# Analyse en Composantes Principales

Benoit Gaüzère, Stéphane Canu
benoit.gauzere@insa-rouen.fr

INSA Rouen Normandie - ITI

March 6, 2025

# Résumer l'information ?

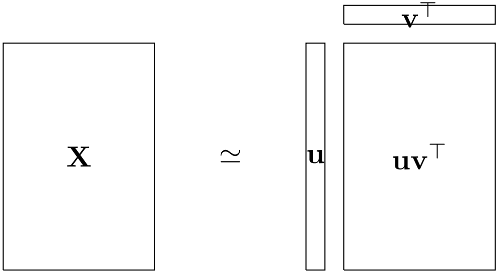
- Calculer une représentation avec moins de données mais un maximum d'informations
- Compression
- Suppression du bruit
- Visualisation en 2D ou 3D

# Résumer un tableau de données

## Comment résumer l'information ?

- Résumer un tableau de données par deux vecteurs $u$ et $v$
- $\mathbf{X} \in \mathbb{R}^{n \times p}$

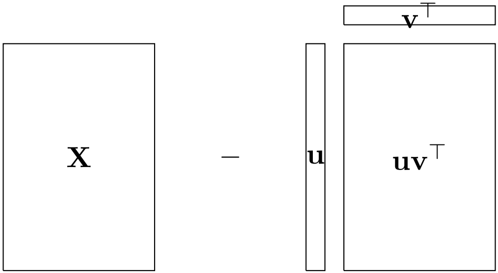
La meilleure représentation linéaire des observations est donnée par le couple de vecteurs $\mathbf{u} \in \mathbb{R}^n$ et $\mathbf{v} \in \mathbb{R}^p$ permettant au mieux de reconstruire la matrice $\mathbf{X}$.



# Reconstruction de $\mathbf{X}$

$$\min_{\mathbf{u},\mathbf{v}} \; J(\mathbf{u},\mathbf{v}) \quad \text{avec} \quad J(\mathbf{u},\mathbf{v}) \;=\; \sum_{i}^{n}\sum_{j}^{p}(x_{ij} - u_i v_j)^2$$

Aussi noté : $J(\mathbf{u},\mathbf{v}) \;= \|\mathbf{X} - \mathbf{u}\mathbf{v}^\top\|_F^2$



# Fonction de coût

La fonction coût $\sum_i^n \sum_j^p (x_{ij} - u_i v_j)^2$ peut se réécrire :

$$
\begin{aligned}
J(\mathbf{u},\mathbf{v}) &= \sum_i^n \sum_j^p x_{ij}^2 - 2 \sum_i^n \sum_j^p x_{ij} u_i v_j + \sum_i^n \sum_j^p (u_i v_j)^2 \\
&= \sum_i^n \sum_j^p x_{ij}^2 - 2 \sum_i^n \left( \sum_j^p x_{ij} v_j \right) u_i + \sum_i^n u_i^2 \ \sum_j^p v_j^2 \\
&= \sum_i^n \sum_j^p x_{ij}^2 - 2 (\mathbf{X}\mathbf{v})^\top \mathbf{u} + \|\mathbf{u}\|^2 \ \|\mathbf{v}\|^2
\end{aligned}
$$

et donc

$$
\min_{\mathbf{u},\mathbf{v}} \underbrace{\|\mathbf{X} - \mathbf{u}\mathbf{v}^\top\|_F^2}_{J(\mathbf{u},\mathbf{v})} \Leftrightarrow \min_{\mathbf{u},\mathbf{v}} \underbrace{-2(\mathbf{X}\mathbf{v})^\top \mathbf{u} + \|\mathbf{u}\|^2 \ \|\mathbf{v}\|^2}_{\mathcal{J}(\mathbf{u},\mathbf{v})}
$$

# Comment résoudre

$$\min_{\mathbf{u},\mathbf{v}} J(\mathbf{u},\mathbf{v})?$$

# Comment résoudre

$$\min_{\mathbf{u},\mathbf{v}} J(\mathbf{u},\mathbf{v})?$$

$\Rightarrow$ La méthode du gradient

# Méthode du gradient

# Minimisation d'une fonction de plusieurs variables

$$\min_{\mathbf{u},\mathbf{v}} J(\mathbf{u},\mathbf{v}) \; = \; \|\mathbf{X} - \mathbf{u}\mathbf{v}^\top\|_F^2$$

**Définition : Gradient**

soit $F$ une fonction de plusieurs $(d)$ variables :

$$\begin{array}{rccc} F: & \mathbb{R}^d & \longmapsto & \mathbb{R} \\ & \mathbf{x} & \longrightarrow & F(\mathbf{x}) \end{array}$$

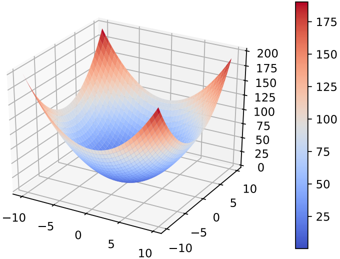
On appelle gradient de $F$ au point $\mathbf{x}$ la fonction des dérivées partielles

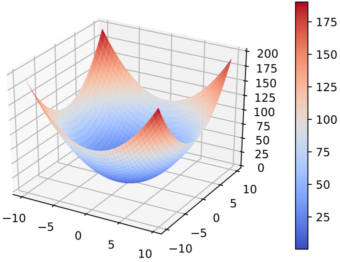
$$\begin{array}{rccl} \nabla_{\mathbf{x}} F: & \mathbb{R}^d & \longmapsto & \mathbb{R}^d \\ & \mathbf{x} & \longrightarrow & \nabla_{\mathbf{x}} F(\mathbf{x}) = \left( \frac{\partial F(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial F(\mathbf{x})}{\partial x_i}, \dots, \frac{\partial F(\mathbf{x})}{\partial x_d} \right)^\top \end{array}$$

# Minimisation d'une fonction de plusieurs variables

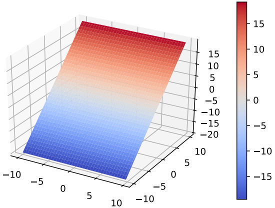
## Condition d'optimalité

- $(\mathbf{u}^\star, \mathbf{v}^\star)$ est solution du problème de minimisation ssi le gradient de la fonction $J$ s'annule en ce point
- $$\begin{cases} \nabla_{\mathbf{u}} J(\mathbf{u}^\star, \mathbf{v}^\star) = 0 \\ \nabla_{\mathbf{v}} J(\mathbf{u}^\star, \mathbf{v}^\star) = 0 \end{cases}$$
- $F$ doit être convexe et différentiable
- Si $F$ est strictement convexe : solution unique

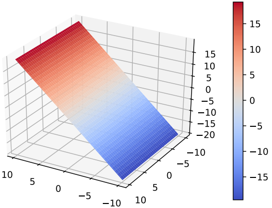




200
175
150
125
100
75
50
25
0
−10
−5
0
5
10

15
10
5
0
−5
−10
−15
−20

15
10
5
0
−5
−10
−15
−20

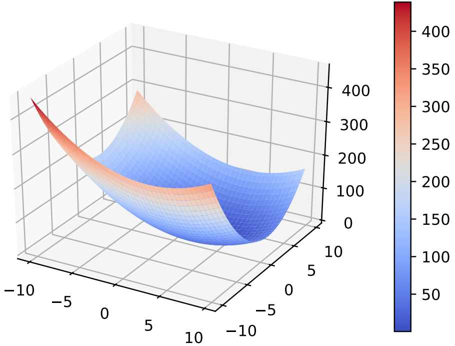
# Minimisation d'une fonction de plusieurs variables

## Exemple

$$\min_{x,y} J(x,y) \ = 2(x-a)^2 + (y-b)^2$$

## Méthode de résolution du problème

1. Calcul du gradient $\nabla_{\mathbf{x}} F(x,y)$ et $\nabla_{\mathbf{y}} F(x,y)$
   - $\nabla_x = 4(x-a)$
   - $\nabla_y = 2(y-b)$
2. Résolution des équations $\nabla_{\mathbf{x}} F(x^\star, y^\star) = 0$ (2 équations à 2 inconnues)
   - $\nabla_x = 0 \Leftrightarrow x^\star = a$
   - $\nabla_y = 0 \Leftrightarrow y^\star = b$



−10
−5
0
5
10
−10
−5
0
5
10
0
100
200
300
400
50
100
150
200
250
300
350
400

# Introduction à l'ACP

# Reconstruction de $\mathbf{X}$

## Rappel du problème

$$\min_{\mathbf{u},\mathbf{v}} \underbrace{\|\mathbf{X} - \mathbf{u}\mathbf{v}^\top\|_F^2}_{J(\mathbf{u},\mathbf{v})} \Leftrightarrow \min_{\mathbf{u},\mathbf{v}} \underbrace{-2(\mathbf{X}\mathbf{v})^\top \mathbf{u} + \|\mathbf{u}\|^2 \, \|\mathbf{v}\|^2}_{\mathcal{J}(\mathbf{u},\mathbf{v})}$$

## Résolution

Minimiser le coût c'est trouver $\mathbf{u}$ et $\mathbf{v}$ qui annulent le gradient :

$$\min_{\mathbf{u},\mathbf{v}} J(\mathbf{u},\mathbf{v}) = \|\mathbf{X} - \mathbf{u}\mathbf{v}^\top\|_F^2 \Leftrightarrow \begin{cases} \nabla_{\mathbf{u}}(\mathbf{u}) = 0 \\ \nabla_{\mathbf{v}}(\mathbf{v}) = 0 \end{cases}$$

# Calcul du gradient

$$J(\mathbf{u},\mathbf{v}) \; = \; \sum_i^n \sum_j^p x_{ij}^2 - 2 \sum_i^n \left( \sum_j^p x_{ij} v_j \right) u_i + \sum_i^n u_i^2 \; \sum_j^p v_j^2$$

## Gradient

$$\begin{cases} \dfrac{\partial J(\mathbf{u},\mathbf{v})}{\partial u_i} \; = \; -2 \displaystyle\sum_j^p x_{ij} v_j + 2 u_i \; \sum_j^p v_j^2 \\ \dfrac{\partial J(\mathbf{u},\mathbf{v})}{\partial v_j} \; = \; -2 \displaystyle\sum_i^n x_{ij} u_i + 2 v_j \; \sum_i^n u_i^2 \end{cases}$$

## Écriture matricielle du gradient[1]

$$\begin{cases} \nabla_{\mathbf{u}} J(\mathbf{u}) = -2\mathbf{X}\mathbf{v} + 2\|\mathbf{v}\|^2 \mathbf{u} \\ \nabla_{\mathbf{v}} J(\mathbf{v}) = -2\mathbf{X}^\top \mathbf{u} + 2\|\mathbf{u}\|^2 \mathbf{v} \end{cases}$$

[1] `https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`

## $\min_{\mathbf{u},\mathbf{v}} \ \|\mathbf{X} - \mathbf{u}\mathbf{v}^\top\|_F^2$

### Conditions d'optimalité

$$\left\{ \begin{array}{lcl} \nabla_{\mathbf{u}}\mathcal{J}(\mathbf{u}) = 0 & \Leftrightarrow & -\mathbf{X}\mathbf{v} + \|\mathbf{v}\|^2\mathbf{u} = 0 \\ \nabla_{\mathbf{v}}\mathcal{J}(\mathbf{v}) = 0 & \Leftrightarrow & -\mathbf{X}^\top\mathbf{u} + \|\mathbf{u}\|^2\mathbf{v} = 0 \end{array} \right.$$

### Solutions

$$\left\{ \begin{array}{l} \mathbf{X}\mathbf{v} = \|\mathbf{v}\|^2\mathbf{u} \\ \mathbf{X}^\top\mathbf{u} = \|\mathbf{u}\|^2\mathbf{v} \end{array} \right. \Rightarrow \mathbf{X}^\top\mathbf{X}\mathbf{v} = \|\mathbf{v}\|^2\mathbf{X}^\top\mathbf{u} = \underbrace{\|\mathbf{v}\|^2\|\mathbf{u}\|^2}_{\lambda}\ \mathbf{v}$$

- Solution $A\mathbf{v} = \lambda\mathbf{v}$
- $p$ solutions $(\mathbf{v}_i,\ \lambda_i)$.

# Quel vecteur propre choisir ?

## À l'optimum

$$\mathcal{J}(\mathbf{u},\mathbf{v}) \;=\; -2(\mathbf{X}\mathbf{v})^\top \mathbf{u} + \|\mathbf{u}\|^2\;\|\mathbf{v}\|^2 \qquad \text{et} \qquad \mathbf{X}\mathbf{v} = \|\mathbf{v}\|^2 \mathbf{u}$$

$$\begin{aligned}
\Rightarrow \mathcal{J}(\mathbf{u},\mathbf{v}) &= \; -2\|\mathbf{v}\|^2\mathbf{u}^\top\mathbf{u} + \|\mathbf{u}\|^2\;\|\mathbf{v}\|^2 \\
&= \; -2\|\mathbf{v}\|^2\|\mathbf{u}\|^2 + \|\mathbf{u}\|^2\;\|\mathbf{v}\|^2 \\
&= \; -\|\mathbf{u}\|^2\;\|\mathbf{v}\|^2 \qquad\qquad\qquad = -\lambda
\end{aligned}$$

$$\Rightarrow J(\mathbf{u},\mathbf{v}) \;= \|\mathbf{X}\|_F^2 - \lambda$$

## Solution

La solution du problème est donc donnée par le vecteur propre associé à la plus grande valeur propre de la matrice $\mathbf{X}^\top\mathbf{X}$ car :

$$\nabla_J(\mathbf{v}) = 0 \quad \Leftrightarrow \quad \mathbf{X}^\top\mathbf{X}\mathbf{v} - \lambda\mathbf{v} = 0 \quad \text{avec} \quad J(\mathbf{u},\mathbf{v}) \;= \|\mathbf{X}\|^2 - \lambda$$

Toutes les v.p. de $\mathbf{X}^\top\mathbf{X}$ (sdp.) sont positives.

# Résumé d'un tableau de données : résultat principal

**Théorème : (Eckart & Young, 1936)**

La solution unique du problème d'optimisation

$$\min_{\mathbf{u},\mathbf{v}} J(\mathbf{u},\mathbf{v}) \qquad \text{avec} \qquad J(\mathbf{u},\mathbf{v}) = \|\mathbf{X} - \mathbf{u}\mathbf{v}^\top\|_F^2$$

avec $\|\mathbf{v}^\star\| = 1$, est donnée par

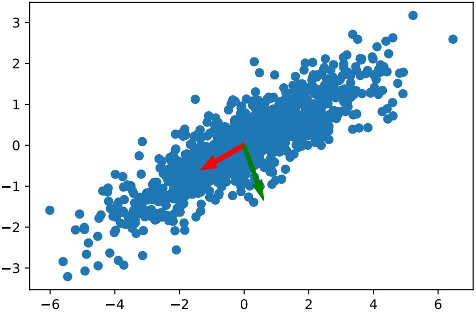
$$\mathbf{v}^\star \qquad \text{et} \qquad \mathbf{u}^\star = \mathbf{X}\,\mathbf{v}^\star,$$

où $\mathbf{v}^\star$ est le vecteur propre normé associé à $\lambda$ la plus grande valeur propre de la matrice $\mathbf{X}^\top\mathbf{X}$. De plus on a $\|\mathbf{u}^\star\| = \sqrt{\lambda}$.

# En résumé

$\min_{\mathbf{u},\mathbf{v}} \|\mathbf{X} - \mathbf{u}\mathbf{v}^\top\|_F^2$

- $\mathbf{u} = \mathbf{X}\mathbf{v}$ : résumé de $\mathbf{X}$ en 1D
- $\mathbf{v}$ : vecteur propre de $\lambda_1$ de $\Sigma$
- $\mathbf{v}$ : projection de $\mathbb{R}^{n\times p} \to \mathbb{R}^{n\times 1}$
- $\mathbf{v}^\top$ : projection de $\mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times p}$



# Analyse de la signification de $\lambda$

## Quelle est la variance de la projection $\mathbf{u}$ ?

- Considérons $\mathbf{X}$ centrée
- $\mathbf{u} = \mathbf{X}\mathbf{v}$, $\bar{\mathbf{u}} = 0$
- Variance de $\mathbf{u}$ : $s_{\mathbf{u}}^2 = \frac{1}{n}\|\mathbf{u}\|^2 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{u}_i^2$
- $s_{\mathbf{u}}^2 = \frac{1}{n}(\mathbf{X}\mathbf{v})^\top\mathbf{X}\mathbf{v} = \frac{1}{n}\mathbf{v}^\top\mathbf{X}^\top\mathbf{X}\mathbf{v}$

$$\Rightarrow s_{\mathbf{u}}^2 = \frac{1}{n}\lambda, \text{ avec } \mathbf{X}^\top\mathbf{X}\mathbf{v} = \lambda\mathbf{v}$$

## Meilleure projection ?

- $\mathbf{v}$ est le premier axe factoriel de $\mathbf{X}$
- $\mathbf{X}\mathbf{v}$ est la projection linéaire dont la variance est maximale
- Variance $\Leftrightarrow$ Information

# Et ensuite ?

## Augmenter la quantité d'infos

- La matrice de projection peut être étendue à plusieurs dimensions
- Reconstructions de plus en plus précise

Comment faire ?

# Les Résidus I

## Matrice des résidus

Construisons la matrice des résidus $\mathbf{R} = \mathbf{X} - \mathbf{u}\mathbf{v}^\top$

- Contient toute l'info "inexpliquée" par $\mathbf{u}$

## Spectre de $\mathbf{R}$

- $\mathbf{R}^\top\mathbf{R}$ admet les mêmes valeurs/vecteurs propres que $\mathbf{X}^\top\mathbf{X}$

$$
\begin{aligned}
R^\top R\mathbf{v}_i &= (\mathbf{X} - \mathbf{u}\mathbf{v}^\top)^\top(\mathbf{X} - \mathbf{u}\mathbf{v}^\top)\mathbf{v}_i \\
&= \mathbf{X}^\top\mathbf{X}\mathbf{v}_i - 2\mathbf{X}^\top\mathbf{u}\mathbf{v}^\top\mathbf{v}_i + \mathbf{v}\mathbf{u}^\top\mathbf{u}\mathbf{v}^\top\mathbf{v}_i \\
&= \lambda_i\mathbf{v}_i
\end{aligned}
$$

NB: $\mathbf{v}^\top\mathbf{v}_i = 0$ puisque les vecteurs propres sont orthogonaux entre eux.

# Les Résidus II

## Spectre de $\mathbf{R}$

- $\mathbf{R}^\top\mathbf{R}$ admet les mêmes valeurs/vecteurs propres que $\mathbf{X}^\top\mathbf{X}$
- …Sauf la plus grande $\lambda$:
  - $\mathbf{X}^\top\mathbf{X}\mathbf{v} = \lambda\mathbf{v}, \mathbf{X}^\top\mathbf{X}\mathbf{v}_2 = \lambda_2\mathbf{v}_2$ et $\lambda_2 < \lambda$
  - On a :

$$\begin{array}{rcll}\mathbf{R}^\top\mathbf{R}\mathbf{v} & = & (\mathbf{X}-\mathbf{u}\mathbf{v}^\top)^\top(\mathbf{X}-\mathbf{u}\mathbf{v}^\top)\mathbf{v} & \\ & = & \mathbf{X}^\top\mathbf{X}\mathbf{v} - 2\mathbf{v}\mathbf{u}^\top\mathbf{X}\mathbf{v} + \mathbf{v}\mathbf{u}^\top\mathbf{u}\mathbf{v}^\top\mathbf{v} & \\ & = & \lambda\mathbf{v} - 2\|\mathbf{u}\|^2\mathbf{v} + \|\mathbf{u}\|^2\|\mathbf{v}\|^2\mathbf{v} & \\ & = & \lambda\mathbf{v} - \|\mathbf{u}\|^2\mathbf{v} & = 0\end{array}$$

- NB: $\mathbf{u} = \mathbf{X}\mathbf{v}$, $\|\mathbf{v}\|^2 = 1$ et $\|\mathbf{u}\|^2 = \|\mathbf{X}\mathbf{v}\|^2 = \mathbf{v}^\top\mathbf{X}^\top\mathbf{X}\mathbf{v} = \lambda$

# 2ème axe factoriel

## $\mathbf{v}_2$

- $\mathbf{X}^\top \mathbf{X} \mathbf{v}_2 = \lambda_2 \mathbf{v}_2$, $\lambda_1 > \lambda_2 > \lambda_3 \ldots$
- $\mathbf{R}\mathbf{v}_2$ est la projection linéaire de $R$ qui maximise la variance

## Matrice de projection

- $\mathbf{P} = [\mathbf{v}_1, \mathbf{v}_2]$
- $\mathbf{XP}$ : projection de $\mathbf{X}$ en 2D
- Maximise l'information encodée
- et ainsi de suite ...

$$\begin{array}{cc} & 2 \\ p & \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{pmatrix} \end{array} = P$$

$$X = \begin{array}{c} p \\ \begin{pmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_i^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{pmatrix} \end{array} \begin{array}{cc} \begin{pmatrix} \\ \\ XP \\ \\ \\ \end{pmatrix} & n \\ 2 & \end{array}$$

# $k$ premiers axes factoriels

**Lemme**

Le sous-espace de dimension $k$ maximisant la variance des projections contient nécessairement le sous-espace de dimension $k-1$.

## $k$-ième axe factoriel

- $\mathbf{X}^\top \mathbf{X} \mathbf{v}_k = \lambda_k \mathbf{v}_k$, $\lambda_k$ : $k$-ième valeur propre.
- Matrice de projection $\mathbf{P}_k : \mathbb{R}^p \to \mathbb{R}^k$:

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_k \end{bmatrix}$$

- Les $k$ vecteurs propres liées aux $k$ plus grandes valeurs

# Algorithme

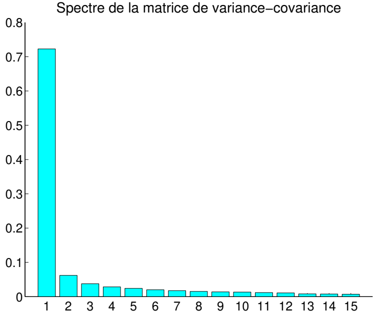
1. Centrer les données : $\{\mathbf{x}_i \in \mathbb{R}^p\}_{i=1}^N \longrightarrow \{\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}} \in \mathbb{R}^p\}_{i=1}^N$
2. Calculer la matrice de covariance $\boldsymbol{\Sigma} = \frac{1}{N}\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}$ avec $\tilde{\mathbf{X}}^\top = \begin{pmatrix} \tilde{\mathbf{x}}_1 & \cdots & \tilde{\mathbf{x}}_N \end{pmatrix}$
3. Calculer la décomposition en valeurs propres $\{\mathbf{v}_j \in \mathbb{R}^p, \lambda_j \in \mathbb{R}\}_{j=1}^p$ de $\boldsymbol{\Sigma}$
4. Ordonner les valeurs propres $\lambda_j$ par ordre décroissant
5. Nouvelle base de représentation des données :
$$\mathbf{P} = (\mathbf{v}_1, \cdots, \mathbf{v}_k) \in \mathbb{R}^{p \times k}$$
$\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ sont les $k$ vecteurs propres associés aux $k$ plus grandes valeurs propres $\lambda_j$.
6. Projection de tous les points sur $\mathbf{P}$ :
$$\mathbf{C} = \mathbf{X}\mathbf{P} \text{ avec } \mathbf{C} = \begin{pmatrix} \mathrm{t}_1^\top \\ \vdots \\ \mathrm{t}_N^\top \end{pmatrix} \quad , \quad \mathrm{t}_i^\top = \tilde{\mathbf{X}}(i,:)\mathbf{P}$$

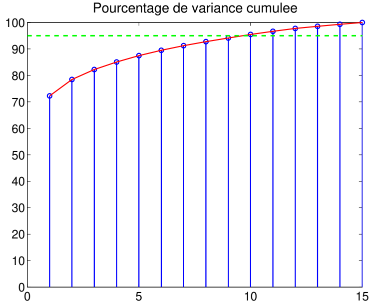
# Propriétés des axes factoriels

- Les valeurs propres de $\mathbf{\Sigma}$ sont positives car $\mathbf{\Sigma}$ est une matrice semi-définie positive
- Le nombre d'axes factoriels est égal au nombre de valeurs propres non-nulles de $\mathbf{\Sigma}$.
- Variance expliquée par l'axe factoriel $\mathbf{v}_k$ : $I_k = \mathbf{v}_k^\top \mathbf{\Sigma} \mathbf{v}_k = \lambda_k$.
- Variance totale des axes factoriels : $I = \sum_{k=1}^{p} \lambda_k$
- Pourcentage de variance expliquée par les $d$ premiers axes

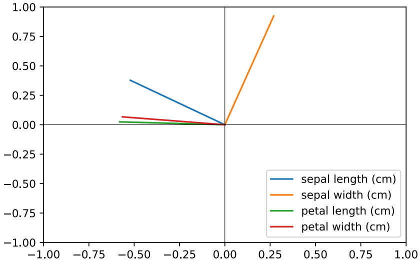
$$\frac{\sum_{k=1}^{d} \lambda_k}{\sum_{k=1}^{D} \lambda_k} \cdot 100$$

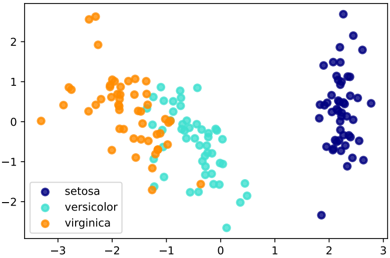
# Propriétés des axes factoriels



# Exemple sur Iris



# Reconstruction

## Revenir dans l'espace de départ

- $\mathbf{XP}$ : passage de $\mathbb{R}^p \to \mathbb{R}^d$
- Comment revenir dans l'espace de départ $\mathbb{R}^p$ ?
- Application sur la compression d'images
- Suppression du bruit

$$\begin{array}{cc} & p \\ 2 & \begin{pmatrix} \boldsymbol{v}_1^\top \\ \boldsymbol{v}_2^\top \end{pmatrix} \end{array} = P^\top$$

## Matrice de projection inverse

- Inverse de $\mathbf{P}$ : $\mathbf{P}^{-1}$
- $\mathbf{P}$ est orthogonale : $\mathbf{P}^{-1} = \mathbf{P}^\top$
- $\mathbf{P}^\top \in \mathbb{R}^{p \times d}$
- Attention à la normalisation en reconstruction

$$\begin{array}{ccc} n & \begin{pmatrix} \\ XP \\ \\ \end{pmatrix} & \begin{pmatrix} \tilde{\boldsymbol{x}}_1^\top \\ \tilde{\boldsymbol{x}}_2^\top \\ \vdots \\ \tilde{\boldsymbol{x}}_i^\top \\ \vdots \\ \tilde{\boldsymbol{x}}_n^\top \end{pmatrix} \\ & 2 & p \end{array} = \tilde{X}$$

# Reconstruction

$$\begin{array}{cc} & p \\ 2 & \begin{pmatrix} \boldsymbol{v}_1^\top \\ \boldsymbol{v}_2^\top \end{pmatrix} \end{array} = P^\top$$

$$\begin{array}{ccc} n & \begin{pmatrix} \\ \\ XP \\ \\ \\ \end{pmatrix} & \begin{pmatrix} \widetilde{\boldsymbol{x}}_1^\top \\ \widetilde{\boldsymbol{x}}_2^\top \\ \vdots \\ \widetilde{\boldsymbol{x}}_i^\top \\ \vdots \\ \widetilde{\boldsymbol{x}}_n^\top \end{pmatrix} \\ & 2 & p \end{array} = \tilde{X}$$

# ACP : Conclusion

## Méthode de réduction de dimension

- Matrice de projection $\mathbf{P}$ : $d$ premiers vecteurs propres de $\mathbf{\Sigma}$
- Centrer/réduire les données
- Projection : $\mathbf{X}_d = \mathbf{X}\mathbf{P}$
- Information de l'axe $\mathbf{v}_k$ : $\lambda_k$
- Reconstruction : $\tilde{\mathbf{X}} = \mathbf{X}_d\mathbf{P}^\top$
- Les derniers vecteurs propres contiennent le “bruit”

## Limites

- Méthode de projection linéaire
- Hypothèse de distribution gaussienne

# Bonus

# Un autre mode de calcul du gradient

$\mathcal{J}(\mathbf{u},\mathbf{v}) \;=\; -2(\mathbf{X}\mathbf{v})^\top\mathbf{u} + \|\mathbf{u}\|^2\;\|\mathbf{v}\|^2$, le minimum est atteint lorsque le gradient s'annule.

## gradient d'une fonction : $J : \mathbb{R}^n \longrightarrow \mathbb{R}$

par définition le gradient $\nabla_J(\mathbf{v})$ s'obtient en posant $\phi(t) = J(\mathbf{v}+t\mathbf{d})$ où $\mathbf{d}$ est un vecteur de $\mathbb{R}^p$ et en calculant la valeur de la dérivée de $\phi$ au point zéro qui vérifie $\phi'(0) = \mathbf{d}^\top\nabla_J(\mathbf{v})$.

Faisons le calcul

$$\begin{aligned}
\phi(t) &= \mathcal{J}(\mathbf{u},\mathbf{v}+t\mathbf{d})\\
&= -2(\mathbf{X}(\mathbf{v}+t\mathbf{d}))^\top\mathbf{u} + \|\mathbf{u}\|^2\;\|\mathbf{v}+t\mathbf{d}\|^2\\
&= -2(\mathbf{v}^\top\mathbf{X}^\top\mathbf{u} + t\mathbf{d}^\top\mathbf{X}^\top\mathbf{u}) + \|\mathbf{u}\|^2(\|\mathbf{v}\|^2 + t^2\|\mathbf{d}\|^2 + 2t\;\mathbf{v}^\top\mathbf{d})
\end{aligned}$$

$$\begin{aligned}
\phi'(t) &= -2\mathbf{d}^\top\mathbf{X}^\top\mathbf{u} + \|\mathbf{u}\|^2(2t\|\mathbf{d}\|^2 + 2\mathbf{v}^\top\mathbf{d})\\
\phi'(0) &= -2\mathbf{d}^\top\mathbf{X}^\top\mathbf{u} + 2\|\mathbf{u}\|^2\;\mathbf{v}^\top\mathbf{d}
\end{aligned}$$

$$= d^\top\left(-2\mathbf{X}^\top\mathbf{u} + 2\|\mathbf{u}\|^2\mathbf{v}\right) \quad\Rightarrow\quad \nabla_{\mathbf{v}}\mathcal{J}(\mathbf{v}) = -2\mathbf{X}^\top\mathbf{u} + 2\|\mathbf{u}\|^2\mathbf{v}$$