

Chrysostome Baskiotis

Laurence Lamoulié

Fascicule 3

**ANALYSE NUMÉRIQUE**

# Valeurs et vecteurs propres



Année 2009 – 2010



# 1

## VALEURS ET VECTEURS PROPRES

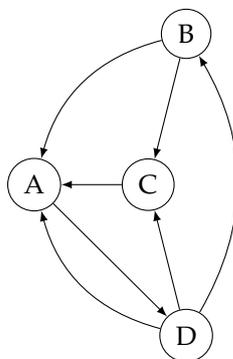
---

1.1	Introduction . . . . .	2
1.1.1	Exercices . . . . .	5
1.2	Calcul des valeurs et vecteurs propres . . . . .	5
1.2.1	Calcul par factorisation $\mathbf{LU}$ . . . . .	8
1.3	Localisation des valeurs propres . . . . .	10
1.3.1	Exercices . . . . .	11
1.4	Estimation des erreurs en cas des perturbations . . . . .	12
1.5	Références . . . . .	14

---

Le calcul des valeurs et vecteurs propres est un domaine de mathématiques d'une très grande importance théorique. En même temps ses applications pratiques sont aussi, et même plus, importantes. Citons par exemple, les systèmes dynamiques, le calcul des structures, l'analyse de données, la cryptologie, le traitement d'images, l'évaluation de la pertinence des pages web, etc.

À titre d'exemple, nous présentons ci-après de façon simplifiée, l'algorithme *pagerank* de la pondération des pages web selon leur pertinence, utilisé par le moteur Google. La pertinence d'une page est fondée sur le nombre de pages web qui ont un lien vers cette page, pondéré par le nombre de liens qu'elle a chacune de ces pages. Considérons quatre pages web, A, B, C et D représentées par le graphe suivant



où les fleches indiquent les liens. Par exemple la fleche de la page B vers la page A signifie que la page B a un lien vers la page A. Cette liaison sera pondérée par le nombre de liens vers d'autres

pages qui sont présents dans la page B, à savoir 2. Donc la case (A,B) de la matrice des liaisons aura la valeur 1/2. Nous avons ainsi la matrice

$$\mathbf{H} = \begin{bmatrix} 0 & 1/2 & 1 & 1/3 \\ 0 & 0 & 0 & 1/3 \\ 0 & 1/2 & 0 & 1/3 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

La pertinence de quatre sites est fournie par le vecteur propre  $\mathbf{r}$  de la matrice  $\mathbf{H}$  qui correspond à la valeur propre 1. On a donc

$$\mathbf{H}\mathbf{r} = \mathbf{r}$$

En calculant on trouve  $\mathbf{r} = [0.6507914, 0.2169305, 0.3253957, 0.6507914]^T$  comme pertinence pour les quatre sites respectivement. Ainsi la pertinence des pages A et D est la même car la page A est référencée par les trois autres et la page D est référencée par A dont elle est la seule et unique référence. La page C est référencée par deux pages différentes, elle a donc un meilleur rang que la page B qui elle référencée par une page seulement.

L'objectif de ce chapitre est de présenter de façon succincte les aspects théoriques des valeurs propres. En ce qui concerne les méthodes numériques du calcul des valeurs et vecteurs propres on présente l'esquisse d'une méthode en renvoyant, pour des méthodes plus complexes, le lecteur intéressé à la bibliographie.

Ce partie pris d'occulter les méthodes de résolution est dû au fait que nous voulons, dans un temps court, présenter deux autres sujets intimement reliés avec les valeurs propres et qui sont d'une importance considérable pour les applications. Il s'agit de la décomposition d'une matrice en valeurs singulières et de la pseudo-inverse d'une matrice et qui sont utilisées abondamment en analyse de données, en automatique et en traitement d'images. En particulier la notion de la pseudo-inverse permet d'obtenir la solution du problème des moindres carrés et de l'analyse de données<sup>1</sup>. La notion de pseudo-inverse est aussi un outil puissant pour la modélisation des mémoires hétéro-associatives en réseaux de neurones et traitement d'images<sup>2</sup>.

Il va de soi que pour la présentation des différentes méthodes, et dans la mesure où on veut établir des programmes sur ordinateur, on utilisera le langage du calcul matriciel. Mais afin de donner une approche plus intuitive des concepts et des méthodes utilisés nous ferons aussi une présentation géométrique

## 1.1 Introduction

Soit  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$  et un nombre  $\lambda \in \mathbb{R}$ . Le *problème algébrique des valeurs propres* consiste en la solution du système de  $n + 1$  équation linéaires

$$(\mathbf{A} - \lambda\mathbf{B})\mathbf{u} = \mathbf{0} \quad \Rightarrow \quad \mathbf{A}\mathbf{u} = \lambda\mathbf{B}\mathbf{u} \quad (1.1.1)$$

1. Voir cours de statistique 2eme année et cours de l'analyse de données en filière TSI pour une description de ces méthodes.

2. voir les cours correspondants en 2eme année tronc commun et filière TSI



l'aide des relations

$$\mathbf{u}' = (\mathbf{A}^\top)^{-1} \mathbf{u} \quad (1.1.8)$$

et

$$\mathbf{u} = \mathbf{A}^\top \mathbf{u}' \quad (1.1.9)$$

Supposons aussi que dans l'espace vectoriel nous avons deux bases  $(\mathbf{f}_1, \dots, \mathbf{f}_n)$  et  $(\mathbf{f}'_1, \dots, \mathbf{f}'_n)$  et soit  $A$  l'application linéaire de  $E$  vers  $F$  avec comme matrice associée  $\mathbf{A}$ . Si on note cette application  $B$  on a par rapport aux nouvelles bases :

$$\mathbf{B} = \mathbf{T}^{-1} \mathbf{A} \mathbf{S} \quad \Rightarrow \quad \mathbf{A} = \mathbf{T} \mathbf{B} \mathbf{S}^{-1} \quad (1.1.10)$$

où  $\mathbf{T}$  est la matrice de changement de base dans  $F$  et  $\mathbf{B}$  la matrice associée à l'application  $B$ . Dans le cas où  $E = F$ , on a

$$\mathbf{B} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S} \quad \Rightarrow \quad \mathbf{A} = \mathbf{S} \mathbf{B} \mathbf{S}^{-1} \quad (1.1.11)$$

Deux matrices  $\mathbf{A}, \mathbf{B}$  qui sont liées par de telles relations sont appelées *semblables*. On conclut donc que les matrices, par rapport à différentes bases, associées à une application linéaire sont semblables.

**Projections.** Considérons un espace vectoriel  $E$  qui se décompose en somme directe

$$E = F \oplus G \quad (1.1.12)$$

où  $F, G$  sont deux sous-espaces vectoriels de  $E$ . La projection linéaire de  $E$  sur  $F$  est une application linéaire  $p : E \rightarrow E$  définie par la relation

$$p(\mathbf{y} + \mathbf{z}) = \mathbf{y}; \quad \forall \mathbf{y} \in F \text{ et } \forall \mathbf{z} \in G \quad (1.1.13)$$

Remarquons que nous pouvons avoir plusieurs projections possibles pour le même résultat.

Avant de terminer cette section nous donnons une liste avec les principales propriétés des valeurs et vecteurs propres.

## PROPRIÉTÉS DES VALEURS ET VECTEURS PROPRES

**Soit  $\mathbf{A}$  matrice carrée d'ordre  $n$ . On note par  $\lambda$  une valeur propre de  $\mathbf{A}$  et  $\mathbf{u}$  le vecteur propre associé. On a**

- (1)  $\lambda^k$  est valeur propre de  $\mathbf{A}^k$  avec vecteur propre  $\mathbf{u}$ .**
- (2)  $1/\lambda$  est valeur propre de  $\mathbf{A}^{-1}$  avec vecteur propre  $\mathbf{u}$ .**
- (3)  $\lambda$  est valeur propre de  $\mathbf{A}'$ .**
- (4)  $\lambda + r$  est valeur propre de  $\mathbf{A} + r\mathbf{I}_n$ .**
- (5)  $c\lambda$  est valeur propre de  $c\mathbf{A}$  avec vecteur propre  $\mathbf{u}$ .**
- (6) Si  $\lambda_1, \dots, \lambda_n$  sont les valeurs propres de la matrice  $\mathbf{A}$ , nous avons la relation  $\text{trace}(\mathbf{A}) = \lambda_1 + \dots + \lambda_n$ .**
- (7) Si  $\mathbf{u}$  est un vecteur propre de  $\mathbf{A}$ , alors le vecteur propre  $\mathbf{u}' = \frac{1}{\|\mathbf{u}\|}\mathbf{u}$  l'est aussi. De plus  $\mathbf{u}'$  est unitaire, c'est-à-dire  $\|\mathbf{u}'\| = 1$ .**
- (8) Si  $\mathbf{B}$  matrice régulière,  $\lambda$  est valeur propre de  $\mathbf{B}^{-1}\mathbf{A}\mathbf{B}$ .**
- (9) Si  $\mathbf{B}$  matrice orthogonale ou unitaire,  $\lambda$  est valeur propre de  $\mathbf{B}^\top\mathbf{A}\mathbf{B}$ .**
- (10) Si  $\mathbf{A}$  matrice symétrique, définie positive (resp. définie semi-positive), toutes les valeurs propres sont réelles, positives (resp. positives ou nulles).**
- (11) Si  $\mathbf{A}$  est une matrice triangulaire, alors les éléments diagonaux  $a_{ii}$  sont ses valeurs propres.**

## 1.1.1 Exercices

EXERCICE 1.1 Soit  $\mathbf{A} \in \mathbb{R}^{n \times n}$  matrice régulière et soient  $\lambda$  valeur propre associée au vecteur propre  $\mathbf{u}$ . Montrer que  $1/\lambda$  est valeur propre de  $\mathbf{A}^{-1}$  associée au même vecteur propre  $\mathbf{u}$ .

EXERCICE 1.2 Soit  $\mathbf{A} \in \mathbb{R}^{n \times n}$  matrice régulière et soient  $\lambda$  valeur propre associée au vecteur propre  $\mathbf{u}$ . Montrer que  $\lambda^k$  est valeur propre de  $\mathbf{A}^k$  associée au même vecteur propre  $\mathbf{u}$ ,  $k > 1$ .

EXERCICE 1.3 Soit  $\mathbf{A} \in \mathbb{R}^{n \times n}$  matrice régulière.  $\mathbf{u}$  est un vecteur propre de  $\mathbf{A}$  si et seulement si  $\mathbf{u}$  est un vecteur propre de  $(\mathbf{A} - \alpha\mathbf{I})^{-1}$ .

EXERCICE 1.4 Si  $\mathbf{A} \in \mathbb{R}^{n \times n}$  est nilpotente (c'est-à-dire il existe  $k > 0$  tel que  $\mathbf{A}^k = \mathbf{0}$ ), alors  $\text{trace}(\mathbf{A}) = 0$ .

## 1.2 Calcul des valeurs et vecteurs propres

Rappelons que le problème des valeurs propres est le suivant :

$$\text{Trouver } \lambda \in \mathbb{R} \text{ tel que } \mathbf{A}\mathbf{u} = \lambda\mathbf{u} \text{ pour un } \mathbf{u} \in \mathbb{R}^n, \mathbf{u} \neq \mathbf{0} \quad (1.2.1)$$

On appelle  $\lambda$  valeur propre de la matrice  $\mathbf{A}$  et  $\mathbf{u}$  vecteur propre de  $\mathbf{A}$  associé à  $\lambda$ . (1.2.1) est un système de  $n$  équations linéaires avec  $n + 1$  inconnues : les  $n$  composantes du vecteur  $\mathbf{u}$  et  $\lambda$ . Bien

sûr dans ce cas le vecteur  $\mathbf{u}$  n'est pas déterminé de façon unique. Si on impose une condition supplémentaire pour le vecteur  $\mathbf{u}$ , à savoir

$$\mathbf{u}^\top \mathbf{u} = 1 \quad (1.2.2)$$

alors nous avons un système avec  $n + 1$  équations mais qu'il n'est plus linéaire.

Nous pouvons facilement comprendre que, dans le cas des vecteurs propres de la matrice  $\mathbf{A}$ , leur multiplication par  $\mathbf{A}$  se réduit à une opération scalaire le long du vecteur propre.

Nous consacrons le reste de cette section à la recherche d'une méthode efficace pour le calcul des valeurs et vecteurs propres.

L'ensemble de toutes les valeurs propres de  $\mathbf{A}$  forme *le spectre* de  $\mathbf{A}$ , noté  $\sigma(\mathbf{A})$ . L'ensemble  $\mathbf{A}(\lambda) = \{\mathbf{u} \in \mathbb{R}^n / \mathbf{u} \neq \mathbf{0}, (\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0}\}$  forme un sous-espace de  $\mathbb{R}^n$  de dimension  $\delta(\lambda) = n - \text{rang}(\mathbf{A} - \lambda\mathbf{I})$ .  $\mathbf{A}(\lambda)$  est formé pour chaque valeur propre  $\lambda$  de  $\mathbf{A}$ . Donc  $\lambda$  est une valeur propre de  $\mathbf{A}$  si  $\mathbf{A}(\lambda) \neq \emptyset$  c'est-à-dire si  $\delta(\lambda) = n$  et  $\mathbf{A} - \lambda\mathbf{I}$  est singulière ou, encore,  $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ . On pose  $P_{\mathbf{A}} = \det(\mathbf{A} - \lambda\mathbf{I})$ .  $P_{\mathbf{A}}$  est le *polynôme caractéristique* de  $\mathbf{A}$ . Ses racines sont les valeurs propres de  $\mathbf{A}$ .

Considérons maintenant une matrice  $\mathbf{U} \in \mathbb{R}^{n \times n}$  régulière et posons  $\mathbf{v} = \mathbf{U}^{-1}\mathbf{u}$ . Alors de (1.2.1) nous avons  $\mathbf{A}\mathbf{U}\mathbf{v} = \lambda\mathbf{U}\mathbf{v}$ , d'où

$$\mathbf{U}^{-1}\mathbf{A}\mathbf{U}\mathbf{v} = \lambda\mathbf{v}, \mathbf{v} \neq \mathbf{0} \quad (1.2.3)$$

Donc  $\mathbf{v}$  est un vecteur propre de la matrice  $\mathbf{B} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$  associé à la valeur propre  $\lambda$ . Les matrices  $\mathbf{A}$  et  $\mathbf{B}$  sont dites *semblables* ou *similaires* et cette relation sera notée par  $\mathbf{A} \sim \mathbf{B}$ . La transformation  $\mathbf{A} \mapsto \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$  est appelée *transformation de similitude*.

On voit ainsi que les matrices semblables ont les mêmes valeurs propres. Il y a deux conséquences importantes de ce fait :

- Les transformation de similitude d'une matrice laissent son déterminant inchangé, car le déterminant d'une matrice est égal au produit de ses valeurs propres.
- Les transformation de similitude d'une matrice laissent sa trace inchangée, car la trace d'une matrice est égale à la somme de ses valeurs propres.

Étant donné qu'à la matrice  $\mathbf{A} \in \mathbb{R}^{n \times n}$  est associée une application linéaire  $A$  de  $\mathbb{R}^n$  dans  $\mathbb{R}^n$ , nous avons que l'ensemble  $\mathcal{S}(\mathbf{A}) = \{\mathbf{U}^{-1}\mathbf{A}\mathbf{U} / \mathbf{U} \in \mathbb{R}^{n \times n}; \mathbf{U} \text{ régulière}\}$  contient toutes les matrices qui représentent cette application et il est, en même temps, l'ensemble de toutes les matrices qui sont semblables à  $\mathbf{A}$ . On peut, ainsi, voir que les matrices semblables forment une relation d'équivalence dans l'espace des applications linéaires de  $\mathbb{R}^n$  dans  $\mathbb{R}^n$  qui est le même que l'espace des matrices carrées ( $n \times n$ ). Le théorème suivant précise les éléments d'une matrice  $\mathbf{A}$  qui sont invariants pour toutes les matrices de sa classe d'équivalence  $\mathcal{S}(\mathbf{A})$ .

**THÉORÈME 1.2.1** *Soit  $\mathbf{A} \in \mathbb{R}^{n \times n}$  une matrice carrée. Toute matrice  $\mathbf{B} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U} \in \mathcal{S}(\mathbf{A})$  a le même polynôme caractéristique que  $\mathbf{A}$ , et donc aussi les mêmes valeurs propres et le même déterminant.  
Si  $\mathbf{u}$  est un vecteur propre de  $\mathbf{A}$  correspondant à une valeur propre  $\lambda$ , alors  $\mathbf{v} = \mathbf{U}^{-1}\mathbf{u}$  est le vecteur propre de  $\mathbf{B}$  qui correspond à la même valeur propre.*

Parmi les matrices semblables d'une matrice  $\mathbf{A}$ , les matrices diagonales sont très importantes pour les applications. Si une matrice  $\mathbf{A}$  est semblable à une matrice diagonale, on dit que  $\mathbf{A}$  est *diagonalisable* et les théorèmes suivants fournissent les conditions à cet effet.

**THÉORÈME 1.2.2** Une matrice  $\mathbf{A} \in \mathbb{R}^{n \times n}$  est diagonalisable s'il existe un ensemble de  $n$  vecteurs  $\mathbf{u}_1, \dots, \mathbf{u}_n$  linéairement indépendants et qui sont des vecteurs propres de  $\mathbf{A}$ .

**THÉORÈME 1.2.3** Si une matrice  $\mathbf{A} \in \mathbb{R}^{n \times n}$  a toutes ses valeurs propres distinctes, alors elle est diagonalisable.

Dans le cas des matrices diagonalisables nous avons la relation

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1} \quad \Rightarrow \quad \mathbf{\Lambda} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U} \quad (1.2.4)$$

où  $\mathbf{U}$  la matrice des vecteurs propres de  $\mathbf{A}$  et  $\mathbf{\Lambda}$  matrice diagonale avec comme éléments diagonaux les valeurs propres de  $\mathbf{A}$ . Notons que puisque les vecteurs propres sont linéairement indépendants, la matrice  $\mathbf{U}$  est régulière.

L'ennui avec les théorèmes mathématiques est que nous donnent des informations très intéressantes, mais très souvent ne sont pas constructifs. C'est le cas avec les deux théorèmes ci-dessus. Ils affirment qu'une matrice régulière est diagonalisable à l'aide d'une transformation de similitude mais ils ne disent rien sur la manière de construire cette transformation et nous sommes donc obligés à la rechercher. Ici la première étape de recherche passe par la découverte d'une forme pour la matrice  $\mathbf{U}$  qui rendra la transformation de similitude "sympathique", c'est-à-dire telle que l'inversion de  $\mathbf{U}$  soit facile.

Nous pouvons, par exemple, envisager que  $\mathbf{U}$  soit une matrice orthogonale<sup>3</sup>, c'est-à-dire que  $\mathbf{U}\mathbf{U}^\top = \mathbf{I}$ . Dans ce cas, on a pour la transformation de similarité  $\mathbf{A} \mapsto \mathbf{B} = \mathbf{U}^\top \mathbf{A}\mathbf{U}$  et on évite ainsi l'inversion de  $\mathbf{U}$ . Une telle transformation s'appelle *transformation unitaire de similitude* et les matrices  $\mathbf{A}$  et  $\mathbf{B}$  sont *orthogonalement similaires*. Dans ce cas, nous avons les deux propriétés suivantes ~ :

- $\kappa_2(\mathbf{A}) = \kappa_2(\mathbf{B})$  et  $\|\mathbf{A}\|_2 = \|\mathbf{B}\|_2$ .
- Si  $\mathbf{B} + \mathbf{\Delta B} = \mathbf{U}^\top (\mathbf{A} + \mathbf{\Delta A}) \mathbf{U}$ , alors  $\|\mathbf{\Delta B}\|_2 = \|\mathbf{\Delta A}\|_2$ .

La deuxième étape de recherche consiste à trouver parmi la myriade de matrices orthogonales, une qui convient au mieux, c'est-à-dire qui fournit une solution unique. Dans des problèmes de cette nature, il faut réfléchir sur le nombre de degrés de liberté que le système possède et, si on veut que la solution soit unique, il faut faire en sorte que ce système n'en possède aucun.

Une matrice  $\mathbf{U} \in \mathbb{R}^{n \times n}$  orthogonale a  $n(n-1)/2$  conditions d'orthogonalité et  $n$  conditions de normalité. Au total elle a  $n(n+1)/2$  conditions et donc  $n^2 - n(n+1)/2 = n(n-1)/2$  degrés de liberté. Par conséquent pour annuler ces degrés de libertés, il faut envisager une matrice triangulaire qui a justement  $n(n-1)/2$  éléments nuls.

Conclusion : on cherche pour  $\mathbf{A}$  une transformation de similitude avec matrice  $\mathbf{U}$  orthogonale et triangulaire.

Nous présentons maintenant le théorème de Schur qui établit que chaque matrice carrée est orthogonalement similaire à une matrice triangulaire supérieure.

**THÉORÈME 1.2.4 (THÉORÈME DE SCHUR)** Soit une matrice  $\mathbf{A} \in \mathbb{R}^{n \times n}$ . Il existe une matrice unitaire  $\mathbf{U} \in \mathbb{C}^{n \times n}$  et une matrice triangulaire supérieure  $\mathbf{T} \in \mathbb{R}^{n \times n}$  telles que  $\mathbf{T} = \mathbf{U}^* \mathbf{A}\mathbf{U}$ .

3. ou matrice unitaire si  $\mathbf{U}$  est une matrice avec nombres complexes comme éléments.

D'après ce théorème, nous pouvons aussi écrire  $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$  qui est la *décomposition de Schur* de la matrice  $\mathbf{A}$ . Nous pouvons prouver que les éléments diagonaux de la matrice  $\mathbf{T}$  sont les valeurs propres de  $\mathbf{A}$  et que la première colonne de  $\mathbf{U}$  est le vecteur propre associé à la valeur propre qui est le premier élément diagonal de  $\mathbf{T}$ . Les autres vecteurs propres peuvent être calculés facilement. Si toutes les valeurs propres ne sont pas distinctes, la décomposition de Schur n'est pas unique.

Bien évidemment, le théorème de Schur n'est pas non plus constructif. Nous présentons à la section suivante, une méthode qui, sous des hypothèses particulières, permet la décomposition de Schur ;

Nous avons un cas particulier de transformation unitaire de similitude si  $\mathbf{A}$  est une matrice normale, c'est-à-dire telle que  $\mathbf{A}^\top \mathbf{A} = \mathbf{A}\mathbf{A}^\top$  qui est reflétée par le théorème suivant :

**THÉORÈME 1.2.5** Soit  $\mathbf{A} \mapsto \mathbf{B} = \mathbf{U}^\top \mathbf{A}\mathbf{U}$  une transformation unitaire de similitude et  $\mathbf{A}$  matrice normale. Alors

$$\mathbf{U}^* \mathbf{A}\mathbf{U} = \mathbf{\Lambda} = \text{diag} (\lambda_1, \dots, \lambda_n)$$

où  $\lambda_i$  les valeurs propres de  $\mathbf{A}$ .

Un cas particulier de matrice normale, est la matrice symétrique. Nous avons, d'après ce théorème, que si  $\mathbf{A}$  est une matrice symétrique, la décomposition de Schur donne :  $\mathbf{A} = \mathbf{U}^\top \mathbf{\Lambda}\mathbf{U}$ . Ici les colonnes de  $\mathbf{U}$  sont les vecteurs propres correspondants. Dans ce cas particulier, la décomposition de Schur s'appelle *décomposition spectrale* de  $\mathbf{A}$ .

Nous donnons en annexe, à la fin du chapitre les principales propriétés des valeurs et vecteurs propres d'une matrice.

Pour calculer les valeurs et vecteurs propres il y a plusieurs méthodes qui sont soit générales, c'est-à-dire applicables sur n'importe quelle forme de matrice carrée, soit particulières, c'est-à-dire tirant profit de la forme spécifique de la matrice (symétrique, définie positive, orthogonale, ...). Nous en mentionnons ici deux afin d'inviter l'utilisateur à les éviter. Il s'agit, d'une part, de la méthode de la puissance, qui est une méthode itérative et qui produit une suite de scalaires et des vecteurs, susceptible de converger, quand la méthode marche, vers une valeur propre et le vecteur propre associé. En règle générale, la précision numérique n'est pas acceptable. D'autre part, nous pouvons calculer les valeurs propres en résolvant le polynôme caractéristique de la matrice. Cette méthode ne fournit pas les vecteurs propres et, de surcroît, la précision numérique n'est pas bonne.

### 1.2.1 Calcul par factorisation LU

La méthode de factorisation LU peut être utilisée pour calculer toutes les valeurs propres d'une matrice  $\mathbf{A}$ . On effectue une suite de décompositions LU afin de trouver une matrice triangulaire supérieurs semblable à la matrice  $\mathbf{A}$ . Les éléments diagonaux de la matrice triangulaire sont les valeurs propres de  $\mathbf{A}$  et les vecteurs propres peuvent être ensuite calculés facilement. Cette méthode proposée par Rutishauser en 1958, a été ensuite améliorée par Francis qui a utilisé la transformation de Hessenberg. Nous donnons ci-après l'algorithme LU pour le calcul des valeurs propres d'une matrice  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , en faisant deux hypothèses :

- La matrice  $\mathbf{A}$  est définie positive.
- Les valeurs propres de la matrice  $\mathbf{A}$  sont toutes distinctes les unes des autres et telles que

$$|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| \quad (1.2.5)$$

L'algorithme est le suivant :

**ALGORITHME DE FACTORISATION LU**

1° :  $\mathbf{A}_1 = \mathbf{A}$

2° : Pour  $k = 1, \dots$

- décomposition LU de la matrice  $\mathbf{A}_k$  :  $\mathbf{A}_k = \mathbf{L}_k \mathbf{U}_k$  ; où  $\mathbf{L}_k$  est une matrice triangulaire inférieure, unitaire et  $\mathbf{U}_k$  est une matrice triangulaire supérieure.
- $\mathbf{A}_{k+1} = \mathbf{R}_k \mathbf{U}_k$ .

3° :  $\lim_{k \rightarrow \infty} \mathbf{A}_k = \begin{bmatrix} \lambda_1 & \dots & * \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}$

4° : On suppose qu'on a atteint la limite si les éléments diagonaux des matrices  $\mathbf{A}_k$  – qui sont les valeurs propres que l'on cherche – ne varient pas sensiblement au cours des itérations.

Pour le calcul des valeurs propres on observe que, puisque  $\mathbf{L}_k$  doit être unitaire, nous avons  $\lim_{k \rightarrow \infty} \mathbf{L}_k = \mathbf{I}_n$ . Donc  $\lim_{k \rightarrow \infty} \mathbf{A}_k = \mathbf{U}_k = \mathbf{U}$ . D'autre part nous avons

$$\mathbf{A}_{k+1} = (\mathbf{L}_1 \cdots \mathbf{L}_k)^{-1} \mathbf{A} (\mathbf{L}_1 \cdots \mathbf{L}_k) = (\mathbf{U}_1 \cdots \mathbf{U}_k) \mathbf{A} (\mathbf{U}_1 \cdots \mathbf{U}_k)^{-1} \quad (1.2.6)$$

c'est-à-dire les matrices  $\mathbf{A}$  et  $\mathbf{A}_k$  sont similaires. Donc si  $\mathbf{U}\mathbf{v} = \lambda\mathbf{v}$ , alors  $\mathbf{A}(\mathbf{L}_1 \cdots \mathbf{L}_k)\mathbf{v} = \lambda(\mathbf{L}_1 \cdots \mathbf{L}_k)\mathbf{v}$ .

Par conséquent les vecteurs propres de  $\mathbf{U}$  sont aussi vecteurs propres de  $\mathbf{A}$ . À cause de la forme

triangulaire supérieure de  $\mathbf{U}$  :  $\mathbf{U} = \begin{bmatrix} \lambda_1 & u_{12} & \dots & u_{1n} \\ 0 & \lambda_2 & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ , les vecteurs propres peuvent être calculés

par substitution inverse, en commençant par le vecteur propre  $\mathbf{v}_1^\top = [1, 0, \dots, 0]$ , ce qui permet d'obtenir  $\mathbf{v}_2^\top = \left[ \frac{r_{12}}{\lambda_2 - \lambda_1}, 1, 0, \dots, 0 \right]$ ,  $\mathbf{v}_3^\top = \left[ \left( \frac{r_{23} \cdot r_{12}}{\lambda_3 - \lambda_2} + r_{13} \right) \frac{1}{\lambda_2 - \lambda_1}, 1, 0, \dots, 0 \right]$ , etc.

Notons que si lors de factorisations successives LU on tombe sur un pivot nul, on modifie la matrice en rajoutant une matrice diagonale dont tous les éléments diagonaux sont égaux à un nombre réel quelconque. On peut, par exemple, rajouter la matrice identité. Sachant que les valeurs propres de  $\mathbf{A} + \mathbf{I}$  sont  $\lambda_i + 1$ , il suffit, quand on a convergé, de soustraire de  $\mathbf{A}_k$  autant de fois la matrice identité que l'on a rajoutée lors du déroulement de l'algorithme.

La vitesse de convergence vers une matrice triangulaire supérieure dépend de la vitesse avec laquelle les termes sous la diagonale s'annulent. Cette vitesse est proportionnelle au rapport des valeurs propres correspondantes. Par exemple pour la matrice  $\mathbf{A}_k$  le terme  $a_{ij}$  avec  $i > j$  tend vers zéro comme le terme  $\left| \frac{\lambda_i}{\lambda_j} \right|^k$  étant bien entendu que  $|\lambda_i| < |\lambda_j|$  d'après (1.2.5).

### 1.3 Localisation des valeurs propres

Nous présentons ci-après une méthode de localisation des valeurs propres d'une matrice fondée sur les disques de Geršgorin dont voici la définition :

**DÉFINITION 1.3.1** Soit  $\mathbf{A}$  une matrice carrée d'ordre  $n$ . On associe à chaque ligne  $i$  de la matrice le disque fermé, appelé disque (de ligne) de Geršgorin,  $D_i$  dont le centre est l'élément diagonal  $a_{ii}$  et le rayon la somme des modules des éléments non diagonaux de la ligne :

$$D_i^{(l)} = \left\{ z \in \mathbb{C} \mid |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \right\}$$

La quantité  $R_i^{(l)} = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$  est le rayon du disque  $D_i^{(l)}$

Nous pouvons appliquer la définition des disques de Geršgorin à la transposée de la matrice  $\mathbf{A}$ . On obtient ainsi les disques (des colonnes) de Geršgorin

$$D_j^{(c)} = \left\{ z \in \mathbb{C} \mid |z - a_{jj}| \leq \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}| \right\}$$

qui sont relatifs aux colonnes de la matrice.

Nous avons le théorème suivant :

**THÉORÈME 1.3.1 (GERŠGORIN)** Soit  $\mathbf{A}$  une matrice carrée d'ordre  $n$ .

(1) Toute valeur propre  $\lambda$  de la matrice  $\mathbf{A}$  appartient à l'un au moins de disques (des lignes) de Geršgorin, c'est-à-dire

$$\forall \lambda \in \sigma(\mathbf{A}) \exists i = 1, \dots, n : \lambda \in D_i^{(l)}$$

où  $\sigma(\mathbf{A})$  le spectre de la matrice  $\mathbf{A}$ .

(2) Toute valeur propre  $\lambda$  de la matrice  $\mathbf{A}$  appartient à l'un au moins de disques (des colonnes) de Geršgorin, c'est-à-dire

$$\forall \lambda \in \sigma(\mathbf{A}) \exists j = 1, \dots, n : \lambda \in D_j^{(c)}$$

De ce théorème on peut induire que

$$\sigma(\mathbf{A}) \subset \bigcup_{i=1}^n D_i^{(l)}$$

Nous avons la proposition suivante :

**PROPRIÉTÉ 1.3.1** Toutes les valeurs propres d'une matrice  $\mathbf{A} \in \mathbb{C}^{n \times n}$  appartiennent à la région du plan complexe définie par l'intersection des deux régions constituées respectivement de la réunion des disques des lignes et des colonnes.

Si de plus  $m$  disques des lignes (ou des colonnes),  $1 \leq m \leq n$ , sont disjoints de la réunion des  $n - m$  autres disques, alors leur réunion contient exactement  $m$  valeurs propres (où pour le calcul de  $m$  il faut tenir compte de la multiplicité des valeurs propres et des disques).

Cette proposition montre que si les disques de Geršgorin sont tous disjoints, alors chacun contient exactement une valeur propre.

Concernant les valeurs propres complexes nous avons que si  $\lambda$  est une valeur propre complexe, son complexe conjugué l'est aussi. Noter que dans cette situation, les valeurs propres conjuguées appartiennent au même disque de Geršgorin.

### 1.3.1 Exercices

EXERCICE 1.5 Soit  $\mathbf{A}$  une matrice symétrique et  $\mathbf{U}$  matrice unitaire telle que

$$\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \text{diag}(\lambda_1, \dots, \lambda_n)$$

Supposons que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ . Montrer que

$$\lambda_1 = \max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \mathbf{x} \neq \mathbf{0}}} \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}; \lambda_n = \min_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \mathbf{x} \neq \mathbf{0}}} \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$$

De plus, si  $\mathbf{A}$  est définie positive (c'est-à-dire  $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0 \forall \mathbf{x} \in \mathbb{R}^n$ ), alors montrer que toutes ses valeurs propres sont positives.

EXERCICE 1.6 En utilisant les disques de Geršgorin, donner une estimation du nombre maximal de valeurs propres complexes des matrices suivantes :

$$\mathbf{A} = \begin{bmatrix} 2 & -1/2 & 0 & -1/2 \\ 0 & 4 & 0 & 2 \\ -1/2 & 0 & 6 & 1/2 \\ 0 & 0 & 1 & 9 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} -5 & 0 & 1/2 & 1/2 \\ 1/2 & 2 & 1/2 & 0 \\ 0 & 1 & 0 & 1/2 \\ 0 & 1/4 & 1/2 & 3 \end{bmatrix}$$

EXERCICE 1.7 Soit la matrice  $\mathbf{A} \in \mathbb{R}^{n \times n}$  avec  $|a_{ij}| < \varepsilon; \forall i = 1, \dots, n; i \neq j$ .

Supposons qu'il existe une ligne  $r \in \{1, \dots, n\}$  de la matrice  $\mathbf{A}$  telle que  $|a_{rr} - a_{ii}| \geq \delta; \forall i = 1, \dots, n; i \neq r$ .

Considérons la matrice  $\mathbf{W}_r \in \mathbb{R}^{n \times n}$  avec

$$w_{ij} = \begin{cases} 0, & \text{si } i \neq j \\ 1, & \text{si } i = j \text{ et } i \neq r \\ \alpha, & \text{si } i = j = r \end{cases}$$

avec  $\alpha > 0$ .

Soit la transformation de similitude  $\mathbf{B}_r = \mathbf{W}_r \mathbf{A} \mathbf{W}_r^{-1}$

- (1) Évaluer les éléments de  $\mathbf{B}_r$  en fonction des éléments de la matrice  $\mathbf{A}$ .
- (2) Calculer une borne supérieure pour les rayons des disques  $D_r^{(l)}$  et  $D_i^{(l)}; i = 1, \dots, n; i \neq r$  de la matrice  $\mathbf{B}_r$ .
- (3) Montrer que sous l'hypothèse  $\varepsilon < \frac{\delta}{2(n-1)}$  et en choisissant  $\alpha \leq \frac{2\varepsilon}{\delta}$ , on réduit le rayon du disque  $D_r^{(l)}$  de  $\mathbf{B}_r$  et on le rend disjoint des autres disques de la même matrice.

(4) Application : Soit la matrice

$$\mathbf{A} = \begin{bmatrix} 4 & 0.5 & -0.4 & 0.4 \\ 0.2 & -1 & -0.1 & 0.05 \\ -0.5 & -0.4 & 2.5 & 0.4 \\ 0.1 & 0.05 & 0.1 & -3 \end{bmatrix}$$

avec valeurs propres  $4.15, 2.39, -1.027, -3.01$ .

Nous avons  $\varepsilon = 0.5$ .

- Calculer les disques de Geršgorin et leurs rayons pour la matrice  $\mathbf{A}$ .
- Pour un indice  $r$  pour lequel le disque de Geršgorin  $D_r^{(l)}$  n'est pas isolé, calculer  $\delta$ .
- Prendre un  $\alpha \leq \frac{\delta}{2\varepsilon}$  et construire la matrice  $\mathbf{W}_r$ .
- Calculer la matrice  $\mathbf{B}_r$ .
- Calculer les disques de Geršgorin et leurs rayons pour la matrice  $\mathbf{B}_r$ .

Le disque  $D_r^{(l)}$  a maintenant une taille réduite.

S'il y a une autre ligne  $r'$  telle que  $D_r^{(l)} \cap D_{r'}^{(l)} \neq \emptyset$ , répéter cette opération pour la ligne  $r'$ .

#### 1.4 Estimation des erreurs en cas des perturbations

On commence en introduisant deux types des vecteurs propres. Celui,  $\mathbf{u}$ , que nous avons déjà vu comme solution du problème (1.2.1) qu'on appellera *vecteur propre droit* de  $\mathbf{A}$  et le vecteur  $\mathbf{v}$  solution du problème

$$\mathbf{v}^\top \mathbf{A} = \mu \mathbf{v}^\top ; \mathbf{v} \neq \mathbf{0}$$

qui sera appelé *vecteur propre gauche*. Par conséquent  $\mathbf{v}$  est un vecteur propre droit de la matrice  $\mathbf{A}^\top$ , c'est-à-dire  $\mathbf{A}\mathbf{v}^\top = \bar{\mu}\mathbf{v}^\top$ , où  $\bar{\mu}$  le conjugué de  $\mu$ . Pour cette raison  $\mathbf{v}^\top$  est appelé vecteur propre ligne de  $\mathbf{A}$ .

Notons que les valeurs propres  $\lambda$  et  $\mu$  sont identiques. Si, de plus,  $\mathbf{A}$  est symétrique, les vecteurs propre droits et gauches sont identiques.

La relation entre vecteur propre droit et gauche correspondant à deux valeurs propres différentes est donnée par le lemme suivant :

**LEMME 1.4.1** Soit un vecteur propre droit  $\mathbf{u}$  correspondant à une valeur propre  $\lambda$  et un vecteur propre gauche  $\mathbf{v}$  correspondant à une valeur propre  $\lambda'$  différente de  $\lambda$ . Alors  $\mathbf{u}$  et  $\mathbf{v}$  sont orthogonaux :

$$\mathbf{v}^\top \mathbf{u} = 0$$

L'étude des erreurs dans le calcul des valeurs propres quand la matrice  $\mathbf{A}$  est perturbée, n'est pas facile du fait que les expressions des valeurs propres en fonction de la perturbation ne sont pas toujours différentiables. Par exemple si on considère la matrice  $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  et la perturbation

$\Delta \mathbf{A} = \begin{bmatrix} 0 & 0 \\ \varepsilon & 0 \end{bmatrix}$ , les valeurs propres de  $\mathbf{A} + \Delta \mathbf{A}$  sont  $1 \pm \sqrt{\varepsilon}$  fonction de  $\varepsilon$  qui n'est pas différentiable au point  $\varepsilon = 0$ . Néanmoins les valeurs propres sont des fonctions continues des éléments de la matrice comme il est montré par le théorème suivant.

**THÉORÈME 1.4.1** *Soit  $\mathbf{A}$  une matrice carrée d'ordre  $n$ , avec  $n$  valeurs propres,  $\lambda_1, \dots, \lambda_n$  pas forcément toutes distinctes les unes des autres. Considérons la matrice perturbée  $\mathbf{A} + \Delta \mathbf{A}$  avec valeurs propres  $\mu_1, \dots, \mu_n$ . Pour chaque valeur propre  $\mu_i$  il existe une valeur propre  $\lambda_j$  telle que*

$$|\mu_i - \lambda_j| \leq 4(\|\mathbf{A}\|_2 + \|\mathbf{A} + \Delta \mathbf{A}\|_2)^{1-1/n} \|\Delta \mathbf{A}\|_2 \quad (1.4.1)$$

Le théorème suivant fournit une borne supérieure pour les erreurs du calcul des valeurs et vecteurs propres.

**THÉORÈME 1.4.2** *Soit  $\mathbf{A}$  une matrice carrée d'ordre  $n$ , avec  $n$  vecteurs propres linéairement indépendants. Considérons la perturbation  $\Delta \mathbf{A} = \varepsilon \mathbf{A}$  et la matrice perturbée*

$$\mathbf{A}(\varepsilon) = \mathbf{A} + \Delta \mathbf{A} \quad (1.4.2)$$

*Soit  $\lambda$  une valeur propre  $\mathbf{A}$ , alors la valeur propre  $\lambda(\varepsilon)$  de  $\mathbf{A}(\varepsilon)$  est telle que*

$$|\lambda(\varepsilon) - \lambda| = \mathcal{O}(\varepsilon^{1/m}) \quad (1.4.3)$$

*où  $m$  est la multiplicité de  $\lambda$ . De plus, si  $\lambda$  est simple, on a*

$$\lambda = \lim_{|\varepsilon| \rightarrow 0} \frac{\lambda(\varepsilon) - \lambda}{\varepsilon} = \frac{\mathbf{v}^\top \mathbf{A} \mathbf{u}}{\mathbf{v}^\top \mathbf{u}} \quad (1.4.4)$$

*où  $\mathbf{u}$  et  $\mathbf{v}$  sont les vecteurs propres droit et gauche de  $\mathbf{A}$  correspondant à  $\lambda$ .*

De ce théorème, on en conclut que l'erreur du calcul de la valeur propre  $\lambda$  dépend de la valeur de  $\left| \frac{\mathbf{v}^\top \mathbf{A} \mathbf{u}}{\mathbf{v}^\top \mathbf{u}} \right|$ . Si les vecteurs propres sont normalisés, c'est-à-dire si  $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2$ , on a

$$\left| \frac{\mathbf{v}^\top \mathbf{A} \mathbf{u}}{\mathbf{v}^\top \mathbf{u}} \right| \leq \frac{\|\mathbf{A}\|_2}{|\mathbf{v}^\top \mathbf{u}|} \quad (1.4.5)$$

Par conséquent, en vertu de (1.2.2) et (1.2.3), on obtient

$$\max_{|\varepsilon| \rightarrow 0} \lim \frac{\lambda(\varepsilon) - \lambda}{\varepsilon} = \frac{\|\mathbf{A}\|_2}{|\mathbf{v}^\top \mathbf{u}|} \quad (1.4.6)$$

Notons que ces résultats restent valables si on remplace dans (??) la matrice de perturbations  $\Delta \mathbf{A}$  par  $\Delta \mathbf{C} = \varepsilon \mathbf{C}$  où  $\mathbf{C}$  matrice carrée d'ordre  $n$ , avec  $\|\mathbf{C}\| = \|\mathbf{A}\|$ .

Sur le même sujet nous avons aussi le théorème suivant :

**THÉORÈME 1.4.3** *Soit  $\mathbf{A}$  matrice carrée, diagonalisable et soit  $\mathbf{U}$  matrice telle que  $\mathbf{U}^{-1} \mathbf{A} \mathbf{U} = \text{diag}(\lambda_1, \dots, \lambda_n)$ , où  $\lambda_i$  valeurs propres de  $\mathbf{A}$ . Considérons la matrice perturbée  $\mathbf{B} = \mathbf{A} + \Delta \mathbf{A}$ . Alors pour chaque valeur propre  $\mu$  de  $\mathbf{B}$  il existe une valeur propre  $\lambda$  de  $\mathbf{A}$  telle que*

$$|\mu - \lambda| \leq \kappa_\infty(\mathbf{U}) \cdot \|\Delta \mathbf{A}\|_\infty \quad (1.4.7)$$

D'après ce théorème la perturbation des valeurs propres due à une perturbation de la matrice dépend du conditionnement de la matrice  $\mathbf{U}$  et non pas de  $\mathbf{A}$ .

Pour les matrices symétriques nous avons aussi le

THÉORÈME 1.4.4 Si les matrices  $\mathbf{A}$  et  $\mathbf{A} + \Delta\mathbf{A}$  sont symétriques, avec valeurs propres  $\lambda_1, \dots, \lambda_n$  et  $\mu_1, \dots, \mu_n$  respectivement, alors

$$|\mu_i - \lambda_j| \leq \|\Delta\mathbf{A}\|_2 \quad (1.4.8)$$

## 1.5 Références

---

Les livres utilisés pour la rédaction de ce chapitre sont les suivants :

- [ 1 ] V. N. FADDEEVA : *Computational methods of linear algebra*, Dover, 1959
- [ 2 ] F. R. GANTMACHER : *Théorie des matrices*, tome 1, Théorie générale, Dunod, 1966
- [ 3 ] I. C. F. IPSEN : *Numerical matrix analysis*, SIAM 2009
- [ 4 ] E. ISAACSON, H. B. KELLER : *Analysis of numerical methods*, Wiley, 1966
- [ 5 ] H. LÜTKEPOHL : *Handbook of matrices*, Wiley, 1996
- [ 6 ] G. W. STEWART : *Matrix algorithms, Vol. II : Eigensystems*, SIAM, 2001
- [ 7 ] J. STOER, R. BULIRSCH : *Introduction to numerical analysis*, Second ed., Springer-Verlag, 1992
- [ 8 ] E. SÜLI, D. F. MAYERS : *An introduction to numerical analysis*, Cambridge Univ. Pr., 2003
- [ 9 ] L. N. TREFETHEN, D. BAU III : *Numerical linear algebra*, SIAM, 1997
- [10] D. S. WATKINS : *Fundamentals of matrix computations*, Second edition, Wiley, 2002
- [11] J. H. WILKINSON : *The algebraic eigenvalue problem*, Clarendon Press, 1965

# 2

## DÉCOMPOSITION EN VALEURS SINGULIÈRES

---

1.1	Introduction . . . . .	2
1.1.1	Exercices . . . . .	5
1.2	Calcul des valeurs et vecteurs propres . . . . .	5
1.2.1	Calcul par factorisation <b>LU</b> . . . . .	8
1.3	Localisation des valeurs propres . . . . .	10
1.3.1	Exercices . . . . .	11
1.4	Estimation des erreurs en cas des perturbations . . . . .	12
1.5	Références . . . . .	14

---

### 2.1 Étude géométrique des transformations linéaires

---

#### 2.1.1 Les espaces fondamentaux d'une application linéaire

Rappelons d'abord qu'une matrice  $\mathbf{A}$  est orthogonale si  $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$ , d'où on a  $\mathbf{A}^{-1} = \mathbf{A}^\top$ . Dans ce cas ses colonnes et ses lignes sont orthonormales. De plus pour chaque vecteur  $\mathbf{x} \in \mathbb{R}^n$ , nous avons  $\|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ , c'est-à-dire l'application linéaire  $A$  associée à la matrice  $\mathbf{A}$  est une *isométrie*.

Considérons une application linéaire  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  et soit  $\mathbf{A} \in \mathbb{R}^{m \times n}$  la matrice associée avec rang égal à  $r \leq \min\{m, n\}$ . À cette application sont associés quatre sous-espaces, deux de  $\mathbb{R}^n$  et deux de  $\mathbb{R}^m$ . Le premier espace correspond à l'image de l'application qui est un sous-espace de  $\mathbb{R}^m$  que l'on note  $R(A)$  et qui est définie comme suit :

$$R(A) = \{\mathbf{y} \in \mathbb{R}^m / \mathbf{y} = A(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^m$$

Bien évidemment la matrice associée  $\mathbf{A}$  engendre, par l'intermédiaire de l'application  $A$ , un sous-espace dans  $\mathbb{R}^m$  qui, par abus de langage, on appellera l'*image* de  $\mathbf{A}$  et l'on notera  $R(\mathbf{A})$ . Nous avons

$$R(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m / \mathbf{y} = \mathbf{A}\mathbf{x}\} \subset \mathbb{R}^m \tag{2.1.1}$$

et  $\dim R(\mathbf{A}) = r$

Le deuxième espace est engendré par la transposée de  $\mathbf{A}$ . Il s'agit de l'image  $R(\mathbf{A}^\top)$  définie par

$$R(\mathbf{A}^\top) = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \mathbf{A}^\top \mathbf{y}, \mathbf{y} \in \mathbb{R}^m \right\} \subset \mathbb{R}^n \quad (2.1.2)$$

Nous avons  $\dim R(\mathbf{A}^\top) = r$ .

Notons que  $R(\mathbf{A})$  est le sous-espace engendré par les colonnes de  $\mathbf{A}$  et  $R(\mathbf{A}^\top)$  est le sous-espace engendré par les lignes de  $\mathbf{A}$ .

Le troisième espace est engendré par le *noyau* de  $\mathbf{A}$  qui est l'ensemble des solutions du système homogène  $\mathbf{A}\mathbf{x} = \mathbf{0}$  :

$$N(\mathbf{A}) = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0} \right\} \subset \mathbb{R}^n \quad (2.1.3)$$

Nous avons  $\dim N(\mathbf{A}) = n - r$ . Le quatrième espace est le noyau de la matrice transposée :

$$N(\mathbf{A}^\top) = \left\{ \mathbf{y} \in \mathbb{R}^m \mid \mathbf{A}^\top \mathbf{y} = \mathbf{0} \right\} \subset \mathbb{R}^m \quad (2.1.4)$$

avec  $\dim N(\mathbf{A}^\top) = m - r$ .

Nous pouvons constater que  $\dim R(\mathbf{A}) + \dim N(\mathbf{A}) = n$ .

Nous avons la propriété suivante :

**PROPOSITION 2.1.1** Soit  $\mathbf{A} \in \mathbb{R}^{m \times n}$  une matrice. Nous avons

- $N(\mathbf{A}) = \{\mathbf{0}\}$  si et seulement si  $\text{rang } \mathbf{A} = n$  ;
- $N(\mathbf{A}^\top) = \{\mathbf{0}\}$  si et seulement si  $\text{rang } \mathbf{A} = m$

**PROPOSITION 2.1.2** Soient deux matrices  $\mathbf{A} \in \mathbb{R}^{m \times n}$  et  $\mathbf{B} \in \mathbb{R}^{n \times p}$ . Nous avons

-

$$\text{rang}(\mathbf{AB}) = \text{rang}(\mathbf{B}) - \dim N(\mathbf{A}) \cap R(\mathbf{B}) \quad (2.1.5)$$

-

$$\text{rang}(\mathbf{AB}) \leq \min \{ \text{rang}(\mathbf{A}), \text{rang}(\mathbf{B}) \} \quad (2.1.6)$$

-

$$\text{rang}(\mathbf{A}) + \text{rang}(\mathbf{B}) \leq \text{rang}(\mathbf{AB}) + n \quad (2.1.7)$$

-

$$\text{rang}(\mathbf{A}^\top \mathbf{A}) = \text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A} \mathbf{A}^\top) \quad (2.1.8)$$

**PROPOSITION 2.1.3** Soit  $\mathbf{A} \in \mathbb{R}^{m \times n}$  une matrice. Nous avons

-

$$R(\mathbf{A}^\top \mathbf{A}) = R(\mathbf{A}^\top) \text{ et } R(\mathbf{A} \mathbf{A}^\top) = R(\mathbf{A}) \quad (2.1.9)$$

-

$$N(\mathbf{A}^\top \mathbf{A}) = N(\mathbf{A}) \text{ et } N(\mathbf{A} \mathbf{A}^\top) = N(\mathbf{A}^\top) \quad (2.1.10)$$

### 2.1.2 Espaces complémentaires et projections

Pour la suite nous avons besoin des sous-espaces complémentaires et des projections dont nous donnons un bref rappel.

**Definition** Deux sous-espaces  $G$  et  $H$  de l'espace  $E \subseteq \mathbb{R}^n$  sont complémentaires si

$$E = G + H \text{ et } G \cap H = \emptyset$$

Dans ce cas  $E$  est la somme directe de  $G$  et  $H$  qui sera notée  $E = G \oplus H$ . De plus les sous-espaces  $F$  et  $G$  sont orthogonaux,  $F \perp G$  et si  $\dim F = r$ , alors  $\dim G = n - r$ .

Nous avons aussi que tout vecteur de  $E$  a une décomposition unique en deux vecteurs de  $G$  et  $H$  :

$$\forall \mathbf{x} \in E \exists \text{ un seul } \mathbf{y} \in F \text{ et un seul } \mathbf{z} \in H : \mathbf{x} = \mathbf{y} + \mathbf{z} \quad (2.1.11)$$

Cette remarque, nous conduit maintenant à la notion de la projection.

**Definition** Soit la somme directe  $E = G \oplus H$  et la décomposition  $\mathbf{x} = \mathbf{y} + \mathbf{z}$  avec  $\mathbf{x} \in E, \mathbf{y} \in F$  et  $\mathbf{z} \in H$ . Nous avons que

- le vecteur  $\mathbf{y}$  est la projection de  $\mathbf{x}$  dans  $F$ , et
- le vecteur  $\mathbf{z}$  est la projection de  $\mathbf{x}$  dans  $G$ .

Ce qui est intéressant en analyse numérique c'est de savoir calculer les projections d'un vecteur. Pour ce faire, il faut construire l'opérateur de projection qui est une matrice  $\mathbf{P} \in \mathbb{R}^{n \times n}$  telle que  $\forall \mathbf{x} \in E, \mathbf{P}\mathbf{x}$  est la projection de  $\mathbf{x}$  dans  $F$ . Pour cette construction, on considère les bases  $\{\mathbf{f}_1, \dots, \mathbf{f}_r\}$  de  $F$  et  $\{\mathbf{g}_1, \dots, \mathbf{g}_{n-r}\}$  de  $G$  et on forme la matrice  $\mathbf{B} = [\mathbf{f}_1, \dots, \mathbf{f}_r, \mathbf{g}_1, \dots, \mathbf{g}_{n-r}] = [\mathbf{F} | \mathbf{G}]$ . Comme les vecteurs de deux bases sont linéairement indépendants et les deux sous-espaces sont orthogonaux, la matrice  $\mathbf{B}$  est régulière. Si donc  $\mathbf{P}\mathbf{x}$  est la projection de  $\mathbf{x}$  dans  $F$ , on a d'après (2.1.11)  $\mathbf{P}\mathbf{B} = \mathbf{P}[\mathbf{F} | \mathbf{G}] = [\mathbf{F} | \mathbf{0}]$  et par conséquent on a pour l'opérateur de projection de  $E$  dans  $F$ , la formule

$$\mathbf{P} = [\mathbf{F} | \mathbf{0}] \mathbf{B}^{-1} = \mathbf{B} \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{B}^{-1} \quad (2.1.12)$$

Si  $\mathbf{P}$  est l'opérateur de projection de  $E$  dans  $F$ , alors l'opérateur de projection de  $E$  dans  $G$  est  $\mathbf{I} - \mathbf{P}$  et il est donné par la formule

$$\mathbf{Q} = (\mathbf{I} - \mathbf{P}) = \mathbf{B} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-r} \end{bmatrix} \mathbf{B}^{-1} \quad (2.1.13)$$

Notons que les opérateurs de projection  $\mathbf{P}$  et  $\mathbf{Q}$  sont uniques. Ils sont aussi idempotents, à savoir

$$\mathbf{P}^2 = \mathbf{P} \text{ et } \mathbf{Q}^2 = \mathbf{Q} \quad (2.1.14)$$

Le contraire aussi est vrai, c'est-à-dire un opérateur linéaire est un opérateur de projection si et seulement s'il est idempotent.

Nous pouvons relier la notion de l'opérateur de projection avec celle de l'image et du noyau d'une matrice régulière  $\mathbf{A}$ . En effet, si on considère la décomposition  $\mathbf{A}\mathbf{x} = \mathbf{y} + \mathbf{z} = \mathbf{y} + \mathbf{0}$ , nous

avons que  $\mathbf{y} \in R(\mathbf{A})$  et  $\mathbf{z} \in N(\mathbf{A})$ . Par conséquent dans ce cas, les deux espaces complémentaires  $F$  et  $G$  sont  $R(\mathbf{A})$  et  $N(\mathbf{A})$  respectivement. Nous insistons sur le fait que la matrice  $\mathbf{A}$  doit être régulière.

### 2.1.3 Décomposition orthogonale

Nous avons maintenant tous les éléments qui permettent d'étudier le rôle des quatre espaces associés à une application linéaire. Nous avons le

THÉORÈME 2.1.1 Soit une matrice  $\mathbf{A} \in \mathbb{R}^{m \times n}$ . Nous avons

$$R(\mathbf{A})^\perp = N(\mathbf{A}^\top) \text{ et } N(\mathbf{A})^\perp = R(\mathbf{A}^\top) \quad (2.1.15)$$

La matrice  $\mathbf{A}$  engendre une décomposition de  $\mathbb{R}^m$  et  $\mathbb{R}^n$  suivant les schémas suivants

$$\mathbb{R}^m = R(\mathbf{A}) \oplus R(\mathbf{A})^\perp = R(\mathbf{A}) \oplus N(\mathbf{A}^\top) \quad (2.1.16)$$

et

$$\mathbb{R}^n = N(\mathbf{A}) \oplus N(\mathbf{A})^\perp = N(\mathbf{A}) \oplus R(\mathbf{A}^\top) \quad (2.1.17)$$

La notion de l'opérateur de projection, nous conduit à la notion de la projection orthogonale (élémentaire) d'un vecteur unitaire  $\mathbf{u} \in \mathbb{R}^n$  donnée par la matrice

$$\mathbf{P} = \mathbf{I} - \mathbf{u}\mathbf{u}^\top \quad (2.1.18)$$

Le nom de cette matrice est dû au fait que  $\mathbf{P}$  projette chaque vecteur  $\mathbf{x} \in \mathbb{R}^n$  dans l'espace qui contient tous les vecteurs qui sont perpendiculaires à  $\mathbf{u}$ . En effet  $\mathbf{x} \in \mathbb{R}^n$  peut être décomposé en deux parties :

$$\mathbf{x} = (\mathbf{I} - \mathbf{P})\mathbf{x} + \mathbf{P}\mathbf{x}, \text{ où } (\mathbf{I} - \mathbf{P})\mathbf{x} \perp \mathbf{P}\mathbf{x}$$

### 2.1.4 Exercices

EXERCICE 2.1 Déterminer la projection orthogonale  $\mathbf{x} = [2, 0, 1]^\top$  dans  $\text{vect}\{\mathbf{u}\}$  ainsi que la projection orthogonale de  $\mathbf{x}$  dans  $\mathbf{u}^\perp$ . On prendra  $\mathbf{u} = [2, -1, 3]^\top$ . Préciser le projecteur orthogonal  $\mathbf{P}_{\mathbf{u}}$  dans  $\text{vect}\{\mathbf{u}\}$  et le projecteur orthogonal  $\mathbf{P}_{\mathbf{u}^\perp}$  dans  $\mathbf{u}^\perp$ .

EXERCICE 2.2 Toute matrice  $\mathbf{A} \in \mathbb{R}^{n \times n}$  carrée, diagonalement dominante strictement (c'est-à-dire telle que  $|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$ ;  $i = 1, \dots, n$ ) est régulière.

## 2.2 Étude géométrique des transformations linéaires

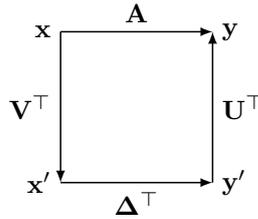
La décomposition en valeurs singulières  $\mathbf{A} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^\top$  permet de représenter l'application linéaire  $\mathbf{y} = \mathbf{A}\mathbf{x}$  par l'application linéaire  $\mathbf{y}' = \mathbf{\Delta}\mathbf{x}'$ . Ce changement d'espace se passe de la façon suivante : considérons les espaces linéaires

- $R(\mathbf{U}) = \{\mathbf{y} \in \mathbb{R}^m / \mathbf{y} = \mathbf{U}\mathbf{y}' ; \mathbf{y}' \in \mathbb{R}^m\}$  engendré par  $\mathbf{U}$  ;
- $R(\mathbf{V}) = \{\mathbf{x} \in \mathbb{R}^n / \mathbf{x} = \mathbf{V}\mathbf{x}' ; \mathbf{x}' \in \mathbb{R}^n\}$  engendré par  $\mathbf{V}$ .

Nous avons  $\mathbf{y} = \mathbf{A}\mathbf{x}$ . Si on remplace  $\mathbf{A}$  par sa transformée, on a  $\mathbf{y} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^\top \mathbf{x}$ . Dans l'espace  $R(\mathbf{V})$  on a  $\mathbf{V}^\top \mathbf{x} = \mathbf{x}'$ , d'où  $\mathbf{y} = \mathbf{U}\mathbf{\Delta}\mathbf{x}'$ . Si donc on pose

$$\mathbf{y}' = \mathbf{\Delta}\mathbf{x}' \quad (2.2.1)$$

on obtient  $\mathbf{y}' = \mathbf{U}^\top \mathbf{y}$  dans l'espace  $R(\mathbf{U})$ . Le diagramme suivant illustre cette transformation.



Ainsi, en utilisant (2.2.1), on a que les coordonnées du vecteur  $\mathbf{y}' \in \mathbb{R}^m$  sont  $y'_1 = \sigma_1 x'_1, \dots, y'_r = \sigma_r x'_r, y'_{r+1} = 0, \dots, y'_m = 0$ , où  $x'_i$  sont les coordonnées du vecteur  $\mathbf{x}' \in \mathbb{R}^n$ . Par conséquent nous pouvons représenter un vecteur  $\mathbf{y} \in \mathbb{R}^m$  par un vecteur  $\mathbf{y}' \in \mathbb{R}^m$  en utilisant  $r \leq m$  coordonnées non-nulles.

Nous présentons ici une étude complète de cette transformation qui sert aussi comme introduction aux sections suivantes de ce chapitre.

Nous avons considéré la matrice  $\mathbf{A} \in \mathbb{R}^{m \times n}$  comme une application linéaire  $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Nous avons vu qu'autour de cette application gravitent en orbite quatre espaces qui jouent un rôle important en algèbre linéaire. Nous allons maintenant déterminer ces espaces en fonction des matrices  $\mathbf{U}$  et  $\mathbf{V}$ .

- Nous avons d'abord l'espace engendré par les colonnes de la matrice  $\mathbf{A}$  qui est l'image de  $\mathbf{A}$  et qui sera notée par  $R(\mathbf{A}) = \{\mathbf{A}\mathbf{x} \in \mathbb{R}^m / \mathbf{x} \in \mathbb{R}^n\}$ . La dimension de cet espace est le rang  $r$  de la matrice. Si on utilise la DVS, alors on a que  $R(\mathbf{A})$  peut être engendré par les  $r$  vecteurs propres  $\mathbf{u}_1, \dots, \mathbf{u}_r$  correspondant aux  $r$  valeurs propres non nulles de la matrice  $\mathbf{A}\mathbf{A}^\top$ . Ces vecteurs sont les  $r$  premières colonnes de la matrice  $\mathbf{U}$ .
- Un autre espace est associé avec les vecteurs  $\mathbf{x} \in \mathbb{R}^n$  qui sont solution de l'équation homogène  $\mathbf{A}\mathbf{x} = \mathbf{0}$ . Cet espace est le noyau de  $\mathbf{A}$  et sera noté par  $N(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n / \mathbf{A}\mathbf{x} = \mathbf{0}\}$ . Sa dimension est égale à  $n - r$ . Si on utilise la DVS, alors on a que  $N(\mathbf{A})$  peut être engendré par les  $n - r$  vecteurs propres  $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$  correspondant aux  $n - r$  valeurs propres nulles de la matrice  $\mathbf{A}^\top \mathbf{A}$ .

Les deux autres espaces sont issus de la matrice  $\mathbf{A}^\top$  selon la même procédure.

- Nous avons ainsi l'espace engendré par les colonnes de  $\mathbf{A}^\top$ , c'est-à-dire par les lignes de  $\mathbf{A}$  qui est l'image de  $\mathbf{A}^\top$  et qui sera notée par  $R(\mathbf{A}^\top) = \{\mathbf{A}^\top \mathbf{y} \in \mathbb{R}^n / \mathbf{y} \in \mathbb{R}^m\}$ . Cet espace est orthogonal à  $N(\mathbf{A})$  et donc  $R(\mathbf{A}^\top) = N(\mathbf{A})^\perp$  avec  $N(\mathbf{A})^\perp = \mathbb{R}^n - N(\mathbf{A})$ . Sa dimension est égale à  $r$ .
- Nous avons enfin l'espace engendré par les vecteurs du système homogène  $\mathbf{A}^\top \mathbf{y} = \mathbf{0}$ . Cet espace est le noyau de  $\mathbf{A}^\top$  et sera noté par  $N(\mathbf{A}^\top) = \{\mathbf{y} \in \mathbb{R}^m / \mathbf{A}^\top \mathbf{y} = \mathbf{0}\}$ . Cet espace est

orthogonal à  $R(\mathbf{A})$  et donc  $N(\mathbf{A}^\top) = R(\mathbf{A})^\perp$  avec  $R(\mathbf{A})^\perp = \mathbb{R}^m - R(\mathbf{A})$ . Sa dimension est égale à  $m - r$ .

D'après la définition, nous avons  $N(\mathbf{A}) \cap R(\mathbf{A}^\top) = N(\mathbf{A}) \cap N(\mathbf{A})^\perp = \{\mathbf{0}\}$  et  $R(\mathbf{A}) \cap N(\mathbf{A}^\top) = R(\mathbf{A}) \cap R(\mathbf{A})^\perp = \{\mathbf{0}\}$ . Donc nous pouvons envisager la décomposition de  $\mathbb{R}^n$  et  $\mathbb{R}^m$  en sommes directes<sup>1</sup> :

$$\mathbb{R}^n = N(\mathbf{A}) \oplus R(\mathbf{A}^\top) \quad (2.2.2)$$

$$\mathbb{R}^m = R(\mathbf{A}) \oplus N(\mathbf{A}^\top) \quad (2.2.3)$$

Nous pouvons ainsi considérer toute application linéaire  $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  comme étant la composition de deux applications : une application de  $\mathbb{R}^n$  dans  $N(\mathbf{A})^\perp$ , qui est en fait la projection orthogonale<sup>2</sup> de tout vecteur de  $\mathbb{R}^n$  dans  $N(\mathbf{A})^\perp = R(\mathbf{A}^\top)$  et une application de  $R(\mathbf{A}^\top)$  dans  $\mathbb{R}^m$  qui est en fait la restriction de  $\mathbf{A}$  dans  $N(\mathbf{A})^\perp$ . Nous avons ainsi le diagramme suivant :

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{\mathbf{P}_{R(\mathbf{A}^\top)}} & N(\mathbf{A})^\perp & \xrightarrow{\mathbf{A}_{N(\mathbf{A})^\perp}} & \mathbb{R}^m \\ & \searrow & \mathbf{A} & \nearrow & \end{array}$$

dans lequel on a noté par  $\mathbf{P}_{R(\mathbf{A}^\top)} : \mathbb{R}^n \rightarrow N(\mathbf{A})^\perp = R(\mathbf{A}^\top)$  la projection orthogonale et par  $\mathbf{A}_{N(\mathbf{A})^\perp} : N(\mathbf{A})^\perp \rightarrow R(\mathbf{A}) \subseteq \mathbb{R}^m$  la restriction de  $\mathbf{A}$  dans  $N(\mathbf{A})^\perp$ .

Avant de terminer cette section nous allons examiner les projections

–  $\mathbf{P}_{R(\mathbf{A}^\top)} : \mathbb{R}^n \rightarrow N(\mathbf{A})^\perp = R(\mathbf{A}^\top)$  qui est une matrice  $(n, n)$  avec les propriétés suivantes :

$$1.- \mathbf{P}_{R(\mathbf{A}^\top)} = \mathbf{P}_{R(\mathbf{A}^\top)}^\top = \mathbf{P}_{R(\mathbf{A}^\top)}^2$$

$$2.- \mathbf{P}_{R(\mathbf{A}^\top)} \mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{x} \in N(\mathbf{A})$$

–  $\mathbf{P}_{R(\mathbf{A})} : \mathbb{R}^m \rightarrow R(\mathbf{A})$ .  $\mathbf{P}_{R(\mathbf{A})}$  qui est une matrice  $(m, m)$  avec les propriétés suivantes :

$$(1'.) \mathbf{P}_{R(\mathbf{A})} = \mathbf{P}_{R(\mathbf{A})}^\top = \mathbf{P}_{R(\mathbf{A})}^2$$

$$(2'.) \mathbf{P}_{R(\mathbf{A})} \mathbf{b} = \mathbf{b} \Leftrightarrow \mathbf{b} \in R(\mathbf{A})$$

Notons aussi qu'à cause de sa construction  $\mathbf{A}_{N(\mathbf{A})^\perp}$  est injective et, par conséquent, son inverse

$$\mathbf{A}_{N(\mathbf{A})^\perp}^{-1} : R(\mathbf{A}) \rightarrow N(\mathbf{A})^\perp \quad (2.2.4)$$

existe. L'existence de cette application inverse permet la définition de la matrice inverse généralisée, qui sera étudiée au chapitre suivant.

1. Si  $E_1, E_2$  sont des sous-espaces de  $\mathbb{R}^n$  avec  $E_1 \cap E_2 = \{\mathbf{0}\}$ , alors leur somme directe est

$$E_1 \oplus E_2 = \{\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 / \mathbf{x}_1 \in E_1; \mathbf{x}_2 \in E_2\}$$

2. Soit  $\mathbb{R}^n = E_1 \oplus E_2$ . Alors tout vecteur  $\mathbf{x} \in \mathbb{R}^n$  peut s'écrire de façon unique comme une somme  $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$  avec  $\mathbf{x}_1 \in E_1; \mathbf{x}_2 \in E_2$ . Le vecteur  $\mathbf{x}_i$  est appelé la *projection orthogonale* de  $\mathbf{x}$  dans  $E_i$ .

### 2.2.1 Exercices

EXERCICE 2.3 Soit l'application linéaire  $A : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  définie sur les vecteurs de la base canonique par les relations

$$A(\mathbf{e}_1) = 2\mathbf{e}_1 + 6\mathbf{e}_2; A(\mathbf{e}_2) = -\mathbf{e}_2; A(\mathbf{e}_3) = \mathbf{e}_1 + \mathbf{e}_3$$

où  $\mathbf{e}_i$  vecteur de la base canonique  $\mathcal{B}$ .

- (1) Écrire la matrice  $\mathbf{A}$  de l'application linéaire  $A$  dans la base  $\mathcal{B}$ .
- (2) Déterminer  $N(\mathbf{A})$  et  $R(\mathbf{A})$  ainsi que leurs dimensions.
- (3) Montrer que les vecteurs

$$\mathbf{f}_1 = \mathbf{e}_2; \mathbf{f}_2 = \mathbf{e}_1 + 2\mathbf{e}_2; \mathbf{f}_3 = -\mathbf{e}_1 - 3\mathbf{e}_2 + \mathbf{e}_3$$

forment une nouvelle base  $\mathcal{B}'$ .

- (4) Écrire la matrice  $\mathbf{A}$  dans la nouvelle base  $\mathcal{B}'$ .

EXERCICE 2.4 Soit la matrice

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \quad (2.2.5)$$

- (1) Calculer les matrices  $\mathbf{L} \in \mathbf{R}^{2 \times 2}$ ,  $\mathbf{U} \in \mathbf{R}^{2 \times 3}$  telles que  $\mathbf{A} = \mathbf{L}\mathbf{U}$ .
- (2) Déterminer les quatre espaces associés avec  $\mathbf{A}$  et préciser leur base.
- (3) En utilisant la réponse à la question précédente, expliquer pourquoi le système  $\mathbf{A}\mathbf{x} = \mathbf{b}$ , avec  $\mathbf{b} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$  n'a pas de solution. Trouver le vecteur  $\mathbf{b}'$  le plus proche de  $\mathbf{b}$  pour lequel le système a une solution.

## 2.3 Décomposition en valeurs singulières

Fréquemment dans les applications, nous avons à résoudre de systèmes dont la matrice  $\mathbf{A}$  est rectangulaire, le plus souvent ayant plus de lignes que de colonnes. Dans ces cas on utilise les *valeurs singulières*.

**Definition** Soit une matrice  $\mathbf{A} \in \mathbb{R}^{n \times m}$ . Les valeurs singulières  $\sigma$  de  $\mathbf{A}$  sont les racines carrées des valeurs propres de la matrice  $\mathbf{A}^\top \mathbf{A}$  (ou  $\mathbf{A}\mathbf{A}^\top$ ), c'est-à-dire  $\sigma_i = \sqrt{\lambda_i}$ ;  $i = 1, \dots, r$  où  $r$  le rang de la matrice  $\mathbf{A}$ .

La notion de la forme quadratique définie positive<sup>3</sup>, associée à une matrice symétrique, dont nous pouvons calculer les vecteurs propres, permet de définir la *décomposition en valeurs singulières* (DVS ou SVD - singular value decomposition) d'une matrice rectangulaire quelconque. Nous avons le théorème suivant :

3. Une forme quadratique  $\mathbf{q}(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$  est dite définie positive si, pour tout vecteur  $\mathbf{x}$ , on a  $\mathbf{q}(\mathbf{x}) \geq 0$  et  $\mathbf{q}(\mathbf{x}) = 0$  si et seulement si  $\mathbf{x} = \mathbf{0}$ .

**THÉORÈME 2.3.1** Toute matrice rectangulaire  $\mathbf{A} \in \mathbb{R}^{m \times n}$  de rang  $r \leq \min(m, n)$ , admet une factorisation :

$$\mathbf{A} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^\top \quad (2.3.1)$$

avec :

- $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}$  : matrice orthogonale dont les colonnes  $\mathbf{u}_i$  sont les vecteurs propres de la matrice  $\mathbf{A}\mathbf{A}^\top = \mathbf{U}\mathbf{\Delta}\mathbf{V}^\top (\mathbf{U}\mathbf{\Delta}\mathbf{V}^\top)^\top = \mathbf{U}\mathbf{\Delta}\mathbf{\Delta}^\top \mathbf{U}^\top$ .
- $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n] \in \mathbb{R}^{n \times n}$  : matrice orthogonale dont les colonnes  $\mathbf{v}_i$  sont les vecteurs propres de la matrice  $\mathbf{A}^\top \mathbf{A} = (\mathbf{U}\mathbf{\Delta}\mathbf{V}^\top)^\top \mathbf{U}\mathbf{\Delta}\mathbf{V}^\top = \mathbf{V}\mathbf{\Delta}^\top \mathbf{\Delta}\mathbf{V}$ .
- $\mathbf{\Delta} \in \mathbb{R}^{m \times n}$  : la matrice de valeurs singulières,  $\mathbf{\Delta} = \mathbf{\Lambda}^{\frac{1}{2}}$  avec  $\mathbf{\Lambda}$  étant la matrice diagonale des valeurs propres non nulles de la matrice  $\mathbf{A}^\top \mathbf{A}$  et, aussi, des valeurs propres non nulles de la matrice  $\mathbf{A}\mathbf{A}^\top$ .

Cette factorisation de la matrice  $\mathbf{A}$  s'appelle *décomposition en valeurs singulières (DVS)*. Remarquons que si  $r \leq \min(m, n)$  est le rang de la matrice  $\mathbf{A}$ , alors il y a  $r$  valeurs singulières non nulles. Donc si  $r < \min(m, n)$ , on a

$$\mathbf{\Delta} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

avec  $\mathbf{D} = \text{diag}(\sigma_1, \dots, \sigma_r)$ . Ainsi nous avons que la matrice  $\mathbf{\Delta}$  a  $n - r$  colonnes nulles et  $m - r$  lignes nulles. En exploitant le fait que le carré des valeurs singulières sont les valeurs propres de  $\mathbf{A}^\top \mathbf{A}$  (et de  $\mathbf{A}\mathbf{A}^\top$ ), nous avons la diagonalisation de  $\mathbf{A}^\top \mathbf{A}$

$$\mathbf{V}^\top \mathbf{A}^\top \mathbf{A} \mathbf{V} = \begin{bmatrix} \mathbf{D}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (2.3.2)$$

On voit ainsi que  $\mathbf{V}$  est la matrice des vecteurs propres droits de  $\mathbf{A}^\top \mathbf{A}$  et donc la matrice des *vecteurs singuliers droits* de  $\mathbf{A}$ . De même  $\mathbf{U}$  est la matrice des *vecteurs singuliers gauches* de  $\mathbf{A}$ .

Si on tient compte de la matrice  $\mathbf{D}$ , nous pouvons exprimer la factorisation (2.3.1) comme suit

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top \quad (2.3.3)$$

où les matrices  $\mathbf{U}, \mathbf{D}$  ont les dimensions appropriées.

Notons que la détermination des valeurs singulières est unique mais pas celle des vecteurs singuliers. Néanmoins, si nous avons déterminé la matrice orthogonale  $\mathbf{V}$  des vecteurs singuliers droits, alors on ne peut pas prendre pour  $\mathbf{U}$  n'importe quelle matrice orthogonale. En effet les colonnes de la matrice  $\mathbf{U}$  sont déterminées par la relation

$$\mathbf{u}_i = \frac{\mathbf{A}\mathbf{v}_i}{\sigma_i} \text{ si } \mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i; \text{ pour } i = 1, \dots, r$$

$$\text{vect} \{ \mathbf{u}_{r+1}, \dots, \mathbf{u}_m \} = N(\mathbf{A}^\top) \text{ si } \mathbf{u}_i^\top \mathbf{A} = 0; \text{ pour } i = r+1, \dots, m$$

Les valeurs propres de la matrice  $\mathbf{A}$  correspondent aux directions invariantes de l'application linéaire  $A$ . Les valeurs singulières contiennent des informations sur la forme géométrique de cette application. En effet, considérons la sphère unité  $S = \{ \mathbf{x} \in \mathbb{R}^n / \|\mathbf{A}\mathbf{x}\| = 1 \}$ . L'image  $A(S)$  de la sphère unité par  $A$ , donnée par  $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$  est un ellipsoïde dont les longueurs des demi-axes principaux sont égales aux valeurs singulières de  $\mathbf{A}$  (voir la figure dans le cas du cercle unité).

En effet si  $\mathbf{Ax} = \mathbf{UDV}^\top \mathbf{x}$ , nous avons que l'application  $A$  agit sur les vecteurs de  $S$  de en trois temps :

- les colonnes de  $\mathbf{V}$  – vecteurs singuliers droits de  $\mathbf{A}$  – sont des éléments de la sphère unité et, par conséquent, on peut les considérer comme une base pour l'espace dans lequel se trouve la sphère ;
- la matrice diagonale  $\mathbf{D}$  transforme la sphère en une ellipse suivant la base canonique, et le vecteur  $\mathbf{x}$  se transforme en  $[\sigma_1 \mathbf{v}_1^\top \mathbf{x}, \dots, \sigma_r \mathbf{v}_r^\top \mathbf{x}]^\top$  ;
- l'application unitaire  $U$  opère une rotation de l'ellipse.

Nous pouvons donner un aspect géométrique à la DVS à l'aide du théorème suivant qui est la version géométrique du théorème 2.3.1 :

**THÉORÈME 2.3.2** Soit une matrice rectangulaire  $\mathbf{A} \in \mathbb{R}^{m \times n}$  de rang  $r \leq \min\{n, m\}$ . Alors  $\mathbb{R}^n$  a une base orthonormale  $\mathbf{u}_1, \dots, \mathbf{u}_n$ ,  $\mathbb{R}^m$  a une base orthonormale  $\mathbf{v}_1, \dots, \mathbf{v}_m$  et il existe des nombres  $\sigma_1 \geq \dots \geq \sigma_r > 0$  tels que

$$\mathbf{A}\mathbf{v}_i = \begin{cases} \sigma_i \mathbf{u}_i, & \text{pour } i = 1, \dots, r \\ \mathbf{0}, & \text{pour } i = r+1, \dots, m \end{cases}, \quad \mathbf{A}^\top \mathbf{u}_i = \begin{cases} \sigma_i \mathbf{v}_i, & \text{pour } i = 1, \dots, r \\ \mathbf{0}, & \text{pour } i = r+1, \dots, m \end{cases} \quad (2.3.4)$$

Une extension de ce théorème est le suivant :

**THÉORÈME 2.3.3** Soit une matrice rectangulaire  $\mathbf{A} \in \mathbb{R}^{m \times n}$  de rang  $r = \min\{n, m\}$ . On définit la matrice

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^\top & \mathbf{0} \end{bmatrix} \quad (2.3.5)$$

Les valeurs singulières de  $\mathbf{A}$  sont  $\sigma_1, \dots, \sigma_r$  si et seulement si les valeurs singulières non nulles de  $\mathbf{B}$  sont  $\sigma_1, \dots, \sigma_r, -\sigma_1, \dots, -\sigma_r$ .

### 2.3.1 DVS réduite

Considérons la matrice  $\mathbf{A} \in \mathbb{R}^{m \times n}$  et supposons que  $n \geq m$  et  $\text{rang } \mathbf{A} = m$ . La factorisation (2.3.3) montre que si le nombre  $m$  de lignes de  $\mathbf{A}$  est supérieure au nombre  $n$  de colonnes, nous pouvons, pour faire la DVS de  $\mathbf{A}$ , à la place des matrices  $\mathbf{U} \in \mathbb{R}^{m \times m}$  et  $\mathbf{\Delta} \in \mathbb{R}^{m \times n}$  utiliser les matrices réduites  $\hat{\mathbf{U}} \in \mathbb{R}^{m \times n}$  et  $\hat{\mathbf{\Delta}} \in \mathbb{R}^{n \times n}$ .

Du point de vue théorique, bien que les colonnes de  $\hat{\mathbf{U}}$  sont orthonormales,  $\hat{\mathbf{U}}$  n'est pas une matrice unitaire. En rajoutant  $n - m$  colonnes orthonormales on peut obtenir de  $\hat{\mathbf{U}} \in \mathbb{R}^{n \times m}$ , la matrice  $\mathbf{U} \in \mathbb{R}^{n \times n}$  qui est unitaire. Dans ce cas, pour pouvoir calculer la DVS, il faut aussi ajouter à la matrice  $\hat{\mathbf{\Delta}} \in \mathbb{R}^{m \times m}$ ,  $n - m$  lignes. Afin que le résultat des calculs ne soit pas affecté, il faut que ces lignes contiennent des 0. On obtient donc la matrice  $\mathbf{\Delta} \in \mathbb{R}^{n \times m}$  dont les  $n - m$  dernières lignes sont nulles. On retrouve ainsi la DVS de la matrice  $\mathbf{A}$ .

**EXEMPLE 2.3.1** Soit la matrice :

$$\mathbf{A} = \begin{bmatrix} 1.1547 & -1.1547 \\ -1.0774 & 0.0774 \\ -0.0774 & 1.0774 \end{bmatrix}$$

peut s'exprimer comme :

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\mathbf{\Delta}\mathbf{V}^\top \\ &= \begin{bmatrix} 0.8165 & 0 & u_1 \\ 0.4083 & -0.7071 & u_2 \\ 0.4083 & 0.7071 & u_3 \end{bmatrix} \times \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} -0.7071 & 0.7071 \\ 0.7071 & 0.7071 \end{bmatrix} \\ &= \begin{bmatrix} 1.1547 & -1.1547 \\ -1.0774 & 0.0774 \\ -0.0774 & 1.0774 \end{bmatrix} \end{aligned}$$

où  $u_i$  représente une valeur réelle quelconque. On peut vérifier que :

$$\begin{aligned} \mathbf{A}\mathbf{A}^\top &= \begin{bmatrix} 0.8165 & 0 & u_1 \\ 0.4083 & -0.7071 & u_2 \\ 0.4083 & 0.7071 & u_3 \end{bmatrix} \times \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 0.8165 & 0.4083 & 0.4083 \\ 0 & -0.7071 & 0.7071 \\ u_1 & u_2 & u_3 \end{bmatrix} \\ &= \begin{bmatrix} 2.6666 & -1.3333 & -1.3333 \\ -1.3333 & 1.1668 & 0.1668 \\ -1.3333 & 0.1668 & 1.1668 \end{bmatrix} = \mathbf{U}\mathbf{\Delta}\mathbf{\Delta}^\top\mathbf{U}^\top \end{aligned}$$

et que :

$$\begin{aligned} \mathbf{A}^\top\mathbf{A} &= \begin{bmatrix} -0.7071 & 0.7071 \\ 0.7071 & 0.7071 \end{bmatrix} \times \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} -0.7071 & 0.7071 \\ 0.7071 & 0.7071 \end{bmatrix} \\ &= \begin{bmatrix} 2.5 & -1.5 \\ -1.5 & 2.5 \end{bmatrix} = \mathbf{V}\mathbf{\Delta}^\top\mathbf{\Delta}\mathbf{V} \end{aligned}$$

### 2.3.2 Exercices

---

EXERCICE 2.5 Utiliser la DVS pour calculer l'inverse d'une matrice régulière.

Si au moins une de valeurs singulières est nulle, que peut-on conclure pour l'inversion de la matrice ?

EXERCICE 2.6 Soit la matrice  $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ . Calculer sa DVS.

### 2.4 Bases de l'image et du noyau

---

Soit la matrice  $\mathbf{A} \in \mathbb{R}^{m \times n}$  et la DVS  $\mathbf{A} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^\top$  avec colonnes  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m]$  et  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ . Si le rang de  $\mathbf{A}$  est  $r$ , alors nous avons que les bases de sous-espaces  $R(\mathbf{A})$  et  $N(\mathbf{A})$  sont données par

$$R(\mathbf{A}) = \text{vect} \{ \mathbf{u}_1, \dots, \mathbf{u}_r \}$$

$$N(\mathbf{A}) = \text{vect} \{ \mathbf{v}_{r+1}, \dots, \mathbf{v}_n \}$$

COROLLAIRE 2.4.1 Si la matrice carrée  $\mathbf{A} \in \mathbb{R}^{n \times n}$  est régulière, alors  $\sigma_i > 0 \forall i = 1, \dots, n$ .

Le théorème suivant établit une relation entre norme matricielle et valeurs singulières.

THÉORÈME 2.4.1 Soit  $\mathbf{A} \in \mathbb{R}^{m \times n}$  et la DVS  $\mathbf{A} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^\top$  avec valeurs singulières  $\sigma_1 \geq \dots \geq \sigma_r > 0$ . Alors on a

$$\|\mathbf{A}\| = \sigma_1$$

Si  $m = n$  et  $\mathbf{A}$  matrice régulière, alors

$$\|\mathbf{A}^{-1}\| = \sigma_n$$

## 2.5 Distance d'une matrice régulière de la singularité

Soit  $\mathbf{A} \in \mathbb{R}^{m \times n}$  avec valeurs singulières  $\sigma_1 \geq \dots \geq \sigma_r > 0$  où  $r$  le rang de la matrice. Nous avons

$$\sigma_1 = \|\mathbf{A}\|_2$$

Cette propriété permet d'avoir des informations sur le nombre condition. En effet soit  $\mathbf{A} \in \mathbb{R}^{n \times n}$  matrice carrée régulière avec valeurs singulières  $\sigma_1 \geq \dots \geq \sigma_n > 0$ . On a

$$\sigma_n = 1/\|\mathbf{A}^{-1}\|_2$$

car la matrice  $\mathbf{A}^{-1}$  a comme valeurs singulières  $1/\sigma_i$ , dans la mesure où ses valeurs propres sont les valeurs propres inverses de  $\mathbf{A}$ .

Nous avons donc pour le conditionnement de la matrice  $\mathbf{A}$  la relation suivante

$$\kappa(\mathbf{A}) = \frac{\sigma_1}{\sigma_n} \quad (2.5.1)$$

Nous allons maintenant évaluer la "distance" qui sépare une matrice régulière  $\mathbf{A}$  de la singularité, c'est-à-dire évaluer la matrice  $\mathbf{E}$  avec la plus petite norme et qui, rajoutée à  $\mathbf{A}$ , la rend singulière. Comme  $\mathbf{A} + \mathbf{E}$  est singulière il existe un vecteur  $\mathbf{x} \neq \mathbf{0}$  tel que  $(\mathbf{A} + \mathbf{E})\mathbf{x} = \mathbf{0}$  d'où  $\|\mathbf{x}\|_2 = \|\mathbf{A}^{-1}\mathbf{E}\mathbf{x}\|_2 \leq \|\mathbf{A}^{-1}\|_2 \|\mathbf{E}\|_2 \|\mathbf{x}\|_2$  et finalement  $\|\mathbf{E}\|_2 \geq 1/\|\mathbf{A}^{-1}\|_2$ . Par conséquent, on a

$$\min_{\mathbf{E}} \{ \|\mathbf{E}\|_2 / \mathbf{A} + \mathbf{E} \text{ singulière} \} \geq \frac{1}{\|\mathbf{A}^{-1}\|_2}$$

La distance de  $\mathbf{A}$  de la singularité est donc

$$\sigma_n = \min_{\mathbf{E}} \{ \|\mathbf{E}\|_2 / \mathbf{A} + \mathbf{E} \text{ singulière} \}$$

et la distance relative

$$\frac{\sigma_n}{\sigma_1} = \min_{\mathbf{E}} \left\{ \frac{\|\mathbf{E}\|_2}{\|\mathbf{A}\|_2} / \mathbf{A} + \mathbf{E} \text{ singulière} \right\}$$

En ce qui concerne l'erreur de calcul des valeurs singulières, nous avons le théorème suivant :

**THÉORÈME 2.5.1** Soient  $\mathbf{A}, \mathbf{A} + \Delta\mathbf{A} \in \mathbb{R}^{m \times n}$ , où  $\Delta$  perturbation de  $\mathbf{A}$ ,  $r = \text{rang } \mathbf{A} = \text{rang } \mathbf{A} + \Delta\mathbf{A} \leq \min\{n, m\}$ ,  $\sigma_1 \geq \dots \geq \sigma_r > 0$  valeurs singulières de  $\mathbf{A}$  et  $\tau_1 \geq \dots \geq \tau_r > 0$  valeurs singulières de  $\mathbf{A} + \Delta\mathbf{A}$ . Nous avons

$$|\sigma_k - \tau_k| \leq \|\Delta\mathbf{A}\|_2; k = 1, \dots, r \quad (2.5.2)$$

et

$$\left( \sum_{i=1}^r (\sigma_i - \tau_i)^2 \right)^{1/2} \leq \|\Delta\mathbf{A}\|_2 \quad (2.5.3)$$

À cause de la relation (2.5.2) il n'est pas recommandé de calculer le rang d'une matrice en utilisant le nombre de valeurs singulières non nulles car la valeur singulière perturbée correspondant à une valeur singulière non perturbée nulle, peut, en module, être égale à  $\|\Delta\mathbf{A}\|_2$ . En Scilab on peut utiliser la fonction `rank(A)` qui calcule le rang de la matrice  $\mathbf{A}$  en utilisant les valeurs singulières mais en considérant une valeur singulière comme nulle si elle est inférieure en module de  $m \times \text{eps} \times \|\mathbf{A}\|_2$ .

### 2.5.1 Exercices

EXERCICE 2.7 Considérons la matrice

$$\mathbf{A} = \begin{bmatrix} 1/3 & 1/3 & 2/3 \\ 2/3 & 2/3 & 1/3 \\ 1/3 & 2/3 & 1.000000001 \\ 0.4 & 0.4 & 0.8 \\ 0.6 & 0.2 & 0.8 \end{bmatrix}$$

et la matrice perturbée

$$\mathbf{B} = \mathbf{A} + \Delta\mathbf{A} = \begin{bmatrix} 1/3 & 1/3 & 2/3 \\ 2/3 & 2/3 & 4/3 \\ 1/3 & 2/3 & 1 \\ 2/5 & 2/5 & 4/5 \\ 3/5 & 1/5 & 4/5 \end{bmatrix}$$

Manifestement la matrice perturbée est de rang 2 parce que la troisième colonne est la somme de deux premières colonnes.

- (1) En utilisant Scilab vérifier le rang des matrices  $\mathbf{A}$  et  $\mathbf{B}$ .
- (2) Comparer les résultats avec les nombre de valeurs singulières, en utilisant la fonction de Scilab `svd`.
- (3) Examiner si la formule (2.5.2) est vérifiée.

EXERCICE 2.8 Soit une matrice  $\mathbf{A} \in \mathbb{R}^{m \times n}$  avec  $r = \text{rang } \mathbf{A} < \min\{n, m\}$ . Montrer que pour tout  $\varepsilon > 0$ , il existe une matrice  $\mathbf{A}_\varepsilon \in \mathbb{R}^{m \times n}$  de rang plein, c'est-à-dire  $\text{rang } \mathbf{A}_\varepsilon = \min\{n, m\}$  telle que  $\|\mathbf{A} - \mathbf{A}_\varepsilon\| = \varepsilon$ .

## 2.6 Applications de la DVS

La DVS est une technique largement utilisée par des diverses applications. Nous présentons dans la suite trois de ces applications.

### 2.6.1 Filtrage du bruit des images et compression des données

Une image numérique est considérée comme un tableau bi-dimensionnel  $\mathbf{A}$  de pixels. Supposons que cette image contient du bruit qu'on considère comme un processus aléatoire de moyenne nulle. On effectue la DVS de  $\mathbf{A}$  selon (2.3.1) qui s'écrit

$$\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^\top \quad (2.6.1)$$

avec  $\sigma_1 \geq \dots \geq \sigma_r > 0$  les valeurs singulières de  $\mathbf{A}$ . Il s'agit donc de la décomposition de l'image en une somme de  $r$  images constituées par les produits  $\mathbf{u}_i \mathbf{v}_i^\top$ ;  $i = 1, \dots, r$ , chacune de ces images étant pondérée par la valeur singulière correspondante  $\sigma_i$ . Si le bruit est uniformément repartit à chaque image élémentaire  $\mathbf{u}_i \mathbf{v}_i^\top$ , alors les images pondérées par des petites valeurs singulières contiennent proportionnellement plus de bruit que les autres. Par conséquent nous pouvons tronquer la somme (2.6.1) à un indice  $p < r$  si  $\sigma_p \gg \sigma_r$ . On obtient donc l'image

$$\mathbf{A}_p = \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

dans laquelle on espère avoir éliminé une partie du bruit.

Si l'image n'a pas de bruit, alors la matrice  $\mathbf{A}_p$  contient la meilleure approximation de la matrice  $\mathbf{A}$  en utilisant  $p$  valeurs singulières. Elle constitue donc une compression optimale (non réversible) d'une image en vertu du théorème suivant :

**THÉORÈME 2.6.1** Soit  $\mathbf{A} \in \mathbb{R}^{n \times m}$  avec rang  $r$  DVS  $\mathbf{A} = \mathbf{U} \mathbf{\Delta} \mathbf{V}^\top$ . Pour un naturel  $p < r$  la matrice  $\mathbf{A}_p$  de rang  $p$ , qui minimise la norme de Frobenius  $\|\mathbf{A} - \mathbf{A}_p\|_F$  par rapport à toute matrice  $\mathbf{B} \in \mathbb{R}^{n \times m}$  et de rang  $p$  est formée par les  $p$  premiers vecteurs singuliers de  $\mathbf{A}$  :

$$\mathbf{A}_p = \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

De plus on a  $\|\mathbf{A} - \mathbf{A}_p\|_F^2 = \sum_{i=p+1}^r \sigma_i^2$ .

### 2.6.2 Analyse canonique

La visualisation à l'aide des plans successifs à deux dimensions d'un nuage de points de  $\mathbb{R}^n$  avec  $n > 3$  pose des sérieux problèmes d'interprétation en ce qui concerne la structure du nuage des points. Une idée est de projeter l'ensemble de points dans un espace de dimension inférieure sous la condition que le nuage subit une déformation minimale. La solution à ce problème est donnée par la *transformation canonique* de Hotelling<sup>4</sup>. Il s'agit de procéder à la décomposition en valeurs singulières de la matrice  $\mathbf{A}$  dans laquelle sont stockées les coordonnées des points du nuage. La déformation minimale du nuage s'obtient si on projette les points dans l'espace qui a comme

4. En traitement du signal cette transformation s'appelle *transformation de Karhunen-Loève*.

base les vecteurs singuliers de la DVS. Chaque plan est formé par deux vecteurs singuliers, par exemple  $\mathbf{u}_i$  et  $\mathbf{u}_j$ . L'information apportée par ce plan est égale à  $\frac{\sigma_i + \sigma_j}{\sigma_{TOT}}$  avec  $\sigma_{TOT} = \sum_{i=1}^r \sigma_i$ .

Nous allons maintenant montrer que nous pouvons utiliser pour la SVD, seulement  $r$  vecteurs propres des matrices  $\mathbf{A}\mathbf{A}^\top$  et  $\mathbf{A}^\top\mathbf{A}$ , où  $r$  le rang de la matrice  $\mathbf{A}$ . On construit à cet effet la matrice

$$\mathbf{S} = \begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^\top & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(m+n) \times (m+n)} \quad (2.6.2)$$

Soit  $\mathbf{w}_i \in \mathbb{R}^{(m+n)}$  un vecteur propre de  $\mathbf{S}$ . Nous pouvons le décomposer en deux vecteurs  $\mathbf{w}_i = \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}$  avec  $\mathbf{u}_i \in \mathbb{R}^m$  et  $\mathbf{v}_i \in \mathbb{R}^n$ . Ainsi le problème du calcul des vecteurs propres de  $\mathbf{S}$  se décompose en un couple de problèmes des vecteurs propres :

$$\begin{aligned} \mathbf{A}\mathbf{v}_i &= \lambda_i \mathbf{u}_i \\ \mathbf{A}^\top \mathbf{u}_i &= \lambda_i \mathbf{v}_i \end{aligned} \quad (2.6.3)$$

En multipliant la première relation par  $\mathbf{A}^\top$  et la seconde par  $\mathbf{A}$  nous obtenons

$$\begin{aligned} \mathbf{A}^\top \mathbf{A}\mathbf{v}_i &= \lambda_i^2 \mathbf{v}_i \\ \mathbf{A}\mathbf{A}^\top \mathbf{u}_i &= \lambda_i^2 \mathbf{u}_i \end{aligned} \quad (2.6.4)$$

$\mathbf{U}$  et  $\mathbf{V}$  sont les matrices dont les colonnes sont les vecteurs propres  $\mathbf{u}_i$  et  $\mathbf{v}_i$  respectivement. Notons par  $\mathbf{U}_r$  et  $\mathbf{V}_r$  les matrices dont les colonnes sont les vecteurs propres  $\mathbf{u}_i$  et  $\mathbf{v}_i$  qui correspondent aux  $r$  valeurs propres non nulles. De même notons par  $\mathbf{U}_0$  et  $\mathbf{V}_0$  les matrices dont les colonnes sont les vecteurs propres  $\mathbf{u}_i$  et  $\mathbf{v}_i$  qui correspondent aux valeurs propres nulles. Notons encore par  $\mathbf{\Delta}_r = \text{diag}(\lambda_1, \dots, \lambda_r)$ . Nous avons

$$\mathbf{A}\mathbf{V}_r = \mathbf{U}_r \mathbf{\Delta}_r, \quad \mathbf{A}^\top \mathbf{U}_r = \mathbf{V}_r \mathbf{\Delta}_r, \quad \mathbf{A}\mathbf{V}_0 = \mathbf{0}, \quad \mathbf{A}^\top \mathbf{U}_0 = \mathbf{0} \quad (2.6.5)$$

Du fait que  $\mathbf{U}\mathbf{U}^\top = \mathbf{U}^\top\mathbf{U} = \mathbf{I}_m$  et  $\mathbf{V}\mathbf{V}^\top = \mathbf{V}^\top\mathbf{V} = \mathbf{I}_n$ , nous avons  $\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Delta}$  et par conséquent :

$$\mathbf{A} = [\mathbf{U}_r, \mathbf{U}_0] \begin{bmatrix} \mathbf{\Delta}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_r^\top \\ \mathbf{V}_0^\top \end{bmatrix} = \mathbf{U}_r \mathbf{\Delta}_r \mathbf{V}_r^\top \quad (2.6.6)$$

qui est la décomposition en valeurs singulières en utilisant les  $r$  vecteurs propres des matrices  $\mathbf{A}\mathbf{A}^\top$  et  $\mathbf{A}^\top\mathbf{A}$  correspondants aux valeurs propres non nulles.

Cette démarche conduit à l'analyse canonique de Hotelling (1936) qui peut se présenter de la façon suivante. Supposons que les valeurs singulières non nulles de la matrice  $\mathbf{A}$  sont rangées par ordre décroissant :  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ . Si à partir d'un indice  $p < r$  ces valeurs  $\sigma_{p+1}, \sigma_{p+2}, \dots \geq \sigma_r$  sont négligeables par rapport aux précédentes, nous pouvons avoir une approximation assez fidèle de la matrice  $\mathbf{A}$  en utilisant  $p < r$  vecteurs propres à la place de  $r$  vecteurs propres :

$$\hat{\mathbf{A}} = \mathbf{U}_p \mathbf{\Delta}_p \mathbf{V}_p^\top$$

L'analyse canonique est à la base des méthodes de l'analyse de données et on la retrouve aussi au codage et à la reconstitution d'images.

### 2.6.3 Moteur de recherche dans un ensemble des documents

Pour retrouver des documents qui filtrent une question (par exemple des pages sur internet), on peut utiliser la méthode Latent Semantic Analysis de Deerwester et al. Pour cela il faut d'abord avoir indexer les documents  $D_1, \dots, D_n$  selon un ensemble de termes  $T_1, \dots, T_m$ . Ainsi pour un document  $D_j$ , on construit le vecteur  $\mathbf{d}_j = [d_{1j}, \dots, d_{mj}]^\top$ . Ces vecteurs forment la matrice  $\mathbf{A} = [\mathbf{d}_1, \dots, \mathbf{d}_n] \in \mathbb{R}^{m \times n}$ . Une question est un vecteur  $\mathbf{q} = [q_1, \dots, q_m]^\top$  avec

$q_i = \begin{cases} 1, & \text{si le terme } T_i \text{ fait partie de la question} \\ 0, & \text{sinon} \end{cases}$ . Pour mesurer la similarité de la question avec

un document  $\mathbf{d}_j$ , on calcula la valeur de

$$\cos \theta_j = \frac{\mathbf{q}^\top \mathbf{d}_j}{\|\mathbf{q}\|_2 \|\mathbf{d}_j\|_2} = \frac{\mathbf{q}^\top \mathbf{A} \mathbf{e}_j}{\|\mathbf{q}\|_2 \|\mathbf{A} \mathbf{e}_j\|_2}$$

et en fonction de la valeur de  $\cos \theta_j$  on considère que le document  $\mathbf{d}_j$  est pertinent ou non pour la question  $\mathbf{q}$ . À cause des ambiguïtés de la langue naturelle, des imprécisions de l'indexation et d'autres causes encore, il y a une quantité non négligeable de bruit dans la matrice  $\mathbf{t}$ . Ainsi il convient de filtrer la réponse afin d'éliminer une partie du bruit. En supposant que  $\text{rang } \mathbf{A} = r$ , on procède en une DVS en utilisant  $p < r$  termes, à savoir

$$\mathbf{A}_p = \mathbf{U}_p \mathbf{D}_p \mathbf{V}_p^\top = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

Dans ce cas, à la place de  $\cos \theta_j$ , on calcule  $\cos \phi_j$  :

$$\cos \phi_j = \frac{\mathbf{q}^\top \mathbf{A}_p \mathbf{e}_j}{\|\mathbf{q}\|_2 \|\mathbf{A}_p \mathbf{e}_j\|_2}$$

Si on pose  $\mathbf{S}_p = \mathbf{D}_p \mathbf{V}_p^\top = [\mathbf{s}_1, \dots, \mathbf{s}_p]$ , on a  $\|\mathbf{A}_p \mathbf{e}_j\|_2 = \|\mathbf{U}_p \mathbf{D}_p \mathbf{V}_p^\top \mathbf{e}_j\|_2 = \|\mathbf{U}_p \mathbf{S}_p \mathbf{e}_j\|_2 = \|\mathbf{U}_p \mathbf{s}_j\|_2 = \|\mathbf{s}_j\|_2$  et, par conséquent,

$$\cos \phi_j = \frac{\mathbf{q}^\top \mathbf{U}_p \mathbf{s}_j}{\|\mathbf{q}\|_2 \|\mathbf{s}_j\|_2}$$

Dans cette formule les matrices  $\mathbf{U}_p$  et  $\mathbf{S}_p$ , et donc le produit  $\mathbf{U}_p \mathbf{S}_p$  sont calculés une seule fois, ce qui fait que le calcul de  $\cos \phi_j; j = 1, \dots, n$  est rapide.

## 2.7 Références

Les livres consultés lors de la rédaction de ce chapitre sont les suivants :

R. A. HORN, C. R. JOHNSON : *Matrix analysis*, Cambridge Un. Press, 1990

I. C. F. IPSEN : *Numerical matrix analysis*, SIAM 2009

E. ISAACSON, H. B. KELLER : *Analysis of numerical methods*, Wiley, 1966

C. L. LAWSON, R. J. HANSON : *Solving least squares problems*, Prentice-Hall, 1974

H. LÜTKEPOHL : *Handbook of matrices*, Wiley, 1996

- W. PRESS et al. : *Numerical recipies in Fortran*, Second edition, Cambridge Un.P., 1992
- G. W. STEWART : *Matrix algorithms, Vol. II : Eigensystems*, SIAM, 2001
- J. STOER, R. BULIRSCH : *Introduction to numerical analysis*, Second ed., Springer-Verlag, 1992
- A. THIKHONOV, V. ARSENINE : *Méthode de résolution des problèmes mal posés*, MIR, 1976
- L. N. TREFETHEN, D. BAU III : *Numerical linear algebra*, SIAM, 1997
- D. S. WATKINS : *Fundamentals of matrix computations*, Second edition, Wiley, 2002

# 3

## PSEUDOÏNVERSE ET MOINDRES CARRÉS

---

1.1	Introduction	2
1.1.1	Exercices	5
1.2	Calcul des valeurs et vecteurs propres	5
1.2.1	Calcul par factorisation <b>LU</b>	8
1.3	Localisation des valeurs propres	10
1.3.1	Exercices	11
1.4	Estimation des erreurs en cas des perturbations	12
1.5	Références	14

---

### 3.1 Pseudoïinverse d'une matrice quelconque

---

La notion de *l'inverse généralisée* est une notion essentielle pour la mise en place de certaines méthodes en statistique descriptive ou pour l'étude des réseaux de neurones. Comme son nom l'indique, elle permet de généraliser la notion d'inverse d'une matrice à des cas pour lesquels la définition classique ne peut s'appliquer. Rappelons que l'inverse d'une matrice n'est définie que pour les matrices carrées et de rang plein, ou, autrement dit, l'inverse d'une matrice carrée n'existe que si les lignes et les colonnes de cette matrice sont linéairement indépendantes. Lorsqu'elle existe, cette inverse est unique.

Le problème de généraliser la notion d'inverse à l'ensemble des matrices, carrées ou non, est apparu dans le cadre de la résolution de systèmes linéaires, et en statistique, dans le cadre de l'approximation par les moindres carrés de systèmes d'équations. Moore en 1920, et plus tard indépendamment Penrose ont proposé une généralisation connue sous le nom de *pseudoïinverse* ou *inverse de Moore-Penrose*. La pseudoïinverse de la matrice  $\mathbf{A}$  sera notée  $\mathbf{A}^+$  (on rencontre aussi les notations  $\mathbf{A}^\dagger$ ,  $\mathbf{A}^\ddagger$  ou même  $\mathbf{A}^-$  mais cette dernière notation est à éviter pour des raisons qui deviendront claires plus loin). Elle vérifie les propriétés suivantes :

$$\begin{aligned}
(i) \quad & \mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} \\
(ii) \quad & \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ \\
(iii) \quad & (\mathbf{A}\mathbf{A}^+)^\top = \mathbf{A}\mathbf{A}^+ \\
(iv) \quad & (\mathbf{A}^+\mathbf{A})^\top = \mathbf{A}^+\mathbf{A}
\end{aligned}$$

Cette définition de la pseudoinverse entraîne son unicité.

Nous allons maintenant examiner l'utilisation de la pseudoinverse pour la résolution des systèmes linéaires généralisés. Soit un tel système :

$$\mathbf{A}\mathbf{x} = \mathbf{b} ; \mathbf{A} \in \mathbb{R}^{m \times n}; \mathbf{x}, \mathbf{b} \in \mathbb{R}^n \quad (\text{P})$$

On cherche toujours à calculer  $\mathbf{x}$ , malgré le fait que la matrice  $\mathbf{A}$  n'est pas carrée. Nous considérons encore ici la matrice  $\mathbf{A}$  comme une application  $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  avec les deux projections  $\mathbf{P}_{R(\mathbf{A}^\top)}$  et  $\mathbf{P}_{R(\mathbf{A})}$ .

Pour résoudre le problème (P) il faut pouvoir établir une application

$$\mathbf{A}^+ : \mathbb{R}^m \rightarrow \mathbb{R}^n \quad (3.1.1)$$

telle que pour tout  $\mathbf{b} \in \mathbb{R}^m$  nous pouvons construire, à l'aide de  $\mathbf{A}^+$ , un  $\mathbf{x} \in \mathbb{R}^n$  avec  $\mathbf{A}\mathbf{x} = \mathbf{b}$ . Pour ce faire on construit une application composée

$$\mathbf{A}^+ = \mathbf{A}_{N(\mathbf{A})^\perp}^{-1} \mathbf{P}_{R(\mathbf{A})} : \mathbb{R}^m \rightarrow N(\mathbf{A})^\perp = R(\mathbf{A}^\top) \subseteq \mathbb{R}^n \quad (3.1.2)$$

définie par

$$\mathbf{A}^+\mathbf{y} = \begin{cases} \mathbf{A}_{N(\mathbf{A})^\perp}^{-1}\mathbf{y}, & \text{si } \mathbf{y} \in R(\mathbf{A}) \\ \mathbf{0}, & \text{sinon} \end{cases} \quad (3.1.3)$$

$\mathbf{A}^+$  est l'inverse généralisée ou pseudoinverse de la matrice  $\mathbf{A}$ .  $\mathbf{A}^+$  satisfait aux conditions de Moore-Penrose.

Le construction pratique de cette pseudoinverse se fait comme suit. On considère un vecteur  $\mathbf{x} \in N(\mathbf{A})^\perp$  et aussi un vecteur  $\mathbf{b} \in R(\mathbf{A})$ , plutôt que  $\mathbf{b} \in \mathbb{R}^m$ . Ainsi nous pouvons déjà établir l'application

$$\mathbf{A}_{N(\mathbf{A})^\perp}^{-1} : R(\mathbf{A}) \rightarrow N(\mathbf{A})^\perp \quad (3.1.4)$$

définie par  $R(\mathbf{A}) \ni \mathbf{b} \mapsto f(\mathbf{b}) \in N(\mathbf{A})^\perp$ . Par conséquent

$$\forall \mathbf{b} \in R(\mathbf{A}) \text{ tel que } \mathbf{A}_{N(\mathbf{A})^\perp}^{-1}(\mathbf{b}) \in N(\mathbf{A})^\perp \text{ on a } \mathbf{A}\mathbf{A}_{N(\mathbf{A})^\perp}^{-1}(\mathbf{b}) = \mathbf{b} \quad (3.1.5)$$

Considérons un  $\mathbf{b} \in R(\mathbf{A})$  et posons  $\mathbf{x} = \mathbf{A}_{N(\mathbf{A})^\perp}^{-1}(\mathbf{b})$ . Alors

$\mathbf{b} = \mathbf{A}\mathbf{x} = \mathbf{A} \left( \mathbf{P}_{R(\mathbf{A}^\top)} + \left( \mathbf{I} - \mathbf{P}_{R(\mathbf{A}^\top)} \right) \right) \mathbf{x} = \mathbf{A}\mathbf{P}_{R(\mathbf{A}^\top)}\mathbf{x} + \mathbf{A} \left( \mathbf{I} - \mathbf{P}_{R(\mathbf{A}^\top)} \right) \mathbf{x}$ . Comme  $\mathbf{P}_{R(\mathbf{A}^\top)}\mathbf{x} \in N(\mathbf{A})^\perp$ , alors  $\mathbf{P}_{R(\mathbf{A}^\top)}\mathbf{x} \neq \mathbf{0}$ . Donc  $\left( \mathbf{I} - \mathbf{P}_{R(\mathbf{A}^\top)} \right) \mathbf{x} \in N(\mathbf{A}) \Rightarrow \left( \mathbf{I} - \mathbf{P}_{R(\mathbf{A}^\top)} \right) \mathbf{x} = \mathbf{0}$ . Par conséquent  $\mathbf{b} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}_1$  où  $\mathbf{x}_1 = \mathbf{P}_{R(\mathbf{A}^\top)}\mathbf{x} \in N(\mathbf{A})^\perp$  projection de  $\mathbf{x}$  dans  $N(\mathbf{A})^\perp$ .

Notons que pour le même  $\mathbf{b}$  cette projection est unique, car si  $\mathbf{x}_1, \mathbf{x}_2 \in N(\mathbf{A})^\perp$  avec  $\mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{x}_2$ , alors  $\mathbf{x}_1 - \mathbf{x}_2 \in N(\mathbf{A}) \cap N(\mathbf{A})^\perp = \{\mathbf{0}\}$ .

L'extension de l'ensemble de définition de  $\mathbf{A}_{N(\mathbf{A})^\perp}^{-1}$  à tout l'espace  $\mathbb{R}^m$  peut se faire en utilisant la projection  $\mathbf{P}_{R(\mathbf{A})}$  car  $\forall \mathbf{b} \in \mathbb{R}^m : \mathbf{P}_{R(\mathbf{A})}\mathbf{b} \in R(\mathbf{A})$ . Par conséquent la pseudo-inverse  $\mathbf{A}^+$  est l'application composée  $\mathbf{A}_{N(\mathbf{A})^\perp}^{-1} \circ \mathbf{P}_{R(\mathbf{A})}$ . Nous avons donc

$$\forall \mathbf{b} \in \mathbb{R}^m : \mathbf{A}^+\mathbf{b} = \mathbf{A}_{N(\mathbf{A})^\perp}^{-1}(\mathbf{P}_{R(\mathbf{A})}(\mathbf{b})) \quad (3.1.6)$$

$\mathbf{A}^+$  est une matrice  $(n, m)$ .

En plus de l'inverse de Moore-Penrose, il existe d'autres inverses généralisées qui satisfont des conditions plus faibles. Pour résoudre des équations linéaires, on définit une classe d'inverses parfois appelée inverses généralisées, inverses de Rao, semi-inverses, etc. Ces matrices notées en général  $\mathbf{A}^-$  (mais les notations sont loin d'être établies dans ce domaine) vérifient la propriété (i) de la pseudo-inverse :

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A} \quad (3.1.7)$$

Il est évident, d'après la définition, que l'inverse de Moore-Penrose est une inverse de Rao, mais qu'il existe des inverses de Rao qui, ne satisfaisant que la condition (i), ne sont pas des inverses de Moore-Penrose.

### 3.1.1 Calcul de la pseudo-inverse à l'aide de la DVS

Pour calculer la pseudo-inverse d'une matrice nous pouvons utiliser sa décomposition en valeurs singulières. Précisément, la pseudo-inverse d'une matrice quelconque  $\mathbf{A} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^\top$  peut s'exprimer comme :

$$\mathbf{A}^+ = \mathbf{V}\mathbf{\Delta}^{-1}\mathbf{U}^\top \quad (3.1.8)$$

Si la matrice  $\mathbf{\Delta}$  est une matrice diagonale  $(m \times n)$ , alors la matrice  $\mathbf{\Delta}^{-1}$  est une matrice diagonale  $(n \times m)$  dont les éléments sont l'inverse des éléments de  $\mathbf{\Delta}$ .

EXEMPLE 3.1.1 Pour l'exemple 2.3.1 on obtient :

$$\begin{aligned} \mathbf{A}^+ &= \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \times \begin{bmatrix} 2^{-1} & 0 & 0 \\ 0 & 1^{-1} & 0 \end{bmatrix} \times \begin{bmatrix} 0.8165 & 0.4083 & 0.4083 \\ 0 & -0.7071 & 0.7071 \\ u'_1 & u'_2 & u'_3 \end{bmatrix} \\ &= \begin{bmatrix} 0.2887 & -0.6443 & 0.3557 \\ -0.2887 & -0.3557 & 0.6443 \end{bmatrix} \end{aligned}$$

Il est facile de vérifier que  $\mathbf{A}^+$  satisfait les quatre conditions de Moore-Penrose.

Comme nous avons vu au chapitre précédent, les colonnes de  $\mathbf{U}$  et de  $\mathbf{V}$  engendrent les quatre espaces fondamentaux. L'image de  $\mathbf{A}$  est engendré par les  $r$  premières colonnes de  $\mathbf{U}$ . Le noyau de  $\mathbf{A}$  est engendré par les  $n - r$  dernières colonnes de  $\mathbf{V}$ . De même les  $r$  premières colonnes de  $\mathbf{V}$  engendrent l'image de  $\mathbf{A}^\top$  et les  $m - r$  dernières colonnes de  $\mathbf{U}$  engendrent le noyau gauche de  $\mathbf{A}$ .

### 3.1.2 Calcul de la pseudoinverse pour des matrices perturbées

Le calcul de la pseudoinverse pose beaucoup de problèmes numériques dans la mesure où pseudoinverse n'a pas un comportement continu. Par exemple pour la matrice  $\mathbf{A}(x) = \begin{bmatrix} 1 & 0 \\ 0 & x \end{bmatrix}$  on a  $\mathbf{A}^+(x) = \begin{bmatrix} 1 & 0 \\ 0 & 1/x \end{bmatrix}$  et  $\mathbf{A}^+(0) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ . Donc le calcul de la pseudoinverse n'est pas stable, c'est-à-dire des petites variations des valeurs des éléments de la matrice  $\mathbf{A}$  peuvent provoquer des grandes variations des valeurs de  $\mathbf{A}^+$ . Ce type de problème s'appelle *problème mal posé*.

Nous pouvons généraliser la notion du conditionnement des matrices dans le cas des matrices rectangulaires, en utilisant la pseudoinverse. On a donc

$$\kappa^+(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A}^+\|$$

$\kappa^+(\mathbf{A})$  est le nombre-condition de  $\mathbf{A}$  pour l'inversion gauche.

PROPOSITION 3.1.1 Soit le système linéaire  $\mathbf{Ax} = \mathbf{b}$  avec  $\mathbf{A} \in \mathbb{R}^{m \times n}$  et la perturbation  $\mathbf{A}(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b}$ . L'erreur relative est bornée comme suit  $\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \kappa^+(\mathbf{A}) \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}$

### 3.1.3 Exercices

EXERCICE 3.1 Soit  $\mathbf{A}$  matrice  $(m, n)$ . La pseudoinverse  $\mathbf{A}^+$  est une matrice  $(n, m)$  qui satisfait aux propriétés suivantes :

- (1)  $\mathbf{A}^+\mathbf{A} = \mathbf{P}$  ;  $\mathbf{AA}^+ = \mathbf{Q}$
- (2)  $\mathbf{A}^+\mathbf{A} = (\mathbf{A}^+\mathbf{A})^\top$
- (3)  $\mathbf{AA}^+ = (\mathbf{AA}^+)^\top$
- (4)  $\mathbf{AA}^+\mathbf{A} = \mathbf{A}$
- (5)  $\mathbf{A}^+\mathbf{AA}^+ = \mathbf{A}^+$

EXERCICE 3.2 (Conditions de Moore-Penrose) Si  $\mathbf{Z}$  est une matrice telle que

- (1)  $\mathbf{ZA} = (\mathbf{ZA})^\top$
- (2)  $\mathbf{AZ} = (\mathbf{AZ})^\top$
- (3)  $\mathbf{AZA} = \mathbf{A}$
- (4)  $\mathbf{ZAZ} = \mathbf{Z}$

alors  $\mathbf{Z} = \mathbf{A}^+$ .

EXERCICE 3.3 Montrer que  $(\mathbf{A}^+)^+ = \mathbf{A}$  et  $(\mathbf{A}^+)^\top = (\mathbf{A}^\top)^+$ .

EXERCICE 3.4 Montrer que le vecteur  $\mathbf{x}^* = \mathbf{A}^+\mathbf{b}$  satisfait à

$$\|\mathbf{Ax} - \mathbf{b}\|_2 \geq \|\mathbf{Ax}^* - \mathbf{b}\|_2 \quad \forall \mathbf{x} \in \mathbb{R}^n$$

De plus

$$\text{si } \|\mathbf{Ax} - \mathbf{b}\|_2 = \|\mathbf{Ax}^* - \mathbf{b}\|_2 \text{ et } \mathbf{x}^* \neq \mathbf{x}, \text{ alors } \|\mathbf{x}\|_2 > \|\mathbf{x}^*\|_2$$

i.e.  $\mathbf{x}^*$  est le point de norme minimale qui minimise la norme  $\|\mathbf{Ax} - \mathbf{b}\|_2$ .

EXERCICE 3.5 Soit la matrice  $\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

- (1) Calculer les valeurs propres et les vecteurs propres des matrices  $\mathbf{AA}^\top$  et  $\mathbf{A}^\top\mathbf{A}$ .
- (2) Calculer les quatre espace fondamentaux associés à la matrice  $\mathbf{A}$ .
- (3) Calculer la pseudo-inverse de  $\mathbf{A}$  sans faire usage de la DVS.

### 3.2 Approximation linéaire par moindres carrés

Le problème de l'approximation linéaire de la valeur d'une grandeur physique  $y$  inaccessible est un problème important du calcul scientifique. Supposons que sa valeur dépend des valeurs des  $n$  autres variables  $x_1, \dots, x_n$  selon la relation

$$y = f(x_1, \dots, x_n) \quad (3.2.1)$$

où  $f$  fonction dont la forme nous est connue mais pas la valeur de ses paramètres.

Le problème de l'approximation des données est le suivant : Si nous avons réalisé  $m$  ( $m > n$ ) relevés de mesures  $y_k = f(x_{k1}, \dots, x_{kn})$ ,  $k = 1, \dots, m$ , on cherche à calculer les paramètres inconnus de la fonction  $f$ . La solution à ce problème est donnée par la minimisation de l'expression

$$\sum_{k=1}^m (y_k - f(x_{k1}, \dots, x_{kn}))^2 \quad (3.2.2)$$

Si la fonction  $f$  a de dérivées partielles continues par rapport à toutes les variables  $x_i$ , alors une condition nécessaire pour minimiser (3.2.2) est d'avoir :

$$\frac{\partial}{\partial x_i} \sum_{k=1}^m (y_k - f(x_{k1}, \dots, x_{kn}))^2 = 0; \quad i = 1, \dots, n \quad (3.2.3)$$

qui sont appelées *équations normales*. Si la fonction  $f$  est linéaire, c'est-à-dire si

$$f(x_{k1}, \dots, x_{kn}) = a_1 x_{k1} + \dots + a_n x_{kn}$$

alors nous avons :

$$\begin{bmatrix} f(x_{11}, \dots, x_{1n}) \\ \vdots \\ f(x_{m1}, \dots, x_{mn}) \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \mathbf{Xa} \quad (3.2.4)$$

et le problème (3.2.2) devient

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2 \quad (3.2.5)$$

et, par conséquent, les conditions nécessaires (3.2.3) s'écrivent :

$$\frac{\partial}{\partial \mathbf{a}} \left[ (\mathbf{y} - \mathbf{X}\mathbf{a})^\top (\mathbf{y} - \mathbf{X}\mathbf{a}) \right] = 2\mathbf{X}^\top \mathbf{X} - 2\mathbf{X}^\top \mathbf{y} = \mathbf{0} \quad (3.2.6)$$

d'où, finalement, nous avons les équations normales dans le cas linéaire :

$$\mathbf{X}^\top \mathbf{X} = \mathbf{X}^\top \mathbf{y} \quad (3.2.7)$$

Le théorème suivant fournit les conditions d'existence et d'unicité du problème des moindres carrés.

**THÉORÈME 3.2.1** *Le problème des moindres carrés (3.2.5) a au moins une solution  $\mathbf{a}_0$ . Si  $\mathbf{a}_1$  est une autre solution, alors  $\mathbf{X}\mathbf{a}_0 = \mathbf{X}\mathbf{a}_1$ . Le résidu  $\mathbf{r} = \mathbf{y} - \mathbf{X}\mathbf{a}_0$  est unique et il est solution de l'équation  $\mathbf{X}^\top \mathbf{r} = \mathbf{0}$ . Chaque point  $\mathbf{a}_0$  qui réalise le minimum de (3.2.5) est aussi solution du système d'équations normales, i.e.*

$$\mathbf{a}_0 = \left( \mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{y} = \mathbf{X}^+ \mathbf{y} \quad (3.2.8)$$

où  $\mathbf{X}^+$  la pseudoinverse de Moore-Penrose.

### 3.2.1 Conditionnement des moindres carrés

Le calcul de la solution d'un problème par la méthode des moindres carrés est beaucoup plus sensible aux perturbations que le calcul de la solution d'un système linéaire. La raison vient d'être invoquée : c'est le comportement non continu de la pseudoinverse. Pour éviter que le problème des moindres carrés soit mal posé, il faut imposer que le rang de la matrice  $\mathbf{X}$  soit égal au nombre  $n$  de ses colonnes. En effet, supposons que les observations  $\mathbf{y}$ , contiennent des erreurs  $\Delta\mathbf{y}$ . Supposons aussi que  $\hat{\mathbf{a}}$  soit la solution du problème  $\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2$  et  $\tilde{\mathbf{a}}$  celle du problème

$\min_{\mathbf{x} \in \mathbb{R}^n} \|(\mathbf{y} + \Delta\mathbf{y}) - \mathbf{X}\mathbf{a}\|_2$ . Si  $\hat{\mathbf{a}} \neq \mathbf{0}$ , alors

$$\frac{\|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\|_2}{\|\hat{\mathbf{a}}\|_2} \leq \kappa_2^+(\mathbf{X}) \frac{\|\Delta\mathbf{y}\|_2}{\|\mathbf{X}\|_2 \|\hat{\mathbf{a}}\|_2} \quad (3.2.9)$$

et si  $\tilde{\mathbf{a}} \neq \mathbf{0}$ , alors

$$\frac{\|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\|_2}{\|\tilde{\mathbf{a}}\|_2} \leq \kappa_2^+(\mathbf{X}) \frac{\|\Delta\mathbf{y}\|_2}{\|\mathbf{X}\|_2 \|\tilde{\mathbf{a}}\|_2} \quad (3.2.10)$$

On voit donc que l'erreur relative de la solution est amplifiée par le nombre condition  $\kappa_2^+(\mathbf{X})$  de la matrice  $(\mathbf{X})$ .

Supposons maintenant que les mesures de la matrice  $\mathbf{X}$  contiennent aussi des petites erreurs  $\Delta\mathbf{X}$ . On a que  $\text{rang } \mathbf{X} = \text{rang } (\mathbf{X} + \Delta\mathbf{X}) = n = \min\{m, n\}$ . Soit  $\tilde{\mathbf{a}}$  la solution de  $\min_{\mathbf{x} \in \mathbb{R}^n} \|(\mathbf{y} + \Delta\mathbf{y}) - (\mathbf{X} + \Delta\mathbf{X})\mathbf{a}\|_2$ ,

avec  $\tilde{\mathbf{a}} \neq \mathbf{0}$ . Alors

$$\frac{\|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\|_2}{\|\tilde{\mathbf{a}}\|_2} \leq \kappa_2^+(\mathbf{X}) \cdot \left( \frac{\|\Delta \mathbf{X}\|_2}{\|\mathbf{X}\|_2} + \frac{\|\Delta \mathbf{y}\|_2}{\|\mathbf{X}\|_2 \|\tilde{\mathbf{a}}\|_2} \right) + (\kappa_2^+(\mathbf{X}))^2 \cdot \frac{\|(\mathbf{X} + \Delta \mathbf{X})\tilde{\mathbf{a}} - (\mathbf{y} + \Delta \mathbf{y})\|_2}{\|\mathbf{X}\|_2 \|\tilde{\mathbf{a}}\|_2} \cdot \frac{\|\Delta \mathbf{X}\|_2}{\|\mathbf{X}\|_2} \quad (3.2.11)$$

Si  $\Delta \mathbf{X} = \mathbf{0}$ , le terme de droite de (3.2.11) est identique au terme de droite de (3.2.10). On peut donc en conclure que la méthode des moindres carrés est plus sensible aux perturbations de la matrice des données  $\mathbf{X}$  qu'aux perturbations des observations  $\mathbf{y}$ .

### 3.2.2 Exercice

EXERCICE 3.6 (PERTURBATION DE LA MATRICE DE DONNÉES  $\mathbf{X}$ ) Soit  $\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 0 & x \\ 0 & 0 \end{bmatrix}$ ,  $\mathbf{y} = \begin{bmatrix} y_1 \\ 0 \\ y_2 \end{bmatrix}$  avec

$x, y_1, y_2 > 0$ . Considérons la matrice perturbée  $\mathbf{X} + \Delta \mathbf{X} = \begin{bmatrix} 1 & 0 \\ 0 & x \\ 0 & \delta x \end{bmatrix}$ . Supposons que  $\hat{\mathbf{a}}$  est la solution des

moindres carrés non perturbés et  $\tilde{\mathbf{a}}$  la solution des moindres carrés perturbés. Montrer que

$$\frac{\|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\|_2}{\|\hat{\mathbf{a}}\|_2} \leq \frac{y_3}{a^2 y_1} \delta x$$

## 3.3 Problèmes inverses

Le problème des moindres carrés peut être vu de façon « inverse ». Nous pouvons ainsi chercher à estimer ou à reconstruire  $\mathbf{x} = [x_1, \dots, x_n]^\top$  à partir d'un vecteur d'observations  $\mathbf{y} = [y_1, \dots, y_m]^\top$ . La détermination d'un modèle  $\mathbf{x}$  s'appelle *problème inverse* et, dans notre cas, *problème inverse linéaire*, car il s'agit à inverser la matrice  $\mathbf{A}$ .

Si  $m < n$ , alors l'espace noyau n'est pas vide et, donc, il y a une infinité de modèles qui sont une solution exacte du problème. Dans le cas contraire ( $m > n$ ) il n'y a pas de solution exacte au problème (sauf si  $m - n$  équations sont linéairement dépendantes) et on trouve une solution en minimisant un critère prédéfini.

Dans le cas où on ne peut pas inverser  $\mathbf{A}$ , on parle, d'après J. Hadamard, des *problèmes mal posés*.

Pour caractériser un problème inverse nous utiliserons la DVS. Nous savons que la matrice  $\mathbf{A}$  peut être factorisée comme suit :

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^\top \quad (3.3.1)$$

Si  $r \leq \min\{m, n\}$  est le rang de la matrice  $\mathbf{A}$ , alors la matrice  $\mathbf{\Lambda}$  de format  $(m \times n)$  est composée comme suit

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.3.2)$$

avec  $\mathbf{\Lambda}_r = \text{diag}(\lambda_1, \dots, \lambda_r)$ . Par conséquent (3.3.1) s'écrit

$$\mathbf{A} = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{V}_r^\top \quad (3.3.3)$$

Nous cherchons maintenant à évaluer la signification des autres vecteurs propres lorsque  $r < \min\{m, n\}$ . Nous avons

$$\mathbf{y} = \mathbf{Ax} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top \mathbf{x} = [\mathbf{U}_r, \mathbf{U}_0] \begin{bmatrix} \mathbf{\Lambda}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [\mathbf{V}_r, \mathbf{V}_0]^\top \mathbf{x} \quad (3.3.4)$$

Si nous construisons un modèle, e.g.  $\mathbf{x}_0$  comme combinaison linéaire des vecteurs propres qui font partie seulement de  $\mathbf{V}_0$ , alors nous aurons  $\mathbf{y} = \mathbf{0}$ . Par conséquent

$$\mathbf{y} + \mathbf{0} = \mathbf{y} = \mathbf{Ax} + \mathbf{Ax}_0 = \mathbf{A}(\mathbf{x} + \mathbf{x}_0) \quad (3.3.5)$$

C'est justement l'existence des solutions du modèle dues à l'espace noyau qui provoque la non unicité de la solution.

Les problèmes inverses mal posés présentent une difficulté concernant le calcul de leur solution. En effet les vecteurs propres correspondants aux valeurs propres  $\lambda_1, \dots, \lambda_r$  sont de norme minimale. Mais comme le problème est mal posé, la solution n'est pas exacte, elle est soumise à des erreurs. On peut comprendre que plus la norme d'un vecteur est petite, plus l'erreur relative est grande. Donc la solution par moindres carrés du problème inverse peut présenter des incohérences physiques, bien que du point de vue mathématique, tout semble bien se passer. Ainsi pour résoudre un problème inverse mal posé il faut chercher une solution de norme faible mais pas minimale. Tikhonov a proposé une solution qui porte le nom de *régularisation* et qui est une extension sous-optimale des moindres carrés. On cherche à trouver  $\mathbf{x} \in \mathbb{R}^n$  qui, à la place de (3.2.5), minimise l'expression

$$\|\mathbf{Ax} - \mathbf{y}\|_2 + \alpha \|\mathbf{Lx}\|_2$$

où  $\alpha > 0$  est un nombre petit qui détermine l'*amplitude de la régularisation*. En suivant un raisonnement analogue à celui du § 3.2, on trouve comme solution

$$\mathbf{x} = (\mathbf{A}^\top \mathbf{A} + \mathbf{L}^\top \mathbf{L})^{-1} \mathbf{A}^\top \mathbf{y}$$

$\mathbf{L}$  doit être choisie de sorte que  $\mathbf{A}^\top \mathbf{A} + \mathbf{L}^\top \mathbf{L}$  soit bien conditionnée. Deux choix de  $\mathbf{L}$  sont assez souvent utilisés :

$$- \mathbf{L} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \text{ qui correspond à } \|\mathbf{x}\|_2 \text{ minimum et qui s'appelle } \textit{régularisation d'ordre 0},$$

et

$$- \mathbf{L} = \begin{bmatrix} 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \text{ qui correspond à } \|\mathbf{x}_i - x_{i+1}\|_2 \text{ minimum et qui s'appelle } \textit{régularisation d'ordre 1}.$$

Une autre technique de régularisation est la troncature de la DVS, c'est-à-dire l'analyse canonique.

### 3.4 Références

---

Les livres consultés lors de la rédaction de ce chapitre sont les suivants :

- R. A. HORN, C. R. JOHNSON : *Matrix analysis*, Cambridge Un. Press, 1990  
I. C. F. IPSEN : *Numerical matrix analysis*, SIAM 2009  
E. ISAACSON, H. B. KELLER : *Analysis of numerical methods*, Wiley, 1966  
C. L. LAWSON, R. J. HANSON : *Solving least squares problems*, Prentice-Hall, 1974  
H. LÜTKEPOHL : *Handbook of matrices*, Wiley, 1996  
W. PRESS et al. : *Numerical recipes in Fortran*, Second edition, Cambridge Un.P., 1992  
G. W. STEWART : *Matrix algorithms, Vol. II : Eigensystems*, SIAM, 2001  
J. STOER, R. BULIRSCH : *Introduction to numerical analysis*, Second ed., Springer-Verlag, 1992  
A. THIKHONOV, V. ARSENINE : *Méthode de résolution des problèmes mal posés*, MIR, 1976  
L. N. TREFETHEN, D. BAU III : *Numerical linear algebra*, SIAM, 1997  
D. S. WATKINS : *Fundamentals of matrix computations*, Second edition, Wiley, 2002



# Table des matières

<b>1</b>	<b>VALEURS ET VECTEURS PROPRES</b>	<b>1</b>
1.1	Introduction . . . . .	2
1.1.1	Exercices . . . . .	5
1.2	Calcul des valeurs et vecteurs propres . . . . .	5
1.2.1	Calcul par factorisation <b>LU</b> . . . . .	8
1.3	Localisation des valeurs propres . . . . .	10
1.3.1	Exercices . . . . .	11
1.4	Estimation des erreurs en cas des perturbations . . . . .	12
1.5	Références . . . . .	14
<b>2</b>	<b>DÉCOMPOSITION EN VALEURS SINGULIÈRES</b>	<b>15</b>
2.1	Étude géométrique des transformations linéaires . . . . .	15
2.1.1	Les espaces fondamentaux d'une application linéaire . . . . .	15
2.1.2	Espaces complémentaires et projections . . . . .	17
2.1.3	Décomposition orthogonale . . . . .	18
2.1.4	Exercices . . . . .	18
2.2	Étude géométrique des transformations linéaires . . . . .	18
2.2.1	Exercices . . . . .	21
2.3	Décomposition en valeurs singulières . . . . .	21
2.3.1	DVS réduite . . . . .	23
2.3.2	Exercices . . . . .	24
2.4	Bases de l'image et du noyau . . . . .	24
2.5	Distance d'une matrice régulière de la singularité . . . . .	25
2.5.1	Exercices . . . . .	26
2.6	Applications de la DVS . . . . .	26
2.6.1	Filtrage du bruit des images et compression des données . . . . .	27
2.6.2	Analyse canonique . . . . .	27
2.6.3	Moteur de recherche dans un ensemble des documents . . . . .	29
2.7	Références . . . . .	29
<b>3</b>	<b>PSEUDOÏNVERSE ET MOINDRES CARRÉS</b>	<b>31</b>
3.1	Pseudoïinverse d'une matrice quelconque . . . . .	31
3.1.1	Calcul de la pseudoïinverse à l'aide de la DVS . . . . .	33
3.1.2	Calcul de la pseudoïinverse pour des matrices perturbées . . . . .	34
3.1.3	Exercices . . . . .	34
3.2	Approximation linéaire par moindres carrés . . . . .	35
3.2.1	Conditionnement des moindres carrés . . . . .	36
3.2.2	Exercice . . . . .	37
3.3	Problèmes inverses . . . . .	37

3.4	Références . . . . .	39
-----	----------------------	----