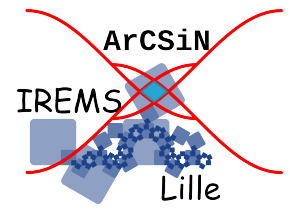


L'Analyse en Composantes Principales : découverte et applications possibles.

Fabrice Eudes- Lycée Gaston Berger Lille
Jean-Marc Duquesnoy- Lycée André Malraux Béthune
Groupe ArCSiN - IREMS de Lille
Dernière version : 8 mars 2026



« Celui qui ne bouge pas ne sent pas ses chaînes. »

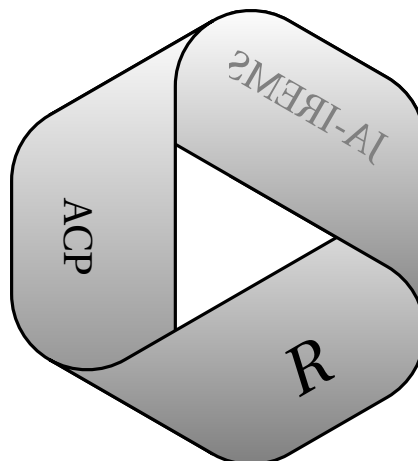


Rosa Luxemburg

« S'il y a des miséreux dans la société, des gens sans asile, sans vêtements et sans pain, c'est que la société dans laquelle nous vivons est mal organisée. On ne peut pas admettre qu'il y ait encore des gens qui crèvent la faim quand d'autres ont des millions à dépenser en turpitudes. C'est cette pensée qui me révolte! »



Louise Michel



Une pensée émue pour Belaïd Djellali

mon ami

mon frère de cœur

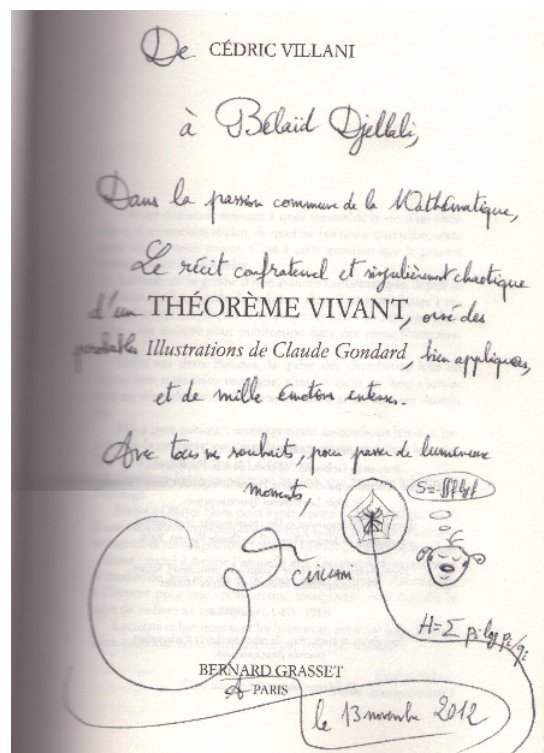


Table des matières

1	Introduction	4
2	Déroulé de l'atelier	5
3	Exemples d'ACP réalisées sur deux fichiers comportant 3 variables et 20 individus	6
4	Présentation théorique de l'analyse en composantes principales	6
4.1	Introduction	6
4.2	Notations	6
4.3	Les données	6
4.4	Matrice de variance-covariance et matrice de corrélation	7
4.5	Métrique sur l'espace des individus	9
4.6	Avertissement	10
4.7	Inertie du nuage	10
4.8	Inertie d'un sous-espace F	11
4.9	Inertie d'un axe	12
5	ACP	14
5.1	Recherche de l'axe principal 1	14
5.2	Notations	16
6	Interprétation des résultats d'une ACP	16
6.1	Qualité de représentation	16
6.2	Qualité de représentation des individus	17
6.2.1	Sur un axe factoriel	17
6.2.2	Sur un plan factoriel	17
6.3	Qualité de représentation des variables	20
6.3.1	Sur un axe factoriel	20
6.3.2	Sur un plan factoriel	21
6.4	Sélection du nombre k d'axes factoriels à retenir	22
7	Clustering par la méthode du k-means	23
7.1	Introduction	23
7.2	Mise en Place de l'Algorithme	23
7.3	Choix du Nombre de Clusters k	24
7.4	Avantage de l'Algorithme k-means	24
7.5	Limites de l'Algorithme k-means	24
8	ANNEXE 1 : Langage utilisé pour réaliser les différentes ACP présentées et sous quelles interfaces?	24
8.1	Langage	24
8.2	Interfaces	25
9	ANNEXE 2 : Outils Mathématiques	27
9.1	Espaces Métriques	27
9.2	Espaces Vectoriels	27
9.3	Espaces Vectoriels Normés	28
9.4	De la Norme à la Métrique	28
9.5	Espace vectoriel de dimension finie	29
9.6	Matrices de Passage et Changement de Base	30
9.7	Produit Scalaire et Espaces Euclidiens	31
9.8	Orthogonalité	32
9.9	Interprétation géométrique de l'orthogonalité	32
9.10	Matrices Associées aux Produits Scalaires	33
9.11	Changement de Base et Matrices de Passage	34

9.12	Supplémentaire orthogonal d'un sous-espace vectoriel	34
9.12.1	Définition et existence	34
9.12.2	Projection orthogonale	36
9.13	Diagonalisation	36
9.13.1	Lien Endomorphismes–Matrices	36
9.13.2	Diagonalisation des matrices symétriques	38

Table des figures

1	Théorème de Huygens, Inerties I_F et $I_{F^\perp}$	11
2	Inertie portée par un axe Δ_u	12
3	Projections de deux individus sur l'axe factoriel D_k	17
4	Projections de deux individus sur le plan factoriel P_{k-l}	18
5	Exemple de comparaison de la qualité de représentation de l'individu z_k	18
6	Projections orthogonales de l'individu z_i sur les axes D_k et D_l , ainsi que sur le plan P_{k-l}	19
7	p-sphère et cercle de corrélation dans le plan factoriel P_{k-l}	21
8	Projections de deux variables z^i et z^j dans le cercle de corrélation associé au plan factoriel P_{k-l}	21
9	Éboulis des valeurs propres	22
10	Tentative d'explication graphique du changement de bases.	31

Liste des tableaux

1	Tableau X de données brutes : n individus $\times$ p variables	6
2	Tableau Z de données centrées et réduites : n individus $\times$ p variables	8

1 Introduction

Dans un monde où la data est devenue omniprésente, nous sommes souvent confrontés à des ensembles de données complexes et multidimensionnels.

Une question fondamentale revient sans cesse : comment extraire une information pertinente et synthétique à partir d'une multitude de variables souvent corrélées, donc redondantes ? Comment visualiser la structure intrinsèque de ces données pour en extraire une information la plus exhaustive possible ?

C'est pour répondre à ces défis que l'**Analyse en Composantes Principales (ACP)** s'impose comme l'un des outils les plus puissants et fondamentaux de l'analyse exploratoire de données.

Bien plus qu'une simple technique statistique, l'**ACP** est une méthode de projection et de réduction de dimensionnalité qui permet de révéler la « simplicité » cachée au sein de données « complexes ».

Vous découvrirez comment l'**ACP** construit de nouveaux axes d'analyse, appelés composantes principales, qui capturent le maximum d'informations contenues dans les données originales.

Au-delà de la théorie, ce document illustrera la puissance de l'**ACP** à travers quelques exemples d'applications concrètes, et qui vous fourniront des clés pour comprendre le fonctionnement de l'**ACP**, interpréter ses résultats et l'appliquer à vos propres données pour en révéler toute la richesse informationnelle.

 L'atelier n'a la prétention de présenter l'**ACP** de façon exhaustive.

L'**ANNEXE 1** (voir [8](#)) donne des méthodes pour installer un compilateur interprétant le langage R, mais aussi quelques interfaces possibles pour réaliser une **ACP** sous R.

L'**ANNEXE 2** (voir [9](#)) permettra de rappeler, si besoin, les notions mathématiques indispensables à la compréhension de la partie théorique qui va suivre. Mais cette Annexe a été écrite avec toute l'humilité qui s'impose. Un lecteur averti constatera qu'il manque des pans entiers d'algèbre linéaire et bilinéaire pour constituer une Annexe rigoureuse et exhaustive.

Mais le document est là, le lecteur en fera ce que bon lui semble. 

Plongeons maintenant ensemble dans l'univers de la réduction de dimension.

2 Déroulé de l'atelier

L'atelier va être découpé en trois séquences, à savoir :

- **Séquence 1** : On présentera trois exemples d'application de l'**Analyse en Composantes Principales (ACP)** qui permettront de montrer comment exploiter un fichier de données, mais aussi d'interpréter les résultats obtenus.
Le troisième exemple appliquera aussi une méthode de **Clustering**, la méthode **k-means** en l'occurrence.
- **Séquence 2** : On donnera un script, écrit en langage PYTHON, qui permettra de visualiser graphiquement que toute matrice symétrique à coefficients réels est diagonalisable et qu'alors, ses vecteurs propres constituent une base orthonormée de $\mathbb{R}^n$. On se placera dans le cas où $n = 3$.
- **Séquence 3** : On proposera enfin des scripts.R qui permettront aux stagiaires de pratiquer l'**ACP**, à l'aide du logiciel R, sur des fichiers de données mis à disposition.

3 Exemples d'ACP réalisées sur deux fichiers comportant 3 variables et 20 individus

Deux exemples d'ACP sont présentés dans ce qui suit dans le but de constater, par le calcul et graphiquement, que les vecteurs propres associés aux valeurs propres d'une matrice de corrélation, carrée d'ordre 3, qui est symétrique, forment une base orthonormée de $\mathbb{R}^3$.

[Ouvrir le fichier ExemplesTroisVariables.pdf](#)

4 Présentation théorique de l'analyse en composantes principales

4.1 Introduction

L'analyse en composantes principales (ACP), développée en France dans les années 1960 par *J-P Benzécri*, est une méthode de statistique exploratoire permettant de décrire un tableau de données de type « individus-variables ». Lorsque les individus sont décrits par un nombre important de variables, aucune représentation graphique simple ne permet de visualiser le nuage de points formé par les données. L'ACP propose une représentation dans un espace de dimension réduite, permettant ainsi de mettre en évidence d'éventuelles structures au sein des données. Pour cela, nous recherchons les sous-espaces dans lesquels la projection du nuage déforme le moins possible le nuage initial.

4.2 Notations

 **Notations :**

Dans la suite du document :

- Si D est une matrice carrée diagonale d'ordre p telle que

$$D = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_p \end{pmatrix}$$

On la notera par la suite $D = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_p)$.

- La matrice carrée unité d'ordre p sera notée I_p .
- Si $A = (a_i^j)$ est une matrice de taille $n \times p$, on notera A^t la matrice transposée de la matrice A . A^t est de taille $p \times n$.

4.3 Les données

variables \ individus	variables				
	x^1	$\dots$	x^j	$\dots$	x^p
x_1			$\vdots$		
$\vdots$			$\vdots$		
x_i	$\dots$	$\dots$	x_i^j	$\dots$	$\dots$
$\vdots$			$\vdots$		
x_n			$\vdots$		

TABLE 1 – Tableau X de données brutes : n individus $\times$ p variables

Les données sont présentées sous la forme d'un tableau à n lignes et p colonnes, que l'on stocke sous la forme d'une matrice X de taille $n \times p$:

$$X = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^j & \cdots & x_1^p \\ x_2^1 & x_2^2 & \cdots & x_2^j & \cdots & x_2^p \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_i^1 & x_i^2 & \cdots & x_i^j & \cdots & x_i^p \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^j & \cdots & x_n^p \end{pmatrix}$$

- Chaque ligne $(x_i^1 \dots x_i^p)$ de X représente les valeurs des p variables de l'individu i .
- Chaque colonne $\begin{pmatrix} x_1^j \\ \vdots \\ x_n^j \end{pmatrix}$ de X représente les valeurs de la variable j des n individus.
- Par simplicité de langage, l'individu i sera identifié à la matrice ligne $x_i = (x_i^1 \dots x_i^p)$, tandis que la variable j sera identifiée à la matrice colonne $x^j = \begin{pmatrix} x_1^j \\ x_2^j \\ \vdots \\ x_n^j \end{pmatrix} = (x_1^j \dots x_n^j)^t$.

Remarque :

Les p variables peuvent avoir des unités différentes.

Dans ce cas, il sera utile de centrer et réduire chaque variable.

Pour ce faire, on définit, pour tout j variant de 1 à p , la moyenne arithmétique, la variance et l'écart-type de la variable x^j respectivement par :

$$\begin{cases} \overline{x^j} = \frac{1}{n} \sum_{i=1}^n x_i^j \\ (s^j)^2 = \frac{1}{n} \sum_{i=1}^n (x_i^j - \overline{x^j})^2 \\ s^j = \sqrt{(s^j)^2} \end{cases}$$

On pose alors, pour tout j variant de 1 à p et i variant de 1 à n ,

$$z_i^j = \frac{x_i^j - \overline{x^j}}{s^j},$$

Le tableau Z contenant les z_i^j est le tableau de données centrées et réduites associé au tableau X des données brutes.

Notations :

On notera aussi Z la matrice, de taille $n \times p$, associée au tableau Z des données centrées réduites.

Conséquences : Les variables (colonnes) de la matrices Z sont de moyenne 0 et de variance 1 et, de fait, centrer et réduire les données revient donc à effectuer une translation et une normalisation du nuage des individus.

4.4 Matrice de variance-covariance et matrice de corrélation

Les moyennes arithmétiques de chacune des p variables définissent le point moyen g du nuage des individus.

variables					
	z^1	$\dots$	z^j	$\dots$	z^p
individus					
z_1			$\vdots$		
$\vdots$			$\vdots$		
z_i	$\dots$	$\dots$	z_i^j	$\dots$	$\dots$
$\vdots$			$\vdots$		
z_n			$\vdots$		

TABLE 2 – Tableau Z de données centrées et réduites : n individus $\times$ p variables

Autrement dit, $g = (\overline{x^1} \dots \overline{x^p})$, où $\overline{x^j} = \frac{1}{n} \sum_{i=1}^n x_i^j$, pour tout j variant de 1 à p , est la moyenne des valeurs prises par les n individus pour la variable x^j .

La matrice Y définie par $y_i^j = x_i^j - \overline{x^j}$, pour tout j variant de 1 à p , est la matrice des données centrées obtenues à partir de la matrice X des données initiales.

Rappel : On a noté précédemment $(s^j)^2 = \frac{1}{n} \sum_{i=1}^n (x_i^j - \overline{x^j})^2$ la variance de la variable x^j , sa racine carrée s^j étant alors

son écart-type et $v_{kl} = \frac{1}{n} \sum_{i=1}^n (x_i^k - \overline{x^k})(x_i^l - \overline{x^l})$ est la covariance des variables x^k et x^l .

Définition :

On appelle matrice de variance la matrice carrée d'ordre p , symétrique, notée V, contenant, pour tout j variant de 1 à p , les variances $(s^j)^2$ des variables x^j sur la diagonale et, pour tout k et l variant de 1 à p , avec $l \neq k$, les covariances v_{kl} des variables x^k et x^l , en dehors de la diagonale principale.

Cette matrice s'exprime aussi $V = \frac{1}{n} Y^t Y$.

$$V = \begin{pmatrix} (s^1)^2 & v_{12} & \dots & v_{1p} \\ v_{21} & (s^2)^2 & \dots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \dots & (s^p)^2 \end{pmatrix}$$

Le coefficient de corrélation linéaire entre les variables x^k et x^l , pour tout k et l variant de 1 à p , est alors défini par $r_{kl} = \frac{v_{kl}}{s^k s^l}$.

On appelle matrice de corrélation la matrice carrée d'ordre p , notée R, telle que, pour tous i et j compris entre 1 et p , $R = (r_{ij}^j)$:

Ce qui donne :

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}$$

Propriétés :

- Pour tout i et j compris entre 1 et p , $r_{ij} = r_{ji}$. La matrice R est donc symétrique, propriété que l'on peut traduire par l'égalité

$$R^t = R$$

- Pour tout j variant de 1 à p ,

$$r_{jj} = 1$$

- $\det(R) \geq 0$.
- Pour tout i et j compris entre 1 et p ,

$$-1 \leq r_{ij} \leq 1$$

- La matrice R peut s'écrire $R = D_{\frac{1}{s}} V D_{\frac{1}{s}}$, où $D_{\frac{1}{s}}$ est la matrice diagonale dont la diagonale principale est constituée des $\frac{1}{s^j}$, j variant de 1 à p . D'où l'égalité

$$D_{\frac{1}{s}} = \text{diag}\left(\frac{1}{s^1}, \frac{1}{s^2}, \dots, \frac{1}{s^p}\right)$$

4.5 Métrique sur l'espace des individus

Dans l'espace des individus, que nous supposons ici être $\mathbb{R}^p$, chaque individu est représenté par un point. L'objectif de l'ACP est de visualiser le plus fidèlement possible le nuage de ces points dans un espace de faible dimension.

Une distance permettant de mesurer l'allongement du nuage des individus, ainsi que la proximité plus ou moins grande entre individus doit être choisie.

Notons d cette distance choisie.

On note alors M la matrice associée à la distance d . M est nécessairement symétrique définie positive de taille p .

Par exemple, la distance euclidienne dans $\mathbb{R}^p$ est associée à la matrice I_p .

Dans la suite du document, on fera l'abus de langage en parlant de la métrique M .

$\langle \dots | \dots \rangle_M$ et $\| \dots \|_M$ sont respectivement le produit scalaire et la norme associés à la métrique d , donc à la matrice M .

Le carré de la distance $d^2(x_i, x_j)$ entre deux individus x_i et x_j peut s'écrire :

$$d^2(x_i, x_j) = \|x_i - x_j\|_M^2 = \langle x_i - x_j | x_i - x_j \rangle_M = (x_i - x_j) M (x_i - x_j)^t$$

Dans le cas où toutes les variables sont exprimées dans la même unité, la métrique euclidienne convient. Dans ce cas $M = I_p$.

On peut alors écrire, pour deux individus x_i et $x_{i'}$:

$$\begin{aligned} d^2(x_i, x_{i'}) &= (x_i - x_{i'}) I_p (x_i - x_{i'})^t \\ &= (x_i^1 - x_{i'}^1, \dots, x_i^p - x_{i'}^p) I_p (x_i^1 - x_{i'}^1, \dots, x_i^p - x_{i'}^p)^t \\ &= \sum_{j=1}^p (x_i^j - x_{i'}^j)^2 \end{aligned}$$

Dans le cas où les variables ne sont pas toutes exprimées dans la même unité, la métrique la plus couramment utilisée est la métrique des inverses des variances $M = D_{1/s^2}$, où $D_{1/s^2} = \text{diag}\left(\frac{1}{(s^1)^2}, \frac{1}{(s^2)^2}, \dots, \frac{1}{(s^p)^2}\right)$.

Cette métrique permet à la fois de s'affranchir de l'unité de mesure et de donner à chaque variable la même importance dans le calcul de la distance.

Elle vérifie donc :

$$d^2(x_i, x_{i'}) = (x_i - x_{i'}) D_{1/s^2} (x_i - x_{i'})^t$$

Propriété :

Travailler avec la métrique $M = D_{1/s^2}$ sur le tableau X revient donc à travailler avec la métrique identité sur le tableau centré réduit Z .

Justification :

On peut écrire :

$$d^2(z_i, z_{i'}) = \sum_{j=1}^p \left(\frac{x_i^j - \bar{x}^j}{s^j} - \frac{x_{i'}^j - \bar{x}^j}{s^j} \right)^2 = \sum_{j=1}^p \frac{1}{(s^j)^2} \left((x_i^j - \bar{x}^j) - (x_{i'}^j - \bar{x}^j) \right)^2 = \sum_{j=1}^p \frac{1}{(s^j)^2} (x_i^j - x_{i'}^j)^2$$

□

On va donc centrer et réduire les variables, puis utiliser la métrique identité : cela permettra alors de réaliser une ACP que l'on appellera **ACP normée**.

4.6 Avertissement

AVERTISSEMENT

De ce qui précède, le choix est fait de travailler avec les données centrées réduites, et donc de présenter dans la suite du document l'**ACP normée**.

- La matrice des données sera dorénavant la matrice $Z = (z_i^j)$ de taille $n \times p$.
- Les individus seront, pour tout i variant de 1 à n , z_i .
- Les variables seront, pour tout j variant de 1 à p , z^j .
- Le centre de gravité du nuage des individus sera donc le point O^p de $\mathbb{R}^p$ de coordonnées $(0, 0, \dots, 0)$.
- On pourra parler aussi bien du point z_i que du vecteur $\overrightarrow{O^p z_i}$ assimilé parfois à z_i .

4.7 Inertie du nuage

L'inertie est une notion fondamentale en **ACP**, puisqu'elle est une mesure de la dispersion du nuage de points autour de son centre de gravité O^p . Plus l'inertie est grande, plus le nuage est dispersé, et au contraire, plus elle est petite, plus le nuage est concentré sur son centre de gravité.

Définition :

L'inertie totale du nuage autour de son centre de gravité g est :

$$I_g = \frac{1}{n} \sum_{i=1}^n d^2(z_i, O^p)$$

Propriété :

Dans le cas d'une **ACP normée**, l'inertie totale du nuage d'individus est donc égale au nombre p de variables, donc p .

Justification :

L'inertie totale peut s'écrire :

$$I_g = \frac{1}{n} \sum_{i=1}^n d^2(z_i, O^p) = \frac{1}{n} \sum_{i=1}^n \|z_i\|^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (z_i^j)^2$$

Par conséquent, rappelant que les variables z^j sont centrées et réduites, on a alors :

$$I_g = \sum_{j=1}^p \left(\frac{1}{n} \sum_{i=1}^n (z_i^j)^2 \right) = \sum_{j=1}^p 1 = p$$

□

4.8 Inertie d'un sous-espace F

Soit F un sous-espace vectoriel de $\mathbb{R}^p$, alors l'inertie du nuage constitué par les n individus par rapport au sous-espace F est égale à

$$I_F = \frac{1}{n} \sum_{i=1}^n d^2(z_i, h_F(z_i))$$

où $h_F(z_i)$ est la projection orthogonale de l'individu z_i sur le sous-espace F.

Propriété :

Soit $F^\perp$ le supplémentaire orthogonal de F dans $\mathbb{R}^p$, alors le théorème de Pythagore donne l'égalité

$$I_F + I_{F^\perp} = I_g$$

C'est une expression du théorème de Huygens.

AVERTISSEMENT

La figure qui suit représente le supplémentaire orthogonal de F, noté $F^\perp$ comme un axe. Il ne faudrait pas croire que la dimension de ce sous-espace vectoriel est nécessairement égale à 1.

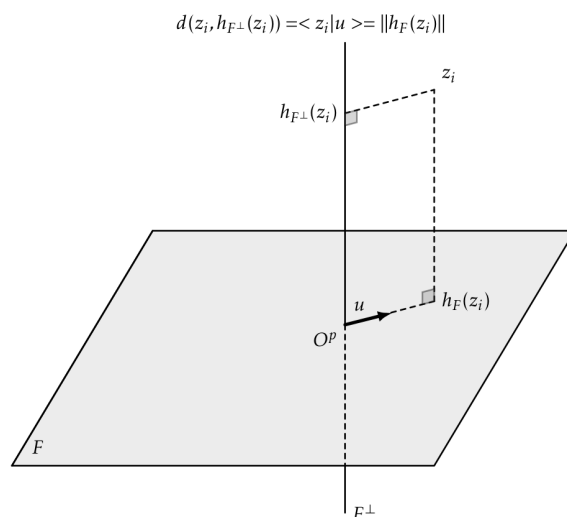


FIGURE 1 – Théorème de Huygens, Inerties I_F et $I_{F^\perp}$

Justification :

le théorème de Pythagore permet d'écrire, pour tout i variant de 1 à n :

$$d^2(O^p, z_i) = d^2(O^p, h_F(z_i)) + d^2(O^p, h_{F^\perp}(z_i))$$

donc

$$\frac{1}{n} \sum_{i=1}^n d^2(z_i, O^p) = \frac{1}{n} \sum_{i=1}^n (d^2(O^p, h_F(z_i)) + d^2(O^p, h_{F^\perp}(z_i)))$$

donc

$$\frac{1}{n} \sum_{i=1}^n d^2(z_i, O^p) = \frac{1}{n} \sum_{i=1}^n d^2(O^p, h_F(z_i)) + \frac{1}{n} \sum_{i=1}^n d^2(O^p, h_{F^\perp}(z_i))$$

d'où l'égalité :

$$I_g = I_F + I_{F^\perp}$$

□

Cas particulier : Lorsque F est un axe, autrement dit un sous-espace vectoriel de dimension 1, l'inertie $I_{F^\perp}$ par rapport à son supplémentaire orthogonal mesure alors l'allongement du nuage de points selon cet axe, et est appelée **l'inertie portée par l'axe** ou **inertie expliquée par l'axe**.

4.9 Inertie d'un axe

Propriété :

Si Δ_u est un axe, de vecteur directeur le vecteur unitaire u , alors, R étant la matrice de corrélation, l'inertie $I_{\Delta_u^\perp}$ portée par l'axe vérifie :

$$I_{\Delta_u^\perp} = u^t R u$$

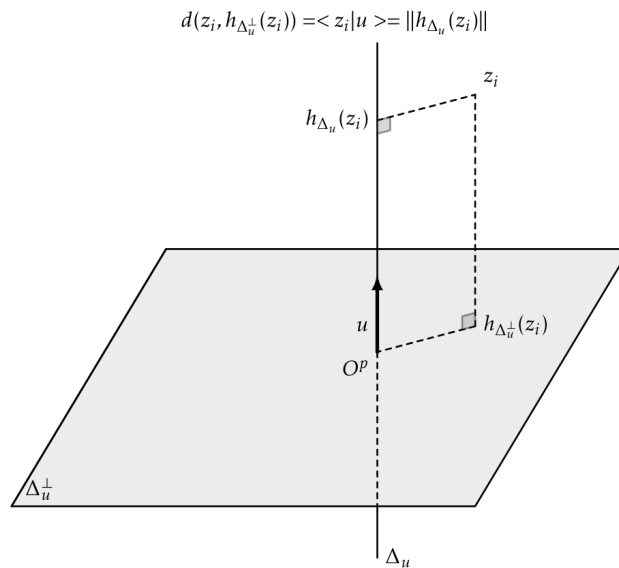


FIGURE 2 – Inertie portée par un axe Δ_u

⚡ AVERTISSEMENT

La figure précédente représente le supplémentaire orthogonal de Δ_u , noté $\Delta_u^\perp$ comme un plan. Il ne faudrait pas croire que la dimension de ce sous-espace vectoriel est nécessairement égale à 2.

Justification :

$$I_{\Delta_u^\perp} = \frac{1}{n} \sum_{i=1}^n d^2(z_i, \Delta_u^\perp)$$

donc

$$I_{\Delta_u^\perp} = \frac{1}{n} \sum_{i=1}^n \|h_{\Delta_u}(z_i)\|^2$$

où $h_{\Delta_u}(z_i)$ est le projeté orthogonal de z_i sur l'axe Δ_u .

Par conséquent

$$I_{\Delta_u^\perp} = \frac{1}{n} \sum_{i=1}^n \|\langle z_i | u \rangle \cdot u\|^2 = \frac{1}{n} \sum_{i=1}^n \langle z_i | u \rangle^2 \|u\|^2 = \frac{1}{n} \sum_{i=1}^n \langle z_i | u \rangle^2$$

donc

$$I_{\Delta_u^\perp} = \frac{1}{n} \sum_{i=1}^n (z_i u)^t (z_i u) = \frac{1}{n} \sum_{i=1}^n u^t z_i^t z_i u = u^t \left(\frac{1}{n} \sum_{i=1}^n z_i^t z_i \right) u$$

Par conséquent,

$$I_{\Delta_u^\perp} = u^t \left(\frac{1}{n} Z^t Z \right) u = u^t R u$$

R étant la matrice de corrélation. □

Propriétés :

Pour tous k et l compris entre 1 et p

- Le produit scalaire entre deux variables z^k et z^l est

$$\langle z^k | z^l \rangle = n \times r_{kl}$$

- Le carré de la norme d'une variable z^k est égal à n :

$$\|z^k\|^2 = (s^k)^2 = n$$

- Le cosinus de l'angle θ_{kl} entre deux variables z^k et z^l est égal à leur coefficient de corrélation linéaire :

$$\cos \theta_{kl} = \frac{\langle z^k | z^l \rangle}{\|z^k\| \times \|z^l\|} = \frac{r_{kl}}{s^k s^l} = r_{kl}$$

Justification :

Rappelons uniquement pourquoi la première égalité est vraie.

Le produit scalaire entre deux variables z^k et z^l est

$$\langle z^k | z^l \rangle = \sum_{i=1}^n z_i^k z_i^l = \sum_{i=1}^n \frac{(z_i^k - 0)(z_i^l - 0)}{1 \times 1} = n \times \frac{1}{n} \sum_{i=1}^n \frac{(z_i^k - \bar{z}^k)(z_i^l - \bar{z}^l)}{s^k \times s^l} = n \times r_{kl}$$

□

5 ACP

5.1 Recherche de l'axe principal 1

On cherche l'axe Δ_u tel que l'inertie $I_{\Delta_u^\perp}$ expliquée par cet axe soit maximale. Cela revient donc à chercher Δ_u tel que I_{Δ_u} soit minimale d'après le théorème de Huygens.

Puisqu'en projetant sur l'axe Δ_u , on perd l'inertie I_{Δ_u} , on aura bien une inertie restante $I_{\Delta_u^\perp}$ maximale, ce qui revient à déformer le moins possible le nuage des individus.

On rappelle qu'en **ACP normée**, l'inertie expliquée par l'axe Δ_u est égale à $I_{\Delta_u^\perp} = u^t R u$.

Il faut donc trouver le vecteur unitaire u solution du problème d'optimisation suivant :

$$\begin{cases} \max u^t R u \\ u^t u = 1 \end{cases}$$

Propriété :

La matrice de corrélation R étant symétrique et définie positive, elle est diagonalisable, et ses valeurs propres sont positives.

Par conséquent, $R = P \Lambda P^t$, où Λ est la matrice diagonale composée des valeurs propres $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, et où la matrice de passage P est la matrice orthogonale dont les p colonnes sont les vecteurs propres v_j , pour tout j compris entre 1 et p , associés aux valeurs propres λ_j de la matrice R .

Il vient alors que

$$u^t R u = \sum_{j=1}^p \lambda_j \langle u | v_j \rangle^2$$

Justification :

R est une matrice de corrélation, carrée d'ordre p , symétrique et définie positive, donc diagonalisable.

R est définie positive car, pour tout $y \in \mathbb{R}^p$, on a :

$$y^t R y = y^t \frac{1}{n} Z^t Z y = \frac{1}{n} y^t Z^t Z y = \frac{1}{n} (Z y)^t Z y = \frac{1}{n} \|Z y\|^2 \geq 0$$

Il existe alors une matrice P , orthogonale, qui vérifie l'égalité :

$$R = P \Lambda P^t$$

où :

- $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ dans laquelle les valeurs propres, λ_j , j variant de 1 à p , sont rangées dans l'ordre décroissant : $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.
- $P = [v_1 | v_2 | \dots | v_p]$, où P est une matrice orthogonale formée des vecteurs propres de la matrice R .
- $P^t P = P P^t = I_p$ car la matrice P est orthogonale.
- $(v_1, v_2, \dots, v_p)$ constitue une base orthonormée de $\mathbb{R}^p$.

De l'égalité

$$R = P \Lambda P^t$$

on déduit l'égalité

$$u^t R u = u^t (P \Lambda P^t) u$$

Soit $w = P^t u$. Alors $u = P w$ (car P est orthogonale et représente la matrice de changement de base).

Les composantes de w sont alors, pour tout j variant de 1 à p

$$w_j = v_j^t u = \langle u | v_j \rangle$$

et

$$u^t R u = u^t P \Lambda P^t u = (P^t u)^t \Lambda (P^t u) = w^t \Lambda w$$

Comme la matrice Λ est diagonale, on peut alors en déduire les égalités

$$w^t \Lambda w = \sum_{j=1}^p \lambda_j w_j^2 = \sum_{j=1}^p \lambda_j (v_j^t u)^2$$

D'où

$$u^t R u = \sum_{j=1}^p \lambda_j \langle u | v_j \rangle^2$$

□

Application à la recherche du premier axe principal

En **ACP normée**, on cherche le vecteur u qui maximise $u^t R u$ sous la contrainte $\|u\| = 1$.

De ce qui précède, on peut écrire :

$$\max_{\|u\|=1} u^t R u = \max_{\|u\|=1} \sum_{j=1}^p \lambda_j \langle u | v_j \rangle^2$$

On peut écrire ce qui suit, en ayant remarqué au préalable, la base $(v_1, \dots, v_p)$, formée des vecteurs propres, étant orthonormée, que $\sum_{j=1}^p \langle u | v_j \rangle^2 = \|u\|^2$:

$$u^t R u = \sum_{j=1}^p \lambda_j \langle u | v_j \rangle^2 \leq \lambda_1 \sum_{j=1}^p \langle u | v_j \rangle^2 = \lambda_1$$

Si on prend $u = v_1$, alors

$$u^t R u = v_1^t R v_1$$

qui donne alors

$$u^t R u = \sum_{j=1}^p \lambda_j \langle v_1 | v_j \rangle^2 = \lambda_1 \langle v_1 | v_1 \rangle^2 + \sum_{j=2}^p \lambda_j \langle v_1 | v_j \rangle^2$$

or $(v_1, v_2, \dots, v_p)$ est une base orthonormée, donc

$$u^t R u = \lambda_1$$

Conclusion :

Le vecteur v_1 , de norme 1, réalise donc le maximum de $u^t R u$.

Le vecteur propre associé à la plus grande valeur propre λ_1 de R , vérifie donc $I_{\Delta_{v_1}^\perp} = \lambda_1$.

Recherche des axes principaux suivants

Nous cherchons cette fois un vecteur unitaire u , orthogonal au vecteur (v_1) précédent, et maximisant la quantité $u^t R u$.

En nommant u_j les composantes du vecteur u dans la base $(v_1, v_2, \dots, v_p)$ des vecteurs propres, et en remarquant que $\langle u | v_1 \rangle = 0$, on a alors

$$u^t R u = \sum_{j=1}^p \lambda_j u_j^2 = \sum_{j=2}^p \lambda_j u_j^2 \leq \lambda_2 (u_2^2 + \dots + u_p^2) \leq \lambda_2$$

Le vecteur u maximisant cette quantité n'est autre que v_2 , le vecteur propre associé à la seconde plus grande valeur propre λ_2 de la matrice R , pour lequel on a donc $I_{\Delta_{v_2}^\perp} = \lambda_2$.

Conclusion :

On vient ainsi de construire un processus de recherche itérative des axes principaux D_{v_k} successifs, pour k variant de 1 à p .

AVERTISSEMENT

Si une valeur propre λ_k a un ordre de multiplicité supérieur ou égale à 2, le choix du vecteur propre associé est arbitraire, la part d'inertie expliquée étant la même, quel que soit le vecteur propre choisi.

5.2 Notations

Notations :

- Les axes D_{v_k} sont appelés **axes factoriels** ou **axes principaux**. Dans la suite du document, on les notera D_k .
- Pour tous k et l compris entre 1 et p , $k \neq l$, $D_k \perp D_l$.
- L'inertie expliquée par l'axe D_{v_k} est la valeur propre λ_k .
- L'inertie expliquée par le sous-espace, appelé **sous-espace factoriel**, engendré par les k axes factoriels $D_1, \dots, D_k$, noté F_k , est

$$I_{F_k} = \sum_{i=1}^k I_{\Delta_i^\perp} = \lambda_1 + \dots + \lambda_k$$

- Le **pourcentage d'inertie** expliquée par ce sous-espace est donc

$$\frac{\lambda_1 + \dots + \lambda_k}{p}$$

6 Interprétation des résultats d'une ACP

6.1 Qualité de représentation

Un des points les plus délicats de l'ACP est d'apprécier la perte d'information engendrée par la réduction de la dimension.

On a vu précédemment que la qualité de représentation du nuage dans le sous-espace factoriel F_k est exprimée par le pourcentage d'inertie expliquée $\frac{\lambda_1 + \dots + \lambda_k}{p}$.

Cette mesure globale doit être complétée par d'autres considérations comme la représentation des individus et des variables dans ce sous-espace factoriel.

Il peut en effet arriver que les axes retenus, bien que représentant une part importante de l'inertie globale du nuage, ne suffisent pas à expliquer correctement certains individus ou variables.

6.2 Qualité de représentation des individus

6.2.1 Sur un axe factoriel

La qualité de représentation d'un individu z_i sur un axe factoriel dépend de l'angle α entre l'axe factoriel et le vecteur z_i , autrement dit, entre les vecteurs z_i et $h_{D_k}(z_i)$, $h_{D_k}(z_i)$ étant la projection orthogonale de z_i sur l'axe D_k . Plus le cosinus de cet angle est grand et proche de 1 en valeur absolue, plus l'individu z_i sera proche de l'axe factoriel et sera donc bien représenté sur cet axe.

Pour cela, il suffit donc d'évaluer la valeur de $\cos^2(\alpha)$.

La figure qui suit permet de comparer la qualité de représentation de deux individus, z_i et z_j , sur l'axe factoriel D_k :

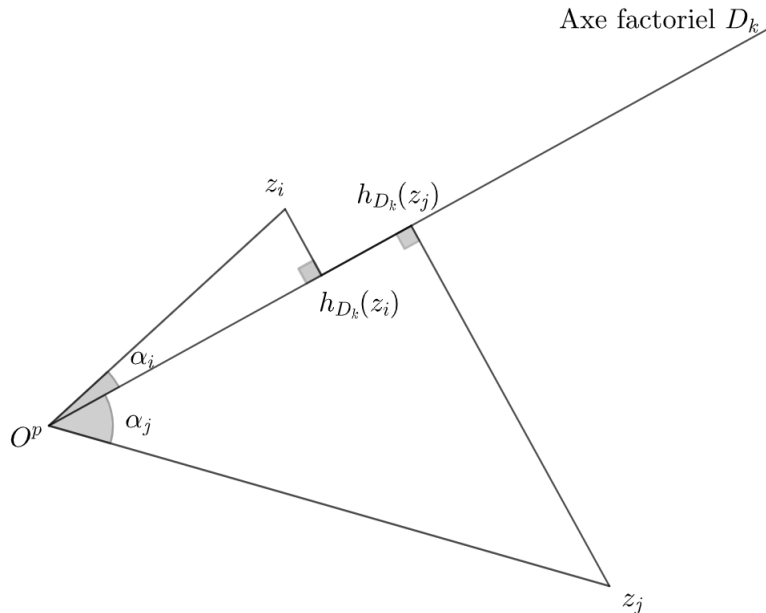


FIGURE 3 – Projections de deux individus sur l'axe factoriel D_k

L'individu z_i est mieux représenté que l'individu z_j .

La valeur de $\cos^2(\alpha_i)$ est en effet plus proche de 1 que celle de $\cos^2(\alpha_j)$.

6.2.2 Sur un plan factoriel

Deux individus, z_i et z_j peuvent être « éloignés » alors que leurs projections respectives $h(z_i)$ et $h(z_j)$ sur le plan factoriel P_{k-l} sont proches.

La figure qui suit illustre la remarque précédente :

D'autre part, de même que l'on mesure la qualité de représentation d'un individu sur un axe factoriel, on mesure sa qualité de représentation sur un plan factoriel.

Pour cela, on détermine la valeur de $\cos^2(\alpha)$, où α est l'angle entre les vecteur z_k et $h(z_k)$, $h(z_k)$ étant la projection orthogonale de z_k sur le plan factoriel.

Notations :

- z_i : individu dans l'espace $\mathbb{R}^p$
- D_k : k -ième axe factoriel.
- D_l : l -ième axe factoriel.
- $h_{D_k}(z_i)$ et $h_{D_l}(z_i)$: projections orthogonales respectives de z_i sur D_k et D_l .
- $h_{P_{k-l}}(z_i)$: projection orthogonale de z_i sur le plan factoriel P_{k-l} défini par les axes orthogonaux D_k et D_l .
- α_k : angle entre z_i et D_k .
- α_l : angle entre z_i et D_l .

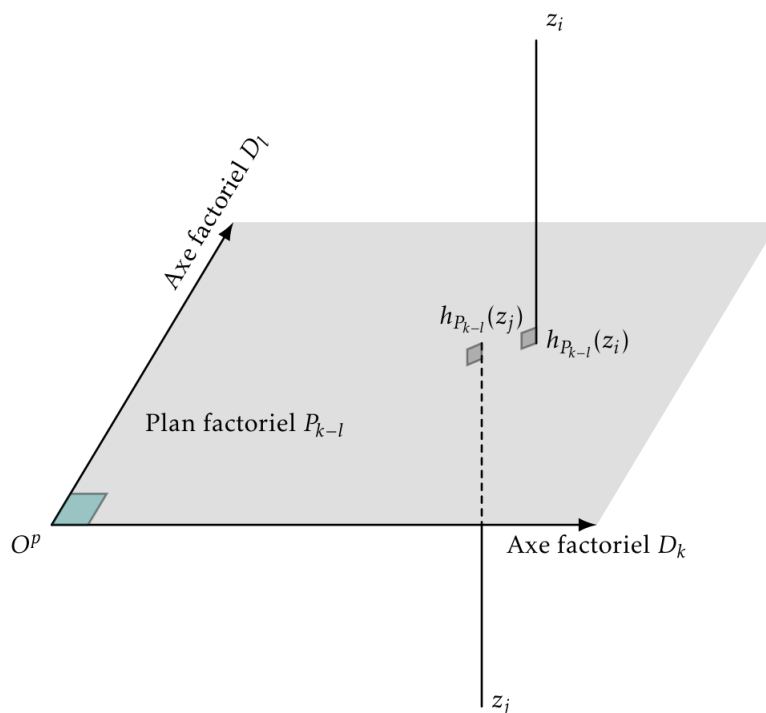


FIGURE 4 – Projections de deux individus sur le plan factoriel P_{k-l}

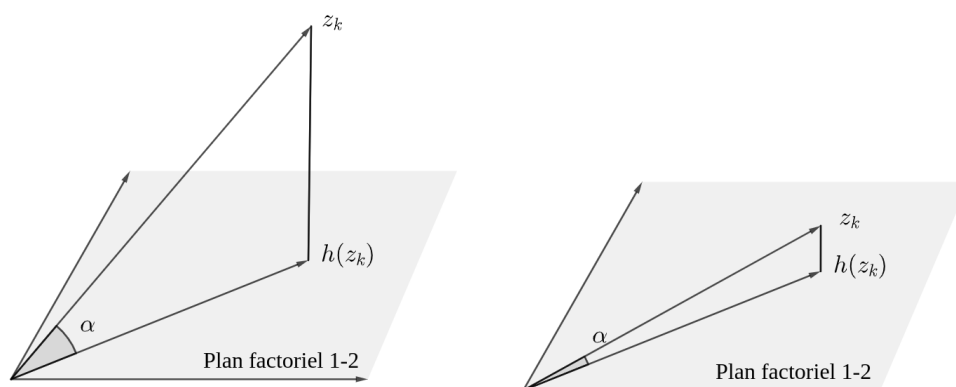


FIGURE 5 – Exemple de comparaison de la qualité de représentation de l'individu z_k

- α : angle entre z_i et sa projection orthogonale sur le plan P_{k-l} .

Propriétés :

On a les égalités suivantes :

- $\cos^2(\alpha_k) = \frac{\|h_{D_k}(z_i)\|^2}{\|z_i\|^2}$;
- $\cos^2(\alpha_l) = \frac{\|h_{D_l}(z_i)\|^2}{\|z_i\|^2}$;
- $\cos^2(\alpha) = \frac{\|h_{P_{k-l}}(z_i)\|^2}{\|z_i\|^2}$;
- $\cos^2(\alpha) = \cos^2(\alpha_k) + \cos^2(\alpha_l)$.

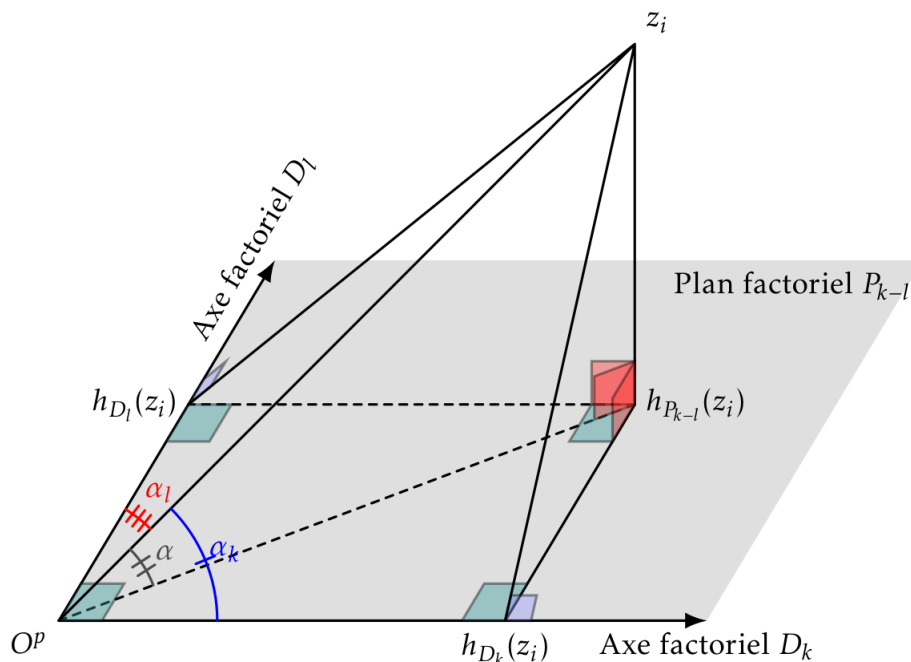


FIGURE 6 – Projections orthogonales de l'individu z_i sur les axes D_k et D_l , ainsi que sur le plan P_{k-l}

Justification :

On remarquera dans la suite de cette démonstration que l'on peut parler aussi bien du point z_i que du vecteur z_i (même constat pour les autres points).

Les triangles $O^p z_i h_{D_k}(z_i)$ et $O^p z_i h_{D_l}(z_i)$ sont rectangles respectivement en $h_{D_k}(z_i)$ et $h_{D_l}(z_i)$.

On a donc les deux égalités

$$\cos^2(\alpha_k) = \frac{\|h_{D_k}(z_i)\|^2}{\|z_i\|^2}$$

et

$$\cos^2(\alpha_l) = \frac{\|h_{D_l}(z_i)\|^2}{\|z_i\|^2}.$$

Le triangle $O^p z_i h_{P_{k-l}}(z_i)$ est rectangle en $h_{P_{k-l}}(z_i)$, donc

$$\cos^2(\alpha) = \frac{\|h_{P_{k-l}}(z_i)\|^2}{\|z_i\|^2}.$$

Or, le quadrilatère $O^p h_{D_k}(z_i) h_{P_{k-l}}(z_i) h_{D_l}(z_i)$ est un rectangle.

Par conséquent, le théorème de Pythagore permet d'écrire l'égalité

$$\|h_{P_{k-l}}(z_i)\|^2 = \|h_{D_k}(z_i)\|^2 + \|h_{D_l}(z_i)\|^2,$$

donc l'égalité

$$\cos^2(\alpha) = \frac{\|h_{D_k}(z_i)\|^2 + \|h_{D_l}(z_i)\|^2}{\|z_i\|^2},$$

par conséquent, on a l'égalité

$$\cos^2(\alpha) = \frac{\|h_{D_k}(z_i)\|^2}{\|z_i\|^2} + \frac{\|h_{D_l}(z_i)\|^2}{\|z_i\|^2}.$$

Conclusion :

$$\cos^2(\alpha) = \cos^2(\alpha_k) + \cos^2(\alpha_l).$$

□

6.3 Qualité de représentation des variables

6.3.1 Sur un axe factoriel

⚡ AVERTISSEMENT

La recherche des axes factoriels nous a permis d'identifier une nouvelle base orthonormée de l'espace $\mathbb{R}^p$. Les nouveaux axes obtenus sont des combinaisons linéaires des axes initiaux, et on peut donc associer aux variables initiales les combinaisons linéaires correspondantes afin de former de « nouvelles variables », appelées **composantes principales**. On peut alors s'intéresser à l'interprétation de ces « nouvelles variables », en cherchant à les relier aux « anciennes variables ».

On a, pour tout k variant entre 1 et p ,

$$c^k = \sum_{j=1}^p v_{kj} Z_j = Z v_k$$

Afin d'étudier le lien entre les composantes principales et les variables initiales, on s'intéresse aux corrélations entre les nouvelles et les anciennes variables. On définit d'abord la covariance entre une variable initiale z^j et une composante principale c^k :

$$\text{Cov}(z^j, c^k) = \frac{1}{n} v_k^t Z^t z^j = \frac{1}{n} v_k^t Z^t Z \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = v_k^t R \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \lambda_k v_k^t \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \lambda_k v_{kj}$$

On en déduit la relation entre z^j et c^k :

$$r(z^j, c^k) = \frac{\text{Cov}(z^j, c^k)}{\sqrt{\text{Var}(z^j)\text{Var}(c^k)}} = \frac{\lambda_k v_{kj}}{\sqrt{(s^j)^2 \lambda_k}} = \sqrt{\lambda_k} \frac{v_{kj}}{s^j},$$

Si les variables sont réduites, alors $r(z^j, c^k) = \sqrt{\lambda_k} v_{kj}$.

Cette quantité donne la qualité de représentation de la variable z^j sur l'axe D_k . Plus elle est proche de 1 en valeur absolue, plus la variable est bien représentée sur l'axe D_k . Le signe de la corrélation permet de savoir si la variable contribue positivement ou négativement à la définition de cet axe.

On représente souvent graphiquement ces corrélations en dimension 2, à l'aide de ce que l'on appelle un **cercle des corrélations**.

Pour tout k variant entre 1 et p , on pose :

$$f^k = \frac{c^k}{\sqrt{n \lambda_k}}$$

On admet que, dans l'espace des variables $\mathbb{R}^n$, les composantes principales f^k forment une base orthonormée. La variable z^j s'y décompose selon la formule $z^j = \sum_{k=1}^p \text{Cor}(z^j, c^k) f^k$, avec $\text{Cor}(z^j, c^k) = \sqrt{\lambda_k} u_{kj}$. Puisque $\|z^j\|^2 = n$, les extrémités des vecteurs $\tilde{z}^j = \frac{1}{\|z^j\|} z^j$ se situent sur une p -sphère de rayon 1. Plus précisément, on choisit deux axes D_k et D_l , et on définit le cercle de corrélation associé comme l'intersection du plan factoriel engendré par D_k et D_l et la p -sphère de centre O^p et de rayon 1. On trace alors les points dont les coordonnées sont les corrélations de chacune des variables avec les k -ème et l -ème composantes principales. La qualité de représentation d'une variable z^j sur le k -ième axe factoriel est exprimée par le coefficient de corrélation linéaire $r(c^k, z^j)$

6.3.2 Sur un plan factoriel

Définition :

Le cercle de corrélation associé à un plan factoriel est l'intersection de ce plan avec la p -sphère de centre O^p et de rayon 1.

Plus la projection orthogonale d'une variable est proche du cercle de corrélation associé à un plan factoriel, meilleure est la qualité de représentation de cette variable sur ce plan factoriel.

Les figures qui suivent donnent l'exemple de deux variables, z^i et z^j , projetées sur le plan factoriel 1-2.

On observe que la qualité de représentation de la variable z^i est meilleure que celle de la variable z^j .

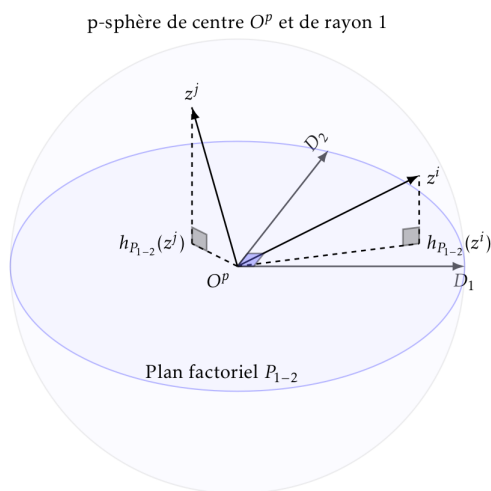


FIGURE 7 – p -sphère et cercle de corrélation dans le plan factoriel P_{k-l}

Cercle de corrélation de centre O^p et de rayon 1 dans le plan P_{1-2}

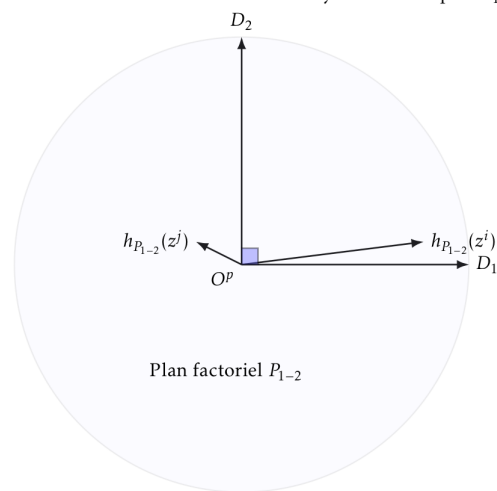


FIGURE 8 – Projections de deux variables z^i et z^j dans le cercle de corrélation associé au plan factoriel P_{k-l}

Conclusion :

Ainsi, bien qu'ayant retenu un sous-espace factoriel expliquant une part importante de l'inertie totale, il est possible que certaines variables ou individus soient mal représentés dans ce sous-espace.

Il sera alors intéressant de compléter le sous-espace factoriel en ajoutant des axes factoriels supplémentaires de sorte que ces variables ou individus soient alors bien ou mieux représentés.

6.4 Sélection du nombre k d'axes factoriels à retenir

La sélection du nombre d'axes à retenir, abordée précédemment, est une étape importante d'une **ACP**.

La **méthode du Coude** (ou *scree plot* en anglais) est une approche graphique utilisée pour déterminer le nombre optimal d'axes factoriels (ou composantes principales) à retenir dans une **Analyse en Composantes Principales**. Cette méthode repose sur l'analyse des valeurs propres associées à chaque composante principale.

Propriété :

- L'ACP transforme les variables initiales en un ensemble de **composantes principales**, classées par ordre décroissant de variance expliquée.
- Chaque composante principale est associée à une **valeur propre**, qui représente la quantité de variance expliquée par cette composante.
- La **méthode du Coude** consiste à tracer un graphique représentant les valeurs propres en fonction du numéro de la composante principale :
 - **Axe des abscisses** : Numéro de la composante principale (1, 2, 3, ...).
 - **Axe des ordonnées** : Valeur propre associée à chaque composante.

Voici un exemple de graphique affichant l'éboulis des valeurs propres :

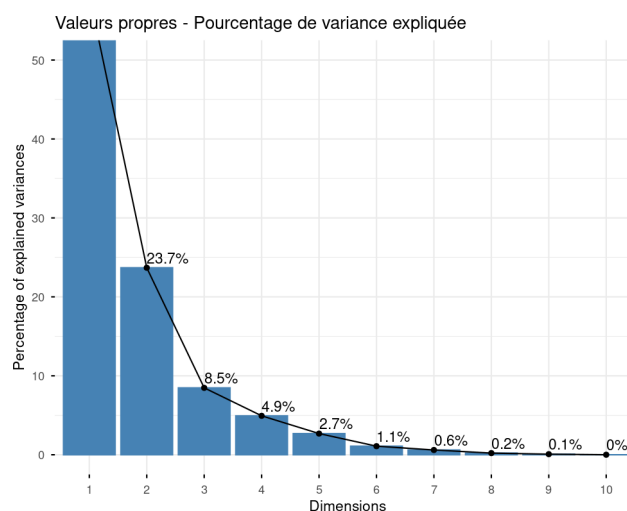


FIGURE 9 – Éboulis des valeurs propres

Interprétation du Graphique

- **Forme typique** : Les valeurs propres décroissent rapidement au début, puis plus lentement, formant une courbe qui ressemble à un « coude ».
- **Règle du Coude** : Le nombre optimal de composantes à retenir correspond au point où la courbe des valeurs propres commence à s'aplatir, c'est-à-dire juste avant que la décroissance ne devienne beaucoup moins forte.
- **Exemple** : Si le graphique montre une forte décroissance des valeurs propres jusqu'à la 3^e composante, puis une décroissance plus lente à partir de la 4^e, on retient généralement **3 composantes principales**.

Pourquoi utiliser la méthode du Coude ?

- **Équilibre** : Elle permet de trouver un compromis entre :
 - Retenir suffisamment de composantes pour expliquer une grande partie de la variance totale.
 - Éviter de retenir trop de composantes (ce qui compliquerait l'interprétation sans apporter d'information significative).

- **Simplicité** : C'est une méthode visuelle et intuitive, accessible même sans connaissances statistiques avancées.

Limites de la méthode

- **Subjectivité** : Le choix du « coude » peut parfois être ambigu, surtout si la courbe ne présente pas de point d'inflexion clair.
- **Complémentarité** : Elle est souvent utilisée en combinaison avec d'autres critères, comme le **pourcentage cumulé de variance expliquée** (par exemple, retenir les composantes qui expliquent au moins 80% de la variance totale).

7 Clustering par la méthode du k-means

7.1 Introduction

Le **k-means** est un algorithme d'apprentissage non supervisé utilisé pour le **clustering** (ou partitionnement de données).

Son objectif est de regrouper des données similaires en k clusters, où chaque cluster est représenté par un **centroïde**. Cet algorithme est largement utilisé en raison de sa simplicité et de son efficacité pour des jeux de données de taille moyenne.

Cette partie ne sera pas développée.

Les collègues qui désireront obtenir des explications pourront consulter

<https://www.lemagit.fr/conseil/Le-clustering-k-means-expliques-avec-des-exemples>

<https://www.ibm.com/fr-fr/think/topics/k-means-clustering>

<https://mrmint.fr/algorithmes-k-means>

<https://datascientest.com/algorithmes-des-k-means>

7.2 Mise en Place de l'Algorithme

L'algorithme k-means suit les étapes suivantes :

```

1  Algorithme Algorithme du k-means
2  Entrées :
3  Un ensemble de données  $X = \{x_1, x_2, \dots, x_n\}$ , un nombre de clusters  $k$ .
4  Étape 1 : Initialisation
5  Choisir aléatoirement  $k$  points parmi les données comme centroïdes initiaux
    $\mu_1, \mu_2, \dots, \mu_k$ .
6  Étape 2 : Affectation des points aux clusters
7  Pour chaque point  $x_i$  dans  $X$  Faire
8  |   Calculer la distance entre  $x_i$  et chaque centroïde  $\mu_j$  (pour  $j = 1$  à  $k$ ).
9  |   Affecter  $x_i$  au cluster  $C_j$  dont le centroïde  $\mu_j$  est le plus proche.
10 Fin du Pour
11 Étape 3 : Mise à jour des centroïdes
12 Pour chaque cluster  $C_j$  (pour  $j = 1$  à  $k$ ) Faire
13 |   Recalculer le centroïde  $\mu_j$  comme la moyenne des points affectés à  $C_j$ .
14 Fin du Pour
15 Étape 4 : Vérification de la convergence
16 Si les centroïdes n'ont pas changé et deviennent stables Alors
17 |   Arrêter l'algorithme.
18 Sinon
19 |   Retourner à l'Étape 2.
20 Fin du Si
21 Sorties :
22  $k$  clusters  $C_1, C_2, \dots, C_k$  et leurs centroïdes  $\mu_1, \mu_2, \dots, \mu_k$ .

```

7.3 Choix du Nombre de Clusters k

Le choix de k est crucial pour obtenir des résultats pertinents. Plusieurs méthodes existent pour déterminer k :

- **Méthode du Coude** : Tracer la courbe représentant la somme des carrés des distances intra-clusters en fonction de k et choisir le k où la courbe forme un « coude ».
- **Connaissances professionnelles** : Utiliser des informations spécifiques au domaine pour choisir k .

7.4 Avantage de l'Algorithme k-means

L'algorithme est facile à comprendre et à implémenter.

7.5 Limites de l'Algorithme k-means

- **Sensibilité à l'initialisation** : Les résultats peuvent varier en fonction des **centroïdes** initiaux.
- **Nombre de clusters k à définir** : Le choix de k peut être subjectif.
- **Sensibilité aux points aberrants** : Ces points, parfois appelés outliers, peuvent influencer les **centroïdes**.

8 ANNEXE 1 : Langage utilisé pour réaliser les différentes ACP présentées et sous quelles interfaces?

8.1 Langage

Les scripts permettant la réalisation des différentes **ACP** seront écrits en langage R.

Les stagiaires et lecteurs désirant découvrir, expérimenter et s'approprier l'outil, devront installer R sur leur machine en suivant l'une des méthodes décrites ci-dessous.

- Sous Linux, en ligne de commandes dans un Terminal :

Script Bash

```
# 1. Ajouter le dépôt CRAN
sudo apt install -y software-properties-common
sudo add-apt-repository 'deb https://cloud.r-project.org/bin/linux/ubuntu
  ~ focal-cran40/'
# Remplacez 'focal' par votre version d'Ubuntu si nécessaire

# 2. Ajouter la clé GPG
sudo apt-key adv --keyserver keyserver.ubuntu.com --recv-keys
  ~ E298A3A825C0D65DFD57CBB651716619E084DAB9

# 3. Mettre à jour et installer
sudo apt update
sudo apt install -y r-base r-base-dev
```

- Sous macOS, en ligne de commande :

Script Bash

```
brew install --cask r
```

- Sous MS Windows
 - Allez sur le site du CRAN
 - Cliquez sur « Download R for Windows »
 - « Download R-4.x.x for Windows » (x.x.x = version)
 - Exécuter le fichier *.exe

Pour toute information, voir le site CRAN ((The Comprehensive R Archive Network) (adresse : <https://cran.r-project.org/>)

8.2 Interfaces

Différentes interfaces permettant d'interpréter des scripts écrits en langage R sont mises à disposition de potentiels utilisateurs. Il est possible de travailler :

- en ligne de commande après avoir ouvert un Terminal :

Script Bash

```
(base) jm@jm-ThinkPad-T450s:~ $ R

R version 4.5.2 (2025-10-31) -- "[Not] Part in a Rumble"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.
```

R est un projet collaboratif avec de nombreux contributeurs.
Tapez `'contributors()'` pour plus d'information et
`'citation()'` pour la façon de le citer dans les publications.

Tapez `'demo()'` pour des démonstrations, `'help()'` pour l'aide
en ligne ou `'help.start()'` pour obtenir l'aide au format HTML.
Tapez `'q()'` pour quitter R.

[Sauvegarde de la session précédente restaurée]

>

- sous Rstudio.

Comment installer RStudio?

- **Téléchargement** : Aller sur <https://cran.rstudio.com/>
- **Choisir votre version** :
RStudio Desktop (gratuite) - pour la plupart des utilisateurs
- **Sélectionner votre système d'exploitation** :
Windows : .exe (double-cliquer sur le fichier téléchargé pour l'installer)
Mac : .dmg (glisser le fichier téléchargé dans dans « Applications »)
Linux : .deb (Ubuntu/Debian) ou .rpm (Fedora) (exécuter les commandes qui suivent dans un Terminal) :

Script Bash

```
(base) jm@jm-ThinkPad-T450s:~$  
# Pour .deb (Ubuntu/Debian)  
sudo dpkg -i rstudio-x.yy.zzz-amd64.deb  
sudo apt-get install -f # Pour les dépendances  
  
# Pour .rpm (Fedora/CentOS)  
sudo dnf install rstudio-x.yy.zzz-x86_64.rpm
```

9 ANNEXE 2 : Outils Mathématiques

9.1 Espaces Métriques

Définition :

Espace métrique Un espace métrique est un couple (X, d) où X est un ensemble et $d : X \times X \rightarrow \mathbb{R}^+$ est une application (distance) vérifiant :

1. pour tous $x \in X$ et $y \in X$, $d(x, y) = 0 \iff x = y$ (séparation)
2. pour tous $x \in X$ et $y \in X$, $d(x, y) = d(y, x)$ (symétrie)
3. pour tous $x \in X$, $y \in X$ et $z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$ (inégalité triangulaire)

Exemples :

1. Si $\forall (x, y) \in \mathbb{R}^2$, $d(x, y) = |x - y|$ alors $(\mathbb{R}, d)$ est un espace métrique.
2. Soit $n \in \mathbb{N}^*$. Si, pour tous $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ et $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$: $d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$, alors $(\mathbb{R}^n, d)$ est un espace métrique.
 d est alors appelée **distance euclidienne**.

9.2 Espaces Vectoriels

Définition :

Espace vectoriel

Un $\mathbb{R}$ -espace vectoriel est un ensemble E (on dit aussi que E est un espace vectoriel sur $\mathbb{R}$) muni de deux opérations :

- **une addition vectorielle**, notée $+$, application : $E \times E \rightarrow E$;
- **une multiplication scalaire**, notée $\cdot$, application : $\mathbb{R} \times E \rightarrow E$;

vérifiant les axiomes suivants :

- **Axiomes de l'addition vectorielle**

1. **Associativité** : $\forall (u, v, w) \in E^3$, $(u + v) + w = u + (v + w)$;
2. **Commutativité** : $\forall (u, v) \in E^2$, $u + v = v + u$;
3. **Élément neutre** : $\exists 0_E \in E$, $\forall u \in E$, $u + 0_E = u$;
4. **Opposé** : $\forall u \in E$, $\exists (-u) \in E$, $u + (-u) = 0_E$;

- **Axiomes de la multiplication scalaire**

1. **Distributivité à gauche** : $\forall \lambda, \mu \in \mathbb{R}$, $\forall u \in E$, $(\lambda + \mu) \cdot u = \lambda \cdot u + \mu \cdot u$;
2. **Distributivité à droite** : $\forall \lambda \in \mathbb{R}$, $\forall (u, v) \in E^2$, $\lambda \cdot (u + v) = \lambda \cdot u + \lambda \cdot v$;
3. **Compatibilité** : $\forall \lambda, \mu \in \mathbb{R}$, $\forall u \in E$, $(\lambda \mu) \cdot u = \lambda \cdot (\mu \cdot u)$;
4. **Élément neutre** : $\forall u \in E$, $1 \cdot u = u$.

Exemples :

- $\mathbb{R}^n$ est un espace vectoriel sur $\mathbb{R}$.
- $\mathcal{M}_{n,p}(\mathbb{R})$, ensemble des matrices de taille $n \times p$ à coefficients réels, est un espace vectoriel sur $\mathbb{R}$.

Définition :

Sous-espace vectoriel

Une partie F de E est un sous-espace vectoriel si :

1. $0_E \in F$;
2. $\forall (u, v) \in F^2, u + v \in F$ (stabilité par addition);
3. $\forall \lambda \in \mathbb{R}, \forall u \in F, \lambda \cdot u \in F$ (stabilité par multiplication scalaire).

Exemples :

- Droites et plans vectoriels dans $\mathbb{R}^3$
- Matrices symétriques dans $\mathcal{M}_n(\mathbb{R})$

9.3 Espaces Vectoriels Normés

Définition :

Espace vectoriel normé

Un espace vectoriel normé est un couple $(E, \|\cdot\|)$ où E est un espace vectoriel sur $\mathbb{R}$ et $\|\cdot\| : E \rightarrow \mathbb{R}^+$ est une application, appelée norme, vérifiant :

1. $\|x\| = 0 \iff x = 0$ (séparation)
2. $\forall \lambda \in \mathbb{R}, \forall x \in E, \|\lambda x\| = |\lambda| \|x\|$ (homogénéité)
3. $\forall (x, y) \in E^2, \|x + y\| \leq \|x\| + \|y\|$ (inégalité triangulaire)

Exemples :

Si $E = \mathbb{R}^n$, alors, si $\forall x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$,

1. $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ (norme euclidienne);
2. $\|x\|_1 = \sum_{i=1}^n |x_i|$ (norme 1);
3. $\|x\|_\infty = \max |x_i|$ (norme infinie);

alors $(E, \|\cdot\|_2)$, $(E, \|\cdot\|_1)$ et $(E, \|\cdot\|_\infty)$ sont des espaces vectoriels normés.

9.4 De la Norme à la Métrique

Propriété :

Tout espace vectoriel normé $(E, \|\cdot\|)$ est un espace métrique pour la distance d définie par :

$$\forall (x, y) \in E^2, d(x, y) = \|x - y\|$$

Justification :

1. $\forall (x, y) \in E^2, d(x, y) = 0 \iff \|x - y\| = 0 \iff x - y = 0 \iff x = y$.
2. $\forall (x, y) \in E^2, d(x, y) = \|x - y\| = \|(y - x)\| = |-1| \|y - x\| = d(y, x)$.
3. $\forall (x, y, z) \in E^3, d(x, z) = \|x - z\| = \|(x - y) + (y - z)\| \leq \|x - y\| + \|y - z\| = d(x, y) + d(y, z)$.

□

Propriétés :

Dans un espace vectoriel normé, la distance hérite de propriétés supplémentaires :

- Invariance par translation : $\forall (x, y, z) \in E^3, d(x+z, y+z) = d(x, y)$
- Homogénéité : $\forall (x, y) \in E^2, \forall \lambda \in \mathbb{R}, d(\lambda x, \lambda y) = |\lambda| d(x, y)$

Justification :

- $\forall (x, y, z) \in E^3, d(x+z, y+z) = \|(x+z) - (y+z)\| = \|x - y\| = d(x, y)$
- $\forall (x, y) \in E^2, \forall \lambda \in \mathbb{R}, d(\lambda x, \lambda y) = \|\lambda x - \lambda y\| = |\lambda| \|x - y\| = |\lambda| d(x, y)$

□

9.5 Espace vectoriel de dimension finie

AVERTISSEMENT

Ce qui suit sera sans doute sujet à de nombreuses critiques, son contenu ne faisant pas preuve d'une rigueur et d'une cohérence à toute épreuve.

Son objectif est simplement de fixer les notations qui seront utilisées par la suite.

Définition :

Soit E un espace vectoriel sur $\mathbb{R}$. E est de dimension finie s'il existe une famille $\mathcal{B} = (e_1, \dots, e_n)$ de vecteurs de E telle que :

1. $\mathcal{B}$ est une famille libre : $\forall (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n,$

$$\sum_{i=1}^n \lambda_i e_i = 0 \implies \forall i \in \{1, \dots, n\}, \lambda_i = 0$$

2. $\mathcal{B}$ est une famille génératrice : $\forall x \in E, \exists (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n, x = \sum_{i=1}^n \lambda_i e_i.$

Dans ce cas, on dit que la famille $\mathcal{B}$ est une base de E .

Dans ce qui suit, E est un espace vectoriel sur $\mathbb{R}$, de dimension finie.

Propriété :

Admise

Toutes les bases de E ont le même nombre d'éléments, appelé dimension de E .

Exemples :

Base canonique de $\mathbb{R}^n$

La base canonique de $\mathbb{R}^n$ est $\mathcal{B}_c = (e_1, e_2, \dots, e_n)$ où :

$$e_1 = (1, 0, 0, \dots, 0), e_2 = (0, 1, 0, \dots, 0), \dots, e_n = (0, 0, \dots, 0, 1)$$

Tout vecteur $x = (x_1, x_2, \dots, x_n)$ s'écrit de façon unique : $x = x_1 e_1 + x_2 e_2 + \dots + x_n e_n.$

Définition :

Coordonnées dans une base

Si $\mathcal{B} = (e_1, \dots, e_n)$ est une base de E , alors pour tout $x \in E$, il existe un unique n -uplet $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ tel que :

$$x = \sum_{i=1}^n \lambda_i e_i$$

Les scalaires $\lambda_1, \dots, \lambda_n$ sont appelés les **coordonnées** de x dans la base $\mathcal{B}$.

9.6 Matrices de Passage et Changement de Base

Définition :

Matrice de passage

Soient $\mathcal{B} = (e_1, \dots, e_n)$ et $\mathcal{B}' = (e'_1, \dots, e'_n)$ deux bases de E . La matrice de passage de la base $\mathcal{B}$ à la base $\mathcal{B}'$ est la matrice carrée d'ordre n , notée $P_{\mathcal{B}-\mathcal{B}'} = (p_i^j)$ définie par :

$$\forall j \in \{1, \dots, n\}, e'_j = \sum_{i=1}^n p_i^j e_i$$

Propriétés :

Soient $\mathcal{B}, \mathcal{B}', \mathcal{B}''$ trois bases de E , alors :

1. $P_{\mathcal{B}-\mathcal{B}'}$ est inversible.
2. $P_{\mathcal{B}-\mathcal{B}'} = P_{\mathcal{B}'-\mathcal{B}}^{-1}$.
3. $P_{\mathcal{B}-\mathcal{B}''} = P_{\mathcal{B}-\mathcal{B}'} \cdot P_{\mathcal{B}'-\mathcal{B}''}$.

Propriété :

Soit $x \in E$ de coordonnées $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ dans la base $\mathcal{B}$ et $X' = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}$ dans la base $\mathcal{B}'$, alors :

$$X' = P_{\mathcal{B}'-\mathcal{B}} \cdot X \text{ et } X = P_{\mathcal{B}-\mathcal{B}'} \cdot X'$$

Justification :

On a $x = \sum_{i=1}^n x_i e_i = \sum_{j=1}^n x'_j e'_j = \sum_{j=1}^n x'_j \left(\sum_{i=1}^n p_i^j e_i \right) = \sum_{i=1}^n \left(\sum_{j=1}^n p_i^j x'_j \right) e_i$, donc $x_i = \sum_{j=1}^n p_i^j x'_j$. □

Exemple :

$E = \mathbb{R}^2$. $\mathcal{B}$ est la base canonique de E . Les vecteurs $e'_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{2}{\sqrt{2}} \end{pmatrix}$ et $e'_2 = \begin{pmatrix} \frac{-\sqrt{2}}{2} \\ \frac{2}{\sqrt{2}} \end{pmatrix}$ forment une base $\mathcal{B}'$.

La matrice de passage de la base $\mathcal{B}$ à la base $\mathcal{B}'$ est

$$P_{\mathcal{B}-\mathcal{B}'} = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{2}{\sqrt{2}} & \frac{2}{\sqrt{2}} \end{pmatrix}$$

On montre facilement que

$$P_{\mathcal{B}'-\mathcal{B}} = P_{\mathcal{B}-\mathcal{B}'}^{-1} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{2}{\sqrt{2}} & \frac{2}{\sqrt{2}} \end{pmatrix}$$

Soit u le vecteur de coordonnées $X = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ dans la base $\mathcal{B}$.

Les coordonnées X' du vecteur u dans la base $\mathcal{B}'$ sont alors données par le produit matriciel

$$X' = P_{\mathcal{B}'-\mathcal{B}} X = P_{\mathcal{B}'-\mathcal{B}}^{-1} X$$

Par conséquent le vecteur u a pour coordonnées

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

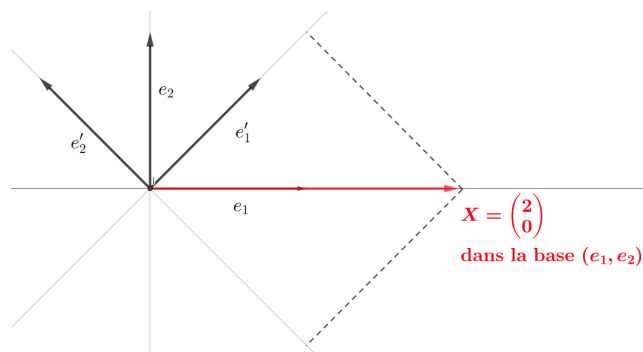


FIGURE 10 – Tentative d'explication graphique du changement de bases.

9.7 Produit Scalaire et Espaces Euclidiens

Définition :

Produit scalaire

Soit E un espace vectoriel sur $\mathbb{R}$. Un produit scalaire est une forme bilinéaire symétrique définie positive sur $E \times E$, autrement dit, une application, notée par la suite $\langle \cdot, \cdot \rangle : E \times E \rightarrow \mathbb{R}$ vérifiant :

1. **Symétrie** : $\forall (x, y) \in E^2, \langle x, y \rangle = \langle y, x \rangle$.
2. **Bilinéarité** :
 - Linéaire à gauche : $\forall (x, y, z) \in E^3, \forall (\lambda, \mu) \in \mathbb{R}^2, \langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$.
 - Linéaire à droite : $\forall (x, y, z) \in E^3, \forall (\lambda, \mu) \in \mathbb{R}^2, \langle x, \lambda y + \mu z \rangle = \lambda \langle x, y \rangle + \mu \langle x, z \rangle$.
3. **Définie positive** : $\forall x \in E, \langle x, x \rangle \geq 0$.
4. **Séparation** : $\langle x, x \rangle = 0 \iff x = 0_E$.

Exemples :

Produit scalaire canonique sur $\mathbb{R}^n$

Pour tous $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$ appartenant à $\mathbb{R}^n$:

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$$

Définition :

Espace euclidien

E est un espace euclidien si E est un espace vectoriel de dimension finie muni d'un produit scalaire.
La norme associée est, $\forall x \in E, \|x\| = \sqrt{\langle x, x \rangle}$.

Propriétés :

1. $\forall x \in E, \|x\| = 0 \iff x = 0$.
2. $\forall x \in E, \forall \lambda \in \mathbb{R}, \|\lambda x\| = |\lambda| \|x\|$.
3. $\|x + y\| \leq \|x\| + \|y\|$ (inégalité triangulaire).

Définition :

Base orthonormée

Une base $\mathcal{B} = (e_1, \dots, e_n)$ de E est orthonormée si :

$$\forall i \in \{1, \dots, n\}, \forall j \in \{1, \dots, n\}, \langle e_i, e_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

9.8 Orthogonalité

Définition :

Vecteurs orthogonaux

Deux vecteurs x et y de l'espace euclidien E sont orthogonaux si $\langle x, y \rangle = 0$. On note $x \perp y$.

Propriétés :

Expression du produit scalaire dans une base orthonormée

Si $\mathcal{B} = (e_1, \dots, e_n)$ est une base orthonormée de l'espace euclidien E, alors pour tout $x = \sum_{i=1}^n x_i e_i \in E$:

- $\forall i \in \{1, \dots, n\}, x_i = \langle x, e_i \rangle$ (coordonnées par projection) ;
- $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$;
- $\|x\|^2 = \sum_{i=1}^n x_i^2$ (théorème de Pythagore).

Justification :

Pour tout $j \in \{1, \dots, n\}$,

$$\langle x, e_j \rangle = \left\langle \sum_{i=1}^n x_i e_i, e_j \right\rangle = \sum_{i=1}^n x_i \langle e_i, e_j \rangle = \sum_{i=1}^n x_i \delta_{ij} = x_j$$

□

9.9 Interprétation géométrique de l'orthogonalité

Définition :

Angle entre deux vecteurs

Dans un espace euclidien E , on définit l'angle θ entre deux vecteurs x et y non nuls par :

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|}$$

Cette définition est justifiée par l'inégalité de Cauchy-Schwarz.

Définition :

Projection orthogonale sur une droite vectorielle

Soit E un espace euclidien, u un vecteur unitaire ($\|u\| = 1$), et Δ_u la droite engendrée par u . La projection orthogonale sur Δ_u est l'application $p_u : E \rightarrow \Delta_u$ qui à tout $x \in E$ associe l'unique vecteur $p_u(x) \in \Delta_u$ tel que $x - p_u(x) \perp \Delta_u$.

La projection orthogonale d'un vecteur x sur un vecteur unitaire u ($\|u\| = 1$), notée $p_u(x)$, est :

$$p_u(x) = \langle x, u \rangle u$$

Justification :

- **Condition d'orthogonalité** : On cherche λ tel que $x - \lambda u \perp u$
- **Calcul** : $\langle x - \lambda u, u \rangle = 0 \Rightarrow \langle x, u \rangle - \lambda = 0 \Rightarrow \lambda = \langle x, u \rangle$
- **Vérification** : Avec $\lambda = \langle x, u \rangle$, on a bien $x - \langle x, u \rangle u \perp u$

□

Propriété :

Théorème de Pythagore

E est un espace euclidien. Soient x et y deux vecteurs de E .

Si $x \perp y$, alors $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Justification :

Si $x \perp y$, alors $\langle x, y \rangle = 0$, d'où

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + 2\langle x, y \rangle + \langle y, y \rangle = \|x\|^2 + \|y\|^2$$

□

9.10 Matrices Associées aux Produits Scalaires

Définition :

Matrice de Gram

Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base de E . La matrice de Gram du produit scalaire dans la base $\mathcal{B}$ est la matrice, carrée d'ordre n , $G = (g_i^j)$, où :

$$\forall i \in \{1, \dots, n\}, \forall j \in \{1, \dots, n\}, g_i^j = \langle e_i, e_j \rangle$$

Propriété :

Expression du produit scalaire

Si $x = \sum_{i=1}^n x_i e_i$ et $y = \sum_{j=1}^n y_j e_j$, alors :

$$\langle x, y \rangle = \sum_{i=1}^n \sum_{j=1}^n g_i^j x_i y_j = X^t G Y$$

où X, Y sont les matrices colonnes associées aux vecteurs x et y dans la base $\mathcal{B}$ et X^t la matrice transposée de X .

Justification :



Par bilinéarité : $\langle x, y \rangle = \langle \sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j \rangle = \sum_{i=1}^n \sum_{j=1}^n x_i y_j \langle e_i, e_j \rangle$.

□

Propriétés :

Propriétés de la matrice de Gram

1. G est symétrique : $G^t = G$.
2. G est une matrice définie positive : $X^T G X = \langle x, x \rangle > 0$ pour $x \neq 0$, donc $X^t G X > 0$ pour tout $X \neq 0$.
3. $\det G > 0$.

Propriété :

Matrice de Gram dans une base orthonormée

Si E est un espace euclidien de dimension n et $\mathcal{B}$ une base orthonormée, alors $G = I_n$ et :

$$\forall (x, y) \in E, \langle x, y \rangle = \sum_{i=1}^n x_i y_i = X^t Y$$

Justification :



Immédiat puisque $\forall i \in \{1, \dots, n\} \forall j \in \{1, \dots, n\} \langle e_i, e_j \rangle = \delta_{ij}$.

□

9.11 Changement de Base et Matrices de Passage

Propriété :

Soient u et v deux vecteurs de E , de coordonnées respectives X et Y dans la base $\mathcal{B}$, et X' et Y' dans $\mathcal{B}'$.

On a $X = P X'$ et $Y = P Y'$, alors :

$$\langle x, y \rangle = X^t G Y = (P X')^t G (P Y') = X'^t (P^t G P) Y'$$

Conséquences :

- Dans une base orthonormée pour le produit scalaire, la matrice de Gram est l'identité.
- Si $\mathcal{B}$ est orthonormée, $G = I_n$. Alors pour toute autre base $\mathcal{B}'$: $G' = P^T P$.

9.12 Supplémentaire orthogonal d'un sous-espace vectoriel

9.12.1 Définition et existence

Définition :

Orthogonal d'un sous-ensemble

Soit E un espace euclidien et $A \subset E$ une partie non vide. L'orthogonal de A est l'ensemble, noté $A^\perp$ défini par :

$$A^\perp = \{x \in E \mid \forall a \in A, \langle x, a \rangle = 0\}$$

Définition :

Supplémentaire orthogonal

Soit F un sous-espace vectoriel de E . Le supplémentaire orthogonal de F est l'orthogonal $F^\perp$.

Propriétés :

Structure de l'orthogonal

Pour tout sous-espace vectoriel F de E :

1. $F^\perp$ est un sous-espace vectoriel de E .
2. $F \cap F^\perp = \{0\}$.
3. $F \subset (F^\perp)^\perp$.

Justification :

1. Soient $(x, y) \in F^\perp \times F^\perp$, $\lambda \in \mathbb{R}$. Pour tout $z \in F$:

$$\langle x + \lambda y, z \rangle = \langle x, z \rangle + \lambda \langle y, z \rangle = 0 + 0 = 0$$

Donc $x + \lambda y \in F^\perp$.

2. Si $x \in F \cap F^\perp$, alors $x \perp x$ donc $\langle x, x \rangle = 0$ d'où $x = 0$.
3. Si $x \in F$, alors pour tout $y \in F^\perp$, $\langle x, y \rangle = 0$ donc $x \in (F^\perp)^\perp$.

□

Propriétés :

Soit E un espace euclidien de dimension finie et F un sous-espace vectoriel de E , alors :

$$E = F \oplus F^\perp \text{ (signifie que } E \text{ est la somme directe de } F \text{ et de } F^\perp)$$

De plus, on a les propriétés :

1. $\dim F + \dim F^\perp = \dim E$.
2. $(F^\perp)^\perp = F$.
3. La projection orthogonale sur F est bien définie.

Justification :

Soit $(e_1, \dots, e_k)$ une base orthonormée de F . On la complète (voir le théorème de la base incomplète) en une base orthonormée $(e_1, \dots, e_k, e_{k+1}, \dots, e_n)$ de E .

Preuve que $E = F \oplus F^\perp$:

$$\text{— Tout } x \in E \text{ s'écrit : } x = \underbrace{\sum_{i=1}^k \langle x, e_i \rangle e_i}_{\in F} + \underbrace{\sum_{i=k+1}^n \langle x, e_i \rangle e_i}_{\in F^\perp}$$

— La somme est directe car $F \cap F^\perp = \{0\}$

Preuve de $(F^\perp)^\perp = F$:

On a déjà $F \subset (F^\perp)^\perp$. Montrons l'inclusion inverse :

— Soit $x \in (F^\perp)^\perp$. D'après la décomposition précédente :

$$x = x_F + x_{F^\perp} \quad \text{avec } x_F \in F, x_{F^\perp} \in F^\perp$$

— Comme $x \in (F^\perp)^\perp$ et $x_{F^\perp} \in F^\perp$, on a $\langle x, x_{F^\perp} \rangle = 0$.

— Mais $\langle x, x_{F^\perp} \rangle = \langle x_F, x_{F^\perp} \rangle + \langle x_{F^\perp}, x_{F^\perp} \rangle = 0 + \|x_{F^\perp}\|^2$.

— Donc $\|x_{F^\perp}\|^2 = 0$ et $x_{F^\perp} = 0$, d'où $x = x_F$, $x_F \in F$.

□

9.12.2 Projection orthogonale

Définition :

Projection orthogonale

Soit F un sous-espace vectoriel de E . La projection orthogonale sur F est l'application linéaire $p_F : E \rightarrow F$ qui à $x \in E$ associe l'unique vecteur $p_F(x) \in F$ tel que $x - p_F(x) \in F^\perp$.

Propriété :

Caractérisation de la projection

Soit $(e_1, \dots, e_k)$ une base orthonormée de F . Alors :

$$p_F(x) = \sum_{i=1}^k \langle x, e_i \rangle e_i$$

De plus, $p_F(x)$ est l'unique vecteur de F tel que :

$$\|x - p_F(x)\| = \min_{y \in F} \|x - y\|$$

Justification :

Expression de la projection :

Soit $x \in E$. On écrit $x = x_F + x_{F^\perp}$ avec $x_F \in F$, $x_{F^\perp} \in F^\perp$. Alors :

$$\langle x, e_i \rangle = \langle x_F, e_i \rangle + \langle x_{F^\perp}, e_i \rangle = \langle x_F, e_i \rangle + 0$$

$$\text{Donc } x_F = \sum_{i=1}^k \langle x, e_i \rangle e_i.$$

Propriété de minimisation :

Soit $y \in F$. Alors :

$$\|x - y\|^2 = \|(x - p_F(x)) + (p_F(x) - y)\|^2$$

Comme $x - p_F(x) \in F^\perp$ et $p_F(x) - y \in F$, le théorème de Pythagore donne :

$$\|x - y\|^2 = \|x - p_F(x)\|^2 + \|p_F(x) - y\|^2 \geq \|x - p_F(x)\|^2$$

avec égalité si et seulement si $y = p_F(x)$.

□

9.13 Diagonalisation

9.13.1 Lien Endomorphismes–Matrices

⚡ AVERTISSEMENT

Soit E un $\mathbb{R}$ -espace vectoriel de dimension n , de base $\mathcal{B} = (e_1, e_2, \dots, e_n)$, et $f : E \rightarrow E$ un endomorphisme (application linéaire de E dans E).

Un scalaire $\lambda \in \mathbb{R}$ est une **valeur propre** de f s'il existe un vecteur $x \neq 0$ de E tel que :

$$f(x) = \lambda x$$

Le vecteur x est appelé **vecteur propre** associé à λ .

La **matrice de f dans la base $\mathcal{B}$** , notée $M_{\mathcal{B}}(f)$, est la matrice $(a_i^j) \in \mathcal{M}_n(\mathbb{R})$

$$M_{\mathcal{B}}(f) = \begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^n \\ a_2^1 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^1 & a_n^2 & \cdots & a_n^n \end{pmatrix}$$

dont la j -ème colonne est formée des coordonnées de $f(e_j)$ dans la base $\mathcal{B}$:

$$f(e_j) = \sum_{i=1}^n a_i^j e_i \quad \text{pour } j = 1, \dots, n$$

Autrement dit, a_i^j est la i -ème coordonnée de $f(e_j)$.

Plus explicitement :

$$M_{\mathcal{B}}(f) = \begin{array}{c|cccc} & f(e_1) & f(e_2) & \cdots & f(e_n) \\ \hline e_1 & a_1^1 & a_1^2 & \cdots & a_1^n \\ e_2 & a_2^1 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_n & a_n^1 & a_n^2 & \cdots & a_n^n \end{array}$$

Dans ce qui suit, on ne s'intéressera qu'à la diagonalisation des matrices carrées d'ordre n , sans se préoccuper du lien existentiel avec la diagonalisation des endomorphismes sur un espace vectoriel de dimension n .

📖 Définition :

Valeur propre et vecteur propre d'une matrice carrée

Soit E un $\mathbb{R}$ -espace vectoriel et A une matrice carrée d'ordre n , $A \in \mathcal{M}_n(\mathbb{R})$. Un scalaire $\lambda \in \mathbb{R}$ est une **valeur propre** de A s'il existe un vecteur $x \neq 0$ de E tel que :

$$Ax = \lambda x,$$

Le vecteur x est appelé **vecteur propre** associé à λ .

Le sous-espace propre associé à la valeur propre λ est :

$$E_{\lambda} = \{x \in E \mid Ax = \lambda x\},$$

📖 Définition :

Matrice diagonalisable

Une matrice $A \in \mathcal{M}_n(\mathbb{R})$ est **diagonalisable** s'il existe une matrice inversible P et une matrice diagonale D telles que :

$$A = PDP^{-1},$$

9.13.2 Diagonalisation des matrices symétriques

Définition :

A est une matrice carrée d'ordre n , $A \in \mathcal{M}_n(\mathbb{R})$.

A est symétrique si

$$A^t = A$$

Définition :

A est une matrice carrée d'ordre n , $A \in \mathcal{M}_n(\mathbb{R})$.

A est orthogonale si

$$A^t A = I_n,$$

dans ce cas $A^{-1} = A^t$.

Propriété :

Théorème spectral

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

1. Les valeurs propres de A sont réelles.
2. A est diagonalisable dans une base orthonormée.
3. Les sous-espaces propres sont orthogonaux deux à deux.

Plus précisément, il existe une matrice orthogonale P ($P^t P = I_n$) et une matrice diagonale D telles que :

$$A = P D P^t$$

Justification :

On admet la propriété 1. et la propriété 2. se démontre par récurrence.

Démonstration de l'orthogonalité des sous-espaces propres

Soient $\lambda \neq \mu$ deux valeurs propres distinctes de A. Alors $E_\lambda \perp E_\mu$.

Soient $X \in E_\lambda$, $Y \in E_\mu$. Alors :

$$\lambda \langle X, Y \rangle = \langle \lambda X, Y \rangle = \langle AX, Y \rangle = X^t A Y = X^t (\mu Y) = \mu \langle X, Y \rangle$$

Donc $(\lambda - \mu) \langle X, Y \rangle = 0$ et comme $\lambda \neq \mu$, on a $\langle X, Y \rangle = 0$.

□