Fondements mathématiques de l'Analyse en Composantes Principales (ACP)

Démonstration, extensions et exemples approfondis

BOUADDI ADAM

Étudiant en génie mécanique

14 novembre 2025

Résumé

Ce document expose de manière entièrement détaillée le fondement mathématique de l'Analyse en Composantes Principales (ACP) et de sa version non linéaire (ACP à noyau, ou *kernel PCA*).

Aucune formule n'est posée sans être reliée à une définition précise. Nous partons des notions élémentaires de moyenne, variance et covariance empiriques, puis nous montrons étape par étape comment ces définitions conduisent à la matrice de covariance $\Sigma = \frac{1}{n} X^\top X$, pour une matrice de données centrée X.

Nous formulons ensuite le problème d'ACP comme un problème d'optimisation quadratique sous contrainte, et nous le résolvons complètement à l'aide des multiplicateurs de Lagrange. Nous démontrons que le problème se ramène à une équation aux valeurs propres $\Sigma w = \lambda w$, et nous établissons le lien avec la décomposition en valeurs singulières (SVD).

Une section de rappels sur les espaces de Hilbert permet ensuite de formuler rigoureusement l'ACP à noyau : nous expliquons comment l'ACP linéaire dans un espace de Hilbert (potentiellement de dimension infinie) se ramène à un problème aux valeurs propres sur la matrice noyau K, via les coefficients α.

Enfin, nous présentons un exemple numérique complet en dimension 2 et discutons les limites de l'ACP linéaire (nature linéaire, sensibilité aux valeurs aberrantes).

Table des matières

Introduction

L'Analyse en Composantes Principales (ACP, ou *Principal Component Analysis*, PCA) est une méthode fondamentale en statistique et en apprentissage automatique pour :

— réduire la dimension des données tout en conservant l'essentiel de l'information (au sens de la variance) ;

— visualiser des données de grande dimension dans des espaces de dimension 2 ou 3 ;

— prétraiter des données avant des tâches de régression ou de classification.

L'idée intuitive est la suivante : on observe des points dans $\mathbb{R}^p$ et on cherche une direction w telle que la projection des points sur w ait une variance maximale. Le résultat central est que les meilleures directions (au sens de la variance maximisée) sont les vecteurs propres de la matrice de covariance des données centrées.

L'objectif de ce document est de donner :

— une **construction précise** de la matrice de covariance à partir des définitions de base ;

— une **démonstration complète** du lien entre maximisation de la variance projetée et vecteurs propres ;

— une **mise en perspective** avec la SVD et les espaces de Hilbert pour l'ACP à noyau.

Nous commençons par rappeler les outils d'algèbre linéaire et de probabilité utilisés.

1 Rappels d'algèbre linéaire

1.1 Espaces vectoriels, produit scalaire et norme

Définition 1.1 (Espace vectoriel réel)**.** Un *espace vectoriel réel* est un ensemble E muni de deux opérations :

— une addition $(x, y) \mapsto x + y$ de $E \times E$ dans E ;

— une multiplication par un scalaire $(\lambda, x) \mapsto \lambda x$ de $\mathbb{R} \times E$ dans E,

satisfaisant les axiomes usuels (associativité, commutativité de l'addition, existence d'un vecteur nul, existence d'opposés, distributivité, etc.).

Dans tout le texte, l'espace vectoriel de base sera $\mathbb{R}^p$.

Définition 1.2 (Produit scalaire canonique sur $\mathbb{R}^p$)**.** Pour $u, v \in \mathbb{R}^p$, le *produit scalaire canonique* est

$$\langle u, v \rangle = u^\top v = \sum_{j=1}^{p} u_j v_j.$$

Définition 1.3 (Norme euclidienne)**.** La *norme euclidienne* associée au produit scalaire canonique est

$$\|u\| = \sqrt{\langle u, u \rangle} = \left(\sum_{j=1}^{p} u_j^2 \right)^{1/2}.$$

1.2 Matrices symétriques et valeurs propres

Définition 1.4 (Matrice symétrique réelle)**.** Une matrice $A \in \mathbb{R}^{p \times p}$ est dite *symétrique* si

$$A^\top = A,$$

c'est-à-dire $A_{jk} = A_{kj}$ pour tous j, k.

Définition 1.5 (Valeur propre et vecteur propre)**.** Un réel λ est une *valeur propre* de A s'il existe un vecteur non nul $w \in \mathbb{R}^p$ tel que

$$Aw = \lambda w.$$

Un tel w est appelé *vecteur propre* associé à λ.

Théorème 1.1 (Théorème spectral pour les matrices symétriques, admis)**.** *Soit $A \in \mathbb{R}^{p\times p}$ une matrice symétrique réelle. Alors :*

(i) toutes les valeurs propres de A sont réelles ;

(ii) il existe une base orthonormée de $\mathbb{R}^p$ formée de vecteurs propres de A ;

(iii) il existe une matrice orthogonale Q ($Q^\top Q = I_p$) et une matrice diagonale réelle Λ telles que

$$A = Q\Lambda Q^\top.$$

Ce résultat est standard en algèbre linéaire et sera utilisé sans preuve.

2 Rappels de probabilités et définitions empiriques

2.1 Variance et covariance empiriques (scalaires)

Dans toute la suite, nous adoptons un point de vue empirique : nous travaillons avec un nombre fini d'observations.

Définition 2.1 (Moyenne empirique, variance empirique)**.** Soient $z_1, \dots, z_n \in \mathbb{R}$ des observations d'une quantité réelle.

— La *moyenne empirique* est

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i.$$

— La *variance empirique* (avec normalisation $1/n$) est

$$\mathrm{Var}_{\mathrm{emp}}(z) = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2.$$

Définition 2.2 (Covariance empirique)**.** Soient deux suites d'observations réelles $(z_i)_{1\le i\le n}$ et $(y_i)_{1\le i\le n}$. La *covariance empirique* (avec normalisation $1/n$) est

$$\mathrm{Cov}_{\mathrm{emp}}(z, y) = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y}),$$

où

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Remarque : en statistique, on rencontre aussi la normalisation par $1/(n-1)$. Ici, pour simplifier l'écriture, on adopte systématiquement $1/n$.

2.2 Vecteurs de données et matrices de données

Nous passons maintenant en dimension $p \geq 1$.

Définition 2.3 (Vecteurs de données, matrice de données brutes)**.** Nous avons n observations, chacune étant un vecteur de $\mathbb{R}^p$:

$$x_i = (x_{i1}, \ldots, x_{ip})^\top \in \mathbb{R}^p, \quad i = 1, \ldots, n.$$

La *matrice de données brutes* est

$$X^{\text{brut}} = \begin{pmatrix} x_1^\top \\ x_2^\top \\ \vdots \\ x_n^\top \end{pmatrix} \in \mathbb{R}^{n \times p}.$$

La colonne j de X^{brut} contient les n observations de la variable j :

$$X_{ij}^{\text{brut}} = x_{ij}.$$

2.3 Centrage des données : définition détaillée

Pour définir la covariance de manière cohérente, on commence par centrer chaque variable.

Définition 2.4 (Moyenne empirique par variable)**.** Pour chaque variable $j \in \{1, \ldots, p\}$, la moyenne empirique est

$$\bar{x}^{(j)} = \frac{1}{n} \sum_{i=1}^{n} x_{ij}.$$

Définition 2.5 (Matrice centrée)**.** La *matrice centrée* $X \in \mathbb{R}^{n \times p}$ est définie par

$$X_{ij} = x_{ij} - \bar{x}^{(j)}, \quad 1 \leq i \leq n,\ 1 \leq j \leq p.$$

Autrement dit, on soustrait à chaque colonne de X^{brut} sa moyenne. Ainsi, la colonne j de X a moyenne empirique nulle :

$$\frac{1}{n} \sum_{i=1}^{n} X_{ij} = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}^{(j)}) = \bar{x}^{(j)} - \bar{x}^{(j)} = 0.$$

Dans toute la suite, pour alléger les notations, nous travaillerons directement avec la matrice centrée X (et non plus X^{brut}).

3 Construction détaillée de la matrice de covariance

3.1 Définition coordonnée par coordonnée

Pour chaque paire de variables (j, k), nous voulons définir la covariance empirique entre la variable j et la variable k. Les observations de ces variables, après centrage, sont :

$$(X_{1j}, \ldots, X_{nj}) \quad \text{et} \quad (X_{1k}, \ldots, X_{nk}).$$

La covariance empirique (définition scalaire) donne donc :

$$\Sigma_{jk} = \text{Cov}_{\text{emp}}\Big(\text{var } j, \text{var } k\Big) = \frac{1}{n} \sum_{i=1}^{n} X_{ij} X_{ik}. \tag{3.1}$$

Définition 3.1 (Matrice de covariance empirique)**.** La *matrice de covariance empirique* $\Sigma \in \mathbb{R}^{p\times p}$ est définie par

$$\Sigma = (\Sigma_{jk})_{1\leq j,k\leq p}, \quad \text{où } \Sigma_{jk} \text{ est donné par (3.1).}$$

Nous allons maintenant montrer explicitement que cette matrice peut s'écrire de manière compacte sous la forme

$$\Sigma = \frac{1}{n} X^\top X,$$

et expliquer ce que cette écriture signifie.

3.2 Calcul explicite de $X^\top X$

La matrice centrée X est de taille $n \times p$:

$$X = \begin{pmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{np} \end{pmatrix}.$$

Sa transposée $X^\top$ est de taille $p \times n$:

$$X^\top = \begin{pmatrix} X_{11} & \cdots & X_{n1} \\ \vdots & & \vdots \\ X_{1p} & \cdots & X_{np} \end{pmatrix}.$$

Le produit $X^\top X$ est donc de taille $p \times p$. Par définition du produit matriciel, le coefficient (j, k) de $X^\top X$ est

$$\begin{aligned}(X^\top X)_{jk} &= \sum_{i=1}^{n} (X^\top)_{ji}\, X_{ik} \\ &= \sum_{i=1}^{n} X_{ij} X_{ik},\end{aligned}$$

car $(X^\top)_{ji} = X_{ij}$.

En comparant avec la définition (3.1), on obtient :

$$\Sigma_{jk} = \frac{1}{n} \sum_{i=1}^{n} X_{ij} X_{ik} = \frac{1}{n} (X^\top X)_{jk}.$$

Autrement dit, pour chaque paire (j, k),

$$\Sigma_{jk} = \left(\frac{1}{n} X^\top X \right)_{jk}.$$

Par égalité coordonnée par coordonnée, nous avons donc

$$\boxed{\Sigma = \frac{1}{n} X^\top X \in \mathbb{R}^{p\times p}.} \tag{3.2}$$

3.3 Symétrie et semi-définie positivité de Σ

Nous justifions maintenant deux propriétés importantes de Σ.

Proposition 3.1. *La matrice $\Sigma = \frac{1}{n}X^\top X$ vérifie :*

(i) Σ est symétrique : $\Sigma^\top = \Sigma$;

(ii) Σ est semi-définie positive : pour tout $w \in \mathbb{R}^p$,

$$w^\top \Sigma w \geq 0.$$

Démonstration. (i) Comme $(AB)^\top = B^\top A^\top$ pour des matrices A, B compatibles, on a

$$\Sigma^\top = \left(\frac{1}{n}X^\top X\right)^\top = \frac{1}{n}(X^\top X)^\top = \frac{1}{n}X^\top (X^\top)^\top = \frac{1}{n}X^\top X = \Sigma.$$

(ii) Soit $w \in \mathbb{R}^p$. On calcule

$$w^\top \Sigma w = w^\top \left(\frac{1}{n}X^\top X\right) w = \frac{1}{n}w^\top X^\top X w.$$

Par associativité,

$$w^\top X^\top X w = (Xw)^\top (Xw).$$

Or, pour tout vecteur u, $u^\top u = \|u\|^2 \geq 0$. Donc

$$w^\top \Sigma w = \frac{1}{n}\|Xw\|^2 \geq 0.$$

□

4 Variance de la projection sur une direction

4.1 Définition de la projection

Soit $w \in \mathbb{R}^p$ une direction. Pour chaque observation $x_i \in \mathbb{R}^p$, on considère la projection scalaire

$$z_i = w^\top x_i.$$

On regroupe ces valeurs dans un vecteur

$$z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \in \mathbb{R}^n.$$

Proposition 4.1. *En notation matricielle, on a*

$$z = Xw.$$

Démonstration. Par définition,

$$(Xw)_i = \sum_{j=1}^{p} X_{ij} w_j = x_i^\top w = w^\top x_i = z_i,$$

pour tout $i = 1, \ldots, n$. Donc Xw et z ont les mêmes composantes, d'où $z = Xw$. □

4.2 Calcul détaillé de la variance projetée

Nous supposons que X est déjà centrée, donc les moyennes de chaque variable sont nulles. Montrons que la moyenne empirique des z_i est également nulle.

Proposition 4.2. *Si les données x_i sont centrées, alors*

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = 0.$$

Démonstration. En remplaçant z_i par sa définition :

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n} w^\top x_i = w^\top \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right).$$

Mais, comme les données sont centrées, la moyenne empirique est

$$\frac{1}{n}\sum_{i=1}^{n} x_i = 0,$$

d'où

$$\bar{z} = w^\top 0 = 0.$$

□

La variance empirique de (z_i) est donc

$$\mathrm{Var}_{\mathrm{emp}}(z) = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2.$$

Proposition 4.3. *Pour toute direction $w \in \mathbb{R}^p$,*

$$\mathrm{Var}_{\mathrm{emp}}(z) = w^\top \Sigma w. \tag{4.1}$$

Démonstration. On remplace $z_i = w^\top x_i$:

$$\begin{aligned}\mathrm{Var}_{\mathrm{emp}}(z) &= \frac{1}{n}\sum_{i=1}^{n}(w^\top x_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(w^\top x_i)(w^\top x_i).\end{aligned}$$

Pour chaque i, on peut écrire

$$(w^\top x_i)(w^\top x_i) = w^\top x_i x_i^\top w,$$

car un scalaire est invariant par transposition. Donc

$$\mathrm{Var}_{\mathrm{emp}}(z) = \frac{1}{n}\sum_{i=1}^{n} w^\top x_i x_i^\top w = w^\top \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top\right) w.$$

Or, par la construction de la section précédente,

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top = \frac{1}{n} X^\top X = \Sigma.$$

D'où

$$\mathrm{Var}_{\mathrm{emp}}(z) = w^\top \Sigma w.$$

□

La formule (4.1) relie donc directement la variance projetée et la matrice de covariance.

5 Formulation du problème d'ACP comme optimisation

5.1 Problème de la première composante principale

L'idée centrale de l'ACP est :

> Trouver la direction w (de norme 1) qui maximise la variance des données projetées sur w.

D'après la section précédente, cette variance est $w^\top \Sigma w$. La contraintes « norme 1 » s'écrit

$$\|w\|^2 = w^\top w = 1.$$

Définition 5.1 (Problème d'optimisation de la première composante principale)**.** La première composante principale est obtenue en résolvant :

$$\max_{w \in \mathbb{R}^p} w^\top \Sigma w \quad \text{sous la contrainte} \quad w^\top w = 1. \tag{5.1}$$

Nous allons maintenant résoudre ce problème de manière détaillée à l'aide des multiplicateurs de Lagrange.

6 Résolution par multiplicateurs de Lagrange

6.1 Construction de la fonction de Lagrange

On introduit un multiplicateur scalaire $\lambda \in \mathbb{R}$ pour la contrainte $w^\top w = 1$, et on définit la fonction de Lagrange

$$\mathcal{L}(w, \lambda) = w^\top \Sigma w - \lambda(w^\top w - 1). \tag{6.1}$$

L'objectif est de trouver les points critiques de $\mathcal{L}$, c'est-à-dire les paires (w, λ) pour lesquelles :

$$\nabla_w \mathcal{L}(w, \lambda) = 0 \quad \text{et} \quad \frac{\partial \mathcal{L}}{\partial \lambda}(w, \lambda) = 0.$$

6.2 Dérivée par rapport à λ

La dérivée partielle par rapport à λ est immédiate :

$$\frac{\partial \mathcal{L}}{\partial \lambda}(w, \lambda) = -(w^\top w - 1).$$

L'équation

$$\frac{\partial \mathcal{L}}{\partial \lambda}(w, \lambda) = 0$$

donne donc

$$w^\top w - 1 = 0 \quad \Leftrightarrow \quad w^\top w = 1.$$

La contrainte est ainsi réinjectée dans les conditions d'optimalité.

6.3 Gradient par rapport à w : calcul coordonnée par coordonnée

Écrivons $w = (w_1, \dots, w_p)^\top$. Nous calculons d'abord $\frac{\partial}{\partial w_\ell}(w^\top \Sigma w)$.
Écrivons cette quantité sous forme de somme :

$$w^\top \Sigma w = \sum_{j=1}^{p} \sum_{k=1}^{p} w_j \Sigma_{jk} w_k.$$

La dérivée partielle par rapport à w_ℓ est :

$$\frac{\partial}{\partial w_\ell} \left(\sum_{j,k} w_j \Sigma_{jk} w_k \right) = \sum_{j,k} \frac{\partial}{\partial w_\ell} \big(w_j \Sigma_{jk} w_k \big).$$

On distingue trois cas :
— si $j = \ell$ et $k \neq \ell$, le terme est $w_\ell \Sigma_{\ell k} w_k$, dont la dérivée vaut $\Sigma_{\ell k} w_k$;
— si $j \neq \ell$ et $k = \ell$, le terme est $w_j \Sigma_{j\ell} w_\ell$, dont la dérivée vaut $w_j \Sigma_{j\ell}$;
— si $j = \ell$ et $k = \ell$, le terme est $w_\ell \Sigma_{\ell\ell} w_\ell = \Sigma_{\ell\ell} w_\ell^2$, dont la dérivée vaut $2\Sigma_{\ell\ell} w_\ell$;
— si $j \neq \ell$ et $k \neq \ell$, la dérivée est 0.
On peut rassembler ces termes :

$$\frac{\partial}{\partial w_\ell}(w^\top \Sigma w) = \sum_{k=1}^{p} \Sigma_{\ell k} w_k + \sum_{j=1}^{p} w_j \Sigma_{j\ell}.$$

Comme Σ est symétrique, on a $\Sigma_{j\ell} = \Sigma_{\ell j}$, donc :

$$\sum_{j=1}^{p} w_j \Sigma_{j\ell} = \sum_{j=1}^{p} \Sigma_{\ell j} w_j = \sum_{k=1}^{p} \Sigma_{\ell k} w_k.$$

On obtient finalement

$$\frac{\partial}{\partial w_\ell}(w^\top \Sigma w) = 2 \sum_{k=1}^{p} \Sigma_{\ell k} w_k.$$

En écriture vectorielle, cela signifie :

$$\nabla_w (w^\top \Sigma w) = 2\Sigma w.$$

De manière similaire, on a

$$w^\top w = \sum_{j=1}^{p} w_j^2 \quad \Rightarrow \quad \frac{\partial}{\partial w_\ell}(w^\top w) = 2w_\ell,$$

donc

$$\nabla_w (w^\top w) = 2w.$$

Ainsi, le gradient de la Lagrangienne (6.1) est

$$\nabla_w \mathcal{L}(w, \lambda) = 2\Sigma w - 2\lambda w.$$

6.4 Équation aux valeurs propres

La condition

$$\nabla_w \mathcal{L}(w, \lambda) = 0$$

donne

$$2\Sigma w - 2\lambda w = 0 \quad \Leftrightarrow \quad \Sigma w = \lambda w.$$

Nous avons donc obtenu l'équation

$$\Sigma w = \lambda w, \tag{6.2}$$

accompagnée de la contrainte

$$w^\top w = 1. \tag{6.3}$$

L'équation (6.2) est exactement l'équation aux valeurs propres pour la matrice Σ. Les points critiques de (5.1) sont donc des vecteurs propres unitaires de Σ.

6.5 Choix de la solution maximale

Notons les valeurs propres de Σ par

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0,$$

et $w^{(1)}, \ldots, w^{(p)}$ des vecteurs propres associés, normalisés ($\|w^{(j)}\| = 1$).

D'après le Théorème spectral 1.1, il existe une matrice orthogonale Q telle que

$$\Sigma = Q\Lambda Q^\top,$$

où $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ et les colonnes de Q sont les vecteurs propres $w^{(j)}$.

Théorème 6.1 (Solution du problème d'ACP)**.** *Les solutions du problème*

$$\max_{\|w\|=1} w^\top \Sigma w$$

sont les vecteurs propres unitaires de Σ associés à la plus grande valeur propre λ_1. La valeur maximale de $w^\top \Sigma w$ est λ_1.

Démonstration. Pour tout $w \neq 0$, posons $y = Q^\top w$. Comme Q est orthogonale,

$$\|y\| = \|w\|.$$

On a

$$\begin{aligned} w^\top \Sigma w &= w^\top Q \Lambda Q^\top w \\ &= (Q^\top w)^\top \Lambda (Q^\top w) \\ &= y^\top \Lambda y. \end{aligned}$$

Comme Λ est diagonale, cela donne

$$y^\top \Lambda y = \sum_{j=1}^{p} \lambda_j y_j^2.$$

De plus,

$$\|w\|^2 = \|y\|^2 = \sum_{j=1}^{p} y_j^2.$$

Sous la contrainte $\|w\| = 1$, on a $\sum_j y_j^2 = 1$. La quantité

$$w^\top \Sigma w = \sum_{j=1}^{p} \lambda_j y_j^2$$

est donc une combinaison convexe des λ_j. Elle est maximum lorsque tout le poids est sur la plus grande valeur propre, c'est-à-dire lorsque y est colinéaire à le vecteur de base canonique associé à λ_1, ce qui signifie que w est colinéaire à $w^{(1)}$.

Comme on impose $\|w\| = 1$, la solution maximale est $w = w^{(1)}$ (ou son opposé, ce qui ne change pas la variance). La valeur maximale est alors

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

□

Ainsi, la première composante principale est le vecteur propre de Σ associé à la plus grande valeur propre.

Les composantes suivantes se construisent en ajoutant des contraintes d'orthogonalité (on maximise $w^\top \Sigma w$ sous les contraintes $\|w\| = 1$ et $w \perp w^{(1)}, \ldots, w^{(k-1)}$), ce qui conduit aux vecteurs propres associés à $\lambda_2, \lambda_3, \ldots$.

7 Exemple numérique détaillé en dimension 2

Nous donnons maintenant un exemple complet où tous les calculs sont faits jusqu'au bout.

7.1 Données brutes

Considérons les 4 points suivants dans le plan :

$$(x, y)_1 = (0, 0), \quad (x, y)_2 = (1, 1), \quad (x, y)_3 = (2, 1), \quad (x, y)_4 = (3, 4).$$

La matrice de données brutes est donc

$$X^{\text{brut}} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 2 & 1 \\ 3 & 4 \end{pmatrix}.$$

7.2 Centrage des colonnes

Calculons les moyennes de chaque variable :

$$\bar{x} = \frac{0+1+2+3}{4} = \frac{6}{4} = 1{,}5, \qquad \bar{y} = \frac{0+1+1+4}{4} = \frac{6}{4} = 1{,}5.$$

On soustrait ces moyennes à chaque ligne :

$$X = \begin{pmatrix} 0-1{,}5 & 0-1{,}5 \\ 1-1{,}5 & 1-1{,}5 \\ 2-1{,}5 & 1-1{,}5 \\ 3-1{,}5 & 4-1{,}5 \end{pmatrix} = \begin{pmatrix} -1{,}5 & -1{,}5 \\ -0{,}5 & -0{,}5 \\ 0{,}5 & -0{,}5 \\ 1{,}5 & 2{,}5 \end{pmatrix}.$$

7.3 Calcul de $X^\top X$ et de Σ

La transposée est

$$X^\top = \begin{pmatrix} -1{,}5 & -0{,}5 & 0{,}5 & 1{,}5 \\ -1{,}5 & -0{,}5 & -0{,}5 & 2{,}5 \end{pmatrix}.$$

Le produit $X^\top X$ est une matrice 2×2 dont les coefficients sont :

$$(X^\top X)_{jk} = \sum_{i=1}^{4} X_{ij} X_{ik}.$$

Coefficient $(1,1)$**.**

$$(X^\top X)_{11} = \sum_{i=1}^{4} X_{i1}^2 = (-1{,}5)^2 + (-0{,}5)^2 + (0{,}5)^2 + (1{,}5)^2 = 2{,}25 + 0{,}25 + 0{,}25 + 2{,}25 = 5.$$

Coefficient $(2,2)$**.**

$$(X^\top X)_{22} = \sum_{i=1}^{4} X_{i2}^2 = (-1{,}5)^2 + (-0{,}5)^2 + (-0{,}5)^2 + (2{,}5)^2 = 2{,}25 + 0{,}25 + 0{,}25 + 6{,}25 = 9.$$

Coefficients $(1,2)$ **et** $(2,1)$**.**

$$(X^\top X)_{12} = \sum_{i=1}^{4} X_{i1} X_{i2} = (-1{,}5)(-1{,}5) + (-0{,}5)(-0{,}5) + (0{,}5)(-0{,}5) + (1{,}5)(2{,}5).$$

On obtient :

$$2{,}25 + 0{,}25 - 0{,}25 + 3{,}75 = 6.$$

Par symétrie, $(X^\top X)_{21} = 6$.

Ainsi,

$$X^\top X = \begin{pmatrix} 5 & 6 \\ 6 & 9 \end{pmatrix}.$$

La matrice de covariance est

$$\Sigma = \frac{1}{4} X^\top X = \frac{1}{4} \begin{pmatrix} 5 & 6 \\ 6 & 9 \end{pmatrix} = \begin{pmatrix} 1{,}25 & 1{,}50 \\ 1{,}50 & 2{,}25 \end{pmatrix}.$$

7.4 Valeurs propres et vecteurs propres

Le polynôme caractéristique de Σ est

$$\det(\Sigma - \lambda I_2) = \begin{vmatrix} 1{,}25 - \lambda & 1{,}50 \\ 1{,}50 & 2{,}25 - \lambda \end{vmatrix} = (1{,}25 - \lambda)(2{,}25 - \lambda) - 1{,}50^2.$$

Développons :

$$(1{,}25 - \lambda)(2{,}25 - \lambda) = 1{,}25 \cdot 2{,}25 - 1{,}25\lambda - 2{,}25\lambda + \lambda^2.$$

Le produit $1{,}25 \cdot 2{,}25 = 2{,}8125$, donc

$$(1{,}25 - \lambda)(2{,}25 - \lambda) = \lambda^2 - 3{,}5\lambda + 2{,}8125.$$

Par ailleurs, $1{,}50^2 = 2{,}25$, ce qui donne

$$\det(\Sigma - \lambda I_2) = \lambda^2 - 3{,}5\lambda + (2{,}8125 - 2{,}25) = \lambda^2 - 3{,}5\lambda + 0{,}5625.$$

On résout l'équation

$$\lambda^2 - 3{,}5\lambda + 0{,}5625 = 0.$$

Le discriminant est

$$\Delta = 3{,}5^2 - 4 \cdot 0{,}5625 = 12{,}25 - 2{,}25 = 10.$$

Les solutions sont

$$\lambda_{1,2} = \frac{3{,}5 \pm \sqrt{10}}{2} \approx 3{,}33 \quad \text{et} \quad 0{,}17.$$

Pour λ_1, on résout

$$(\Sigma - \lambda_1 I_2)w = 0.$$

Cela donne le système

$$\begin{cases} (1{,}25 - \lambda_1)w_1 + 1{,}50 w_2 = 0, \\ 1{,}50 w_1 + (2{,}25 - \lambda_1)w_2 = 0. \end{cases}$$

Ces deux équations sont linéairement dépendantes (par définition d'un vecteur propre), il suffit donc d'en utiliser une, par exemple la première :

$$(1{,}25 - \lambda_1)w_1 + 1{,}50 w_2 = 0.$$

Avec $\lambda_1 \approx 3{,}33$, on a $1{,}25 - \lambda_1 \approx -2{,}08$, donc

$$-2{,}08 w_1 + 1{,}50 w_2 = 0 \quad \Rightarrow \quad w_2 \approx \frac{2{,}08}{1{,}50} w_1 \approx 1{,}39 w_1.$$

On peut choisir $w_1 = 1$, ce qui donne un vecteur propre non normalisé :

$$u^{(1)} = \begin{pmatrix} 1 \\ 1{,}39 \end{pmatrix}.$$

Sa norme est

$$\|u^{(1)}\|^2 = 1^2 + 1{,}39^2 \approx 1 + 1{,}93 = 2{,}93, \quad \|u^{(1)}\| \approx 1{,}71.$$

Le vecteur propre unitaire est

$$w^{(1)} = \frac{1}{\|u^{(1)}\|} \begin{pmatrix} 1 \\ 1{,}39 \end{pmatrix} \approx \begin{pmatrix} 0{,}585 \\ 0{,}811 \end{pmatrix}.$$

On procède de manière analogue pour λ_2 et on trouve

$$w^{(2)} \approx \begin{pmatrix} -0{,}811 \\ 0{,}585 \end{pmatrix}.$$

7.5 Variance expliquée

La variance totale (somme des variances sur toutes les directions orthogonales) est la trace de Σ, c'est-à-dire la somme des valeurs propres :

$$\lambda_1 + \lambda_2 \approx 3{,}33 + 0{,}17 = 3{,}50.$$

La proportion de variance expliquée par la première composante est :

$$\mathrm{PVE}_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2} \approx \frac{3{,}33}{3{,}50} \approx 0{,}95.$$

Ici, une seule composante explique déjà environ 95% de la variance totale, ce qui illustre l'intérêt de la réduction de dimension par ACP.

8 ACP et décomposition en valeurs singulières (SVD)

8.1 Rappel de la SVD (théorème admis)

Théorème 8.1 (Décomposition en valeurs singulières, admis)**.** *Pour toute matrice réelle $X \in \mathbb{R}^{n\times p}$ de rang r, il existe des matrices*

— *$U \in \mathbb{R}^{n\times r}$, à colonnes orthonormées ($U^\top U = I_r$) ;*

— *$V \in \mathbb{R}^{p\times r}$, à colonnes orthonormées ($V^\top V = I_r$) ;*

— *$\Delta \in \mathbb{R}^{r\times r}$, diagonale à coefficients $\delta_1 \geq \cdots \geq \delta_r > 0$,*

telles que

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

Les scalaires δ_j sont les valeurs singulières *de X.*

8.2 Lien détaillé avec la matrice de covariance

Partons de la SVD de la matrice centrée X :

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

Calculons $X^\top X$:

$$X^\top X = (U\Delta V^\top)^\top (U\Delta V^\top).$$

En utilisant $(AB)^\top = B^\top A^\top$,

$$(U\Delta V^\top)^\top = V\Delta^\top U^\top.$$

Donc

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

Comme les colonnes de U sont orthonormées, on a $U^\top U = I_r$. De plus, Δ est diagonale, donc $\Delta^\top = \Delta$. On obtient :

$$X^\top X = V\Delta^2 V^\top.$$

La matrice de covariance est

$$\Sigma = \frac{1}{n} X^\top X = V\left(\frac{1}{n}\Delta^2\right)V^\top.$$

Proposition 8.1. *Les colonnes de V sont des vecteurs propres de Σ, avec pour valeurs propres*

$$\lambda_j = \frac{\delta_j^2}{n}.$$

Démonstration. Notons v_j la j-ème colonne de V. On a

$$\Sigma V = V\left(\frac{1}{n}\Delta^2\right)V^\top V = V\left(\frac{1}{n}\Delta^2\right),$$

car $V^\top V = I_r$. La j-ème colonne de ΣV est donc

$$\Sigma v_j = \frac{\delta_j^2}{n} v_j.$$

Chaque v_j est donc un vecteur propre de Σ de valeur propre δ_j^2/n. □

Ainsi, l'ACP linéaire est directement reliée à la SVD de la matrice de données centrée X.

9 Rappels sur les espaces de Hilbert

Pour discuter de l'ACP à noyau, nous avons besoin du cadre des espaces de Hilbert, qui généralise l'ACP à des espaces de dimension potentiellement infinie.

9.1 Produit scalaire et norme

Définition 9.1 (Produit scalaire sur un espace vectoriel réel)**.** Soit H un espace vectoriel réel. Un *produit scalaire* sur H est une application

$$\langle \cdot, \cdot \rangle : H \times H \to \mathbb{R}$$

telle que, pour tous $u, v, w \in H$ et $\alpha, \beta \in \mathbb{R}$:

(i) $\langle u, v\rangle = \langle v, u\rangle$ (symétrie) ;

(ii) $\langle \alpha u + \beta v, w\rangle = \alpha\langle u, w\rangle + \beta\langle v, w\rangle$ (bilinéarité) ;

(iii) $\langle u, u\rangle \geq 0$ et $\langle u, u\rangle = 0 \Rightarrow u = 0$ (positivité).

Définition 9.2 (Norme induite)**.** La norme induite par un produit scalaire est

$$\|u\| = \sqrt{\langle u, u\rangle}.$$

Définition 9.3 (Espace préhilbertien)**.** Un *espace préhilbertien réel* est un espace vectoriel réel muni d'un produit scalaire.

9.2 Complétude et espaces de Hilbert

Définition 9.4 (Suite de Cauchy)**.** Dans un espace normé $(H, \|\cdot\|)$, une suite (u_n) est *de Cauchy* si

$$\forall \varepsilon > 0,\ \exists N \text{ tel que } \forall m, n \geq N,\ \|u_n - u_m\| < \varepsilon.$$

Définition 9.5 (Espace de Hilbert)**.** Un *espace de Hilbert réel* est un espace préhilbertien réel complet pour la norme induite, c'est-à-dire dans lequel toute suite de Cauchy converge.

9.3 Orthogonalité et projection orthogonale

Définition 9.6 (Orthogonalité)**.** Dans $(H, \langle\cdot,\cdot\rangle)$, deux vecteurs $u, v \in H$ sont *orthogonaux* si

$$\langle u, v\rangle = 0.$$

On note alors $u \perp v$.

Définition 9.7 (Sous-espace engendré)**.** Pour un ensemble $S \subset H$, le *sous-espace engendré par* S est

$$\operatorname{span}(S) = \left\{ \sum_{k=1}^{m} \alpha_k s_k \ : \ m \in \mathbb{N}^*,\ \alpha_k \in \mathbb{R},\ s_k \in S \right\}.$$

Définition 9.8 (Complément orthogonal)**.** Si F est un sous-espace de H, son *complément orthogonal* est

$$F^{\perp} = \{h \in H \ : \ \forall f \in F,\ \langle h, f\rangle = 0\}.$$

Théorème 9.1 (Projection orthogonale sur un sous-espace de dimension finie)**.** *Soit $(H, \langle\cdot,\cdot\rangle)$ un espace de Hilbert réel et F un sous-espace de dimension finie de H. Alors :*

(i) pour tout $h \in H$, il existe un unique couple $(f, g) \in F \times F^{\perp}$ tel que $h = f + g$;

(ii) f est appelé la projection orthogonale *de h sur F.*

10 ACP à noyau (kernel PCA)

Nous pouvons maintenant formaliser l'ACP à noyau dans ce cadre hilbertien.

10.1 Application de caractéristiques et noyau

Soient $x_1, \ldots, x_n \in \mathbb{R}^p$ les observations d'origine. On considère une application (éventuellement non linéaire)

$$\Phi : \mathbb{R}^p \to \mathcal{H},$$

où $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ est un espace de Hilbert réel. On note

$$\phi_i = \Phi(x_i) \in \mathcal{H}.$$

Définition 10.1 (Noyau)**.** On définit le *noyau* $K : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ associé à Φ par

$$K(x, y) = \langle \Phi(x), \Phi(y)\rangle_{\mathcal{H}}.$$

Définition 10.2 (Matrice noyau)**.** La *matrice noyau* (ou matrice de Gram) est

$$K = (K_{ij})_{1 \le i,j \le n} \quad \text{avec} \quad K_{ij} = K(x_i, x_j) = \langle \phi_i, \phi_j\rangle_{\mathcal{H}}.$$

10.2 Centrage dans l'espace de Hilbert

Comme dans l'ACP linéaire, nous supposons que les données sont centrées dans l'espace de Hilbert :

$$\frac{1}{n} \sum_{i=1}^{n} \phi_i = 0.$$

En pratique, cela peut se faire en transformant K par une opération de centrage (formule classique de centrage de la matrice de Gram), mais nous considérons ici que ce centrage a déjà été effectué.

10.3 Formulation du problème d'ACP dans $\mathcal{H}$

Nous cherchons un vecteur $v \in \mathcal{H}$ tel que :
— $\|v\|_{\mathcal{H}} = 1$;
— la variance des projections

$$z_i = \langle v, \phi_i \rangle_{\mathcal{H}}$$

est maximale.

Comme les ϕ_i sont centrés, la moyenne des z_i est nulle :

$$\frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n} \langle v, \phi_i \rangle_{\mathcal{H}} = \left\langle v, \frac{1}{n}\sum_{i=1}^{n} \phi_i \right\rangle_{\mathcal{H}} = \langle v, 0 \rangle_{\mathcal{H}} = 0.$$

La variance empirique des z_i est donc

$$\mathrm{Var}_{\mathrm{emp}}(z) = \frac{1}{n}\sum_{i=1}^{n} z_i^2 = \frac{1}{n}\sum_{i=1}^{n} \langle v, \phi_i \rangle_{\mathcal{H}}^2.$$

Définition 10.3 (Problème d'ACP dans $\mathcal{H}$)**.** Le problème d'ACP dans l'espace de Hilbert $\mathcal{H}$ est

$$\max_{v \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \langle v, \phi_i \rangle_{\mathcal{H}}^2 \quad \text{sous la contrainte} \quad \|v\|_{\mathcal{H}} = 1.$$

10.4 Réduction au sous-espace engendré par les ϕ_i

On considère le sous-espace de dimension finie

$$\mathcal{H}_0 = \mathrm{span}\{\phi_1, \ldots, \phi_n\} \subset \mathcal{H}.$$

Proposition 10.1. *Pour résoudre le problème d'ACP dans $\mathcal{H}$, il suffit de chercher v dans $\mathcal{H}_0$.*

Démonstration. Par le Théorème 9.1, tout $v \in \mathcal{H}$ se décompose

$$v = v_0 + v_\perp,$$

avec $v_0 \in \mathcal{H}_0$ et $v_\perp \in \mathcal{H}_0^\perp$.

Pour tout i, $\phi_i \in \mathcal{H}_0$, donc

$$\langle v_\perp, \phi_i \rangle_{\mathcal{H}} = 0.$$

Donc

$$\langle v, \phi_i \rangle_{\mathcal{H}} = \langle v_0, \phi_i \rangle_{\mathcal{H}} + \langle v_\perp, \phi_i \rangle_{\mathcal{H}} = \langle v_0, \phi_i \rangle_{\mathcal{H}}.$$

La variance devient

$$\frac{1}{n}\sum_{i=1}^{n} \langle v, \phi_i \rangle_{\mathcal{H}}^2 = \frac{1}{n}\sum_{i=1}^{n} \langle v_0, \phi_i \rangle_{\mathcal{H}}^2.$$

Elle ne dépend donc pas de $v_\perp$.

De plus, par Pythagore,

$$\|v\|_{\mathcal{H}}^2 = \|v_0\|_{\mathcal{H}}^2 + \|v_\perp\|_{\mathcal{H}}^2 \geq \|v_0\|_{\mathcal{H}}^2.$$

Si $\|v\|_{\mathcal{H}} = 1$, alors $\|v_0\|_{\mathcal{H}} \leq 1$.

Si $v_0 = 0$, la variance est nulle. Sinon, on peut normaliser v_0 en posant $\tilde{v}_0 = v_0/\|v_0\|_{\mathcal{H}}$; alors $\|\tilde{v}_0\|_{\mathcal{H}} = 1$ et la variance correspondante est

$$\frac{1}{n}\sum_{i=1}^{n}\langle \tilde{v}_0, \phi_i\rangle_{\mathcal{H}}^2 = \frac{1}{\|v_0\|_{\mathcal{H}}^2} \cdot \frac{1}{n}\sum_{i=1}^{n}\langle v_0, \phi_i\rangle_{\mathcal{H}}^2,$$

qui est au moins aussi grande que celle obtenue avec v. On peut donc se restreindre à $\mathcal{H}_0$ sans perte de généralité. □

Ainsi, on cherche v sous la forme

$$v = \sum_{i=1}^{n} \alpha_i \phi_i,$$

avec $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$.

10.5 Expression de la variance et de la norme en fonction de α

Notons $\alpha = (\alpha_1, \ldots, \alpha_n)^\top$. Pour chaque i,

$$\begin{aligned} z_i &= \langle v, \phi_i\rangle_{\mathcal{H}} \\ &= \left\langle \sum_{j=1}^{n} \alpha_j \phi_j, \phi_i \right\rangle_{\mathcal{H}} \\ &= \sum_{j=1}^{n} \alpha_j \langle \phi_j, \phi_i\rangle_{\mathcal{H}} \\ &= \sum_{j=1}^{n} \alpha_j K_{ji}. \end{aligned}$$

Comme K est symétrique, $K_{ji} = K_{ij}$, donc

$$z_i = \sum_{j=1}^{n} K_{ij}\alpha_j.$$

En notation matricielle, si $z = (z_1, \ldots, z_n)^\top$, on obtient :

$$z = K\alpha.$$

La variance est donc

$$\mathrm{Var}_{\mathrm{emp}}(z) = \frac{1}{n}\sum_{i=1}^{n} z_i^2 = \frac{1}{n}\|z\|^2 = \frac{1}{n}(K\alpha)^\top(K\alpha) = \frac{1}{n}\alpha^\top K^\top K\alpha.$$

Comme K est symétrique, $K^\top = K$ et $K^\top K = K^2$, d'où

$$\mathrm{Var}_{\mathrm{emp}}(z) = \frac{1}{n}\alpha^\top K^2 \alpha.$$

Calculons maintenant la norme de v :

$$\begin{aligned} \|v\|_{\mathcal{H}}^2 &= \left\langle \sum_{i=1}^{n} \alpha_i \phi_i, \sum_{j=1}^{n} \alpha_j \phi_j \right\rangle_{\mathcal{H}} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \langle \phi_i, \phi_j\rangle_{\mathcal{H}} \\ &= \sum_{i,j=1}^{n} \alpha_i \alpha_j K_{ij} \\ &= \alpha^\top K \alpha. \end{aligned}$$

La contrainte $\|v\|_{\mathcal{H}} = 1$ devient donc

$$\alpha^\top K\alpha = 1.$$

10.6 Problème d'optimisation sur α et équation aux valeurs propres

Le problème d'ACP dans $\mathcal{H}$ se réécrit exactement comme un problème sur α :

$$\max_{\alpha\in\mathbb{R}^n} \frac{1}{n}\alpha^\top K^2\alpha \quad \text{sous la contrainte} \quad \alpha^\top K\alpha = 1.$$

On forme alors la Lagrangienne

$$\mathcal{L}(\alpha,\lambda) = \frac{1}{n}\alpha^\top K^2\alpha - \lambda(\alpha^\top K\alpha - 1).$$

Comme K et K^2 sont symétriques,

$$\nabla_\alpha(\alpha^\top K^2\alpha) = 2K^2\alpha, \quad \nabla_\alpha(\alpha^\top K\alpha) = 2K\alpha.$$

Donc

$$\nabla_\alpha\mathcal{L}(\alpha,\lambda) = \frac{2}{n}K^2\alpha - 2\lambda K\alpha.$$

La condition $\nabla_\alpha\mathcal{L} = 0$ donne

$$\frac{2}{n}K^2\alpha - 2\lambda K\alpha = 0 \quad \Leftrightarrow \quad K^2\alpha = n\lambda K\alpha.$$

Si $K\alpha = 0$, alors la contrainte $\alpha^\top K\alpha = 1$ est violée (dans ce cas $\alpha^\top K\alpha = 0$). On suppose donc $K\alpha \neq 0$.

On peut réécrire

$$K(K\alpha) = n\lambda(K\alpha).$$

Cela signifie que $K\alpha$ est un vecteur propre de K de valeur propre $n\lambda$.

En pratique, on résout directement

$$K\alpha = n\lambda\alpha. \tag{10.1}$$

10.7 Variance et normalisation

Si α vérifie (10.1), alors

$$\begin{aligned}\mathrm{Var}_{\mathrm{emp}}(z) &= \frac{1}{n}\alpha^\top K^2\alpha \\ &= \frac{1}{n}\alpha^\top K(K\alpha) \\ &= \frac{1}{n}\alpha^\top K(n\lambda\alpha) \\ &= \lambda(\alpha^\top K\alpha).\end{aligned}$$

Si l'on impose la contrainte $\alpha^\top K\alpha = 1$, alors

$$\mathrm{Var}_{\mathrm{emp}}(z) = \lambda.$$

Comme dans l'ACP linéaire, la valeur propre correspond à la variance expliquée par la composante principale correspondante.

Dans la pratique, on calcule les couples (λ,α) solutions de $K\alpha = n\lambda\alpha$, puis on normalise α de sorte que $\alpha^\top K\alpha = 1$.

10.8 Projection d'un nouveau point

Soit $x \in \mathbb{R}^p$ un nouveau point. Sa représentation dans $\mathcal{H}$ est $\Phi(x)$, et sa projection sur la composante principale associée à $v = \sum_i \alpha_i \phi_i$ est

$$\begin{aligned} z(x) &= \langle v, \Phi(x) \rangle_{\mathcal{H}} \\ &= \left\langle \sum_{i=1}^{n} \alpha_i \phi_i, \Phi(x) \right\rangle_{\mathcal{H}} \\ &= \sum_{i=1}^{n} \alpha_i \langle \phi_i, \Phi(x) \rangle_{\mathcal{H}} \\ &= \sum_{i=1}^{n} \alpha_i K(x_i, x). \end{aligned}$$

Ainsi, la projection se calcule uniquement à partir du noyau K et des coefficients α ; il n'est pas nécessaire de connaître explicitement l'application Φ (c'est le *truc du noyau*).

11 Limites de l'ACP et remarques finales

11.1 Nature linéaire de l'ACP classique

L'ACP linéaire cherche des combinaisons linéaires des variables d'origine. Si la structure des données est fortement non linéaire (par exemple, points alignés sur un cercle, une spirale, une variété intrinsèquement courbe), aucune combinaison linéaire ne peut la « redresser » globalement. L'ACP peut alors écraser la structure, même si elle reste utile pour donner une approximation globale.

L'ACP à noyau répond partiellement à ce problème : en travaillant dans un espace de Hilbert de grande dimension via Φ, une ACP linéaire dans $\mathcal{H}$ devient non linéaire dans l'espace d'origine.

11.2 Sensibilité aux valeurs aberrantes

La matrice de covariance Σ est basée sur des moyennes de produits $X_{ij}X_{ik}$. Des observations extrêmes (outliers) peuvent influencer fortement ces moyennes et donc orienter les vecteurs propres. L'ACP classique peut donc être très sensible aux outliers.

Des variantes robustes de l'ACP (basées sur des estimateurs de covariance robustes) existent pour atténuer ce problème.

11.3 Structure globale vs structure locale

L'ACP maximise une quantité globale (variance totale projetée). Elle ne préserve pas forcément les relations de voisinage locales entre les points. Des méthodes comme t-SNE ou UMAP visent plutôt à préserver la structure locale au prix d'une perte d'interprétation linéaire.

12 Conclusion

Dans ce document, nous avons :

— construit la matrice de covariance empirique Σ à partir des définitions classiques de la covariance, en démontrant en détail la formule compacte $\Sigma = \frac{1}{n} X^\top X$;
— montré que la variance de la projection sur une direction w est $w^\top \Sigma w$, et formulé l'ACP comme un problème d'optimisation sous contrainte ;
— résolu ce problème à l'aide des multiplicateurs de Lagrange, ce qui conduit à l'équation aux valeurs propres $\Sigma w = \lambda w$;
— relié l'ACP à la décomposition en valeurs singulières (SVD) de la matrice centrée X ;
— introduit les espaces de Hilbert et formulé rigoureusement l'ACP à noyau, en montrant que le problème se ramène à une équation aux valeurs propres sur la matrice noyau K ;
— présenté un exemple numérique complet en dimension 2.

Références

[1] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11) :559–572, 1901.

[2] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24 :417–441, 1933.

[3] I. T. Jolliffe. *Principal Component Analysis*, 2ème édition. Springer, 2002.

[4] I. T. Jolliffe and J. Cadima. Principal component analysis : a review and recent developments. *Philosophical Transactions of the Royal Society A*, 374(2065) :20150202, 2016.

[5] B. Schölkopf, A. Smola, and K. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5) :1299–1319, 1998.

[6] B. Escofier and J. Pagès. *Analyses factorielles simples et multiples : objectifs, méthodes et interprétation*, 4ème édition. Dunod, 2008.

[7] L. Lebart, A. Morineau, and M. Piron. *Statistique exploratoire multidimensionnelle*. Dunod, 1995.