

**Université Aix Marseille**

**Licence de mathématiques**

**Cours d'Analyse numérique**

Raphaèle Herbin

10 septembre 2015

# Table des matières

<b>1</b>	<b>Systèmes linéaires</b>	<b>5</b>
1.1	Objectifs . . . . .	5
1.2	Pourquoi et comment ? . . . . .	5
1.2.1	Quelques rappels d’algèbre linéaire . . . . .	5
1.2.2	Discrétisation de l’équation de la chaleur . . . . .	11
1.2.3	Exercices . . . . .	15
1.2.4	Suggestions pour les exercices . . . . .	21
1.2.5	Corrigés des exercices . . . . .	21
1.3	Les méthodes directes . . . . .	28
1.3.1	Définition . . . . .	28
1.3.2	Méthode de Gauss, méthode <i>LU</i> . . . . .	28
1.3.3	Méthode de Choleski . . . . .	36
1.3.4	Quelques propriétés . . . . .	42
1.3.5	Exercices . . . . .	44
1.3.6	Suggestions . . . . .	48
1.3.7	Corrigés . . . . .	48
1.4	Normes et conditionnement d’une matrice . . . . .	60
1.4.1	Normes, rayon spectral . . . . .	61
1.4.2	Le problème des erreurs d’arrondis . . . . .	66
1.4.3	Conditionnement et majoration de l’erreur d’arrondi . . . . .	66
1.4.4	Discrétisation d’équations différentielles, conditionnement “efficace” . . . . .	70
1.4.5	Exercices . . . . .	71
1.4.6	Suggestions pour les exercices . . . . .	78
1.4.7	Corrigés . . . . .	79
1.5	Méthodes itératives . . . . .	95
1.5.1	Définition et propriétés . . . . .	95
1.5.2	Quelques exemples de méthodes itératives . . . . .	97
1.5.3	Les méthodes par blocs . . . . .	103
1.5.4	Exercices, énoncés . . . . .	106
1.5.5	Exercices, suggestions . . . . .	113
1.5.6	Exercices, corrigés . . . . .	114
1.6	Valeurs propres et vecteurs propres . . . . .	133
1.6.1	Méthode de la puissance et de la puissance inverse . . . . .	133
1.6.2	Méthode QR . . . . .	134
1.6.3	Exercices . . . . .	135
1.6.4	Suggestions . . . . .	139
1.6.5	Corrigés . . . . .	139

<b>2</b>	<b>Systèmes non linéaires</b>	<b>144</b>
2.1	Rappels et notations de calcul différentiel . . . . .	144
2.2	Les méthodes de point fixe . . . . .	147
2.2.1	Point fixe de contraction . . . . .	147
2.2.2	Point fixe de monotonie . . . . .	151
2.2.3	Vitesse de convergence . . . . .	153
2.2.4	Méthode de Newton dans $\mathbb{R}$ . . . . .	155
2.2.5	Exercices . . . . .	156
2.3	Méthode de Newton dans $\mathbb{R}^n$ . . . . .	164
2.3.1	Construction et convergence de la méthode . . . . .	164
2.3.2	Variantes de la méthode de Newton . . . . .	166
2.3.3	Exercices . . . . .	169
<b>3</b>	<b>Optimisation</b>	<b>204</b>
3.1	Définitions et rappels . . . . .	204
3.1.1	Extrema, points critiques et points selle. . . . .	204
3.1.2	Convexité . . . . .	206
3.1.3	Exercices . . . . .	208
3.2	Optimisation sans contrainte . . . . .	210
3.2.1	Définition et condition d'optimalité . . . . .	210
3.2.2	Résultats d'existence et d'unicité . . . . .	211
3.2.3	Exercices . . . . .	214
3.3	Algorithmes d'optimisation sans contrainte . . . . .	220
3.3.1	Méthodes de descente . . . . .	220
3.3.2	Algorithme du gradient conjugué . . . . .	224
3.3.3	Méthodes de Newton et Quasi-Newton . . . . .	227
3.3.4	Résumé sur les méthodes d'optimisation . . . . .	230
3.3.5	Exercices . . . . .	231
3.4	Optimisation sous contraintes . . . . .	258
3.4.1	Définitions . . . . .	258
3.4.2	Existence – Unicité – Conditions d'optimalité simple . . . . .	258
3.4.3	Conditions d'optimalité dans le cas de contraintes égalité . . . . .	259
3.4.4	Contraintes inégalités . . . . .	262
3.4.5	Exercices . . . . .	263
3.5	Algorithmes d'optimisation sous contraintes . . . . .	267
3.5.1	Méthodes de gradient avec projection . . . . .	267
3.5.2	Méthodes de dualité . . . . .	269
3.5.3	Exercices . . . . .	272
<b>4</b>	<b>Equations différentielles</b>	<b>275</b>
4.1	Introduction . . . . .	275
4.2	Consistance, stabilité et convergence . . . . .	278
4.3	Théorème général de convergence . . . . .	280
4.4	Exemples . . . . .	282
4.5	Explicite ou implicite ? . . . . .	283
4.5.1	L'implicite gagne... . . . .	283
4.5.2	L'implicite perd... . . . .	284
4.5.3	Match nul . . . . .	284
4.6	Etude du schéma d'Euler implicite . . . . .	284
4.7	Exercices . . . . .	286
4.8	Corrigés . . . . .	294

# Introduction

L'objet de l'analyse numérique est de concevoir et d'étudier des méthodes de résolution de certains problèmes mathématiques, en général issus de la modélisation de problèmes "réels", et dont on cherche à calculer la solution à l'aide d'un ordinateur.

Le cours est structuré en quatre grands chapitres :

- Systèmes linéaires
- Systèmes non linéaires
- Optimisation
- Equations différentielles.

On pourra consulter les ouvrages suivants pour ces différentes parties (ceci est une liste non exhaustive !) :

- A. Quarteroni, R. Sacco et F. Saleri, Méthodes Numériques : Algorithmes, Analyse et Applications, Springer 2006.
- P.G. Ciarlet, Introduction à l'analyse numérique et à l'optimisation, Masson, 1982, (pour les chapitre 1 à 3 de ce polycopié).
- M. Crouzeix, A.L. Mignot, Analyse numérique des équations différentielles, Collection mathématiques appliquées pour la maîtrise, Masson, (pour le chapitre 4 de ce polycopié).
- J.P. Demailly, Analyse numérique et équations différentielles Collection Grenoble sciences Presses Universitaires de Grenoble
- L. Dumas, Modélisation à l'oral de l'agrégation, calcul scientifique, Collection CAPES/Agrégation, Ellipses, 1999.
- E. Hairer, polycopié du cours "Analyse Numérique", <http://www.unige.ch/hairer/polycop.html>
- J. Hubbard, B. West, Equations différentielles et systèmes dynamiques, Cassini.
- J. Hubbard et F. Hubert, Calcul Scientifique, Vuibert.
- P. Lascaux et R. Théodor, Analyse numérique matricielle appliquée à l'art de l'ingénieur, tomes 1 et 2, Masson, 1987
- L. Sainsaulieu, Calcul scientifique cours et exercices corrigés pour le 2ème cycle et les écoles d'ingénieurs, Enseignement des mathématiques, Masson, 1996.
- M. Schatzman, Analyse numérique, cours et exercices, (chapitres 1,2 et 4).
- D. Serre, Les matrices, Masson, (2000). (chapitres 1,2 et 4).
- P. Lascaux et R. Theodor, Analyse numérique appliquée aux sciences de l'ingénieur, Paris, (1994)
- R. Temam, Analyse numérique, Collection SUP le mathématicien, Presses Universitaires de France, 1970.

Et pour les anglophiles...

- M. Braun, Differential Equations and their applications, Springer, New York, 1984 (chapitre 4).
- G. Dahlquist and A. Björck, Numerical Methods, Prentice Hall, Series in Automatic Computation, 1974, Englewood Cliffs, NJ.

- R. Fletcher, Practical methods of optimization, J. Wiley, New York, 1980 (chapitre 3).
- G. Golub and C. Van Loan, Matrix computations, The John Hopkins University Press, Baltimore (chapitre 1).
- R.S. Varga, Matrix iterative analysis, Prentice Hall, Englewood Cliffs, NJ 1962.

Pour des rappels d'algèbre linéaire :

- Poly d'algèbre linéaire de première année, P. Bousquet, R. Herbin et F. Hubert, <http://www.cmi.univ-mrs.fr/herbin/PUBLI/L1alg.pdf>
- Introduction to linear algebra, Gilbert Strang, Wellesley Cambridge Press, 2008

# Chapitre 1

## Systemes linéaires

### 1.1 Objectifs

On note  $\mathcal{M}_n(\mathbb{R})$  l'ensemble des matrices carrées d'ordre  $n$ . Soit  $A \in \mathcal{M}_n(\mathbb{R})$  une matrice inversible, et  $b \in \mathbb{R}^n$ , on a comme objectif de résoudre le système linéaire  $Ax = b$ , c'est-à-dire de trouver  $x$  solution de :

$$\begin{cases} x \in \mathbb{R}^n \\ Ax = b \end{cases} \quad (1.1)$$

Comme  $A$  est inversible, il existe un unique vecteur  $x \in \mathbb{R}^n$  solution de (1.1). Nous allons étudier dans les deux chapitres suivants des méthodes de calcul de ce vecteur  $x$  : la première partie de ce chapitre sera consacrée aux méthodes "directes" et la deuxième aux méthodes "itératives". Nous aborderons ensuite en troisième partie les méthodes de résolution de problèmes aux valeurs propres.

Un des points essentiels dans l'efficacité des méthodes envisagées concerne la taille des systèmes à résoudre. La taille de la mémoire des ordinateurs a augmenté de façon drastique de 1980 à nos jours.

Le développement des méthodes de résolution de systèmes linéaires est liée à l'évolution des machines informatiques. C'est un domaine de recherche très actif que de concevoir des méthodes qui permettent de profiter au mieux de l'architecture des machines (méthodes de décomposition en sous domaines pour profiter des architectures parallèles, par exemple).

Dans la suite de ce chapitre, nous verrons deux types de méthodes pour résoudre les systèmes linéaires : les méthodes directes et les méthodes itératives. Pour faciliter la compréhension de leur étude, nous commençons par quelques rappels d'algèbre linéaire.

### 1.2 Pourquoi et comment ?

Nous donnons dans ce paragraphe un exemple de problème dont la résolution numérique requiert la résolution d'un système linéaire, et qui nous permet d'introduire des matrices que nous allons beaucoup étudier par la suite. Nous commençons par donner ci-après après quelques rappels succincts d'algèbre linéaire, outil fondamental pour la résolution de ces systèmes linéaires.

#### 1.2.1 Quelques rappels d'algèbre linéaire

##### Quelques notions de base

Ce paragraphe rappelle des notions fondamentales que vous devriez connaître à l'issue du cours d'algèbre linéaire de première année. On va commencer par revisiter le **produit matriciel**, dont la vision combinaison linéaire de lignes est fondamentale pour bien comprendre la forme matricielle de la procédure d'élimination de Gauss.

Soient  $A$  et  $B$  deux matrices carrées d'ordre  $n$ , et  $M = AB$ . Prenons comme exemple d'illustration

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} -1 & 0 \\ 3 & 2 \end{bmatrix} \text{ et } M = \begin{bmatrix} 5 & 4 \\ 3 & 2 \end{bmatrix}$$

On note  $a_{i,j}$ ,  $b_{i,j}$  et  $m_{i,j}$ ,  $i, j = 1, \dots, n$  les coefficients respectifs de  $A$ ,  $B$  et  $M$ . Vous savez bien sûr que

$$m_{i,j} = \sum_{k=1}^n a_{i,k} b_{k,j}. \quad (1.2)$$

Si on écrit les matrices  $A$  et  $B$  sous forme de lignes (notées  $\ell_i$ ) et colonnes (notées  $\mathbf{c}_j$ ) :

$$A = \begin{bmatrix} \ell_1(A) \\ \dots \\ \ell_n(A) \end{bmatrix} \text{ et } B = [\mathbf{c}_1(B) \quad \dots \quad \mathbf{c}_n(B)]$$

Dans nos exemples, on a donc

$$\ell_1(A) = [1 \quad 2], \ell_2(A) = [0 \quad 1], \mathbf{c}_1(B) = \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \mathbf{c}_2(B) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

L'expression (1.2) s'écrit encore

$$m_{i,j} = \ell_i(A) \mathbf{c}_j(B),$$

qui est le produit d'une matrice  $1 \times n$  par une matrice  $n \times 1$ , qu'on peut aussi écrire sous forme d'un produit scalaire :

$$m_{i,j} = (\ell_i(A))^t \cdot \mathbf{c}_j(B)$$

où  $(\ell_i(A))^t$  désigne la matrice transposée, qui est donc maintenant une matrice  $n \times 1$  qu'on peut identifier à un vecteur de  $\mathbb{R}^n$ . C'est la technique "habituelle" de calcul du produit de deux matrices. On a dans notre exemple :

$$\begin{aligned} m_{1,2} &= \ell_1(A) \mathbf{c}_2(B) = [1 \quad 2] \begin{bmatrix} 0 \\ 2 \end{bmatrix}. \\ &= (\ell_1(A))^t \cdot \mathbf{c}_2(B) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} \\ &= 4. \end{aligned}$$

Mais de l'expression (1.2), on peut aussi avoir l'expression des lignes et des colonnes de  $M = AB$  en fonction des lignes de  $B$  ou des colonnes de  $A$  :

$$\ell_i(AB) = \sum_{k=1}^n a_{i,k} \ell_k(B) \quad (1.3)$$

$$\mathbf{c}_j(AB) = \sum_{k=1}^n b_{k,j} \mathbf{c}_k(A) \quad (1.4)$$

Dans notre exemple, on a donc :

$$\ell_1(AB) = [-1 \quad 0] + 2 [3 \quad 2] = [5 \quad 4]$$

ce qui montre que la ligne 1 de  $AB$  est combinaison linéaire des lignes de  $B$ . Les colonnes de  $AB$ , par contre, sont des combinaisons linéaires de colonnes de  $A$ . Par exemple :

$$\mathbf{c}_2(AB) = 0 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

Il faut donc retenir que dans un produit matriciel  $AB$ ,

les colonnes de  $AB$  sont des combinaisons linéaires des colonnes de  $A$   
 les lignes de  $AB$  sont des combinaisons linéaires des lignes de  $B$ .

Cette remarque est très importante pour la représentation matricielle de l'élimination de Gauss : lorsqu'on calcule des systèmes équivalents, on effectue des combinaisons linéaires de lignes, et donc on multiplie à gauche par une matrice d'élimination.

Le tableau ci-dessous est la traduction littérale de "Linear algebra in a nutshell", par Gilbert Strang<sup>1</sup> Pour une matrice carrée  $A$ , on donne les caractérisations du fait qu'elle est inversible ou non.

$A$ inversible	$A$ non inversible
Les vecteurs colonne sont indépendants	Les vecteurs colonne sont liés
Les vecteurs ligne sont indépendants	Les vecteurs ligne sont liés
Le déterminant est non nul	Le déterminant est nul
$Ax = 0$ a une unique solution $x = 0$	$Ax = 0$ a une infinité de solutions.
Le noyau de $A$ est réduit à $\{0\}$	Le noyau de $A$ contient au moins un vecteur non nul.
$Ax = b$ a une solution unique $x = A^{-1}b$	$Ax = b$ a soit aucune solution, soit une infinité.
$A$ a $n$ (nonzero) pivots	$A$ a $r < n$ pivots
$A$ est de rang maximal : $\text{rg}(A) = n$ .	$\text{rg}(A) = r < n$
La forme totalement échelonnée $R$ de $A$ est la matrice identité	$R$ a au moins une ligne de zéros.
L'image de $A$ est tout $\mathbb{R}^n$ .	L'image de $A$ est strictement incluse dans $\mathbb{R}^n$ .
L'espace $L(A)$ engendré par les lignes de $A$ est tout $\mathbb{R}^n$ .	$L(A)$ est de dimension $r < n$
Toutes les valeurs propres de $A$ sont non nulles	Zero est valeur propre de $A$ .
$A^t A$ is symétrique définie positive <sup>2</sup>	$A^t A$ n'est que semi- définie .

TABLE 1.1: Extrait de "Linear algebra in a nutshell", G. Strang

On rappelle pour une bonne lecture de ce tableau les quelques définitions suivantes :

**Définition 1.1** (Pivot). Soit  $A \in \mathcal{M}_n(\mathbb{R})$  une matrice carrée d'ordre  $n$ . On appelle pivot de  $A$  le premier élément non nul de chaque ligne dans la forme échelonnée de  $A$  obtenue par élimination de Gauss. Si la matrice est inversible, elle a donc  $n$  pivots (non nuls).

**Définition 1.2** (Valeurs propres). Soit  $A \in \mathcal{M}_n(\mathbb{R})$  une matrice carrée d'ordre  $n$ . On appelle valeur propre de  $A$  tout  $\lambda \in \mathbb{C}$  tel qu'il existe  $x \in \mathbb{C}^n$ ,  $x \neq 0$  tel que  $Ax = \lambda x$ . L'élément  $x$  est appelé vecteur propre de  $A$  associé à  $\lambda$ .

**Définition 1.3** (Déterminant). Il existe une unique application, notée  $\det$  de  $\mathcal{M}_n(\mathbb{R})$  dans  $\mathbb{R}$  qui vérifie les propriétés suivantes

(D1) Le déterminant de la matrice identité est égal à 1.

(D2) Si la matrice  $\tilde{A}$  est obtenue à partir de  $A$  par échange de deux lignes, alors  $\det \tilde{A} = -\det A$ .

1. Voir la page web de Strang [www.mit.edu/~gs](http://www.mit.edu/~gs) pour une foule d'informations et de cours sur l'algèbre linéaire.

(D3) Le déterminant est une fonction linéaire de chacune des lignes de la matrice  $A$ .

(D3a) (multiplication par un scalaire) si  $\tilde{A}$  est obtenue à partir de  $A$  en multipliant tous les coefficients d'une ligne par  $\lambda \in \mathbb{R}$ , alors  $\det(\tilde{A}) = \lambda \det(A)$ .

(D3b) (addition) si  $A = \begin{bmatrix} \ell_1(A) \\ \vdots \\ \ell_k(A) \\ \vdots \\ \ell_n(A) \end{bmatrix}$ ,  $\tilde{A} = \begin{bmatrix} \ell_1(A) \\ \vdots \\ \tilde{\ell}_k(A) \\ \vdots \\ \ell_n(A) \end{bmatrix}$  et  $B = \begin{bmatrix} \ell_1(A) \\ \vdots \\ \ell_k(A) + \tilde{\ell}_k(A) \\ \vdots \\ \ell_n(A) \end{bmatrix}$ , alors

$$\det(B) = \det(A) + \det(\tilde{A}).$$

On peut déduire de ces trois propriétés fondamentales un grand nombre de propriétés importantes, en particulier le fait que  $\det(AB) = \det A \det B$  et que le déterminant d'une matrice inversible est le produit des pivots : c'est de cette manière qu'on le calcule sur les ordinateurs. En particulier on n'utilise jamais la formule de Cramer, beaucoup trop coûteuse en termes de nombre d'opérations.

On rappelle que si  $A \in \mathcal{M}_n(\mathbb{R})$  une matrice carrée d'ordre  $n$ , les valeurs propres sont les racines du **polynôme caractéristique**  $P_A$  de degré  $n$ , qui s'écrit :

$$P_A(\lambda) = \det(A - \lambda I).$$

### Matrices diagonalisables

Un point important de l'algèbre linéaire, appelé "réduction des endomorphismes" dans les programmes français, consiste à se demander s'il existe une base de l'espace dans laquelle la matrice de l'application linéaire est diagonale ou tout au moins triangulaire (on dit aussi trigonale).

**Définition 1.4** (Matrice diagonalisable dans  $\mathbb{R}$ ). Soit  $A$  une matrice réelle carrée d'ordre  $n$ . On dit que  $A$  est diagonalisable dans  $\mathbb{R}$  s'il existe une base  $(\mathbf{u}_1, \dots, \mathbf{u}_n)$  de  $\mathbb{R}^n$  et des réels  $\lambda_1, \dots, \lambda_n$  (pas forcément distincts) tels que  $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$  pour  $i = 1, \dots, n$ . Les réels  $\lambda_1, \dots, \lambda_n$  sont les valeurs propres de  $A$ , et les vecteurs  $\mathbf{u}_1, \dots, \mathbf{u}_n$  sont les vecteurs propres associés.

Vous connaissez sûrement aussi la diagonalisation dans  $\mathbb{C}$  : une matrice réelle carrée d'ordre  $n$  admet toujours  $n$  valeurs propres dans  $\mathbb{C}$ , qui ne sont pas forcément identiques. Une matrice est diagonalisable dans  $\mathbb{C}$  s'il existe une base  $(\mathbf{u}_1, \dots, \mathbf{u}_n)$  de  $\mathbb{C}^n$  et des nombres complexes  $\lambda_1, \dots, \lambda_n$  (pas forcément distincts) tels que  $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$  pour  $i = 1, \dots, n$ . Ceci est vérifié si la dimension de chaque sous espace propre  $E_i = \text{Ker}(A - \lambda_i \text{Id})$  (appelée multiplicité géométrique) est égale à la multiplicité algébrique de  $\lambda_i$ , c.à.d. son ordre de multiplicité en tant que racine du polynôme caractéristique.

Par exemple la matrice  $A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$  n'est pas diagonalisable dans  $\mathbb{C}$  (ni évidemment, dans  $\mathbb{R}$ ). Le polynôme caractéristique de  $A$  est  $P_A(\lambda) = \lambda^2$ , l'unique valeur propre est donc 0, qui est de multiplicité algébrique 2, et de multiplicité géométrique 1, car le sous espace propre associé à la valeur propre nulle est  $F = \{\mathbf{x} \in \mathbb{R}^2 ; A\mathbf{x} = 0\} = \{\mathbf{x} = (0, t), t \in \mathbb{R}\}$ , qui est de dimension 1.

Ici et dans toute la suite, comme on résout des systèmes linéaires réels, on préfère travailler avec la diagonalisation dans  $\mathbb{R}$  ; cependant il y a des cas où la diagonalisation dans  $\mathbb{C}$  est utile et même nécessaire (étude de stabilité des

systèmes différentiels, par exemple). Par souci de clarté, nous préciserons toujours si la diagonalisation considérée est dans  $\mathbb{R}$  ou dans  $\mathbb{C}$ .

**Lemme 1.5.** *Soit  $A$  une matrice réelle carrée d'ordre  $n$ , diagonalisable dans  $\mathbb{R}$ . Alors*

$$A = P \operatorname{diag}(\lambda_1, \dots, \lambda_n) P^{-1},$$

où  $P$  est la matrice dont les vecteurs colonnes sont égaux aux vecteurs propres  $\mathbf{u}_1, \dots, \mathbf{u}_n$  associées aux valeurs propres  $\lambda_1, \dots, \lambda_n$

DÉMONSTRATION – Par définition d'un vecteur propre, on a  $A\mathbf{u}_i = \lambda_i\mathbf{u}_i$  pour  $i = 1, \dots, n$ , et donc, en notant  $P$  la matrice dont les colonnes sont les vecteurs propres  $\mathbf{u}_i$ ,

$$[A\mathbf{u}_1 \ \dots \ A\mathbf{u}_n] = A [\mathbf{u}_1 \ \dots \ \mathbf{u}_n] = AP$$

et donc

$$AP = [\lambda_1\mathbf{u}_1 \ \dots \ \lambda_n\mathbf{u}_n] = [\mathbf{u}_1 \ \dots \ \mathbf{u}_n] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{bmatrix} = P \operatorname{diag}(\lambda_1, \dots, \lambda_n).$$

Notons que dans ce calcul, on a fortement utilisé la multiplication des matrices par colonnes, c.à.d.

$$\mathbf{c}_i(AB) = \sum_{j=1, n} a_{i,j} \mathbf{c}_j(B).$$

Remarquons que  $P$  est aussi la matrice définie (de manière unique) par  $P\mathbf{e}_i = \mathbf{u}_i$ , où  $(\mathbf{e}_i)_{i=1, \dots, n}$  est la base canonique de  $\mathbb{R}^n$ , c'est-à-dire que  $(\mathbf{e}_i)_j = \delta_{i,j}$ . La matrice  $P$  est appelée matrice de passage de la base  $(\mathbf{e}_i)_{i=1, \dots, n}$  à la base  $(\mathbf{u}_i)_{i=1, \dots, n}$ ; (il est bien clair que la  $i$ -ème colonne de  $P$  est constituée des composantes de  $\mathbf{u}_i$  dans la base canonique  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ ).

La matrice  $P$  est inversible car les vecteurs propres forment une base, et on peut donc aussi écrire :

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \text{ ou } A = P \operatorname{diag}(\lambda_1, \dots, \lambda_n) P^{-1}.$$

■

La diagonalisation des matrices réelles symétriques est un outil qu'on utilisera souvent dans la suite, en particulier dans les exercices. Il s'agit d'un résultat extrêmement important.

**Lemme 1.6** (Une matrice symétrique est diagonalisable dans  $\mathbb{R}$ ). *Soit  $E$  un espace vectoriel sur  $\mathbb{R}$  de dimension finie :  $\dim E = n$ ,  $n \in \mathbb{N}^*$ , muni d'un produit scalaire i.e. d'une application*

$$\begin{aligned} E \times E &\rightarrow \mathbb{R}, \\ (x, y) &\rightarrow (x | y)_E, \end{aligned}$$

qui vérifie :

$$\begin{aligned} \forall x \in E, (x | x)_E &\geq 0 \text{ et } (x | x)_E = 0 \Leftrightarrow x = 0, \\ \forall (x, y) \in E^2, (x | y)_E &= (y | x)_E, \\ \forall y \in E, \text{ l'application de } E \text{ dans } \mathbb{R}, \text{ définie par } x &\rightarrow (x | y)_E \text{ est linéaire.} \end{aligned}$$

Ce produit scalaire induit une norme sur  $E$ ,  $\|x\| = \sqrt{(x | x)_E}$ .

Soit  $T$  une application linéaire de  $E$  dans  $E$ . On suppose que  $T$  est symétrique, c.à.d. que  $(T(x) | y)_E = (x | T(y))_E$ ,  $\forall (x, y) \in E^2$ . Alors il existe une base orthonormée  $(\mathbf{f}_1, \dots, \mathbf{f}_n)$  de  $E$  (c.à.d. telle que  $(\mathbf{f}_i, \mathbf{f}_j)_E = \delta_{i,j}$ ) et  $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$  tels que  $T(\mathbf{f}_i) = \lambda_i \mathbf{f}_i$  pour tout  $i \in \{1, \dots, n\}$ .

**Conséquence immédiate :** Dans le cas où  $E = \mathbb{R}^n$ , le produit scalaire canonique de  $x = (x_1, \dots, x_n)^t$  et  $y = (y_1, \dots, y_n)^t$  est défini par  $(x | y)_E = x \cdot y = \sum_{i=1}^n x_i y_i$ . Si  $A \in \mathcal{M}_n(\mathbb{R})$  est une matrice symétrique, alors l'application  $T$  définie de  $E$  dans  $E$  par  $T(x) = Ax$  est linéaire, et :  $(Tx | y) = Ax \cdot y = x \cdot A^t y = x \cdot Ay = (x | Ty)$ . Donc  $T$  est linéaire symétrique. Par le lemme précédent, il existe  $(f_1, \dots, f_n)$  et  $(\lambda_1 \dots \lambda_n) \in \mathbb{R}$  tels que  $Tf_i = Af_i = \lambda_i f_i \forall i \in \{1, \dots, n\}$  et  $f_i \cdot f_j = \delta_{i,j}, \forall (i, j) \in \{1, \dots, n\}^2$ .

**Interprétation algébrique :** Il existe une matrice de passage  $P$  de  $(e_1, \dots, e_n)$  base canonique dans  $(f_1, \dots, f_n)$  dont la première colonne de  $P$  est constituée des coordonnées de  $f_i$  dans  $(e_1 \dots e_n)$ . On a :  $Pe_i = f_i$ . On a alors  $P^{-1}APe_i = P^{-1}Af_i = P^{-1}(\lambda_i f_i) = \lambda_i e_i = \text{diag}(\lambda_1, \dots, \lambda_n)e_i$ , où  $\text{diag}(\lambda_1, \dots, \lambda_n)$  désigne la matrice diagonale de coefficients diagonaux  $\lambda_1, \dots, \lambda_n$ . On a donc :

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = D.$$

De plus  $P$  est orthogonale, i.e.  $P^{-1} = P^t$ . En effet,

$$P^t Pe_i \cdot e_j = Pe_i \cdot Pe_j = (f_i | f_j) = \delta_{i,j} \quad \forall i, j \in \{1 \dots n\},$$

et donc  $(P^t Pe_i - e_i) \cdot e_j = 0 \quad \forall j \in \{1 \dots n\} \quad \forall i \in \{1, \dots, n\}$ . On en déduit  $P^t Pe_i = e_i$  pour tout  $i = 1, \dots, n$ , i.e.  $P^t P = PP^t = Id$ .

**DÉMONSTRATION du lemme 1.6** Cette démonstration se fait par récurrence sur la dimension de  $E$ .

1ère étape.

On suppose  $\dim E = 1$ . Soit  $e \in E, e \neq 0$ , alors  $E = \mathbb{R}e = f_1$  avec  $f_1 = \frac{e}{\|e\|}$ . Soit  $T : E \rightarrow E$  linéaire symétrique, on a :  $Tf_1 \in \mathbb{R}f_1$  donc il existe  $\lambda_1 \in \mathbb{R}$  tel que  $Tf_1 = \lambda_1 f_1$ .

2ème étape.

On suppose le lemme vrai si  $\dim E < n$ . On montre alors le lemme si  $\dim E = n$ . Soit  $E$  un espace vectoriel normé sur  $\mathbb{R}$  tel que  $\dim E = n$  et  $T : E \rightarrow E$  linéaire symétrique. Soit  $\varphi$  l'application définie par :

$$\begin{aligned} \varphi : E &\rightarrow \mathbb{R} \\ x &\rightarrow (Tx | x). \end{aligned}$$

L'application  $\varphi$  est continue sur la sphère unité  $S_1 = \{x \in E | \|x\| = 1\}$  qui est compacte car  $\dim E < +\infty$ ; il existe donc  $e \in S_1$  tel que  $\varphi(x) \leq \varphi(e) = (Te | e) = \lambda$  pour tout  $x \in E$ . Soit  $y \in E \setminus \{0\}$ , et soit  $t \in ]0, \frac{1}{\|y\|}[$  alors  $e + ty \neq 0$ . On en déduit que :

$$\frac{e + ty}{\|e + ty\|} \in S_1 \text{ et donc } \varphi(e) = \lambda \geq \left( T \left( \frac{e + ty}{\|e + ty\|} \right) \middle| \frac{e + ty}{\|e + ty\|} \right)_E$$

donc  $\lambda(e + ty | e + ty)_E \geq (T(e + ty) | e + ty)$ . En développant on obtient :

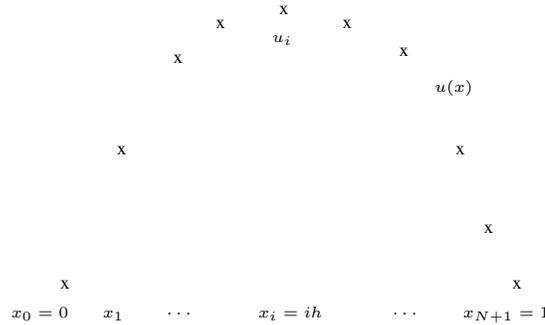
$$\lambda[2t(e | y) + t^2(y | y)_E] \geq 2t(T(e) | y) + t^2(T(y) | y)_E.$$

Comme  $t > 0$ , ceci donne :

$$\lambda[2(e | y) + t(y | y)_E] \geq 2(T(e) | y) + t(T(y) | y)_E.$$

En faisant tendre  $t$  vers  $0^+$ , on obtient  $2\lambda(e | y)_E \geq 2(T(e) | y)$ , Soit  $0 \geq (T(e) - \lambda e | y)$  pour tout  $y \in E \setminus \{0\}$ . De même pour  $z = -y$  on a  $0 \geq (T(e) - \lambda e | z)$  donc  $(T(e) - \lambda e | y) \geq 0$ . D'où  $(T(e) - \lambda e | y) = 0$  pour tout  $y \in E$ . On en déduit que  $T(e) = \lambda e$ . On pose  $f_n = e$  et  $\lambda_n = \lambda$ .

Soit  $F = \{x \in E; (x | e) = 0\}$ , on a donc  $F \neq E$ , et  $E = F \oplus \mathbb{R}e$  : on peut décomposer  $x \in E$  comme  $(x = x - (x | e)e + (x | e)e)$ . L'application  $S = T|_F$  est linéaire symétrique et on a  $\dim F = n - 1$ . et  $S(F) \subset F$ . On peut donc utiliser l'hypothèse de récurrence :  $\exists(\lambda_1 \dots \lambda_{n-1}) \in \mathbb{R}^n$  et  $\exists(f_1 \dots f_{n-1}) \in E^n$  tels que  $\forall i \in \{1 \dots n - 1\}$ ,  $Sf_i = Tf_i = \lambda_i f_i$ , et  $\forall i, j \in \{1 \dots n - 1\}$ ,  $f_i \cdot f_j = \delta_{i,j}$ . Et donc  $(\lambda_1 \dots \lambda_n)$  et  $(f_1, \dots, f_n)$  conviennent. ■

FIGURE 1.1: Solution exacte et approchée de  $-u'' = f$ 

## 1.2.2 Discrétisation de l'équation de la chaleur

Dans ce paragraphe, nous prenons un exemple très simple pour obtenir un système linéaire à partir de la discrétisation d'un problème continu.

### L'équation de la chaleur unidimensionnelle

**Discrétisation par différences finies de  $-u'' = f$**  Soit  $f \in C([0, 1], \mathbb{R})$ . On cherche  $u$  tel que

$$-u''(x) = f(x) \quad (1.5a)$$

$$u(0) = u(1) = 0. \quad (1.5b)$$

**Remarque 1.7** (Problèmes aux limites, problèmes à conditions initiales). *L'équation différentielle  $-u'' = f$  admet une infinité de solutions. Pour avoir existence et unicité, il est nécessaire d'avoir des conditions supplémentaires. Si l'on considère deux conditions en 0 (ou en 1, l'origine importe peu) on a ce qu'on appelle un problème de Cauchy, ou problème à conditions initiales. Le problème (1.5) est lui un problème aux limites : il y a une condition pour chaque bord du domaine. En dimension supérieure, le problème  $-\Delta u = f$  nécessite une condition sur au moins "un bout" de frontière pour être bien posé : voir le cours d'équations aux dérivées partielles de master pour plus de détails à ce propos.*

On peut montrer (on l'admettra ici) qu'il existe une unique solution  $u \in C^2([0, 1], \mathbb{R})$ . On cherche à calculer  $u$  de manière approchée. On va pour cela introduire la méthode de discrétisation dite *par différences finies*. Soit  $n \in \mathbb{N}^*$ , on définit  $h = 1/(n + 1)$  le *pas de discrétisation*, c.à.d. la distance entre deux points de discrétisation, et pour  $i = 0, \dots, n + 1$  on définit les points de discrétisation  $x_i = ih$  (voir Figure 1.1), qui sont les points où l'on va écrire l'équation  $-u'' = f$  en vue de se ramener à un système discret, c.à.d. à un système avec un nombre fini d'inconnues  $u_1, \dots, u_n$ . Remarquons que  $x_0 = 0$  et  $x_{n+1} = 1$ , et qu'en ces points,  $u$  est spécifiée par les conditions limites (1.5b). Soit  $u(x_i)$  la valeur exacte de  $u$  en  $x_i$ . On écrit la première équation de (1.5a) en chaque point  $x_i$ , pour  $i = 1 \dots n$ .

$$-u''(x_i) = f(x_i) = b_i \quad \forall i \in \{1 \dots n\}. \quad (1.6)$$

Supposons que  $u \in C^4([0, 1], \mathbb{R})$  (ce qui est vrai si  $f \in C^2$ ). Par développement de Taylor, on a :

$$\begin{aligned} u(x_{i+1}) &= u(x_i) + hu'(x_i) + \frac{h^2}{2}u''(x_i) + \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\xi_i), \\ u(x_{i-1}) &= u(x_i) - hu'(x_i) + \frac{h^2}{2}u''(x_i) - \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\eta_i), \end{aligned}$$

avec  $\xi_i \in (x_i, x_{i+1})$  et  $\eta_i \in (x_i, x_{i+1})$ . En sommant ces deux égalités, on en déduit que :

$$u(x_{i+1}) + u(x_{i-1}) = 2u(x_i) + h^2 u''(x_i) + \frac{h^4}{24} u^{(4)}(\xi_i) + \frac{h^4}{24} u^{(4)}(\eta_i).$$

On définit l'erreur de consistance, qui mesure la manière dont on a approché  $-u''(x_i)$  ; l'erreur de consistance  $R_i$  au point  $x_i$  est définie par

$$R_i = -u''(x_i) - \frac{u(x_{i+1}) + u(x_{i-1}) - 2u(x_i)}{h^2}. \quad (1.7)$$

On a donc :

$$\begin{aligned} |R_i| &= \left| -\frac{u(x_{i+1}) + u(x_{i-1}) - 2u(x_i)}{h^2} + u''(x_i) \right| \\ &\leq \left| \frac{h^4}{24} u^{(4)}(\xi_i) + \frac{h^4}{24} u^{(4)}(\eta_i) \right| \\ &\leq \frac{h^2}{12} \|u^{(4)}\|_\infty. \end{aligned} \quad (1.8)$$

où  $\|u^{(4)}\|_\infty = \sup_{x \in ]0,1[} |u^{(4)}(x)|$ . Cette majoration nous montre que l'erreur de consistance tend vers 0 comme  $h^2$  : on dit que le schéma est *consistant d'ordre 2*.

On introduit alors les inconnues  $(u_i)_{i=1, \dots, n}$  qu'on espère être des valeurs approchées de  $u$  aux points  $x_i$  et qui sont les composantes de la solution (si elle existe) du système suivant

$$\begin{cases} -\frac{u_{i+1} + u_{i-1} - 2u_i}{h^2} = b_i, & \forall i \in \llbracket 1; n \rrbracket, \\ u_0 = u_{n+1} = 0. \end{cases} \quad (1.9)$$

On cherche donc  $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n$  solution de (1.9). Ce système peut s'écrire sous forme matricielle :  $K_n \mathbf{u} = \mathbf{b}$

où  $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$  et  $K_n$  est la matrice carrée d'ordre  $n$  de coefficients  $(k_{i,j})_{i,j=1,n}$  définis par :

$$\begin{cases} k_{i,i} &= \frac{2}{h^2}, \forall i = 1, \dots, n, \\ k_{i,j} &= -\frac{1}{h^2}, \forall i = 1, \dots, n, j = i \pm 1, \\ k_{i,j} &= 0, \forall i = 1, \dots, n, |i - j| > 1. \end{cases} \quad (1.10)$$

On remarque immédiatement que  $K_n$  est tridiagonale.

On peut montrer que  $K_n$  est symétrique définie positive (voir exercice 10 page 19), et elle est donc inversible. Le système  $K_n \mathbf{u} = \mathbf{b}$  admet donc une unique solution. C'est bien, mais encore faut-il que cette solution soit ce qu'on espérait, c.à.d. que chaque valeur  $u_i$  soit une approximation pas trop mauvaise de  $u(x_i)$ . On appelle erreur de discrétisation en  $x_i$  la différence de ces deux valeurs :

$$e_i = u(x_i) - u_i, \quad i = 1, \dots, n. \quad (1.11)$$

Si on appelle  $\mathbf{e}$  le vecteur de composantes  $e_i$ , on déduit de la définition 1.11 de l'erreur de consistance et des équations (exactes) 1.6 que

$$K_n \mathbf{e} = \mathbf{R} \text{ et donc } \mathbf{e} = K_n^{-1} \mathbf{R}. \quad (1.12)$$

Le fait que le schéma soit consistant est une bonne chose, mais cela ne suffit pas à montrer que le schéma est convergent, c.à.d. que l'erreur entre  $\max_{i=1, \dots, n} e_i$  tend vers 0 lorsque  $h$  tend vers 0, parce que  $A$  dépend de

$h$  ! Pour cela, il faut de plus que le schéma soit *stable*, au sens où l'on puisse montrer que  $\|K_n^{-1}\|$  est borné indépendamment de  $h$ , ce qui revient à trouver une estimation sur les valeurs approchées  $u_i$  indépendante de  $h$ . La stabilité et la convergence font l'objet de l'exercice 47, où l'on montre que le schéma est convergent, et qu'on a l'estimation d'erreur suivante :

$$\max_{i=1\dots n} \{|u_i - u(x_i)|\} \leq \frac{h^2}{96} \|u^{(4)}\|_\infty.$$

Cette inégalité donne la précision de la méthode (c'est une méthode dite d'ordre 2). On remarque en particulier que si on raffine la discrétisation, c'est-à-dire si on augmente le nombre de points  $n$  ou, ce qui revient au même, si on diminue le pas de discrétisation  $h$ , on augmente la précision avec laquelle on calcule la solution approchée.

### L'équation de la chaleur bidimensionnelle

Prenons maintenant le cas d'une discrétisation du Laplacien sur un carré par différences finies. Si  $u$  est une fonction de deux variables  $x$  et  $y$  à valeurs dans  $\mathbb{R}$ , et si  $u$  admet des dérivées partielles d'ordre 2 en  $x$  et  $y$ , l'opérateur laplacien est défini par  $\Delta u = \partial_{xx}u + \partial_{yy}u$ . L'équation de la chaleur bidimensionnelle s'écrit avec cet opérateur. On cherche à résoudre le problème :

$$\begin{aligned} -\Delta u &= f \text{ sur } \Omega = ]0, 1[ \times ]0, 1[, \\ u &= 0 \text{ sur } \partial\Omega, \end{aligned} \quad (1.13)$$

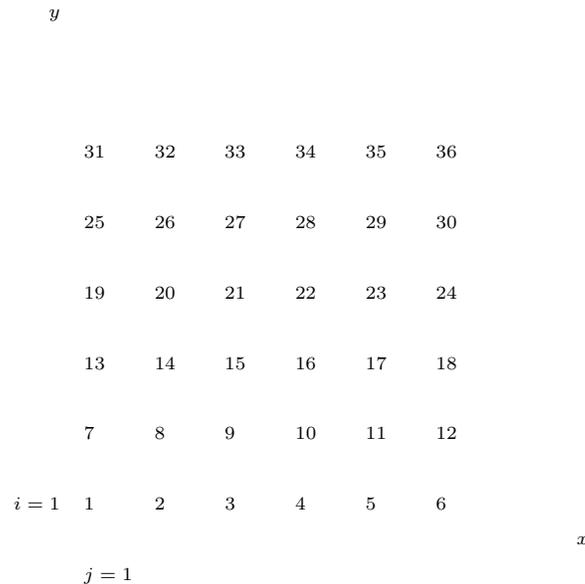
On rappelle que l'opérateur Laplacien est défini pour  $u \in C^2(\Omega)$ , où  $\Omega$  est un ouvert de  $\mathbb{R}^2$ , par

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

Définissons une discrétisation uniforme du carré par les points  $(x_i, y_j)$ , pour  $i = 1, \dots, M$  et  $j = 1, \dots, M$  avec  $x_i = ih$ ,  $y_j = jh$  et  $h = 1/(M+1)$ , représentée en figure 1.2 pour  $M = 6$ . On peut alors approcher les dérivées secondes par des quotients différentiels comme dans le cas unidimensionnel (voir page 11), pour obtenir un système linéaire :  $Au = b$  où  $A \in \mathcal{M}_n(\mathbb{R})$  et  $b \in \mathbb{R}^n$  avec  $n = M^2$ . Utilisons l'ordre "lexicographique" pour numéroter les inconnues, c.à.d. de bas en haut et de gauche à droite : les inconnues sont alors numérotées de 1 à  $n = M^2$  et le second membre s'écrit  $b = (b_1, \dots, b_n)^t$ . Les composantes  $b_1, \dots, b_n$  sont définies par : pour  $i, j = 1, \dots, M$ , on pose  $k = j + (i-1)M$  et  $b_k = f(x_i, y_j)$ .

Les coefficients de  $A = (a_{k,\ell})_{k,\ell=1,n}$  peuvent être calculés de la manière suivante :

$$\left\{ \begin{array}{l} \text{Pour } i, j = 1, \dots, M, \text{ on pose } k = j + (i-1)M, \\ a_{k,k} = \frac{4}{h^2}, \\ a_{k,k+1} = \begin{cases} -\frac{1}{h^2} & \text{si } j \neq M, \\ 0 & \text{sinon,} \end{cases} \\ a_{k,k-1} = \begin{cases} -\frac{1}{h^2} & \text{si } j \neq 1, \\ 0 & \text{sinon,} \end{cases} \\ a_{k,k+M} = \begin{cases} -\frac{1}{h^2} & \text{si } i < M, \\ 0 & \text{sinon,} \end{cases} \\ a_{k,k-M} = \begin{cases} -\frac{1}{h^2} & \text{si } i > 1, \\ 0 & \text{sinon,} \end{cases} \\ \text{Pour } k = 1, \dots, n, \text{ et } \ell = 1, \dots, n; \\ a_{k,\ell} = 0, \forall k = 1, \dots, n, 1 < |k - \ell| < n \text{ ou } |k - \ell| > n. \end{array} \right.$$

FIGURE 1.2: Ordre lexicographique des inconnues, exemple dans le cas  $M = 6$ 

La matrice est donc tridiagonale par blocs, plus précisément si on note

$$D = \begin{bmatrix} 4 & -1 & 0 & \dots & \dots & 0 \\ -1 & 4 & -1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & & \\ 0 & & \ddots & \ddots & \ddots & -1 \\ 0 & \dots & & 0 & -1 & 4 \end{bmatrix},$$

les blocs diagonaux (qui sont des matrices de dimension  $M \times M$ ), on a :

$$A = \begin{bmatrix} D & -\text{Id} & 0 & \dots & \dots & 0 \\ -\text{Id} & D & -\text{Id} & 0 & \dots & 0 \\ 0 & -\text{Id} & D & -\text{Id} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & -\text{Id} & D & -\text{Id} \\ 0 & \dots & & 0 & -\text{Id} & D \end{bmatrix}, \quad (1.14)$$

où  $\text{Id}$  désigne la matrice identité d'ordre  $M$ , et  $0$  la matrice nulle d'ordre  $M$ .

**Matrices monotones, ou à inverse positive** Une propriété qui revient souvent dans l'étude des matrices issues de la discrétisation d'équations différentielles est le fait que si leur action sur un vecteur  $u$  donne un vecteur positif  $v$  (composante par composante) alors le vecteur  $u$  de départ doit être positif (composante par composante) ; on dit souvent que la matrice est "monotone", ce qui n'est pas un terme très évocateur... Dans ce cours, on lui préférera le terme "à inverse positive" ; en effet, on montre à la proposition qu'une matrice  $A$  est monotone si et seulement si elle inversible et à inverse positive.

**Définition 1.8** (IP-matrice ou matrice monotone). Si  $\mathbf{x} \in \mathbb{R}^n$ , on dit que  $\mathbf{x} \geq 0$  [resp.  $\mathbf{x} > 0$ ] si toutes les composantes de  $\mathbf{x}$  sont positives [resp. strictement positives].

Soit  $A \in \mathcal{M}_n(\mathbb{R})$ , on dit que  $A$  est une matrice monotone si elle vérifie la propriété suivante :

$$\text{Si } \mathbf{x} \in \mathbb{R}^n \text{ est tel que } A\mathbf{x} \geq 0, \text{ alors } \mathbf{x} \geq 0,$$

ce qui peut encore s'écrire :  $\{\mathbf{x} \in \mathbb{R}^n \text{ t.q. } A\mathbf{x} \geq 0\} \subset \{\mathbf{x} \in \mathbb{R}^n \text{ t.q. } \mathbf{x} \geq 0\}$ .

**Proposition 1.9** (Caractérisation des matrices monotones). Une matrice  $A$  est monotone si et seulement si elle est inversible et à inverse positive (c.à.d. dont tous les coefficients sont positifs).

La démonstration de ce résultat est l'objet de l'exercice 8. Retenez que toute matrice monotone est inversible et d'inverse positive que cette propriété de monotonie est utilisée pour établir une borne de  $\|A^{-1}\|$  pour la matrice de discrétisation du Laplacien, dont on a besoin pour montrer la convergence du schéma. C'est donc une propriété qui est importante au niveau de l'analyse numérique.

### 1.2.3 Exercices

**Exercice 1** (Théorème du rang). Soit  $A \in \mathcal{M}_{n,p}(\mathbb{R})$  ( $n, p \geq 1$ ). On rappelle que  $\text{Ker}(A) = \{x \in \mathbb{R}^p; Ax = 0\}$ ,  $\text{Im}(A) = \{Ax, x \in \mathbb{R}^p\}$  et  $\text{rang}(A) = \dim(\text{Im}(A))$ . Noter que  $\text{Ker}(A) \subset \mathbb{R}^p$  et  $\text{Im}(A) \subset \mathbb{R}^n$ .

Soit  $f_1, \dots, f_r$  une base de  $\text{Im}(A)$  (donc  $r \leq n$ ) et, pour  $i \in \{1, \dots, r\}$ ,  $a_i$  tel que  $Aa_i = f_i$ .

1. Montrer que la famille  $a_1, \dots, a_r$  est une famille libre de  $\mathbb{R}^p$  (et donc  $r \leq p$ ).
2. On note  $G$  le sous espace vectoriel de  $\mathbb{R}^p$  engendré par  $a_1, \dots, a_r$ . Montrer que  $\mathbb{R}^p = G \oplus \text{Ker}(A)$ . En déduire que (théorème du rang)

$$p = \dim(\text{Ker}(A)) + \dim(\text{Im}(A)).$$

3. On suppose ici que  $n = p$ . Montrer que l'application  $x \mapsto Ax$  (de  $\mathbb{R}^n$  dans  $\mathbb{R}^n$ ) est injective si et seulement si elle est surjective.

**Exercice 2** ( $\text{rang}(A) = \text{rang}(A^t)$ ). Soit  $A \in \mathcal{M}_{n,p}(\mathbb{R})$  ( $n, p \geq 1$ ).

1. Soient  $P$  une matrice inversible de  $\mathcal{M}_n(\mathbb{R})$  et  $Q$  une matrice inversible de  $\mathcal{M}_p(\mathbb{R})$ .

Montrer que  $\dim(\text{Im}(PA)) = \dim(\text{Im}(AQ)) = \dim(\text{Im}(A))$ .

Montrer aussi que les matrices  $P^t$  et  $Q^t$  sont inversibles.

Soit  $f_1, \dots, f_r$  une base de  $\text{Im}(A)$  (donc  $r \leq p$ ) et, pour  $i \in \{1, \dots, r\}$ ,  $a_i$  tel que  $Aa_i = f_i$ . Soit  $a_{r+1}, \dots, a_p$  une base de  $\text{Ker}(A)$  (si  $\text{Ker}(A) \neq \{0\}$ ). La famille  $a_1, \dots, a_n$  est une base de  $\mathbb{R}^p$  (voir l'exercice 1). De même, on complète (si  $r < n$ )  $f_1, \dots, f_r$  par  $f_{r+1}, \dots, f_n$  de manière à avoir une base  $f_1, \dots, f_n$  de  $\mathbb{R}^n$ .

Enfin, on note  $P$  et  $Q$  les matrices telles que  $Pe_i = a_i$  (pour tout  $i = 1, \dots, p$ ) et  $Qf_j = \bar{e}_j$  (pour tout  $j = 1, \dots, n$ ) ou  $e_1, \dots, e_p$  est la base canonique de  $\mathbb{R}^p$  et  $\bar{e}_1, \dots, \bar{e}_n$  est la base canonique de  $\mathbb{R}^n$ . On pose  $J = QAP$ .

2. Montrer que  $P$  et  $Q$  sont inversibles.
3. calculer les colonnes de  $J$  et de  $J^t$  et en déduire que les matrices  $J$  et  $J^t$  sont de même rang.
4. Montrer que  $A$  et  $A^t$  sont de même rang.

**Exercice 3** (Vrai ou faux ? Motiver les réponses...). *Corrigé en page 21*

On suppose dans toutes les questions suivantes que  $n \geq 2$ .

1. Soit  $Z \in \mathbb{R}^n$  un vecteur non nul. La matrice  $ZZ^t$  est inversible.
2. La matrice inverse d'une matrice triangulaire inférieure est triangulaire supérieure.

3. Les valeurs propres sont les racines du polynôme caractéristique.
4. Toute matrice inversible est diagonalisable dans  $\mathbb{R}$ .
5. Toute matrice inversible est diagonalisable dans  $\mathbb{C}$ .
6. Le déterminant d'une matrice  $A$  est égal au produit de ses valeurs propres.
7. Soit  $A$  une matrice carrée telle que  $A\mathbf{x} = \mathbf{0} \implies \mathbf{x} = \mathbf{0}$ , alors  $A$  est inversible.
8. Soit  $A$  une matrice carrée telle que  $A\mathbf{x} \geq \mathbf{0} \implies \mathbf{x} \geq \mathbf{0}$ , alors  $A$  est inversible.
9. Une matrice symétrique est inversible.
10. Une matrice symétrique définie positive est inversible. Le système linéaire

$$\sum_{j=1}^{n+1} a_{i,j}x_j = 0 \text{ pour tout } i = 1, \dots, n$$

admet toujours une solution non nulle.

**Exercice 4** (Sur quelques notions connues). *Corrigé en page 21*

1. Soit  $A$  une matrice carrée d'ordre  $n$  et  $\mathbf{b} \in \mathbb{R}^3$ . Peut-il exister exactement deux solutions distinctes au système  $A\mathbf{x} = \mathbf{b}$  ?
2. Soient  $A, B$  et  $C$  de dimensions telles que  $AB$  et  $BC$  existent. Montrer que si  $AB = Id$  et  $BC = Id$ , alors  $A = C$ .
3. Combien y a-t-il de matrices carrées d'ordre 2 ne comportant que des 1 ou des 0 comme coefficients ? Combien d'entre elles sont inversibles ?
4. Soit  $B = \begin{bmatrix} 3 & 2 \\ -5 & -3 \end{bmatrix}$ . Montrer que  $B^{1024} = Id$ .

**Exercice 5** (La matrice  $K_3$ ). *Suggestions en page 21. Corrigé en page 22*

Soit  $f \in C([0, 1], \mathbb{R})$ . On cherche  $u$  tel que

$$-u''(x) = f(x), \quad \forall x \in (0, 1), \tag{1.15a}$$

$$u(0) = u(1) = 0. \tag{1.15b}$$

1. Calculer la solution exacte  $u(x)$  du problème lorsque  $f$  est la fonction identiquement égale à 1 (on admettra que cette solution est unique), et vérifier que  $u(x) \geq 0$  pour tout  $x \in [0, 1]$ .

On discrétise le problème suivant par différences finies, avec un pas  $h = \frac{1}{4}$  avec la technique vue en cours.

2. A l'aide de développements de Taylor, écrire l'approximation de  $u''(x_i)$  au deuxième ordre en fonction de  $u(x_i)$ ,  $u(x_{i-1})$  et  $u(x_{i+1})$ . En déduire le schéma aux différences finies pour l'approximation de (1.15), qu'on écrira sous la forme :

$$K_3 \mathbf{u} = \mathbf{b}, \tag{1.16}$$

où  $K_3$  est la matrice de discrétisation qu'on explicitera,  $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$  et  $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix}$ .

3. Résoudre le système linéaire (1.16) par la méthode de Gauss. Comparer  $u_i$  et  $u(x_i)$  pour  $i = 1, 2, 3$ , et expliquer pourquoi l'erreur de discrétisation  $u(x_i) - u_i$  est nulle.
4. Reprendre les questions précédentes en remplaçant les conditions limites (1.15b) par :

$$u(0) = 0, \quad u'(1) = 0. \tag{1.17}$$

5. Soit  $c \in \mathbb{R}$ . On considère maintenant le problème suivant :

$$-u''(x) = c, \quad \forall x \in (0, 1), \quad (1.18a)$$

$$u'(0) = u'(1) = 0, \quad (1.18b)$$

- (a) Montrer que le problème (1.18) admet soit une infinité de solutions, soit pas de solution.  
 (b) Ecrire la discrétisation du problème (1.18), toujours avec  $h = \frac{1}{4}$ , sous la forme  $\tilde{K}_3 \mathbf{u} = \tilde{\mathbf{b}}$  en explicitant  $\tilde{K}_3$  et  $\tilde{\mathbf{b}}$ .  
 (c) Montrer que la matrice  $\tilde{K}_3$  n'est pas inversible : on part d'un problème continu mal posé, et on obtient par discrétisation un problème discret mal posé...

**Exercice 6** (Matrices symétriques définies positives). *Suggestions en page 21, corrigé en page 23. On rappelle que toute matrice  $A \in \mathcal{M}_n(\mathbb{R})$  symétrique est diagonalisable dans  $\mathbb{R}$  (cf. lemme 1.6 page 9). Plus précisément, on a montré en cours que, si  $A \in \mathcal{M}_n(\mathbb{R})$  est une matrice symétrique, il existe une base de  $\mathbb{R}^n$ , notée  $\{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ , et il existe  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$  t.q.  $A\mathbf{f}_i = \lambda_i \mathbf{f}_i$ , pour tout  $i \in \{1, \dots, n\}$ , et  $\mathbf{f}_i \cdot \mathbf{f}_j = \delta_{i,j}$  pour tout  $i, j \in \{1, \dots, n\}$  ( $x \cdot y$  désigne le produit scalaire de  $x$  avec  $y$  dans  $\mathbb{R}^n$ ).*

1. Soit  $A \in \mathcal{M}_n(\mathbb{R})$ . On suppose que  $A$  est symétrique définie positive, montrer que les éléments diagonaux de  $A$  sont strictement positifs.
2. Soit  $A \in \mathcal{M}_n(\mathbb{R})$  une matrice symétrique. Montrer que  $A$  est symétrique définie positive si et seulement si toutes les valeurs propres de  $A$  sont strictement positives.
3. Soit  $A \in \mathcal{M}_n(\mathbb{R})$ . On suppose que  $A$  est symétrique définie positive. Montrer qu'on peut définir une unique matrice  $B \in \mathcal{M}_n(\mathbb{R})$ , symétrique définie positive t.q.  $B^2 = A$  (on note  $B = A^{\frac{1}{2}}$ ).

**Exercice 7** (Diagonalisation dans  $\mathbb{R}$ ). *Corrigé en page 24.*

Soit  $E$  un espace vectoriel réel de dimension  $n \in \mathbb{N}$  muni d'un produit scalaire, noté  $(\cdot, \cdot)$ . Soient  $T$  et  $S$  deux applications linéaires symétriques de  $E$  dans  $E$  ( $T$  symétrique signifie  $(Tx, y) = (x, Ty)$  pour tous  $x, y \in E$ ). On suppose que  $T$  est définie positive (c'est-à-dire  $(Tx, x) > 0$  pour tout  $x \in E \setminus \{0\}$ ).

1. Montrer que  $T$  est inversible. Pour  $x, y \in E$ , on pose  $(x, y)_T = (Tx, y)$ . Montrer que l'application  $(x, y) \mapsto (x, y)_T$  définit un nouveau produit scalaire sur  $E$ .
2. Montrer que  $T^{-1}S$  est symétrique pour le produit scalaire défini à la question précédente. En déduire, avec le lemme 1.6 page 9, qu'il existe une base de  $E$ , notée  $\{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  et une famille  $\{\lambda_1, \dots, \lambda_n\} \subset \mathbb{R}$  telles que  $T^{-1}S\mathbf{f}_i = \lambda_i \mathbf{f}_i$  pour tout  $i \in \{1, \dots, n\}$  et t.q.  $(T\mathbf{f}_i, \mathbf{f}_j) = \delta_{i,j}$  pour tout  $i, j \in \{1, \dots, n\}$ .

**Exercice 8** (IP-matrice). *Corrigé en page 25*

Soit  $n \in \mathbb{N}^*$ , on note  $\mathcal{M}_n(\mathbb{R})$  l'ensemble des matrices de  $n$  lignes et  $n$  colonnes et à coefficients réels. Si  $x \in \mathbb{R}^n$ , on dit que  $x \geq 0$  [resp.  $x > 0$ ] si toutes les composantes de  $x$  sont positives [resp. strictement positives].

Soit  $A \in \mathcal{M}_n(\mathbb{R})$ , on dit que  $A$  est une IP-matrice si elle vérifie la propriété suivante :

$$\text{Si } x \in \mathbb{R}^n \text{ est tel que } Ax \geq 0, \text{ alors } x \geq 0,$$

ce qui peut encore s'écrire :  $\{x \in \mathbb{R}^n \text{ t.q. } Ax \geq 0\} \subset \{x \in \mathbb{R}^n \text{ t.q. } x \geq 0\}$ .

1. Soit  $A = (a_{i,j})_{i,j=1,\dots,n} \in \mathcal{M}_n(\mathbb{R})$ . Montrer que  $A$  est une IP-matrice si et seulement si  $A$  est inversible et  $A^{-1} \geq 0$  (c'est-à-dire que tous les coefficients de  $A^{-1}$  sont positifs).

2. Soit  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  une matrice réelle d'ordre 2. Montrer que  $A$  est une IP-matrice si et seulement si :

$$\begin{cases} ad < bc, \\ a < 0, d < 0 \\ b \geq 0, c \geq 0 \end{cases} \quad \text{ou} \quad \begin{cases} ad > bc, \\ a > 0, d > 0, \\ b \leq 0, c \leq 0. \end{cases} \quad (1.19)$$

3. Montrer que si  $A \in \mathcal{M}_n(\mathbb{R})$  est une IP-matrice alors  $A^t$  (la transposée de  $A$ ) est une IP-matrice.

4. Montrer que si  $A$  est telle que

$$a_{i,j} \leq 0, \text{ pour tout } i, j = 1, \dots, n, i \neq j, \text{ et } a_{i,i} > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{i,j}|, \text{ pour tout } i = 1, \dots, n, \quad (1.20)$$

alors  $A$  est une IP-matrice.

5. En déduire que si  $A$  est telle que

$$a_{i,j} \leq 0, \text{ pour tout } i, j = 1, \dots, n, i \neq j, \text{ et } a_{i,i} > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{j,i}|, \text{ pour tout } i = 1, \dots, n, \quad (1.21)$$

alors  $A$  est une IP-matrice.

6. Montrer que si  $A \in \mathcal{M}_n(\mathbb{R})$  est une IP-matrice et si  $x \in \mathbb{R}^n$  alors :

$$Ax > 0 \Rightarrow x > 0.$$

c'est-à-dire que  $\{x \in \mathbb{R}^n \text{ t.q. } Ax > 0\} \subset \{x \in \mathbb{R}^n \text{ t.q. } x > 0\}$ .

7. Montrer, en donnant un exemple, qu'une matrice  $A$  de  $\mathcal{M}_n(\mathbb{R})$  peut vérifier  $\{x \in \mathbb{R}^n \text{ t.q. } Ax > 0\} \subset \{x \in \mathbb{R}^n \text{ t.q. } x > 0\}$  et ne pas être une IP-matrice.

8. On suppose dans cette question que  $A \in \mathcal{M}_n(\mathbb{R})$  est inversible et que  $\{x \in \mathbb{R}^n \text{ t.q. } Ax > 0\} \subset \{x \in \mathbb{R}^n \text{ t.q. } x > 0\}$ . Montrer que  $A$  est une IP-matrice.

9. (Question plus difficile) Soit  $E$  l'espace des fonctions continues sur  $\mathbb{R}$  et admettant la même limite finie en  $+\infty$  et  $-\infty$ . Soit  $\mathcal{L}(E)$  l'ensemble des applications linéaires continues de  $E$  dans  $E$ . Pour  $f \in E$ , on dit que  $f > 0$  (resp.  $f \geq 0$ ) si  $f(x) > 0$  (resp.  $f(x) \geq 0$ ) pour tout  $x \in \mathbb{R}$ . Montrer qu'il existe  $T \in \mathcal{L}(E)$  tel que  $Tf \geq 0 \implies f \geq 0$ , et  $g \in E$  tel que  $Tg > 0$  et  $g \not\geq 0$  (ceci démontre que le raisonnement utilisé en 2 (b) ne marche pas en dimension infinie).

**Exercice 9 (M-matrice).** *Corrigé en page 26* Dans ce qui suit, toutes les inégalités écrites sur des vecteurs ou des matrices sont à entendre au sens composante par composante.

Soit  $A = (a_{i,j})_{i,j=1,\dots,n}$  une matrice carrée d'ordre  $n$ . On dit que  $A$  est une  $M$ -matrice si  $A$  est une IP-matrice ( $A$  est inversible et  $A^{-1} \geq 0$ , voir exercice 8) qui vérifie de plus

(a)  $a_{i,i} > 0$  pour  $i = 1, \dots, n$ ;

(b)  $a_{i,j} \leq 0$  pour  $i, j = 1, \dots, n, i \neq j$ ;

1. Soit  $A$  une IP-matrice ; montrer que  $A$  est une  $M$ -matrice si et seulement si la propriété (b) est vérifiée.

2. Soit  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  une matrice réelle d'ordre 2. Montrer que  $A$  est une  $M$ -matrice si et seulement si :

$$\begin{cases} ad > bc, \\ a > 0, d > 0, \\ b \leq 0, c \leq 0. \end{cases} \quad (1.22)$$

3. Les matrices  $A = \begin{pmatrix} -1 & 2 \\ 2 & -1 \end{pmatrix}$  et  $B = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$  sont-elles des IP-matrices ? des  $M$ -matrices ?

4. Soit  $A$  la matrice carrée d'ordre 3 définie par :

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

Montrer que  $A^{-1} \geq 0$  mais que  $A$  n'est pas une  $M$ -matrice.

5. Soit  $A$  une matrice carrée d'ordre  $n = m + p$ , avec  $m, p \in \mathbb{N}$  tels que  $m \geq 1$  et  $p \geq 1$ , vérifiant :

$$\left. \begin{array}{l} a_{i,i} \geq 0, \\ a_{i,j} \leq 0, \text{ pour } j = 1, \dots, n, j \neq i, \\ a_{i,i} + \sum_{\substack{j=1 \\ j \neq i}}^n a_{i,j} = 0 \end{array} \right\} \text{ pour } i = 1, \dots, m, \quad (1.23)$$

$$\left. \begin{array}{l} a_{i,i} = 1, \\ a_{i,j} = 0, \text{ pour } j = 1, \dots, n, j \neq i, \end{array} \right\} \text{ pour } i = m + 1, \dots, n. \quad (1.24)$$

$$\forall i \leq m, \exists (k_\ell)_{\ell=1, \dots, L_i}; k_1 = i, k_{L_i} > m, \text{ et } a_{k_\ell, k_{\ell+1}} < 0, \forall \ell = 1, \dots, L_i. \quad (1.25)$$

Soit  $b \in \mathbb{R}^n$  tel que  $b_i = 0$  pour  $i = 1, \dots, m$ . On considère le système linéaire

$$Au = b \quad (1.26)$$

5.1 Montrer que le système (1.26) admet une et une seule solution.

5.2 Montrer que  $u$  est tel que  $\min_{k=m+1, n} b_k \leq u_i \leq \max_{k=m+1, n} b_k$ . (On pourra pour simplifier supposer que les équations sont numérotées de telle sorte que  $\min_{k=m+1, n} b_k = b_{m+2} \leq b_2 \leq \dots \leq b_n = \max_{k=m+1, n} b_k$ .)

6. On considère le problème de Dirichlet suivant :

$$-u'' = 0 \text{ sur } [0, 1] \quad (1.27a)$$

$$u(0) = -1 \quad (1.27b)$$

$$u(1) = 1. \quad (1.27c)$$

6.1 Calculer la solution exacte  $u$  de ce problème et vérifier qu'elle reste comprise entre -1 et 1.

6.2 Soit  $m > 1$  et soient  $A$  et  $b$  et la matrice et le second membre du système linéaire d'ordre  $n = m + 2$  obtenu par discrétisation par différences finies avec un pas uniforme  $h = \frac{1}{m}$  du problème (1.27) (en écrivant les conditions aux limites dans le système). Montrer que la solution  $\mathbf{u} = (u_1, \dots, u_n)^t \in \mathbb{R}^n$  du système  $A\mathbf{u} = \mathbf{b}$  vérifie  $-1 \leq u_i \leq 1$ .

**Exercice 10** (Matrice du Laplacien discret 1D). *Corrigé détaillé en page 23.*

Soit  $f \in C([0, 1])$ . Soit  $n \in \mathbb{N}^*$ ,  $n$  impair. On pose  $h = 1/(n + 1)$ . Soit  $K_n$  la matrice définie par (1.10) page 12, issue d'une discrétisation par différences finies avec pas constant du problème (1.5a) page 11.

Montrer que  $K_n$  est symétrique définie positive.

**Exercice 11** (Pas non constant). *Reprendre la discrétisation vue en cours avec un pas  $h_i = x_{i+1} - x_i$  non constant, et montrer que dans ce cas, le schéma est consistant d'ordre 1 seulement.*

**Exercice 12** (Réaction diffusion 1d.). *Corrigé détaillé en page 24. On s'intéresse à la discrétisation par Différences Finies du problème aux limites suivant :*

$$\begin{aligned} -u''(x) + u(x) &= f(x), \quad x \in ]0, 1[, \\ u(0) &= u(1) = 0. \end{aligned} \quad (1.28)$$

Soit  $n \in \mathbb{N}^*$ . On note  $U = (u_j)_{j=1, \dots, n}$  une "valeur approchée" de la solution  $u$  du problème (1.28) aux points  $\left(\frac{j}{n+1}\right)_{j=1, \dots, n}$ . Donner la discrétisation par différences finies de ce problème sous la forme  $AU = b$ .

**Exercice 13** (Discrétisation). *Corrigé en page 27*

On considère la discrétisation à pas constant par le schéma aux différences finies symétrique à trois points du problème (1.5a) page 11, avec  $f \in C([0, 1])$ . Soit  $n \in \mathbb{N}^*$ ,  $n$  impair. On pose  $h = 1/(n + 1)$ . On note  $u$  est la solution exacte,  $x_i = ih$ , pour  $i = 1, \dots, n$  les points de discrétisation, et  $(u_i)_{i=1, \dots, n}$  la solution du système discrétisé (1.9).

1. Montrer que si  $u \in C^4([0, 1])$ , alors la propriété (1.7) est vérifiée, c.à.d. :

$$-\frac{u(x_{i+1}) + u(x_{i-1}) - 2u(x_i)}{h^2} = -u''(x_i) + R_i \text{ avec } |R_i| \leq \frac{h^2}{12} \|u^{(4)}\|_\infty.$$

2. Montrer que si  $f$  est constante, alors

$$\max_{1 \leq i \leq n} |u_i - u(x_i)| = 0.$$

3. Soit  $n$  fixé, et  $\max_{1 \leq i \leq n} |u_i - u(x_i)| = 0$ . A-t-on forcément que  $f$  est constante sur  $[0, 1]$  ?

## 1.3 Les méthodes directes

### 1.3.1 Définition

**Définition 1.10** (Méthode directe). On appelle méthode directe de résolution de (1.1) une méthode qui donne exactement  $x$  ( $A$  et  $b$  étant connus) solution de (1.1) après un nombre fini d'opérations élémentaires ( $+$ ,  $-$ ,  $\times$ ,  $/$ ).

Parmi les méthodes de résolution du système (1.1), la plus connue est la *méthode de Gauss* (avec pivot), encore appelée *méthode d'échelonnement* ou *méthode LU* dans sa forme matricielle.

Nous rappelons la méthode de Gauss et sa réécriture matricielle qui donne la méthode *LU* et nous étudierons plus en détails la méthode de Choleski, qui est adaptée aux matrices symétriques.

### 1.3.2 Méthode de Gauss, méthode LU

Soit  $A \in \mathcal{M}_n(\mathbb{R})$  une matrice inversible, et  $b \in \mathbb{R}^n$ . On cherche à calculer  $x \in \mathbb{R}^n$  tel que  $Ax = b$ . Le principe de la méthode de Gauss est de se ramener, par des opérations simples (combinaisons linéaires), à un système triangulaire équivalent, qui sera donc facile à inverser.

Commençons par un exemple pour une matrice  $3 \times 3$ . Nous donnerons ensuite la méthode pour une matrice  $n \times n$ .

#### Un exemple $3 \times 3$

On considère le système  $Ax = b$ , avec

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & -1 \\ -1 & 1 & -2 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}.$$

On écrit la **matrice augmentée**, constituée de la matrice  $A$  et du second membre  $b$ .

$$\tilde{A} = [A \quad b] = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 2 & -1 & 1 \\ -1 & 1 & -2 & -2 \end{bmatrix}.$$

**Gauss et opérations matricielles** Allons y pour Gauss :

La première ligne a un 1 en première position (en gras dans la matrice), ce coefficient est non nul, et c'est un **pivot**. On va pouvoir diviser toute la première ligne par ce nombre pour en soustraire un multiple à toutes les lignes d'après, dans le but de faire apparaître des 0 dans tout le bas de la colonne.

La deuxième équation a déjà un 0 dessous, donc on n'a rien besoin de faire. On veut ensuite annuler le premier coefficient de la troisième ligne. On retranche donc  $(-1)$  fois la première ligne à la troisième<sup>3</sup> :

$$\begin{bmatrix} \mathbf{1} & 0 & 1 & 2 \\ 0 & 2 & -1 & 1 \\ -1 & 1 & -2 & -2 \end{bmatrix} \xrightarrow{\ell_3 \leftarrow -\ell_3 + \ell_1} \begin{bmatrix} \mathbf{1} & 0 & 1 & 2 \\ 0 & \mathbf{2} & -1 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

Ceci revient à multiplier  $\tilde{A}$  à gauche par la matrice  $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ .

La deuxième ligne a un terme non nul en deuxième position (2) : c'est un pivot. On va maintenant annuler le deuxième terme de la troisième ligne ; pour cela, on retranche  $1/2$  fois la ligne 2 à la ligne 3 :

3. Bien sûr, ceci revient à ajouter la première ligne ! Il est cependant préférable de parler systématiquement de "retrancher" quitte à utiliser un coefficient négatif, car c'est ce qu'on fait conceptuellement : pour l'élimination on enlève un multiple de la ligne du pivot à la ligne courante.

$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 2 & -1 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \xrightarrow{\ell_3 \leftarrow \ell_3 - 1/2 \ell_2} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 2 & -1 & 1 \\ 0 & 0 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

Ceci revient à multiplier la matrice précédente à gauche par la matrice  $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{2} & 1 \end{bmatrix}$ . On a ici obtenu une

matrice sous forme triangulaire supérieure à trois pivots : on peut donc faire la remontée pour obtenir la solution du système, et on obtient (en notant  $x_i$  les composantes de  $\mathbf{x}$ ) :  $x_3 = 1$  puis  $x_2 = 1$  et enfin  $x_1 = 1$ .

On a ainsi résolu le système linéaire.

On rappelle que le fait de travailler sur la matrice augmentée est extrêmement pratique car il permet de travailler simultanément sur les coefficients du système linéaire et sur le second membre.

Finalement, au moyen des opérations décrites ci-dessus, on a transformé le système linéaire

$$A\mathbf{x} = \mathbf{b} \text{ en } U\mathbf{x} = E_2 E_1 \mathbf{b}, \text{ où } U = E_2 E_1 A$$

est une matrice triangulaire supérieure.

**Factorisation LU** Tout va donc très bien pour ce système, mais supposons maintenant qu'on ait à résoudre 3089 systèmes, avec la même matrice  $A$  mais 3089 seconds membres  $\mathbf{b}$  différents<sup>4</sup>. Il serait un peu dommage de recommencer les opérations ci-dessus 3089 fois, alors qu'on peut en éviter une bonne partie. Comment faire ? L'idée est de "factoriser" la matrice  $A$ , c.à.d de l'écrire comme un produit  $A = LU$ , où  $L$  est triangulaire inférieure (lower triangular) et  $U$  triangulaire supérieure (upper triangular). On reformule alors le système  $A\mathbf{x} = \mathbf{b}$  sous la forme  $LU\mathbf{x} = \mathbf{b}$  et on résout maintenant deux systèmes faciles à résoudre car triangulaires :  $L\mathbf{y} = \mathbf{b}$  et  $U\mathbf{x} = \mathbf{y}$ . La factorisation  $LU$  de la matrice découle immédiatement de l'algorithme de Gauss. Voyons comment sur l'exemple précédent.

1/ On remarque que  $U = E_2 E_1 A$  peut aussi s'écrire  $A = LU$ , avec  $L = (E_2 E_1)^{-1}$ .

2/ On sait que  $(E_2 E_1)^{-1} = (E_1)^{-1} (E_2)^{-1}$ .

3/ Les matrices inverses  $E_1^{-1}$  et  $E_2^{-1}$  sont faciles à déterminer : comme  $E_2$  consiste à retrancher 1/2 fois la ligne 2 à la ligne 3, l'opération inverse consiste à ajouter 1/2 fois la ligne 2 à la ligne 3, et donc

$$E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & 1 \end{bmatrix}.$$

Il est facile de voir que  $E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$  et donc  $L = E_1^{-1} E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & \frac{1}{2} & 1 \end{bmatrix}$ .

La matrice  $L$  est une matrice triangulaire inférieure (et c'est d'ailleurs pour cela qu'on l'appelle  $L$ , pour "lower" in English...) dont les coefficients sont particulièrement simples à trouver : les termes diagonaux sont tous égaux à un, et **chaque terme non nul sous-diagonal  $\ell_{i,j}$  est égal au coefficient par lequel on a multiplié la ligne pivot  $i$  avant de la retrancher à la ligne  $j$ .**

4/ On a bien donc  $A = LU$  avec  $L$  triangulaire inférieure (lower triangular) et  $U$  triangulaire supérieure (upper triangular).

La procédure qu'on vient d'expliquer s'appelle **méthode LU** pour la résolution des systèmes linéaires, et elle est d'une importance considérable dans les sciences de l'ingénieur, puisqu'elle est utilisée dans les programmes informatiques pour la résolution des systèmes linéaires.

Dans l'exemple que nous avons étudié, tout se passait très bien car nous n'avons pas eu de zéro en position pivotale. Si on a un zéro en position pivotale, la factorisation peut quand même se faire, mais au prix d'une permutation.

<sup>4</sup>. Ceci est courant dans les applications. Par exemple on peut vouloir calculer la réponse d'une structure de génie civil à 3089 chargements différents.

$$\begin{array}{cccc}
 & & a_{1,1}^{(1)} & & a_{1,N}^{(1)} \\
 & & 0 & & \\
 & & & & \\
 A^{(i+1)} = & & 0 & a_{i+1,i+1}^{(i+1)} & \\
 & & & a_{i+2,i+1}^{(i+1)} & \\
 & & & & \\
 & & 0 & 0 & a_{N,i+1}^{(i+1)} & a_{N,N}^{(i+1)}
 \end{array}$$

FIGURE 1.3: Allure de la matrice de Gauss à l'étape  $i + 1$ 

Le résultat général que l'on peut démontrer est que si la matrice  $A$  est inversible, alors il existe une matrice de permutation  $P$ , une matrice triangulaire inférieure  $L$  et une matrice triangulaire supérieure  $U$  telles que  $PA = LU$  : voir le théorème 1.19.

### Le cas général d'une matrice $n \times n$

De manière plus générale, pour une matrice  $A$  carrée d'ordre  $n$ , la méthode de Gauss s'écrit :

On pose  $A^{(1)} = A$  et  $b^{(1)} = b$ . Pour  $i = 1, \dots, n - 1$ , on cherche à calculer  $A^{(i+1)}$  et  $b^{(i+1)}$  tels que les systèmes  $A^{(i)}x = b^{(i)}$  et  $A^{(i+1)}x = b^{(i+1)}$  soient équivalents, où  $A^{(i+1)}$  est une matrice dont les coefficients sous-diagonaux des colonnes 1 à  $i$  sont tous nuls, voir figure 1.3 :

Une fois la matrice  $A^{(n)}$  (triangulaire supérieure) et le vecteur  $b^{(n)}$  calculés, il sera facile de résoudre le système  $A^{(n)}x = b^{(n)}$ . Le calcul de  $A^{(n)}$  est l'étape de "factorisation", le calcul de  $b^{(n)}$  l'étape de "descente", et le calcul de  $x$  l'étape de "remontée". Donnons les détails de ces trois étapes.

**Etape de factorisation et descente** Pour passer de la matrice  $A^{(i)}$  à la matrice  $A^{(i+1)}$ , on va effectuer des combinaisons linéaires entre lignes qui permettront d'annuler les coefficients de la  $i$ -ème colonne situés en dessous de la ligne  $i$  (dans le but de se rapprocher d'une matrice triangulaire supérieure). Evidemment, lorsqu'on fait ceci, il faut également modifier le second membre  $b$  en conséquence. L'étape de factorisation et descente s'écrit donc :

1. Pour  $k \leq i$  et pour  $j = 1, \dots, n$ , on pose  $a_{k,j}^{(i+1)} = a_{k,j}^{(i)}$  et  $b_k^{(i+1)} = b_k^{(i)}$ .
2. Pour  $k > i$ , si  $a_{i,i}^{(i)} \neq 0$ , on pose :

$$a_{k,j}^{(i+1)} = a_{k,j}^{(i)} - \frac{a_{k,i}^{(i)}}{a_{i,i}^{(i)}} a_{i,j}^{(i)}, \text{ pour } k = j, \dots, n, \quad (1.33)$$

$$b_k^{(i+1)} = b_k^{(i)} - \frac{a_{k,i}^{(i)}}{a_{i,i}^{(i)}} b_i^{(i)}. \quad (1.34)$$

La matrice  $A^{(i+1)}$  est de la forme donnée sur la figure 1.3. Remarquons que le système  $A^{(i+1)}x = b^{(i+1)}$  est bien équivalent au système  $A^{(i)}x = b^{(i)}$ .

Si la condition  $a_{i,i}^{(i)} \neq 0$  est vérifiée pour  $i = 1$  à  $n$ , on obtient par le procédé de calcul ci-dessus un système linéaire  $A^{(n)}x = b^{(n)}$  équivalent au système  $Ax = b$ , avec une matrice  $A^{(n)}$  triangulaire supérieure facile à inverser. On verra un peu plus loin les techniques de pivot qui permettent de régler le cas où la condition  $a_{i,i}^{(i)} \neq 0$  n'est pas vérifiée.

**Étape de remontée** Il reste à résoudre le système  $A^{(n)}x = b^{(n)}$ . Ceci est une étape facile. Comme  $A^{(n)}$  est une matrice inversible, on a  $a_{i,i}^{(i)} \neq 0$  pour tout  $i = 1, \dots, n$ , et comme  $A^{(n)}$  est une matrice triangulaire supérieure, on peut donc calculer les composantes de  $x$  en "remontant", c'est-à-dire de la composante  $x_n$  à la composante  $x_1$  :

$$x_n = \frac{b_n^{(n)}}{a_{n,n}^{(n)}},$$

$$x_i = \frac{1}{a_{i,i}^{(i)}} \left[ b^{(i)} - \sum_{j=i+1, n} a_{i,j}^{(i)} x_j \right], i = n-1, \dots, 1.$$

Il est important de savoir mettre sous forme algorithmique les opérations que nous venons de décrire : c'est l'étape clef avant l'écriture d'un programme informatique qui nous permettra de faire faire le boulot par l'ordinateur !

**Algorithme 1.11** (Gauss sans permutation).

1. (Factorisation et descente)

Pour  $i$  allant de 1 à  $n$ , on effectue les calculs suivants :

(a) On ne change pas la  $i$ -ème ligne (qui est la ligne du pivot)

$$u_{i,j} = a_{i,j} \text{ pour } j = i, \dots, n,$$

$$y_i = b_i$$

(b) On calcule les lignes  $i+1$  à  $n$  de  $U$  et le second membre  $y$  en utilisant la ligne  $i$ .

Pour  $k$  allant de  $i+1$  à  $n$  :

$$\ell_{k,i} = \frac{a_{k,i}}{a_{i,i}} \text{ (si } a_{i,i} = 0, \text{ prendre la méthode avec pivot partiel)}$$

pour  $j$  allant de  $i+1$  à  $n$ ,

$$u_{k,j} = a_{k,j} - \ell_{k,i} u_{i,j} \text{ (noter que } a_{k,i} = 0)$$

Fin pour

$$y_k = b_k - \ell_{k,i} y_i$$

Fin pour

2. (Remontée) On calcule  $x$  :

$$x_n = \frac{y_n}{u_{n,n}}$$

Pour  $i$  allant de  $n-1$  à 1,

$$x_i = y_i$$

Pour  $j$  allant de  $i+1$  à  $n$ ,

$$x_i = x_i - u_{i,j} x_j$$

Fin pour

$$x_i = \frac{1}{u_{i,i}} x_i$$

Fin pour

**Coût de la méthode de Gauss (nombre d'opérations)** On peut montrer (on fera le calcul de manière détaillée pour la méthode de Choleski dans la section suivante, le calcul pour Gauss est similaire) que le nombre d'opérations nécessaires  $n_G$  pour effectuer les étapes de factorisation, descente et remontée est  $\frac{2}{3}n^3 + O(n^2)$ ; on rappelle qu'une fonction  $f$  de  $\mathbb{N}$  dans  $\mathbb{N}$  est  $O(n^2)$  veut dire qu'il existe un réel constant  $C$  tel que  $f(n) \leq Cn^2$ . On a donc  $\lim_{n \rightarrow +\infty} \frac{n_G}{n^3} = \frac{2}{3}$  : lorsque  $n$  est grand, le nombre d'opérations se comporte comme  $n^3$ .

En ce qui concerne la place mémoire, on peut très bien stocker les itérés  $A^{(i)}$  dans la matrice  $A$  de départ, ce qu'on n'a pas voulu faire dans le calcul précédent, par souci de clarté.

**Décomposition LU** Si le système  $Ax = b$  doit être résolu pour plusieurs second membres  $b$ , on a déjà dit qu'on a intérêt à ne faire l'étape de factorisation (*i.e.* le calcul de  $A^{(n)}$ ), qu'une seule fois, alors que les étapes de descente et remontée (*i.e.* le calcul de  $b^{(n)}$  et  $x$ ) seront faits pour chaque vecteur  $b$ . L'étape de factorisation peut se faire en décomposant la matrice  $A$  sous la forme  $LU$ . Supposons toujours pour l'instant que lors de l'algorithme de Gauss, la condition  $a_{i,i}^{(i)} \neq 0$  est vérifiée pour tout  $i = 1, \dots, n$ . La matrice  $L$  a comme coefficients  $\ell_{k,i} = \frac{a_{k,i}^{(i)}}{a_{i,i}^{(i)}}$  pour  $k > i$ ,  $\ell_{i,i} = 1$  pour tout  $i = 1, \dots, n$ , et  $\ell_{i,j} = 0$  pour  $j > i$ , et la matrice  $U$  est égale à la matrice  $A^{(n)}$ . On peut vérifier que  $A = LU$  grâce au fait que le système  $A^{(n)}x = b^{(n)}$  est équivalent au système  $Ax = b$ . En effet, comme  $A^{(n)}x = b^{(n)}$  et  $b^{(n)} = L^{-1}b$ , on en déduit que  $LUx = b$ , et comme  $A$  et  $LU$  sont inversibles, on en déduit que  $A^{-1}b = (LU)^{-1}b$  pour tout  $b \in \mathbb{R}^n$ . Ceci démontre que  $A = LU$ . La méthode  $LU$  se déduit donc de la méthode de Gauss en remarquant simplement que, ayant conservé la matrice  $L$ , on peut effectuer les calculs sur  $b$  après les calculs sur  $A$ , ce qui donne :

**Algorithme 1.12** ( $LU$  simple (sans permutation)).

1. (Factorisation) Pour  $i$  allant de 1 à  $n$ , on effectue les calculs suivants :

(a) On ne change pas la  $i$ -ème ligne

$$u_{i,j} = a_{i,j} \text{ pour } j = i, \dots, n,$$

(b) On calcule les lignes  $i + 1$  à  $n$  de  $U$  en utilisant la ligne  $i$  (mais pas le second membre).

Pour  $k$  allant de  $i + 1$  à  $n$  :

$$\ell_{k,i} = \frac{a_{k,i}}{a_{i,i}} \text{ (si } a_{i,i} = 0, \text{ prendre la méthode avec pivot partiel)}$$

pour  $j$  allant de  $i + 1$  à  $n$ ,

$$u_{k,j} = a_{k,j} - \ell_{k,i}u_{i,j} \text{ (noter que } a_{k,i} = 0)$$

Fin pour

2. (Descente) On calcule  $y$  (avec  $Ly = b$ )

Pour  $i$  allant de 1 à  $n$ ,

$$y_i = b_i - \sum_{k=1}^{i-1} \ell_{i,k}y_k \text{ (on a ainsi implicitement } \ell_{i,i} = 1)$$

3. (Remontée) On calcule  $x$  (avec  $Ux = y$ )

Pour  $i$  allant de  $n$  à 1,

$$x_i = \frac{1}{u_{i,i}}(y_i - \sum_{j=i+1}^n u_{i,j}x_j)$$

**Remarque 1.13** (Optimisation mémoire). L'introduction des matrices  $L$  et  $U$  et des vecteurs  $y$  et  $x$  n'est pas nécessaire (tout peut s'écrire avec la matrice  $A$  et le vecteur  $b$ , que l'on modifie au cours de l'algorithme). L'introduction de  $L$ ,  $U$ ,  $x$  et  $y$  peut toutefois aider à comprendre la méthode. Le principe retenu est que, dans les algorithmes (Gauss ou  $LU$ ), on modifie la matrice  $A$  et le second membre  $b$  (en remplaçant le système à résoudre par un système équivalent) mais on ne modifie jamais  $L$ ,  $U$ ,  $y$  et  $x$  (qui sont définis au cours de l'algorithme).

Nous allons maintenant donner une condition nécessaire et suffisante (CNS) pour qu'une matrice  $A$  admette une décomposition  $LU$  avec  $U$  inversible et sans permutation. Commençons par un petit lemme technique qui va nous permettre de prouver cette CNS.

**Lemme 1.14** (Décomposition  $LU$  de la matrice principale d'ordre  $k$ ). Soit  $n \in \mathbb{N}$ ,  $A \in \mathcal{M}_n(\mathbb{R})$  et  $k \in \{1, \dots, N\}$ . On appelle matrice principale d'ordre  $k$  de  $A$  la matrice  $A_k \in \mathcal{M}_k(\mathbb{R})$  définie par  $(A_k)_{i,j} = a_{i,j}$  pour  $i = 1, \dots, k$  et  $j = 1, \dots, k$ . On suppose qu'il existe une matrice  $L_k \in \mathcal{M}_k(\mathbb{R})$  triangulaire inférieure de coefficients diagonaux tous égaux à 1 et une matrice triangulaire supérieure  $U_k \in \mathcal{M}_k(\mathbb{R})$  inversible, telles que  $A_k = L_k U_k$ . Alors  $A$  s'écrit sous la forme "par blocs" suivante :

$$A = \begin{bmatrix} L_k & 0_{k \times (n-k)} \\ C_k & \text{Id}_{n-k} \end{bmatrix} \begin{bmatrix} U_k & B_k \\ 0_{(n-k) \times k} & D_k \end{bmatrix}, \quad (1.35)$$

où  $0_{p,q}$  désigne la matrice nulle de dimension  $p \times q$ ,  $B_k \in \mathcal{M}_{k,n-k}(\mathbb{R})$  et  $C_k \in \mathcal{M}_{n-k,k}(\mathbb{R})$  et  $D_k \in \mathcal{M}_{n-k,n-k}(\mathbb{R})$ ; de plus, la matrice principale d'ordre  $k+1$  s'écrit sous la forme

$$A_{k+1} = \begin{bmatrix} L_k & 0_{1 \times k} \\ c_k & 1 \end{bmatrix} \begin{bmatrix} U_k & b_k \\ 0_{k \times 1} & d_k \end{bmatrix} \quad (1.36)$$

où  $b \in \mathcal{M}_{k,1}(\mathbb{R})$  est la première colonne de la matrice  $B_k$ ,  $c_k \in \mathcal{M}_{1,k}$  est la première ligne de la matrice  $C_k$ , et  $d_k$  est le coefficient de la ligne 1 et colonne 1 de  $D_k$ .

DÉMONSTRATION – On écrit la décomposition par blocs de  $A$  :

$$A = \begin{bmatrix} A_k & E_k \\ F_k & G_k \end{bmatrix},$$

avec  $A_k \in \mathcal{M}_k(\mathbb{R})$ ,  $E_k \in \mathcal{M}_{k,n-k}(\mathbb{R})$ ,  $F_k \in \mathcal{M}_{n-k,k}(\mathbb{R})$  et  $G_k \in \mathcal{M}_{n-k,n-k}(\mathbb{R})$ . Par hypothèse, on a  $A_k = L_k U_k$ . De plus  $L_k$  et  $U_k$  sont inversibles, et il existe donc une unique matrice  $B_k \in \mathcal{M}_k(\mathbb{R})$  (resp.  $C_k \in \mathcal{M}_k(\mathbb{R})$ ) telle que  $L_k B_k = E_k$  (resp.  $C_k U_k = G_k$ ). En posant  $D_k = G_k - B_k C_k$ , on obtient (1.35). L'égalité (1.36) en découle immédiatement. ■

**Proposition 1.15** (CNS pour  $LU$  sans permutation). Soit  $n \in \mathbb{N}$ ,  $A \in \mathcal{M}_n(\mathbb{R})$ . Les deux propriétés suivantes sont équivalentes.

(P1) Il existe un unique couple  $(L, U)$ , avec  $L$  matrice triangulaire inférieure de coefficients égaux à 1 et  $U$  une matrice inversible triangulaire supérieure, tel que  $A = LU$ .

(P2) Les mineurs principaux<sup>5</sup> de  $A$  sont tous non nuls.

DÉMONSTRATION – Si  $A = LU$  avec  $L$  triangulaire inférieure de coefficients égaux à 1 et  $U$  inversible triangulaire supérieure, alors  $A_k = L_k U_k$  où les matrices  $L_k$  et  $U_k$  les matrices principales d'ordre  $k$  de  $L$  et  $U$ , qui sont encore respectivement triangulaire inférieure de coefficients égaux à 1 et inversible triangulaire supérieure. On a donc

$$\det(A_k) = \det(L_k) \det(U_k) \neq 0 \text{ pour tout } k = 1, \dots, n,$$

et donc (P1)  $\Rightarrow$  (P2).

Montrons maintenant la réciproque. On suppose que les mineurs sont non nuls, et on va montrer que  $A = LU$ . On va en fait montrer que pour tout  $k = 1, \dots, n$ , on a  $A_k = L_k U_k$  où  $L_k$  triangulaire inférieure de coefficients égaux à 1 et  $U_k$  inversible triangulaire supérieure. Le premier mineur est non nul, donc  $a_{11} = 1 \times a_{11}$ , et la récurrence est bien initialisée. On la suppose vraie à l'étape  $k$ . Par le lemme 1.14, on a donc  $A_{k+1}$  qui est de la forme (1.36), et donc une  $A_{k+1} = L_{k+1} U_{k+1}$ . Comme  $\det(A_{k+1}) \neq 0$ , la matrice  $U_{k+1}$  est inversible, et l'hypothèse de récurrence est vérifiée à l'ordre  $k+1$ . On a donc bien (P2)  $\Rightarrow$  (P1). ■

**Que faire en cas de pivot nul : la technique de permutation** La caractérisation que nous venons de donner pour qu'une matrice admette une décomposition  $LU$  sans permutation est intéressante mathématiquement, mais de peu d'intérêt en pratique. On ne va en effet jamais calculer  $n$  déterminants pour savoir si non peut ou non permuter. En pratique, on effectue la décomposition  $LU$  sans savoir si on a le droit ou non de le faire, avec ou sans permutation. Au cours de l'élimination, si  $a_{i,i}^{(i)} = 0$ , on va permuter la ligne  $i$  avec une des lignes suivantes telle que  $a_{k,i}^{(i)} \neq 0$ . Notons que si le "pivot"  $a_{i,i}^{(i)}$  est très petit, son utilisation peut entraîner des erreurs d'arrondi importantes dans les calculs et on va là encore permuter. En fait, même dans le cas où la CNS donnée par la proposition 1.15 est vérifiée, la plupart des fonctions de bibliothèques scientifiques vont permuter.

Plaçons-nous à l'itération  $i$  de la méthode de Gauss. Comme la matrice  $A^{(i)}$  est forcément non singulière, on a :

$$\det(A^{(i)}) = a_{1,1}^{(i)} a_{2,2}^{(i)} \cdots a_{i-1,i-1}^{(i)} \det \begin{pmatrix} a_{i,i}^{(i)} & \cdots & a_{i,n}^{(i)} \\ \vdots & \ddots & \vdots \\ a_{n,i}^{(i)} & \cdots & a_{n,n}^{(i)} \end{pmatrix} \neq 0.$$

5. On rappelle que le mineur principal d'ordre  $k$  est le déterminant de la matrice principale d'ordre  $k$ .

On a donc en particulier

$$\det \begin{pmatrix} a_{i,i}^{(i)} & \dots & a_{i,n}^{(i)} \\ \vdots & \ddots & \vdots \\ a_{n,i}^{(i)} & \dots & a_{n,n}^{(i)} \end{pmatrix} \neq 0.$$

On déduit qu'il existe  $i_0 \in \{i, \dots, n\}$  tel que  $a_{i_0,i}^{(i)} \neq 0$ . On choisit alors  $i_0 \in \{i, \dots, n\}$  tel que  $|a_{i_0,i}^{(i)}| = \max\{|a_{k,i}^{(i)}|, k = i, \dots, n\}$ . Le choix de ce max est motivé par le fait qu'on aura ainsi moins d'erreur d'arrondi. On échange alors les lignes  $i$  et  $i_0$  (dans la matrice  $A$  et le second membre  $\mathbf{b}$ ) et on continue la procédure de Gauss décrite plus haut.

L'intérêt de cette stratégie de pivot est qu'on aboutit toujours à la résolution du système (dès que  $A$  est inversible).

**Remarque 1.16** (Pivot total). *La méthode que nous venons de d'écrire est souvent nommée technique de pivot "partiel". On peut vouloir rendre la norme du pivot encore plus grande en considérant tous les coefficients restants et pas uniquement ceux de la colonne  $i$ . A l'étape  $i$ , on choisit maintenant  $i_0$  et  $j_0 \in \{i, \dots, n\}$  tels que  $|a_{i_0,j_0}^{(i)}| = \max\{|a_{k,j}^{(i)}|, k = i, \dots, n, j = i, \dots, n\}$ , et on échange alors les lignes  $i$  et  $i_0$  (dans la matrice  $A$  et le second membre  $\mathbf{b}$ ), les colonnes  $i$  et  $j_0$  de  $A$  et les inconnues  $x_i$  et  $x_{j_0}$ . La stratégie du pivot total permet une moins grande sensibilité aux erreurs d'arrondi. L'inconvénient majeur est qu'on change la structure de  $A$  : si, par exemple la matrice avait tous ses termes non nuls sur quelques diagonales seulement, ceci n'est plus vrai pour la matrice  $A^{(n)}$ .*

Ecrivons maintenant l'algorithme de la méthode  $LU$  avec pivot partiel ; pour ce faire, on va simplement remarquer que l'ordre dans lequel les équations sont prises n'a aucune importance pour l'algorithme. Au départ de l'algorithme, on initialise la bijection  $t$  de  $\{1, \dots, n\}$  dans  $\{1, \dots, n\}$  par l'identité, c.à.d.  $t(i) = i$  ; cette bijection  $t$  va être modifiée au cours de l'algorithme pour tenir compte du choix du pivot.

**Algorithme 1.17** ( $LU$  avec pivot partiel).

1. (Initialisation de  $t$ ) Pour  $i$  allant de 1 à  $n$ ,  $t(i) = i$ . Fin pour
2. (Factorisation)

Pour  $i$  allant de 1 à  $n$ , on effectue les calculs suivants :

- (a) Choix du pivot (et de  $t(i)$ ) : on cherche  $i^* \in \{i, \dots, n\}$  t.q.  $|a_{t(i^*),i}| = \max\{|a_{t(k),i}|, k \in \{i, \dots, n\}\}$  (noter que ce max est forcément non nul car la matrice est inversible).

On modifie alors  $t$  en inversant les valeurs de  $t(i)$  et  $t(i^*)$ .

$$p = t(i^*) ; t(i^*) = t(i) ; t(i) = p.$$

On ne change pas la ligne  $t(i)$  :

$$u_{t(i),j} = a_{t(i),j} \text{ pour } j = i, \dots, n,$$

- (b) On modifie les lignes  $t(k)$ ,  $k > i$  (et le second membre), en utilisant la ligne  $t(i)$ .

Pour  $k = i + 1, \dots, n$  (noter qu'on a uniquement besoin de connaître l'ensemble, et pas l'ordre) :

$$\ell_{t(k),i} = \frac{a_{t(k),i}}{a_{t(i),i}}$$

Pour  $j$  allant de  $i + 1$  à  $n$ ,

$$u_{t(k),j} = a_{t(k),j} - \ell_{t(k),i} u_{t(i),j} \text{ (noter que } u_{t(k),i} = 0)$$

Fin pour

Fin pour

3. (Descente) On calcule  $\mathbf{y}$

Pour  $i$  allant de 1 à  $n$ ,

$$y_{t(i)} = b_{t(i)} - \sum_{j=1}^{i-1} \ell_{t(j),k} y_j$$

Fin pour

4. (Remontée) On calcule  $x$

Pour  $i$  allant de  $n$  à  $1$ ,

$$x_{t(i)} = \frac{1}{u_{t(i),i}} (y_i - \sum_{j=i+1}^n u_{t(i),j} x_j)$$

Fin pour

NB : On a changé l'ordre dans lequel les équations sont considérées (le tableau  $t$  donne cet ordre, et donc la matrice  $P$ ). Donc, on a aussi changé l'ordre dans lequel interviennent les composantes du second membre : le système  $Ax = b$  est devenu  $PAx = Pb$ . Par contre, on n'a pas touché à l'ordre dans lequel interviennent les composantes de  $x$  et  $y$ .

Il reste maintenant à signaler la propriété magnifique de cet algorithme... Il est inutile de connaître *a priori* la bijection pour cet algorithme. A l'étape  $i$  de l'item 1 (et d'ailleurs aussi à l'étape  $i$  de l'item 2), il suffit de connaître  $t(j)$  pour  $j$  allant de  $1$  à  $i$ , les opérations de 1(b) se faisant alors sur toutes les autres lignes (dans un ordre quelconque). Il suffit donc de partir d'une bijection arbitraire de  $\{1, \dots, n\}$  dans  $\{1, \dots, n\}$  (par exemple l'identité) et de la modifier à chaque étape. Pour que l'algorithme aboutisse, il suffit que  $a_{t(i),i} \neq 0$  (ce qui toujours possible car  $A$  est inversible).

**Remarque 1.18** (Ordre des équations et des inconnues). *L'algorithme se ramène donc à résoudre  $LUx = b$ , en résolvant d'abord  $Ly = b$  puis  $Ux = y$ . Notons que lors de la résolution du système  $Ly = b$ , les équations sont dans l'ordre  $t(1), \dots, t(k)$  (les composantes de  $b$  sont donc aussi prises dans cet ordre), mais le vecteur  $y$  est bien le vecteur de composantes  $(y_1, \dots, y_n)$ , dans l'ordre initial. Puis, on résout  $Ux = y$ , et les équations sont encore dans l'ordre  $t(1), \dots, t(k)$  mais les vecteurs  $x$  et  $y$  ont comme composantes respectives  $(x_1, \dots, x_n)$  et  $(y_1, \dots, y_n)$ .*

**Le théorème d'existence** L'algorithme LU avec pivot partiel nous permet de démontrer le théorème d'existence de la décomposition LU pour une matrice inversible.

**Théorème 1.19** (Décomposition LU d'une matrice). *Soit  $A \in \mathcal{M}_n(\mathbb{R})$  une matrice inversible, il existe une matrice de permutation  $P$  telle que, pour cette matrice de permutation, il existe un et un seul couple de matrices  $(L, U)$  où  $L$  est triangulaire inférieure de termes diagonaux égaux à  $1$  et  $U$  est triangulaire supérieure, vérifiant*

$$PA = LU.$$

DÉMONSTRATION –

1. **L'existence** de la matrice  $P$  et des matrices  $LU$  peut s'effectuer en s'inspirant de l'algorithme "LU avec pivot partiel" 1.17). Posons  $A^{(0)} = A$ .

À chaque étape  $i$  de l'algorithme 1.17 peut s'écrire comme  $A^{(i)} = E^{(i)} P^{(i)} A^{(i-1)}$ , où  $P^{(i)}$  est la matrice de permutation qui permet le choix du pivot partiel, et  $E^{(i)}$  est une matrice d'élimination qui effectue les combinaisons linéaires de lignes permettant de mettre à zéro tous les coefficients de la colonne  $i$  situés en dessous de la ligne  $i$ . Pour simplifier, raisonnons sur une matrice  $4 \times 4$  (le raisonnement est le même pour une matrice  $n \times n$ . On a donc en appliquant l'algorithme de Gauss :

$$E^{(3)} P^{(3)} E^{(2)} P^{(2)} E^{(1)} P^{(1)} A = U$$

Les matrices  $P^{(i+1)}$  et  $E^{(i)}$  ne permutent en gén'ral pas. Prenons par exemple  $E_2$ , qui est de la forme

$$E^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & a & 1 & 0 \\ 0 & b & 0 & 1 \end{bmatrix}$$

Si  $P^{(3)}$  est la matrice qui échange les lignes 3 et 4, alors

$$P^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ et } P^{(3)} E^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & b & 0 & 1 \\ 0 & a & 1 & 0 \end{bmatrix}, \text{ alors que } E^{(2)} P^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & a & 0 & 1 \\ 0 & b & 1 & 0 \end{bmatrix}$$

Mais par contre, comme la multiplication à gauche par  $P^{(i+1)}$  permute les lignes  $i + 1$  et  $i + k$ , pour un certain  $k \geq 1$ , et que la multiplication à droite permute les colonnes  $i + 1$  et  $i + k$ , la matrice  $\widetilde{E}^{(i)} = P^{(i+1)} E^{(i)} P^{(i+1)}$  est encore une matrice triangulaire inférieure avec la même structure que  $E^{(i)}$  : on a juste échangé les coefficients extradiagonaux des lignes  $i + 1$  et  $i + k$ . On a donc

$$P^{(i+1)} E^{(i)} = \widetilde{E}^{(i)} P^{(i+1)}. \quad (1.37)$$

Dans l'exemple précédent, on effectue le calcul :

$$P^{(3)} E^{(2)} P^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & b & 1 & 0 \\ 0 & a & 0 & 1 \end{bmatrix} = \widetilde{E}^{(2)},$$

qui est une matrice triangulaire inférieure de coefficients tous égaux à 1, et comme  $P^{(3)} P^{(3)} = \text{Id}$ , on a donc :

$$P^{(3)} E^{(2)} = \widetilde{E}^{(2)} P^{(3)}.$$

Pour revenir à notre exemple  $n = 4$ , on peut donc écrire :

$$E^{(3)} \widetilde{E}^{(2)} P^{(3)} \widetilde{E}^{(1)} P^{(2)} P^{(1)} A = U$$

Mais par le même raisonnement que précédemment, on a  $P^{(3)} \widetilde{E}^{(1)} = \widetilde{\widetilde{E}}^{(1)} P^{(3)}$  où  $\widetilde{\widetilde{E}}^{(1)}$  est encore une matrice triangulaire inférieure avec des 1 sur la diagonale. On en déduit que

$$E^{(3)} \widetilde{E}^{(2)} \widetilde{\widetilde{E}}^{(1)} P^{(3)} P^{(2)} P^{(1)} A = U, \text{ soit encore } PA = LU$$

où  $P = P^{(3)} P^{(2)} P^{(1)}$  bien une matrice de permutation, et  $L = (E^{(3)} \widetilde{E}^{(2)} \widetilde{\widetilde{E}}^{(1)})^{-1}$  est une matrice triangulaire inférieure avec des 1 sur la diagonale.

Le raisonnement que nous venons de faire pour  $n = 3$  se généralise facilement à  $n$  quelconque. Dans ce cas, l'échelonnement de la matrice s'écrit sous la forme

$$U = E^{(n-1)} P^{(n-1)} \dots E^{(2)} P^{(2)} E^{(1)} P^{(1)} A,$$

et se transforme grâce à (1.37) en

$$U = F^{(n-1)} \dots F^{(2)} F^{(1)} P^{(n-1)} \dots P^{(2)} P^{(1)} A,$$

où les matrices  $F^{(i)}$  sont des matrices triangulaires inférieures de coefficients diagonaux tous égaux à 1. Plus précisément,  $F^{(n-1)} = E^{(n-1)}$ ,  $F^{(n-2)} = \widetilde{E}^{(n-2)}$ ,  $F^{(n-3)} = \widetilde{\widetilde{E}}^{(n-3)}$ , etc... On a ainsi démontré l'existence de la décomposition  $LU$ .

2. Pour montrer l'unicité du couple  $(L, U)$  à  $P$  donnée, supposons qu'il existe une matrice  $P$  et des matrices  $L_1, L_2$ , triangulaires inférieures et  $U_1, U_2$ , triangulaires supérieures, telles que

$$PA = L_1 U_1 = L_2 U_2$$

Dans ce cas, on a donc  $L_2^{-1} L_1 = U_2 U_1^{-1}$ . Or la matrice  $L_2^{-1} L_1$  est une matrice triangulaire inférieure dont les coefficients diagonaux sont tous égaux à 1, et la matrice  $U_2 U_1^{-1}$  est une matrice triangulaire supérieure. On en déduit que  $L_2^{-1} L_1 = U_2 U_1^{-1} = \text{Id}$ , et donc que  $L_1 = L_2$  et  $U_1 = U_2$ . ■

**Remarque 1.20** (Décomposition LU pour les matrices non inversibles). *En fait n'importe quelle matrice carrée admet une décomposition de la forme  $PA = LU$ . Mais si la matrice  $A$  n'est pas inversible, son échelonnement va nous donner des lignes de zéros pour les dernières lignes. La décomposition LU n'est dans ce cas pas unique. Cette remarque fait l'objet de l'exercice 22.*

### 1.3.3 Méthode de Choleski

On va maintenant étudier la méthode de Choleski, qui est une méthode directe adaptée au cas où  $A$  est symétrique définie positive. On rappelle qu'une matrice  $A \in \mathcal{M}_n(\mathbb{R})$  de coefficients  $(a_{i,j})_{i=1,n,j=1,n}$  est symétrique si  $A = A^t$ , où  $A^t$  désigne la transposée de  $A$ , définie par les coefficients  $(a_{j,i})_{i=1,n,j=1,n}$ , et que  $A$  est définie positive si  $Ax \cdot x > 0$  pour tout  $x \in \mathbb{R}^n$  tel que  $x \neq 0$ . Dans toute la suite,  $x \cdot y$  désigne le produit scalaire des deux vecteurs  $x$  et  $y$  de  $\mathbb{R}^n$ . On rappelle (exercice) que si  $A$  est symétrique définie positive elle est en particulier inversible.

**Description de la méthode**

Commençons par un exemple. On considère la matrice  $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ , qui est symétrique. Calculons sa décomposition  $LU$ . Par échelonnement, on obtient

$$A = LU = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & 0 & \frac{4}{3} \end{bmatrix}$$

La structure  $LU$  ne conserve pas la symétrie de la matrice  $A$ . Pour des raisons de coût mémoire, il est important de pouvoir la conserver. Une façon de faire est de décomposer  $U$  en sa partie diagonale fois une matrice triangulaire. On obtient

$$U = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix}$$

On a donc  $U = DL^t$ , et comme tous les coefficients de  $D$  sont positifs, on peut écrire  $D = \sqrt{D}\sqrt{D}$ , où  $\sqrt{D}$  est la matrice diagonale dont les éléments diagonaux sont les racines carrées des éléments diagonaux de  $A$ . On a donc  $A = L\sqrt{D}\sqrt{D}L^t = \tilde{L}\tilde{L}^t$ , avec  $\tilde{L} = L\sqrt{D}$ . Notons que la matrice  $\tilde{L}$  est toujours triangulaire inférieure, mais ses coefficients diagonaux ne sont plus astreints à être égaux à 1. C'est la décomposition de Choleski de la matrice  $A$ .

De fait, la méthode de Choleski consiste donc à trouver une décomposition d'une matrice  $A$  symétrique définie positive de la forme  $A = LL^t$ , où  $L$  est triangulaire inférieure de coefficients diagonaux strictement positifs. On résout alors le système  $Ax = b$  en résolvant d'abord  $Ly = b$  puis le système  $L^tx = y$ . Une fois la matrice  $A$  "factorisée", c'est-à-dire la décomposition  $LL^t$  obtenue (voir paragraphe suivant), on effectue les étapes de "descente" et "remontée" :

1. Etape 1 : "descente" Le système  $Ly = b$  s'écrit :

$$Ly = \begin{bmatrix} \ell_{1,1} & 0 & & \\ \vdots & \ddots & \vdots & \\ \ell_{n,1} & \dots & \ell_{n,n} & \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Ce système s'écrit composante par composante en partant de  $i = 1$ .

$$\begin{aligned} \ell_{1,1}y_1 &= b_1, \text{ donc} & y_1 &= \frac{b_1}{\ell_{1,1}} \\ \ell_{2,1}y_1 + \ell_{2,2}y_2 &= b_2, \text{ donc} & y_2 &= \frac{1}{\ell_{2,2}}(b_2 - \ell_{2,1}y_1) \\ &\vdots & &\vdots \\ \sum_{j=1,i} \ell_{i,j}y_j &= b_i, \text{ donc} & y_i &= \frac{1}{\ell_{i,i}}(b_i - \sum_{j=1,i-1} \ell_{i,j}y_j) \\ &\vdots & &\vdots \\ \sum_{j=1,n} \ell_{n,j}y_j &= b_n, \text{ donc} & y_n &= \frac{1}{\ell_{n,n}}(b_n - \sum_{j=1,n-1} \ell_{n,j}y_j). \end{aligned}$$

On calcule ainsi  $y_1, y_2, \dots, y_n$ .

2. Etape 2 : "remontée" On calcule maintenant  $x$  solution de  $L^t x = y$ .

$$L^t x = \begin{bmatrix} \ell_{1,1} & \ell_{2,1} & \dots & \ell_{n,1} \\ 0 & \ddots & & \\ \vdots & & & \\ 0 & \dots & & \ell_{n,n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

On a donc :

$$\ell_{n,n} x_n = y_n \text{ donc } x_n = \frac{y_n}{\ell_{n,n}}$$

$$\ell_{n-1,n-1} x_{n-1} + \ell_{n,n-1} x_n = y_{n-1} \text{ donc } x_{n-1} = \frac{y_{n-1} - \ell_{n,n-1} x_n}{\ell_{n-1,n-1}}$$

$\vdots$

$$\sum_{j=1,n} \ell_{j,1} x_j = y_1 \text{ donc } x_1 = \frac{y_1 - \sum_{j=2,n} \ell_{j,1} x_j}{\ell_{1,1}}.$$

On calcule ainsi  $x_n, x_{n-1}, \dots, x_1$ .

### Existence et unicité de la décomposition

Soit  $A$  une matrice symétrique définie positive. On sait déjà par le théorème 1.19 page 35, qu'il existe une matrice de permutation et  $L$  triangulaire inférieure et  $U$  triangulaire supérieure telles que  $PA = LU$ . L'avantage dans le cas où la matrice est symétrique définie positive, est que la décomposition est toujours possible sans permutation. On prouve l'existence et unicité en construisant la décomposition, c.à.d. en construisant la matrice  $L$ .

Pour comprendre le principe de la preuve, commençons d'abord par le cas  $n = 2$ . Dans ce cas on peut écrire

$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ . On sait que  $a > 0$  car  $A$  est s.d.p. . L'échelonnement de  $A$  donne donc

$$A = LU = \begin{bmatrix} 1 & 0 \\ \frac{b}{a} & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & c - \frac{b^2}{a} \end{bmatrix}$$

En extrayant la diagonale de  $U$ , on obtient :

$$A = LU = \begin{bmatrix} 1 & 0 \\ \frac{b}{a} & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & c - \frac{b^2}{a} \end{bmatrix} \begin{bmatrix} a & \frac{b}{a} \\ 0 & 1 \end{bmatrix}.$$

Et donc

$$A = \tilde{L}\tilde{L}^t \text{ avec } \tilde{L} = \begin{bmatrix} \sqrt{a} & 0 \\ b\sqrt{\frac{ac-b^2}{a}} & 1 \end{bmatrix}.$$

**Théorème 1.21** (Décomposition de Choleski). *Soit  $A \in \mathcal{M}_n(\mathbb{R})$  ( $n \geq 1$ ) une matrice symétrique définie positive. Alors il existe une unique matrice  $L \in \mathcal{M}_n(\mathbb{R})$ ,  $L = (\ell_{i,j})_{i,j=1}^n$ , telle que :*

1.  $L$  est triangulaire inférieure (c'est-à-dire  $\ell_{i,j} = 0$  si  $j > i$ ),
2.  $\ell_{i,i} > 0$ , pour tout  $i \in \{1, \dots, n\}$ ,
3.  $A = LL^t$ .

DÉMONSTRATION –

I- Existence de  $L$  : démonstration par récurrence sur  $n$

1. Dans le cas  $n = 1$ , on a  $A = (a_{1,1})$ . Comme  $A$  est symétrique définie positive, on a  $a_{1,1} > 0$ . On peut donc définir  $L = (\ell_{1,1})$  où  $\ell_{1,1} = \sqrt{a_{1,1}}$ , et on a bien  $A = LL^t$ .
2. On suppose que la décomposition de Choleski s'obtient pour  $A \in \mathcal{M}_p(\mathbb{R})$  symétrique définie positive, pour  $1 \leq p \leq n$  et on va démontrer que la propriété est encore vraie pour  $A \in \mathcal{M}_{n+1}(\mathbb{R})$  symétrique définie positive. Soit donc  $A \in \mathcal{M}_{n+1}(\mathbb{R})$  symétrique définie positive ; on peut écrire  $A$  sous la forme :

$$A = \left[ \begin{array}{c|c} B & a \\ \hline a^t & \alpha \end{array} \right] \quad (1.38)$$

où  $B \in \mathcal{M}_n(\mathbb{R})$  est symétrique,  $a \in \mathbb{R}^n$  et  $\alpha \in \mathbb{R}$ . Montrons que  $B$  est définie positive, c.à.d. que  $By \cdot y > 0$ , pour tout  $y \in \mathbb{R}^n$  tel que  $y \neq 0$ . Soit donc  $y \in \mathbb{R}^n \setminus \{0\}$ , et  $x = \begin{bmatrix} y \\ 0 \end{bmatrix} \in \mathbb{R}^{n+1}$ . Comme  $A$  est symétrique définie positive, on a :

$$0 < Ax \cdot x = \left[ \begin{array}{c|c} B & a \\ \hline a^t & \alpha \end{array} \right] \begin{bmatrix} y \\ 0 \end{bmatrix} \cdot \begin{bmatrix} y \\ 0 \end{bmatrix} = \left[ \begin{array}{c|c} By & \\ \hline a^t y & \end{array} \right] \begin{bmatrix} y \\ 0 \end{bmatrix} = By \cdot y$$

donc  $B$  est définie positive. Par hypothèse de récurrence, il existe une matrice  $M \in \mathcal{M}_n(\mathbb{R})$   $M = (m_{i,j})_{i,j=1}^n$  telle que :

- (a)  $m_{i,j} = 0$  si  $j > i$
- (b)  $m_{i,i} > 0$
- (c)  $B = MM^t$ .

On va chercher  $L$  sous la forme :

$$L = \left[ \begin{array}{c|c} M & 0 \\ \hline b^t & \lambda \end{array} \right] \quad (1.39)$$

avec  $b \in \mathbb{R}^n$ ,  $\lambda \in \mathbb{R}_+^*$  tels que  $LL^t = A$ . Pour déterminer  $b$  et  $\lambda$ , calculons  $LL^t$  où  $L$  est de la forme (1.39) et identifions avec  $A$  :

$$LL^t = \left[ \begin{array}{c|c} M & 0 \\ \hline b^t & \lambda \end{array} \right] \left[ \begin{array}{c|c} M^t & b \\ \hline 0 & \lambda \end{array} \right] = \left[ \begin{array}{c|c} MM^t & Mb \\ \hline b^t M^t & b^t b + \lambda^2 \end{array} \right]$$

On cherche  $b \in \mathbb{R}^n$  et  $\lambda \in \mathbb{R}_+^*$  tels que  $LL^t = A$ , et on veut donc que les égalités suivantes soient vérifiées :

$$Mb = a \text{ et } b^t b + \lambda^2 = \alpha.$$

Comme  $M$  est inversible (en effet, le déterminant de  $M$  s'écrit  $\det(M) = \prod_{i=1}^n m_{i,i} > 0$ ), la première égalité ci-dessus donne :  $b = M^{-1}a$  et en remplaçant dans la deuxième égalité, on obtient :  $(M^{-1}a)^t (M^{-1}a) + \lambda^2 = \alpha$ , donc  $a^t (M^t)^{-1} M^{-1} a + \lambda^2 = \alpha$  soit encore  $a^t (MM^t)^{-1} a + \lambda^2 = \alpha$ , c'est-à-dire :

$$a^t B^{-1} a + \lambda^2 = \alpha \quad (1.40)$$

Pour que (1.40) soit vérifiée, il faut que

$$\alpha - a^t B^{-1} a > 0 \quad (1.41)$$

Montrons que la condition (1.41) est effectivement vérifiée : Soit  $z = \begin{pmatrix} B^{-1}a \\ -1 \end{pmatrix} \in \mathbb{R}^{n+1}$ . On a  $z \neq 0$  et donc  $Az \cdot z > 0$  car  $A$  est symétrique définie positive. Calculons  $Az$  :

$$Az = \left[ \begin{array}{c|c} B & a \\ \hline a^t & \alpha \end{array} \right] \begin{bmatrix} B^{-1}a \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ a^t B^{-1}a - \alpha \end{bmatrix}.$$

On a donc  $Az \cdot z = \alpha - a^t B^{-1} a > 0$  ce qui montre que (1.41) est vérifiée. On peut ainsi choisir  $\lambda = \sqrt{\alpha - a^t B^{-1} a}$  ( $> 0$ ) de telle sorte que (1.40) est vérifiée. Posons :

$$L = \left[ \begin{array}{c|c} M & 0 \\ \hline (M^{-1}a)^t & \lambda \end{array} \right].$$

La matrice  $L$  est bien triangulaire inférieure et vérifie  $\ell_{i,i} > 0$  et  $A = LL^t$ .

On a terminé ainsi la partie “existence”.

**II- Unicité et calcul de  $L$ .** Soit  $A \in \mathcal{M}_n(\mathbb{R})$  symétrique définie positive ; on vient de montrer qu’il existe  $L \in \mathcal{M}_n(\mathbb{R})$  triangulaire inférieure telle que  $\ell_{i,j} = 0$  si  $j > i$ ,  $\ell_{i,i} > 0$  et  $A = LL^t$ . On a donc :

$$a_{i,j} = \sum_{k=1}^n \ell_{i,k} \ell_{j,k}, \quad \forall (i,j) \in \{1 \dots n\}^2. \quad (1.42)$$

1. Calculons la 1-ère colonne de  $L$  ; pour  $j = 1$ , on a :

$$\begin{aligned} a_{1,1} &= \ell_{1,1} \ell_{1,1} \text{ donc } \ell_{1,1} = \sqrt{a_{1,1}} \quad (a_{1,1} > 0 \text{ car } \ell_{1,1} \text{ existe}), \\ a_{2,1} &= \ell_{2,1} \ell_{1,1} \text{ donc } \ell_{2,1} = \frac{a_{2,1}}{\ell_{1,1}}, \\ a_{i,1} &= \ell_{i,1} \ell_{1,1} \text{ donc } \ell_{i,1} = \frac{a_{i,1}}{\ell_{1,1}} \quad \forall i \in \{2, \dots, n\}. \end{aligned}$$

2. On suppose avoir calculé les  $q$  premières colonnes de  $L$ . On calcule la colonne  $(q+1)$  en prenant  $j = q+1$  dans (1.42)

$$\text{Pour } i = q+1, a_{q+1,q+1} = \sum_{k=1}^{q+1} \ell_{q+1,k} \ell_{q+1,k} \text{ donc}$$

$$\ell_{q+1,q+1} = (a_{q+1,q+1} - \sum_{k=1}^q \ell_{q+1,k}^2)^{1/2} > 0. \quad (1.43)$$

Notons que  $a_{q+1,q+1} - \sum_{k=1}^q \ell_{q+1,k}^2 > 0$  car  $L$  existe : il est indispensable d’avoir d’abord montré l’existence de  $L$  pour pouvoir exhiber le coefficient  $\ell_{q+1,q+1}$ .

On procède de la même manière pour  $i = q+2, \dots, n$  ; on a :

$$a_{i,q+1} = \sum_{k=1}^{q+1} \ell_{i,k} \ell_{q+1,k} = \sum_{k=1}^q \ell_{i,k} \ell_{q+1,k} + \ell_{i,q+1} \ell_{q+1,q+1}$$

et donc

$$\ell_{i,q+1} = \left( a_{i,q+1} - \sum_{k=1}^q \ell_{i,k} \ell_{q+1,k} \right) \frac{1}{\ell_{q+1,q+1}}. \quad (1.44)$$

On calcule ainsi toutes les colonnes de  $L$ . On a donc montré que  $L$  est unique par un moyen constructif de calcul de  $L$ .

■

**Remarque 1.22** (Choleski et  $LU$ ). *Considérons une matrice  $A$  symétrique définie positive. Alors une matrice  $P$  de permutation dans le théorème 1.21 possible n’est autre que l’identité. Il suffit pour s’en convaincre de remarquer qu’une fois qu’on s’est donné la bijection  $t = \text{Id}$  dans l’algorithme 1.17, celle-ci n’est jamais modifiée et donc on a  $P = \text{Id}$ . Les théorèmes d’existence et d’unicité 1.19 et 1.21 nous permettent alors de remarquer que  $A = LU = \tilde{L}\tilde{L}^t$  avec  $\tilde{L} = L\sqrt{D}$ , où  $D$  est la matrice diagonale extraite de  $U$ , et  $\sqrt{D}$  désigne la matrice dont les coefficients sont les racines carrées des coefficients de  $D$  (qui sont tous positifs). Voir à ce sujet l’exercice 23 46.*

La décomposition  $LU$  permet de caractériser les matrices symétriques définies positives.

**Proposition 1.23** (Caractérisation des matrices symétriques définies positives par la décomposition  $LU$ ). *Soit  $A$  une matrice symétrique admettant une décomposition  $LU$  sans permutation, c'est-à-dire qu'on suppose qu'il existe  $L$  triangulaire inférieure de coefficients diagonaux tous égaux à 1, et  $U$  triangulaire supérieure telle que  $A = LU$ . Alors  $A$  est symétrique définie positive si et seulement si tous les pivots (c'est-à-dire les coefficients diagonaux de la matrice  $U$ ) sont strictement positifs.*

DÉMONSTRATION – Soit  $A$  une matrice symétrique admettant une décomposition  $LU$  sans permutation. Si  $A$  est symétrique définie positive, le théorème 1.21 de décomposition de Choleski donne immédiatement le résultat.

Montrons maintenant la réciproque : supposons que  $A = LU$  avec tous les pivots strictement positifs. On a donc  $A = LU$ , et  $U$  est inversible car c'est une matrice triangulaire supérieure dont tous les coefficients diagonaux sont strictement positifs. Donc  $A$  est aussi inversible, et la décomposition  $LU$  est donc unique, par le théorème 1.19 de décomposition  $LU$  d'une matrice inversible. On a donc  $A = LU = LD\tilde{L}^t$  où  $D$  est la matrice diagonale dont la diagonale est celle de  $U$ , et  $\tilde{L}$  est la matrice triangulaire inférieure de coefficients diagonaux tous égaux à 1 définie par  $\tilde{L}^t = D^{-1}U$ . On a donc aussi par symétrie de  $A$

$$A^t = \tilde{L}DL^t = A = LU$$

et par unicité de la décomposition  $LU$ , on en déduit que  $\tilde{L} = L$  et  $DL^t = U$ , ce qui entraîne que  $A = LDL^t = CC^t$  avec  $C = L\sqrt{D}$ . On a donc pour tout  $\mathbf{x} \in \mathbb{R}^n$ ,  $A\mathbf{x} \cdot \mathbf{x} = CC^t\mathbf{x} \cdot \mathbf{x} = \|C\mathbf{x}\|^2$  et donc que  $A$  est symétrique définie positive. ■

Attention : la proposition précédente est fautive si la décomposition est avec permutation, méditer pour s'en convaincre l'exemple  $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ .

**Remarque 1.24** (Pivot partiel et Choleski.). *Considérons une matrice  $A$  symétrique définie positive. On a vu dans le théorème qu'on n'a pas besoin de permutation pour obtenir la décomposition  $LL^t$  d'une matrice symétrique définie positive. Par contre, on utilise malgré tout la technique de pivot partiel pour minimiser les erreurs d'arrondi. On peut illustrer cette raison par l'exemple suivant :*

$$A = \begin{bmatrix} -10^{-n} & 1 \\ 1 & 1 \end{bmatrix}$$

À titre d'illustration, pour  $n = 12$  en FORTRAN (double précision), on obtient la bonne solution, c.à.d.  $(-1, 1)$ , avec le programme `gausslupivot` donné plus haut, alors que le programme sans pivot `gausslu` donne comme solution  $(0, 1)$ .

### Calcul du coût de la méthode de Choleski

**Calcul du coût de calcul de la matrice  $L$**  Dans le procédé de calcul de  $L$  exposé ci-dessus, le nombre d'opérations pour calculer la première colonne est  $n$ . Calculons, pour  $p = 0, \dots, n-1$ , le nombre d'opérations pour calculer la  $(p+1)$ -ième colonne : pour la colonne  $(p+1)$ , le nombre d'opérations par ligne est  $2p+1$ , car le calcul de  $\ell_{p+1,p+1}$  par la formule (1.43) nécessite  $p$  multiplications,  $p$  soustractions et une extraction de racine, soit  $2p+1$  opérations ; le calcul de  $\ell_{i,p+1}$  par la formule (1.44) nécessite  $p$  multiplications,  $p$  soustractions et une division, soit encore  $2p+1$  opérations. Comme les calculs se font des lignes  $p+1$  à  $n$  (car  $\ell_{i,p+1} = 0$  pour  $i \leq p$ ), le nombre d'opérations pour calculer la  $(p+1)$ -ième colonne est donc  $(2p+1)(n-p)$ . On en déduit que le nombre d'opérations  $N_L$  nécessaires au calcul de  $L$  est :

$$\begin{aligned} N_L &= \sum_{p=0}^{n-1} (2p+1)(n-p) = 2n \sum_{p=0}^{n-1} p - 2 \sum_{p=0}^{n-1} p^2 + n \sum_{p=0}^{n-1} 1 - \sum_{p=0}^{n-1} p \\ &= (2n-1) \frac{n(n-1)}{2} + n^2 - 2 \sum_{p=0}^{n-1} p^2. \end{aligned}$$

(On rappelle que  $2 \sum_{p=0}^{n-1} p = n(n-1)$ .) Il reste à calculer  $C_n = \sum_{p=0}^n p^2$ , en remarquant par exemple que

$$\begin{aligned} \sum_{p=0}^n (1+p)^3 &= \sum_{p=0}^n 1 + p^3 + 3p^2 + 3p = \sum_{p=0}^n 1 + \sum_{p=0}^n p^3 + 3 \sum_{p=0}^n p^2 + 3 \sum_{p=0}^n p \\ &= \sum_{p=1}^{n+1} p^3 = \sum_{p=0}^n p^3 + (n+1)^3. \end{aligned}$$

On a donc  $3C_n + 3\frac{n(n+1)}{2} + n + 1 = (n+1)^3$ , d'où on déduit que

$$C_n = \frac{n(n+1)(2n+1)}{6}.$$

On a donc :

$$\begin{aligned} N_L &= (2n-1) \frac{n(n-1)}{2} - 2C_{n-1} + n^2 \\ &= n \left( \frac{2n^2 + 3n + 1}{6} \right) = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} = \frac{n^3}{3} + 0(n^2). \end{aligned}$$

**Coût de la résolution d'un système linéaire par la méthode  $LL^t$**  Nous pouvons maintenant calculer le coût (en termes de nombre d'opérations élémentaires) nécessaire à la résolution de (1.1) par la méthode de Choleski pour  $A \in \mathcal{M}_n(\mathbb{R})$  symétrique définie positive. On a besoin de  $N_L$  opérations pour le calcul de  $L$ , auquel il faut rajouter le nombre d'opérations nécessaires pour les étapes de descente et remontée. Le calcul de  $y$  solution de  $Ly = b$  s'effectue en résolvant le système :

$$\begin{bmatrix} \ell_{1,1} & & 0 \\ \vdots & \ddots & \vdots \\ \ell_{n,1} & \dots & \ell_{n,n} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Pour la ligne 1, le calcul  $y_1 = \frac{b_1}{\ell_{1,1}}$  s'effectue en une opération.

Pour les lignes  $p = 2$  à  $n$ , le calcul  $y_p = (b_p - \sum_{i=1}^{p-1} \ell_{i,p} y_i) / \ell_{p,p}$  s'effectue en  $(p-1)$  (multiplications) +  $(p-2)$  (additions) + 1 soustraction + 1 (division) =  $2p-1$  opérations. Le calcul de  $y$  (descente) s'effectue donc en  $N_1 = \sum_{p=1}^n (2p-1) = n(n+1) - n = n^2$ . On peut calculer de manière similaire le nombre d'opérations nécessaires pour l'étape de remontée  $N_2 = n^2$ . Le nombre total d'opérations pour calculer  $x$  solution de (1.1) par la méthode de Choleski est  $N_C = N_L + N_1 + N_2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} + 2n^2 = \frac{n^3}{3} + \frac{5n^2}{2} + \frac{n}{6}$ . L'étape la plus coûteuse est donc la factorisation de  $A$ .

**Remarque 1.25** (Décomposition  $LDL^t$ ). Dans les programmes informatiques, on préfère implanter la variante suivante de la décomposition de Choleski :  $A = \tilde{L}D\tilde{L}^t$  où  $D$  est la matrice diagonale définie par  $d_{i,i} = \ell_{i,i}^2$ ,  $\tilde{L}_{i,i} = L\tilde{D}^{-1}$ , où  $\tilde{D}$  est la matrice diagonale définie par  $d_{i,i} = \ell_{i,i}$ . Cette décomposition a l'avantage de ne pas faire intervenir le calcul de racines carrées, qui est une opération plus compliquée que les opérations "élémentaires" ( $\times$ ,  $+$ ,  $-$ ).

### 1.3.4 Quelques propriétés

#### Comparaison Gauss/Choleski

Soit  $A \in \mathcal{M}_n(\mathbb{R})$  inversible, la résolution de (1.1) par la méthode de Gauss demande  $2n^3/3 + 0(n^2)$  opérations (exercice). Dans le cas d'une matrice symétrique définie positive, la méthode de Choleski est donc environ deux fois moins chère.

### Et la méthode de Cramer ?

Soit  $A \in \mathcal{M}_n(\mathbb{R})$  inversible. On rappelle que la méthode de Cramer pour la résolution de (1.1) consiste à calculer les composantes de  $x$  par les formules :

$$x_i = \frac{\det(A_i)}{\det(A)}, \quad i = 1, \dots, n,$$

où  $A_i$  est la matrice carrée d'ordre  $n$  obtenue à partir de  $A$  en remplaçant la  $i$ -ème colonne de  $A$  par le vecteur  $b$ , et  $\det(A)$  désigne le déterminant de  $A$ .

Chaque calcul de déterminant d'une matrice carrée d'ordre  $n$  nécessite au moins  $n!$  opérations (voir cours L1-L2, ou livres d'algèbre linéaire proposés en avant-propos). Par exemple, pour  $n = 10$ , la méthode de Gauss nécessite environ 700 opérations, la méthode de Choleski environ 350 et la méthode de Cramer plus de 4 000 000. . . . Cette dernière méthode est donc à proscrire.

### Conservation du profil de $A$

Dans de nombreuses applications, par exemple lors de la résolution de systèmes linéaires issus de la discrétisation<sup>6</sup> de problèmes réels, la matrice  $A \in \mathcal{M}_n(\mathbb{R})$  est "creuse", au sens où un grand nombre de ses coefficients sont nuls. Il est intéressant dans ce cas pour des raisons d'économie de mémoire de connaître le "profil" de la matrice, donné dans le cas où la matrice est symétrique, par les indices  $j_i = \min\{j \in \{1, \dots, n\} \text{ tels que } a_{i,j} \neq 0\}$ . Le profil de la matrice est donc déterminé par les diagonales contenant des coefficients non nuls qui sont les plus éloignées de la diagonale principale. Dans le cas d'une matrice creuse, il est avantageux de faire un stockage "profil" de  $A$ , en stockant, pour chaque ligne  $i$  la valeur de  $j_i$  et des coefficients  $a_{i,k}$ , pour  $k = i - j_i, \dots, i$ , ce qui peut permettre un large gain de place mémoire.

Une propriété intéressante de la méthode de Choleski est de conserver le profil. On peut montrer (en reprenant les calculs effectués dans la deuxième partie de la démonstration du théorème 1.21) que  $\ell_{i,j} = 0$  si  $j < j_i$ . Donc si on a adopté un stockage "profil" de  $A$ , on peut utiliser le même stockage pour  $L$ .

### Matrices non symétriques

Soit  $A \in \mathcal{M}_n(\mathbb{R})$  inversible. On ne suppose plus ici que  $A$  est symétrique. On cherche à calculer  $x \in \mathbb{R}^n$  solution de (1.1) par la méthode de Choleski. Ceci est possible en remarquant que :  $Ax = b \Leftrightarrow A^t Ax = A^t b$  car  $\det(A) = \det(A^t) \neq 0$ . Il ne reste alors plus qu'à vérifier que  $A^t A$  est symétrique définie positive. Remarquons d'abord que pour toute matrice  $A \in \mathcal{M}_n(\mathbb{R})$ , la matrice  $AA^t$  est symétrique. Pour cela on utilise le fait que si  $B \in \mathcal{M}_n(\mathbb{R})$ , alors  $B$  est symétrique si et seulement si  $Bx \cdot y = x \cdot By$  et  $Bx \cdot y = x \cdot B^t y$  pour tout  $(x, y) \in (\mathbb{R}^n)^2$ . En prenant  $B = A^t A$ , on en déduit que  $A^t A$  est symétrique. De plus, comme  $A$  est inversible,  $A^t Ax \cdot x = Ax \cdot Ax = |Ax|^2 > 0$  si  $x \neq 0$ . La matrice  $A^t A$  est donc bien symétrique définie positive.

La méthode de Choleski dans le cas d'une matrice non symétrique consiste donc à calculer  $A^t A$  et  $A^t b$ , puis à résoudre le système linéaire  $A^t A \cdot x = A^t b$  par la méthode de Choleski "symétrique".

Cette manière de faire est plutôt moins efficace que la décomposition  $LU$  puisque le coût de la décomposition  $LU$  est de  $2n^3/3$  alors que la méthode de Choleski dans le cas d'une matrice non symétrique nécessite au moins  $4n^3/3$  opérations (voir exercice 24).

### Systèmes linéaires non carrés

On considère ici des matrices qui ne sont plus carrées. On désigne par  $\mathcal{M}_{M,n}(\mathbb{R})$  l'ensemble des matrices réelles à  $M$  lignes et  $n$  colonnes. Pour  $A \in \mathcal{M}_{M,n}(\mathbb{R})$ ,  $M > n$  et  $b \in \mathbb{R}^M$ , on cherche  $x \in \mathbb{R}^n$  tel que

$$Ax = b. \tag{1.45}$$

6. On appelle discrétisation le fait de se ramener d'un problème où l'inconnue est une fonction en un problème ayant un nombre fini d'inconnues.

Ce système contient plus d'équations que d'inconnues et n'admet donc en général pas de solution. On cherche  $x \in \mathbb{R}^n$  qui vérifie le système (1.45) "au mieux". On introduit pour cela une fonction  $f$  définie de  $\mathbb{R}^n$  dans  $\mathbb{R}$  par :

$$f(x) = |Ax - b|^2,$$

où  $|x| = \sqrt{x \cdot x}$  désigne la norme euclidienne sur  $\mathbb{R}^n$ . La fonction  $f$  ainsi définie est évidemment positive, et s'il existe  $x$  qui annule  $f$ , alors  $x$  est solution du système (1.45). Comme on l'a dit, un tel  $x$  n'existe pas forcément, et on cherche alors un vecteur  $x$  qui vérifie (1.45) "au mieux", au sens où  $f(x)$  soit le plus proche de 0. On cherche donc  $x \in \mathbb{R}^n$  satisfaisant (1.45) en minimisant  $f$ , c.à.d. en cherchant  $x \in \mathbb{R}^n$  solution du problème d'optimisation :

$$f(x) \leq f(y) \quad \forall y \in \mathbb{R}^n \quad (1.46)$$

On peut réécrire  $f$  sous la forme :  $f(x) = A^t Ax \cdot x - 2b \cdot Ax + b \cdot b$ . On montrera au chapitre III que s'il existe une solution au problème (1.46), elle est donnée par la résolution du système linéaire suivant :

$$AA^t x = A^t b \in \mathbb{R}^n, \quad (1.47)$$

qu'on appelle équations normales du problème de minimisation. La résolution approchée du problème (1.45) par cette procédure est appelée méthode des moindres carrés. La matrice  $AA^t$  étant symétrique, on peut alors employer la méthode de Choleski pour la résolution du système (1.47).

### 1.3.5 Exercices

**Exercice 14** (Vrai ou faux ?). *Corrigé en page 48*

Les propositions suivantes sont-elles vraies ou fausses ?

1. La matrice  $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$  admet une décomposition de Choleski.
2. La matrice  $B = \begin{bmatrix} 1 & -2 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$  est symétrique définie positive.
3. La matrice  $B$  ci-dessus admet une décomposition  $LU$ .
4. La matrice  $\begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix}$  s'écrit  $C^t C$ .
5. La matrice  $A = \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix}$  admet une décomposition de Choleski  $A = C^t C$  avec  $C = \begin{bmatrix} -1 & -1 \\ 0 & -2 \end{bmatrix}$ .
6. Soit  $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$  (a) La matrice  $AA^t$  admet une décomposition de Choleski.  
(b) La matrice  $A^t A$  admet une décomposition de Choleski.

**Exercice 15** (LU). *Corrigé en page 48*

1. Donner la décomposition  $LU$  de la matrice  $A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 2 & 1 & 0 \end{bmatrix}$ .
2. Montrer que la matrice  $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$  vérifie  $PA = LU$  avec  $P$  une matrice de permutation,  $L$  triangulaire inférieure et  $U$  triangulaire supérieure à déterminer.
3. Calculer la décomposition  $LU$  de la matrice  $\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$

**Exercice 16** (Echelonnement et factorisation  $LU$  et  $LDU$ ). *Corrigé en page 49.*

Echelonner les matrices suivantes (c.à.d. appliquer l'algorithme de Gauss), et lorsqu'elle existe, donner leur décomposition  $LU$  et  $LDU$

$$A = \begin{bmatrix} 2 & -1 & 4 & 0 \\ 4 & -1 & 5 & 1 \\ -2 & 2 & -2 & 3 \\ 0 & 3 & -9 & 4 \end{bmatrix}; \quad B = \begin{bmatrix} 1. & 2. & 1. & 2. \\ -1. & -1. & 0. & -2. \\ 1. & 2. & 2. & 3. \\ -1. & -1. & 1. & 0. \end{bmatrix}.$$

**Exercice 17** (Décomposition  $LU$  d'une matrice à paramètres). *Corrigé en page 50.*

Soient  $a, b, c$  et  $d$  des nombres réels. On considère la matrice suivante :

$$A = \begin{bmatrix} a & a & a & a \\ a & b & b & b \\ a & b & c & c \\ a & b & c & d \end{bmatrix}.$$

Appliquer l'algorithme d'élimination de Gauss à  $A$  pour obtenir sa décomposition  $LU$ .

Donner les conditions sur  $a, b, c$  et  $d$  pour que la matrice  $A$  soit inversible.

**Exercice 18** (Décomposition de Choleski d'une matrice particulière). *Corrigé en page 51*

Soit  $n \in \mathbb{N} \setminus \{0\}$ . On considère la matrice  $A_n$  carrée d'ordre  $n$  dont les coefficients sont donnés par  $(A_n)_{i,j} : \min(i, j)$ , et qui s'écrit donc :

$$A_n = \begin{bmatrix} 1 & 1 & \cdots & \cdots & 1 \\ 1 & 2 & \cdots & \cdots & 2 \\ \vdots & \vdots & & & \\ \vdots & \vdots & & n-1 & n-1 \\ 1 & 2 & & n-1 & n \end{bmatrix}$$

1. Écrire et échelonner les matrices  $A_2$  et  $A_3$ . Montrer que  $A_2$  et  $A_3$  sont des matrices symétriques définies positives et donner leur décomposition de Choleski.
2. En déduire la décomposition de Choleski de la matrice  $A_n$ .

**Exercice 19** (Décomposition  $LU$  et mineurs principaux). *Montrer que la matrice  $A$  de coefficients*

$$a_{ij} = \begin{cases} -1 & \text{si } i > j \\ 1 & \text{si } i = j \text{ et } j = n \\ 0 & \text{sinon} \end{cases}$$

admet une décomposition  $LU$  (sans permutation préalable). Calculer les coefficients diagonaux de la matrice  $U$ .

**Exercice 20** (Conservation du profil). *On considère des matrices  $A$  et  $B \in \mathcal{M}_4(\mathbb{R})$  de la forme suivante, où  $x$  en position  $(i, j)$  de la matrice signifie que le coefficient  $a_{i,j}$  est non nul) et 0 en position  $(i, j)$  de la matrice signifie que  $a_{i,j} = 0$ .*

$$A = \begin{bmatrix} x & x & x & x \\ x & x & x & 0 \\ 0 & x & x & 0 \\ 0 & 0. & x. & x. \end{bmatrix} \quad \text{et} \quad B = \begin{bmatrix} x & x & x & 0 \\ x & x & 0 & x \\ 0 & x & x & x \\ 0 & x. & x. & x. \end{bmatrix}.$$

Quels sont les coefficients nuls (notés 0 dans les matrices) qui resteront nuls dans les matrices  $L$  et  $U$  de la factorisation  $LU$  (si elle existe) ?

**Exercice 21** (Matrices définies positives et décomposition  $LU$ ). *On rappelle que les mineurs principaux d'une matrice  $A \in \mathcal{M}_n(\mathbb{R})$ , sont les déterminants  $\Delta_p$  des matrices  $A_p = A(1:p, 1:p)$  extraites de la matrice  $A$ .*

1. Montrer qu'une matrice symétrique définie positive a tous ses mineurs principaux strictement positifs.
2. En déduire que toute matrice symétrique définie positive admet une décomposition LU.

**Exercice 22** (Matrices non inversibles et décomposition LU).

Corrigé en page 51

1. Matrices  $2 \times 2$

(a) Soit  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$  On suppose que  $a_{11} \neq 0$ .

i. Echelonner la matrice  $A$  et en déduire qu'il existe une matrice  $\tilde{L}$  triangulaire inférieure dont les coefficients diagonaux sont égaux à 1, et une matrice  $\tilde{U}$  triangulaire supérieure telles que  $A = \tilde{L}\tilde{U}$ .

ii. Montrer que  $\tilde{L}$  et  $\tilde{U}$  sont uniques.

iii. Donner une condition nécessaire et suffisante sur les coefficients de  $A$  pour que la matrice  $\tilde{U}$  soit inversible.

(b) On pose maintenant  $A = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$ . Trouver deux matrices  $\tilde{L}_1$  et  $\tilde{L}_2$  distinctes, toutes deux triangulaires inférieures et dont les coefficients diagonaux sont égaux à 1, et des matrices  $\tilde{U}_1$  et  $\tilde{U}_2$  triangulaires supérieures avec  $A = \tilde{L}_1\tilde{U}_1 = \tilde{L}_2\tilde{U}_2$ .

2. Matrices  $3 \times 3$

(a) Echelonner la matrice  $A = \begin{bmatrix} 1. & 2. & 3. \\ 5. & 7. & 9. \\ 12. & 15. & 18. \end{bmatrix}$  et en déduire que la matrice  $A$  peut se décomposer en

$A = \tilde{L}\tilde{U}$  où  $\tilde{L}$  est une matrice triangulaire inférieure dont les coefficients diagonaux sont égaux à 1, et  $\tilde{U}$  est une matrice triangulaire supérieure.

(b) Soit  $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ . Montrer que si  $a_{11} \neq 0$  et que la matrice  $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$  est inversible,

alors il existe un unique couple de matrices  $(\tilde{L}, \tilde{U})$  tel que  $A = \tilde{L}\tilde{U}$ , où  $\tilde{L}$  est triangulaire inférieure dont les coefficients diagonaux sont égaux à 1, et une matrice  $\tilde{U}$  triangulaire supérieure.

3. Matrices  $n \times n$ .

(a) Généraliser le résultat de la question précédente à une matrice de dimension  $n$  : donner le résultat espéré sous forme de théorème et le démontrer.

(b) Soit maintenant  $A$  une matrice de dimensions  $n \times n$ . Montrer qu'il existe une matrice de permutation  $P$  et des matrices  $\tilde{L}$  triangulaire inférieure et de coefficients diagonaux égaux à 1, et  $\tilde{U}$  triangulaire supérieure, telles que  $PA = LU$ . (On pourra commencer par le cas où est de rang égal à  $n - 1$ .)

**Exercice 23** (Décomposition  $LL^t$  "pratique"). Corrigé en page 53.

1. Soit  $A$  une matrice symétrique définie positive. Montrer que la décomposition de Choleski  $\tilde{L}\tilde{L}^t$  de la matrice  $A$  est obtenue à partir de sa décomposition LU en posant  $\tilde{L} = L\sqrt{D}$  où  $D$  est la matrice diagonale extraite de  $U$ . (Voir remarque 1.22.)

En déduire la décomposition  $LL^t$  de la matrice particulière  $A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$ .

2. Que deviennent les coefficients nuls dans la décomposition  $LL^t$  ci-dessus ? Quelle est la propriété vue en cours qui est ainsi vérifiée ?

**Exercice 24** (Sur la méthode  $LL^t$ ). Corrigé détaillé en page 55.

Soit  $A$  une matrice carrée d'ordre  $n$ , symétrique définie positive et pleine. On cherche à résoudre le système  $A^2x = b$ .

On propose deux méthodes de résolution de ce système :

1. Calculer  $A^2$ , effectuer la décomposition  $LL^t$  de  $A^2$ , résoudre le système  $LL^t x = b$ .
2. Calculer la décomposition  $LL^t$  de  $A$ , résoudre les systèmes  $LL^t y = b$  et  $LL^t x = y$ .

Calculer le nombre d'opérations élémentaires nécessaires pour chacune des deux méthodes et comparer.

**Exercice 25** (Décomposition  $LDL^t$  et  $LL^t$ ). Corrigé en page 56

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

Calculer la décomposition  $LDL^t$  de  $A$ . Existe-t-il une décomposition  $LL^t$  de  $A$  ?

2. Montrer que toute matrice de  $\mathcal{M}_n(\mathbb{R})$  symétrique définie positive admet une décomposition  $LDL^t$ .
3. Ecrire l'algorithme de décomposition  $LDL^t$ . La matrice  $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  admet-elle une décomposition  $LDL^t$  ?

**Exercice 26** (Décomposition  $LL^t$  d'une matrice tridiagonale symétrique). Corrigé en page 58

Soit  $A \in \mathcal{M}_n(\mathbb{R})$  symétrique définie positive et tridiagonale (i.e.  $a_{i,j} = 0$  si  $i - j > 1$ ).

1. Montrer que  $A$  admet une décomposition  $LL^t$ , où  $L$  est de la forme

$$L = \begin{pmatrix} \alpha_1 & 0 & \dots & & 0 \\ \beta_2 & \alpha_2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \dots & 0 \\ \vdots & \ddots & \ddots & \dots & \vdots \\ 0 & \dots & 0 & \beta_n & \alpha_n \end{pmatrix}.$$

2. Donner un algorithme de calcul des coefficients  $\alpha_i$  et  $\beta_i$ , en fonction des coefficients  $a_{i,j}$ , et calculer le nombre d'opérations élémentaires nécessaires dans ce cas.
3. En déduire la décomposition  $LL^t$  de la matrice :

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

4. L'inverse d'une matrice inversible tridiagonale est elle tridiagonale ?

**Exercice 27** (Choleski pour matrice bande). Suggestions en page 48, corrigé en page 59

Soit  $A \in \mathcal{M}_n(\mathbb{R})$  une matrice symétrique définie positive.

1. On suppose ici que  $A$  est tridiagonale. Estimer le nombre d'opérations de la factorisation  $LL^t$  dans ce cas.
2. Même question si  $A$  est une matrice bande (c'est-à-dire  $p$  diagonales non nulles).
3. En déduire une estimation du nombre d'opérations nécessaires pour la discrétisation de l'équation  $-u'' = f$  vue page 11. Même question pour la discrétisation de l'équation  $-\Delta u = f$  présentée page 13.