Analyse en Composantes Principales (ACP)

Serge Mazauric

CPE – Lyon

4ETI

2021 – 2022

Sommaire

1. Introduction

1. Introduction

- L'analyse en composantes principales est une méthode qui permet transformer un tableau de données de façon à fournir une représentation graphique simple dans le but de faciliter l'interprétation des données.

- Plus précisément, l'ACP transforme un certain nombre de variables statistiques liées entre-elles (corrélées) en un nombre réduit de variables non corrélées mais contenant tout de même les informations les plus significatives présentes dans les variables initiales : ces nouvelles variables sont appelées « **composantes principales** ».

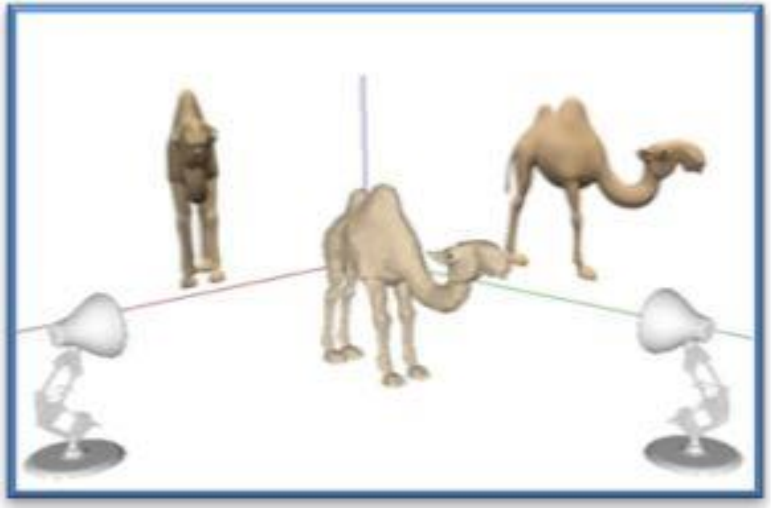


Figure 1 : Selon le plan sur lequel on projette, l'image projetée contient des informations plus ou moins significatives sur « l'objet » projeté (source : http://www.deenov.com/analyse-de-donnees/e-learning.aspx)

1. Introduction

Exemple

Le tableau ci-dessous représente les notes obtenues par 9 étudiants dans 5 matières différentes

	Electronique	Semi-conducteur	Analyse	Python	Traitement du signal
Théo	6	6	8	5.5	8
Tom	8	8	8	8	9
Paul	6	7	11	9.5	11
Léa	14.5	14.5	15.5	15	8
Jules	14	14	12	12	10
Emilie	11	10	5.5	7	13
Eva	5.5	7	14	11.5	10
Hugo	13	12	8.5	9.5	12
Lydie	9	9.5	12.5	12	18

- $m = 5$ variables statistiques contenant chacune $n = 9$ données.
- Le nuage de points associé à ce tableau contient donc 9 points dans un espace à 5 dimension !
- On ne peut donc pas visualiser graphiquement un tel nuage de points. Le tableau est alors difficilement interprétable : on ne peut pas se rendre compte si 2 points du nuage sont proches ou au contraire éloignés.

2. Introduction

Exemple (suite)

- Toutefois, si on ne retient que les 2 premières composantes principales (nous verrons comment plus tard), c'est-à-dire celles qui contiennent la quasi-totalité des informations du tableau, alors il est possible de les représenter par un nuage de points dans un espace à 2 dimensions.

- Tout se passe comme si l'on avait fait une **projection** des données d'un espace de dimension 5 dans un espace de dimension 2 : on parle alors de **réduction de dimensionnalité**.

- On passe donc d'une base de dimension 5 à une nouvelle base de dimension 2 constitués de 2 nouveaux axes appelés **axes factoriels** : les variables associées aux deux nouveaux axes sont dé-corrélées.

- Cette projection est faite de façon à ce que la **variance** des données projetées soit **maximale**.

Exemple (suite et fin)

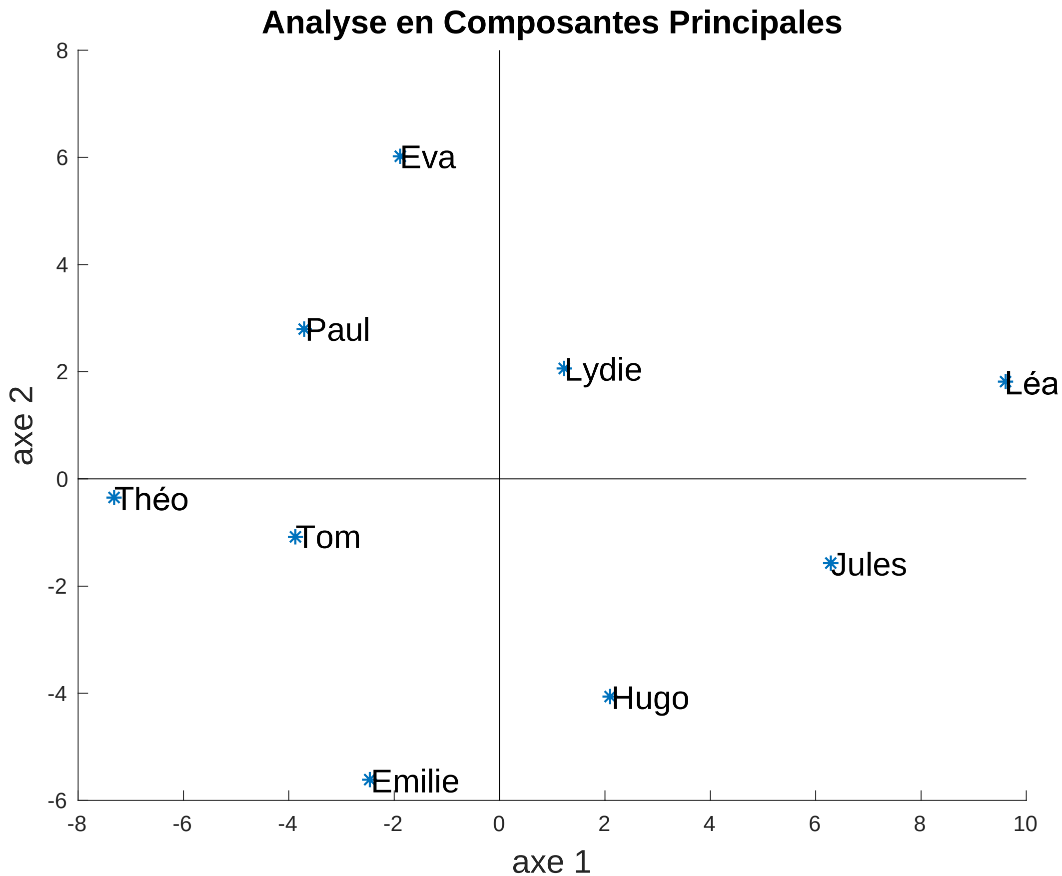


Figure 2 : Nuage de points après réduction de dimensionnalité par ACP, c'est-à-dire après projection du nuage initial (celui dans l'espace de dimension 5) sur les 2 premières composantes principales.

2. Rappels d’algèbre linéaire (3ETI)

2. Rappels d'algèbre linéaire (3ETI)

2.1. Produit scalaire

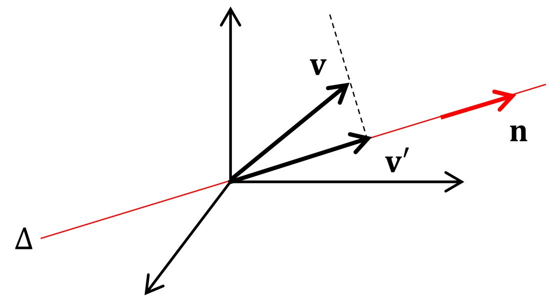
Le produit scalaire de $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ par $\mathbf{y} = (y_1, y_2, \dots, y_n)^T \in \mathbb{R}^n$ s'écrit : $\boxed{\mathbf{x}^T\mathbf{y}}$

En effet :

$$\mathbf{x}^T\mathbf{y} = (x_1, x_2, \dots, x_n)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \sum_{i=1}^{n} x_iy_i$$

En particulier : $\boxed{\mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2}$

2.2 Projection orthogonale sur une droite



On considère :

- une droite Δ dans $\mathbb{R}^3$
- un vecteur directeur normé de Δ : $\mathbf{n} = (a, b, c)^T$

  avec $\|\mathbf{n}\|^2 = \mathbf{n}^T\mathbf{n} = a^2 + b^2 + c^2 = 1$
- un vecteur $\mathbf{v}$ de $\mathbb{R}^3$

La projection orthogonale de V sur la droite Δ est le vecteur : $\boxed{\mathbf{v}' = \mathbf{n}\mathbf{n}^T\mathbf{v}}$

avec :

$$\mathbf{n}\mathbf{n}^T = \begin{pmatrix} a \\ b \\ c \end{pmatrix} (a, b, c) = \begin{pmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{pmatrix}$$

matrice de projection

Démonstration

a) Montrons que $\mathbf{v}'$ et $\mathbf{n}$ sont colinéaires : $\mathbf{v}' = \mathbf{n}(\mathbf{n}^T\mathbf{v}) = \mathbf{n}\,\underbrace{(a, b, c)\begin{pmatrix} x \\ y \\ z \end{pmatrix}}_{\text{scalaire}} = (ax + by + cz)\mathbf{n}$

b) Montrons que $\mathbf{v}' - \mathbf{v}$ et $\mathbf{n}$ sont orthogonaux, pour cela calculons le produit scalaire de $\mathbf{n}$ par $\mathbf{v}' - \mathbf{v}$

$$\mathbf{n}^T(\mathbf{v}' - \mathbf{v}) = \mathbf{n}^T\mathbf{v}' - \mathbf{n}^T\mathbf{v} = \underbrace{\mathbf{n}^T\mathbf{n}}_{1}\,\mathbf{n}^T\mathbf{v} - \mathbf{n}^T\mathbf{v} = 0$$

2.3. Forme quadratique définie par une matrice symétrique

Soit $\mathbf{A}$ une matrice symétrique de dimension n et $\mathbf{x} \in \mathbb{R}^n$ alors $\mathbf{x}^T\mathbf{A}\mathbf{x}$ est une forme quadratique.

Exemple : $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$ et $\mathbf{x} = (x, y)^T$

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = (x, y)\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = (x + 2y, 2x + 3y)\begin{pmatrix} x \\ y \end{pmatrix} = x^2 + 4xy + 3y^2$$

On note :

$$f(x, y) = f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = x^2 + 4xy + 3y^2$$

2.4. Gradient d'une forme quadratique définie par une matrice symétrique

Le gradient d'une forme quadratique $f(\mathbf{x})$ avec $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ est le vecteur de $\mathbb{R}^n$ dont la $i^{\text{ème}}$ composante est la dérivée partielle de f par rapport à x_i :

$$\overrightarrow{\mathrm{grad}} f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right)^T$$

Si $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ où $\mathbf{A}$ est une matrice symétrique alors :

$$\boxed{\overrightarrow{\mathrm{grad}}(\mathbf{x}^T \mathbf{A} \mathbf{x}) = 2\mathbf{A}\mathbf{x}}$$

Reprenons l'exemple précédent : $f(x, y) = x^2 + 4xy + 3y^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$ avec $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$

$$\overrightarrow{\mathrm{grad}} f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x + 4y \\ 4x + 6y \end{pmatrix} = 2 \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2\mathbf{A}\mathbf{x}$$

Cas particulier : $\mathbf{A} = \mathbf{I}$

$$\boxed{\overrightarrow{\mathrm{grad}}(\mathbf{x}^T \mathbf{x}) = \overrightarrow{\mathrm{grad}}(\|\mathbf{x}\|^2) = 2\mathbf{x}}$$

3. Description des données

3. Description des données

3.1. Matrice des données

On considère m variables statistiques notées $\mathbf{x}_{\bullet 1}, \mathbf{x}_{\bullet 2}, \dots \mathbf{x}_{\bullet m}$ contenant chacune n données, les variables sont alors représentées par des vecteurs colonnes à n composantes, c'est-à-dire :

$$\forall j \in [\![1, m]\!], \qquad \mathbf{x}_{\bullet j} = \left(x_{1j}, x_{2j}, \dots, x_{nj}\right)^T \qquad (1)$$

Lorsqu'on rassemble côte à côte ces m vecteurs colonnes on obtient la **matrice des données** :

Nuage de points (individus)

$$\begin{array}{cc}
\begin{array}{c} \mathbf{x}_{1\bullet} \longrightarrow \\ \mathbf{x}_{2\bullet} \longrightarrow \\ \\ \mathbf{x}_{n\bullet} \longrightarrow \end{array} &
\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \qquad (2) \\
& \begin{array}{cccc} \uparrow & \uparrow & & \uparrow \\ \mathbf{x}_{\bullet 1} & \mathbf{x}_{\bullet 2} & & \mathbf{x}_{\bullet m} \end{array}
\end{array}$$

Variables (attributs)

Les lignes de cette matrice sont des points dans un espace à m dimensions (on parle alors de nuage de points) ; on dit qu'on a n **individus** décrits par m **attributs**. Ces points sont représentés pas un vecteur notés $\mathbf{x}_{i\bullet}$:

$$\forall i \in [\![1, n]\!], \qquad \mathbf{x}_{i\bullet} = (x_{i1}, x_{i2}, \dots, x_{im}) \qquad (3)$$

3. Description des données

On retiendra :

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

matrice des données

Variable (ou attribut) = une colonne de la matrice des données

$$\mathbf{x}_{\bullet j} = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}$$

Individu (ou point) = une ligne de la matrice des données

$$\mathbf{x}_{i\bullet} = (x_{i1}, x_{i2}, \dots, x_{im})$$

3. Description des données

Moyenne d'une variable $\mathbf{x}_{\bullet j}$:

$$\forall j \in [\![1, m]\!], \qquad \bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \qquad (4)$$

Le **centre de gravité** (on dit aussi **Individu moyen**) du nuage de points est :

$$\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m)^T \qquad (5)$$

Variance d'une variable $x_{\bullet j}$:

$$\forall j \in [\![1, m]\!], \qquad \mathrm{var}(\mathbf{x}_{\bullet j}) = \sigma_j^2 = \frac{1}{n}\sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right)^2 \qquad (6)$$

On rappelle que plus la variance est grande (resp. faible), plus les valeurs de la variable sont dispersées (resp. rassemblées) autour de la moyenne.

3. Description des données

Covariance de deux variables $\mathbf{x}_{\bullet j}$ et $\mathbf{x}_{\bullet k}$:

$$\operatorname{cov}\left(\mathbf{x}_{\bullet j}, \mathbf{x}_{\bullet k}\right) = \sigma_{jk} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)\left(x_{ik} - \bar{x}_k\right) \qquad (7)$$

On rappelle que si la covariance est grande (en valeur absolue) alors les variables sont dépendantes (corrélées) et que si la covariance est nulle alors les variables sont indépendantes (dé-corrélées).

3. Description des données

3.2. Matrice des données centrées

Afin de centrer le nuage de point sur l'origine, on soustrait à chaque élément x_{ij} de la matrice (2) la moyenne $\bar{x}_j$, on obtient alors la matrice des données centrées :

$$\mathbf{X} = \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1m} - \bar{x}_m \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2m} - \bar{x}_m \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \dots & x_{nm} - \bar{x}_m \end{pmatrix} \quad (8)$$

matrice des données centrées

3.3. Matrice de covariance

En multipliant à gauche la matrice $\mathbf{X}$ par sa transposée, on obtient la matrice $(m \times m)$ suivante :

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_1 & \dots & x_{n1}-\bar{x}_1 \\ x_{12}-\bar{x}_2 & x_{22}-\bar{x}_2 & \dots & x_{n2}-\bar{x}_2 \\ \vdots & \vdots & & \vdots \\ x_{1m}-\bar{x}_m & x_{2m}-\bar{x}_m & \dots & x_{nm}-\bar{x}_m \end{pmatrix} \begin{pmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \dots & x_{1m}-\bar{x}_m \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \dots & x_{2m}-\bar{x}_m \\ \vdots & \vdots & & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \dots & x_{nm}-\bar{x}_m \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2 & \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2) & \dots & \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{im}-\bar{x}_m) \\ \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2) & \sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2 & \dots & \sum_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{im}-\bar{x}_m) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{im}-\bar{x}_m) & \sum_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{im}-\bar{x}_m) & \dots & \sum_{i=1}^{n}(x_{im}-\bar{x}_m)^2 \end{pmatrix}$$

$$= n \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1m} & \sigma_{2m} & \dots & \sigma_m^2 \end{pmatrix}$$

3. Description des données

En divisant par n on obtient :

$$\mathbf{M} = \frac{1}{n}\mathbf{X}^T\mathbf{X} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1m} & \sigma_{2m} & \cdots & \sigma_m^2 \end{pmatrix} \qquad (9)$$

matrice de covariance

La matrice $\mathbf{M}$ est une matrice symétrique dont les coefficients sont les covariances des couples de variables $(\mathbf{x}_{\bullet j}, \mathbf{x}_{\bullet k})$. $\mathbf{M}$ est définie positive ($\mathbf{u}^T\mathbf{M}\mathbf{u} > 0$ pour tout vecteur $\mathbf{u}$ non nul).

4. Méthode ACP

4. Méthode ACP

4.1. Objectif

Projeter le nuage centré, i.e. les n points $\mathbf{x}_{i\bullet} - \bar{\mathbf{x}} = (x_{i1} - \bar{x}_1, x_{i2} - \bar{x}_2, \ldots, x_{im} - \bar{x}_m)^T \in \mathbb{R}^m$, dans un espace de dimension $p \leq m$, ceci de façon à ce que les pertes d'informations soient minimales.

- On suppose que ce nouvel espace est muni d'une base orthonormale. Le problème revient donc à rechercher des vecteurs $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$ tels que $\mathbf{e}_i^T \mathbf{e}_j = \delta_{ij}$. Ces vecteurs sont appelés **axes factoriels.**

- Attention, même si l'espace de projection est de dimension $p \leq m$, les vecteurs de base $\mathbf{e}_j$ sont exprimés dans la base de l'espace initial qui est de dimension $m \Rightarrow$ les vecteurs $\mathbf{e}_j$ sont donc de dimension $(m \times 1)$.

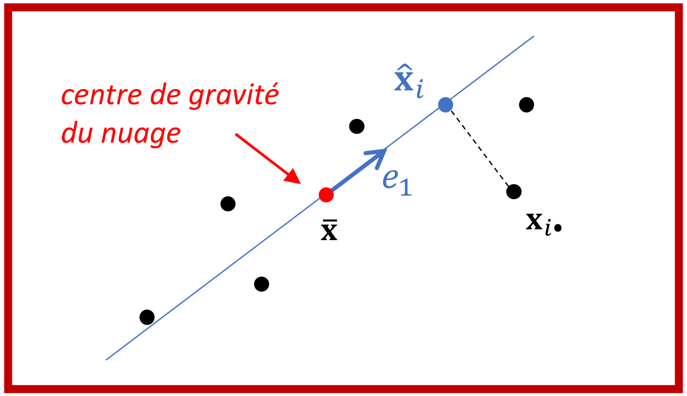
4. Méthode ACP

4.2. Premier axe factoriel

Le premier axe vectoriel est le vecteur $\mathbf{e}_1$: il est tel que la projection des points

$$\mathbf{x}_{i\bullet} - \bar{\mathbf{x}} = (x_{i1} - \bar{x}_1, x_{i2} - \bar{x}_2, \dots, x_{im} - \bar{x}_m)^T$$

dans sa direction soit **maximale** (voir figure ci-dessous dans le cas particulier de données en dimension 2).



On cherche donc le vecteur $\mathbf{e}_1$ qui maximise la quantité :

$$\boxed{J(\mathbf{e}_1) = \frac{1}{n}\sum_{i=1}^{n} \|\hat{\mathbf{x}}_i - \bar{\mathbf{x}}\|^2 \quad (10)}$$

Inertie du nuage de points par rapport à l'axe $\mathbf{e}_1$

avec

$$\hat{\mathbf{x}}_i - \bar{\mathbf{x}} = \mathbf{e}_1 \mathbf{e}_1^T (\mathbf{x}_{i\bullet} - \bar{\mathbf{x}}) \quad (11)$$

4. Méthode ACP

Théorème

Le premier axe factoriel $\mathbf{e}_1$ est le vecteur propre de la matrice de covariance $\mathbf{M}$ associé à la plus grande valeur propre.

Démonstration

1ère étape : montrons que $J(\mathbf{e}_1) = \mathbf{e}_1^T \mathbf{M} \mathbf{e}_1$

$$J(\mathbf{e}_1) = \frac{1}{n} \sum_{i=1}^{n} \|\hat{\mathbf{x}}_i - \bar{\mathbf{x}}\|^2$$

4. Méthode ACP

Démonstration (suite)

Chaque terme de la somme qui apparait dans la formule (12) est une matrice :

Démonstration (suite)

2ème étape : montrons que le vecteur $\mathbf{e}_1$ qui maximise $J(\mathbf{e}_1) = \mathbf{e}_1^T\mathbf{M}\mathbf{e}_1$ est le vecteur propre de $\mathbf{M}$ associé à la plus grande valeur propre (avec en plus la contrainte $\mathbf{e}_1^T\mathbf{e}_1 = 1$)

On peut montrer que maximiser $\mathbf{e}_1^T\mathbf{M}\mathbf{e}_1$ sous la contrainte $\mathbf{e}_1^T\mathbf{e}_1 - 1 = 0$ revient à maximiser la quantité :

$$\boxed{\mathcal{L}(\mathbf{e}_1) = \mathbf{e}_1^T\mathbf{M}\mathbf{e}_1 - \lambda(\mathbf{e}_1^T\mathbf{e}_1 - 1) \quad (14)}$$

C'est la **méthode des multiplicateurs de Lagrange** ($\mathcal{L}$ est appelé le **lagrangien** et λ est appelé **multiplicateur de Lagrange**).

Il faut donc chercher le vecteur $\mathbf{e}_1$ tels que le gradient de $\mathcal{L}(\mathbf{e}_1)$ soit nul : $\overrightarrow{\mathrm{grad}}\,\mathcal{L}(\mathbf{e}_1) = 0 \quad (15)$

Simplifions d'abord l'expression de $\mathcal{L}(\mathbf{e}_1)$: $\mathcal{L}(\mathbf{e}_1) = \mathbf{e}_1^T(\mathbf{M} - \lambda\mathbf{I})\mathbf{e}_1 + \lambda = \mathbf{e}_1^T\mathbf{A}\mathbf{e}_1 + \lambda \quad$ avec $\mathbf{A} = \mathbf{M} - \lambda\mathbf{I}$

Démonstration (suite et fin)

On a alors (cf. rappels page 12) :

$$\overrightarrow{\text{grad}}\,\mathcal{L}(\mathbf{e}_1) = 0 \quad \Leftrightarrow \quad 2\mathbf{A}\mathbf{e}_1 = 0 \quad \Leftrightarrow \quad (\mathbf{M} - \lambda\mathbf{I})\mathbf{e}_1 = 0 \quad \Leftrightarrow \quad \mathbf{M}\mathbf{e}_1 = \lambda\mathbf{e}_1$$

Le vecteur $\mathbf{e}_1$ recherché est donc bien un vecteur propre de la matrice de covariance.

La valeur maximale de $J(\mathbf{e}_1)$ vaut alors : $J(\mathbf{e}_1) = \mathbf{e}_1^T\lambda\mathbf{e}_1 = \lambda$

Le premier axe factoriel $\mathbf{e}_1$ est donc le vecteur propre de la matrice de covariance $\mathbf{M}$ associé à la plus grande valeur propre. La matrice de covariance étant symétrique définie positive, ses valeurs propres sont réelles positives.

Commentaires

- Il faut donc projeter les données sur le vecteur propre de $\mathbf{M}$ ayant la plus grande valeur propre pour obtenir le premier axe factoriel.

- Le second axe factoriel $\mathbf{e}_2$ est donc égal au vecteur propre associé à la deuxième valeur propre maximale de $\mathbf{M}$, et ainsi de suite jusqu'à l'axe factoriel $\mathbf{e}_p$.

- Le nombre d'axes factoriels est égal au nombre de valeurs propres non nulle. Ainsi, lorsqu'on diagonalise $\mathbf{M}$ il est conseillé d'écrire, dans la matrice diagonale, les valeurs propres dans l'ordre décroissant.

- On rappelle que la trace de la matrice $\mathbf{M}$ est égale à la somme de ses valeurs propres :

$$\mathrm{tr}(\mathbf{M}) = \sum_{j=1}^{m} \sigma_j^2 = \sum_{j=1}^{m} \lambda_j$$

4.3. Composantes principales

La $j^{\text{ème}}$ composante principale, notées $\mathbf{c}_j$ (avec $j = 1,2, \dots, p$), est définie par : $\boxed{\mathbf{c}_j = \mathbf{X}\mathbf{e}_j}$

avec :

- $\mathbf{X}$ est la matrice des données centrées de dimension $(n \times m)$
- $\mathbf{e}_j$ est le $j^{\text{ème}}$ axe factoriel de dimension $(m \times 1)$

On introduit les notations suivantes :

- Vecteurs $\mathbf{e}_j$
  - dimension : $(m \times 1)$
  - il y en a p (dimension de l'espace sur lequel on projette les points $\mathbf{x}_{i\bullet}$)
  - $\mathbf{e}_j = \left(e_{1j}, e_{2j}, \dots, e_{mj}\right)^T$
- Vecteurs $\mathbf{c}_j$
  - dimension : $(n \times 1)$, même dimension que les variables $\mathbf{x}_{\bullet j}$
  - il y en a p (dimension de l'espace sur lequel on projette les points $\mathbf{x}_{i\bullet}$)
  - $\mathbf{c}_j = \left(c_{1j}, c_{2j}, \dots, c_{nj}\right)^T$

4. Méthode ACP

On a alors :

$$\mathbf{c}_j = \begin{pmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{nj} \end{pmatrix} = \underbrace{\begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1m} - \bar{x}_m \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2m} - \bar{x}_m \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \dots & x_{nm} - \bar{x}_m \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} e_{1j} \\ e_{2j} \\ \vdots \\ e_{mj} \end{pmatrix}}_{\mathbf{e}_j} = \begin{pmatrix} \sum_{l=1}^{m} (x_{1l} - \bar{x}_l) e_{lj} \\ \sum_{l=1}^{m} (x_{2l} - \bar{x}_l) e_{lj} \\ \vdots \\ \sum_{l=1}^{m} (x_{nl} - \bar{x}_l) e_{lj} \end{pmatrix}$$

Ainsi, $\forall j \in [\![1, p]\!]$ et $\forall i \in [\![1, n]\!]$:

$$\boxed{c_{ij} = \sum_{l=1}^{m} (x_{il} - \bar{x}_l) e_{lj} = (\mathbf{x}_{i\bullet} - \bar{\mathbf{x}})^T \mathbf{e}_j}$$

La composante c_{ij} du vecteur $\mathbf{c}_j$ est la projection du point centré $\mathbf{x}_{i\bullet} - \bar{\mathbf{x}}$ sur l'axe factoriel $\mathbf{e}_j$.

4. Méthode ACP

Les données, initialement représentées par un nuage de n points dans un espace de dimension m (cf. matrice $\mathbf{X}$),

sont alors représentées par un nuage de n points également mais dans un espace de dimension $p \leq m$.

La matrice associée au nouveau nuage de points, notée $\mathbf{X}^*$, est définie par :

$$\boxed{\mathbf{X}^* = \mathbf{XP}}$$

où $\mathbf{P}$ est la matrice $(m \times p)$ dont les colonnes sont les p axes factoriels $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$:

$$\boxed{\mathbf{P} = \left(\mathbf{e}_1 | \mathbf{e}_2 | \dots | \mathbf{e}_p\right)}$$

Une écriture et un calcul matriciels par blocs donnent : $\mathbf{X}^* = \mathbf{X}\left(\mathbf{e}_1 | \mathbf{e}_2 | \dots | \mathbf{e}_p\right) = \left(\mathbf{X}\mathbf{e}_1 | \mathbf{X}\mathbf{e}_2 | \dots | \mathbf{X}\mathbf{e}_p\right)$

C'est-à-dire :

$$\boxed{\mathbf{X}^* = \left(\mathbf{c}_1 | \mathbf{c}_2 | \dots | \mathbf{c}_p\right)}$$

$\mathbf{X}^*$ est donc la matrice de dimension $(n \times p)$ dont les colonnes sont les composantes principales :

- la 1$^{\text{ère}}$ colonne est l'ensemble des projections des n points sur l'axe factoriel principal $\mathbf{e}_1$
- la 2$^{\text{ème}}$ colonne est l'ensemble des projections des n points sur l'axe factoriel secondaire $\mathbf{e}_2$
- etc …

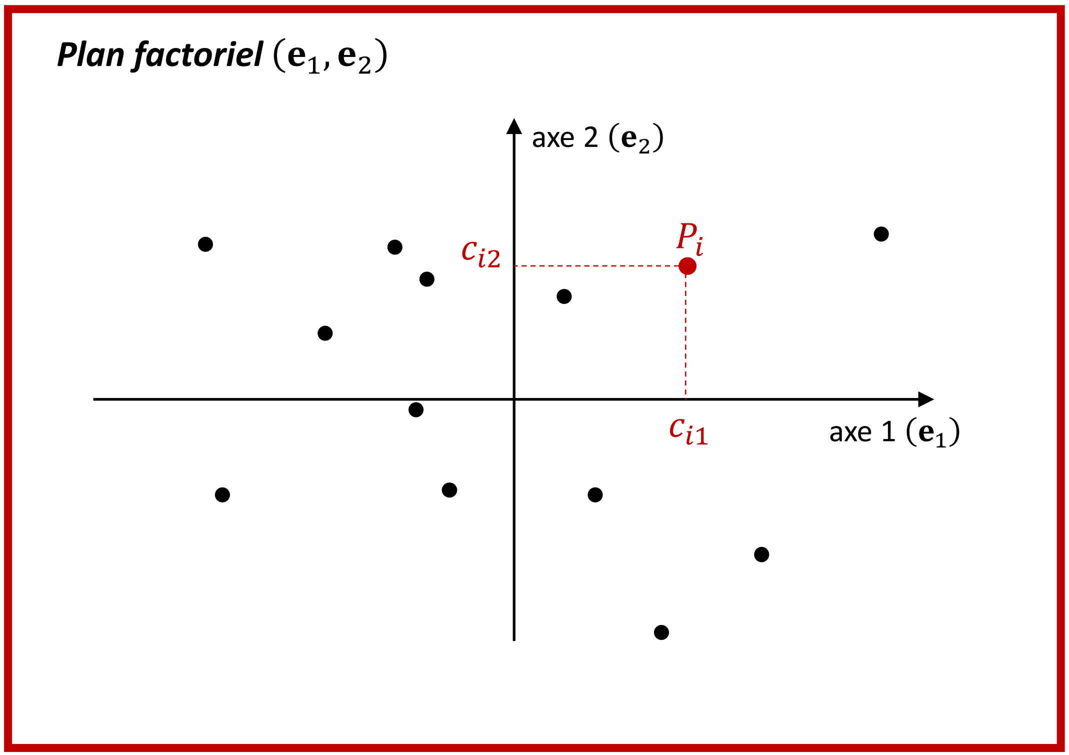
4. Méthode ACP

L'ACP permet donc de représenter les données dans un espace de dimension réduite ; on parle alors de **réduction de dimensionnalité.** Pour cela on a construit la matrice des composantes principales $\mathbf{c}_j$ (matrice $\mathbf{X}^*$) en projetant les données centrées (matrice $\mathbf{X}$) sur les axes factoriels $\mathbf{e}_j$ (matrice $\mathbf{P}$).

$$\mathbf{X} = \underbrace{\begin{pmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \dots & x_{1m}-\bar{x}_m \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \dots & x_{2m}-\bar{x}_m \\ \vdots & \vdots & & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \dots & x_{nm}-\bar{x}_m \end{pmatrix}}_{n\times m} \xrightarrow{ACP} \mathbf{X}^* = \mathbf{XP} = \underbrace{\begin{pmatrix} (\mathbf{x}_{1\bullet}-\bar{\mathbf{x}})^T\mathbf{e}_1 & (\mathbf{x}_{1\bullet}-\bar{\mathbf{x}})^T\mathbf{e}_2 & \dots & (\mathbf{x}_{1\bullet}-\bar{\mathbf{x}})^T\mathbf{e}_p \\ (\mathbf{x}_{2\bullet}-\bar{\mathbf{x}})^T\mathbf{e}_1 & (\mathbf{x}_{2\bullet}-\bar{\mathbf{x}})^T\mathbf{e}_2 & \dots & (\mathbf{x}_{2\bullet}-\bar{\mathbf{x}})^T\mathbf{e}_p \\ \vdots & \vdots & & \vdots \\ (\mathbf{x}_{n\bullet}-\bar{\mathbf{x}})^T\mathbf{e}_1 & (\mathbf{x}_{n\bullet}-\bar{\mathbf{x}})^T\mathbf{e}_2 & \dots & (\mathbf{x}_{n\bullet}-\bar{\mathbf{x}})^T\mathbf{e}_p \end{pmatrix}}_{n\times p}$$

4. Méthode ACP

Représentation géométrique (en dimension 2)



$$\begin{array}{ccccccc} & P_1 & P_2 & & P_i & & P_n \\ \mathbf{c}_1 = (& c_{11}, & c_{21}, & \dots, & c_{i1}, & \dots, & c_{n1})^T \\ \mathbf{c}_2 = (& c_{12}, & c_{22}, & \dots, & c_{i2}, & \dots, & c_{n2})^T \end{array}$$

avec $c_{i1} = (\mathbf{x}_{i\bullet} - \bar{\mathbf{x}})^T \mathbf{e}_1$ et $c_{i2} = (\mathbf{x}_{i\bullet} - \bar{\mathbf{x}})^T \mathbf{e}_2$

4. Méthode ACP

- La **variance expliquée** (on dit aussi l'**inertie expliquée**) par l'axe factoriel $\mathbf{e}_j$ est :

$$\boxed{I_j = \mathbf{e}_j^T \mathbf{M} \mathbf{e}_j = \mathbf{e}_j^T \lambda_j \mathbf{e}_j = \lambda_j}$$

Chaque valeur propre λ_j représente la partie de variance expliquée par l'axe factoriel correspondant $\mathbf{e}_j$, sa contribution relative est égale à :

$$\boxed{\tau_j = \frac{\lambda_j}{\sum_{j=1}^{m} \lambda_j}}$$

On dit que τ_j représente le taux d'**inertie expliquée** par l'axe factoriel e_j.

- Le taux d'**inertie totale** des axes factoriels est alors définie par :

$$\boxed{\tau = \frac{\sum_{j=1}^{p} \lambda_j}{\sum_{j=1}^{m} \lambda_j}}$$

Ainsi, si les variables originales sont très corrélées entre elles, alors un nombre réduits de composantes principales permet d'obtenir un taux de variance de l'ordre de 80% voire 90%.

4. Méthode ACP

Les composantes principales $\mathbf{c}_j$ avec $j = 1,2, \dots, p$ peuvent être vues comme de nouvelles « variables » statistiques dont les caractéristiques sont les suivantes (rappel : $\mathbf{c}_j = \mathbf{X}\mathbf{e}_j$) :

Moyenne : $\bar{\mathbf{c}}_j = 0$

Variance : $\mathrm{var}(\mathbf{c}_j) = \lambda_j$

Covariance : $\mathrm{cov}(\mathbf{c}_j, \mathbf{c}_k) = 0$

Démonstrations

Moyenne

$$\bar{\mathbf{c}}_j = \frac{1}{n}\sum_{i=1}^{n} c_{ij} =$$

Variance

$$\mathrm{var}(\mathbf{c}_j) = \frac{1}{n}\sum_{i=1}^{n}\left(c_{ij} - \bar{\mathbf{c}}_j\right)^2 =$$

Covariance

$$\mathrm{cov}(\mathbf{c}_j, \mathbf{c}_\mathbf{k}) = \frac{1}{n}\sum_{i=1}^{n}(c_{ij} - \bar{\mathbf{c}}_j)(c_{ik} - \bar{\mathbf{c}}_k) =$$

4. Méthode ACP

4.4. Cercle des corrélations

On définit les **données initiales réduites** de la façon suivante :

$$\boxed{\mathbf{z}_j = \frac{\mathbf{x}_{\bullet j} - \overline{\mathbf{x}}_j}{\sigma_j}}$$

On a alors des variables centrées et réduites, c'est-à-dire : $\overline{\mathbf{z}}_j = 0$ et $\mathrm{var}(\mathbf{z}_j) = 1$

Ces nouvelles variables ne privilégient pas les données dispersées et les distances sont indépendantes des unités de mesure. Le coefficient de corrélation entre les variables $\mathbf{z}_j$ et les composantes principales $\mathbf{c}_j$ s'écrit :

$$\boxed{\rho(\mathbf{z}_j, \mathbf{c}_k) = \frac{\mathbf{z}_j^T \mathbf{c}_k}{n\sqrt{\lambda_k}}}$$

En effet : $\rho(\mathbf{z}_j, \mathbf{c}_k) \stackrel{\text{def}}{=} \dfrac{\mathrm{cov}(\mathbf{z}_j, \mathbf{c}_k)}{\sqrt{\mathrm{var}(\mathbf{z}_j)\mathrm{var}(\mathbf{c}_k)}} = \dfrac{1}{n\sqrt{\lambda_k}} \displaystyle\sum_{i=1}^{n} z_{ij} c_{ik} = \dfrac{1}{n\sqrt{\lambda_k}} \mathbf{z}_j^T \mathbf{c}_k$

On représente alors chaque variable initiale réduite $\mathbf{z}_j$ par un point dans l'espace de projection dont les coordonnées sont les corrélations $\rho(\mathbf{z}_j, \mathbf{c}_k)$:

$$\left(\rho(\mathbf{z}_j, \mathbf{c}_1), \rho(\mathbf{z}_j, \mathbf{c}_2), \ldots, \rho(\mathbf{z}_j, \mathbf{c}_p)\right)$$

4. Méthode ACP

Montrons que, dans le cas d'une projection sur un plan factoriel (c'est-à-dire $p = 2$) , ces points sont à l'intérieur du cercle unité.

Pour cela, on calcule d'abord la covariance puis la corrélation entre les variables initiales centrées réduites $\mathbf{z}_j$ et les composantes principales $\mathbf{c}_k$:

$$\mathrm{cov}(\mathbf{z}_j, \mathbf{c}_k) = \mathrm{cov}\left(\sum_{l=1}^{p} a_{lj}\mathbf{c}_l \,, \mathbf{c}_k\right) = \sum_{l=1}^{p} a_{lj}\mathrm{cov}(\mathbf{c}_l, \mathbf{c}_k) = a_{kj}\lambda_k$$

$$\rho(\mathbf{z}_j, \mathbf{c}_k) \stackrel{\text{def}}{=} \frac{\mathrm{cov}(\mathbf{z}_j, \mathbf{c}_k)}{\sqrt{\mathrm{var}(\mathbf{z}_j)\mathrm{var}(\mathbf{c}_k)}} = a_{kj}\sqrt{\lambda_k}$$

Par ailleurs :

$$1 = \mathrm{var}(\mathbf{z}_j) = \mathrm{cov}(\mathbf{z}_j, \mathbf{z}_j) = \mathrm{cov}\left(\mathbf{z}_j, \sum_{k=1}^{p} a_{kj}\mathbf{c}_k\right)$$

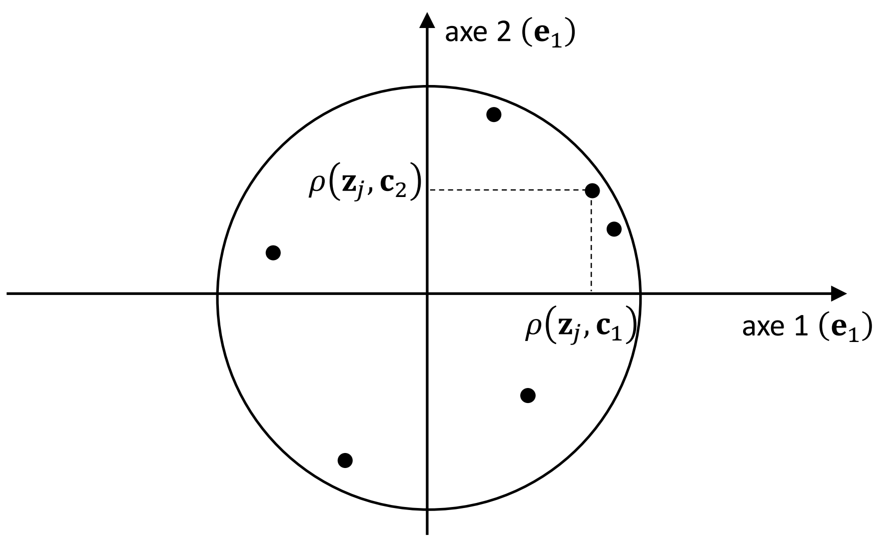
$$= \sum_{k=1}^{p} a_{kj}\mathrm{cov}(\mathbf{z}_j, \mathbf{c}_k) = \sum_{k=1}^{p} a_{kj}^2\lambda_k = \sum_{k=1}^{p} \rho(\mathbf{z}_j, \mathbf{c}_k)^2 \geq \rho(\mathbf{z}_j, \mathbf{c}_1)^2 + \rho(\mathbf{z}_j, \mathbf{c}_2)^2$$

4. Méthode ACP

Par conséquent, si on ne considère que les 2 premières composantes principales (c'est-à-dire $p = 2$) , les points dont le coordonnées sont $\left(\rho(\mathbf{z}_j, \mathbf{c}_1), \rho(\mathbf{z}_j, \mathbf{c}_2)\right)$ sont à l'intérieur du disque de rayon 1 :

$$\rho(\mathbf{z}_j, \mathbf{c}_1)^2 + \rho(\mathbf{z}_j, \mathbf{c}_2)^2 \leq 1$$

Plan factoriel $(\mathbf{e}_1, \mathbf{e}_2)$



- Attention, ici chaque point représente une variable (et non pas un individu).
- Un point proche de la circonférence signifie que la variable (réduite) associée est très corrélée avec l'une et/ou l'autre des deux composantes principales associées au plan factoriel (ici $\mathbf{c}_1$ et $\mathbf{c}_2$)

4.5. Reconstruction des données

La décomposition en valeurs singulière de la matrice des données centrées s'écrit :

$$\mathbf{X} = \mathbf{U}.\mathbf{S}.\mathbf{V}^T$$

$(n \times m)$

avec : $\mathbf{U} = (\mathbf{u}_1|\mathbf{u}_2|\dots|\,\mathbf{u}_n)$ $(n \times n)$ $\mathbf{V} = (\mathbf{v}_1|\mathbf{v}_2|\dots|\,\mathbf{v}_m)$ $(m \times m)$ $\mathbf{S} = \begin{pmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_p & & \\ & & & & \ddots & \\ & & & & & \sigma_m \\ & & & & & \end{pmatrix}$ $(n \times m)$

Les valeurs singulières $\sigma_1, \sigma_2, \dots, \sigma_m$ sont les racines carrées des valeurs propres de la matrice $\mathbf{X}^T.\mathbf{X}$:

$$\sigma_j = \sqrt{\mu_j} = \sqrt{n\lambda_j}$$

$\mathbf{X}^T.\mathbf{X} = n\mathbf{M}$

4. Méthode ACP

On montre que la matrice $\mathbf{X}$ peut aussi s'écrire comme une combinaison linéaire de matrices :

$$\mathbf{X} = \sum_{j=1}^{m} \sigma_j \mathbf{u}_j \mathbf{v}_j^T$$

En ne tenant compte que des p premières valeurs singulières (c'est-à-dire celles correspondantes aux p premiers axes factoriels), on obtient la matrice de reconstruction des données :

$$\underset{(n \times m)}{\mathbf{X}_{\boldsymbol{p}}} = \sum_{j=1}^{p} \sqrt{n\lambda_j}\, \mathbf{u}_j \mathbf{v}_j^T$$

5. Exemple

5. Exemple

Reprenons le tableau de l'exemple de la page 2 et faisons l'analyse en composantes principale de ces données. Ici

$n = 9$ (individu = étudiant) et $m = 5$ (variable = matière) $\Rightarrow$ nuage de 9 points dans un espace à 5 dimensions

- Matrice des données (colonne = variables $\mathbf{x}_{\bullet j}$ avec $j = 1,2,..,5$)

ELA	SC	M-ANA	Python	SSL	
6.0000	6.0000	8.0000	5.5000	8.0000	Théo
8.0000	8.0000	8.0000	8.0000	9.0000	Tom
6.0000	7.0000	11.0000	9.5000	11.0000	Paul
14.5000	14.5000	15.5000	15.0000	8.0000	Léa
14.0000	14.0000	12.0000	12.0000	10.0000	Jules
11.0000	10.0000	5.5000	7.0000	13.0000	Emile
5.5000	7.0000	14.0000	11.5000	10.0000	Eva
13.0000	12.0000	8.5000	9.5000	12.0000	Hugo
9.0000	9.5000	12.5000	12.0000	18.0000	Lydie

- Centre de gravité : $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_5)^T$

9.6667	9.7778	10.5556	10.0000	11.0000

- Matrice des données centrées $\mathbf{X}$

ELA	SC	M-ANA	Python	SSL	
-3.6667	-3.7778	-2.5556	-4.5000	-3.0000	Théo
-1.6667	-1.7778	-2.5556	-2.0000	-2.0000	Tom
-3.6667	-2.7778	0.4444	-0.5000	0	Paul
4.8333	4.7222	4.9444	5.0000	-3.0000	Léa
4.3333	4.2222	1.4444	2.0000	-1.0000	Jules
1.3333	0.2222	-5.0556	-3.0000	2.0000	Emile
-4.1667	-2.7778	3.4444	1.5000	-1.0000	Eva
3.3333	2.2222	-2.0556	-0.5000	1.0000	Hugo
-0.6667	-0.2778	1.9444	2.0000	7.0000	Lydie

5. Exemple

- Matrice de covariance $\mathbf{M} = \frac{1}{9}\mathbf{X}^T\mathbf{X}$

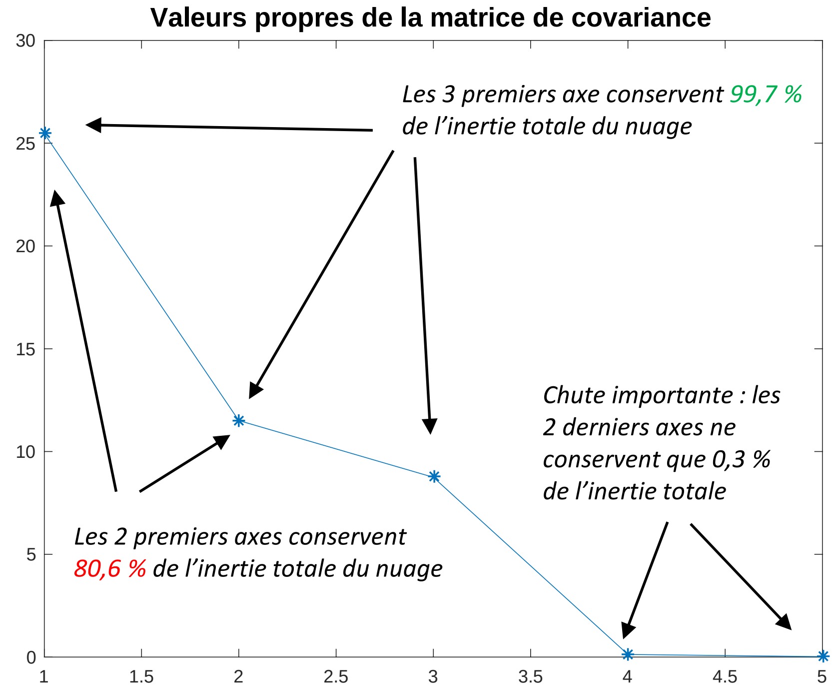
11.3889	9.7315	1.4352	4.5833	0.1111
9.7315	8.6728	2.9568	5.2778	0
1.4352	2.9568	9.4691	7.6944	-0.6111
4.5833	5.2778	7.6944	7.6667	0.7222
0.1111	0	-0.6111	0.7222	8.6667

- Réduction de dimensionnalité et taux de variance expliquée

si $p = 2,\ \tau = 0.8061$
si $p = 3,\ \tau = 0.9969$

- Valeurs propres de $\mathbf{M}$ (ordre décroissant)

```
25.4651
11.5074
 8.7480
 0.1254
 0.0183
```



⇒ On peut choisir de ne retenir que les 2 premiers axes $\mathbf{e}_1$ et $\mathbf{e}_2$ car le plan $(\mathbf{e}_1, \mathbf{e}_2)$ ne déforme pas trop le nuage (puisqu'il explique 80,6 % de l'inertie).

5. Exemple

- Les $p = 3$ premiers axes factoriels : matrice $\mathbf{P}$

axe 1 ($\mathbf{e}_1$)	axe 2 ($\mathbf{e}_2$)	axe 3 ($\mathbf{e}_3$)	
0.5728	-0.5103	-0.0542	ELA
0.5484	-0.2925	-0.0451	SC
0.3810	0.7054	-0.0147	M-ANA
0.4753	0.3887	0.1076	Python
0.0104	-0.0729	0.9916	SSL

Les axes factoriels sont des combinaisons linéaires des axes d'origine $\boldsymbol{\varepsilon}_1$ (ELA), $\boldsymbol{\varepsilon}_2$ (SC), $\boldsymbol{\varepsilon}_3$ (M-ANA), $\boldsymbol{\varepsilon}_4$ (Python) et $\boldsymbol{\varepsilon}_5$ (SSL). Par exemple :

$$\mathbf{e}_1 = 0.5728\ \boldsymbol{\varepsilon}_1 + 0.5484\ \boldsymbol{\varepsilon}_2 + 0.3810\ \boldsymbol{\varepsilon}_3 + 0.4753\ \boldsymbol{\varepsilon}_4 + 0.0104\ \boldsymbol{\varepsilon}_5$$

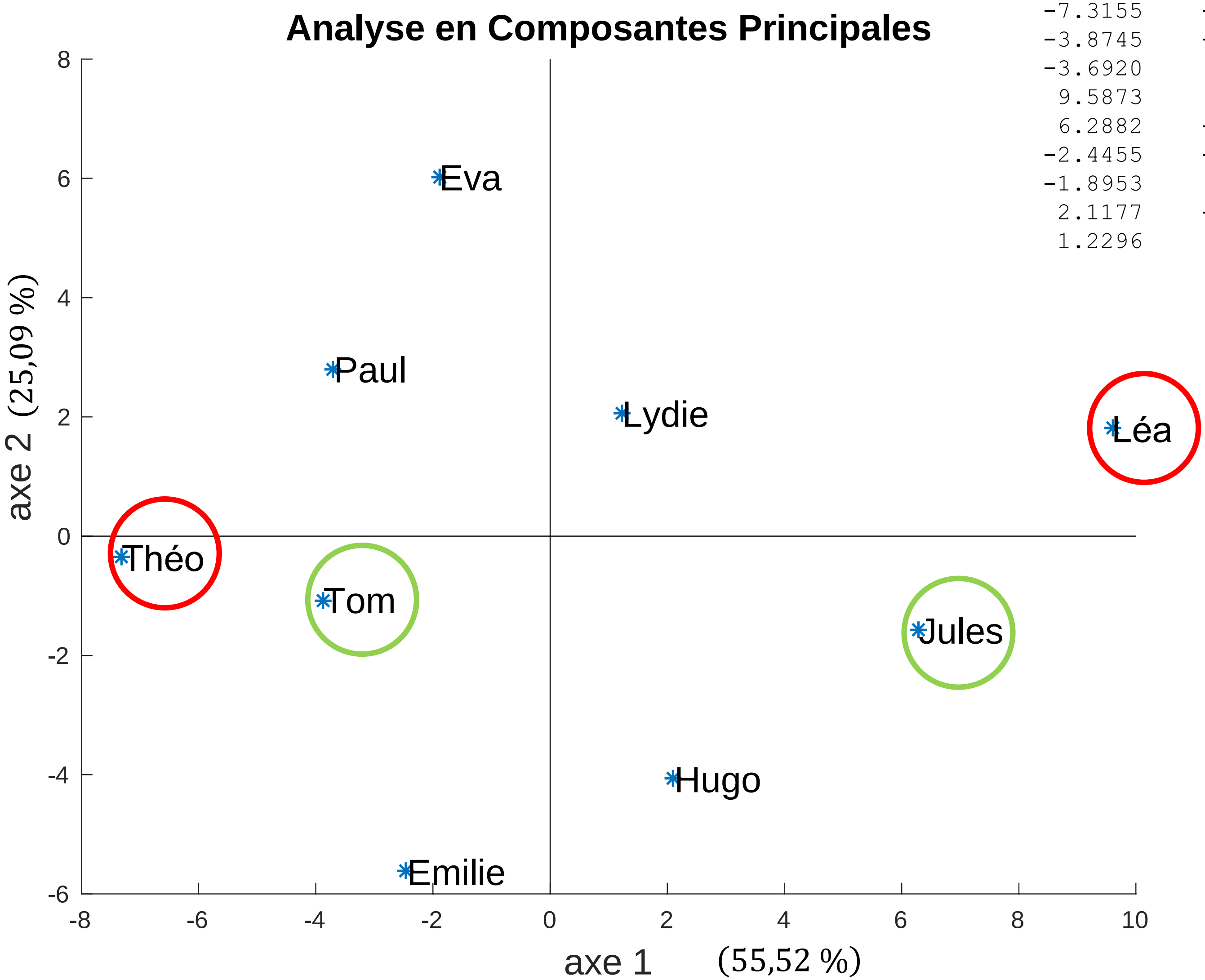
  - *Les axes $\mathbf{e}_1$ et $\mathbf{e}_2$ sont plutôt définis par les 4 premiers attributs (ELA, SC, M-ANA, Python) et très peu par le dernier (SSL)*
  - *L'axe $\mathbf{e}_3$ est essentiellement défini par le dernier attribut (SSL)*

- Matrice des composantes principales $\mathbf{X}^* = \mathbf{XP}$

$\mathbf{c}_1 = \mathbf{Xe}_1$	$\mathbf{c}_2 = \mathbf{Xe}_2$	$\mathbf{c}_3 = \mathbf{Xe}_3$	
-7.3155	-0.3570	-3.0520	Théo
-3.8745	-1.0638	-1.9902	Tom
-3.6920	2.8028	0.2638	Paul
9.5873	1.8024	-2.9847	Léa
6.2882	-1.5771	-1.2231	Jules
-2.4455	-5.6237	1.6525	Emile
-1.8953	6.0245	-0.5297	Eva
2.1177	-4.0683	0.6871	Hugo
1.2296	2.0602	7.1763	Lydie

5. Exemple

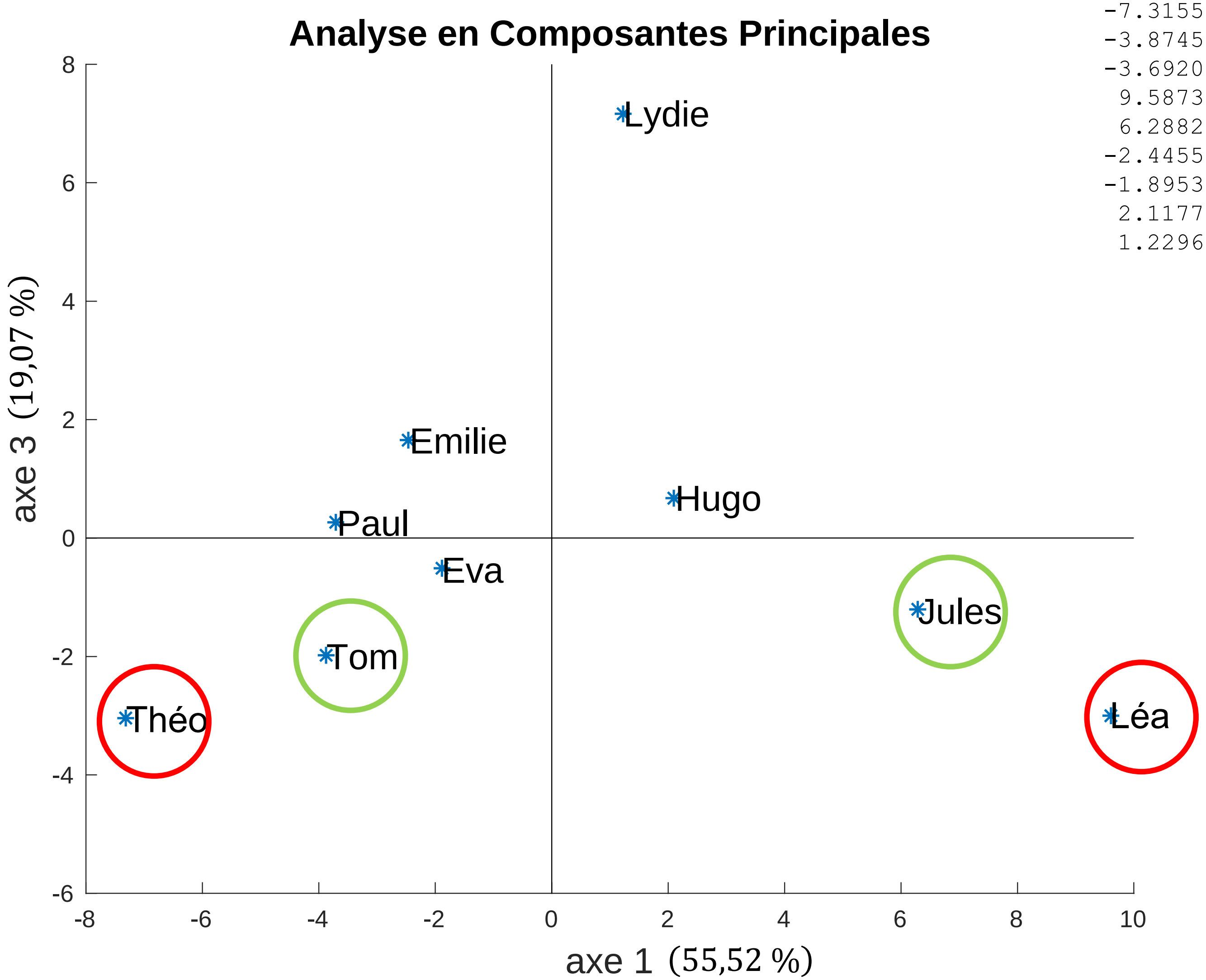
- Représentation des 12 points (individus) dans le repère $(\mathbf{e}_1, \mathbf{e}_2)$



$\mathbf{c}_1 = \mathbf{X}\mathbf{e}_1$	$\mathbf{c}_2 = \mathbf{X}\mathbf{e}_2$	$\mathbf{c}_3 = \mathbf{X}\mathbf{e}_3$	
-7.3155	-0.3570	-3.0520	Théo
-3.8745	-1.0638	-1.9902	Tom
-3.6920	2.8028	0.2638	Paul
9.5873	1.8024	-2.9847	Léa
6.2882	-1.5771	-1.2231	Jules
-2.4455	-5.6237	1.6525	Emile
-1.8953	6.0245	-0.5297	Eva
2.1177	-4.0683	0.6871	Hugo
1.2296	2.0602	7.1763	Lydie

5. Exemple

- Représentation des 12 points (individus) dans le repère $(\mathbf{e}_1, \mathbf{e}_3)$

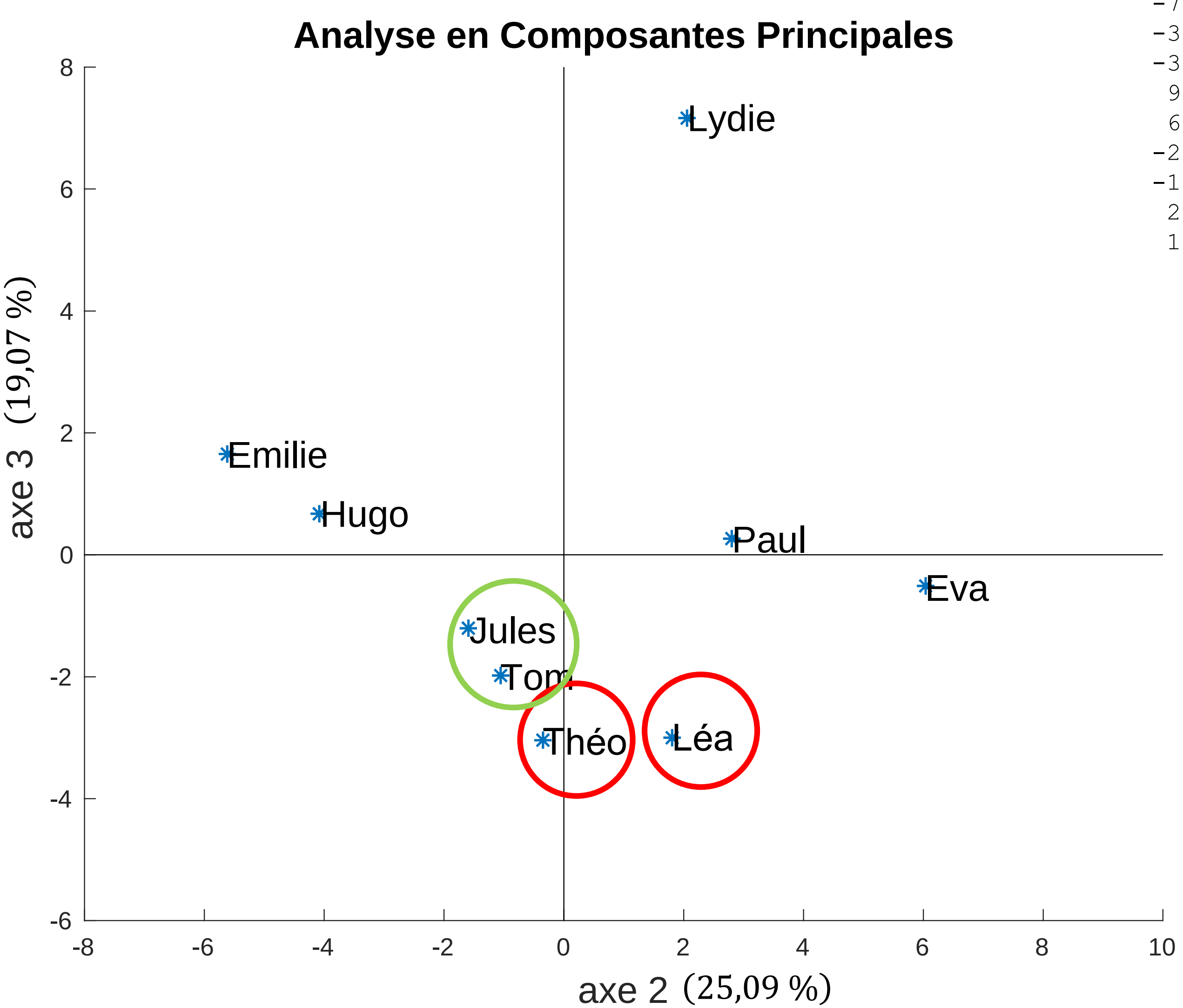


$\mathbf{c}_1 = \mathbf{X}\mathbf{e}_1$	$\mathbf{c}_2 = \mathbf{X}\mathbf{e}_2$	$\mathbf{c}_3 = \mathbf{X}\mathbf{e}_3$	
-7.3155	-0.3570	-3.0520	Théo
-3.8745	-1.0638	-1.9902	Tom
-3.6920	2.8028	0.2638	Paul
9.5873	1.8024	-2.9847	Léa
6.2882	-1.5771	-1.2231	Jules
-2.4455	-5.6237	1.6525	Emile
-1.8953	6.0245	-0.5297	Eva
2.1177	-4.0683	0.6871	Hugo
1.2296	2.0602	7.1763	Lydie

5. Exemple

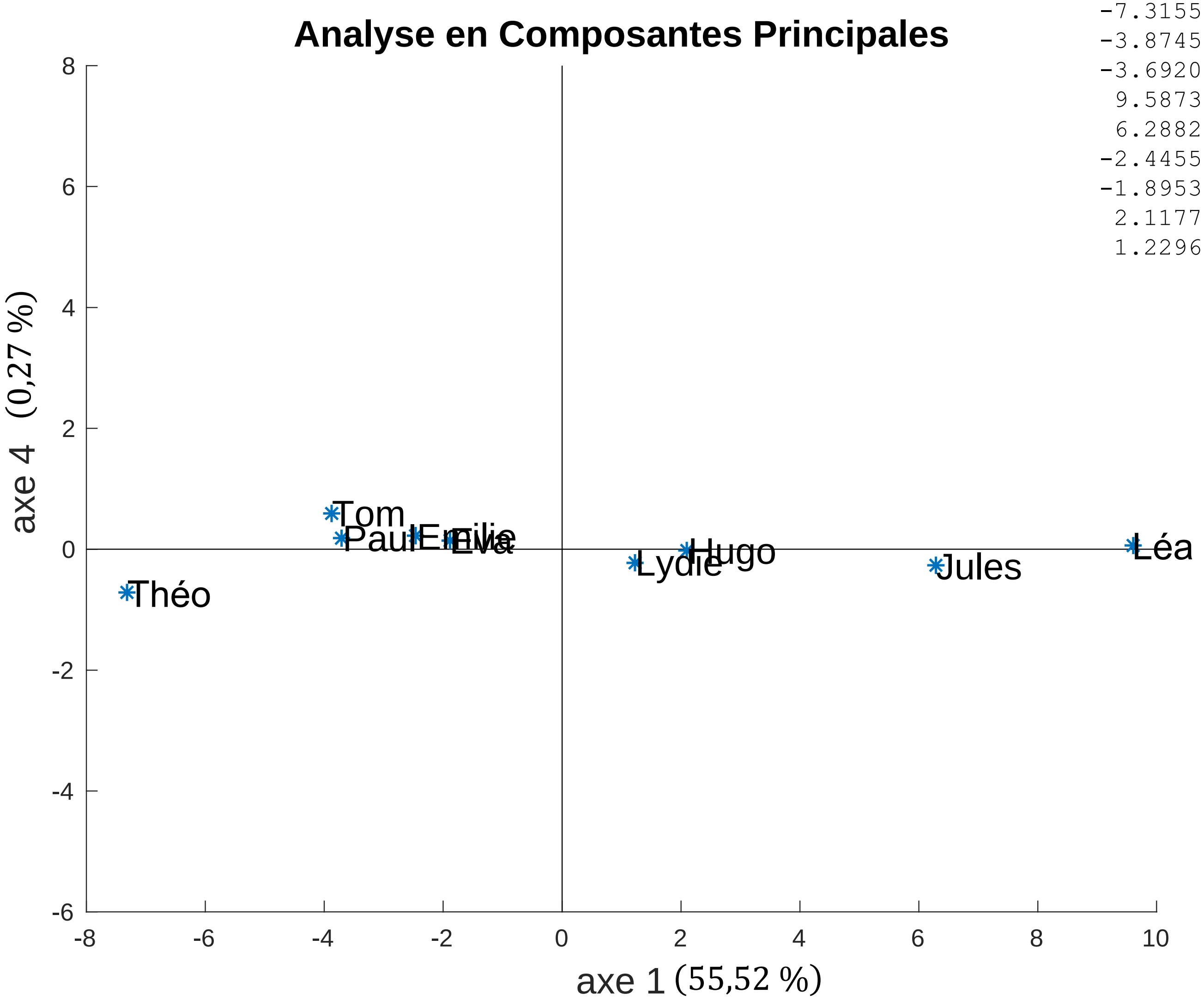
- Représentation des 12 points (individus) dans le repère $(\mathbf{e}_2, \mathbf{e}_3)$

$\mathbf{c}_1 = \mathbf{X}\mathbf{e}_1$	$\mathbf{c}_2 = \mathbf{X}\mathbf{e}_2$	$\mathbf{c}_3 = \mathbf{X}\mathbf{e}_3$	
-7.3155	-0.3570	-3.0520	Théo
-3.8745	-1.0638	-1.9902	Tom
-3.6920	2.8028	0.2638	Paul
9.5873	1.8024	-2.9847	Léa
6.2882	-1.5771	-1.2231	Jules
-2.4455	-5.6237	1.6525	Emile
-1.8953	6.0245	-0.5297	Eva
2.1177	-4.0683	0.6871	Hugo
1.2296	2.0602	7.1763	Lydie



5. Exemple

- Par curiosité, représentation des 12 points dans le repère $(\mathbf{e}_1, \mathbf{e}_4)$

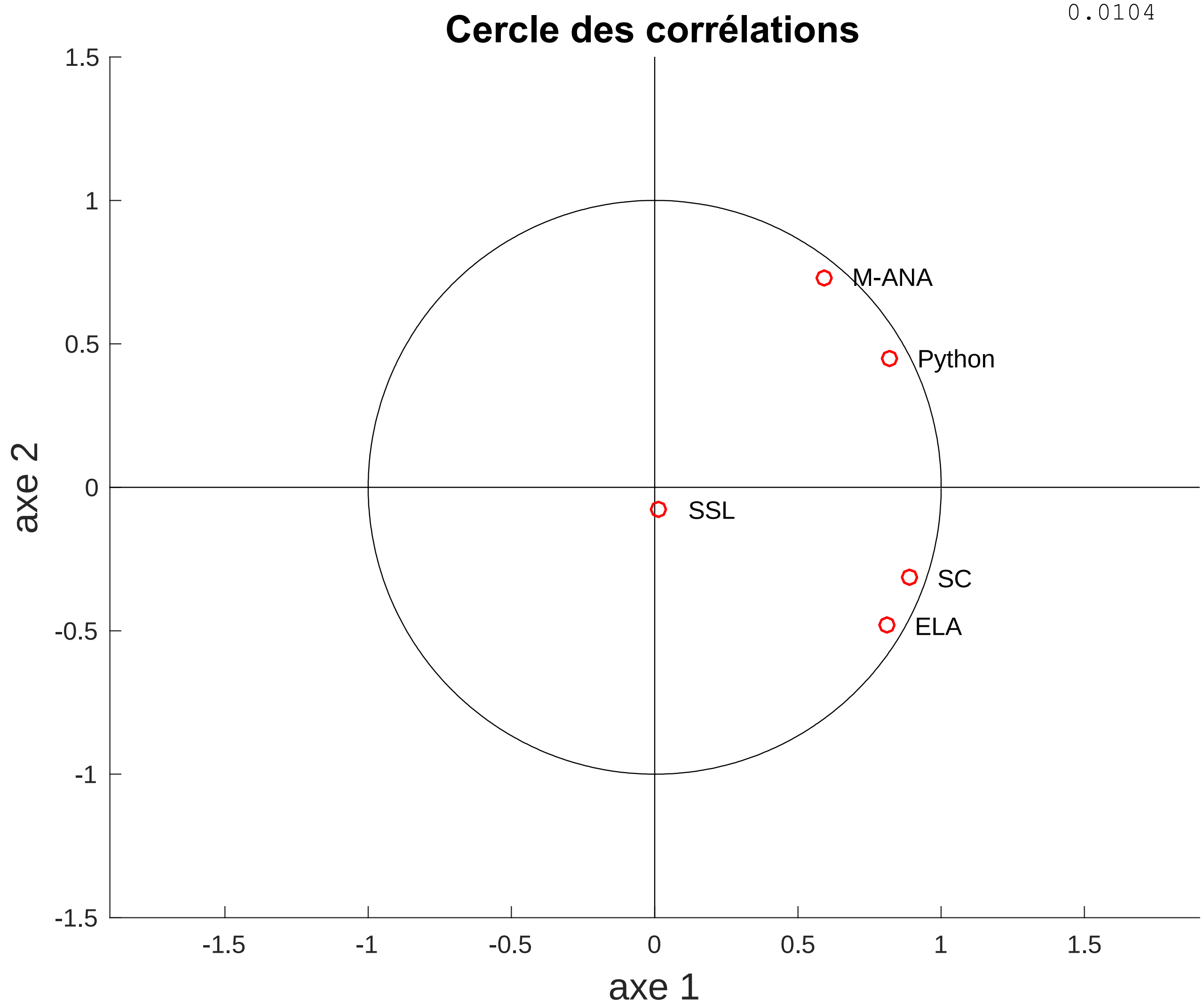


$\mathbf{c}_1 = \mathbf{X}\mathbf{e}_1$	$\mathbf{c}_2 = \mathbf{X}\mathbf{e}_2$	$\mathbf{c}_3 = \mathbf{X}\mathbf{e}_3$	
-7.3155	-0.3570	-3.0520	Théo
-3.8745	-1.0638	-1.9902	Tom
-3.6920	2.8028	0.2638	Paul
9.5873	1.8024	-2.9847	Léa
6.2882	-1.5771	-1.2231	Jules
-2.4455	-5.6237	1.6525	Emile
-1.8953	6.0245	-0.5297	Eva
2.1177	-4.0683	0.6871	Hugo
1.2296	2.0602	7.1763	Lydie

5. Exemple

- Cercle des corrélations dans le repère $(\mathbf{e}_1, \mathbf{e}_2)$

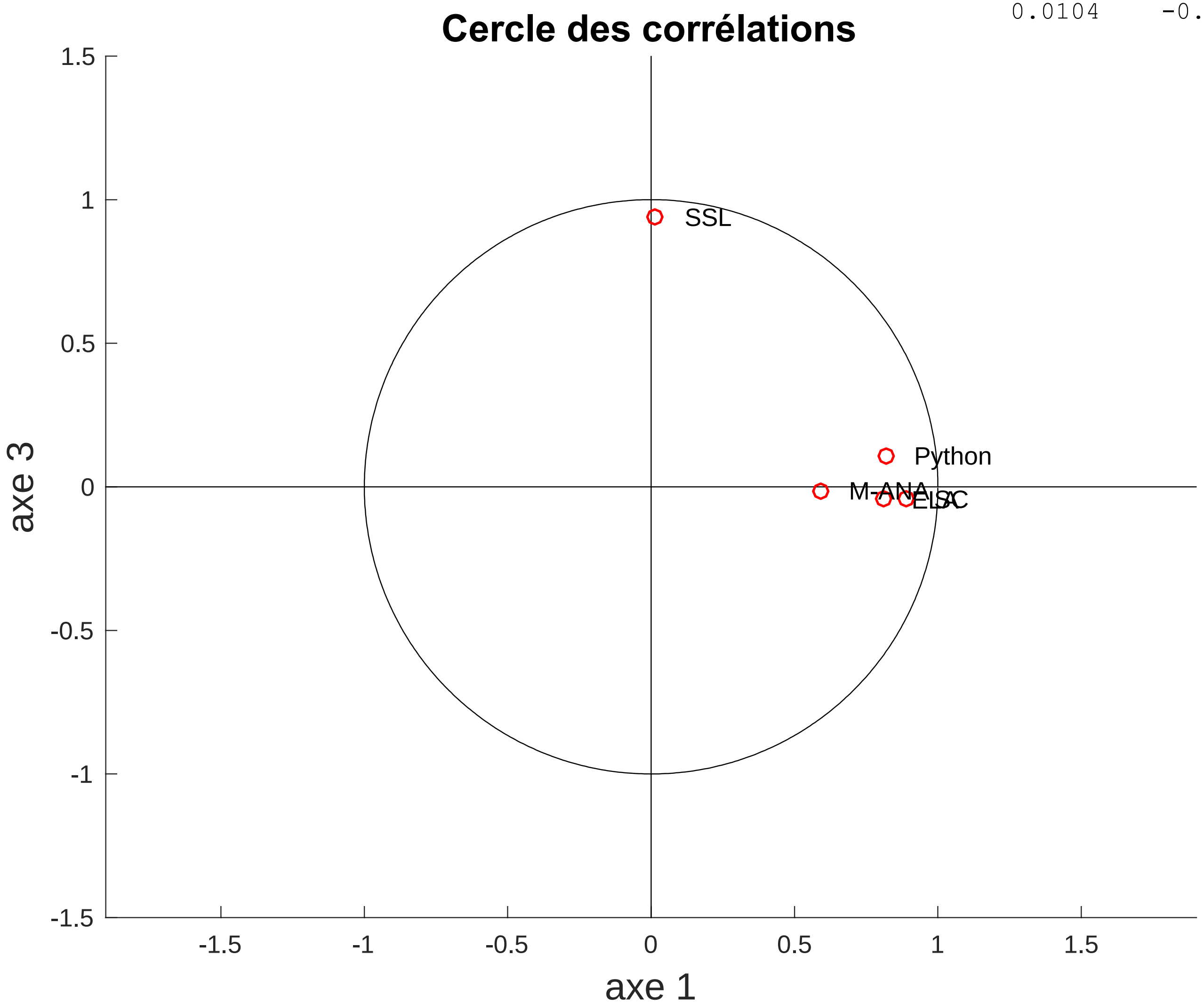
axe 1 ($\mathbf{e}_1$)	axe 2 ($\mathbf{e}_2$)	axe 3 ($\mathbf{e}_3$)	
0.5728	-0.5103	-0.0542	ELA
0.5484	-0.2925	-0.0451	SC
0.3810	0.7054	-0.0147	M-ANA
0.4753	0.3887	0.1076	Python
0.0104	-0.0729	0.9916	SSL



5. Exemple

- Cercle des corrélations dans le repère $(\mathbf{e}_1, \mathbf{e}_3)$

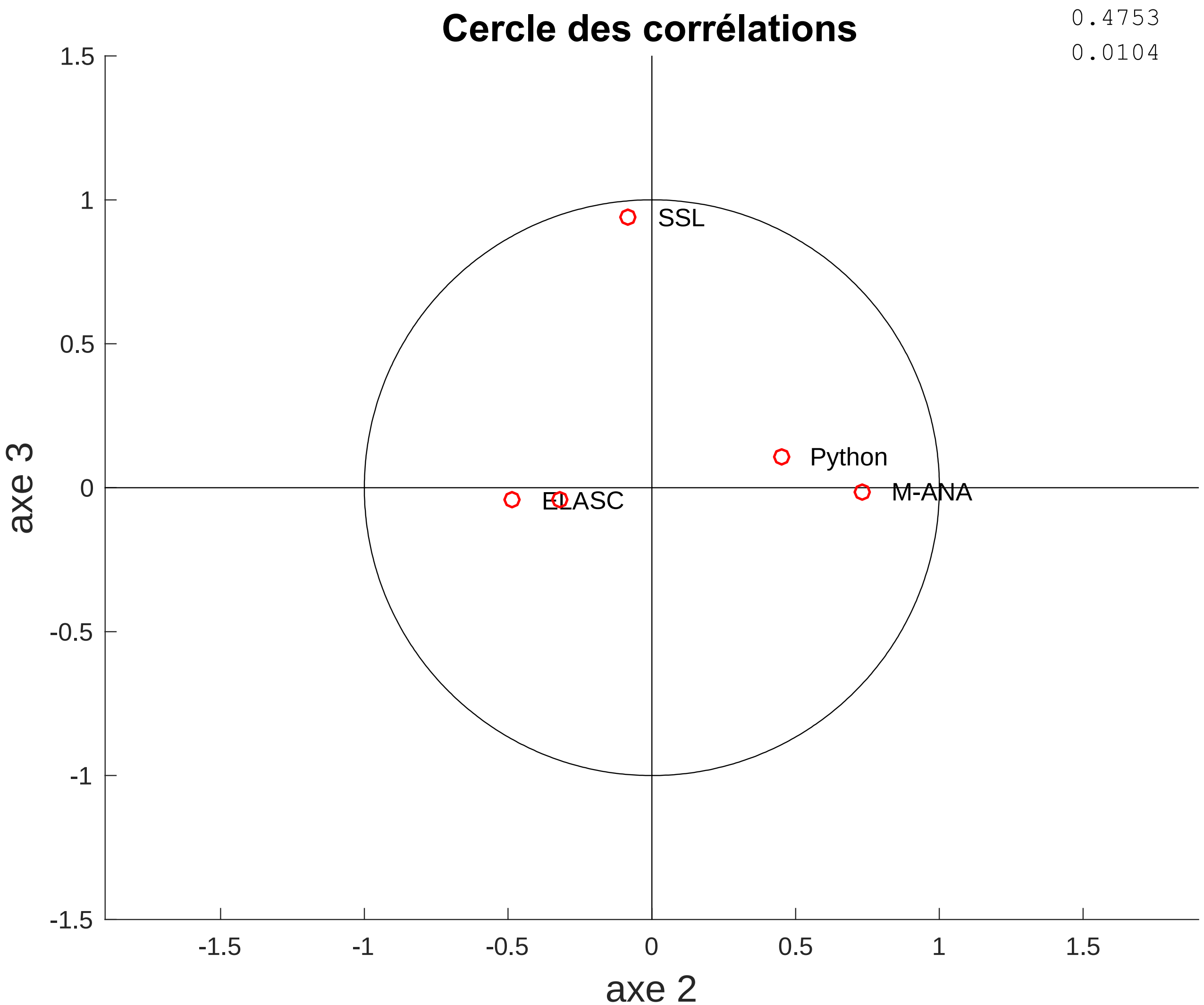
axe 1 ($\mathbf{e}_1$)	axe 2 ($\mathbf{e}_2$)	axe 3 ($\mathbf{e}_3$)	
0.5728	-0.5103	-0.0542	ELA
0.5484	-0.2925	-0.0451	SC
0.3810	0.7054	-0.0147	M-ANA
0.4753	0.3887	0.1076	Python
0.0104	-0.0729	0.9916	SSL



5. Exemple

- Cercle des corrélations dans le repère $(\mathbf{e}_2, \mathbf{e}_3)$

axe 1 ($\mathbf{e}_1$)	axe 2 ($\mathbf{e}_2$)	axe 3 ($\mathbf{e}_3$)	
0.5728	-0.5103	-0.0542	ELA
0.5484	-0.2925	-0.0451	SC
0.3810	0.7054	-0.0147	M-ANA
0.4753	0.3887	0.1076	Python
0.0104	-0.0729	0.9916	SSL



5. Exemple

- Reconstruction des données à partir de $p = 3$ composantes principales