho(J_A)$. Que peut-on en conclure?
3°) Écrire l'algorithme de Jacobi associé à la matrice A.
4°) Calculer les vecteurs $X^{(1)}$ et $X^{(2)}$ en partant du vecteur initial $X^{(0)} = (0,0,0,0)^t$.

Solution

1°) Les éléments de la matrice J_A sont donnés par

$$J_A = (j_{ij})_{1 \leq i,j \leq 4} = \begin{cases} -\dfrac{a_{ij}}{a_{ii}} & \text{si} \quad i \neq j \\ \\ 0 & \text{si} \quad i = j \end{cases} \qquad \text{d'où} \qquad J_A = \begin{pmatrix} 0 & 0 & 0 & 1/4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \\ 1/5 & 0 & 2/5 & 0 \end{pmatrix}$$

2°) Le polynôme caractéristique $P(\lambda) = det(J_A - \lambda Id_4) = -\lambda^2(-\lambda^2 + 1/4)$, donc les valeurs propres de J_A sont $\lambda_1 = \lambda_2 = 0$, $\lambda_3 = 1/2$ et $\lambda_4 = -1/2$.
Ainsi, $\rho(J_A) = \max(0,1/2) = 1/2$ et comme $\rho(J_A) = 1/2 < 1$ donc d'après le théorème (2.2), la méthode de Jacobi associée à A est convergente.
3°) L'algorithme de Jacobi est donné par la formule suivante:

$$\begin{cases} X^{(k+1)} = J_A X^{(k)} + D^{-1}b \\ X^{(0)} \text{ donné} \end{cases} \Leftrightarrow \begin{cases} x_1^{(k+1)} = 1/4x_4^{(k)} + 1/2 \\ x_2^{(k+1)} = 1 \\ x_3^{(k+1)} = 1/2x_4^{(k)} \\ x_4^{(k+1)} = 1/5x_1^{(k)} + 2/5x_3^{(k)} + 7/5 \end{cases} \quad D^{-1}b = \begin{pmatrix} 1/2 \\ 1 \\ 0 \\ 7/5 \end{pmatrix}$$

4°) A partir de $X^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ et pour k=1,2 on a $X^{(1)} = \begin{pmatrix} 1/2 \\ 1 \\ 0 \\ 7/5 \end{pmatrix}$, $X^{(2)} = \begin{pmatrix} 17/20 \\ 1 \\ 7/10 \\ 3/2 \end{pmatrix}$.

2.4.2 Méthode de Gauss-Seidel

Une autre décomposition de A nous permet d'avoir $AX = b \Leftrightarrow (D - E - F)X = b \Leftrightarrow (D - E))X = FX + b$. Comme $a_{ii} \neq 0$, $\forall i = 1,...,n$ et $D - E$ triangulaire inférieure donc inversible d'où $X = (D - E)^{-1}FX + (D - E)^{-1}b$.
La matrice $G = (D - E)^{-1}F$ est appelée matrice de Gauss-Seidel associée à la matrice A. Ainsi, l'algorithme ou méthode de Gauss-Seidel est défini par:

$$\begin{cases} X^{(k+1)} = GX^{(k)} + (D - E)^{-1}b \\ \\ X^{(0)} \text{ donné} \end{cases} \tag{2.8}$$

et en développant la formule (2.8), on obtient

$$\begin{cases} x_i^{(k+1)} = \dfrac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij}.x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}.x_j^{(k)} \right), \qquad i = 1,...,n \text{ et } k = 1,2,... \\ x^{(0)} \text{ donné} \end{cases} \tag{2.9}$$

Remarque 2.3.
La méthode de Gauss-Seidel est une amélioration de la méthode de Jacobi. En effet, la méthode de jacobi conserve la totalité des composantes du vecteur $X^{(k)}$ pour calculer le vecteur $X^{(k+1)}$ (cette méthode a besoin de 2n cases mémoires) alors que la méthode de Gauss-Seidel substitue au fur et à mesure ces composantes (cette méthode a besoin juste de n cases mémoires).
Ce qui donne l'avantage à la méthode de Gauss-seidel de converger plus rapidement que la méthode de Jacobi.

Exemple d'application sur la méthode de Gauss-Seidel

On considère la matrice A et le vecteur b tels que:

$$A = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & -1/2 \\ 1 & 1 & 2 \end{pmatrix} \qquad \text{et} \qquad b = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

1°) Déterminer la matrice de Gauss-Seidel G_A associée à la matrice A.
2°) Calculer le rayon spectral $\rho(G_A)$. La méthode de Gauss-Seidel est-elle convergente?
3°) Donner l'algorithme de Gauss-seidel associé à la matrice A.

Solution

1°) La matrice de Gauss-Seidel est définie par $G_A = (D-E)^{-1}.F$ où

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \qquad E = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & -1 & 0 \end{pmatrix}, \qquad F = \begin{pmatrix} 0 & -1 & 1 \\ 0 & 0 & 1/2 \\ 0 & 0 & 0 \end{pmatrix}$$

Le calcul de l'inverse de $(D-E)^{-1}$ peut se faire ,par exemple, par la méthode de Gauss-Jordan et on obtient

$$(D-E)^{-1} = \begin{pmatrix} 1/2 & 0 & 0 \\ -1/4 & 1/2 & 0 \\ -1/8 & -1/4 & 1/2 \end{pmatrix} \quad \text{et} \quad G_A = (D-E)^{-1}.F = \begin{pmatrix} 0 & -1/2 & 1/2 \\ 0 & 1/4 & 0 \\ 0 & 1/8 & -1/4 \end{pmatrix}$$

2°) Le polynôme caractéristique de G_A s'écrit $P(\lambda) = det(G_A - \lambda Id_3)$
$= -\lambda(1/4-\lambda)(-1/4-\lambda)$ et les valeurs propres de G_A sont $\lambda_1 = 0$, $\lambda_2 = 1/4$ et $\lambda_3 = -1/4$ d'où $\rho(G_A) = 1/4 = 0{,}25 < 1$ donc la méthode de Gauss-Seidel associée à A est convergente.
3°) L'algorithme de Gauss-seidel est donné par la formule suivante:

$$\begin{cases} X^{(k+1)} = G_A X^{(k)} + (D-E)^{-1}b \\ \\ X^{(0)} \text{ donné} \end{cases} \Leftrightarrow \begin{cases} x_1^{(k+1)} = -1/2x_2^{(k)} + 1/2x_3^{(k)} + 1/2 \\ x_2^{(k+1)} = 1/4x_2^{(k)} - 1/4 \\ x_3^{(k+1)} = 1/8x_2^{(k)} - 1/4x_3^{(k)} + 3/8 \end{cases}$$

2.4.3 Convergence des méthodes de Jacobi et de Gauss-Seidel

Théorème 2.3. *Condition suffisante de convergence des méthodes de Jacobi et G-Seidel*
Si A est une matrice à diagonale strictement dominante alors les méthodes de Jacobi et Gauss-seidel sont convergentes.

Démonstration.
1°) Prouvons la convergence concernant la méthode de Jacobi. La matrice de Jacobi associée à la matrice A étant définie par $J = (j_{ij})_{1\leq i,j\leq n} = \begin{cases} -\dfrac{a_{ij}}{a_{ii}} & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases}$

Comme A est à diagonale strictement dominante alors $|a_{ii}| > \sum_{j=1,j\neq i}^{j=n} |a_{ij}|$, $(\forall i = 1,...,n)$
ou bien $\sum_{j=1,j\neq i}^{j=n} |\frac{a_{ij}}{a_{ii}}| < 1$, $(\forall i = 1,...,n) \Rightarrow r = \max_{1\leq i\leq n} \sum_{j=1,j\neq i}^{j=n} |\frac{a_{ij}}{a_{ii}}| < 1 \Rightarrow$
$r = \max_{1\leq i\leq n} \sum_{j=1}^{j=n} |j_{ij}| < 1$ car $j_{ii} = 0 \Rightarrow \|J\|_\infty < 1 \Rightarrow$ la méthode de Jacobi converge d'après la proposition (2.3).

2°) Prouvons maintenant la convergence concernant la méthode de Gauss-Seidel.
Considérons l'erreur à l'étape $(k+1)$ pour $k \in \mathbb{N}$ de la méthode de Gauss-Seidel qui vérifie

$$e_i^{(k+1)} = -\sum_{j=1}^{i-1} \frac{aij}{aii} e_j^{(k+1)} - \sum_{j=i+1}^{n} \frac{aij}{aii} e_j^{(k)}, \quad i = 1,...,n$$

En raisonnant par récurrence sur l'indice $i = 1,...,n$ des composantes du vecteur $e_i^{(k+1)}$, montrons que $\|e^{(k+1)}\|_\infty \leq r.\|e^{(k)}\|_\infty$ pour $\forall k \in \mathbb{N}$.
Supposons que $\left|e_j^{(k+1)}\right| \leq r.\|e^k\|_\infty$ pour $j = 1,...,i-1$, on a alors

$$\left|e_j^{(k+1)}\right| \leq \sum_{j=1}^{i-1} \left|\frac{aij}{aii}\right| . \left|e_j^{(k+1)}\right| + \sum_{j=i+1}^{n} \left|\frac{aij}{aii}\right| . \left|e_j^{(k)}\right| \leq r\|e^{(k)}\|_\infty$$

et par conséquent $\|e^{(k+1)}\|_\infty \leq r\|e^{(k)}\|_\infty$ d'où la formule

$$\|e^{(k)}\|_\infty \leq r\|e^{(k-1)}\|_\infty \leq r^2\|e^{(k-2)}\|_\infty \leq ... \leq r^k.\|e^{(0)}\|_\infty \;\Rightarrow\; \|e^{(k)}\|_\infty \leq r^k.\|e^{(0)}\|_\infty$$

Par passage à la limite quand $k \mapsto \infty$ et comme $r < 1$ on obtient $\lim_{k\mapsto+\infty} \|e^{(k)}\|_\infty = 0$ d'où la convergence du processus de Gauss-seidel. □

Remarque 2.4. Condition suffisante de convergence de la méthode de Gauss-Seidel
Si A est une matrice symétrique définie positive, alors la méthode de Gauss- Seidel converge.

2.5 Nombre d'itérations nécessaires

Théorème 2.4.
Le nombre d'itérations nécessaires à une méthode itérative pour réduire l'erreur d'un facteur ε ($\varepsilon > 0$) est donné par la formule:

$$k \geq \frac{Ln\varepsilon}{Ln\rho(C)} \tag{2.10}$$

Démonstration.
On a déjà montré que si on approche le vecteur X par le $X^{(k)}$ alors l'erreur commise vérifie $\|X^{(k)} - X\| = \|e^{(k)}\| = \|C^k.e^{(0)}\| \leq \|C^k\|\|e^{(0)}\|$.
On cherche à déterminer k tel que $\|X^{(k)} - X\| \leq \varepsilon$ ce qui revient à prendre k tel que $\|C^k\| \leq \varepsilon$.
D'après la proposition (2.1), on a $\rho(C) \leq \|C\|$ pour $\forall C$ d'où $\rho(C^k) \leq \|C^k\|$ et comme $\rho(C^k) = (\rho(C))^k$ donc $(\rho(C))^k \leq \|C^k\| \leq \varepsilon \Rightarrow Ln(\rho(C))^k \leq Ln\varepsilon \Rightarrow kLn(\rho(C)) \leq Ln\varepsilon$
$\Rightarrow k \geq \dfrac{Ln\varepsilon}{Ln\rho(C)}$ car $\rho(C) < 1$.
Donc le nombre d'itérations nécessaires N pour approcher la solution exacte d'un système par une méthode itérative $X = CX + l$ est donné par

$$N = E\left[\frac{Ln\varepsilon}{Ln\rho(C)}\right] + 1 \text{.....Formule à utiliser quand } \rho(C) \text{ est connue} \tag{2.11}$$

□

2.5.1 Estimation de l'erreur d'un processus itératif

Soient $X^{(k+1)}$ et $X^{(k)}$ ($k \geq 1$) deux approximations successives de la solution du système linéaire $X = CX + l$. Pour $p \geq 1$, on peut écrire

$$\|X^{(k+p)} - X^{(k)}\| \leq \|X^{(k+p)} - X^{(k+p-1)}\| + ... + \|X^{(k+2)} - X^{(k+1)}\| + \|X^{(k+1)} - X^{(k)}\| \tag{2.12}$$

Comme $X^{(m+1)} = CX^{(m)} + l$ et $X^{(m)} = CX^{(m-1)} + l$, on a alors
$X^{(m+1)} - X^{(m)} = C(X^{(m)} - X^{(m-1)})$ et donc

$$\|X^{(m+1)} - X^{(m)}\| \leq \|C\|\|X^{(m)} - X^{(m-1)}\| \leq \|C\|^{m-k}\|X^{(k+1)} - X^{(k)}\|, \quad \text{pour} \quad m > k \geq 1$$

La formule (2.12) devient

$$\|X^{(k+p)} - X^{(k)}\| \leq \|C\|^{p-1}\|X^{(k+1)} - X^{(k)}\| + ... + \|C\|\|X^{(k+1)} - X^{(k)}\| + \|X^{(k+1)} - X^{(k)}\|$$

Et $\|X^{(k+p)} - X^{(k)}\| \leq [1 + \|C\| + ... + \|C\|^{p-1}] \|X^{(k+1)} - X^{(k)}\|$

$$\|X^{(k+p)} - X^{(k)}\| \leq \left[\frac{1 - \|C\|^p}{1 - \|C\|}\right] \|X^{(k+1)} - X^{(k)}\|$$

En passant à la limite dans la dernière inégalité quand $p \mapsto \infty$, on obtient finalement

$$\|X - X^{(k)}\| \leq \frac{\|X^{(k+1)} - X^{(k)}\|}{1 - \|C\|} \tag{2.13}$$

On a aussi

$$\|X - X^{(k)}\| \leq \frac{\|C\|^k}{1 - \|C\|}\|X^{(1)} - X^{(0)}\| \tag{2.14}$$

En particulier, si l'on choisit $X^{(0)} = l$, il vient $X^{(1)} = Cl + l$ et aussi
$\|X^{(1)} - X^{(0)}\| = \|Cl\| \leq \|C\|\|l\|$ et par conséquent

$$\|X - X^{(k)}\| \leq \frac{\|C\|^{k+1}}{1 - \|C\|}\|l\| \tag{2.15}$$

La convergence des méthodes itératives se teste généralement en utilisant un paramètre de tolérance ε. On arrête ainsi les calculs dès que l'erreur relative est assez petite.
$\|X - X^{(k)}\| \leq \frac{\|C\|^{k+1}}{1 - \|C\|}\|l\| \leq \varepsilon \Rightarrow \|C\|^k \leq \frac{(1 - \|C\|)\varepsilon}{\|l\|.\|C\|}$. Comme $\|C\| < 1$ donc le nombre d'itérations nécessaires N pour approcher la solution exacte $X = A^{-1}b$ d'un système par une méthode itérative $X = CX + l$ est donné par

$$N = E\left[\frac{Ln\left(\frac{(1 - \|C\|)\varepsilon}{\|l\|.\|C\|}\right)}{Ln\|C\|}\right] + 1 \text{.....Formule à utiliser quand } \|C\| \text{ est connue} \tag{2.16}$$

Exemple d'application

1°) Montrer que pour le système $AX = b$ suivant, le processus itératif converge.

$$\begin{cases} 10x_1 - x_2 + 2x_3 - 3x_4 = 0 \\ x_1 + 10x_2 - x_3 + 2x_4 = 5 \\ 2x_1 + 3x_2 + 20x_3 - x_4 = -10 \\ 3x_1 + 2x_2 + x_3 + 20x_4 = 15 \end{cases} \tag{2.17}$$

2°) Combien d'itérations faut-il réaliser pour trouver la solution du système (2.17) à 10^{-4} près.

Solution

1°) En réduisant le système (2.17) à la forme itérative, on obtient

$$\begin{cases} x_1 = 0{,}1x_2 - 0{,}2x_3 + 0{,}3x_4 \\ x_2 = -0{,}1x_1 + 0{,}1x_3 - 0{,}2x_4 + 0{,}5 \\ x_3 = -0{,}1x_1 - 0{,}15x_2 + 0{,}05x_4 - 0{,}5 \\ x_4 = -0{,}15x_1 - 0{,}1x_2 - 0{,}05x_3 + 0{,}75 \end{cases} \tag{2.18}$$

On en tire la matrice d'itération du système $C = \begin{pmatrix} 0 & 0{,}1 & -0{,}2 & 0{,}3 \\ -0{,}1 & 0 & 0{,}1 & -0{,}2 \\ -0{,}1 & -0{,}15 & 0 & 0{,}05 \\ -0{,}15 & -0{,}1 & -0{,}05 & 0 \end{pmatrix}$

En utilisant par exemple, la norme infinie, on obtient
$\|C\|_\infty = \max(0{,}35; 0{,}35; 0{,}35; 0{,}55) = 0{,}55 < 1$, donc le processus itératif converge.

2°) Si on choisit pour approximation initiale de la solution X le vecteur $X^{(0)} = l = (0; 0{,}5; -0{,}5; 0{,}75)^t$ on aura $\|l\| = 0 + 0{,}5 + 0{,}5 + 0{,}75 = 1{,}75$.
Soit N le nombre d'itérations nécessaires pour obtenir la précision donnée. En appliquant la formule (2.16), on aura $N = E\left[\dfrac{Ln\left(\dfrac{(1-0{,}55)10^{-4}}{1{,}75.0{,}55}\right)}{Ln0{,}55}\right] + 1 = 17$ itérations.

2.6 Comparaison de deux méthodes itératives

Soient les deux méthodes itératives suivantes

$$(1)\begin{cases} X^{(k+1)} = C_1X^{(k)} + l_1 \\ X^{(0)} \text{ donné} \end{cases} \quad \text{et} \quad (2)\begin{cases} X^{(k+1)} = C_2X^{(k)} + l_2 \\ X^{(0)} \text{ donné} \end{cases}$$

Soient $\rho(C_1)$ (resp. $\rho(C_2)$) le rayon spectral de la matrice C_1 (resp. C_1).
Supposons que $\rho(C_1) < \rho(C_2) \Rightarrow Ln\rho(C_1) < Ln\rho(C_2) \Rightarrow \dfrac{1}{Ln\rho(C_1)} > \dfrac{1}{Ln\rho(C_2)} \Rightarrow$
$\dfrac{Ln\varepsilon}{Ln\rho(C_1)} > \dfrac{Ln\varepsilon}{Ln\rho(C_2)}$ avec $0 < \varepsilon < 1$.

On note $N_1 = E\left[\dfrac{Ln\varepsilon}{Ln\rho(C_1)}\right] + 1$ et $N_2 = E\left[\dfrac{Ln\varepsilon}{Ln\rho(C_2)}\right] + 1$ donc $N_1 \leq N_2$.

Comme N_1 représente le nombre d'itérations nécessaires pour approcher la solution exacte d'un système par la méthode itérative (1) et N_2 représente le nombre d'itérations nécessaires pour approcher la solution exacte d'un système par la méthode itérative (2) donc la méthode itérative (1) utilise moins d'itérations donc converge plus rapidement.

Conclusion

La méthode itérative qui converge le plus rapidement est celle dont le rayon spectral de la matrice associée à la méthode est le plus petit.

2.7 Exercices corrigés sur le chapitre

Exercice 1 Soit le système $AX = b$ suivant $\begin{cases} 10x_1 + x_2 + x_3 = 12 \\ 2x_1 + 10x_2 + x_3 = 13 \\ 2x_1 + 2x_2 + 10x_3 = 14 \end{cases}$

(1) Transformer le système $AX = b$ en un système équivalent $X = CX + l$ (C et l à déterminer).

(2) Montrer que ce processus itératif est convergent.

(3) Écrire et détailler l'algorithme de Gauss-Seidel associé à ce processus.

(4) Calculer par la méthode de Gauss-Seidel les différentes approximations en commençant par $X^{(0)} = (1.2,0,0)^t$ (tous les calculs se feront avec quatre chiffres après la virgule).

(5) Quelle remarque faites-vous à la $4^{ème}$ et $5^{ème}$ itération?

(6) Peut-on déduire à ce stade la solution exacte du système $AX = b$. Si oui, laquelle?

Solution de l'exercice 1: (1)

$$AX = b \Leftrightarrow \begin{cases} 10x_1 + x_2 + x_3 = 12 \\ 2x_1 + 10x_2 + x_3 = 13 \\ 2x_1 + 2x_2 + 10x_3 = 14 \end{cases} \Leftrightarrow \begin{cases} x_1 = 1.2 - 0.1x_2 - 0.1x_3 \\ x_2 = 1.3 - 0.2x_1 - 0.1x_3 \\ x_3 = 1.4 - 0.2x_1 - 0.2x_2 \end{cases} \Leftrightarrow X = CX + l$$

Où $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $C = \begin{pmatrix} 0 & -0.1 & -0.1 \\ -0.2 & 0 & -0.1 \\ -0.2 & -0.2 & 0 \end{pmatrix}$, $l = \begin{pmatrix} 1.2 \\ 1.3 \\ 1.4 \end{pmatrix}$.

(2) Le processus itératif est convergent ssi il existe une norme matricielle telle que $\|C\| < 1$. Calculons par exemple, $\|C\|_\infty = \max\limits_{1 \le i \le 3}(\sum\limits_{j=1}^{3} |c_{ij}|) = \max(0.2,0.3,0.4) = 0.4 < 1$ donc le processus $X^{(k+1)} = CX^{(k)} + l$ est convergent.

(3) L'algorithme de ce processus est défini par:

$$\begin{cases} X^{(k+1)} = CX^k + l \\ \\ X^{(0)} \text{ donné} \end{cases} \Leftrightarrow \begin{pmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{pmatrix} = \begin{pmatrix} 0 & -0.1 & -0.1 \\ -0.2 & 0 & -0.1 \\ -0.2 & -0.2 & 0 \end{pmatrix} . \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{pmatrix} + \begin{pmatrix} 1.2 \\ 1.3 \\ 1.4 \end{pmatrix}$$

Ainsi, l'algorithme de Gauss-Seidel associé à ce processus s'écrit:

$$\begin{cases} x_1^{(k+1)} = -0.1x_2^{(k)} - 0.1x_3^{(k)} + 1.2 \\ x_2^{(k+1)} = -0.2\underline{x_1^{(k+1)}} - 0.1x_3^{(k)} + 1.3 \\ x_3^{(k+1)} = -0.2\underline{x_1^{(k+1)}} - 0.2\underline{x_2^{(k+1)}} + 1.4 \end{cases}$$

(4) Pour $X^{(0)} = (x_1^{(0)},x_2^{(0)},x_3^{(0)})^t = (1.2,0,0)^t$, la première itération obtenue pour $k = 0$

donne $X^{(1)} = \begin{cases} x_1^{(1)} = -0.1x_2^{(0)} - 0.1x_3^{(0)} + 1.2 = 1.2 \\ x_2^{(1)} = -0.2x_1^{(1)} - 0.1x_3^{(0)} + 1.3 = 1.06 \\ x_3^{(1)} = -0.2x_1^{(1)} - 0.2x_2^{(1)} + 1.4 = 0.948 \end{cases} \Rightarrow X^{(1)} = \begin{pmatrix} x_1^{(1)} = 1.2 \\ x_2^{(1)} = 1.06 \\ x_3^{(1)} = 0.948 \end{pmatrix}$

Pour $k = 1$, $X^{(2)} = \begin{cases} x_1^{(2)} = -0.1x_2^{(1)} - 0.1x_3^{(1)} + 1.2 = 0.9992 \\ x_2^{(2)} = -0.2x_1^{(2)} - 0.1x_3^{(1)} + 1.3 = 1.0053 \\ x_3^{(2)} = -0.2x_1^{(2)} - 0.2x_2^{(2)} + 1.4 = 0.9991 \end{cases} \Rightarrow X^{(2)} = \begin{pmatrix} x_1^{(2)} = 0.9992 \\ x_2^{(2)} = 1.0053 \\ x_3^{(2)} = 0.9991 \end{pmatrix}$

Par le même procédé, on obtient les autres approximations

$$X^{(3)} = \begin{pmatrix} x_1^{(3)} = 0.9996 \\ x_2^{(3)} = 1.0001 \\ x_3^{(3)} = 1.0001 \end{pmatrix}, \quad X^{(4)} = \begin{pmatrix} x_1^{(4)} = 1.0000 \\ x_2^{(4)} = 1.0000 \\ x_3^{(4)} = 1.0000 \end{pmatrix}, \quad X^{(5)} = \begin{pmatrix} x_1^{(5)} = 1.0000 \\ x_2^{(5)} = 1.0000 \\ x_3^{(5)} = 1.0000 \end{pmatrix}$$

(5) On remarque qu'à la quatrième et cinquième itération, les résultats coïncident à 10^{-4} près d'où la solution du système initial est donnée par $X = \begin{pmatrix} x_1 = 1.0000 \pm 0.0001 \\ x_2 = 1.0000 \pm 0.0001 \\ x_3 = 1.0000 \pm 0.0001 \end{pmatrix}$

(6) La solution exacte du système $AX = b$ est déduite par $X = (1,1,1)^t$.

Exercice 2 Soit A une matrice inversible de $\mathcal{M}_n(\mathbb{R})$. On considère la processus itératif suivant:

$$A_k = A_{k-1}.(2Id - A.A_{k-1}), \quad k \geq 1 \tag{2.19}$$

On pose $B_k = Id - A.A_k$, où Id où désigne la matrice identité d'ordre n.

(1) Calculer B_k en fonction de B_0.

(2) Montrer que la norme $\|A^{-1} - A_k\| \leq \|A^{-1}\|.\|B_k\|$.

(3) En déduire une condition suffisante de convergence du processus (2.19)

Solution de l'exercice 2:

(1) $B_k = Id - A.A_k = Id - A.(A_{k-1}.(2Id - A.A_{k-1})) = Id - 2A.A_{k-1} + (A.A_{k-1})^2 = (Id - A.A_{k-1})^2 = (B_{k-1})^2$

A partir de $B_k = (B_{k-1})^2 = (B_{k-2}^2)^2 = ((B_{k-3}^2)^2)^2 = B_{k-3}^6 = ... = B_0^{2k}$.

(2) $A^{-1} - A_k = A^{-1} - A_{k-1}.(2Id - A.A_{k-1}) = A^{-1}(Id - 2A.A_{k-1} + (A.A_{k-1})^2) = A^{-1}.B_k$

On obtient donc $A^{-1} - A_k = A^{-1}.B_k$ et par passage à la norme, on a

$$\|A^{-1} - A_k\| = \|A^{-1}.B_k\| \leq \|A^{-1}\|.\|B_k\|.$$

(3) De (1) et (2), on a $\|A^{-1} - A_k\| \leq \|A^{-1}\|.\|B_k\| \leq \|A^{-1}\|.\|B_0^{2k}\| \leq \|A^{-1}\|.\|B_0\|^{2k}$

Si $\lim_{k\to\infty} \|B_0\|^{2k} = 0$ alors $\lim_{k\to\infty} \|A^{-1} - A_k\| = 0$ c'est à dire $\lim_{k\to\infty} A_k = A^{-1}$.

La condition suffisante de convergence du processus (2.19) est que $\lim_{k\to\infty} \|B_0\|^{2k} = 0$ qui est équivalent à $\lim_{k\to\infty} B_0^{2k} = O_{\mathcal{M}_n(\mathbb{R})}$ qui est équivalent aussi à $\rho(B_0^2) < 1$ d'après le théorème (2.3).

Exercice 3 Soit à résoudre le système $AX = b$ où $A = \begin{pmatrix} \alpha & 1 & 1 \\ 1 & \alpha & 1 \\ 1 & 1 & \alpha \end{pmatrix}$, $b \in \mathbb{R}^n$,$\alpha \in \mathbb{R}^*$

(1) Donner une condition suffisante à laquelle doit satisfaire le paramètre α pour que la méthode itérative de Jacobi converge.

(2) Dans cette question, on prendra $\alpha \in \mathbb{R}_+^*$.

- **(a)** Déterminer la matrice J_A de Jacobi associée à A.
- **(b)** Montrer que les valeurs propres de J_A sont $\lambda_1 = 1/\alpha$ et $\lambda_2 = -2/\alpha$.
- **(c)** Donner les valeurs de α pour lesquelles la méthode Jacobi converge.

(3) Soient $P = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix}$, et $H = \begin{pmatrix} 0 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

- **(a)** La matrice $(P - H)$ est-elle inversible?

(b) Vérifier que $(P-H)^{-1} = \begin{pmatrix} 1/3 & -1/12 & -1/12 \\ 0 & 3/8 & -1/8 \\ 0 & -1/8 & 3/8 \end{pmatrix}$

(c) Montrer que $(P-H) - H^t = A$ pour $\alpha = 3$.

(d) Montrer que $AX = b$ est équivalent au système $X = KX + d$ où $K = (P-H)^{-1}.H^t$ et d un vecteur à déterminer.

(e) Calculer le rayon spectral de la matrice K. Conclure.

(4) Pour $\alpha = 3$

(a) Montrer que la matrice G de Gauss-Seidel associée à A s'écrit

$$G = \begin{pmatrix} 0 & -1/3 & -1/3 \\ 0 & 1/9 & -2/9 \\ 0 & 2/27 & 5/27 \end{pmatrix}$$

(b) Calculer le rayon spectral de G. Conclure.

(5) Comparer les trois méthodes.

Solution de l'exercice 3:

(1) Une condition suffisante pour que la méthode itérative de Jacobi converge est que la matrice A soit à diagonale strictement dominante c'est à dire $|a_{ii}| > \sum_{j=1,j\neq i}^{3} |a_{ij}|$, $\forall i = 1,2,3$, ce qui nous donne $|\alpha| > 1+1 = 2$ c'est à dire $|\alpha| > 2$.

$(2)_{(a)}$ La matrice de Jacobi associée à la matrice A est définie par:

$$J_A = (j_{ij})_{1\leq i,j\leq 3} = \begin{cases} -\dfrac{a_{ij}}{a_{ii}} \text{ si } i \neq j \\ \\ 0 \qquad \text{si } i = j \end{cases} \Leftrightarrow J_A = \begin{pmatrix} 0 & -1/\alpha & -1/\alpha \\ -1/\alpha & 0 & -1/\alpha \\ -1/\alpha & -1/\alpha & 0 \end{pmatrix}, \ \alpha \in \mathbb{R}_+^*.$$

$(2)_{(b)}$ Les valeurs propres de J sont les zéros du polynôme caractéristique $P(\lambda)$ ou les racines de l'équation $P(\lambda) = det(J_A - \lambda.Id_3) = 0$. Après calcul du déterminant, on trouve $P(\lambda) = det(J_A - \lambda.Id_3) = \left(-\lambda + \dfrac{1}{\alpha}\right)\left(\lambda^2 + \dfrac{\lambda}{\alpha} - \dfrac{2}{\alpha^2}\right)$ qui s'annule pour $\lambda_1 = \dfrac{1}{\alpha}$, $\lambda_2 = \dfrac{1}{\alpha}$ et $\lambda_3 = \dfrac{-2}{\alpha}$ qui représentent les valeurs propres de la matrice J_A.

$(2)_{(c)}$ La méthode de Jacobi converge ssi le rayon spectral $\rho(J_A) < 1$ où $\rho(J_A) = \max(|\lambda_1|,|\lambda_2|,|\lambda_3|) = \max\left(\left|\dfrac{1}{\alpha}\right|, \left|\dfrac{2}{\alpha}\right|\right) = \dfrac{2}{\alpha}$ car $\alpha > 0$ d'où $\rho(J_A) = \dfrac{2}{\alpha} < 1$ si

$\alpha > 2$ et on conclut que la méthode de Jacobi associée à A converge si $\alpha > 2$.

$(3)_{(a)}$ La matrice $(P - H)$ est inversible ssi son déterminant est non nul. On a $det(P-H) = \begin{vmatrix} 3 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 1 & 3 \end{vmatrix} = 24 \neq 0$ donc la matrice inverse $(P-H)^{-1}$ existe.

$(3)_{(b)}$ On vérifie facilement que $(P-H)^{-1}.(P-H) = (P-H).(P-H)^{-1} = Id_3$.

$(3)_{(c)}$ Par un calcul simple, on a bien $(P-H) - H^t = A$ pour $\alpha = 3$.

$(3)_{(d)}$ $A.X = b \Leftrightarrow ((P-H) - H^t).X = b \Leftrightarrow (P-H).X = H^t.X + b$. En tirant X, on obtient $X = (P-H)^{-1}.H^t.X + (P-H)^{-1}.b = K.X + d$ où $K = (P-H)^{-1}.H^t$ et $d = (P-H)^{-1}.b$.

$(3)_{(e)}$ Calcul de $K = (P-H)^{-1}.H^t = \begin{pmatrix} 1/6 & 0 & 0 \\ -1/4 & 0 & 0 \\ -1/4 & 0 & 0 \end{pmatrix}$

Les valeurs propres de K sont données par $P(\lambda) = det(K - \lambda.Id_3) = \lambda^2.\left(\frac{1}{6} - \lambda\right) = 0$ si $\lambda_1 = \lambda_2 = 0$ et $\lambda_3 = \frac{1}{6}$ d'où $\rho(K) = \max\left(0, \frac{1}{6}\right) = \frac{1}{6}$.

Comme $\rho(K) = \frac{1}{6} < 1$ donc la méthode itérative $X = K.X + d$ est convergente.

$(4)_{(a)}$ La matrice de Gauss-Seidel est donnée par la formule $G_A = (D-E)^{-1}.F$ où

$$D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & -1 & 0 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & -1 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} \quad \alpha = 3$$

En utilisant, par exemple la méthode de Gauss-Jordan, l'inverse de $(D-E)$ sera

$$(D-E)^{-1} = \begin{pmatrix} 1/3 & 0 & 0 \\ -1/9 & 1/3 & 0 \\ -2/27 & -1/9 & 1/3 \end{pmatrix} \quad \text{d'où} \quad G_A = \begin{pmatrix} 0 & -1/3 & -1/3 \\ 0 & 1/9 & -2/9 \\ 0 & 2/27 & 5/27 \end{pmatrix}$$

$(4)_{(b)}$ Les valeurs propres de G sont les solutions de
$P(\lambda) = det(G - \lambda.Id_3) = \frac{-\lambda}{27}(27\lambda^2 - 8\lambda + 1) = 0$, ce qui donne comme valeurs propres

$\lambda_1 = 0$, $\lambda_2 = \dfrac{4 + i\sqrt{11}}{27}$ et $\lambda_3 = \dfrac{4 - i\sqrt{11}}{27}$.

Le rayon spectral $\rho(G) = \max\left(0, \left|\dfrac{4 + i\sqrt{11}}{27}\right|, \left|\dfrac{4 - i\sqrt{11}}{27}\right|\right) = \max\left(0, \sqrt{\dfrac{1}{27}}\right) = \sqrt{\dfrac{1}{27}}$

Comme $\rho(G) = \sqrt{\dfrac{1}{27}} \approx 0.19 < 1$ ce qui implique que la méthode de Gauss-seidel associée à A est convergente pour $\alpha = 3$.

(5) La comparaison des trois méthodes se fait sur la base des rayons spectraux.
En effet, on a $\rho(J_A) = \dfrac{2}{\alpha} = \dfrac{2}{3} \approx 0.66$, $\rho(K) = \dfrac{1}{6} \approx 0.16$ et $\rho(G) = \sqrt{\dfrac{1}{27}} \approx 0.19$
donc on a l'inégalité suivante $\rho(K) < \rho(G) < \rho(J_A)$ et on conclut que la méthode itérative associée à la matrice K converge plus rapidement que les méthodes de Gauss-seidel et Jacobi associées à A pour $\alpha = 3$ (car elle a le plus petit rayon spectral).

Chapitre 3

Résolution d'équations non linéaires f(x)=0

3.1 Rappels de quelques éléments d'analyse

Cette partie est consacrée à des rappels de quelques notions et résultats d'analyse.

Définition 3.1.

Soit A une partie de $\mathbb{R}$ et x un nombre réel. On dit que x est

1. Un majorant (resp. minorant) de A dans $\mathbb{R}$ si et seulement si
   $\forall a \in A,\ a \leq x$ (resp. $x \leq a$)
2. Un plus grand élément (resp. un plus petit élément) de A dans $\mathbb{R}$ si et seulement si c'est un majorant (resp. minorant) de A dans $\mathbb{R}$ appartenant à A.

Définition 3.2.

On appelle borne supérieure (resp. borne inférieure) de A dans $\mathbb{R}$ le plus petit des majorants (resp. le plus grand des minorants) de A dans $\mathbb{R}$. Si cette borne existe, elle est alors unique et notée $supA$ (resp. $infA$).

Définition 3.3.

Une partie A de $\mathbb{R}$ est dite bornée lorsqu'il existe deux réels m et M tels que, pour tout $x \in A$, $m \leq x \leq M$.

Proposition 3.1. *Axiome de la borne supérieure et de la borne inférieure*
Toute partie non vide et majorée (resp. minorée) de $\mathbb{R}$ *admet une borne supérieure (resp.*

inférieure) dans $\mathbb{R}$.

Théorème 3.1. *Soit f une fonction continue sur [a,b]. Si $f(a)$ et $f(b)$ sont de signes contraires alors il existe au moins un point $c \in]a,b[$ tel que $f(c) = 0$.*

Démonstration.
Supposons par exemple que $f(a) > 0$ et $f(b) < 0$.
Considérons l'ensemble $E = \{x \in [a,b] / f(x) > 0\}$, cette partie E étant non vide (car $a \in E$) et majorée donc admet une borne supérieure notée $c = supE$ d'après l'axiome précédent et montrons que $f(c) = 0$ avec $a < c < b$.
Si $f(c) \neq 0$, il existe alors un intervalle $]c - \delta, c + \delta[$ $(\delta > 0)$ où f conserve le signe de $f(c)$ d'après la continuité de f et ceci est en contradiction avec l'hypothèse que $c = supE$.
En effet,
— Si $f(x) > 0$ dans l'intervalle $]c - \delta, c + \delta[$ alors $c + \frac{\delta}{2} \in E$ et $c = supE \geq c + \frac{\delta}{2} \in E$, ce qui est impossible.
— Si $f(x) < 0$ dans l'intervalle $]c - \delta, c + \delta[$ alors $c = supE \leq c - \delta$, ce qui est impossible.
Ainsi $f(c) = 0$. □

Remarque 3.1. Ce théorème (3.1) est utilisé pour encadrer les solutions d'une équation.

Théorème 3.2. *Théorème des valeurs intermédiaires*
Soit $[a,b]$ un intervalle non vide de $\mathbb{R}$ et f une application définie et continue sur $[a,b]$ à valeurs dans $\mathbb{R}$. Alors, pour tout réel k strictement compris entre $f(a)$ et $f(b)$, il existe au moins un réel $c \in]a,b[$ tel que $f(c) = k$.

Démonstration.
Supposons que $f(a) < f(b)$ et soit un réel k tel que $f(a) < k < f(b)$. Considérons la fonction $g : [a,b] \mapsto \mathbb{R}$ définie par $g(x) = f(x) - k$. Comme la fonction g vérifie les hypothèses du théorème (3.1) et donc s'annule en un point $c \in]a,b[$ d'où $f(c) = k$.

□

Théorème 3.3. *Théorème de la bijection*
Soit f une fonction continue et strictement monotone sur $[a,b]$. Alors, pour tout réel k compris entre $f(a)$ et $f(b)$, l'équation $f(x) = k$ admet une solution unique c dans $]a,b[$.

Démonstration.
D'après le théorème des valeurs intermédiaires (3.2), l'équation $f(x) = k$ admet au moins

une solution c dans $]a,b[$. La fonction f étant strictement monotone dans $[a,b]$ donc pour tout $x \in [a,b]$ et $x \neq c$ on a $f(x) \neq f(c)$. Le réel c est donc l'unique solution de $f(x) = k$ dans $[a,b]$.

□

Corollaire. *Si f est continue et strictement monotone sur $[a,b]$ et si $f(a).f(b) < 0$ alors l'équation $f(x) = 0$ a une solution unique dans $[a,b]$.*

Démonstration.
L'hypothèse $f(a).f(b) < 0$ implique $f(a) \neq 0$, $f(b) \neq 0$ et $f(a)$, $f(b)$ sont de signes contraires donc zéro appartient à l'intervalle $[f(a),f(b)]$ ou $[f(b),f(a)]$ d'où le résultat.

□

Théorème 3.4. *Théorème de Rolle*
Soit $f : [a,b] \longmapsto \mathbb{R}$ une fonction continue sur $[a,b]$, dérivable sur $]a,b[$ et telle que $f(a) = f(b)$. Alors il existe un point $c \in]a,b[$ tel que $f'(c) = 0$.

Démonstration.
La démonstration consiste à montrer que f admet un extrémum local en un point de $]a,b[$.
1°) Si f est constante, le résultat est trivial alors $\forall x \in [a,b]$, $f'(x) = 0$.
2°) Sinon, comme f est continue sur l'intervalle fermé borné $[a,b] \subset \mathbb{R}$ donc atteint ses bornes. On pose $M = \sup\limits_{[a,b]} f(x)$ et $m = \inf\limits_{[a,b]} f(x)$, on a donc $m < M$ et une, au moins, des inégalités strictes $m < f(a)$ ou $M > f(a)$ est vérifiée.
On Suppose que $M > f(a)$, comme f est continue sur $[a,b]$, il $\exists c \in [a,b]$ tel que $f(c) = M$ et $c \in]a,b[$ car $f(a) \neq M$ et $f(b) \neq M$. La fonction f admet en c un maximum local, la fonction f étant dérivable sur $]a,b[$, on en déduit que $f'(c) = 0$ (d'après le théorème, si f admet un extremum au point c et si f est dérivable alors $f'(c) = 0$). □

Théorème 3.5. *Théorème des accroissements finis*
Soit f une fonction continue sur $[a,b]$ et dérivable sur $]a,b[$, alors il existe au moins un nombre c de $]a,b[$ tel que $f(b) - f(a) = (b-a)f'(c)$.

Démonstration.
Considérons la fonction auxiliaire suivante $\phi : [a,b] \longmapsto \mathbb{R}$ définie par
$\phi(x) = f(x) - f(a) - \dfrac{f(b) - f(a)}{b - a}(x - a)$.
La fonction ϕ est continue dans $[a,b]$, dérivable dans $]a,b[$ et $\phi(a) = \phi(b) = 0$. Les hypothèses

du théorème de Rolle sont vérifiées, ainsi, il existe au moins un nombre c dans $]a,b[$ tel que $\phi'(c) = f'(c) - \dfrac{f(b)-f(a)}{b-a} = 0$, ce qui donne $f(b) - f(a) = (b-a)f'(c)$. □

3.2 Résolution d'équations non linéaires f(x)=0

3.2.1 Position du problème

Soit f une fonction définie sur $D \subset \mathbb{R}$ dans $\mathbb{R}$, le problème est de trouver tous les éléments réels $x \in D$ qui vérifient

$$f(x) = 0 \tag{3.1}$$

Il n'est pas toujours possible de résoudre l'équation (3.1) quelle que soit la forme de la fonction f (complexité de f et la non linéarité). Pour la résoudre, nous allons construire des méthodes numériques (de dichotomie, du point fixe et de Newton) nous permettant d'obtenir une valeur approchée des racines.

Définition 3.4. Tout réel $\overline{x}$ qui vérifie $f(\overline{x}) = 0$ s'appelle racine de l'équation (3.1) ou zéro de la fonction f.

Définition 3.5.
La racine $\overline{x}$ de l'équation (3.1) est dite racine séparée ou racine isolée dans l'intervalle $[a,b]$ si $\overline{x}$ est l'unique racine réelle de cette équation dans $[a,b]$.

3.3 Localisation et séparation des racines

Toutes les méthodes que nous allons étudier nécessitent d'abord la séparation des racines. Les deux méthodes les plus classiques pour séparer les racines d'une équation réelle sont la méthode graphique et la méthode algébrique de balayage.

3.3.1 Méthode graphique

Les racines réelles de l'équation (3.1) peuvent être déterminées approximativement
– Comme les abscisses des points d'intersection de la courbe de la fonction $f(x)$ avec l'axe Ox,

– Ou comme les abscisses des points d'intersection des courbes des fonctions $y = f_1(x)$ et $y = f_2(x)$ telles que $f(x) = f_1(x) - f_2(x)$ où f_1 et f_2 sont deux fonctions plus simples à étudier que f.

Exemple d'application

Localiser et séparer graphiquement les racines de l'équation

$$f(x) = x^3 + 2x - 5 = 0 \tag{3.2}$$

Solution

Cette équation peut être décomposée sous la forme $f(x) = f_1(x) - f_2(x)$ avec $f_1(x) = x^3$ et $f_2(x) = -2x + 5$. L'intersection de la parabole cubique $y = x^3$ et de la droite d'équation $y = -2x + 5$ fournit approximativement l'unique racine réelle $\overline{x} \in]1; 1{,}5[$ de l'équation (3.2).

3.3.2 Méthode algébrique de balayage

Cette méthode est basée sur le théorème (3.2) des valeurs intermédiaires. Soit f une fonction définie et continue sur l'intervalle $[a,b]$, on détermine le signe de $f(a)$ et $f(b)$.

— Si $f(a).f(b) > 0$, deux cas peuvent se présenter:
  (1) Soit il n'existe pas de racines de $f(x) = 0$ dans $[a,b]$,
  (2) Soit le nombre de racines de $f(x) = 0$ dans $[a,b]$ est pair.

— Sinon, si $f(a).f(b) < 0$, le nombre de racines de $f(x) = 0$ est impair dans $[a,b]$. Dans ce cas, on détermine le signe de $f(x)$ en quelques points quelconques λ_1,λ_2,...etc de $[a,b]$. S'il existe des points λ_k et λ_{k+1} tels que $f(\lambda_k).f(\lambda_{k+1}) < 0$ alors il existe au moins une racine $\overline{x} \in]\lambda_k,\lambda_{k+1}[$ de $f(x) = 0$. Il reste juste à définir l'unicité de cette racine en étudiant la monotonie de f sur $]\lambda_k,\lambda_{k+1}[$. (les intervalles $]\lambda_k,\lambda_{k+1}[$ doivent être les plus réduits possibles).

Remarque 3.2.
En pratique, on opère par une bipartition. Cette dernière consiste à diviser l'intervalle $[a,b]$ en 2,4,8,...parties égales et à déterminer les signes de la fonction f aux points de division.

Exemple d'application

Localiser et séparer les racines de l'équation $f(x) = x^3 - 6x + 2 = 0$ sur $\mathbb{R}$.

Solution

L'équation f(x)=0 admet au plus trois racines réelles sur $\mathbb{R}$ car le degré de f est de 3. On choisit des points intermédiaires λ_i quelconques.

points intermédiaires λ_i	-∞	-3	-1	0	1	3	+∞
signe de $f(\lambda_i)$	(-)	(-)	(+)	(+)	(-)	(+)	(+)

Comme $f(-3)f(-1) < 0$ donc $\exists\lambda_1 \in]-3,-1[$ telle que $f(\lambda_1) = 0$, $f(0)f(1) < 0$ donc $\exists\lambda_2 \in$]0,1[telle que $f(\lambda_2) = 0$ et $f(1)f(3) < 0$ donc $\exists\lambda_3 \in$]1,3[telle que $f(\lambda_3) = 0$.
En conclusion, l'équation $f(x) = x^3 - 6x + 2 = 0$ admet trois racines réelles telles que $\lambda_1 \in]-3,-1[$, $\lambda_2 \in$]0,1[et $\lambda_3 \in$]1,3[.

Remarque 3.3.
Si la fonction f est dérivable et que les zéros de sa dérivée sont faciles à déterminer, le processus de séparation des racines peut être ordonné. Il suffit de considérer comme points intermédiaires les zéros de la dérivée et les points frontières a et b de l'intervalle [a,b].

Exemple d'application

Localiser et séparer les racines de l'équation $f(x) = x^3 - 12x + 5 = 0$.
La dérivée $f'(x) = 3x^2 - 12$ s'annule aux points $x = \pm 2$.

Solution

points intermédiaires λ_i	-∞	-4	-2	0	2	4	+∞
signe de $f(\lambda_i)$	(-)		(+)		(-)		(+)
signe autres points		(-)		(+)		(+)	

L'équation $f(x) = x^3-12x+5 = 0$ admet trois racines réelles telles que $\lambda_1 \in]-\infty,-2[$, $\lambda_2 \in]-2,+2[$ et $\lambda_3 \in]2,+\infty[$, et mieux encore nous pouvons réduire les intervalles à $\lambda_1 \in]-4,-2[$, $\lambda_2 \in]0,+2[$ et $\lambda_3 \in]2,+4[$.

3.3.3 Évaluation de l'erreur d'une racine approchée

On suppose que s est racine exacte de l'équation (3.1) et que s est racine séparée dans [a,b]. L'approximation de s à ε près consiste à trouver une valeur $\widetilde{s}$ telle que $|s-\widetilde{s}| \leq \varepsilon$ (ε étant un nombre positif donné).

Théorème 3.6.
Soit f une fonction de classe $C^1([a,b])$, s une racine exacte et $\widetilde{s}$ une racine approchée de l'équation $f(x)=0$ sur [a,b]. Si $|f'(x)| \geq m > 0$ pour tout $x \in [a,b]$ alors $|s-\widetilde{s}| \leq \dfrac{|f(\widetilde{s})|}{m}$ (on peut prendre $m = \min\limits_{a\leq x\leq b} |f'(x)|$).

Démonstration.
F étant continue et dérivable sur [a,b], on peut appliquer le théorème (3.5) des accroissements finis sur l'intervalle [s,$\widetilde{s}$] (ou [$\widetilde{s}$,s] $\subset$ [a,b]), il existe $c \in]s,\widetilde{s}[$ tel que
$f(\widetilde{s}) - f(s) = (\widetilde{s}-s)f'(c)$. Comme s est racine exacte donc $f(s)=0$ et donc
$f(\widetilde{s}) = (\widetilde{s}-s)f'(c) \Rightarrow |f'(c)| = \dfrac{|f(\widetilde{s}|}{|\widetilde{s}-s|} \geq m$ car m est le minimum de $|f'(x)| \Rightarrow$
$|f(\widetilde{s}| \geq m|\widetilde{s}-s| \Rightarrow |s-\widetilde{s}| \leq \dfrac{|f(\widetilde{s})|}{m}$. □

Exemple d'application

Une racine approchée de l'équation $f(x) = x^4 - x - 1 = 0$ est $\widetilde{s} = 1.22$
Évaluer l'erreur absolue de cette approximation.

Solution

Il faut d'abord déterminer l'intervalle [a,b] en choisissant des points très proches de la valeur 1.22. On a $f(1.21) < 0$, $f(1.22) < 0$ et $f(1.23 > 0$, comme la fonction f est continue strictement monotone sur $[1.22; 1.23]$ donc la racine exacte s est unique dans $[1.22; 1.23]$.
La dérivée $f'(x) = 4x^3 - 1$ étant positive strictement croissante sur $[1.22; 1.23]$ donc son minimum est atteint au point 1.22 d'où $m = f'(1.22) = 6.264$. On conclut que l'erreur $e = \dfrac{|f(\widetilde{s})|}{m} = \dfrac{|f(1.22)|}{6.264} \simeq 0.0007$. Comme $s \simeq \widetilde{s} \pm e$ donc $s \simeq 1.22 \pm 0.0007$ et plus exactement $s \simeq 1.22 + 0.0007$ car $s > 1.22$

3.4 Méthode de dichotomie

La méthode de dichotomie encore appelée méthode de bissection ou bipartition est basée sur le théorème (3.2) des valeurs intermédiaires. Elle suppose que la fonction f est continue vérifiant $f(a).f(b) < 0$ et f n'admet qu'une seule racine s dans l'intervalle réel $[a,b]$. Cette méthode fournit une suite d'intervalles emboîtés encadrant la racine s de l'équation (3.1) que l'on veut résoudre.

3.4.1 Principe de la méthode de dichotomie

On pose $[a,b] = [a^{(0)},b^{(0)}]$ avec $a^{(0)} = a$ et $b^{(0)} = b$. On divise le segment $[a^{(0)},b^{(0)}]$ en deux segments $[a^{(0)},x^{(0)}]$ et $[x^{(0)},b^{(0)}]$ où $x^{(0)} = \dfrac{a^{(0)} + b^{(0)}}{2}$ étant son milieu.

— Si $f(x^{(0)}) = 0$ alors $s = x^{(0)} = \dfrac{a^{(0)} + b^{(0)}}{2}$ est racine de l'équation (3.1),

— Si $f(a^{(0)}).f(x^{(0)}) < 0$ alors $s \in]a^{(0)},x^{(0)}[$ et de plus on a $\left|s - x^{(0)}\right| < \dfrac{b-a}{2}$

car $\left|s - x^{(0)}\right| < \left|x^{(0)} - a^{(0)}\right| = \dfrac{b^{(0)} - a^{(0)}}{2} = \dfrac{b-a}{2}$,

— Si $f(x^{(0)}).f(b^{(0)}) < 0$ alors la racine $s \in]x^{(0)},b^{(0)}[$.

On suppose que la racine $s \in]a^{(0)},x^{(0)}[$, on divise à nouveau cet intervalle en deux et soit $x^{(1)} = \dfrac{a^{(0)} + x^{(0)}}{2}$ son milieu.
D'après le raisonnement précédent la racine $s \in]a^{(0)},x^{(1)}[$ ou $s \in]x^{(1)},x^{(0)}[$.
On Suppose que $f(a^{(0)}).f(x^{(1)}) < 0$ alors $s \in]a^{(0)},x^{(1)}[$ et de plus on a $\left|s - x^{(1)}\right| < \dfrac{b-a}{2^2}$
car $\left|s - x^{(1)}\right| < \left|x^{(1)} - a^{(0)}\right| = \left|\dfrac{a^{(0)} + x^{(0)}}{2} - a^{(0)}\right| = \left|\dfrac{x^{(0)} - a^{(0)}}{2}\right| = \dfrac{b^{(0)} - a^{(0)}}{2^2} = \dfrac{b-a}{2^2}$.
En poursuivant ainsi la subdivision, on construit de manière récurrente trois suites $(a^{(k)})_{k\in\mathbb{N}}$, $(b^{(k)})_{k\in\mathbb{N}}$ et $(x^{(k)})_{k\in\mathbb{N}}$ définies par:

$$\begin{cases} a^{(0)} = a \quad b^{(0)} = b \\ x^{(k)} = (a^{(k)} + b^{(k)})/2 \quad \forall k \in \mathbb{N} \\ a^{(k+1)} = a^{(k)} \text{ et } b^{(k+1)} = x^{(k)} \text{ si } s \in]a^{(k)},x^{(k)}[\\ a^{(k+1)} = x^{(k)} \text{ et } b^{(k+1)} = b^{(k)} \text{ et } s \in]x^{(k)},b^{(k)}[\end{cases} \tag{3.3}$$

A l'ordre (n), on obtient une valeur approchée de s notée $x^{(n)}$ et l'erreur que l'on commet sur cette approximation est donnée par:

$$\left|s - x^{(n)}\right| < \frac{b-a}{2^{n+1}}, \quad n = 0,1,2,... \tag{3.4}$$

3.4.2 Convergence de la méthode de dichotomie

L'inégalité (3.4) prouve que la suite $(x^{(n)})_{n\in\mathbb{N}}$, ainsi construite dans la méthode de dichotomie, converge vers la solution exacte s de l'équation $f(x) = 0$.
En effet, $0 \longleftarrow 0 < \left|s - x^{(n)}\right| < \frac{b-a}{2^{n+1}} \longrightarrow 0$ d'où $\lim\limits_{n\mapsto+\infty}(s - x^{(n)}) = 0$ et $\lim\limits_{n\mapsto+\infty} x^{(n)} = s$.
On conclut que la méthode de dichotomie converge vers la solution exacte s de l'équation (3.1) de manière certaine.

3.4.3 Critère d'arrêt de la méthode de dichotomie

L'inégalité (3.4) fournit un critère d'arrêt de la méthode de dichotomie à une précision ε près donnée.
L'inégalité (3.4) nous permet aussi d'estimer à l'avance le nombre d'itérations nécessaires pour approcher s par $x^{(n)}$ avec la précision ε.

En effet, pour avoir une précision ne dépassant pas ε, il suffit d'avoir $\frac{b-a}{2^{n+1}} \leq \varepsilon \Rightarrow$

Formule du nombre d'itérations nécessaires à dichotomie $$n \geq \frac{Ln\left(\frac{b-a}{2\varepsilon}\right)}{Ln2} \tag{3.5}$$

Exemple d'application

Approcher à 10^{-2} près la racine exacte de $F(x) = x^2 - 4 = 0$ comprise entre 1.7 et 2.2 par la méthode de dichotomie.

Solution

La fonction F étant continue, strictement croissante et vérifiant $F(1.7).F(2.2) < 0$ donc il existe une racine séparée dans l'intervalle $[1.7; 2.2]$.
Avant d'appliquer la méthode de dichotomie, calculons d'abord le nombre d'itérations nécessaires donné par la formule (3.5).

$n \geq \frac{Ln\left(\frac{b-a}{2\varepsilon}\right)}{Ln2} \geq \frac{Ln\left(\frac{2.2-1.7}{2.10^{-2}}\right)}{Ln2} \simeq 4.64$ pour n = 0,1,...,$E[4.64]+1$ où $E[x]$ représente la partie entière du nombre réel x. Donc la méthode de dichotomie nécessite 6 itérations de 0 à 5 (voir tableau ci-dessous).

n	$a^{(n)}$	$b^{(n)}$	$x^{(n)}$	$F(x^{(n)})$
0	1.7	2.2	1.95	-0.1975
1	1.95	2.2	2.075	0.30
2	1.95	2.075	2.0125	0.05
3	1.95	2.0125	1.98125	-0.07
4	1.98125	2.0125	1.99687	-0.01
5	1.99687	2.0125	2.00468	0.01

Toutes les valeurs du dernier intervalle $[x^{(4)},x^{(5)}] = [1.99687; 2.00468]$ peuvent être considérées comme racines approchées de la racine exacte $s = 2$, en particulier on prend pour $\widetilde{s} = x^{(6)} = \frac{x^{(4)} + x^{(5)}}{2} \simeq 2.0007$ et on conclut que $s \simeq 2.0007 \pm 0.01$.

3.5 Méthode du point fixe

La méthode du point fixe, dite aussi méthode des approximations successives, est l'une des méthodes les plus importantes de résolution numérique des équations du type (3.1). C'est une méthode itérative puisque à partir de l'équation (3.1), on construit la suite itérative suivante définie par:

$$\begin{cases} x^{(0)} \text{ donné dans } [a,b] \\ \\ x^{(n+1)} = g(x^{(n)}), \quad n \in \mathbb{N} \end{cases} \tag{3.6}$$

Définition 3.6.
Soit g une fonction définie de $\mathbb{R}$ dans $\mathbb{R}$, le réel $\overline{x}$ est un point fixe de g s'il vérifie $g(\overline{x}) = \overline{x}$.

3.5.1 Principe de la méthode du point fixe

On remplace l'équation $f(x) = 0$ à résoudre où f est une fonction continue par une équation équivalente sous la forme $x = g(x)$. Ceci est toujours possible en posant $g(x) = x + f(x)$. Ainsi, résoudre l'équation (3.1) revient à déterminer les points fixes de la fonction g. Le problème sera alors de déterminer les conditions que doit vérifier la fonction g pour que la suite $(x^{(n)})_{n\in\mathbb{N}}$ converge vers la racine exacte s de l'équation (3.1).

Exemple d'application

L'équation $F(x) = 0$ où $F(x) = x^5 - 2x + 1$ peut se mettre sous la forme $x = g(x)$ avec $g_1(x) = x^5 - x + 1 = x$ ou $g_2(x) = \dfrac{x^5+1}{2} = x$ ou $g_3(x) = \sqrt[5]{2x-1} = x$ ou $g_4(x) = \dfrac{2}{x^3} - \dfrac{1}{x^4} = x$ ou $g_5(x) = \dfrac{1}{2-x^4}$...etc.

Théorème 3.7. *Soit l'intervalle réel non vide* $[a,b]$ *et* $g : [a,b] \mapsto [a,b]$ *une application continue. Alors, il existe un point* $\overline{x} \in [a,b]$ *vérifiant* $g(\overline{x}) = \overline{x}$.

Démonstration.
De l'équation $g(x) = x + f(x)$, on obtient $f(x) = g(x) - x$ et on a $f(a) = g(a) - a \geq 0$ et $f(b) = g(b) - b \leq 0$ car $g(x) \in [a,b]$ pour tout $x \in [a,b]$. Comme la fonction f est continue sur $[a,b]$ et vérifiant la relation $f(a).f(b) < 0$ donc le théorème des valeurs intermédiaires (3.2) assure l'existence d'au moins une racine $\overline{x}$ de l'équation $f(x) = 0$ c'est à dire $g(\overline{x}) = \overline{x}$. □

Exemple d'application

L'équation $F(x) = e^x - 2x - 1 = 0$ admet une racine séparée dans l'intervalle [1,2] car F est continue strictement croissante sur [1,2] et vérifiant $F(1).F(2) < 0$.
Les différentes fonctions g obtenues à partir de $F(x) = 0$ sont $g_1(x) = Ln(2x+1)$, $g_2(x) = e^x - x - 1$ et $g_3(x) = (e^x - 1)/2$.
Appliquons le théorème précédent à ces différentes fonctions g_i $(i = 1,2,3)$.
1°) $g_1(x) = Ln(2x+1)$ satisfait les hypothèses car g_1 continue et croissante sur [1,2] avec $g_1(1) = 1.09 \in [1,2]$ et $g_1(2) = 1.61 \in [1,2]$.
On conclut l'existence d'un point fixe $\overline{x}$ tel que $g_1(\overline{x}) = \overline{x}$.
2°) $g_2(x) = e^x - x - 1$ ne satisfait pas les hypothèses du théorème (3.7) car $g_2(1) = 0.71 \notin [1,2]$.
3°) $g_3(x) = (e^x - 1)/2$ ne satisfait pas les hypothèses du théorème (3.7) car $g_3(2) = 3.19 \notin [1,2]$.
On conclut que les suites définies par $x^{(n+1)} = g_2(x^n)$ et $x^{(n+1)} = g_3(x^n)$ ne convergent pas vers la racine de l'équation $F(x) = 0$.

Définition 3.7.
Soit $[a,b]$ un intervalle non vide de $\mathbb{R}$ et g une application de $[a,b]$ dans $\mathbb{R}$. La fonction g est dite contractante sur $[a,b]$ s'il existe une constante $k \in]0,1[$ telle que $|g(x) - g(y)| \leq k|x - y|$,

pour tout x,$y \in [a,b]$.

Théorème 3.8. *Théorème du point fixe*
Soit g une fonction de classe $C^1([a,b])$ vérifiant les conditions suivantes:
(i) $g([a,b]) \subset [a,b] \Leftrightarrow a \leq g(x) \leq b$ *pour tout* $x \in [a,b]$,
(ii) g *est contractante sur* $[a,b] \Leftrightarrow \exists k \in]0,1[$ *tel que* $|g'(x)| \leq k$, $\forall x \in [a,b]$,
Alors la suite définie par $x^{(n+1)} = g(x^{(n)})$ *converge vers l'unique point fixe* $\overline{x}$ *de* g *pour toute valeur initiale* $x^{(0)} \in [a,b]$.

Remarque 3.4. On peut prendre pour la constante $k = \max\limits_{a \leq x \leq b} \left|g'(x)\right|$.

Démonstration. du théorème (3.8)
1°) Montrons que la suite $(x^{(n)})_n$ est de Cauchy $\Leftrightarrow$
$(\forall \varepsilon > 0), (\exists N_{(\varepsilon)} \in \mathbb{N}) : \forall n \geq N_{(\varepsilon)}, \forall p > q > N_{(\varepsilon)} \Rightarrow \left|x^{(p)} - x^{(q)}\right| \leq \varepsilon$.
Calculons l'expression $\left|x^{(n)} - x^{(n-1)}\right|$ en utilisant le théorème des accroissements finis (3.5).
$\left|x^{(n)} - x^{(n-1)}\right| = \left|g(x^{(n-1)}) - g(x^{(n-2)})\right| = \left|g'(\theta)(x^{(n-1)} - x^{(n-2)})\right| = \left|g'(\theta)\right| \left|x^{(n-1)} - x^{(n-2)}\right|$
$\leq k.\left|x^{(n-1)} - x^{(n-2)}\right| = k\left|g(x^{(n-2)}) - g(x^{(n-3)})\right| \leq k^2.\left|x^{(n-2)} - x^{(n-3)}\right|$
et par récurrence, on obtient $\left|x^{(n)} - x^{(n-1)}\right| \leq k^{n-1}.\left|x^{(1)} - x^{(0)}\right|$.

Soit le couple $(p,q) \in \mathbb{N}^* \times \mathbb{N}^*$ tel que $p > q$ et $p = q + n$ avec $n \in \mathbb{N}^* \Rightarrow q = p - n$.
$\left|x^{(p)} - x^{(q)}\right| = \left|x^{(p)} - x^{(p-1)} + x^{(p-1)} - x^{(p-2)} + ... + x^{(q+1)} - x^{(q)}\right| =$
$\left|x^{(p)} - x^{(p-1)} + x^{(p-1)} - x^{(p-2)} + ... + x^{(p-(n-1))} - x^{(p-n)}\right|$
$\leq \left|x^{(p)} - x^{(p-1)}\right| + \left|x^{(p-1)} - x^{(p-2)}\right| + ... + \left|x^{(p-(n-1))} - x^{(p-n)}\right|$
$\leq k^{p-1}.\left|x^{(1)} - x^{(0)}\right| + k^{p-2}.\left|x^{(1)} - x^{(0)}\right| + ... + k^{p-n}.\left|x^{(1)} - x^{(0)}\right|$
$\leq k^{p-1}\left[1 + k^{-1} + k^{-2} + ... + k^{-n+1}\right].\left|x^{(1)} - x^{(0)}\right|$
$$\leq k^{p-1}\left[\frac{1 - \left(\frac{1}{k}\right)^n}{1 - \frac{1}{k}}\right].\left|x^{(1)} - x^{(0)}\right|$$
$$\leq k^{p-n=q}.\left(\frac{1 - k^n}{1 - k}\right).\left|x^{(1)} - x^{(0)}\right|$$
$$\leq k^q.\left(\frac{1 - k^n}{1 - k}\right).\left|x^{(1)} - x^{(0)}\right|$$
Par hypothèses, $0 < k < 1$ donc $0 < 1 - k^n < 1$ d'où la relation

$$\left|x^{(p)} - x^{(q)}\right| \leq k^q.\frac{\left|x^{(1)} - x^{(0)}\right|}{1 - k} \quad \text{pour tout } p > q \tag{3.7}$$

Comme $0 < k < 1$ donc $\lim\limits_{q\to\infty} k^q = 0$ et donc $\lim\limits_{q\to\infty} k^q.\dfrac{\left|x^{(1)} - x^{(0)}\right|}{1-k} = 0 \Leftrightarrow$

$$(\forall \varepsilon > 0),(\exists N \in \mathbb{N})(\forall q > N) \;\Rightarrow\; |k^q - 0| \leq \varepsilon \text{ et } \left| k^q.\frac{\left|x^{(1)} - x^{(0)}\right|}{1-k} - 0 \right| \leq \varepsilon \tag{3.8}$$

D'après les relations (3.7) et (3.8), on conclut que:
$(\forall \varepsilon > 0),(\exists N_{(\varepsilon)} = N \in \mathbb{N})/\ \forall p > q > N_{(\varepsilon)} \;\Rightarrow\; \left|x^{(p)} - x^{(q)}\right| \leq \varepsilon.$
et la suite $(x^{(n)})_n$ est de Cauchy donc convergente.

2°) Calcul de la limite de la suite $(x^{(n)})_n$. Notons $\overline{x}$ cette limite, à partir de
$x^{(n+1)} = g(x^{(n)}) \Rightarrow \lim\limits_{n\to\infty} x^{(n+1)} = \lim\limits_{n\to\infty} g(x^{(n)}) = g(\lim\limits_{n\to\infty} x^{(n)}) \Rightarrow l = g(l).$
D'après les hypothèses, g est continue et $g([a,b]) \subset [a,b]$ donc d'après le théorème (3.7), $\overline{x}$ est racine de l'équation $x = g(x)$ et de plus $\overline{x} \in [a,b]$.

3°) Montrons l'unicité de la racine. Supposons que l'équation $x = g(x)$ admet deux racines distinctes $\overline{x}_1$ et $\overline{x}_2$ alors $\overline{x}_1 = g(\overline{x}_1)$ et $\overline{x}_2 = g(\overline{x}_2)$ d'où $|g(\overline{x}_1) - g(\overline{x}_2)| = |\overline{x}_1 - \overline{x}_2|$, d'autres part $|g(\overline{x}_1) - g(\overline{x}_2)| = \left|g'(\theta)\right| |\overline{x}_1 - \overline{x}_2| \leq k\,|\overline{x}_1 - \overline{x}_2|$, par conséquent
$|\overline{x}_1 - \overline{x}_2| \leq k\,|\overline{x}_1 - \overline{x}_2| \;\Rightarrow\; k \geq 1$ ce qui est absurde car $k < 1$ d'où l'unicité de la racine. □

3.5.2 Estimation de l'erreur de la méthode du point fixe

Pour $\forall p > n$, la formule (3.7) devient $\left|x^{(p)} - x^{(n)}\right| \leq k^n.\dfrac{\left|x^{(1)} - x^{(0)}\right|}{1-k}$
Comme la suite $(x^{(n)})_n$ converge vers la racine exacte $\overline{x}$, en faisant un passage à la limite (quand $p \to \infty$) dans la relation (3.7) on obtient la relation de l'estimation de l'erreur suivante:

$$\left|\overline{x} - x^{(n)}\right| \leq k^n.\frac{\left|x^{(1)} - x^{(0)}\right|}{1-k} \quad \text{pour tout } n \tag{3.9}$$

3.5.3 Critère d'arrêt de la méthode du point fixe

L'inégalité (3.9) nous permet d'estimer à l'avance le nombre d'itérations nécessaires pour approximer la racine exacte $\overline{x}$ avec une précision donnée ε.
En effet, ce nombre d'itérations est donnée par:

$$n \geq \frac{Ln\left(\dfrac{\varepsilon(1-k)}{\left|x^{(1)} - x^{(0)}\right|}\right)}{Lnk} \quad \text{avec} \quad k = \max_{a\leq x\leq b}\left|g'(x)\right| \tag{3.10}$$

Remarque 3.5. Dans la pratique, un autre critère d'arrêt pour cette méthode consiste à surveiller les chiffres significatifs d'une itération à l'autre. Si ceux-ci se maintiennent pendant plusieurs itérations, on peut supposer que la racine a été isolée à ce nombre de chiffres significatifs exacts près. Ce critère s'applique quand aucune précision n'est donnée.

3.5.4 Interprétation géométrique de la méthode du point fixe

Définition 3.8.
Soit g une fonction continue d'un intervalle non vide $[a,b]$ dans lui même et $\overline{x}$ un point fixe de g dans $[a,b]$.

- Le point $\overline{x}$ est dit point fixe attractif si la suite $(x^{(n)})_{n\in\mathbb{N}}$ définie par la relation (3.6) converge pour toute valeur initiale $x^{(0)}$ suffisamment proche de $\overline{x}$,
- Le point $\overline{x}$ est dit point fixe répulsif si la suite définie par la relation (3.6) ne converge pour aucune valeur initiale $x^{(0)}$ situé dans un voisinage de $\overline{x}$ avec $\overline{x} \neq x^{(0)}$.

Géométriquement, la méthode des approximations successives s'explique comme suit: on construit dans le plan xOy les courbes des fonctions $y = x$ et $y = g(x)$ ainsi, toute racine réelle $\overline{x}$ de l'équation $x = g(x)$ est l'abscisse d'un point d'intersection de ces deux courbes.
On part d'un point $(x^{(0)},g(x^{(0)}))$, on trace une parallèle à l'axe $\overrightarrow{Ox}$ qui coupe $y = x$ en un point M_0. Par M_0, on trace une perpendiculaire à $\overrightarrow{Ox}$ qui coupe la courbe $y = g(x)$ en M_1 puis on recommence le même processus à partir de M_1 jusqu'au point fixe $(\overline{x},g(\overline{x}))$.

- Si $-1 < g^{'}(x) \leq 0$, on a une convergence en spirale pour tout $x^{(0)}$ suffisamment proche de $\overline{x}$ et ce point est attractif, voir Figure 1,
- Si $0 \leq g^{'}(x) < 1$, on a une convergence en marche d'escalier pour tout $x^{(0)}$ suffisamment proche de $\overline{x}$ et ce point est attractif, voir Figure 3,
- Si $\left|g^{'}(x)\right| \geq 1$, on ne peut rien dire quand à la convergence ou la divergence de la suite, voir Figure 2 et Figure 4.

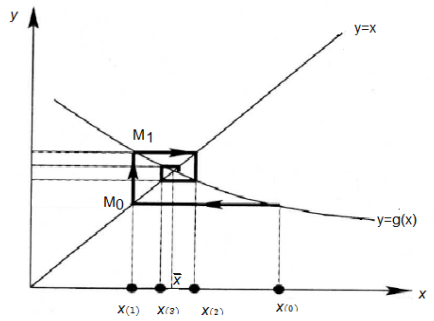

Figure 1: Convergence en spirale (-1<g'(x)⩽0)

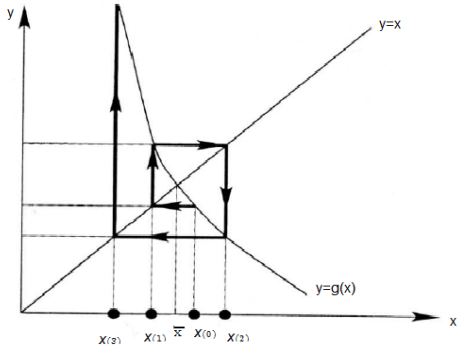

Figure 2: Divergence en spirale (g'(x)⩽-1)

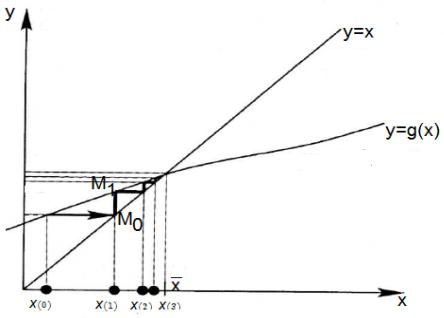

Figure 3: Convergence en escalier (0≤g'(x)<1)

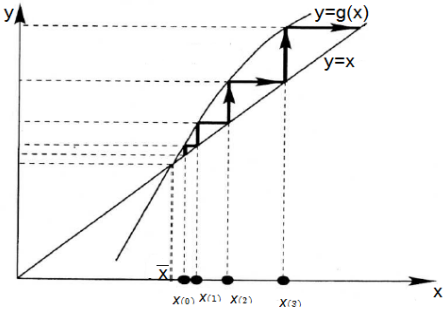

Figure 4: Divergence en escalier (1≤g'(x))

Exemple d'application

Utiliser la méthode du point fixe pour trouver la racine de l'équation $F(x) = x^3 - x - 1 = 0$ comprise dans l'intervalle [1,2] avec la précision $\varepsilon = 10^{-3}$.

Solution

1°) La racine $\overline{x}$ est séparée dans l'intervalle [1,2] car F est continue croissante sur [1,2] avec $F(1).F(2) < 0$.

2°) On détermine les fonctions g_i telles que $g_i(x) = x + F(x)$. On a $g_1(x) = (x+1)^{1/3}$, $g_2(x) = \dfrac{1}{x^2-1}$, $g_3(x) = x^3 - 1$ et $g_4(x) = \dfrac{1}{x} + \dfrac{1}{x^2}$.

3°) Les fonctions g_2, g_3 et g_4 ne vérifient pas les conditions du théorème (3.8) car $g_2(1.5) = 0.8 \notin [1,2]$, $g_3(1.5) = 2.37 \notin [1,2]$ et $g_4(2) = 0.75 \notin [1,2]$ donc on conclut que les suites définies par $x^{(n+1)} = g_2(x^{(n)})$, $x^{(n+1)} = g_3(x^{(n)})$ et $x^{(n+1)} = g_4(x^{(n)})$ ne convergent pas vers la racine exacte $\overline{x}$ de l'équation $F(x) = 0$.

4°) Appliquons le théorème (3.8) à la fonction $g_1(x) = (x+1)^{1/3}$.
-(a)- La fonction $g_1 \in C^1([1,2])$ car g est dérivable et g' continue sur [1,2].
-(b)- $g_1(1) = 1.25 \in [1,2]$, $g_1(2) = 1.44 \in [1,2]$ et g_1 est strictement croissante pour tout $x \in [1,2]$ donc $g_1([1,2]) = [g_1(1), g_1(2)] = [1.25; 1.44] \subset [1,2]$.
-(c)- $g_1'(x) = \frac{1}{3}.(x+1)^{-2/3}$ et $g_1''(x) = -\frac{2}{9}.(x+1)^{-5/3} < 0 \quad \forall x \in [1,2]$ donc la fonction g_1' est strictement décroissante sur [1,2] et son maximum est atteint au point $x = 1$ ainsi $k = \max\limits_{1 \le x \le 2} \left|g_1'(x)\right| = \left|g'(1)\right| \simeq 0.21$ d'où il existe $0 < k \simeq 0.21 < 1$ tel que $\left|g_1'(x)\right| \le 0.21$ pour tout $x \in [1,2]$.

Conclusion: De (a), (b) et (c), on déduit que d'après le théorème du point fixe (3.8), la suite définie par $x^{(n+1)} = g_1(x^{(n)}) = (x^{(n)}+1)^{1/3}, \quad n \in \mathbb{N}$ converge vers la racine exacte $\overline{x}$ de $g_1(x) = x$ ou $F(x) = 0$ pour toute approximation initiale $x^{(0)} \in [1,2]$.

5°) Déterminons le nombre d'itérations nécessaires pour approcher $\overline{x}$ avec la précision $\varepsilon = 10^{-3}$. La formule (3.10) nous donne avec par exemple $x^{(0)} = 1.5 \in [1,2]$ et $x^{(1)} = g_1(x^{(0)}) = 1.3572$.

$$n \geq \frac{Ln\left(\frac{\varepsilon(1-k)}{|x^{(1)}-x^{(0)}|}\right)}{Lnk} = \frac{Ln\left(\frac{10^{-3}(1-0.21)}{|1.3572-1.5|}\right)}{Ln0.21} \simeq 3.33 \text{ pour } n = 0,...,E[3.33]+1 = 4,$$

donc il faut effectuer 5 itérations.

6°) Première itération $x^{(1)} = g_1(x^{(0)}) = 1.3572$.
Deuxième itération $x^{(2)} = g_1(x^{(1)}) = 1.3308$.
Troisième itération $x^{(3)} = g_1(x^{(2)}) = 1.3258$.
Quatrième itération $x^{(4)} = g_1(x^{(3)}) = \underline{1.3249}$.
Cinquième itération $x^{(5)} = g_1(x^{(4)}) = \underline{1.3248}$.

On arrête, $x^{(5)}$ est la racine approchée de $\overline{x}$, d'où $\overline{x} \simeq x^{(5)} \pm \varepsilon \Rightarrow \overline{x} \simeq 1.324 \pm 0.001$.

Remarque 3.6. Si on avait à calculer $x^{(6)} = g_1(x^{(5)}) = \underline{1.3247}3$, $x^{(7)} = g_1(x^{(6)}) = \underline{1.3247}1$, $x^{(8)} = g_1(x^{(7)}) = \underline{1.3247}16$...etc. On remarque qu'à partir de la quatrième itération les chiffres 1.324 reviennent à chaque itération donc on peut conclure que la racine sera $\overline{x} \simeq 1.324$ (voir remarque (3.5)).

3.6 Méthode de Newton-Raphson

3.6.1 Principe de la méthode de Newton-Raphson

La méthode de Newton dite aussi méthode des tangentes est une méthode itérative. Notons par $\overline{x}$ une racine exacte de $f(x) = 0$ et $x^{(0)}$ une valeur approchée de $\overline{x}$. On suppose que f est de classe C^2 au voisinage de $\overline{x}$. En appliquant le développement de Taylor d'ordre un de f on obtient:

$$f(\overline{x}) = f(x^{(0)}) + f'(x^{(0)})(\overline{x}-x^{(0)}) + \frac{f''(\xi)}{2}(\overline{x}-x^{(0)})^2 \quad \text{où } \xi \in [\overline{x},x^{(0)}] \text{ ou } [x^{(0)},\overline{x}] \quad (3.11)$$

comme $f(\overline{x}) = 0$ et en supposant $f'(x^{(0)}) \neq 0$, la formule (3.11) devient:

$$\overline{x} = x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})} - \frac{f''(\xi)}{2f'(x^{(0)})}(\overline{x}-x^{(0)})^2 \quad (3.12)$$

En négligeant le reste $R = -\frac{f''(\xi)}{2f'(x^{(0)})}(\overline{x}-x^{(0)})^2$, la quantité $x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})}$ dans la relation (3.12) notée $x^{(1)}$ constitue alors une valeur approchée améliorée de $\overline{x}$. En

itérant le procédé, on trouve la formule de récurrence suivante:

$$\begin{cases} x^{(0)} \text{ donné} \\ x^{(k+1)} = x^{(k)} - \dfrac{f(x^{(k)})}{f'(x^{(k)})}, \quad k \in \mathbb{N} \quad \text{appelée algorithme de Newton-Raphson} \end{cases} \tag{3.13}$$

On vérifie immédiatement que $\overline{x}$ est un point fixe de la fonction $g(x) = x - \dfrac{f(x)}{f'(x)}$ car $g(\overline{x}) = \overline{x}$ avec supposition que $f'(x) \neq 0$.

Théorème 3.9. *Théorème de Newton-Raphson*
Soit f une fonction définie dans l'intervalle non vide $[a,b]$ de $\mathbb{R}$ et vérifiant:
(1) $f \in C^2([a,b])$,
(2) $f(a).f(b) < 0$,
(3) $f'(x) \neq 0 \quad \forall x \in [a,b]$,
(4) $f''(x) \neq 0 \quad \forall x \in [a,b]$,
Alors, pour toute initialisation $x^{(0)} \in [a,b]$ vérifiant $f(x^{(0)}).f''(x^{(0)}) > 0$, la suite $(x^{(n)})_n$ définie par la relation (3.13) converge vers l'unique racine $\overline{x}$ de l'équation $f(x) = 0$.

Démonstration.
Les hypothèses (2) et (3) vérifiées par la fonction f continue impliquent qu'il existe une unique racine $\overline{x} \in [a,b]$.
On suppose que $f(a) < 0$, $f(b) > 0$, $f'(x) > 0$, $f''(x) > 0$ et on prend $x^{(0)} = b$ (car $f(x^{(0)}).f''(x^{(0)}) > 0$).

1°) Montrons que la suite $(x^{(n)})_n$ est minorée par $\overline{x}$.
Comme $f \in C^2([a,b])$, on peut utiliser le développement de Taylor à l'ordre 1 au voisinage de $x^{(n)}$ dans l'intervalle $[\overline{x},x^{(n)}]$ ou $[x^{(n)},\overline{x}]$, on obtient:

$$f(\overline{x}) = f(x^{(n)}) + (\overline{x} - x^{(n)})f'(x^{(n)}) + \frac{(\overline{x} - x^{(n)})^2}{2} f''(\theta) \text{ où } \theta \in [\overline{x},x^{(n)}] \Rightarrow$$

$$0 = f(x^{(n)}) + (\overline{x} - x^{(n)})f'(x^{(n)}) + \frac{(\overline{x} - x^{(n)})^2}{2} f''(\theta) \quad \Rightarrow$$

$$f(x^{(n)}) + (\overline{x} - x^{(n)})f'(x^{(n)}) = -\frac{(\overline{x} - x^{(n)})^2}{2} f''(\theta) < 0 \text{ car } f''(x) > 0 \ \forall x \Rightarrow$$

$$f(x^{(n)}) < -(\overline{x} - x^{(n)})f'(x^{(n)}) \Rightarrow x^{(n)} - \frac{f(x^{(n)})}{f'(x^{(n)})} > \overline{x}$$

c'est à dire $x^{(n+1)} > \overline{x}$ pour tout n donc $(x_n)_n$ est minorée par $\overline{x}$.

2°) Montrons que la suite $(x^{(n)})_n$ est décroissante.
$x^{(n+1)} - x^{(n)} = \left(x^{(n)} - \dfrac{f(x^{(n)})}{f'(x^{(n)})}\right) - x^{(n)} = -\dfrac{f(x^{(n)})}{f'(x^{(n)})} < 0$ car pour $x^{(n)} > \overline{x}$ on a $f(x^{(n)}) > f(\overline{x}) = 0$ car f est continue croissante et de plus $f'(x) > 0$ pour tout x.
Par conséquent, $x^{(n+1)} - x^{(n)} < 0$ et la suite $(x^{(n)})_n$ est décroissante.

La suite $(x^{(n)})_n$ étant minorée décroissante donc convergente. On note l sa limite.

3°) Calcul de la limite de la suite $(x^{(n)})_n$ quand $n \to \infty$.
Par passage à la limite dans la formule $x^{(n+1)} = x^{(n)} - \dfrac{f(x^{(n)})}{f'(x^{(n)})}$, on obtient $l = l - \dfrac{f(l)}{f'(l)} \Rightarrow$
$\dfrac{f(l)}{f'(l)} = 0 \Rightarrow f(l) = 0$ car $f'(l) \neq 0 \Rightarrow l$ est une racine de l'équation $f(x) = 0$.
Comme cette équation admet une racine unique $\overline{x}$ dans [a,b] donc $l = \overline{x}$. □

3.6.2 Estimation de l'erreur de la méthode de Newton

Généralement, on utilise la relation suivante donnée au théorème (3.6)

$$\left|\overline{x} - x^{(n)}\right| \leq \frac{\left|f(x^{(n)})\right|}{m} \tag{3.14}$$

pour estimer l'erreur commise avec la méthode de Newton où $m = \min\limits_{a \leq x \leq b} \left|f'(x)\right|$.

Autre formule d'approximation de l'erreur

Écrivons le développement de Taylor d'ordre 1 au voisinage de $x^{(n-1)}$ sur l'intervalle [$x^{(n-1)}$,$x^{(n)}$].
$f(x^{(n)}) = f(x^{(n-1)}) + \left(x^{(n)} - x^{(n-1)}\right) f'(x^{(n-1)}) + \dfrac{\left(x^{(n)} - x^{(n-1)}\right)^2}{2}.f''(\theta)$ où $\theta \in [x^{(n-1)},x^{(n)}]$.
or $x^{(n)} = x^{(n-1)} - \dfrac{f(x^{(n-1)})}{f'(x^{(n-1)})}$ d'où
$f(x^{(n)}) = f(x^{(n-1)}) + \left(x^{(n-1)} - \dfrac{f(x^{(n-1)})}{f'(x^{(n-1)})} - x^{(n-1)}\right) f'(x^{(n-1)}) + \dfrac{\left(x^{(n)} - x^{(n-1)}\right)^2}{2}.f''(\theta)$
$f(x^{(n)}) = \dfrac{\left(x^{(n)} - x^{(n-1)}\right)^2}{2}.f''(\theta) \quad \Rightarrow$
$\left|f(x^{(n)})\right| = \dfrac{\left(x^{(n)} - x^{(n-1)}\right)^2}{2}.\left|f''(\theta)\right| \leq \dfrac{M}{2}\left(x^{(n)} - x^{(n-1)}\right)^2$

avec $M = \max\limits_{a\leq x\leq b} \left|f''(x)\right| \geq \left|f''(\theta)\right|$ et en utilisant la formule (3.14), on obtient une deuxième formule d'approximation de l'erreur de la méthode de Newton-Raphson.

$$\left|\overline{x} - x^{(n)}\right| \leq \frac{M}{2m}\left(x^{(n)} - x^{(n-1)}\right)^2 \tag{3.15}$$

Avec $M = \max\limits_{a\leq x\leq b} \left|f''(x)\right|$ et $m = \min\limits_{a\leq x\leq b} \left|f'(x)\right|$.

3.6.3 Critère d'arrêt de la méthode de Newton

On arrête le processus itératif de Newton dès qu'on a la relation suivante:

$$\left|x^{(n)} - x^{(n-1)}\right| \leq \varepsilon \quad (\varepsilon > 0) \tag{3.16}$$

3.6.4 Interprétation géométrique de la méthode de Newton

Dans cette méthode, il s'agit de remplacer la fonction f dont on cherche une racine par sa tangente en un point voisin de cette racine. Ainsi, à partir d'un point $x^{(0)}$ proche de la solution $\overline{x}$ de l'équation $f(x) = 0$, on construit la tangente (d_0) à la courbe $y = f(x)$ au point $\left(x^{(0)},f(x^{(0)}\right)$. Cette tangente coupe l'axe $\overrightarrow{Ox}$ au point $x^{(1)}$, l'équation de cette tangente s'écrit $y - f(x^{(0)}) = (x - x^{(0)})f'(x^{(0)})$ d'où les coordonnées du point $x^{(1)}$ sont $y = 0$ et $y - f(x^{(0)}) = (x^{(1)} - x^{(0)})f'(x^{(0)})$ c'est à dire
$x^{(1)} = x^{(0)} - \dfrac{f(x^{(0)})}{f'(x^{(0)})}$. En appliquant encore ce procédé une fois au point $\left(x^{(1)},f(x^{(1)})\right)$, on trouvera $x^{(2)} = x^{(1)} - \dfrac{f(x^{(1)})}{f'(x^{(1)})}$ et ainsi de suite (voir Figure 5).
La limite de la suite $(x^{(n)})_n$ ainsi obtenue donne la racine cherchée $\overline{x}$.

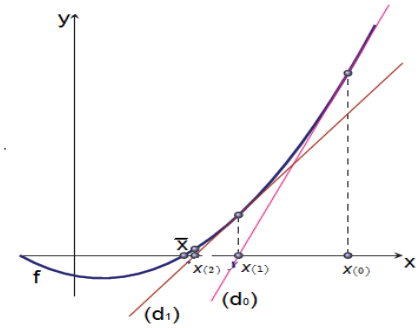

Figure 5: Interprétation géométrique de la méthode de Newton

Exemple d'application

Approcher la racine s de l'équation $F(x) = x^3 + 2x^2 + 10x - 20 = 0$ sur l'intervalle [1,2] par la méthode de Newton avec la précision $\varepsilon = 2.10^{-4}$.

Solution

Vérifions les hypothèses du théorème (3.9) pour la fonction F.
(1) $F \in C^2([1,2])$ car F est un polynôme,
(2) $F(1).F(2) = (-7)16 < 0$,
(3) $F'(x) = 3x^2 + 4x + 10 > 0$ pour tout $x \in [1,2]$,
(4) $F''(x) = 6x + 4 > 0$ pour tout $x \in [1,2]$,
(5) Pour $x^{(0)} = 1.5 \in [1,2]$ on a bien $F(1.5).F''(1.5) > 0$ donc $x^{(0)} = 1.5$ est une bonne approximation initiale.
La fonction F vérifie les hypothèses du théorème (3.9) donc la suite définie par:

$$\begin{cases} x^{(0)} = 1.5 \\ x^{(n+1)} = x^{(n)} - \dfrac{f(x^{(n)})}{f'(x^{(n)})}, \quad n \in \mathbb{N} \end{cases}$$ converge vers la racine exacte s de $F(x) = 0$.

Approximation de s

Pour $x^{(0)} = 1.5$, on calcule $x^{(1)} = x^{(0)} - \dfrac{F(x^{(0)})}{F'(x^{(0)})} = 1.37362$ et on applique le test d'arrêt (3.19) pour avoir $\left|x^{(1)} - x^{(0)}\right| = 0.126 > \varepsilon$ donc on continue les itérations.

Pour $x^{(1)} = 1.37362$, on calcule $x^{(2)} = x^{(1)} - \dfrac{F(x^{(1)})}{F'(x^{(1)})} = 1.36889$ et on applique le test d'arrêt (3.19) pour avoir $\left|x^{(2)} - x^{(1)}\right| = 0.0047 > \varepsilon$ donc on continue les itérations.
Pour $x^{(2)} = 1.36889$, on calcule $x^{(3)} = x^{(2)} - \dfrac{F(x^{(2)})}{F'(x^{(2)})} = 1.36880$ et on applique le test d'arrêt (3.19) pour avoir $\left|x^{(3)} - x^{(2)}\right| = 9.10^{-5} < \varepsilon$ et on arrête les itérations ($x^{(3)}$ est la racine approchée de s).
On conclut que $\left|s - x^{(3)}\right| \leq \varepsilon \;\Rightarrow\; s \simeq x^{(3)} \pm \varepsilon$ d'où $s \simeq 1.3688 \pm 0.0002$.

3.7 Méthode de la sécante

3.7.1 Principe de la méthode de la sécante

La méthode de Newton exige le calcul de la dérivée de la fonction f et dans certaines situations, la dérivée f' est très compliquée ou même impossible à expliciter. Il est possible d'obtenir une variante de la méthode de Newton appelée "méthode de la sécante" encore appelée "méthode de la fausse position", son principe est d'approximer, dans l'algorithme de newton donné par la relation (3.13), la dérivée $f'(x^{(k)})$ par le taux d'accroissement de f sur $[x^{(k-1)},x^{(k)}]$ c'est à dire:

$$f'(x^{(k)}) \simeq \frac{f(x^{(k)}) - f(x^{(k-1)})}{x^{(k)} - x^{(k-1)}} \tag{3.17}$$

Ainsi l'algorithme itératif de la méthode de la sécante est donné par:

$$\begin{cases} x^{(0)} \text{ , } x^{(1)} \text{ donnés} \\ x^{(k+1)} = x^{(k)} - \dfrac{f(x^{(k)}).(x^{(k)} - x^{(k-1)})}{f(x^{(k)}) - f(x^{(k-1)})}, \quad k \geq 1 \end{cases} \tag{3.18}$$

Remarque 3.7.
1°) L'algorithme itératif de la méthode de la sécante ne peut démarrer que si on dispose déjà de deux valeurs approchées $x^{(0)}$ et $x^{(1)}$ (avec $x^{(0)} \neq x^{(1)}$) de la racine exacte $\overline{x}$ de l'équation à résoudre $f(x) = 0$.
2°) La méthode de la sécante n'est définie que si on a toujours $f(x^{(k)}) \neq f(x^{(k-1)})$, $\forall k \geq 1$.

3.7.2 Critère d'arrêt de la méthode de la sécante

On arrête le processus itératif de la méthode de la sécante dès qu'on a la relation suivante:

$$\left|x^{(n)} - x^{(n-1)}\right| \leq \varepsilon \quad (\varepsilon > 0) \tag{3.19}$$

3.7.3 Interprétation géométrique de la méthode de la sécante

La méthode de la sécante consiste à se donner deux points d'initialisation $x^{(0)}$ et $x^{(1)}$ comme valeurs approchées de la racine exacte $\overline{x}$, tracer la droite qui passe par les points $\left(x^{(0)},f(x^{(0)})\right)$ et $\left(x^{(1)},f(x^{(1)})\right)$, cette droite coupe l'axe des x en $x^{(2)}$ puis on recommence le principe avec les itérés $x^{(1)}$ et $x^{(2)}$ pour obtenir $x^{(3)}$ et ainsi de suite. En généralisant le processus, pour tout entier positif k, le point $x^{(k+1)}$ est le point d'intersection de l'axe des abscisses avec la droite passant par les points $\left(x^{(k-1)},f(x^{(k-1)})\right)$ et $\left(x^{(k)},f(x^{(k)})\right)$ de la courbe représentative de la fonction f (voir Figure 6).

Exemple d'application

Reprenons l'exemple d'application fait avec la méthode de Dichotomie à la section (3.4), en approchant à 10^{-2} près la racine exacte s de $F(x) = x^2 - 4 = 0$ comprise entre 1.7 et 2.2 par la méthode de la sécante.

Solution

La fonction F étant continue, strictement croissante et vérifiant $F(1.7).F(2.2) < 0$ donc il existe une racine séparée dans l'intervalle $[1.7; 2.2]$. En utilisant la relation (3.18), on obtient le tableau suivant:

n	$x^{(n)}$	$x^{(n+1)}$	$x^{(n+2)}$	$F(x^{(n+2)})$
0	1.7	2.2	1.984615	-0.061303
1	2.2	1.984615	1.999264	-0.002943
2	1.984615	1.999264	2.000002	0.000008

Comme $|x^{(4)} - x^{(3)}| < 0.01 = \varepsilon$ donc on arrête la méthode après trois itérations et on conclut que $s = x^{(4)} \pm 0.01$ c'est à dire $s = 2.000002 \pm 0.01$ (La méthode de Dichotomie a nécessité 06 itérations).

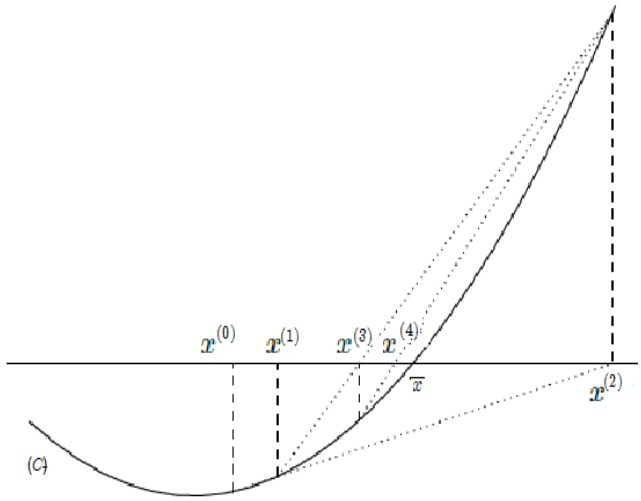

Figure 6: Interprétation géométrique de la méthode de la sécante

3.8 Exercices corrigés sur le chapitre

Exercice 1 Soit f une fonction de classe C^1 sur $[a,b]$ vérifiant:

$$f(b) = a \quad \text{et} \quad -\frac{1}{2} \leq f'(x) \leq 0 \quad \forall x \in [a,b]. \tag{3.20}$$

(1) Montrer que $a \leq f(x) \quad \forall x \in [a,b]$.

(2) En utilisant le théorème des accroissements finis sur $[a,b]$, prouver que $f(x) \leq b \quad \forall x \in [a,b]$.

(3) Montrer que la suite définie par $x_{n+1} = f(x_n)$ converge vers la solution α de $x = f(x)$ pour tout $x_0 \in [a,b]$.

(4) En supposant que $\alpha > x_n$, montrer que $|x_n - \alpha| \leq |x_n - x_{n-1}|$.

Solution de l'exercice 1:

(1) Comme $f'(x) \leq 0, \quad \forall x \in [a,b]$ donc la fonction f est décroissante sur l'intervalle $[a,b]$ donc $f(b) \leq f(a)$. Ainsi, $f(b) = a \leq f(x) \leq f(a)$, $\forall x \in [a,b]$ et on conclut que $a \leq f(x)$, $\forall x \in [a,b]$.

(2) La fonction f étant continue et dérivable sur $[a,b]$ donc on peut lui appliquer le théorème des accroissements finis sur $[a,b]$ en écrivant
$\exists c \in]a,b[$ tel que $f(b) - f(a) = (b-a)f'(c)$. En utilisant la question (1) et la relation (3.20), on obtient
$a - f(a) \geq (b-a)(\frac{-1}{2}) = \frac{a}{2} - \frac{b}{2}$ et donc $\frac{a+b}{2} \geq f(a)$ avec $\frac{a+b}{2} < b$ et comme $f(x) \leq f(a) \leq \frac{a+b}{2} < b$, $\forall x \in [a,b]$ d'où $f(x) \leq b, \quad \forall x \in [a,b]$.

(3) Appliquons le théorème du point fixe à la fonction f sur l'intervalle $[a,b]$.
a) $f \in C^1([a,b]$.
b) $\frac{-1}{2} \leq f'(x) \leq 0 \leq \frac{1}{2}$ donc $|f'(x)| \leq \frac{1}{2} = k$, $\forall x \in [a,b]$.
c) D'après les questions (1) et (2), on a $a \leq f(x) \leq b$, $\forall x \in [a,b]$ c'est à dire $f([a,b]) \subset [a,b]$.

Les hypothèses du théorème fixe étant vérifiées donc la suite définie par $x_{n+1} = f(x_n)$ converge vers la solution α de $x = f(x)$ pour tout $x_0 \in [a,b]$.

(4) Supposons que la racine exacte $\alpha > x_n$ et appliquons le théorème des accroissements finis sur $[x_{n-1},\alpha]$ ou $[\alpha,x_{n-1}]$. On obtient, $f(\alpha) - f(x_{n-1}) = (\alpha - x_{n-1})f'(\theta)$ avec $\theta \in]x_{n-1},\alpha[$. Comme $\alpha = f(\alpha)$, $f(x_{n-1}) = x_n$ et $f'(\theta) \leq 0$, on obtient
$\alpha - x_n = (\alpha - x_{n-1})f'(\theta)$ d'où $\alpha - x_{n-1} \leq 0$ ainsi $x_n < \alpha < x_{n-1}$ et que
$|x_n - \alpha| \leq |x_n - x_{n-1}|$.

Exercice 2 Soit F: $\mathbb{R} \mapsto \mathbb{R}$ définie par $F(x) = cosx + 2x$.
On veut trouver une solution approchée de l'équation $F(x) = 0$.

(1) Montrer graphiquement que $F(x) = 0$ admet une solution unique l sur $\mathbb{R}$ et que $l \in]a,b[= \left]\frac{-\Pi}{2},0\right[$.

(2) Appliquer la méthode de Dichotomie quatre fois pour réduire l'intervalle $]a,b[$. Soit I cet intervalle réduit.

(3) Appliquer la méthode du point fixe à la fonction $g(x) = -\frac{cosx}{2}$ sur I.

Donner dans chaque cas la solution approchée de l avec la précision $\varepsilon = 0{,}1$ près.

Solution de l'exercice 2:

(1) Graphiquement, l'intersection des courbes des fonctions $y = cosx$ et $y = -2x$ fournit approximativement une racine unique $l < 0$ et comme $F(0).F(\frac{-\Pi}{2}) < 0$ donc il existe une racine unique $l \in]a,b[= \left]\frac{-\Pi}{2},0\right[$ de l'équation $F(x) = 0$.

(2) Appliquons quatre fois la méthode de Dichotomie sur l'intervalle $]a,b[= \left]\frac{-\Pi}{2},0\right[$.

Première itération: On divise l'intervalle $]a,b[= \left]\frac{-\Pi}{2},0\right[$ en deux, son milieu est $x_0 = \frac{-\Pi}{4}$ et $F\left(\frac{-\Pi}{4}\right) < 0$ donc le nouvel intervalle à diviser est $\left]\frac{-\Pi}{4},0\right[$.

Deuxième itération: On divise l'intervalle $\left]\frac{-\Pi}{4},0\right[$ en deux, son milieu est $x_1 = \frac{-\Pi}{8}$ et $F\left(\frac{-\Pi}{8}\right) > 0$ donc le nouvel intervalle à diviser est $\left]\frac{-\Pi}{4},\frac{-\Pi}{8}\right[$.

Troisième itération: On divise l'intervalle $\left]\frac{-\Pi}{4},\frac{-\Pi}{8}\right[$ en deux, son milieu est $x_2=\frac{-3\Pi}{16}$ et $F\left(\frac{-3\Pi}{16}\right)<0$ donc le nouvel intervalle à diviser est $\left]\frac{-3\Pi}{16},\frac{-\Pi}{8}\right[$.

Quatrième itération: On divise l'intervalle $\left]\frac{-3\Pi}{16},\frac{-\Pi}{8}\right[$ en deux, son milieu est $x_3=\frac{-5\Pi}{32}$ et $F\left(\frac{-5\Pi}{32}\right)<0$ et le dernier intervalle à diviser est $I=\left]\frac{-5\Pi}{32},\frac{-\Pi}{8}\right[$.

On peut prendre par exemple, comme racine approchée de l, le milieu de l'intervalle I c'est à dire $\widetilde{l}=\frac{-9\Pi}{64}\approx -0.4418$ d'où $l=\widetilde{l}\pm\varepsilon=-0.44\pm 0.1$

(3) Application de la méthode du point fixe à la fonction $g(x)=-\frac{cosx}{2}$ sur I.
a) La fonction $g\in C^1(I)$ car g continue, $g'(x)=\frac{sinx}{2}$ existe et est continue sur I.
b) $g\left(-\frac{5\Pi}{32}\right)=-0.44\in I$, $g\left(\frac{-\Pi}{8}\right)=-0.46\in I$ et $g'(x)=\frac{sinx}{2}<0$ sur I donc g est décroissante sur I et on a $g(I)=]-0.46,-0.46[\subset I$ d'où la stabilité de I par g.
c) $k=\max\limits_{x\in I}|g'(x)|=\max\limits_{x\in I}\left(-\frac{sinx}{2}\right)=\frac{-sin\left(\frac{-5\Pi}{32}\right)}{2}\approx 0.23$ car $\left(-\frac{sinx}{2}\right)$ est décroissante sur I donc il existe $0<k\approx 0.23<1$ tel que $|g'(x)|\leq k$.

La fonction g vérifie les hypothèses du théorème du point fixe donc la suite définie par $x_{n+1}=g(x_n)$ converge vers la racine exacte l pour $\forall x_0\in I$.

Calculons le nombre d'itérations nécessaires à la méthode du point fixe pour approcher la racine l à 10^{-1} près.
Pour $x_0=\frac{-\Pi}{8}\approx -0.3927$, on a $x_1=g(x_0)\approx -0.4619$ et le nombres d'itérations est donné par la formule:

$$n\geq\frac{Ln\left(\frac{\varepsilon(1-k)}{|x_1-x_0|}\right)}{Lnk}=\frac{Ln\left(\frac{10^{-1}(1-0.23)}{|-0.4619+03927|}\right)}{Ln0.23}\approx 1.5$$

Il faut faire deux itérations. On calcule $x_2=g(x_1)=-0.4476$ et donc $l\approx x_2\pm\varepsilon$ c'est à dire $l\approx -0.44\pm 0.1$.

Exercice 4 Soit F la fonction définie par $F(x) = x^2.e^x + 2x - 1$.

(1) Vérifier que l'équation $F(x) = 0$ admet une racine séparée dans l'intervalle $[0.3; 0.5]$.

(2) **(a)** Montrer que l'équation $F(x) = 0$ est équivalente à l'équation $x = f(x)$ où $f(x) = x - \alpha F(x)$ et $\alpha \in \mathbb{R}_+^*$.

(b) Calculer $\lambda = \dfrac{1}{\max\limits_{0.3 \leq x \leq 0.5} |F'(x)|}$.

(3) Dans cette question, on prend $\alpha = \lambda + 0.11$.

(a) Montrer que la méthode des approximations successives est applicable dans l'intervalle $[0.3; 0.5]$.

(b) Quel est le nombre d'itérations nécessaires pour avoir une valeur approchée de s à 10^{-3} près avec $x_0 = 0{,}4$.

(4) **(a)** Vérifier que la méthode de Newton est applicable sur $[0.3; 0.5]$.

(b) Approcher la racine s à 10^{-3} près.

(5) Quelle est la méthode la plus avantageuse. Dites pourquoi?.

Solution de l'exercice 3:

(1) $F(0.3) \approx -0.27 < 0$, $F(0.5) \approx 0.41 > 0$ et F monotone sur $[0.3; 0.5]$ donc l'équation $F(x) = 0$ admet une racine séparée $s \in [0.3; 0.5]$.

$(2)_{(a)}$ $F(x) = 0 \Leftrightarrow x = f(x)$ où $f(x) = x - \alpha.F(x)$ et $\alpha \in \mathbb{R}_+^*$.
($\Rightarrow$) Si $F(x) = 0 \quad \Rightarrow \quad f(x) = x$.
($\Leftarrow$) Si $f(x) = x \quad \Rightarrow \quad f(x) = x - \alpha.F(x) = x \quad \Rightarrow \quad -\alpha.F(x) = 0 \quad \Rightarrow \quad F(x) = 0$.

$(2)_{(b)}$ Calcul de $\max\limits_{0.3 \leq x \leq 0.5} |F'(x)| = \max(F'(x)) = F'(0.5) \approx 4.06$ car F' est positive et croissante sur $[0.3; 0.5]$. Donc $\lambda = \dfrac{1}{\max\limits_{0.3 \leq x \leq 0.5} |F'(x)|} = \dfrac{1}{4.06} \approx 0.24$.

$(3)_{(a)}$ On prend $\alpha = \lambda + 0.11 \approx 0.24 + 0.11 = 0.35$.
Appliquons la méthode des approximations successives à la fonction
$f(x) = x - 0.35.F(x) = x - 0.35(x^2.e^x + 2x - 1)$ sur l'intervalle $[0.3; 0.5]$.
a) $f \in C^1([0.3; 0.5])$ car f est continue, dérivable et f' est continue sur $[0.3; 0.5]$.

b) $f([0.3;0.5]) = [0.35;0.39] \subset [0.3;0.5]$ car f est décroissante sur $[0.3;0.5]$.
c) $k = \max\limits_{x\in[0.3;0.5]} |f'(x)| = \max\limits_{x\in[0.3;0.5]} (-f'(x)) = (-f'(0.5)) \approx 0.42$ car $(-f')$ est croissante sur $[0.3;0.5]$. Donc il existe $k \in]0,1[$ tel que $|f'(x)| \leq k \approx 0.42$ pour tout $x \in [0.3;0.5]$.

Les hypothèses du théorème du point fixe étant vérifiées donc la suite définie par $x_{n+1} = f(x_n)$ converge vers la racine exacte s de $f(x) = x$ (ou de $F(x) = 0$) quelque soit $x_0 \in [0.3;0.5]$.

$(3)_{(b)}$ Pour $x_0 = 0.4$, on a $x_1 = f(x_0) \approx 0.38$ d'où

$$n \geq \frac{Ln\left(\frac{\varepsilon(1-k)}{|x_1 - x_0|}\right)}{Lnk} = \frac{Ln\left(\frac{10^{-3}(1-0.42)}{|0.38-0.4|}\right)}{Ln0.42} \approx 4.08.$$

On peut donc prendre $n = 4$ itérations nécessaires pour la méthode du point fixe.

$(4)_{(a)}$ Application de la méthode de Newton à la fonction F sur $[0.3;0.5]$.
1) $F \in C^2([0.3;0.5])$.
2) s est séparée dans $[0.3;0.5]$.
3) $F''(x) = e^x\,(x^2 + 4x + 2) > 0, \ \ \forall x \in [0.3;0.5]$.
4) Pour $x_0 = 0.5$, on a bien $F(0.5).F''(0.5) > 0$.

La fonction F vérifie les hypothèses du théorème de Newton donc la suite définie par $x_{n+1} = x_n - \dfrac{F(x_n)}{F'(x_n)}$ converge vers la racine exacte s de l'équation $f(x) = 0$.

$(4)_{(b)}$ Pour $x_0 = 0.5$,

Première itération: $x_1 = x_0 - \dfrac{F(x_0)}{F'(x_0)} \approx 0.3990$ et $|x_1 - x_0| > \varepsilon = 10^{-3}$ donc on passe à l'itération suivante.

Deuxième itération: $x_2 = x_1 - \dfrac{F(x_1)}{F'(x_1)} \approx 0.3887$ et $|x_2 - x_1| > \varepsilon = 10^{-3}$ donc on passe à l'itération suivante.

Troisième itération: $x_3 = x_2 - \dfrac{F(x_2)}{F'(x_2)} \approx 0.3886$ et $|x_3 - x_2| < \varepsilon = 10^{-3}$ donc on arrête les itérations.

La racine approchée de la racine exacte s est x_3 d'où $s \approx x_3 \pm \varepsilon$ c'est à dire $s \approx 0.388 \pm 0.001$.

(5) La méthode de Newton est plus avantageuse que la méthode du point fixe car elle nécessite moins d'itérations.

Chapitre 4

Interpolation polynomiale

4.1 Position du problème

Soit une fonction $y = f(x)$ connue uniquement aux points x_0, x_1,..., x_n. On désire connaître y pour x différent de x_i $(i = 0,...,n)$.
— Si $x \in [x_0, x_n]$, on dit qu'on doit faire une interpolation.
— Si $x \notin [x_0, x_n]$, on dit qu'on doit faire une extrapolation.
L'interpolation de f est une méthode d'approximation qui consiste à construire une fonction ϕ proche de f dans un intervalle $[a,b]$ qui contient les points x_i $(i = 0,...,n)$. Cette fonction ϕ doit vérifier la relation

$$\phi(x_i) = f(x_i) \qquad i = 0,...,n \tag{4.1}$$

Les points x_i $(i = 0,...,n)$ sont appelés points d'interpolation.

Définition 4.1. Toute fonction vérifiant la relation (4.1) est appelée fonction d'interpolation de f aux points x_i $(i = 0,...,n)$.

Exemple de problème d'interpolation

Un certain pays recense sa population tous les 10 ans depuis 1930 jusqu'en 2000.

année	1930	1940	1950	1960	1970	1980	1990	2000
population en millions	3.5	6	10.3	14	19.5	24	30	37

Question:

En analysant ces données, peut-on savoir quelle était la population en 1975 ou quelle sera la population en 2020.
Pour $x \in [a,b]$ un point différent des x_i $(i = 0,...,n)$, peut-on avoir une idée de la valeur de f en ce point x?
Considérons alors une fonction ϕ dépendant de plusieurs paramètres et telle que $\phi(x_i) = f(x_i)$ $i = 0,...,n$. On peut aussi écrire $f = \phi + \epsilon$ ou bien $f(x_i) = \phi(x_i) + \epsilon(x_i)$ avec $\epsilon(x_i) = 0$ pour $i = 0,...,n$. La fonction ϵ est alors appelée erreur d'interpolation donc l'interpolation consiste à remplacer $f(x)$ par $\phi(x)$ en commettant une erreur d'interpolation.

Le choix de la fonction ϕ va être dicté par les considérations suivantes:

1. Les paramètres dont elle dépend doivent être simples à calculer, en général la dépendance sera linéaire.
2. Simplicité du calcul de ϕ.
3. L'erreur d'interpolation $\epsilon(x)$ doit être la plus petite possible.

Remarque 4.1. En pratique la fonction ϕ est une combinaison linéaire de $(n+1)$ fonctions de base $(1,x,x^2,...,x^n)$ ou de $(1,e^x,e^{2x},...,e^{nx})$ ou de $(1,sinx,sin2x,...,sinnx)$...etc. Il peut arriver aussi que ϕ soit une fraction rationnelle.

4.2 Interpolation polynomiale

Soit f une fonction donnée aux points x_i $(i = 0,...,n)$ où les x_i sont des points distincts de $[a,b]$ c'est à dire $x_i \neq x_j$ pour $i \neq j$.
On cherche à déterminer un polynôme P_n de degré inférieur ou égal à n tel que $P_n(x_i) = f(x_i)$ pour $i = 0,...,n$ et

$$P_n(x) = \sum_{k=0}^{k=n} a_k x^k \tag{4.2}$$

Le polynôme P_n ainsi défini est appelé polynôme d'interpolation de f.

Théorème 4.1. *Interpolation polynomiale*
Il existe un unique polynôme de degré inférieur ou égal à n qui interpole la fonction f aux points x_i ($i = 0,...,n$) distincts deux à deux dans $[a,b]$.

Démonstration.
Soit $\mathbb{R}_n[x]$ l'espace des polynômes de degré inférieur ou égal à n qui est un espace vectoriel de dimension finie $(n+1)$. On considère $\varphi_0(x) = 1$, $\varphi_1(x) = x$, $\varphi_2(x) = x^2$, ..., $\varphi_n(x) = x^n$ la base canonique de $\mathbb{R}_n[x]$. On cherche à déterminer un polynôme P_n tel que $d^\circ P_n \leq n$ et $P_n(x_i) = f(x_i)$ pour $i = 0,...,n$ d'où $\sum_{k=0}^{k=n} a_k x_i^k = f(x_i)$ $(i = 0,...,n) \Leftrightarrow$

$$(\mathrm{S}')\begin{cases} a_0 + a_1x_0 + a_2x_0^2 + \cdots + a_nx_0^n = f(x_0) \\ \qquad\qquad\qquad \vdots \\ a_0 + a_1x_i + a_2x_i^2 + \cdots + a_nx_i^n = f(x_i) \\ \qquad\qquad\qquad \vdots \\ a_0 + a_1x_n + a_2x_n^2 + \cdots + a_nx_n^n = f(x_n) \end{cases}$$

Pour déterminer les coefficients a_k, il suffit de résoudre le système linéaire (S') de $(n+1)$ équations à $(n+1)$ inconnues. On note Δ son déterminant tel que

$$\Delta = \begin{vmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & x_i & x_i^2 & \cdots & x_i^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{vmatrix}$$

Ce déterminant connu sous le nom du déterminant de Vandermonde est égal à
$\Delta = \prod_{i>j}(x_i - x_j) = \prod_{j=0}^{n-1}\prod_{i=j+1}^{n}(x_i - x_j)$ avec $0 \leq i,j \leq n$.
Comme $x_i \neq x_j$ pour $i \neq j$ donc $\Delta \neq 0$ et le système (S') admet une solution unique et donc le polynôme P_n est unique. □

Remarque 4.2. Le théorème (4.1) montre que par $(n+1)$ points x_i $(i = 0,1,...,n)$, il ne peut passer qu'un seul polynôme de degré inférieur ou égal à n et donc par deux points, on ne peut faire passer qu'une seule droite $y = ax + b$ et par trois points, on ne peut faire passer qu'une seule parabole $y = ax^2 + bx + c$.

4.2.1 Méthode directe pour la recherche du polynôme P_n

Quand n est petit, une manière simple de déterminer le polynôme d'interpolation P_n, de degré inférieur ou égal à n est d'écrire ce polynôme dans la base canonique

de $\mathbb{R}_n[x]$ sous la forme $P_n(x) = \sum_{k=0}^{k=n} a_k x^k$ puis d'appliquer la relation $P_n(x_i) = f(x_i)$ pour $i = 0,...,n$ avec les x_i distincts dans $[a,b]$. On doit résoudre un système de $(n+1)$ équations à $(n+1)$ inconnues pour déterminer les inconnues a_i, i=0,...,n.

Exemple d'application

Trouver le polynôme $P_2(x)$ tel que $P_2(0) = -2$, $P_2(-1) = -2$ et $P_2(1/2) = -5/4$.

Solution

On a $P_2(x) = \sum_{k=0}^{2} a_k x^k = a_0 + a_1 x + a_2 x^2$ et $P_2(x_i) = \sum_{k=0}^{2} a_k x_i^k$.
Cherchons les coefficients a_k $(k = 0,1,2)$ en résolvant le système suivant:

$$\begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 1 \\ 1 & 1/2 & 1/4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ -5/4 \end{bmatrix}$$

La résolution de ce système donne $a_0 = -2$, $a_1 = a_2 = 1$ et le polynôme d'interpolation s'écrit $P_2(x) = x^2 + x - 2$.

Remarque 4.3. Pour déterminer le polynôme d'interpolation P_n, on doit résoudre un système de $(n + 1)$ équations à $(n + 1)$ inconnues, ce qui est trop compliqué quand n est grand. Pour éviter ce problème, on va utiliser des polynômes particuliers qui sont les polynômes de Lagrange et les polynômes de Newton.

4.3 Polynôme d'interpolation de Lagrange

Définition 4.2. On appelle polynômes de Lagrange que l'on note $L_i(x)$, les polynômes de degré inférieur ou égal à n, définis par:

$$L_i(x) = \prod_{j=0, j\neq i}^{j=n} \frac{(x - x_j)}{(x_i - x_j)} \tag{4.3}$$

4.3.1 Détermination pratique des polynômes de Lagrange

Considérons le tableau carré suivant:

$$\begin{array}{ccccc} x-x_0 & x_0-x_1 & x_0-x_2 & \cdots & x_0-x_n \\ x_1-x_0 & x-x_1 & x_1-x_2 & \cdots & x_1-x_n \\ x_2-x_0 & x_2-x_1 & x-x_2 & \cdots & x_2-x_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n-x_0 & x_n-x_1 & x_n-x_2 & \cdots & x-x_n \end{array} \tag{4.4}$$

Le polynôme $L_i(x)$ est le quotient du produit des termes diagonaux du tableau sur le produit des termes de la $(i+1)^{ème}$ ligne du tableau.

Exemple d'application

Déterminer les cinq polynômes $L_i(x)$ attachés aux valeurs $x_0 = 0$, $x_1 = 4$, $x_2 = -4$, $x_3 = 2$ et $x_4 = 1$.

Solution

Dressons le tableau (4.4) pour $n = 4$.

$$\begin{array}{ccccc} x-x_0 & x_0-x_1 & x_0-x_2 & x_0-x_3 & x_0-x_4 \\ x_1-x_0 & x-x_1 & x_1-x_2 & x_1-x_3 & x_1-x_4 \\ x_2-x_0 & x_2-x_1 & x-x_2 & x_2-x_3 & x_2-x_4 \\ x_3-x_0 & x_3-x_1 & x_3-x_2 & x-x_3 & x_3-x_4 \\ x_4-x_0 & x_4-x_1 & x_4-x_2 & x_4-x_3 & x-x_4 \end{array} = \begin{array}{ccccc} x & -4 & 4 & -2 & -1 \\ 4 & x-4 & 8 & 2 & 3 \\ -4 & -8 & x+4 & -6 & -5 \\ 2 & -2 & 6 & x-2 & 1 \\ 1 & -3 & 5 & -1 & x-1 \end{array}$$

D'où $L_0(x) = \frac{\text{Produit des termes diagonaux}}{\text{Produit des termes de la } 1^{ère} \text{ ligne}} = \dfrac{x(x-4)(x+4)(x-2)(x-1)}{x(-4)(4)(-2)(-1)}$

$$L_1(x) = \frac{\text{Produit des termes diagonaux}}{\text{Produit des termes de la } 2^{ème} \text{ ligne}} = \frac{x(x-4)(x+4)(x-2)(x-1)}{4(x-4)(8)(2)(3)}$$

$$L_2(x) = \frac{\text{Produit des termes diagonaux}}{\text{Produit des termes de la } 3^{ème} \text{ ligne}} = \frac{x(x-4)(x+4)(x-2)(x-1)}{(-4)(-8)(x+4)(-6)(-5)}$$

$$L_3(x) = \frac{\text{Produit des termes diagonaux}}{\text{Produit des termes de la } 4^{ème} \text{ ligne}} = \frac{x(x-4)(x+4)(x-2)(x-1)}{(2)(-2)(6)(x-2)(1)}$$

$$L_4(x) = \frac{\text{Produit des termes diagonaux}}{\text{Produit des termes de la } 5^{ème} \text{ ligne}} = \frac{x(x-4)(x+4)(x-2)(x-1)}{(1)(-3)(5)(-1)(x-1)}$$

Remarque 4.4.

$$L_i(x_k) = \prod_{j=0, j\neq i}^{j=n} \frac{(x_k - x_j)}{(x_i - x_j)} = \begin{cases} 1 \text{ si } k = i \\ 0 \text{ si } k \neq i \end{cases} \tag{4.5}$$

$L_i(x_k)$ s'identifient au symbole de Kronecker.

Proposition 4.1.

Les polynômes $L_i(x)$ forment une base de l'espace $\mathcal{P}_n$.

Démonstration.

Comme dimension de $\mathcal{P}_n$ est finie et égale à $(n+1)$ donc il suffit de montrer que la famille $(L_0(x), L_1(x),..., L_n(x))$ est libre c'est à dire
$\forall \lambda_i \in \mathbb{R}$, $i = 0,...,n$ si $\sum_{i=0}^{n} \lambda_i L_i(x) = 0$, $\forall x$ alors $\lambda_i = 0$ $\forall i = 0,...,n$. En particulier cette somme est nulle pour $x = x_k$ donc $\sum_{i=0}^{n} \lambda_i L_i(x_k) = 0$ et d'après la relation (4.5) on a $\sum_{i=0}^{n} \lambda_i L_i(x_k) = \lambda_k = 0$, $\forall k = 0,...,n$ donc la famille $(L_0(x), L_1(x),..., L_n(x))$ est une base de $\mathcal{P}_n$. □

Pour que le polynôme P_n soit un polynôme d'interpolation de la fonction f aux points x_k, $k = 0,...,n$ il faut et il suffit qu'il vérifie la relation (4.1) c'est à dire $P_n(x_k) = f(x_k)$ $k = 0,...,n$ d'où $\sum_{i=0}^{n} \lambda_i L_i(x_k) = f(x_k)$ $k = 0,...,n$ et d'après la relation (4.5) on obtient $\lambda_k = f(x_k)$ pour $k = 0,...,n$. Donc la formule du polynôme devient $P_n(x) = \sum_{i=0}^{n} f(x_i) L_i(x)$.

Définition 4.3. Le polynôme d'interpolation mis sous la forme

$$P_n(x) = \sum_{i=0}^{n} f(x_i) L_i(x) \tag{4.6}$$

est appelé polynôme d'interpolation de Lagrange aux points x_i, $i = 0,...,n$.

Exemple d'application où la fonction f est inconnue

Soit f une fonction donnée aux points $x_0 = 0$ avec $f(x_0) = -1$ et $x_1 = 1$ avec $f(x_1) = 2$. Écrire le polynôme d'interpolation de Lagrange de la fonction f aux points x_1 et x_2 et en déduire la valeur de $f(\frac{1}{2})$.

Solution

D'après la relation (4.6), le polynôme d'interpolation de Lagrange est donné par $P_1(x) = \sum_{i=0}^{1} f(x_i)L_i(x) = f(x_0)L_0(x) + f(x_1)L_1(x) = -L_0(x) + 2L_1(x)$ et d'après la relation (4.3), nous avons $L_0(x) = \prod_{j=0,j\neq 0}^{j=1} \frac{(x-x_j)}{(x_0-x_j)} = \frac{(x-x_1)}{(x_0-x_1)} = \frac{x-1}{0-1} = -x+1$ et $L_1(x) = \prod_{j=0,j\neq 1}^{j=1} \frac{(x-x_j)}{(x_1-x_j)} = \frac{(x-x_0)}{(x_1-x_0)} = \frac{x-0}{1-0} = x$ et par conséquent
$P_1(x) = -L_0(x) + 2L_1(x) = x - 1 + 2x = 3x - 1$ donc le polynôme d'interpolation de Lagrange de f aux points x_i, i = 1,2 est donné par $P_1(x) = 3x - 1$ et la valeur approximative de $f(\frac{1}{2}) \simeq P_1(\frac{1}{2}) = \frac{1}{2}$.

Exemple d'application où la fonction f est connue

Construire le polynôme d'interpolation de Lagrange de la fonction $f(x) = \sin \pi x$ pour les points $x_0 = 0$, $x_1 = \frac{1}{6}$ et $x_2 = \frac{1}{2}$.

Solution

Calculons d'abord les valeurs correspondantes de la fonction aux points x_i, i = 0,1,2.
$f(x_0) = f(0) = 0$, $f(x_1) = f(\frac{1}{6}) = \frac{1}{2}$ et $f(x_2) = f(\frac{1}{2}) = 1$.
$P_2(x) = \sum_{i=0}^{2} f(x_i)L_i(x) = 0.\frac{(x-\frac{1}{6})(x-\frac{1}{2})}{(-\frac{1}{6})(-\frac{1}{2})} + \frac{1}{2}\frac{x(x-\frac{1}{2})}{\frac{1}{6}(\frac{1}{6}-\frac{1}{2})} + 1.\frac{x(x-\frac{1}{6})}{\frac{1}{2}(\frac{1}{2}-\frac{1}{6})} = -3x^2 + \frac{7}{2}x$

4.3.2 Erreur d'interpolation de la formule de Lagrange

Théorème 4.2. *Soit f une fonction de classe $\mathcal{C}^{n+1}([a,b])$ alors il existe θ_x appartenant au plus petit intervalle contenant les points x_i, i = 0,...,n tel que:*

$$\epsilon(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\theta_x)}{(n+1)!} \prod_{j=0}^{j=n} (x - x_j) \tag{4.7}$$

Remarque 4.5. La formule (4.7) ne permet évidemment pas de calculer la valeur exacte de l'erreur parce que, en général, θ_x est inconnu. Elle permet par contre, d'en calculer une majoration.

Remarque 4.6. En posant $M = \max_{a\leq x\leq b} \left|f^{(n+1)}(x)\right|$, la majoration de l'erreur sera

$$|\epsilon(x)| \leq \frac{M}{(n+1)!} \left| \prod_{j=0}^{j=n} (x - x_j) \right| \tag{4.8}$$

Démonstration. du théorème (4.2)
— Cas 1: Si $x = x_i$, on a $\epsilon(x_i) = f(x_i) - P_n(x_i) = 0\ \forall i = 0,...,n$ par définition du polynôme d'interpolation. Or $\prod_{j=0}^{j=n}(x - x_j) = 0$ pour $x = x_i$, $i = 0,...,n$ d'où
$\frac{f^{(n+1)}(\theta_x)}{(n+1)!}\prod_{j=0}^{j=n}(x - x_j) = 0$ et on déduit que $\epsilon(x_i) = \frac{f^{(n+1)}(\theta_x)}{(n+1)!}\prod_{j=0}^{j=n}(x - x_j) = 0.$

— Cas 2: Si $x \neq x_i$, considérons la fonction $F : [a,b] \mapsto \mathbb{R}$ définie par
$F(t) = f(t) - P_n(t) - \frac{f(x) - P_n(x)}{\Pi_n(x)}\Pi_n(t)$ où $\Pi_n(x) = \prod_{j=0}^{j=n}(x - x_j)$.
On a d'abord $F(x) = 0$ et comme P_n est un polynôme d'interpolation, on a aussi
$F(x_i) = f(x_i) - P_n(x_i) - \frac{f(x) - P_n(x)}{\Pi_n(x)}\Pi_n(x_i) = 0$, $i = 0,...,n$. Donc $F(x) = 0$ et $F(x_i) = 0$ pour $i = 0,...,n$ et on conclut que la fonction F admet au moins (n+2) racines et on déduit que F' admet au moins $(n + 1)$ racines puis F'' admet au moins n racines puis $...F^{(n+1)}$ admet au moins une racine qu'on notera θ_x d'où $F^{(n+1)}(\theta_x) = 0$.
Calculons à présent la dérivée d'ordre (n+1) de la fonction F, on obtient:
$F^{(n+1)}(t) = f^{(n+1)}(t) - P_n^{(n+1)}(t) - \frac{f(x) - P_n(x)}{\Pi_n(x)}\Pi_n^{(n+1)}(t).$
Comme $P_n \in \mathcal{P}_n$ donc $P_n^{(n+1)}(t) = 0$ et $\Pi_n(t)$ est un polynôme de degré inférieur ou égal à n+1 qui s'écrit sous la forme
$\Pi_n(t) = \prod_{j=0}^{j=n}(t - t_j) = [t^{n+1} + Q_n(t)]$ et sa dérivée d'ordre (n+1) s'écrit
$\Pi_n^{(n+1)}(t) = \left(\prod_{j=0}^{j=n}(t - t_j)\right)^{(n+1)} = [t^{n+1} + Q_n(t)]^{(n+1)} = (t^{n+1})^{(n+1)} + (Q_n(t))^{(n+1)} =$
$(n + 1)! + 0 = (n + 1)!$ où $Q_n(t) \in \mathcal{P}_n$
Remplaçons cette expression dans $F^{(n+1)}(t)$ et on obtient
$F^{(n+1)}(t) = f^{(n+1)}(t) - \frac{f(x) - P_n(x)}{\Pi_n(x)}.(n + 1)!$, or $F^{(n+1)}(\theta_x) = 0$ alors
$F^{(n+1)}(\theta_x) = f^{(n+1)}(\theta_x) - \frac{f(x) - P_n(x)}{\Pi_n(x)}.(n + 1)! = 0$ ce qui donne
$\epsilon(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\theta_x)}{(n+1)!}\Pi_n(x) = \frac{f^{(n+1)}(\theta_x)}{(n+1)!}\prod_{j=0}^{j=n}(x - x_j).$ □

Exemple d'application

Avec quelle précision peut-on calculer $\sqrt{115}$ à l'aide de la formule de Lagrange pour la fonction $y = \sqrt{x}$ si l'on prend les points d'interpolation $x_0 = 100$, $x_1 = 121$ et $x_2 = 144$?

Solution

Dans la formule (4.8) de la majoration de l'erreur, on remplace $x = 115$, $x_0 = 100$, $x_1 = 121$, $x_2 = 144$ et $n = 2$, on obtient $|\epsilon(115)| \leq \frac{M}{3!} |(115-100)(115-121)(115-144)|$.
Il reste juste à déterminer la constante $M = \max_{100 \leq x \leq 144} \left|y^{(3)}(x)\right|$, pour cela calculons la dérivée $y^{(3)}(x)$. Comme $y = \sqrt{x} \Rightarrow y' = \frac{1}{2}x^{-\frac{1}{2}} \Rightarrow y'' = -\frac{1}{4}x^{-\frac{3}{2}} \Rightarrow y^{(3)}(x) = \frac{3}{8}x^{-\frac{5}{2}}$.
Comme $y^{(3)}(x)$ est strictement décroissante sur [100,144] donc son maximum est atteint en $x = 100$ et $M = \left|y^{(3)}(100)\right| = \frac{3}{8}100^{-\frac{5}{2}} \simeq 0.37.10^{-5}$.
Conclusion $|\epsilon(115)| \leq \frac{0.37.10^{-5}}{3!} |(115-100)(115-121)(115-144)| \simeq 1.6.10^{-3}$
d'où la précision est $|\epsilon(115)| \simeq 1.6.10^{-3}$.

4.3.3 Cas particulier où les points d'interpolation sont équidistants

Les points sont équidistants si $x_{i+1} - x_i = h$ pour tout $i = 0,...,n-1$ et on a aussi $x_n = x_0 + n.h$. Dans ce cas, on effectue le changement de variables
$x = x_0 + h.t$ dans la formule (4.3) et on obtient comme polynômes de Lagrange
$$\mathrm{L}_i(x) = \prod_{j=0, j\neq i}^{j=n} \frac{(x - x_j)}{(x_i - x_j)} = L_i(x_0 + h.t) = \prod_{j=0, j\neq i}^{j=n} \frac{(x_0 + h.t - x_0 - j.h)}{(x_0 + i.h - x_0 - j.h)} = \prod_{j=0, j\neq i}^{j=n} \frac{t-j}{i-j}$$

Remarque 4.7. Dans le cas où les points d'interpolation x_i, $i = 0,...,n$ sont équidistants, les polynômes de lagrange deviennent:

$$\mathcal{L}_i(t) = \prod_{j=0, j\neq i}^{j=n} \frac{t-j}{i-j} \tag{4.9}$$

Et le polynôme d'interpolation de Lagrange devient alors un polynôme en t sous la forme:

$$\mathcal{P}_n(t) = \sum_{i=0}^{n} f(x_i)\mathcal{L}_i(t) \tag{4.10}$$

Exemple d'application

Écrire le polynôme de Lagrange de la fonction f définie par

x_i	2	4	6
$f(x_i)$	1	-1	-2

Solution

Les points x_i, $i = 0,1,2$ étant équidistants avec $h = 2$, posons alors le changement de variables $x = x_0 + h.t = 2 + 2t$ d'où $t = \frac{x-2}{2}$. D'après la formule (4.10) on a

$\mathcal{P}_2(t) = \sum_{i=0}^{2} f(x_i)\mathcal{L}_i(t) = f(x_0)\mathcal{L}_0(t) + f(x_1)\mathcal{L}_1(t) + f(x_2)\mathcal{L}_2(t) = 1.\mathcal{L}_0(t) - 1.\mathcal{L}_1(t) - 2\mathcal{L}_2(t)$
En utilisant la formule (4.9), déterminons les polynômes de Lagrange $\mathcal{L}_i(t)$, $i = 0,1,2$.
$\mathcal{L}_0(t) = \prod_{j=0,j\neq 0}^{j=2} \frac{t-j}{0-j} = \frac{(t-1)(t-2)}{(-1)(-2)} = \frac{1}{2}(t^2 - 3t + 2)$
$\mathcal{L}_1(t) = \prod_{j=0,j\neq 1}^{j=2} \frac{t-j}{1-j} = \frac{(t)(t-2)}{(1)(-1)} = -t^2 + 2t$
$\mathcal{L}_2(t) = \prod_{j=0,j\neq 2}^{j=2} \frac{t-j}{2-j} = \frac{(t)(t-1)}{(2)(1)} = \frac{t^2}{2} - \frac{t}{2}$
Donc $\mathcal{P}_2(t) = \mathcal{L}_0(t) - \mathcal{L}_1(t) - 2\mathcal{L}_2(t) = \frac{t^2}{2} - \frac{5t}{2} + 1$
Revenons à la variable x en remplaçant $t = \frac{x-2}{2}$ dans l'expression de $\mathcal{P}_2(t)$ et le polynôme d'interpolation de Lagrange de f aux points x_i, $i = 0,1,2$ est donné par $\mathcal{P}_2(t) = P_2(x) = \frac{x^2 - 14x + 32}{8}$.

4.4 Polynôme d'interpolation de Newton

4.4.1 Différences divisées

Définition 4.4. Différences divisées
On appelle différence divisée d'ordre 1,2,...,k aux points x_i, $i = 0,...,n$ les quantités définies par:

$$\delta f[x_i] = f(x_i) \quad \text{et} \quad \delta f[x_i,x_j] = \frac{f(x_j) - f(x_i)}{x_j - x_i}, \quad i \neq j \tag{4.11}$$

$$\delta f[x_l,x_i,x_j] = \frac{\delta f[x_l,x_j] - \delta f[x_l,x_i]}{x_j - x_i}, \quad i \neq j \neq l \tag{4.12}$$

$$\delta f[x_0,x_1,...,x_k] = \frac{\delta f[x_1,x_2,...,x_k] - \delta f[x_0,x_1,...,x_{k-1}]}{x_k - x_0} \tag{4.13}$$

Remarque 4.8.
Les différences divisées sont indépendantes de l'ordre des points x_i c'est à dire
$\delta f[x_{p(1)},x_{p(2)},...,x_{p(n)}] = \delta f[x_1,x_2,...,x_n]$ où $p(i)$ est une permutation des indices 1,2,...,n.

Exemple

$\delta f[x_0,x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$

$$\delta f[x_0,x_1,x_2] = \frac{\delta f[x_0,x_2] - \delta f[x_0,x_1]}{x_2 - x_1} = \frac{\dfrac{f(x_2)-f(x_0)}{x_2-x_0} - \dfrac{f(x_1)-f(x_0)}{x_1-x_0}}{x_2-x_1}$$

Définition 4.5. Le polynôme d'interpolation de la fonction f aux points x_i, $i = 0,...,n$ mis sous la forme

$$P_n(x) = f(x_0) + \sum_{i=1}^{i=n} \delta f[x_0,x_1,...,x_i].\prod_{j=0}^{i-1}(x - x_j) \tag{4.14}$$

est appelé polynôme d'interpolation de Newton.
La formule (4.14) développée s'écrit:

$$\begin{aligned} P_n(x) = f(x_0) &+ (x - x_0)\delta f[x_0,x_1] + (x - x_0)(x - x_1)\delta f[x_0,x_1,x_2] + ... \\ &+(x - x_0)(x - x_1)...(x - x_{n-1})\delta f[x_0,x_1,...,x_n]. \end{aligned}$$

Exemple d'application

Trouver le polynôme d'interpolation de Newton pour la fonction f définie par

x_i	0	1	2
$f(x_i)$	-1	0	2

Solution

Appliquons la formule (4.14) pour $n = 2$,
$P_2(x) = f(x_0) + (x - x_0)\delta f[x_0,x_1] + (x - x_0)(x - x_1)\delta f[x_0,x_1,x_2]$ avec
$\delta f[x_0,x_1] = \dfrac{f(x_1) - f(x_0)}{x_1 - x_0} = 1$ et

$$\delta f[x_0,x_1,x_2] = \frac{\delta f[x_0,x_2] - \delta f[x_0,x_1]}{x_2 - x_1} = \frac{\dfrac{f(x_2)-f(x_0)}{x_2-x_0} - \dfrac{f(x_1)-f(x_0)}{x_1-x_0}}{x_2-x_1} = \frac{1}{2}$$

Donc $P_2(x) = -1 + x + \dfrac{1}{2}x(x-1) = \dfrac{1}{2}x^2 + \dfrac{1}{2}x - 1.$

4.4.2 Formule explicite des différences divisées

Le polynôme d'interpolation d'une fonction f aux points x_i, $i = 0,...,n$ étant unique alors le polynôme d'interpolation de Lagrange est égal au polynôme d'interpolation de Newton c'est à dire $P_n(x) = \sum_{i=0}^{n} f(x_i)L_i(x) = f(x_0) + \sum_{i=1}^{i=n} \delta f[x_0,x_1,...,x_i].\prod_{j=0}^{i-1}(x - x_j)$.

Identifions les coefficients du plus haut degré en n des deux côtés de cette égalité,
$\sum_{i=0}^{n} f(x_i)L_i(x) = \sum_{i=0}^{n} f(x_i).\prod_{j=0,j\neq i}^{j=n}\frac{(x-x_j)}{(x_i-x_j)} = \sum_{i=0}^{n}\frac{f(x_i)}{\prod_{j=0,j\neq i}^{j=n}(x_i-x_j)}.x^n + Q_{n-1}(x)$
Et $f(x_0) + \sum_{i=1}^{i=n}\delta f[x_0,x_1,...,x_i].\prod_{j=0}^{i-1}(x-x_j) = \delta f[x_0,x_1,...,x_n].x^n + P_{n-1}(x)$.
où Q_{n-1} et P_{n-1} sont deux polynômes de degré inférieur ou égal à $(n-1)$.
En égalisant les coefficients du plus haut degré en n on obtient l'égalité suivante:

$$\delta f[x_0,x_1,...,x_n] = \sum_{i=0}^{n}\frac{f(x_i)}{\prod_{j=0,j\neq i}^{j=n}(x_i-x_j)} \tag{4.15}$$

4.4.3 Erreur d'interpolation de la formule de Newton

Déterminons l'erreur $\epsilon(x) = f(x) - P_n(x)$, pour cela on tire $f(x)$ de la relation
$\delta f[x_0,x] = \frac{f(x)-f(x_0)}{x-x_0} \Rightarrow f(x) = f(x_0)+(x-x_0)\delta f[x_0,x]$.....relation $(*)$
On a aussi $\delta f[x_0,x_1,x] = \frac{\delta f[x_0,x]-\delta f[x_0,x_1]}{x-x_1} \Rightarrow \delta f[x_0,x] = \delta f[x_0,x_1]+(x-x_1)\delta f[x_0,x_1,x]$
On remplace $\delta f[x_0,x]$ dans la relation $(*)$ et on obtient
$f(x) = f(x_0) + (x-x_0)\delta f[x_0,x_1] + (x-x_0)(x-x_1)\delta f[x_0,x_1,x]$ puis on continue le raisonnement jusqu'à l'ordre n pour obtenir
$f(x) = f(x_0) + (x-x_0)\delta f[x_0,x_1] + ... + (x-x_0)(x-x_1)...(x-x_{n-1})\delta f[x_0,...,x_n] +$
$(x-x_0)(x-x_1)...(x-x_n)\delta f[x_0,x_1,...,x_n,x]$.
D'où l'expression de l'erreur de Newton

$$\epsilon(x) = f(x) - P_n(x) = \prod_{j=0}^{n}(x-x_j)\delta f[x_0,x_1,...,x_n,x] \tag{4.16}$$

Proposition 4.2.
Si f est une fonction de classe $\mathcal{C}^{n+1}[a,b]$ alors il existe θ_x appartenant au plus petit intervalle contenant les points x_i, $i = 0,...,n$ tel que:

$$\delta f[x_0,x_1,...,x_n,x] = \frac{f^{(n+1)}(\theta_x)}{(n+1)!} \tag{4.17}$$

Démonstration.
Puisque le polynôme d'interpolation de f aux points x_i, $i = 0,...,n$ est unique alors l'erreur (4.7) de Lagrange et l'erreur (4.16) de Newton est la même donc

$$\epsilon(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\theta_x)}{(n+1)!} \prod_{j=0}^{j=n} (x - x_j) = \delta f[x_0,x_1,...,x_n,x]. \prod_{j=0}^{n} (x - x_j).$$

En simplifiant par l'expression $\prod_{j=0}^{n} (x - x_j)$ des deux côtés, on obtient donc

$$\delta f[x_0,x_1,...,x_n,x] = \frac{f^{(n+1)}(\theta_x)}{(n+1)!}.$$

4.4.4 Détermination pratique des différences divisées

Exemple

On veut déterminer le polynôme de degré inférieur ou égal à 2 qui interpole la fonction f aux points x_i, $i = 0,1,2$ avec la méthode de Newton tel que

x_i	1	2	3
$f(x_i)$	-1	1	0

Solution

La relation (4.14) pour $n = 2$ donne

$P_2(x) = \underbrace{f(x_0)} + (x - x_0)\,\underbrace{\delta f[x_0,x_1]} + (x - x_0)(x - x_1)\,\underbrace{\delta f[x_0,x_1,x_2]}$

Dressons le tableau suivant pour avoir directement les valeurs de $\delta f[x_0,x_1]$ et de $\delta f[x_0,x_1,x_2]$.

x_i	$f(x_i)$	δ_1	δ_2
$x_0 = 1$	$\underbrace{f(x_0)} = -1$		
		$\underbrace{\delta f[x_0,x_1]} = \frac{f(x_1)-f(x_0)}{x_1-x_0} = \frac{1+1}{2-1} = 2$	
$x_1 = 2$	$f(x_1) = 1$		$\underbrace{\delta f[x_0,x_1,x_2]} = \frac{\delta f[x_1,x_2]-\delta f[x_0,x_1]}{x_2-x_0} = -\frac{3}{2}$
		$\delta f[x_1,x_2] = \frac{f(x_2)-f(x_1)}{x_2-x_1} = \frac{0-1}{3-2} = -1$	
$x_2 = 3$	$f(x_2) = 0$		

Donc $P_2(x)$ devient $P_2(x) = (-1) + (x - 1)(2) + (x - 1)(x - 2)(-\frac{3}{2}) = -\frac{3}{2}x^2 + \frac{13}{2}x - 6$.

Tableau des différences divisées à l'ordre n

Pour déterminer les différences divisées d'ordre n de la fonction f aux points x_i, $i = 0,...,n$, il suffit de dresser le tableau suivant et d'appliquer les relations (4.11) et (4.13).

x_i	$f(x_i)$	δ_1	δ_2	$\cdots$	δ_{n-1}	δ_n
x_0	$\underbrace{f(x_0)}$					
		$\underbrace{\delta f[x_0,x_1]}$				
x_1	$f(x_1)$		$\underbrace{\delta f[x_0,x_1,x_2]}$			
		$\delta f[x_1,x_2]$				
x_2	$f(x_2)$		$\delta f[x_1,x_2,x_3]$			
		$\delta f[x_2,x_3]$				
x_3	$f(x_3)$					
$\vdots$	$\vdots$				$\underbrace{\delta f[x_0,x_1,\ldots,x_{n-1}]}$	
						$\underbrace{\delta f[x_0,x_1,\ldots,x_n]}$
$\vdots$	$\vdots$				$\delta f[x_1,x_2,\ldots,x_n]$	
x_{n-1}	$f(x_{n-1})$		$\delta f[x_{n-2},x_{n-1},x_n]$			
		$\delta f[x_{n-1},x_n]$				
x_n	$f(x_n)$					

Exemple d'application

Déterminer, par trois manières différentes, le polynôme d'interpolation de la fonction f

telle que

x_i	$x_0=-2$	$x_1=0$	$x_2=1$	$x_3=5$	$x_4=6$
$f(x_i)$	$f(x_0)=3$	$f(x_1)=-2$	$f(x_2)=5$	$f(x_3)=1$	$f(x_4)=7$

Solution

Cas 1: Polynôme d'interpolation de Newton

Formons le tableau des différences divisées d'ordre 4 suivant:

x_i	$f(x_i)$	δ_1	δ_2	δ_3	δ_4
x_0	$\underbrace{f(-2)}=3$				
		$\underbrace{\delta f[x_0,x_1]}=-\frac{5}{2}$			
x_1	$f(0)=-2$		$\underbrace{\delta f[x_0,x_1,x_2]}=\frac{19}{6}$		
		$\delta f[x_1,x_2]=7$		$\underbrace{\delta f[x_0,\ldots,x_3]}=-\frac{143}{210}$	
x_2	$f(1)=5$		$\delta f[x_1,x_2,x_3]=-\frac{8}{5}$		$\underbrace{\delta f[x_0,\ldots,x_4]}=\frac{31}{210}$
		$\delta f[x_2,x_3]=-1$		$\delta f[x_1,\ldots,x_4]=\frac{1}{2}$	
x_3	$f(5)=1$		$\delta f[x_2,x_3,x_4]=\frac{7}{5}$		
		$\delta f[x_3,x_4]=6$			
x_4	$f(6)=7$				

$P_4(x) = 3 - \frac{5}{2}(x+2) + \frac{19}{6}(x+2)x - \frac{143}{210}(x+2)x(x-1) + \frac{31}{210}(x+2)x(x-1)(x-5) = P_{Newton}(x)$

Cas 2: Polynôme d'interpolation de Lagrange

Appliquons les relations (4.3), (4.6) et $n = 4$ pour avoir le polynôme d'interpolation de Lagrange, on a $P_4(x) = \sum_{i=0}^{4} f(x_i).L_i(x) = \sum_{i=0}^{4} f(x_i). \prod_{j=0, j\neq i}^{j=4} \frac{(x - x_j)}{(x_i - x_j)}$

$$P_4(x) = 3\frac{x(x-1)(x-5)(x-6)}{336} + 2\frac{(x+2)(x-1)(x-5)(x-6)}{60} + 5\frac{(x+2)x(x-5)(x-6)}{60} - \frac{(x+2)x(x-1)(x-6)}{140} + 7\frac{(x+2)x(x-1)(x-5)}{240} = P_{Lagrange}(x)$$

Cas 3: En utilisant la méthode directe (4.2.1)

Les relations (4.1) et (4.2) donnent le système suivant $\sum_{k=0}^{4} a_k x_i^k = f(x_i)$ pour $i = 0,...,4$.

$$\begin{bmatrix} 1 & -2 & 4 & -8 & 16 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 5 & 25 & 125 & 625 \\ 1 & 6 & 36 & 216 & 1296 \end{bmatrix} . \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ 5 \\ 1 \\ 7 \end{bmatrix}$$

Comme les points x_i, $i = 0,...,4$ sont distincts donc ce système admet une solution unique qui est $a_0 = -2$, $a_1 = \frac{1401}{210}$, $a_2 = \frac{305}{210}$, $a_3 = \frac{-267}{210}$ et $a_4 = \frac{31}{210}$. Donc d'après la relation (4.2), la forme classique du polynôme d'interpolation de f est $P_4(x) = \sum_{k=0}^{4} a_k x^k$ c'est à dire

$P_4(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 = -2 + \frac{1401}{210}x + \frac{305}{210}x^2 - \frac{267}{210}x^3 + \frac{31}{210}x^4 = P_{méthode\ directe}(x)$

D'après l'unicité du polynôme d'interpolation, il est clair que

$P_{Newton}(x) = P_{Lagrange}(x) = P_{méthode\ directe}(x)$

4.5 Différences finies ascendantes

Dans le cas où les points d'interpolation x_i, $i = 0,...,n$ sont équidistants c'est à dire $x_k = x_0 + k.h$, $k = 0,...,n$, on préfère souvent utiliser les différences finies que les différences divisées.

Définition 4.6.
La différence $f(x_{i+1}) - f(x_i) = y_{i+1} - y_i$ est appelée différence finie ascendante (ou progressive) du 1^{er} ordre. On la note

$$\Delta f(x_i) = \Delta y_i = y_{i+1} - y_i \tag{4.18}$$

Définition 4.7.
La différence finie ascendante d'ordre n est définie par:

$$\Delta^n f(x_i) = \Delta^n y_i = \Delta(\Delta^{n-1} y_i) \quad n = 2,3,... \tag{4.19}$$

Remarque 4.9.
Le symbole "Δ", appelé opérateur différence finie ascendant, vérifie les propriétés suivantes:

1. $\Delta^0 y = y$ par convention,
2. $\Delta(y + z) = \Delta y + \Delta z$,
3. $\Delta(c^{cste} y) = c^{cste} \Delta y$,
4. $\Delta^m(\Delta^n y) = \Delta^{m+n} y$ où $y = f(x)$, $z = g(x)$, c^{cste} = constante, m,$n \in \mathbb{N}$

4.5.1 Tableau des différences finies ascendantes

Soit le tableau suivant, pour $n = 5$, par exemple

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$	$\Delta^5 y_i$
x_0	y_0	Δy_0	$\Delta^2 y_0$	$\Delta^3 y_0$	$\Delta^4 y_0$	$\Delta^5 y_0$
x_1	y_1	Δy_1	$\Delta^2 y_1$	$\Delta^3 y_1$	$\Delta^4 y_1$	$\nearrow$
x_2	y_2	Δy_2	$\Delta^2 y_2$	$\Delta^3 y_2$	$\nearrow$	
x_3	y_3	Δy_3	$\Delta^2 y_3$	$\nearrow$		
x_4	y_4	Δy_4	$\nearrow$			
x_5	y_5	$\nearrow$				

Les calculs se font sur la base, par exemple, de
$\Delta y_1 = y_2 - y_1$,
$\Delta^2 y_1 = \Delta(\Delta y_1) = \Delta(y_2 - y_1) = \Delta y_2 - \Delta y_1$,
$\Delta^3 y_1 = \Delta^2(\Delta y_1) = \Delta^2(y_2 - y_1) = \Delta^2 y_2 - \Delta^2 y_1$,
$\Delta^4 y_1 = \Delta^3(\Delta y_1) = \Delta^3(y_2 - y_1) = \Delta^3 y_2 - \Delta^3 y_1$,
$\Delta^5 y_0 = \Delta^4(\Delta y_0) = \Delta^4(y_1 - y_0) = \Delta^4 y_1 - \Delta^4 y_0$ et $y_i = f(x_i)$,...$i = 1$,...,5. D'où

y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$	$\Delta^5 y_i$
y_0	$\Delta y_0 = y_1 - y_0$	$\Delta^2 y_0 = \Delta y_1 - \Delta y_0$	$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0$	$\Delta^4 y_0 = \Delta^3 y_1 - \Delta^3 y_0$	$\Delta^5 y_0$
y_1	$\Delta y_1 = y_2 - y_1$	$\Delta^2 y_1 = \Delta y_2 - \Delta y_1$	$\Delta^3 y_1 = \Delta^2 y_2 - \Delta^2 y_1$	$\Delta^4 y_1 = \Delta^3 y_2 - \Delta^3 y_1$	
y_2	$\Delta y_2 = y_3 - y_2$	$\Delta^2 y_2 = \Delta y_3 - \Delta y_2$	$\Delta^3 y_2 = \Delta^2 y_3 - \Delta^2 y_2$		
y_3	$\Delta y_3 = y_4 - y_3$	$\Delta^2 y_3 = \Delta y_4 - \Delta y_3$			
y_4	$\Delta y_4 = y_5 - y_4$				
y_5					

Exemple d'application

Soit le tableau suivant

x_i	0	1	2	3	4	5
$f(x_i) = y_i$	-7	-3	6	25	62	129

Former le tableau des différences ascendantes de ces données.

Solution

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$	$\Delta^5 y_i$
0	-7	4	5	5	3	1
1	-3	9	10	8	4	↗
2	6	19	18	12	↗	
3	25	37	30	↗		
4	62	67	↗			
5	129	↗				

4.5.2 Polynôme d'interpolation ascendant de Newton

Définition 4.8.

Le polynôme d'interpolation de la fonction f aux points x_i, $i = 0,...,n$ mis sous la forme

$$P_n(x) = y_0 + \frac{q}{1!}\Delta y_0 + \frac{q(q-1)}{2!}\Delta^2 y_0 + ... + \frac{q(q-1)...(q-(n-1))}{n!}\Delta^n y_0 \qquad (4.20)$$

où $q = \dfrac{x - x_0}{h}$ et $h = x_i - x_{i-1}$, $i = 1,...,n$ est appelé "polynôme d'interpolation ascendant (ou progressif) de Newton" ou bien aussi "première formule d'interpolation de Newton".

Exemple d'application

En partant des données du tableau suivant

x_i	1	2	3	4	5	6	7
y_i	3	7	13	21	31	43	57

,

construire le polynôme d'interpolation progressif de Newton et calculer sa valeur pour $x = 1.5$.

Solution

Le tableau des différences ascendantes de ces données étant:

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$
1	$y_0 = 3$	$\Delta y_0 = 4$	$\Delta^2 y_0 = 2$	$\Delta^3 y_0 = 0$
2	7	6	2	0
3	13	8	2	0
4	21	10	2	0
5	31	12	2	
6	43	14		
7	57			

D'après la formule (4.20) pour $n = 2$, nous avons alors le polynôme d'interpolation progressif de Newton

$P_2(x) = y_0 + \frac{q}{1!}\Delta y_0 + \frac{q(q-1)}{2!}\Delta^2 y_0 + \frac{q(q-1)(q-2)}{3!}\Delta^3 y_0 = 3 + 4q + q(q-1) + 0 = q^2 + 3q + 3$

avec $q = \frac{x - x_0}{h} = x - 1$ et $h = 1$.

En remplaçant q dans $P_2(x)$, nous obtenons $P_2(x) = x^2 + x + 1$ et $P_2(1.5) = 4.75$.

Remarque 4.10.

Le polynôme d'interpolation ascendant de Newton est utilisé surtout lorsque la fonction $y = f(x)$ est interpolée dans un voisinage de x_0 (x voisin de x_0) où q est petit en valeur absolue.

4.5.3 Évaluation de l'erreur de la première formule de Newton

D'après la formule (4.7), on a $\epsilon(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\theta_x)}{(n+1)!} \prod_{j=0}^{j=n} (x - x_j)$ avec $\theta_x \in [x_0, x_n]$.

Comme les points x_i,...$i = 0,...,n$ sont équidistants donc $x_j = x_0 + j.h$ et $x = x_0 + q.h$ alors $\epsilon(x) = \frac{f^{(n+1)}(\theta_x)}{(n+1)!} \prod_{j=0}^{j=n} (x_0 + q.h - x_0 - j.h) = \frac{f^{(n+1)}(\theta_x)}{(n+1)!} \prod_{j=0}^{j=n} h(q - j)$ c'est à dire

$\epsilon(x) = \dfrac{h^{n+1}.f^{(n+1)}(\theta_x)}{(n+1)!}(q(q-1)(q-2)...(q-n))$.
En utilisant la remarque (4.6) et $M = \max\limits_{a\leq x\leq b}\left|f^{(n+1)}(x)\right|$, la majoration de cette erreur devient:

$$|\epsilon(x)| \leq \frac{M.h^{n+1}}{(n+1)!}\left|\prod_{j=0}^{j=n}(q-j)\right| \tag{4.21}$$

En supposant que les $\Delta^{n+1}y$ soient quasi-constantes pour la fonction $y = f(x)$, h suffisamment petit avec approximativement $f^{(n+1)}(\theta_x) \simeq \dfrac{\Delta^{n+1}y_0}{h^{n+1}}$ donc on peut écrire aussi:

$$|\epsilon(x)| \simeq \left|\frac{q(q-1)(q-2)...(q-n)}{(n+1)!}\Delta^{n+1}y_0\right| \tag{4.22}$$

Exemple d'application

Soient les données suivantes

x_i	0	1	2	3	4	5
y_i	-7	-3	6	25	62	129

Trouver $f(1.1)$ et évaluer le résultat obtenu.

Solution

Formons d'abord le tableau des différences ascendantes de ces données:

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$	$\Delta^5 y_i$
0	-7	4	5	5	3	1
$x_0 = 1$	$y_0 = -3$	$\Delta y_0 = 9$	$\Delta^2 y_0 = 10$	$\Delta^3 y_0 = 8$	$\Delta^4 y_0 = 4$	
2	6	19	18	12		
3	25	37	30			
4	62	67				
5	129					

Remarquons que les différences ascendantes d'ordre 4 sont "presque" les mêmes donc on peut prendre $n = 4$ et posons pour $x_0 = 1$ un point très proche de $x = 1.1$ avec $q = \dfrac{x - x_0}{h} = x - 1 = 1.1 - 1 = 0{,}1$ et $h = 1$. Donc,
$P_4(x) = y_0 + \dfrac{q}{1!}\Delta y_0 + \dfrac{q(q-1)}{2!}\Delta^2 y_0 + \dfrac{q(q-1)(q-2)}{3!}\Delta^3 y_0 + \dfrac{q(q-1)(q-2)(q-3)}{4!}\Delta^4 y_0$
$\mathrm{P}_4(1.1) = -3 + 9\dfrac{q}{1} + 10\dfrac{q(q-1)}{2} + 8\dfrac{q(q-1)(q-2)}{6} + 4\dfrac{q(q-1)(q-2)(q-3)}{24} = -2.4047$
d'où $f(1.1) \simeq -2.4047$.

L'évaluation de l'erreur est donnée par la relation (4.22) pour $n = 4$, $x = 1.1$ et $q = 0.1$
d'où $|\epsilon(1.1)| \simeq \left|\frac{0.1(0.1-1)(0.1-2)(0.1-3)(0.1-4)}{5!}\Delta^5 y_0\right| \simeq 0.016$
Ainsi, $f(1.1) = -2.4047 \pm 0.016$.

4.6 Différences finies descendantes

Définition 4.9.
La différence $f(x_i) - f(x_{i-1}) = y_i - y_{i-1}$ est appelée différence finie descendante (ou régressive) du 1^{er} ordre. On la note

$$\nabla f(x_i) = \nabla y_i = y_i - y_{i-1} \tag{4.23}$$

Définition 4.10.
La différence finie descendante d'ordre n est définie par:

$$\nabla^n f(x_i) = \nabla^n y_i = \nabla(\nabla^{n-1} y_i) \quad n = 2,3,... \tag{4.24}$$

4.6.1 Tableau des différences finies descendantes

Soit le tableau suivant, pour $n = 5$, par exemple

x_i	y_i	∇y_i	$\nabla^2 y_i$	$\nabla^3 y_i$	$\nabla^4 y_i$	$\nabla^5 y_i$
x_0	y_0	$\searrow$				
x_1	y_1	∇y_1	$\searrow$			
x_2	y_2	∇y_2	$\nabla^2 y_2$	$\searrow$		
x_3	y_3	∇y_3	$\nabla^2 y_3$	$\nabla^3 y_3$	$\searrow$	
x_4	y_4	∇y_4	$\nabla^2 y_4$	$\nabla^3 y_4$	$\nabla^4 y_4$	$\searrow$
x_5	y_5	∇y_5	$\nabla^2 y_5$	$\nabla^3 y_5$	$\nabla^4 y_5$	$\nabla^5 y_5$

Exemple d'application

Former le tableau des différences finies descendantes des données suivantes

x_i	0	1	2	3	4	5
$f(x_i) = y_i$	-7	-3	6	25	62	129

Solution

x_i	y_i	∇y_i	$\nabla^2 y_i$	$\nabla^3 y_i$	$\nabla^4 y_i$	$\nabla^5 y_i$
0	-7	$\searrow$				
1	-3	4	$\searrow$			
2	6	9	5	$\searrow$		
3	25	19	10	5	$\searrow$	
4	62	37	18	8	3	$\searrow$
5	129	67	30	12	4	1

4.6.2 Polynôme d'interpolation descendant de Newton

Définition 4.11.

Le polynôme d'interpolation de la fonction f aux points x_i, $i = 0,...,n$ mis sous la forme

$$P_n(t) = y_n + \frac{t}{1!}\nabla y_n + \frac{t(t+1)}{2!}\nabla^2 y_n + ... + \frac{t(t+1)(t+2)...(t+n-1)}{n!}\nabla^n y_n \qquad (4.25)$$

où $t = \dfrac{x - x_n}{h}$ et $h = x_i - x_{i-1}$, $i = 1,...,n$ est appelé "polynôme d'interpolation descendant (ou régressif) de Newton " ou bien aussi "deuxième formule d'interpolation de Newton".

Remarque 4.11.

Le polynôme d'interpolation descendant de Newton est utilisé surtout lorsque la fonction $y = f(x)$ est interpolée dans un voisinage de x_n (x voisin de x_n).

Exemple d'application

Construire le polynôme d'interpolation, en utilisant la deuxième formule de Newton avec

x_i	0	1	2	3	4
$f(x_i) = y_i$	-1	3	17	47	99

Solution

Le tableau des différences finies descendantes est donné par:

x_i	y_i	∇y_i	$\nabla^2 y_i$	$\nabla^3 y_i$	$\nabla^4 y_i$
0	-1	$\searrow$			
1	3	4	$\searrow$		
2	17	14	10	$\searrow$	
3	47	30	16	6	$\searrow$
4	99	52	22	6	0

De ce tableau, on conclut que $n = 3$, $x_n = x_3 = 3$ et $h = 1$ d'où $t = \dfrac{x - x_3}{h} = x - 3$.
D'après la relation (4.25), le polynôme d'interpolation pour $n = 3$ s'écrit:
$P_3(t) = y_3 + \dfrac{t}{1!}\nabla y_3 + \dfrac{t(t+1)}{2!}\nabla^2 y_3 + \dfrac{t(t+1)(t+2)}{3!}\nabla^3 y_3$
d'où $P_3(t) = 47 + 30t + 8t(t+1) + t(t+1)(t+2)$ et enfin
$P_3(x) = 47 + 30(x-3) + 8(x-3)(x-2) + (x-3)(x-2)(x-1) = x^3 + 2x^2 + x - 1$.

4.7 Polynôme d'interpolation d'Hermite

4.7.1 Principe de l'interpolation d'Hermite

Soient x_0,x_1,...,x_n, $(n+1)$ points deux à deux distincts de l'intervalle $[a,b]$ et soit $f \in C^1([a,b])$ une fonction dont on connaît les valeurs $f(x_i)$ et $f'(x_i)$ pour $i = 0,...,n$. On cherche à déterminer un polynôme $P(x)$ qui interpole f et f' aux points x_i, $i = 0,...,n$. Ainsi, d'après la définition d'un polynôme d'interpolation $P(x)$ doit satisfaire simultanément les deux conditions suivantes:

$$P(x_i) = f(x_i), \quad i = 0,...,n \tag{4.26}$$

$$P'(x_i) = f'(x_i), \quad i = 0,...,n \tag{4.27}$$

Le théorème suivant prouve l'existence et l'unicité d'un tel polynôme.

Théorème 4.3. *Une condition nécessaire et suffisante pour qu'il existe un unique polynôme d'interpolation $P(x)$ de f aux points x_i, $i = 0,...,n$ vérifiant les relations* (4.26) *et* (4.27)*, de degré inférieur ou égal à $(2n+1)$, est que les points x_i, $i = 0,...,n$ soient tous distincts deux à deux.*

Définition 4.12. Le polynôme d'interpolation d'Hermite est défini par la formule suivante:

$$P_{2n+1}(x) = \sum_{i=0}^{n} f(x_i).H_i(x) + \sum_{i=0}^{n} f'(x_i).V_i(x) \tag{4.28}$$

Où les polynômes $H_i(x)$ et $V_i(x)$ sont donnés par les formules suivantes:

$$H_i(x) = \left[1 - 2\left(x - x_i\right).L_i'(x_i)\right].L_i^2(x) \tag{4.29}$$

$$V_i(x) = \left(x - x_i\right).L_i^2(x) \tag{4.30}$$

Où $L_i(x) = \prod_{j=0, j\neq i}^{j=n} \frac{(x - x_j)}{(x_i - x_j)}$ sont les polynômes de Lagrange définis dans la relation (4.3) et $L_i'(x_i) = \sum_{j=0, j\neq i}^{j=n} \frac{1}{x_i - x_j}$ avec $L_i'(x)$ la dérivée du polynôme $L_i(x)$.

4.7.2 Erreur d'interpolation d'Hermite

Nous pouvons établir un résultat d'expression de l'erreur d'interpolation similaire à celui obtenu pour les polynômes de Lagrange.

Théorème 4.4. *Soit f une fonction de classe $\mathcal{C}^{2n+2}([a,b])$ alors il existe un réel θ_x appartenant au plus petit intervalle contenant les points x_i, $i = 0,...,n$ deux à deux distincts tel que:*

$$\epsilon(x) = f(x) - P_{2n+1}(x) = \frac{f^{(2n+2)}(\theta_x)}{(2n+2)!}\left(\prod_{j=0}^{j=n}(x - x_j)\right)^2 \tag{4.31}$$

Remarque 4.12. La formule (4.31) ne permet évidemment pas de calculer la valeur exacte de l'erreur parce que, en général, θ_x est inconnu. Elle permet par contre, d'en déduire la majoration suivante en posant $M = \max_{a\leq x\leq b} \left|f^{(2n+2)}(x)\right|$.

$$|\epsilon(x)| \leq \frac{M}{(2n+2)!}\left(\prod_{j=0}^{j=n}(x - x_j)\right)^2 \tag{4.32}$$

Exemple d'application

Soit $f \in C^2([0,1])$, trouver le polynôme $P_3(x)$ de degré ≤ 3 tel que $P_3(0) = f(0)$, $P_3(1) = f(1)$, $P_3'(0) = f'(0)$ et $P_3'(1) = f'(1)$.

Solution

Commençons par calculer les polynômes de Lagrange $L_i(x)$ aux points $x_0 = 0$ et $x_1 = 1$ d'où $L_0(x) = 1-x$; $L_1(x) = x$; $L_0'(x) = -1$ et $L_1'(x) = 1$. En appliquant les formules (4.29) et (4.30) de la définition (4.12) avec les points $x_0 = 0$ et $x_1 = 1$, on obtient:
$H_0(x) = [1 - 2(x - x_0)L_0'(x_0)]L_0^2(x) = 1 + 2x$ et $H_1(x) = [1 - 2(x - x_1)L_1'(x_1]L_1^2(x)$
$V_0(x) = (x - x_0)L_0^2(x) = x(1 - x)^2$ et $V_1(x) = (x - x_1)L_1^2(x) = (x - 1)x^2$.
En remplaçant toutes ses expressions dans la relation (4.28) du polynôme d'hermite, on obtient:
$P_3(x) = f(0)(1 + 2x)(1 - x)^2 + f(1)(3 - 2x)x^2 + f'(0)x(1 - x)^2 + f'(1)(x - 1)x^2$.
En développant, le polynôme d'interpolation d'Hermite aux points $x_0 = 0$ et $x_1 = 1$ est donné par:
$P_3(x) = (2f(0)-2f(1)+f'(0)+f'(1))x^3+(-3f(0)+3f(1)-2f'(0)-f'(1))x^2+f'(0)x+f(0)$

4.8 Exercices corrigés sur le chapitre

Exercice 1 Considérons les points d'interpolation suivants:
$(x_i,f(x_i))$, $i = 0,...,n$; $(a,f(a))$ et $(b,f(b))$. Soient
$P(x)$ le polynôme d'interpolation de f aux points $(x_i,f(x_i))$, $i = 0,...,n$ et $(a,f(a))$, et
$Q(x)$ le polynôme d'interpolation de f aux points $(x_i,f(x_i))$, $i = 0,...,n$ et $(b,f(b))$.
Montrer que $M(x) = \dfrac{(x-b)P(x) - (x-a)Q(x)}{a-b}$ est le polynôme d'interpolation de f associé à tous les points $(x_i,f(x_i))$, $i = 0,...,n$; $(a,f(a))$ et $(b,f(b))$.

Solution de l'exercice 1:

$M(x)$ est un polynôme d'interpolation de f aux points $(x_i,f(x_i))$, $i = 0,...,n$; $(a,f(a))$ et $(b,f(b))$ si et seulement si $M(x_i) = f(x_i)$ $\ i = 0,...,n$, $M(a) = f(a)$ et $M(b) = f(b)$.

(1) $M(x_i) = \dfrac{(x_i-b)P(x_i) - (x_i-a)Q(x_i)}{a-b} = \dfrac{(x_i-b)f(x_i) - (x_i-a)f(x_i)}{a-b} = f(x_i)$,
$i = 0,...,n$ car $P(x_i) = Q(x_i) = f(x_i)$, $\ i = 0,...,n$.

(2) $M(a) = \dfrac{(a-b)P(a) - (a-a)Q(a)}{a-b} = \dfrac{(a-b)f(a)}{a-b} = f(a)$.

(3) $M(b) = \dfrac{(b-b)P(b) - (b-a)Q(b)}{a-b} = \dfrac{(a-b)f(b)}{a-b} = f(b)$.

(1), (2) et (3) impliquent que $M(x)$ est un polynôme d'interpolation aux $(n+3)$ points.

Exercice 2 Soit f une fonction dont les valeurs numériques sont données par $f(-2) = 0$; $f(-1) = -1$; $f(1) = 1$ et $f(2) = 0$.

(1) Calculer le polynôme d'interpolation $P_2(x)$ de la fonction f sous forme de Lagrange aux points -2, -1 et 1.

(2) Calculer le polynôme d'interpolation $Q_2(x)$ de la fonction f sous forme de Newton aux points -1, 1 et 2.

(3) **(a)** Montrer que le polynôme $P(x) = \dfrac{(x+2)Q_2(x) - (x-2)P_2(x)}{4}$ est un polynôme d'interpolation de f aux points -2, -1, 1 et 2.

(b) Donner une valeur approchée de f au point $x = 0$.

(c) Donner l'expression de l'erreur commise au point $x = 0$.

Solution de l'exercice 2:

(1) Le polynôme d'interpolation $P_2(x)$ de la fonction f sous forme de Lagrange aux points -2, -1 et 1 s'écrit $P_2(x) = f(-2)L_0(x) + f(-1)L_1(x) + f(1)L_2(x)$ où $L_0(x) = \dfrac{(x+1)(x-1)}{3}$; $L_1(x) = \dfrac{-(x+2)(x-1)}{2}$ et $L_2(x) = \dfrac{(x+1)(x+2)}{6}$ sont les polynômes de Lagrange d'où $P_2(x) = \dfrac{2x^2+3x-2}{3}$.

(2) Le polynôme d'interpolation $Q_2(x)$ de la fonction f sous forme de Newton aux points -1, 1 et 2 est donné par $Q_2(x) = f(x_0) + (x-x_0)\delta f[x_0,x_1] + (x-x_0)(x-x_1)\delta f[x_0,x_1,x_2] = f(-1) + (x+1)\delta f[-1,1] + (x+1)(x-1)\delta f[-1,1,2] = -1 + (x+1) + (x+1)(x-1)(-2/3)$ où $\delta f[-1,1]$ et $\delta f[-1,1,2]$ sont les différences divisées d'ordre 2 et 3 de f d'où $Q_2(x) = \dfrac{-2x^2+3x+2}{3}$.

$(3)_{(a)}$ D'après l'exercice 1, le polynôme $P(x) = \dfrac{(x+2)Q_2(x) - (x-2)P_2(x)}{4}$ est bien le polynôme d'interpolation de f aux points $x_0 = -1$; $x_1 = 1$; $a = 2$ et $b = -2$.

$(3)_{(b)}$ La valeur approchée de $f(0) \approx P(0) = \dfrac{(0+2)Q_2(0) - (0-2)P_2(0)}{4} = 0$ d'où $f(0) \approx 0$.

$(3)_{(c)}$ Une majoration de l'erreur commise au point $x = 0$ est donnée par la formule suivante:

$$|\epsilon(0)| = |f(0) - P(0)| = \left| \frac{f^{(4)}(\theta_x)}{(4)!} \prod_{j=0}^{j=3} (0 - x_j) \right| \text{ où } f^{(4)} \text{ est la dérivée d'ordre 4 de } f.$$

$$|\epsilon(0)| = \frac{\left|f^{(4)}(\theta_x)\right|}{4!} |(0+2).(0+1).(0-1).(0-2)| \text{ avec } \theta_x \in [-2,2].$$

$$|\epsilon(0)| \leq \frac{M}{6} \text{ avec } M = \max_x \left|f^{(4)}(x)\right|.$$

Exercice 3 En utilisant l'erreur d'interpolation de lagrange, montrer que $\sum_{i=0}^{i=n} x_i^{n+1}.L_i(x) = x^{n+1} - \prod_{j=0}^{j=n}(x - x_j)$ où $L_i(x)$, $i = 0,...,n$ représentent les polynômes de Lagrange.

Solution de l'exercice 3:

En posant $f(x) = x^{n+1}$ et en calculant ses dérivées jusqu'à l'ordre $(n+1)$, on obtient $f'(x) = (n+1)x^n$, $f''(x) = (n+1)nx^{n-1}$, $f^{(3)}(x) = (n+1)n(n-1)x^{n-2}$,..., et $f^{(n+1)}(x) = (n+1)n(n-1)(n-2)...3 \times 2 \times 1 = (n+1)!$ et on a aussi $f^{(n+1)}(\theta_x) = (n+1)!$

L'erreur d'interpolation de Lagrange s'écrit:

$$\epsilon(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\theta_x)}{(n+1)!}\prod_{j=0}^{j=n}(x - x_j) = \frac{(n+1)!}{(n+1)!}\prod_{j=0}^{j=n}(x - x_j) = \prod_{j=0}^{j=n}(x - x_j)$$

d'où $\epsilon(x) = \prod_{j=0}^{j=n}(x - x_j)$.

Le polynôme d'interpolation de Lagrange de la fonction $f(x) = x^{n+1}$ aux points x_i,

$i = 0,...,n$ s'écrit $P_n(x) = \sum_{i=0}^{n} f(x_i).L_i(x) = \sum_{i=0}^{n} x_i^{n+1}.L_i(x)$.

Comme $\epsilon(x) = f(x) - P_n(x)$ et en remplaçant $\epsilon(x)$ et $f(x)$ par leur valeur,

on obtient $P_n(x) = f(x) - \epsilon(x) = x^{n+1} - \prod_{j=0}^{j=n}(x - x_j)$ et finalement on a

$$P_n(x) = \sum_{i=0}^{n} x_i^{n+1}.L_i(x) = x^{n+1} - \prod_{j=0}^{j=n}(x - x_j).$$

Bibliographie

[1] Allab A., *É*léments d'analyse: Fonction d'une variable réelle. Office des Publications Universitaires, 1986.

[2] Atteia M. et Pradel M., *É*léments d'analyse numérique. Editions Cépaduès, 2000.

[3] Bakhvalov N., Méthodes numériques. Editions Mir, Moscou, 1976.

[4] Baranger J., Introduction à l'Analyse Numérique. Editions Herman, Paris, 1977.

[5] Boumahrat M. et Bourdin A., Méthodes numériques appliquées: Avec nombreux problèmes résolus en Fortran IV. Office des Publications Universitaires, 1983.

[6] Ciarlet P.G., Introduction à l'analyse numérique matricielle et à l'optimisation. Dunod, Paris, 1998.

[7] Curtis F. Gerald and Patrick O. Wheatley, Applied Numerical Analysis. Addison-Wesley, 1999.

[8] Demidovitch B. et Maron I., *É*léments de Calcul Numérique. Editions Mir, Moscou, 1979.

[9] Lahrib M., Cours d'analyse numérique. Office des Publications Universitaires, 2003.

[10] Lascaux P. et Théodor R., Analyse numérique matricielle appliquée à l'art de l'ingénieur. tomes I et II, Masson, 1987.

[11] Martin V., Brancherie D. et Smaoui H., Cours MT09: Qu'est-ce que l'analyse numérique?. Laboratoire de Mathématiques Appliquées de Compiègne, 2013.