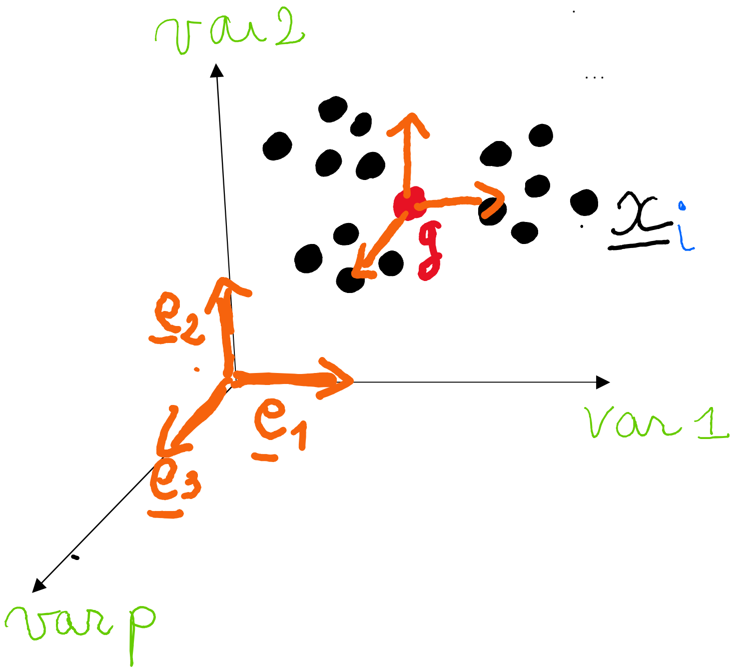
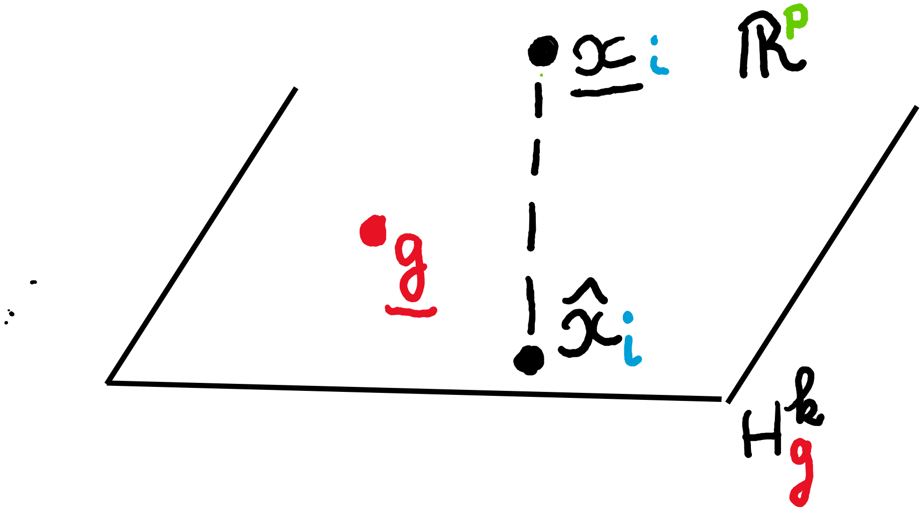
Analyse de données

Cours 3
Caractérisation des axes factoriels

Données





$$I_g = \frac{1}{n}\sum_{i=1}^{n} \|\hat{x}_i\|^2 + \frac{1}{n}\sum_{i=1}^{n} \|\hat{x}_i - x_i\|^2$$

inertie totale du nuage

inertie expliquée

inertie résiduelle

$$I_g = \frac{1}{n}\sum_{i=1}^{n}\|\underline{\hat{x}}_i\|^2 + \frac{1}{n}\sum_{i=1}^{n}\|\underline{\hat{x}}_i - \underline{x}_i\|^2$$

inertie totale du nuage

inertie expliquée

inertie résiduelle

Maximiser l'inertie expliquée $\Longleftrightarrow$ Minimiser l'inertie résiduelle

Conséquence : pour des raisons de calcul on cherche à maximiser l'inertie expliquée (formule plus simple !)

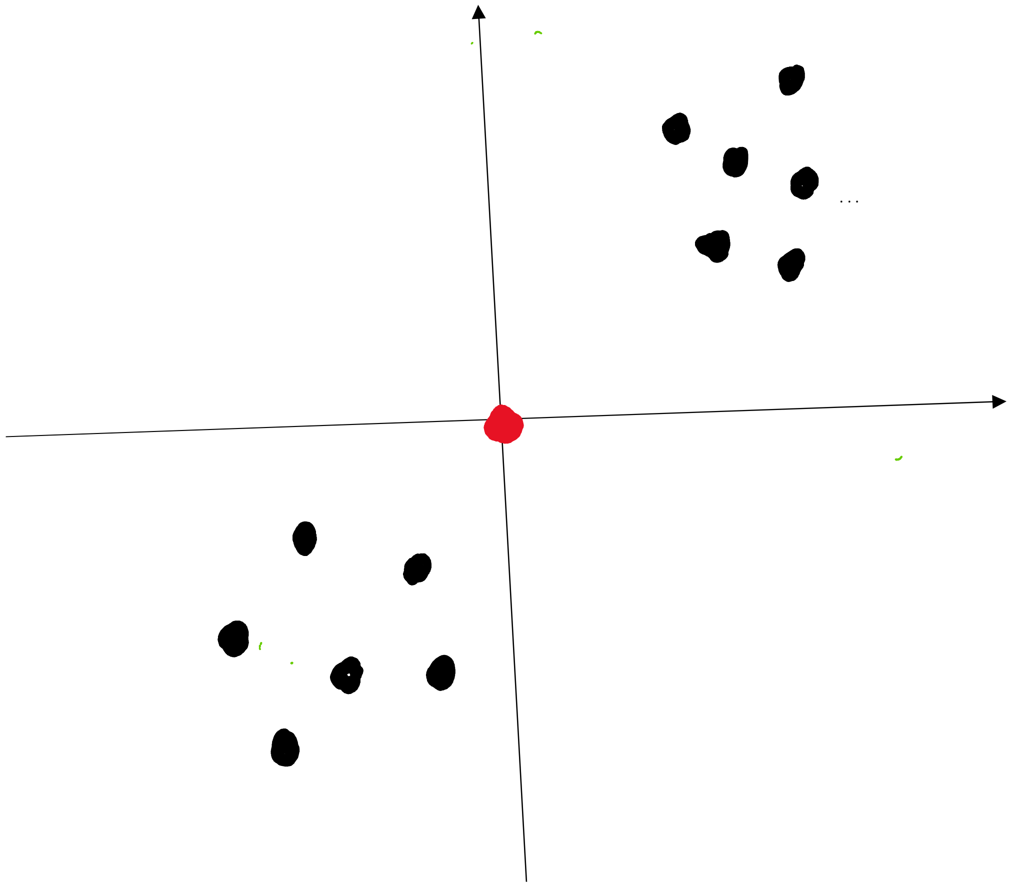
Lemme fondamental

Si H^k est le sous-espace de dimension k d'inertie expliquée maximale alors le sous-espace de dimension $k+1$ d'inertie expliquée maximale est

$$H^{k+1} = H^k \oplus \Delta_{\underline{u}}$$

où $\Delta_{\underline{u}}$ est la droite vectorielle orthogonale à H^k d'inertie expliquée maximale

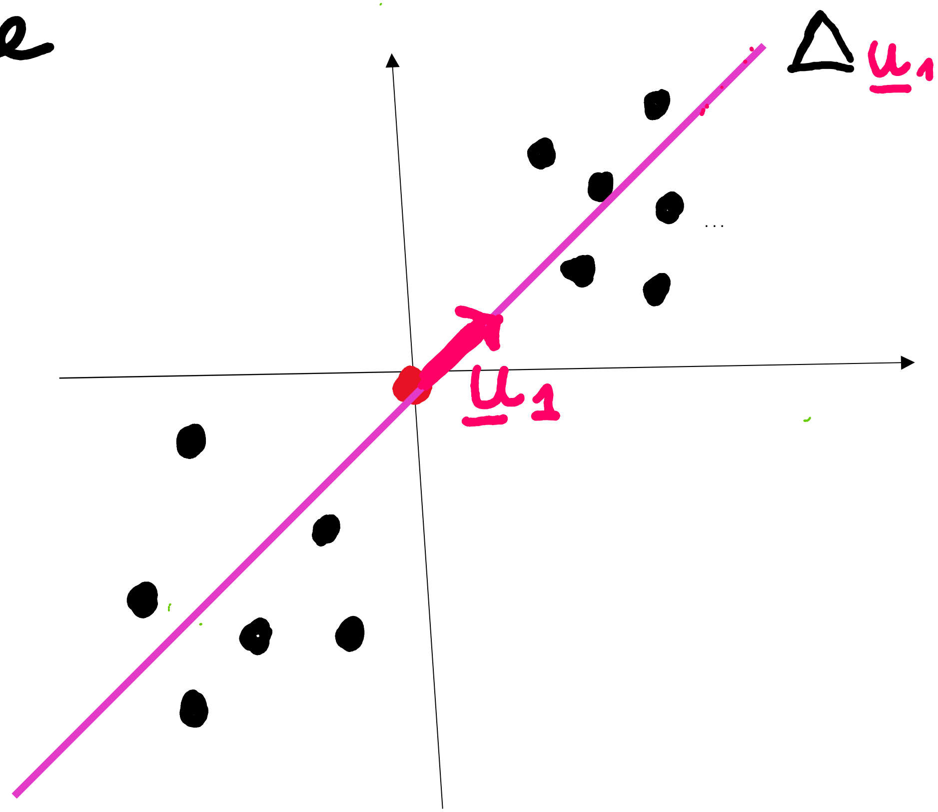
$k = 2$ $\qquad$ $H^{k+1} = H^k \oplus \Delta \underline{u}$



Algorithme de recherche des sous-espaces (1)

① · Déterminer la droite $H^1 = \Delta_{\underline{u}_1}$ d'inertie expliquée maximale

$$\operatorname*{Max}_{\underline{u}_1 \in \mathbb{R}} I(\Delta_{\underline{u}_1})$$



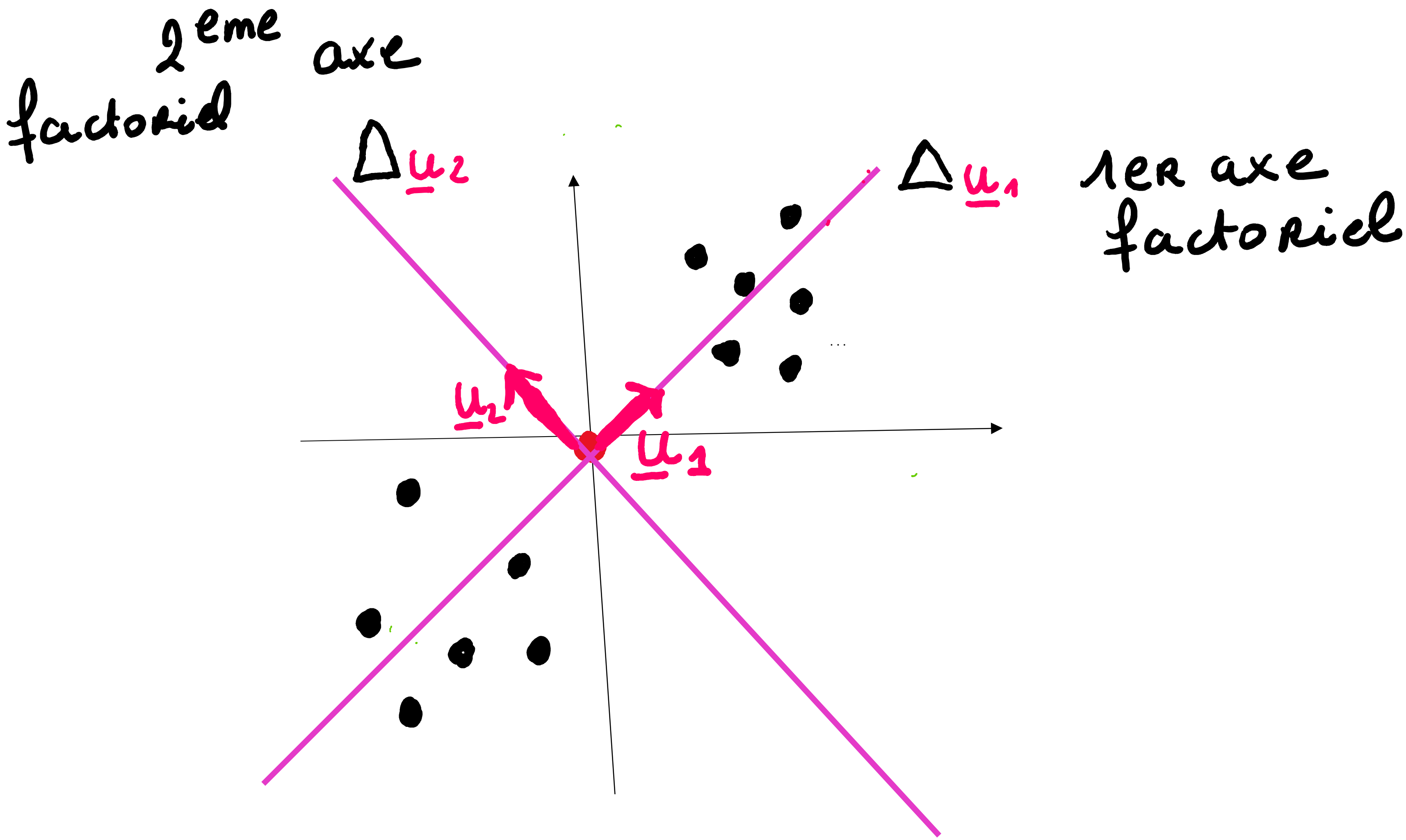
Algorithme de recherche des sous-espaces (2)

(2) Pour déterminer H^2 on sait d'après le lemme que :

$H^2 = \Delta_{\underline{u_1}} \oplus \Delta_{\underline{u_2}}$ où $\underline{u_2}$ est le vecteur normé orthogonal à $\underline{u_1}$ et $\underline{u_2}$ est solution du problème

$$\begin{array}{l} \text{Max } I(\Delta_{\underline{u_2}}) \\ \underline{u_2} \perp \underline{u_1} \\ \underline{u_2} \in \mathbb{R}^p \end{array}$$

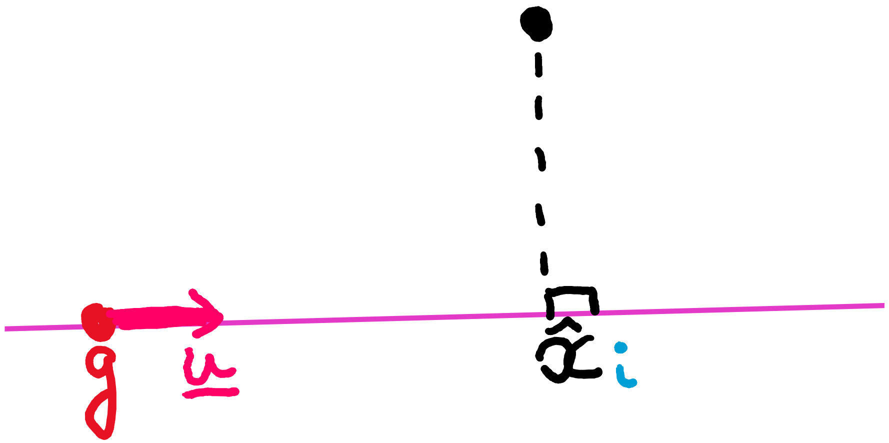
etc ...



2eme axe factoriel
$\Delta_{\underline{u}_2}$
$\Delta_{\underline{u}_1}$
1er axe factoriel
$\underline{u}_2$
$\underline{u}_1$

Détermination de $\Delta_{\underline{u}_1} \in \mathbb{R}^p$ qui maximise $I(\Delta_{\underline{u}_1})$

$$I(\Delta_{\underline{u}}) = \frac{1}{n} \sum_{i=1}^{n} \|\hat{x}_i\|^2$$



Réécriture de l'inertie expliquée (1)

$$I(\Delta_{\underline{u}}) = \frac{1}{n}\sum_{i=1}^{n} \|\hat{\underline{x}}_i\|^2 \qquad \text{où } \hat{\underline{x}}_i = \langle \underline{x}_i, \underline{u} \rangle$$

$$= ({}^t\underline{x}_i \cdot \underline{u})$$

$$= \frac{1}{n}\sum_{i=1}^{n} \langle ({}^t\underline{x}_i \cdot \underline{u})\underline{u}, ({}^t\underline{x}_i \cdot \underline{u})\underline{u} \rangle \qquad \|\underline{u}\|^2 = 1$$

$$= \frac{1}{n}\sum_{i=1}^{n} ({}^t\underline{x}_i \cdot \underline{u})^2 \langle \underline{u}, \underline{u} \rangle$$

$$= \frac{1}{n}\sum_{i=1}^{n} ({}^t\underline{u} \cdot \underline{x}_i)({}^t\underline{x}_i \cdot \underline{u})$$

$$= {}^t\underline{u}\left(\sum_{i=1}^{n} \underline{x}_i\, {}^t x_i\right)\underline{u}$$

Réécriture de l'inertie expliquée (1)

$$I(\Delta_{\underline{u}}) = {}^t\underline{u}\left(\sum_{i=1}^{n} \frac{1}{n}\, \underline{x}_i\, {}^t\underline{x}_i\right)\underline{u}$$

$$= {}^t\underline{u}\; {}^tX D X\, \underline{u}$$

$$\underset{n\times p}{X} = \begin{array}{c c} & \text{variables} \\ \text{individus} & \begin{pmatrix} x_i^1 & x_i^2 & \cdots & x_i^p \end{pmatrix} \end{array} \qquad \underset{n\times n}{D} = \begin{pmatrix} 1/n & & 0 \\ & \ddots & \\ 0 & & 1/n \end{pmatrix}$$

$V = {}^tXDX$ matrice d'inertie

Caractérisation des axes $\Delta_{\underline{u}_1}, \Delta_{\underline{u}_2}, \Delta_{\underline{u}_3} \ldots$
$H^1 \quad H^2 \quad H^3$

qui maximisent l'inertie expliquée

$$I(\Delta_{\underline{u}}) = {}^t\underline{u}\, V\, \underline{u} \qquad V = {}^tX D X$$

V est symétrique et positive. Elle est diagonalisable

Il existe une base orthonormée $(\underline{u}_1, \underline{u}_2, \ldots, u_p) \in \mathbb{R}^p$
formée des vecteurs propres de V telle que

$$V \underline{u}_k = \lambda_k \underline{u}_k \qquad k = 1 \text{ à } p$$

$$I(\Delta_{\underline{u}}) = {}^t\underline{u}\,\mathbf{V}\,\underline{u}$$

? $\underline{u} \in \mathbb{R}^p$ qui maximise $I(\Delta_{\underline{u}})$

$$\mathbf{V} = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots & \lambda_p \end{pmatrix}$$

λ_i valeur propre associée au vecteur propre $\underline{u}_i$ de $\mathbf{V}$

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$$

$I(\Delta_{\underline{u}})$ est maximale pour $\underline{u} = \underline{u}_1$

$$I(\Delta_{\underline{u}_1}) = \lambda_1$$