

Aide mémoire d'analyse de données

B. Rousselet et J.P. Labrousse *

12 octobre 2007

* *Laboratoire de Mathématiques, Parc Valrose, F 06108 Nice, Cédex 2, email : br@math.unice.fr*

Table des matières

1	Orientation	4
1.1	Contenu	4
1.2	Méthodologie ; classe WIMS	4
1.3	Réponse numérique avec WIMS	5
2	Utilisation de scilab	6
2.1	Introduction	6
2.2	Mise en train	6
2.2.1	Les constantes	7
2.2.2	Fonctions usuelles	7
2.2.3	L'utilisation de matrices est très souple.	7
2.3	Programmation	8
2.3.1	Fonctions	8
2.3.2	Boucles	9
2.3.3	Conditions	10
2.3.4	Lectures dans fichiers	10
2.4	Les bibliothèques	10
2.5	Statistiques et analyse de données	10
2.6	Equations différentielles, optimisation	11
2.7	Divers	11
2.7.1	Résultat matrice	11
3	Equations, intersections, projections	12
3.1	Introduction	12
3.1.1	Références à utiliser	12
3.1.2	Guide d'activité	12
3.1.3	Prérequis	12
3.1.4	Méthodologie ; classe WIMS	14
3.2	Droite du plan	15
3.2.1	Droite du plan :	15
3.2.2	Équation paramétrique ou cartésienne (implicite) :	15
3.2.3	Droites parallèles, orthogonales	16
3.2.4	Projection orthogonale d'un point	16
3.2.5	De la géométrie au calcul :	17
3.2.6	Vecteur orthogonal à un vecteur donné par projection	18
3.3	Plan et droite dans l'espace tridimensionnel	18
3.3.1	Plan dans l'espace	18
3.3.2	Droite dans l'espace	20
3.3.3	Intersection de droite et plan	21
3.3.4	Projection sur un plan	22
3.3.5	Projection sur une droite	23
3.4	Droite et plan dans un espace de grande dimension	23
3.4.1	Vecteurs et matrices	23

3.4.2	Plan dans un espace de dimension $N > 3$	24
3.4.3	Droite dans un espace de dimension $N > 3$	24
3.4.4	Passage d'équation cartésienne à paramétrique et réciproquement	25
3.4.5	Projections orthogonales	26
4	Notations d'algèbre linéaire	28
5	Inertie d'un système	30
5.1	Définition, propriétés	30
5.2	Maximum, minimum de l'inertie	32
6	Analyse en composantes principales (A.C.P.)	34
6.1	Introduction	34
6.2	Pratique de l'ACP	34
6.3	Interprétation de l'ACP	35
6.3.1	Qualités de la représentation	35
6.3.2	Anciens et nouveaux caractères	35
6.4	Calcul de valeurs et vecteurs propres	36
7	Analyse factorielle des correspondances	37
7.1	Compléments d'algèbre linéaire	37
7.2	Notations, définitions	37
7.3	Résultats	38
7.4	Interprétation	38
8	Régression	40
8.1	La méthode	40
8.1.1	Introduction	40
8.1.2	Calcul	40
8.2	Introduction, suppression de variables	41
8.2.1	Introduction d'une variable supplémentaire	41
8.2.2	Suppression d'une variable	41
9	Devoirs	42

Chapitre 1

Orientation

1.1 Contenu

L'analyse de données est une branche de la statistique descriptive ; elle cherche à représenter de façon synthétique de grands tableaux de données. Si l'on fait remonter l'analyse factorielle aux psychologues du début du XXème siècle c' est dans la deuxième moitié de ce siècle que l' utilisation des ordinateurs a permis une utilisation commode de ces techniques. Le principal objectif est une représentation graphique ; *je choisis donc de faire une introduction à la géométrie affine euclidienne*, en particulier les projections orthogonales sur un sous-espace affine ; les concepts sont introduits dans un but opérationnel en vue du calcul numérique dans des espaces de grande dimension.

Principe général de la démarche en analyse de données

- Données : grands tableaux de nombre ;
- organisation en matrice rectangulaire ;
- a chaque colonne on associe un point d'un espace (affine) de grande dimension) ;
- transformation géométrique sur ces points (projection...) ;
- écriture vectorielle de la transformation ;
- écriture matricielle de la transformation ;
- écriture ou utilisation d'un programme informatique qui réalise les opérations matricielles (scilab,...)
- on obtient de petits tableaux ;
- on associe des points dans un espace affine de petite dimension (2,3...)
- dessin des points
- interprétation

Pour l' analyse de données, on pourra consulter : [3, 5, 2] ainsi que des sites internet : [6, 1]

1.2 Méthodologie ; classe WIMS

Cet enseignement est accompagné d'exercices à données aléatoires qui peuvent être réalisés avec le logiciel WIMS (wims.unice.fr). De nombreux exemples numériques sont proposés : leur solution peut s'obtenir à l'aide du logiciel de calcul numérique SCILAB (www.inria.fr,[8]), octave (www.octave.org, [11]) ou Maple (calcul symbolique avec extensions numériques). Le logiciel **WIMS** (wims.unice.fr, [12]) d'enseignement assisté par ordinateur permet d'entrainer les étudiants de façon systématique ; ce logiciel est du domaine public.

Dans la classe WIMS analyse de donnees, vous trouvez des exercices en rapport avec ce chapitre ; pour un usage indépendant, consultez l'auteur.

Aucun logiciel spécialisé en analyse de données n'est utilisé dans ce module qui vise à maîtriser les fondements des méthodes.

1.3 Réponse numérique avec WIMS

Pour une précision relative $\frac{1}{M}$, la réponse numérique r est acceptée pour la solution demandée s si et seulement si :

$$\frac{|s - r|}{\text{Max}(|s + r|, \frac{1}{M})} \leq \frac{1}{M}$$

en d'autres termes,

- si $|s + r| \leq \frac{1}{M}$ on doit verifier $|s - r| \leq \frac{1}{M^2}$
- si $|s + r| > \frac{1}{M}$ on doit verifier $\frac{|s - r|}{|s + r|} \leq \frac{1}{M}$

Chapitre 2

Utilisation de scilab

2.1 Introduction

Scilab est un logiciel intégré dédié au calcul numérique ; c'est un logiciel du domaine public ; il comprend à la fois un langage de programmation et des bibliothèques de sous programmes ; à noter qu'à la différence du **fortran** c'est un langage *interprété* ; l'utilisation est assez proche d'un système de calcul formel comme **maple** ou **maxima**. **Scilab** est développé par l'INRIA et est du domaine public : il peut être trouvé sur le site (téléchargement gratuit) : <http://www-rocq.inria.fr/scilab> ; on peut y trouver documentation et renseignements ; on peut aussi y accéder à partir du site www.inria.fr. Citons une référence pédagogique récente [4]. D'autres systèmes similaires sont **octave** (du domaine public) et **matlab** (commercial).

Utilisez de préférence un ordinateur utilisant le système linux bien que

scilab

fonctionne aussi avec le système windows. Le système linux est aussi du domaine public et peut être téléchargé gratuitement mais vous le trouvez aussi en vente, à un prix modique, avec des mensuels consacrés à Linux chez votre marchand de journaux !

2.2 Mise en train

Si **scilab** est dans vos onglets, cliquez sur l'icône, sinon mettez dans votre chemin d'accès (par exemple fichier `.zshenv` si vous utilisez le "z shell" avec linux) : `/usr/local/bin/scilab` ou un chemin adapté à votre ordinateur. Pour appeler **scilab**, tapez dans une console ou `xterm`

scilab&

Il s'ouvre une fenêtre dédiée à **scilab**, on peut y utiliser les commandes de base de l'éditeur de texte **EMACS** ; en particulier :

1. **ctrl b** (ou flèche vers gauche ←) pour se déplacer à gauche ;
2. **ctrl f** (ou →) pour aller à droite
3. **ctrl p** (ou touche flèche vers haut ↑) pour instruction précédente ;
4. **ctrl n** (ou ↓) pour retourner à l'instruction suivante

Les commandes sont séparées par la touche `'return'` ou par `'`, si elles sont tapées sur la même ligne ; une commande terminée par `'` ; ne donne aucune sortie de résultats à écran à l'exécution.

```
\\ marque le d'ebut d'un commentaire
```

Exercice 2.1 *taper quelques instructions simples*

2.2.1 Les constantes

```
%i; %pi; %e; %eps; // num\eriques
%t; %f // logiques
```

Exercice 2.2 *manipuler quelques constantes reelles et complexes*

2.2.2 Fonctions usuelles

```
abs(-5), x=10^(-4), sqrt(x), exp(3), %e^3, log(%e^3), sin(%pi)
```

2.2.3 L'utilisation de matrices est très souple.

```
n=10
x=[1:n]
size(x) // vecteur ligne (1 ligne et 10 colonnes)
//multiplication terme a terme
y=x.*x // pour tout i: y(i)=x(i)*x(i)
//produit scalaire
z=x' //vecteur colonne
px=z'*z // attention z*z' est une matrice carr\ee!
//transposition
x'
//saisie; voir aussi plus loin avec des fichiers
v=[1 +3]; w=[1,3]; ww=[1+ 3];
v=5:-.25:3;
v=ones(1:4); v=zeros(1:6)
a=[1 2; 6 -3]; b=ones(2,2); d=eye(3,4);
p=zeros(size(a,'c'),size(a,'c'));
//transformer matrice de nombres en matrice chaine de caracteres
as=string(a)

//Attention '*' designe le produit matriciel
a*b
//
g=sum(a,'c') // somme des colonnes de a
```

Attention, pour des matrices, `'*'` designe le produit matriciel

'/' entre matrices effectue un calcul mais pas toujours celui souhaite !

'./' effectue une division terme à terme et '.*' le produit terme à terme

Exercice 2.3 *Faire quelques manipulations de matrices*

2.3 Programmation

Attention, il est vivement déconseillé de taper un grand nombre d'instructions directement dans la fenetre **scilab** !

- taper les instructions avec l'éditeur de textes emacs ; *installez le mode scilab très commode* ; ou utilisez l'éditeur de texte intégré scilab.
- pour exécuter le fichier
`'exec('nom_fichier')'`
- Réaliser des fonctions ! voir plus bas.

Pour nettoyer la cession : `'clear'`.
Si le comportement de **scilab** est bizarre, quitter et relancer **scilab** !

Pour les matrices

- lecture :
`a=read ('plot.dat',3,5)` (3 lignes et 5 colonnes)
- écriture
`write('plot4.dat',b)`
- pour des fichiers sur lesquels on écrit plusieurs fois, il vaut mieux utiliser un numéro d'unité logique (sinon effacer le fichier avant toute écriture) :
`ua=file('open', status='old','plot4'); write(ua, a');`

Dialogues

```
b=read(%io(1),2,1)
exo=read(%io(1),1,1,'(a)')
write(%io(2),[exo, 'cas pas prévu'])
//afficher 2 variables x et y:
disp(x,y)
```

2.3.1 Fonctions

Les fonctions se mettent dans un fichier et se tapent avec un editeur de texte (emacs ...).

```
function [mu,sig2]=calcmoystd(xx)
mu=mean(xx)
sig2=st_deviation(xx)
//end function calcmoystd
```



```
//
function ydot=f(t,y)
yp(1)=y(3)
yp(2)=y(4)
r=sqrt(y(1)**2 + y(2)**2)
k=1; m=1;
yp(3)=-(k*(m/r)/r)*(y(1)/r)
yp(4)=-(k*(m/r)/r)*(y(2)/r)
ydot=yp
// end
```

Pour utiliser une fonction : on va chercher le fichier par :

```
getf('fonc-rand')
// appel fonction:
xx=rand(nnb,1,'uniform');
[m,s]=calcmoystd(xx)
```

Noter : pas de end en fin de fonction.
c'est différent d'autres langages.
Il est commode de le mettre en commentaire.

Exercice 2.4 *Utiliser ces fonctions; faire de petites modifications.*

Réaliser une fonction pour calculer la fonction caractéristique du complémentaire d'un cercle dans un carré; calculer espérance et écart type.

On peut aussi définir des fonctions simples en ligne :

```
deff(' [y]=fsol2(x)', 'y=a*x+b+1.e-3*x');
```

Mise au point, utiliser

```
//dans la fonction
pause
// a l'execution
return
//ou
resume
```

2.3.2 Boucles

```
ninterv=22
for i=1:ninterv ,
    l(i)=(i)/ninterv - 1/(2*ninterv)
end
//
while x<14, x=3*x, end
//
for i=1:ninterv ,
```

```

    if ( ((i-1)/ninterv <= xx(n) ) & (xx(n) <= (i)/ninterv) ) then,
        nfreq(i)=nfreq(i) +1;
// noter break: sortie de boucle
        break,
    end
end

```

2.3.3 Conditions

```

//noter & logique
if ( ((i-1)/ninterv <= xx(n) ) & (xx(n) <= (i)/ninterv) ) then,
    nfreq(i)=nfreq(i) +1;
    break,
end
//
select lecas,
case 1, y=.3,
case 2, y=x^2,
case 3, y=x*sin(x)
end

```

Exercice 2.5 *histogrammes, moyennes, écart-type, loi des grands nombres.*

2.3.4 Lectures dans fichiers

Pour lire des nombres dans des fichiers, 2 méthodes :

1. tableaux de nombres utiliser : $a=read('nomfich')$; le résultat est rangé dans a
2. Si le fichiers contient des enregistrement du type : $a(1,1) = 1.3e - 4$ etc, il s'agit d'instructions **scilab**; il suffit de : $exec('nomfich2')$

2.4 Les bibliothèques

En cliquant sur bouton 'help', on a accès à une aide de toutes les fonctions prédéfinies; pour des raisons d'efficacité, elles sont le plus souvent codées en fortran! Elles sont souvent issues de la bibliothèque netlib (www.netlib.org).

Pour obtenir le temps d'exécution de la fonction `calcmoystd` définie ci dessus :

```
timer(); [m,s]=calcmoystd(xx); timer()
```

Exercice 2.6 *A l'aide de l'aide en ligne, réaliser des calculs : sommes des termes d'un vecteur, racines carrés, sinus; utiliser des boutons de dialogue 'x'; tracer une courbe simple*

2.5 Statistiques et analyse de données

Scilab contient une librairie de statistiques et d'analyse de données.

2.6 Equations différentielles, optimisation

Scilab fournit différents algorithmes pour résoudre des équations différentielles et optimiser des fonctions.

La plupart de ces programmes peuvent être utilisés pour résoudre de manière interactive des exercices avec WIMS (voir la section 1.2).

2.7 Divers

2.7.1 Résultat matrice

Si l'on demande à Scilab d'afficher à l'écran, une matrice *proj* par une des trois méthodes :

- *proj*
- *disp(proj)*
- *print(%io(2),proj)*

la sortie obtenue n'est pas acceptée par wims ; principalement parceque il y a un blanc entre le signe “-” et le nombre !

Il convient d'utiliser : *write(%io(2),proj)* et de copier les lignes une à une.

Noton que *%io(2)* désigne la sortie à l'écran ; en général on peut mettre un nom de fichier ou d'unité logique associée.

Chapitre 3

Equations, intersections, projections

en géométrie affine euclidienne

3.1 Introduction

3.1.1 Références à utiliser

1. Pour une initiation on pourra consulter dans le site de l'Université en ligne, (<http://www.uel-pcsm.education.fr>), [10] le chapitre introduction géométrique de l'algèbre linéaire.
2. Enfin travailler dans les documents de WIMS ([12, 7]) : “Comment construire des sous espaces vectoriels” de B. Perrin-Riou et M. Deschamps. A noter aussi un site de documents pour les enseignants du secondaire : [9].

3.1.2 Guide d'activité

Le but de ce chapitre est de rappeler et manipuler les différents types d'équations pour représenter des sous espaces affines et les utiliser pour déterminer des projections orthogonales.

Ce point de vue géométrique pourra permettre de comprendre des méthodes très utilisées en pratique comme *la méthode des moindres carrés* ou *l'analyse en composantes principales*.

3.1.3 Prérequis

1. Notions géométriques de T.S : droites du plan ou de l'espace affine ; droites et plans de l'espace affine ; barycentres ; produits scalaires.
2. Espaces vectoriels : sous-espaces vectoriels, bases ; applications linéaires, matrices ; produit scalaire.
3. On considère une espace affine \mathbf{E} de direction l'espace vectoriel $\vec{\mathbf{E}}$. A tout couple de points A et B de l'espace affine \mathbf{E} , on peut associer un vecteur unique \vec{u} de l'espace vectoriel $\vec{\mathbf{E}}$ tel que $\vec{u} = \overrightarrow{AB}$. L'application qui associe \vec{u} à un couple de points A, B n'est pas injective.

4. A ce sujet, on rappelle la règle du parallélogramme :

$$\overrightarrow{AB} = \overrightarrow{CD}$$

si et seulement si le quadrilatère $ABCD$ est un parallélogramme.

5. Si l'on fixe un point O , on peut associer à tout point A , un vecteur unique \vec{u} tel que $\vec{u} = \overrightarrow{OA}$ et réciproquement ; en d'autres termes, on peut identifier un espace affine muni d'un point origine et un espace vectoriel. *Intuitivement on peut dire que si l'on oublie l'origine d'un espace vectoriel, il reste un espace affine*
6. **Représentation graphique** : en *algèbre linéaire*, vous avez dû prendre l'habitude de représenter les vecteurs partant tous de la même origine. L'espace dans lequel nous vivons est affine : on ne peut pas dessiner de vecteurs mais des couples de points auxquels on peut associer des vecteurs selon la formule ci-dessus. Si l'on se place dans un espace vectoriel, on peut associer à 2 vecteurs \vec{p} et \vec{q} , un vecteur $\vec{u} = \vec{q} - \vec{p}$; on notera aussi ce vecteur \overrightarrow{pq} .

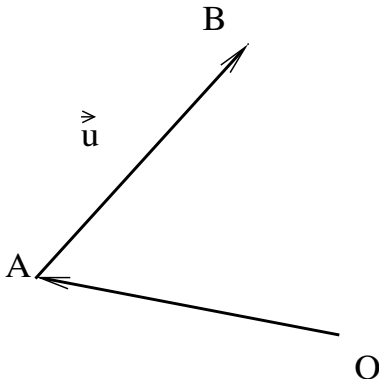


FIG. 3.1 – vecteur

7. **Relation de Chasles** : $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$, $\overrightarrow{AB} + \overrightarrow{AC} = \overrightarrow{AD}$
 $\overrightarrow{AD} - \overrightarrow{AB} = \overrightarrow{BD}$

Remarque 1 En particulier $\overrightarrow{OD} - \overrightarrow{OB}$ est un vecteur indépendant du point O !

Exercice 3.1 Dessin (triangle et parallélogramme) de somme et soustraction de vecteurs

Un exemple avec des données numériques et dessin.

8. Quand on considère un sous espace affine \mathbf{F} d'un espace vectoriel $\vec{\mathbf{E}}$: la différence de deux vecteurs de ce sous espace affine est dans un sous espace vectoriel $\vec{\mathbf{F}}$.

Dans un sous espace affine \mathbf{F} d'un espace affine \mathbf{E} : pour deux points a et m de ce sous espace, le vecteur \overrightarrow{am} est dans un sous espace vectoriel $\vec{\mathbf{F}}$.

Dans les 2 cas on a une caractérisation d'un sous espace affine à l'aide d'un point $a \in \mathbf{F}$ fixé :

$$m \in \mathbf{F} \Leftrightarrow \overrightarrow{am} \in \vec{\mathbf{F}} \quad (3.1)$$

Remarque 2 Conformément à la remarque 1, cette définition de sous espace affine est bien **indépendante** de la position du point a choisi !

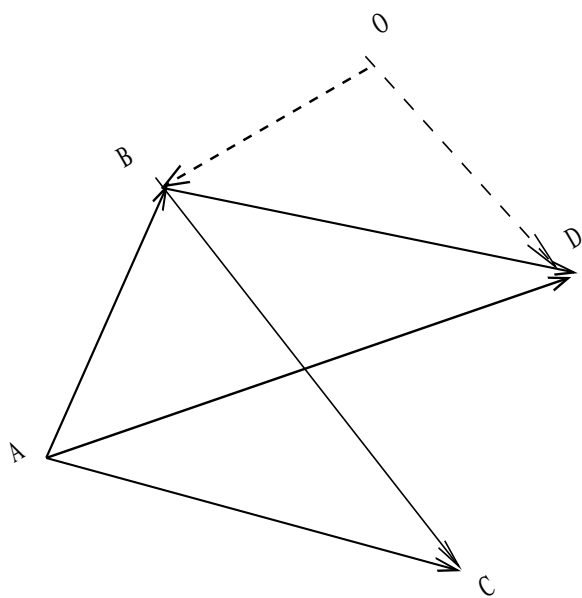
$$\overrightarrow{a'm} = \overrightarrow{a'a} + \overrightarrow{am}$$

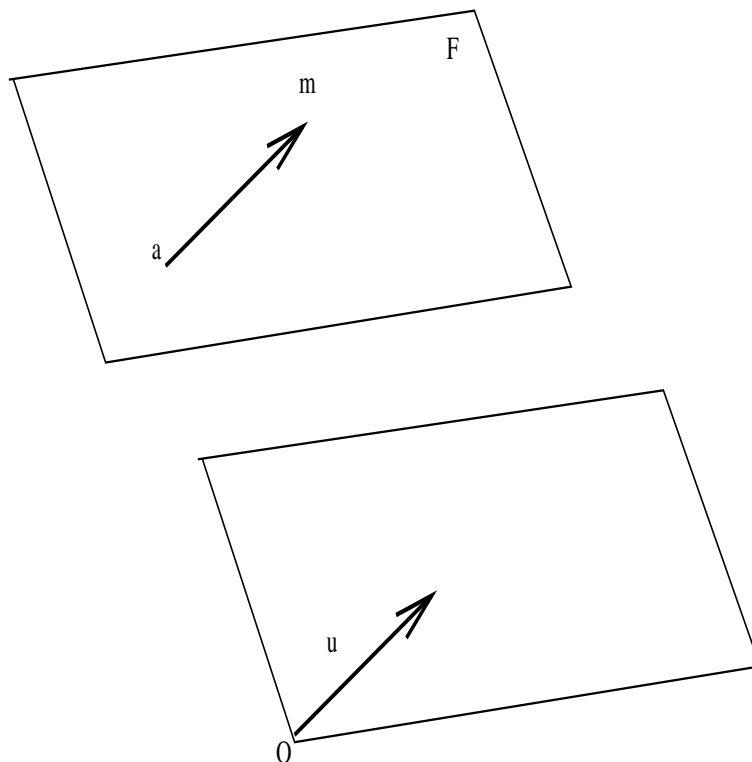
Exercice 3.2 Justifier la remarque précédente

9. Espace affine **euclidien** : l'espace vectoriel associé est euclidien, c'est à dire muni d'un produit scalaire. Dans un tel espace affine, on dispose de notions de distance et d'angle par suite l'orthogonalité et les repères orthonormés sont définis !

Dans toute la suite quand on parle d'**orthogonalité**, on suppose que l'espace est **euclidien**.

Remarque 3 Les équations posées sont le plus souvent écrites vectoriellement sans utiliser de repère. Ceci est très utile dans le plan mais devient presque indispensable à partir de la dimension 3.





Programme du module d'analyse de données de licence MASS

Prérequis : espaces vectoriels euclidiens, calcul matriciel ; notions élémentaires de géométrie affine euclidienne.

Programme

- Compléments de géométrie affine euclidienne :
 - équations paramétrées et implicites (ou cartésiennes) de droites et plans affines dans un espace vectoriel de grandes dimensions.
 - Projection orthogonale sur un plan affine (dans un espace vectoriel de grandes dimensions) ; conservation des barycentres par projection.
 - Inertie par rapport à un plan affine ; détermination du plan qui minimise l'inertie d'un système de points par rapport à ce plan
- analyse en composantes principales ; interprétation des graphiques.
- analyse factorielle des correspondances ; interprétation des graphiques.
- Régression linéaire multiple
- Classification : quelques algorithmes (nuées dynamiques etc).

3.1.4 Méthodologie ; classe WIMS

De nombreux exemples numériques sont proposés à l'aide du logiciel de calcul numérique SCILAB (www.inria.fr, [8]), octave (www.octave.org, [11]) ou Maple (calcul symbolique avec extensions numériques). Le logiciel **WIMS** (wims.unice.fr, [12]) d'enseignement assisté par ordinateur permet d'entraîner les étudiants de façon systématique ; ce logiciel est du domaine public.

Dans la classe WIMS analyse de donnees, vous trouvez des exercices en rapport avec ce chapitre.

Aucun logiciel spécialisé en analyse de données n'est utilisé dans ce module qui vise à maîtriser les fondements des méthodes.

3.2 Droite du plan

On se place dans un espace affine associé à un espace vectoriel de dimension 2. Quand on parle de vecteur **normal**, le plan est supposé **euclidien**. Les rappels et exercices ont pour but de réviser mais surtout

d'utiliser des méthodes qui passent en plus grande dimension.

Dans tout ce paragraphe, O est un point fixé choisi comme origine dans le plan affine.

3.2.1 Droite du plan :

, Vecteur directeur et vecteur normal d'une droite vectorielle et affine du plan.

	vecteur directeur	vect. directeur unitaire	vect. normal	vect. normal unitaire
droite vectorielle $\overrightarrow{D_0}$	$b \in D_0, \overrightarrow{u} = \overrightarrow{ob}$	$\overrightarrow{\tau} = \frac{\overrightarrow{u}}{\ \overrightarrow{u}\ }$	$\overrightarrow{v} \cdot \overrightarrow{\tau} = 0$	$\overrightarrow{\nu} = \frac{\overrightarrow{v}}{\ \overrightarrow{v}\ }$
droite affine D_a	$a, b \in D_a, \overrightarrow{u} = \overrightarrow{ab}$	“	“	“

3.2.2 Équation paramétrique ou cartésienne (implicite) :

1. Eq. paramétrique avec 1 point et 1 vecteur directeur

$$\overrightarrow{am} = \lambda \overrightarrow{\tau} \quad \text{ou} \quad m(\lambda) = a + \lambda \overrightarrow{\tau} \quad (3.2)$$

$$\text{ou avec 2 points : } m(\mu) = a + \mu \overrightarrow{ab} \quad (3.3)$$

$$(3.4)$$

2. équation cartésienne (1 point et 1 vecteur normal).

$$(\overrightarrow{am}) \cdot \overrightarrow{\nu} = 0 \quad (3.5)$$

	eq. paramétrique	eq. cartésienne (implicite)
droite vect $\overrightarrow{D_0}$	$\overrightarrow{om} = \lambda \overrightarrow{u}$	$\overrightarrow{om} \cdot \overrightarrow{\nu} = 0$
droite affine D_a	$m = a + \lambda \overrightarrow{u}$	$\overrightarrow{am} \cdot \overrightarrow{\nu} = 0$

3. Avec un repère cartésien, $(O, \vec{e}_1, \vec{e}_2)$ et le vecteur directeur :

$$\overrightarrow{u} = \begin{pmatrix} \beta_1 - \alpha_1 \\ \beta_2 - \alpha_2 \end{pmatrix}, \quad (3.6)$$

l' équation paramétrique fournit 2 équations :

$$x = \alpha_1 + \lambda(\beta_1 - \alpha_1) \quad (3.7)$$

$$y = \alpha_2 + \lambda(\beta_2 - \alpha_2) \quad (3.8)$$

$$\text{en particulier avec} \quad (3.10)$$

$$\vec{u} = \begin{pmatrix} 1 \\ u_2 \end{pmatrix}, a = \begin{pmatrix} 0 \\ \beta_2 \end{pmatrix}, \quad (3.11)$$

$$x = \lambda \quad (3.12)$$

$$y = b_2 + \lambda u_2 \text{ ou l'équation bien connue } y = u_2 x + b_2 \quad (3.13)$$

$$(3.14)$$

4. Avec un repère orthonormé , 1 équation cartésienne (implicite)

$$(x - a_1)v_1 + (y - a_2)v_2 = 0 \quad \text{ou} \quad (3.15)$$

$$v_1 x + v_2 y = v_1 a_1 + v_2 a_2 \quad (3.16)$$

Exercice 3.3 1. Exemples d'équation paramétrique avec des données numériques et dessin.

2. Exemples d'équation cartésienne avec des données numériques et dessin.

Exercice 3.4 1. Rappeler l' équation barycentrique d' une droite

2. Rappeler comment passer d'une représentation paramétrique à un représentation barycentrique.

Exercice 3.5 Sujet interdisciplinaire : Barycentre et moment d'un système de forces en mécanique.

3.2.3 Droites parallles, orthogonales

(avec représentation paramétrique ou cartésienne)

– Droites parallèles :

$$\overrightarrow{am} = \lambda_1 \vec{u}, \quad \overrightarrow{bm} = \lambda_2 \vec{u}$$

ou

$$\overrightarrow{am} \cdot \vec{v} = 0, \quad \overrightarrow{bm} \cdot \vec{v} = 0$$

– Droites orthogonales :

$$\overrightarrow{am} = \lambda_1 \vec{u}, \quad \overrightarrow{bm} \cdot \vec{u} = 0$$

ou

$$\overrightarrow{am} = \lambda_1 \vec{u}, \quad \overrightarrow{bm} = \lambda_2 \vec{v}$$

avec $\vec{u} \cdot \vec{v} = 0$

Utiliser des vecteurs directeurs ou normaux suivant les cas.

Exercice 3.6 Un exemple avec des données numériques et dessin.

3.2.4 Projection orthogonale d'un point

b sur une droite vectorielle $\overrightarrow{D_0}$ et affine D_a (équation paramétrique décrite avec 1 vecteur directeur).

Rappel : la projection orthogonale de b est caractérisée par :

$(\vec{b} - \tilde{P}_{D_a} \vec{b}) \cdot \vec{\tau} = 0$ et $\tilde{P}_{D_a} \vec{b} \in D_a$ cette appartenance qui peut se traduire par :
 $(\tilde{P}_{D_a} \vec{b} - a) \cdot \vec{\nu} = 0$ ou bien $\tilde{P}_{D_a} \vec{b} = a + \lambda \vec{u}$.

Dans le cas où $\vec{\tau}^2 = 1$ on a :

$$\tilde{P}_{D_0} \vec{b} = \lambda \vec{\tau} \text{ d'où} \quad (3.17)$$

$$P_{D_0} \vec{b} = (\vec{b} \cdot \vec{\tau}) \vec{\tau} \quad (3.18)$$

$$\tilde{P}_{D_a} \vec{b} = a + \lambda \vec{\tau} \text{ d'où} \quad (3.19)$$

$$\tilde{P}_{D_a} \vec{b} = \vec{a} + [(\vec{b} - \vec{a}) \cdot \vec{\tau}] \vec{\tau} \quad (3.20)$$

$$\tilde{P}_{D_a} \vec{b} = \vec{a} + P_{D_0}(\vec{b} - \vec{a}) \quad (3.21)$$

$$\tilde{P}_{D_a} \vec{b} = \vec{a} - P_{D_0} \vec{a} + P_{D_0} \vec{b} \quad \text{on utilise parfois un vecteur pivot :} \quad (3.22)$$

$$p = \vec{a} - P_{D_0} \vec{a} \quad \text{alors} \quad \tilde{P}_{D_a} \vec{b} = p + P_{D_0} \vec{b} \quad (3.23)$$

pour le cas affine : 2 interprétations avec dessin.

Exercice 3.7 – *Un exemple avec des données numériques et dessin.*

- Projection orthogonale sur une droite // à un axe.
- Généraliser au cas où l'on dispose de vecteur directeur (ou normal) non unitaire.
- Projection sur une droite d'un nuage de points du plan ; exemple avec 2 droites bien placées.
- Retrouver la formule de distance d'un point à une droite.

Remarque 4 Projection sur D_{a^\perp} : $\tilde{P}_{D_{a^\perp}} \vec{b} = \vec{a} + P_{D_0^\perp}(\vec{b} - \vec{a})$: et l'on a :

$$b - a = \tilde{P}_{D_a} \vec{b} - a + \tilde{P}_{D_{a^\perp}} \vec{b} - a$$

Cette dernière formule n'est autre que celle de la décomposition d'un vecteur selon deux axes orthogonaux !

3.2.5 De la géométrie au calcul :

Voici pour 2 exemples la démarche de la position du problème au calcul de la solution :

Fait géométrique : projection orthogonale point sur une droite du plan.

formule vectorielle :

$$\tilde{P}_{D_a} \vec{b} = \vec{a} + [(\vec{b} - \vec{a}) \cdot \vec{\tau}] \vec{\tau}$$

formule matricielle :

$$pb = a + ({}^t(b - a)\tau)\tau$$

formule en scilab `pb=a+((b-a)'*tau)*tau`

noter la présence du `*` pour désigner le produit de nombres ou bien le produit matriciel !

Fait géométrique : projection orthogonale de plusieurs points sur une droite du plan.

formule vectorielle :

$$\tilde{P}_{D_a} \vec{b}^j = \vec{a} + \left[(\vec{b}^j - \vec{a}) \cdot \vec{\tau} \right] \vec{\tau}, j = 1, \dots, n$$

formule matricielle : si l'on décide de mettre tous les vecteurs b^j dans une matrice bb à n colonnes, il convient de remplacer a par une matrice à n colonnes identiques $aa = [a \dots a]$; dans ce cas le produit de la matrice rectangulaire $bb - aa$ par τ n'a pas de sens ! toutefois, remarquons que l'on peut écrire pour un point :

$$pb = a + \tau({}^t\tau(b - a)) = a + (\tau {}^t\tau)(b - a)$$

ce qui nous permet d'obtenir pour plusieurs points :

$$pb = aa + (\tau {}^t\tau)(bb - aa)$$

formule en scilab s'en déduit directement ; noter que l'on peut l'écrire en posant :

`aa=a*ones(1,n)`

3.2.6 Vecteur orthogonal à un vecteur donné par projection

Pour calculer un vecteur orthogonal à un vecteur unitaire τ : Calculer pour un vecteur w , sa projection orthogonale sur la droite vectorielle de vecteur directeur τ : $P_{\vec{\tau}}\vec{w} = (\vec{w} \cdot \vec{\tau})\vec{\tau}$; si $w \neq P_{\vec{\tau}}\vec{w}$, un vecteur orthogonal : $\vec{v} = \frac{w - P_{\vec{\tau}}w}{\|w - P_{\vec{\tau}}w\|}$

Exercice 3.8 *Autres méthodes pour vecteur orthogonal : ex : rotation ; $\nu \cdot \tau = 0$; en repère orthonormé, vérifier : $\nu = \begin{pmatrix} -\tau_1 \\ \tau_2 \end{pmatrix}$ est un vecteur normal !*

Exercice 3.9 Faire la feuille "droites du plan" de la classe analyse de données sur wims.

3.3 Plan et droite dans l'espace tridimensionnel

On se place dans un espace affine associé à un espace vectoriel de dimension 3. Quand on parle de vecteur **normal**, l'espace est supposé **euclidien**. Les rappels et exercices ont pour but de réviser mais

surtout d'utiliser des méthodes qui passent en plus grande dimension.

Dans tout ce paragraphe, O est un point fixé choisi comme origine dans l'espace affine.

3.3.1 Plan dans l'espace

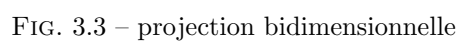
1. **Plan dans l'espace** : vecteurs directeurs et normal.

	vecteurs directeurs	vect.normal	v. norm. unitaire
plan vectoriel $\vec{\mathcal{P}}_0$	$b_1, b_2 \in \mathcal{P}_0, \vec{u}_i = \vec{ob}_i, i = 1, 2$	$\vec{v} \cdot \vec{u}_i = 0, i = 1, 2$	$\vec{v} = \frac{\vec{v}}{\ \vec{v}\ }$
plan affine \mathcal{P}_a	$a, b_1, b_2 \in \mathcal{P}_a, \vec{u}_i = \vec{ab}_i, i = 1, 2$	"	"

Les vecteurs u_i peuvent être orthonormalisés.

2. **Equation paramétrique et implicite**

	eq. paramétrique	eq. cartésienne (implicite)
plan vect $\vec{\mathcal{P}}_0$	$\vec{om} = \lambda_1 \vec{u}_1 + \lambda_2 \vec{u}_2$	$\vec{om} \cdot \vec{v} = 0$
plan affine \mathcal{P}_a	$m = a + \lambda_1 \vec{u}_1 + \lambda_2 \vec{u}_2$	$\vec{am} \cdot \vec{v} = 0$



3. **Equation paramétrique** \mapsto **Equation cartésienne** Données : on dispose de 2 vecteurs directeurs : u_1, u_2 ; en trois dimension, on peut trouver un vecteur orthogonal $v = u_1 \wedge u_2$.
Voici une autre *démarche* qui passe en plus grande dimension et qui redonne ce vecteur : on écrit que les vecteurs $u_1, u_2, \overrightarrow{am}$ sont liés : le déterminant de leurs composantes est nul $\det(u_1, u_2, \overrightarrow{am}) = 0$; ceci est identique car : $\det(u_1, u_2, \overrightarrow{am}) = (u_1 \wedge u_2) \cdot \overrightarrow{am}$
4. **Equation cartésienne d'un plan défini par 3 points** a_1, a_2, a_3 : on prend un des points a_1 et on définit 2 vecteurs directeurs : $u_1 = a_2 - a_1, u_2 = a_3 - a_1$ et on utilise une des méthodes ci-dessus.

Remarque 5 Avec un repère cartésien, on a 3 équations paramétriques avec 2 paramètres ; 1 équation cartésienne avec 3 inconnues.

Remarque 6 En algèbre linéaire les équations cartésiennes forment un système sous déterminé qui admet une infinité de solutions ; les équations paramétriques fournissent toutes ces solutions.

- équation cartésienne d'un plan vectoriel orthogonal à un vecteur donné
Exercice 3.10 Un exemple avec des données numériques et dessin.
- équation paramétrique d'un plan vectoriel défini par 2 vecteurs du plan.
Exercice 3.11 Un exemple avec des données numériques et dessin.
- équation cartésienne d'un plan affine orthogonal à un vecteur donné et passant par le point a.
Exercice 3.12 Un exemple avec des données numériques et dessin.
- équation paramétrique d'un plan affine défini par 2 vecteurs du plan et passant par le point a.
Exercice 3.13 Un exemple avec des données numériques et dessin.

3.3.2 Droite dans l'espace

- Vecteur directeur et vecteurs normaux.**

	vecteur directeur	vect. dir. unitaire	vect. normaux
droite vectorielle D_0	$b \in D_0, \overrightarrow{u} = \overrightarrow{ob}$	$\overrightarrow{\tau} = \frac{\overrightarrow{u}}{\ \overrightarrow{u}\ }$	$\vec{v}_i \cdot \overrightarrow{\tau} = 0, i = 1, 2$
droite affine D_a	$a, b \in D_a, \overrightarrow{u} = \overrightarrow{ab}$	"	"

- Equation paramétrique et cartésienne (implicite).**

	eq. paramétrique	eq. cartésienne (implicite)
dte vect D_0	$\overrightarrow{om} = \lambda \overrightarrow{u}$	$\overrightarrow{om} \cdot \vec{v}_i = 0, i = 1, 2$
dte affine D_a	$m = a + \lambda \overrightarrow{u}$	$\overrightarrow{am} \cdot \vec{v}_i = 0, i = 1, 2$

- Equation cartésienne** \mapsto **équation paramétrique** Situation "duale" de celle du plan, on dispose de v_1, v_2 et l'on a besoin d'un vecteur directeur orthogonal à ces 2 vecteurs ; mêmes méthodes.
- Equation cartésienne d'une droite définie par 2 points b_1, b_2 : on prend $a = b_1$ (par exemple) et on pose $\overrightarrow{u} = \overrightarrow{b_1 b_2}$

Les vecteurs $v_i, i = 1, 2$ peuvent être orthonormalisés.

Remarque 7 Avec un repère cartésien, on 3 équations paramétriques avec 1 paramètre ; mais 2 équations cartésiennes avec 3 inconnues. Les équation cartésiennes s'écrivent par exemple :

$$\nu_{11}(m_1 - a_1) + \nu_{21}(m_2 - a_2) + \nu_{31}(m_3 - a_3) = 0 \quad (3.24)$$

$$\nu_{12}(m_1 - a_1) + \nu_{22}(m_2 - a_2) + \nu_{32}(m_3 - a_3) = 0 \quad (3.25)$$

- équation paramétrique : 1 point et 1 vecteur ; 2 points.

Exercice 3.14 *Un exemple avec des données numériques et dessin.*

2. équation cartésienne avec 1 point et 2 vecteurs normaux ; intersection de 2 plans passant par le point.

Exercice 3.15 *Un exemple avec des données numériques et dessin.*

Exercice 3.16 *Trouver une équation paramétrique d'une droite définie par équation cartésienne. (lien avec le théorème des fonctions implicites).*

Exercice 3.17 feuille droites et plans (equations, intersections, projections) **de la classe capes-br de WIMS**

3.3.3 Intersection de droite et plan

- Le plus simple est de disposer d'une équation paramétrique de la droite et d'une équation cartésienne du plan :

$$\overrightarrow{am} = \lambda \vec{u} \quad \text{et} \quad \overrightarrow{a'm} \cdot \vec{v} = 0 \quad (3.26)$$

On reporte l'équation de la droite dans celle du plan ce qui donne une équation pour déterminer λ ; on en déduit \overrightarrow{am} .

- Dans le cas d'une droite définie par une équation catésienne et d'un plan paramétré :

$$\overrightarrow{a'm} = \lambda_1 \vec{u}_1 + \lambda_2 \vec{u}_2 \quad \text{et} \quad \overrightarrow{am} \cdot \vec{v}_i = 0 \quad \text{pour } i = 1, 2 \quad (3.27)$$

On reporte l'équation du plan dans celle de la droite et l'on obtient un système de 2 équation à 2 inconnues (λ_1, λ_2).

- Deux droites paramétriques :

$$\overrightarrow{am} = \lambda \vec{u}; \quad \overrightarrow{a'm} = \lambda' \vec{u}'; \quad \overrightarrow{aa'} = \lambda \vec{u} - \lambda' \vec{u}'; \quad (3.28)$$

Démarche : $\overrightarrow{aa'}$ appartient au plan engendré par $\vec{u}, \vec{u'}$ ssi $\det(\overrightarrow{aa'}, \vec{u}, \vec{u'}) = 0$. On peut plus directement calculer λ, λ' avec 2 équations et reporter dans la troisième pour vérifier la compatibilité.

- Deux droites cartésiennes :

$$\overrightarrow{am} \cdot \vec{v}_i = 0; \quad \overrightarrow{a'm} \cdot \vec{v}_i' = 0; \quad i = 1, 2 \quad (3.29)$$

$$\overrightarrow{om} \cdot \vec{v}_i = \overrightarrow{oa} \cdot \vec{v}_i; \quad i = 1, 2 \quad \overrightarrow{om} \cdot \vec{v}_i' = \overrightarrow{oa'} \cdot \vec{v}_i'; \quad i = 1, 2 \quad (3.30)$$

Pour déterminer le point m d'intersection, on a 4 équations pour 3 inconnues ; on prend 3 équations pour déterminer m et on vérifie la comptabilité avec un bordant (revoir les systèmes surdéterminés) ou on reporte dans la troisième équation.

- L'intersection d'une droite et d'un plan tous deux paramétriques ou tous deux cartésiens conduit à un système de 3 équations à 3 inconnues qui a une solution unique si les 3 vecteurs impliqués sont linéairement indépendants (droite et plan non parallèle)

3.3.4 Projection sur un plan

La projection est caractérisée par :

avec vecteurs directeurs	avec vecteur normal
$(\vec{b} - \tilde{P}_{\mathcal{P}_a} \vec{b}) \cdot \vec{u}_\alpha = 0$	$(\vec{b} - \tilde{P}_{\mathcal{P}_a} \vec{b}) \cdot \vec{v} = \beta \vec{v}$
$\tilde{P}_{\mathcal{P}_a} \vec{b} = a + \lambda_1 \vec{u}_1 + \lambda_2 \vec{u}_2$	$((\tilde{P}_{\mathcal{P}_a} \vec{b}) - a) \cdot \vec{v} = 0$

En pratique, on la détermine par l'une des 2 méthodes suivantes.

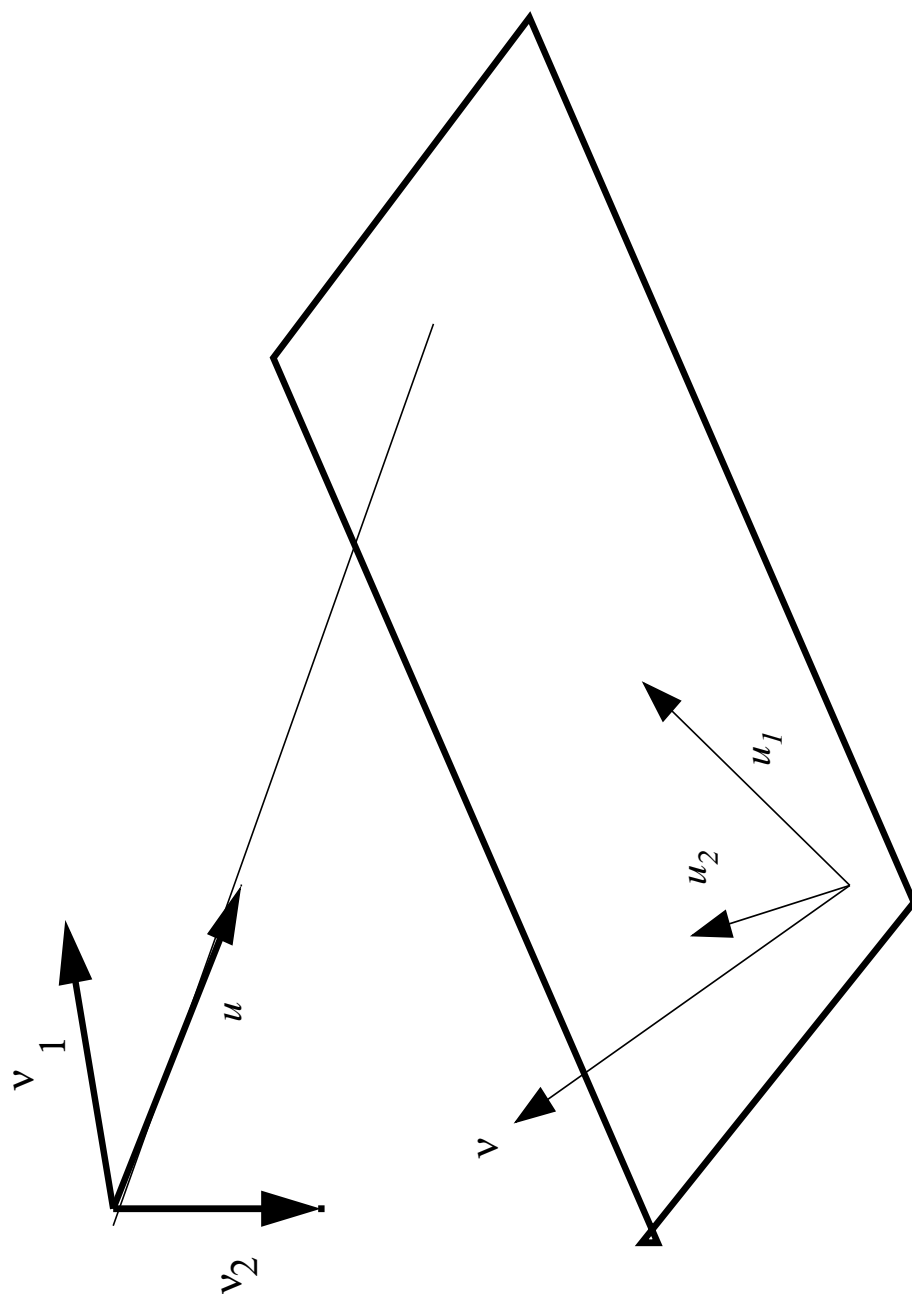


FIG. 3.4 – Intersection de droite et plan

Première méthode, avec les vecteurs directeurs.

$\tilde{P}_{\mathcal{P}_a} \vec{b} = a + \alpha_1 u_1 + \alpha_2 u_2$ et $(\vec{b} - \tilde{P}_{\mathcal{P}_a} \vec{b}) \cdot u_\alpha = 0$ ce qui fournit les paramètres solutions du système linéaire :

$$G \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} u_1 \cdot (b - a) \\ u_2 \cdot (b - a) \end{pmatrix}$$

où la matrice G dite de Gramm est

$$G = \begin{pmatrix} u_1 \cdot u_1 & u_1 \cdot u_2 \\ u_2 \cdot u_1 & u_2 \cdot u_2 \end{pmatrix}$$

Remarque 8 Dans le cas où u_1, u_2 est orthonormée, G est la matrice identité et la formule revient à calculer les composantes sur ces deux vecteurs.

Deuxième méthode, avec le vecteur normal, On utilise

$$(\vec{b} - \tilde{P}_{\mathcal{P}_a} \vec{b}) = \beta \vec{v}$$

et $\tilde{P}_{\mathcal{P}_a} \vec{b} \in \mathcal{P}_a$ ou $((\tilde{P}_{\mathcal{P}_a} \vec{b}) - a) \cdot \vec{v} = 0$

$$\tilde{P}_{\mathcal{P}_a} \vec{b} = b - ((b - a) \cdot \vec{v}) \vec{v}$$

Vecteur normal à 2 vecteurs : méthode de la projection En dehors des méthodes indiquées ci dessus, on peut aussi prendre un vecteur quelconque \vec{w} non situé dans le plan de ces 2 vecteurs u_1, u_2 , le projeter sur le plan vectoriel engendré par ces 2 vecteurs ; un vecteur orthogonal est alors :

$$\vec{v} = \vec{w} - \tilde{P}_{[\vec{u}_1, \vec{u}_2]} \vec{w} \quad (3.31)$$

Remarque 9 On a bien sûr

$$\tilde{P}_{\mathcal{P}_a} \vec{b} = a + P_{\mathcal{P}_a} (b - a)$$

.

Exercice 3.18 Un exemple avec des données numériques et dessin.

Exercice 3.19 Soit $\vec{v}^1 \vec{v}^2$ avec $\vec{v}^1 = {}^t [1, 0, \epsilon]$ $\vec{v}^2 = [0, 1, \epsilon]$.

1. Pour $\epsilon = 0$, équation paramétrique et cartésienne du plan vectoriel Π_0 engendré par $\vec{v}^1 \vec{v}^2$.
2. Pour $\epsilon = 0$, équation paramétrique et cartésienne de la droite engendrée par le vecteur normal à ces 2 vecteurs.
3. Calculer la projection sur ce plan Π_0 du point $b = [b_1, b_2, b_3]$
4. Pour $\epsilon \neq 0$, équation paramétrique et cartésienne du plan vectoriel Π_ϵ engendré par $\vec{v}^1 \vec{v}^2$.
5. Pour $\epsilon \neq 0$, équation paramétrique et cartésienne de la droite engendré par le vecteur normal à ces 2 vecteurs.
6. Calculer la projection sur ce plan Π_ϵ du point $b = [b_1, b_2, b_3]$
7. Calculer un développement limité au premier ordre de la projection sur le plan Π_ϵ pour $\epsilon \rightarrow 0$. Interprétation.

3.3.5 Projection sur une droite

Avec les mêmes idées, on peut utiliser

$$(b - \tilde{P}_{D_a} b) \cdot \vec{u} = 0$$

et

$$\tilde{P}_{D_a} b = a + \lambda \vec{u}$$

on trouve :

$$\tilde{P}_{D_a} b = a + \alpha \vec{u} = a + \frac{\vec{u} \cdot (\vec{ab})}{\|\vec{u}\|^2} \vec{u}$$

3.4 Droite et plan dans un espace de grande dimension

3.4.1 Vecteurs et matrices

Pour la suite en grande dimension, nous allons utiliser des matrices. C'est ainsi que le produit scalaire de 2 vecteurs \vec{u} et \vec{v} est noté

$${}^t\vec{v} \cdot \vec{u} \quad (3.32)$$

Nous notons V , une matrice à m lignes et 2 colonnes ; nous utiliserons souvent sa décomposition en vecteurs colonnes (notez la position de l'indice de colonne dans ce cours)

$$\vec{V} = [\vec{v}^1 \vec{v}^2] \quad (3.33)$$

$$[{}^tVV]_i^j = {}^t v^i v^j \quad (3.34)$$

Une combinaison linéaire de vecteurs s'écrit :

$$\sum_i \alpha_i \vec{v}^i = V[\alpha] \quad \text{avec } [\alpha] = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \quad (3.35)$$

On se place dans un espace affine associé à un espace vectoriel de dimension N strictement plus grand que 3. Quand on parle de vecteur **normal**, l'espace est supposé **euclidien**. On utilise des méthodes adaptées à la grande dimension.

Dans tout ce paragraphe, O est un point fixé choisi comme origine dans l'espace affine.

3.4.2 Plan dans un espace de dimension $N > 3$

1. Plan : vecteurs directeurs et normaux.

	2 vecteurs directeurs	N-2 vect.normaux
plan vectoriel $\vec{\mathcal{P}}_0$	$b_1, b_2 \in \vec{\mathcal{P}}_0, \vec{u}_i = \vec{ob}_i, i = 1, 2$	$\vec{v}_j \cdot \vec{u}_i = 0, j = 1, \dots, N-2, i = 1, 2$
plan affine \mathcal{P}_a	$a, b_1, b_2 \in \mathcal{P}_a, \vec{u}_i = \vec{ab}_i, i = 1, 2$	"

Les vecteurs u_i peuvent être orthonormalisés.

2. equation paramétrique et implicite

	eq. param	eq. cartésienne (implicite)
plan vect $\vec{\mathcal{P}}_0$	$\vec{om} = \lambda_1 \vec{u}_1 + \lambda_2 \vec{u}_2$	$\vec{om} \cdot \vec{v}_j = 0, j = 1, \dots, N-2$
plan affine \mathcal{P}_a	$m = a + \lambda_1 \vec{u}_1 + \lambda_2 \vec{u}_2$	$\vec{am} \cdot \vec{v}_j = 0, j = 1, \dots, N-2$

Remarque 10 Avec un repère cartésien, on a N équations paramétriques avec 2 paramètres ; mais $N-2$ équations cartésiennes pour N inconnues.

3.4.3 Droite dans un espace de dimension $N > 3$

1. Vecteur directeur et vecteurs normaux.

	vecteur directeur	vect. dir. unitaire	N-1 vect.normaux
droite vectorielle \vec{D}_0	$b \in \vec{D}_0, \vec{u} = \vec{ob}$	$\vec{\tau} = \frac{\vec{u}}{\ \vec{u}\ }$	$\vec{v}_i \cdot \vec{\tau} = 0, i = 1, \dots, N-1$
droite affine D_a	$a, b \in D_a, \vec{u} = \vec{ab}$	“	“

2. Equation paramétrique et implicite.

	eq. param	eq. cartésienne (implicite)
dte vect \vec{D}_0	$\vec{om} = \lambda \vec{u}$	$\vec{om} \cdot \vec{v}_i = 0, i = 1, \dots, N-1$
dte affine D_a	$m = a + \lambda \vec{u}$	$\vec{am} \cdot \vec{v}_i = 0, i = 1, \dots, N-1$

Les vecteurs $v_i, i = 1, 2$ peuvent être orthonormalisés.

Remarque 11 Avec un repère cartésien, on a N équations paramétriques avec 1 paramètre ; mais $N-1$ équations cartésiennes pour N inconnues.

3.4.4 Passage d'équation cartésienne à paramétrique et réciproquement

Equations cartésiennes données

Considérons par exemple, le cas $N = 5$, les équations cartésiennes

$$\vec{am} \cdot \vec{v}_i = 0, i = 1, 3$$

; pour simplifier considérons une base $e_i, i = 1, \dots, 5$ et supposons (quite à permuter les colonnes) que la matrice 3×3

$${}^t v' = \begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix}$$

soit inversible ; notons :

$${}^t v'' = \begin{pmatrix} v_{14} & v_{15} \\ v_{24} & v_{25} \\ v_{34} & v_{35} \end{pmatrix}$$

notons x' (resp x'') les 3 premières (resp. les 2 dernières) composantes de \vec{om} alors on a une équation *paramétrique* en inversant cette matrice v' dans le système suivant :

$${}^t v' x' = -{}^t v'' x'' + {}^t v a$$

avec les 2 paramètres x_4, x_5 .

Equations paramétriques données

Réciproquement, considérons les équations paramétriques $m = a + \lambda_1 \vec{u}_1 + \lambda_2 \vec{u}_2$; pour obtenir des équations cartésiennes, il convient d'éliminer λ_1, λ_2 ; pour cela, nous pouvons faire appel à la théorie des systèmes linéaires : à quelle condition, peut-on déterminer λ_1, λ_2 solution du système linéaire

$$\vec{u}_1 \lambda_1 + \vec{u}_2 \lambda_2 = \vec{am}$$

Supposons pour simplifier (quite à permuter les lignes) , que la matrice 2×2

$$u' = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}$$

est inversible ; le système linéaire a une solution si et seulement si **les matrices bordantes** sont non inversibles ; ce qui peut s'exprimer avec les déterminants de ces matrices bordantes sont nuls. Ceci nous donne les 3 équations :

$$\det \begin{pmatrix} u_{11} & u_{12} & m_1 - a_1 \\ u_{21} & u_{22} & m_2 - a_2 \\ u_{31} & u_{32} & m_3 - a_3 \end{pmatrix} = 0$$

$$\det \begin{pmatrix} u_{11} & u_{12} & m_1 - a_1 \\ u_{21} & u_{22} & m_2 - a_2 \\ u_{41} & u_{42} & m_4 - a_4 \end{pmatrix} = 0$$

$$\det \begin{pmatrix} u_{11} & u_{12} & m_1 - a_1 \\ u_{21} & u_{22} & m_2 - a_2 \\ u_{51} & u_{52} & m_5 - a_5 \end{pmatrix} = 0$$

Remarque 12 Notons que cela revient à vérifier que les projections de $m - a$ sur tous les sous espaces vectoriels de dimension 3 sont dans les sous espaces vectoriels engendrés par les projections de \vec{u}_1, \vec{u}_2 sur ces mêmes sous espaces tridimensionnels. Cela revient à vérifier que ces projections tridimensionnelles sont orthogonales aux produits vectoriels des projections tridimensionnelles de \vec{u}_1, \vec{u}_2 .

3.4.5 Projections orthogonales

Pour des raisons pratiques, le produit scalaire utilisé n'est pas toujours le produit usuel (ie les bases ne sont pas toujours orthonormées) ; nous utiliserons une matrice Q définie positive :

$${}^t v Q v \text{ strictement positif} \quad (3.36)$$

elle sera souvent diagonale à coefficients strictement positifs.

Proposition 3.1 Soit M_0 un sous espace vectoriel de dimension d de \mathbb{R}^m défini par un système libre de vecteurs : $\vec{v}^1, \vec{v}^2 \dots \vec{v}^d$; soit $x \in \mathbb{R}^m$; soit $V = [\vec{v}^i]$; la projection Q orthogonale de x sur M_0 est donné par :

$$P_{M_0} x = \sum_i \alpha_i \vec{v}^i \text{ avec} \quad (3.37)$$

$$G[\alpha] = {}^t V Q x \text{ où } G = {}^t V Q V \text{ ou encore :} \quad (3.38)$$

$$P_{M_0} x = V \alpha = V G^{-1} {}^t V Q x \quad (3.39)$$

Démonstration Nous nous limitons au seul cas pratique : $d=2$. Notons :

$$m = P_{M_0} x$$

Il suffit d'écrire que la projection m se trouve dans le plan et satisfait : le vecteur $\overrightarrow{x m}$ est orthogonal aux 2 vecteurs \vec{v}^1, \vec{v}^2 .

$${}^t V Q m = {}^t V Q x \quad \text{avec} \quad m = V[\alpha] \quad (3.40)$$

D'où le résultat.

Proposition 3.2 La projection affine Q -orthogonale sur le sous espace affine M_a passant par a et parallèle à M_0 est donné par :

$$\tilde{P}_{M_a} \vec{x} = \vec{a} + P_{M_0}(\vec{x} - \vec{a}) \quad (3.41)$$

$$\text{ou matriciellement } G[\alpha] = {}^t V Q(x - a), \tilde{P}_{M_a} x = a + V\alpha = V G^{-1}[{}^t V Q(x - a)] \quad (3.42)$$

$$\tilde{P}_{M_a} \vec{x} = \vec{a} - P_{M_0} \vec{a} + P_{M_0} \vec{x} \quad (3.43)$$

$$\text{ou matriciellement } G[\alpha_a] = {}^t V Q(a), G[\alpha_x] = {}^t V Q(x), \tilde{P}_{M_a} x = a - V\alpha_a + V\alpha_x \quad (3.44)$$

ou en définissant le **pivot**

$$\vec{p} = \vec{a} - P_{M_0} \vec{a} \quad (3.45)$$

$$\tilde{P}_{M_a} \vec{x} = \vec{p} + P_{M_0} \vec{x} \quad (3.46)$$

Démonstration La différence avec la projection sur un plan vectoriel : $\tilde{P}_{M_a} \vec{x} = m$ se trouve dans le plan affine ssi $m - a$ se trouve dans le plan vectoriel parallèle ; et l'on a encore le vecteur $\vec{x}\vec{m}$ est orthogonal aux 2 vecteurs $\vec{v}^1 \vec{v}^2$.

$${}^t V Q m = {}^t V Q x \quad \text{avec } m - a = V[\alpha] \quad \text{ou} \quad (3.47)$$

$$m - a = P_{M_0}(\vec{x} - \vec{a}) \quad (3.48)$$

Proposition 3.3 La projection affine Q -orthogonale n'est pas une application linéaire mais c'est une application affine : elle conserve le barycentre (voir exercice).

Exercice 3.20 1. Démontrer que la projection affine Q -orthogonale conserve le barycentre.

2. Pour un nuage de points du plan, dessiner (avec scilab) : le barycentre ainsi que les projections sur une droite avec le barycentre des projections.

Proposition 3.4 Soit M et M^\perp deux sous espaces vectoriels orthogonaux, alors on a

$$\vec{x} = P_M \vec{x} + P_{M^\perp} \vec{x} \quad (3.49)$$

$$\vec{x} = \vec{a} + P_M(\vec{x} - \vec{a}) + P_{M^\perp}(\vec{x} - \vec{a}) \quad (3.50)$$

$$\vec{x} = -\vec{a} + \tilde{P}_{M_a} \vec{x} + \tilde{P}_{M_a^\perp} \vec{x} \quad (3.51)$$

$$(3.52)$$

Exercice 3.21 Interpréter la proposition ci dessus dans le cas où M_0 est // à des axes de coordonnées.

Chapitre 4

Notations d'algèbre linéaire

L'analyse de données utilise abondamment de l'algèbre linéaire et de la géométrie affine euclidienne. Dans le polycopié, les matrices sont notées par une majuscule et les coefficients par la minuscule correspondante :

$$A \quad \text{de coefficients } a_i^j \text{ parfois noté } [A]_i^j \quad (4.1)$$

avec j indice de colonne et i indice de ligne. Bien distinguer de la notation A_{ij} où i est indice de ligne et j de colonne. Le produit de 2 matrices se note :

$$[AB]_i^j = \sum_{k=1}^{k=m} a_i^k b_k^j \quad (4.2)$$

Attention en analyse de données, on utilise beaucoup de matrices rectangulaires : $A \in L(m, n)$ à m lignes et n colonnes. On utilise aussi la décomposition en blocs et surtout en colonnes ou en lignes :

$$A = [\vec{a}^1, \vec{a}^2, \dots, \vec{a}^j, \dots, \vec{a}^n] \quad (4.3)$$

que je note souvent :

$$A = [\underline{a}^1, \underline{a}^2, \dots, \underline{a}^j, \dots, \underline{a}^n] \quad (4.4)$$

En ligne :

$$\begin{bmatrix} \underline{a}_1 \\ \underline{a}_2 \\ \dots \\ \underline{a}_i \\ \dots \\ \underline{a}_m \end{bmatrix} \quad (4.5)$$

Cela donne une écriture du produit matriciel avec un produit scalaire :

$$[AB]_i^j = < \underline{a}_i, \underline{b}^j > \quad (4.6)$$

Pour tout produit de matrices rectangulaires, dessiner le schéma :

$$\boxed{A} \quad \boxed{B} = \boxed{AB}$$

Transposition, nous utilisons la notation française tA à comparer avec une autre notation plutt anglaise : A^T .

Chapitre 5

Inertie d'un système

5.1 Définition, propriétés

A partir de ce chapitre

Notation 1 : $\boxed{p^j \in \mathbb{R}, \quad p^j > 0, \sum_{j=1}^n p^j = 1 \text{ et } \underline{\underline{P}} \text{ matrice diagonale}}$ de coefficients diagonaux p^j

La dispersion des points par rapport à un sous espace affine peut être mesurée par l'inertie ; ce qui généralise l'inertie par rapport à un axe utilisée dans la dynamique d'un solide rigide en rotation autour d'un axe fixe.

Définition 5.1 Avec les notations $\underline{\underline{A}} = [\underline{a}^1 \dots \underline{a}^n]$ où $\underline{a}^j \in \mathbb{R}^m$; M un sous espace vectoriel de dimension d de \mathbb{R}^m , posons $M_{\underline{a}} = \underline{a} + M$ (sous espace affine passant par \underline{a} et // à M) ; on appelle inertie de $\underline{\underline{A}}$ par rapport à $M_{\underline{a}}$

$$I_{M_{\underline{a}}}(\underline{\underline{A}}) = \sum_{j=1}^n p^j \|\underline{a}^j - \underline{\underline{P}}_{M_{\underline{a}}} \underline{a}^j\|_{\underline{\underline{Q}}}^2 \quad (5.1)$$

en particulier, on peut considérer l'inertie par rapport à un sous espace vectoriel :

$$I_M(\underline{\underline{A}}) = \sum_{j=1}^n p^j \|\underline{a}^j - \underline{\underline{P}}_M \underline{a}^j\|_{\underline{\underline{Q}}}^2 \quad (5.2)$$

Remarque 13 L'inertie par rapport à un sous-espace affine n'est autre que la somme pondérée des carrés des distances des points \underline{a}^j au sous espace affine considéré.

Exercice 5.1 Calculer l'inertie de 2 points $\underline{a}^1, \underline{a}^2$ du plan par rapport à une droite. Discuter suivant la position de la droite par rapport au centre de gravité et par rapport à son angle avec la droite défini par les 2 points $\underline{a}^1, \underline{a}^2$.

Proposition 5.1 (Huygens) Soit le barycentre g des points \underline{a}^j affectés des poids p^j , $\underline{g} = \sum_{i=1}^n p^i \underline{a}_i$ on a :

$$I_{M_{\underline{a}}}(\underline{\underline{A}}) = I_{M_{\underline{g}}}(\underline{\underline{A}}) + d^2(\underline{g}, M_{\underline{a}}) = I_{M_{\underline{g}}}(\underline{\underline{A}}) + d^2(\underline{a}, M_{\underline{g}})$$

Remarque 14

$$d^2(\underline{g}, M_{\underline{a}}) = I_{M_{\underline{a}}}(g); \quad d^2(\underline{a}, M_{\underline{g}}) = I_{M_{\underline{g}}}(a)$$

Démonstration Remarquons :

$$\tilde{P}_{M_a} \underline{a}^j = \underline{a} + P_M(\underline{a}^j - \underline{a}) \quad (5.3)$$

$$\tilde{P}_{M_a} \underline{g} = \underline{a} + P_M(\underline{g} - \underline{a}) \quad (5.4)$$

$$\tilde{P}_{M_g} \underline{a}^j = \underline{g} + P_M(\underline{a}^j - \underline{g}) \quad (5.5)$$

$$\underline{a}^j - \tilde{P}_{M_a} \underline{a}^j = \underline{a}^j - \tilde{P}_{M_g} \underline{a}^j + \underline{g} - \tilde{P}_{M_a} \underline{g} \quad (5.6)$$

L'utilisation de la conservation du barycentre par projection (proposition 3.3) va donner la première égalité :

$$\sum_{j=1}^n p^j \tilde{P}_{M_g} \underline{a}^j = \tilde{P}_{M_g} \left(\sum_{j=1}^n p^j \underline{a}^j \right) = \tilde{P}_{M_g} \underline{g} = \underline{g} \text{ ce qui fournit} \quad (5.7)$$

$$\sum_{j=1}^n p^j \|\underline{a}^j - \tilde{P}_{M_a} \underline{a}^j\|_{\underline{\underline{Q}}}^2 = \quad (5.8)$$

$$\sum_{j=1}^n p^j \|\underline{a}^j - \tilde{P}_{M_g} \underline{a}^j\|_{\underline{\underline{Q}}}^2 + \sum_{j=1}^n p^j \|\underline{g} - \tilde{P}_{M_a} \underline{g}\|_{\underline{\underline{Q}}}^2 + 2 \sum_{j=1}^n p^j (\underline{a}^j - \tilde{P}_{M_g} \underline{a}^j) \cdot (\underline{g} - \tilde{P}_{M_a} \underline{g}) \quad (5.9)$$

et le double produit est nul avec l'égalité 5.7 Enfin comme $\underline{g} - \tilde{P}_{M_a} \underline{g} = -(a - \tilde{P}_{M_g} \underline{a})$, on a la dernière égalité de la proposition.

Définition 5.2 Il est commode d'introduire $d^j = a^j - g$ et $\underline{\underline{D}} = \underline{\underline{A}} - [\underline{g} \dots \underline{g}]$

Proposition 5.2

$$I_{M_{\underline{g}}}(\underline{\underline{A}}) = I_M(\underline{\underline{D}}) \quad (5.10)$$

$$I_{\{0\}}(\underline{\underline{D}}) = I_M(\underline{\underline{D}}) + I_{M^\perp}(\underline{\underline{D}}) \quad \text{ou} \quad (5.11)$$

$$I_{\{g\}}(\underline{\underline{A}}) = I_{M_g}(\underline{\underline{A}}) + I_{M_g^\perp}(\underline{\underline{A}}) \quad \text{avec} \quad (5.12)$$

$$I_{M^\perp}(\underline{\underline{D}}) = \sum_{j=1}^n p^j \|\underline{P}_M d^j\|_{\underline{\underline{Q}}}^2 \quad (5.13)$$

le théorème de Huygens peut encore s'écrire :

$$I_{M_{\underline{a}}}(\underline{\underline{A}}) = I_M(\underline{\underline{D}}) + d^2(\underline{g}, M_{\underline{a}}) \quad (5.14)$$

Démonstration Utiliser la formule $\tilde{P}_{M_a} \underline{x} = \underline{a} + P_{M_0}(\underline{x} - \underline{a})$, et $\underline{a}^j - \tilde{P}_{M_g} \underline{a}^j = \underline{d}^j - \underline{P}_M \underline{d}^j$ le résultat s'en déduit.

Proposition 5.3 Soit $\underline{\underline{V}} = [\underline{v}^1 \dots \underline{v}^d]$ une base $\underline{\underline{Q}}$ -orthonormée de M , alors : $I_{M^\perp}(\underline{\underline{D}}) = \text{tr}({}^t \underline{\underline{V}} \underline{\underline{Q}} \underline{\underline{D}} \underline{\underline{P}} {}^t \underline{\underline{D}} \underline{\underline{Q}} \underline{\underline{V}})$

La démonstration repose sur les lemmes :

Lemme 1

$$\text{tr}(\underline{h} {}^t \underline{k}^l) = {}^t \underline{h} \underline{k} \quad (\text{attention à la position des transposées!}) \quad (5.15)$$

$$\text{tr}(\underline{h} {}^t \underline{k}) = \sum_l {}^t \underline{h} \underline{k}^l \quad \text{et} \quad \sum_{j=1}^n p^j {}^t \underline{x}^j \underline{x}^j = \text{tr}(\underline{x} \underline{P} {}^t \underline{x}) \quad (5.16)$$

d'où

$$\sum_j p^j {}^t(P_M \underline{d}^j) Q(P_M \underline{d}^j) = \quad (5.17)$$

$$\sum_j p^j {}^t d^j Q V {}^t V Q V {}^t V Q d^j = \sum_j p^j {}^t d^j Q V {}^t V Q d^j = \quad (5.18)$$

$$tr({}^t \underline{\underline{V Q D}} \underline{\underline{P}} {}^t \underline{\underline{D Q V}}) \quad (5.19)$$

Remarque 15

$$[\underline{\underline{D}} \underline{\underline{P}} {}^t \underline{\underline{D}}]_j^k = \sum_i d_j^i p^i d_k^i = cov(a_j, a_k) \quad (5.20)$$

où $cov(a_j, a_k)$ est la covariance empirique des caractères a_j et a_k pondérés par les poids p^i .

Pour démontrer la proposition ci-dessous, nous aurons besoin de lemmes.

Lemme 2 Si K est une matrice symétrique, Q une matrice symétrique définie positive, alors

1. Soit u^i les vecteurs propres de $\sqrt{Q}K\sqrt{Q}$; ces vecteurs propres sont orthonormés : ${}^t u^k u^i = 0$ pour $k \neq i$ et ${}^t u^k u^k = 1$
2. Soit $u^i = \sqrt{Q^{-1}} u^i$ ce sont les vecteurs propres

$$K Q u^i = \lambda_i u^i$$

et ils sont Q orthogonaux : ${}^t u^k Q u^i = 0$ pour $k \neq i$

Remarque 16 Pour une éventuelle valeur propre multiple, on peut (théoriquement) trouver une base orthogonale du sous espace propre; pour de grandes matrices, dans la pratique numérique cela n'est pas toujours facile.

5.2 Maximum, minimum de l'inertie

Proposition 5.4 Soit M un s.e.v. de \mathbb{R}^m de dimension d , alors l'application

$$M \longrightarrow I_{M^\perp}(\underline{\underline{D}}) \quad (5.21)$$

atteint son maximum lorsque M est engendré par les d vecteurs propres de $\underline{\underline{D P}} {}^t \underline{\underline{D Q}}$ associés aux d plus grandes valeurs propres $\lambda_1 \dots \lambda_d$ et ce maximum est égal à

$$I_{M^\perp}(\underline{\underline{D}}) = \sum_{i=1}^d \lambda_i \quad (5.22)$$

En effet :

$$I_{M^\perp}(\underline{\underline{D}}) = \sum_{i=1}^d {}^t \underline{\underline{v}}^i Q D P {}^t D Q \underline{\underline{v}}^i \quad (5.23)$$

Une propriété due à Lord Rayleigh dit que pour une matrice symétrique K , on a :

$${}^t v K v \leq {}^t u^1 K u^1 = \lambda_1$$

parmi les ${}^t v v = 1$ et avec u^1 vecteur propre normalisé associé à la plus grande valeur propre λ_1 . Nous allons adapter cette idée. Nous nous limitons pour simplifier au cas $d = 2$, et posons pour une base Q orthonormée :

$$[v^1, v^2] = [u^1, \dots, u^m][y^1, y^2] \text{ avec } y^1, y^2 \in \mathbf{R}^m \text{ d'où} \quad (5.24)$$

$$DP {}^t DQ[v^1, v^2] = [\lambda_1 u^1, \dots, \lambda_m u^m][y^1, y^2] \quad (5.25)$$

$$\begin{bmatrix} {}^t v^1 \\ {}^t v^2 \end{bmatrix} QDP {}^t DQ[v^1, v^2] = \begin{bmatrix} {}^t y^1 \\ {}^t y^2 \end{bmatrix} \begin{bmatrix} {}^t u^1 \\ \vdots \\ {}^t u^m \end{bmatrix} [\lambda_1 Q u^1, \dots, \lambda_m Q u^m][y^1, y^2] = \quad (5.26)$$

$$\begin{bmatrix} {}^t y^1 \\ {}^t y^2 \end{bmatrix} \Lambda[y^1, y^2] = \sum_{k=1}^m \lambda_k (y_k^1)^2 + \sum_{k=1}^m \lambda_k (y_k^2)^2 \quad (5.27)$$

comme v^1, v^2 sont Q orthonormés, on en déduit facilement que y^1, y^2 sont orthonormés et que le maximum est atteint pour :

$$y^1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad y^2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (5.28)$$

ceci permet de conclure que les v^i doivent être les vecteurs propres de $\underline{\underline{DP}} {}^t \underline{\underline{DQ}}$.

Corollaire 1 *Sous les hypothèses de la proposition précédente*

$$M_{\underline{a}} \longrightarrow I_{M_{\underline{a}}}(\underline{A}) \quad (5.29)$$

atteint son minimum quand $a = g$ et quand M est le sous espace propres associés aux d plus grandes valeurs propres $\lambda_1 \dots \lambda_d$ de $\underline{\underline{DP}} {}^t \underline{\underline{DQ}}$; ce minimum est égal à la somme des $m - d$ plus petites valeurs propres de $\underline{\underline{DP}} {}^t \underline{\underline{DQ}}$:

$$I_{M_{\underline{a}}}(\underline{A}) = \sum_{i=d+1}^m \lambda_i \quad (5.30)$$

Chapitre 6

Analyse en composantes principales (A.C.P.)

6.1 Introduction

L'objet d'étude est un tableau A dont chaque colonne A^j correspond à un *individu* (au sens de la statistique descriptive) et la ligne A_i un *caractère*; pour se faire une idée des ressemblances entre les individus ou entre les caractères, on va projeter ces individus sur un plan affine (ou un sous-espace de plus grande dimension d); ce plan est choisi de telle façon que ces projections soient assez proches du nuage d'individus A^j .

Une façon de rendre compte de cette proximité est l'inertie des colonnes de A par rapport à ce sous-espace affine. Nous sommes conduits en utilisant le corollaire 1 à projeter sur le sous-espace propre de dimension d associé aux d plus grandes valeurs propres de $\underline{DP}^t \underline{DQ}$.

6.2 Pratique de l'ACP

Soient n individus et m caractères rangés dans une matrice A (m lignes et n colonnes; des poids $p^j > 0$, $j = 1, \dots, n$; $\sum_{j=1}^n p^j = 1$ pour les colonnes (individus) et des poids $q_i > 0$, $i = 1, \dots, m$ pour les lignes; pour réaliser une A.C.P. :

1. Calculer le barycentre \underline{g} des colonnes de A ;
2. Calculer la matrice centrée : $D = A - [\underline{g} \dots \dots \dots]$;
3. Calculer la matrice de covariance $DP^t D$, puis $\sqrt{Q}DP^t D\sqrt{Q}$ et $DP^t DQ$;
4. Calculer les d plus grandes valeurs propres λ_k , $k = 1, \dots, d$ de $\sqrt{Q}DP^t D\sqrt{Q}$ et les vecteurs propres orthonormés associés \underline{u}'^k , $k = 1, \dots, d$.
5. On peut en déduire l'inertie par rapport au sous-espace affine M_g passant par g et parallèle au sous-espace propre engendré par \underline{u}^k avec le corollaire 1.
6. En déduire les vecteurs propres Q -orthonormés $\underline{u}^k = \sqrt{Q}^{-1} \underline{u}'^k$ de $DP^t DQ$; posons $U = [u^1, \dots, u^d]$.
7. Projeter (pour le produit scalaire de matrice Q) les individus sur le sous-espace affine M_g passant par g et parallèle au sous-espace propre engendré par \underline{u}^k .

Ceci revient à remplacer le tableau D par le tableau $U^t U Q D$.

Pour la **représentation graphique** on utilise le tableau à d lignes des composantes des projections dans la base de vecteurs propres : ${}^t U Q D$ (**solution retenue**)

Remarque 17 Attention les vecteurs propres $\underline{u}^k, k = 1, \dots, d$ doivent être orthonormés ; de plus les vecteurs propres sont définis au signe près ; le dessin des projections obtenu pourra avoir 4 dispositions ; elles peuvent être obtenues à partir d'une d'elle par symétrie par rapport à chacun des 2 axes et d'autre part par rapport à l'origine.

6.3 Interprétation de l'ACP

6.3.1 Qualités de la représentation

Qualité globale de la représentation

$$q_g = \frac{I_{M_g}(\underline{A})}{I_g(\underline{A})} = \frac{\sum_1^d \lambda_j}{\sum_1^m \lambda_k} \quad (6.1)$$

on a $0 \leq q_g \leq 1$.

En effet, avec la proposition 5.2 :

$$I_{\{g\}}(\underline{A}) = I_{M_g}(\underline{A}) + I_{M_g^\perp}(\underline{A}) = I_M(\underline{D}) + I_{M^\perp}(\underline{D}).$$

D'autre part avec la proposition 5.4,

$$I_{M^\perp}(\underline{D}) = \sum_{i=1}^d \lambda_i \text{ et } I_M(\underline{D}) = \sum_{i=d+1}^m \lambda_i$$

$$I_g(\underline{A}) = I_0(\underline{D}) = \sum_{i=1}^m \lambda_i$$

En pratique, on regarde souvent si la valeur propre λ_{d+1} est bien plus petite que la dernière valeur propre retenue λ_d .

Autres critères de qualité

On peut aussi considérer des critères de qualité relatifs à un individu donné.

6.3.2 Anciens et nouveaux caractères

Une fois que la représentation graphique des points sur le meilleur plan est obtenue, il reste à interpréter ce graphique, en particulier les 2 axes du graphique. Pour cela, on va d'abord calculer le cosinus entre les anciens et les nouveaux caractères : ce n'est autre que le coefficient de corrélation entre ces anciens et nouveaux caractères.

- Les anciens caractères sont les lignes de la matrice A , notés a_k ; cet ancien caractère centré autour de sa moyenne (la k ème composante g_k du barycentre) est la k ème ligne de la matrice centrée $D = A - [g \dots g]$, notée d_k ; en désignant par

$${}^t e_k = (0 \dots 1 \dots 0)$$

avec le 1 à la k ème colonne, on a $d_k = {}^t e_k D$.

- Les nouveaux caractères :

$$v_j = {}^t u^j Q D$$

; le cosinus de l'angle entre ancien et nouveaux caractères doit se calculer avec le produit scalaire défini par les poids p^i rangés dans la matrice diagonale P ; c'est encore le coefficient de corrélation entre l'ancien et le nouveau caractère :

$$\cos(\theta_{jk}) = \rho_{jk} = \frac{v_j P^t d_k}{\|v_j\|_P \|d_k\|_P} \quad (6.2)$$

Le coefficient de corrélation multiple $R_{kjj'}$:

$$R_{kjj'}^2 = \rho_{jk}^2 + \rho_{j'k}^2$$

; le théorème de Pythagore montre que ce coefficient est le cosinus de l'angle entre l'ancien caractère et le plan des nouveaux caractères; Plus ce nombre est proche de 1, plus l'ancien caractère d_k est bien expliqué par les 2 nouveaux caractères.

– Représentation graphique des anciens caractères : on représente les points

$$C_k = (\rho_{jk}, \rho_{j'k})$$

; ces points peuvent être calculés par (6.2) :

La proximité de certains de ces points avec l'axe 1 ou 2 permet d'interpréter ces axes à partir des anciens caractères. Notons que cette interprétation n'est pas de nature mathématique mais relève du savoir faire et de la connaissance de la situation.

6.4 Calcul de valeurs et vecteurs propres

Le Calcul de valeurs et vecteurs propres est un problème plus difficile que la résolution de systèmes linéaires. De nombreux algorithmes existent adaptés à des situations diverses.

Une des méthodes les plus simples à mettre en œuvre pour calculer les plus grandes valeurs propres est la méthode de la puissance. Supposons pour simplifier que la matrice A est symétrique et que ses valeurs propres de plus grand module sont toutes de module distinct. Soit x^0 quelconque non orthogonal à x^1 est un vecteur propre associé à la valeur propre de plus grand module ;

$$y^{n+1} = Ax^n, \quad x^{n+1} = \frac{y^{n+1}}{\|y^{n+1}\|}$$

x^n converge vers le vecteur propre associé à la valeur propre de plus grand module. Si x^1 est un vecteur propre associé à la valeur propre de plus grand module : soit

$$u^0 = v^0 - ({}^t x^1 v^0) x^1$$

avec v^0 quelconque mais non colinéaire à x^1 et non orthogonal à x^2 vecteur propre associé à la deuxième valeur propre dans le classement par module. :

$$v^{n+1} = Au^n, \quad u^{n+1} = v^{n+1} - ({}^t x^1 u^{n+1}) x^1, \quad u^{n+1} = \frac{u^{n+1}}{\|u^{n+1}\|}$$

u^n converge vers x^2 .

Chapitre 7

Analyse factorielle des correspondances

7.1 Compléments d'algèbre linéaire

Nous utiliserons le lemme suivant avec les 2 matrices de covariance utilisées en A.F.C.

Lemme 3 Soit H une matrice rectangulaire $m \times n$ (pour fixer les idées : $m \leq n$) et les matrices carrées :

$${}^XC = H {}^tH, \quad {}^YC = {}^tH H;$$

(matrices symétriques !) soient leurs valeurs propres : $0 \leq \lambda_m \leq \dots \leq \lambda_1$ et $0 \leq \mu_n \leq \dots \leq \mu_1$ avec :

$${}^XC u^j = \lambda_j u^j, \quad {}^YC v^k = \mu_k v^k$$

alors pour $j = 1, \dots, m$, $\lambda_j = \mu_j$ et pour $k = m+1, \dots, n$, $\mu_k = 0$ et

$$v^k = \frac{{}^tH u^k}{\sqrt{\lambda_k}}$$

est un vecteur propre orthonormé de YC quand u^k est un vecteur propre orthonormé de XC .

Démonstration On remarque d'abord que ${}^XC, {}^YC$ sont positives ; leurs valeurs propres sont donc positives (réelles puisque symétriques). Pour simplifier, supposons que H soit de rang m ; on peut donc en extraire une sous matrice carrée H_m de rang m ; en partionant YC avec cette matrice, on voit facilement qu'elle est de rang m ; ses plus petites valeurs propres sont donc nulles. Le lien entre les vecteurs propres découle facilement de :

$${}^YC {}^tH u^k = {}^tH H {}^tH u^k = {}^tH {}^XC u^k = \lambda_k {}^tH u^k$$

7.2 Notations, définitions

Deux caractères \mathcal{X} et \mathcal{Y} sur une population ; \mathcal{X} a n modalités et \mathcal{Y} m modalités ; dans le tableau N , le coefficient N_j^i représente le nombre d'individus ayant la i^{eme} modalité de \mathcal{X} et la j^{eme} modalité de \mathcal{Y} .

$$N = \left| \begin{array}{ccc|c} N_1^1 & \dots N_1^i & \dots N_1^n & N_1 = \sum_{i=1}^n N_1^i \\ \vdots & \vdots & \vdots & \vdots \\ N_j^1 & \dots N_j^i & \dots N_j^n & N_j = \sum_{i=1}^n N_j^i \\ \vdots & \vdots & \vdots & \vdots \\ N_m^1 & \dots N_m^i & \dots N_m^n & N_m = \sum_{i=1}^n N_m^i \\ N^1 = \sum_{j=1}^m N_j^1 & N^i = \sum_{j=1}^m N_j^i & N^n = \sum_{j=1}^m N_j^n & \bar{N} = \sum_{i=1}^n N^i = \sum_{j=1}^m N_j \end{array} \right|$$

On désigne le tableau des profils des colonnes par X avec $X_j^i = \frac{N_j^i}{N^i}$ et le tableau des profils des lignes Y par $Y_j^i = \frac{N_j^i}{N_j}$.

L'AFC en pratique consiste à réaliser d'une part une analyse en composantes principales du tableau X avec les poids $^X p^i = f^i$ et $^X q_j = \frac{1}{f_j}$, avec les fréquences $f^i = \frac{N^i}{N}$ et $f_j = \frac{N_j}{N}$.

On réalise d'autre part une ACP du tableau Y par lignes avec les poids $^Y p^i = \frac{1}{f^i}$ et $^Y q_j = f_j$.

On représente sur le même dessin les projections des colonnes de X et les lignes de Y chacun sur son plan d'ACP.

7.3 Résultats

Lemme 4 *Le barycentre des colonnes de X est $^X g_j = \frac{N_j}{N} = f_j$*

Le barycentre des lignes de Y est $^Y g^i = \frac{N^i}{N} = f^i$

Proposition 7.1 *Les matrices de covariance*

$$^X C = \sqrt{^X Q} \ ^X D \ ^X P \ ^t X D \ \sqrt{^X Q} \quad (7.1)$$

$$^Y C = \sqrt{^Y P} \ ^t Y D \ ^Y Q \ ^Y D \ \sqrt{^Y P} \quad (7.2)$$

peuvent s'exprimer :

$$^X C = H \ ^t H \quad ^Y C = {}^t H H \text{ avec} \quad (7.3)$$

$$H = \sqrt{^X Q} \ ^X D \ \sqrt{^X P} = \sqrt{^Y Q} \ ^Y D \ \sqrt{^Y P} \quad (7.4)$$

$$H_j^i = \frac{N_j^i}{\sqrt{N^i N_j}} - \frac{\sqrt{N_j} \sqrt{N^i}}{N} \quad (7.5)$$

Corollaire 2 *Les vecteurs propres orthonormés u^j de $^X C$ fournissent*

$$v^k = \frac{{}^t H u^k}{\sqrt{\lambda_k}}$$

vecteurs propres orthonormés de $^Y C$.

7.4 Interprétation

L'intérêt de l'AFC est de pouvoir interpréter la position d'un point colonne (resp. ligne) par rapport { à l'ensemble des points lignes (resp.colonnes) représentés sur le graphique ; cette interprétation donne une correspondance entre les modalités des 2 caractères \mathcal{X} et \mathcal{Y} . Une bonne interprétation demande une connaissance du sujet d'application.

Chapitre 8

Régression

8.1 La méthode

8.1.1 Introduction

Données : un vecteur $y \in \mathcal{R}^m$ et un tableau A constitué de n vecteurs colonnes de \mathcal{R}^m avec $m > n$. $A = [\underline{a}^1, \dots, \underline{a}^n]$.

On souhaite approcher y par une fonction affine des n vecteurs \underline{a}^j , $j = 1, \dots, n$ (les statisticiens disent qu'ils cherchent à expliquer y par les variables \underline{a}^j , $j = 1, \dots, n$, ce qui revient à trouver une fonction affine des variables \underline{a}^j qui approche le nuage de points de $\mathcal{R}^{(n+1)}$, $(y, \underline{a}^1, \dots, \underline{a}^n)$. Il s'agit donc de projeter y sur le sous espace vectoriel d'équation paramétrique écrite sous forme matricielle :

$$z = \alpha_0 \underline{a}^0 + A \underline{\alpha}$$

avec $\underline{a}^0 = {}^t [1, \dots, 1]$ En pratique, les lignes de A correspondent à des observations affectées de poids q_i qui vérifient $q_i \geq 0$, $\sum_i q_i = 1$. La projection se fait donc avec le produit scalaire de matrice diagonale Q avec $Q_{ii} = q_i$

8.1.2 Calcul

Introduisons

$$\tilde{A} = [\underline{a}^0, A]$$

, la projection peut se calculer en introduisant la matrice de Gramm :

$$\tilde{G} = {}^t \tilde{A} Q \tilde{A}$$

et en résolvant :

$$\tilde{G} \tilde{\alpha} = {}^t \tilde{A} Q y \tag{8.1}$$

Vu la valeur de \underline{a}^0 , on a $\bar{y} = {}^t \underline{a}^0 Q y$ où \bar{y} désigne la moyenne de y ; de même, $\overline{a^j}$ désigne la moyenne de a^j . alors :

$$\alpha_0 = \bar{y} - [\overline{a^1}, \dots, \overline{a^j}, \dots, \overline{a^n}] \alpha$$

posons $\bar{A} = \underline{a}^0 [\overline{a^1} \dots \overline{a^n}]$, en introduisant la matrice centrée

$$B = A - \bar{A}$$

, et la matrice de Gramm

$$G = {}^t B Q B$$

on a :

$$G\alpha = {}^t BQy$$

La matrice G et le second membre peuvent s'interpréter comme des covariances.

La solution cherchée est :

$$y^* = \alpha_0 a^0 + A\alpha$$

ou on a la relation fonctionnelle entre y^* et les variables a^1, \dots, a^n :

$$y^* = \alpha_0 + \sum_j a^j$$

; ici les a^j sont des variables dont le tableau A est un jeu de valeurs !

Qualité

On peut la mesurer avec le coefficient de corrélation entre y et y^* ou l'angle ϕ entre $y - \bar{y}$ et $y^* - \bar{y}^*$:

$$\rho(y, y^*) = \cos(\phi) = \frac{{}^t(y - \bar{y})Q(y^* - \bar{y}^*)}{\|y - \bar{y}\|_Q \|y^* - \bar{y}^*\|_Q}$$

.

8.2 Introduction, suppression de variables

8.2.1 Introduction d'une variable supplémentaire

On cherche à introduire une variable supplémentaire a , on désigne par y^+ , la régression selon les variables a^1, \dots, a^n, a .

L'amélioration de la qualité peut se mesurer par :

$$\rho(y - y^*, y - y^+) = \cos(\gamma) = \frac{{}^t(y - y^*)Q(y^+ - y)}{\|(y - y^*)\|_Q \|(y^+ - y^*)\|_Q}$$

ce cosinus peut se calculer sans effectuer la régression par rapport aux $n + 1$ données ; il suffit de calculer la régression a^* de la nouvelle variable a par rapport aux anciennes ; en effet :

$$\cos(\gamma) = \rho(a - a^*, y - y^*) = \frac{{}^t(a - a^*)Q(y - y^*)}{\|a - a^*\|_Q \|y - y^*\|_Q}$$

On retient la variable si $|\cos(\gamma)|$ est proche de 1.

8.2.2 Suppression d'une variable

supposons que nous souhaitons éliminer la variable a^j ; nous allons calculer

$$\cos(\theta) = \rho(y^* - y^-, y - y^-)$$

avec y^- la régression par rapport à $n - 1$ variables ; en comparant ce coefficient de corrélation pour tous les choix de a^j , on retire la variable pour lequel le cosinus est le plus petit.

Chapitre 9

Devoirs

Contrôle analyse de données, licence MASS 2000-2001

1. On considère la matrice :

$$S = \begin{pmatrix} 2.4 & 2. & 5. \\ 2. & 5.4 & 10. \\ 5. & 10. & 26.4 \end{pmatrix}$$

- (a) Calculer avec scilab valeurs et vecteurs propres de A ; vérifier la main que les valeurs propres sont correctes ; avec scilab effectuer les vérifications qui s'imposent.
- (b) Calculer l'opérateur de projection sur le sous espace propre M associé aux deux plus petites valeurs propres ; (avec scilab)
- (c) Déterminer une base orthonormée de vecteurs propres ; (avec scilab).
- (d) Avec cette nouvelle base, recalculer l'opérateur de projection sur le sous espace M associé aux mêmes valeurs propres ; comparer les 2 résultats.
- (e) Fournir une équation du plan affine M_a passant par

$$a = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

et parallèle M .

- (f) Trouver la projection sur M_a des points x^1, x^2 de composantes :

$$\begin{pmatrix} 3 & 5 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$$

avec scilab effectuer les vérifications qui s'imposent.

- (a) On considère les points dont les composantes sont dans la matrice

$$A = \begin{pmatrix} 610 & 140 & 60 & 10 & 120 & 95 & 115 \\ 475 & 90 & 250 & 30 & 140 & 120 & 100 \\ 10 & 0 & 495 & 110 & 170 & 110 & 130 \\ 615 & 140 & 65 & 10 & 115 & 90 & 115 \\ 179 & 29 & 421 & 87 & 161 & 112 & 119 \end{pmatrix}$$

on prend les p_i tous égaux et les $q_j = 1$; calculer le centre de gravité de ces points et la matrice de covariance associée.

- (b) Calculer ses 2 plus grandes valeurs propres et le sous espace propre associé.
- (c) Calculer l'inertie absorbée par M , sous espace propre associé aux 2 plus grandes valeurs propres.

1. Notre disquaire prend une classification plus fine des acheteurs et obtient le tableau de contingence :

$$S = \begin{pmatrix} & \text{chanson} & \text{rock} & \text{classique} \\ j.f. & 41 & 16 & 8 \\ j.h. & 28 & 25 & 10 \\ f. & 172 & 84 & 127 \\ h. & 133 & 118 & 157 \\ v.f. & 16 & 5 & 19 \\ v.h. & 11 & 6 & 24 \end{pmatrix}$$

avec : j. pour les jeunes, v. pour les vieux, f. pour les femmes, h. pour les hommes.

- Réaliser une analyse factorielle des correspondances sur ce tableau (on pourra récupérer le programme `afc1.sce` dans mon répertoire situé sur `uranie /u/profs/br`).
- Reporter sur votre feuille le dessin exécuté avec `scilab` ; y mettre les numéros des points et leur signification (type d'acheteur et de musique).
- Donner une interprétation des correspondances constatées ; comparer avec ce que le disquaire avait déjà obtenu dans l'étude réalisée en TD.
- Qualité globale de la représentation.
- Interprétation des axes.
- Par rapport au programme utilisé en TD, j'ai complété avec :

```
ulhuc=ul'*h'*uc;
for i=1:size(n,'c'),
    for j=1:size(n,'r'),
        if ulhuc(i,j) < 0 then
            ul(:,i)=-ul(:,i);
        end,
    end,
end,
[sl,kl]=sort(vp1);
ul2d=zeros(2,size(d,'c'));
for i=1:2
    ul2d(i,:)=ul(:,kl(i))*d
end
```

A quoi peut servir ce morceau de programme ? (la matrice h est celle appelée H dans le polycopi).

- Donner l'équation du plan passant par le centre de gravité de la matrice X des profils de colonnes et parallèle aux 2 vecteurs propres associés aux 2 plus grandes valeurs propres de la matrice de covariance.
2. On reprend l'étude sur les occupations quotidiennes suivant l'état civil.
- On considère le tableau de contingence ci-dessous. La population est américaine dans les années 1965/66.

$$A = \begin{pmatrix} & \text{prof} & \text{tran} & \text{mena} & \text{enfa} & \text{cour} & \text{toil} & \text{repas} & \text{somm} & \text{tele} & \text{lois} \\ ha & 610 & 140 & 60 & 10 & 120 & 95 & 115 & 760 & 175 & 315 \\ fa & 475 & 90 & 250 & 30 & 140 & 120 & 100 & 775 & 115 & 305 \\ fna & 10 & 0 & 495 & 110 & 170 & 110 & 130 & 785 & 160 & 430 \\ hm & 615 & 140 & 65 & 10 & 115 & 90 & 115 & 765 & 180 & 305 \\ fm & 179 & 29 & 421 & 87 & 161 & 112 & 119 & 776 & 143 & 373 \\ hc & 585 & 115 & 50 & 0 & 150 & 105 & 100 & 760 & 150 & 385 \\ fc & 482 & 94 & 196 & 18 & 141 & 130 & 96 & 775 & 132 & 336 \end{pmatrix}$$

avec : h : homme, f : femme, a : actif, na : non actif, m : mari, c : célibataire ; prof : activité professionnelle, tran : transport, mena : travail ménager, enfa : s'occuper des enfants, cour : courses, toil : toilette, repas : consommation des repas, somm : sommeil, tele : télévision, lois : loisirs autres que tl. Les nombres donnent le temps passé en centimes d'heure.

- i. Raliser une analyse factorielle des correspondances sur ce tableau (on pourra rcuprer le programme `afc.sce` dans mon rpertoire /br).
- ii. Reporter sur votre feuille le dessin excut avec `scilab` ; y mettre les numros des points et leur signification (type d'acheteur et de musique).
- iii. Donner une interprétation des correspondances constatées.
- iv. Qualit globale de la représentation.
- v. Interprtation des axes.

Contrôle analyse de données, licence MASS septembre 2001

Notes manuscrites et polycopi du cours autorisés.

1. (a) On considère le plan M de \mathbb{R}^3 engendré par les vecteurs v^1, v^2 de composantes :

$$[v^1 \ v^2] = \begin{bmatrix} 2 & 1 \\ -1 & 1 \\ 2 & 2 \end{bmatrix}$$

Calculer la matrice de la projection orthogonale sur le plan M .

- (b) Donner une équation paramétrique du plan affine M_a parallèle au plan M passant par le point $a = {}^t[1, 1, 1]$.
 (c) Calculer la projection du point a sur le plan vectoriel M .
 (d) Calculer le pivot de M_a .
 (e) Déterminer la projection de l'origine sur M_a .
 (f) Calculer l'inertie de $A = [a^1 \ a^2]$ par rapport à M_a avec les poids $p^1 = p^2$ avec la matrice

$$A = \begin{bmatrix} 0 & 2 \\ 0 & 2 \\ 0 & 2 \end{bmatrix}$$

2. On considère la matrice

$$A = \begin{bmatrix} rev/educ & non\ reponse & scol.obligat. & cap/bep & bepc & bac & superieur \\ non\ reponse & 54 & 134 & 70 & 51 & 65 & 66 \\ < 4000 & 5 & 66 & 29 & 15 & 14 & 17 \\ 4000 - 7000 & 16 & 155 & 127 & 53 & 82 & 48 \\ 7000 - 10000 & 7 & 75 & 107 & 41 & 74 & 102 \\ 10000 - 20000 & 3 & 30 & 62 & 29 & 80 & 147 \\ 20000 - 30000 & 0 & 6 & 4 & 1 & 4 & 24 \\ > 30000 & 0 & 0 & 0 & 1 & 0 & 2 \end{bmatrix}$$

Cette matrice correspond aux effectifs lors d'une enquête effectuée sur les diplômes et revenus d'un échantillon de population. On a réalisé une *analyse factorielle des correspondances*. On a tracé le dessin ci-dessous dans le plan des 2 premiers vecteurs propres des matrices de covariance.

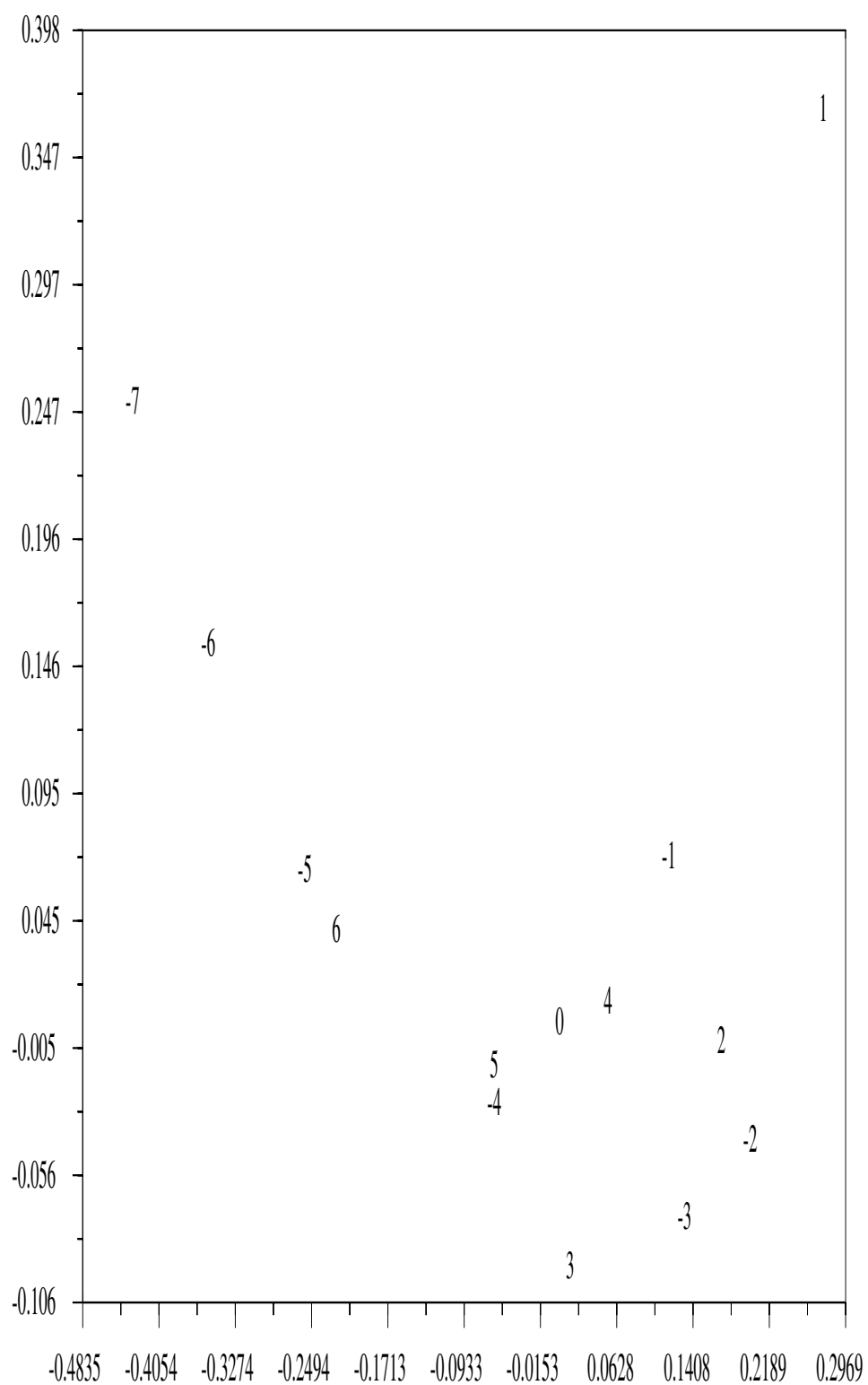
- (a) Indiquer brièvement de quelles matrices il s'agit ;
 (b) Comment calculer ces matrices (calcul effectif non demandé) ;
 (c) Comment calculer valeurs propres et vecteurs propres (calcul effectif non demandé) ;
 (d) On a les correspondances suivantes entre les nombres représentés sur le graphique et leur signification pour l'enquête :

$$\begin{bmatrix} non\ reponse & scol.obligat. & cap/bep & bepc & bac & superieur \\ 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}$$

$$\begin{bmatrix} non\ reponse & < 4000 & 4000 - 7000 & 7000 - 10000 & 10000 - 20000 & 20000 - 30000 & > 30000 \\ -1 & -2 & -3 & -4 & -5 & -6 & -7 \end{bmatrix}$$

Interpréter le premier axe obtenu sur le dessin ;

- (e) Quelles conclusions (sur les revenus et diplômes) relatives à l'enquête pouvez-vous tirer de l'analyse factorielle des correspondances ?



Contrôle continu : licence mass, analyse de données novembre 2001

Calculatrices interdites ; notes de cours, TD et photocopi distribués autorisés. tD désigne comme dans le poly la matrice transposée de D .

1. (a) En partant de la définition de l'inertie totale, montrer que l'on a bien la formule utilisée en TD (avec P et Q des matrices diagonales coefficients strictement positifs) :

$$I_T(D) = \text{Tr}(\sqrt{Q} DP {}^tD \sqrt{Q}) = \text{Tr}(DP {}^tD Q) = \text{Tr}(P {}^tD Q D) \quad (9.1)$$

- (b) Comment peut-on retrouver 9.1 avec la formule (0.28) du photocopi :

$$I_{M^\perp}(D) = \text{Tr}({}^tV \sqrt{Q} C \sqrt{Q} V) \quad (9.2)$$

2. On considère la matrice :

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 6 & 1 & 11 & 5 & 3 & 2 & 9 & 7 & 8 & 10 & 4 \\ 2 & 6 & 5 & 3 & 4 & 1 & 8 & 9 & 7 & 10 & 11 \end{bmatrix}$$

les lignes représentent le classement de 11 étudiants dans les matières suivantes : mathématiques, économie et anglais.

- (a) Dans une analyse en composantes principales du tableau A avec une matrice $Q = Id$, de quelle matrice calcule-t-on les valeurs et vecteurs propres ?
- (b) Comme tous les étudiants ont le même "poids", quelle matrice P suggérez-vous ?
- (c) Donner les formules à utiliser dans le logiciel **scilab** pour calculer cette matrice et ses valeurs propres et vecteurs propres ;
- (d) On trouve comme valeurs propres : 19.555122; -2.3283122; -8.1165657 quelle est la qualité globale de la représentation ?
- (e) Indiquer les calculs à effectuer et les formules à utiliser dans le logiciel **scilab** pour obtenir la figure 9.2 dans le plan des vecteurs propres associés aux 2 plus grandes valeurs propres (le choix des 2 plus grandes valeurs propres est fait manuellement).
- (f) La figure 9.3 représente les coefficients de corrélation des anciens caractères avec les nouveaux caractères ; quelles formules utiliser pour obtenir les coordonnées de ces points ?
- (g) Interpréter les nouveaux axes.
- (h) Le lendemain on dispose des classements en informatique :

$$A = \begin{bmatrix} 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 1 & 2 \end{bmatrix}$$

On réalise une ACP et on obtient la projection de la figure 9.4 et les corrélations de la figure 9.5. Interpréter les nouveaux axes.

- (i) on trouve comme valeurs propres : 19.557224; -1.9290897; -8.0815937; -10.432093; quelle est la qualité globale de la représentation ?
- (j) Commenter la position des étudiants avec 3 et 4 matières.
- (k) On réalise une nouvelle ACP dans laquelle on souhaite donner beaucoup d'importance au coefficient de l'informatique ; on prend une matrice Q avec $Q_{ii} = 1$ $i = 1, 2, 3$ et $Q_{44} = 9$. On trouve la représentation de la figure 9.6 pour les individus et de la figure 9.7 pour les caractères. Interpréter les nouveaux axes.
- (l) on trouve comme valeurs propres : 90.340847; -19.552669; -1.9988868; -8.1075973; quelle est la qualité globale de la représentation ?

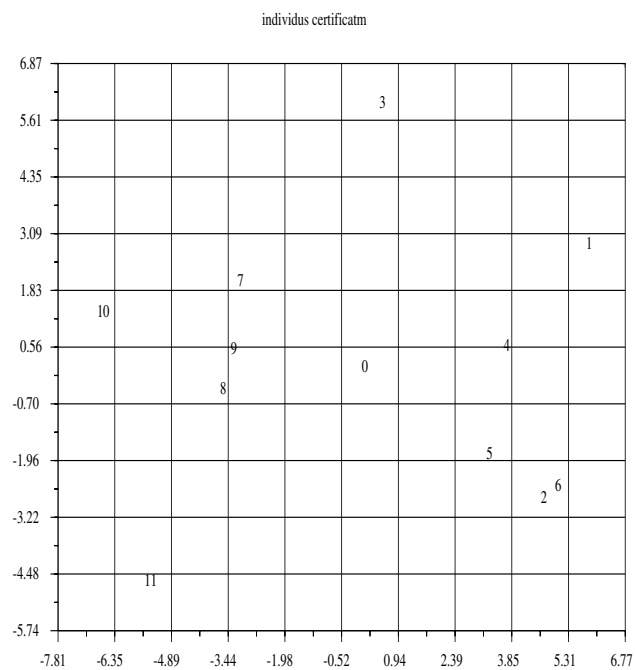


FIG. 9.2 – classement avec 3 matieres

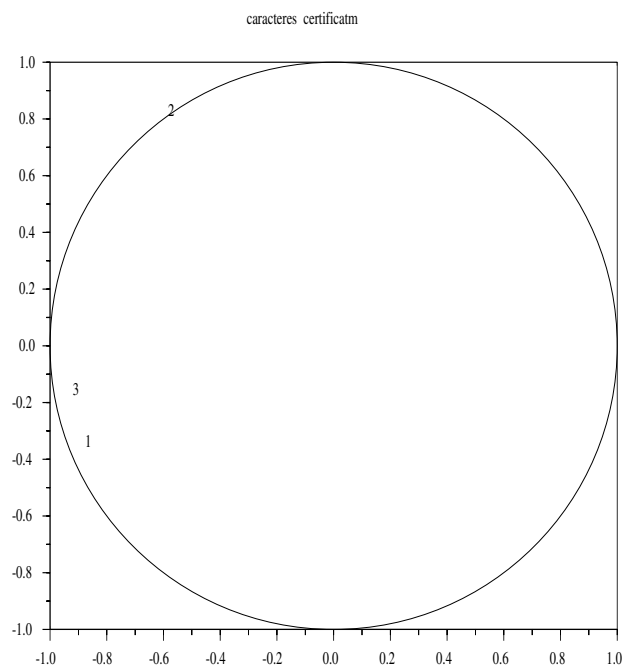


FIG. 9.3 – correlations pour 3 matieres

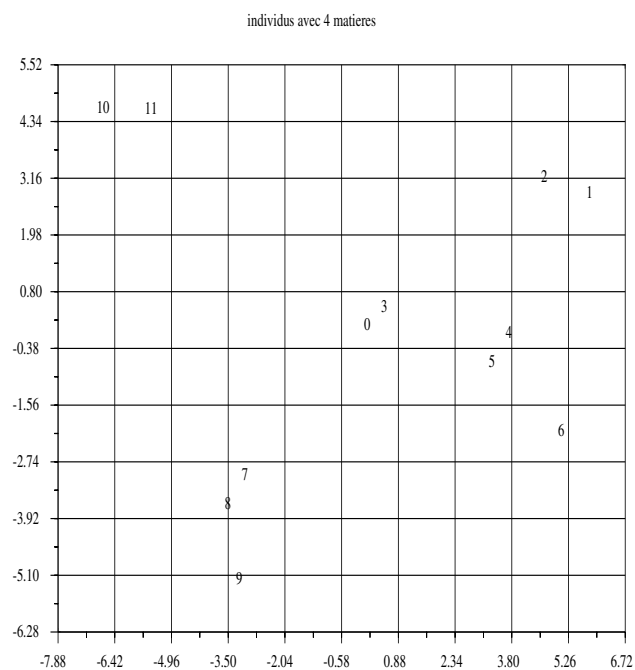


FIG. 9.4 – classement avec 4 matieres

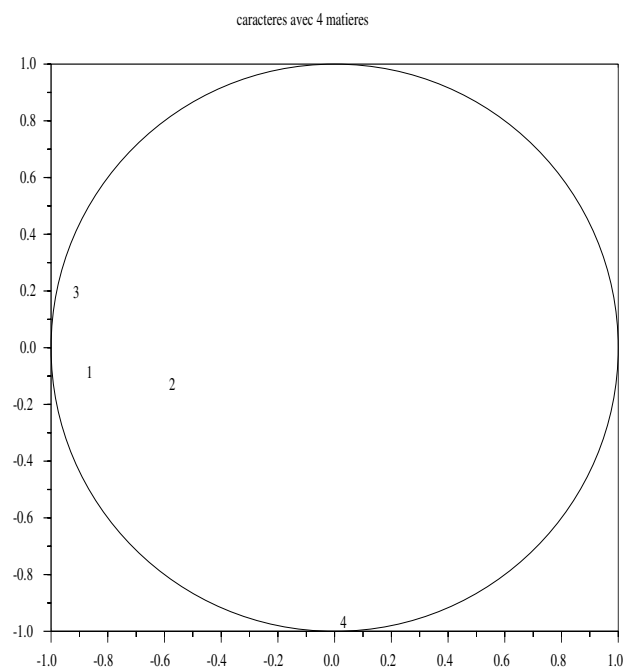


FIG. 9.5 – correlations avec 4 matieres

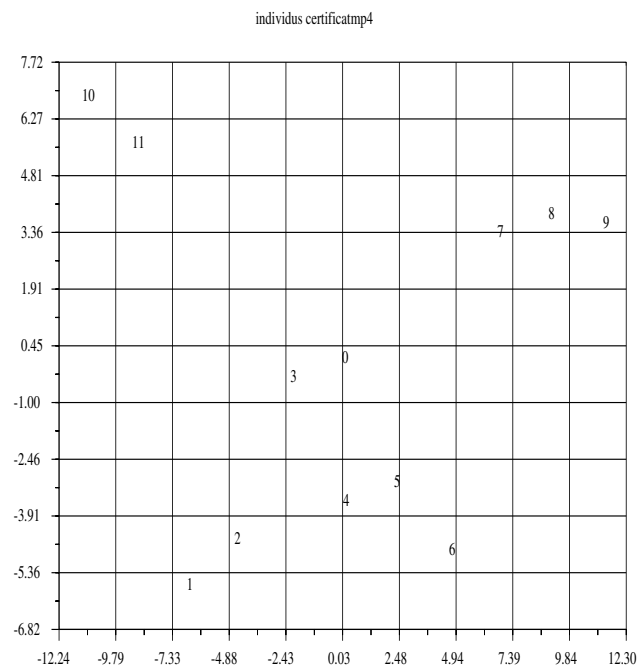


FIG. 9.6 – classement avec 4 matires pondres : 1,1,1,9

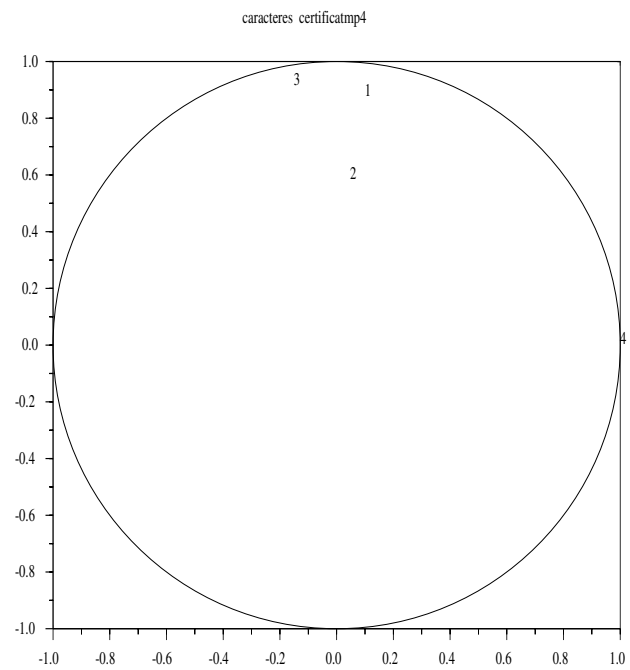


FIG. 9.7 – classement avec 4 matires pondres : 1,1,1,9

Examen analyse de données, lic. MASS 2001-2002. Notes manuscrites et polycopi du cours autorisés ; calculatrices interdites ; soigner la présentation.

1. Soit $\underline{y} \in \mathbb{R}^m$ une variable expliquée et $\underline{a}^1 \in \mathbb{R}^m$ une variable explicative ; les poids $q_j > 0$ vérifient : $\sum_{j=1}^m q_j = 1$ et l'on note $Q = \text{diag}(q_j)$
 - (a) Notons $\text{cov}(\underline{y}, \underline{a}^1) = \langle \underline{y}, \underline{a}^1 \rangle_Q - \bar{y} \bar{a}^1$ avec le produit scalaire $\langle \underline{y}, \underline{a}^1 \rangle_Q = \sum_{j=1}^m y_j q_j a_j^1$ et la moyenne $\bar{y} = \langle \underline{y}, \underline{a}^0 \rangle_Q$ avec ${}^t \underline{a}_0 = [1, \dots, 1]$ et idem pour les autres vecteurs considérés. Démontrer que $\text{cov}(\underline{y}, \underline{a}^1) = \langle \underline{y} - \bar{y} \underline{a}^0, \underline{a}^1 - \bar{a}^1 \underline{a}^0 \rangle_Q$
 - (b) Rappeler l'expression des 2 coefficients α_0 et α_1 de la droite des moindres carrs (droite de régression) : $\underline{y}^* = \alpha_0 \underline{a}^0 + \alpha_1 \underline{a}^1$ avec : ${}^t \underline{a}_0 = [1, \dots, 1]$ où l'on note t la transposée d'une matrice ou d'un vecteur.
 - (c) En notant $\text{var}(\underline{y}) = \text{cov}(\underline{y}, \underline{y})$, en déduire que

$$\text{var}(\underline{y}) = \text{var}(\underline{y} - \underline{y}^*) + \text{var}(\underline{y}^*) + 2\text{cov}(\underline{y} - \underline{y}^*, \underline{y}^*) \quad (9.3)$$

- (d) Montrer que $\text{cov}(\underline{y} - \underline{y}^*, \underline{y}^*) = 0$
- (e) Montrer que $\text{var}(\underline{y}^*) = \frac{\text{cov}(\underline{a}^1, \underline{y})^2}{\text{var}(\underline{a}^1)}$
- (f) En déduire : $\text{var}(\underline{y}) = \text{var}(\underline{y} - \underline{y}^*) + \rho(\underline{a}^1, \underline{y})^2 \text{var}(\underline{y})$
- (g) À l'aide de la formule précédente donner une nouvelle justification du critère retenu en cours pour la qualité de la régression.
- (h) On considère le tableau des 2 variables : population totale (variable explicative) et population active (variable expliquée) de quelques pays en 1970 :

pays	population	actifs
<i>Etatsunis</i>	204.8	78.62
<i>Japon</i>	103.5	50.94
<i>Allemagne</i>	61.5	26.7
<i>Angleterre</i>	55.8	24.7
<i>France</i>	50.7	20.41
<i>Italie</i>	54.4	18.77
<i>Canada</i>	21.4	7.87
<i>Paysbas</i>	13.0	4.56
<i>Belgique</i>	9.67	3.74

Dessiner main levée ces points sur un graphique avec la droite des moindres carrs

- (i) On a $\rho(\underline{a}^1, \underline{y}) = 0.988$; préciser la qualité de la régression.
2. On considère la matrice suivante

$$N = \begin{bmatrix} 20 & 20 & 1 & 1 & 2 & 18 & 9 \\ 20 & 10 & 20 & 10 & 9 & 3 & 17 \\ 20 & 1 & 10 & 20 & 16 & 8 & 3 \end{bmatrix}$$

Les lignes représentent les notes obtenues par 7 étudiants en mathématiques, économie, comptabilité ; chaque colonne fournit les 3 notes de chaque étudiant.

- (a) Pourquoi ce tableau est-il susceptible d'une analyse factorielle des correspondances ?
- (b) Calculer (ventuellement en partie) la matrice des profils des colonnes.
- (c) Calculer (ventuellement en partie) la matrice des profils des lignes.
- (d) Lors d'une analyse factorielle des correspondances, donner une équation paramétrique du plan sur lequel on projette (aucune valeur numérique n'est demandée).
- (e) Que dessine-t-on en pratique ?
- (f) La figure 9.12 représente les projections des étudiants et des 3 matières sur le plan précis ci-dessus. Donner un programme scilab pour le calcul de ces projections
- (g) Interpréter les correspondances relevées sur le graphique.

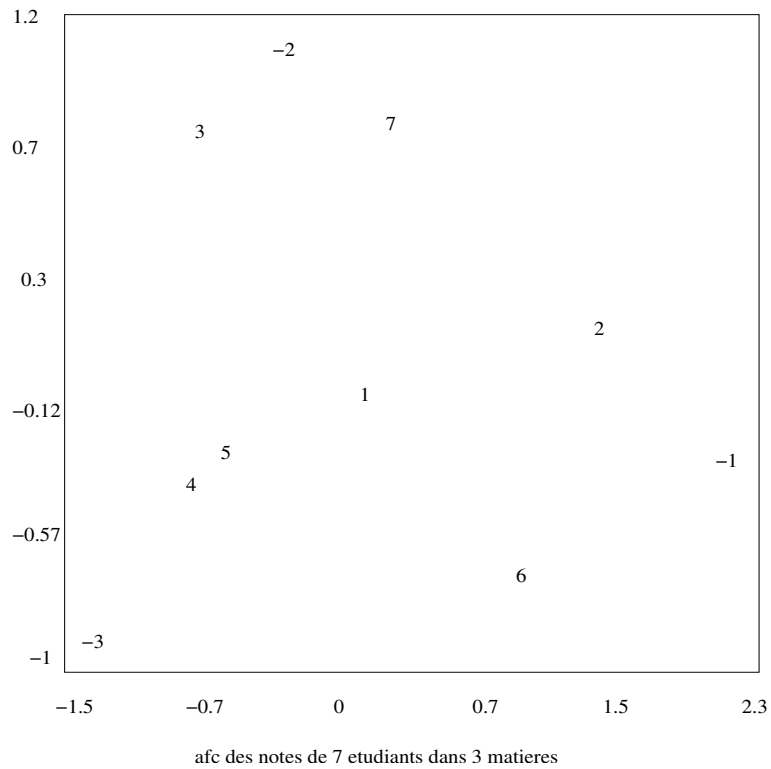


FIG. 9.8 –

- (h) Interpréter les axes.
3. On considère le tableau des individus suivants : $A = [-4, -3, -2, 1, 2, 3, 4, 5, 6]$ On souhaite répartir ces individus en 2 classes.
- En partant de la partition $\Pi_1 = \{C_1^1, C_1^2\}$ avec $C_1^1 = [-4, 2, 3, 4]$ et $C_1^2 = [-3, -2, 1, 5, 6]$ effectuer une première itération de la méthode des nuées dynamiques.
 - Effectuer une deuxième itération.
 - Évaluer la qualité de la partition

Examen analyse de données, lic. MASS 2001-2002, septembre 2002.

Notes manuscrites et polycopi du cours autorisés ; calculatrices interdites ; soigner la présentation et la rédaction.

1. On considère le tableau des individus suivants : $A = [-4, -3, -2, 1, 2, 3, 4, 5, 6]$
On souhaite répartir ces individus en 2 classes.
 - (a) En partant de la partition $\Pi_1 = \{C_1^1, C_1^2\}$ avec $C_1^1 = [-4, 2, 3, 4]$ et $C_1^2 = [-3, -2, 1, 5, 6]$ effectuer une première itération de la méthode des nues dynamiques.
 - (b) Effectuer une deuxième itération.
 - (c) Évaluer la qualité de la partition
2. (a) On considère le plan M de \mathbb{R}^2 engendré par les vecteurs v^1, v^2 de composantes :

$$[v^1 \ v^2] = \begin{bmatrix} 2 & 1 \\ -1 & 1 \\ 2 & 2 \end{bmatrix}$$

Calculer la matrice de la projection orthogonale sur le plan M .

- (b) Donner une équation paramétrique du plan affine M_a parallèle au plan M passant par le point $a = {}^t[1, 1, 1]$.
- (c) Calculer la projection du point a sur le plan vectoriel M .
- (d) Calculer le pivot de M_a .
- (e) Déterminer la projection de l'origine sur M_a .
- (f) Calculer l'inertie de $A = [a^1 \ a^2]$ par rapport à M_a avec les poids $p^1 = p^2$ avec la matrice

$$A = \begin{bmatrix} 0 & 2 \\ 0 & 2 \\ 0 & 2 \end{bmatrix}$$

3. On considère la matrice suivante

$$N = \begin{bmatrix} 20 & 20 & 1 & 1 & 2 & 18 & 9 \\ 20 & 10 & 20 & 10 & 9 & 3 & 17 \\ 20 & 1 & 10 & 20 & 16 & 8 & 3 \end{bmatrix}$$

Les lignes représentent les notes obtenues par 7 étudiants en mathématiques, économie, comptabilité ; chaque colonne fournit les 3 notes de chaque étudiant.

- (a) Pourquoi ce tableau est-il susceptible d'une analyse factorielle des correspondances ?
- (b) Calculer (ventuellement en partie) la matrice des profils des colonnes.
- (c) Calculer (ventuellement en partie) la matrice des profils des lignes.
- (d) Lors d'une analyse factorielle des correspondances, donner une équation paramétrique du plan sur lequel on projette (aucune valeur numérique n'est demandée).
- (e) Que dessine-t-on en pratique ?
- (f) La figure 9.12 représente les projections des étudiants et des 3 matières sur le plan précis ci-dessus. Donner un programme Scilab pour le calcul de ces projections
- (g) Interpréter les correspondances relevées sur le graphique.
- (h) Interpréter les axes.

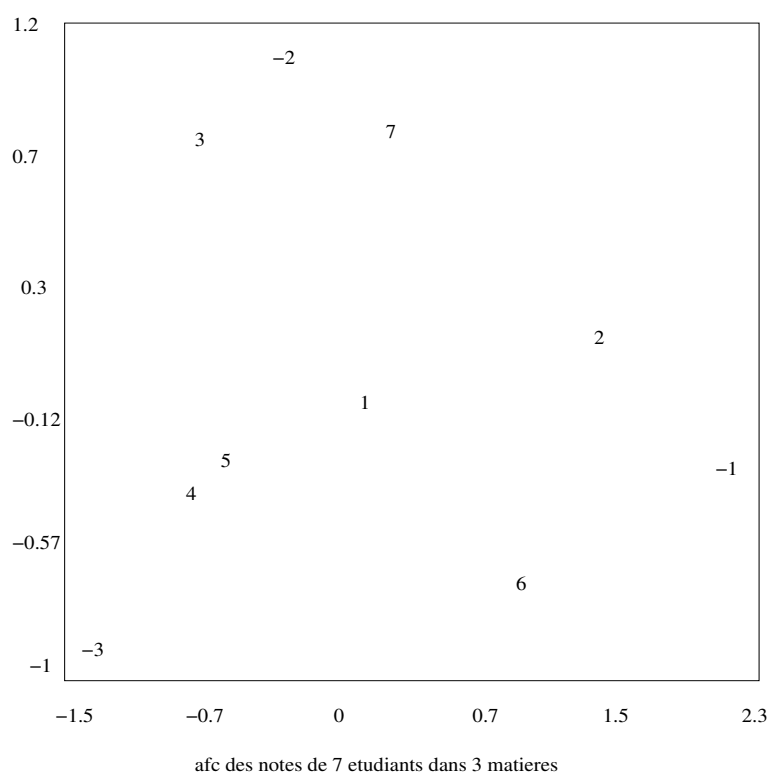


FIG. 9.9 –

Contrôle continu : licence mass, analyse de données novembre 2002

Calculatrices interdites ; notes de cours, TD et polycopi distribu autoriss.

tD dsigne comme dans le poly la matrice transpose de $\underline{\underline{D}}$.

1. (a) On considere 2 vecteurs $\underline{v}^1, \underline{v}^2$ de \mathbb{R}^m ; on note $\underline{\underline{V}} = [\underline{v}^1 \ \underline{v}^2]$ et soit $\underline{\underline{Q}}$ une matrice symtrique non ncessairement diagonale valeurs propres strictement positives.
Montrer que

$$\forall \underline{w} \in \mathbb{R}^m \quad {}^t\underline{\underline{w}}\underline{\underline{Q}}\underline{w} \geq \alpha {}^t\underline{w}\underline{w} \quad (9.4)$$

avec $\alpha > 0$ que l'on explicitera avec les valeurs propres de $\underline{\underline{Q}}$.

- (b) Soit $\underline{\underline{G}} = {}^t\underline{\underline{V}}\underline{\underline{Q}}\underline{\underline{V}}$; quelles sont les dimensions de $\underline{\underline{G}}$?
- (c) Peut-on multiplier $\underline{\underline{G}}$ et ${}^t\underline{\underline{V}}$?
- (d) On considere la matrice : $\underline{\underline{V}}^* = \underline{\underline{V}}\underline{\underline{G}}^{-1}$; quelles sont ses dimensions ?
- (e) Soit :

$$\underline{v}^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \underline{v}^2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (9.5)$$

Avec $\underline{\underline{Q}} = Id$, Calculer $\underline{\underline{G}}, \underline{\underline{G}}^{-1}, \underline{\underline{V}}^*$; dessiner $\underline{v}^1, \underline{v}^2, \underline{v}^{1*}, \underline{v}^{2*}$

- (f) On reprend pour la suite $\underline{v}^1, \underline{v}^2$ quelconques ; on note $\underline{\underline{V}}^* = [\underline{v}^{1*}, \underline{v}^{2*}]$. Montrer les identits $\underline{\underline{G}}^{-1} = {}^t\underline{\underline{V}}^*\underline{\underline{Q}}\underline{\underline{V}}^*$ et $\underline{\underline{G}} {}^t\underline{\underline{V}}^* = {}^t\underline{\underline{V}}$.
- (g) Explicier les coefficients de ${}^t\underline{\underline{V}}^*\underline{\underline{Q}}\underline{x}$ avec $\underline{x} \in \mathbb{R}^m$.
- (h) Preciser la dimension et expliciter les coefficients de $\underline{\underline{V}} {}^t\underline{\underline{V}}^*\underline{\underline{Q}}\underline{x}$; quel rapport avec :

$$\langle \underline{v}^{1*}, \underline{x} \rangle_Q \underline{v}^1 + \langle \underline{v}^{2*}, \underline{x} \rangle_Q \underline{v}^2 \quad (9.6)$$

- (i) Dans le cas o \underline{x} est dans le sous espace engendr par $\underline{v}^1, \underline{v}^2$ que vaut $\underline{\underline{V}} {}^t\underline{\underline{V}}^*\underline{\underline{Q}}\underline{x}$?
- (j) Montrer que $\underline{x} - \underline{\underline{V}} {}^t\underline{\underline{V}}^*\underline{\underline{Q}}\underline{x}$ est orthogonal $\underline{v}^1, \underline{v}^2$.

2. On considere la matrice :

$$A = \begin{bmatrix} 10. & 9. & 9. & 11. & 10. & 10. & 10. & 12. & 18. & 3. \\ 18. & 18. & 18. & 18. & 20. & 3. & 3. & 3. & 10. & 10. \\ 10.5 & 10. & 19. & 10. & 15. & 19. & 10. & 10. & 3. & 18. \end{bmatrix}$$

les lignes representent les notes de 10 tudians dans les matires suivantes : mathmatiques, conomie et anglais.

- (a) Dans une analyse en composantes principales du tableau $\underline{\underline{A}}$ avec une matrice $\underline{\underline{Q}} = Id$, de quelle matrice calcule-t-on les valeurs et vecteurs propres ?
- (b) Comme tous les tudians ont le mme "poids", quelle matrice $\underline{\underline{P}}$ suggrez vous ?
- (c) Donner les formules utiliser dans le logiciel **scilab** pour calculer la matrice demande questions plus haut et ses valeurs propres et vecteurs propres ;
- (d) On trouve comme valeurs propres : 0.1757001; 0.0253222; 0.0989777 quelle est la qualit globale de la representation ?
- (e) Indiquer les calculs effectuer et les formules utiliser dans le logiciel **scilab** pour obtenir la figure 9.10 dans le plan des vecteurs propres associs aux 2 plus grandes valeurs propres (le choix des 2 plus grandes valeurs propres est fait manuellement). On a reprsent les caractres avec un nombre ngatif et les individus avec un nombre positif .
- (f) Interprter les nouveaux axes.
- (g) discuter la position de quelques individus typiques.
- (h) On considere une 4me matire EPS ; on choisit les coefficients diagonaux de $\underline{\underline{Q}}$: $q_1 = q_2 = q_3 = 0.3$ et $q_4 = .1$; reprendre la question (a) avec ces donnes.
- (i) Reprendre la question(e).

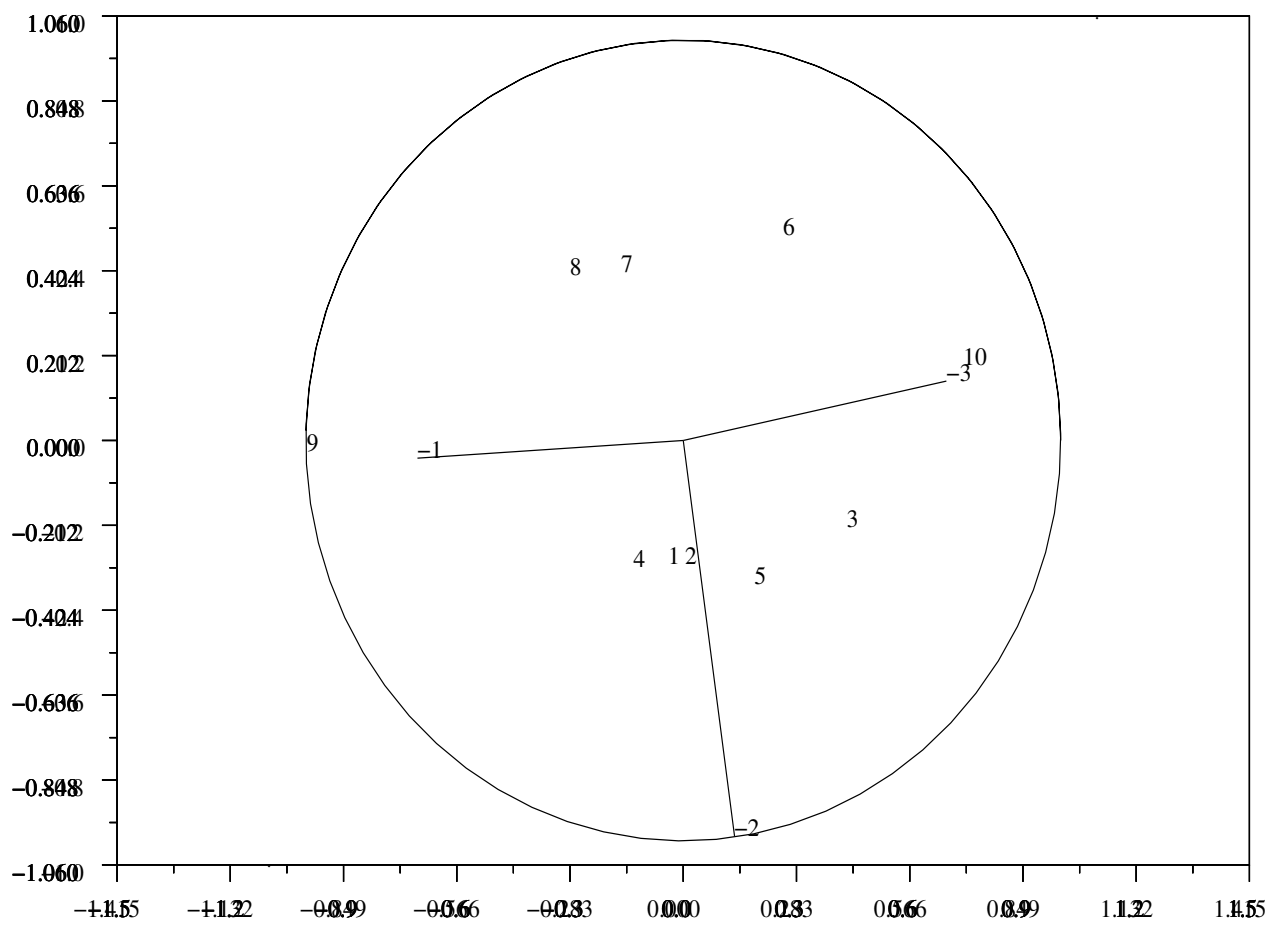


FIG. 9.10 – classement avec 3 matieres

Examen analyse de données, lic. MASS 2002-2003, janvier 03. Notes manuscrites et polycopi du cours autorisés ; calculatrices interdites ; soigner la présentation.

1. On considère les matrices suivantes :

$$\underline{\underline{A_1}} = \begin{bmatrix} 1 & 0 & 0 & 2 & -1 \\ 3 & 0 & -3 & 0 & 5 \\ 0 & 0 & 1 & 0 & -3 \end{bmatrix} \quad \underline{\underline{A_2}} = \begin{bmatrix} 0 & 0 & -4 & 0 & 0 \\ 1 & 0 & 0 & -1 & 3 \end{bmatrix} \quad \underline{\underline{A_3}} = \begin{bmatrix} 0 & 0 & 7 \\ 0 & 1 & -1 \\ 5 & 0 & -3 \\ 1 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix} \quad (9.7)$$

$$\underline{\underline{A_4}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 3 \end{bmatrix} \quad \underline{\underline{A_5}} = \begin{bmatrix} 0 & 1 \\ 1 & 5 \\ 1 & -3 \\ -2 & 0 \\ -3 & 0 \end{bmatrix} \quad \underline{\underline{A_6}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -5 \end{bmatrix} \quad (9.8)$$

- Indiquer dans un tableau tous les produits matriciels qui ont un sens ; indiquer la taille de ces produits ; indiquer aussi les produits qui n'ont pas de sens.
 - Effectuer 3 de ces produits matriciels.
2. Pour réaliser une approximation par moindres carrés, on peut considérer d'autres fonctions qu'une droite $a \rightarrow \alpha + \beta a$, par exemple une parabole ; on considère donc les 3 vecteurs $\underline{\underline{A}} = [\underline{a}^0, \underline{a}, \underline{a}^2]$ où $\underline{a}^0 \in \mathbb{R}^m$, $\underline{a} \in \mathbb{R}^m$, $\underline{a}^2 \in \mathbb{R}^m$ avec toutes les composantes de a distinctes, $m \geq 4$, ${}^t \underline{a}^0 = [1 \dots 1]$; \underline{a}^2 désigne le vecteur dont les composantes sont les carrés des composantes de \underline{a} ; dans la suite on désigne par \underline{y} , le vecteur "expliquer".
- Montrer que la matrice $\underline{\underline{A}}$ est de rang 3.
 - Ecrire une relation en escalier pour calculer \underline{a}^2 à partir de \underline{a} .
 - On considère le vecteur ${}^t \underline{a} = [20 \ 30 \ 40 \ 50 \ 60 \ 70]$, calculer le vecteur \underline{a}^2 .
 - On considère la fonction

$$F(\alpha, \beta, \gamma) = \sum_{j=1}^m q_j (y_j - \alpha a_j^0 - \beta a_j - \gamma a_j^2)^2 \quad (9.9)$$

avec tous les $q_j > 0$; quelle est la condition nécessaire pour un minimum de F ?

- Expliciter cette condition et faire apparaître un système linéaire pour déterminer : α, β, γ .
 - Vérifier que la matrice K de ce système est égale à ${}^t \underline{\underline{A}} \underline{\underline{Q}} \underline{\underline{A}}$ avec $\underline{\underline{Q}} = \text{diag}(q_j)$.
 - En déduire que cette matrice vérifie ${}^t \underline{x} K \underline{x} > 0$ pour tout $x \in \mathbb{R}^3$.
 - La solution du système linéaire vérifie-t-elle une condition suffisante pour réaliser le minimum de F dans \mathbb{R}^3 .
 - Pour ${}^t \underline{y} = [54 \ 90 \ 138 \ 206 \ 292 \ 396]$, on trouve $\alpha = -0.91$ $\beta = 0.92$ $\gamma = 0.066$, dessiner ces données et cette parabole des moindres carrés. Tracer main levée la droite des moindres carrés et comparer.
3. On considère les individus :

$$A = \begin{bmatrix} -2 & -1 & 1 & 4 & 6 & 7 & 8 \end{bmatrix} \quad (9.10)$$

Le tableau ci-dessous fournit une approximation numérique des barycentres (calculs avec des poids égaux) $\overline{a_{i,j}}$ de $a^i \dots a^j$ avec valeur nulle si $j < i$.

-2.000	-1.500	-0.667	0.500	1.600	2.500	3.286
0.000	-1.000	0.000	1.333	2.500	3.400	4.167
0.000	0.000	1.000	2.500	3.667	4.500	5.200
0.000	0.000	0.000	4.000	5.000	5.667	6.250
0.000	0.000	0.000	0.000	6.000	6.500	7.000
0.000	0.000	0.000	0.000	0.000	7.000	7.500
0.000	0.000	0.000	0.000	0.000	0.000	8.000

(9.11)

Le tableau ci-dessous donne une approximation des inerties $I_{i,j}$ de $a^i \dots a^j$ utilisés dans l'algorithme de Fischer ; ainsi $I_{1,3} = 1.566$

0.000	0.250	1.556	5.250	9.040	11.583	13.633
0.000	0.000	1.000	4.222	7.250	9.040	10.472
0.000	0.000	0.000	2.250	4.222	5.250	6.160
0.000	0.000	0.000	0.000	1.000	1.556	2.188
0.000	0.000	0.000	0.000	0.000	0.250	0.667
0.000	0.000	0.000	0.000	0.000	0.000	0.250
0.000	0.000	0.000	0.000	0.000	0.000	0.000

(9.12)

Le tableau ci-dessous, fournit une approximation les sommes $I_{i,d} + I_{d+1,7}$

10.472	6.410	3.743	5.917	9.290	11.583
0.000	6.160	3.188	4.889	7.500	9.040
0.000	0.000	2.188	2.917	4.472	5.250
0.000	0.000	0.000	0.667	1.250	1.556
0.000	0.000	0.000	0.000	0.250	0.250
0.000	0.000	0.000	0.000	0.000	0.000

(9.13)

- (a) Interpréter les zéros de ces 3 tableaux et vérifier 3 coefficients non nuls.
- (b) De ces tableaux, déduire la meilleure partition en 2 classes par l'algorithme de Fischer.
- (c) Déterminer ensuite les meilleures partitions en 2 classes des ensembles $a^i \dots a^7$.
- (d) Déterminer la meilleure partition en 3 classes des individus proposés.
- (e) En partant de la partition (-2) , $(-1 \ 1 \ 4)$ $(6 \ 7 \ 8)$, utiliser l'algorithme des nues dynamiques.
- (f) Comparer les inerties des partitions trouvées avec les 2 algorithmes.

Examen analyse de données, lic. MASS 2002-2003, septembre 03.

Notes manuscrites et polycopi du cours autorisés ; calculatrices interdites ; soigner la présentation.

1. Soit $\underline{y} \in \mathbb{R}^m$ une variable à expliquer et $\underline{a}^1 \in \mathbb{R}^m$ une variable explicative ; les poids $q_j > 0$ vérifient : $\sum_{j=1}^m q_j = 1$ et l'on note $Q = \text{diag}(q_j)$
 - (a) Notons $\text{cov}(\underline{y}, \underline{a}^1) = \langle \underline{y}, \underline{a}^1 \rangle_Q - \bar{y} \bar{a}^1$ avec le produit scalaire $\langle \underline{y}, \underline{a}^1 \rangle_Q = \sum_{j=1}^m y_j q_j a_j^1$ et la moyenne $\bar{y} = \langle \underline{y}, \underline{a}^0 \rangle_Q$ avec ${}^t \underline{a}_0 = [1, \dots, 1]$ et idem pour les autres vecteurs considérés. Démontrer que $\text{cov}(\underline{y}, \underline{a}^1) = \langle \underline{y} - \bar{y} \underline{a}^0, \underline{a}^1 - \bar{a}^1 \underline{a}^0 \rangle_Q$
 - (b) Rappeler l'expression des 2 coefficients α_0 et α_1 de la droite des moindres carrés (droite de régression) : $\underline{y}^* = \alpha_0 \underline{a}^0 + \alpha_1 \underline{a}^1$ avec : ${}^t \underline{a}_0 = [1, \dots, 1]$ où l'on note t la transposée d'une matrice ou d'un vecteur.
 - (c) En notant $\text{var}(\underline{y}) = \text{cov}(\underline{y}, \underline{y})$, en déduire que

$$\text{var}(\underline{y}) = \text{var}(\underline{y} - \underline{y}^*) + \text{var}(\underline{y}^*) + 2\text{cov}(\underline{y} - \underline{y}^*, \underline{y}^*) \quad (9.14)$$

- (d) Montrer que $\text{cov}(\underline{y} - \underline{y}^*, \underline{y}^*) = 0$
- (e) Montrer que $\text{var}(\underline{y}^*) = \frac{\text{cov}(\underline{a}^1, \underline{y})^2}{\text{var}(\underline{a}^1)}$
- (f) En déduire : $\text{var}(\underline{y}) = \text{var}(\underline{y} - \underline{y}^*) + \rho(\underline{a}^1, \underline{y})^2 \text{var}(\underline{y})$
- (g) À l'aide de la formule précédente donner une nouvelle justification du critère retenu en cours pour la qualité de la régression.
- (h) On considère le tableau des 2 variables : population totale (variable explicative) et population active (variable à expliquer) de quelques pays en 1970 :

pays	population	actifs
Etats-unis	204.8	78.62
Japon	103.5	50.94
Allemagne	61.5	26.7
Angleterre	55.8	24.7
France	50.7	20.41
Italie	54.4	18.77
Canada	21.4	7.87
Pays-bas	13.0	4.56
Belgique	9.67	3.74

Dessiner main levée ces points sur un graphique avec la droite des moindres carrés

- (i) On a $\rho(\underline{a}^1, \underline{y}) = 0.988$; préciser la qualité de la régression.

2. On considère la matrice

$$A = \begin{bmatrix} \text{rev/educ} & \text{non reponse} & \text{scol.obligat.} & \text{cap/bep} & \text{bepc} & \text{bac} & \text{supérieur} \\ \text{non reponse} & 54 & 134 & 70 & 51 & 65 & 66 \\ < 4000 & 5 & 66 & 29 & 15 & 14 & 17 \\ 4000 - 7000 & 16 & 155 & 127 & 53 & 82 & 48 \\ 7000 - 10000 & 7 & 75 & 107 & 41 & 74 & 102 \\ 10000 - 20000 & 3 & 30 & 62 & 29 & 80 & 147 \\ 20000 - 30000 & 0 & 6 & 4 & 1 & 4 & 24 \\ > 30000 & 0 & 0 & 0 & 1 & 0 & 2 \end{bmatrix}$$

Cette matrice correspond aux effectifs lors d'une enquête effectuée sur les diplômes et revenus d'un échantillon de population. On a réalisé une *analyse factorielle des correspondances*. On a tracé le dessin ci-dessous dans le plan des 2 premiers vecteurs propres des matrices de covariance.

- (a) Indiquer brièvement de quelles matrices il s'agit ;

- (b) Comment calculer ces matrices (calcul effectif non demand) ;
 (c) Comment calculer valeurs propres et vecteurs propres (calcul effectif non demand) ;
 (d) On a les correspondances suivantes entre les nombres represents sur le graphique et leur signification pour l'enquete :

$$\begin{bmatrix} non\ reponse & scol.obligat. & cap/ bep & bepc & bac & superieur \\ 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}$$

$$\begin{bmatrix} non\ reponse & < 4000 & 4000 - 7000 & 7000 - 10000 & 10000 - 20000 & 20000 - 30000 & > 30000 \\ -1 & -2 & -3 & -4 & -5 & -6 & -7 \end{bmatrix}$$

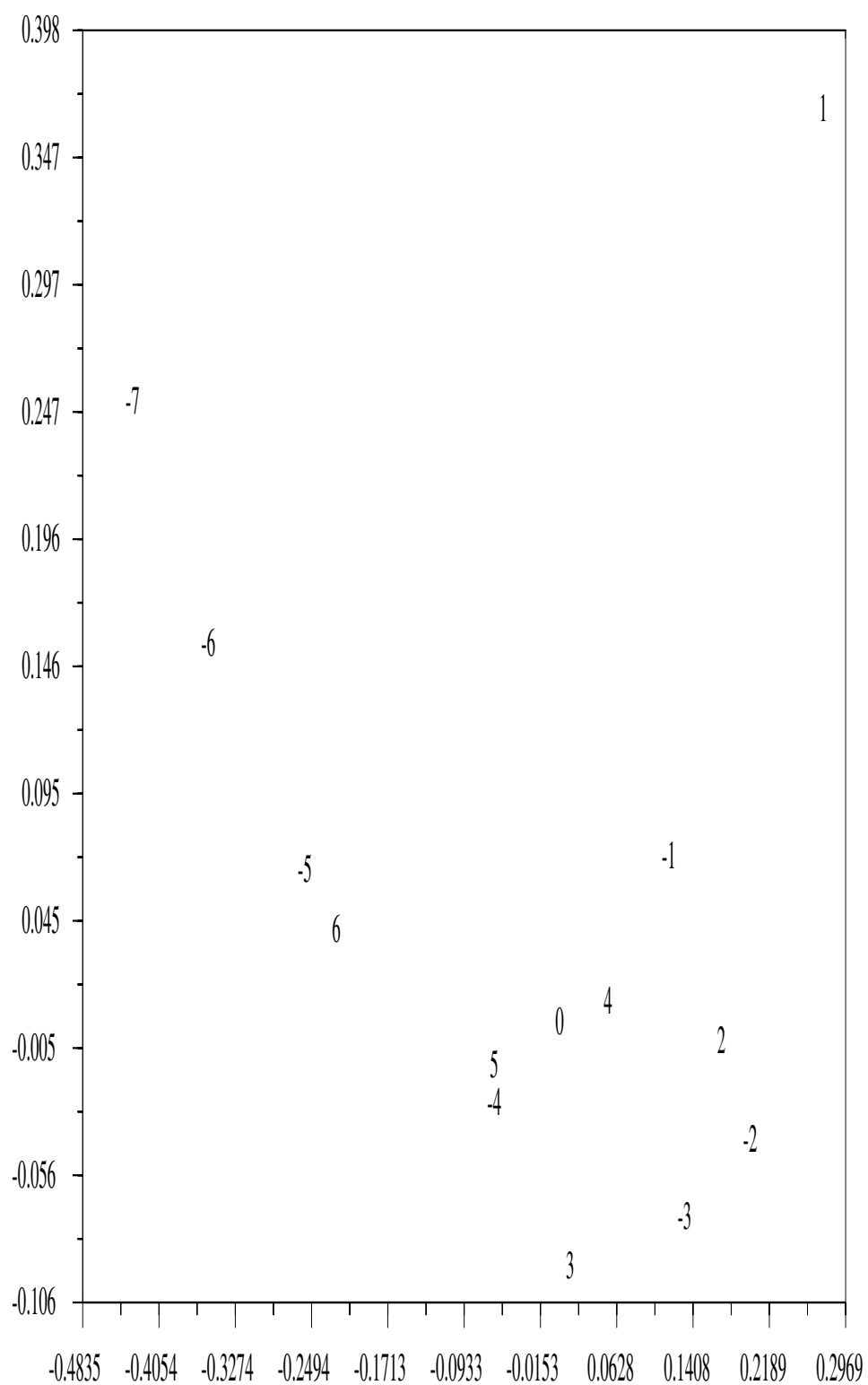
Interprter le premier axe obtenu sur le dessin ;

- (e) Quelles conclusions (sur les revenus et diplmes) relatives l'enquete pouvez vous tirer de l'analyse factorielle des correspondances ?
3. (a) On considere le plan M de \mathbb{R}^3 engendr par les vecteurs v^1 v^2 de composantes :

$$[v^1 \ v^2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \epsilon & \epsilon \end{bmatrix}$$

Calculer la matrice de la projection orthogonale sur le plan M .

- (b) Dans le cas $\epsilon = 0$, que trouvez vous ? Interpretation.



Examen analyse de donnees, lic. MASS 2003-2004, janvier 03. Notes manuscrites et polycopi du cours autoriss; calculettes autorises, micro-ordinateurs interdits; soigner la prsentation.

1. (a) On considere, dans l'espace 3 dimensions, le plan passant par le point $a = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$ et parallle aux 2

vecteurs $[v_1 \ v_2] = \begin{pmatrix} 1 & 0 \\ -1 & -1 \\ -1 & 1 \end{pmatrix}$. Vrifier que ce systme est libre. Donner une quation paramtrique de ce plan.

- (b) En dduire le systme linare qui fournit les composantes dans la base $[v_1, v_2]$ de $x - a$ avec x , projection orthogonale du vecteur $b = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$

(c) Vrifier votre rsultat.

2. (a) On considere le tableau A de 6 individus et 3 caractres :

$$A = \begin{bmatrix} 1 & 0 & 0 & 2 & 1 & 2 \\ 0 & 0 & 1 & 2 & 0 & 3 \\ 0 & 1 & 2 & 1 & 0 & 2 \end{bmatrix}$$

On suppose que tous les individus ont le mme poids et l'on met des poids units sur les lignes (caractres) : $\underline{\underline{Q}} = \underline{\underline{Id}}$.

- (b) Quelle matrice $\underline{\underline{P}}$ suggrez vous ?
(c) Dterminer le barycentre des colonnes.
(d) Calculer la matrice centre $\underline{\underline{D}}$.
(e) Dans une analyse en composantes principales du tableau $\underline{\underline{A}}$ avec une matrice $\underline{\underline{Q}} = \underline{\underline{Id}}$, de quelle matrice $\underline{\underline{C}}$ calcule-t-on les valeurs et vecteurs propres ?
(f) Vrifier que 0 est valeur propre de $\underline{\underline{C}}$ et qu'un vecteur propre est :

$$\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

- (g) Dterminer les valeurs propres et vecteurs propres $\underline{\underline{u}}^j$ de $\underline{\underline{C}}$.
(h) Calculer les projections utilises dans une analyse en composantes principales avec 2 vecteurs propres.
(i) Ecrire un programme en scilab qui calcule ces projections.
(j) Dessiner ces projections.
(k) Qualit globale de la reprsentation.
(l) Preciser les nouveaux caractres (caractres dans la base de vecteurs propres) $({}^t\underline{\underline{u}}^j \underline{\underline{Q}} \underline{\underline{D}})$.
(m) Calculer les cosinus entre les anciens caractres ${}^t e_j \underline{\underline{D}}$ et les nouveaux caractres (coefficients de corrlation).
(n) Dessin et commentaires.

Examen analyse de données, lic. MASS 2003-2004, septembre 2004.

Notes manuscrites et polycopié du cours autorisés ; calculatrices interdites ; soigner la présentation et la rédaction.

1. (a) On considère le plan M de \mathbb{R}^3 engendré par les vecteurs v^1 v^2 de composantes :

$$[v^1 \ v^2] = \begin{bmatrix} 2 & 1 \\ -1 & 1 \\ 2 & 2 \end{bmatrix}$$

Calculer la matrice de la projection orthogonale sur le plan M .

- (b) Donner une équation paramétrique du plan affine M_a parallèle au plan M passant par le point $a = {}^t[1, 1, 1]$.
(c) Calculer la projection du point a sur le plan vectoriel M .
(d) Calculer le pivot de M_a .
(e) Déterminer la projection de l'origine sur M_a .
(f) Calculer l'inertie de $A = [a^1 \ a^2]$ par rapport M_a avec les poids $p^1 = p^2$ avec la matrice

$$A = \begin{bmatrix} 0 & 2 \\ 0 & 2 \\ 0 & 2 \end{bmatrix}$$

2. On considère la matrice suivante

$$N = \begin{bmatrix} 20 & 20 & 1 & 1 & 2 & 18 & 9 \\ 20 & 10 & 20 & 10 & 9 & 3 & 17 \\ 20 & 1 & 10 & 20 & 16 & 8 & 3 \end{bmatrix}$$

Les lignes représentent les notes obtenues par 7 étudiants en mathématiques, économie, comptabilité ; chaque colonne fournit les 3 notes de chaque étudiant.

- (a) Pourquoi ce tableau est-il susceptible d' une analyse factorielle des correspondances ?
(b) Calculer (éventuellement en partie) la matrice des profils des colonnes.
(c) Calculer (éventuellement en partie) la matrice des profils des lignes.
(d) Lors d' une analyse factorielle des correspondances, donner une équation paramétrique du plan sur lequel on projette (aucune valeur numérique n' est demandée).
(e) Que dessine-t-on en pratique ?
(f) La figure 9.12 représente les projections des étudiants et des 3 matières sur le plan précisé ci dessus. Donner un programme scilab pour le calcul de ces projections
(g) Interpréter les correspondances relevées sur le graphique.
(h) Interpréter les axes.
3. On considère les matrices suivantes :

$$\underline{\underline{A_1}} = \begin{bmatrix} 1 & 0 & 0 & 2 & -1 \\ 3 & 0 & -3 & 0 & 5 \\ 0 & 0 & 1 & 0 & -3 \end{bmatrix} \quad \underline{\underline{A_2}} = \begin{bmatrix} 0 & 0 & -4 & 0 & 0 \\ 1 & 0 & 0 & -1 & 3 \end{bmatrix} \quad \underline{\underline{A_3}} = \begin{bmatrix} 0 & 0 & 7 \\ 0 & 1 & -1 \\ 5 & 0 & -3 \\ 1 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix} \quad (9.15)$$

$$\underline{\underline{A_4}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 3 \end{bmatrix} \quad \underline{\underline{A_5}} = \begin{bmatrix} 0 & 1 \\ 1 & 5 \\ 1 & -3 \\ -2 & 0 \\ -3 & 0 \end{bmatrix} \quad \underline{\underline{A_6}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -5 \end{bmatrix} \quad (9.16)$$

- (a) Indiquer dans un tableau tous les produits matriciels qui ont un sens ; indiquer la taille de ces produits ; indiquer aussi les produits qui n'ont pas de sens.
(b) Effectuer 3 de ces produits matriciels.

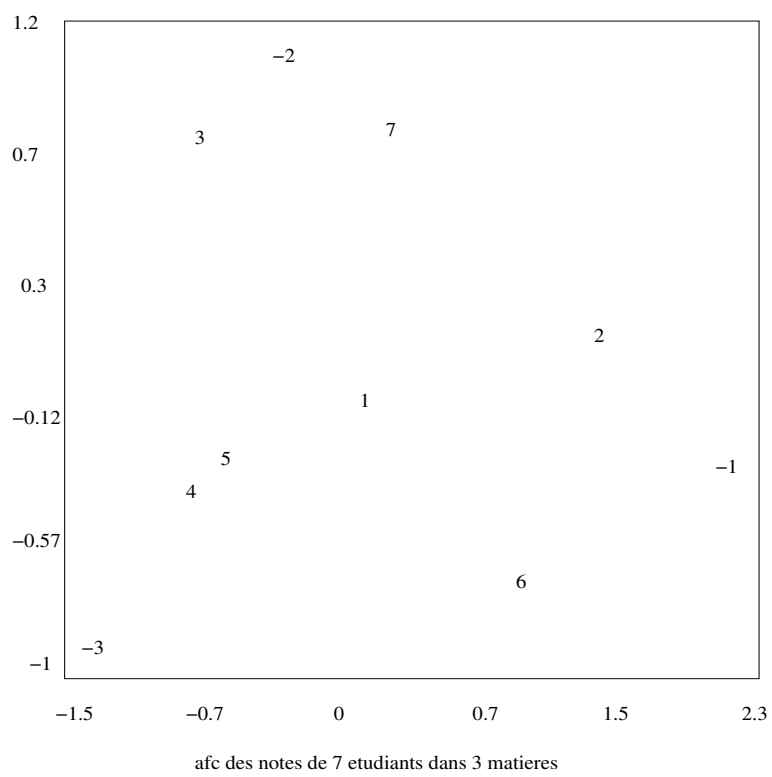


FIG. 9.12 –

Bibliographie

- [1] J.P. Cabannes. site. <http://www.chez.com/deaecodemo/>.
- [2] J. de Lagarde. *site : Initiation l'analyse des données*. Dunod, 1995.
- [3] Pouget Testu Diday, Lemaire. *Elements d'analyse de données*. Dunod, 1982.
- [4] M. de Lara J-Ph Chancelier. Introduction scilab, manipulations vectorielles. cermics.enpc.fr/scilab, 2005.
- [5] M. Jambu. *Méthodes de base de l'analyse de données*. Eyrolles, 1999.
- [6] L. Lebart. site. <http://www.enst.fr/egsh/lebart/>.
- [7] B. Perrin. Wims paris-sud. <http://wims.auto.u-psud.fr/>.
- [8] Inria.
- [9] Ressources scientifiques pour les enseignants de mathématiques. <http://www.dma.ens.fr/culturemath/>.
- [10] Université en ligne. Site internet, <http://www.uel-pcsn.education.fr>.
- [11] [//www.octave.org](http://www.octave.org).
- [12] G. Xiao. Wims. <http://wims.unice.fr>.

Index

vecteur normal, 8

Chasles, 6

equation parametrique, 8

Vecteur directeur, 8