

MDI 210

Analyse numérique et optimisation continue

Partie 1 : Éléments d'analyse numérique

Irène Charon      Olivier Hudry

21 février 2017



# Table des matières

<b>1 Analyse matricielle - Généralités</b>	<b>5</b>
1.1 Rappels d'algèbre linéaire . . . . .	5
1.1.1 Adjointes . . . . .	5
1.1.2 Types de matrices . . . . .	6
1.1.3 Spectre d'une matrice . . . . .	6
1.1.4 Réduction d'une matrice . . . . .	7
1.1.5 Valeurs singulières . . . . .	8
1.2 Normes . . . . .	8
1.2.1 Convergence de suites de matrices . . . . .	11
<b>2 Problèmes de l'analyse numérique</b>	<b>13</b>
2.1 Erreurs . . . . .	13
2.2 Conditionnement . . . . .	14
2.2.1 Conditionnement d'un système linéaire . . . . .	14
2.2.2 Conditionnement d'un problème de recherche de valeurs propres . . . . .	17
<b>3 Résolution de systèmes linéaires</b>	<b>19</b>
3.1 Généralités . . . . .	19
3.2 Méthode de Gauss . . . . .	20
3.2.1 Étape d'élimination . . . . .	20
3.2.2 Choix du pivot . . . . .	22
3.2.3 Complexité . . . . .	23
3.2.4 Variante : la méthode de Gauss-Jordan . . . . .	23
3.3 Factorisation LU . . . . .	24
3.4 Méthode de Cholesky . . . . .	27

<b>4 Valeurs propres et vecteurs propres</b>	<b>31</b>
4.1 Méthode de Jacobi. . . . .	32
4.2 Bibliographie . . . . .	39

# Chapitre 1

## Analyse matricielle - Généralités

### 1.1 Rappels d'algèbre linéaire

#### 1.1.1 Adjoints

Dans toute la suite, on considère  $\mathbb{R}$  ou  $\mathbb{C}$  comme corps de base. Rappelons d'abord quelques définitions.

Étant donné un vecteur  $x$  (représenté généralement par une matrice colonne), on appelle *adjoint* de  $x$  et on note  $x^*$  le vecteur transposé du vecteur

conjugué de  $x$  : si  $x = \begin{pmatrix} x_1 \\ \dots \\ x_i \\ \dots \\ x_n \end{pmatrix}$ , alors  $x^* = (\overline{x_1}, \dots, \overline{x_i}, \dots, \overline{x_n})$ .

*Remarque*

Si on se place dans  $\mathbb{R}$ , on a :  $x^* = (x_1, \dots, x_i, \dots, x_n)$ .

Le *produit hermitien* de deux vecteurs  $x$  et  $y$  de dimension  $n$  est défini par :  $(x, y) = \sum_{i=1}^n \overline{x_i} y_i$ . Si les vecteurs sont représentés par des vecteurs colonnes, on a :  $(x, y) = x^* y$ , où le produit est le produit matriciel. Si les vecteurs sont à composantes réelles, le produit hermitien devient le *produit scalaire euclidien* :  $(x, y) = \sum_{i=1}^n x_i y_i = x^t y$ .

À une matrice  $A$ , on peut associer sa *matrice adjointe*, notée  $A^*$ . Celle-ci est définie comme suit : si  $A = (a_{i,j})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}$ , alors  $A^* = \overline{A^t} = (\overline{a_{j,i}})_{\substack{1 \leq j \leq p \\ 1 \leq i \leq n}}$ .

On a :  $(A^*)^* = A$ .

Si  $x$  et  $y$  sont deux vecteurs colonnes ayant respectivement  $n$  lignes et  $p$  lignes et  $A$  une matrice à  $n$  lignes et  $p$  colonnes, on peut vérifier la propriété :  $(x, Ay) = (A^*x, y)$ .

### 1.1.2 Types de matrices

Une matrice carrée réelle  $A$  est dite :

- *symétrique* si  $A^t = A$ ,
- *normale* si  $AA^t = A^tA$ ,
- *orthogonale* si  $AA^t = A^tA = I$ , où  $I$  désigne la matrice identité.

Dans la suite, les qualificatifs « symétrique » et « orthogonale » ne s'appliquent qu'à des matrices réelles.

Une matrice carrée complexe  $A$  est dite :

- *hermitienne* si  $A^* = A$ ,
- *normale* si  $AA^* = A^*A$ .
- *unitaire* si  $A^*A = AA^* = I$ .

On remarquera qu'une matrice symétrique ou hermitienne est normale. De même pour une matrice orthogonale ou unitaire. On rappelle que les valeurs propres d'une matrice réelle symétrique ou d'une matrice hermitienne sont réelles.

### 1.1.3 Spectre d'une matrice

Une *valeur propre* d'une matrice carrée  $A$  est un scalaire  $\lambda$  tel qu'il existe un vecteur  $x$  non nul vérifiant :  $Ax = \lambda x$ . Le vecteur  $x$  est alors dit *vecteur propre* de  $A$ .

Soit  $A$  une matrice carrée. Le *spectre* de  $A$  est l'ensemble des valeurs propres de  $A$ . Le *rayon spectral* de  $A$  est le plus grand des modules des valeurs propres de  $A$ ; il est noté  $\rho(A)$ .

### 1.1.4 Réduction d'une matrice

Deux matrices carrées sont dites *semblables* si elles sont susceptibles de représenter la même application linéaire sur deux bases différentes. Si  $A$  et  $B$  sont deux matrices semblables, il existe une matrice inversible  $P$  vérifiant  $A = P^{-1}BP$ ; la matrice  $P$  s'appelle *matrice de passage*. Une matrice est *diagonalisable* si elle est semblable à une matrice diagonale; cette matrice diagonale est constituée des valeurs propres de  $A$  comptées avec leur ordre de multiplicité.

Toute matrice symétrique réelle est semblable à une matrice diagonale réelle.

Une matrice carrée est inversible si et seulement si elle ne possède aucune valeur propre nulle.

Toute matrice symétrique réelle est semblable à une matrice diagonale réelle.

En fait on peut aussi démontrer, pour les matrices carrées, les résultats suivants :

**Théorème :**

1. Soit  $A$  une matrice carrée quelconque ; il existe une matrice unitaire  $U$  telle que  $U^{-1}AU$  soit triangulaire.
2. Soit  $A$  une matrice normale ; il existe une matrice unitaire  $U$  telle que  $U^{-1}AU$  soit diagonale.
3. Soit  $A$  une matrice symétrique ; il existe une matrice orthogonale  $O$  telle que  $O^{-1}AO$  soit diagonale.

**Corollaires des définitions et de ce théorème :**

1. Le module des valeurs propres d'une matrice orthogonale ou unitaire vaut 1.
2. Une matrice hermitienne (resp. symétrique) ou unitaire est diagonalisable par une matrice de passage unitaire (resp. orthogonale).
3. Une matrice orthogonale  $O$  est diagonalisable par une matrice  $U$ , en général non réelle, unitaire ( $O = U^*DU$ ), les éléments diagonaux de  $D$  étant de module 1.

### 1.1.5 Valeurs singulières

La matrice  $A^*A$  est normale, elle est donc diagonalisable. On peut montrer facilement que ses valeurs propres sont positives ou nulles. On appelle *valeurs singulières* de  $A$  les racines carrées positives des valeurs propres de  $A^*A$ . La matrice  $A$  est inversible si et seulement si ses valeurs singulières sont toutes strictement positives.

Deux matrices  $A$  et  $B$  sont dites *équivalentes* s'il existe deux matrices inversibles  $U$  et  $V$  telles que  $B = U^{-1}AV$ .

Soit  $A$  une matrice carrée ;  $A$  est équivalente à une matrice diagonale dont la diagonale est constituée des valeurs singulières de  $A$ . Plus précisément :

- si  $A$  est réelle, il existe deux matrices carrées orthogonales  $U$  et  $V$  et une matrice diagonale  $D$  constituée des valeurs singulières de  $A$  telles que :  $A = U^tDV$  ;
- si  $A$  est complexe, il existe deux matrices carrées unitaires  $U$  et  $V$  et une matrice diagonale  $D$  constituée des valeurs singulières de  $A$  telles que :  $A = U^*DV$ .

## 1.2 Normes

Nous aurons besoin dans ces chapitres non seulement de la notion de norme vectorielle mais également de norme matricielle : nous allons donc rappeler quelques notions concernant les premières et définir les secondes. Soit  $x = (x_i)_{1 \leq i \leq n}$  un vecteur. Les trois normes vectorielles les plus usuelles sont les suivantes :

- $\|x\|_1 = \sum_{i=1}^n |x_i|$  (norme 1)
- $\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}}$  (norme 2, ou norme euclidienne)
- $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$  (norme infinie)

Plus généralement :  $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$  (norme  $p$ ).

La démonstration du fait qu'il s'agit d'une norme utilise les inégalités suivantes :



- Inégalité de Hölder : si  $p$  et  $q$  sont deux nombres vérifiant  $p > 1$  et l'égalité  $\frac{1}{p} + \frac{1}{q} = 1$  (ce qui entraîne  $q > 1$ ), alors :

$$\sum_{i=1}^n |x_i y_i| \leq \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \left( \sum_{i=1}^n |y_i|^q \right)^{\frac{1}{q}}.$$

Pour  $p = q = 2$ , cela redonne l'inégalité de Cauchy-Schwarz.

- Inégalité de Minkowski :

$$\left( \sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{1}{p}} \leq \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left( \sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}}.$$

Dans  $\mathbb{R}^n$  et  $\mathbb{C}^n$ , toutes les les normes sont *équivalentes* (deux normes  $\| \cdot \|$  et  $\| \cdot \|'$  sont équivalentes sur un espace vectoriel  $E$  s'il existe deux constantes strictement positives  $C$  et  $C'$  telles que, pour tout  $x$  dans  $E$  :  $C\|x\| \leq \|x\|' \leq C'\|x\|$ ).

On peut également s'intéresser aux normes matricielles. On appelle  $\mathcal{A}_n$  l'anneau des matrices carrées d'ordre  $n$  à coefficients dans  $\mathbb{R}$  ou  $\mathbb{C}$ . On appelle *norme matricielle* une application de  $\mathcal{A}_n$  dans  $\mathbb{R}^+$  notée  $\| \cdot \|$  qui vérifie les propriétés suivantes :

- pour toute matrice  $A$  de  $\mathcal{A}_n$ ,  $\|A\| = 0 \Leftrightarrow A = 0$
- pour tout  $\alpha$  de  $\mathbb{R}$  (ou  $\mathbb{C}$ ) et pour tout  $A$  de  $\mathcal{A}_n$ ,  $\|\alpha A\| = |\alpha| \|A\|$
- pour toutes matrices  $A$  et  $B$  de  $\mathcal{A}_n$ ,  $\|A + B\| \leq \|A\| + \|B\|$
- pour toutes matrices  $A$  et  $B$  de  $\mathcal{A}_n$ ,  $\|A \times B\| \leq \|A\| \times \|B\|$ .

On peut très facilement construire des normes matricielles à partir de normes vectorielles : elles sont dites alors *normes matricielles subordonnées*. Pour cela, on peut définir  $\|A\|$  par les formules équivalentes suivantes :

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\| = \sup_{0 < \|x\| \leq 1} \frac{\|Ax\|}{\|x\|}.$$

On a :  $\|Ax\| \leq \|A\| \|x\|$ .

Les normes matricielles subordonnées aux normes les plus usuelles que nous avons décrites plus haut sont donc, pour  $A = (a_{i,j})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$  :

- $\|A\|_1 = \sup_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|$
- $= \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sqrt{\rho(A^*A)} = \|A^*\|_2$  où  $\rho(A^*A)$  représente le plus grand module des valeurs propres de  $A^*A$  (rayon spectral de  $A^*A$ )
- $\|A\|_\infty = \sup_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$ .

La norme  $\|\cdot\|_2$  est invariante par transformation unitaire : si  $U$  est une matrice unitaire, c'est-à-dire vérifie  $U^*U = I$ , on a alors

$$\|A\|_2 = \|AU\|_2 = \|UA\|_2 = \|U^*AU\|_2.$$

Si  $A$  est normale, c'est-à-dire si  $A$  vérifie  $A^*A = AA^*$  (en particulier si  $A$  est hermitienne ou symétrique), alors  $\|A\|_2 = \rho(A)$ .

Si  $A$  est unitaire ou orthogonale,  $\|A\|_2 = 1$ .

*Remarque :*  $\|A\|_1$  et  $\|A\|_\infty$  sont faciles à calculer mais pas  $\|A\|_2$ .

### Théorème

- Soit  $\|\cdot\|$  une norme subordonnée ; soit  $B$  vérifiant  $\|B\| < 1$ . Alors  $I + B$  est inversible et  $\|(I + B)^{-1}\| \leq \frac{1}{1 - \|B\|}$ .
- Si une matrice de la forme  $I + B$  n'est pas inversible, alors, pour toute norme, subordonnée ou non,  $\|B\| \geq 1$ .

*Exemple de norme non subordonnée : la norme euclidienne*

Cette norme est définie par :  $\|A\|_E = \left( \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}} = \sqrt{\text{trace}(A^*A)}$  (on rappelle que la trace d'une matrice est la somme de ses termes diagonaux). La norme  $\|A\|_E$  est invariante par transformation unitaire ; autrement dit, si  $U^*U = I$ , alors  $\|A\|_E = \|AU\|_E = \|UA\|_E = \|U^*AU\|_E$ .

De plus :  $\|A\|_2 \leq \|A\|_E \leq \sqrt{n}\|A\|_2$ .

**Théorème.** Soit  $\|\cdot\|$  une norme quelconque (subordonnée ou non) ; on a :  $\rho(A) \leq \|A\|$  et, pour tout  $\epsilon > 0$ , il existe une norme subordonnée  $\|\cdot\|_{A,\epsilon}$  vérifiant  $\|A\|_{A,\epsilon} \leq \rho(A) + \epsilon$ .

### 1.2.1 Convergence de suites de matrices

Si l'espace est de dimension finie, toutes les normes sont équivalentes. Pour qu'une suite  $(x^k)$  de vecteurs converge, il faut et il suffit que les composantes de  $(x^k)$  convergent. Il en est de même pour les suites de matrices. On a en particulier le théorème suivant pour la suite des puissances d'une matrice :

**Théorème :** Soit  $B$  une matrice carrée.

1.  $\lim_{k \rightarrow \infty} B^k = 0 \Leftrightarrow \forall x, \lim_{k \rightarrow \infty} B^k x = 0 \Leftrightarrow \rho(B) < 1 \Leftrightarrow$   
pour au moins une norme subordonnée,  $\|B\| < 1$ .
2. Soit  $\|\cdot\|$  une norme quelconque ; alors  $\lim_{k \rightarrow \infty} \|B^k\|^{\frac{1}{k}} = \rho(B)$ .



# Chapitre 2

## Problèmes de l'analyse numérique

Les deux problèmes principaux que nous allons étudier dans la suite de ce cours sont la résolution des systèmes linéaires et le calcul des valeurs propres et vecteurs propres des matrices. Lorsqu'on applique les méthodes de l'analyse numérique à des problèmes de calcul, il faut prendre en compte deux types de « qualité ». Il s'agit d'une part de l'aspect que l'on appelle *complexité*, c'est-à-dire du nombre d'opérations élémentaires à effectuer pour obtenir un résultat, mais aussi il faut savoir déterminer si la solution est acceptable ou non ; en effet, on peut commettre deux sortes d'erreurs : d'une part, les erreurs d'arrondi, dues à la précision des calculs et, d'autre part, les erreurs dites de troncature, lorsque l'on utilise des méthodes itératives, alors que l'on s'arrête bien sûr après un nombre fini d'itérations.

### 2.1 Erreurs

*Erreur d'arrondi* : erreur due au codage où le nombre de chiffres représentant un réel est limité. Si le nombre est codé sur  $t$  bits pour la mantisse, l'erreur sur la mantisse est majorée par  $2^{-t}$ .

*Erreur de troncature* : dans les méthodes itératives, le calcul de la limite nécessiterait *a priori* un nombre infini d'itérations. Comme on arrête forcément les calculs après un nombre  $k_0$  d'itérations, on commet une erreur de troncature mesurée par  $\|x^\infty - x^{k_0}\|$ , où  $x^\infty$  représente la limite,  $x^{k_0}$  le résultat obtenu à la  $k_0^{\text{e}}$  itération et  $\|\cdot\|$  une norme donnée, quand on arrête la méthode itérative (en fait,  $x^\infty$  est inconnu, ce qui ne permet pas d'estimer l'erreur).

## 2.2 Conditionnement

Dans tout ce qui suit, on considère un système linéaire écrit sous la forme matricielle  $Ax = b$ . Avant de rentrer dans le détail des méthodes, qui feront l'objet du chapitre suivant, nous allons traiter d'un paramètre important des systèmes linéaires : il s'agit de leur *conditionnement*, lequel est attaché à la matrice  $A$  du système. Le plus souvent, dans la pratique, les coefficients de  $A$ , comme les composantes du vecteur  $b$ , sont les résultats de mesure et sont donc entachés d'une certaine erreur. Il est essentiel de voir comment une petite modification de  $A$  ou de  $b$  influe, indépendamment de la méthode utilisée, sur la solution supposée exacte du système.

### 2.2.1 Conditionnement d'un système linéaire

Considérons le système suivant :

$$\begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix} \text{ de solution } \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Considérons maintenant le système perturbé en modifiant légèrement le vecteur du second membre, la matrice restant inchangée :

$$\begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \begin{pmatrix} x_1 + \delta x_1 \\ x_2 + \delta x_2 \\ x_3 + \delta x_3 \\ x_4 + \delta x_4 \end{pmatrix} = \begin{pmatrix} 32,1 \\ 22,9 \\ 33,1 \\ 30,9 \end{pmatrix} \text{ de solution } \begin{pmatrix} 9,2 \\ -12,6 \\ 4,5 \\ -1,1 \end{pmatrix}.$$

On constate qu'une erreur relative de l'ordre de  $1/300$  sur le second membre entraîne une erreur relative de l'ordre de 10 sur plusieurs coordonnées de la solution du système, et donc une amplification des erreurs relatives de l'ordre de 3000.

Considérons maintenant de légères modifications sur la matrice avec le système :

$$\begin{pmatrix} 10 & 7 & 8,1 & 7,2 \\ 7,08 & 5,04 & 6 & 5 \\ 8 & 5,98 & 9,89 & 9 \\ 6,99 & 4,99 & 9 & 9,98 \end{pmatrix} \begin{pmatrix} x_1 + \delta x_1 \\ x_2 + \delta x_2 \\ x_3 + \delta x_3 \\ x_4 + \delta x_4 \end{pmatrix} = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix} \text{ de solution } \begin{pmatrix} -81 \\ 137 \\ -34 \\ 22 \end{pmatrix}.$$

On constate ici aussi que de petites variations des éléments de la matrice modifient considérablement la solution du système linéaire.

Supposons, toutes choses étant égales par ailleurs, que l'on considère le système :  $A(x + \delta x) = b + \delta b$ , et supposons la matrice  $A$  inversible. On voit que l'on a  $\delta x = A^{-1}\delta b$ . Si on choisit alors une norme matricielle  $\| \cdot \|$ , subordonnée à une norme vectorielle, on trouve  $\|\delta x\| \leq \|A^{-1}\| \|\delta b\|$  et, de plus,  $\|b\| \leq \|A\| \|x\|$  de sorte que l'on a sur  $x$  une erreur relative  $\frac{\|\delta x\|}{\|x\|}$

majorée par  $\|A\| \|A^{-1}\| \frac{\|\delta b\|}{\|b\|}$ . On appelle *conditionnement* de la matrice  $A$  (relativement à la norme  $\| \cdot \|$ ) la quantité  $\|A\| \|A^{-1}\|$ , ce que l'on note  $\text{cond}_{\| \cdot \|}(A)$  ou, plus simplement,  $\text{cond}(A)$ .

On pourrait prouver de même que si l'on apporte maintenant une petite variation aux coefficients de  $A$ , de sorte que cette matrice devienne  $A + \delta A$ , alors  $\frac{\|\delta x\|}{\|x + \delta x\|}$  est majorée par  $\|A\| \|A^{-1}\| \frac{\|\delta A\|}{\|A\|}$ .

Ces deux majorations prouvent l'intérêt du conditionnement. Une matrice est d'autant mieux conditionnée que son conditionnement (qui est toujours supérieur ou égal à 1) est proche de 1.

Nous n'avons évoqué ici que le conditionnement d'une matrice par rapport à la résolution d'un système linéaire. Nous verrons ultérieurement ce qu'est le conditionnement pour un problème de valeurs propres. Une même matrice peut en fait être mal conditionnée en tant que matrice d'un système linéaire et l'être bien pour le problème de la recherche des valeurs propres, et *vice versa*.

Le théorème suivant donne d'autres renseignements sur le conditionnement d'une matrice au sens des systèmes.

**Théorème :** Soit  $A$  une matrice inversible. On a alors :

1.  $\text{cond}(A) \geq 1$
2.  $\text{cond}(A) = \text{cond}(A^{-1})$
3. pour tout  $\alpha \neq 0$ ,  $\text{cond}(\alpha A) = \text{cond}(A)$
4. en notant  $\text{cond}_2$  le conditionnement associé à  $\| \cdot \|_2$  et en notant respectivement  $\mu_1(A)$  et  $\mu_n(A)$  la plus petite et la plus grande des valeurs singulières de  $A$ ,  $\text{cond}_2(A) = \frac{\mu_n(A)}{\mu_1(A)}$
5. si  $A$  est normale (c'est-à-dire vérifie  $AA^* = A^*A$ ),  $\text{cond}_2(A) = \frac{\max_i |\lambda_i(A)|}{\min_i |\lambda_i(A)|}$   
où les  $\lambda_i(A)$  représentent les valeurs propres de  $A$
6. si  $A$  est unitaire ou orthogonale,  $\text{cond}_2(A) = 1$
7.  $\text{cond}_2(A)$  est invariant par transformation unitaire ou orthogonale : si  $UU^* = I$  (respectivement  $OO^t = I$ ), alors

$$\text{cond}_2(A) = \text{cond}_2(AU) = \text{cond}_2(UA) = \text{cond}_2(U^*AU) \\ (\text{respectivement } \text{cond}_2(O^tAO)).$$

Calculons par exemple le conditionnement de la matrice utilisée précédemment :

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}.$$

Cette matrice a pour valeurs propres approchées :

$$\lambda_1 \approx 0,01015 < \lambda_2 \approx 0,8431 < \lambda_3 \approx 3,858 < \lambda_4 \approx 30,2887.$$

On a ainsi :  $\text{cond}_2(A) = \frac{\lambda_4}{\lambda_1} \approx 2984$ . La matrice  $A$  a donc un très mauvais conditionnement, ce qui explique la sensibilité aux erreurs des systèmes linéaires définis avec la matrice  $A$ .

Comme pour tout  $\alpha \neq 0$ ,  $\text{cond}(\alpha A) = \text{cond}(A)$ , on ne peut espérer diminuer le conditionnement de  $A$  en multipliant tous ses éléments par un même



nombre. En revanche, on peut le faire en multipliant par exemple chaque ligne (et/ou chaque colonne) par un coefficient approprié ; c'est là le problème de l'*équilibrage d'une matrice*, qui peut s'énoncer comme suit : étant donnée une matrice  $A$ , déterminer deux matrices diagonales inversibles  $D_1$  et  $D_2$  vérifiant :  $\text{cond}(D_1 A D_2) = \inf_{\Delta_1, \Delta_2 \text{ diagonales inversibles}} \text{cond}(\Delta_1 A \Delta_2)$ .

On résout alors  $Ax = b$  en deux étapes :

- résolution de  $D_1 A D_2 y = D_1 b$
- résolution de  $x = D_2 y$ .

En pratique, le conditionnement n'est pas une fonction simple des éléments de  $D_1$  et  $D_2$  ; on essaie plutôt de minimiser le rapport entre le plus grand et le plus petit élément non nul de  $A' = \Delta_1 A \Delta_2$ . Posons  $E = \{(i, j) \text{ avec } 1 \leq i \leq n, 1 \leq j \leq n \text{ et } a_{ij} \neq 0\}$ . On cherche deux matrices  $\Delta_1$  et  $\Delta_2$  diagonales et inversibles qui minimisent le rapport :

$$\frac{\max_{(i,j) \in E} |a'_{ij}|}{\min_{(i,j) \in E} |a'_{ij}|}.$$

En notant  $x_i$  le  $i^e$  élément de la diagonale de  $\Delta_1$  et  $y_i$  le  $i^e$  élément de la diagonale de  $\Delta_2$ , on a :  $a'_{ij} = x_i a_{ij} y_j$ . On passe aux logarithmes en posant  $\alpha_{ij} = \ln |a_{ij}|$ ,  $u_i = \ln |x_i|$ ,  $v_j = \ln |y_j|$ . Le problème devient :

$$\text{minimiser}_{u_i, v_j \text{ avec } (i,j) \in E} \left[ \max_{(i,j) \in E} (\alpha_{ij} + u_i + v_j) - \min_{(i,j) \in E} (\alpha_{ij} + u_i + v_j) \right],$$

ce qui se réécrit comme le programme linéaire suivant (car on peut, par une translation des valeurs, se restreindre aux solutions où le minimum sur les  $u_i$  et  $v_j$  de  $\alpha_{ij} + u_i + v_j$  vaut 0) :

$$\begin{cases} \text{minimiser } z \\ \text{avec, pour tout } (i, j) \in E, & 0 \leq \alpha_{ij} + u_i + v_j \leq z \\ u_i \text{ et } v_j \text{ de signes quelconques.} \end{cases}$$

### 2.2.2 Conditionnement d'un problème de recherche de valeurs propres

Dans un problème de recherche de valeurs propres, il est à nouveau important de connaître l'influence d'une petite modification des coefficients de la

matrice  $A$  sur les valeurs propres calculées. Ce conditionnement fait intervenir le conditionnement des matrices de passage de  $A$  à une forme diagonale, et non  $A$  directement. Le théorème suivant permet de définir ce nouveau conditionnement que l'on notera  $\Gamma(A)$ .

**Théorème :** soit  $A$  une matrice diagonalisable et  $P$  une matrice telle que  $P^{-1}AP$  soit diagonale de termes diagonaux  $\lambda_i$ . Soit  $\|\cdot\|$  une norme matricielle telle que, pour toute matrice diagonale  $\text{diag}(\delta_i)$  :

$$\|\text{diag}(\delta_i)\| = \max_i |\delta_i|.$$

Alors, pour toute matrice  $\delta A$  :

$$\text{spectre}(A + \delta A) \subset \bigcup_{i=1}^n D_i,$$

avec  $D_i = \{z \in \mathbb{C} \text{ tels que } |z - \lambda_i| \leq \text{cond}_{\|\cdot\|}(P) \|\delta A\|\}$ .

Ceci veut dire que, si  $A$  est diagonalisable, la perturbation  $\delta A$  laisse globalement les valeurs propres dans des disques complexes, centrés en les anciennes valeurs propres et de rayon  $\text{cond}_{\|\cdot\|}(P) \|\delta A\|$ .

Pour  $A$  diagonalisable, le conditionnement  $\Gamma(A)$  relativement à la recherche des valeurs propres est défini comme étant le minimum de  $\text{cond}_{\|\cdot\|}(P)$  pris sur les matrices  $P$  telles que  $P^{-1}AP$  soit diagonale. Le théorème ci-dessus indique ainsi que, pour  $A$  diagonalisable, on a l'inclusion :

$$\text{spectre}(A + \delta A) \subset \bigcup_{i=1}^n \{z \in \mathbb{C} \text{ tels que } |z - \lambda_i| \leq \Gamma(A) \|\delta A\|\}.$$

Les matrices normales étant diagonalisables grâce à des matrices  $P$  unitaires, elles ont un conditionnement  $\Gamma(A)$  égal à 1 pour  $\|\cdot\|_2$ . Ceci est donc en particulier le cas pour les matrices symétriques. Dans ce dernier cas on a de plus le théorème :

**Théorème :** Soit  $A$  une matrice symétrique et  $B = A + \delta A$ , où la perturbation  $\delta A$  est également symétrique. Soient  $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_n$  les valeurs propres de  $A$  et  $\beta_1 \leq \beta_2 \leq \dots \leq \beta_n$  les valeurs propres de  $B$ . Alors, on a pour  $1 \leq i \leq n$  :  $|\alpha_i - \beta_i| \leq \|\delta A\|_2$ .

Ce théorème exprime que, si  $A$  et  $\delta A$  sont toutes deux symétriques, chaque valeur propre de  $A + \delta A$  reste dans un intervalle réel centré sur l'ancienne valeur propre et de rayon  $\|\delta A\|_2$ .

# Chapitre 3

## Résolution de systèmes linéaires

### 3.1 Généralités

Le problème auquel on s'intéresse peut se formuler de la façon suivante.

*Problème :* Soient  $A = (a_{i,j})$  une matrice carrée inversible de dimension  $n$ ,  $x = (x_i)$  et  $b = (b_i)$  deux vecteurs colonnes de dimension  $n$  ; résoudre par rapport à  $x$  le système  $Ax = b$ .

*Remarques :*

1. Les méthodes numériques de résolution n'utilisent généralement pas le calcul de  $A^{-1}$ .
2. Si  $A$  est sous forme triangulaire supérieure (les éléments sous la diagonale principale sont tous nuls) avec des termes diagonaux non nuls, alors la résolution est aisée. On commence par résoudre la dernière relation, qui est une équation linéaire en la seule variable  $x_n$ , on reporte cette valeur dans l'avant-dernière relation qui devient une équation en  $x_{n-1}$ , et on continue ainsi de proche en proche jusqu'à  $x_1$ . Cette méthode, dite *méthode de remontée* et résumée ci-dessous, nécessite  $n(n-1)/2$  additions,  $n(n-1)/2$  multiplications et  $n$  divisions.

$$\left\{ \begin{array}{lclclcl} a_{1,1}x_1 + & \dots & + & a_{1,n-1}x_{n-1} & + & a_{1,n}x_n & = & b_1 \\ & & & \dots & & & & \\ & & & a_{n-1,n-1}x_{n-1} & + & a_{n-1,n}x_n & = & b_{n-1} \\ & & & & & a_{n,n}x_n & = & b_n \end{array} \right.$$

$$\Rightarrow \begin{cases} x_n = \frac{b_n}{a_{n,n}} \\ x_{n-1} = \frac{b_{n-1} - a_{n-1,n}x_n}{a_{n-1,n-1}} \\ \dots \\ x_1 = \frac{b_1 - a_{1,2}x_2 - \dots - a_{1,n}x_n}{a_{1,1}} \end{cases}$$

## 3.2 Méthode de Gauss

La méthode de Gauss est utilisée lorsque la matrice  $A$  est inversible quelconque. Le principe en est le suivant :

- À l'aide de combinaisons linéaires entre les lignes de  $A$ , on élimine successivement certaines inconnues des relations, pour obtenir une forme  $(MA)x = Mb$  où  $MA$  est une matrice triangulaire supérieure. Remarquons qu'en fait on ne calcule pas  $M$ , mais qu'on construit directement  $MA$  et  $Mb$ .
- On résout  $(MA)x = Mb$  par une méthode de remontée.

### 3.2.1 Étape d'élimination

- On choisit dans la première colonne un coefficient  $a_{i,1}$  différent de 0 ; il en existe toujours puisque la matrice est inversible. Cet élément constitue le *pivot*.
- Si le pivot n'est pas en première ligne, on échange la ligne du pivot avec la première ligne.
- Par des combinaisons linéaires bien choisies, obtenues en retranchant à chaque ligne la première ligne multipliée par le bon coefficient, on annule tous les termes de la colonne du pivot situés sous la diagonale.
- On obtient alors une matrice  $A'$  dont la première colonne n'a que des 0 sous le premier terme qui, lui, est non nul.
- On considère la matrice obtenue en supprimant la première ligne et la première colonne de  $A'$ . On réitère le procédé sur cette nouvelle matrice.

- On arrête ce procédé quand la matrice obtenue est de dimension 1.
- En remplaçant les lignes et colonnes supprimées au fur et à mesure, on obtient une matrice triangulaire.

*Remarque* : Le déterminant de  $A$  s'obtient par le produit des pivots multiplié par  $(-1)^p$ , où  $p$  représente le nombre de fois que le pivot n'était pas sur la diagonale.

*Exemple 1*

On considère le système :

$$\begin{cases} 2x_1 + x_2 - 3x_3 = 5 \\ 4x_1 + x_2 + 5x_3 = -1 \\ 10x_1 - 7x_2 + 13x_3 = -3 \end{cases}$$

Après la première itération, en ayant choisi comme pivot la valeur 2, en gras ci-dessus, on obtient :

$$\begin{cases} 2x_1 + x_2 - 3x_3 = 5 \\ -1x_2 + 11x_3 = -11 \\ -12x_2 + 28x_3 = -28 \end{cases}$$

Après la seconde itération (le pivot est le coefficient de la deuxième ligne, deuxième colonne et vaut  $-1$ ), on obtient :

$$\begin{cases} 2x_1 + x_2 - 3x_3 = 5 \\ -x_2 + 11x_3 = -11 \\ -104x_3 = 104 \end{cases}$$

On applique alors une méthode de remontée, et l'on obtient successivement :

$$x_3 = -1, x_2 = \frac{-11 - 11x_3}{-1} = 0, x_1 = \frac{5 - x_2 + 3x_3}{2} = 1.$$

*Remarque* : comme les lignes n'ont pas été échangées, le déterminant de  $A$  est égal au déterminant de la matrice correspondant au dernier système. On a donc :  $\det(A) = 2 \times (-1) \times (-104) = 208$ .

*Exemple 2*

On considère le système :

$$\begin{cases} \mathbf{2}x_1 + x_2 - 3x_3 = -3 \\ 4x_1 + 2x_2 - x_3 = 4 \\ 6x_1 + 5x_2 + 8x_3 = 27 \end{cases}$$

Après la première itération, en ayant choisi comme pivot la valeur 2, en gras ci-dessus, on obtient :

$$\begin{cases} 2x_1 + x_2 - 3x_3 = -3 \\ \phantom{2x_1} 0x_2 + 5x_3 = 10 \\ \phantom{2x_1} \mathbf{2}x_2 + 17x_3 = 36 \end{cases}$$

Le pivot est maintenant nécessairement le coefficient de  $x_2$  dans la dernière ligne (de valeur 2) ; on échange la deuxième et la troisième ligne ; on obtient :

$$\begin{cases} 2x_1 + x_2 - 3x_3 = -3 \\ \phantom{2x_1} \mathbf{2}x_2 + 17x_3 = 36 \\ \phantom{2x_1} 0x_2 + 5x_3 = 10 \end{cases}$$

Le coefficient de  $x_2$  dans la dernière ligne étant nul, il ne reste plus qu'à effectuer la remontée :

$$x_3 = 10/5 = 2, x_2 = \frac{36 - 17x_3}{2} = 1, x_1 = \frac{-3 - x_2 + 3x_3}{2} = 1.$$

*Remarque* : les lignes ayant été échangées une fois, le déterminant de  $A$  est égal au déterminant de la matrice correspondant au dernier système multiplié par  $-1$ . On a donc :  $\det(A) = (-1) \times 2 \times 2 \times 5 = -20$ .

### 3.2.2 Choix du pivot

À cause des erreurs d'arrondi, le choix du pivot est important ; en effet, un pivot trop petit en module peut conduire à de mauvaises solutions du fait de la division par le pivot. Deux stratégies sont en fait possibles.

- *Pivot partiel* : on choisit dans la colonne courante le terme de plus grand module situé sous la diagonale ou sur celle-ci.
- *Pivot total* : on choisit le terme de plus grand module de la matrice résiduelle, c'est-à-dire, si on est à l'étape  $n - k + 1$ , la matrice constituée des  $k$  dernières lignes et des  $k$  dernières colonnes. Cette méthode est plus coûteuse en temps.

### 3.2.3 Complexité

On peut évaluer le nombre d'opérations nécessaires pour la méthode de Gauss ; dans le cas où on ne choisit pas le pivot, on effectue en tout environ  $\frac{n^3}{3}$  additions, autant de multiplications,  $\frac{n^2}{2}$  divisions et donc au total un nombre d'opérations arithmétiques équivalent à  $\frac{2n^3}{3}$ .

### 3.2.4 Variante : la méthode de Gauss-Jordan

Par rapport à la méthode de Gauss, la seule différence apportée par la méthode de Gauss-Jordan est que, dans la phase d'élimination, on élimine également les termes situés au-dessus de la diagonale. On obtient ainsi une matrice diagonale. Cette méthode est notamment utilisée pour le calcul de l'inverse d'une matrice. On résout alors simultanément les  $n$  systèmes linéaires  $Ax_j = e_j$ , l'inconnue étant le vecteur colonne  $x_j$  (le  $j^{\text{e}}$  vecteur de la matrice inverse), les  $e_j$  constituant les vecteurs de la base canonique de  $\mathbb{R}^n$ .

*Exemple* : Calcul de l'inverse de  $A = \begin{pmatrix} 1 & -3 & 14 \\ 1 & -2 & 10 \\ -2 & 4 & -19 \end{pmatrix}$ .

On résout les trois systèmes :

$$\left\{ \begin{array}{rrrr} x_1 & - & 3x_2 & + & 14x_3 & = & 1 & | & 0 & | & 0 \\ x_1 & - & 2x_2 & + & 10x_3 & = & 0 & | & 1 & | & 0 \\ -2x_1 & + & 4x_2 & - & 19x_3 & = & 0 & | & 0 & | & 1 \end{array} \right.$$

*Première itération* (ici, avec pivot partiel) : on échange la première et la troisième lignes, ce qui donne, avec le pivot en haut à gauche (en gras) :

$$\left\{ \begin{array}{rrrr} -\mathbf{2}x_1 & + & 4x_2 & - & 19x_3 & = & 0 & | & 0 & | & 1 \\ x_1 & - & 2x_2 & + & 10x_3 & = & 0 & | & 1 & | & 0 \\ x_1 & - & 3x_2 & + & 14x_3 & = & 1 & | & 0 & | & 0 \end{array} \right.$$

On élimine les termes de la première colonne sauf le terme diagonal. On obtient :

$$\left\{ \begin{array}{rrrr} -\mathbf{2}x_1 & + & 4x_2 & - & 19x_3 & = & 0 & | & 0 & | & 1 \\ & & 0x_2 & + & 1/2 x_3 & = & 0 & | & 1 & | & 1/2 \\ & & - & x_2 & + & 9/2 x_3 & = & 1 & | & 0 & | & 1/2 \end{array} \right.$$

*Deuxième itération* (maintenant, avec pivot total pour illustrer cette variante) : le plus grand coefficient en valeur absolue étant  $9/2$ , en bas à droite,

on échange la deuxième et la troisième lignes ainsi que la deuxième et la troisième colonnes. On obtient :

$$\left\{ \begin{array}{rrrr} -2x_1 & - & 19x_3 & + & 4x_2 & = & 0 & \left| \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right| & \begin{array}{c} 1 \\ 1/2 \\ 1/2 \end{array} \\ & & \mathbf{9/2} & x_3 & - & x_2 & = & 1 & \\ & & 1/2 & x_3 & + & 0x_2 & = & 0 & \end{array} \right.$$

On élimine les termes de la deuxième colonne sauf le terme diagonal. On obtient :

$$\left\{ \begin{array}{rrrr} -2x_1 & & - & 2/9 & x_2 & = & 38/9 & \left| \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right| & \begin{array}{c} 28/9 \\ 1/2 \\ 4/9 \end{array} \\ & 9/2 & x_3 & - & x_2 & = & 1 & \\ & & & \mathbf{1/9} & x_2 & = & -1/9 & \end{array} \right.$$

*Troisième et dernière itération* ; le pivot ne peut être que l'élément de la ligne non encore traitée : il s'agit du 1/9 en bas à droite. On obtient :

$$\left\{ \begin{array}{rrrr} -2x_1 & & & = & 4 & \left| \begin{array}{c} 2 \\ 9 \\ 1 \end{array} \right| & \begin{array}{c} 4 \\ 9/2 \\ 4/9 \end{array} \\ & 9/2 & x_3 & = & 0 & \\ & & & 1/9 & x_2 & = & -1/9 & \end{array} \right.$$

On peut maintenant résoudre les trois systèmes immédiatement :

$$\left\{ \begin{array}{rrr} x_1 & = & -2 \\ x_2 & = & -1 \\ x_3 & = & 0 \end{array} \left| \begin{array}{c} -1 \\ 9 \\ 2 \end{array} \right| \begin{array}{c} -2 \\ 4 \\ 1 \end{array} \right.$$

On déduit l'inverse de  $A$  de ces calculs :  $A^{-1} = \begin{pmatrix} -2 & -1 & -2 \\ -1 & 9 & 4 \\ 0 & 2 & 1 \end{pmatrix}$ .

Comme il y a deux échanges de lignes et un échange de colonnes, le déterminant de  $A$  vaut :  $(-1)^3 \times (-2) \times \frac{9}{2} \times \frac{1}{9} = 1$ .

### 3.3 Factorisation LU

Dans la méthode de Gauss, on transforme  $Ax = b$  en  $MAx = Mb$  où  $MA$  est une matrice triangulaire supérieure, que nous noterons  $U$  (pour *upper*). Supposons que, dans cette construction, le pivot se trouve sur la diagonale, ce qui équivaut à dire que le terme qui apparaît dans la case d'indices  $(k, k)$



après  $k - 1$  ( $1 \leq k \leq n$ ) étapes ne vaut jamais 0 (c'est largement le cas général).

Soit  $k$  un indice vérifiant  $1 \leq k \leq n - 1$ . Notons  $M_k$  la matrice du système obtenue après  $k - 1$  itérations, avec  $M_1 = A$ ; cette matrice a des 0 sous les  $(k - 1)$  premières valeurs de la diagonale (i.e. pour tout couple d'indices  $(s, t)$  avec  $1 \leq t \leq k - 1, s \geq t$ ) et, par hypothèse,  $(M_k)_{k,k}$  est non nul. On a  $M = M_n$ . Pour  $1 \leq i \leq n$ , posons  $\alpha_i = (M_k)_{i,k}$ ; ainsi, la  $k^e$  colonne de  $M_k$  (la colonne du pivot) est :

$$\begin{pmatrix} \alpha_1 \\ \dots \\ \alpha_k \neq 0 \\ \dots \\ \alpha_n \end{pmatrix}.$$

Soit  $E_k$  la matrice qui possède des 1 sur la diagonale et ailleurs des 0 sauf pour  $(E_k)_{i,k}$  avec  $i > k$  (partie de la  $k^e$  colonne située sous la diagonale) : pour  $i > k$ , on pose  $(E_k)_{i,k} = -\frac{\alpha_i}{\alpha_k}$ . Cette matrice  $E_k$  est donc triangulaire inférieure et ne diffère de la matrice identité que par sa  $k^e$  colonne :

$$E_k = \begin{pmatrix} 1 & 0 & \dots & & & \\ 0 & 1 & 0 & \dots & & \\ & & \dots & 0 & \dots & \\ & & & 1 & 0 & \dots \\ & & & -\frac{\alpha_{k+1}}{\alpha_k} & 1 & 0 & \dots \\ & & & \dots & & & \\ & & & -\frac{\alpha_n}{\alpha_k} & 0 & \dots & 1 \end{pmatrix}.$$

Par la méthode de Gauss, on passe alors de la matrice  $M_k$  à la matrice  $M_{k+1}$  en multipliant  $M_k$  à gauche par la matrice  $E_k$ . D'où  $M_{k+1} = E_k M_k$  et  $M = E_{n-1} E_{n-2} \dots E_1$ . Le produit de matrices triangulaires inférieures dont la diagonale ne contient que des 1 étant aussi une matrice triangulaire inférieure avec une diagonale de 1, la matrice  $M$  est elle-même triangulaire inférieure avec une diagonale de 1. Elle est donc inversible.

La relation  $MA = U$  donne :  $A = M^{-1}U$ . On pose  $L = M^{-1}$ . D'où  $A = LU$ . L'inverse d'une matrice triangulaire inférieure avec une diagonale de 1 étant aussi une matrice triangulaire inférieure avec une diagonale de 1, on conclut que  $L$  est elle-même une matrice triangulaire inférieure avec une diagonale de 1 (d'où la notation  $L$ , pour *lower*).

On pose maintenant, pour  $1 \leq k < i \leq n$ ,  $\beta_{i,k} = \frac{\alpha_i}{\alpha_k}$  :  $\beta_{i,k}$  est l'opposé du facteur par lequel, à la  $k^e$  étape, on multiplie la  $k^e$  ligne du système (la ligne du pivot) pour obtenir, en soustrayant cette  $k^e$  ligne ainsi multipliée à la  $i^e$  ligne, la nouvelle  $i^e$  ligne. On vérifie facilement l'expression de  $E_k^{-1}$  :

$$E_k^{-1} = \begin{pmatrix} 1 & 0 & \dots & & & \\ 0 & 1 & 0 & \dots & & \\ & & \dots & 0 & \dots & \\ & \dots & 0 & 1 & 0 & \dots \\ & & & \beta_{k+1,k} & 1 & 0 & \dots \\ & & & \dots & & & \\ & & & \beta_{n,k} & 0 & \dots & 1 \end{pmatrix}.$$

Or, on a  $L = M^{-1} = E_1^{-1} \dots E_{n-2}^{-1} E_{n-1}^{-1}$ . On vérifie facilement aussi le résultat suivant :

$$L = E_1^{-1} \dots E_{n-2}^{-1} E_{n-1}^{-1} = \begin{pmatrix} 1 & 0 & \dots & & & \\ \beta_{2,1} & 1 & 0 & \dots & & \\ \beta_{3,1} & \beta_{3,2} & 1 & 0 & \dots & \\ & & & \dots & & \\ \beta_{n,1} & \beta_{n,2} & \beta_{n,3} & \dots & \beta_{n,n-1} & 1 \end{pmatrix}.$$

On obtient ainsi la décomposition dite *factorisation LU* de  $A$  :  $A = LU$ , avec  $L$  matrice triangulaire inférieure dont la diagonale ne contient que des 1 et  $U$  matrice triangulaire supérieure.

*Exemple*

On reprend l'exemple 1 de la méthode de Gauss avec :  $A = \begin{pmatrix} 2 & 1 & -3 \\ 4 & 1 & 5 \\ 10 & -7 & 13 \end{pmatrix}$ .

La résolution du système donne l'expression suivante pour  $U$  :

$$U = \begin{pmatrix} 2 & 1 & -3 \\ 0 & -1 & 11 \\ 0 & 0 & -104 \end{pmatrix}.$$

Les paragraphes ci-dessus et l'observation des facteurs utilisés pour faire apparaître les 0 sous la diagonale donne l'expression suivante pour  $L$  :

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 5 & 12 & 1 \end{pmatrix}.$$

Les considérations précédentes reposent sur l'hypothèse selon laquelle, dans l'application de la méthode de Gauss, le pivot se trouve sur la diagonale (*cf.* plus haut). Le théorème suivant donne une condition suffisante pour que cette hypothèse soit vérifiée.

### **Théorème d'existence de la factorisation LU**

Soit  $A = (a_{ij})$  une matrice carrée (inversible) telle que, pour tout  $k$  compris

entre 1 et  $n$ , la sous-matrice  $\begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{pmatrix}$  soit inversible. Alors, la facto-

risation  $A = LU$  est possible (plus précisément, les pivots successifs peuvent toujours être pris sur la diagonale, sans échange de lignes). De plus, on peut choisir  $(L)_{ii} = 1$  et la décomposition est alors unique.

En fait, on peut montrer que si la factorisation LU échoue (c'est-à-dire si les pivots ne peuvent pas être toujours choisis sur la diagonale sans échange de lignes), on peut permuter au départ les lignes de la matrice  $A$  pour obtenir une matrice  $A'$  pour laquelle la factorisation LU est possible.

Lorsque l'on doit résoudre plusieurs systèmes linéaires de même matrice  $A$ , on calcule la factorisation LU lors de la résolution du premier de ces systèmes. La résolution de tout système ultérieur  $Ax = b$  se ramène à la résolution de deux systèmes de matrices triangulaires : le système  $Ly = b$  puis le système  $Ux = y$  (on notera ainsi qu'il est inutile de connaître  $M$  explicitement, dont le calcul n'est pas nécessairement aisé). Chaque système ne prend plus alors que  $n(n-1)$  additions,  $n(n-1)$  multiplications et  $2n$  divisions.

## **3.4 Méthode de Cholesky**

La méthode de Cholesky donne une factorisation intéressante dans le cas des matrices symétriques définies positives. Dans ce cas, on peut choisir une factorisation LU avec  $U = L^t$  en renonçant néanmoins à avoir des termes diagonaux tous égaux à 1 dans  $L$ .

**Théorème.** Soit  $A$  une matrice symétrique définie positive. Il existe une matrice triangulaire inférieure  $B$  vérifiant  $A = BB^t$ . De plus, on peut imposer que les éléments diagonaux de la matrice  $B$  soient tous strictement positifs et la factorisation  $A = BB^t$  est alors unique.

En pratique, on calcule la matrice  $B = \begin{pmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & 0 & \dots & 0 \\ & & \dots & & \\ b_{n1} & b_{n2} & \dots & & b_{nn} \end{pmatrix}$  colonne par colonne, à partir des égalités la définissant : pour  $1 \leq i \leq j \leq n$ ,

$$a_{ij} = \sum_{k=1}^i b_{ik} b_{jk} = a_{ji}.$$

- Pour la première colonne, la formule donne

$$\begin{aligned} \star \quad b_{11} &= \sqrt{a_{11}} \\ \star \quad \text{pour } 2 \leq i \leq n, \quad b_{i1} &= \frac{a_{i1}}{b_{11}}. \end{aligned}$$

- Pour  $2 \leq j \leq n$ ,

$$\begin{aligned} \star \quad \text{sur la diagonale, } b_{jj} &= \sqrt{a_{jj} - \sum_{k=1}^{j-1} b_{jk}^2} \\ \star \quad \text{pour } j < i \leq n, \quad b_{ij} &= \frac{a_{ij} - \sum_{k=1}^{j-1} b_{ik} b_{jk}}{b_{jj}}. \end{aligned}$$

*Remarques.*

1. La preuve du théorème précédent permettrait de montrer que les  $b_{ij}$  ainsi obtenus sont bien définis, grâce au fait que  $A$  est définie positive.
2. Le déterminant de la matrice  $A$  peut se calculer facilement :

$$\det(A) = (b_{11} b_{22} \dots b_{nn})^2.$$

Un système  $Ax = b$  devient alors  $BB^t x = b$ . Pour résoudre le système, on résout  $By = b$  puis  $B^t x = y$ .

*Complexité.* Au total (la factorisation et les deux résolutions), on a effectué de l'ordre de  $n^3/6$  additions,  $n^3/6$  multiplications,  $n^2/2$  divisions,  $n$  extractions

de racines carrées, soit de l'ordre de  $n^3/3$  opérations, c'est-à-dire environ la moitié des opérations mises en œuvre par la méthode de Gauss. On a donc intérêt à appliquer la méthode de Cholesky plutôt que la méthode de Gauss quand  $A$  est symétrique définie positive.

*Exemple.*

Considérons le système suivant :

$$\begin{cases} 4x_1 - 2x_2 &= 4 \\ -2x_1 + 2x_2 + 3x_3 &= -8 \\ 3x_2 + 10x_3 &= -20 \end{cases}$$

La matrice  $A$  correspondante est :  $A = \begin{pmatrix} 4 & -2 & 0 \\ -2 & 2 & 3 \\ 0 & 3 & 10 \end{pmatrix}$ . Cette matrice est symétrique définie positive. En effet, soit  $x$  un vecteur de  $\mathbb{R}^n$  représenté comme un vecteur colonne. On a alors :

$$\begin{aligned} x^t A x &= 4x_1^2 - 4x_1x_2 + 2x_2^2 + 6x_2x_3 + 10x_3^2 \\ &= (2x_1 - x_2)^2 + (x_2 + 3x_3)^2 + x_3^2. \end{aligned}$$

Par conséquent, si  $x$  est non nul,  $x^t A x$  est un réel strictement positif.

*Première étape :* on calcule  $B$  telle que  $A = BB^t$  avec  $B$  triangulaire supérieure. L'application des formules précédentes donne :

$$\begin{aligned} b_{11} &= \sqrt{a_{11}} = 2 \\ b_{21} &= \frac{a_{21}}{b_{11}} = -1 \\ b_{31} &= \frac{a_{31}}{b_{11}} = 0 \\ b_{22} &= \sqrt{a_{22} - \sum_{k=1}^1 b_{2k}^2} = \sqrt{2 - 1} = 1 \\ b_{32} &= \frac{a_{32} - \sum_{k=1}^1 b_{3k}b_{2k}}{b_{22}} = \frac{3 - 0 \times (-1)}{1} = 3 \\ b_{33} &= \sqrt{a_{33} - \sum_{k=1}^2 b_{3k}^2} = \sqrt{10 - 0 - 9} = 1. \end{aligned}$$

D'où :  $B = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}$  et  $\det(A) = (2 \times 1 \times 1)^2 = 4$ .

*Seconde étape* : par la méthode de remontée, on résout les deux systèmes  $By = b$  et  $B^t x = y$ .

Le système  $By = b$  s'écrit :

$$\begin{cases} 2y_1 & & & = & 4 \\ -y_1 & + & y_2 & & = & -8 \\ & & 3y_2 & + & y_3 & = & -20 \end{cases}$$

qui a pour solution  $y_1 = 2, y_2 = -6, y_3 = -2$ .

Le système  $B^t x = y$  s'écrit :

$$\begin{cases} 2x_1 & - & x_2 & & = & 2 \\ & & x_2 & + & 3x_3 & = & -6 \\ & & & & x_3 & = & -2 \end{cases}$$

qui a pour solution  $x_3 = -2, x_2 = 0, x_1 = 1$ .

La solution du système est donc :  $x = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$ .

Si on avait un autre système à résoudre avec la même matrice  $A$ , seule la seconde étape serait appliquée.

## Chapitre 4

# Valeurs propres et vecteurs propres

Remarquons d'abord que la recherche des valeurs propres d'une matrice, au contraire du calcul de son inverse, est un problème difficile. Étant donné le polynôme  $P(\lambda) = \lambda^n + a_1\lambda^{n-1} + \dots + a_{n-1}\lambda + a_n$ , définissons la matrice :

$$\begin{pmatrix} -a_1 & -a_2 & -a_3 & \dots & -a_{n-1} & -a_n \\ 1 & 0 & & & & \\ 0 & 1 & 0 & & & \\ & 0 & 1 & 0 & & \\ & & & \dots & & \\ & & & 0 & 1 & 0 \\ & & & & 0 & 1 & 0 \end{pmatrix}$$

Cette matrice est dite « compagne du polynôme »  $P$ . Son polynôme caractéristique vaut  $(-1)^n P(\lambda)$  ; la matrice a donc pour valeurs propres les racines de  $P$ . Or, d'après le théorème d'Abel, il est impossible de calculer les racines de tout polynôme à partir du degré 5 à l'aide d'un nombre fini d'applications des quatre opérations arithmétiques usuelles plus l'extraction de racines. Si une méthode de recherche de valeurs propres convergerait toujours en un nombre fini de ces opérations, il en serait alors de même de la recherche des racines d'une équation polynomiale quelconque, ce qui est contraire au résultat d'Abel. En revanche, il est courant, pour déterminer les racines d'un polynôme  $P$ , de chercher les valeurs propres de la compagne de  $P$ .

Pour calculer une approximation des valeurs propres d'une matrice  $A$ , l'idée de base est de rechercher une matrice semblable à  $A$ , c'est-à-dire de

la forme  $P^{-1}AP$ , triangulaire ou diagonale, et dont la diagonale sera donc constituée des valeurs propres de  $A$ . Nous étudierons dans ce chapitre une seule méthode, la méthode de Jacobi, qui s'applique au cas des matrices symétriques réelles. Rappelons que les valeurs propres d'une telle matrice sont réelles.

## 4.1 Méthode de Jacobi

Soit  $A$  une matrice symétrique réelle, soient deux indices  $p$  et  $q$  vérifiant  $p < q$  tels que l'élément (non diagonal)  $a_{pq}$  soit non nul (s'il n'en existe pas,  $A$  est diagonale et les valeurs propres de  $A$  sont précisément les valeurs de la diagonale).

Soit  $\theta$  un nombre réel ; on définit une matrice  $\Omega$  dépendant de  $\theta$ . La matrice  $\Omega$  diffère de la matrice identité d'ordre  $n$  uniquement par les quatre coefficients suivants :

$$\Omega_{pp} = \Omega_{qq} = \cos \theta, \Omega_{pq} = \sin \theta, \Omega_{qp} = -\sin \theta.$$

La matrice  $\Omega$  est représentée ci-dessous.

$$\Omega = \begin{pmatrix} 1 & 0 & & \dots & & 0 & 0 \\ 0 & 1 & & \dots & & 0 & 0 \\ & & \dots & & & & \\ & & & \cos \theta & & \sin \theta & \\ & & & & 1 & & \\ & & & & & \dots & \\ & & & & & & 1 \\ & & & -\sin \theta & & \cos \theta & \\ & & \dots & & & & \dots \\ 0 & 0 & & \dots & & 1 & 0 \\ 0 & 0 & & \dots & & 0 & 1 \end{pmatrix}.$$

La matrice  $\Omega$  est orthogonale. C'est la matrice de rotation d'angle  $-\theta$  dans le plan défini par les  $p^e$  et  $q^e$  vecteurs de base.

On pose :  $B = \Omega^t A \Omega$ . La matrice  $B$ , elle aussi symétrique, est semblable à la matrice  $A$  et admet donc les mêmes valeurs propres que  $A$ . On établit



facilement les égalités suivantes :

$$\left\{ \begin{array}{l} \text{si } i \notin \{p, q\} \text{ et } j \notin \{p, q\}, \quad b_{ij} = b_{ji} = a_{ij} \\ \text{si } i \notin \{p, q\}, \quad b_{pi} = b_{ip} = a_{pi} \cos \theta - a_{qi} \sin \theta \\ \text{si } i \notin \{p, q\}, \quad b_{qi} = b_{iq} = a_{pi} \sin \theta + a_{qi} \cos \theta \\ b_{pp} = a_{pp} \cos^2 \theta + a_{qq} \sin^2 \theta - a_{pq} \sin 2\theta \\ b_{qq} = a_{pp} \sin^2 \theta + a_{qq} \cos^2 \theta + a_{pq} \sin 2\theta \\ b_{pq} = b_{qp} = a_{pq} \cos 2\theta + \frac{a_{pp} - a_{qq}}{2} \sin 2\theta. \end{array} \right.$$

On remarque l'équivalence  $b_{pq} = 0 \Leftrightarrow \cot 2\theta = \frac{a_{qq} - a_{pp}}{2a_{pq}}$  (où  $\cot$  désigne la fonction trigonométrique cotangente). On essaie de faire en sorte d'avoir  $b_{pq} = 0$  et on choisit donc  $\theta$  pour qu'il vérifie la formule ci-dessus. Il y a quatre solutions dans l'intervalle  $]-\pi, \pi]$ , deux solutions successives différant de  $\pi/2$ . Il y a donc une unique solution dans l'intervalle  $]-\frac{\pi}{4}, \frac{\pi}{4}]$ , c'est la solution retenue.

Posons maintenant :  $x = \frac{a_{qq} - a_{pp}}{2a_{pq}}, t = \tan \theta, s = \sin \theta, c = \cos \theta$ . On rappelle les relations trigonométriques suivantes :

$$\cot 2\theta = \frac{\cos 2\theta}{\sin 2\theta} = \frac{\cos^2 \theta - \sin^2 \theta}{2 \sin \theta \cos \theta} = \frac{1 - t^2}{2t}.$$

On cherche à avoir :  $x = \frac{1 - t^2}{2t}$  ; il en résulte que  $t$  doit vérifier l'équation :  $t^2 + 2xt - 1 = 0$ . Comme le produit des racines vaut  $-1$  et que  $\theta$  est dans l'intervalle  $]-\frac{\pi}{4}, \frac{\pi}{4}]$ ,  $t$  est la racine de l'équation de plus petite valeur absolue si les racines ne sont pas  $1$  et  $-1$ , et vaut  $1$  si  $x = 0$ .

Comme on a  $c > 0$ , il vient  $c = \frac{1}{\sqrt{1 + t^2}}$  et  $s = ct = \frac{t}{\sqrt{1 + t^2}}$ .

Les coefficients de la matrice  $B$  peuvent finalement être calculés par les

formules suivantes, dans lesquelles  $t$ ,  $c$  et  $s$  sont définis comme ci-dessus :

$$\begin{cases} \text{si } i \notin \{p, q\} \text{ et } j \notin \{p, q\}, b_{ij} = b_{ji} = a_{ij} \\ \text{si } i \notin \{p, q\}, b_{pi} = b_{ip} = ca_{pi} - sa_{qi} \\ \text{si } i \notin \{p, q\}, b_{qi} = b_{iq} = sa_{pi} + ca_{qi} \\ b_{pp} = a_{pp} - ta_{pq} \\ b_{qq} = a_{qq} + ta_{pq}. \end{cases}$$

*Une étape de la méthode de Jacobi, résumé*

- On choisit dans la matrice courante deux indices  $p$  et  $q$ , avec  $p < q$ .
- On pose  $x = \frac{a_{qq} - a_{pp}}{2a_{pq}}$ .
- On résout :  $t^2 + 2xt - 1 = 0$ . On retient pour  $t$  la **racine de plus petite valeur absolue** si les racines ne sont pas 1 et  $-1$ , et 1 **si  $x = 0$** .
- On calcule :  $c = \frac{1}{\sqrt{1+t^2}}$  et  $s = \frac{t}{\sqrt{1+t^2}}$ .
- On calcule les nouveaux coefficients avec les formules ci-dessus qui utilisent  $c$ ,  $s$  et  $t$ .

*Remarque.* Il est naturel de se demander si, en faisant cette transformation qui a le mérite d'annuler des éléments non diagonaux, on ne risque pas, ce faisant, de rendre en même temps non nuls des éléments qui étaient précédemment nuls. Cela est non seulement vrai, mais même inévitable, puisque, dans le cas contraire, on diagonaliserait la matrice  $A$  avec environ  $n^3$  opérations élémentaires, alors que nous avons indiqué dans la remarque initiale que cela était impossible. Il y a cependant de bonnes raisons d'espérer, en réitérant le procédé, une convergence des matrices  $B$  obtenues vers une matrice diagonale, comme nous allons l'expliquer ci-dessous.

**Théorème.** Soit  $A$  une matrice symétrique réelle et soit  $B$  la matrice obtenue à l'aide du procédé précédent. On a alors les relations :

$$\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 = \sum_{i=1}^n \sum_{j=1}^n b_{ij}^2,$$

$$\sum_{i=1}^n a_{ii}^2 + 2a_{pq}^2 = \sum_{i=1}^n b_{ii}^2.$$

*Preuve du théorème.* La première relation résulte de la conservation de la norme  $\| \cdot \|_E$  par une transformation unitaire. Quant à la seconde, seuls les éléments des lignes et colonnes  $p$  et  $q$  sont modifiés. Les éléments diagonaux

autres que  $a_{pp}$  et  $a_{qq}$  sont donc invariants ainsi que leurs carrés. On a :

$$\begin{aligned} b_{pp}^2 + b_{qq}^2 &= a_{pp}^2 + a_{qq}^2 + 2t^2 a_{pq}^2 + 2ta_{pq}(a_{qq} - a_{pp}) \\ &= a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2 + 2a_{pq}(t^2 a_{pq} + t(a_{qq} - a_{pp}) - a_{pq}). \end{aligned}$$

Or, le choix de  $t$  fait que l'on a  $t^2 + t \frac{a_{qq} - a_{pp}}{a_{pq}} - 1 = 0$ .

D'où le résultat énoncé :  $b_{pp}^2 + b_{qq}^2 = a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2$ . ■

Ce théorème montre que le poids de la matrice se déporte, au cours des itérations de la méthode de Jacobi, sur la diagonale de la matrice et, par conséquent, que les éléments non diagonaux, eux, ont un poids qui diminue. Par ailleurs, il semble que pour accélérer la convergence du procédé, on ait intérêt à choisir comme couple  $(p, q)$  les indices d'un élément non diagonal de valeur absolue maximum. C'est effectivement ce choix qui est fait dans la méthode de Jacobi dite classique.

**Théorème.** La suite des matrices obtenues par la méthode de Jacobi est convergente et converge vers une matrice diagonale contenant les valeurs propres de  $A$ .

La méthode de Jacobi permet aussi d'obtenir une approximation des vecteurs propres d'une matrice  $A$ , au moins quand les valeurs propres de  $A$  sont distinctes. C'est ce que précise le théorème suivant.

**Théorème.** Si toutes les valeurs propres de la matrice  $A$  sont distinctes, alors la suite des produits des matrices  $\Omega$  (en mettant à chaque étape la nouvelle matrice  $\Omega$  à droite du produit) converge vers une matrice orthogonale dont les vecteurs colonnes constituent un ensemble orthonormal de vecteurs propres de la matrice  $A$ .

*Exemple 1.* Appliquons la méthode de Jacobi à la recherche d'approximations

des valeurs propres et des vecteurs propres de la matrice  $A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 5 \end{pmatrix}$ . Il

n'y a que les coefficients  $p = 1$  et  $q = 2$  qui sont à considérer. Avec les notations précédentes, on a  $x = 0$  et donc  $t = 1, s = c = \frac{\sqrt{2}}{2}$ . Par conséquent, la matrice

$$\Omega \text{ vaut } \Omega = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

L'application des formules précédentes donne :

- terme inchangé (ici, un seul *a priori*) :  $b_{33} = a_{33} = 5$

- première ligne et première colonne, sauf diagonale :

$$b_{12} = b_{21} = 0$$

$$b_{13} = b_{31} = ca_{13} - sa_{23} = 0$$

- deuxième ligne et deuxième colonne, sauf diagonale :

$$b_{23} = b_{32} = sa_{13} + ca_{23} = 0$$

- termes diagonaux qui changent *a priori* :

$$b_{11} = a_{11} - ta_{12} = 1 - 2 = -1$$

$$b_{22} = a_{22} + ta_{12} = 1 + 2 = 3.$$

$$\text{On obtient donc : } B = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

La matrice est diagonale, la méthode de Jacobi converge ici en une itération (l'exemple est très simple) et nous donne les valeurs propres (exactement) ainsi que les vecteurs propres de  $A$ . Les valeurs propres de  $A$  valent :  $-1$ ,  $3$  et  $5$ . La base orthonormale de vecteurs propres est constituée des vecteurs :  $(\sqrt{2}/2, -\sqrt{2}/2, 0)^t$ ,  $(\sqrt{2}/2, \sqrt{2}/2, 0)^t$ ,  $(0, 0, 1)^t$ .

*Exemple 2.* Appliquons la méthode de Jacobi à la recherche d'approximations des valeurs propres et des vecteurs propres de la matrice  $A = \begin{pmatrix} 1 & 2 & 4 \\ 2 & -3 & -1 \\ 4 & -1 & 7 \end{pmatrix}$ .

*Première étape.*

Choisissons la plus grande valeur absolue d'un coefficient non diagonal : il s'agit de la valeur 4, avec  $p = 1, q = 3$ .

$$\text{On calcule } x : x = \frac{7-1}{2 \times 4} = \frac{3}{4}.$$

On résout l'équation  $t^2 + 2xt - 1 = 0$ , c'est-à-dire :  $t^2 + \frac{3}{2}t - 1 = 0$ , qui a pour racines  $t = 1/2$  et  $t = -2$ . On retient la plus petite racine en valeur absolue :  $t = 1/2$ .

$$\text{On calcule } c \text{ et } s : c = \frac{1}{\sqrt{1+t^2}} = \frac{2}{\sqrt{5}} = \frac{2\sqrt{5}}{5} \text{ et } s = tc = \frac{\sqrt{5}}{5}.$$

Puis on applique les formules donnant les coefficients de  $B$ , avec bien sûr

$b_{13} = b_{31} = 0 :$

$b_{22}$  reste inchangé :  $b_{22} = -3$

$$b_{12} = b_{21} = ca_{12} - sa_{32} = \frac{2\sqrt{5}}{5} \times 2 - \frac{\sqrt{5}}{5} \times (-1) = \sqrt{5}$$

$$b_{32} = b_{23} = sa_{12} + ca_{32} = \frac{\sqrt{5}}{5} \times 2 + \frac{2\sqrt{5}}{5} \times (-1) = 0$$

$$b_{11} = a_{11} - ta_{13} = 1 - \frac{1}{2} \times 4 = -1$$

$$b_{33} = a_{33} + ta_{13} = 7 + \frac{1}{2} \times 4 = 9.$$

On obtient ainsi la matrice  $B$  :

$$B = \begin{pmatrix} -1 & \sqrt{5} & 0 \\ \sqrt{5} & -3 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

et la matrice de passage  $\Omega_1$  :

$$\Omega_1 = \begin{pmatrix} 2\frac{\sqrt{5}}{5} & 0 & \frac{\sqrt{5}}{5} \\ 0 & 1 & 0 \\ -\frac{\sqrt{5}}{5} & 0 & 2\frac{\sqrt{5}}{5} \end{pmatrix} \approx \begin{pmatrix} 0,894 & 0 & 0,447 \\ 0 & 1 & 0 \\ -0,447 & 0 & 0,894 \end{pmatrix}.$$

*Seconde étape.* On repart de la matrice  $B$  pour passer à une matrice  $C$  calculée avec la méthode de Jacobi.

On pose :  $p = 1, q = 2$ .

On calcule  $x$  :  $x = \frac{-3+1}{2\sqrt{5}} = -\frac{\sqrt{5}}{5}$ .

On résout l'équation :  $t^2 - 2\frac{\sqrt{5}}{5}t - 1 = 0$  qui a pour racines :  $t = \frac{\sqrt{5}}{5}(1 + \sqrt{6})$  et  $t = \frac{\sqrt{5}}{5}(1 - \sqrt{6})$ . On retient la plus petite racine en valeur absolue :

$t = \frac{\sqrt{5}}{5}(1 - \sqrt{6}) \approx -0,648$ . En conséquence :  $c = \frac{1}{\sqrt{1+t^2}} \approx 0,839$  et

$s = ct \approx -0,544$ . On obtient alors :

$$c_{33} = b_{33} = 9$$

$$c_{12} = c_{21} = 0$$

$$c_{11} = b_{11} - tb_{12} = -1 - \frac{\sqrt{5}}{5}(1 - \sqrt{6})\sqrt{5} = -2 + \sqrt{6}$$

$$c_{22} = b_{22} + tb_{12} = -3 + \frac{\sqrt{5}}{5}(1 - \sqrt{6})\sqrt{5} = -2 - \sqrt{6}$$

$$c_{13} = c_{31} = cb_{13} - sb_{23} = 0$$

$$c_{23} = c_{32} = sb_{13} + cb_{23} = 0.$$

D'où  $C$  :

$$C = \begin{pmatrix} -2 + \sqrt{6} & 0 & 0 \\ 0 & -2 - \sqrt{6} & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

La matrice de passage  $\Omega_2$  approchée est donnée par :

$$\Omega_2 \approx \begin{pmatrix} 0,839 & -0,544 & 0 \\ 0,544 & 0,839 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

La matrice  $C$  étant diagonale, la méthode est terminée. Les valeurs propres de  $A$  valent :  $-2 + \sqrt{6}$ ,  $-2 - \sqrt{6}$ , 9.

Une base orthonormale approchée de vecteurs propres s'obtient en calculant le produit  $\Omega_1 \Omega_2$  :  $\Omega_1 \Omega_2 \approx \begin{pmatrix} 0,75 & -0,486 & 0,447 \\ 0,544 & 0,839 & 0 \\ -0,375 & 0,243 & 0,894 \end{pmatrix}$ .

*Exemple 3.* Appliquons la méthode de Jacobi à la recherche d'approximations des valeurs propres et des vecteurs propres de la matrice  $A = \begin{pmatrix} 1 & 2 & 4 \\ 2 & -3 & 0 \\ 4 & 0 & 7 \end{pmatrix}$ .

Nous choisissons la plus grande valeur absolue d'un coefficient non diagonal. Il s'agit de la valeur 4. On pose :  $p = 1, q = 3$ .

On calcule comme dans l'exemple 2 :  $t = 1/2$ ,  $c = \frac{2\sqrt{5}}{5}$  et  $s = \frac{\sqrt{5}}{5}$ .

On a :

$$b_{22} = -3$$

$$b_{13} = b_{31} = 0$$

$$b_{12} = b_{21} = ca_{12} - sa_{32} = \frac{2\sqrt{5}}{5} \times 2 - \frac{\sqrt{5}}{5} \times 0 = \frac{4\sqrt{5}}{5}$$

$$b_{32} = b_{23} = sa_{12} + ca_{32} = \frac{\sqrt{5}}{5} \times 2 + \frac{2\sqrt{5}}{5} \times 0 = \frac{2\sqrt{5}}{5}$$

$$b_{11} = a_{11} - ta_{13} = -1$$

$$b_{33} = a_{33} + ta_{13} = 9.$$

On obtient donc :  $B = \begin{pmatrix} -1 & 4\frac{\sqrt{5}}{5} & 0 \\ 4\frac{\sqrt{5}}{5} & -3 & 2\frac{\sqrt{5}}{5} \\ 0 & 2\frac{\sqrt{5}}{5} & 9 \end{pmatrix}.$

Cet exemple montre que des coefficients peuvent passer de nuls à non nuls. Néanmoins, en passant de  $A$  à  $B$ , le poids de la matrice s'est concentré sur la diagonale. Il faudrait poursuivre la méthode pour calculer une approximation des valeurs propres et des vecteurs propres de la matrice  $A$ .

## 4.2 Bibliographie

Nous nous sommes inspirés de l'ouvrage suivant pour rédiger ce polycopié :

P. G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, Masson, 1985.

# Index

- complexité, 13
- conditionnement
  - d'un système linéaire, 14
  - d'une matrice, 15
  - pour des valeurs propres, 17
- erreur, 13
  - d'arrondi, 13
  - de troncature, 13
- factorisation LU, 24, 26
- inégalité
  - de Cauchy-Schwarz, 9
  - de Hölder, 9
  - de Minkowski, 9
- matrice
  - adjointe, 5
  - convergence, 11
  - de passage, 7
  - diagonalisable, 7
  - équilibrage, 17
  - équivalente, 8
  - hermitienne, 6
  - normale, 6
  - orthogonale, 6
  - semblable, 7
  - symétrique, 6
  - trace, 10
  - unitaire, 6
- méthode
  - de Cholesky, 27
  - de Gauss, 20
  - de Gauss-Jordan, 23
  - de remontée, 19
- norme, 8
  - équivalente, 9
  - matricielle, 9
    - euclidienne, 10
    - subordonnée, 9
  - vectorielle, 8
    - euclidienne, 8
    - infinie, 8
    - norme 1, 8
- pivot, 20
- produit
  - hermitien, 5
  - scalaire euclidien, 5
- rayon spectral, 6
- spectre, 6
- systèmes linéaires, 19
- valeur
  - propre, 6, 31
  - singulière, 8
- vecteur
  - adjoint, 5
  - propre, 6, 31