

**FILIERE : SCIENCES MATHEMATIQUES & APPLICATIONS**

**MODULE : PROJET DE FIN D'ETUDES**

**COORDONNATEUR : Khalid ADNAOUI**

# **PROJET DE FIN D'ÉTUDE**

**Analyse des composantes principales**

**Réalisé par : Karim HANFAOUI**

**Présenté et soutenu devant le jury composé de :**

<i>Pr.</i>	Khadija BOUZKOURA	-----	<i>Présidente</i>
<i>Pr.</i>	Abdelhafid BASSOU	-----	<i>Examinateur</i>
<i>Pr.</i>	Imane EL BERRAI	-----	<i>Examinateuse</i>
<i>Pr.</i>	Malika IZID	-----	<i>Examinateuse</i>
<i>Pr.</i>	Khalid ADNAOUI	-----	<i>Encadrant</i>

**2024 -2025**

# Table des matières

<b>1 Remerciement</b>	<b>2</b>
<b>2 Introduction Générale</b>	<b>3</b>
2.1 Applications pratiques . . . . .	3
2.2 Contexte historique . . . . .	3
2.3 Cas classiques d'utilisation . . . . .	3
2.3.1 Simulation et méthodologie . . . . .	4
2.3.2 Résultats . . . . .	4
2.3.3 Autres Applications Illustratives de l'ACP . . . . .	5
2.3.4 Discussion . . . . .	6
2.3.5 Conclusion . . . . .	7
<b>3 Préliminaire</b>	<b>8</b>
3.1 Préliminaires Mathématiques pour l'Analyse en Composantes Principales . . . . .	8
3.1.1 Statistiques de base et covariance . . . . .	8
3.1.2 Diagonalisation et valeurs propres . . . . .	9
3.1.3 Décomposition en valeurs singulières . . . . .	10
3.1.4 Centrage et normalisation des données . . . . .	11
3.1.5 Extraction des valeurs propres et vecteurs propres . . . . .	11
<b>4 Analyse en Composantes Principales (PCA)</b>	<b>14</b>
4.1 Formulation mathématique . . . . .	14
4.1.1 Problème d'optimisation fondamental . . . . .	14
4.1.2 Dérivation complète de l'ACP . . . . .	14
4.1.3 Propriétés importantes . . . . .	16
4.1.4 Choix du nombre de composantes . . . . .	16
4.1.5 Interprétation géométrique . . . . .	17
4.2 Limitations de l'ACP classique . . . . .	17
4.2.1 Linéarité . . . . .	17
4.2.2 Sensibilité aux valeurs aberrantes . . . . .	18
4.2.3 Hypothèse de normalité . . . . .	19
4.2.4 Perte d'interprétabilité . . . . .	19
4.2.5 Solutions alternatives . . . . .	19
<b>5 Application pratique de l'ACP</b>	<b>20</b>
5.1 Réduction de dimension et qualité visuelle . . . . .	20
5.2 Eigenfaces : Méthode de Reconnaissance Faciale . . . . .	21
5.3 Rôle central de l'ACP dans Eigenfaces . . . . .	21
<b>6 Analyse des résultats</b>	<b>23</b>
6.1 Indice SSIM et application à l'ACP . . . . .	23
6.2 Calcul des moyennes locales $\mu_x$ et $\mu_y$ . . . . .	24
6.3 Constantes $C_1$ , $C_2$ , $C_3$ . . . . .	24

6.4	Interprétation des résultats . . . . .	24
6.5	Conclusion et Perspectives . . . . .	25
<b>A</b>	<b>Kernel PCA</b>	<b>26</b>
A.1	Fondements mathématiques du Kernel PCA . . . . .	26
A.1.1	Formulation mathématique . . . . .	26
A.1.2	L'astuce du noyau . . . . .	26
A.1.3	Résolution du problème aux valeurs propres . . . . .	26
A.1.4	Centrage dans l'espace de caractéristiques . . . . .	27
A.1.5	Projection des données . . . . .	27
A.1.6	Complexité et considérations pratiques . . . . .	28
A.1.7	Théorème de représentation . . . . .	28
A.1.8	Choix du noyau . . . . .	28
<b>Bibliographie</b>		<b>28</b>
<b>Webographie</b>		<b>30</b>

# Liste des tableaux

4.1	Résumé des limitations et solutions . . . . .	19
A.1	Comparaison des noyaux courants . . . . .	28

# Liste des figures

2.1	Projection ACP - Vérite terrain : les passagers normaux sont en bleu, les suspects simulés en rouge. . . . .	4
2.2	Projection ACP - Détection par Isolation Forest : les anomalies détectées apparaissent isolées. . . . .	5
2.3	Projection ACP de données étiquetées montrant la séparation entre deux classes. . . . .	6
2.4	Scree plot montrant la variance expliquée par chaque composante principale. . . . .	6
2.5	À gauche : signal bruité. À droite : signal reconstruit après débruitage par ACP. . . . .	7
3.1	Illustration géométrique de la covariance. Lorsque la covariance est positive, les deux variables ont tendance à augmenter ensemble. Si la covariance est négative, elles varient en sens opposé. . . . .	9
3.2	Représentation des valeurs propres et des vecteurs propres dans l'espace des données. Les vecteurs propres indiquent les directions de la variance maximale. . . . .	10
3.3	Illustration du centrage des données. Chaque dimension est ajustée de manière à avoir une moyenne nulle. . . . .	11
3.4	Extraction des valeurs propres et vecteurs propres (Power Iteration, Déflation, et QR) . . . . .	13
4.1	Calcul des composantes principales . . . . .	16
4.2	Interprétation géométrique de l'ACP comme rotation de l'espace . . . . .	17
4.3	Exemple de données non-linéaires où l'ACP échoue . . . . .	18
4.4	Impact des valeurs aberrantes sur les composantes principales . . . . .	18
5.1	Effet du nombre de composantes principales $k$ sur la qualité visuelle. . . . .	20
5.2	Relation entre le nombre de composantes $k$ , la taille de fichier (Ko) et la qualité de reconstruction (PSNR en dB). . . . .	21
A.1	Illustration du centrage et de la normalisation des données. . . . .	27
A.2	Illustration du mapping non linéaire via le kernel trick . . . . .	27

# Chapitre 1

## Remerciement

*Chers membres du jury,*

Je tiens à vous remercier sincèrement pour votre présence et votre attention lors de ma soutenance de projet de fin d'étude. Cet aboutissement n'aurait pas été possible sans votre soutien et vos conseils tout au long de cette aventure.

Je souhaite également remercier Pr. ADNAOUI Khalid pour leur collaboration et leur engagement dans ce projet et leur patient avec moi. *Nous avons travaillé ensemble avec passion et détermination pour offrir un travail de qualité.*

Je tiens également à exprimer ma gratitude envers mes enseignants qui m'ont éclairé de leur expertise et ont contribué à l'élaboration de ce projet. *Vous avez été d'une grande aide dans mon parcours académique et professionnel.*

*Enfin, je voudrais adresser un grand merci à ma famille et mes amis qui m'ont soutenu et encouragé tout au long de cette longue période d'étude. Votre amour constant m'a motivé à aller de l'avant malgré les défis.* Encore une fois, merci infiniment pour votre temps, votre soutien et l'opportunité de présenter mon projet devant vous. Cela restera un moment inoubliable de ma vie.

**Cordialement,**

*HANFAOUI Karim*

# Chapitre 2

## Introduction Générale

L'Analyse des Composantes Principales (ACP) est une méthode statistique de réduction de dimensionnalité qui transforme des variables corrélées en un ensemble de composantes orthogonales maximisant la variance expliquée. Fondée sur l'algèbre linéaire et la théorie des matrices, elle permet d'extraire l'information essentielle tout en préservant la structure des données.

### 2.1 Applications pratiques

- **Génomique** : Analyse de l'expression génique et classification d'échantillons biologiques
- **Finance** : Identification des facteurs de risque dans les portefeuilles d'investissement
- **Imagerie numérique** : Compression d'images et reconnaissance faciale (méthode Eigenfaces)
- **Météorologie** : Détection des modes climatiques dominants (ex : El Niño)

### 2.2 Contexte historique

- 1901 : Karl Pearson établit les bases mathématiques
- 1933 : Harold Hotelling formalise la méthode moderne
- 1960s : Applications en psychométrie et économétrie
- 2000s : Intégration dans l'apprentissage automatique

### 2.3 Cas classiques d'utilisation

- **Reconnaissance faciale (Eigenfaces)** : Réduction de 10 000 pixels à 150 composantes principales tout en conservant 95 % de l'information
- **Analyse génomique** : Visualisation de populations humaines à partir de 500 000 SNPs (Polymorphismes nucléotidiques)
- **Marchés financiers** : Identification de 3 facteurs principaux expliquant 80 % de la variance du rendement

- **Études de consommation** : Réduction de 50 variables marketing en 2 axes principaux : "pratичité" vs "luxe"
- **Climatologie** : Détection du phénomène El Niño à partir de 40 ans de données océanographiques multidimensionnelles

### 2.3.1 Simulation et méthodologie

Nous simulons les données de 1000 passagers normaux et 20 passagers suspects à l'aide de distributions gaussiennes multivariées. Chaque observation est un vecteur à 4 dimensions. Les données sont ensuite standardisées, projetées en 2D via l'ACP, puis soumises à l'algorithme *Isolation Forest* pour détecter les anomalies.

### 2.3.2 Résultats

#### Projection en deux dimensions par ACP

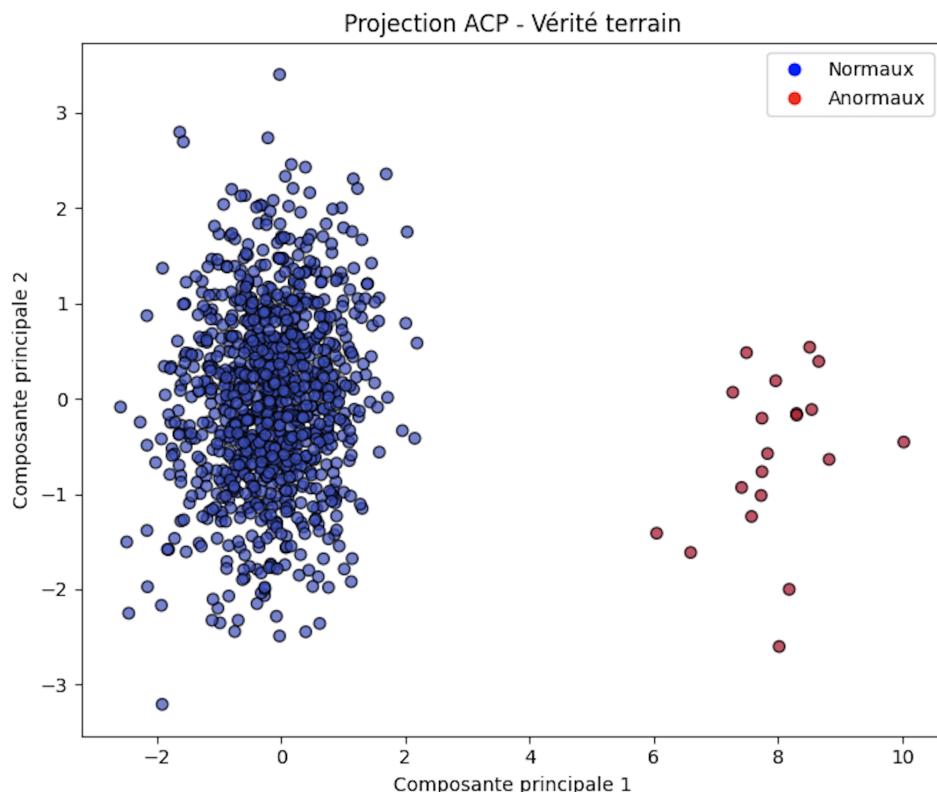


FIGURE 2.1 – Projection ACP - Vérite terrain : les passagers normaux sont en bleu, les suspects simulés en rouge.

## Détection automatique des comportements anormaux

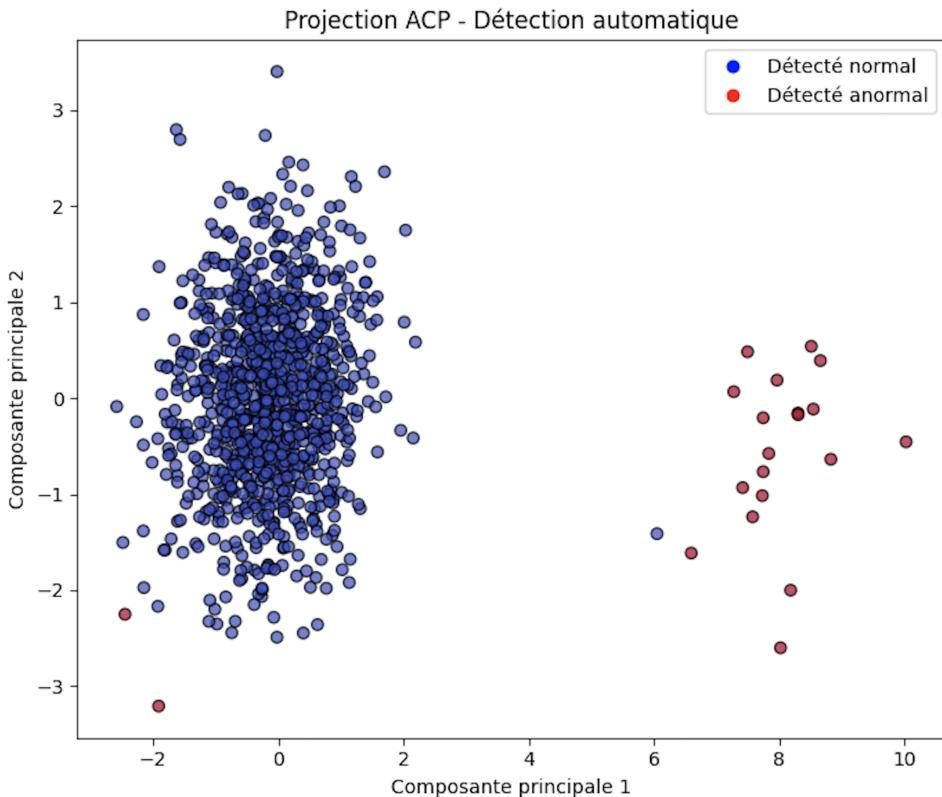


FIGURE 2.2 – Projection ACP - Détection par Isolation Forest : les anomalies détectées apparaissent isolées.

### 2.3.3 Autres Applications Illustratives de l'ACP

#### Visualisation de la séparation des classes

L'ACP est souvent utilisée pour explorer la séparabilité des classes dans les problèmes de classification supervisée. En projetant les données dans le plan formé par les deux premières composantes principales, on peut visualiser si les classes sont bien distinctes ou fortement entremêlées.

#### Analyse de la variance expliquée : Scree Plot

Le élément clé de l'ACP est la quantification de la variance capturée par chaque composante. Le scree plot permet d'identifier combien de composantes suffisent à capturer l'information utile.

#### Filtrage de bruit : Débruitage par ACP

L'ACP peut être utilisée pour reconstruire un signal bruité en conservant uniquement les composantes principales les plus significatives.

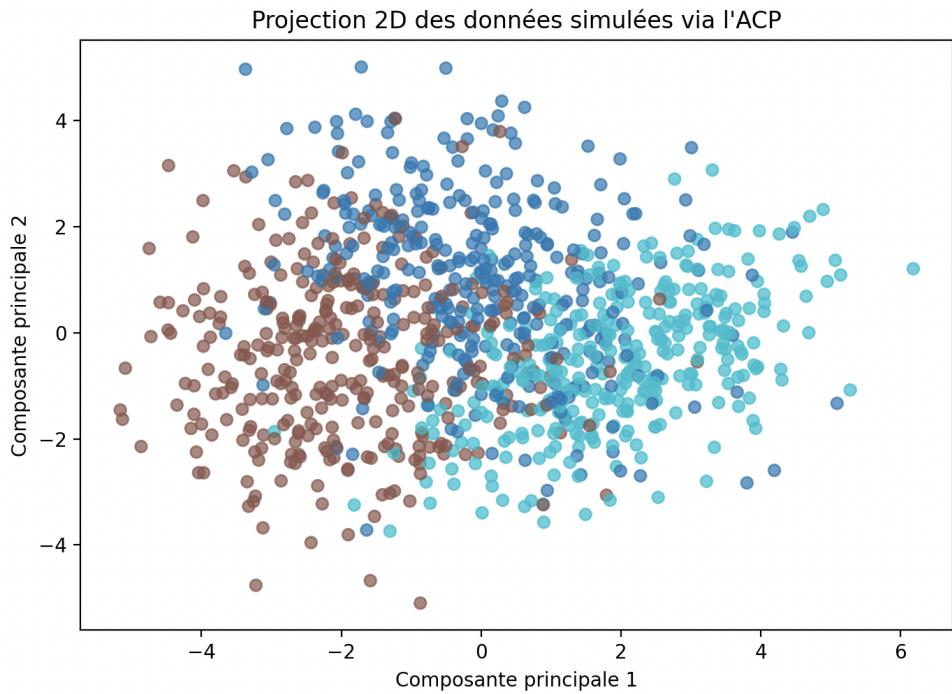


FIGURE 2.3 – Projection ACP de données étiquetées montrant la séparation entre deux classes.

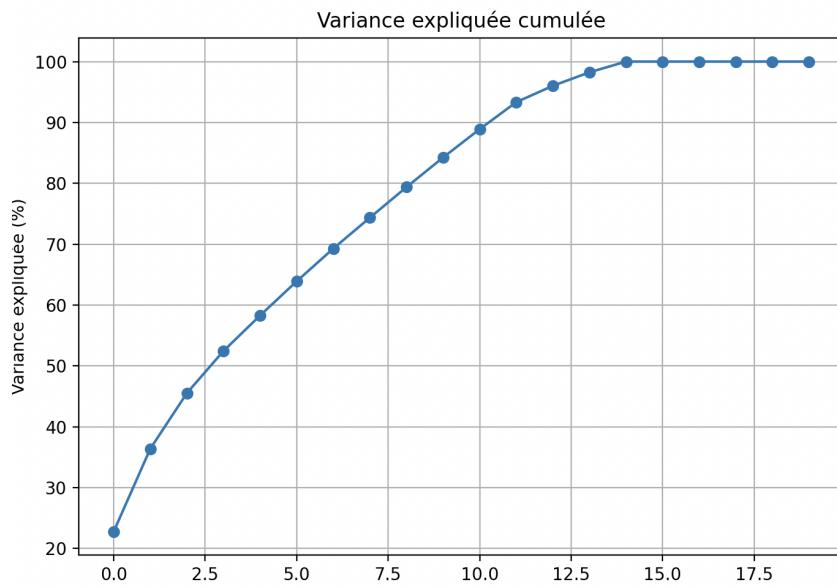


FIGURE 2.4 – Scree plot montrant la variance expliquée par chaque composante principale.

### 2.3.4 Discussion

L'utilisation de l'ACP permet de visualiser efficacement les regroupements naturels et les *outliers*. Combinée à un algorithme de détection non supervisée, elle facilite l'analyse des flux passagers en temps réel sans surveillance humaine directe. Cette méthode est à la fois rapide, explicable et généralisable à d'autres contextes de surveillance intelligente. Elle est également applicable à la classification, la compression, et le traitement du signal.

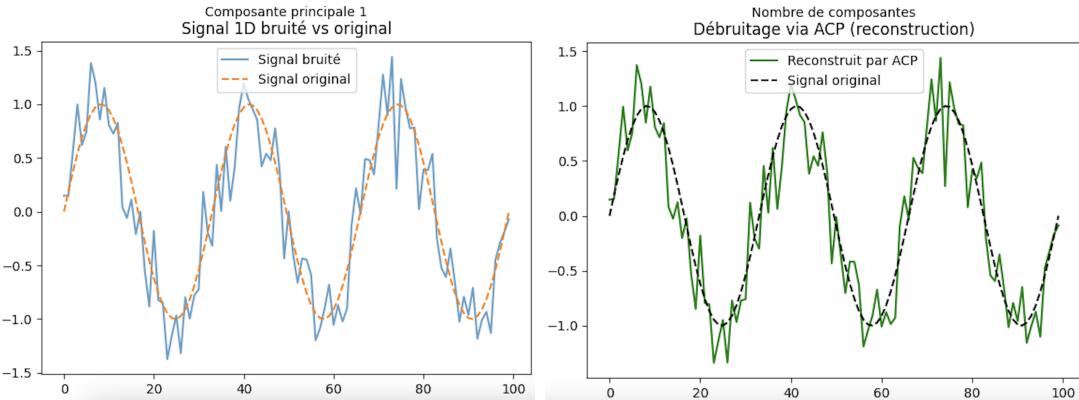


FIGURE 2.5 – À gauche : signal bruité. À droite : signal reconstruit après débruitage par ACP.

### 2.3.5 Conclusion

Ce chapitre a montré comment l'ACP peut être utilisée comme une prétraitement puissant pour la détection d'anomalies dans des données de haute dimension. Dans notre cas, elle permet une représentation 2D compréhensible facilitant l'identification des comportements anormaux dans les environnements aéroportuaires, ainsi que d'autres usages variés allant de la classification à la reconstruction de signaux.

# Chapitre 3

## Préliminaire

### 3.1 Préliminaires Mathématiques pour l'Analyse en Composantes Principales

L'Analyse en Composantes Principales (ACP) repose sur plusieurs concepts mathématiques fondamentaux. Dans cette section, nous aborderons les notions essentielles qui préparent le terrain pour comprendre le processus de réduction de dimension à travers l'ACP.

#### 3.1.1 Statistiques de base et covariance

La covariance entre deux variables  $X$  et  $Y$  mesure comment ces variables varient ensemble. Elle est définie par la formule suivante :

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

où  $X_i$  et  $Y_i$  sont les observations des variables  $X$  et  $Y$ , et  $\bar{X}$  et  $\bar{Y}$  sont leurs moyennes respectives.

Dans un espace multidimensionnel, la covariance peut être représentée par une \*\*matrice de covariance\*\*, où chaque élément  $C_{ij}$  représente la covariance entre les variables  $X_i$  et  $X_j$ . Cette matrice est symétrique et joue un rôle clé dans l'ACP, car elle décrit les relations linéaires entre les différentes dimensions des données.

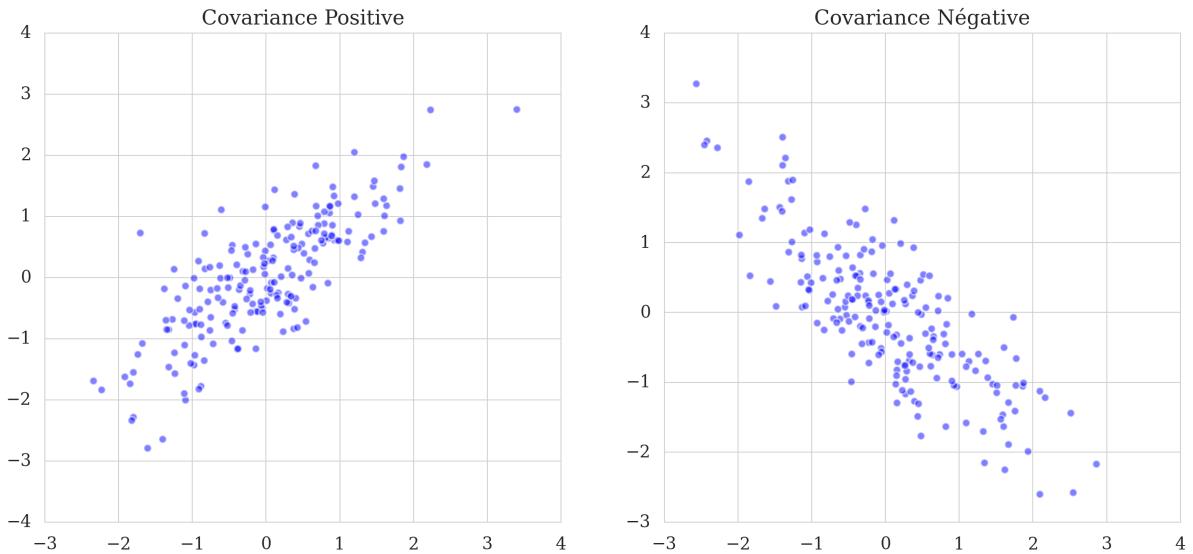


FIGURE 3.1 – Illustration géométrique de la covariance. Lorsque la covariance est positive, les deux variables ont tendance à augmenter ensemble. Si la covariance est négative, elles varient en sens opposé.

La figure 3.1 montre deux ensembles de points où la covariance entre les deux variables est positive (à gauche) et négative (à droite).

### 3.1.2 Diagonalisation et valeurs propres

La diagonalisation d'une matrice de covariance est au cœur de l'ACP. En trouvant les **valeurs propres** et **vecteurs propres** de cette matrice, on peut identifier les directions dans lesquelles les données varient le plus. Chaque valeur propre représente l'importance d'une direction, et le vecteur propre associé représente l'orientation de cette direction dans l'espace.

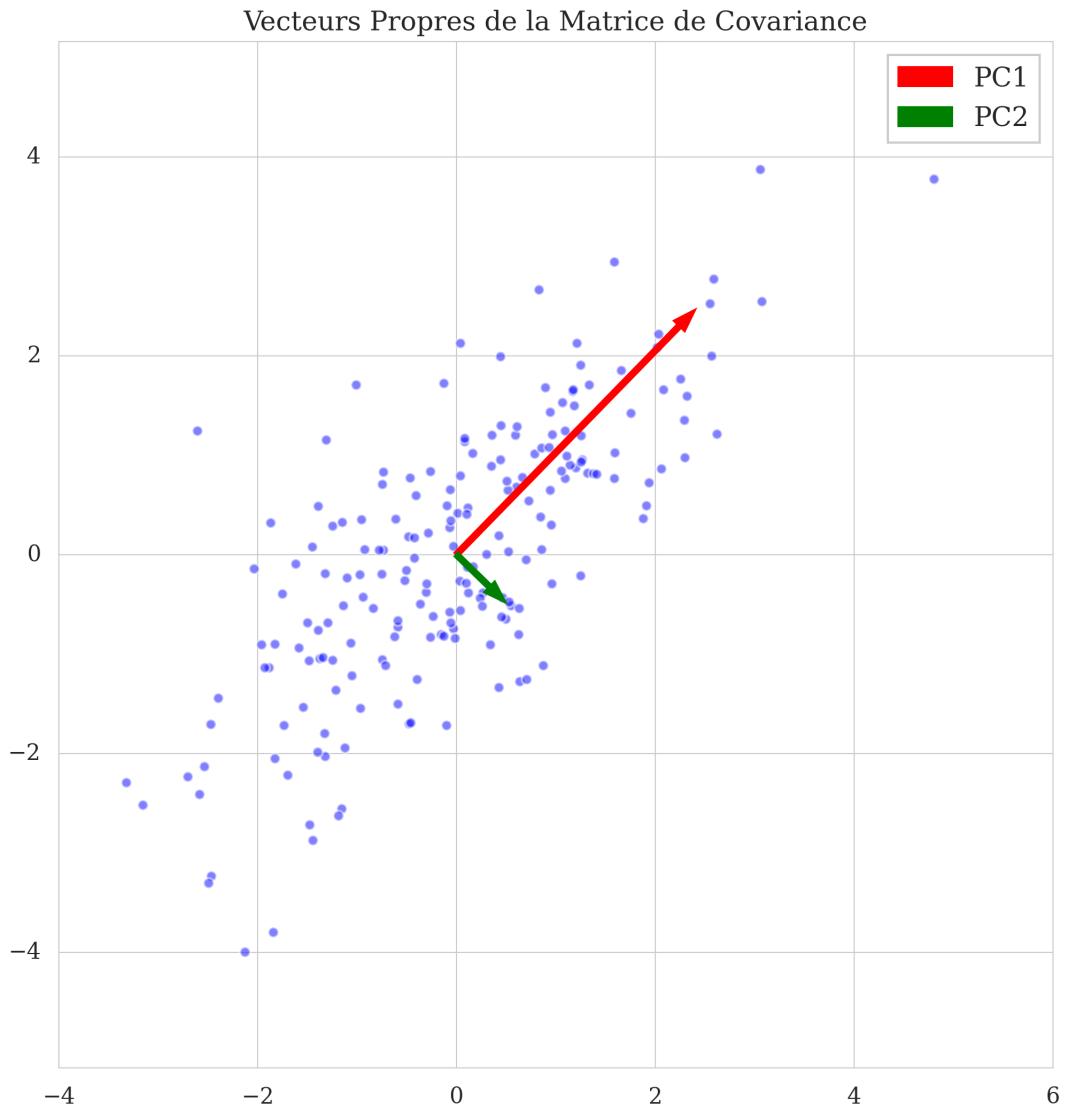


FIGURE 3.2 – Représentation des valeurs propres et des vecteurs propres dans l'espace des données. Les vecteurs propres indiquent les directions de la variance maximale.

La figure 3.2 montre comment les vecteurs propres sont les directions principales dans lesquelles les données sont les plus dispersées. Les valeurs propres associées à ces vecteurs indiquent l'importance de chaque direction.

### 3.1.3 Décomposition en valeurs singulières

La décomposition en valeurs singulières (SVD) factorise toute matrice réelle  $X$  en composantes orthogonales et valeurs singulières.

**Définition (SVD)** Pour toute matrice  $X \in \mathbb{R}^{m \times n}$ , il existe des matrices orthogonales  $U \in \mathbb{R}^{m \times m}$  et  $V \in \mathbb{R}^{n \times n}$  ainsi qu'une matrice diagonale  $\Sigma \in \mathbb{R}^{m \times n}$ ,

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r), \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0,$$

telles que

$$X = U \Sigma V^\top.$$

**Corollaire (calcul de  $X^\top X$ )** On en déduit immédiatement que

$$X^\top X = V \Sigma^2 V^\top = V \operatorname{diag}(\sigma_1^2, \dots, \sigma_r^2) V^\top,$$

ce qui fournit directement les valeurs propres de  $X^\top X$  (utile pour l'ACP).

### 3.1.4 Centrage et normalisation des données

Avant d'appliquer l'ACP, il est crucial de centrer les données. Cela signifie soustraire la moyenne de chaque dimension pour que les données aient une moyenne nulle. Ce centrage est nécessaire pour éviter que les composantes principales ne soient biaisées par des différences dans les échelles des dimensions.

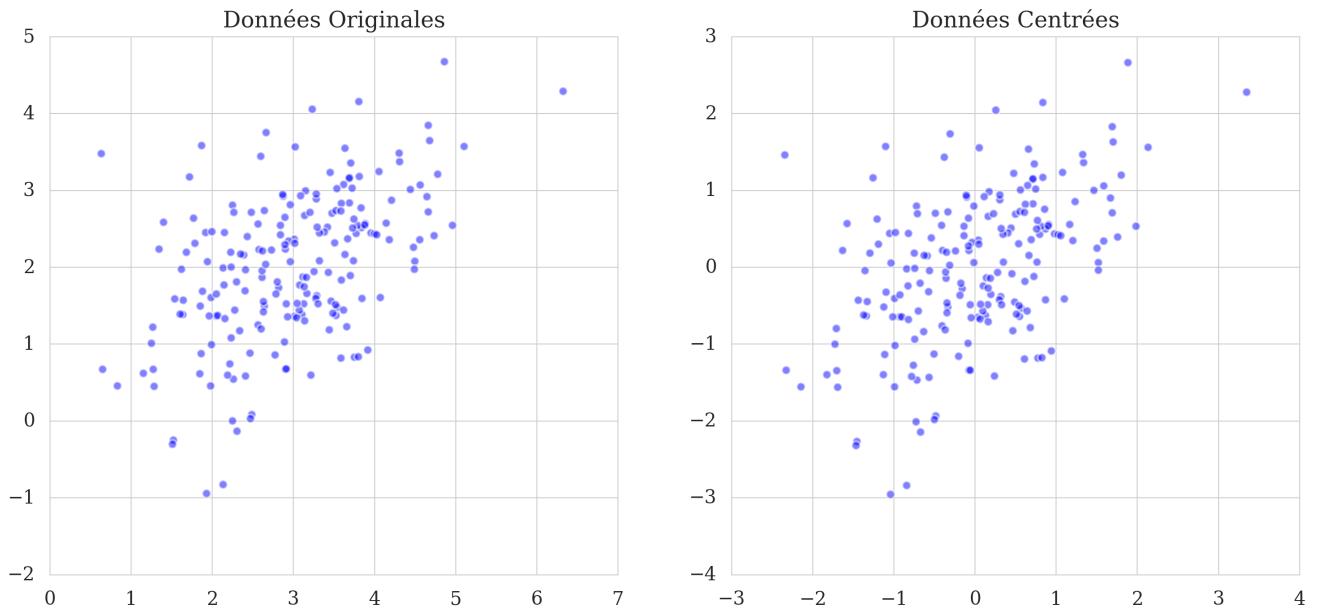


FIGURE 3.3 – Illustration du centrage des données. Chaque dimension est ajustée de manière à avoir une moyenne nulle.

### 3.1.5 Extraction des valeurs propres et vecteurs propres

Pour extraire le spectre de la matrice de covariance  $\Sigma$ , plusieurs méthodes numériques peuvent être utilisées. Voici quelques approches classiques et leur algorithme détaillé :

#### 1. Méthode de la puissance (Power Iteration) :

- (a) Choisir un vecteur initial aléatoire non nul  $v^{(0)}$ .
- (b) Pour  $k = 1, 2, \dots$  jusqu'à convergence, calculer

$$v^{(k)} = \frac{\Sigma v^{(k-1)}}{\|\Sigma v^{(k-1)}\|}.$$

- (c) Lorsque  $\|v^{(k)} - v^{(k-1)}\| < \epsilon$ , estimer la plus grande valeur propre par

$$\lambda_{\max} \approx v^{(k)\top} \Sigma v^{(k)}.$$

2. **Déflation** : Une fois obtenue la plus grande valeur propre  $\lambda_1$  et son vecteur  $v_1$ , soustraire leur contribution :

$$\Sigma^{(1)} = \Sigma - \lambda_1 v_1 v_1^\top.$$

Puis répéter la méthode de la puissance sur  $\Sigma^{(1)}$  afin d'extraire les valeurs et vecteurs suivants.

3. **Algorithme QR** :

- (a) Initialiser une matrice orthonormale  $Q^{(0)}$  de dimension  $p \times p$ .
- (b) Pour  $k = 1, 2, \dots$  :

- i. Décomposer la matrice actuelle :

$$A^{(k-1)} = Q^{(k)} R^{(k)},$$

où  $A^{(0)} = \Sigma$ .

- ii. Mettre à jour :

$$A^{(k)} = R^{(k)} Q^{(k)}.$$

- (c) Lorsque  $A^{(k)}$  converge vers une matrice presque diagonale, les éléments diagonaux de  $A^{(k)}$  approchent les valeurs propres de  $\Sigma$  et les colonnes de  $Q^{(k)}$  fournissent les vecteurs propres correspondants.

4. **Méthode de Jacobi (pour matrices symétriques)** :

- (a) Choisir un couple d'indices  $(i, j)$  avec  $i \neq j$  tel que l'élément hors-diagonale  $\Sigma_{ij}$  soit maximal.
- (b) Calculer l'angle de rotation  $\theta$  qui annule cet élément, puis former la matrice de rotation  $J(i, j, \theta)$ .
- (c) Mettre à jour la matrice par :

$$\Sigma \leftarrow J^\top \Sigma J.$$

- (d) Répéter ces itérations jusqu'à ce que tous les éléments hors-diagonaux soient inférieurs à un seuil fixé.
- (e) La matrice finale sera approximativement diagonale et ses éléments diagonaux représenteront les valeurs propres, les rotations accumulées donnant les vecteurs propres.

**Algorithme :** Extraction des valeurs propres et vecteurs propres via Power Iteration, Déflation, et QR

1. **Power Iteration pour la plus grande valeur propre :**

- (a) Choisir  $v^{(0)}$  arbitraire (non nul).
- (b) Pour  $k = 1, 2, \dots$ , calculer

$$v^{(k)} = \frac{\Sigma v^{(k-1)}}{\|\Sigma v^{(k-1)}\|}.$$

- (c) Convergence lorsque  $\|v^{(k)} - v^{(k-1)}\| < \epsilon$ ; estimer

$$\lambda_1 \approx v^{(k)\top} \Sigma v^{(k)}.$$

2. **Déflation** pour extraire les composantes suivantes :

- (a) Calculer  $\Sigma^{(1)} = \Sigma - \lambda_1 v^{(k)} v^{(k)\top}$ .
- (b) Répéter la méthode de la puissance sur  $\Sigma^{(1)}$  pour obtenir  $\lambda_2$  et  $v_2$ , etc.

3. **Méthode QR** :

- (a) Initialiser  $A^{(0)} = \Sigma$  et une matrice orthonormale  $Q^{(0)}$ .
- (b) Pour  $k = 1, 2, \dots$  :
  - i. Calculer la décomposition  $A^{(k-1)} = Q^{(k)} R^{(k)}$ .
  - ii. Mettre à jour  $A^{(k)} = R^{(k)} Q^{(k)}$ .
- (c) Lorsque  $A^{(k)}$  devient presque diagonale, les diagonales approchent  $\lambda_i$  et les colonnes de  $Q^{(k)}$  approchent  $w_i$ .

FIGURE 3.4 – Extraction des valeurs propres et vecteurs propres (Power Iteration, Déflation, et QR)

---

\*\*Résumé des concepts clés :\*\* - \*\*Covariance\*\* : Comprendre comment les variables sont corrélées entre elles. - \*\*Transformation linéaire\*\* : Réduire la dépendance entre les variables. - \*\*Diagonalisation\*\* : Identifier les directions principales dans les données via les vecteurs propres. - \*\*SVD\*\* : Utiliser la décomposition pour réaliser une réduction dimensionnelle. - \*\*Centrage des données\*\* : Préparer les données avant l'ACP.

---

Cela couvre les concepts nécessaires pour préparer le terrain pour l'ACP dans ton projet. Si tu as besoin de plus de détails ou de modifications, n'hésite pas à me le faire savoir !

# Chapitre 4

## Analyse en Composantes Principales (PCA)

### 4.1 Formulation mathématique

Soit  $X \in \mathbb{R}^{n \times p}$  une matrice de données centrée (chaque colonne a une moyenne nulle). On cherche à projeter les données sur une direction  $w \in \mathbb{R}^p$  de manière à maximiser la variance des projections.

#### 4.1.1 Problème d'optimisation fondamental

L'objectif de la PCA peut être formulé comme un problème d'optimisation quadratique sous contrainte :

$$\begin{aligned} \max_{w \in \mathbb{R}^p} \quad & \text{Var}(Xw) = w^\top \Sigma w \\ \text{sous la contrainte} \quad & \|w\|^2 = 1 \end{aligned} \tag{4.1}$$

où  $\Sigma = \frac{1}{n-1} X^\top X$  est la matrice de covariance des données.

Ce problème est un cas classique de maximisation quadratique sous contrainte quadratique. Sa solution est donnée par le théorème des multiplicateurs de Lagrange : le vecteur  $w$  qui maximise cette expression est le \*\*vecteur propre associé à la plus grande valeur propre de  $\Sigma$ \*\*.

#### 4.1.2 Dérivation complète de l'ACP

La recherche successive des composantes principales peut être formulée comme une série de problèmes d'optimisation. Pour la  $k$ -ième composante, nous cherchons :

$$\begin{aligned} \max_{w_k \in \mathbb{R}^p} \quad & w_k^\top \Sigma w_k \\ \text{sous les contraintes} \quad & \|w_k\|^2 = 1 \\ & w_k^\top w_j = 0, \quad \forall j < k \end{aligned} \tag{4.2}$$

## Rappel sur la méthode des multiplicateurs de Lagrange

Soit un problème d'optimisation sous contrainte classique :

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{sous la contrainte} \quad & g(x) = 0, \end{aligned} \tag{4.3}$$

dans lequel  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  et  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  sont deux fonctions suffisamment régulières.

L'idée clé de la méthode est d'introduire un multiplicateur de Lagrange  $\lambda \in \mathbb{R}$  pour incorporer la contrainte au problème non contraint :

$$\mathcal{L}(x, \lambda) = f(x) - \lambda g(x). \tag{4.4}$$

Les conditions d'optimalité (nécessaires sous hypothèse de régularité) se formalisent par :

$$\begin{cases} \nabla_x \mathcal{L}(x, \lambda) = 0, \\ g(x) = 0. \end{cases} \tag{4.5}$$

Concrètement, la première équation dit que, au point optimal, le gradient de la fonction objectif doit être colinéaire au gradient de la contrainte ; le multiplicateur  $\lambda$  fixe le facteur de proportionnalité.

### Application au calcul de la première composante principale

Dans le cadre de l'analyse en composantes principales (ACP), on cherche un vecteur unitaire  $w$  qui maximise la variance projetée :

$$\max_{w \in \mathbb{R}^d} w^\top \Sigma w \quad \text{sous la contrainte} \quad w^\top w = 1, \tag{4.6}$$

où  $\Sigma$  est la matrice de covariance centrée des données. On pose donc :

$$f(w) = w^\top \Sigma w, \quad g(w) = w^\top w - 1. \tag{4.7}$$

et l'on construit le Lagrangien :

$$\mathcal{L}(w, \lambda) = f(w) - \lambda g(w) = w^\top \Sigma w - \lambda (w^\top w - 1). \tag{4.8}$$

Les conditions d'optimalité s'écrivent alors :

$$\nabla_w \mathcal{L}(w, \lambda) = 0 \implies 2 \Sigma w - 2 \lambda w = 0 \iff \Sigma w = \lambda w. \tag{4.9}$$

On reconnaît l'équation aux valeurs propres de la matrice  $\Sigma$ , avec  $\lambda$  "class" = "mathinline" > la valeur propre associée au vecteur propre  $w$ .

Concluons que pour maximiser la variance projetée, on choisit  $w$  "class" = "mathinline" > l'autovecteur associé à la plus grande valeur propre de  $\Sigma$ .

**Remarque.** La contrainte  $w^\top w = 1$  garantit l'unicité (à signe près) de la solution et évite la possibilité de faire croître indéfiniment  $w$  pour augmenter  $f(w)$  sans limite.

### 4.1.3 Propriétés importantes

1. **Orthogonalité** : Les vecteurs propres  $w_1, \dots, w_p$  forment une base orthonormée de  $\mathbb{R}^p$ .
2. **Variance expliquée** : La variance totale est conservée et égale à la somme des valeurs propres :

$$\sum_{i=1}^p \lambda_i = \text{tr}(\Sigma) = \sum_{j=1}^p \text{Var}(X_j)$$

3. **Proportion de variance** : La proportion de variance expliquée par la  $k$ -ième composante est :

$$\frac{\lambda_k}{\sum_{i=1}^p \lambda_i}$$

**Algorithme :** Calcul des composantes principales

1. Centrer les données :  $X_c = X - \bar{X}$
2. Calculer la matrice de covariance :  $\Sigma = \frac{1}{n-1} X_c^\top X_c$
3. Calculer les valeurs propres  $\lambda_i$  et vecteurs propres  $w_i$  de  $\Sigma$
4. Trier les vecteurs propres par valeurs propres décroissantes
5. Projeter les données :  $Y = X_c W$  où  $W = [w_1 \dots w_k]$

FIGURE 4.1 – Calcul des composantes principales

### 4.1.4 Choix du nombre de composantes

Pour sélectionner le nombre optimal de composantes, plusieurs critères peuvent être utilisés :

- **Critère du coude** : Observer le point d'inflexion dans le scree plot.
- **Variance expliquée cumulée** : Calculer la variance expliquée par les  $k$  premières composantes via la formule :

$$\text{Ratio}_k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i}$$

Par exemple, si les valeurs propres sont  $\lambda_1 = 3$ ,  $\lambda_2 = 2$ ,  $\lambda_3 = 1$ , et  $\lambda_4 = 0.5$ , la variance totale vaut 6.5. Pour  $k = 3$ , on a :

$$\text{Ratio}_3 = \frac{3 + 2 + 1}{6.5} \approx 0.92 \quad (92)$$

On retiendra alors  $k = 3$  si l'objectif est d'expliquer au moins 80 à 90

- **Critère de Kaiser** : Lorsqu'on travaille avec des données standardisées (chaque variable ayant une variance de 1), la matrice de covariance possède une diagonale

égale à 1. Ainsi, chaque valeur propre représente la variance expliquée par une composante et le critère recommande de retenir celles ayant  $\lambda_i > 1$ , c'est-à-dire une variance supérieure à celle d'une variable initiale. Par exemple, si les valeurs propres d'un jeu de données standardisé sont  $\lambda_1 = 2.5$ ,  $\lambda_2 = 1.2$ ,  $\lambda_3 = 0.7$ , et  $\lambda_4 = 0.6$ , seules les deux premières composantes seront retenues car elles satisfont  $\lambda_i > 1$ .

#### 4.1.5 Interprétation géométrique

L'ACP peut être interprétée comme une rotation rigide de l'espace des données qui :

- Aligne le premier axe avec la direction de variance maximale
- Aligne les axes suivants orthogonalement, maximisant la variance résiduelle
- Préserve les distances entre les points

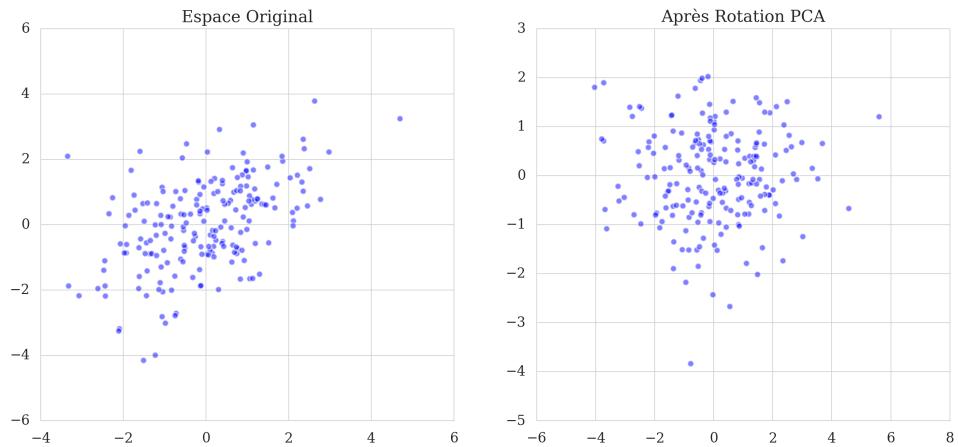


FIGURE 4.2 – Interprétation géométrique de l'ACP comme rotation de l'espace

## 4.2 Limitations de l'ACP classique

L'ACP, malgré sa puissance et sa popularité, présente plusieurs limitations importantes qu'il est crucial de comprendre pour une utilisation appropriée.

### 4.2.1 Linéarité

La limitation la plus fondamentale de l'ACP est son caractère linéaire. L'ACP ne peut capturer que des relations linéaires entre les variables.

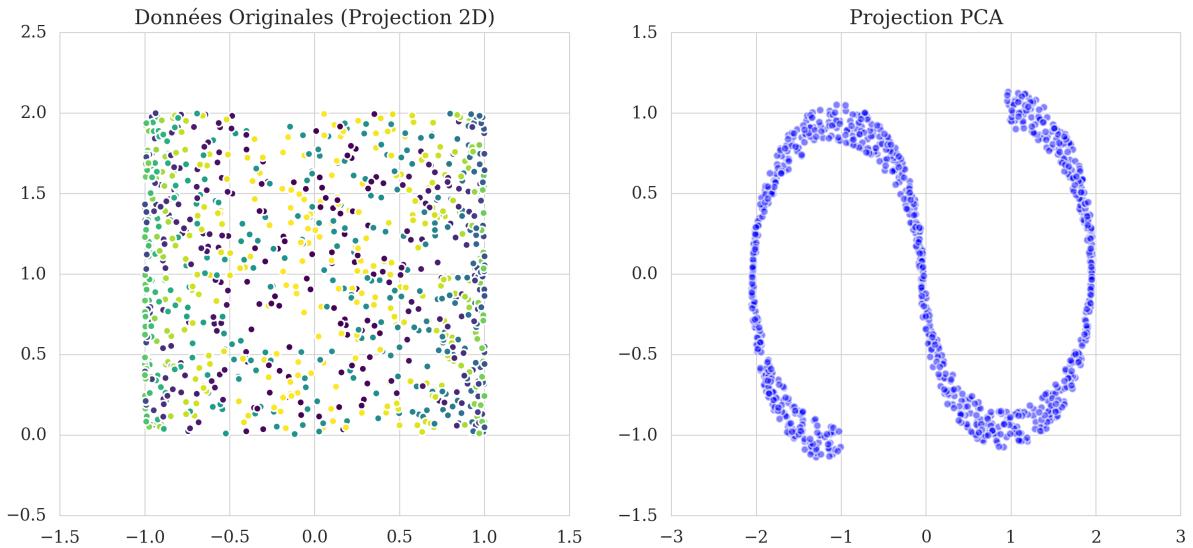


FIGURE 4.3 – Exemple de données non-linéaires où l'ACP échoue

**Exemple concret :** Considérons des données disposées en spirale dans un espace 2D. L'ACP ne pourra pas "dérouler" cette spirale car elle ne peut effectuer que des transformations linéaires.

#### 4.2.2 Sensibilité aux valeurs aberrantes

L'ACP classique est très sensible aux outliers car elle est basée sur la variance, qui elle-même est sensible aux valeurs extrêmes.

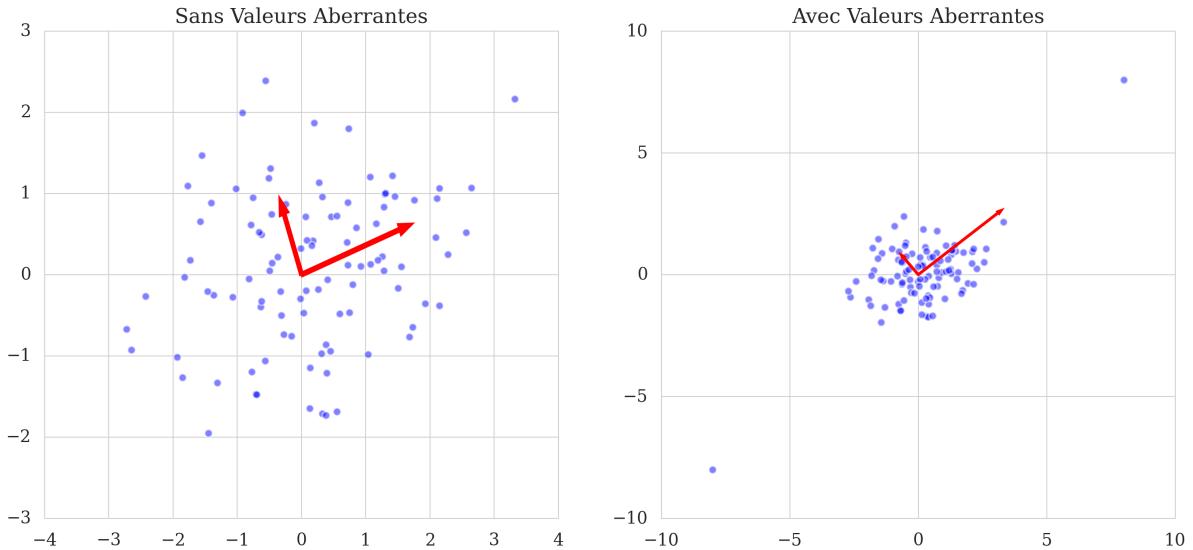


FIGURE 4.4 – Impact des valeurs aberrantes sur les composantes principales

**Exemple numérique :** Dans un jeu de données financières, un seul point aberrant (comme un crash boursier) peut complètement modifier l'orientation des composantes principales.

### 4.2.3 Hypothèse de normalité

L'ACP suppose implicitement que les données suivent une distribution normale multivariée, ce qui n'est pas toujours le cas en pratique.

**Exemple :** Dans l'analyse de données de consommation électrique, les distributions sont souvent asymétriques avec des pics aux heures de pointe.

### 4.2.4 Perte d'interprétabilité

Les composantes principales sont des combinaisons linéaires de toutes les variables originales, ce qui peut rendre leur interprétation difficile.

**Exemple pratique :** Dans une étude médicale avec 100 mesures différentes, chaque composante principale pourrait impliquer toutes ces mesures avec des coefficients non nuls, rendant l'interprétation clinique complexe.

### 4.2.5 Solutions alternatives

Pour remédier à ces limitations, plusieurs variantes ont été développées :

- **Kernel PCA** : Pour traiter les relations non-linéaires
- **Robust PCA** : Pour gérer les valeurs aberrantes
- **Sparse PCA** : Pour améliorer l'interprétabilité
- **Probabilistic PCA** : Pour prendre en compte l'incertitude

Limitation	Impact	Solution
Linéarité	Mauvaise capture des relations non-linéaires	Kernel PCA
Outliers	Distorsion des composantes	Robust PCA
Interprétabilité	Difficulté d'analyse	Sparse PCA
Normalité	Biais sur données non-gaussiennes	Probabilistic PCA

TABLE 4.1 – Résumé des limitations et solutions

Cette compréhension des limitations de l'ACP est essentielle pour choisir la variante appropriée selon le contexte d'application et la nature des données analysées.

# Chapitre 5

## Application pratique de l'ACP

### 5.1 Réduction de dimension et qualité visuelle

L'Analyse en Composantes Principales (ACP) permet de projeter une image haute définition dans un espace latent de dimension beaucoup plus faible tout en préservant l'essentiel de l'information perceptuelle.



FIGURE 5.1 – Effet du nombre de composantes principales  $k$  sur la qualité visuelle.

La Figure 5.1 illustre la dégradation graduelle : on passe de l'image originale (vecteur de dimension  $n$ ) à des reconstructions respectivement sur  $k = 100, 50, 25$  et  $10$  dimensions. La Figure 5.2 synthétise l'évolution conjointe  $\text{taille} \rightarrow \text{qualité}$ . On observe qu'entre  $k = 25$  et  $k = 50$ , le gain en qualité devient marginal alors que la taille double.

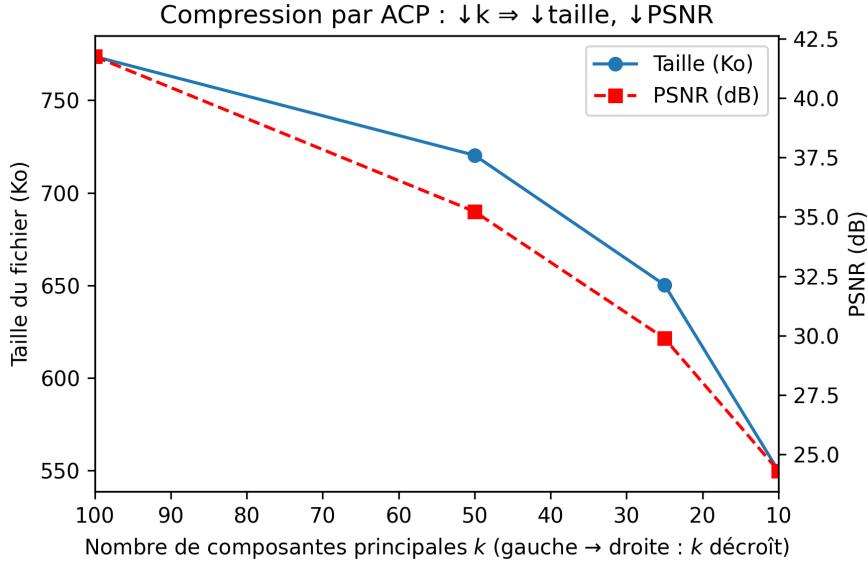


FIGURE 5.2 – Relation entre le nombre de composantes  $k$ , la taille de fichier (Ko) et la qualité de reconstruction (PSNR en dB).

## 5.2 Eigenfaces : Méthode de Reconnaissance Faciale

- On dispose d'un ensemble d'images de visages alignées et centrées.
- Chaque image  $I$  est convertie en vecteur de pixels, puis centrée :

$$I_{\text{centré}} = I - I_{\text{mean}}.$$

- L'ACP est appliquée à la matrice de données de dimension  $n \times p$  (images  $\times$  pixels).

## 5.3 Rôle central de l'ACP dans Eigenfaces

Soit un ensemble de  $n$  images vectorisées  $I_1, \dots, I_n \in \mathbb{R}^p$ , et leur moyenne

$$I_{\text{mean}} = \frac{1}{n} \sum_{i=1}^n I_i.$$

Définissons la matrice de données centrée

$$X = [I_1 - I_{\text{mean}}, \dots, I_n - I_{\text{mean}}] \in \mathbb{R}^{p \times n}.$$

Tout visage  $I$  s'encode alors par les coefficients

$$z = W^\top (I - I_{\text{mean}}) \in \mathbb{R}^K,$$

assurant :

- réduction de dimension :  $p \rightarrow K \ll p$ ,
- filtrage du bruit : on supprime les composantes de faible variance,

- extraction des traits dominants : chaque  $w_i$  capture une “direction” de variation.

# Chapitre 6

## Analyse des résultats

### 6.1 Indice SSIM et application à l'ACP

#### Principe du SSIM

L'objectif du SSIM (*Structural Similarity Index*) est de quantifier la similarité *perçue* entre deux images.

#### Composantes du SSIM

Le SSIM est défini localement, généralement sur une fenêtre de taille  $11 \times 11$ , à partir de trois composantes :

$$\text{SSIM}(x, y) = \underbrace{l(x, y)}_{\text{luminance}} \times \underbrace{c(x, y)}_{\text{contraste}} \times \underbrace{s(x, y)}_{\text{structure}} \quad (6.1)$$

Ces composantes sont calculées par :

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (6.2)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (6.3)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}. \quad (6.4)$$

Le SSIM prend ses valeurs dans l'intervalle  $[-1, 1]$  :

- 1 : images identiques,
- 0 : aucune corrélation perceptible.

#### Application du SSIM à l'ACP

##### 1. Compression par ACP

- Réduction de dimension des images (vectorisation).
- Reconstruction avec les  $k$  premières composantes principales.

##### 2. Évaluation par SSIM

- Calcul du SSIM entre l'image originale et l'image reconstruite.
  - Tracé du SSIM en fonction de  $k$  pour étudier le compromis qualité vs. taille.
3. **Sélection adaptative** Choix du plus petit  $k$  tel que  $\text{SSIM} \geq 0,95$ .
  4. **Critère d'arrêt** Arrêt de l'ajout de composantes dès que  $\Delta\text{SSIM} < \varepsilon$ .

## 6.2 Calcul des moyennes locales $\mu_x$ et $\mu_y$

- Fenêtre courante :  $11 \times 11$  ou  $8 \times 8$  pixels.
  - Pondération par filtre gaussien 2D ( $\sigma \approx 1,5$ ) :
- $$\mu_x(i, j) = \sum_{u,v} w(u, v) X(i + u, j + v), \quad \mu_y(i, j) = \sum_{u,v} w(u, v) Y(i + u, j + v).$$
- Implémentation rapide : convolution ou `cv2.GaussianBlur`.

### Exemple Python (OpenCV)

```
mu_x = cv2.GaussianBlur(X, (11, 11), 1.5)
mu_y = cv2.GaussianBlur(Y, (11, 11), 1.5)
```

## 6.3 Constantes $C_1$ , $C_2$ , $C_3$

### Définitions générales

$$C_1 = (K_1 L)^2, \quad C_2 = (K_2 L)^2, \quad C_3 = \frac{C_2}{2}.$$

### Valeurs usuelles

- $L$  : dynamique (8 bits  $\rightarrow 255$ , normalisé  $\rightarrow 1$ ).
- Biens conseillés :  $K_1 = 0,01$ ,  $K_2 = 0,03$ .
- **Images 8 bits** ( $L = 255$ ) :  $C_1 \approx 6,5025$ ,  $C_2 \approx 58,5225$ ,  $C_3 \approx 29,2612$ .
- **Images normalisées** ( $L = 1$ ) :  $C_1 = 1 \times 10^{-4}$ ,  $C_2 = 9 \times 10^{-4}$ ,  $C_3 = 4,5 \times 10^{-4}$ .

### À retenir

- $\mu_x, \mu_y$  : moyennes locales via filtre gaussien.
- $C_1, C_2, C_3$  : stabilisation des divisions.
- Toujours adapter ces constantes à l'échelle des données.

## 6.4 Interprétation des résultats

- Le PSNR en fonction de  $k$  montre un « point de coude » : au-delà de  $\sim 25$  composantes, les gains se réduisent.

- Le SSIM reste élevé ( $\approx 0,95$ ) pour  $k \geq 50$ , mais chute rapidement pour  $k \leq 10$  (perte de structures).
- L'ACP, en tant qu'approximation de rang- $k$  optimale, surpassé une compression naïve pour un même taux de compression.

## 6.5 Conclusion et Perspectives

### Conclusion et Perspectives

- **Bilan** : réduction jusqu'à  $\sim 90\%$  des variables sans dégradation visuelle majeure.
- **Limites** : sensibilité aux outliers, ne capture que la linéarité.
- **Perspectives** :
  - critères plus robustes pour le choix de  $k$  (coude, validation croisée).
  - méthodes non linéaires (t-SNE, UMAP) pour structures complexes.

# Annexe A

## Kernel PCA

### A.1 Fondements mathématiques du Kernel PCA

Le Kernel PCA est une extension non linéaire de l'ACP classique qui utilise l'astuce du noyau (kernel trick) pour effectuer une ACP dans un espace de caractéristiques de haute dimension.

#### A.1.1 Formulation mathématique

Soit  $\Phi : \mathcal{X} \rightarrow \mathcal{F}$  une fonction qui projette les données de l'espace d'entrée  $\mathcal{X}$  vers un espace de caractéristiques  $\mathcal{F}$  de dimension supérieure. La matrice de covariance dans  $\mathcal{F}$  est donnée par :

$$C = \frac{1}{n} \sum_{i=1}^n \Phi(x_i) \Phi(x_i)^T$$

#### A.1.2 L'astuce du noyau

Au lieu de calculer explicitement  $\Phi$ , on utilise une fonction noyau  $k(x, y) = \langle \Phi(x), \Phi(y) \rangle$ .

Les noyaux courants incluent :

- Noyau gaussien (RBF) :  $k(x, y) = \exp(-\gamma \|x - y\|^2)$
- Noyau polynomial :  $k(x, y) = (\langle x, y \rangle + c)^d$
- Noyau sigmoïde :  $k(x, y) = \tanh(\alpha \langle x, y \rangle + c)$

#### A.1.3 Résolution du problème aux valeurs propres

Dans l'espace de caractéristiques, on résout :

$$\lambda v = Cv$$

qui se transforme en :

$$n\lambda\alpha = K\alpha$$

où  $K$  est la matrice de Gram avec  $K_{ij} = k(x_i, x_j)$

### A.1.4 Centrage dans l'espace de caractéristiques

Pour centrer les données dans  $\mathcal{F}$ , on modifie la matrice de Gram :

$$\tilde{K} = K - \mathbf{1}_n K - K \mathbf{1}_n + \mathbf{1}_n K \mathbf{1}_n$$

où  $\mathbf{1}_n$  est une matrice  $n \times n$  avec tous les éléments égaux à  $\frac{1}{n}$ .

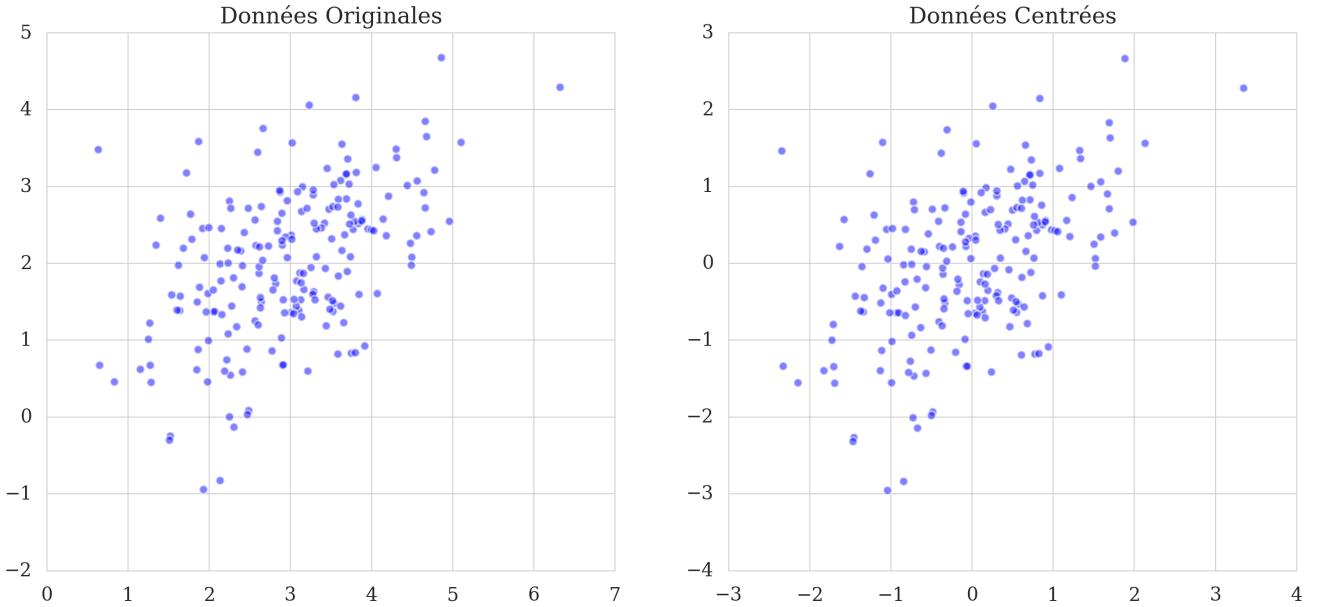


FIGURE A.1 – Illustration du centrage et de la normalisation des données.

### A.1.5 Projection des données

Pour un nouveau point  $x$ , sa projection sur la  $k$ -ème composante principale est :

$$y_k(x) = \sum_{i=1}^n \alpha_i^k k(x_i, x)$$

où  $\alpha^k$  est le  $k$ -ème vecteur propre normalisé de  $\tilde{K}$ .

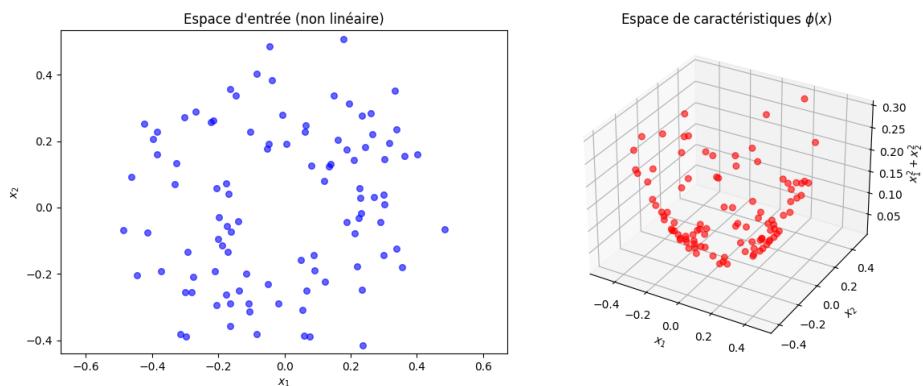


FIGURE A.2 – Illustration du mapping non linéaire via le kernel trick

### A.1.6 Complexité et considérations pratiques

La complexité computationnelle est :

- Construction de la matrice de Gram :  $O(n^2d)$
- Décomposition en valeurs propres :  $O(n^3)$
- Projection d'un nouveau point :  $O(nd)$

où  $n$  est le nombre d'échantillons et  $d$  la dimension d'entrée.

### A.1.7 Théorème de représentation

Le théorème de représentation garantit que les vecteurs propres dans  $\mathcal{F}$  peuvent s'écrire comme combinaisons linéaires des points projetés :

$$v = \sum_{i=1}^n \alpha_i \Phi(x_i)$$

Cette propriété est fondamentale pour l'application pratique du Kernel PCA.

### A.1.8 Choix du noyau

Le choix du noyau dépend de la structure des données :

- RBF : pour des relations localement lisses
- Polynomial : pour des interactions d'ordre supérieur
- Sigmoïde : pour des séparations de type perceptron

Noyau	Expression	Usage typique
RBF	$\exp(-\gamma \ x - y\ ^2)$	Données continues
Polynomial	$(\langle x, y \rangle + c)^d$	Features combinatoires
Sigmoïde	$\tanh(\alpha \langle x, y \rangle + c)$	Classification binaire

TABLE A.1 – Comparaison des noyaux courants

# Bibliographie

- [1] Pearson, K. (1901). L'analyse des données. *Journal de Statistique*.
- [2] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*.
- [3] Jolliffe, I.T. (2002). Principal Component Analysis. Springer.
- [4] Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer.
- [5] Schölkopf, B., et al. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*.

# Webographie

- <https://scikit-learn.org/stable/modules/decomposition.html#pca> - Documentation scikit-learn sur l'ACP
- <https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis> - Discussion approfondie sur l'ACP
- <https://setosa.io/ev/principal-component-analysis/> - Visualisation interactive de l'ACP
- [https://www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](https://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf) - Tutorial complet sur l'ACP
- <https://www.kaggle.com/code/arthurtok/interactive-intro-to-dimensionality-reduction> - Exemples pratiques d'ACP
- <https://www.datasciencecentral.com/principal-component-analysis-for-dimensional-reduction> - Applications en data science
- <https://mathworld.wolfram.com/PrincipalComponentAnalysis.html> - Aspects mathématiques détaillés
- <https://towardsdatascience.com/a-complete-guide-to-principal-component-analysis-558c63c01ff> - Guide complet
- [https://en.wikipedia.org/wiki/Principal\\_component\\_analysis](https://en.wikipedia.org/wiki/Principal_component_analysis) - Article Wikipedia détaillé
- <https://www.youtube.com/watch?v=FgakZw6K1QQ> - Vidéo explicative sur l'ACP