

République Algérienne Démocratique et Populaire
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOULOUDE MAMMERI DE TIZI-OUZOU



FACULTE DE GENIE ELECTRIQUE ET D'INFORMATIQUE
DEPARTEMENT D'ELECTRONIQUE

**Mémoire de Fin
d'Etudes
En vue de l'obtention
du diplôme
de MASTER ACADEMIQUE en Electronique
Spécialité : Télécommunications et Réseaux
Filière : Génie Electrique**

Thème

Réduction des paramètres de données avec
l'ACP en traitement d'images

Proposé et dirigé par :

Mr. ATTAF Youcef

Présenté par :

M^{elle} AMRAR Nassima

Mr BELHOUARI Amine

Année universitaire : 2013/2014

Remerciements

Nous tenons à remercier notre promoteur

*Mr Attaf- Youcef qui a su nous encadrer et nous guider tout au
long de notre travail*

*Nos remerciements s'adressent aussi aux membres du jury pour avoir
examiné ce mémoire et pour avoir accepté de nous juger.*

*Nous adressons un remerciement tout particulier à tous nos amis, ainsi
qu'à tous les professeurs qui nous ont encadrés durant notre cursus
universitaire et à nos deux familles respectives.*

Dédicace

*Je dédie ce mémoire à mes très chers parents qui m'ont aimés
et soutenus pendant des moments difficiles dans ma vie
que dieu le tout puissant veille sur eux et les protèges.*

A mes frères et sœurs, mes neveux et à toute la famille.

Sans oublié bien sur, mes meilleurs amis pour leur amitié.

A tout les gents qui me connaissent.

A mon binôme Amine.

A tous ceux que j'aime.

A vous tous.

Je vous dédie le fruit de mon travail.

Nassima

Dédicace

Je dédie ce modeste mémoire à :

Mes très chers parents : cherif et zahia.

Mon frère : Islam.

Mes sœurs : Siham et Lamia.

Mes grands parents.

Tous mes amis.

Mon binôme Nassima.

Amine

Sommaire

Introduction générale... .. 1**Chapitre I : Notions générales sur le traitement d'images**

I.1 Introduction	2
I.2 Définition de l'image	2
I.2.1 Image binaire	2
I.2.1.1 Traitement de l'image binaire.....	3
I.2.2 Image en niveau de gris.....	3
I.2.3 Image en couleur.....	4
I.3 Image numérique	5
I.3 Caractéristique d'une image numérique.....	6
I.3.1.1 Pixel	6
I.3.1.2 Pixel écran-Pixel numérique.....	7
I.3.2 Qualité de l'image numérique	7
I.3.2.1 Images bitmaps et images vectorielles.....	8
I.3 Transformation de l'image numérique en image binaire.....	8
I.4 Dimension	8
1.5 Résolution	8
I.6 Bruit	8
I.7 Contour et textures	9
I.8 Luminance	9
I.9 Contraste	9

I.9.1	Augmentation du contraste...	10
I.10	Histogramme	11
I.10.1	Egalisation d'histogramme	11
I.11	Les systèmes de traitement d'images	12
I.11.1	Acquisition des données images.....	12
I.11.2	prétraitement et post-traitement.....	13
I.11.3	Traitement numérique des images	13
I.11.3.1	Le seuillage.....	13
I.11.3.1.1	Le filtrages numérique	14
	1. Le filtre passe haut (accentuation)	14
	2. Le filtre passe-bas (lissage).....	15
	3. Filtre passe-bande (différentiation).....	15
	4. Filtre directionnel.....	15
I.11.3.1.2	La convolution	16
I.11.3.1.3	La segmentation	16
I.11.3.1.4	Simplification morphologique.....	16
I.12	Visualisation	18
I.13	Domaines d'application	18
I.14	Discussion	18

Chapitre II : Les différentes techniques pour la réduction des paramètres de données.

II.1 Introduction	20
II.2 Analyse en Composantes Principales (ACP)	20
II.2.1 Introduction à L'ACP.....	20
II.2.2 Principe de l'ACP.....	22
II.2.2.1 Les objectifs	22
II.2.2.2 La transformation des données	24
II.2.2.3 L'analyse des nuages.....	29
II.2.2.4 L'ajustement	33
II.2.3 Représentation simultanée.....	39
II.1.4 Interprétation	41
II.1.5 Conclusion	43

Chapitre III : Application et résultats de la méthode sous Matlab

III.1 Introduction a Matlab	46
III.2 Gestion des images sous Matlab	47
III.2.1 Les fonctions Matlab.....	47
III.2.2 Les toolboxes de Matlab	48
III.2.3 Les images tests	48
III.2.3.1 Images fournies avec l'image processing toolbox	48
III.2.3.2 Images au format JPG	48

III.2.3.3 Images au format TIF... ..48

III.2.3.4 Images au format PNG... ..49

III.3 Présentations des images... .. 49

III.3.1 Tableau de pixel49

III.4 Application de L'ACP dans le traitement d'image... .. 49

III.4.1 Les utilisations les plus courantes de l'ACP... ..49

III.4.2 Les termes employés dans l'ACP... ..50

III.4.3 Principe de la transformation de Karhunen-Loeve (TKL)... ..50

III.4.4 Propriétés de l'algorithme51

III.4.5 Algorithme... ..51

III.5 Résultats expérimentaux et interprétations... .. 52

III.6 Conclusion... .. 57

Annexes

Bibliographie

Liste des figures

Figure I.1 : Image binaire : $I(x,y) \in \{0,1\}$	2
Figure I.2 : Image en niveau de gris: $I(x,y) \in \{0, 255\}$	3
Figure I.3 : Images couleurs : $IB(x,y), IR(x,y), IV(x,y)$	5
Figure I.4 : la lettre A affichée sous forme d'un groupe de pixels.....	6
Figure I.5 : le voisinage d'un pixel.....	7
Figure I.6 : Modification de l'échelle des niveaux de gris. Ici 0 = noir et 1 = blanc....	10
Figure I.7 : Etalement d'histogramme.....	10
Figure I.8 : Egalisation d'histogramme	11
Figure I.9 : Schéma d'un système de traitement d'images	12
Figure I.10 : Traitement d'images et vision.....	17
Figure II.1 : Les nuages de points	30
Figure II.2 : Les formes de nuages de points	30
Figure II.3 : Nuage des individus N_I dans \mathbb{R}^K	31
Figure II.4 : Différents types de nuages.....	31
Figure II.5 : Nuage des variables N_K dans \mathbb{R}^I	33
Figure II.6 : Ajustement du nuage N_I des individus pour l'AC.....	34
Figure II.7 : Représentation des individus	36
Figure II.8 : Ajustement du nuage N_K des variables pour l'ACP.....	37
Figure II.9 : Représentation des variables.....	38
Figure II.10 : Forme de dualité exprimant le nuage N_I en fonction du nuage. N_K	40
Figure II.11 : Forme de dualité exprimant le nuage N_K en fonction du nuage N_I	40
Figure II.12 : Représentation simultanée.....	41
Figure II.13 : Résumé de l'ACP.....	45

Figure II.14 : Principe de l'ACI... ..60

Figure III.1 : Les fonctions de Matlab... ..46

Introduction générale

L'évolution que connaissent l'acquisition et le stockage d'images de très haute qualité et donc à très grandes dimensions, a rendu leur exploitation complexe; ce qui redirige les recherches vers des moyens d'en extraire les informations utiles.

Plusieurs manières de réduction des paramètres existent dans la littérature. Elles visent à extraire et à utiliser uniquement la structure intrinsèque de ces données qui contient juste l'information utile pour les traitements postérieurs. Il s'agira, dans le cas présent, d'une classification.

Cette dernière nécessite la connaissance préalable du nombre de classes. Nous vous proposons d'utiliser les méthodes issues de l'analyse discriminatoire pour réduire les dimensions des données et pour optimiser ce paramètre de classification.

Dans ce cadre on propose d'étudier les différentes méthodes susceptibles d'être utilisées pour réduire et exploiter des images avant de pouvoir les traiter. On s'intéressera principalement à l'analyse en composantes principales (ACP), et en passera en revue d'une manière générale l'analyse en composantes indépendantes (ACI), l'analyse linéaire discriminante (LDA), la diversité marginale maximale (MMD).

Pour ce fait on a organisé ce manuscrit en trois chapitres :

- Dans le premier chapitre on va aborder des généralités sur le traitement d'images.
- Le deuxième chapitre traitera des différentes Techniques pour la réduction des paramètres de données.
- Le dernier chapitre, est dédié aux tests et résultats.
Un Programme ACP sous Matlab a été développé afin de réaliser une réduction de données sur plusieurs types d'images.
- Nous achèverons ce travail par une conclusion générale, et une bibliographie.

Chapitre I

Notions générales sur le traitement d'images

I.1 Introduction :

Le traitement d'image est une méthode qui est née de l'idée et de la nécessité de remplacer l'observateur humain par la machine. Il désigne l'ensemble des méthodes et techniques dont l'objectif est d'améliorer la qualité de l'image. En corrigeant les dégradations subies lors de son acquisition et extraction des informations permettant une interprétation visuelle ou automatique, faisant appel à des techniques destinées à mieux exploiter les informations qu'elle contient. Dans ce chapitre, nous aborderons les notions fondamentales et indispensables à toute opération de traitement d'images.

I.2 Définition de l'image :

Sur l'échelle d'observation visuelle, une image est peut être considérée comme étant un signal bidimensionnel d'objets variant dans l'espace suivant les directions horizontale X et vertical Y et tridimensionnelles de natures diverses.

L'image peut être présentée sous forme d'un ensemble de points répartis dans une surface donnée.

La valeur de chaque point est la réalisation d'une variable aléatoire liée à l'apparition de la luminance dans un site donné de l'image.

Il existe trois formes physiques d'existences d'une image : forme binaire, en niveau de gris et couleur.

I.2.1 Image binaire :

Une image binaire est une matrice rectangulaire dont le nombre de niveaux de gris est réduit aux deux éléments 0 et 1, ou le niveau 0 représente le noir absolu et le niveau 1 représente le blanc. On n'utilise dans ce cas qu'un seul bit pour coder le niveau de gris.

(Voir **Figure I.1**)

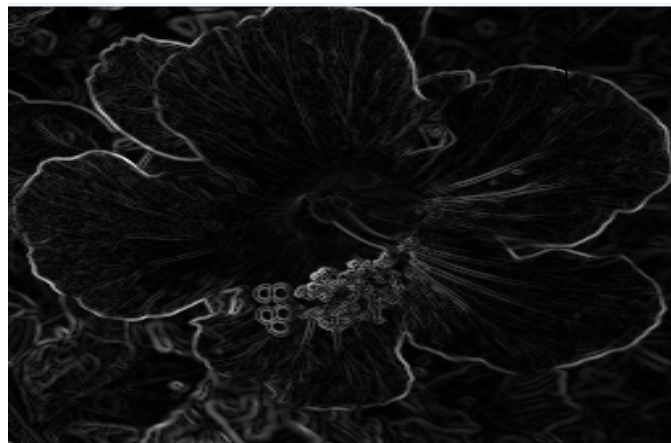


Figure I.1 : Image binaire : $I(x,y) \in \{0,1\}$.

I.2.1.1 Traitement de l'image binaire :

Les structures de cette image binaire peuvent être étudiées sous deux aspects: la morphométrie ou la stéréologie.

La morphométrie : est une analyse quantitative de l'image d'un objet ou d'une structure, elle définit des grandeurs mesurables.

La stéréologie : est un ensemble de méthodes mathématiques qui permet de passer d'une analyse bidimensionnelle sur coupe à une analyse tridimensionnelle.

I.2.2 Image en niveau de gris :

Le niveau de gris et la valeur de l'intensité lumineuse en un point. La couleur du pixel peut prendre des valeurs allant du noir au blanc en passant par un nombre fini de niveaux intermédiaires. Donc pour représenter les images à niveaux de gris, on peut attribuer à chaque pixel de l'image une valeur correspondante à la quantité de lumière renvoyée. Cette valeur peut être comprise entre 0 et 255. Chaque pixel n'est donc plus représenté par un bit, mais par un octet. Pour cela, il faut que le matériel utilisé pour afficher l'image soit capable de produire les différents niveaux de gris correspondant [Har 95].

Le nombre de niveaux de gris dépend du nombre de bit utilisés pour décrire la "couleur" de chaque pixel de l'image. Plus ce nombre est important, plus les niveaux possibles sont nombreux. (**Figure I.2**)



Figure I.2 : Image en niveau de gris: $I(x,y) \in \{0, 255\}$.

Codage en bit des niveaux de gris :

- Un bit : Un bit est la plus petite donnée qu'un ordinateur peut utiliser. De la même manière qu'une lampe, il ne peut être qu'allumé ou éteint. Si nous décidons d'utiliser un bit pour décrire notre image, nous pouvons utiliser cet état allumé ou éteint pour représenter le noir ou le blanc sans avoir d'état intermédiaire possible (pas de gris).
- Deux bits : Si nous utilisons désormais deux bits pour décrire notre image, nous avons désormais quatre états possibles : noir, gris foncé, gris clair, blanc.
- n bits : A chaque fois que nous augmenterons de un le nombre de bit, nous doublerons le nombre total de combinaisons que nous utiliserons pour définir le nombre total de niveau de gris. Une échelle de 0 à 63 sera ainsi codée sur 6 bits [Rec 96].

I.2.3 Image en couleur :

La commission internationale de l'éclairage (CIE) a fixée conventionnellement un système de référence dit RVB (rouge, vert, bleu), qui consiste en la reproduction de presque toutes les couleurs qui existent par la combinaison de ces trois composantes primaires. La représentation des couleurs s'effectue de la même manière que les images monochromes avec cependant quelques particularités. En effet, il faut tout d'abord choisir un modèle de représentation [Har 95].

Dans Une image couleur chaque plan est codé comme une image en niveau de gris avec des valeurs allant de 0 à 255 lorsque R=V=B, la valeur associée est un niveau de gris. D'autre part pour aller d'une image couleur à une image en niveau de gris on utilise la fonction suivante :

$$I(I, j) = R(I, j) + V(I, j) + B(I, j) / 3 \quad (1)$$

Avec:

$I(I, j)$: niveau de gris du pixel situé à la ligne I et la colonne j.

$R(i, j)$: intensité de la couleur rouge du pixel (i, j).

$V(i, j)$: intensité de la couleur verte du pixel (i, j).

$B(i, j)$: intensité de la couleur bleu du pixel (i, j).

Chacun de ces trois plans peut être codé sur 8, 24, 32, et 64 bits.

Pour la représentation en couleur réelle, nous pouvons donc utiliser 24 bit pour chaque point de l'image. 8 bits pour le rouge(R), 8 bits pour vert (V) et 8 bits pour le bleu (B). Il est ainsi possible de représenter environ 16.7 million de couleurs différentes. Cela est cependant théorique, car aucun écran n'est capable d'afficher les 16 million de couleurs. Par ailleurs, l'œil humain n'est pas capable de distinguer autant de couleurs.

Selon la représentation couleur choisie, on distingue trois types d'images comme indiquées sur la figure suivante : **(Figure I.3)**



Figure I.3: Images couleurs :

$IB(x,y)$, $IR(x,y)$, $IV(x,y)$.

I.3 Image numérique :

Contrairement aux images obtenues à l'aide d'un appareil photo, ou dessinées sur du papier, les images manipulées par un ordinateur sont numériques (représentées par une série de bits). L'image numérique est l'image dont la surface est divisée en éléments de tailles fixes appelés cellules ou pixels, ayant chacun comme caractéristique un niveau de gris ou de

couleurs prélevé à l'emplacement correspondant dans l'image réelle, ou calculé à partir d'une description interne de la scène à représenter [GW 77]. La numérisation d'une image est la conversion de celle-ci de son état analogique (distribution continue d'intensités lumineuses dans un plans xOy) en une image numérique représentée par une matrice bidimensionnelle de valeurs numériques $f(x,y)$ où :

- x, y : coordonnées cartésiennes d'un point de l'image.
- $f(x, y)$: niveau de gris en ce point.

Pour des raisons de commodité de représentation pour l'affichage et l'adressage, les données images sont généralement rangées sous formes de tableau I de n lignes et p colonnes. Chaque élément $I(x, y)$ représente un pixel de l'image et à sa valeur est associé un niveau de gris codé sur m bits (2^m niveaux de gris ; $0 = \text{noir}$; $2^m-1 = \text{blanc}$). La valeur en chaque point exprime la mesure d'intensité lumineuse perçue par le capteur.

I.3.1 Caractéristique d'une image numérique :

L'image est un ensemble structuré d'informations caractérisé par les paramètres suivants :

I.3.1.1 Pixel :

Contraction de l'expression anglaise " Picture elements " : Eléments d'image, le pixel est le plus petit point de l'image. C'est une entité calculable qui peut recevoir une structure et une quantification. Si le bit est la plus petite unité d'information que peut traiter un ordinateur, le pixel est le plus petit élément que peuvent manipuler les matériels et logiciels d'affichage ou d'impression [Tou87].

La lettre A, par exemple, peut être affichée comme un groupe de pixels dans la (**Figure I.4**) suivante :

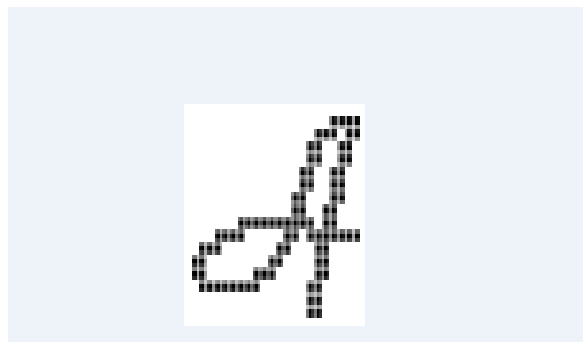


Figure I.4 : la lettre A affichée sous forme d'un groupe de pixels.

I.3.1.2 Pixel écran-Pixel numérique :

Il est important de bien dissocier le pixel d'un écran de télévision, qui a une taille physique, et le pixel numérique. Le pixel de l'écran de télévision est déterminé une bonne fois pour toutes lors de sa fabrication alors que le pixel digital à une taille est simplement une définition mathématique à l'intérieur de l'ordinateur que l'on caractérise en choisissant une taille [CC 85].

La figure suivante illustre la notion de pixel et son voisinage (**Figure I.5**).

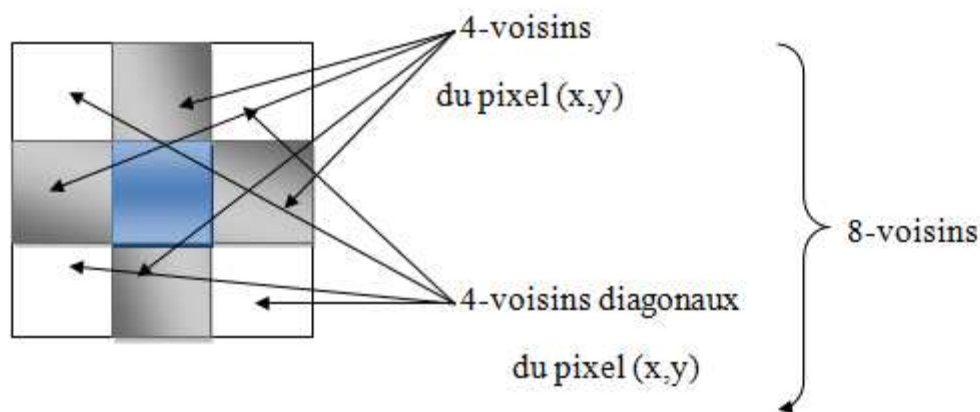


Figure I.5 : le voisinage d'un pixel.

I.3.2 Qualité de l'image numérique :

Elle dépend, d'une part, de la qualité des images d'origine et, d'autre part, des moyens mis en œuvre pour convertir un signal analogique en signal numérique. Elle dépend aussi de :

- la qualité des périphériques de numérisation de l'image, du nombre de niveaux de gris ou de couleurs enregistrées, etc...
- la qualité de l'affichage à l'écran : définition de l'écran, nombre de teintes disponibles simultanément, calibrage de l'écran, etc...

Les critères d'appréciation de la qualité d'une image, tels que cités succinctement ci-dessus, dépendent largement de la structure même de l'image réaliste ou conceptuelle et de son mode de représentation (bitmap ou vectorielle) [Tou 87].

I.3.2.1 Images bitmaps et images vectorielles :

Les images appartiennent à deux grandes familles : bitmap (image-bit) et vectorielle. Dans l'image vectorielle, les données sont représentées par des formes géométriques simples qui sont décrites d'un point de vue mathématique. L'image bitmap est formée d'un tableau de points ou pixels, plus la densité des points est élevée plus le nombre de l'information est grand et donc plus la résolution de l'image est élevée.

Si les images vectorielles peuvent être manipulées avec beaucoup de facilité, les modifications de taille, par exemple, apportées à une image bitmap ne sont pas sans incidence [Har 95].

I.3.3 Transformation de l'image numérique en image binaire :

En fonction des formes recherchées (grains de tailles, zones de couleurs particulières), on va séparer les structures intéressantes des autres par l'opération de seuillage. Elle consiste à donner la valeur 1 aux structures qui vérifient les critères sélectionnés et 0 aux autres. On obtient donc une matrice binaire [Mig 03].

I.4 Dimension :

C'est la taille de l'image. Celle-ci se présente sous forme d'une matrice de n ligne et p colonnes. Le nombre de ligne de cette matrice multiplié par le nombre de colonnes nous donne la dimension ou la taille ($n \times p$) qui est le nombre de pixel de l'image.

1.5 Résolution :

C'est la clarté ou la finesse de détails atteinte par un moniteur ou une imprimante dans la production d'images. Sur les moniteurs d'ordinateurs, la résolution est exprimée en nombre de pixels par unité de mesure (pouce ou centimètre). On utilise aussi le mot résolution pour désigner le nombre total de pixels affichables horizontalement ou verticalement sur un moniteur; plus grand est ce nombre, meilleure est la résolution [Tou 87].

I.6 Bruit :

Un bruit (parasite) dans une image est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins il peut provenir de l'éclairage des dispositifs optiques et électronique du capteur [GW 77].

Si le niveau du signal est suffisant la proportion du bruit dans le signal utile (le rapport signal/bruit) reste insignifiante. Par contre si le niveau du bruit prend le pied sur l'information principale, le bruit sera présent.

I.7 Contour et textures :

Les contours représentent la frontière entre les objets de l'image, ou la limite entre deux pixels dont les niveaux de gris représentent une différence significative [Gra 84]. Les textures décrivent la structure de ceux-ci. L'extraction du contour consiste à identifier dans l'image, les points qui séparent deux textures différentes [Kun 93].

I.8 Luminance :

C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface, pour un observateur lointain, le mot luminance est substitué au mot brillance, qui correspond à l'éclat d'un objet [GW 77]. Une bonne luminance se caractérise par :

- Des images lumineuses (brillantes).
- Un bon contraste : il faut éviter les images où la gamme de contraste tend vers le blanc ou le noir; ces images entraînent des pertes de détails dans les zones sombres ou lumineuses.
- L'absence de parasites.

I.9 Contraste :

C'est l'opposition marquée entre deux régions d'une image, plus précisément entre les régions sombres et les régions claires de cette image. Le contraste est défini en fonction des luminances de deux zones d'images [GW 77]. Si L_1 et L_2 sont les degrés de luminosité respectivement de deux zones voisines A_1 et A_2 d'une image, le contraste c 'est défini par le rapport suivant :

$$C = (L_1 - L_2) / (L_1 + L_2) \quad (2)$$

I.9.1 Augmentation du contraste :

On peut agir directement sur la table des niveaux de gris

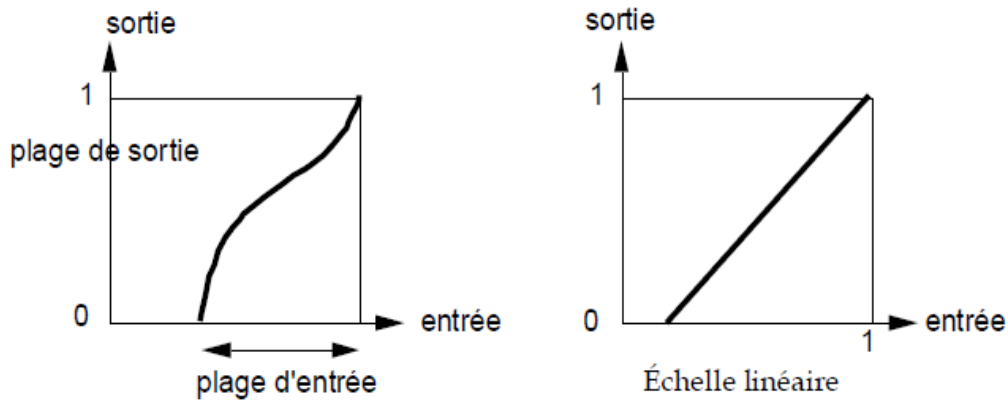


Figure I.6 : Modification de l'échelle des niveaux de gris. Ici 0 = noir et 1 = blanc.

On cherche à obtenir une courbe monotone qui occupe toute la plage de dynamique de l'image : l'intervalle $[0, 1]$. Si la plage d'entrée est plus petite que l'intervalle $[0,1]$ on augmente le contraste de l'image ; sinon, c'est l'inverse.

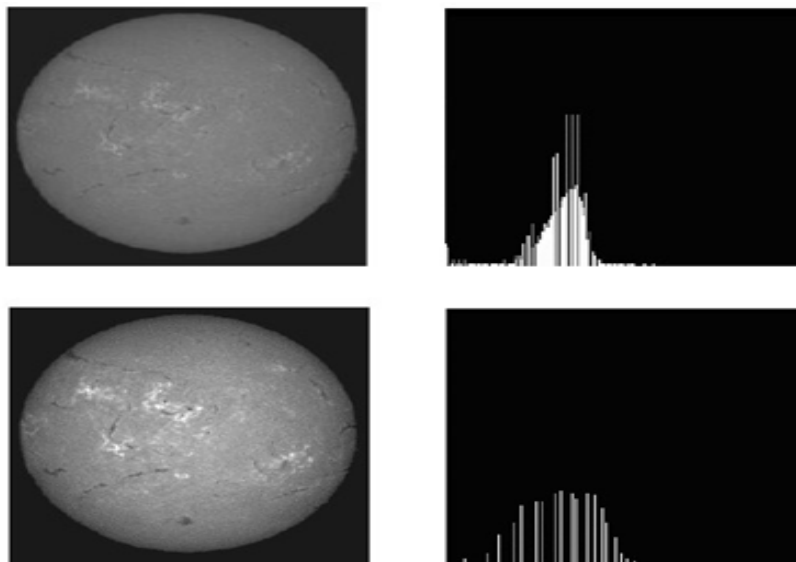


Figure I.7 : Etalement d'histogramme.

La (Figure I.7) est composée de 4 images. L'image supérieure gauche est une image du soleil de l'observatoire du Kanzelhohe (Autriche) très peu contrastée, les détails sont peu visibles. L'histogramme de l'image est montré à sa droite. On s'aperçoit qu'une petite plage des niveaux de gris est occupée (valeurs non nulles dans l'histogramme).

Dans l'image située en dessous, on a utilisé toute la plage de la dynamique en employant une courbe de transformation du type de celle visualisée au-dessus. L'histogramme équivalent est représenté sur la droite. L'image est plus contrastée, les niveaux de gris très sombres et très clairs sont peu occupés, ce qui donne l'impression que l'histogramme est encore nul pour ces valeurs.

I.10 Histogramme :

L'histogramme des niveaux de gris ou des couleurs d'une image est une fonction qui donne la fréquence d'apparition de chaque niveau de gris (couleur) dans l'image. Pour diminuer l'erreur de quantification, pour comparer deux images obtenues sous des éclairages différents, ou encore pour mesurer certaines propriétés sur une image, on modifie souvent l'histogramme correspondant [Kun 93] [GW 77].

Il permet de donner un grand nombre d'information sur la distribution des niveaux de gris (couleur) et de voir entre quelles bornes est répartie la majorité des niveaux de gris (couleur) dans le cas d'une image trop claire ou trop foncée.

Il peut être utilisé pour améliorer la qualité d'une image (Rehaussement d'image) en introduisant quelques modifications, pour pouvoir extraire les informations utiles de celle-ci.

I.10.1 Egalisation d'histogramme :

L'idée qui a servi de base à cette technique est de renforcer le contraste des parties intéressantes de l'image.

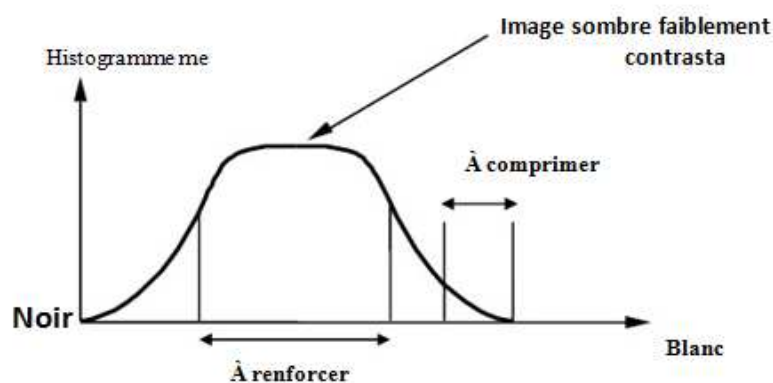


Figure I.8 : Egalisation d'histogramme.

Si on observe l'histogramme de la (**Figure I.8**) (cas classique), on s'aperçoit que la plupart des pixels (partie ventrue, haute de l'histogramme) occupent une plage faible de l'histogramme. En revanche, les deux queues (avant et arrière) de l'histogramme correspondent à des pixels peu nombreux, mais qui occupent une partie importante de l'histogramme. Il faudrait donc étaler la bosse principale de l'histogramme en largeur et, au contraire, comprimer les deux queues. On conçoit, à la limite, que les problèmes vont disparaître, si on dispose d'un histogramme plat, ne comportant pas de bosse. Cela correspond à une utilisation équilibrée de la plage de dynamique de l'image. Il s'avère que cette situation correspond également à une amélioration de la qualité visuelle de l'image.

I.11 Les systèmes de traitement d'images :

Un système de traitement numérique d'image est composé de plusieurs étapes, comme indiqué sur la (**Figure I.9**) suivante :

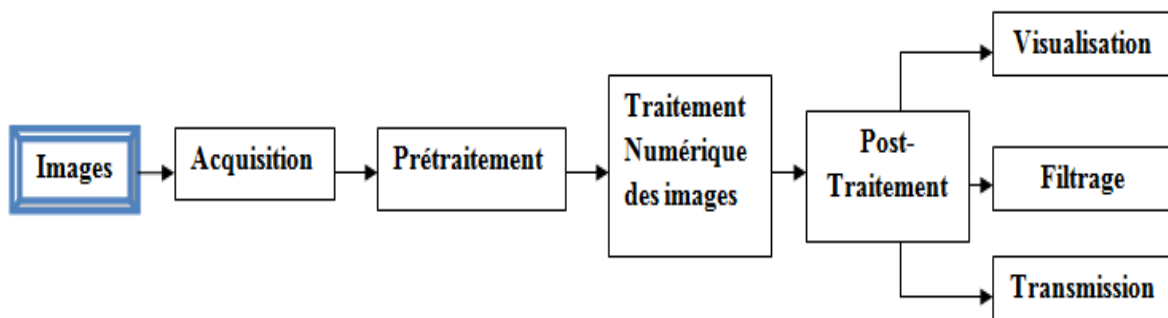


Figure I.9 : Schéma d'un système de traitement d'images.

I.11.1 Acquisition des données images :

L'acquisition consiste d'abord en la transformation d'ondes optiques représentant la scène qu'on souhaite prendre en image, en signaux électriques capable d'être traité par un système électronique, puis en numérisation de ces signaux pour être traité par des systèmes informatiques. La numérisation se décompose en deux opérations, l'échantillonnage et la quantification.

- L'échantillonnage consiste à transformer le signal continu en une suite d'échantillons ou points élémentaires pixels.
- La quantification consiste à mesurer les amplitudes des échantillons, et les coder sur un certain nombre de bits, selon un choix limité de valeurs appelées

niveaux de gris. La valeur de cette mesure est comprise entre 0 et 255 (codée sur 8 bits) et se trouve dans l'image sous forme d'un point ou pixel de couleur variant du blanc au noir.

I.11.2 Prétraitement et post-traitement :

Le prétraitement est la première étape du traitement de l'information visuelle, son but est d'améliorer la perception de certains détails et de réduire le bruit de certains défauts notamment par les capteurs. Il permet aussi de faciliter l'analyse d'une image en renforçant la ressemblance entre pixels appartenant à des régions différentes.

Le prétraitement est nécessaire notamment dans les cas suivants :

- L'image est bruitée.
- Le contraste n'est pas suffisant.
- L'éclairage de l'image n'est pas uniforme.

Les méthodes les plus utilisées sont :

- Modification de l'histogramme.
- La réduction du bruit.
- Le rehaussement de contraste.
- Le filtrage.

Le post-traitement concerne les images prétraitées. Toutes les opérations utilisables en prétraitement, le sont aussi en post-traitement.

I.11.3 Traitement numérique des images :

On peut citer les traitements numériques suivants [And 87] :

I.11.3.1 Le seuillage :

Une image numérique, pour pouvoir être exploitée a généralement besoin d'être simplifiée, c'est le but du seuillage. Ce dernier consiste à transformer l'image codée sur 6, 8 ou 16 bits, en une image binaire où les pixels à 1 correspondent aux objets et les pixels à 0 au fond de l'image. A la différence des différents traitements de l'image numérique, le seuillage est un passage obligatoire pour toute analyse morphologique ultérieure. Le seuillage permet de sélectionner les parties de l'image qui intéressent l'opérateur, par exemple 2 types de grains (blancs et sombres) dans un mélange. On peut donc, par exemple, attribuer à tous les pixels de

l'image numérique qui ont un niveau de gris compris entre deux valeurs i_1 et i_2 , choisies par l'opérateur, la valeur 1; à tous les autres pixels est attribuée la valeur 0.

Après seuillage, les parties de l'image sélectionnées seront traduites en noir et blanc. L'image digitalisée par l'ordinateur (0 et 1), est appelée image binaire. Cette dernière, tout comme l'image numérique contient des informations superfétatoires, qu'il convient d'éviter, ou masquées qu'il faut révéler. L'image obtenue après seuillage est donc appelée **image binaire** (les pixels ont tous un état 0 ou 1). Les traitements de l'image binaire ont plusieurs buts :

- Filtrage (passe haut ou passe bas...).
- Convolution.
- Segmentation.
- Simplification morphologique.

I.11.3.1.1 Le filtrage numérique :

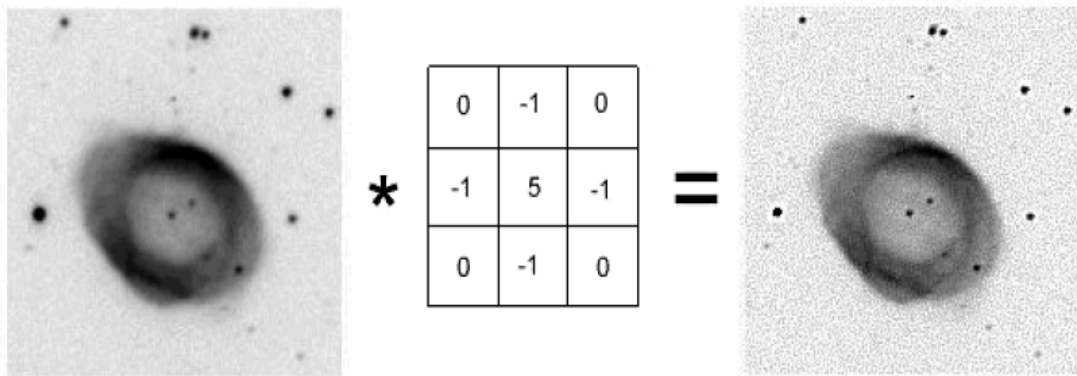
Pour améliorer la qualité de l'image visuelle de l'image, on doit éliminer les effets des bruits (parasites) en lui faisant subir un prétraitement appelé filtrage. Le filtrage consiste à modifier la distribution fréquentielle des composantes d'un signal selon des spécifications données [And 87].

- Le système linéaire utilisé est appelé filtre numérique.
- Parmi ces systèmes, nous distinguons : les filtres passe-bas (lissage), filtres passe-haut (accentuation), filtres passe-bande (différenciation) et filtres directionnels.

1. Le filtre passe haut (accentuation) :

Le renforcement des contours et leur extraction s'obtiennent dans le domaine fréquentiel par l'application d'un filtre passe-haut.

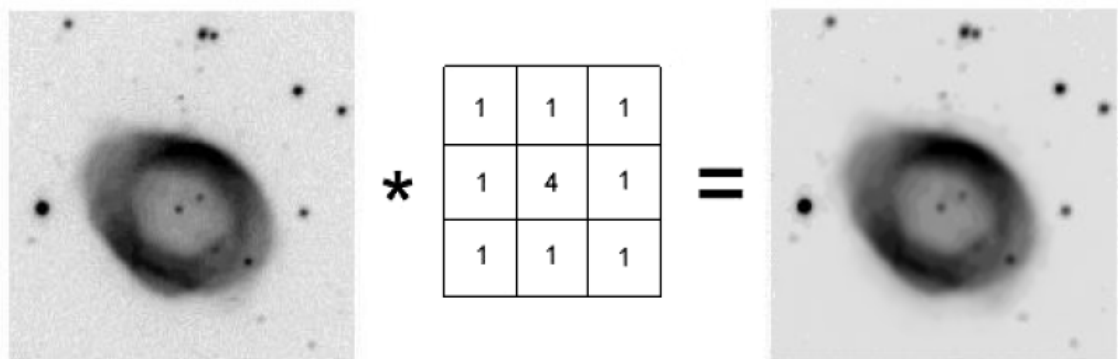
L'application principale des produits de convolution est la création des filtres « passe haut » et « passe bas ». Un filtre « passe haut » favorise les hautes fréquences spatiales, comme les détails, et de ce fait, il améliore le contraste. Un filtre « passe haut » est caractérisé par un noyau comportant des valeurs négatives autour du pixel central, comme dans l'exemple ci-dessous:



Dans cette image de ciel profond, le filtre passe-haut a été utilisé pour améliorer les détails de la nébuleuse planétaire M 57, mais un examen attentif montre des effets secondaires au traitement:

2. le filtre passe-bas (lissage) :

Les filtres « passe bas » agissent en sens inverse des filtres « passe haut » et le résultat est, un adoucissement des détails (impression de léger flou), ainsi qu'une réduction du bruit granuleux. Il est appelé filtrage d'adoucissement ou lissage car il adoucit considérablement les contours.



3. filtre passe-bande (différentiation) :

Cette opération est une dérivée du filtre passe-bas. Elle consiste à éliminer la redondance d'information entre l'image originale et l'image obtenue par filtre passe-bas. Seule la différence entre l'image source et l'image traitée est conservée.

4. filtre directionnel :

Dans certain cas, on cherche à faire apparaître des détails de l'image dans une direction bien déterminée. Pour cela, on utilise des filtres qui opèrent suivants des directions (horizontales, verticales et diagonales).

I.11.3.1.2 La convolution :

La convolution est le remplacement de la valeur d'un pixel par une combinaison linéaire de ses voisins. Elle consiste à faire balayer une fenêtre (masque) sur l'ensemble des points de l'image [HM 95].

I.11.3.1.3 La segmentation :

On est conduit pour analyser une image à la partager en un certain nombre de domaines connexes (unis) : ce processus joue un rôle de plus en plus important en traitement numérique des images [Tou 87].

Deux points appartiennent à un même domaine s'ils possèdent une même propriété caractéristique P et si on peut les joindre par une chaîne continue de points voisins ayant cette propriété P. une segmentation est donc définie par le type de propriété et par la définition de voisinage.

La segmentation d'une image binaire a la même finalité que les seuillage; d'ailleurs, certains auteurs classent le seuillage parmi les méthodes de segmentation. Elle consiste à séparer l'image en plusieurs zones. La segmentation la plus simple consiste à séparer les particules convexes apparues comme connexes à la suite des opérations précédentes. On peut utiliser la morphologie mathématique. La segmentation fait appel à la notion de connexité.

I.11.3.1.4 Simplification morphologique :

Pour étudier la morphologie des objets à analyser, on est souvent conduit à simplifier encore plus l'image. Cette simplification peut être faite par des ouvertures ou fermetures morphologiques qui lissent le contour des objets et permettent ainsi de mettre en évidence, par exemple, les protubérances ou les cavités. La simplification extrême est le squelette.

On vient d'exposer les méthodes de traitement d'images les plus utilisées, qui peuvent être complétées par d'autres. Le schéma suivant (**Figure I.10**) les met en évidence tout en classant les méthodes de traitement d'images par rapport à celle de la vision.

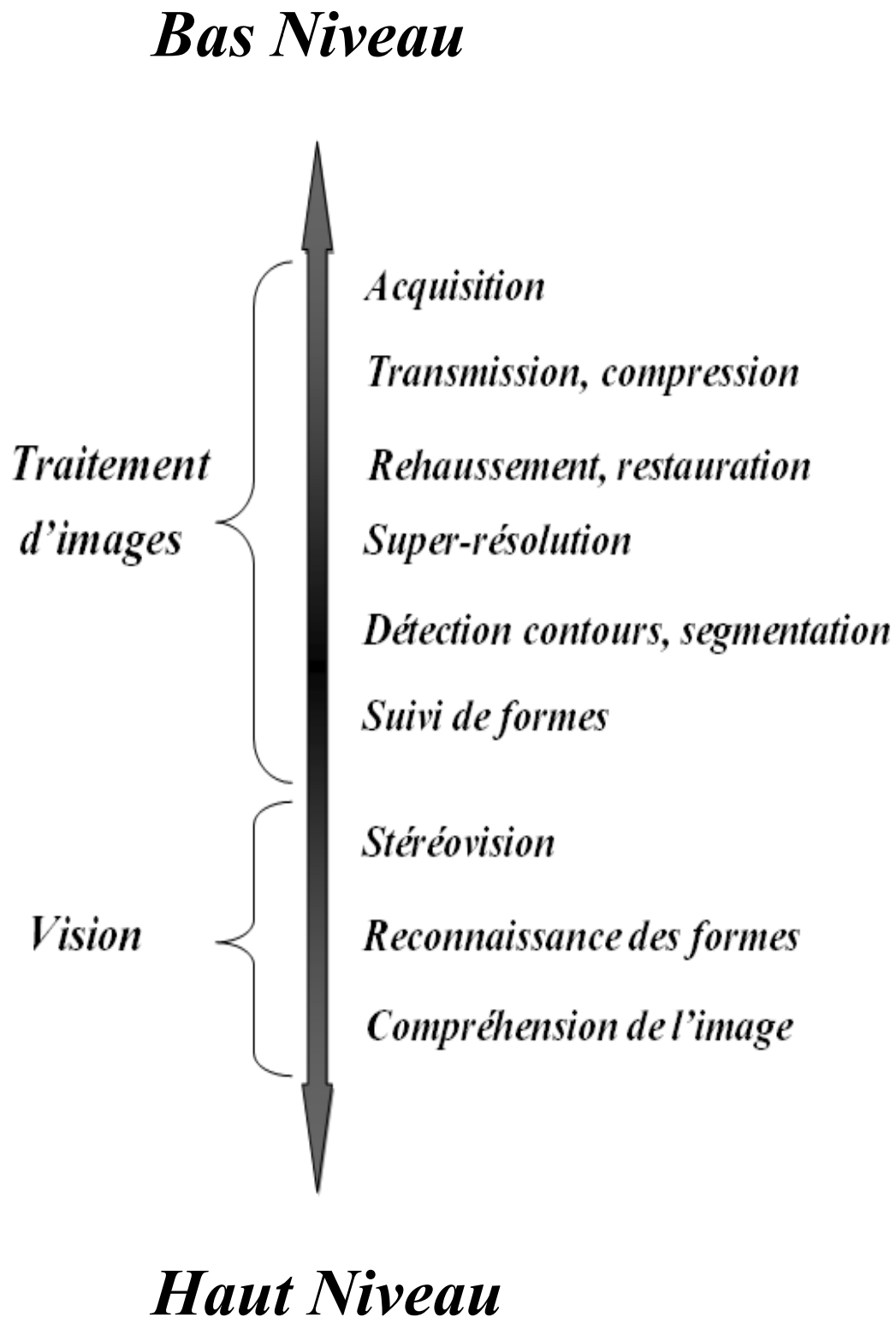


Figure I.10 : Traitement d'images et vision.

I.12 Visualisation :

Tout système de traitement d'image est doté d'un dispositif de visualisation qui permet l'affichage des images.

L'utilisation de différents types de restitutions permet de transformer le signal numérique qu'est la matrice image en un signal analogique visible par l'œil de l'observateur. Pour cela, différents types de support peuvent être employés : moniteur vidéo, clichés photographiques, impression sur papier. Dans tous les cas et pour chaque échantillon de l'image numérique, on recrée un nouvel élément d'image ou un nouveau pixel dont on choisit la forme de façon à reconstituer une image analogique qui soit la plus proche possible de l'image avant numérisation compte tenu des erreurs introduites lors de l'acquisition, de la numérisation et de la transmission [Har 95].

I.13 Domaines d'application :

Le traitement d'image possède l'aspect multidisciplinaire. On trouve ses applications dans des domaines très variés tels que les télécommunications (T.V, vidéo, publicité, etc...) [Sym 00], la médecine (radiographie, ultrasons, etc...) [FAB 97, RBB+ 99, biologie, astronomie, géologie, l'industrie (robotique, sécurité) [Dra 93]. La météorologie, l'architecture [KSR 01], l'imprimerie, l'armement (application militaire) [Urb 65].

De nouvelles applications pratiques sont possibles aujourd'hui et touchent tous les domaines d'activités, tels que : Métiers du spectacle, créations artistiques, etc...

I.14 Conclusion :

La représentation des images fixes est l'un des éléments essentiels des applications multimédias, comme dans la plupart des systèmes de communication. La manipulation des images pose cependant des problèmes beaucoup plus complexes que celle du texte. En effet, l'image est un objet à deux dimensions, censé représenter un espace à trois dimensions, ce qui à deux conséquences majeures :

- Le volume des données à traiter est beaucoup plus important.
- La structure de ces données est nettement plus complexe.

Il en résulte que la manipulation, le stockage et la représentation de ces données se heurtent à certaines limitations.

Grâce au traitement d'image, ces contraintes sont levées ou contournées. En effet, ce domaine cherche à détecter la présence de certaines formes, certains contours ou certaines textures d'un modèle connu, c'est le problème de la détection.

Un autre aspect de traitement d'image concerne l'analyse et la compréhension de l'image dans le but d'en extraire des informations utiles. Il cherche aussi parfois à comprimer l'image, afin de gagner en vitesse lors de la transmission de l'information, et en capacité de stockage, tout en dégradant le moins possible les images considérées. C'est le domaine du codage ou compression des images.

Chapitre II

Réduction des paramètres de données avec L'ACP

II.1 Introduction :

Les méthodes d'analyse de données ont commencées à être développées dans les années 50 poussées par le développement de l'informatique et du stockage des données qui depuis n'a cessé de croître. L'analyse de données a surtout été développée en France par J.P. Benzécri qui a su par l'analyse des correspondances représenter les données de manière simple et interprétable.

L'analyse de données fait toujours l'objet de recherche pour s'adapter à tout type de données et faire face à des considérations de traitements en temps réel en dépit de la quantité de données toujours plus importante. Les méthodes développées (et l'analyse de données) sont maintenant souvent intégrées avec des méthodes issues de l'informatique et de l'intelligence artificielle (apprentissage numérique et symbolique) dans le data mining traduit en français par fouille de données ou encore extraction de connaissance à partir de données, Parmi ces méthodes de nombreuses analyses sont issues de l'analyse factorielle, telles que l'analyse en composantes principales L'ACP, L'ACI, LDA et la MMD, l'analyse en composantes principales est l'une des méthodes les plus employées. Elle est particulièrement adaptée aux variables quantitatives, continues, a priori corrélées entre elles, et qui fait l'objet de notre travail, les autres méthodes de réduction sont données en (annexes A).

II.2 Analyse en Composantes Principales (ACP) :

II.2.1 Introduction à L'ACP :

Le principe de base de l'ACP est de réduire les dimensions de la matrice des données, en conservant toute fois les variations présentes dans le jeu de données de départ. En effet, cette technique permet de privilégier les directions d'un espace de données porteuses du maximum d'informations au sens de la maximisation des variances de projection. L'amplitude des valeurs propres de la matrice de covariance des données quantifie pour chacune de ces directions la quantité d'information encodée.

Les variables initiales sont ainsi transformées en de nouvelles variables, appelées composantes principales. Elles sont obtenues par combinaison linéaire des composantes originales, ordonnées et non corrélées. L'ACP cherche à identifier les vecteurs propres et les valeurs propres de la matrice de corrélation des variables de départ.

Les données pour l'ACP sont généralement présentées sous forme d'un tableau (**Tableau1**).

- soit pour réduire les dimensions des données sans perte importante d'information, par exemple en traitement du signal et des images, où l'ACP intervient souvent en prétraitement pour réduire la quantité de données issues de traitements analogiques.

II.2.2 Principe de l'ACP :

II.2.2.1 Les objectifs :

Dans un premier temps reprenons les objectifs de l'ACP et détaillons-les. Nous avons vu que pour atteindre les objectifs de l'ACP il faut évaluer les ressemblances entre individus ainsi que les liaisons entre variables. Ces deux notions peuvent être interprétées de différentes façons, il est donc important de bien les définir.

Définition : Deux individus se ressemblent, ou sont proches, s'ils possèdent des valeurs proches pour l'ensemble des variables.

Cette définition sous-entend une notion de proximité qui se traduit par une distance. Ainsi, nous définissons la distance entre deux individus i et j par :

$$d^2(i, j) = \sum_{k \in K} (x_{ik} - x_{jk})^2 \quad (3)$$

La métrique ici utilisée est donc euclidienne, mais de manière plus générale nous pouvons définir cette distance par :

$$d^2(i, j) = (X_i - X_j)^T M (X_i - X_j) \quad (4)$$

Où M est une matrice symétrique définie positive de taille K .

Pour établir un bilan des ressemblances entre individus, nous cherchons à répondre à des questions du type :

- Quels sont les individus qui se ressemblent ?
- Quelles sont ceux qui sont différents ?
- Existe-t-il des groupes homogènes d'individus ?
- Est-il possible de mettre en évidence une typologie des individus ?

De la même façon que nous avons défini la ressemblance entre individus, il est essentiel de définir la liaison entre des variables.

Définition : Deux variables sont liées si elles ont un fort coefficient de corrélation linéaire.

Le coefficient de corrélation linéaire est donné par :

$$r(\mathbf{k}, \mathbf{h}) = \frac{\text{cov}(\mathbf{k}, \mathbf{h})}{\sqrt{\text{var}(\mathbf{k})\text{var}(\mathbf{h})}} = \frac{1}{I} \sum_{i \in I} \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right) \left(\frac{x_{ih} - \bar{x}_h}{s_h} \right) (5)$$

Où \bar{x}_k et s_k sont respectivement la moyenne et l'écart-type de la variable k .

la moyenne :

$$\bar{x}_k = \frac{1}{I} \sum_{i=1}^I x_{ik} (6)$$

l'écart type :

$$s_k = \sqrt{\frac{1}{I} \sum_{i=1}^I (x_{ik} - \bar{x}_k)^2} (7)$$

Remarque :

Dans le cadre de l'ACP normée que nous présentons ici, le coefficient de Corrélation est défini à partir de la covariance, cependant dans de rare cas l'ACP peut être fondée sur la matrice de covariance (ACP non-normée) ou encore sur la matrice des coefficients de corrélations des rangs. A partir du coefficient de corrélation de l'équation (3), il est possible de définir une distance entre deux variables k et h :

$$d(k, h) = \frac{1}{I} \sum_{i \in I} \left(\frac{x_{ik} - \bar{x}_k}{s_k} - \frac{x_{ih} - \bar{x}_h}{s_h} \right)^2 = 2(1 - r(k, h)) (8)$$

De même que pour les individus, nous cherchons à établir un bilan des liaisons entre variables en répondant à des questions du type :

- Quelles sont les variables qui sont liées positivement entre elles ?
- Quelles sont celles qui s'opposent (i.e. liées négativement) ?
- Existe-t-il des groupes de variables corrélées entre elles ?
- Est-il possible de mettre en évidence une typologie des variables ?

Et le coefficient de corrélation devient :

$$r(k, h) = \sum_{i \in I} p_i \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right) \left(\frac{x_{ih} - \bar{x}_h}{s_h} \right) (9)$$

Nous retrouvons le cas particulier dans lequel les individus ont le même poids lorsque $p_i = \frac{1}{I}$

De même, il est possible de ne pas accorder la même importance aux différentes variables. Soit m_k le poids associé à la variable k , la distance de l'équation entre deux individus i et j devient :

$$d^2(i, j) = \sum_{k \in K} m_k (x_{ik} - x_{jk})^2 \quad (10)$$

En fait, ces poids ne modifient en rien les principes de l'ACP, nous considérons donc par la suite les cas où les individus et variables ont le même poids.

II.2.2.2 La transformation des données :

Il existe plusieurs transformations utilisées. L'analyse centrée consiste à modifier les données du tableau X en remplaçant les valeurs des x_{ik} par $x_{ik} - \bar{x}_k$. Le fait de centrer les données présente dans le cas de l'ACP des propriétés intéressantes que nous exposons à la section

II.2.2.3 L'analyse centrée réduite ou encore normée, que nous présentons ici, est liée à la transformation des données du tableau X en remplaçant les valeurs des x_{ik} par $\frac{x_{ik} - \bar{x}_k}{s_k}$. Réduire les données permet d'uniformiser les unités de mesures.

Ainsi le tableau de données X présenté sur le tableau 1 devient celui donné par le deuxième tableau. (**Tableau 2**)

		VARIABLES		
		1 k K
INDIVIDUS	1	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> \vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots </div> <div style="text-align: center;"> $x_{ik} - \bar{x}_k$ <hr style="width: 50%; margin: 0 auto;"/> s_k </div> <div style="text-align: center;"> \vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots </div> </div>		
			
	i			
			
			
	I			

Tableau2 :Représentation des données centrée-réduites pour l'ACP.

❖ Exemple illustratif pour l'A.C.P :

Considérons les notes (de 0 à 20) obtenues par 9 élèves dans 4 disciplines (mathématiques, physique, française, anglais) :

	Maths.	Physique	Français	Anglais
Anne	6	6	5	5,5
Bertrand	8	8	8	8
Carine	6	7	11	9,5
Didier	14,5	14,5	15,5	15
Elsa	14	14	12	12,5
Fabienne	11	10	5,5	7
Gustave	5,5	7	14	11,5
Hélène	13	12,5	8,5	9,5
Isabelle	9	9,5	12,5	12

Questions: Quelles structures de corrélation entre les notes dans les diverses matières? Comment caractériser les élèves au travers de leurs notes ?

On note :

$$x = \begin{pmatrix} 6 & 6 & 5 & 5.5 \\ 8 & 8 & 8 & 8 \\ 6 & 7 & 11 & 9.5 \\ 14.5 & 14.5 & 15.5 & 15 \\ 14 & 14 & 12 & 12.5 \\ 11 & 10 & 5.5 & 7 \\ 5.5 & 7 & 14 & 11.5 \\ 13 & 12.5 & 8.5 & 9.5 \\ 9 & 9.5 & 12.5 & 12 \end{pmatrix}$$

Letableau de donnéesde n lignes (**individus**) et p colonnes (**variables**).

$$x = \begin{pmatrix} 6 & 6 & 5 & 5.5 \\ 8 & 8 & 8 & 8 \\ 6 & 7 & 11 & 9.5 \\ \mathbf{14.5} & \mathbf{14.5} & \mathbf{15.5} & \mathbf{15} \\ 14 & 14 & 12 & 12.5 \\ 11 & 10 & 5.5 & 7 \\ 5.5 & 7 & 14 & 11.5 \\ 13 & 12.5 & 8.5 & 9.5 \\ 9 & 9.5 & 12.5 & 12 \end{pmatrix} X_i = \begin{pmatrix} 14.5 \\ 14.5 \\ 15.5 \\ 15 \end{pmatrix}$$

L'individu de la ligne i pour i variant de 1 à n.

$$x = \begin{pmatrix} 6 & 6 & 5 & \mathbf{5.5} \\ 8 & 8 & 8 & \mathbf{8} \\ 6 & 7 & 11 & \mathbf{9.5} \\ 14.5 & 14.5 & 15.5 & \mathbf{15} \\ 14 & 14 & 12 & \mathbf{12.5} \\ 11 & 10 & 5.5 & \mathbf{7} \\ 5.5 & 7 & 14 & \mathbf{11.5} \\ 13 & 12.5 & 8.5 & \mathbf{9.5} \\ 9 & 9.5 & 12.5 & \mathbf{12} \end{pmatrix} X^j = \begin{pmatrix} 5.5 \\ 8 \\ 9.5 \\ 15 \\ 12.5 \\ 7 \\ 11.5 \\ 9.5 \\ 12 \end{pmatrix}$$

La variable de la colonne j pour j variant de 1 à p.

○ **Statistiques élémentaires : moyennes**

Pour tout j = 1, , n, on note

$$\bar{X}^j = \frac{1}{n} \sum_{i=1}^n X_i^j \quad (11)$$

La moyenne empirique de la variable X^j .

J	1	2	3	4
\bar{X}^j	9.67	9.83	10.22	10.06

Il existe des différences de niveaux de notation entre les matières, de la plus exigeante (Mathématiques) à la moins (Français).

○ **Données centrées :**

On appelle **données centrées** le tableau de données

$$\tilde{X} = (X_i^j - \bar{X}^j)_{i=1,\dots,n, j=1,\dots,p} \quad (12)$$

Dans ce tableau, la somme des valeurs d'une même colonne est nulle.

$$\tilde{X} = \begin{pmatrix} -3.67 & -3.83 & -5.22 & -4.56 \\ -1.67 & 1.83 & -2.22 & 2.06 \\ -3.67 & 2.83 & 0.78 & 0.56 \\ 4.83 & 4.67 & 5.28 & 4.94 \\ 4.33 & 4.17 & 1.78 & 2.44 \\ 1.33 & 0.17 & -4.72 & -3.06 \\ -4.17 & -2.83 & 3.78 & 1.44 \\ 3.33 & 2.67 & -1.72 & -0.56 \\ 0.67 & 0.33 & 2.28 & 1.94 \end{pmatrix}$$

○ **Statistiques : Écarts types :**

Pour tout $j = 1, \dots, n$, on note

$$\sigma^j = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i^j - \bar{X}^j)^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\tilde{X}_i^j)^2} \quad (13)$$

L'écart type empirique de la variable X^j .

J	1	2	3	4
σ^j	3.37	2.99	3.47	2.81

Il existe des différences de variation des notations entre les différentes matières, de la matière dont les notes sont le plus dispersées (Français) à la matière dont les notes le sont le moins (Anglais).

○ **Données centrées réduites :**

On appelle **données centrées réduites** le tableau de données

$$\tilde{X}^1 = \left(\frac{X_i^j - \bar{X}^j}{\sigma^j} \right)_{i=1,\dots,n, j=1,\dots,p} = \left(\frac{\tilde{X}_i^j}{\sigma^j} \right)_{i=1,\dots,n, j=1,\dots,p} \quad (14)$$

Dans ce tableau, la somme des valeurs d'une même colonne est nulle, et la somme des carrés des valeurs d'une même colonne vaut 1.

$$\tilde{X}^1 = \begin{pmatrix} -1.087 & -1.282 & -1.504 & -1.619 \\ -0.494 & -0.613 & -0.640 & -0.731 \\ -1.087 & -0.947 & 0.224 & -0.197 \\ 1.432 & 1.560 & 1.520 & 1.758 \\ 1.284 & 1.393 & 0.512 & 0.869 \\ 0.395 & 0.056 & -1.360 & -1.086 \\ -1.235 & -0.947 & 1.088 & 0.513 \\ 0.988 & 0.892 & -0.496 & -0.197 \\ -0.198 & -0.111 & 0.656 & 0.691 \end{pmatrix}$$

○ **Matrice de variances/covariances :**

La matrice des variances/covariances, matrice de même nature que celle des corrélations, bien que moins parlante. La diagonale de cette matrice fournit les variances des 4 variables considérées (on notera qu'au niveau des calculs, il est plus commode de manipuler la variance que l'écart-type ; pour cette raison, dans de nombreuses méthodes statistiques, comme l'ACP, on utilise la variance pour prendre en compte la dispersion d'une variable quantitative).

Matrice de variances/covariances de X la matrice p x p :

$$cov(X) = \left(\frac{1}{n} \sum_{i=1}^n (X_i^j - \bar{X}^j) (X_i^{j'} - \bar{X}^{j'}) \right) j, j' = 1, \dots, p = \frac{1}{n} \tilde{X}^T \tilde{X} \quad (15)$$

Remarque: La diagonale de cette matrice contient les variances de chaque variable étudiée.

$$cov(X) = \begin{pmatrix} 11.39 & 9.92 & 2.66 & 4.82 \\ 9.92 & 8.94 & 4.12 & 5.48 \\ 2.66 & 4.12 & 12.06 & 9.29 \\ 4.82 & 5.48 & 9.29 & 7.91 \end{pmatrix}$$

○ **Matrice de corrélations :**

La matrice des corrélations. Elle donne les coefficients de corrélation linéaire des variables prises deux a deux. C'est une succession d'analyses bi variées, constituant un premier pas vers l'analyse multi variée.

Matrice de corrélations de X la matrice p x p :

$$cor(x) = \left(\frac{1}{n} \sum_{i=1}^n \frac{(X_i^j - \bar{X}^j)(X_i^{j'} - \bar{X}^{j'})}{\sigma^j \sigma^{j'}} \right) j, j' = 1, \dots, p = \frac{1}{n} (\tilde{X}^1)^T = cov(\tilde{X}^1) \quad (16)$$

Remarque: La diagonale de cette matrice contient des 1.

$$\text{cor}(X) = \begin{pmatrix} 1 & \mathbf{0.983} & 0.227 & 0.508 \\ \mathbf{0.983} & 1 & 0.397 & 0.652 \\ 0.227 & 0.397 & 1 & \mathbf{0.951} \\ 0.508 & 0.652 & \mathbf{0.951} & 1 \end{pmatrix}$$

Remarquons que toutes les corrélations linéaires sont positives (ce qui signifie que toutes les variables varient, en moyenne, dans le même sens), certaines étant très fortes (0.983 et 0.951), d'autres moyennes (0.652 et 0.508), d'autres enfin plutôt faibles (0.397 et 0.227).

II.2.2.3 L'analyse des nuages :

En considère deux nuages de points ou deux représentations géométriques associées (**Figure II.1**). Nous obtenons ainsi I points dans l'espace \mathbb{R}^K et K points dans l'espace \mathbb{R}^I . Plusieurs formes de nuages sont remarquables aussi bien pour les projections de l'espace des individus que pour celui des variables (**Figure II.2**). Par exemple, nous pouvons distinguer des formes sphériques ne traduisant aucune direction privilégiée, des formes allongées donnant une direction privilégiée des dépendances, ou encore plusieurs sous-nuages faisant ainsi apparaître plusieurs sous-groupes de la population. D'autres formes sont remarquables telles que les formes triangulaires ou paraboliques [LMP95]. Le problème est de pouvoir rendre compte visuellement de la forme des nuages, pour ce faire l'idée est d'étudier les projections sur des droites ou mieux des plans (les projections dans un espace à 3 dimensions seraient intéressantes si l'œil humain n'était pas souvent trompé). Il faut donc chercher le sous-espace qui ajuste au mieux le nuage de points et cherché à minimiser les déformations que la projection implique.

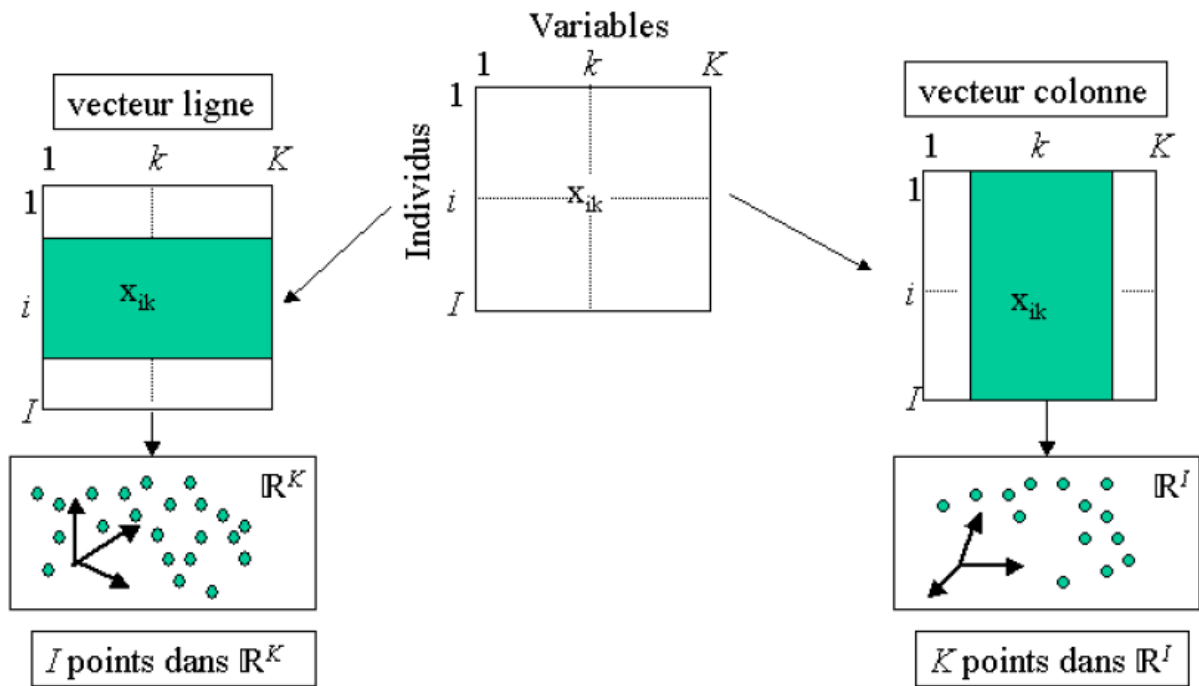


Figure II.1: Les nuages de points.

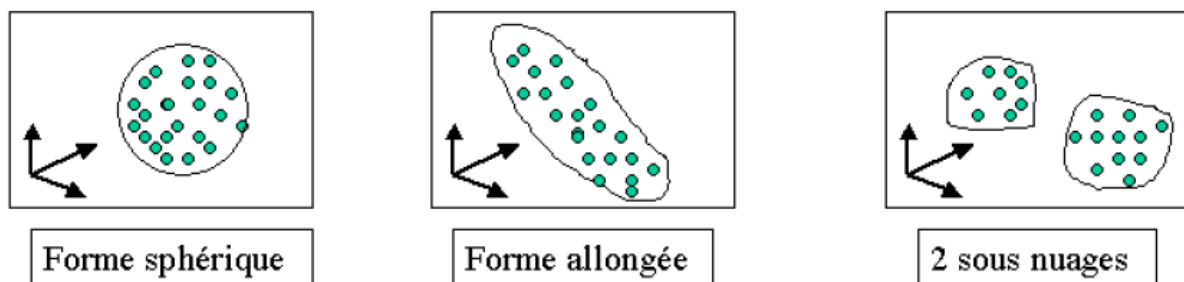


Figure II.2 : Les formes de nuages de points.

- **Analyse du nuage des individus :**

Pour l'analyse du nuage N_I des individus, nous considérons donc le **tableau 2** des données centrées réduites par ligne.

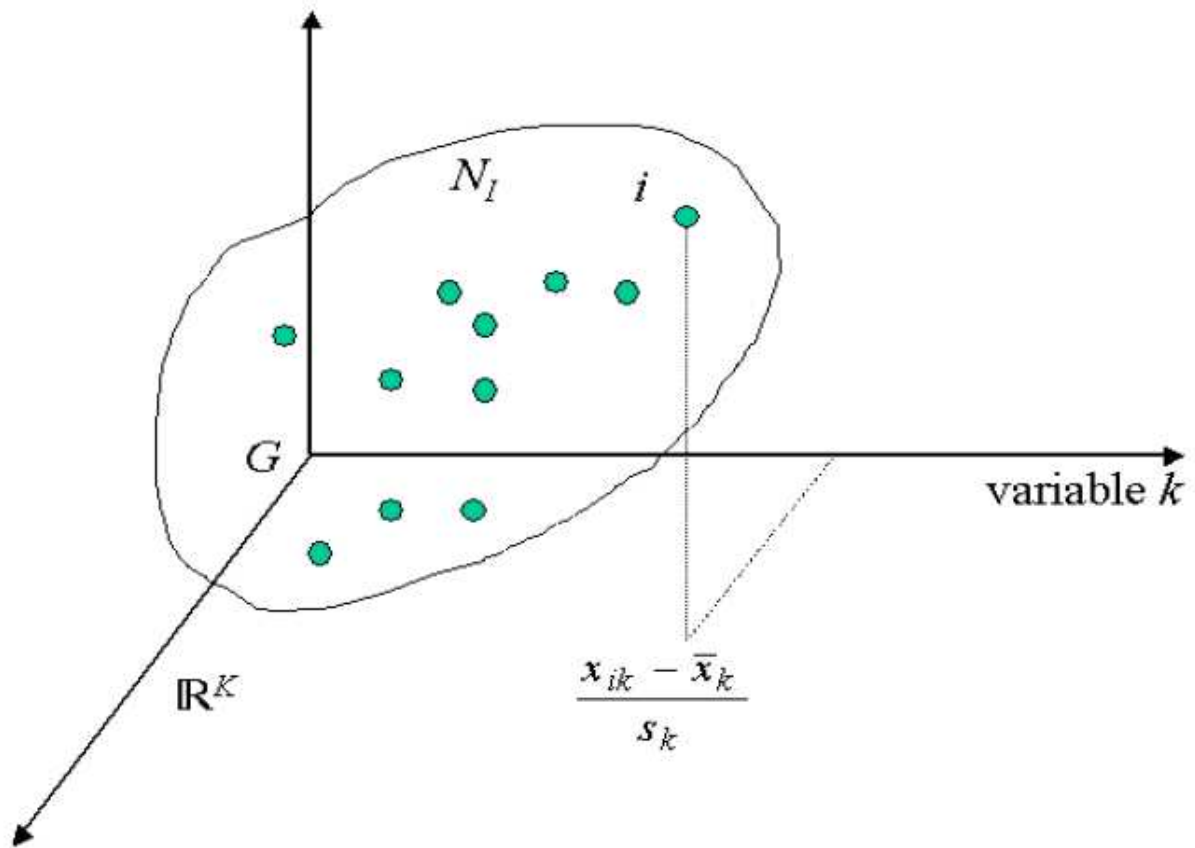


Figure II.3 : Nuage des individus N_I dans \mathbb{R}^K .

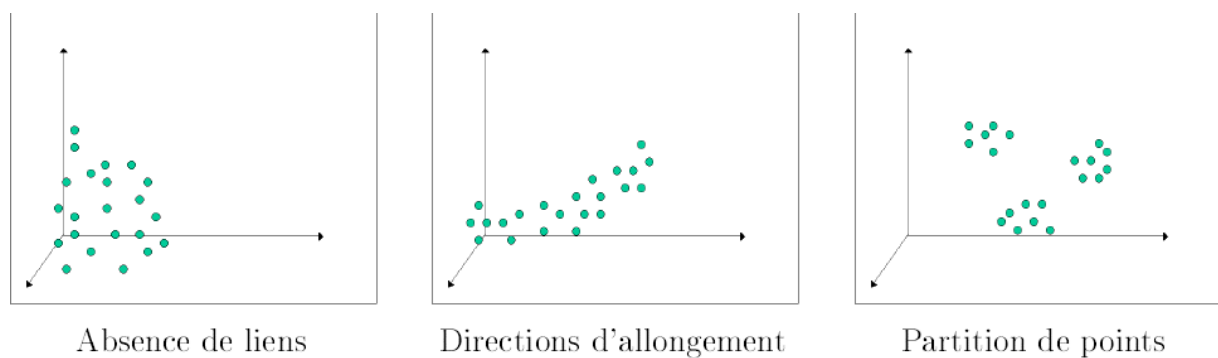


Figure II.4 : Différents types de nuages.

Ainsi le nuage N_I des individus est un espace vectoriel à K dimensions, dont chaque dimension représente une variable (**Figure II.3**). Le fait d'avoir centré les données entraîne que l'origine des axes est confondu avec le centre de gravité G . Ce centre de gravité G peut s'interpréter comme l'individu moyen de la population. L'interprétation de ce nuage N_I va se faire en décelant d'une part une partition de points et d'autre part des directions d'allongement. Ainsi sur la (**Figure II.4**) nous représentons différents types de nuages possibles. Nous pouvons observer une absence de liens, ou bien par exemple une direction d'allongement suivant plutôt le premier axe, ou encore une partition des points en trois groupes. Si l'étude directe est envisageable dans un espace à trois dimensions, dès lors que $K > 3$ elle devient impossible. Nous avons donc recours à l'approche factorielle à partir de laquelle nous pouvons étudier différents plans de projection.

- **Analyse du nuage des variables :**

L'analyse du nuage N_K des variables se fait toujours à partir du **tableau 2** des données centrées réduites, que nous considérons ici par colonne. La représentation du nuage N_K des variables se situe dans un espace vectoriel à I dimensions, chaque dimension représentant un individu de la population totale. La norme de chaque variable k est telle que :

$$\sum_{i \in I} \frac{1}{I} \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right)^2 = 1 \quad (17)$$

Cette norme correspond également au coefficient de corrélation de la variable k avec elle-même, donc $r(k, k) = 1$. Ainsi le nuage N_K est situé sur la sphère unité (de rayon 1) dans l'espace \mathbb{R}^I (**Figure II.5**). Il est intéressant de noter que le cosinus de l'angle entre les vecteurs représentant deux variables k et h est le produit scalaire $\langle k; h \rangle = r(k, h)$. L'interprétation du coefficient de corrélation comme un cosinus est une propriété très importante puisqu'elle donne un support géométrique, donc visuel, au coefficient de corrélation. Cette propriété nécessite d'avoir au préalable centré les données, ce qui justifie une nouvelle fois cette transformation.

L'analyse du nuage N_K des variables se fera donc par l'étude des angles formés pour une variable k par Ok et les axes factoriels. Il est bon de noter que le centre de gravité du nuage N_K n'est pas l'origine de la sphère unité, à la différence du nuage N_I où le centre de gravité correspond au centre du repère lorsque les données sont centrées. Ainsi, ce sont les angles entre les vecteurs représentant les variables qui sont peu déformés par les projections et non pas les distances entre les points du nuage N_K . Cette étude des angles est impossible à réaliser

directement à cause de la dimension de \mathbb{R}^I . Elle se fera donc dans les plans de projection issus de l'approche factorielle.

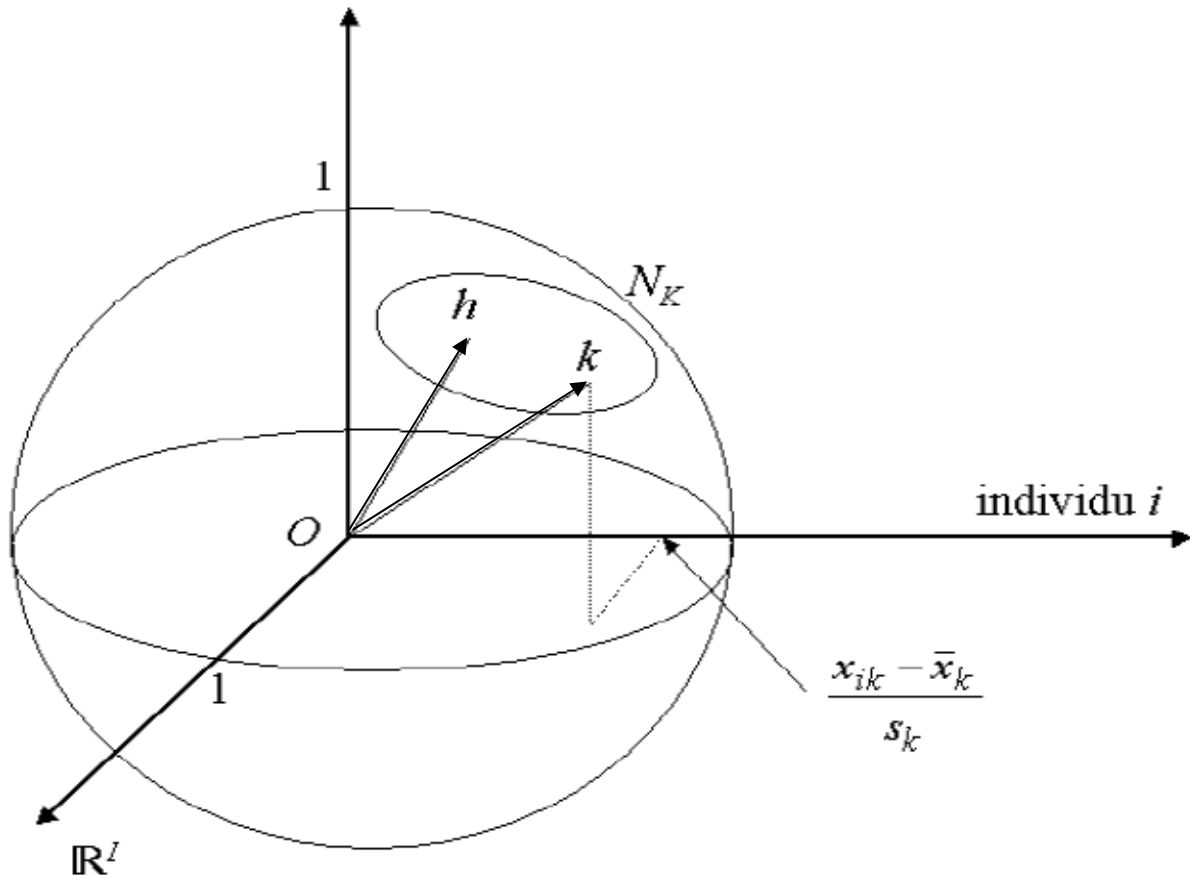


Figure II.5 : Nuage des variables N_K dans \mathbb{R}^I .

II.2.2.4 L'ajustement :

L'approche factorielle consiste donc à approcher ces nuages N_I et N_K dans des sous-espaces vectoriels permettant de fournir quelques images planes de ces nuages.

- **Ajustement du nuage des individus :**

L'objectif est de fournir des images approchées du nuage des individus - que nous noterons N_I dans \mathbb{R}^K . Nous considérons pour la visualisation des images planes de N_I . Nous faisons l'hypothèse que le nuage N_I est contenu dans un sous-espace vectoriel de dimension S inférieure à K . Plus généralement, nous supposons que le nuage N_I est reconstitué de manière satisfaisante dans un sous-espace de dimension S . Nous pouvons ainsi reconstruire les I individus, et donc l'ensemble de la population et du tableau X associé à partir des coordonnées

des individus sur S nouveaux axes. Les $I:K$ valeurs du tableau X sont donc remplacées par $I:S$ (coordonnées) + $K:S$ (composantes).

Donc il faut chercher une suite ; $\{u_s; s = 1, \dots, S\}$ de directions privilégiées les axes factoriels afin de fournir une représentation simplifiée du nuage N_I . Chaque direction u_s rend maximum l'inertie par rapport au centre de gravité G de la projection du nuage N_I sur l'axe factoriel. De plus les directions u_s sont orthogonales deux à deux.

Avec les notations de la (**Figure II.6**) l'individu i se projette en H_i sur u_1 . Nous cherchons donc u_1 qui rend maximum l'inertie $\sum_{i \in I} GH_i^2$. En effet, rendre maximum $\sum_{i \in I} GH_i^2$ revient à rendre minimum l'écart entre le nuage des individus et sa projection ($\sum_{i \in I} iH_i^2$), critère classique des moindres carrés. Ensuite, il faut trouver u_2 orthogonal à u_1 qui satisfait le même critère. Nous pouvons procéder ainsi jusqu'à l'obtention des S axes factoriels donnant une représentation suffisamment bonne.

Définition : Les S axes factoriels ; $\{u_s; s = 1, \dots, S\}$ sont appelées les facteurs principaux.

Du fait d'avoir centré les données, ce critère permet d'interpréter les axes factoriels comme des directions d'allongement maximum du nuage N_K .

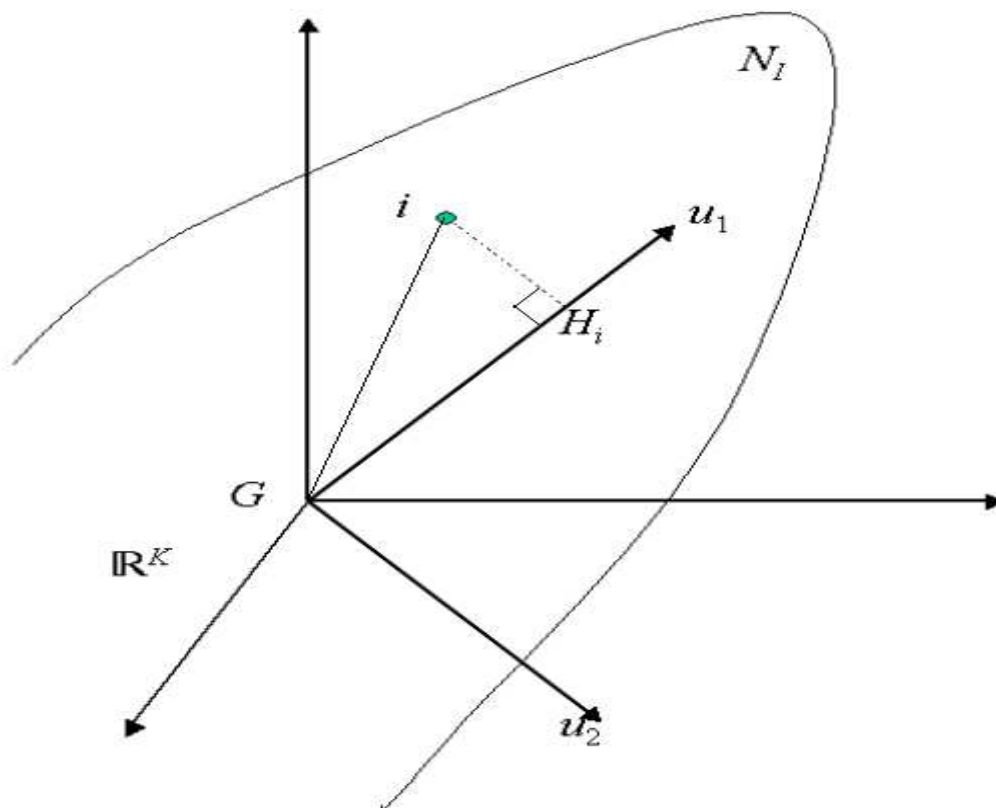
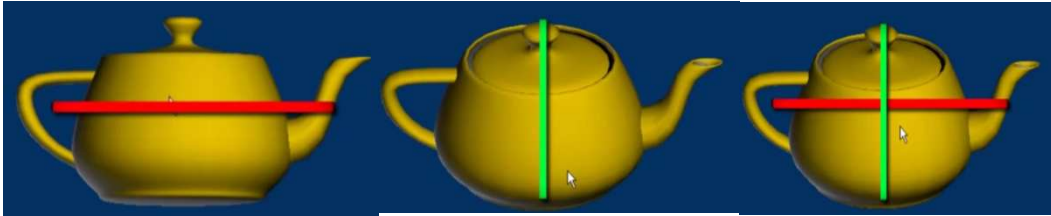


Figure II.6: Ajustement du nuage N_I des individus pour l'ACP.

❖ **Illustration de la construction des axes :**

On a une théière en 3D on va essayer de trouver les axes qui nous permettent de visualiser au mieux cette théière on trouve le premier axe, on cherche un 2em axe orthogonal au 1er qui permet de visualiser au mieux la théière donc en a trouver le premier axe en rouge est le 2em en vert.



○ **Représentation des individus sur notre jeu de donnée et l'interprété :**

	Maths.	Physique	Français	Anglais
Anne	6	6	5	5,5
Bertrand	8	8	8	8
Carine	6	7	11	9,5
Didier	14,5	14,5	15,5	15
Elsa	14	14	12	12,5
Fabienne	11	10	5,5	7
Gustave	5,5	7	14	11,5
Hélène	13	12,5	8,5	9,5
Isabelle	9	9,5	12,5	12

C^1, \dots, C^2 Sont appelées les composantes principales. Les deux premières composantes principales sont :

$$C^{1,2} = \begin{pmatrix} \text{Anne} & -8.61 & -1.41 \\ \text{Bertrand} & -3.88 & -0.5 \\ \text{Carine} & -3.21 & 3.45 \\ \text{Didier} & 9.85 & 0.6 \\ \text{Elsa} & -6.41 & -2.05 \\ \text{Fabienne} & -3.03 & -4.92 \\ \text{Gustave} & -1.03 & 6.38 \\ \text{Helene} & 1.95 & -4.20 \\ \text{Isabelle} & 1.55 & 2.63 \end{pmatrix}$$

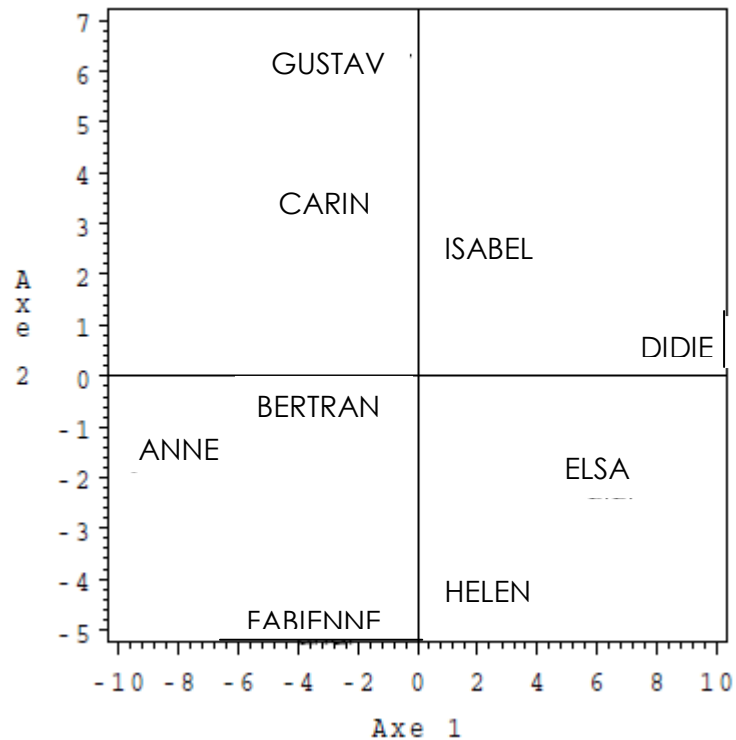


Figure II.7 : Représentation des individus.

On confirme ainsi que l'axe 1 représente le résultat d'ensemble des élèves : si on prend leur score (ou coordonnée) sur l'axe 1, on obtient le même classement que si on prend leur moyenne générale. Par ailleurs, l'élève « le plus haut » sur le graphique, celui qui a la coordonnée la plus élevée sur l'axe 1, est Gustave dont les résultats sont les plus contrastés en faveur des disciplines littéraires (14 et 11.5 contre 7 et 5.5). C'est exactement le contraire pour Fabienne qui obtient la moyenne dans les disciplines scientifiques (11 et 10) mais des résultats très faibles dans les disciplines littéraires (7 et 5.5). On notera que Didier et Bertrand ont un score voisin de 0 sur l'axe 2 car ils ont des résultats très homogènes dans les 4 disciplines (mais a des niveaux très distincts, ce qu'a déjà révélé l'axe 1). L'axe 2 oppose bien les « littéraires » (en haut) aux « scientifiques » (en bas).

- **Ajustement du nuage des variables :**

Nous cherchons ici à obtenir des variables synthétiques ; $\{v_s; s = 1, \dots, S\}$ et une représentation approchée des corrélations entre les variables. La démarche pour le nuage N_K reste la même que pour le nuage N_I des individus. Ainsi, avec les notations de la **Figure**

II.8, la variable k se projette en H_K sur v_1 . Nous cherchons le premier axe factoriel en déterminant le vecteur v_1 qui rend maximum $\sum_{k \in K} OH_k^2$. Puis, nous cherchons le vecteur v_2 orthogonal à v_1 qui satisfait ce même critère.

Nous poursuivons cette démarche jusqu'à l'obtention des S vecteurs recherchés.

Le vecteur v_1 définit une nouvelle variable qui est la combinaison linéaire la plus liée à l'ensemble des variables initiales du tableau X . Ainsi les S vecteurs $\{v_s; s = 1, \dots, S\}$ étant orthogonaux deux à deux, les S nouvelles variables correspondantes sont non corrélées entre elles.

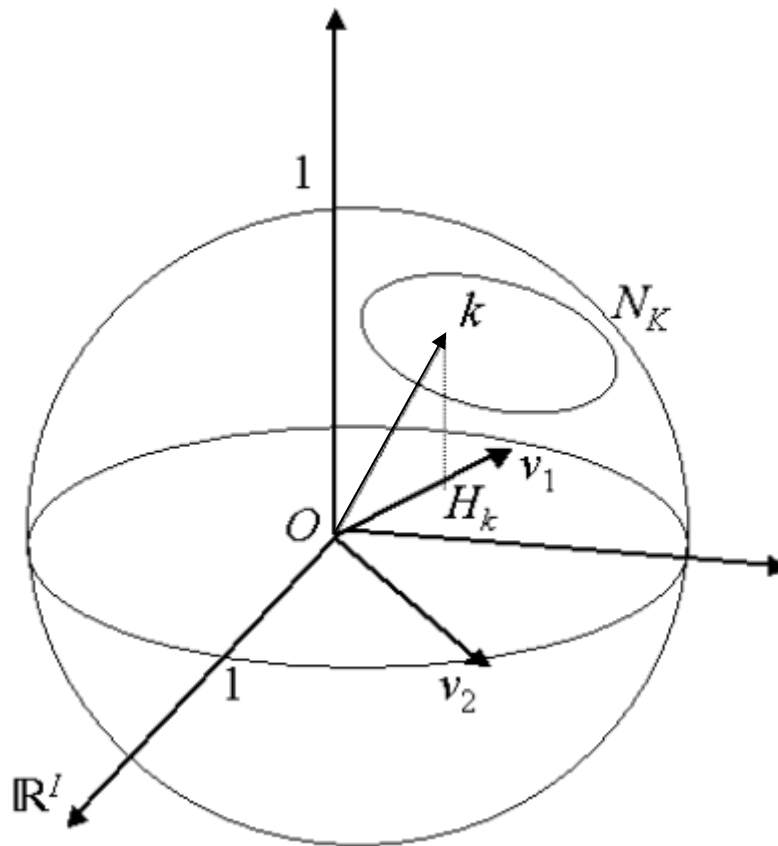


Figure II.8 : Ajustement du nuage N_K des variables pour l'ACP.

Définition : Les S nouvelles variables (axes factoriels) $\{v_s; s = 1, \dots, S\}$ sont appelées les composantes principales.

Ce sont ces vecteurs qui sont à l'origine du nom de cette analyse factorielle.

La coordonnée d'une variable initiale de X sur v_s est son coefficient de corrélation avec v_s du fait que les variables étudiées sont centrées réduites. Ainsi le vecteur v_1 qui rend maximum $\sum_{k \in K} OH_k^2$ équivaut à la combinaison linéaire la plus liée à l'ensemble des variables initiales (la liaison étant entendue au sens du critère maximisant la somme des moindres carrés des corrélations). C'est donc la variable qui synthétise le mieux l'ensemble des variables initiales. Les axes factoriels résument donc l'ensemble des variables initiales du tableau X .

○ **Représentation des variables sur notre jeu de donnée et l'interprété :**

Le résultat fondamental concernant les variables est le tableau des corrélations variablesfacteurs. Il s'agit des coefficients de corrélation linéaire entre les variables initiales et les facteurs. Ce sont ces corrélations qui vont permettre de donner un sens aux facteurs (de les interpréter).

<i>Math</i>	0.81	-0.58	0.01	-0.02
<i>Phys</i>	0.90	-0.43	-0.03	0.02
<i>Fran</i>	0.75	0.66	-0.02	-0.01
<i>Angl</i>	0.91	0.40	0.05	0.01

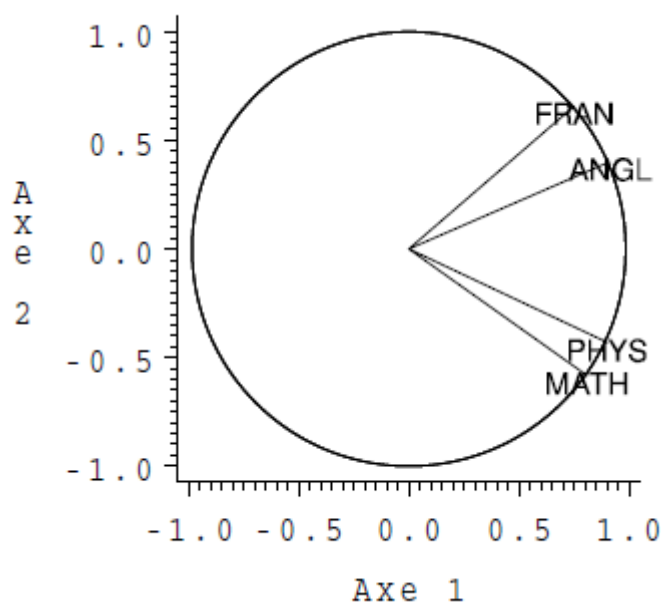


Figure II.9: Représentation des variables.

Interprétation :

Ainsi, on voit que le premier facteur est corrèle positivement, et assez fortement, avec chacune des 4 variables initiales : plus un l'élève obtient de bonnes notes dans chacune des 4 disciplines, plus il a un score élevé sur l'axe 1 ; réciproquement, plus ses notes sont mauvaises, plus son score est négatif ; l'axe 1 représente donc, en quelques sortes, le résultat global (dans l'ensemble des 4 disciplines considérées) des élèves. En ce qui concerne l'axe 2, il oppose, d'une part, le français et l'anglais (corrélations positives), d'autre part, les mathématiques et la physique (corrélations négatives). Il s'agit donc d'un axe d'opposition entre disciplines littéraires et disciplines scientifiques, surtout marqué par l'opposition entre le français et les mathématiques.

II.2.3 Représentation simultanée :

L'ACP permet pour une interprétation simultanée du nuage N_I et du nuage N_K de représenter ces deux nuages simultanément sur les plans issus des premiers axes factoriels. Nous devons cependant prendre garde au fait que les deux nuages ne sont en réalité pas dans les mêmes espaces qui ont des dimensions différentes. Cette représentation simultanée est essentiellement pragmatique.

En effet, le nuage des individus N_I et le nuage des variables N_K sont deux représentations du même tableau de données X . Ainsi des relations fortes (relation de dualité) lient ces deux nuages. Tout d'abord, l'inertie totale des deux nuages est la même :

$$\lambda = \frac{1}{I} \sum_{i \in I} \sum_{k \in K} \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right)^2 \quad (18)$$

De plus, les projections de tous les points du nuage des individus N_I sur le premier axe factoriel u_1 constituent une nouvelle variable (appelée premier facteur, notée F_1) qui se confond à la norme près à la première composante principale (illustrées sur la **Figure II.10**). Ainsi le vecteur F_1 dans \mathbb{R}^I est colinéaire à v_1 (axe factoriel de N_K). Il en est de même pour les projections sur les autres facteurs qui correspondent aux composantes principales de même rang. De manière symétrique, les coordonnées des projections du nuage N_K sur l'axe factoriel v_1 constituent un nouvel individu (premier facteur, noté G_1), ce que nous représentons sur la **Figure II.11**. Ce vecteur G_1 de \mathbb{R}^K est colinéaire à u_1 (axe factoriel de N_I). Cette notion d'individu type est moins employée que celle de composante principale. Il est souvent plus facile de tenter de se ramener à des individus réels comme individu type.

Les relations algébriques des deux dualités précédentes au rang s sont données par :

$$\begin{cases} F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} \frac{x_{ik} - \bar{x}_k}{s_k} G_s(k) \\ G_s(i) = \frac{1}{I} \frac{1}{\sqrt{\lambda_s}} \sum_{i \in I} \frac{x_{ik} - \bar{x}_k}{s_k} F_s(k) \end{cases} \quad (19)$$

Où λ_s est l'inertie projetée du nuage N_I (ou du nuage N_K) sur l'axe factoriel au rang s. Notons que les facteurs peuvent être négatifs.

Cette représentation est donc essentiellement pour l'interprétation.

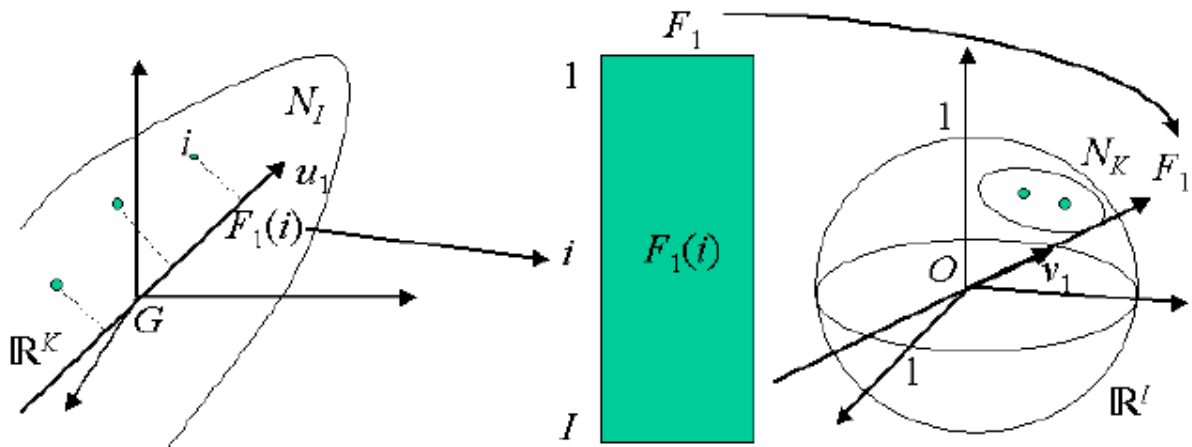


Figure II.10 : Forme de dualité exprimant le nuage N_I en fonction du nuage N_K .

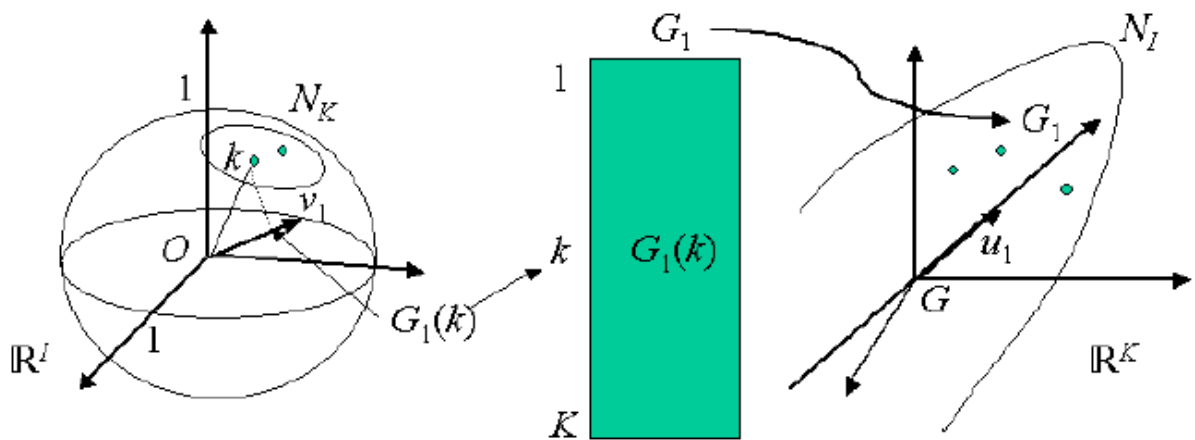


Figure II.11 : Forme de dualité exprimant le nuage N_K en fonction du nuage N_I .

- Représentation simultanée sur notre joue de donnée :

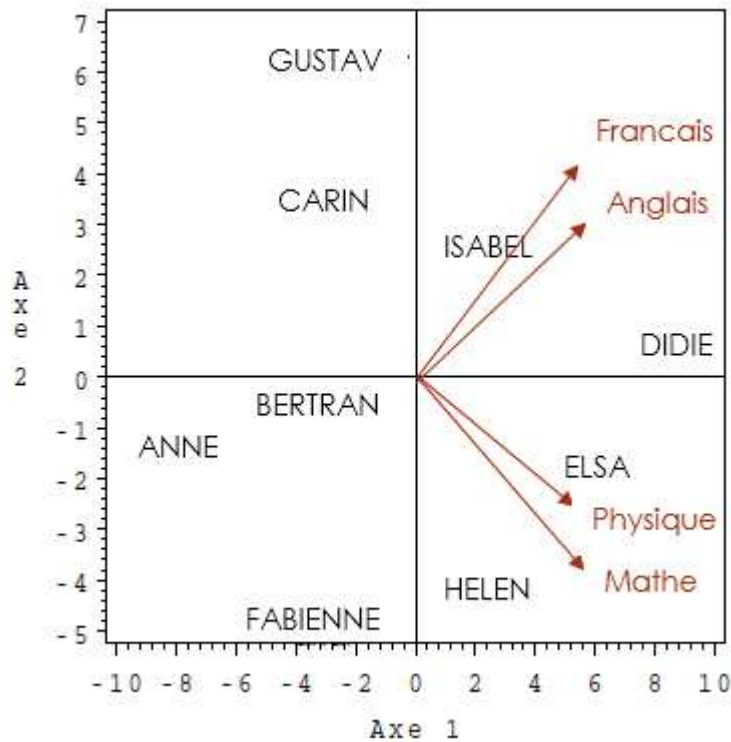


Figure II.12 : Représentation simultanée.

II.1.4 Interprétation :

Les axes factoriels donnent des images approchées des nuages de points N_I et N_K . Il est donc nécessaire de définir des indicateurs pour mesurer la qualité de l'approximation. L'étude d'un plan de projection des sous-espaces vectoriels doit toujours se faire conjointement avec l'étude des indicateurs. En effet, deux points (individus ou variables) peuvent se trouver très proches dans un plan de projection, alors qu'en réalité ils sont éloignés. Nous présentons ici les principales aides à l'interprétation que nous retrouvons dans [EP90].

Qualité de représentation d'un nuage par un axe :

Cette qualité est donnée par le pourcentage d'inertie associé à un axe, c'est-à-dire le rapport de l'inertie de la projection du nuage sur l'axe et de l'inertie totale du nuage :

$$QLT_N = \frac{\sum_{i \in N} (OH_i^S)^2}{\sum_{i \in N} (Oi)^2} (20)$$

Cette qualité mesure " l'importance " d'un axe factoriel. Bien sûr les premiers axes auront plus d'importance que les suivants. Nous devons juger ces pourcentages en fonction de la taille du tableau. Par exemple, 10% est une valeur faible si le tableau comporte 10 variables ; c'est une valeur forte dans le cas de 100 variables. Du fait de la dualité, il est équivalent de calculer ces pourcentages d'inertie à partir du nuage N_I des individus ou N_K des variables.

Contribution d'un élément à l'inertie d'un axe :

La contribution d'un élément i à l'inertie d'un axe s est donnée par le rapport de l'inertie de la projection de i sur l'axe s et de l'inertie de la projection de l'ensemble du nuage sur l'axe s :

$$CT_s(i) = \frac{(OH_i^s)^2}{\sum_{i \in N} (O_i)^2} \quad (21)$$

La contribution est importante si elle est proche de 1 pour les variables et doit être rapportée au tableau pour les individus. Ce rapport permet de mettre en évidence le sous-ensemble d'éléments ayant participé essentiellement à la construction de l'axe. L'interprétation devra en premier lieu s'appuyer sur ces éléments.

Pour aider à l'interprétation nous proposons de suivre le plan suivant :

Choisir le nombre d'axes, notons que le choix du nombre d'axes à retenir reste un problème car il n'y a pas de solutions rigoureuses. Les valeurs propres permettent de choisir ce nombre par exemple de telle sorte que le pourcentage d'information cumulée soit compris en 80% et 90% environ ou tel que toutes les valeurs propres soient supérieures à 1 ou encore lorsque un saut important sur l'histogramme des valeurs propres ou sur les recherches de paliers de celles-ci est observé. De plus le nombre d'axes ne doit pas être trop grand.

- Etudier les indicateurs de la qualité des approximations.
- Interpréter les facteurs simultanément :
 - à l'aide des contributions des individus,
 - à l'aide des coordonnées des variables (interpréter par axe et par plan),
 - à l'aide des coordonnées des individus.
- Mettre en évidence des typologies.

Il est possible de faire intervenir des éléments illustratifs (appelés également supplémentaires) afin d'aider l'opérateur à interpréter. Ces éléments, individus ou variables, n'interviennent pas dans la construction des axes factoriels, mais sont représentés pour l'étape d'interprétation.

Dans le cas des variables, il s'agit de variables quantitatives qui peuvent être continues ou nominales. L'ajout d'éléments illustratifs doit rester exceptionnel, car ils n'appartiennent normalement pas au champ strict de l'étude. Il peut cependant parfois être intéressant de supprimer un individu provoquant un effet de taille dans le calcul des axes, et de le faire apparaître pour interpréter ses projections en fonction des autres individus.

II.1.5 Conclusion :

Dans un premier temps résumons l'analyse en composantes principales à l'aide des neuf étapes de la **Figure II.13**.

- 1 : La première étape concerne la mise en forme des données brutes.
- 2 : La deuxième étape consiste à centrer et réduire les données. Elles sont centrées afin d'obtenir des propriétés intéressantes, et réduites pour uniformiser les unités de mesure.
- 3 : Le tableau est considéré comme juxtaposition de lignes.
- 4 : Le tableau est considéré comme juxtaposition de colonnes.
- 5 : Les individus sont représentés dans un espace à K dimensions. Dans le nuage N_I nous nous intéressons aux distances interindividuelles qui déterminent les ressemblances. Le centre de gravité G représente un individu moyen.
- 6 : Les variables sont représentées dans un espace à I dimensions. Nous nous intéressons ici aux angles des points. Le cosinus de l'angle est le coefficient de corrélation. Toutes les variables sont équidistantes de l'origine car les données ont été réduites, ainsi le nuage N_K se situe sur une hyper sphère.
- AF : Analyse Factorielle. Cette phase permet de mettre en évidence une suite de directions. Dans l'étape 7 ces directions sont des directions d'allongement, et dans l'étape 8 les axes s'interprètent comme des variables synthétiques.
- 7 : Cette étape consiste à projeter les points du nuage N_I sur le premier plan factoriel. C'est un premier ajustement, il peut y en avoir d'autres à suivre. Les distances s'interprètent alors comme des ressemblances entre les individus.
- 8 : Cette étape consiste à projeter les points du nuage N_K sur le premier plan factoriel. Ici aussi, c'est un premier ajustement, et il peut y en avoir d'autres à suivre. Les coordonnées représentent les coefficients de corrélation avec les facteurs sur les individus.

- Les relations de transition expriment les résultats d'une analyse factorielle (AF) dans un espace en fonction des résultats de l'autre.
- 9 : Cette étape est la représentation simultanée de nuages de points qui se trouvent initialement dans des espaces de dimensions différentes. Cette représentation issue des relations de transition permet des interprétations des axes simultanées.

L'ACP est une méthode puissante pour synthétiser et résumer de vastes populations décrites par plusieurs variables quantitatives. Elle permet entre autre de dégager de grandes catégories d'individus et de réaliser un bilan des liaisons entre les variables. Par cette analyse nous pouvons mettre en évidence de grandes tendances dans les données telles que des regroupements d'individus ou des oppositions entre individus (ce qui traduit un comportement radicalement différent de ces individus) ou entre variables (ce qui traduit le fait que les variables sont inversement corrélées). Les représentations graphiques fournies par l'ACP sont simples et riches d'informations. L'ACP peut être une première analyse pour l'étude d'une population dont les résultats seront enrichis par une autre analyse factorielle ou encore une classification automatique des données.

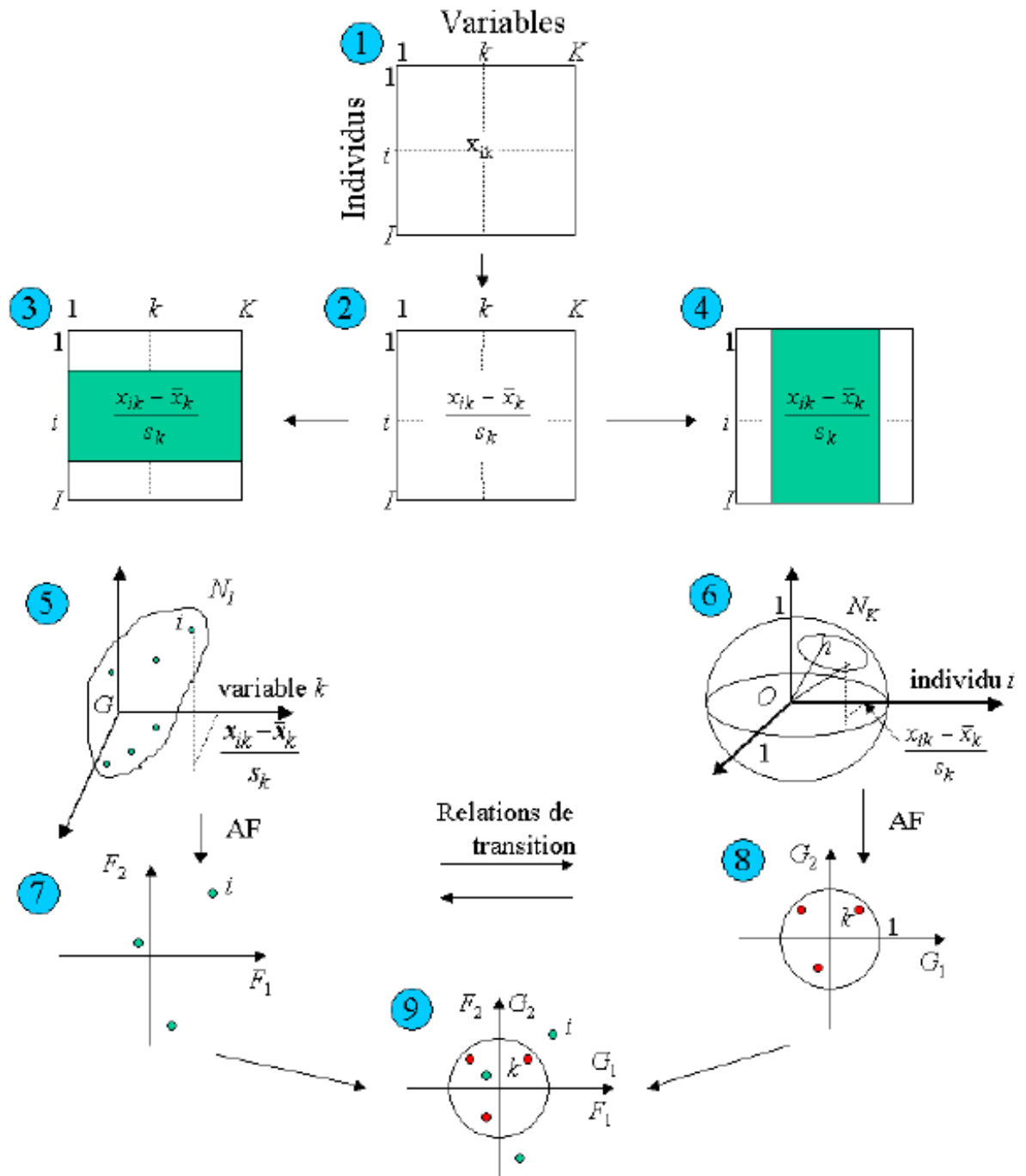


Figure II.13 : Résumé de l'ACP.

Chapitre III

Application et résultats sous Matlab

III.1 Introduction a Matlab :

Matlab est un puissant outil de calcul numérique, de programmation et de visualisation graphique. Son nom signifie *matrix laboratory*, c'est à dire un environnement interactif de travail avec des matrices. La facilité de développement des applications dans son langage fait qu'il est pratiquement devenu le standard dans son domaine.

Il fournit de nombreuses fonctions préprogrammées regroupées en boîtes à outils (toolbox) qui contiennent des fonctions spécialisées permettant d'utiliser l'environnement Matlab pour résoudre des classes spécifiques de problèmes. (Par ex : signal processing, statistics, control theory, optimization, ...).

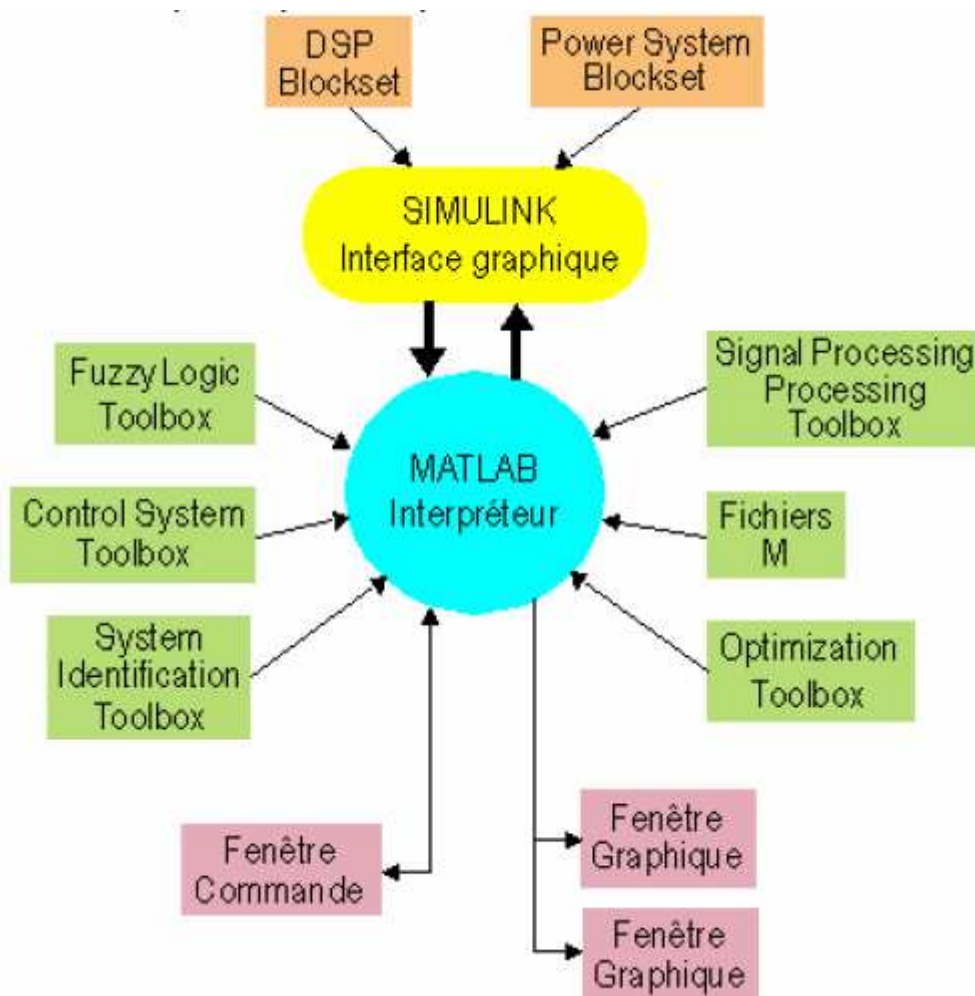


Figure III.1 : Les fonctions de Matlab.

Avec ses fonctions spécialisées, Matlab peut être aussi considéré comme un langage de programmation adapté pour les problèmes scientifiques. Matlab est un interpréteur :

Les instructions sont interprétées et exécutées ligne par ligne. Matlab fonctionne dans plusieurs environnements tels que Windows, Macintosh.

Il existe deux modes de fonctionnement :

- **Mode interactif:** Matlab exécute les instructions au fur et à mesure qu'elles sont données par l'utilisateur.
- **Mode exécutif:** Matlab exécute ligne par ligne un "fichier M" (programme en langage Matlab).

Les blocks de la (**Figure III.1**) précédente sont présentés comme suit :

Fenêtre Commande : Dans cette fenêtre, l'utilisateur donne les instructions et Matlab retourne les résultats.

Fenêtres Graphique : Matlab trace les graphiques dans ces fenêtres.

Fichiers M : Ce sont des programmes en langage Matlab (écrits par l'utilisateur).

Toolboxes : Ce sont des collections de fichiers M développés pour des domaines d'application spécifiques (Signal Processing Toolbox, System Identification Toolbox, Control System Toolbox, u- Synthesis and Analysis Toolbox ,Robust Control Toolbox, Optimization Toolbox, Neural Network Toolbox, Spline Toolbox, Chemometrics Toolbox, Fuzzy Logic Toolbox, etc.)

Simulink : C'est l'extension graphique de Matlab permettant de travailler avec des diagrammes en blocs.

Blocksets : Ce sont des collections de blocs Simulink développés pour des domaines d'application spécifiques (DSP Blockset, Power System Blockset, etc.).

III.2 Gestion des images sous Matlab :

III.2.1 Les fonctions Matlab :

Les fonctions Matlab utiles pour gérer les images, sont les suivantes :

- **image :** affiche une image (objet graphique Image)
- **imshow :** affiche une image (objet graphique Image) avec interpolation des couleurs ;
- **imread :** lit une image d'un fichier (formats standards) ;
- **imwrite :** écrit une image dans fichier (formats standards) ;
- **iminfo :** extrait des informations d'un fichier (formats standards) ;
- **print :** exporte une image (formats standards).

III.2.2 Les toolboxes de Matlab :

Le toolbox "image" est une collection de fonctions spécialisées dans le traitement des images appliqué à des domaines aussi variés que l'imagerie médicale, la télédétection et le contrôle non destructif.

Il existe trois Toolboxes Matlab spécifiques liées à l'acquisition, à l'analyse et au traitement des images :

- **Image Acquisition Toolbox** : acquisition d'images et de vidéos à partir de matériel industriel ;
- **Image Processing Toolbox** : analyse et traitement d'images ;
- **Mapping Toolbox** : analyse et visualisation de données géographiques.

III.2.3 Les images tests :

III.2.3.1 Images fournies avec l'image processing toolbox :

Cette toolbox est livrée avec plusieurs images dans les formats courants (jpeg, tiff, png).

III.2.3.2 Images au format JPEG :



football.jpg



greens.jpg

III.2.3.3 Images au format TIFF :



AT3_1m4_01.tif,
AT3_1m4_02.tif...
AT3_1m4_10.tif



autumn.tif



board.tif



cameraman.tif

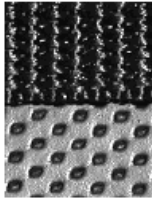


canoe.tif

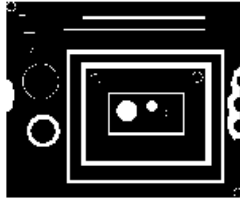


cell.tif

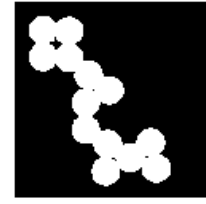
III.2.3.4 Images au format PNG :



bag.png



blobs.png



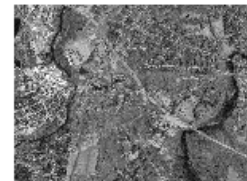
circles.png



coins.png



concordaerial.png



concordorthophoto.png

III.3 Présentations des images :

III.3.1 Tableau de pixels :

Matlab peut lire des images codées sur 8, 16, 24 ou 32 bits. Mais le stockage de ces données ne peut être fait qu'avec trois types de variables :

- le type uint8 (entier non signé de 8 bits) de plage [0 255] ;
- le type uint16 (entier non signé de 16 bits) de plage [0 65535] ;
- le type double (réel 64 bits) de plage [0 1].

III.4 Application de L'ACP dans le traitement d'image :

Dans le traitement d'image, on analyse des données représentées par des pixels sur les images. Dans la pratique on peut être confronté à des dimensionnalités élevées. L'ACP (Analyse en Composantes Principales) permet de réduire cette dimensionnalité en ne gardant que les attributs décorrélés.

III.4.1 Les utilisations les plus courantes de l'ACP :

On utilise généralement cette réduction de dimensionnalité pour deux utilisations :

- Une projection « visuelle » : elle va nous permettre, à partir de plusieurs images (dizaine, centaines, voire plus), d'obtenir le maximum d'informations contenu dans seulement quelques images (généralement une). Le traitement qui suit sera

d'autant plus facile que toutes les informations importantes dispersées dans les images se retrouveront mis en commun dans une seule image. Ces informations seront maintenant visibles pour un humain alors, qu'elles étaient dissimulées dans le panel d'images initiales, dû notamment à l'importance du nombre de ces images.

- D'un point de vue de l'optimisation : réduire un grand panel d'images à seulement quelques images ne supprime que très peu d'informations. En générale on obtient facilement 95% voir plus avec très peu de composantes au final. Donc pour des algorithmes lourds, il est plus intéressant de pratiquer ces algorithmes sur des images en sortie de l'ACP que sur les grandes quantités d'images initiales; surtout que l'ACP est un algorithme assez rapide.

III.4.2 Les termes employés dans l'ACP :

L'ACP est initialement appliqué pour la statistique, les termes employés ne sont donc pas empruntés à des termes d'imagerie. C'est pour cela qu'en utilise la transformation de **Karhunen-Loeve (TKL)**, qui est similaire à l'ACP ou les spectres de l'image (Rouge, Vert, Bleu pour une image couleur) sont appelées individus et les pixels sont des variables.

III.4.3 Principe de la transformation de Karhunen-Loeve (TKL) :

La TKL a pour but de transformer un tableau de données de i individus et de r réalisations pour pouvoir en donner une meilleure interprétation. L'idée de la transformation est de réaliser un changement de base pour obtenir de nouveaux axes où l'information contenue sur chaque axe est distribuée de façon optimale, c'est-à-dire de façon à éloigner au maximum le comportement des individus pour pouvoir mieux cerner leur comportement. Dans le cas d'une image décomposée en RGB, une image ($n \times p$ pixels) peut être considérée comme un tableau de données ayant 3 individus (R, G et B) et np réalisations. Le coefficient correspondant à la $i^{\text{ème}}$ réalisation du $j^{\text{ème}}$ individu est la valeur du $i^{\text{ème}}$ pixel de la $j^{\text{ème}}$ composante (comprise Entre 0 et 255).

On peut directement analyser les données brutes. L'origine du nuage est définie par l'observation qui a comme coordonnées (0, 0,0). Cette origine est très rarement intéressante car elle n'a aucun sens physique. Le plus intéressant est de choisir comme référence le centre de gravité (m_1, m_2, m_3). L'analyse devient une analyse centrée sur le nuage de points. Elle revient à remplacer chaque observation par l'écart de ses coordonnées avec les écarts avec la

moyenne. La TKL garde la prépondérance de certaines variables sur d'autres (on ne normalise pas le nuage de point).

Pour maximiser l'information sur chacun des axes, il faut se mettre dans le référentiel qui permet la comparaison de ces variables. Pour cela on se place dans le référentiel dont la base est constituée des vecteurs propres de la matrice de corrélation de tableau de variable centré.

III.4.4 Propriétés de l'algorithme :

La première propriété importante de la TKL est la totale décorrélation des nouvelles données. De cette propriété émerge un corollaire important du traitement d'image: la transformation des images peut se faire séparément et indépendamment. De plus il est très intéressant de séparer les différentes composantes de l'analyse colorimétrique, les informations sur la luminance sont contenues sur l'axe 1, les informations de chrominances sont contenues sur les axes 2 et 3.

La seconde propriété fondamentale de la transformation est qu'elle est optimale dans le sens où l'on obtient la plus grande distribution de variances sur les axes et que l'énergie du signal est entièrement conservée. Dans l'espace KL, toute l'énergie est contenue dans la somme des valeurs propres. Mieux, l'énergie de chaque axe est donnée par la valeur propre associée. Comme par définition $\lambda_1 > \lambda_2 > \lambda_3$, l'axe 1 est énergétiquement prépondérant sur les deux autres. Si on regarde sur les diverses réalisations, on peut remarquer que ce dernier est très prépondérant « dominant » vis-à-vis des deux autres.

III.4.5 Algorithme :

On commence par charger l'image, puis on la décompose en trois matrices représentant les trois composantes de l'image. Pour pouvoir utiliser la transformation de Karhunen-Loeve, nous devons réaliser un tableau de données. Pour cela, on considère les images comme des vecteurs de pixels ($\text{img}(i,j) = v((i-1)*p+j)$).

Ensuite, on fait la moyenne de chaque vecteur et on crée la matrice ayant 3 individus et $n * p$ réalisations X composée sur chaque colonne le vecteur centré des réalisations de chaque couleur.

On représente ensuite la matrice d'inertie $V = (1 / (n*p)) .* (X'*X)$. On calcule les valeurs propres et les vecteurs propres. Grâce à cela, nous avons la contribution des différents axes à l'image. L'énergie d'un axe est le ratio de la valeur propre de l'axe sur la somme des valeurs propres. Cette énergie est la quantité d'information expliquée par cet axe. Plus la qualité sera proche de 100%, plus la reconstruction sera bonne, ce qui est équivalent à dire que le dernier

axe ne contient que très peu d'information. Pour avoir des chances d'obtenir une bonne reconstruction, il est souhaitable d'avoir une qualité supérieure à 90%.

III.5 Résultats expérimentaux et interprétations :

Nous avons appliqué l'analyse en composantes principales sur différentes images pour récupérer les informations les plus pertinentes de celles-ci.

Il s'agira de faire une projection dans un nouvel espace, où les différentes directions (les nouveaux axes) seront les vecteurs propres de la matrice de covariance de l'image elle-même. De ces axes qui sont orthonormaux, nous garderons que ceux dont les valeurs propres sont les plus importants. Et donc ceux qui contiennent l'information la plus pertinente.

❖ Images test :

Image reel

Dans cette image, l'axe KL1 a récupéré 90.8596% de l'information et au niveau de deuxième axe KL2 il a récupéré 8.324 %.

On utilisant les deux axes principaux KL1 et KL2 ont à récupérer 99.1836 % et on a perdu 0.8164 % d'information; donc on a réussi à reconstitue notre image son perdre trop d'information.

Image réelle



KL2 : 8.324 %



KL1 : 90.8596 %



KL3 : 0.81637 %



Reconstitution avec deux axe KL1 et KL2 :

Image de départ

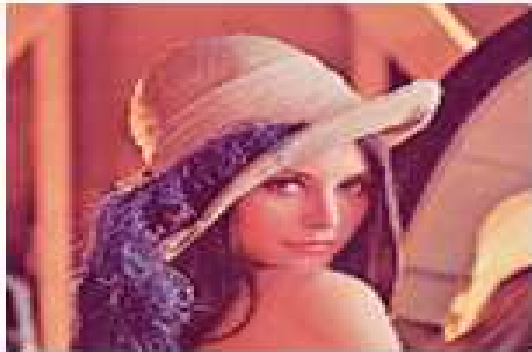


Image reconstituée en utilisant les deux axes



Image noir et blanc

Une image en noir et blanc introduit une très grande corrélation. De ce fait, la transformation donne un excellent résultat. Il suffirait de garder le premier axe pour reconstruire de façon parfaite l'image de départ.

Image en noir et blanc



KL1 : 100 %



KL2 : 1.2692e-14 %



KL3 : -8.145e-15 %



Reconstitution avec un seul axe KL1 :**Image de départ****Image reconstituée en utilisant un seul axe KL1****En va Confirmer que l'axe 1 représente la luminosité :**

Cette image permet de constater la nécessité des deux premiers axes pour reconstruire avec précision l'image. Dans ce cas, l'image représente un contraste entre deux objets (ici, le ciel et la fleur) qui sont tous les deux assez uniformes. Le premier axe explique très convenablement la luminosité, mais on a besoin du deuxième axe pour reconstruire car il apporte la majorité des détails.



KL2 : 10.0832 %



KL1 : 89.1312 %

KL3 : 0.78556 %



Reconstitution avec deux axes KL1 et KL2 :

Image de départ



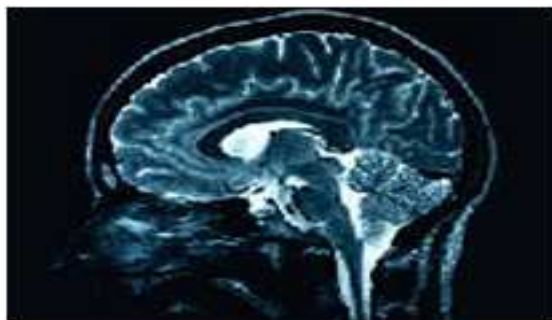
Image reconstituée en utilisant les deux axes



Image médicale

Cette image médicale une radio d'un cerveau humain est très lumineuse et possède un fort gradient de luminosité. Cela reflète parfaitement une propriété de la transformation qui insiste sur le fait que le premier axe possède toute l'information sur la luminance.

Image médicale



KL2 :2.179 %



KL1 :97.7592 %



KL3 :0.061798 %



Reconstitution avec deux axes KL1 et KL2 :

Image de départ

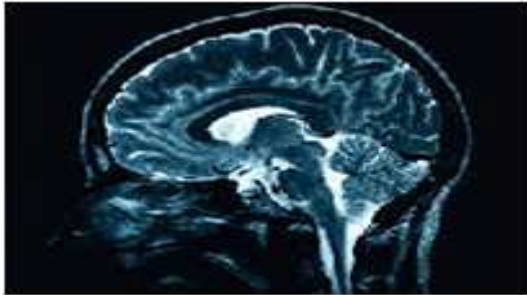
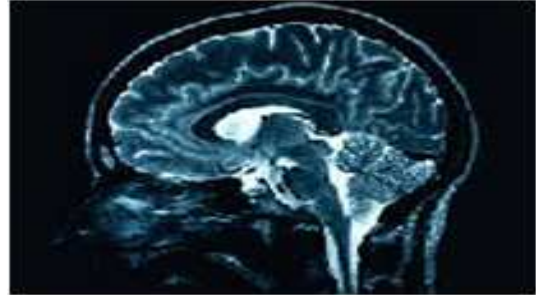


Image reconstituée en utilisant les deux axes



III.6 Conclusion :

L'ACP est un outil qui devient rapidement indispensable en traitement d'image dès lors que l'on travaille sur des images multi-spectrales. Il permet de réduire d'une manière significative le nombre d'images avec très peu de perte d'informations.

On a vu plusieurs utilisations de l'ACP notamment pour l'optimisation, mais on peut l'appliquer dans d'autres domaines que les images multi-spectrales. On peut penser à des images prises à un intervalle régulier.

Conclusion générale

Dans le cadre de cette thèse, nous avons pu identifier le problème lié à la réduction de dimension. Dans ce contexte, l'évaluation de ce paramètre est un enjeu majeur, car il nous est délicat et ardu de définir ce qui est opportun de ce qui ne l'est pas puisqu'on est dépourvu d'une référence. En général toute approche qui a pour but la sélection de variables peut s'appuyer sur la définition de la pertinence d'un sous ensemble de variables, et on peut dire qu'un sous ensemble est pertinent du moment où il participe au jaillissement d'une organisation en groupe d'un ensemble d'objet. Cependant, ceci n'est qu'une amorce de la résolution du problème.

Alors que l'un des enjeux actuels le plus prépondérant est la capacité à traiter de grandes masses de données, on a pensé à une première approche consistant à obtenir les images réduites afin de faciliter la classification. Nous avons utilisé le traitement de données comme l'analyse en composantes principales dans ce but de réduction en premier lieu, et surtout pour estimer un facteur crucial pour l'étape de classification –qui est le nombre de classes- en second lieu.

Après synthèse des résultats obtenus, par l'utilisation de l'analyse en composantes principales pour l'estimation du nombre des classes des images, il a été observé que la garantie d'une bonne reconstitution de l'image en tolérant une marge d'erreur maximum permet de dire que la dimension de l'image dans l'espace de projection est un multiples du nombre de classes de l'image. Le choix est donc assez subjectif puisqu'il nous revient de décider de la qualité de l'image reconstituée reste tributaire du choix heuristique des paramètres. Malgré les résultats encourageants ; on ne peut pas parler de la validation de la méthode avant de l'avoir essayé sur une grande quantité de données. Nous espérons que ce modeste travail reste un outil et un support de travail à tout étudiant qui veut s'intéresser à ce domaine.

Annexe

Les différentes techniques pour la réduction des paramètres de données.

II.2 Analyse en Composantes Indépendantes (ACI) :

Alors que l'ACP est une méthode qui recherche les axes (décorrélés et orthogonaux) qui représentent le mieux les données, l'analyse en composantes indépendantes (ACI) est une méthode non-supervisée qui recherche les axes qui sont les plus statistiquement indépendants les uns des autres (et donc décorrélés, mais pas forcément orthogonaux). Habituellement en traitement du signal, on suppose que les signaux soient seulement décorrélés, l'indépendance est une notion plus forte que la décorrélation. Dans le cas de variables gaussiennes, l'indépendance et la décorrélation sont équivalentes.

Historiquement, l'ACI a été introduite par Jeanny Herault et Christian Jutten en 1985 [Jutten, 1987]. Utilisée initialement pour des problèmes de séparation aveugle de sources, elle peut être également utilisée pour l'analyse et la compression de données, la détection bayésienne, la localisation de sources, et l'identification et la déconvolution aveugle. L'ACI n'est pas un outil de réduction du nombre de dimensions à proprement parlé, car dans certaines variantes de l'ACI, le nombre de dimensions de l'espace de projection est plus grand que le nombre de dimensions de l'espace initial. Elle est d'ailleurs, dans la pratique, souvent utilisée conjointement à une ACP. Mais nous nous intéressons à cette technique, car en plus de permettre la réduction du nombre de dimensions, elle a été utilisée récemment pour la classification de scènes naturelles. En effet, les neurophysiologistes se sont intéressés aux représentations efficaces des images et en particulier aux images de scènes naturelles. Leur théorie est la suivante : si les cellules du cortex visuel primaire sont affectées à des tâches spécifiques de reconnaissance de certaines structures géométriques, c'est peut-être parce que cette solution biologique au problème de la lecture d'une image est optimale en termes de l'analyse et de compression des données. Pour accéder efficacement à l'information pertinente, il faut optimiser le contraste, c'est-à-dire disposer de plusieurs points de vue. Par exemple, pour étudier un objet tridimensionnel, il vaut mieux prendre trois angles de vue les plus différents possible que trois angles de vue proches. L'ACI a également été utilisée pour analyser conjointement le texte et l'image.

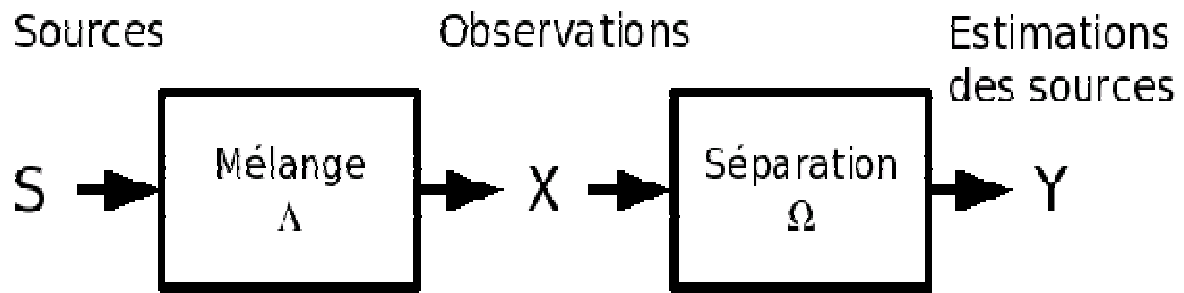


Figure II.14 : Principe de l'ACI.

II.2.1 Principe :

On suppose que les données ont été générées par r sources distinctes :

$$x_{j,i} = a_{j,1} \cdot s_{1,i} + a_{j,2} \cdot s_{2,i} + \dots + a_{j,r} \cdot s_{r,i} \quad \forall j \in \{1, 2, \dots, m\}, \forall i \in \{1, 2, \dots, n\} \quad (22)$$

Ou en notation matricielle :

$$\mathbf{X} = \mathbf{A}\mathbf{S}$$

On cherche alors la matrice $\mathbf{\Omega}$ telle que :

$$\mathbf{Y} = \mathbf{\Omega}\mathbf{X} = \mathbf{\Omega}\mathbf{A}\mathbf{S} \text{ et } \mathbf{Y} \text{ est une bonne estimation de } \mathbf{S}$$

La figure 1 résume le principe de l'ACI.

Par construction, les sources $\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_r$ sont indépendantes, c'est pourquoi on souhaite que les composantes $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_r$ soient les plus indépendantes possibles. Rappelons d'abord que l'indépendance statistique nécessite que la densité de probabilité conjointe soit le produit des probabilités des lois marginales, c'est-à-dire :

$$\mathbf{P}(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_r) = \mathbf{P}(\mathbf{Y}_1)\mathbf{P}(\mathbf{Y}_2) \dots \dots \mathbf{P}(\mathbf{Y}_r) \quad (23)$$

Soit $\Phi(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_r)$ une mesure de l'indépendance des composantes $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_r$, trouver les

composantes Y_1, Y_2, \dots, Y_r les plus indépendantes revient à trouver la matrice Ω tel que :

$$\min \Phi(\Omega X)$$

Il existe un grand nombre de méthodes pour mesurer l'indépendance, la méthode choisie détermine le principe de l'algorithme d'ACI. Il existe deux principales familles de mesures d'indépendance : celle basée sur le maximum de vraisemblance, et celle basée sur la non-gaussianité.

Posons, $\vec{y}_i = (Y_{1,i}, Y_{2,i}, \dots, Y_{r,i})^T = \Omega \vec{x}_i$ Une première méthode, appartenant à la première famille, définit $\Phi(\vec{y}_i)$ tel que :

$$\Phi_{IL}(\vec{y}_i) = \text{divKL}(P_{\vec{y}_i}, \prod_{k=1}^r P_{y_{k,i}}) \quad (24)$$

Où $P_{\vec{y}_i} = (P_{(y_1, y_2, \dots, y_r)})$ est la densité de probabilité (ou distribution) de \vec{y}_i , $P_{y_{k,i}}$ est la densité de probabilité de Y_k et divKL est la divergence de Kullback-Leibler. Une autre méthode, basée sur la non-gaussianité, c'est-à-dire l'écart entre la distribution de \vec{y}_i et la distribution gaussienne de mêmes moyenne $\mu_{\vec{y}_i}$ et écart-type $\sigma_{\vec{y}_i}$, définie par :

$$\Phi_{G(\vec{y}_i)} = \text{divKL}(P_{\vec{y}_i}, N(\mu_{\vec{y}_i}, \sigma_{\vec{y}_i})) \quad (25)$$

Trois conditions sont requises pour l'estimation du modèle ACI :

- au plus une composante peut être gaussienne,
- la matrice de mélange doit être de rang plein,
- le nombre de vecteurs d'attributs doit être supérieur au nombre de composantes à estimer (sources). Le nombre de composantes indépendantes correspond souvent au nombre de dimensions des vecteurs attributs originaux.

Dans la pratique, l'ACI nécessite que les données soient centrées et on utilise une ACP avant l'ACI pour obtenir une matrice diagonale. L'ACP transforme les données dans un espace réduit et décorrèle les données, puis l'ACI transforme les données de l'espace réduit vers un nouvel espace dans lesquels les dimensions sont indépendantes.

II.3 Analyse Linéaire Discriminante (LDA) :

Nous venons de voir que l'ACP et l'ACI peuvent être utilisés pour réduire le nombre de dimensions. Cependant, ce sont des techniques non-supervisées, elles n'utilisent donc pas les classes associées aux données. Contrairement à l'ACP, l'objectif de l'Analyse Linéaire Discriminante (LDA) est de réduire le nombre de dimensions tout en préservant au maximum les classes. Pour cela, elle cherche les axes tels que la projection des données dans l'espace engendré par ces axes permette une plus grande séparation des classes. La LDA est utile particulièrement dans les cas où les fréquences interclasses sont inégales. La **Figure II.15** compare les axes choisis par la LDA et l'ACP pour les mêmes données.

Nous donnons, dans une première partie, le principe de l'analyse de Fisher dans le cas de la séparation en deux classes par une seule droite. Puis, dans la deuxième partie, nous résumons les formules de la LDA dans le cas général.

II.3.1 Cas à deux classes :

Supposons que nous disposons d'un ensemble d'apprentissage où chaque document appartient à l'une des deux classes C_1 ou C_2 ($k = 1$ ou $k = 2$). Nous cherchons l'axe de direction $\vec{w} = (w_1, w_2, \dots, w_m)^2$ tel que les projetés des éléments soient le mieux séparés en classes. Soit le scalaire y_i représentant le projeté de \vec{x}_i sur cet axe, on a :

$$y_i = \vec{w}^T \vec{x}_i$$

Une mesure de la séparation entre les points projetés est la différence des vecteurs moyens de chaque classe. Le vecteur moyen de la classe C_k est : $\vec{\mu}_k = \frac{1}{n_{C_k}} \sum_{\vec{x}_i \in C_k} \vec{x}_i$ vecteur moyen de la classe C_k dans l'espace engendré par \vec{w} est le scalaire :

$$\mu_k = \frac{1}{n_{C_k}} \sum_{y_i \in C_k} y_i = \frac{1}{n_{C_k}} \sum_{\vec{x}_i \in C_k} \vec{w}^T \cdot \vec{x}_i = \vec{w}^T \cdot \vec{\mu}_k \quad (26)$$

C'est le projeté du point $\vec{\mu}_k$ sur \vec{w} . La distance au carré entre les projetés des vecteurs moyens des classes est :

$$|u'_1 - u'_2|^2 = (\vec{w}^T \cdot \vec{u}_1 - \vec{w}^T \cdot \vec{u}_2)^2 = \vec{w}^T (\vec{u}_1 - \vec{u}_2) (\vec{u}_1 - \vec{u}_2)^T \vec{w} = \vec{w}^T S_{B\vec{w}} \quad (27)$$

Où $S_B = (\bar{\mu}_1 - \bar{\mu}_2)(\bar{\mu}_1 - \bar{\mu}_2)^T$ est la matrice de dispersion interclasse. $S_B \bar{w}$ Est toujours dans la direction de $\bar{\mu}_1 - \bar{\mu}_2$. Pour séparer au mieux les projetés en classes, la valeur de $|\bar{\mu}_1 - \bar{\mu}_2|^2$ doit être grande par rapport à la dispersion des projetés de chaque classe C_k , définie par :

$$S_k^2 = \sum_{y_i \in C_k} (y_i - \hat{\mu}_k)^2 = \sum_{\bar{x}_i \in C_k} (\bar{w}^T \cdot \bar{x}_i - \bar{w}^T \cdot \bar{\mu}_k)^2 = \sum_{\bar{x}_i} \bar{w}^T (\bar{x}_i - \bar{\mu}_k) (\bar{x}_i - \bar{\mu}_k)^T \bar{w} = \bar{w}^T S_k \bar{w} \quad (28)$$

Avec \hat{S}_k^2 la dispersion (proportionnelle au carré de l'écart-type) de la classe C_k et S_k la matrice de dispersion de la classe C_k . On appelle matrice de dispersion intra-classe

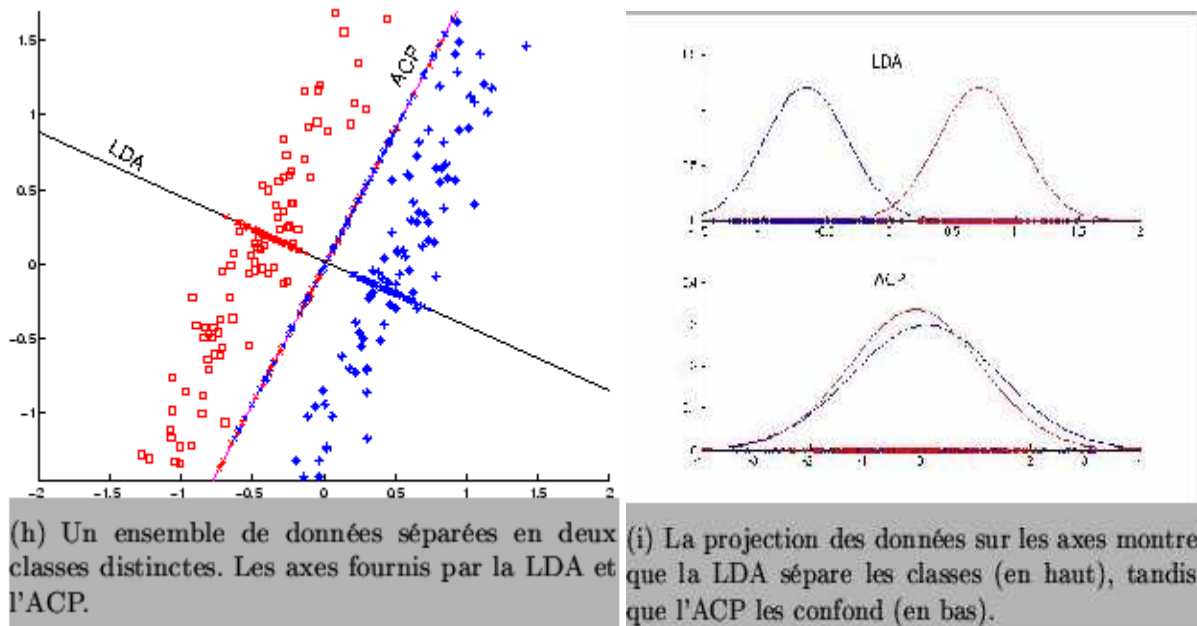


Figure II. 15 : Exemple de réduction de dimensions par LDA et par ACP. Les deux méthodes permettent de réduire de 2 à 1 dimension en trouvant, soit l'axe qui décrit le mieux les données pour l'ACP, soit l'axe qui sépare le mieux les classes pour la LDA.

Finalement, le critère de Fisher qui vise à mesurer la séparation des classes dans l'espace de projection engendré par \bar{w} est défini par :

$$J_{LDA}(\bar{w}) = \frac{|u'_1 - u'_2|^2}{s_1'^2 + s_2'^2} = \frac{\bar{w}^T S_B \bar{w}}{\bar{w}^T S_W \bar{w}} \quad (29)$$

Remarquons que ce critère suppose que les données de chaque classe sont réparties dans l'espace de projection sous la forme d'une gaussienne de vecteur moyen $\vec{\mu}_k$ et d'écart-type σ_k (proportionnel à \mathbf{S}_k).

On cherche \vec{w} qui maximise le critère de Fisher J_{LDA} . On peut exprimer ce problème sous la forme d'un problème de valeurs propres $S_B \vec{w} = \lambda \vec{w}$. Si S_w n'est pas singulière, on montre que dans notre cas à deux classes, les solutions sont de la forme :

$$\vec{w} = \mathbf{S}_w^{-1}(\vec{\mu}_1 - \vec{\mu}_2) \quad (30)$$

On peut utiliser la LDA pour construire un classifieur. Pour cela, il suffit en plus de trouver la borne qui sépare les \vec{x}_i de C_1 et ceux de C_2 .

II.3.2 Cas général :

Dans le cas général, la matrice de dispersion totale (*total scatter matrix*) est définie par :

$$\mathbf{S}_T = \sum_{\vec{x}_i \in A} (\vec{x}_i - \vec{\mu})(\vec{x}_i - \vec{\mu})^T \quad (31)$$

Où $\vec{\mu}$ est le vecteur moyen des données. La matrice de dispersion interclasse (*within-class scatter matrix*) est définie comme la dispersion des moyennes de chaque classe :

$$\mathbf{S}_w = \sum_{k=1}^{n_c} \sum_{\vec{x}_i \in C_k} (\vec{x}_i - \vec{\mu}_k)(\vec{x}_i - \vec{\mu}_k)^T = \sum_{k=1}^{n_c} \mathbf{S}_k \quad (32)$$

Où S_k est la matrice de dispersion des éléments de la classe C_k . La matrice de dispersion interclasse (*between-class scatter matrix*) est définie comme la moyenne des dispersions de chaque classe :

$$\mathbf{S}_B = \sum_{k=1}^{n_c} n_{C_k} (\vec{\mu}_k - \vec{\mu})(\vec{\mu}_k - \vec{\mu})^T \quad (33)$$

Où $\vec{\mu}_k$ est le vecteur moyen des éléments de la classe C_k .

Nous cherchons une transformation linéaire de matrice Ω qui maximise la dispersion des classes (mesurée par la matrice de dispersion interclasse) par rapport à la dispersion des éléments au sein des classes (mesurée par la matrice de dispersion interclasse). Une mesure

simple de dispersion est la valeur du déterminant des matrices de dispersion. Le critère de Fisher propose de choisir la transformation linéaire qui maximise le ratio :

$$J_{LDA}(\Omega) = \frac{|S'_B|}{|S'_W|} = \frac{|\Omega^T S_B \Omega|}{|\Omega^T S_W \Omega|} \quad (34)$$

Où \hat{S}_B et \hat{S}_W sont les matrices de dispersion interclasse et intra classe des projetés. On montre que les vecteurs colonnes de Ω sont les vecteurs propres \mathbf{w}_k qui correspondent aux plus grandes valeurs propres λ_k solutions de l'équation $\mathbf{S}_B \mathbf{w}_k = \lambda_k \mathbf{S}_W \mathbf{w}_k$. Si le déterminant de S_W vaut 1 (matrice isotropique), les vecteurs propres sont les vecteurs propres de S_B et sont dans la direction de $\vec{\mu}_k - \vec{\mu}$. La LDA revient à faire une ACP sur les vecteurs moyens des classes.

II.4 Diversité Marginale Maximale (MMD) :

La LDA suppose que les classes suivent une distribution gaussienne. Or, pour la plupart des problèmes, cette hypothèse n'est pas réaliste. Le critère classique «infomax» ne fait pas cette hypothèse. L'infomax suppose que les dimensions les plus discriminantes sont celles qui maximisent l'information mutuelle entre les descripteurs et les classes. [Vasconcelos, 2003] montre que l'infomax donne en général des résultats optimaux dans le sens de la minimisation de l'erreur bayésienne. Cependant, cette technique est complexe à mettre en œuvre, et ne peut être résolue que par approximation [Torkkola & Campbell, 2000].

D'après [Vasconcelos, 2003], dans le contexte de la reconnaissance visuelle, la sélection des axes discriminants par le critère de la diversité marginale maximale (MMD) (*Maximum Marginal Diversity*) conduit à des solutions optimales au sens de l'infomax. Plus exactement, il montre que MMD et infomax donnent des solutions identiques quand l'information mutuelle entre descripteurs n'est pas affectée par la connaissance de la classe. Or le critère MMD est facile à calculer.

Définition (Diversité Marginale) : la diversité marginale (MD) est la moyenne des divergences de Kullback-Leibler entre les probabilités conditionnelles des classes $p(\vec{w} | C_k)$ et leur moyenne $p(\vec{w})$:

$$J_{MD}(\vec{w}) = \sum_{C_k \in \mathcal{C}} p(\vec{w} | C_k) \log \frac{p(\vec{w} | C_k)}{p(\vec{w})} \quad (35)$$

Définition : (Diversité Marginale Maximale (Vasconcelos, 2003))

La meilleure solution pour un problème de sélection des axes les plus discriminants est de choisir les axes qui maximisent la diversité marginale maximale (MMD).

[Vasconcelos, 2003] montre expérimentalement sur des données synthétiques que la distance MD est, en plus d'être plus facile à calculer, plus efficace que la distance de Mahalanobis. La raison est peut-être que MD ne nécessite que le calcul des lois marginales, tandis que la distance de Mahalanobis nécessite d'estimer les distributions jointes, estimation qui est plus sensible à la malédiction de la dimension.

II.4.1 Algorithme :

Pour un problème de classification parmi n_c classes dans un espace à m dimensions, et en supposant que l'on connaît les probabilités a priori des classes (C_K), nous pouvons appliquer l'algorithme suivant pour déterminer les r dimensions qui ont la plus forte diversité marginale :

- pour chaque dimension \vec{w} :
 - pour chaque classe C_K :
 - calculer l'histogramme $\vec{h}_{w,k}$ estimant la probabilité $P(\vec{w}/C_K)$.
 - calculer l'histogramme $\vec{h}_w = \sum_{C_K \in c} \vec{h}_{w,k}$
 - calculer la diversité marginale :

$$J_{MD}(\vec{w}) = \sum_{C_k \in c} P(C_k) \vec{h}_{w,k}^T \log(\vec{h}_{w,k}/\vec{h}_w) \quad (36)$$

- où l'on suppose que le logarithme et la division sont effectués élément par élément de l'histogramme.
- ordonner les dimensions dans l'ordre décroissant de leur valeur $J_{MD}(\vec{w})$.
- sélectionner les r premières dimensions.

La difficulté est de déterminer le pas de la quantification de l'histogramme [Moddemeijer, 1989]. Un critère classique est de choisir ce pas de telle sorte que le nombre de dimensions n_h de l'histogramme soit l'arrondi supérieur de la racine carrée du nombre n .

$$n_h = \sqrt{n}$$

Un autre critère classique propose de choisir le nombre de dimensions de l'histogramme par :

$$n_h = 1 + \frac{10 \times \log n}{3} \quad (37)$$

La MMD, bien que récente (2002), a déjà été utilisée efficacement pour la reconnaissance de visage, et pour l'extraction et la segmentation d'objets dans des vidéos.

II.5 Discussion :

Tableau II.5 : Tableau comparatif des principales caractéristiques de 4 techniques de réduction des dimensions.

	ACP	ACI	LDA	MMD
Type	non-supervisé	non-supervisé	supervisé	Supervisé
Principe : recherche les axes...	... qui ont la plus grande variance	... qui sont les plus indépendants	... qui sont les plus discriminants	... qui sont les plus discriminants
Avantages	permet de bien représenter les données, supprime les redondances	recherche les "sources", supprime les redondances	méthode simple	plus rapide que l'infomax, pas d'hypothèse de gaussianité
Inconvénients	ne préserve pas les classes	ne préserve pas les classes	hypothèse de gaussianité, garde les redondances	garde les redondances

Chacun des descripteurs apporte de l'information (sa capacité propre à discriminer les classes) et du bruit (fluctuation) d'échantillonnage. Il est important de ne retenir dans l'analyse que les descripteurs dont l'information discriminante surpasse le bruit. Par exemple, si l'on doit séparer les individus entre hommes et femmes, la taille est un descripteur utile, le poids est un descripteur redondant par rapport à la taille, la couleur des yeux est un descripteur nuisible. De plus, réduire le nombre de dimensions permet d'éviter le problème des grandes dimensions. Les techniques telles que l'ACP et l'ACI permettent de réduire les redondances, les techniques telles que la LDA et la MMD permettent de bien séparer les classes.

II.5.1 Avantages et inconvénients

Toutes ces techniques permettent de passer outre la malédiction de la dimension. Cependant, si le nombre de dimensions est très grand (plus de 512), il se peut que la réduction ne permette pas de réduire suffisamment le nombre de dimensions de l'espace sans perdre de l'information. De plus, ces techniques nécessitent de recalculer les matrices A et Ω si l'on ajoute de nouvelles données. Ensuite, elles sont très sensibles aux valeurs aberrantes, c'est pourquoi il est préférable de détecter ces erreurs, puis de normaliser les données avant d'effectuer ces opérations.

Lorsque l'ensemble d'apprentissage est petit, l'ACP donne de meilleurs résultats que la LDA, et est moins sensible aux données d'apprentissage [Martinez & Kak, 2001]. Lorsque le nombre de dimensions est très grand, la LDA nécessite de longs calculs [Yang & Yang, 2003]. Pour éviter ce problème, une technique appelée «ACP plus LDA» [Belhumeur *et al.* 1997, Swets & Weng, 1996] propose de combiner les deux opérations en utilisant certaines de leurs caractéristiques.

D'après [Comon, 1994], le concept d'ACI peut être vu en réalité comme une extension de l'ACP qui, elle, ne peut imposer l'indépendance qu'au second ordre et définit par conséquent des directions orthogonales.

Le tableau II.5 résume quelques-unes des caractéristiques de l'ACP, l'ACI, la LDA et la MMD.

Matlab

Toolbox traitement de signal :

L'un des premiers principes du traitement du signal est la représentation du signal dans une base vectorielle de fonctions (signaux simples) dans le but d'une meilleure extraction et interprétation des propriétés de ce signal.

$$f(t) = \sum_i F_i g_i(t) \quad (38)$$

$f(t)$ est le signal à analyser et $g_i(t)$ sont les fonctions de représentation auxquelles on exige de former une base vectorielle orthonormées.

Les fonctions sinus et cosinus sont les fonctions idéales pour cette représentation. En effet, Joseph Fourier a démontré que tout signal pouvait être représenté sous la forme d'une somme (souvent infinie) de signaux sinusoïdaux de fréquences et de phases différentes.

Si le signal est périodique, on parle de série de Fourier. Pour un signal quelconque (cas général), c'est plutôt la transformée de Fourier qui est utilisée.

Le signal et sa transformée de Fourier sont liés par les expressions suivantes.

$$f(t) = \int_i f(w) \cdot e^{j\omega t} \cdot d\omega \quad \text{et} \quad F(\omega) = \int_i f(t) \cdot e^{-j\omega t} \cdot dt \quad (39)$$

Notons que cette somme est plus souvent représentée sous forme d'intégrale.

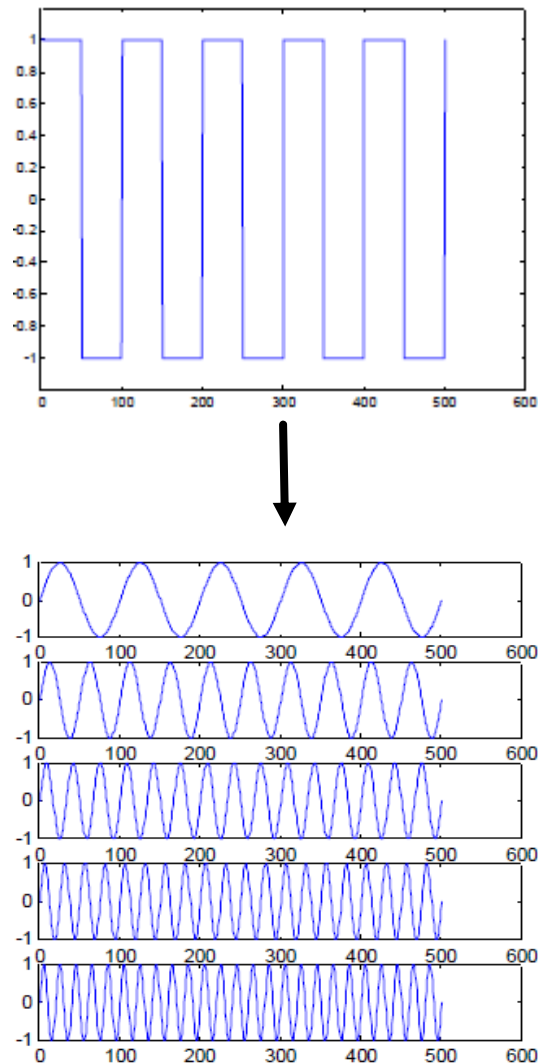
Exemple : TF de signal carré

```
t = 0:0.01:5;
```

```
y = square(2*pi*t);, plot(t,y)
```

```
z=fft(y);
```

```
y=abs(z(1))*cos(2*pi*t+angle(z(1)))+ abs(z(2))*cos(2*pi*t+angle(z(2)))+ ...
```



Le traitement de signal comporte outre les techniques basées sur la représentation de Fourier, d'autres transformations qui ne sont que, comme la TF, des représentations dans des espaces de fonctions orthonormées. Le TS comporte aussi des techniques de traitement statistique du signal.

Le toolbox "signal processing" de Matlab comporte une collection très riche de fonctions de traitement du signal. On peut citer les sous familles suivantes :

- Transformées : `dct`, `fft`, `hilbert`, `idct`, `ifft`, `czt`, ...
- -Analyse et implémentation de filtres : FIR, IIR, ...
- -Translation fréquentielle (modulation),
- -Traitement statistique du signal : `corrcoef`, `cov`, `pcov`, `spectrum`, `tfe`, `xcorr`, `xcov`, ..
- Modélisation paramétrique de série temporelle : `ar`, `arima`, `arburg`, `arcov`, `aryule`, `ident`

❖ **Exemple de calcul d'une ACP avec MATLAB :**

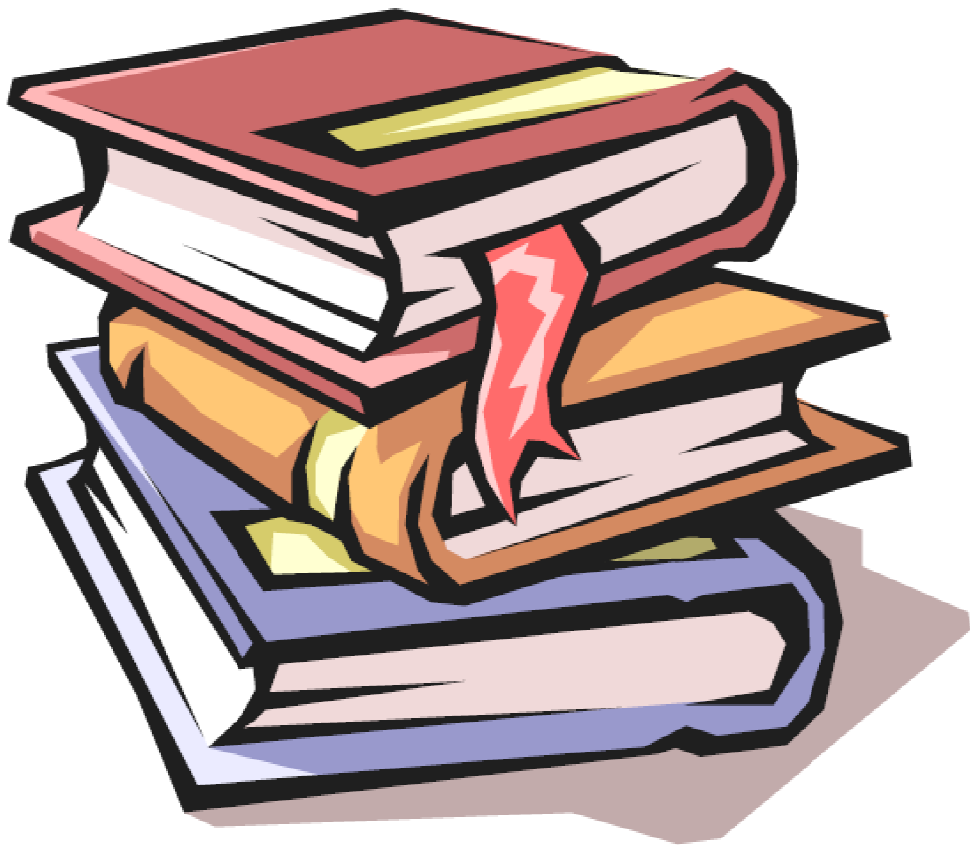
MATLAB	Résultat
<pre>load hald [n ligne n col] =size (hald);</pre>	<p>Lecture du tableau et calcul des dimensions</p> <p>n ligne =13 n col =5</p>
<pre>moy = mean (hald)'</pre>	<p>vecteur des moyennes</p> <p>moy = 7.4615 48.1538 11.7692 30.0000 95.4231</p>
<pre>ect=std(hald)'</pre>	<p>vecteur des ecarts-types</p> <p>ect = 5.8824 15.5609 6.4051 16.7382 15.0437</p>
<pre>Vunit= linspace (1,1,n ligne)'; y=hald-Vunit*moy';</pre>	<p>Centrage des données</p> <p>y = -0.4615 -22.1538 -5.7692 30.0000 -16.9231 -6.4615 -19.1538 3.2308 22.0000 -21.1231 3.5385 7.8462 -3.7692 -10.0000 8.8769 3.5385 -17.1538 -3.7692 17.0000 -7.8231 -0.4615 3.8462 -5.7692 3.0000 0.4769 3.5385 6.8462 -2.7692 -8.0000 13.7769</p>

	-4.4615 22.8462 5.2308 -24.0000 7.2769 -6.4615 -17.1538 10.2308 14.0000 -22.9231 -5.4615 5.8462 6.2308 -8.0000 -2.3231 13.5385 -1.1538 -7.7692 -4.0000 20.4769 -6.4615 -8.1538 11.2308 4.0000 -11.6231 3.5385 17.8462 -2.7692 -18.0000 17.8769 2.5385 19.8462 -3.7692 -18.0000 13.9769
ds=diag(ect)^-1 ;	<p style="text-align: center;">Matrice diagonale des 1/ect</p> ds = 0.1700 0 0 0 0 0 0.0643 0 0 0 0 0 0.1561 0 0 0 0 0 0.0597 0 0 0 0 0 0.0665
z=y*ds ;	<p style="text-align: center;">Matrice des données centrées et réduites</p> z = -0.0785 -1.4237 -0.9007 1.7923 -1.1249 -1.0985 -1.2309 0.5044 1.3144 -1.4041 0.6015 0.5042 -0.5885 -0.5974 0.5901 0.6015 -1.1024 -0.5885 1.0156 -0.5200 -0.0785 0.2472 -0.9007 0.1792 0.0317 0.6015 0.4400 -0.4323 -0.4779 0.9158 -0.7585 1.4682 0.8167 -1.4338 0.4837 -1.0985 -1.1024 1.5973 0.8364 -1.5238 -0.9285 0.3757 0.9728 -0.4779 -0.1544 2.3015 -0.0742 -1.2130 -0.2390 1.3612 -1.0985 -0.5240 1.7534 0.2390 -0.7726 0.6015 1.1469 -0.4323 -1.0754 1.1883 0.4315 1.2754 -0.5885 -1.0754 0.9291

$v = (z' * z) / n \text{ ligne};$	<p style="text-align: center;">Matrice des corrélations</p> <p style="text-align: center;">$v =$</p> <pre> 0.9231 0.2110 -0.7607 -0.2266 0.6745 0.2110 0.9231 -0.1285 -0.8981 0.7535 -0.7607 -0.1285 0.9231 0.0273 -0.4935 -0.2266 -0.8981 0.0273 0.9231 -0.7581 0.6745 0.7535 -0.4935 -0.7581 0.9231 </pre>
$[f \ h] = \text{eig}(v) ;$	<p style="text-align: center;">Vecteurs propres :</p> <p>$f =$</p> <pre> 0.2617 0.4129 0.5809 -0.5125 0.4012 0.6523 0.1884 -0.3899 0.4096 0.4683 0.2657 -0.0531 0.6747 0.6080 -0.3189 0.6586 -0.3791 -0.1039 -0.4471 -0.4603 -0.0408 -0.8047 0.2112 0.0034 0.5534 </pre> <p style="text-align: center;">Valeur propres :</p> <p>$h =$</p> <pre> 0.0015 0 0 0 0 0 0.0108 0 0 0 0 0 0.1837 0 0 0 0 0 1.4549 0 0 0 0 0 2.9646 </pre>
	Coordonnées des individus sur les axes

coord=z*f	<p>coord =</p> <p>0.0378 -0.0270 -0.5220 -1.8957 -1.8584</p> <p>-0.0335 -0.0807 -0.2510 -0.2270 -2.5599</p> <p>-0.0875 0.1263 -0.0575 -0.1904 1.2666</p> <p>-0.0279 0.1054 0.1668 -1.5735 -0.8425</p> <p>0.0181 -0.0314 -0.7616 -0.4862 0.3065</p> <p>-0.0226 -0.2015 0.1293 -0.1741 1.3120</p> <p>0.0121 0.0744 -0.2109 2.1293 1.0505</p> <p>0.0308 0.1629 0.4606 0.7035 -2.6944</p> <p>-0.0480 -0.0588 -0.0125 1.4343 -0.3722</p> <p>0.0189 -0.0039 0.8598 -1.8359 2.1386</p> <p>0.0255 -0.1143 0.5612 1.3049 -1.7827</p> <p>0.0339 -0.0611 -0.0267 0.3834 2.0688</p> <p>0.0424 0.1098 -0.3356 0.4274 1.9671</p>
------------------	--

Bibliographie



- [And 87] M.André. "*Introduction aux Techniques de Traitement d'images*".
Editions: Eyrolles. 1987.
- [Belhumeur *et al.*, 1997] Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J.
1997.
Eigenfaces vs. Fisherfaces: Recognition Using Class
specific Linear Projection.
*IEEE Transactions on Pattern Analysis and Machine
Intelligence*, **19**(7), 711-720.
- [CC 85] M.Coster, M.Chermant. "*Précis d'analyse d'Image*".
Editions : CNRS.1985.
- [Comon, 1994] Independent Component Analysis, a new concept ?
*Signal Processing, Special issue on Higher-Order
Statistics*, **36**(3), 287-314.
Cooper, 1970.
Cooper, W. S. 1970.
The potential usefulness of catalog access points other
than author, title and subject.
- [Dra 93] D.Drascic. "*Stereoscopic Vision and Augmented Reality*".
Scientific Computing et Automation, PP 31-34. Juin
1993.
- [EP90] B. Escoffier et J. Pagès : *Analyses factorielles simples et
multiples – objectifs, méthodes et interprétations*. Dunod, 1990.
- [FAB 97] J.Feldmar, N.Ayache, F.Betting. "*3D-2D Projective
Registration of Free Form Curves and Surfaces*". *Computer
Vision and Image Understanding*, vol. 65 (3), pp 403-
424. 1997.

- [Gra 84] R.M.Gray. "*Vector Quantization*". IEEE ASSP Magazine.
Avril 1984
- [Har 95] A.d'Hardancourt. "*Fou du Multimédia*". Editions: Sybex.
1995.
- [HM 95] R.Horaud, O.Monga. "*Vision par Ordinateur: Outils
Fondamentaux*".
Editions: Hermès. 1995.
- [Hotelling, 1933] Analysis of complex statistical variables into principal
components.
Journal of Educational Psychology, **24**, 417-441.
- [Jutten, 1987] Jutten, Chritian. 1987. *Calcul neuromimétique et
traitement dusignal - Analyse en composantes
indépendantes*.
Thèse d'état, Institut National Polytechnique de
Grenoble.
- [kun 93] M.Kunt. "*Traitement Numérique des images*". Vol.2. 1993.
- [KSR 01] G.Kliner, D.Stricker, D.Reiners. "*Augmented Reality for
Exterior Construst Applications*".Augmented reality and
Wearable Computers, W.Barfield and Caudell.ed.
Lawrence Erlbaum Press.
2001. W.Barfield and Caudell.ed. Lawrence Erlbaum
Press. 2001.
- [Karhunen, 1947] Uber lineare methodem in wahrsche inlichkeits

- rechnung.
Ann. Acad. Sci. Fenn. 37.
- [Loève, 1948] *Fonctions aléatoires de second ordre.*
Hermann, Paris.
- [LMP95] L. Lebart, A. Morineau et M. Piron : *Statistique exploratoire multidimensionnelle.* dunod, 1995.
- [Moddemeijer, 1989] On Estimation of Entropy and Mutual Information of Continuous Distributions.
Signal Processing, **16**(3), 233-246.
- [Martinez &Kak, 2001] Martinez, A. M., & Kak, A. C. 2001.
PCA versus LDA.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **23**(2), 228-233.
- [Rec 96] La Recherche 289. "*L'Image Médicale Morphologie Mathématique*". Dossier du CNRS. Juillet-Août 1996.
- [RBB+99] M.Roth, C.Brack, R.Burgkart, A.Czopf, H.Cotte, A.Schweikard.
"*Multi-View Contourless Registration of Bone Structures Using a Single Calibrated X-Ray Fluoroscope*". Computer Assisted Radiology and Surgery (CARS), pp 756-761. 1999.
- [sym 00] Symah Vision. Epsishttp: // www.epsis.com.
- [Saporta, 1990] *Probabilités, analyse des données et statistique.*
TECHNIP.

- [Swets & Weng, 1996] Using Discriminant Eigenfeatures for Image Retrieval.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **18**(8), 831-836.
- [Torkkola & Campbell, 2000] Torkkola, K., & Campbell, W. 2000. Mutual information in learning feature transformations.
In: Proceedings of International Conference on Machine Learning.
- [Urb 65] E.C.Urban. "The Information Warrior". *IEEE Spectrum* 32, vol 11, pp 66-70. 1995.
- [Vasconcelos, 2003] Vasconcelos, Nuno. 2003. Feature selection maximum Maximum marginal diversity: optimality and implications for visual recognition.
Pages 762-769 of: *Proceedings of IEEE International Conference on Image Processing (ICIP)*.
- [Yang & Yang, 2003] Yang, Jian, & Yang, Jing-Yu. 2003.
Why can LDA be performed in PCA transformed space?
Pattern Recognition, **36**(2), 563-566.
- Philippe Moultou et Henri Mariotte. *Cours d'Analyse de Données de l'Ecole Nationale Supérieure des Mines de Paris*. 1996
- Devaux, Gouton, Truchetet. *Optical Engineering. Karhunen-Loeve transform applied to region-base segmentation of color aerial images*. July 2001
- Bloch, Y. Gousseau, H. Maître, D. Matignon, B. Pesquet-Popescu, F. Schmitt, M. Siguelle, F. Tupin. *Le Traitement des Images*. 2003
- Cours IMN 359. Automne 2010.*