

# Analyse en Composantes Principales

Machine Learning

Cédric RICHARD

Université Nice Sophia Antipolis

# ANALYSE EN COMPOSANTES PRINCIPALES

## Objectifs

---

**Contexte :** Chaque individu  $x_i$  du tableau  $\mathbf{X}$  est considéré comme un point d'un espace vectoriel  $\mathcal{E}$  de dimension  $p$ .

L'ensemble des individus constitue un nuage de points dans  $\mathcal{E}$ .

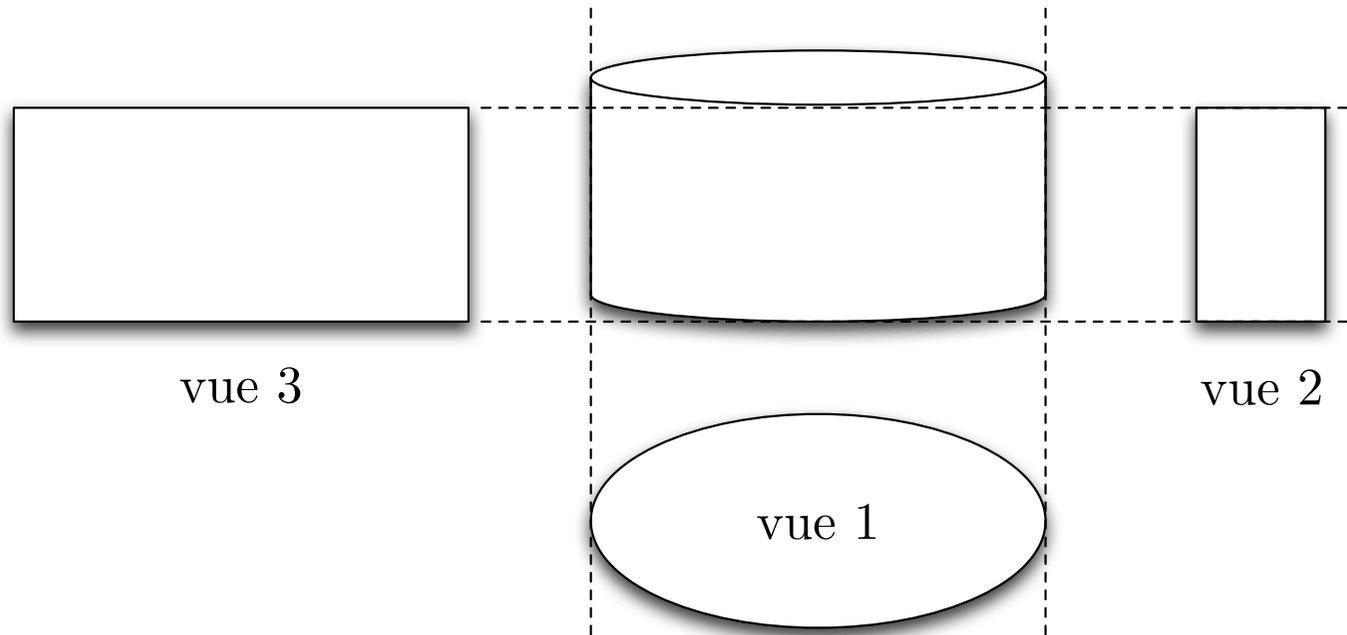
On note  $m$  son point moyen, appelé aussi *centre de gravité*.

**Objectif :** On cherche à réduire le nombre  $p$  de variables en préservant au maximum la structure du nuage, afin de le représenter le plus fidèlement possible.

**Principe :** L'ACP vise à projeter les données dans un sous-espace approprié, de dimension plus faible, préservant la topologie du nuage.

# ANALYSE EN COMPOSANTES PRINCIPALES

Exemple



# ANALYSE EN COMPOSANTES PRINCIPALES

Exemple plus réaliste

---

| cidre | odeur | sucré | acide | amer | astringence | suffocante | piquante | alcool | parfum | fruité |
|-------|-------|-------|-------|------|-------------|------------|----------|--------|--------|--------|
| 1     | 2.14  | 1.86  | 3.29  | 2.29 | 2           | 0.14       | 2.29     | 1.86   | 1.29   | 1.29   |
| 2     | 2.43  | 0.79  | 2.71  | 2.57 | 2           | 0.43       | 2.57     | 2.86   | 0.43   | 0.14   |
| 3     | 2.71  | 3.14  | 2.57  | 2.57 | 1.43        | 0.14       | 2.14     | 0.86   | 2.29   | 1.71   |
| 4     | 3     | 3.71  | 2.14  | 2.07 | 1.57        | 0          | 1.29     | 1      | 3.14   | 3.14   |
| 5     | 3.43  | 1.29  | 2.86  | 3.14 | 2.17        | 1          | 1.86     | 2.86   | 1.14   | 0.29   |
| 6     | 3.14  | 0.86  | 2.86  | 3.79 | 2.57        | 0.14       | 1.71     | 3.29   | 0.14   | 0      |
| 7     | 3.14  | 1.14  | 2.86  | 2.86 | 2           | 0.43       | 1.71     | 1.86   | 0.14   | 0      |
| 8     | 2.43  | 3.71  | 3.21  | 1.57 | 1.71        | 0          | 1        | 0.57   | 2.57   | 2.86   |
| 9     | 5.1   | 2.86  | 2.86  | 3.07 | 1.79        | 1.71       | 0.43     | 1.43   | 0.57   | 2.71   |
| 10    | 3.07  | 3.14  | 2.57  | 3    | 2           | 0          | 0.43     | 1.29   | 2.57   | 3.07   |

Notes obtenues pour 10 cidres selon 10 critères lors d'un concours agricole.

# ESPACE DES INDIVIDUS

## Métrie

---

**Métrie :** L'ACP repose sur les distances entre individus dans  $\mathcal{E}$ . Le choix de la métrie a une influence fondamentale sur le résultat de l'analyse.

**Définition :** Soit  $\mathbf{M}$  une matrice définie positive de dimension  $p$ . La fonction suivante  $d_M : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}^+$  définit une métrie

$$d_M^2(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_M^2 = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j)$$

Cette distance est appelée distance de Mahalanobis lorsque  $\mathbf{M} = \Sigma^{-1}$ , où  $\Sigma$  est la matrice de variance-covariance des données.

**Produit scalaire :** La métrie définie ci-dessus dérive du produit scalaire

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle_M = \mathbf{x}_i^\top \mathbf{M} \mathbf{x}_j$$

On dit que  $\mathbf{x}_i$  et  $\mathbf{x}_j$  sont  $M$ -orthogonaux si  $\langle \mathbf{x}_i, \mathbf{x}_j \rangle_M = 0$ .

# ANALYSE DES DONNÉES

## Métrie

---

Utiliser la métrie  $\mathbf{M} = \mathbf{T}^\top \mathbf{T}$  sur le tableau de données  $\mathbf{X}$  est équivalent à travailler avec la métrie euclidienne sur le tableau transformé  $\mathbf{X}\mathbf{T}^\top$ .

**Tableau transformé :** Lorsqu'on travaille sur le tableau transformé comme ci-dessus, il est alors possible d'utiliser la norme euclidienne. En effet,

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle_M = \mathbf{x}_i^\top (\mathbf{T}^\top \mathbf{T}) \mathbf{x}_j = (\mathbf{T}\mathbf{x}_i)^\top (\mathbf{T}\mathbf{x}_j) = \langle \mathbf{T}\mathbf{x}_i, \mathbf{T}\mathbf{x}_j \rangle_I$$

**Réciproque :** Pour toute matrice définie positive  $\mathbf{M}$ , il existe une matrice définie positive  $\mathbf{T}$  telle que  $\mathbf{M} = \mathbf{T}^\top \mathbf{T}$ . On notera improprement  $\mathbf{T} = \mathbf{M}^{\frac{1}{2}}$ .

Appliquer préalablement la transformation  $\mathbf{X}\mathbf{T}^\top \rightarrow \mathbf{X}$  permet de simplifier les traitements.

# ANALYSE DES DONNÉES

## Métriques particulières

---

**Métrique euclidienne :** Elle est obtenue pour  $M = I$ .

L'une des difficultés rencontrées avec la métrique euclidienne est qu'elle privilégie les variables les plus dispersées et dépend donc de leur unité de mesure.

**Métrique réduite :** Elle consiste à prendre  $M = D_{1/\sigma^2}$ , où  $D_{1/\sigma^2}$  est la matrice diagonale de termes diagonaux les inverses  $\frac{1}{\sigma_i^2}$  des variances des variables.

$$D_{1/\sigma^2} = \begin{pmatrix} \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_p^2} \end{pmatrix}$$

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

# ANALYSE DES DONNÉES

## Tableau de données centrées réduites

---

Utiliser la métrique  $M = D_{1/\sigma^2} = D_{1/\sigma}^\top D_{1/\sigma}$  sur le tableau de données  $X$  revient à travailler avec la métrique euclidienne sur le tableau transformé  $X D_{1/\sigma}^\top$ .

En effet :

$$\begin{aligned} d_M^2(\mathbf{x}_i, \mathbf{x}_j) &= \|\mathbf{x}_i - \mathbf{x}_j\|_M^2 = (\mathbf{x}_i - \mathbf{x}_j)^\top D_{1/\sigma}^\top D_{1/\sigma} (\mathbf{x}_i - \mathbf{x}_j) \\ &= (D_{1/\sigma} \mathbf{x}_i - D_{1/\sigma} \mathbf{x}_j)^\top (D_{1/\sigma} \mathbf{x}_i - D_{1/\sigma} \mathbf{x}_j) \end{aligned}$$

Il est équivalent de travailler avec la métrique  $D_{1/\sigma^2}$  sur le tableau  $X$ , ou avec la métrique euclidienne  $I$  sur le tableau centré réduit  $Z$  composé des données :

$$z_i^j = \frac{x_i^j - \bar{x}^j}{\sigma_j}$$

Le tableau de données centré réduit  $Z$  se calcule matriciellement ainsi :

$$Z = Y D_{1/\sigma} = (X - \mathbf{1} \mathbf{m}^\top) D_{1/\sigma}.$$

# INERTIES

## Inertie par rapport à un point

---

**Définition :** L'inertie du nuage de points  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  en un point quelconque  $\mathbf{a}$  est donnée par

$$I_a = \sum_{i=1}^n p_i \|\mathbf{x}_i - \mathbf{a}\|_M^2$$

**Propriété :** L'inertie du nuage de points  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  en son point moyen  $\mathbf{m}$ , ou *centre de gravité*, est

$$\begin{aligned} I_m &= \sum_{i=1}^n p_i \|\mathbf{x}_i - \mathbf{m}\|_M^2 = \frac{1}{2} \sum_{i,j=1}^n p_i p_j \|\mathbf{x}_i - \mathbf{x}_j\|_M^2 \\ &= \text{Trace}(\Sigma \mathbf{M}) \end{aligned}$$

Cette propriété a été démontrée au chapitre précédent.

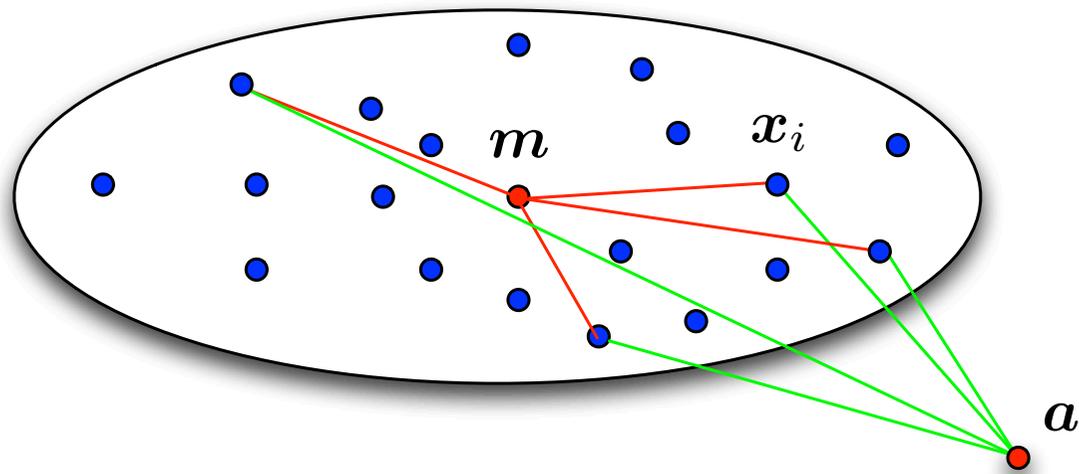
**Propriété :** Pour un tableau de données centrées réduites, on a

$$I_m = \text{Trace}(\mathbf{R}) = p$$

# INERTIES

Inertie par rapport à un point

---



# INERTIES

## Théorème de Huygens

---

**Propriété :** Soit  $\mathbf{m}$  le centre de gravité du nuage de points  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ , et  $\mathbf{a}$  un point quelconque de  $\mathbb{R}^p$ . L'inertie  $I_a$  du nuage au point  $\mathbf{a}$  est donnée par

$$I_a = I_m + d_M^2(\mathbf{a}, \mathbf{m})$$

En conséquence,  $I_a$  est minimum pour  $\mathbf{a} = \mathbf{m}$ .

**Démonstration :**

$$\begin{aligned} I_a &= \sum_{i=1}^n p_i \|\mathbf{x}_i - \mathbf{m} - (\mathbf{a} - \mathbf{m})\|_M^2 \\ &= \sum_{i=1}^n p_i \|\mathbf{x}_i - \mathbf{m}\|_M^2 + \sum_{i=1}^n p_i \|\mathbf{a} - \mathbf{m}\|_M^2 - 2 \sum_{i=1}^n p_i \langle \mathbf{x}_i - \mathbf{m}, \mathbf{a} - \mathbf{m} \rangle_M \\ &= I_m + d_M^2(\mathbf{a}, \mathbf{m}) \end{aligned}$$

# INERTIES

## Inertie par rapport à un axe

---

**Définition :** L'inertie du nuage de points  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  par rapport à un axe  $\Delta$  est définie par

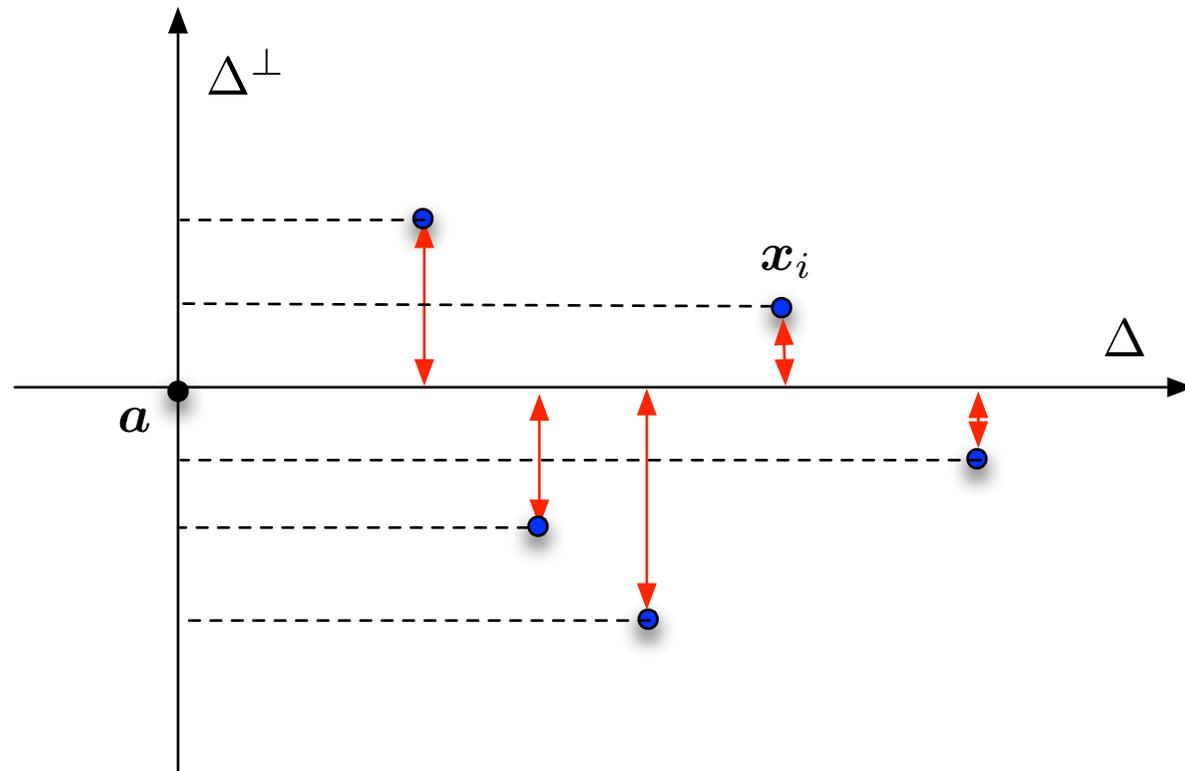
$$I_{\Delta} = \sum_{i=1}^n p_i d_M^2(\mathbf{x}_i, \Delta)$$

Cette inertie quantifie la dispersion du nuage des individus autour de  $\Delta$ .

# INERTIES

Inertie par rapport à un axe

---



# INERTIES

## Inertie par rapport à un sous-espace affine

---

**Définition :** L'inertie du nuage de points  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  par rapport à un sous-espace affine  $\mathcal{F}$  est définie par

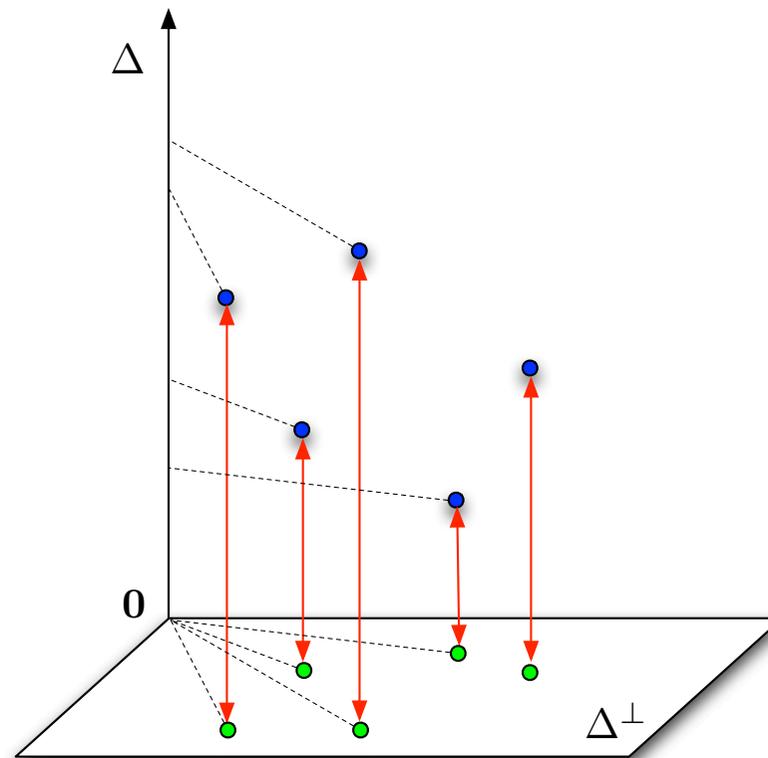
$$I_{\mathcal{F}} = \sum_{i=1}^n p_i d_M^2(\mathbf{x}_i, \mathcal{F})$$

Cette inertie quantifie la dispersion du nuage des individus dans  $\mathcal{F}^\perp$ .

# INERTIES

Inertie par rapport à un sous-espace affine

---



# INERTIES

## Inertie par rapport à un sous-espace affine

---

- ▷ Soit  $\mathcal{E}_1$  un sous-espace vectoriel de  $\mathcal{E} = \mathbb{R}^p$ . On pose  $\mathcal{E}_1 \oplus \mathcal{E}_1^\perp = \mathcal{E}$ , où  $\oplus$  désigne la somme directe.
- ▷ Soient  $\mathcal{F}_1$  et  $\mathcal{F}_1^\perp$  les sous-espaces affines associés à  $\mathcal{E}_1$  et  $\mathcal{E}_1^\perp$  passant par  $\mathbf{a}$ . Celui-ci se décompose selon :

$$\mathbf{a} = \mathbf{a}_\mathcal{E} + \mathbf{a}_{\mathcal{E}^\perp}$$

avec  $\mathbf{a}_\mathcal{E} \in \mathcal{E}_1$  et  $\mathbf{a}_{\mathcal{E}^\perp} \in \mathcal{E}_1^\perp$ .

- ▷ De même, on décompose chaque individu  $\mathbf{x}_i$  sur  $\mathcal{E}_1$  et  $\mathcal{E}_1^\perp$  ainsi

$$\mathbf{x}_i = \mathbf{x}_{i,\mathcal{E}} + \mathbf{x}_{i,\mathcal{E}^\perp}$$

avec  $\mathbf{x}_{i,\mathcal{E}} \in \mathcal{E}_1$  et  $\mathbf{x}_{i,\mathcal{E}^\perp} \in \mathcal{E}_1^\perp$ .

# INERTIES

## Inertie par rapport à un sous-espace affine

---

▷ L'inertie du nuage par rapport à  $\mathcal{F}_1$  et  $\mathcal{F}_1^\perp$  est définie par

$$I_{\mathcal{F}_1} = \sum_{i=1}^n p_i d_M^2(\mathbf{x}_i, \mathcal{F}_1) = \sum_{i=1}^n p_i d_M^2(\mathbf{x}_i, \mathcal{E}^\perp, \mathbf{a}_{\mathcal{E}^\perp})$$
$$I_{\mathcal{F}_1^\perp} = \sum_{i=1}^n p_i d_M^2(\mathbf{x}_i, \mathcal{F}_1^\perp) = \sum_{i=1}^n p_i d_M^2(\mathbf{x}_i, \mathcal{E}, \mathbf{a}_{\mathcal{E}})$$

Voir le schéma ci-après. Par ailleurs, par le théorème de Pythagore, on a

$$d_M^2(\mathbf{x}_i, \mathbf{a}) = d_M^2(\mathbf{x}_i, \mathcal{E}, \mathbf{a}_{\mathcal{E}}) + d_M^2(\mathbf{x}_i, \mathcal{E}^\perp, \mathbf{a}_{\mathcal{E}^\perp})$$

On en déduit la propriété suivante

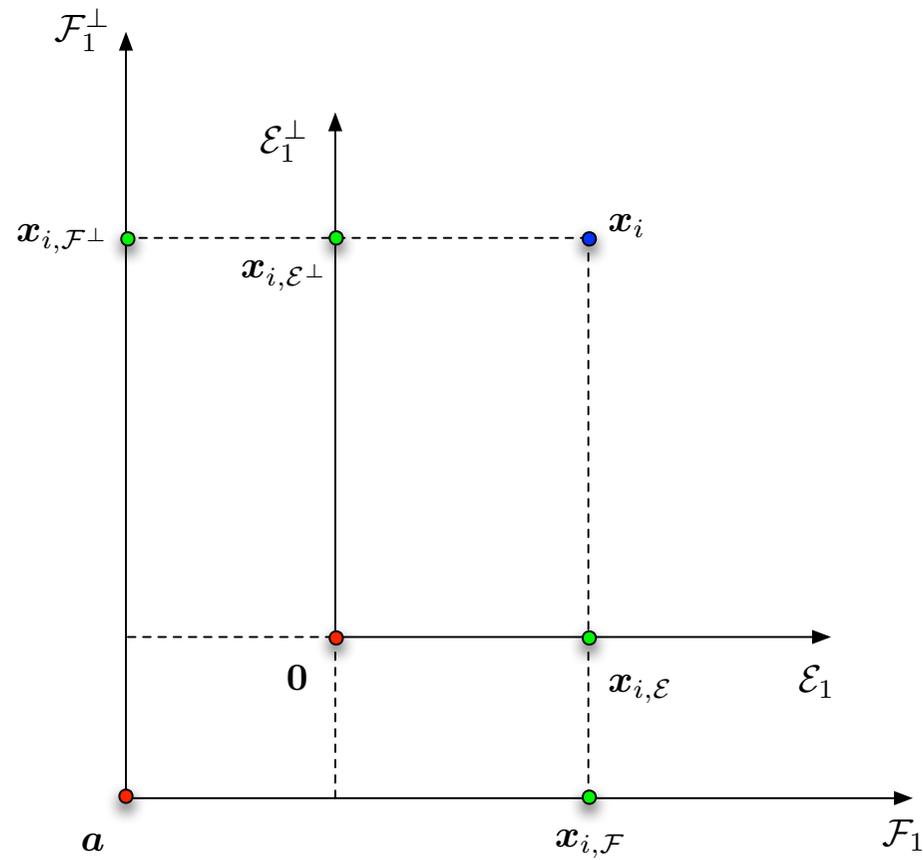
$$\boxed{I_a = I_{\mathcal{F}_1} + I_{\mathcal{F}_1^\perp}}$$

▷ D'après le théorème de Huygens, on a :  $I_a = I_m + d_M^2(\mathbf{a}, \mathbf{m})$ . On en déduit que l'inertie  $I_a$  est minimum pour  $\mathcal{F}_1 = \mathcal{E}_1$  et  $\mathcal{F}_1^\perp = \mathcal{E}_1^\perp$ , et vaut alors  $I_m$ .

# INERTIES

Inertie par rapport à un sous-espace affine

---



# INERTIES

## Inertie par rapport à un sous-espace affine

---

Des deux points vues au transparent précédent, on déduit le résultat suivant :

**Propriété :** Soit  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  un nuage de données centrées dans  $\mathcal{E} = \mathbb{R}^p$ .

Soit  $\mathcal{E}_1$  un sous-espace vectoriel de  $\mathcal{E} = \mathbb{R}^p$ , et  $\mathcal{E}_1^\perp$  son supplémentaire.

On a alors

$$I_m = I_{\mathcal{E}_1} + I_{\mathcal{E}_1^\perp}$$

**Vocabulaire :** Lorsque  $\mathcal{E}_1$  désigne un axe, l'inertie par rapport à son sous-espace supplémentaire  $\mathcal{E}_1^\perp$  est appelée **inertie portée**, ou **inertie expliquée**, par  $\mathcal{E}_1$ .

**Nous considérerons les données centrées et réduites**

$$\mathbf{Z} = \mathbf{Y} \mathbf{D}_{1/\sigma} = (\mathbf{X} - \mathbf{1} \mathbf{m}^\top) \mathbf{D}_{1/\sigma}$$

**Remarque :** toute autre métrique  $\mathbf{M}$  peut être choisie en travaillant avec  $\mathbf{Y} \mathbf{M}^{\frac{1}{2}}$

# INERTIES

## Vocabulaire

---

La propriété  $I_m = I_{\mathcal{E}_1} + I_{\mathcal{E}_1^\perp}$  signifie qu'en projetant les données sur  $\mathcal{E}_1$ , l'inertie perdue par le nuage vaut  $I_{\mathcal{E}_1}$ . Le nuage projeté a alors pour inertie  $I_{\mathcal{E}_1^\perp}$ .

En conséquence :

**Vocabulaire :** Lorsque  $\mathcal{E}_1$  désigne un axe, l'inertie par rapport à son sous-espace supplémentaire  $\mathcal{E}_1^\perp$  est appelée **inertie portée**, ou **inertie expliquée**, par  $\mathcal{E}_1$ .

Minimiser  $I_{\mathcal{E}_1} \Leftrightarrow$  Maximiser  $I_{\mathcal{E}_1^\perp}$

# ANALYSE EN COMPOSANTES PRINCIPALES

## Formulation du problème

---

**Problème :** On recherche le sous-espace  $\mathcal{E}_k$  de dimension  $k < p$  dont l'inertie expliquée  $I_{\mathcal{E}_k^\perp}$  soit maximum, c'est à dire

$$\max_{\mathcal{E}_k} I_{\mathcal{E}_k^\perp} = \sum_{i=1}^n p_i d^2(\mathbf{z}_i, \mathcal{E}_k^\perp)$$

Ceci est équivalent à rechercher le sous-espace  $\mathcal{E}_k$  où la moyenne des distances inter-points projetés y est maximum.

**Théorème d'Inclusion :** Soit  $\mathcal{E}_k$  un sous-espace, de dimension  $k$ , d'inertie portée maximum. Alors le sous-espace optimum  $\mathcal{E}_{k+1}$ , de dimension  $k + 1$ , d'inertie portée maximum est la somme directe de  $\mathcal{E}_k$  et du sous-espace de dimension 1 d'inertie portée maximum.

# ANALYSE EN COMPOSANTES PRINCIPALES

## Résolution du problème

---

**Recherches des axes :** Etant donné le Théorème d'Inclusion, pour déterminer le sous-espace optimum  $\mathcal{E}_k$ , il est possible de procéder séquentiellement :

- recherche du premier axe d'inertie portée maximum
- recherche du deuxième axe, orthogonal au premier, d'inertie portée max.
- etc.

**Rappels :** Soit  $\Delta(\mathbf{u})$  un axe de vecteur directeur normé  $\mathbf{u}$ . On rappelle que

$$d^2(\mathbf{z}_i, \Delta) = \min_{v_i} d^2(\mathbf{z}_i, v_i \mathbf{u})$$

En conséquence, l'inertie du nuage centré  $\{\mathbf{z}_1, \dots, \mathbf{z}_n\}$  par rapport à  $\Delta$  est définie par

$$I_{\Delta}(\mathbf{u}) = \min_{\mathbf{v}} \sum_{i=1}^n p_i d^2(\mathbf{z}_i, v_i \mathbf{u})$$

avec  $\mathbf{v} = [v_1, \dots, v_n]^{\top}$ .

# ANALYSE EN COMPOSANTES PRINCIPALES

## Détermination du premier axe

---

**Définition du problème :** L'axe  $\Delta$  recherché, d'inertie expliquée  $I_{\Delta^\perp}$  maximum, et donc d'inertie  $I_\Delta$  minimum, est en conséquence solution du problème

$$\min_{\mathbf{u}} I_\Delta(\mathbf{u}) \quad \text{avec} \quad I_\Delta(\mathbf{u}) = \min_{\mathbf{v}} \sum_{i=1}^n p_i d^2(\mathbf{z}_i, v_i \mathbf{u})$$

---

**Calcul du gradient :** En développant la fonction coût, notée  $J_\Delta(\mathbf{u}, \mathbf{v})$ , on obtient

$$J_\Delta(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^n p_i [\|\mathbf{z}_i\|^2 - 2v_i \langle \mathbf{z}_i, \mathbf{u} \rangle + v_i^2 \|\mathbf{u}\|^2]$$

En conséquence

$$\nabla_{\mathbf{u}} J_\Delta(\mathbf{u}, \mathbf{v}) = -2\mathbf{Z}^\top \mathbf{D} \mathbf{v} + 2\|\mathbf{v}\|_D^2 \mathbf{u}$$

$$\nabla_{\mathbf{v}} J_\Delta(\mathbf{u}, \mathbf{v}) = -2\mathbf{D} \mathbf{Z} \mathbf{u} + 2\|\mathbf{u}\|^2 \mathbf{D} \mathbf{v}$$

avec  $\mathbf{D}$  la matrice diagonal de termes diagonaux  $p_i$ .

# ANALYSE EN COMPOSANTES PRINCIPALES

## Détermination du premier axe

---

Conditions d'optimalité : Celles-ci s'écrivent

$$\begin{cases} \mathbf{Z}^\top \mathbf{D} \mathbf{v} = \|\mathbf{v}\|_D^2 \mathbf{u} \\ \mathbf{D} \mathbf{Z} \mathbf{u} = \|\mathbf{u}\|^2 \mathbf{D} \mathbf{v} \end{cases}$$

En combinant les 2 équations, on obtient

$$\mathbf{Z}^\top \mathbf{D} \mathbf{Z} \mathbf{u} = \underbrace{\|\mathbf{u}\|^2 \|\mathbf{v}\|_D^2}_{\lambda} \mathbf{u}$$

La solution est un vecteur propre  $\mathbf{u}$  de la matrice de covariance  $\mathbf{\Sigma} = \mathbf{Z}^\top \mathbf{D} \mathbf{Z}$

Lequel ?

# ANALYSE EN COMPOSANTES PRINCIPALES

## Détermination du premier axe

---

Lequel ? En injectant le résultat dans la fonction coût, on obtient

$$\begin{aligned} I_{\Delta}(\mathbf{u}) &= \sum_{i=1}^n p_i [\|\mathbf{z}_i\|^2 - 2 v_i \langle \mathbf{z}_i, \mathbf{u} \rangle + v_i^2 \|\mathbf{u}\|^2] \\ &= \text{Trace}(\boldsymbol{\Sigma}) - 2 \mathbf{u}^{\top} \mathbf{Z}^{\top} \mathbf{D} \mathbf{v} + \|\mathbf{v}\|_D^2 \|\mathbf{u}\|^2 \\ &= \text{Trace}(\boldsymbol{\Sigma}) - 2 \|\mathbf{v}\|_D^2 \|\mathbf{u}\|^2 + \|\mathbf{v}\|_D^2 \|\mathbf{u}\|^2 \\ &= \text{Trace}(\boldsymbol{\Sigma}) - \lambda(\mathbf{u}) \end{aligned}$$

où  $\lambda(\mathbf{u})$  est la valeur propre associée à  $\mathbf{u}$

Le vecteur propre recherché est donc celui associé à la plus grande valeur propre

# ANALYSE EN COMPOSANTES PRINCIPALES

## Détermination du premier axe

---

Soient  $\lambda_1 \geq \dots \geq \lambda_p$  les valeurs propres de la matrice de covariance  $\Sigma$ , de vecteurs propres unitaires associés  $\mathbf{u}_1, \dots, \mathbf{u}_p$

**Choix du premier axe :** L'inertie  $I_{\Delta}(\mathbf{u}) = \text{Trace}(\Sigma) - \lambda(\mathbf{u})$  est minimum pour le couple  $(\mathbf{u}_1, \lambda_1)$

**Inertie correspondante :** L'inertie expliquée par l'axe  $\Delta(\mathbf{u}_1)$  est donnée par

$$\begin{aligned} I_{\Delta^{\perp}}(\mathbf{u}_1) &= I_m - I_{\Delta}(\mathbf{u}_1) \\ &= \text{Trace}(\Sigma) - [\text{Trace}(\Sigma) - \lambda_1] \\ &= \lambda_1 \end{aligned}$$

# ANALYSE EN COMPOSANTES PRINCIPALES

## Détermination du deuxième axe

---

- ▷ On construit le tableau de données résiduel  $\mathbf{Z}_1 = \mathbf{Z} - \mathbf{v}_1 \mathbf{u}_1^\top$ . Il est aisé de montrer que les données correspondantes appartiennent à  $\Delta^\perp$

$$\begin{aligned}\mathbf{Z}_1 \mathbf{u}_1 &= \mathbf{Z} \mathbf{u}_1 - \|\mathbf{u}_1\|^2 \mathbf{v}_1 \\ &= 0\end{aligned}$$

d'après la deuxième condition d'optimalité, car  $\mathbf{D}$  est inversible.

- ▷ La matrice  $\mathbf{Z}_1^\top \mathbf{D} \mathbf{Z}_1$  admet les mêmes couples valeurs/vecteurs propres  $(\lambda_k, \mathbf{u}_k)$  que  $\mathbf{Z}^\top \mathbf{D} \mathbf{Z}$ , pour  $k \neq 1$ ,

$$\begin{aligned}(\mathbf{Z}_1^\top \mathbf{D} \mathbf{Z}_1) \mathbf{u}_k &= (\mathbf{Z}^\top \mathbf{D} \mathbf{Z}) \mathbf{u}_k \quad k \neq 1 \\ &= \lambda_k \mathbf{u}_k\end{aligned}$$

car  $(\mathbf{u}_1 \perp \mathbf{u}_k)$ , excepté  $\lambda_1$  qui devient nulle

$$\begin{aligned}(\mathbf{Z}_1^\top \mathbf{D} \mathbf{Z}_1) \mathbf{u}_1 &= [\mathbf{Z} - \mathbf{v}_1 \mathbf{u}_1^\top]^\top \mathbf{D} [\mathbf{Z} - \mathbf{v}_1 \mathbf{u}_1^\top] \\ &= (1 - 2 + 1) \|\mathbf{u}_1\|^2 \|\mathbf{v}_1\|_D^2 \mathbf{u}_1 = 0\end{aligned}$$

# ANALYSE EN COMPOSANTES PRINCIPALES

## Détermination du deuxième axe

---

Pour déterminer le deuxième axe, on pose le même problème que pour le premier, avec le tableau  $\mathbf{Z}_1$ . On aboutit alors à  $\mathbf{u}_2$ . En effet

**Choix du deuxième axe :** L'inertie  $I_{\Delta}(\mathbf{u}) = \text{Trace}(\mathbf{\Sigma}_1) - \lambda(\mathbf{u})$  est minimum pour le couple  $(\mathbf{u}_2, \lambda_2)$

Remarque :  $\text{Trace}(\mathbf{\Sigma}_1) = \text{Trace}(\mathbf{\Sigma}) - \lambda_1$

**Inertie correspondante :** L'inertie expliquée par l'axe  $\Delta(\mathbf{u}_2)$  est donnée par

$$\begin{aligned} I_{\Delta^{\perp}}(\mathbf{u}_2) &= (I_m - \lambda_1) - I_{\Delta}(\mathbf{u}_2) \\ &= \text{Trace}(\mathbf{\Sigma}_1) - [\text{Trace}(\mathbf{\Sigma}_1) - \lambda_2] \\ &= \lambda_2 \end{aligned}$$

**et ainsi de suite...**

# ANALYSE EN COMPOSANTES PRINCIPALES

## Axes principaux et inertie expliquée

---

- ▷ Les axes  $\Delta(\mathbf{u}_k)$  sont appelés **axes factoriels** ou **axes principaux**
- ▷ L'inertie expliquée par l'axe  $\Delta(\mathbf{u}_k)$  est la valeur propre  $\lambda_k$  de  $\Sigma$
- ▷ L'inertie expliquée par le sous-espace factoriel  $\mathcal{E}_{k_0}$  engendré par les  $k_0$  premiers axes principaux est

$$I_{\mathcal{E}_{k_0}^\perp} = \lambda_1 + \dots + \lambda_{k_0}$$

- ▷ Le **taux d'inertie expliquée** par  $\mathcal{E}_{k_0}$  est donné par

$$\frac{\lambda_1 + \dots + \lambda_{k_0}}{\sum_{k=1}^p \lambda_k}$$

$$\text{car } I_m = \text{Trace}(\Sigma) = \sum_{k=1}^p \lambda_k$$

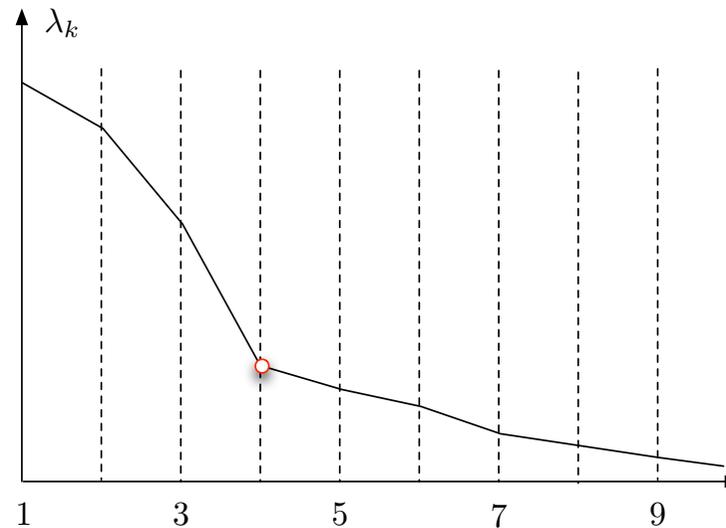
# ANALYSE EN COMPOSANTES PRINCIPALES

## Nombre d'axes à retenir

---

L'ACP vise à réduire la dimension de l'espace des individus. On souhaite conserver aussi peu d'axes principaux que possible. Des critères empiriques :

- ▷ Retenir les axes offrant une corrélation suffisante entre composantes principales et variables initiales. Voir plus loin.
- ▷ Fixer un taux minimum d'inertie expliquée, par exemple 0.80
- ▷ Rechercher un coude dans l'éboulis des valeurs propres  $\lambda_k$



# ANALYSE EN COMPOSANTES PRINCIPALES

## Composantes principales

---

**Coordonnées des individus :** La coordonnée  $c_{\ell k}$  de l'individu  $\mathbf{z}_\ell$  sur l'axe principal  $\Delta(\mathbf{u}_k)$  est donnée par sa projection sur  $\mathbf{u}_k$

$$c_{\ell k} = \mathbf{z}_\ell^\top \mathbf{u}_k$$

**Composantes principales :** Il s'agit des vecteurs  $\mathbf{c}_k$  des coordonnées des individus sur l'axe principal  $\Delta(\mathbf{u}_k)$ , c'est à dire

$$\mathbf{c}_k = \mathbf{Z} \mathbf{u}_k$$

# ANALYSE EN COMPOSANTES PRINCIPALES

## Propriétés des composantes principales

---

**Moyenne arithmétique :**

$$\bar{\mathbf{c}}_k = \mathbf{1}^\top \mathbf{D} \mathbf{c}_k = \mathbf{1}^\top \mathbf{D} \mathbf{Z} \mathbf{u}_k = 0$$

car  $\mathbf{1}^\top \mathbf{D} \mathbf{Z} = \mathbf{0}$ , le tableau étant centré

**Variance :**

$$\text{var}(\mathbf{c}_k) = \mathbf{c}_k^\top \mathbf{D} \mathbf{c}_k = \mathbf{u}_k^\top \mathbf{Z}^\top \mathbf{D} \mathbf{Z} \mathbf{u}_k = \lambda_k$$

car  $\mathbf{u}_k$  est vecteur propre de  $\Sigma = \mathbf{Z}^\top \mathbf{D} \mathbf{Z}$

**Covariance :** Il s'agit des vecteurs  $\mathbf{c}_k$  des coordonnées des individus sur l'axe principal  $\Delta(\mathbf{u}_k)$ , c'est à dire

$$\text{cov}(\mathbf{c}_k, \mathbf{c}_\ell) = \mathbf{c}_k^\top \mathbf{D} \mathbf{c}_\ell = \mathbf{u}_k^\top \mathbf{Z}^\top \mathbf{D} \mathbf{Z} \mathbf{u}_\ell = 0$$

car  $(\mathbf{u}_k \perp \mathbf{u}_\ell)$ . Les composantes principales sont donc décorrélées.

# ANALYSE EN COMPOSANTES PRINCIPALES

## Qualité de la représentation d'un individu

---

**Définition :** La qualité de la représentation d'un individu  $\ell$  sur l'axe principal  $k$  est donnée par le cosinus de leur angle, soit

$$\cos(\mathbf{z}_\ell, \mathbf{u}_k) = \frac{\mathbf{z}_\ell^\top \mathbf{u}_k}{\|\mathbf{z}_\ell\|}$$

Comme  $\mathbf{z}_\ell^\top \mathbf{u}_k = c_{\ell k}$ , et  $\{\mathbf{u}_k\}_k$  est une base orthonormée de  $\mathbb{R}^p$ , on a

$$\cos^2(\mathbf{z}_\ell, \mathbf{u}_k) = \frac{c_{\ell k}^2}{\sum_{k=1}^p c_{\ell k}^2}$$

# ANALYSE EN COMPOSANTES PRINCIPALES

## Contribution d'un individu à une composante principale

---

On a vu que  $\text{var}(\mathbf{c}_k) = \lambda_k = \sum_{i=1}^n p_i c_{ik}^2$ .

**Définition :** On définit la contribution de l'individu  $\ell$  à un axe principal  $k$  par

$$\theta_{\ell k} = \frac{p_{\ell} c_{\ell k}^2}{\lambda_k}$$

**Interprétation :** La contribution d'un individu  $\ell$  doit être relativisée par rapport à son poids  $p_{\ell}$ . Ainsi, si  $\theta_{\ell k} > \alpha p_{\ell}$  avec  $\alpha > 2$ , on peut juger la contribution de l'individu  $\ell$  importante.

**Sur-représentation :** L'individu  $\ell$  peut être considéré comme sur-représenté par l'axe principal  $k$  si  $\theta_{\ell k} > 0.25$ . Il a eu une influence trop importante dans la détermination de celui-ci, risquant de perturber la définition des autres axes.

# ANALYSE EN COMPOSANTES PRINCIPALES

## Espace des variables

---

On s'intéresse à présent aux colonnes  $\mathbf{z}^j$  du tableau  $\mathbf{Z}$ , représentant les variables.

**Métrique  $D$**  : On munit l'espace des variables d'une métrique naturelle,  $D$ , celle des poids des individus

$$\langle \mathbf{z}^i, \mathbf{z}^j \rangle_D = (\mathbf{z}^i)^\top \mathbf{D} \mathbf{z}^j \quad \|\mathbf{z}^i\|_D^2 = (\mathbf{z}^i)^\top \mathbf{D} \mathbf{z}^i$$

**Interprétation** : Pour des variables centrées, ce qui est le cas ici, on a

$$\text{cor}(\mathbf{z}^i, \mathbf{z}^j) = \frac{\langle \mathbf{z}^i, \mathbf{z}^j \rangle_D}{\|\mathbf{z}^i\|_D \|\mathbf{z}^j\|_D} = \cos(\mathbf{z}^i, \mathbf{z}^j)$$

**Remarque** : Les composantes principales normalisée  $\mathbf{c}_k / \sqrt{\lambda_k}$  forment une base  $D$ -orthonormale :

$$\left\langle \frac{\mathbf{c}_i}{\sqrt{\lambda_i}}, \frac{\mathbf{c}_j}{\sqrt{\lambda_j}} \right\rangle_D = \text{cor} \left( \frac{\mathbf{c}_i}{\sqrt{\lambda_i}}, \frac{\mathbf{c}_j}{\sqrt{\lambda_j}} \right) = \delta_{ij}$$

# ANALYSE EN COMPOSANTES PRINCIPALES

## Corrélation des composantes principales et variables

---

▷ Le coefficient de corrélation entre  $\mathbf{c}_k$  et  $\mathbf{z}^i$  est donné par

$$\text{cor}(\mathbf{c}_k, \mathbf{z}^i) = \frac{\langle \mathbf{c}_k, \mathbf{z}^i \rangle_D}{\sqrt{\lambda_k}}$$

▷ Les coefficients de corrélation entre  $\mathbf{c}_k$  et l'ensemble des variables de  $\mathbf{Z}$  peuvent être calculés ainsi

$$\text{cor}(\mathbf{c}_k, \mathbf{Z}) = \frac{\mathbf{Z}^\top \mathbf{D} \mathbf{c}_k}{\sqrt{\lambda_k}}$$

▷ Comme  $\mathbf{Z}^\top \mathbf{D} \mathbf{c}_k = \mathbf{Z}^\top \mathbf{D} \mathbf{Z} \mathbf{u}_k = \boldsymbol{\Sigma} \mathbf{u}_k$ , on a directement

$$\text{cor}(\mathbf{c}_k, \mathbf{Z}) = \sqrt{\lambda_k} \mathbf{u}_k$$

# ANALYSE EN COMPOSANTES PRINCIPALES

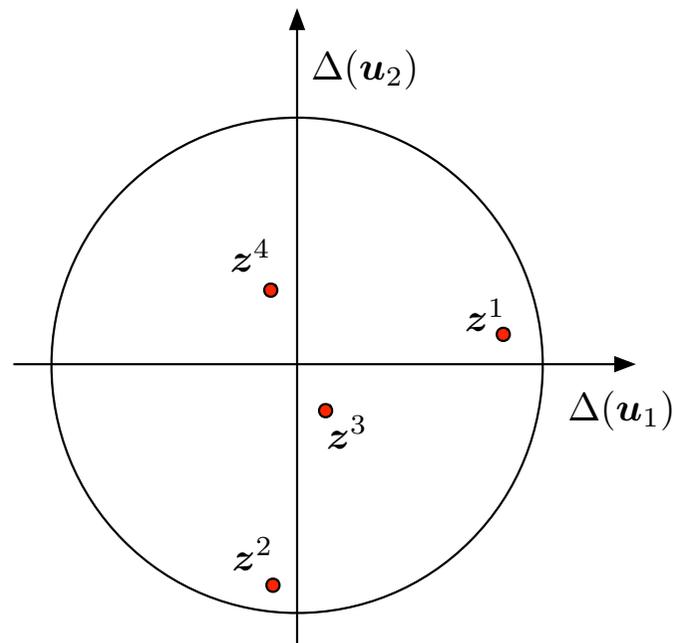
## Cercle des corrélations

---

On représente traditionnellement chaque variable  $z^i$  par un point de coordonnées

$$(\text{cor}(z^i, \mathbf{c}_1), \text{cor}(z^i, \mathbf{c}_2), \text{cor}(z^i, \mathbf{c}_3), \dots)$$

afin d'interpréter les axes principaux obtenus.



# ANALYSE EN COMPOSANTES PRINCIPALES

En résumé

---

- ▷ Centrer et réduire les données  $\mathbf{X} \rightarrow \mathbf{Z}$
- ▷ Calculer la matrice de covariance  $\mathbf{\Sigma} = \mathbf{Z}^\top \mathbf{D} \mathbf{Z}$
- ▷ Diagonaliser  $\mathbf{\Sigma}$  pour obtenir  $\{(\mathbf{u}_k, \lambda_k)\}_{k=1, \dots, p}$
- ▷ Choisir le nombre  $k_0$  d'axes à retenir
- ▷ Préciser le taux d'inertie expliqué par chacun des axes retenus
- ▷ Calculer les composantes principales  $\{\mathbf{c}_k\}_{k=1, \dots, k_0}$
- ▷ Représenter les individus dans le(s) repère(s) des axes principaux
- ▷ Tracer les cercles de corrélation
- ▷ Interpréter