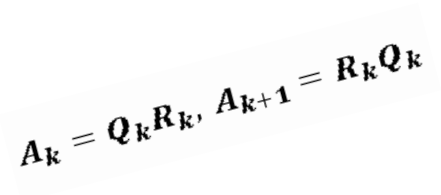
Vasile MORARU
Daniela ISTRATI

ANALYSE NUMERIQUE MATRICIELLE

Notes de cours



Département Informatique et Ingénierie des Systèmes

Maître de conf. dr. Vasile MORARU
Lect. univ. Daniela ISTRATI

ANALYSE NUMERIQUE MATRICIELLE

Notes de cours

CZU

Le présent ouvrage montre les principales méthodes de calcul numérique matriciel pour résoudre les problèmes qui peuvent être rencontrés fréquemment dans la pratique.

Cet ouvrage est destiné premièrement aux étudiants de la spécialité 0613.1 Technologie de l'information, Filière Francophone « Informatique » et apportera un soutien efficace sur l'enseignement des cours *Méthodes et modèles informatiques, méthodes numériques, mathématiques computationnelles, modèles mathématiques et optimisations,* etc.

Cependant, le livre peut également être utilisé par tous ceux qui s'intéressent à l'utilisation de méthodes numériques et des moyens électroniques de calcul pour résoudre des problèmes pratiques.

Auteurs : maître de conf.,dr. Vasile Moraru,
lect. univ. Daniela Istrati

Responsable d'édition - prof. univ., dr. hab. Emilian Guțuleac
Aviz – maître de conf. Liviu Carcea

Descrierea CIP a Camerei Naționale a Cărții

Bun de tipar 07.07.20 Formatul hârtiei 60x84 1/16
Hârtie ofset. Tipar RISO Tirajul 50 ex.
Coli de tipar 6,25 Comanda nr.54

2004, UTM, Chişinău, bd. Ştefan cel Mare şi Sfânt, 168

Préface

Le matériel présenté dans cet ouvrage inclut les méthodes de l'algèbre linéaire et présente des éléments de l'analyse matricielle, les algorithmes les plus représentatifs qui interviennent dans les problèmes de résolution des systèmes des équations linéaires et de calcul des valeurs et des vecteurs propres.

Dans cette élaboration méthodique sont exposées les méthodes directes et itératives de résolution des systèmes des équations linéaires (méthode d'élimination de Gauss, méthode de Cholesky avec ses factorisations triangulaires, méthode de Jacobi, méthode de Gauss-Seidel, méthodes d'orthogonalisation etc.), en faisant simultanément des appréciations sur l'efficacité et la stabilité numérique de celles-ci. On souligne le fait que les méthodes basées sur des transformations de ressemblance orthogonales sont plus efficaces que les méthodes classiques de détermination des valeurs et des vecteurs propres.

Cet ouvrage est destiné premièrement aux étudiants de la spécialité 0613.1 Technologie de l'information, Filière Francophone "Informatique" et apportera un soutien efficace sur l'enseignement des cours *Méthodes et modèles de calcul, méthodes numériques, mathématiques computationnelles, modèles mathématiques et optimisations*, etc.

Cependant, le livre peut également être utilisé par tous ceux qui s'intéressent à l'utilisation de méthodes numériques et des moyens électroniques de calcul pour résoudre des problèmes pratiques.

Cet ouvrage est réalisé avec le support de l'Agence Universitaire de la Francophonie dans le cadre du projet **AUF - Pentalog CHI, ATIC - IUT Rouen -UTM,** *"Première formation universitaire francophone à présence renforcée en entreprise en*

1. Notions introductives

Les méthodes de calcul numérique sont devenues à l'époque actuelle, très importantes. On les applique quasi partout : dans l'ingénierie et dans l'économie, dans les mathématiques et dans la physique, dans la médicine, dans l'astronomie, dans la chimie, dans la géologie etc. C'est tant grâce aux progrès obtenus dans le domaine des ordinateurs qu'aux expériments de plus en plus compliqués dans le modelage mathématique.

Par *méthodes numériques* on sous-entend des méthodes de résolution des problèmes à l'aide des opérations au caractère arithmétique et logique sur les nombres réels, donc à l'aide des opérations qui peuvent être exécutées automatiquement par un ordinateur.

La résolution d'un problème imposé par la pratique commence par la construction du *modèle mathématique*. Le modèle mathématique représente la formulation mathématique du problème énoncé, donc constitue l'expression mathématique des relations et des restrictions d'entre les paramètres du problème.

Après la formulation mathématique du problème on réalise *le choix de la méthode numérique* et *on élabore l'algorithme de calcul*. Ces étapes sont les plus importantes dans le processus de résolution des problèmes. Au choix de la méthode numérique on prend en considération la vitesse de convergence, la précision, la stabilité, le temps d'exécution et le nécessaire de mémoire.

L'algorithme de la méthode numérique consiste d'un nombre limité d'opérations arithmétiques et logiques, qui doivent être effectuées par l'ordinateur pour la résolution du problème donné. Les règles de calcul forment les pas de l'algorithme.

Soulignons le fait que la notion d'algorithme dans sa forme générale se situe parmi les notions fondamentales des

a) *Généralité*. Cela signifie que l'algorithme doit non seulement résoudre un problème, mais également tous les problèmes de la classe respective de problèmes.
b) Finitude. Le nombre des transformations intermédiaires, appliqué aux données d'entrée pour obtenir les données de sortie, est fini.
c) *Unicité*. Toutes les transformations intermédiaires doivent être déterminées sans équivoque des règles de l'algorithme.

Après l'élaboration de l'algorithme de la méthode numérique de calcul on passe à l'écriture du programme de résolution du problème dans un langage de programmation. Puis on passe *au test et la vérification du programme*. Après qu'on teste le programme du point de vue syntaxique, il est nécessaire que le programme soit vérifié par des exemples des problèmes concrets dont les solutions sont connues.

Par conséquence, la résolution d'un problème à l'ordinateur nécessite le parcours des étapes suivantes :

1. énoncé du problème et l'expression mathématique, en soulignant ce qui est donné et ce qui est requis;
2. choix d'une méthode numérique pour obtenir la solution;
3. élaboration de l'algorithme de calcul;
4. écriture du programme de calcul;
5. test et vérification du programme;
6. analyse des résultats obtenus.

La majorité des méthodes de calcul représente des processus itératifs. Cela veut dire que, ayant un x_0 donné, on construit une suite: $x_0, x_1, \dots \dots x_n \dots$ (notée comme d'habitude $\{x_n\}$ qui dans certaines conditions converge vers la solution exacte x_i du problème considéré. Les éléments $x_n, n = 0,1,2, \dots$, peuvent être aussi des nombres réels que des vecteurs ou des matrices.

en fonction de la précision donnée, ainsi que le terme courant x_m constitue une approximation satisfaisante de la solution cherchée x_i. Voilà pourquoi un fait essentiel dans la comparaison des modèles de calcul numérique est constitué par l'appréciation de la vitesse de convergence des méthodes. Dans les méthodes de convergence rapide sont nécessaires un nombre plus petit d'itérations pour atteindre la précision prescrite que dans les méthodes de convergence lente. Des méthodes qui nécessitent le même volume de calcul à chaque itération en pratique est choisie la méthode à convergence plus rapide.

2. Eléments de l'analyse matricielle
2.1. Vecteurs et matrices

Un tableau rectangulaire de $m \times n$ nombres réels disposés sur *m* lignes et *n* colonnes :

$$A = \begin{pmatrix} a_{11}\ a_{12}\ a_{13} \dots a_{1n} \\ a_{21}\ a_{22}\ a_{23} \dots a_{23} \\ \dots\dots\dots\dots\dots\dots\dots \\ a_{m1}\ a_{m2}\ a_{m3} \dots a_{mn} \end{pmatrix}$$

est appelée *matrice*. Les nombres a_{ij} sont nommés les *éléments de la matrice*. La matrice peut se représentée symbolique comme cela :

$$A = (a_{ij}), i = 1,2, \dots m;\ j = 1,2, \dots n; \text{ ou } A = (a_{ij})_{mn}$$

On dira que la matrice *A* est de dimensions $m \times n$. Quand $m = n$ la matrice est nommée *carrée de l'ordre n* et est notée $a = (a_{ij})_n$. Si $m \neq n$ la matrice est nommée *rectangulaire*. Une matrice $1 \times n$ est appelée *vecteur ligne* et une $n \times 1$ est appelée *vecteur colonne*.

Un système ordonné de *n* nombres réels est nommé vecteur *n-dimensionnel*. Un vecteur se représente par une matrice à une seule ligne ou à une seule colonne. Dans cet ouvrage, on comprend le vecteur colonne :

$$x = \begin{pmatrix} x_! \\ x_2 \\ \vdots \end{pmatrix}$$

$$A^T = \begin{pmatrix} a_{11}\, a_{21}\, a_{31} \ldots a_{m1} \\ a_{12}\, a_{22}\, a_{32} \ldots a_{m2} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}\, a_{2n}\, a_{3n} \ldots a_{mn} \end{pmatrix}$$

En particulier, la transposée d'un vecteur colonne *x* c'est un vecteur ligne :

$$x^T = (x_1, x_2, \ldots x_n).$$

La matrice A est nommée matrice symétrique si $A = A^T$, donc $a_{ij} = a_{ji}$.

La totalité des vecteurs *n*-dimensionnels est nommée *espace linéaire n- dimensionnel* et est notée avec R^n.

La somme des matrices A et *B*, toutes les deux de dimensions $m \times n$, est une matrice C à dimensions $m \times n$ avec les éléments $c_{ij} = a_{ij} + b_{ij}$.

Le produit des matrices *A* et *B* se définie seulement dans le cas quand le nombre de colonnes du premier facteur est égal au nombre des lignes du deuxième facteur. Ainsi, si $A = \left(a_{ij}\right)_{mn}$ et $B = \left(b_{ij}\right)_{np}$, alors $C = A \times B$, où $C = \left(c_{ij}\right)_{mp}$ et

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}\,.\, i = 1{,}2, \ldots m\; j = 1{,}2, \ldots p.$$

Considérons deux vecteurs $x, y \in R^n$. Comme un cas particulier on obtient :

$$x^T y = (x_1, x_2, \dots x_n) \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

$$xy^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot (y_1, y_2, \dots y_n) = \begin{pmatrix} x_1 y_1 \; x_1 y_2 \dots x_1 y_n \\ x_2 y_1 \; x_2 y_2 \dots x_2 y_n \\ \dots\dots\dots\dots\dots \\ x_n y_1 \; x_n y_2 \dots x_n y_n \end{pmatrix}.$$

$x^T y$ est nommé *produit scalaire* des vecteurs $x, y \in R^n$ et on le note encore (x, y)

xy^T est nommé *produit dyadique* des vecteurs $x, y \in R^n$ et est une matrice carrée de l'ordre *n* et on le note encore >x,y<.

Pour n'importe quels vecteurs de R^n ont lieu les propriétés :

- *(x,y)=(y,x);*
- *(x+y,z)=(x,z)+(y,z);*
- *(ax,y)=a(x,y);* $(a \in R)$*;*
- *(x,x)≥ 0; (x,x)=0,* si et seulement si *x*=0.

Le produit scalaire est commutatif. En cas général pour les matrices $A \times B \neq B \times A$.

Toute matrice carrée peut se multiplier à elle-même.

$A \times A = AA^2$, $A^2 \times A = A^3$, …

Ont lieu les égalités :

1. $(A \times B) \times C = A \times (B \times C)$, loi associative,
2. $A(B + C) = AB + AC$, loi de distributivité de gauche,
3. $(B + C)A = BA + CA$, loi de distributivité de droite,
4. $\alpha(AB) = (\alpha A)B = A(\alpha B)$, $\alpha \in R$

2.2. Normes de vecteurs et de matrices

La norme d'un vecteur $x \in R^n$ est un nombre réel, noté $\|x\|$, aux propriétés :

1. $\|x\| \geq 0$ pour tout $x \in R^n$
2. $\|x\| = 0$ si et seulement si $x = 0$
3. $\|\alpha x\| = |\alpha|\|x\|$ pour tout $x \in R^n \quad \alpha \in R$
4. $\|x + y\| \leq \|x\| + \|y\|$ pour tous $x, y \in R^n$

Pour n'importe quel vecteur $x \in R^n$ on définit les normes :

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|,$$
$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i{}^2|\right)^{1/2},$$
$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Elles satisfont les propriétés de dessus de la norme. La norme $\|x\|_2$ est nommée *la norme* euclidienne *(de Euclid)*. Elle provient du produit scalaire :

$$\|x\|_2 = \sqrt{(x, x)},$$

et généralise la notion de longueur du vecteur.

Ont lieu les inégalités suivantes :

$$\|x\|_\infty \leq \|x\|_1 \leq n\|x\|_\infty,$$
$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty,$$
$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1.$$

Exemple :

Soit $x = (1, -2, -3)^T$.

Alors $\|x\|_1 = 6$, $\|x\|_2 = \sqrt{14}$ et $\|x\|_\infty = 3$.

Cette relation est appelée *l'inégalité de Schwarz – Cauchy–Buniacovski.*

L'angle de deux vecteurs x ,y du R^n se définie par la formule :

$$cos\,\theta = \frac{(x,y)}{\|x\|_2\|y\|_2}.$$

Deux vecteurs x,y du R^n sont dits *orthogonaux* si $(x,y) = 0$

Sur l'ensemble des matrices carrées, une norme peut être introduite dans le sens défini ci-dessus pour les vecteurs. Les plus importantes et les plus utilisées sont les normes matricielles définies comme cela :

$$\|A\| = \max_{|x|=1} \|Ax\|.$$

Cette norme satisfait la condition suivante :

$$\|A \cdot B\| \leq \|A\| \cdot \|B\|$$

pour toutes matrices *A et B.*

Si en plus, n'importe quel soit le vecteur $x \in R^n$ on a :

$$\|Ax\| \leq \|A\|\|x\|$$

On dit que la norme matricielle est compatible à la norme vectorielle.

Les normes suivantes :

$$\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}|;$$

sont des normes matricielles compatibles à celles vectorielles $\|x\|_\infty$et $\|x\|_1$.

Le nombre λ est nommée valeur propre du A s'il existe un vecteur $x \in R^n$, ainsi que $Ax = \lambda x$. La totalité des valeurs propres de la matrice A forme le *spectre du A* et est noté avec $\sigma(A)$.

Le rayon spectral de A est défini par la relation :

$$\rho(A) = \max_{\lambda\in\sigma(A)} |\lambda|$$

La norme matricielle subordonnée à la norme euclidienne $\|x\|_2$est

$$\|A\|_2 = \sqrt{\rho(A^T A)}.$$

Dans les applications on utilise souvent la norme suivante :

$$\|A\|_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}$$

nommée *la norme du Frobenius*, qui n'est pas subordonnée à une norme vectorielle.

Par I on note la matrice unité :

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

Quelle que soit la matrice *A*, on a *IA=AI=A.* La matrice *A* est appelée *inversible* s'il existe une matrice, notée par A^{-1}, ainsi que $A^{-1}A = AA^{-1} = I$.

Si $\|A\| < 1$, alors la matrice *I-A* est inversible et

$$(I - A)^{-1} = I + A + A^2 + A^3 + \cdots$$

$$\|(I - A)^{-1}\| \leq \frac{1}{1 - \|A\|}.$$

Pour que la matrice *A* soit inversible il suffit que

$$\lim_{n\to\infty} A^n = O.$$

où *O* c'est la matrice nulle : (une matrice aux éléments égaux à zéro).

2.3. Matrices spéciales

Une matrice carrée de la forme :

$$D = \begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & d_n \end{pmatrix}$$

est nommée *matrice diagonale*. Une telle matrice est notée encore

$$D = Diag\ (d_1, d_2, \ldots, d_n).$$

La matrice A est nommée *triangulaire inférieure (triangulaire supérieure)* si ses éléments satisfont les relations :

$$a_{ij} = 0 \text{ pour } i < j (i > j), \quad i, j = 1,2, \ldots, n.$$

Les matrices triangulaires inférieures sont notées d'habitude par L mais celles triangulaires supérieures par U ; par exemple :

$$L = \begin{pmatrix} -1 & 0 & 0 \\ 2 & 3 & 0 \\ 7 & -4 & 5 \end{pmatrix}, \qquad U = \begin{pmatrix} 4 & 2 & 1 \\ 0 & 3 & -7 \\ 0 & 0 & 6 \end{pmatrix}.$$

La matrice A est nommée *tri diagonale* si ses éléments satisfont les relations :

$$a_{ij} = 0 \text{ pour } |i - j| > 1,\ i, j = 1,2, \ldots, n.$$

Ainsi :

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & 0 & \cdots & 0 \\ 0 & a_{32} & a_{33} & a_{34} & \cdots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ & & & & \cdots & a_{n-1,n} \\ 0 & 0 & 0 & 0 & \cdots & a_{n,n-1} a_{nn} \end{pmatrix}$$

Immédiatement on vérifie que

$$(Qx)^T Qy = x^T y, \text{pour } \forall\, x, y \in R^n,$$

$$\|Qx\|_2 = \|x\|_2 \text{ pour } \forall\, x \in R^n,$$

$$\|Q\|_2 = 1, \|QA\|_2 = \|AQ\|_2 = \|A\|_2 \text{ pour } \forall A.$$

Par suite, les matrices orthogonales conservent gardent le produit scalaire, la longueur des vecteurs et la norme des matrices.

Exemple :

$$Q = \begin{pmatrix} cos\,\theta & -sin\,\theta \\ -sin\,\theta & cos\,\theta \end{pmatrix}, \qquad \mathrm{Q}^T = Q^{-1} = \begin{pmatrix} cos\,\theta & sin\,\theta \\ -sin\,\theta & cos\,\theta \end{pmatrix}$$

La matrice *Q* fait tourner tout vecteur d'angle θ, alors que la matrice Q^T le tourne en direction inverse à l'angle -θ.

La matrice obtenue de la matrice unité par la réordonnance de ses colonnes est appelée *matrice de permutation* et elle est notée par *P*.

Par exemple :

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathrm{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Une matrice de la forme uv^T, où $u, v \in R^n$, s'appelle matrice de *premier rang*. Les matrices $I + auv^T$, où *a* est un scalaire, s'intitulent comme *matrices élémentaires*. La matrice

$$\mathrm{H} = I - \frac{2}{v^T v} vv^T$$

On vérifie facilement que les matrices P et H sont orthogonales.

L'identité Sherman-Morrison-Woodbury. Soit $x, y \in R^n$ et supposons que la matrice A est inversible. La matrice $A + xy^T$ sera inversible seulement dans le cas, où $1 + y^T A^{-1} x \neq 0$.

En plus :

$$(A + xy^T)^{-1} = A^{-1} - \frac{A^{-1} x y^T A^{-1}}{1 + y^T A^{-1} x}.$$

Une matrice A symétrique est dénommée positivement définie si

$$(Ax, x) > 0 \text{ pour } \forall\ x \neq 0, x \in R^n.$$

Les éléments sur la diagonale principale de la matrice positivement définie sont positifs. Sur la diagonale principale se trouve aussi l'élément maximal (en module) d'une matrice positivement définie.

Exemple. La matrice

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

est positivement définie, parce que

$$(Ax, x) = x^T Ax = (x_1\ x_2) \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 - 2x_1 x_2 + 2x_2^2$$
$$= (x_1 - x_2)^2 + x_2^2 > 0, \text{pour } \forall x \neq 0.$$

3. Systèmes d'équations algébriques linéaires

Considérons le système d'équations algébriques linéaires :

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n \end{cases} \qquad (3.1)$$

Ce système peut être écrit sous forme matricielle :

$$Ax=b,$$

où

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots & & & \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \ldots\ldots \\ b_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \ldots\ldots \\ x_n \end{pmatrix}$$

Résoudre le système d'équations donné signifie déterminer un vecteur $x^* \in R^n$ qui satisfait les égalités (3.1).

Quel qu'il soit le système d'équations linéaires, il existe seulement trois possibilités.

1. Le système d'équations a une solution unique. En ce cas on dit que le système (3.1) est *compatible déterminé.*

   Par exemple, le système :

$$\begin{cases} x_1 + x_2 = 2 \\ 2x_1 - x_2 = 1 \end{cases}$$

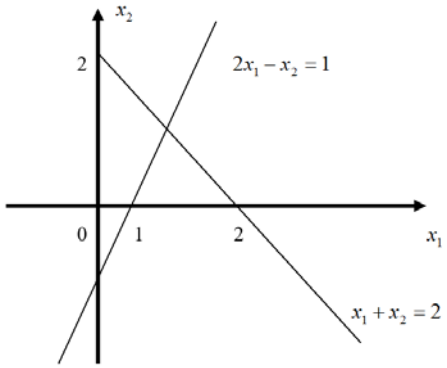


Fig. 3.1 Interprétation géométrique du système compatible déterminé

En faisant l'interprétation géométrique du système considéré, on voit que les droites $x_2 = 2 - x_1$ et $x_2 - 1 + 2x_1$ ne s'entrecroisent que dans un seul point (voir la figure 3.1).

2. Le système d'équations à une infinité de solutions. On dit que ces systèmes sont *comptablement indéterminés*. Sur la figure 3.2 on a l'interprétation géométrique du système :

$$\begin{cases} x_1 + x_2 = 1, \\ 2x_1 + 2x_2 = 2, \end{cases}$$

qui a une infinité de solutions. Ces deux équations décrivent la même droite $x_2 = 1 - x_1$.

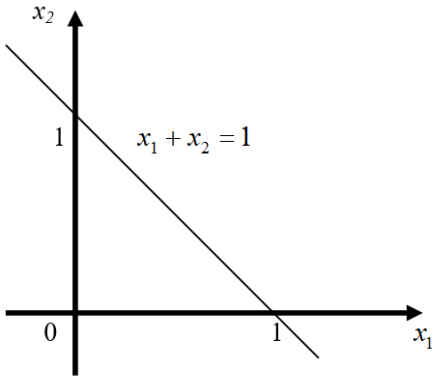


Fig. 3.2 Interprétation géométrique du système compatible indéterminé

3. Le système d'équations n'a pas de solutions, c'est-à-dire il est *incompatible*.
   Par exemple, le système :

$$\begin{cases} x_1 + 2x_2 = 2, \\ 2x_1 + 4x_2 = 7. \end{cases}$$

n'est pas compatible.
Les droites $x_2 = 1 - \frac{1}{2}x_1$ et $x_2 = \frac{7}{4} - \frac{1}{2}x_1$ sont parallèles.

Si la matrice A est non-singulière (detA$\neq 0$), alors quel qu'il soit le vecteur $b \in R^n$ le système $Ax=b$ est compatible déterminé. La solution du système peut être écrite sous la forme :

$$x^* = A^{-1}b \qquad (3.2)$$

où A^{-1} est l'inverse de A.

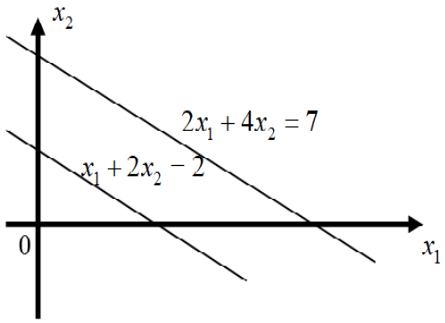


Fig. 3.3 Interprétation géométrique du système *incompatible*

L'inversion des matrices est une opération coûteuse (voir, par ex. [1]) qu'il faut éviter en pratique. Comme exemple démonstratif considérons le " système" d'une équation avec une seule inconnue :

$$7x = 21$$

Le meilleur moyen de résolution de ce problème c'est la division :

$$x^* = \frac{21}{7} = 3$$

L'utilisation de la matrice inverse nous emmènerait vers :

$$x^* = 7^{-1} * 21 = 0.142857 * 21 = 2.99997$$

Le deuxième procédé exige une opération arithmétique davantage et donne un résultat moins précis. La même chose, mais d'une manière plus prononcée est aussi vraie dans le cas de la

exprimer le fait que x^* est la solution unique du système *Ax=b*, et non seulement comme une voie d'obtention de cette solution.

Comme on le sait des mathématiques élémentaires, les systèmes d'équations linéaires peuvent être résolus à l'aide des formules de Cramer :

$$x_i^* = \frac{\Delta_i}{\Delta}, \quad \Delta = det(A), \quad \Delta_i = \sum_{j=1}^{u} A_{ij} b_j, \quad i = 1,2, \dots, n.$$

A_{ij} étant le complément algébrique de a_{ij}.

La méthode de résolution des systèmes par les formules de Cramer du point de vue pratique reste inutilisable, parce qu'elle exige un grand nombre d'opérations arithmétiques, et notamment il est nécessaire de calculer *n+1* déterminants $(\Delta, \Delta_1, \dots, \Delta_n)$ et à effectuer *n* divisions. Pour le calcul d'un déterminant sont nécessaires $((n-1) * n)$ multiplications et *(n-1)* additions. Par exemple, la résolution d'un système de *20* équations à l'aide des formules de Cramer présume l'exécution de *19*201*21* multiplications. On a constaté que si nous accomplissons ces multiplications à une calculatrice électronique avec la vitesse 10^5 opérations par seconde il nous faudra environ $3 * 10^6$ ans !

Les méthodes numériques de résolution des systèmes d'équations linéaires sont de deux types : méthodes directes et méthodes itératives.

Les méthodes directes résident en transformation du système *Ax=b* en un système équivalent pour lequel la résolution est beaucoup plus simple. Dans les méthodes directes la solution exacte s'obtient d'après un nombre limité d'opérations arithmétiques élémentaires (addition, réduction, multiplication, division et racine carrée) et ce nombre d'opérations est de l'ordre n^3. Soulignons que

conséquence, les méthodes directes dans le cas général fournissent seulement une solution approximative. On utilise les méthodes directes pour la résolution des systèmes pas trop « grands » de dimension $n \leq 200$.

La résolution des systèmes d'équations linéaires par une *méthode itérative* signifie la construction d'une chaîne de vecteurs $x^{(k)}$, $k = 0,1, \ldots$ (commençant par un vecteur $x^{(0)}$choisi arbitrairement) qui convergent vers la solution du système considéré. Dans les méthodes itératives une itération exige d'habitude l'exécution d'un nombre de l'ordre n^2 opérations arithmétiques. C'est pourquoi on utilise les méthodes itératives pour la résolution de « grands » systèmes, de dimension $n \geq 10^2$ (dans le cas de l'assurance d'une vitesse augmentée pour un choix de l'approximation initiale adéquate). Le tronquement de la rangée $\{x^{(k)}\}$ a lieu à un indice m ainsi que $x^{(m)}$ constitue une approximation satisfaisante de la solution cherchée x^* (par exemple $\|x^{(m)} - x^*\| < \varepsilon$, où ε>0 est l'erreur admise).

4. La méthode d'élimination de Gauss

La méthode d'élimination de Gauss consiste en amener le système initial à un système équivalent ayant la matrice des coefficients supérieur triangulaire. La transformation du système donné dans un système de forme triangulaire sans modifier la solution du système se réalise à l'aide de trois opérations fondamentales :

1) Réarranger les équations (changer deux équations entre elles) ;
2) Multiplier une équation avec une constante (différente de zéro) ;
3) Soustraire une équation d'une autre et remplacer la seconde par le résultat de la soustraction.

Nous illustrons par des exemples cette méthode pour le système des équations linéaire suivant :

$$\begin{cases} 2x_1 + x_2 + x_3 = 1, \\ 4x_1 + x_2 = -2, \\ -2x_1 + 2x_2 + x_3 = 7. \end{cases}$$

On peut éliminer l'inconnue x_1 des dernières équations en multipliant la première avec les facteurs :

$$\mu_{21} = \frac{a_{21}}{a_{11}} = \frac{4}{2} = 2, \qquad \mu_{31} = \frac{a_{31}}{a_{11}} = \frac{-2}{2} = -1$$

et en soustrayant la première équation de la deuxième et puis de la troisième, nous obtenons le système équivalent suivant :

$$\begin{cases} 2x_1 + x_2 + x_3 = 1, \end{cases}$$

Le coefficient $a_{11} = 2$ de la première équation s'appelle *l'élément pivot* de la première étape d'élimination, et la ligne qui lui correspond s'appelle *la ligne pivot.*

De la même façon, nous pouvons éliminer l'inconnue, x_2 de la dernière équation.

Pour la deuxième étape *l'élément pivot* est $a_{22}^{'} = -1$. On multiplie la deuxième équation avec $\mu_{32} = \frac{a_{32}^{'}}{a_{22}^{'}} = \frac{3}{-1} = -3$ et on la soustrait de la troisième équation. Par conséquent, on obtient le système de forme triangulaire :

$$\begin{cases} 2x_1 + x_2 + x_3 = 1, \\ \quad -x_2 - 2x_3 = -4, \\ \quad\quad -4x_3 = -4. \end{cases}$$

Par la suite, on détermine les inconnues en commençant par la troisième équation : $x_3^* = 1$; en substituant le résultat obtenu dans la deuxième équation on obtient $x_2^* = 2$; à la fin de la première équation on a $x_1^* = -1$.

Généralisons cette méthode. Soit donné le système d'équation linéaires :

$$Ax=b\ ; \qquad (4.1)$$

où $A=(a_{ij})_n$, $x,b \in R^n$, $det\ A \neq 0$.

Supposons que $a_{11} \neq 0$ *; si* $a_{11} = 0$ l'élément non nul de la première colonne est amené à la place (1,1) permutant les équations respectives du système. La première étape consiste à éliminer l'inconnue x_1 des équations du système, en commençant par la deuxième, en multipliant la première équation avec le rapport :

et en soustrayant le résultat obtenu de l'équation *I* pour $\forall i \geq 2$.

On obtient de cette manière le système équivalent :

$$A^{(2)}\, x = b^{(2)}, \tag{4.2}$$

avec les coefficient :

$$\begin{aligned}
&a_{1j}^{(2)} = a_{1j}^{(1)}, \quad j = 1,2,\ldots,n;\\
&a_{i1}^{(2)} = 0, \qquad i = 2,3,\ldots,n;\\
&a_{ij}^{(2)} = a_{ij}^{(1)} - \mu_{i1} a_{1j}^{(1)}, \quad i,j = 2,3,\ldots,n;\\
&b_1^{(2)} = b_1^{(1)}, \quad b_i^{(2)} = b_i^{(1)} - \mu_{i1} b_1^{(1)}, \quad j = 2,3,\ldots,n.
\end{aligned}$$

On a noté plus haut $a_{ij}^{(1)} = a_{ij}; \quad i,j = 1,2,\ldots,n$ et $b_i^{(1)} = = b_i$; i,j=1,2,…,n. La première équation du système (4.2) coïncide avec la première du système (4.1).

Par la suite le procédé ci-dessus est répété pour l'élimination de l'inconnue x_2 du système (4.2) etc. À l'étape k on obtient le système

$$A^{(h)}x = b^{(h)}$$

où :

$$A^{(k)} = \begin{pmatrix}
a_{11}^{(1)} & a_{12}^{(1)} & \ldots & a_{1,k-1}^{(1)} & a_{1k}^{(1)} & \ldots & a_{1n}^{(1)} \\
0 & a_{22}^{(2)} & \ldots & a_{2,k-1}^{(2)} & a_{2k}^{(2)} & \ldots & a_{2n}^{(2)} \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & \ldots & a_{k-1,k-1}^{(k-1)} & a_{k-1,k}^{(k-1)} & \ldots & a_{k-1,n}^{(k-1)} \\
0 & 0 & \ldots & 0 & a_{kk}^{(k)} & \ldots & a_{kn}^{(k)} \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots
\end{pmatrix},$$

$$b^{(k)} == \begin{pmatrix} b_1^{(1)} \\ b_2^{(2)} \\ \vdots \\ b_{k-1}^{(k-1)} \\ b_k^{(k)} \\ \vdots \\ b_n^{(k)} \end{pmatrix}$$

On calcule les éléments $a_{ij}^{(k)}$ de $A^{(k)}$ et $b_i^{(k)}$ de $b^{(k)}$ de façon récursive par les formules :

$$a_{ij}^{(k)} = \begin{cases} a_{ij}^{(k-1)} & ,pour \quad i \le k-1, \\ 0 & ,pour \quad i \ge k, j \le k-1, \\ a_{ij}^{(k-1)} - \mu_{i,k-1} \cdot a_{k-1,j}^{(k-1)} & pour \quad i \ge k, j \ge k, \end{cases}$$

où

$$\mu_{i,k-1} = \frac{a_{i,k-1}^{(k-1)}}{a_{k-1,k-1}^{(k-1)}},$$

et

$$b_i^{(k)} = \begin{cases} b_i^{(k-1)} & ,pour \quad i \le k-1, \\ b_i^{(k-1)} - \mu_{i,k-1} \cdot b_{k-1}^{(k-1)} & ,pour \quad i \ge k. \end{cases}$$

Après *n* étapes, l'inconnue x_{n-1} est éliminée de la dernière équation et on obtient un système avec la matrice supérieure triangulaire :

$$\begin{cases} a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + \ldots + a_{1k}^{(1)}x_k + \ldots + a_{in}^{(1)}x_n = b_1^{(1)}, \\ a_{22}^{(2)}x_2 + \ldots + a_{2k}^{(2)}x_k + \ldots + a_{2n}^{(2)}x_n = b_2^{(2)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{kk}^{(k)}x_k + \ldots + a_{kn}^{(k)}x_n = b_k^{(k)} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{nn}^{(n)}x_n = b_n^{(n)}. \end{cases}$$

Ce système est résolu à partir de la dernière équation en utilisant *le processus d'élimination inverse* qui peut être décrit ainsi :

$$x_n = \frac{b_n^{(n)}}{a_{nn}^{(n)}}, \quad x_{n-1} = \frac{b_{n-1}^{(n-1)} - a_{n-1,n}^{(n-1)}x_n}{a_{n-1,n-1}^{(n-1)}},$$

$$x_k = \frac{b_k^{(k)} - \sum_{j=k+1}^{n} a_{kj}^{(k)}x_j}{a_{kk}^{(k)}}, \quad k = n-2, n-3, \ldots, 2, 1.$$

La méthode d'élimination gaussienne présentée ci-dessus suppose que les éléments *pivot* doivent être non nuls. Si, lors de l'étape *k*, *k* l'élément *pivot* $a_{kk}^{(k)} = 0$, alors au moins un des autres éléments de la colonne *k* et des lignes *k + 1*, *k + 2*, ..., *n* est non nul; sinon la matrice A serait singulière *(det(A)=0).*

En permutant les équations du système nous pouvons amener sur le lieu *(k, k)* l'élément né nul et donc est possible reprise de l'élimination.

Considérons un exemple simple :

$$\begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \end{pmatrix}$$

équation. En changeant les équations les unes avec les autres, on obtient :

$$\begin{cases} 2x_1 + x_2 = 4, \\ \quad 3x_2 = 0, \end{cases}$$

un système en forme triangulaire, qui est résolu immédiatement par l'élimination inverse $x_2^* = 0,\ x_1^* = 2$.
Analysons un autre exemple :

$$\begin{cases} 0.000100x_1 + x_2 = 1, \\ \quad x_1 + x_2 = 2 \end{cases}$$

avec la solution exacte $x_1^* = 1.00010, \quad x_2^* = 0.99990$. On utilise une arithmétique en virgule flottante avec $\beta = 10$ et $t = 3$: seuls trois chiffres décimaux significatifs sont conservés dans les calculs et nous supposons que le résultat est arrondi correctement. En appliquant la méthode d'élimination gaussienne, on obtient le système :

$$\begin{cases} 0{,}000100x_1 + x_2 = 1, \\ \quad -10000x_2 = -10000. \end{cases}$$

De la dernière équation, on a $x_2^* = 1.000$ qui remplacé dans la première équation nous donne $x_1^* = 0.000$, ce qui est évidemment un résultat faux. ***Une catastrophe informatique est survenue !***

En changeant les équations entre elles, on a le système :

$$\begin{cases} \quad x_1 + x_2 = 2, \\ 0{,}000100x_1 + x_2 = 1. \end{cases}$$

et la méthode de l'élimination du Gauss le transforme en :

avec la solution $x_1^* = x_2^* = 1.00$.

Par conséquent, si un élément pivot est exactement égal à zéro, voire proche de zéro, pour des raisons de stabilité numérique, on doit réorganiser les équations.

Il existe deux stratégies pour choisir l'élément pivot afin d'éviter que l'influence des erreurs d'arrondi ne devienne catastrophique. La première stratégie s'appelle *pivotement partiel* et consiste en ce qui suit : à l'étape *k*, le pivot est égal au premier élément maximum dans le mode dans la colonne *k*, sous-diagonal de $A^{(k)}$:

$$\left|a_{rk}^{(k)}\right| = \max_{k \le i \le n} \left|a_{ik}^{(k)}\right|$$

et les lignes *k* et *r* sont permutées.

Une autre stratégie de permutation consiste en un *pivotement complet (total);* les lignes *k* et *r (r≥k)* et les colonnes *k* et *s*, *(s≥k)* sont modifiées de la sorte que le pivot obtenu après la permutation coïncide avec le premier élément maximum dans le mode sous-matrice définie par les *n-k* dernières lignes et colonnes de $A^{(k)}$:

$$\left|a_{rs}^{(k)}\right| = \max_{k \le i,j \le n} \left|a_{ij}^{(k)}\right|.$$

Une matrice *A* est appelée *diagonale dominante* si

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \ne i}}^{n} \left|a_{ij}\right|, \quad i = 1,2, \dots, n.$$

Soit une matrice *A* symétrique et diagonale dominante. Après la première étape d'élimination gaussienne, la matrice $A^{(2)}$ devient :

Où la sous-matrice $\overline{A}_1$ est également en diagonale dominante. On peut démontrer que le processus d'élimination dans le cas des matrices diagonales dominantes ne dépend pas du choix de l'élément *pivot.* Il n'est pas nécessaire le pivotement et dans le cas quand la matrice *A* est positive définie.

Nous pouvons estimer le nombre d'opérations arithmétiques dans la méthode d'élimination gaussienne.

La procédure d'élimination de l'inconnue x_1 demande $n(n-1) = n^2 - n$ opérations arithmétiques. L'élimination de l'inconnue x_k demande $k(k-1) = k^2 - k$ opérations. Par la suite, la procédure directe demande le nombre suivant des opérations arithmétiques :

$$N_1 = (n^2 - n) + \cdots + (k^2 - k) + \cdots + (1^2 - 1) =$$
$$= \sum_{k=1}^{n} k^2 - \sum_{k=1}^{n} k = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2} = \frac{n^2 - n}{3}.$$

Le processus d'élimination inverse est effectué beaucoup plus rapidement. L'inconnue x_n se trouve à l'aide d'une seule opération (division à l'élément *pivot*) ; le calcul x_{n-1} nécessite deux opérations (division - soustraction puis division), etc. Le pas *k* demande seulement *k* opérations. Par suite l'élimination inverse demande :

$$N_2 = \sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

d'opérations arithmétique.

De nombreuses années, on a pensé que la méthode

À l'heure actuelle, on connait la méthode dans laquelle le nombre d'opérations est réduit à Cn^{α} ($2 < \alpha < 3$). Cette méthode se base sur un résultat remarquable obtenu en 1971 *A. Schonhage* et *V. Strassen*, qui ont montrés que théoriquement, la multiplication peut avoir une complexité à peine supérieure à l'addition. Pour montrer que la méthode de Gauss n'est pas optimale examinons l'algorithme de Strassen de multiplication de deux matrices. Soit par exemple :

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

L'algorithme du Strassen se base sur l'identité matricielle :

$$A \times B = \begin{pmatrix} C & D \\ E & F \end{pmatrix},$$

où

$$C = (a_{11} + a_{22})(b_{11} + b_{22}) + a_{22}(-b_{11} + b_{21}) - (a_{11} + a_{12})b_{22} + \\ +(a_{12} - a_{22})(b_{21} + b_{22}),$$

$$F = (a_{11} + a_{22})(b_{11} + b_{22}) + a_{11}(b_{12} - b_{22}) - (a_{21} + a_{22})b_{11} + \\ +(-a_{11} + a_{21})(b_{11} + b_{12}),$$

$$D = a_{11}(b_{12} - b_{22}) + (a_{11} + a_{12})b_{22},$$

$$E = (a_{21} + a_{22})b_{11} + a_{22}(-b_{11} + b_{21}).$$

Ainsi, pour obtenir le produit de deux matrices, il suffit d'effectuer sept multiplications et 18 additions. Si nous multiplions les matrices de manière traditionnelle, nous aurons huit multiplications. Dans l'exemple en haut ne sont pas concrétisés les éléments des matrices *A* et *B* et la propriété de commutativité du produit n'a pas été utilisée. Par suite les éléments a_{ij}, b_{ij} peuvent être considérés matrices et nous avons une procedure de multiplication

matrices à la dimension 2^m, alors en base de l'identité de plus haut, nous avons :

$$N(2^m) = 7 \cdot N(2^{m-1}) + 18 \cdot 2^{2m-2}.$$

Étant donné que *N(2)=7+18*, la dernière relation implique

$$N(2^m) = 7^m + 6(7^m + 4^m)$$

ou

$$N(n) = n^{log_2 7} + 6(n^{log_2 7} + n^{log_2 4}) .$$

Car $\alpha = log_2 7 \approx 2.81 < 3$, l'algorithme de Strassen pour un n suffisamment grand est plus avantageux que la procédure de multiplication matricielle habituelle.

5. Factorisation LU

Prenons exemple numérique du chapitre 4 résolution du système $Ax = b$, avec

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ -2 \\ 7 \end{pmatrix}.$$

Suite au processus d'élimination de l'inconnu x_1, on obtient le système $A^{(2)}x = b^{(2)}$, où

$$A^{(2)} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 3 & 2 \end{pmatrix}, \quad b^{(2)} = \begin{pmatrix} 1 \\ -4 \\ 8 \end{pmatrix}.$$

Notons par M_1 la matrice inférieure triangulaire :

$$M_1 = \begin{pmatrix} 1 & 0 & 0 \\ -\mu_{21} & 1 & 0 \\ -\mu_{31} & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

qu'on obtient de la matrice unité par le remplacement des éléments sous-diagonale de la première colonne avec les multiplicateurs $-\mu_{21}, -\mu_{31}$. Il est facile de vérifier que :

$$M_1 A = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 3 & 2 \end{pmatrix} = A^{(2)},$$

$$M_1 b = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \\ \end{pmatrix} = \begin{pmatrix} 1 \\ -4 \end{pmatrix} = b^{(2)}.$$

Dans l'étape finale de transformation, l'inconnu x_2 serait éliminé de la dernière équation, et on obtient un système sous forme triangulaire $Ux = c$, où

$$U = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{pmatrix}, c = \begin{pmatrix} 1 \\ -4 \\ -4 \end{pmatrix}.$$

On observe que $U = M_2 A^{(2)}$ et $c = M_2 b^{(2)}$, où

$$M_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\mu_{32} & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}.$$

Par conséquence, le procès de transformation à système $Ax = b$ dans un système équivalent de forme triangulaire $Ux = c$ peut être représenté comme multiplication au système initial successif au matrice M_1, M_2

$$M_2 M_1 Ax = M_2 M_1 b.$$

La relation $M_2 M_1 A = U$ permet de prendre une autre interprétation de la méthode de Gauss. En multipliant cette relation à gauche avec la matrice inférieur triangulaire $L = M_1^{-1} M_2^{-1}$on obtient:

$$A = LU$$

Donc, à l'aide de la méthode d'élimination de Gauss la matrice A se décompose en produit de deux facteurs L et U, où L est une matrice inférieur triangulaire, et U est une matrice supérieur

Nous allons montrer que pour toute matrice non singulière il existe une *«factorisation LU»* équivalente à la méthode d'élimination gaussienne. Supposons tout d'abord que la matrice *A* est telle que l'élimination peut se faire sans permutations de lignes ou de colonnes.

Soit

$$m_k = \begin{pmatrix} \overbrace{\begin{matrix}0\\0\\\vdots\\0\end{matrix}}^{k} \\ \mu_{k+1,k} \\ \vdots \\ \mu_{nk} \end{pmatrix}, \quad e_k = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow \text{composante } k$$

où les multiplicateurs $\mu_{ik}, i = k+1, k+2, \ldots n$ sont ceux utilisés dans l'étape $k+1$ pour l'élimination de l'inconnue x_{k+1} (voir le chapitre 4)

Nous défions une matrice M_k ainsi

$$M_k = I - m_k e_k^T.$$

Cette matrice diffère de matrice unité I seulement par des éléments sous diagonale non-nulles de la colonne k:

$$M_k = \begin{pmatrix} 1 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & \ldots & 0 \\ 0 & 0 & \ldots & -\mu_{k+1,k} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & \ldots & 1 \end{pmatrix}$$

La méthode d'élimination du Gauss consiste (voir ch. 4) en détermination du rang de matrice $A = A^{(1)}, A^{(2)}, \dots, A^{(n)}$. Il est facile de voir que

$$A^{(k+1)} = M_k \cdot M_{k-1} \cdot \dots \cdot M_1 A, k = 1,2, \dots n-1.$$

En écrivant cette relation pour $k = n-1$ et désignant $U = A^{(n)}$, on obtient :

$$M_{n-1} \cdot M_{n-2} \cdot \dots \cdot M_2 \cdot M_1 \cdot A = U, ,$$

ou

$$A = M_1^{-1} \cdot M_2^{-1} \cdot \dots \cdot M_{n-1}^{-1} \cdot U..$$

On vérifie immédiatement que

$$M_k^{-1} = I + m_k e_k^T..$$

Nous déduisons de ce qui précède

$$A = L \cdot U, ,$$

ou

$$L = M_1^{-1} \cdot M_2^{-1} \cdot \dots \cdot M_{n-1}^{-1} = I + \sum_{k=1} m_k e_k^T.$$

Par conséquent, la méthode d'élimination gaussienne calcule une factorisation *LU* de la matrice *A*, où

$$L = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ \mu_{21} & 1 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mu_{k1} & \mu_{k2} & \dots. & 1 & \dots & 0 \\ \mu_{k+1,1} & \mu_{k+1,2} & \dots & \mu_{k+1,k} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mu_{n1} & \mu_{n2} & \dots & \mu_{nk} & \dots & 1 \end{pmatrix},$$

$$U = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \dots & a_{2n}^{(2)} \\ \dots. & \dots\dots & \dots & \dots \\ \vdots & \vdots & \dots. & \vdots \\ 0 & 0 & \dots & a_{nn}^{(n)} \end{pmatrix}.$$

Comme indiqué au chapitre 4, pour des raisons de stabilité, il est conseillé d'utiliser une stratégie de pivotement partiel. On montre que toute matrice supporte une factorisation *LU*, éventuellement en effectuant les permutations qui se produisent par pivotement partiel sur les lignes. En d'autres termes, il existe une matrice de permutation *P* telle que

$$PA = LU.$$

La méthode d'élimination gaussienne et la factorisation *LU* sont équivalentes. Une factorisation *LU* de *A* peut être calculée directement, grâce à une procédure compacte appelée *factorisation Crout*. Cette factorisation impose *U* à la diagonale unitaire et est plus avantageuse sur les ordinateurs qui permettent le calcul rapide des produits scalaires. Pour une introduction plus approfondie à la

$$Ly = b,\ Ux = y.$$

Tout d'abord on résout le système inférieur triangulaire $Ly = b$ par une procédure typique de *substitution « directe »* commençant par la première équation :

$$y_1 = \frac{b_1}{l_{11}},\ \ y_i = \frac{b_i - \sum_{k-1}^{i-1} l_{ik} y_k}{l_{ii}}, i = 3,3,\ldots,n.$$

Ici l_{ik} sont les éléments de la matrice L. Ensuite, le système triangulaire supérieur $Ux = y$ est résolu par la procédure de substitution « en arrière », en commençant par la dernière équation:

$$x_n = \frac{y_{nn}}{u_{nn}},\ \ x_i = \frac{y_i - \sum_{k-i+1}^{n} u_{ki} x_k}{u_{ii}}, i = n-1, n-2,\ldots,1,$$

où u_{ik} sont les éléments du matrice U.

En utilisant une factorisation LU de A, nous pouvons résoudre plusieurs systèmes d'équations simultanément, ayant la même matrice A, sans reprendre les calculs depuis le début.

La méthode d'élimination gaussienne permet de calculer le déterminant de la matrice A.

En effet, on observe que

$$det(A) = det(L) \times det(U).$$

Car $det(L) = 1$, on a :

Si l'une des stratégies de pivot est appliquée, alors :

$$det(A) = (-1)^m a_{11}^{(1)} \cdot a_{22}^{(2)} \cdot \ldots \cdot a_{nm}^{(n)}$$

où m est nombre total de permutation effectuer.

6. Factorisation Cholesky

Soit $A = (a_{ij})_{n\times n}$ une matrice symétrique et positive définie :

$$(Ax, x) > 0, \quad \forall x \in R^n, x \neq 0.$$

Nous montrerons dans la factorisation LU de A on peut choisir $U = L^T$.

La décomposition

$$A = L \cdot L^T$$

s'appelle *Factorisation de Cholesky*.

Théorème. Si matrice la A est symétrique et positive définie alors il existe une matrice inférieur triangulaire L, avec les éléments diagonaux positifs, unique, ainsi que $A = L \cdot L^T$.

Démonstration. Nous allons prouver le théorème par induction sur n.

Pour $n = 2$ nous avons :

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad L = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix}$$

Si on forme le produit $L \cdot L^T$ et si on l'identifie avec A_0 nous obtenons :

$$a_{11} = l_{11}^2, \qquad a_{12} = l_{11}l_{21},$$
$$a_{21} = l_{21}l_{11}, \qquad a_{22} = l_{21}^2 + l_{22}^2.$$

Car la matrice A est positive définie, les éléments sur la

de l'équation qui contenant a_{21} ($a_{21} = a_{12}$ car A est une matrice symétrique): $l_{21} = \frac{a_{21}}{\sqrt{a_{11}}}$.

Enfin, l'élément l_{22} se détermine ainsi :

$$l_{22} = \sqrt{a_{22} - l_{21}^2} = \sqrt{a_{22} - \frac{a_{12}^2}{a_{11}}} .$$

Montrons que l'extraction de la racine carrée ci-dessus est possible. En effet, soit le vecteur x avec les composantes a_{12} et $-a_{11}$

$$x = \begin{pmatrix} a_{12} \\ -a_{11} \end{pmatrix}$$

Alors $(Ax, x) = x^T Ax = a_{11}^2 a_{22} - a_{12}^2 a_{11} > 0$ parce que la matrice A est positive définie. En divisant la dernière inégalité à $a_{11}^2 > 0$ nous obtenons :

$$a_{22} - \frac{a_{12}^2}{a_{11}} > 0 .$$

Par conséquence pour $n = 2$ la décomposition $A = L \cdot L^T$ existe et est unique. Supposons que le théorème soit vrai pour $n = k - 1$: où A et L sont matrices de dimension $(k-1) \times (k-1)$. Soit A' et L' ainsi :

$$A' = \begin{pmatrix} A & y \\ y^T & a_{kk} \end{pmatrix}, \quad L' = \begin{pmatrix} L & 0 \\ w^T & l_{kk} \end{pmatrix}$$

On forme le produit $L' \cdot (L')^T$ et on l'identifie avec A'. Alors on obtient :

$$A = LL^T, \qquad y = Lw,$$
$$y^T = w^T L^T, \quad a_{kk} = l_{kk}^2 + w^T w.$$

Par l'hypothèse d'induction mathématique, la matrice L est déterminée en mode unique ainsi que $A = LL^T$. D'ici il résulte que le vecteur w est aussi unique et se calcule comme solution du système $y = Lw$. Puis, l'élément l_{kk} se précise de la formule :

$$l_{kk} = \sqrt{a_{kk} - w^T w}\,.$$

Montrons, aussi comme dans le cas des matrices de dimension 2×2, que l'expression sous la racine est positive. En qualité de vecteur x nous prenons :

$$x = \begin{pmatrix} A^{-1}y \\ -1 \end{pmatrix}.$$

Notons $z = A^{-1}y$; alors

$$(Ax, x) = x^T Ax = z^T Az - 2z^T y + a_{kk} =$$
$$= -z^T y + a_{kk} = a_{kk} - y^T A^{-1} y = a_{kk} - y^T (LL^T)^{-1} y =$$
$$= a_{kk} - (L^{-1}y)^T (L^{-1}y) = a_{kk} - w^T w > 0.$$

Donc la matrice L' avec la propriété $A' = L' \cdot (L')^T$ est uniquement déterminé et **le théorème est prouvé**.

On peut montrer que la factorisation Cholesky d'une matrice $A = (a_{ij})_{n\times n}$ symétrique définie positive nécessite approximativement $n^3/6$ opérations (multiplications et additions) et

La méthode de Cholesky pour résoudre des systèmes d'équations linéaires est également appelée *méthode de la racine carrée* et consiste à décomposer le système $Ax = b$ en systèmes triangulaires :

$$L^T y = b, \quad Lx = y.$$

Les éléments l_{ij} de la matrice inférieure triangulaire L peut être calculés de la façon suivante: on détermine la première colonne de la matrice L

$$l_{11} = \sqrt{a_{11}}, \quad l_{i1} = \frac{a_{i1}}{l_{11}}, \quad i = 2,3,\ldots,n;$$

Après avoir obtenu les premières $(k-1)$ colonnes de la matrice *L*, la colonne *k* est calculée

$$l_{kk} = \sqrt{a_{kk} - \sum_{j=1}^{k-1} l_{kj}^2},$$

$$l_{ik} = \frac{1}{l_{kk}} \left(a_{ik} - \sum_{j=1}^{k-1} l_{ij} l_{kj} \right), \quad i = k+1,\ldots,n.$$

Une caractéristique remarquable de l'algorithme de Cholesky est sa stabilité numérique. Cela résulte du fait que l'élément maximal dans le mode d'une matrice symétrique et positivement définie est situé sur la diagonale principale. De plus, les éléments diagonaux de la matrice *A* et les éléments l_{ij} de la matrice *L* satisfont la relation :

$$l_{1k}^2 + l_{2k}^2 + \ldots + l_{kk}^2 = a_{kk}, \quad k = 1,2,\ldots,n.$$

7. Perturbations. Nombre de conditionnement

Considérons le système linéaire $Ax = b$. La solution exacte du système considéré peut-être écrite sous la forme : $x^* = A^{-1}b$. Supposons que la matrice non-singulière A et le vecteur b souffrent les perturbations δA et δb.

Premièrement analysons le cas quand nous perturbons seulement le terme libre. La solution perturbée $\overline{x}$ du système avec la partie droite $b + \delta b$ satisfait l'égalité :

$$A\overline{x} = b + \delta b.$$

Nous obtenons :

$$\overline{x} - x^* = A^{-1}\delta b.$$

d'où il résulte que :

$$|\overline{x} - x^*| \leq |A^{-1}||\delta b| \qquad (7.1)$$

n'importe quelle est la forme matricielle subordonnée à une norme vectorielle. Pour chaque A et δ il existe une perturbation δb qui satisfait l`égalité (7.1). Par conséquence $\left\|A^{-1}\right\|$ évalue combien peut croître l'erreur fournie par δb.

Pour déterminer l'effet relatif de la perturbation δb, nous observons que

$$\|b\| = \|Ax^*\| \leq \|A\|\|x^*\|,$$

où :

Tenant compte de cela, l'inégalité (7.1) implique

$$\frac{\|\overline{x}-x^*\|}{\|x^*\|} \leq \|A\|\|A^{-1}\| \frac{\|\delta b\|}{\|b\|} \qquad (7.2)$$

Supposons que nous perturbons les éléments de A : dans ce cas la solution perturbée $\overline{x}$ sera l`égalité :

$$(A + \delta A)\overline{x} = b.$$

Il résulte que :

$$\overline{x} - x^* = A^{-1}\delta A\tilde{x}$$

et nous obtenons l'estimation :

$$\|\overline{x} - x^*\| \leq \|A^{-1}\|\|\delta A\|\|\overline{x}\|. \qquad (7.3)$$

Cette inégalité peut être mise sous la forme :

$$\frac{\|\overline{x}-x^*\|}{\|\overline{x}\|} \leq \|A\|\|A^{-1}\| \frac{\|\delta A\|}{\|A\|} \qquad (7.4)$$

Nous observons que dans (7.2) et (7.4) le nombre $\|A\|\|A^{-1}\|$ estime l'erreur relative dans la solution fournie par δb ou δA. Ce nombre est nommée *nombre de conditionnement* de la matrice A par rapport avec la norme matricielle considérée et est noté par :

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

Car pour chaque norme matricielle subordonnée à une

et par conséquence $cond(A) \geq 1$ pour chaque matrice A.

Le nombre de conditionnement caractérise l`effet maximal des perturbations δb et δA à la résolution du système : $Ax = b$. Si le nombre de conditionnement *cond (A)* est grand, des petites perturbations de A et b produisent des grandes perturbations de x^* : dans ce cas on dit que la matrice A est *mal conditionnée*. Les matrices avec le nombre de conditionnement « *petit* » sont nommées *bien conditionnées*.

Soulignons que la dimension de la matrice n'a pas une influence directe sous son nombre de conditionnement : si $A = I$ ou bien $A = \frac{1}{10} I$ on obtient $cond(A) = 1$. Pour comparaison le déterminant de la matrice n'est pas un indice adéquat du conditionnement parce que la valeur du déterminant dépend aussi de la dimension n de la matrice. Si A est une matrice presque singulière, cela ne signifie pas qu'elle est *mal conditionnée*. Dans l'exemple $A = \frac{1}{10} I$ nous avons $\det(A) = 10^{-n}$; cette matrice « presque singulière » est maximalement bien conditionnée.

Exemple : Considérons le système d'équations $Ax = b$, où

$$A = \begin{pmatrix} 1 & -1 & -1 & \dots & -1 & -1 \\ 0 & 1 & -1 & \dots & -1 & -1 \\ 0 & 0 & 1 & \dots & -1 & -1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \quad b = \begin{pmatrix} -1 \\ -1 \\ -1 \\ \vdots \\ -1 \\ 1 \end{pmatrix}$$

Mentionnons que $det(A) = 1 \neq 0$. Le système considéré dans la forme développée est :

$$\begin{cases} x_1 - x_2 - x_3 - \cdots - x_n = -1, \\ \quad x_2 - x_3 - \cdots - x_n = -1, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \quad x_{n-1} - x_n = -1, \\ \quad x_n = 1. \end{cases} \tag{7.5}$$

Le système d'équations (7.5) admet une solution exacte unique $x_1^* = x_2^* = \cdots = x_{n-1}^* = 0, x_n^* = 1$. Suposons que nous perturbons le vecteur libre b avec $\delta b = (0,0,\ldots,0,\varepsilon)^T$. Dans ce cas la solution exacte du système perturbée $A\overline{x} = b + \delta b$ devient $\overline{x} = = x^* + r$ et l'erreur $r = (r_1, r_2, \ldots, r_n)^T$ satisfait les équations :

$$\begin{cases} r_1 - r_2 - r_3 - \cdots - r_n = 0, \\ \quad r_2 - r_3 - \cdots - r_n = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \quad r_{n-1} - r_n = 0, \\ \quad r_n = \varepsilon. \end{cases}$$

D'où il résulte que :

$$\begin{aligned} &r_n = \varepsilon, \\ &r_{n-1} = r_n = \varepsilon, \\ &r_{n-2} = r_n + r_{n-1} = 2\varepsilon, \\ &r_{n-3} = r_n + r_{n-1} + r_{n-2} = 4\varepsilon = 2^2\varepsilon, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &r_{n-k} = r_n + r_{n-1} + \cdots + r_{n-(k+1)} = 2^{k-1}\varepsilon, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &r_1 = r_{n-(n-1)} = 2^{(n-1)-1}\varepsilon = 2^{n-2}\varepsilon. \end{aligned}$$

Ainsi : $\overline{x}_i = 2^{n-i-1}\varepsilon, \quad i = 1,2,\ldots,n-1; \quad \overline{x}_n = 1 + \varepsilon.$

Estimons *cond(A)*, prenant en qualité de norme $\|.\|_\infty$ Nous obtenons :

$$cond(A) = \|A\|_\infty \cdot \|A^{-1}\|_\infty \geq \frac{\|r\|_\infty}{\|x^*\|_\infty} \cdot \frac{\|b\|_\infty}{\|\delta b\|_\infty} = 2^{n-2}.$$

Car $\|A\|_\infty = n$, il résulte que la norme de la matrice inverse est assez grande même si $det(A^{-1}) = \frac{1}{det(A)} = 1$. Par exemple, pour $n = 102$, nous obtenons :

$$\|A\|_\infty = 102, \quad cond(A) \geq 2^{100} > 10^{30}, \quad \text{et } \|A^{-1}\|_\infty > 10^{27}$$

Particulièrement, si $\varepsilon = 10^{-15}$, (une erreure assez petite) nous obtenons $\left\|\overline{x} - x^*\right\|_\infty = \|r\|_\infty > 10^{15}$: une petite perturbation du terme libre a produit une perturbation très grande dans la solution !

La valeur du nombre de conditionnement d'une matrice dépend de la norme matricielle utilisée. Supposons que A est une matrice symétrique et positivement définie avec les valeurs propres positives : $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Analogiquement montrons que pour des telles matrices le nombre de conditionnement est :

$$cond(A) = \frac{\lambda_n}{\lambda_1} = \frac{\lambda_{\max}}{\lambda_{\min}}$$

***Exemple* :** les valeurs propres de la matrice

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1.0001 \end{pmatrix}$$

approximativement sont égales avec : $\lambda_1 = \frac{1}{2}10^{-4}$ et $\lambda_2 = 2$. Voilà pourquoi le nombre de conditionnement $cond(A)$ est

changements dans la solution. Vraiment, soit les systèmes d'équations : $Ax = b$ et $Ax = b + \delta b$où

$$b = \begin{pmatrix} 2 \\ 2.0001 \end{pmatrix}, \delta b = \begin{pmatrix} 0 \\ 0.0001 \end{pmatrix}$$

La solution change de$x^* = (1 \quad 1)^T$ jusqu'a $\overline{x} = (0 \quad 2)^T$;

$$\frac{\left\|x^* - \overline{x}\right\|_2}{\left\|x^*\right\|_2} = \frac{\left\|\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\|_2}{\left\|\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\|_2} = 1, \ \frac{\left\|\delta b\right\|_2}{\left\|b\right\|_2} = \frac{\left\|\begin{pmatrix} 0 \\ 0.0001 \end{pmatrix}\right\|_2}{\left\|\begin{pmatrix} 2 \\ 2.0001 \end{pmatrix}\right\|_2} = \frac{10^{-4}}{2\sqrt{2}}$$

Soit donné un système d'équations linéaires. La représentation avec la virgule mobile des éléments A et B dans l'ordinateur n'est pas exacte. Par conséquence, dans la mémoire de la machine nous avons le système $\overline{A}x = \overline{b}$ ou $\overline{A}$ et $\overline{b}$ sont les arrondis correspondants. Il existe une matrice de perturbation P et D (D est une matrice diagonale) tels que :

$$\overline{A} = A(I + P), \overline{b} = (I + D)b$$

Si nous notons avec ε_M l'unité d'arrondi de la machine, alors $\|P\| \leq \varepsilon_M$ et $\|D\| \leq \varepsilon_M$. Ainsi on obtient:

$$\left\|\delta A\right\| = \left\|\overline{A} - A\right\| \leq \varepsilon_M \left\|A\right\|, \ \left\|\delta b\right\| = \left\|\overline{b} - b\right\| \leq \varepsilon_M \left\|b\right\|$$

Des inégalités de ci-dessus, il résulte que les erreurs de d'arrondi produisent une perturbation estimée par

Soulignons que la détermination du nombre de conditionnement est un problème difficile, parce qu'il contient le calcul du $\|A^{-1}\|$. Le calcul de la matrice inverse et du sa norme nécessite approximativement $n^3 + 2n^2$opérations supplémentaires qui accroit de quatre fois les opérations nécessaires pour résoudre le système $Ax = b$. Un procédé pratique de calcul approximative de $\|A^{-1}\|$est suivant : Observons que si $w = A^{-1}y$. Alors

$$\|w\| \leq \|A^{-1}\|\|y\|.$$

Et par conséquence

$$\|A^{-1}\| \geq \frac{\|w\|}{\|y\|}.$$

Par conséquence nous pouvons choisir *k* vecteurs y_i, après nous résoudrons le système d'équations $Aw_i = y_i$, $i = 1,2,\ldots,k$, et mettons :

$$\left\|A^{-1}\right\| \approx \max_{1\leq i\leq k} \frac{\left\|w_i\right\|}{\left\|y_i\right\|}$$

8. Les calculs des valeurs et des vecteurs propres

8.1 Formulation du problème. Propriétés fondamentales

Soit A une matrice de dimension $n \times n$. Le nombre λ (réel ou complexe) s'appelle *valeur propre* de la matrice A s'il existe un vecteur $x \in R^n$, ainsi que

$$Ax = \lambda x \tag{8.1}$$

Le vecteur $x \neq 0$ s'appelle *vecteur prope* de A associé à la *valeur prope A.*

L'équation (8.1) peut être écrite sous la forme $Ax - \lambda x = 0$ ou

$$(A - \lambda I)x = O, \tag{8.2}$$

où

$$\lambda I = \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ .. & ... & .. & .. \\ .0 & 0 & 0 & \lambda \end{pmatrix}, O = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

L'équation (8.2) peut être écrite détaillée ainsi :

$$\begin{pmatrix} a_{11} - \lambda & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} - \lambda & ... & a_{21} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a - \lambda_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\det(A-\lambda I)=0. \tag{8.3}$$

Ce déterminant est un polynôme de degré *n* avec des coefficients réels et il est note d'habitude :

$$P_n(\lambda)=(-1)^n\lambda^n+p_1\lambda^{n-1}+\ldots+p_{n-1}\lambda+p_n$$

Le polynôme $P_n(\lambda)$ s'appelle *polynôme caractéristique* associe à la matrice *A*, et l'équation (3) s'appelle *l'équation caractéristique* de la matrice *A*.

L'équation caractéristique est une équation algébrique de degré *n* avec des coefficients réels qui en vertu du théorème fondamental de l'algèbre a exactement *n* racines $\lambda_1,\lambda_2,\ldots,\lambda_n$, en général complexes et pas absolument différentes.

La multitude des valeurs caractéristiques de la matrice *A* s'appelle *le spectre* de A et il est note par $\sigma(A)$:

$$\sigma(A)=(\lambda_1,\lambda_2,\ldots,\lambda_n).$$

Le rayon spectral de *A* est note $\rho(A)$ et il est défini par la relation :

$$\rho(A)=\max_{\lambda\in\sigma(A)}|\lambda|.$$

C'est pourquoi la norme matricielle $\|A\|_2=\sqrt{\rho(A^TA)}$ subordonnée à la norme euclidienne $\|x\|_2=\sqrt{\sum {x_t}^2}$ est nommée encore *norme spectrale*.

Plus que cela :

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2$$

où $\|A\|_F$ est la norme de Frobenius (voir la définition dans le compartiment 2.2).

Si A est une matrice symétrique, c'est-à-dire $A = A^T$, alors

$$\|A\|_2 = \max_{\lambda \in \sigma(A)} |\lambda| = \rho(A)$$

Et

$$\|A\|_F = \sqrt{\sum_{i=1}^{n} \lambda_i^2}$$

Si on connait une valeur propre, alors un vecteur propre associe à cette valeur propre est la solution pas nulle du système homogène. (8.2). De l'autre cote si on connait un vecteur propre v alors $Av = \lambda v$, d'où $\lambda(v, v) = (Av, v)$, donc la valeur propre correspondante est obtenue immédiatement :

$$\lambda = \frac{(Av, v)}{(v, v)}.$$

Exemple. Il faut calculer les valeurs propres et les vecteurs propres pour la matrice :

$$A = \begin{pmatrix} 4 & -5 \end{pmatrix}$$

$$P_2(\lambda) = \det(A - \lambda I) = \begin{vmatrix} 4-\lambda & -5 \\ 2 & -3-\lambda \end{vmatrix} =$$
$$= (4-\lambda)(-3-\lambda) + 10 = \lambda^2 - \lambda - 2$$

Pour le calcul des valeurs propres il faut résoudre l'équation caractéristique :

$$P_2(\lambda) = \lambda^2 - \lambda - 2 = 0.$$

On obtient : $\lambda_1 = -1, \lambda_2 = 2$. Par suite la matrice A a deux valeurs propres distinctes. En remplaçant chaque valeur propre dans le système homogène (8.2) on obtient le vecteur propre associé à la valeur propre. Pour $\lambda_1 = -1$ on a:

$$(A - \lambda_1 I)x = \begin{pmatrix} 5 & -5 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

d'où on obtient le vecteur propre :

$$v^1 = c \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad c = const \neq 0.$$

Analogiquement pour $\lambda_2 = 2$ on aura:

$$(A - \lambda_2 I)x = \begin{pmatrix} 2 & -5 \\ 2 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad v^2 = c \begin{pmatrix} 5 \\ 2 \end{pmatrix}, \quad c \neq 0$$

Dans l'exemple considéré le spectre $\sigma(A) = (-1, -2)$ et le

n'importe quel vecteur cx, où c est un scalaire, sera de même un vecteur propre.

On donne quelques résultats remarquables relatif aux valeurs propres des certaines matrices spéciales.

1. Les valeurs propres des matrices symétriques sont réelles.
2. Les valeurs propres des matrices définies positives sont positives.
3. Les valeurs propres des matrices diagonales, inferieur triangulaires et supérieur triangulaires coïncident avec les éléments sur la diagonale principale.

Les propriétés 1-3 sont démontrées immédiatement, en résultant de la définition des valeurs et des vecteurs propres.

4. Deux matrices semblables ont les mêmes valeurs propres.

Démonstration. Soit les matrices A et B sont semblables. Cela signifie qu'il existe une matrice inversible M telle que

$$B = M^{-1}AM.$$

Si $Ax = \lambda x$ alors $M^{-1}Ax = \lambda M^{-1}x$. Nous allons noter $x = My$. Il résulte que :

$$M^{-1}AM = \lambda y \text{ et } By = \lambda y.$$

Par conséquence, pour les matrices semblables les valeurs propres coïncident. De plus, les vecteurs propres sont liés par la relation $x = My$.

5. La somme des valeurs propres d'une matrice A est égale à la somme des éléments sur la diagonale principale :

6. Le produit des valeurs propres de la matrice coïncide avec le déterminant de la matrice :

$$\lambda_{1.} \cdot \lambda_2 \cdot \ldots \cdot \lambda_n = \det(A).$$

7. Si $\lambda_i, 1 \le i \le n$ sont les valeurs propres de la matrice A alors $\lambda_i^k, 1 \le i \le n$ sont les valeurs propres de la matrice $A^k = A. \cdot A \cdot \ldots \cdot A$

Notons

$$Q_n(\lambda) = (-1)^n P_n(\lambda) = \lambda^n + q_1 \lambda^{n-1} + \ldots + q_{n-1}\lambda + q_n$$

Le polynôme $Q_n(\lambda)$ s'appelle *polynôme propre.*

8. *Les formules de Newton* :

$$\mu_k + \sum_{i=1}^{k-1} q_i \mu_{k-1} = -kq_k, \quad k = 1,2,\ldots,n$$

où

$$\mu_k = \sum_{i=1}^{n} \lambda_i^k, \quad k = 1,2,\ldots,n$$

et $\lambda_1, \lambda_2, \ldots, \lambda_n$ sont les racines de l'équation caractéristique.

On observe que $\mu_k = Tr(A^k)$.. La somme $\sum_{i=1}^{n} \lambda_i^k$ il est aussi appelé *moment d'ordre k* des valeurs propres.

9. ***L'identité de Cauley-Hamilton***. Toute matrice carrée *A* est une racine de son polynôme caractéristique :

10. ***Le théorème sur les cercles de Gershgorin***. Toute valeur propre λ de la matrice $A = (a_{ij})_{n\times n}$ se trouve dans le plan complexe, dans la réunion des cercles :

$$\bigcup_{i=1}^{n}\left\{ z : |z - a_{ii}| \le r_i : r_i = \sum_{\substack{j=i \\ j\neq i}}^{n} |a_{ij}| \right\}$$

11. ***Théorème (Rayleigh-Ritz).*** Soit A une matrice symétrique et les valeurs propres de A sont placées en ordre croissant

$$\lambda_{\min} = \lambda_1 \le \lambda_2 \le \ldots \le \lambda_{n-1} \le \lambda_n = \lambda_{\max}$$

Alors

$$\lambda_1 \|x\|_2^2 \le (Ax, x) \le \lambda_n \|x\|_2^2, \quad \forall x \in R^n,$$

$$\lambda_1 = \lambda_{\min} = \min_{x\neq 0} \frac{(Ax, x)}{(x, x)} = \min_{\|x\|_2=1} (Ax, x),$$

$$\lambda_n = \lambda_{\max} = \max_{x\neq 0} \frac{(Ax, x)}{(x, x)} = \max_{\|x\|_2=1} (Ax, x).$$

L'expression $\dfrac{(Ax, x)}{(x, x)}$ s'appelle *le quotient de Rayleigh*.

Les méthodes de calcul des valeurs propres se devisent en deux groupes :

1) des méthodes qui déterminent d'abord les coefficients du polynôme caractéristique, puis résolvent l'équation caractéristique ;
2) méthodes qui déterminent les valeurs propres et les vecteurs

Leverrier, la méthode de Fadéev, la méthode de Lanozos etc. (voir, par exemple [12,16,18,27]). Soulignons que ces méthodes sont recommandables seulement dans les cas où la matrice *A* et de petit ordre et les racines de l'équation caractéristique sont bien séparées. La raison en est que le nombre d'opérations arithmétiques pour déterminer les coefficients du polynôme caractéristique est très grand (par exemple, la méthode de Fadéev exige n^4 opérations). D'autre part, les coefficients sont obtenus avec des erreurs d'arrondi inhérentes qui peuvent conduire à de grandes variations des racines, car le problème de la résolution des équations algébriques est mal conditionné. Soit, par exemple, la matrice *A* de forme [1,4]

$$A(\varepsilon)=\begin{pmatrix} \alpha & 1 & 0 & \dots & 0 & 0 \\ 0 & \alpha & 1 & \dots & 0 & 0 \\ 0 & 0 & \alpha & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \varepsilon & 0 & 0 & \dots & 0 & \alpha \end{pmatrix}$$

où ε est une petite perturbation et α est la valeur propre de multiplicité *n* de la non perturbée *A(0)*. Le polynôme caractéristique de $A(\varepsilon)$ est

$$P_n(\varepsilon,\lambda)=(\lambda-a)^n-(-1)^n\varepsilon$$

On observe que, par exemple, si *n=10*, α *=0* et $\varepsilon=10^{-10}$ alors, pratiquement, les matrices *A(0)* et $A(10^{-10})$ représentent la même matrice dans la mémoire de la machine électronique de calcul, en même temps que le polynôme caractéristique $P_{10}(10^{-10},\lambda)$

8.2. Méthodes basées sur des transformations de ressemblance orthogonale.

Les méthodes pratiques de détermination des valeurs et des vecteurs propres résident en procédés itératifs qui ramènent la matrice considérée à la forme canonique Schur à l'aide des transformations de ressemblance orthogonale.

Deux matrices A et B s'appellent *orthogonalement similaires* s'il existe une matrice orthogonale ainsi que

$$B = Q^T AQ$$

La transformation $Q^T AQ$ de A est appelée *transformation de ressemblance orthogonale*.

Les matrices A et $Q^T AQ$ ont les mêmes valeurs propres, parce que $Q^T = Q^{-1}$ (voir la propriété 4 de 8.1.). Si x est le vecteur propre pour la matrice A, alors $Q^T x$ est vecteur propre pour la matrice $Q^T AQ$.

Théorème (Schur). Quelle que soit la matrice $A = (a_{ij})_n$, il existe une matrice orthogonale Q de dimension $n \times n$ ainsi que :

$$Q^T AQ = S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1k} \\ 0 & s_{22} & \dots & s_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & s_{11} \end{pmatrix}$$

Les blocs s_{ii} de dimension 1×1 représentent les valeurs propres réelles de la matrice *A*, et chaque bloc de dimension 2×2 représente les valeurs propres ensemble conjuguées.

La démonstration de ce théorème peut être trouvée dans [1,18].

La matrice *S* s'appelle *forme canonique Schur réelle* de *A*.

Si la matrice *A* a seulement des valeurs propres réelles, alors la matrice *S* est supérieur triangulaire. Si *A* a encore des valeurs propres complexes alors *S* est *quasi-triangulaire*.

La mise en place d'une forme canonique Shur est précédée d'une préparation initiale de la matrice A, l'amenant à la forme compacte appelée *la forme Hessenberg*.

$$A=\begin{pmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1,n-1} & h_{1n} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2,n-1} & h_{2n} \\ 0 & h_{32} & h_{33} & \dots & h_{3,n-1} & h_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & h_{n-1,n-1} & h_{n-1,n} \\ 0 & 0 & 0 & \dots & h_{n,n-1} & h_{nn} \end{pmatrix}$$

Pour les matrices symétriques, la forme de Hessenberg devient une forme tridiagonale symétrique.

8.2.1 La méthode de Householder

La réduction de chaque matrice *A* à la forme de Householder peut être faite après un nombre limité de rotations

$$H = I - 2\frac{vv^T}{\|v\|_2^2}$$

s'appelle *réflecteur* ou *la transformation de Householder*.

Soit $z = (1,0,\dots,0)^T \in R^n$, $\sigma = \|x\|_2$et $v = x + \sigma z$.
Alors

$$Hx = -\sigma \rightleftarrows z = (-\sigma, 0, \dots, 0)^T.$$

Vraiment,

$$H = x - 2\frac{vv^T x}{\|v\|_2^2} = x - (x + \sigma z)\frac{2(x + \sigma z)^T x}{(x + \sigma z)^T (x + \sigma z)} =$$
$$= x - (x + \sigma z) = -\sigma z, ,$$

vu que $x^T x = \sigma^2$.

La procédure directe pour l'ajout de la matrice *A* à la forme Householder est constitué de *n-2* étapes et est basé sur l'égalité démontré plus haut. Dans la première étape, la matrice est déterminée de sorte que les *n-2* derniers éléments de la première colonne de la matrice soient nuls. Pour cela, nous notons :

$$x = \begin{pmatrix} a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad z = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad Hx = \begin{pmatrix} -\sigma \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Ici *H* est la matrice de taille du Householder $(n-1)\times(n-1)$. Le calcul se fait selon la relation :

$$U_1 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & H & \\ 0 & & & \end{pmatrix} = U_1^T = U_1^{-1}.$$

Parce que l'unité est située dans le coin « en haut à gauche », la multiplication de la matrice U_1 avec la matrice A ne modifie pas la ligne de A, mais modifie les derniers $n-1$ éléments des colonnes suivantes $j = 2,3,\dots,n$. Après avoir obtenu la matrice $U_1 A$, on calcule la matrice $(U_1 A)U_1$. La multiplication de la matrice $U_1 A$ avec U_1 ne modifie pas la première colonne de $U_1 A$, ainsi que la matrice:

$$U_1^{-1} A U_1 = \begin{pmatrix} a_{11} & * & *\dots* \\ -\sigma & * & *\dots* \\ 0 & * & *\dots* \\ \vdots & \vdots & \vdots \quad \vdots \\ 0 & * & *\dots* \end{pmatrix}$$

à la forme Hessenberg dans la première colonne. La première étape est finie.

La deuxième étape est la même que la première : on prend x égal avec le vecteur de $n-2$ éléments de la deuxième colonne de la matrice $U_1^{-1} A U_1$, z un vecteur unité de la dimension correspondante et H_2 sera une matrice de la dimension $(n-2)\times(n-2)$. Analogiquement on obtient :

$$U_2 = U_2^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & H_2 & \\ 0 & 0 & & & \end{pmatrix} \text{ et } U_2(U_1AU_1)U_2 = \begin{pmatrix} * & * & * & \dots & * \\ * & * & * & \dots & * \\ 0 & * & * & \dots & * \\ 0 & 0 & * & \dots & * \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & * & \dots & * \end{pmatrix}$$

Ainsi la matrice :

$$U_{n-2}U_{n-3} \dots U_1AU_1U_2 \dots U_{n-3}U_{n-2}$$

obtenue à la fin, après avoir traversé les $n-2$ étapes de la procédure, est la forme de Hessenberg. Le nombre d'opérations arithmétiques requises est égal à $5n^3/3$.

Si nous appliquons les transformations décrites à une matrice symétrique, alors toutes les matrices $U_kU_{k-1} \dots U_1AU_1U_2 \dots U_{k-1}U_k$ seront également symétriques et finalement on obtient une matrice de Hessenberg tridiagonal symétrique. Dans ce cas il ne faut que $2n^3/3$ opérations arithmétiques.

Exemple. Considérons la matrice [14]:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Nous aurons :

$$x = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad v = x + \sigma z = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$H = I - 2\frac{vv^T}{v^Tv} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad HAH = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

$$U = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \quad U^{-1}AU = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

8.2.2. L'algorithme QR

L'algorithme QR construit itérativement une suite de matrices orthogonales $A_0 = A, A_1, A_2, \ldots$ qui converge vers la forme canonique Schur de la matrice *A*. Par conséquent, cet algorithme résout le problème complet des valeurs propres.

Soit *A* une matrice de dimension $n \times n$. Notons $A_0 = A$ et considérons une factorisation *QR* de la matrice A_0 (voir le compartiment 10.3):

$$A_0 = Q_0 R_0.$$

Q_0 est une matrice orthogonale et R_0 est une matrice supérieur triangulaire. La matrice suivante A_1 est obtenue en multipliant les facteurs Q_0, R_0 dans l'ordre inverse :

$$A_1 = R_0 Q_0.$$

Les matrices A_0 et A_1 sont orthogonales similaires :

$$Q_0^{-1} A_0 Q_0 = Q_0^{-1}(Q_0 R_0) Q_0 = R_0 Q_0 = A_1$$

En général, la chaîne de matrice définie de façon récurrente $A_0, A_1, \ldots,$ est calculée par les formules :

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$A_k = Q_k R_k,\ A_{k+1} = R_k Q_k$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..$$

On peut démontrer (voir, par exemple, [1, 7, 8, 12]) que dans des conditions générales, la chaîne $\{A_k\}_{k=0}^{\infty}$ converge et la matrice limite $A_\infty = S$ coïncide à la forme canonique Schur de *A*. Plus que cela, cette matrice limite A_∞ a sur la diagonale les valeurs caractéristiques de A.

Si certains rapports $\frac{\lambda_{m+1}}{\lambda_m}$ sont près de la valeur 1, alors la convergence est lente. En ce cas l'algorithme QR est modifié ainsi

$$A_k - \alpha_k I = Q_k R_k, \quad A_{k+1} = R_k Q_k + \alpha_k I, \quad k = 0,1, \ldots$$

où α_k est un paramètre d'accélération de la convergence nommé *déplacement.* Car

$$Q_k^{-1} A_k Q_k = Q_k^{-1}(Q_k R_k + \alpha_k I) Q_k = R_k Q_k + \alpha_k I = A_{k+1}$$

chaque pas dans l'algorithme QR avec des déplacements est une transformation de similitude orthogonale.

Si les déplacements α_k sont élus près d'une valeur propre λ de *A*, on aura une convergence rapide de la chaîne $\{A_k\}$ vers la forme canonique Schur.

Une itération de l'algorithme *QR* exige $\frac{4n^3}{3}$ opérations, ce qui est exagéré. En pratique il est mieux d'amener initialement la matrice considérée à la forme Hessenberg à l'aide de la méthode de

des matrices asymétriques et à environ $12n$ dans le cas des matrices symétriques.

8.3. La méthode de la puissance

La méthode de la puissance est la plus simple méthode de détermination de la plus grande valeur propre (en modul) d'une matrice réelle A de dimension $n \times n$ et du vecteur propre associé valeur propre dominante.

Soit A une matrice simple. La matrice A de dimension $n \times n$ s'appelle *simple* si elle a exactement n vecteurs propres linéaires indépendants $e_1, e_2, \ldots, e_n$. Ces vecteurs forment une base de l'espace n-dimensionnel et peuvent être choisi de telle sorte que:

$$\|\boldsymbol{e_i}\|_2 = \sqrt{(\boldsymbol{e_i}, \boldsymbol{e_i})} = \mathbf{1}, \quad \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{n}.$$

Par exemple, toute matrice symétrique est une matrice *simple*. Supposons que la matrice A admet une valeur propre réelle dominante λ_1,
c'est-à-dire:

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$$

Prenons un vecteur initial arbitraire $x^{(0)}$ non nul. Parce que les vecteurs propres $e_1, e_2, \ldots, e_n$ forment une base, il existe les scalaires $c_1, c_2, \ldots, c_n$ pas tous nuls, tel que :

$$x^{(0)} = c_1 e_1 + c_2 e_2 + \cdots + c_n e_n$$

Supposons que $c_1 \neq 0_i$; en cas contraire peut être élu un autre vecteur initial $x^{(0)}$ de sorte que le coefficient correspondant $c_1 \neq 0.$

Formons la rangée des vecteurs :

$$x^{(k)} = Ax^{(k-1)} = A^k x^{(0)}, \quad k = 1,2, \ldots$$

Car $Ae_i = \lambda_i e_i$ on peut écrire:

$$\begin{aligned} x^{(1)} = Ax^{(0)} = A\left(c_1 e_1 + c_2 e_2 + \ldots c_n e_n\right) = \\ = c_1 Ae_1 + c_2 Ae_2 + \ldots + c_n Ae_n = \\ = c_1 \lambda_1 e_1 + c_2 \lambda_2 e_2 + \ldots + c_n \lambda_n e_n. \end{aligned}$$

En général la chaine $\left\{x^{(k)}\right\}$ a la représentation :

$$x^{(k)} = c_1 \lambda_1^k e_1 + c_2 \lambda_2^k e_2 + \ldots + c_n \lambda_n^k e_n = \lambda_1^k \left(c_1 e_1 + \eta^{(k)}\right)$$

Où

$$\eta^{(k)} = c_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k e_2 + \ldots + c_n \left(\frac{\lambda_n}{\lambda_1}\right)^k e_n$$

Par hypothèse $\left|\frac{\lambda_i}{\lambda_1}\right| < 1$ pour $i \geq 2$ donc $\left\|\eta^{(k)}\right\|_2 = 0\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$ et

$$\lim_{k \to \infty} \left\|\eta^{(k)}\right\|_2 = 0 \cdot$$

$$\left(Ax^{(k-1)}, x^{(k-1)}\right) = \left(x^{(k)}, x^{(k-1)}\right) = \lambda_1^{2k-1}\left(c_1 e_1 + \eta^{(k)}, c_1 e_1 + \eta^{(k-1)}\right)$$
$$=$$
$$= \lambda_1^{2k-1}\left[c_1^2(e_1, e_1) + c_1\left(e_1, \eta^{(k-1)}\right) + c_1\left(\eta^{(k)}, e_1\right) + \left(\eta^{(k)}, \eta^{(k-1)}\right)\right].$$

Car $(e_1, e_1) = \|e_1\|_2^2 = 1$ et

$$\left|\left(e_1, \eta^{(k-1)}\right)\right| \le \|e_1\|_2 \left\|\eta^{(k-1)}\right\|_2 = \left\|\eta^{(k-1)}\right\|_2,$$
$$\left|\left(\eta^{(k)}, e_1\right)\right| \le \left\|\eta^{(k)}\right\|_2 \|e_1\|_2 = \left\|\eta^{(k)}\right\|_2,$$
$$\left|\left(\eta^{(k)}, \eta^{(k-1)}\right)\right| \le \left\|\eta^{(k)}\right\|_2 \left\|\eta^{(k-1)}\right\|_2,$$

on en déduit que

$$\left(Ax^{(k-1)}, x^{(k-1)}\right) = \lambda_1^{2k-1}\left[c_1^2 + o\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{k-1}\right)\right]..$$

Analogiquement on obtient :

$$\left(x^{(k-1)}, x^{(k-1)}\right) = \lambda_1^{2k-2}\left[c_1^2 + o\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{k-1}\right)\right]..$$

Donc, résulte que le quotient de Rayleigh :

$$\frac{\left(Ax^{(k-1)}, x^{(k-1)}\right)}{\left(x^{(k-1)}, x^{(k-1)}\right)} = \lambda_1 + o\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{k-1}\right).$$

aura tendance à sa propre valeur (maximum en mode) λ_1.
On voit que

Il en résulte que pour $|\lambda_1| > 1$ nous avons $\|x^{(k)}\|_2 \to \infty$, mais si $|\lambda_1| < 1$ alors $\|x^{(k)}\|_2 \to 0$. En résolvant le problème avec la machine de calcul électronique, dans le premier cas, nous pouvons avoir l'apparence d'un signal de dépassement, après quoi, les calculs sont interrompus. Dans le deuxième cas $\|x^{(k)}\|_2$ peut devenir un zéro-machine. Par conséquent, dans la pratique, il est nécessaire de changer le facteur d'échelle de la chaîne $\{x^{(k)}\} = \{Ax^{(k-1)}\}$; au lieu de la suite $\{x^{(k)}\}$ une autre chaîne vectorielle est formée $\{z^{(k)}\}$ avec $\|z^{(k)}\|_2 = 1$qui est construite de la façon suivante :

$$z^{(k)} = \frac{Az^{(k-1)}}{\|Az^{(k-1)}\|}, \quad k = 1,2, \ldots,$$

d'où

$$\frac{(Az^{(k-1)}, z^{(k-1)})}{(z^{(k-1)}, z^{(k-1)})} = (z^{(k)}, z^{(k-1)}) \to \lambda_1.$$

Exemple. Calculer la valeur propre maximale λ_1 pour la matrice :

$$A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix}$$

Utilisons comme vecteur de départ sur $x^{(0)} = (0, 0, 1)^T$. Nous organisons les calculs dans un schéma facile à suivre :

A	$x^{(0)}$	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$	…	$x^{(\infty)}$
1 –1 –1 -1 2 0 -1 0 2	0 0 1	-1 0 2	-3 1 5	-9 3 13	-25 15 35	-75 45 95	… … …	$-\infty$ $+\infty$ $+\infty$
$z^{(k)}=\frac{x^{(k)}}{\left\|x^{(k)}\right\|_2}$	0 0 1	-1/√5 0 2/√5	-3/√35 1/√35 5/√35	-9/√259 3/√25 13/√259			… … …	-1/√3 1/√3 1/√3
$\frac{\left(x^{k}\right)_i}{\left(x^{(k-1)}\right)_i}$	- - 2	3 - 2.5	3 3 2.6	2.8 3 2.69	3 3 2.71	… … …	3 3 3

Dans le diagramme ci-dessus, dans l'avant-dernier tableau, nous avons les vecteurs $z^{(k)}=\frac{x^{(k)}}{\left\|x^{(k)}\right\|_2}$ et dans le dernier – le quotient $\frac{\left(x^{(k)}\right)_i}{\left(x^{(k-1)}\right)_i}$, où par $\left(x^{(k)}\right)_i$ on a noté la composante i du vecteur $\left(x^{(k)}\right)$. Les valeurs propres de la matrice considérée sont : $\lambda_1=3, \lambda_2=2$ et $\lambda_3=0$. Le quotient $\frac{\left(x^{(k)}\right)_i}{\left(x^{(k-1)}\right)_i}$ tend à la valeur maximale λ_1 et la chaîne de vecteurs $\left\{z^{(k)}\right\}$ converge au vecteur propre $\left(\frac{-1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)^T$ qui correspond à cette valeur propre. La méthode de la puissance peut également être utilisée pour calculer la valeur propre minimale (en mode) λ_n avec la condition qu'on a $|\lambda_n|<|\lambda_{n-1}|$. En effet, si la matrice A est non singulière, alors $A^{-1}x=\lambda^{-1}x$; et sa valeur propre

$$z^{(k)} = \frac{A^{-1} z^{(k-1)}}{\left\| A z^{(k-1)} \right\|}, \quad k = 1,2,\dots .$$

Cette méthode est appelée *méthode de puissance inverse.* Nous soulignons qu'en pratique dans la méthode de puissance inverse le vecteur $z^{(k)}$ est déterminé après avoir résolu le système d'équations linéaires

$$A z^{(k)} = \frac{z^{(k-1)}}{\left\| A z^{(k-1)} \right\|}.$$

Une autre méthode beaucoup plus efficace est *la méthode de puissance inverse avec déplacements.* On observe que la matrice $A - \alpha I$ a les valeurs propres $\lambda_i - \alpha$ et les mêmes vecteurs propres que *A.* Par conséquent, si les déplacements α_m sont choisis assez proche de la valeur propre de λ, la chaîne proches à valeur propre λ de A_1 alors la chaine $\left\{ z^{(k)} \right\}$ défini par:

$$\left(A - \alpha_m I \right) z^{(k)} = \beta_k z^{(k-1)}$$

converge vers la valeur propre e_1 à un rythme assez rapide. Le paramètre β_k est un facteur de normalisation, choisi ainsi que $\left\| z^{(k)} \right\|_2 = 1$. Si *A* est inversible et si nous prenons $\alpha_m = 0$, nous obtenons: $A z^{(k)} = \beta_k z^{(k-1)}$- la méthode de la puissance inverse. En pratique pour accélérer la convergence de la chaîne $\left\{ z^{(k)} \right\}$ vers le

$$\alpha_m = \frac{\left(Az^{(m)}, z^{(m)}\right)}{\left(z^{(m)}, z^{(m)}\right)} \ .$$

On peut démontrer que $\{\alpha_m\}$ converge vers λ_1 extrêmement rapide (avec vitesse de convergence cubique).

Ici, nous concluons la discussion sur les méthodes de calcul des valeurs propres et des vecteurs propres. Soulignons une fois de plus que l'algorithme QR est l'une des méthodes les plus remarquables des mathématiques appliquées. Pour le lecteur qui veut d'approfondir ses connaissances dans ce domaine nous recommandons le travaux (1, 2, 3, 4, 8, 11).

9. Les méthodes itératives de résolution de systèmes des équations linéaires

Considérons le système d'équations linéaires :

$$Ax = b \tag{9.1}$$

où $b \in R^n$ et A est une matrice non singulière ($det(A) \neq 0$) de taille $n \times n$. La méthode d'élimination gaussienne pour résoudre le système linéaire (9.1) nécessite au moins $n^3/3$ opérations arithmétiques et tant que ce nombre d'opérations est acceptable, nous pouvons utiliser cette méthode. D'un autre côté, lorsque $n^3/3$ est élevé, à l'aide de méthodes itératives, une approximation satisfaisante de la solution peut être obtenue après avoir effectué un nombre beaucoup plus petit d'opérations arithmétiques.

Les méthodes itératives sont construites en utilisant le dépliage de la matrice A définie par

$$A = S - T.$$

Alors, le système (9.1) est équivalent avec le suivant :

$$Sx = Tx + b, \tag{9.2}$$

or

$$x = Qx + d, \tag{9.3}$$

où $Q = S^{-1}T$, $d = S^{-1}b$. Par conséquent, nous pouvons construire la suite $\{x^{(k)}\}$, en utilisant la relation récurrente :

or

$$x^{(k+1)} = Qx^{(k)} + d, \quad k = 0,1,2,\ldots, \tag{9.5}$$

où $x^{(0)} \in R^n$ est une approximation initiale de la solution $x*$.

Afin de réduire le système (9.1) pour former (9.2) ou (9.3), adapté pour l'itération, la dissolution de la matrice *A* doit satisfaire aux conditions suivantes :

- Le système (9.4) a une solution unique $x^{(k+1)}$ et est facile à résoudre. C'est pourquoi la matrice *S* est choisie sous une forme simple et est inversible. Elle peut être diagonale ou triangulaire.
- La suite $\left\{x^{(k)}\right\}_{k=0}^{\infty}$ converge vers la solution exacte $x*$ quel que soit $x^{(0)} \in R^n$.

Car $x* = Qx^* + d$ on a

$$x^{(k+1)} - x* = Q\left(x^{(k)} - x*\right), \quad k = 0,1,2, \ldots,$$

d'où il résulte que:

$$x^{(k+1)} - x* = Q^k\left(x^{(0)} - x*\right).$$

Évidemment $\lim\limits_{k\to\infty} x^{(k)} = x*$ si et seulement si

$$\lim_{k\to\infty} Q^k = O$$

Au cours de l'algèbre, il est prouvé que $Q^k \underset{k\to\infty}{\to} O$ si et seulement si le rayon spectral de *Q* est inférieur à l'unité:

vers x^*, il est encore plus grand, autant que le rayon spectral $\rho(Q)$ est plus petit. Il s'ensuit que le théorème suivant est vrai.

Théorème 1. (Condition de convergence nécessaire et suffisante). La suite $\{x^{(k)}\}$ définie par (9.5) converge vers la solution x^* de système unique (9.3) pour toute approximation initiale $x^{(0)} \in R^n$ si et seulement si

$$\rho(Q) = \max_{\lambda \in \sigma(Q)} |\lambda| < 1 \tag{9.6}$$

En pratique, nous ne connaissons pas les valeurs propres de *Q*, donc le théorème 1 est difficile à utiliser. Le théorème suivant est utilisé à la place du théorème 1.

Théorème 2. (Condition de convergence suffisante). S'il existe une norme matricielle subordonnée à une norme vectorielle telle que $\|Q\| \leq q < 1$, alors le système (9.3) a une solution unique x^*, la suite $\{x^{(k)}\}$ définie par (9.5) converge vers x^* quelle que soit l'approximation initiale $x^{(0)} \in R^n$ et l'erreur est évaluée par :

$$\|x^{(k)} - x*\| \leq \frac{q}{1-q}\|x^{(k)} - x^{(k-1)}\| \leq \frac{q^k}{1-q}\|x^{(1)} - x^{(0)}\|$$

Démonstration. La condition $\|Q\| < 1$ implique la condition (9.6) (voir paragraphe 2.2). Pour obtenir une estimation de l'erreur, nous utilisons la relation :

$$x* - x^{(k-1)} = x^{(k)} - x^{(k-1)} + Q\left(x^* - x^{(k-1)}\right)$$

Alors

$$\|x* - x^{(k-1)}\| \leq \|x^{(k)} - x^{(k-1)}\| + \|Q\|\|x^* - x^{(k-1)}\|$$

Car par hypothèse $1 - \|Q\| \geq 1 - q > 0$ on a

$$\left\|x* - x^{(k-1)}\right\| \leq \frac{1}{1-q}\left\|x^{(k)} - x^{(k-1)}\right\|.$$

D'autre part

$$\left\|x* - x^{(k)}\right\| = \left\|Q(x* - x^{(k-1)})\right\| \leq \|Q\|\left\|x* - x^{(k-1)}\right\|$$

Donc

$$\left\|x* - x^{(k)}\right\| \leq \frac{q}{1-q}\left\|x^{(k)} - x^{(k-1)}\right\|.$$

Le théorème est prouvé.

Nous supposons que les éléments diagonaux $a_{ii} \neq 0, i - 1,2,\ldots,n$. Alors comme matrice S on peut prendre la matrice diagonale attachée à la matrice A :

$$S = Diag(a_{11}, a_{22}, \ldots, a_{nn})$$

Nous avons

$$S^{-1} = Diag(\frac{1}{a_{11}}, \frac{1}{a_{22}}, \quad ..... \quad , \frac{1}{a_{nn}})$$

Dans ce cas, le système (9.3) devient :

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

Le processus itératif (9.5) est défini par :

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij} x_j^{(k)} \right), \quad i = 1,2,...,n \tag{9.7}$$

On obtient ainsi une méthode de résolution du système linéaire (9.1) appelée *méthode de Jacobi.*

Exemple. Étant donné le système d'équations linéaires :

$$\left.\begin{aligned} 2x_1 - x_2 &= 1, \\ -x_1 + 2x_2 &= 1 \end{aligned}\right\} \tag{9.8}$$

Nous aurons :

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \quad S = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$Q = S^{-1}T = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \quad d = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

Les valeurs propres de la matrice Q sont : $\lambda_1 = -\frac{1}{2}, \quad \lambda_2 = \frac{1}{2}$. . Donc la méthode de Jacobi :

$$\left.\begin{aligned} x_1^{(k+1)} &= \frac{1}{2} x_2^{(k)} + \frac{1}{2}, \\ x_2^{(k+1)} &= \frac{1}{2} x_1^{(k)} + \frac{1}{2} \end{aligned}\right\}$$

converge vers la solution exacte $x^* = (1,1)^T$ quel que soit $x^{(0)} \in R^2$. En particulier, pour $x^{(0)} = (0,0)^T$ nous obtenons la suite de vectores:

$$x^{(1)} = \left(\frac{1}{2}, \frac{1}{2}\right)^T,\ x^{(2)} = \left(\frac{3}{4}, \frac{3}{4}\right)^T,\ x^{(3)} = \left(\frac{7}{8}, \frac{7}{8}\right)^T,\ x^{(4)} = \left(\frac{15}{16}, \frac{15}{16}\right)^T, \ldots$$

On remarque que pour la méthode de Jacobi la matrice *Q* a les éléments

$$q_{ij} = \begin{cases} 0, & \text{dacă} \quad i = j, \\ -\dfrac{a_{ij}}{a_{ii}}, & \text{dacă} \quad i \neq j \end{cases}$$

En utilisant le théorème 2 avec la norme $\|\bullet\|_\infty$, nous obtenons:

$$\|Q\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |q_{ij}| = \max_{1 \le i \le n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{|a_{ij}|}{|a_{ii}|} < 1$$

Il s'avère que si

$$|a_{ij}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \quad i = 1,2,\ldots,n$$

Dans la méthode de Jacobi, il est nécessaire de conserver dans la mémoire de l'ordinateur toutes les composantes du vecteur $x^{(k)}$ tant que le vecteur $x^{(k+1)}$ est calculé.

Nous pouvons modifier la méthode de Jacobi pour qu'à le pas $(k+1)$ nous utilisons dans le calcul de la composante $x_i^{(k+1)}$, les valeurs déjà calculées au même pas $x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_{i-1}^{(k+1)}$. Cette modification de la méthode de Jacobi est appelée *méthode de Gauss-Seidel*, et la suite itérative (9.7) devient :

$$x_i^{(k+1)} = \frac{1}{a_{ij}}\left(b_{ij} - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k)}\right) \quad \text{i+1,2,…,n}$$

Il est facile de voir que la méthode de Gauss-Seidel correspond au déroulement $A = S - T$ où S est la matrice sous-diagonale attachée à A, et T est la matrice strictement supra diagonale avec les éléments $-a_{ij}$ attachés à la même matrice A :

$$S = \begin{pmatrix} a_{11} & 0 & 0 & \ldots & 0 \\ a_{21} & a_{22} & 0 & \ldots & 0 \\ a_{31} & a_{32} & a_{33} & \ldots & 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots & & & & \\ a_{n1} & a_{n2} & a_{n3} & \ldots & a_{nn} \end{pmatrix}, \quad T = \begin{pmatrix} 0 & -a_{12} & -a_{13} & \ldots & -a_{1n} \\ 0 & 0 & -a_{23} & \ldots & -a_{2n} \\ 0 & 0 & 0 & \ldots & -a_{3n} \\ \ldots\ldots\ldots\ldots\ldots\ldots & & & & \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}.$$

Prenons l'exemple ci-dessus avec le système d'équations (9.8). Pour la méthode de Gauss-Seidel nous avons :

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \quad S = \begin{pmatrix} 2 & 0 \\ -1 & 2 \end{pmatrix}, \quad T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

$$Q = S^{-1}T = \begin{pmatrix} 0 & \frac{1}{2} \\ 0 & \frac{1}{4} \end{pmatrix}, \quad d = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

La suite Gauss-Seidel sera ainsi :

$$\left.\begin{aligned} x_1^{(k+1)} &= \frac{1}{2}x_2^{(k)} + \frac{1}{2}, \\ x_2^{(k+1)} &= \frac{1}{2}x_1^{(k+1)} + \frac{1}{2} \end{aligned}\right\}$$

ou

$$\begin{pmatrix} 2 & 0 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Pour l'approximation initiale $x^{(0)} = (0, 0)^T$ nous obtenons :

$$x^{(1)} = \left(\frac{1}{2}, \frac{1}{2}\right)^T, \; x^{(2)} = \left(\frac{3}{4}, \frac{7}{8}\right)^T, \; x^{(3)} = \left(\frac{15}{16}, \frac{31}{32}\right)^T, \ldots$$

On observe qu'une itération Gauss - Seidel ici équivaut à deux itérations de Jacobi, car les valeurs propres de la matrice Q sont 0 et $\frac{1}{4}$, donc le rayon spectral est $\rho\left(S^{-1}T\right) = \frac{1}{4}$.

Cela signifie que l'erreur dans chaque itération est divisée par 4 ; dans la méthode de Jacobi $\rho\left(S^{-1}T\right) = \frac{1}{2}$, l'erreur est divisée

général meilleure que la méthode de Jacobi. On peut montrer que si *A* est une matrice définie positivement, la méthode de Gauss-Seidel converge deux fois plus vite vers la solution que la méthode de Jacobi.

Il est également démontré ici (voir, par exemple, [15], p.181) que si la matrice *A* a diagonale dominante, alors la méthode de Gauss-Seidel converge. Il existe des modèles connus dans lesquels la méthode de Gauss-Seidel converge, et la méthode de Jacobi ne converge pas et vice versa. À titre d'illustration, nous annonçons le théorème suivant.

Théorème (Reich). Si la matrice *A* est symétrique et a les éléments diagonaux $a_{ii} > 0$ pour tout *i*, alors la méthode Gauss - Seidel converge si et seulement si *A* est une matrice définie positive.

On sait que la méthode de Jacobi ne converge pas toujours pour la matrice définie positivement *A*. Par exemple, si *A* est une matrice définie positive et n'est pas dominante en diagonale, alors il est possible que le rayon spectral $\rho\left(S^{-1}T\right) > 1$ pour la méthode de Jacobi.

Comme vu ci-dessus, si la matrice *A* est dominante en diagonale, alors la méthode de Jacobi et la méthode de Gauss-Seidel vont générer une série d'approximations successives qui convergent vers la solution exacte quelle que soit l'approximation initiale $x^{(0)}$.

Nous soulignons que cette condition est seulement suffisante et non nécessaire. Par exemple pour la matrice

$$A = \begin{pmatrix} 8 & 2 & 1 \\ 10 & 4 & 1 \\ 50 & 25 & 2 \end{pmatrix}$$

qui n'est pas dominante en diagonale, la méthode de Gauss-Seidel

successives. Soit $\overline{x}^{(k)}$ le vecteur obtenu à l'étape $k+1$ par la méthode de Gauss-Seidel. Méthode itérative définie par :

$$x^{(k+1)} = x^{(k)} + \omega\left(\overline{x}^{(k)} - x^{(k)}\right)$$

connue comme *la méthode de surrelaxation successive*. Le paramètre de relaxation ω est choisi de manière à augmenter le taux de convergence. Pour $\omega = 1$ la méthode est réduite à la méthode de Gauss - Seidel.

Il a été constaté que pour un choix approprié du paramètre ω, la convergence de la méthode de sur-relaxation successive est clairement supérieure aux méthodes de Jacobi et Gauss-Seidel. Par exemple, dans le cas du système d'équations linéaires (9.8), une itération par la méthode de surrelaxation successive équivaut (voir [2,8]) à 30 itérations par la méthode de Jacobi.

On peut montrer que $\omega \in (0, 2)$, généralement en pratique:

$$\omega \approx 1.8 \div 1.9$$

Il est montré [10] que la méthode de surtension successive converge pour toutes les matrices *A* symétriques définies positivement. Pour une compréhension plus approfondie des méthodes itératives, nous recommandons les références [8,10].

10. Systèmes linéaires surdéterminés et la méthode des plus petits carrés

10.1 Formulation du problème

Soit le système avec *m* équations et *n* inconnus

$$Ax = b \tag{10.1}$$

où *A* est une matrice de dimension $m \times n$, et $b \in R^m$ est un vecteur avec *m* composants. Si $m > n$ le système (10.1) est appelé *système surdéterminé*. Étant donné que le système (10.1) contient plus d'équations inconnues, nous ne pouvons pas, en général, trouver une solution qui permet de vérifier exactement toutes les équations du système. Par exemple, le système :

$$\begin{cases} 2x_1 = b_1, \\ 3x_1 = b_2, \\ 4x_1 = b_3, \end{cases}$$

aura la solution seulement dans le cas quand les termes libres b_1, b_2 et b_3 se trouvent dans le rapport *2:3:4.*

Bien que la plupart des systèmes surdéterminés ne soient pas compatibles, ils sont souvent rencontrés dans la pratique, par exemple en matière statistique. L'une des façons de résoudre des systèmes surdéterminés consiste à déterminer la pseudo-solution *x** qui minimise l'erreur moyenne pour toutes les équations du système.

Une pseudo-solution dans le sens des moindres carrés (MC) pour le système surdéterminé (10.1) est un vecteur $x * \in R^n$ avec la propriété :

Le vecteur x^* est également appelé *solution généralisée* au sens des MC. Ce vecteur x^* minimise la norme euclidienne du vecteur résiduel $r = Ax - b$, c'est-à-dire minimise l'écart quadratique de Ax par rapport à b :

$$\|r\|_2^2 = (r, r) = \sum_{i-1}^{n} r_i^2$$

D'ici dénomination de MC. Dans l'exemple antérieur

$$\|r\|_2^2 = (2x_1 - b_1)^2 + (3x_1 - b_2)^2 + (4x_1 - b_3)^2$$

Le problème est de déterminer un vecteur $x* \in R^n$ qui atteint l'expression minimale :

$$E(x) = \|Ax - b\|_2^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2 . \qquad (10.3)$$

Ce problème revient à la détermination de x^* avec la propriété (10.2)

10.2. Méthodes basées sur les systèmes normaux

La valeur minimale de la somme (10.3) est obtenue en annulant les dérivées partielles par rapport à $x_1, x_2, \ldots, x_n$, c'est-à-dire en annulant le gradient de la fonction $E(x)$:

Commente

$$A^T(Ax^* - b) = 0 \tag{10.4}$$

ou

$$A^T Ax^* = A^T b \tag{10.5}$$

Le système d'équations (10.5) est appelé le système normal associé au problème (10.1). Dans ce système, $C = A^T A$ c'est une matrice de dimension $n \times n$, symétrique avec les éléments :

$$c_{ij} = \overline{a}_i^T \overline{a}_j = (\overline{a}_i, \overline{a}_j)$$

où $\overline{a}_i = (a_{1i}, a_{2i}, \ldots, a_{mi})^T$ sont les vecteurs colonnes de la matrice A, $i = 1,2,\ldots,n$.

De toute évidence, la matrice $C = A^T A$ est positivement semi-définie car

$$(Cx, x) = x^T Cx = x^T A^T Ax = (Ax)^T Ax = \|Ax\|_2^2 \geq 0.$$

Si les colonne $\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_n$ de la matrice A sont linéaire indépendantes alors de $x \neq 0$ il résulte que $Ax \neq 0$, donc la matrice $C = A^T A$ est définie positive. Par conséquent il est vrai le théorème suivant :

Le théorème d'existence et unicité. Si la matrice A de dimension $m \times n$ a les colonnes linéaires indépendantes alors quel que soit le vecteur $b \in R^m$ le système (10.1) a une pseudo solution dans le sens MC unique $x* \in R^n$et

$$x* = (A^T A)^{-1} A^T b \tag{10.6}$$

$$\left.\begin{aligned} 2x_1 - x_2 &= 9, \\ x_1 + 4x_2 &= 0, \\ 3x_1 + x_2 &= -3. \end{aligned}\right\}$$

Nous avons

$$A = \begin{pmatrix} 2 & -1 \\ 1 & 4 \\ 3 & 1 \end{pmatrix}, \quad A^T = \begin{pmatrix} 2 & 1 & 3 \\ -1 & 4 & 1 \end{pmatrix},$$

$$A^T A = \begin{pmatrix} 14 & 5 \\ 5 & 18 \end{pmatrix}, \quad A^T b = \begin{pmatrix} 9 \\ -12 \end{pmatrix}$$

Le système normal associé au problème proposé devient :

$$\left.\begin{aligned} 14x_1 + 5x_2 &= 9, \\ 5x_1 + 18x_2 &= -12. \end{aligned}\right\}$$

Le système normal permet la détermination de la pseudo solution par les méthodes présentées aux paragraphes 10.3 - 10.6, 10.8. Car la matrice $A^T A$ est symétrique et définie positive, nous pouvons utiliser la factorisation de Cholesky. Pour le calcul de $A^T A$ et $A^T b$ sont nécessaires $m \cdot n(n+3)/2$ opérations arithmétiques pour son calcul, et la méthode de Cholesky pour résoudre des systèmes nécessite approximativement $n^3/3$ opérations. Ainsi, la plupart des efforts sont nécessaires pour former le système normal associé au problème (10.1).

$$cond(A^T A) = [cond(A)]^2$$

Il s'avère qu'en général la matrice $A^T A$ est mal conditionnée et son calcul est donc affecté par des erreurs d'arrondi avec un effet souvent catastrophique. Par conséquent, dans la pratique, la formation de systèmes normaux et leur résolution sont évitées. Il existe de bien meilleures façons de résoudre des systèmes surdéterminés au sens du MC. Ils sont basés sur la factorisation orthogonale de la matrice *A*.

10. 3. Méthodes d'orthogonalisation

Dans le cas où les colonnes $\overline{a}_i^T, i = 1,2,\ldots,n$ de la matrice *A* sont orthogonales nous pouvons facilement déterminer la pseudo solution du système surdéterminé (10.1). Vraiment, si $\overline{a}_i^T \overline{a}_j = 0, i \neq j$ la matrice $A^T A$ devient une matrice diagonale avec les éléments sur diagonale égales avec $\overline{a}_i^T \overline{a}_i \neq 0$ et immédiatement s'obtient la pseudo solution :

$$x_i^* = \frac{b^T \overline{a}_i}{\overline{a}_i^T \overline{a}_i}, . i = 1,2,\ldots\ldots,n.$$

Par conséquent, au lieu de former le système normal, nous pouvons orthogonaliser les colonnes de la matrice *A*. Une méthode d'orthogonalisation classique est *la méthode du Gram-Schmidt.* La suite de vecteurs linéaires indépendants $a_1, a_2, \ldots, a_n$ s'orthogonalise selon les formules :

Il est facile de voir que les vecteurs $v_1, v_2, \dots, v_n$ sont orthogonaux. En divisant chaque vecteur par sa longueur, nous obtenons une série de vecteurs orthonormés :

$$q_1 = \frac{v_1}{\|v_1\|_2}, \quad q_2 = \frac{v_2}{\|v_2\|_2}, \dots, q_n = \frac{v_n}{\|v_n\|_2}$$

Exemple : Soit les vecteurs :

$$a_1 = \begin{pmatrix} 1` \\ 1 \\ 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

Alors $v_1 = a_1$, mais v_2 se calcule conformément (10.7):

$$v_2 = a_2 - \frac{a_2^T v_1}{v_1^T v_1} = a_2 - \frac{1}{2} v_1 = \begin{pmatrix} 1/2 \\ -1/2 \\ 1 \end{pmatrix}$$

Les vecteurs ortho normaux sont :

$$q_1 = \frac{v_1}{\|v_1\|_2} = \sqrt{\frac{1}{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad q_2 = \frac{v_2}{\|v_2\|_2} = \sqrt{\frac{2}{3}} \begin{pmatrix} 1/2 \\ -1/2 \\ 1 \end{pmatrix}$$

Pour la méthode d'élimination gaussienne, un moyen pratique d'écrire le résultat est de factoriser la matrice $A = LU$.

Le processus d'orthonormation de Gram-Schmidt donne une autre factorisation pour la matrice *A*, appelée factorisation *QR*.

$$a_2 = -\frac{1}{2}v_1 + v_2, \quad \text{ou} \quad a_2 = \sqrt{\frac{1}{2}}q_1 + \sqrt{\frac{3}{2}}q_2$$

La représentation matricielle de ces deux équations est :

$$\begin{pmatrix} a_1 & a_2 \end{pmatrix} = \begin{pmatrix} q_1 & q_2 \end{pmatrix} \begin{pmatrix} \sqrt{2} & \sqrt{1/2} \\ 0 & \sqrt{3/2} \end{pmatrix}$$

c'est-à-dire $A = QR$ où Q a des colonnes orthogonales et R est supérieurement triangulaire.

On peut montrer que toute *matrice de dimension* $m \times n, m \geq n$ *, avec des colonnes linéairement indépendantes, admet une factorisation QR, où Q est une matrice (de taille* $m \times n$*) avec des colonnes orthonormées, et R est une matrice (carré de taille* $n \times n$*) triangulaire supérieure.*

Si la factorisation QR de la matrice A est connue, le MC est facilement résolu. De (10.6) on obtient :

$$x^* = \left(A^T A\right)^{-1} A^T = \left(R^T Q^T QR\right)^{-1} R^T Q^T b$$

d'où , en tenant compte que $Q^{\mathrm{T}} Q = I$, il résulte que

$$x^* = \left(R^T R\right)^{-1} R^T Q^T b = R^{-1} Q^T b$$

Par conséquence, la pseudo solution dans le sens MC peut être obtenue facilement en résolvant le système triangulaire :

$$Rx = Q^T b \tag{10.8}$$

système triangulaire (10.8) seulement $n(n+1)/2$ opération. Donc, le nombre total opérations est approximativement de deux fois plus grand que celui pour la formation du système d'équations normales.

Il existe une variante nouvelle de la méthode Gram – Schmidt nommée l'algorithme *de Gram –Schmidt modifié :*

$$
\begin{aligned}
&Pour \quad k = 1,2,\ldots,n \\
&\quad a_h = \frac{a_h}{|a_h|_2} \\
&\quad Pour \quad j = k+1,\ k+2,\ldots,n \\
&\quad a_j = a_j - \left(a_j^T a_h\right) a_h
\end{aligned}
$$

L'algorithme de Gram-Schmidt modifié est numériquement stable en raison du réarrangement de l'ordre des calculs. De plus, elle nécessite moins de mémoire de travail que la méthode classique d'orthogonalisation. Les vecteurs q_k sont calculés et placés dans le même emplacement mémoire que les vecteurs d'origine a_k.

Pour compléter les connaissances avec d'autres méthodes pour résolution du problème MC, il est recommandé [1, 2, 5, 7, 8, 9]. Pour le lecteur qui maîtrise bien les bases de l'algèbre linéaire on recommande le travail fondamental [5].

11. EXERCICES

1. Soit la matrice

$$A = \begin{pmatrix} 0 & 0 & 6 \\ 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{pmatrix}.$$

   À calculer A^3.
2. Déterminer les matrices carrées d'ordre 2, en satisfaisant les relations :
   a) $A^2 = 0$, même si $A \neq 0$
   b) $B^2 = -I$
   c) $CD = -DC$, mais $CD \neq 0$
   d) $EF = 0$, bien que les éléments des matrices E et F soient non nuls.
3. Montrez que

$$(AB)^T = B^T A^T ; (AB)^{-1} = B^{-1} A^{-1} ; \left(A^{-1}\right)^T = \left(A^T\right)^{-1}$$

4. La matrice P est appelée *« idempotente »* si $P^2 = P$. Déterminez toutes les matrices « idempotentes » d'ordre 2.
5. À calculer $\|x\|_1, \|x\|_2, \|x\|_\infty$ pour $x = (0, -1, -2, 0, 1)^T$.
6. Montrer que quelle que soit la norme vectorielle, des inégalités suivantes se produisent :
   a) a) $|\|x\| - \|y\|| \leq \|x - y\|$
   b) b) $\|x \pm y\| \leq \|x\| + \|y\|$.
7. Déterminer l'angle entre les vecteurs $x = (2, -2, 1)^T$ et

9. Soit $A = I - 2xx^T$ où $x \in R^n, (x,x) = 1$. Montrer que la matrice A est orthogonale et $A^2 = I$.
10. Soit A une matrice orthogonale et $Ax = \lambda x, x \neq 0$. Montrer que $|\lambda| = 1$.
11. Soit la matrice de premier rang

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 9 \end{pmatrix}.$$

Mettre la matrice A de la forme uv^T.
12. Soit la matrice :

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Montrer que la matrice A est orthogonale.
13. Montrer que si A est une matrice positive définie, alors les matrices A^2 et A^{-1} sont aussi positives définies.
14. Calculer la factorisation LU de la matrice

$$A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}.$$

15. Calculer la factorisation Cholesky LL^T de la matrice

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

$$\begin{cases} x_1 + 100x_2 = 100 \\ x_2 = 0 \end{cases}$$

est mal conditionné. Calculez le nombre de conditionnement.

18. Soit $x = (3 \quad 4)^T, z = (1 \quad 1)^T$. Calculez $\sigma = \|x\|_2$, $v = z + \sigma x$ et la matrice correspondante Householder.
19. Comparer les méthodes de Jacobi, de Gauss-Seidel et de surrelaxation successive pour la matrice A de l'exercice 15 et $b = (1,0,1)^T$ et $x^{(0)} = (0,0,0)^T$.
20. Soit

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}, \; b = \begin{pmatrix} 0.01 \\ 0.1 \end{pmatrix}, \; \delta b = \begin{pmatrix} 0.0001 \\ 0 \end{pmatrix}$$

Calculer la pré-solution dans le sens des plus petits carrés des systèmes surdéterminés $Ax = b, Ax = b + \delta b$. Comparez le résultat.

21. Calculez les valeurs propres et les vecteurs propres de la matrice.

$$A = \begin{pmatrix} 5 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}.$$

22. Calculer les valeurs propres de la matrice de l'exercice 15

23. Soit le système surdéterminé $Ax = b, A = \left(a_{ij}\right)_{m\times n}$, $b \in R^m$, la pseudo solution $x^* = \left(A^T A\right)^{-1} Ab$. Montrer que le vecteur résiduel $r = Ax^* - b$ est orthogonal sur sous-espace :

$$\mathrm{Im}\, A = \left\{y \middle| y = Ax, x \in R^n\right\} \subset R^m$$

24. La matrice $A^+ = \left(A^T A\right)^{-1} A^T$ de dimension $n \times m$ s'appelle « *la pseudo inverse* » de A ou *« la généralisation de la matrice inverse de More-Fenrose »*. Montrer que

$$AA^+A = A; \quad A^+AA^+ = A^+;$$
$$\left(AA^+\right)^T = AA^+; \quad \left(A^+A\right)^T = A^+A.$$

25. $P_A = AA^+$ est le projecteur orthogonal de A sur l'espace Im A. Montrer que $P_A^2 = P_A$ et $P_A^T = P_A$, c'est-à-dire la matrice P est idempotente et symétrique.

BIBLIOGRAPHIE

1. Brătianu C., Bostan V., Cojocia L., Negreanu G. Metode numerice. Editura tehnică, Bucureşti,1996. -212 p.
2. Воеводин В.В. Вычислительные основы линейной алгебры. М.: Наука, 1977. — 303 p.
3. Икрамов Х. Д. Численое решение матричных уравнений. М.: Наука, 1984. — 190 p.
4. Ланкастер П. Теория матриц. М.: Наука, 1978. — 280 p. (traducere din limba engleză *Lankaster P. Theory of matrices. Academic Press, New-York-London, 1969*)
5. Лоусон Ч., Хенсон Р. Численое решение задач метода наименьших квадратов. М.: Наука, 1986. — 232 p. (traducere din limba engleză *Lawson Ch., Hanson R. Solving least squares problems. Prentice – Hall, 1974*).
6. Парлетт Б. Симметричная проблема собственных значений. Численные методы. М.: Мир, 1983. — 384 p. (traducere din limba engleză *Parlett B. The symmetric eigenvalue problem, 1980*).
7. Райс Дж. Матричные вычисления и математическое обеспечение. М.: Мир, 1984. — 264 p. (traducere din limba engleză *Rice John. Matrix computations and mathematical software, 1981*).
8. Стренг Г. Линейная алгебра и ее применения. М.: Мир, 1980. — 454 p. (traducere din limba engleză *Strang Gilbert. Linear algebra and its applications,Academic Press, 1976*).
9. Форсайт Дж., Малькольи М., Моулер К., Машинные методы математических вычислений. М.: Мир, 1980. - 279 p. (traducere din limba engleză *Fosythe G., Malcolm M., Moler C. Computer Methods for Mathematical Computations, Prentice-Hall, 1982*).

11. Хорн Р., Джонсон Ч. Матричный анализ. М.: Мир, 1989. - 655 p. (traducere din limba engleză *Horn R., Johnson Ch. Matrix analysis. Cambridje University Press, 1986*).
12. Malbos Philipe. Analyse matricielle et algèbre linéaire appliquée. Notes de cours et de travaux dirigés. Université Claude Bernard Lyon 1. http://math.univ-lyon1.fr/homes-www/malbos/Ens/amalaa11.pdf

TABLE DE MATIERES