

Algèbre Linéaire
en vue des applications en
Analyse de Données
et mise en pratique avec le
langage R
L2 MIASHS

Julien Ah-Pine (julien.ah-pine@univ-lyon2.fr)

Université Lyon 2

2016/2017

Motivations

- Les techniques d'analyse de données (ADD) traitent des tables de données composées de plusieurs lignes (individus) et de plusieurs colonnes (variables).
- Le but de ces techniques est de représenter de façon "efficace" et "intelligente", l'information contenue dans une table au travers de **graphiques** présentant les données dans un **espace géométrique de dimension faible**.
- Le terme **information** est ici de nature géométrique et repose principalement sur la notion de **variance** d'un nuage de points.
- Ces méthodes peuvent donc être vues comme des techniques de **visualisation** de données.
- Mais ces méthodes sont utilisées dans d'autres contextes également (compression de données, données en grande dimension...).

Rappel du Sommaire

- 1 Introduction
- 2 Algèbre Linéaire
- 3 Méthodes de réduction de dimension
- 4 Méthodes classiques de classification automatique (introduction)

Motivations (suite)

- Contrairement aux statistiques descriptives de base, ces méthodes visent à caractériser des données décrites par de très nombreuses variables (càd > 2). On parle d'ailleurs de **méthodes statistiques exploratoires multidimensionnelles**.
- Les chercheurs en statistiques français ont contribué énormément au domaine de l'ADD (J-P. Benzécri, B. Escofier, L. Lebart, ...).
- L'utilisation en **sciences humaines et sociales** de l'analyse de données est très fréquente :
 - ▶ Théorie de l'espace social de P. Bourdieu en sociologie,
 - ▶ Analyse de données de recensement de la population en démographie,
 - ▶ Analyse de données textuelles en linguistique,
 - ▶ Facteurs d'aptitude cognitive (facteur g) de Ch. Spearman (le précurseur des méthodes factorielles) en psychométrie,
 - ▶ De façon générale, analyse de résultats d'enquêtes en SHS,
 - ▶ ...

Objectifs du cours

- Nous commencerons par introduire les outils mathématiques nécessaires à l'ADD qui reposent sur l'algèbre linéaire (AL).
- Nous étudierons ensuite des techniques de réduction de dimensions :
 - ▶ Analyse en composantes principales (ACP),
 - ▶ Analyse factorielle des correspondances (AFC),
 - ▶ Analyse des correspondances multiples (ACM).
 - ▶ Analyse factorielle de données mixtes (AFDM) -si le temps le permet-.
- Nous verrons enfin des méthodes de classification automatique :
 - ▶ Méthode de partitionnement des k -moyennes (k -means)
 - ▶ Méthode de classification hiérarchique ascendante (formule de Lance-Williams)
- Pour la mise en pratique, nous utiliserons le **langage R** qui est l'un des outils privilégiés par les statisticiens dans le cadre du data science.

Quelques références faisant partie des sources du cours

- En AL :
 - ▶ G. Grifone. *Algèbre Linéaire (2ème édition)*. Cépadués. 2002.
 - ▶ A. Planche. *Mathématiques pour économistes (3ème édition)*. Dunod. 2004.
- En ADD :
 - ▶ B. Escofier, J. Pagés. *Analyses factorielles simples et multiples*. Dunod. 2008 (4ème Ed.)
 - ▶ L. Lebart, M. Piron, A. Morineau. *Statistique exploratoire multidimensionnelle*. Dunod. 2006 (4ème Ed.)
 - ▶ G. Saporta. *Probabilités, analyse des données et statistique*. Technip. 2011 (3ème Ed.)

Déroulement du cours

- 12 séances CM de 1h45 (~6 en AL et ~6 en ADD).
- 11 séances TD/TP de 1h45 (~5 en AL et ~5 en ADD + 1 Partiel).
- Le support de cours est mis à disposition sur le BV au sein du groupe L2_MIASHS_ADD-1617.

Rappel du Sommaire

- 1 Introduction
 - De la nécessité de l'AL en ADD

Le point de départ en ADD : une table de données

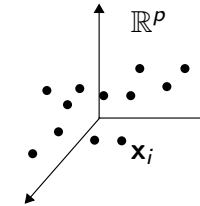
- Notons \mathbf{X} une matrice de données de taille $(n \times p)$:
$$\mathbf{X} = \begin{matrix} & \mathbf{x}^1 & \dots & \mathbf{x}^k & \dots & \mathbf{x}^p \\ \begin{matrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_i \\ \vdots \\ \mathbf{x}_n \end{matrix} & \begin{pmatrix} x_{11} & \dots & x_{1k} & \dots & x_{1p} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{ik} & \dots & x_{ip} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \dots & x_{nk} & \dots & x_{np} \end{pmatrix} \end{matrix}$$
- x_{ik} est le terme général de \mathbf{X} et $\forall i = 1, \dots, n; \forall k = 1, \dots, p : x_{ik} \in \mathbb{R}$.
- Chaque ligne i de \mathbf{X} représente un vecteur \mathbf{x}_i de taille $(p \times 1)$.
- Chaque colonne k de \mathbf{X} représente un vecteur \mathbf{x}^k de taille $(n \times 1)$.

$$\forall i : \mathbf{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix} \quad \text{et} \quad \forall k : \mathbf{x}^k = \begin{pmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{pmatrix}$$

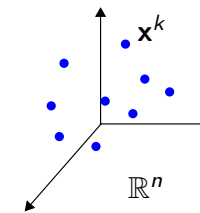
- On notera également un vecteur comme ceci : $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$.

Une table de données mais deux nuages de points

- L'ensemble des vecteurs $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ peut être représenté dans \mathbb{R}^p . On notera ce nuage $\mathbb{N}\mathbb{O}$.



- L'ensemble des vecteurs $\{\mathbf{x}^1, \dots, \mathbf{x}^p\}$ peut être représenté dans \mathbb{R}^n . On notera ce nuage $\mathbb{N}\mathbb{A}$.



Objectifs des méthodes de ce cours

- Pour fixer les idées supposons que les lignes représentent des individus et les colonnes des variables de différentes natures (par exemple sexe, âge, nombre d'enfants, salaire...)
- Pour appréhender l'information contenue dans la table de données, on s'intéresse aux nuages de points formés par les individus d'une part et par les variables d'autre part.
- Le point de vue géométrique de ces données permettrait d'apporter des éléments de réponses aux questions suivantes :
 - Quels sont les individus qui sont proches et qui présentent donc des similitudes ?
 - Quelles sont les variables corrélées entre elles ?
 - Peut-on représenter visuellement ces informations dans un espace dont les axes porteraient une sémantique particulière ?
 - Peut-on dégager des groupes homogènes d'individus et expliquer ces groupes par un comportement particulier de plusieurs variables ?

Objectifs des méthodes de ce cours (suite)

- Toutefois, il n'est pas possible d'appréhender ces données dans des espaces de dimension supérieure à 3 puisqu'on ne sait pas représenter visuellement ce type d'espace.
- On cherche donc à représenter ces données dans des espaces réduits de dimension 2 ou 3. Mais passer d'un espace de dimension p (ou n) à un espace de dimension 2 implique une perte d'information. On cherche donc des espaces réduits qui conservent au mieux l'information initiale.
- Les concepts de données, espace, projection et information sont abordés du point de vue géométrique. Pour cela on a recours à des outils mathématiques issus principalement de l'AL mais également de l'optimisation.

Rappel du Sommaire

1 Introduction

2 Algèbre Linéaire

3 Méthodes de réduction de dimension

4 Méthodes classiques de classification automatique (introduction)

Rappel du Sommaire

2 Algèbre Linéaire

- Espaces vectoriels
 - Définitions
 - Bases d'un espace vectoriel
 - Les théorèmes fondamentaux sur la dimension
 - Somme et somme directe de plusieurs sous-espaces
- La méthode du pivot de Gauss
- Applications linéaires et matrices
- Changement de base
- Déterminants
- Réduction des endomorphismes
- Formes bilinéaires et formes quadratiques
- Espace euclidien

Introduction

- En algèbre linéaire on manipule des objets mathématiques \mathbf{x} et \mathbf{y} appartenant à un ensemble \mathbb{E} tel que :
 - ▶ l'addition $\mathbf{x} + \mathbf{y}$ est un objet de \mathbb{E}
 - ▶ la multiplication, $\lambda \mathbf{x}$ où λ est un réel ou complexe, est également un objet de \mathbb{E} .
- De nombreux domaines scientifiques (différentes branches des mathématiques, physiques, chimie, économie, ...) présentent des problèmes dont le formalisme mathématique met en oeuvre ce type de structure algébrique.
- Ce cadre mathématique commun est la notion d'**espace vectoriel**.

Définition d'un espace vectoriel

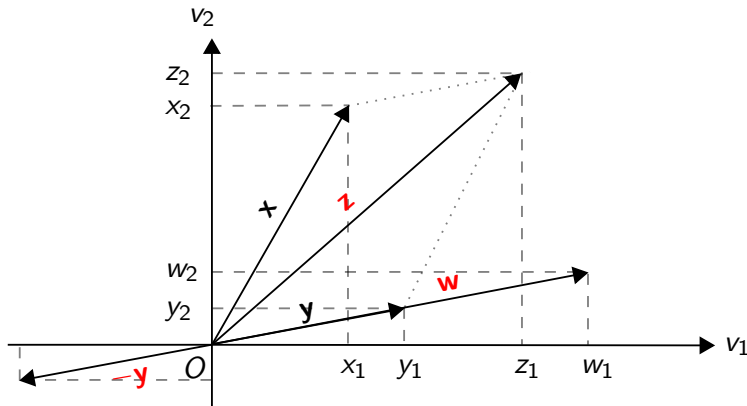
Définition.

Soit \mathbb{K} un corps commutatif (tel que \mathbb{R}). On appelle **espace vectoriel (ev)** sur \mathbb{K} un ensemble \mathbb{E} sur lequel on a **deux lois de composition** :

- 1 Une loi interne dite **addition** $+: \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{E}$ et vérifiant :
 - ▶ $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{E} : (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (associativité).
 - ▶ $\forall \mathbf{x}, \mathbf{y} \in \mathbb{E} : \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (commutativité).
 - ▶ il existe un élément de \mathbb{E} , $\mathbf{0}_{\mathbb{E}}$ ou plus simplement $\mathbf{0}$, dit **neutre** tel que : $\forall \mathbf{x} \in \mathbb{E} : \mathbf{x} + \mathbf{0} = \mathbf{x}$,
 - ▶ pour tout $\mathbf{x} \in \mathbb{E}$, il existe un élément de \mathbb{E} et noté $-\mathbf{x}$, dit **opposé** de \mathbf{x} tel que : $\forall \mathbf{x} \in \mathbb{E} : \mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
- 2 Une loi externe de domaine \mathbb{K} càd une application $\mathbb{K} \times \mathbb{E} \rightarrow \mathbb{E}$, dite **multiplication**, qui vérifie :
 - ▶ $\forall \lambda, \mu \in \mathbb{K}; \forall \mathbf{x} \in \mathbb{E} : \lambda(\mu \mathbf{x}) = (\lambda \mu) \mathbf{x}$ (associativité).
 - ▶ $\forall \lambda, \mu \in \mathbb{K}; \forall \mathbf{x} \in \mathbb{E} : (\lambda + \mu) \mathbf{x} = \lambda \mathbf{x} + \mu \mathbf{x}$ (distributivité).
 - ▶ $\forall \lambda \in \mathbb{K}; \forall \mathbf{x}, \mathbf{y} \in \mathbb{E} : \lambda(\mathbf{x} + \mathbf{y}) = \lambda \mathbf{x} + \lambda \mathbf{y}$ (distributivité).
 - ▶ $\forall \mathbf{x} \in \mathbb{E} : 1 \mathbf{x} = \mathbf{x}$ (1 étant l'élément neutre dans \mathbb{K}).

Exemple et représentation géométrique

- $\mathbf{x} = (x_1, x_2)$ et $\mathbf{y} = (y_1, y_2)$ deux vecteurs de $\mathbb{E} = \mathbb{R}^2$ d'origine $O = (0, 0)$.
- $\mathbf{z} = \mathbf{x} + \mathbf{y}$, $\mathbf{w} = 2\mathbf{y}$ sont deux autres vecteurs de \mathbb{R}^2 et $\mathbf{y} + (-\mathbf{y}) = \mathbf{0}$.



Autres exemples

- $\mathbb{E} = \mathbb{R}^n$ muni des deux lois suivantes :
 - $(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$.
 - $\lambda(x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n)$.
 Ici, $\mathbf{0} = (0, \dots, 0)$ et l'opposé de (x_1, \dots, x_n) est $(-x_1, \dots, -x_n)$.
- $\mathbb{E} = \mathcal{M}_2(\mathbb{K})$, l'ensemble des matrices carrées d'ordre 2 muni des deux lois suivantes :
 - $\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a+a' & b+b' \\ c+c' & d+d' \end{pmatrix}$.
 - $\lambda \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{pmatrix}$.
 Dans ce cas, $\mathbf{0} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ et l'opposé de $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est $\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$.

Sous-espaces vectoriels

Définition.

Soit \mathbb{E} un ev et \mathbb{F} une partie non vide de \mathbb{E} . On dit que \mathbb{F} est un **sous-espace vectoriel (sev)** de \mathbb{E} , si la restriction des lois de \mathbb{E} à \mathbb{F} fait de \mathbb{F} un ev.

Propriété.

Soit \mathbb{E} un ev et $\mathbb{F} \subseteq \mathbb{E}$. \mathbb{F} est un **sev** de \mathbb{E} ssi :

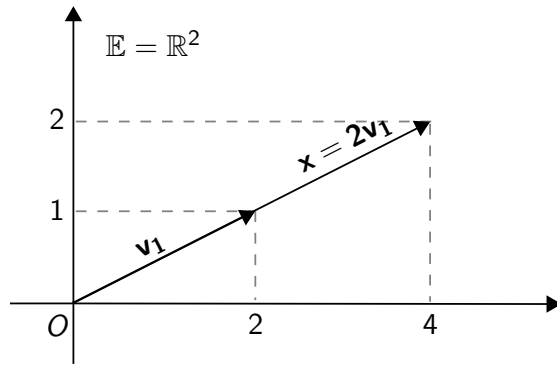
- $\mathbb{F} \neq \emptyset$
- $\mathbf{x}, \mathbf{y} \in \mathbb{F}; \lambda, \mu \in \mathbb{K} \Rightarrow \lambda \mathbf{x} + \mu \mathbf{y} \in \mathbb{F}$.

Exemples fondamentaux de sous-espaces vectoriels

- **Droite vectorielle** : Soit $\mathbf{v}_1 \in \mathbb{E}$, $\mathbf{v}_1 \neq \mathbf{0}$ alors :
 $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} \mid \exists \lambda \in \mathbb{K} : \mathbf{x} = \lambda \mathbf{v}_1\}$ est un sev de \mathbb{E} dit **droite vectorielle engendrée par \mathbf{v}_1** .
- **Plan vectoriel** : Soit $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{E}$, alors :
 $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} \mid \exists \lambda_1, \lambda_2 \in \mathbb{K} : \mathbf{x} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2\}$ est un sev de \mathbb{E} dit **plan vectoriel engendré par \mathbf{v}_1 et \mathbf{v}_2** .
- **Sous-espace engendré** : De manière générale, soient $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p \in \mathbb{E}$ alors
 $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} \mid \exists \lambda_1, \dots, \lambda_p \in \mathbb{K} : \mathbf{x} = \lambda_1 \mathbf{v}_1 + \dots + \lambda_p \mathbf{v}_p\}$ est un sev de \mathbb{E} dit **sous-espace engendré par $\mathbf{v}_1, \dots, \mathbf{v}_p$** .

Exemple et représentation géométrique

- $\mathbb{E} = \mathbb{R}^2$ d'origine $O = (0, 0)$.
- $\mathbf{v}_1 = (2, 1)$ et $\mathbb{F} = \{\mathbf{x} \in \mathbb{E} \mid \exists \lambda \in \mathbb{K} : \mathbf{x} = \lambda \mathbf{v}_1\}$.



- Autre définition de $\mathbb{F} : \mathbb{F} = \{\mathbf{x} \in \mathbb{E} \mid x_1 - 2x_2 = 0\}$.
 - ▶ $\mathbf{0} \in \mathbb{F}$
 - ▶ $\mathbf{x}, \mathbf{y} \in \mathbb{F}; \lambda, \mu \in \mathbb{K}$ sont tels que $\lambda \mathbf{x} + \mu \mathbf{y} \in \mathbb{F}$.

Résultats sur les sous-espaces vectoriels

Propriété.

Soient \mathbb{F} et \mathbb{G} deux sev de \mathbb{E} :

- $\mathbb{F} \cap \mathbb{G}$ est un sev de \mathbb{E}
- $\mathbb{F} \cup \mathbb{G}$ n'est pas en général un sev de \mathbb{E} .
- Le complémentaire $\mathbb{E} \setminus \mathbb{F}$ n'est pas un sev de \mathbb{E} .

Famille de vecteurs génératrice

Définition.

Une famille de vecteurs $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ d'un ev \mathbb{E} est dite **génératrice**, si $\mathbb{E} = \text{Vec}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ ce qui veut dire :

$$\forall \mathbf{x} \in \mathbb{E}, \exists \lambda_1, \dots, \lambda_p \in \mathbb{K} \text{ tel que : } \mathbf{x} = \lambda_1 \mathbf{v}_1 + \dots + \lambda_p \mathbf{v}_p$$

- On dit aussi que tout \mathbf{x} de \mathbb{E} est une **combinaison linéaire** des vecteurs \mathbf{v}_i .

Définition.

Un ev est dit de **dimension finie**, s'il existe une famille génératrice finie. Dans le cas contraire on dit qu'il est de dimension infinie.

Exemple

- Dans \mathbb{R}^2 , la famille $\{\mathbf{e}_1 = (1, 0), \mathbf{e}_2 = (0, 1)\}$ est une famille génératrice car tout $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ s'écrit :
 $(x_1, x_2) = x_1(1, 0) + x_2(0, 1) = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2$.
- Dans \mathbb{R}^2 , la famille $\{\mathbf{v}_1 = (1, 1), \mathbf{v}_2 = (1, -1)\}$ est une famille génératrice. En effet, soit $\mathbf{x} = (a, b) \in \mathbb{R}^2$ avec a et b quelconques. On peut montrer qu'il existe $x_1, x_2 \in \mathbb{R}$ tels que $\mathbf{x} = x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2$.
 En effet :

$$\begin{aligned} (a, b) &= x_1(1, 1) + x_2(1, -1) \\ &= (x_1 + x_2, x_1 - x_2) \end{aligned}$$
 On a donc $\begin{cases} a = x_1 + x_2 \\ b = x_1 - x_2 \end{cases}$ et on montre que tout $\mathbf{x} = (a, b)$ peut être combinaison linéaire de $\mathbf{v}_1, \mathbf{v}_2$. Il suffit pour cela de prendre :

$$\begin{cases} x_1 = \frac{a+b}{2} \\ x_2 = \frac{a-b}{2} \end{cases}.$$

Famille de vecteurs libres et liés

Définition.

Soit $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ une famille de vecteurs de \mathbb{E} . On dit qu'elle est **libre** si :

$$\lambda_1 \mathbf{v}_1 + \dots + \lambda_p \mathbf{v}_p = \mathbf{0} \Rightarrow \lambda_1 = 0, \dots, \lambda_p = 0$$

On dit aussi que les vecteurs $\mathbf{v}_1, \dots, \mathbf{v}_p$ sont **linéairement indépendants**. Une famille qui n'est pas libre est dite **liée**.

- Exemple : dans \mathbb{R}^3 , les vecteurs $\mathbf{v}_1 = (1, 2, 1)$, $\mathbf{v}_2 = (-1, 3, 1)$ et $\mathbf{v}_3 = (-1, 13, 5)$ sont liés car :

$$2 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + 3 \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix} - \begin{pmatrix} -1 \\ 13 \\ 5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Base d'un espace vectoriel

Définition.

$\mathbb{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est une **base** si les vecteurs $\mathbf{v}_1, \dots, \mathbf{v}_p$ forment une famille **libre et génératrice**.

Propriété.

Une famille $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est une **base** de l'espace \mathbb{E} , s.s.i. tout $\mathbf{x} \in \mathbb{E}$ se décompose d'une façon unique sur les \mathbf{v}_i :

$$\forall \mathbf{x} \in \mathbb{E}, \exists! (x_1, \dots, x_p) \in \mathbb{K}^p : \mathbf{x} = x_1 \mathbf{v}_1 + \dots + x_p \mathbf{v}_p$$

Propriété.

Si $\mathbb{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est une **base** de \mathbb{E} alors il existe une **bijection** : $\phi_{\mathbb{B}} : \mathbb{E} \rightarrow \mathbb{K}^p$ qui à tout vecteur $\mathbf{x} = x_1 \mathbf{v}_1 + \dots + x_p \mathbf{v}_p$ de \mathbb{E} lui fait correspondre un p -uplet (x_1, \dots, x_p) de \mathbb{K}^p .

Les scalaires x_i sont appelés les **composantes** de \mathbf{x} dans la base \mathbb{B} .

Résultats sur les familles libres

Propriété.

Une famille $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est **liée** s.s.i. l'un au moins des vecteurs \mathbf{v}_i s'écrit comme une combinaison linéaire des autres vecteurs de la famille (càd si au moins l'un des vecteurs appartient à l'ev engendré par les autres).

Propriété.

Soit $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ une **famille de vecteurs libres** et \mathbf{x} un vecteur quelconque de $\text{Vec}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ (càd \mathbf{x} est une combinaison linéaire des \mathbf{v}_i). Alors la **décomposition de \mathbf{x} sur les \mathbf{v}_i est unique**.

Démonstration.

Soient $\mathbf{x} = \lambda_1 \mathbf{v}_1 + \dots + \lambda_p \mathbf{v}_p$ et $\mathbf{x} = \mu_1 \mathbf{v}_1 + \dots + \mu_p \mathbf{v}_p$ deux décompositions de \mathbf{x} . En soustrayant, on obtient

$\mathbf{0} = (\lambda_1 - \mu_1) \mathbf{v}_1 + \dots + (\lambda_p - \mu_p) \mathbf{v}_p$. Les \mathbf{v}_i formant une famille libre on a : $\lambda_1 - \mu_1 = 0, \dots, \lambda_p - \mu_p = 0$ et donc $\lambda_1 = \mu_1, \dots, \lambda_p = \mu_p$. \square

Exemples de base

- Dans \mathbb{R}^2 , la famille $\{\mathbf{e}_1 = (1, 0), \mathbf{e}_2 = (0, 1)\}$ est une base (appelée **base canonique**). On a vu qu'elle était génératrice. Elle est également libre :

$$\begin{aligned} \lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 = \mathbf{0} &\Leftrightarrow \lambda_1 (1, 0) + \lambda_2 (0, 1) = \mathbf{0} \\ &\Leftrightarrow (\lambda_1 + 0, 0 + \lambda_2) = (0, 0) \\ &\Leftrightarrow \lambda_1 = 0 \text{ et } \lambda_2 = 0 \end{aligned}$$

- Dans $\mathcal{M}_2(\mathbb{K})$, l'ensemble des matrices carrées d'ordre 2, les matrices suivantes forment une base :

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} ; \mathbf{E}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} ; \mathbf{E}_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} ; \mathbf{E}_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Résultats sur les familles génératrices et libres

Propriété.

- Toute famille contenant une famille génératrice est génératrice.
- Toute sous-famille d'une famille libre est libre.
- Toute famille contenant une famille liée est liée.
- Toute famille $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ dont l'un des vecteurs \mathbf{v}_i est nul, est liée.

Dimension d'un espace vectoriel

Théorème.

Dans un ev \mathbb{E} sur \mathbb{K} de dimension finie, toutes les bases ont le même nombre d'éléments. Ce nombre est appelé **dimension** de \mathbb{E} sur \mathbb{K} et est notée $\dim(\mathbb{E})$.

Corollaire.

- Dans un ev de dimension p , toute famille de vecteurs ayant plus de p éléments est liée.
- Dans un ev de dimension p , toute famille de vecteurs ayant moins de p éléments ne peut être génératrice.
- Exemple : Dans \mathbb{R}^3 , la famille $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, (a, b, c)\}$ est liée puisque $(a, b, c) = a\mathbf{e}_1 + b\mathbf{e}_2 + c\mathbf{e}_3$. Par ailleurs, dans \mathbb{R}^3 , la famille $\{(1, 1, 0), (0, 0, 1)\}$ n'est pas génératrice puisque par exemple on ne peut décomposer le vecteur $(1, 2, 3)$ en fonction de ces derniers.

Théorème sur l'existence d'une base

Théorème.

Soit $\mathbb{E} \neq \emptyset$ un ev de dimension finie alors :

- De toute famille génératrice on peut extraire une base.
- Toute famille libre peut être complétée de manière à former une base (théorème de la base incomplète).
- Exemple : Dans \mathbb{R}^3 , les vecteurs $\mathbf{v}_1 = (1, 3, 0)$ et $\mathbf{v}_2 = (2, 1, 0)$ forme une famille libre mais non génératrice. On peut alors ajouter le vecteur canonique $\mathbf{e}_3 = (0, 0, 1)$ ce qui forme une famille libre et génératrice.

Dimension d'un ev et d'un sev

Théorème.

Soit \mathbb{E} un ev de dimension p alors :

- Toute famille génératrice ayant p éléments est une base.
- Toute famille libre ayant p éléments est une base.

Théorème.

Soit \mathbb{E} un ev de dimension finie et \mathbb{F} un sev de \mathbb{E} . Alors \mathbb{F} est de dimension finie et :

- $\dim(\mathbb{F}) \leq \dim(\mathbb{E})$.
- $\dim(\mathbb{F}) = \dim(\mathbb{E}) \Leftrightarrow \mathbb{F} = \mathbb{E}$.

Somme et Somme directe

Définition.

Soient $\mathbb{E}_1, \dots, \mathbb{E}_n$ des sev d'un même ev \mathbb{E} . On note :

$$\mathbb{E}_1 + \dots + \mathbb{E}_n = \{\mathbf{x} \in \mathbb{E} \mid \exists \mathbf{x}_1 \in \mathbb{E}_1, \dots, \exists \mathbf{x}_n \in \mathbb{E}_n : \mathbf{x} = \mathbf{x}_1 + \dots + \mathbf{x}_n\}$$

$\mathbb{E}_1 + \dots + \mathbb{E}_n$ est un sev de \mathbb{E} dit **somme des sous-espaces** \mathbb{E}_i .

Définition.

Soient $\mathbb{E}_1, \dots, \mathbb{E}_n$ des sev d'un même ev \mathbb{E} . On note :

$$\mathbb{E}_1 \oplus \dots \oplus \mathbb{E}_n = \{\mathbf{x} \in \mathbb{E} \mid \exists ! \mathbf{x}_1 \in \mathbb{E}_1, \dots, \exists ! \mathbf{x}_n \in \mathbb{E}_n : \mathbf{x} = \mathbf{x}_1 + \dots + \mathbf{x}_n\}$$

$\mathbb{E}_1 + \dots + \mathbb{E}_n$ est un sev de \mathbb{E} dit **somme directe des sous-espaces** \mathbb{E}_i .

Résultats sur la décomposition d'un ev en somme directe

Théorème.

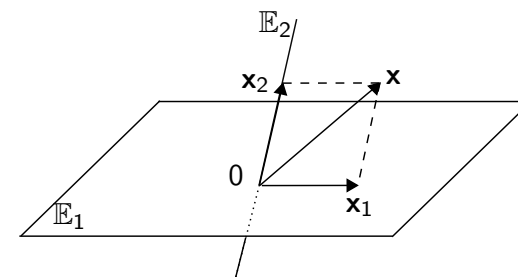
$\mathbb{E} = \mathbb{E}_1 \oplus \dots \oplus \mathbb{E}_n$ s.s.i. pour toute base $\mathbb{B}_1, \dots, \mathbb{B}_n$ de $\mathbb{E}_1, \dots, \mathbb{E}_n$, la famille $\mathbb{B} = \{\mathbb{B}_1, \dots, \mathbb{B}_n\}$ est une base de \mathbb{E} .

Théorème.

Soit \mathbb{E} un ev de dimension finie. Alors $\mathbb{E} = \mathbb{E}_1 \oplus \dots \oplus \mathbb{E}_n$ ssi :

- $\mathbb{E} = \mathbb{E}_1 + \dots + \mathbb{E}_n$.
- $\dim(\mathbb{E}) = \dim(\mathbb{E}_1) + \dots + \dim(\mathbb{E}_n)$.

Exemple



Rappel du Sommaire

2 Algèbre Linéaire

- Espaces vectoriels
- La méthode du pivot de Gauss
 - Systèmes d'équations linéaires et méthode du pivot
 - Espaces vectoriels et méthode du pivot
- Applications linéaires et matrices
- Changement de base
- Déterminants
- Réduction des endomorphismes
- Formes bilinéaires et formes quadratiques
- Espace euclidien

Introduction

- L'étude de la dépendance linéaire ou de l'indépendance d'un système de vecteurs est omniprésente en algèbre linéaire : détermination de bases, calcul de la dimension, mise en évidence de sommes directes, ...
- ⇒ Objectif : rappeler un algorithme simple, la **méthode du pivot de Gauss**, permettant de résoudre ce type de problème.

Matrice échelonnée

Définition.

Soit \mathbf{A} une matrice de dimension $p \times n$ d'éléments de \mathbb{K} . On dit que \mathbf{A} est **échelonnée** si les lignes commencent par un nombre de zéro strictement croissant à mesure que l'indice (des lignes) augmente.

- $\mathbf{A} = \begin{pmatrix} 2 & 1 & -2 & 3 \\ 3 & 2 & -1 & 2 \\ 3 & 3 & 3 & -3 \end{pmatrix}$ n'est pas échelonnée ;
- $\mathbf{A}' = \begin{pmatrix} 2 & 1 & -2 & 3 \\ 0 & 1 & 4 & -5 \\ 0 & 3 & 12 & -15 \end{pmatrix}$ n'est pas échelonnée ;
- $\mathbf{A}'' = \begin{pmatrix} 2 & 1 & -2 & 3 \\ 0 & 1 & 4 & -5 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ est échelonnée.

Système d'équations linéaires

- Un **système d'équations linéaires** est un système du type :

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pn}x_n = b_p \end{cases}$$

- Les a_{ij} et les b_i sont des éléments de \mathbb{K} (\mathbb{R}) donnés.
- Les x_i sont des inconnues de \mathbb{K} et résoudre le système signifie déterminer celles, s'il y en a, qui vérifient toutes les équations.
- On notera \mathbf{A} la matrice de taille $p \times n$ de terme général a_{ij} et \mathbf{b} le vecteur $p \times 1$ de terme général b_i :

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \dots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix}$$

Fondements et principe de la méthode du pivot

- La méthode du pivot est fondée sur les propriétés suivantes :

Propriété.

L'ensemble des solutions d'un système d'équations linéaires ne change pas si l'on effectue sur les équations les opérations élémentaires suivantes :

- ▶ Changer l'ordre des équations.
- ▶ Multiplier une équation (1er et 2nd membre) par un scalaire non nul.
- ▶ Ajouter à une équation une combinaison linéaire des autres équations.

- Principe de la méthode du pivot (connu aussi sous le nom d'élimination de Gauss-Jordan) :
 - ▶ La méthode consiste à mettre le système d'équations linéaires sous forme échelonnée (ie aboutir à une matrice des coefficients du systèmes qui soit échelonnée).
 - ▶ Une fois sous forme échelonnée, on peut partir de la solution de la dernière équation et en remontant résoudre toutes les équations.

$$\begin{matrix} l_1 \\ l_2 \\ l_3 \end{matrix} \left\{ \begin{array}{l} \textcolor{red}{2} \ x_1 + 1 \ x_2 - 2 \ x_3 + 3 \ x_4 = 1 \\ \mathbf{3} \ x_1 + 2 \ x_2 - 1 \ x_3 + 2 \ x_4 = 4 \\ \mathbf{3} \ x_1 + 3 \ x_2 + 3 \ x_3 - 3 \ x_4 = 5 \end{array} \right.$$

- $$\begin{cases} l_1 & 2x_1 + 1x_2 - 2x_3 + 3x_4 = 1 \\ l'_2 = 2l_2 - 3l_1 & \textcolor{red}{1}x_2 + 4x_3 - 5x_4 = 5 \\ l'_3 = 2l_3 - 3l_1 & \mathbf{3}x_2 + 12x_3 - 15x_4 = 7 \end{cases}$$

- $$l_1' = l_3' - 3l_2' \quad \left\{ \begin{array}{l} 2x_1 + 1x_2 - 2x_3 + 3x_4 = 1 \\ 1x_2 + 4x_3 - 5x_4 = 5 \\ 0 = -8 \end{array} \right.$$

- | | | | |
|--------------------------|---------------------------|-----------|----------|
| J. Ah-Pine (Univ-Lyon 2) | Alg. Lin. et Ana. de don. | 2016/2017 | 41 / 419 |
|--------------------------|---------------------------|-----------|----------|

$$\begin{array}{rccccccc} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots & \dots & \dots & \dots & + a_{1n}x_n & = & b_1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{ir}x_r + a_{ir+1}x_{r+1} + \dots & \dots & \dots & \dots & + a_{in}x_n & = & b_i \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{ps}x_s + \dots & \dots & \dots & \dots & + a_{pn}x_n & = & b_p \end{array}$$

J. Ah-Pine (Univ-Lyon 2)	Alg. Lin. et Ana. de don.	2016/2017	43 / 419
--------------------------	---------------------------	-----------	----------

J. Ah-Pine (Univ-Lyon 2)	Alg. Lin. et Ana. de don.	2016/2017	42 / 419
--------------------------	---------------------------	-----------	----------

$$\begin{matrix} l_1 \\ l_2 \\ l_3 \end{matrix} \left\{ \begin{array}{l} \textcolor{red}{1} x_1 + 2 x_2 - 2 x_3 + 3 x_4 = 2 \\ \textcolor{black}{2} x_1 + 4 x_2 - 3 x_3 + 4 x_4 = 5 \\ \textcolor{black}{5} x_1 + 10 x_2 - 8 x_3 + 11 x_4 = 12 \end{array} \right.$$

- $$\begin{matrix} l_1 \\ l'_2 = l_2 - 2l_1 \\ l'_3 = l_3 - 5l_1 \end{matrix} \left\{ \begin{array}{l} 1 \ x_1 + 2 \ x_2 - 2 \ x_3 + 3 \ x_4 = 2 \\ \textcolor{red}{1} \ x_3 - 2 \ x_4 = 1 \\ \textcolor{red}{2} \ x_3 - 4 \ x_4 = 2 \end{array} \right.$$

[illegible]

J. Ah-Pine (Univ-Lyon 2)	Alg. Lin. et Ana. de don.	2016/2017	44 / 419
--------------------------	---------------------------	-----------	----------

Exemple (suite)

- On enlève la dernière ligne :

$$\begin{array}{l} l_1 \\ l'_2 \end{array} \left\{ \begin{array}{l} 1 \ x_1 + 2 \ x_2 - 2 \ x_3 + 3 \ x_4 = 2 \\ + + 1 \ x_3 - 2 \ x_4 = 1 \end{array} \right.$$

x_1 et x_3 sont les inconnues principales et x_2 et x_4 sont des variables libres.

- L'ensemble des solutions s'expriment de la manière suivante :

$$\begin{cases} x_2, x_4 \in \mathbb{R} \\ x_3 - 2x_4 = 1 \\ x_1 + 2x_2 - 2x_3 + 3x_4 = 2 \end{cases} \Leftrightarrow \begin{cases} x_2, x_4 \in \mathbb{R} \\ x_3 = 1 + 2x_4 \\ x_1 = 2 - 2x_2 + 2(1 + 2x_4) - 3x_4 \end{cases} \Leftrightarrow \begin{cases} x_2, x_4 \in \mathbb{R} \\ x_3 = 1 + 2x_4 \\ x_1 = 4 - 2x_2 + x_4 \end{cases}$$

Système d'équations linéaires homogène

- Un **système d'équations linéaires homogène** est de type :

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = 0 \\ \vdots \\ a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pn}x_n = 0 \end{cases}$$

- Il s'agit d'un système d'équation linéaire pour lequel le vecteur $\mathbf{b} = \mathbf{0}$.
- Un tel système a au moins une solution le vecteur $\mathbf{x} = \mathbf{0}$ (solution triviale). Au delà de la solution triviale :

Théorème.

L'ensemble des solutions d'un système d'équations linéaires homogène à n inconnues est un sev de \mathbb{K}^n . Si le système sous forme échelonnée comporte k équations, l'espace des solutions est de dimension $n - k$. En particulier, un système homogène avec plus d'inconnues que d'équations (ie $n > p$) admet des solutions non nulles.

Exemple (suite)

- L'ensemble des solutions s'expriment également tel un sous-espace affine (sous-espace vectoriel que l'on translate par un vecteur donné) :

$$\begin{cases} x_2, x_4 \in \mathbb{R} \\ x_3 = 1 + 2x_4 \\ x_1 = 4 - 2x_2 + x_4 \end{cases} \Leftrightarrow \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 - 2x_2 + x_4 \\ x_2 \\ 1 + 2x_4 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \Leftrightarrow \mathbf{x} \in \text{Vec}\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \right\} + \begin{pmatrix} 4 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

Opérations élémentaires sur une famille de vecteurs

- On peut appliquer directement la méthode du pivot sur les familles de vecteurs.

Théorème.

Soit $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ une famille de vecteurs. L'espace qu'il engendre ne change pas si l'on effectue sur les vecteurs de la famille l'une des opérations élémentaires suivantes :

- Changer l'ordre des vecteurs.
- Multiplier un vecteur par un scalaire non nul.
- Ajouter à un vecteur une combinaison linéaire des autres vecteurs.

Matrice engendrée par une famille et base associée

Définition.

Soit $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ une famille de vecteurs de \mathbb{E} . On appelle matrice engendrée par les \mathbf{v}_i dans la base $\{\mathbf{e}_i\}$ la matrice dont les lignes sont les composantes des vecteurs $\mathbf{v}_1, \dots, \mathbf{v}_p$ dans la base $\{\mathbf{e}_i\}$.

Corollaire.

Les vecteurs lignes d'une matrice et les vecteurs lignes de la matrice correspondant à la mise sous forme échelonnée, engendrent le même espace vectoriel.

Théorème.

Soient $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ une famille de vecteurs, \mathbf{A} la matrice engendrée dans une base quelconque ($\{\mathbf{e}_i\}$ par exemple) et \mathbf{A}' la matrice issue de \mathbf{A} mise sous forme échelonnée. Alors les lignes non nulles de \mathbf{A}' donnent une base de $\text{Vec}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$.

Rappel du Sommaire

2 Algèbre Linéaire

- Espaces vectoriels
- La méthode du pivot de Gauss
- Applications linéaires et matrices
 - Rappels et généralités sur les applications
 - Applications linéaires et propriétés
 - Représentation matricielle des applications linéaires et propriétés
- Changement de base
- Déterminants
- Réduction des endomorphismes
- Formes bilinéaires et formes quadratiques
- Espace euclidien

Exemple

- Détermination d'une base du sous-espace de \mathbb{R}^4 engendré par les vecteurs : $\mathbf{v}_1 = (1, 1, 0, -1)$, $\mathbf{v}_2 = (-1, 1, 1, 0)$, $\mathbf{v}_3 = (0, 2, 1, -1)$.

$$\mathbf{A} = \begin{matrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{matrix} \begin{pmatrix} 1 & 1 & 0 & -1 \\ -1 & 1 & 1 & 0 \\ 0 & 2 & 1 & -1 \end{pmatrix} \rightarrow \begin{matrix} \mathbf{v}_1 \\ \mathbf{v}_2' \\ \mathbf{v}_3 \end{matrix} = \begin{matrix} \mathbf{v}_1 \\ \mathbf{v}_2 + \mathbf{v}_1 \\ \mathbf{v}_3 \end{matrix} \begin{pmatrix} 1 & 1 & 0 & -1 \\ 0 & 2 & 1 & -1 \\ 0 & 2 & 1 & -1 \end{pmatrix}$$

$$\rightarrow \begin{matrix} \mathbf{v}_1 \\ \mathbf{v}_2' \\ \mathbf{v}_3' \end{matrix} = \begin{matrix} \mathbf{v}_1 \\ \mathbf{v}_2' \\ \mathbf{v}_3 - \mathbf{v}_2' \end{matrix} \begin{pmatrix} 1 & 1 & 0 & -1 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \mathbf{A}'$$

- $\mathbf{v}_1 = (1, 1, 0, -1)$ et $\mathbf{v}_2' = (0, 2, 1, -1)$ forment une base de $\text{Vec}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.
- Par ailleurs, $\mathbf{v}_3' = \mathbf{0} \Rightarrow \mathbf{v}_2' - \mathbf{v}_3 = \mathbf{0} \Rightarrow \mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}$ (relation liant les vecteurs).

Introduction

- On rappelle les définitions générales des applications et leurs propriétés.
- On s'intéresse plus particulièrement aux applications linéaires qui sont des applications possédant des propriétés dites de linéarité.
- Une application linéaire entre deux ev \mathbb{E} et \mathbb{E}' permet de transformer les vecteurs de \mathbb{E} en des vecteurs de \mathbb{E}' .
- Toute application linéaire peut être représentée par une matrice et réciproquement toute matrice peut être associée à une application linéaire.
- On étudie donc deux représentations équivalentes des applications linéaires, une représentation algébrique consistant en des équations linéaires et une représentation matricielle.
- Parmi les applications linéaires particulières, on étudie celles qui correspondent à des changements de base.

Applications, injections, surjections

Définition.

Soient \mathbb{A} et \mathbb{B} deux ensembles. On appelle f une **application** de \mathbb{A} dans \mathbb{B} , un sous-ensemble du produit cartésien $\mathbb{A} \times \mathbb{B}$ tel que pour tout élément \mathbf{a} de \mathbb{A} correspond un élément \mathbf{b} de \mathbb{B} . On écrira alors :
$$\left\{ \begin{array}{l} f : \mathbb{A} \rightarrow \mathbb{B} \\ \mathbf{a} \rightarrow \mathbf{b} \end{array} \right.$$

Définition.

Une application $f : \mathbb{A} \rightarrow \mathbb{B}$ est dite **surjective** si $\forall \mathbf{b} \in \mathbb{B}, \exists \mathbf{a} \in \mathbb{A} : f(\mathbf{a}) = \mathbf{b}$.

Définition.

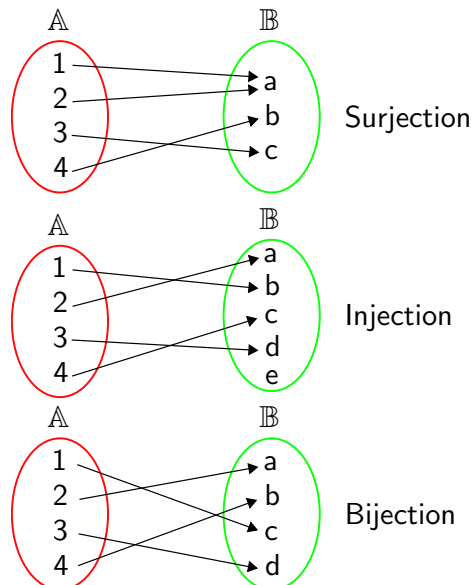
Une application $f : \mathbb{A} \rightarrow \mathbb{B}$ est dite **injective** si $\forall \mathbf{a}, \mathbf{a}' \in \mathbb{A} : \mathbf{a} \neq \mathbf{a}' \Rightarrow f(\mathbf{a}) \neq f(\mathbf{a}')$ ce qui est équivalent à $\forall \mathbf{a}, \mathbf{a}' \in \mathbb{A} : f(\mathbf{a}) = f(\mathbf{a}') \Rightarrow \mathbf{a} = \mathbf{a}'$.

Bijections

Définition.

$f : \mathbb{A} \rightarrow \mathbb{B}$ est dite **bijjective** si elle est surjective et injective. Formellement, f est une bijection si $\forall \mathbf{b} \in \mathbb{B}, \exists ! \mathbf{a} \in \mathbb{A} : f(\mathbf{a}) = \mathbf{b}$.

Illustration



Application linéaire

Définition. (Application linéaire)

Soient \mathbb{E} et \mathbb{E}' deux ev sur le même corps \mathbb{K} et f une **application** de \mathbb{E} dans \mathbb{E}' . On dit que f est **linéaire** si :

- ① $\forall \mathbf{x}, \mathbf{y} \in \mathbb{E} : f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$.
- ② $\forall \mathbf{x} \in \mathbb{E}, \forall \lambda \in \mathbb{K} : f(\lambda \mathbf{x}) = \lambda f(\mathbf{x})$.

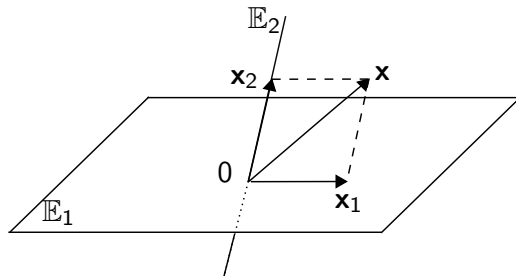
L'ensemble des applications linéaires de \mathbb{E} dans \mathbb{E}' est noté $\mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}')$ ou plus simplement $\mathcal{L}(\mathbb{E}, \mathbb{E}')$.

Définition. (Forme linéaire)

Une application linéaire de \mathbb{E} dans \mathbb{K} (le corps \mathbb{K} étant considéré tel un ev sur lui-même) est appelé plus particulièrement **forme linéaire**.

Exemples

- $\begin{cases} f: \mathbb{E} \rightarrow \mathbb{E}' \\ \mathbf{x} \rightarrow \mathbf{0} \end{cases}$ est l'application linéaire dite nulle :
 - ▶ $f(\mathbf{x} + \mathbf{y}) = \mathbf{0}$ et $f(\mathbf{x}) + f(\mathbf{y}) = \mathbf{0} + \mathbf{0} = \mathbf{0}$
 - ▶ $f(\lambda \mathbf{x}) = \mathbf{0}$ et $\lambda f(\mathbf{x}) = \lambda \mathbf{0} = \mathbf{0}$.
- Soit $\mathbb{E} = \mathbb{E}_1 \oplus \mathbb{E}_2$ donc $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ avec $\mathbf{x}_1 \in \mathbb{E}_1$ et $\mathbf{x}_2 \in \mathbb{E}_2$.
 $\begin{cases} p_1: \mathbb{E} \rightarrow \mathbb{E}_1 \\ \mathbf{x}_1 + \mathbf{x}_2 \rightarrow \mathbf{x}_1 \end{cases}$ est une application linéaire dite **projecteur sur \mathbb{E}_1 parallèlement à \mathbb{E}_2** .



Exemples (suite)

- $\begin{cases} f: \mathbb{R}^3 \rightarrow \mathbb{R}^2 \\ (x_1, x_2, x_3) \rightarrow (2x_1 + x_2, x_2 - x_3) \end{cases}$
- Il s'agit d'une application linéaire car la linéarité est vérifiée :

$$\begin{aligned}
 f(\lambda \mathbf{x} + \mu \mathbf{y}) &= f(\lambda(x_1, x_2, x_3) + \mu(y_1, y_2, y_3)) \\
 &= f(\lambda x_1 + \mu y_1, \lambda x_2 + \mu y_2, \lambda x_3 + \mu y_3) \\
 &= (2(\lambda x_1 + \mu y_1) + (\lambda x_2 + \mu y_2), (\lambda x_2 + \mu y_2) - (\lambda x_3 + \mu y_3)) \\
 &= ((2\lambda x_1 + \lambda x_2) + (2\mu y_1 + \mu y_2), (\lambda x_2 - \lambda x_3) + (\mu y_2 - \mu y_3)) \\
 &= (\lambda(2x_1 + x_2) + \mu(2y_1 + y_2), \lambda(x_2 - x_3) + \mu(y_2 - y_3)) \\
 &= \lambda f(\mathbf{x}) + \mu f(\mathbf{y})
 \end{aligned}$$

Remarque sur la linéarité

- Reprenons l'exemple d'application linéaire précédent :

$$\begin{cases} f: \mathbb{R}^3 \rightarrow \mathbb{R}^2 \\ (x_1, x_2, x_3) \rightarrow (2x_1 + x_2, x_2 - x_3) \end{cases}$$

- La **linéarité** de f tient au fait que **les composantes x_i dans l'espace d'arrivée (ici \mathbb{R}^2), apparaissent toutes à la puissance 1**.
En d'autres termes, chaque composante dans l'espace d'arrivée est un **polynôme homogène de degré 1** en les x_i .
- Ainsi l'exemple suivant n'est pas une application linéaire :

$$\begin{cases} f: \mathbb{R}^3 \rightarrow \mathbb{R} \\ (x_1, x_2, x_3) \rightarrow (x_1 x_2 + x_3) \end{cases}$$

On remarquera par exemple que $f(\lambda \mathbf{x}) \neq \lambda f(\mathbf{x})$.

Endomorphisme, isomorphisme et automorphisme

Définition.

On appelle un **endomorphisme** de \mathbb{E} , une application linéaire de \mathbb{E} dans \mathbb{E} (même espace d'arrivée et de départ). L'ensemble des endomorphismes de \mathbb{E} est noté $\text{End}_{\mathbb{K}}(\mathbb{E})$ ou plus simplement $\text{End}(\mathbb{E})$.

Définition.

On appelle un **isomorphisme** de \mathbb{E} sur \mathbb{E}' , une application linéaire **bijective** de \mathbb{E} dans \mathbb{E}' .

Définition.

On appelle un **automorphisme**, un endomorphisme **bijectif**.

Exemples

- $\begin{cases} id_{\mathbb{E}} : \mathbb{E} \rightarrow \mathbb{E} \\ \mathbf{x} \rightarrow \mathbf{x} \end{cases}$ est un automorphisme.
- $\begin{cases} f : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \\ (x_1, x_2, x_3) \rightarrow (\lambda_1 x_1, \lambda_2 x_2, \lambda_3 x_3) \end{cases}$
où $\lambda_1, \lambda_2, \lambda_3 > 0$ est un automorphisme.

Exemple

- Soit $\mathbb{E} = \mathbb{E}_1 \oplus \mathbb{E}_2$. Tout $\mathbf{x} \in \mathbb{E}$ se décompose donc de manière unique de la manière suivante : $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ avec $\mathbf{x}_1 \in \mathbb{E}_1$ et $\mathbf{x}_2 \in \mathbb{E}_2$. Alors :
 $\begin{cases} p_1 : \mathbb{E} \rightarrow \mathbb{E}_1 \\ \mathbf{x}_1 + \mathbf{x}_2 \rightarrow \mathbf{x}_1 \end{cases}$ est une application linéaire dite **projecteur sur \mathbb{E}_1 parallèlement à \mathbb{E}_2** .
 Dans ce cas, on a $Im(p_1) = \mathbb{E}_1$ et $Ker(p_1) = \mathbb{E}_2$.

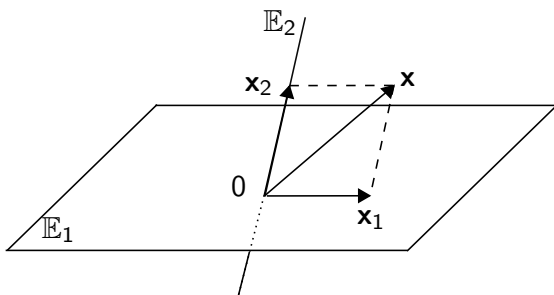


Image et noyau d'une application linéaire

Propriété.

Soit $f : \mathbb{E} \rightarrow \mathbb{E}'$ une application linéaire et \mathbb{F} un sev de \mathbb{E} . Alors $f(\mathbb{F})$ est un sev de \mathbb{E}' .

Définition.

En particulier, $f(\mathbb{E})$ est un sev de \mathbb{E}' appelé **image** de f et noté $Im(f)$. Sa dimension est appelée **rang** de f et est noté $rg(f)$ et on a :

$$rg(f) = \dim(Im(f))$$

Définition.

Soit $f \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$. Le **noyau** de f , noté $Ker(f)$, est un sev de \mathbb{E} défini comme suit :

$$Ker(f) = \{\mathbf{x} \in \mathbb{E} : f(\mathbf{x}) = \mathbf{0}\}$$

Propriété du noyau d'une application linéaire

Propriété.

Soit $f \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$. Alors f est **injective** s.s.i. $Ker(f) = \{\mathbf{0}\}$.

Démonstration.

- ⇐ Supposons $Ker(f) = \{\mathbf{0}\}$. Prenons $\mathbf{x}, \mathbf{y} \in \mathbb{E}$ tels que $f(\mathbf{x}) = f(\mathbf{y})$. On a $f(\mathbf{x}) - f(\mathbf{y}) = \mathbf{0}$ d'où $f(\mathbf{x} - \mathbf{y}) = \mathbf{0}$. Ainsi, $\mathbf{x} - \mathbf{y} \in Ker(f)$ et comme $Ker(f) = \{\mathbf{0}\}$, on a $\mathbf{x} = \mathbf{y}$ ce qui montre que f est injective.
- ⇒ Supposons que f est injective. Prenons $\mathbf{x} \in Ker(f)$, on a donc $f(\mathbf{x}) = \mathbf{0}$. On a par ailleurs $f(\mathbf{0}) = \mathbf{0}$ pour toute application linéaire. Donc $f(\mathbf{x}) = f(\mathbf{0})$. Comme f est injective, on a donc $\mathbf{x} = \mathbf{0}$ et donc $Ker(f) = \{\mathbf{0}\}$. □

Application linéaire et famille de vecteurs

Propriété.

Soient $f \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$ et $\{\mathbf{v}_i\}_{i \in \mathbb{I}}$ une famille de vecteurs de \mathbb{E} alors :

- Si f est injective et la famille $\{\mathbf{v}_i\}_{i \in \mathbb{I}}$ libre, alors la famille $\{f(\mathbf{v}_i)\}_{i \in \mathbb{I}}$ de \mathbb{E}' est libre.
- Si f est surjective et la famille $\{\mathbf{v}_i\}_{i \in \mathbb{I}}$ génératrice de \mathbb{E} , alors la famille $\{f(\mathbf{v}_i)\}_{i \in \mathbb{I}}$ est génératrice de \mathbb{E}' .

En particulier, si f est **bijective** (isomorphisme) alors **l'image d'une base de \mathbb{E} est une base de \mathbb{E}' .**

Théorèmes fondamentaux (suite)

Théorème du rang.

Supposons que $\dim(\mathbb{E}) = n$ et $\dim(\text{Ker}(f)) = r$. Montrons que $\dim(\text{Im}(f)) = n - r$. Soit $\{\mathbf{w}_1, \dots, \mathbf{w}_r\}$ une base de $\text{Ker}(f)$ et $\{\mathbf{v}_1, \dots, \mathbf{v}_{n-r}\}$ une famille de vecteurs tel que $\{\mathbf{w}_1, \dots, \mathbf{w}_r, \mathbf{v}_1, \dots, \mathbf{v}_{n-r}\}$ soit une base de \mathbb{E} .

Montrons alors que $\mathbb{B} = \{f(\mathbf{v}_1), \dots, f(\mathbf{v}_{n-r})\}$ est une base de $\text{Im}(f)$.

- \mathbb{B} engendre $\text{Im}(f)$. Soit $\mathbf{y} = f(\mathbf{x}) \in \text{Im}(f)$ et prenons $\mathbf{x} \in \mathbb{E}$. On a $\mathbf{x} = a_1\mathbf{w}_1 + \dots + a_r\mathbf{w}_r + b_1\mathbf{v}_1 + \dots + b_{n-r}\mathbf{v}_{n-r}$ et \mathbf{y} tel que :

$$\mathbf{y} = a_1f(\mathbf{w}_1) + \dots + a_rf(\mathbf{w}_r) + b_1f(\mathbf{v}_1) + \dots + b_{n-r}f(\mathbf{v}_{n-r})$$

$$= b_1f(\mathbf{v}_1) + \dots + b_{n-r}f(\mathbf{v}_{n-r})$$
- \mathbb{B} est libre. Supposons que $b_1f(\mathbf{v}_1) + \dots + b_{n-r}f(\mathbf{v}_{n-r}) = \mathbf{0}$. On a donc $f(b_1\mathbf{v}_1 + \dots + b_{n-r}\mathbf{v}_{n-r}) = \mathbf{0}$ d'où $b_1\mathbf{v}_1 + \dots + b_{n-r}\mathbf{v}_{n-r} \in \text{Ker}(f)$. Donc, il existe $a_1, \dots, a_r \in \mathbb{K}$ tels que : $b_1\mathbf{v}_1 + \dots + b_{n-r}\mathbf{v}_{n-r} = a_1\mathbf{w}_1 + \dots + a_r\mathbf{w}_r$. D'où $a_1\mathbf{w}_1 + \dots + a_r\mathbf{w}_r - b_1\mathbf{v}_1 - \dots - b_{n-r}\mathbf{v}_{n-r} = \mathbf{0}$. Puisque $\{\mathbf{w}_1, \dots, \mathbf{w}_r, \mathbf{v}_1, \dots, \mathbf{v}_{n-r}\}$ est libre alors $b_1 = \dots = b_{n-r} = 0$.

Théorèmes fondamentaux

Théorème.

Deux espaces vectoriels de dimension finie sont **isomorphes**^a s.s.i. ils ont même dimension.

a. c.à.d. il existe un isomorphisme entre ces deux ev

Théorème. (Théorème du rang)

Soient \mathbb{E} et \mathbb{E}' deux espaces vectoriels de dimension finie et $f : \mathbb{E} \rightarrow \mathbb{E}'$ une application linéaire. On a alors :

$$\dim(\mathbb{E}) = \text{rg}(f) + \dim(\text{Ker}(f))$$

Théorèmes fondamentaux (suite)

Corollaire.

Soit $f \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$ avec \mathbb{E} et \mathbb{E}' étant **deux ev de même dimension finie**^a. Alors les propriétés suivantes sont équivalentes :

- f est injective.
- f est surjective.
- f est bijective.

a. en particulier pour $f \in \text{End}(\mathbb{E})$ avec $\dim(\mathbb{E})$ fini

- Dans le cas où les dimensions des ev sont finies, pour montrer que $f \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$ est bijective il suffit de démontrer qu'elle est injective ou surjective.

Espaces vectoriels d'applications linéaires

Propriété.

$\mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}')$ ^a muni des lois :

$$\begin{cases} f + g : \mathbb{E} \rightarrow \mathbb{E}' \\ \mathbf{x} \rightarrow f(\mathbf{x}) + g(\mathbf{x}) \end{cases} \text{ et } \begin{cases} \lambda f : \mathbb{E} \rightarrow \mathbb{E}' \\ \mathbf{x} \rightarrow \lambda f(\mathbf{x}) \end{cases}$$

est un **espace vectoriel** sur \mathbb{K} .

a. ensemble des applications linéaires de \mathbb{E} dans \mathbb{E}'

- Autrement dit, l'ensemble des applications linéaires de l'ev \mathbb{E} dans l'ev \mathbb{E}' est un espace vectoriel.
- Nous allons montrer que nous pouvons représenter les applications linéaires entre ev par des matrices.

Rappels sur les matrices (suite)

- Sur l'ensemble $\mathcal{M}_{p,n}(\mathbb{K})$ on définit les lois suivantes :
 - ▶ **addition** : si $\mathbf{A} = (a_{ij})$ et $\mathbf{B} = (b_{ij})$ alors on note $\mathbf{C} = \mathbf{A} + \mathbf{B}$ la matrice $\mathbf{C} = (c_{ij})$ tel que $\forall i = 1, \dots, p; \forall j = 1, \dots, n : c_{ij} = a_{ij} + b_{ij}$.
 - ▶ **produit par un scalaire** : si $\mathbf{A} = (a_{ij})$ et $\lambda \in \mathbb{K}$ alors on note $\lambda \mathbf{A}$ la matrice de terme général (λa_{ij}) .

- Exemple :

$$\begin{pmatrix} 2 & 3 & -7 \\ 0 & -6 & 4 \end{pmatrix} + \begin{pmatrix} 1 & -1 & 6 \\ 0 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 2 & -1 \\ 0 & -7 & 4 \end{pmatrix},$$

$$2 \begin{pmatrix} 2 & 3 & -7 \\ 0 & -6 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 6 & -14 \\ 0 & -12 & 8 \end{pmatrix}$$

Rappels sur les matrices

Définition.

On appelle **matrice** de type (p, n) à coefficients dans \mathbb{K} un tableau \mathbf{A} de pn éléments de \mathbb{K} rangés sur p lignes et n colonnes :

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{pmatrix}$$

ou en abrégé $\mathbf{A} = (a_{ij})$.

L'ensemble des matrices à p lignes et n colonnes est noté $\mathcal{M}_{p,n}(\mathbb{K})$. Si $p = n$ on notera $\mathcal{M}_{n,n}(\mathbb{K})$ par $\mathcal{M}_n(\mathbb{K})$.

- Exemple : $\begin{pmatrix} 2 & 3 & -7 \\ 0 & -6 & 4 \end{pmatrix} \in \mathcal{M}_{2,3}(\mathbb{R})$, $\begin{pmatrix} 1-i & 5 \\ 4i & 3+2i \end{pmatrix} \in \mathcal{M}_2(\mathbb{C})$

Rappels sur les matrices (suite)

- $\mathcal{M}_{p,n}(\mathbb{K})$ muni de ces deux lois est un **ev** sur \mathbb{K} d'élément nul la matrice remplie de 0 et où l'opposé de (a_{ij}) est la matrice $(-a_{ij})$. Par ailleurs, on a :

$$\dim(\mathcal{M}_{p,n}(\mathbb{K})) = pn$$

- En effet, on vérifie facilement que les pn matrices élémentaires :

$$\mathbf{E}_{ij} = (e_{i'j'}) \text{ où } e_{i'j'} = \begin{cases} 1 & \text{si } i' = i \text{ et } j' = j \\ 0 & \text{sinon} \end{cases}$$

forment une base de $\mathcal{M}_{p,n}(\mathbb{K})$ dite **base canonique**.

Matrices et applications linéaires

- Soient \mathbb{E} et \mathbb{E}' deux ev sur \mathbb{K} , de dimension n et p respectivement et $f : \mathbb{E} \rightarrow \mathbb{E}'$ une application linéaire. Prenons une base $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ de \mathbb{E} et une base $\{\epsilon_1, \dots, \epsilon_p\}$ de \mathbb{E}' . Les images par f des vecteurs $\mathbf{e}_1, \dots, \mathbf{e}_n$ se décomposent sur la base $\{\epsilon_1, \dots, \epsilon_p\}$:

$$\begin{aligned} f(\mathbf{e}_1) &= a_{11}\epsilon_1 + a_{21}\epsilon_2 + \dots + a_{p1}\epsilon_p \\ f(\mathbf{e}_2) &= a_{12}\epsilon_1 + a_{22}\epsilon_2 + \dots + a_{p2}\epsilon_p \\ &\vdots \\ f(\mathbf{e}_n) &= a_{1n}\epsilon_1 + a_{2n}\epsilon_2 + \dots + a_{pn}\epsilon_p \end{aligned}$$

Définition.

On appelle **matrice de f dans les bases $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, $\{\epsilon_1, \dots, \epsilon_p\}$** , la matrice notée $M(f)_{\mathbf{e}_i, \epsilon_j}$ appartenant à $\mathcal{M}_{p,n}(\mathbb{K})$ dont les colonnes sont les composantes des vecteurs $f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)$ dans la base $\{\epsilon_1, \dots, \epsilon_p\}$.

Matrices et applications linéaires (suite)

Propriété.

Soient \mathbb{E} et \mathbb{E}' deux ev sur \mathbb{K} de dimension n et p respectivement, $\{\mathbf{e}_i\}$ et $\{\epsilon_j\}$ des bases de \mathbb{E} et \mathbb{E}' . Alors l'application :

$$\begin{cases} M : \mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{E}') \rightarrow \mathcal{M}_{p,n}(\mathbb{K}) \\ f \rightarrow M(f)_{\mathbf{e}_i, \epsilon_j} \end{cases}$$

est un **isomorphisme** d'espaces vectoriels càd M est bijective,

$$M(f + g) = M(f) + M(g) \text{ et } M(\lambda f) = \lambda M(f).$$

En particulier $\dim(\mathcal{L}(\mathbb{E}, \mathbb{E}')) = pn$

- Autrement dit :

- On peut représenter l'ensemble des applications linéaires d'un ev dans un autre par le biais de matrices et vice-versa.
- L'ensemble des applications linéaires de l'ev \mathbb{E} de dimension n dans l'ev \mathbb{E}' de dimension p est un ev de dimension pn .

Matrices et applications linéaires (suite)

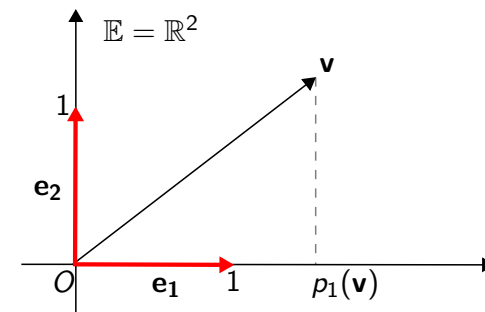
$$\begin{aligned} f(\mathbf{e}_1) &= a_{11}\epsilon_1 + a_{21}\epsilon_2 + \dots + a_{p1}\epsilon_p \\ f(\mathbf{e}_2) &= a_{12}\epsilon_1 + a_{22}\epsilon_2 + \dots + a_{p2}\epsilon_p \\ &\vdots \\ f(\mathbf{e}_n) &= a_{1n}\epsilon_1 + a_{2n}\epsilon_2 + \dots + a_{pn}\epsilon_p \end{aligned}$$

$$\updownarrow$$

$$M(f)_{\mathbf{e}_i, \epsilon_j} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \dots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{pmatrix}$$

- On utilisera également la notation $(f(\mathbf{e}_1), f(\mathbf{e}_2), \dots, f(\mathbf{e}_n))_{\epsilon_j}$. S'il n'y a pas d'ambiguïté on écrira aussi $M(f)$ mais il est clair que la matrice associée à f dépend du choix des bases de \mathbb{E} et \mathbb{E}' .

Exemple



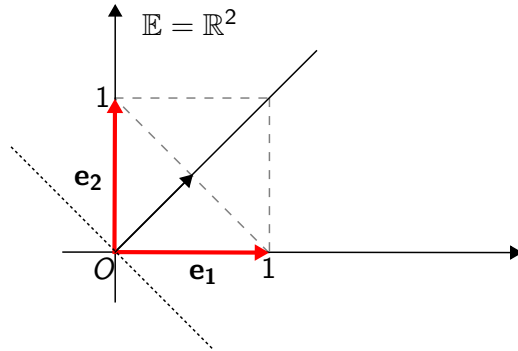
- Soit $\mathbb{E} = \mathbb{R}^2$ et :

$$\begin{cases} p_1 : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \\ (x_1, x_2) \rightarrow (x_1, 0) \end{cases}$$

Considérons la base canonique de \mathbb{R}^2 . On a $p_1(\mathbf{e}_1) = \mathbf{e}_1$ et $p_1(\mathbf{e}_2) = (0, 0)$.

$$\text{Donc : } M(p_1)_{\mathbf{e}_i} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Exemple (suite)



- Soit $E = \mathbb{R}^2$ et f la projection sur la première bissectrice parallèlement à la deuxième bissectrice. Considérons la base canonique de \mathbb{R}^2 . On a $f(e_1) = f(e_2) = \epsilon = \frac{1}{2}(e_1 + e_2)$.
Donc : $M(f)_{e_i} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

Rappels sur les matrices (suite)

- On appelle **produit** de matrices l'application :

$$\begin{cases} \mathcal{M}_{p,n}(\mathbb{K}) \times \mathcal{M}_{n,q}(\mathbb{K}) \rightarrow \mathcal{M}_{p,q}(\mathbb{K}) \\ \mathbf{A} \quad \quad \quad \mathbf{B} \quad \rightarrow \quad \mathbf{C} \end{cases}$$

où $\mathbf{C} = (c_{ij})$ est tel que $\forall i = 1, \dots, p, \forall j = 1, \dots, q :$

$$\begin{aligned} c_{ij} &= a_{i1}b_{1j} + a_{i2}b_{2j} \dots + a_{in}b_{nj} \\ &= \sum_{k=1}^n a_{ik}b_{kj} \end{aligned}$$

- Exemple :

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 0 & 2 \end{pmatrix} \text{ et } \mathbf{B} = \begin{pmatrix} 2 & 0 & 1 & 0 \\ 1 & 1 & 2 & 2 \\ 1 & 2 & 1 & 0 \end{pmatrix} \text{ et } \mathbf{C} = \mathbf{AB} = \begin{pmatrix} 4 & 1 & 1 & -2 \\ 4 & 4 & 3 & 0 \end{pmatrix}$$

Composition d'applications linéaires

Propriété.

- Soit $f \in \mathcal{L}_{\mathbb{K}}(E, E')$ et $g \in \mathcal{L}_{\mathbb{K}}(E', E'')$ alors la **composition** :

$$\begin{cases} g \circ f : E \rightarrow E'' \\ x \rightarrow g(f(x)) \end{cases} \text{ est un élément de } \mathcal{L}_{\mathbb{K}}(E, E'').$$
- Par ailleurs, $\forall f, h \in \mathcal{L}_{\mathbb{K}}(E, E') ; g, k \in \mathcal{L}_{\mathbb{K}}(E', E''), \forall \lambda \in \mathbb{K} :$
 - ▶ $g \circ (f + h) = g \circ f + g \circ h$
 - ▶ $(g + k) \circ f = g \circ f + k \circ f$
 - ▶ $g \circ (\lambda f) = \lambda g \circ f$
- Enfin si f est bijective f^{-1} (l'application réciproque -ou inverse- de f) est linéaire.
- Comme précédemment, nous allons montrer que nous pouvons représenter et déterminer les compositions d'applications linéaires entre ev en utilisant des matrices.

Rappels sur les matrices (suite)

- La multiplication matricielle est :
 - ▶ **Associative** : $\forall \mathbf{A} \in \mathcal{M}_{p,n}, \forall \mathbf{B} \in \mathcal{M}_{n,q}, \forall \mathbf{C} \in \mathcal{M}_{q,m},$

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$$
 - ▶ **Distributive** à gauche et à droite par rapport à l'addition :
 $\forall \mathbf{A}, \mathbf{D} \in \mathcal{M}_{p,n}, \forall \mathbf{B}, \mathbf{C} \in \mathcal{M}_{n,q},$

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \text{ et } (\mathbf{A} + \mathbf{D})\mathbf{B} = \mathbf{AB} + \mathbf{DB}$$

“Matrice” d’un vecteur

Définition.

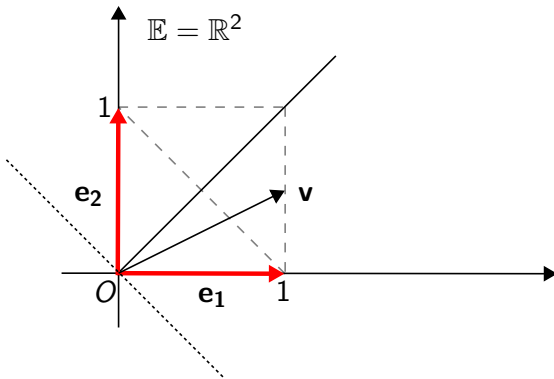
Soit \mathbb{E} un ev de dimension n , $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ une base de \mathbb{E} et $\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n$ un vecteur de \mathbb{E} . On appelle matrice de \mathbf{x} dans la base $\{\mathbf{e}_i\}$ la **matrice colonne des composantes de \mathbf{x} dans $\{\mathbf{e}_i\}$** :

$$M(\mathbf{x})_{\mathbf{e}_i} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

ce qui sera aussi noté plus simplement par $M(\mathbf{x})$ ou \mathbf{x} si pas d’ambiguïté.

- Un vecteur $\mathbf{x} \in \mathbb{E}$ est formellement défini comme combinaison linéaire des éléments de la base $\{\mathbf{e}_i\}$. La représentation $M(\mathbf{x})_{\mathbf{e}_i}$ est la matrice colonne ($n \times 1$) remplie des composantes de la combinaison linéaire définissant \mathbf{x} . En fait, par abus de langage on utilise le terme **vecteur \mathbf{x}** pour désigner la représentation matricielle de cet élément.

Exemple



- Soit $\mathbb{E} = \mathbb{R}^2$ et f la projection sur la première bissectrice parallèlement à la deuxième bissectrice. Considérons la base canonique de \mathbb{R}^2 . L’image du vecteur $\mathbf{v} = (1, \frac{1}{2})$ est donnée par : $M(f)_{\mathbf{e}_i} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} \\ \frac{3}{4} \end{pmatrix}$.

Calcul de l’image d’un vecteur

Propriété.

Soient \mathbb{E} et \mathbb{F} deux ev sur \mathbb{K} , $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ et $\{\epsilon_1, \dots, \epsilon_p\}$ deux bases de \mathbb{E} et \mathbb{F} respectivement. Pour toute application $f \in \mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{F})$ et pour tout vecteur $\mathbf{x} \in \mathbb{E}$, on a :

$$M(f(\mathbf{x}))_{\epsilon_j} = M(f)_{\epsilon_i, \epsilon_j} M(\mathbf{x})_{\mathbf{e}_i}$$

ou plus brièvement $M(f(\mathbf{x})) = M(f)M(\mathbf{x}) = M(f)\mathbf{x}$.

Composition d’applications linéaires et multiplication matricielle

Propriété.

Soient \mathbb{E} , \mathbb{F} et \mathbb{G} trois ev de dimension finie sur \mathbb{K} , $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, $\{\epsilon_1, \dots, \epsilon_p\}$ et $\{\eta_1, \dots, \eta_q\}$ des bases de \mathbb{E} , \mathbb{F} et \mathbb{G} respectivement. Si $g \in \mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{F})$ $f \in \mathcal{L}_{\mathbb{K}}(\mathbb{F}, \mathbb{G})$, alors on a :

$$M(f \circ g)_{\epsilon_i, \eta_k} = M(f)_{\epsilon_j, \eta_k} M(g)_{\mathbf{e}_i, \epsilon_j}$$

ou plus brièvement $M(f \circ g) = M(f)M(g)$.

Inverse d'une matrice carrée

Définition.

Une matrice carrée $\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$ est dite **inversible** s'il existe une matrice $\mathbf{A}' \in \mathcal{M}_n(\mathbb{K})$ telle que :

$$\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$$

où \mathbf{I} est la matrice diagonale d'ordre n .

\mathbf{A}' est dite **inverse** de \mathbf{A} et est notée \mathbf{A}^{-1} .

- Exemple :

L'inverse de $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}$ est $\mathbf{A}^{-1} = \begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix}$.

Rang d'une application linéaire et rang d'une matrice

- On a vu que le rang d'une application linéaire est la dimension de l'image de celle-ci. De plus, on a établi l'isomorphisme d'ev entre $\mathcal{L}_{\mathbb{K}}(\mathbb{E}, \mathbb{F})$ et $\mathcal{M}_{p,n}(\mathbb{K})$. Comment se traduit le rang d'une application linéaire par le biais de sa matrice associée ?

Définition.

- Soit $\{\mathbf{v}_i\}$ une famille de vecteurs. On appelle **rang de la famille** la dimension de l'espace engendré par les vecteurs $\{\mathbf{v}_i\}$
- Soit $\mathbf{A} \in \mathcal{M}_{p,n}(\mathbb{K})$, $\mathbf{A} = (\mathbf{c}_1, \dots, \mathbf{c}_n)$ où les \mathbf{c}_i sont les vecteurs colonnes de \mathbf{A} qui appartiennent à \mathbb{K}^p . On appelle **rang de la matrice \mathbf{A}** , le rang de la famille des vecteurs colonnes de \mathbf{A} :

$$\text{rg}(\mathbf{A}) = \text{rg}(\{\mathbf{c}_1, \dots, \mathbf{c}_n\}) = \dim(\text{Vec}\{\mathbf{c}_1, \dots, \mathbf{c}_n\})$$

Inverse d'une matrice carrée et isomorphisme

Propriété.

Soient \mathbb{E} et \mathbb{F} deux ev de même dimension n sur \mathbb{K} , $\{\mathbf{e}_i\}$ une base de \mathbb{E} , $\{\epsilon_j\}$ une base de \mathbb{F} . Une application linéaire $f : \mathbb{E} \rightarrow \mathbb{F}$ est **bijective** ssi $M(f)_{\mathbf{e}_i, \epsilon_j}$ est **inversible**. De plus :

$$M(f)_{\mathbf{e}_i, \epsilon_j}^{-1} = M(f^{-1})_{\epsilon_j, \mathbf{e}_i}$$

ou d'une manière plus concise : $M(f^{-1}) = M(f)^{-1}$.

Définition.

L'ensemble des matrices inversibles de $\mathcal{M}_n(\mathbb{K})$ est noté $GL(n, \mathbb{K})$ et est appelé **groupe linéaire**.

Rang d'une application linéaire et rang d'une matrice (suite)

Propriété.

Soient \mathbb{E} et \mathbb{F} deux ev de dimension finie et $f \in \mathcal{L}(\mathbb{E}, \mathbb{F})$. Soient $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ et $\{\epsilon_1, \dots, \epsilon_p\}$ deux bases quelconques de \mathbb{E} et \mathbb{F} respectivement et $\mathbf{A} = M(f)_{\mathbf{e}_i, \epsilon_j}$, on a alors :

$$\text{rg}(f) = \text{rg}(\mathbf{A})$$

- Deux matrices qui représentent la même application linéaire en des bases différentes ont même rang.
- Deux matrices semblables ont même rang.

Transposée d'une matrice et rang associé

Définition.

Soit $\mathbf{A} \in \mathcal{M}_{p,n}(\mathbb{K})$ et soit \mathbf{A}^\top la matrice de $\mathcal{M}_{n,p}(\mathbb{K})$ telle que les lignes de \mathbf{A}^\top sont les colonnes de \mathbf{A} . Alors \mathbf{A}^\top est dite la transposée de \mathbf{A} .

- Exemple :

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & -3 \\ 3 & 5 \end{pmatrix} \leftrightarrow \mathbf{A}^\top = \begin{pmatrix} 2 & 0 & 3 \\ 1 & -3 & 5 \end{pmatrix}$$

Propriété.

Pour toute matrice \mathbf{A} , on a :

$$\text{rg}(\mathbf{A}) = \text{rg}(\mathbf{A}^\top)$$

Exemple (suite)

- Mais** comme $\text{rg}(\mathbf{A}) = \text{rg}(\mathbf{A}^\top)$ on voit qu'il est préférable d'échelonner la matrice :

$$\begin{pmatrix} 1 & -1 & 3 & 5 & 1 \\ 2 & 0 & -1 & 3 & 1 \\ 3 & -1 & 2 & 8 & 2 \end{pmatrix}$$

- En échelonnant la matrice précédente, on obtient donc :

$$\text{rg}(\mathbf{A}) = \text{rg}(\mathbf{A}^\top) = 2$$

Exemple

- Calculez le rang de la matrice :

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 3 & 5 & 1 \\ 2 & 0 & -1 & 3 & 1 \\ 3 & -1 & 2 & 8 & 2 \end{pmatrix}$$

- Par définition, il faut calculer le rang des vecteurs colonnes :
 $\text{rg}(\mathbf{A}) = \text{rg}(\{\mathbf{c}_1, \dots, \mathbf{c}_5\}) = \dim(\text{Vec}\{\mathbf{c}_1, \dots, \mathbf{c}_5\})$.
- En **échelonnant** (méthode du **pivot de Gauss**) la matrice ci-dessous, on obtient la dimension de $\text{Vec}\{\mathbf{c}_1, \dots, \mathbf{c}_5\}$:

$$\begin{matrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \\ \mathbf{c}_4 \\ \mathbf{c}_5 \end{matrix} \begin{pmatrix} 1 & 2 & 3 \\ -1 & 0 & -1 \\ 3 & -1 & 2 \\ 5 & 3 & 8 \\ 1 & 1 & 2 \end{pmatrix}$$

Rappel du Sommaire

2 Algèbre Linéaire

- Espaces vectoriels
- La méthode du pivot de Gauss
- Applications linéaires et matrices
- Changement de base
 - Matrice de passage
 - Action des matrices de passage sur vecteurs et applications linéaires
- Déterminants
- Réduction des endomorphismes
- Formes bilinéaires et formes quadratiques
- Espace euclidien

Introduction

- La matrice qui représente une application linéaire a été construite à l'aide d'un choix des bases dans l'espace de départ et dans l'espace d'arrivée.
- Dans ce paragraphe, nous allons voir comment sont reliées deux matrices qui représentent la **même** application linéaire mais dans des **bases différentes**.
- De plus, le changement de base peut être lui même associé à une application linéaire. C'est en fait un **automorphisme**. Nous donnerons donc les écritures matricielles des actions des changements de base à des vecteurs et à des applications linéaires.

Matrice de passage (suite)

$$\begin{aligned}
 \mathbf{e}'_1 &= a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2 + \dots + a_{n1}\mathbf{e}_n \\
 \mathbf{e}'_2 &= a_{12}\mathbf{e}_1 + a_{22}\mathbf{e}_2 + \dots + a_{n2}\mathbf{e}_n \\
 &\vdots \\
 \mathbf{e}'_n &= a_{1n}\mathbf{e}_1 + a_{2n}\mathbf{e}_2 + \dots + a_{nn}\mathbf{e}_n
 \end{aligned}$$

$$\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

$$\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i} = M(\text{id}_{\mathbb{E}})_{\mathbf{e}'_i, \mathbf{e}_i}$$

Matrice de passage

- Soit \mathbb{E} un espace vectoriel de dimension n , $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ et $\{\mathbf{e}'_1, \dots, \mathbf{e}'_n\}$ deux bases de \mathbb{E} .
- On peut exprimer les vecteurs \mathbf{e}'_i comme combinaisons linéaires des vecteurs \mathbf{e}_j .

$$\begin{aligned}
 \mathbf{e}'_1 &= a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2 + \dots + a_{n1}\mathbf{e}_n \\
 \mathbf{e}'_2 &= a_{12}\mathbf{e}_1 + a_{22}\mathbf{e}_2 + \dots + a_{n2}\mathbf{e}_n \\
 &\vdots \\
 \mathbf{e}'_n &= a_{1n}\mathbf{e}_1 + a_{2n}\mathbf{e}_2 + \dots + a_{nn}\mathbf{e}_n
 \end{aligned}$$

Définition.

On appelle **matrice de passage** de la base $\{\mathbf{e}_i\}$ à la base $\{\mathbf{e}'_i\}$, la matrice notée $\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i}$ dont les colonnes sont les composantes des vecteurs \mathbf{e}'_i dans la base \mathbf{e}_j .

Propriétés des matrices de passage

Propriété.

- Propriété de **transitivité** :

$$\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i} \mathbf{P}_{\mathbf{e}'_i \rightarrow \mathbf{e}''_i} = \mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}''_i}$$

- Les matrices de passage sont **inversibles** et on a :

$$\left(\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i} \right)^{-1} = \mathbf{P}_{\mathbf{e}'_i \rightarrow \mathbf{e}_i}$$

Corollaire.

Soit \mathbb{E} un ev de dimension n , $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ et $\{\mathbf{e}'_1, \dots, \mathbf{e}'_n\}$ deux bases de \mathbb{E} . $\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i}$ est carrée d'ordre n et inversible donc, c'est un automorphisme de \mathbb{E} . Réciproquement tout automorphisme de \mathbb{E} correspond à un changement de base.

Action du changement de base sur les composantes d'un vecteur

- Soit $\mathbf{x} \in \mathbb{E}$, de composantes (x_1, \dots, x_n) dans la base $\{\mathbf{e}_i\}$ et de composantes (x'_1, \dots, x'_n) dans la base $\{\mathbf{e}'_i\}$.
- Il est facile de déterminer les relations entre les x_i et les x'_i à l'aide de la matrice de passage $\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i}$.
- Notons : $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = M(\mathbf{x})_{\mathbf{e}_i}$, $\mathbf{x}' = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = M(\mathbf{x})_{\mathbf{e}'_i}$ et $\mathbf{P} = \mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i}$.
- On a : $\mathbf{P}\mathbf{x}' = M(id_{\mathbb{E}})_{\mathbf{e}'_i, \mathbf{e}_i} M(\mathbf{x})_{\mathbf{e}'_i} = M(id_{\mathbb{E}}(\mathbf{x}))_{\mathbf{e}_i} = \mathbf{x}$.
et donc : $\mathbf{P}\mathbf{x}' = \mathbf{x}$ d'où $\mathbf{x}' = \mathbf{P}^{-1}\mathbf{x}$.

Exemple

- Soit \mathbb{R}^2 muni de deux bases canoniques $\{\mathbf{e}_1, \mathbf{e}_2\}$ et la base $\{\mathbf{e}'_1, \mathbf{e}'_2\}$ définie par :

$$\begin{cases} \mathbf{e}'_1 = 2\mathbf{e}_1 + \mathbf{e}_2 \\ \mathbf{e}'_2 = 3\mathbf{e}_1 + 2\mathbf{e}_2 \end{cases}$$
- Soit $\mathbf{x} = 2\mathbf{e}_1 + 3\mathbf{e}_2$. Déterminer les composantes de \mathbf{x} dans $\{\mathbf{e}'_1, \mathbf{e}'_2\}$.
 - 1 Méthode **matricielle** :

$$\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i} = \mathbf{P} = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}, \quad \mathbf{P}^{-1} = \begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix}, \quad \mathbf{x} = M(\mathbf{x})_{\mathbf{e}_i} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

En utilisant l'inverse \mathbf{P}^{-1} , on obtient :

$$\mathbf{x}' = \mathbf{P}^{-1}\mathbf{x} = \begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -5 \\ 4 \end{pmatrix}$$

Action du changement de base sur les composantes d'un vecteur (suite)

Propriété.

Soient $\mathbf{x} \in \mathbb{E}$, $\{\mathbf{e}_i\}$ et $\{\mathbf{e}'_i\}$ deux bases de \mathbb{E} , $\mathbf{P} = \mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i}$ et $\mathbf{x} = M(\mathbf{x})_{\mathbf{e}_i}$, $\mathbf{x}' = M(\mathbf{x})_{\mathbf{e}'_i}$ alors :

$$\mathbf{x}' = \mathbf{P}^{-1}\mathbf{x}$$

Exemple (suite)

- Soit \mathbb{R}^2 muni de deux bases canonique $\{\mathbf{e}_1, \mathbf{e}_2\}$ et la base $\{\mathbf{e}'_1, \mathbf{e}'_2\}$ défini par :

$$\begin{cases} \mathbf{e}'_1 = 2\mathbf{e}_1 + \mathbf{e}_2 \\ \mathbf{e}'_2 = 3\mathbf{e}_1 + 2\mathbf{e}_2 \end{cases}$$
- Soit $\mathbf{x} = 2\mathbf{e}_1 + 3\mathbf{e}_2$. Déterminer les composantes de \mathbf{x} dans $\{\mathbf{e}'_1, \mathbf{e}'_2\}$.
 - 2 Méthode **directe** : on résoud le système précédent,

$$\begin{cases} \mathbf{e}_1 = 2\mathbf{e}'_1 - \mathbf{e}'_2 \\ \mathbf{e}_2 = -3\mathbf{e}'_1 + 2\mathbf{e}'_2 \end{cases}$$

Puis en remplaçant dans l'expression initiale, on obtient :

$$\begin{aligned} \mathbf{x} &= 2\mathbf{e}_1 + 3\mathbf{e}_2 \\ &= 2(2\mathbf{e}'_1 - \mathbf{e}'_2) + 3(-3\mathbf{e}'_1 + 2\mathbf{e}'_2) \\ &= -5\mathbf{e}'_1 + 4\mathbf{e}'_2 \end{aligned}$$

Action du changement de base sur la représentation matricielle d'une application linéaire

Propriété.

Soient $f \in \mathcal{L}(\mathbb{E}, \mathbb{F})$, $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ et $\{\mathbf{e}'_1, \dots, \mathbf{e}'_n\}$ deux bases de \mathbb{E} , et $\{\epsilon_1, \dots, \epsilon_p\}$ et $\{\epsilon'_1, \dots, \epsilon'_p\}$ deux bases de \mathbb{F} .

Notons $\mathbf{A} = M(f)_{\mathbf{e}_i, \epsilon_i}$, $\mathbf{A}' = M(f)_{\mathbf{e}'_i, \epsilon'_i}$, $\mathbf{P} = \mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i}$ et $\mathbf{Q} = \mathbf{Q}_{\epsilon_i \rightarrow \epsilon'_i}$.

On a alors :

$$\mathbf{A}' = \mathbf{Q}^{-1} \mathbf{A} \mathbf{P}$$

Démonstration.

Soit $\mathbf{x} \in \mathbb{E}$ un vecteur arbitraire et notons $\mathbf{s} = M(\mathbf{x})_{\mathbf{e}_i}$. D'après les résultats précédents, on a :

$$M(f(\mathbf{x}))_{\epsilon'_j} = \mathbf{Q}^{-1} M(f(\mathbf{x}))_{\epsilon_j} = \mathbf{Q}^{-1} M(f)_{\mathbf{e}_i, \epsilon_j} M(\mathbf{x})_{\mathbf{e}_i} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{x}.$$

D'autre part, si $\mathbf{x} = M(\mathbf{x})_{\mathbf{e}_i}$:

$$M(f(\mathbf{x}))_{\epsilon'_j} = M(f)_{\mathbf{e}'_i, \epsilon'_j} M(\mathbf{x})_{\mathbf{e}'_i} = \mathbf{A}' \mathbf{x}' = \mathbf{A}' \mathbf{P}^{-1} \mathbf{x}.$$

Donc $\mathbf{A}' \mathbf{P}^{-1} \mathbf{x} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{x}$. □

Cas particulier des endomorphismes

Corollaire.

Soit $f \in \mathcal{L}(\mathbb{E})$ un endomorphisme de \mathbb{E} et $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ et $\{\mathbf{e}'_1, \dots, \mathbf{e}'_n\}$ deux bases de \mathbb{E} .

Notons $\mathbf{A} = M(f)_{\mathbf{e}_i}$, $\mathbf{A}' = M(f)_{\mathbf{e}'_i}$, $\mathbf{P} = \mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i}$.

On a alors :

$$\mathbf{A}' = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$$

Définition.

Deux matrices $\mathbf{A}, \mathbf{A}' \in \mathcal{M}_n(\mathbb{K})$ sont dites **semblables** s'il existe une matrice $\mathbf{P} \in \mathcal{M}_n(\mathbb{K})$ inversible telle que :

$$\mathbf{A}' = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$$

- Donc, deux matrices semblables représentent le même endomorphisme en des bases différentes.

Exemple

- Soit f l'endomorphisme de \mathbb{R}^3 qui dans la base canonique $\{\mathbf{e}_i\}$ est représenté par la matrice :

$$\mathbf{A} = M(f)_{\mathbf{e}_i} = \begin{pmatrix} 3 & -1 & 1 \\ 0 & 2 & 0 \\ 1 & -1 & 3 \end{pmatrix}$$

- Déterminer la matrice \mathbf{A}' qui représente f dans la base $\{\mathbf{e}'_i\}$ où :

$$\begin{cases} \mathbf{e}'_1 = (1, 0, -1) \\ \mathbf{e}'_2 = (0, 1, 1) \\ \mathbf{e}'_3 = (1, 0, 1) \end{cases}$$

- On a $\mathbf{A}' = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ avec $\mathbf{P} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix}$.

- D'après les résultats précédents $\mathbf{P}^{-1} = \mathbf{P}_{\mathbf{e}'_i \rightarrow \mathbf{e}_i}$. Pour déterminer \mathbf{P}^{-1} il faut donc exprimer $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ dans la base $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$.

Exemple (suite)

- On a :

$$\begin{cases} \mathbf{e}'_1 = \mathbf{e}_1 - \mathbf{e}_3 \\ \mathbf{e}'_2 = \mathbf{e}_2 + \mathbf{e}_3 \\ \mathbf{e}'_3 = \mathbf{e}_1 + \mathbf{e}_3 \end{cases}$$

- La résolution de ce système donne :

$$\begin{cases} \mathbf{e}_1 = \frac{1}{2}(\mathbf{e}'_1 + \mathbf{e}'_3) \\ \mathbf{e}_2 = \frac{1}{2}(\mathbf{e}'_1 + 2\mathbf{e}'_2 - \mathbf{e}'_3) \\ \mathbf{e}_3 = \frac{1}{2}(-\mathbf{e}'_1 + \mathbf{e}'_3) \end{cases} \text{ donc } \mathbf{P}^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 & -1 \\ 0 & 2 & 0 \\ 1 & -1 & 1 \end{pmatrix}$$

- En effectuant le produit $\mathbf{A}' = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$, on obtient :

$$\mathbf{A}' = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

- $\mathbf{A}' = M(f)_{\mathbf{e}'_i}$ ce qui veut dire que : $f(\mathbf{e}'_1) = 2\mathbf{e}'_1$, $f(\mathbf{e}'_2) = 2\mathbf{e}'_2$ et $f(\mathbf{e}'_3) = 4\mathbf{e}'_3$.

Rappel du Sommaire

2 Algèbre Linéaire

- Espaces vectoriels
- La méthode du pivot de Gauss
- Applications linéaires et matrices
- Changement de base
- Déterminants
- Réduction des endomorphismes
- Formes bilinéaires et formes quadratiques
- Espace euclidien

Définition des déterminants par récurrence

Définition.

Soit $\mathbf{A} = (a_{ij}) \in \mathcal{M}_n(\mathbb{K})$. On définit, par récurrence, une application $\det : \mathcal{M}_n(\mathbb{K}) \rightarrow \mathbb{K}$ de la manière suivante :

- Si $n = 1$, càd $\mathbf{A} = (a)$ alors $\det(\mathbf{A}) = a$.
- Si $n > 1$, notons $\mathbf{A}_{ij} \in \mathcal{M}_{n-1}(\mathbb{K})$ la matrice obtenue de \mathbf{A} en supprimant la i ème ligne et la j ème colonne (càd la ligne et la colonne passant par l'élément a_{ij}), on pose alors :

$$\det(\mathbf{A}) = a_{11}\det(\mathbf{A}_{11}) + \dots + (-1)^{k+1}a_{1k}\det(\mathbf{A}_{1k}) + \dots + (-1)^{n+1}a_{1n}\det(\mathbf{A}_{1n})$$

Le scalaire $\det(\mathbf{A})$ est dit **déterminant** de \mathbf{A} et :

Si $\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$ alors on dénote $\det(\mathbf{A})$ par $\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$.

Introduction

- Nous avons déjà mentionné que l'étude de la dépendance linéaire ou de l'indépendance d'un système de vecteurs est omniprésente en algèbre linéaire.
 - Nous avons étudié ce type de problème par le biais de la résolution de systèmes d'équations en utilisant la méthode du pivot.
- ⇒ Objectif : rappeler le concept et le calcul de **déterminant** de matrices qui donne une autre méthode permettant d'étudier la dépendance linéaire.
- Toutefois, à la différence de la méthode du pivot qui cherche à résoudre les systèmes, et dont la solution permet de conclure sur la dépendance ou l'indépendance linéaire, les déterminants permettent de répondre **directement** à la question de la dépendance ou de l'indépendance linéaire. Ils permettent ainsi de formaliser des résultats en algèbre linéaire que la méthode du pivot ne permet pas.

Déterminants de matrice de $\mathcal{M}_2(\mathbb{K})$

- Si $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ alors :

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

- Exemple :

$$\begin{aligned} \det \left(\begin{pmatrix} 4 & -1 \\ 3 & 2 \end{pmatrix} \right) &= \begin{vmatrix} 4 & -1 \\ 3 & 2 \end{vmatrix} \\ &= 4 \times 2 - (-1) \times 3 \end{aligned}$$

Déterminants de matrice de $\mathcal{M}_3(\mathbb{K})$

- Soit $\mathbf{A} = \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \end{pmatrix}$. On a :

$$\begin{aligned} \begin{vmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \end{vmatrix} &= a \begin{vmatrix} b' & b'' \\ c' & c'' \end{vmatrix} - a' \begin{vmatrix} b & b'' \\ c & c'' \end{vmatrix} + a'' \begin{vmatrix} b & b' \\ c & c' \end{vmatrix} \\ &= a(b'c'' - b''c') - a'(bc'' - b''c) + a''(bc' - b'c) \\ &= \textcolor{red}{ab'c''} + \textcolor{red}{a'b''c} + \textcolor{red}{a''bc'} - \textcolor{red}{ab''c'} - \textcolor{red}{a'bc''} - \textcolor{red}{a''b'c} \end{aligned}$$

- Règle de Sarrus :

$$\begin{array}{l|ll} l_1 & \textcolor{red}{a} & \textcolor{red}{a'} & \textcolor{red}{a''} \\ l_2 & \textcolor{red}{b} & \textcolor{red}{b'} & \textcolor{red}{b''} \\ l_3 & \textcolor{red}{c} & \textcolor{red}{c'} & \textcolor{red}{c''} \\ \hline l_1 & \textcolor{red}{a} & \textcolor{red}{a'} & \textcolor{red}{a''} \\ l_2 & \textcolor{red}{b} & \textcolor{red}{b'} & \textcolor{red}{b''} \end{array}$$

Déterminants de matrice de $\mathcal{M}_3(\mathbb{K})$ (exemple)

- Exemple : $\mathbf{A} = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 1 & -1 \\ 1 & 5 & 1 \end{pmatrix}$.

- Règle de Sarrus :

$$\begin{array}{l|ll} l_1 & 1 & -2 & 3 \\ l_2 & 2 & 1 & -1 \\ l_3 & 1 & 5 & 1 \\ \hline l_1 & 1 & -2 & 3 \\ l_2 & 2 & 1 & -1 \end{array}$$

- On a donc :

$$\det(\mathbf{A}) = (1 \times 1 \times 1) + (2 \times 5 \times 3) + (1 \times -2 \times -1) - (2 \times -2 \times 1) - (1 \times 5 \times -1) - (1 \times 1 \times 3) = 39.$$

Déterminant d'un ensemble de vecteurs

Théorème.

Soit $\mathbf{A} = (\mathbf{c}_1 \ \dots \ \mathbf{c}_n) \in \mathcal{M}_n(\mathbb{K})$. Alors les vecteurs $\mathbf{c}_1, \dots, \mathbf{c}_n$ forment une base de \mathbb{K}^n ssi $\det(\mathbf{A}) \neq 0$.

Autrement dit, un déterminant est nul ssi l'une des colonnes est combinaison linéaire des autres colonnes.

- Exemple : dans \mathbb{R}^3 , les vecteurs $\mathbf{v}_1 = (1, 2, 1)$, $\mathbf{v}_2 = (-1, 3, 1)$ et $\mathbf{v}_3 = (-1, 13, 5)$ forment-ils une base ?

- On calcule le déterminant :

$$\begin{aligned} \begin{vmatrix} 1 & -1 & -1 \\ 2 & 3 & 13 \\ 1 & 1 & 5 \end{vmatrix} &= 1 \begin{vmatrix} 3 & 13 \\ 1 & 5 \end{vmatrix} - (-1) \begin{vmatrix} 2 & 13 \\ 1 & 5 \end{vmatrix} + (-1) \begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} \\ &= (15 - 13) + (10 - 13) - (2 - 3) \end{aligned}$$

- Le déterminant étant nul, les vecteurs sont liés et ne forment donc pas une base de \mathbb{R}^3 .

Déterminant de la transposée d'une matrice

Théorème.

Pour toute matrice $\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$, on a :

$$\det(\mathbf{A}^\top) = \det(\mathbf{A})$$

Corollaire.

Toute les propriétés des déterminants relatives aux colonnes peuvent être affirmées pour les lignes :

- Le déterminant est une application multilinéaire par rapport à chaque ligne.
- Si une matrice a deux lignes égales, le déterminant est nul.
- Le déterminant d'une matrice est non nul ssi les vecteurs lignes sont indépendants.

Calcul des déterminants

- Nous avons présenté le calcul du déterminant en considérant le développement selon la première ligne. Compte tenu des résultats précédents, on peut considérer le calcul du déterminant en développant selon une ligne ou une colonne quelconque. Dans le cadre de cette généralisation on introduit :

Définition.

Soit $\mathbf{A} = (a_{ij}) \in \mathcal{M}_n(\mathbb{K})$. On appelle **cofacteur** de l'élément a_{ij} le scalaire :

$$\text{cof}(a_{ij}) = (-1)^{i+j} \det(\mathbf{A}_{ij})$$

- Exemple : $\mathbf{A} = \begin{pmatrix} 1 & 0 & -3 \\ 2 & 4 & -2 \\ 5 & -1 & 3 \end{pmatrix}$. Calculez $\text{cof}(a_{12})$.

$$\text{cof}(a_{12}) = \text{cof}(0) = (-1)^{1+2} \begin{vmatrix} 2 & -2 \\ 5 & 3 \end{vmatrix} = -16.$$

Propriétés

Propriété.

Le déterminant ne change pas si à une ligne (resp à une colonne), on ajoute une combinaison linéaire des autres lignes (resp des autres colonnes).

- En pratique, on fait usage de cette propriété afin de faire apparaître le plus de zéros sur les lignes ou les colonnes d'une matrice ce qui facilite le calcul du déterminant.

Calcul des déterminants (suite)

- En termes de cofacteurs, le développement du déterminant selon la première ligne s'écrit :

$$\det(\mathbf{A}) = a_{11}\text{cof}(a_{11}) + \dots + a_{1k}\text{cof}(a_{1k}) + \dots + a_{1n}\text{cof}(a_{1n}).$$
- On peut généraliser le développement à toute ligne ou colonne quelconque de \mathbf{A} :

Théorème.

- Développement de $\det(\mathbf{A})$ selon la i ème ligne :

$$\det(\mathbf{A}) = a_{i1}\text{cof}(a_{i1}) + \dots + a_{ik}\text{cof}(a_{ik}) + \dots + a_{in}\text{cof}(a_{in})$$

- Développement de $\det(\mathbf{A})$ selon la j ème colonne :

$$\det(\mathbf{A}) = a_{1j}\text{cof}(a_{1j}) + \dots + a_{kj}\text{cof}(a_{kj}) + \dots + a_{nj}\text{cof}(a_{nj})$$

Exemple

- Calculez $\det(\mathbf{A}) = \begin{vmatrix} 1 & -2 & 1 & 3 & 4 \\ 1 & -1 & 0 & 2 & 4 \\ 2 & 1 & 3 & 1 & 2 \\ -1 & 0 & 1 & 1 & 3 \\ 0 & 1 & -1 & 1 & 3 \end{vmatrix}$.
- On a $\det(\mathbf{A}) = \begin{vmatrix} 1 & -2 & 1 & 3 & 4 \\ 0 & 1 & -1 & -1 & 0 \\ 0 & 5 & 1 & -5 & -6 \\ 0 & -2 & 2 & 4 & 7 \\ 0 & 1 & -1 & 1 & 3 \end{vmatrix}$.
- Si on développe par rapport à la 1ère colonne on obtient :

$$\det(\mathbf{A}) = \begin{vmatrix} 1 & -1 & -1 & 0 \\ 5 & 1 & -5 & -6 \\ -2 & 2 & 4 & 7 \\ 1 & -1 & 1 & 3 \end{vmatrix}.$$

Exemple (suite)

- On a ensuite : $\det(\mathbf{A}) = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 5 & 6 & 0 & -6 \\ -2 & 0 & 2 & 7 \\ 1 & 0 & 2 & 3 \end{vmatrix}.$

- Si on développe maintenant par rapport à la 1ère ligne on obtient :

$$\det(\mathbf{A}) = \begin{vmatrix} 6 & 0 & -6 \\ 0 & 2 & 7 \\ 0 & 2 & 3 \end{vmatrix}.$$

- On développe par rapport à la 1ère colonne et on obtient :

$$\det(\mathbf{A}) = 6(6 - 14) = -48.$$

Déterminant d'un endomorphisme

Définition.

Soit \mathbb{E} un ev de dimension finie et f un endomorphisme de \mathbb{E} . On appelle **déterminant de f** , le déterminant de la matrice qui représente f dans une base (quelconque) de \mathbb{E} :

$$\det(f) = \det(M(f)_{\mathbf{e}_i})$$

$\{\mathbf{e}_i\}$ étant une base quelconque de \mathbb{E} .

Théorème.

$f \in \text{End}_{\mathbb{K}}(\mathbb{E})$ est un automorphisme ssi $\det(f) \neq 0$.

Déterminant du produit de matrices

Théorème.

Pour deux matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_n(\mathbb{K})$, on a :

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$$

Corollaire.

$\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$ est inversible ssi $\det(\mathbf{A}) \neq 0$ et on a alors :

$$\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$$

Corollaire.

Si $\mathbf{A}, \mathbf{A}' \in \mathcal{M}_n(\mathbb{K})$ sont deux matrices semblables (càd si elles représentent le même endomorphisme dans deux bases différentes) alors $\det(\mathbf{A}) = \det(\mathbf{A}')$.

Calcul de l'inverse d'une matrice

Théorème.

Soit $\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$ et $\text{cof}(\mathbf{A})$ la **comatrice** de \mathbf{A} (matrice des cofacteurs) dont le terme est défini par :

$$[\text{cof}(\mathbf{A})]_{ij} = \text{cof}(a_{ij})$$

On a alors le résultat suivant :

$$\mathbf{A}\text{cof}(\mathbf{A})^{\top} = \text{cof}(\mathbf{A})^{\top}\mathbf{A} = \det(\mathbf{A})\mathbf{I}$$

En particulier, si \mathbf{A} est inversible, on a :

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\text{cof}(\mathbf{A})^{\top}$$

Exemple

- Soit $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}$. On a $\det(\mathbf{A}) = 5$. Donc \mathbf{A} est inversible. On a $\text{cof}(\mathbf{A}) = \begin{pmatrix} 3 & 1 \\ -2 & 1 \end{pmatrix}$ et donc $\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \text{cof}(\mathbf{A})^\top = \frac{1}{5} \begin{pmatrix} 3 & -2 \\ 1 & 1 \end{pmatrix}$.
- Soit $\mathbf{A} = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 3 & 0 \\ 0 & 1 & -1 \end{pmatrix}$. On a $\det(\mathbf{A}) = -5$. Donc \mathbf{A} est inversible.
On a $\text{cof}(\mathbf{A}) = \begin{pmatrix} -3 & -1 & -1 \\ 2 & -1 & -1 \\ 0 & 0 & 5 \end{pmatrix}$ et donc $\mathbf{A}^{-1} = \frac{1}{-5} \begin{pmatrix} -3 & 2 & 0 \\ -1 & -1 & 0 \\ -1 & -1 & 5 \end{pmatrix}$.

Introduction

- Soit \mathbb{E} un ev sur \mathbb{K} et f un endomorphisme de \mathbb{E} . Si \mathbb{E} est de dimension finie et $\{\mathbf{e}_i\}$ est une base de \mathbb{E} , on peut construire la matrice qui représente f dans cette base : $M(f)_{\mathbf{e}_i} = (f(\mathbf{e}_1), \dots, f(\mathbf{e}_n))_{\mathbf{e}_i}$.
 - Comme on l'a vu précédemment, si l'on change de base, la matrice qui représente f change : si $\mathbf{A} = M(f)_{\mathbf{e}_i}$, $\mathbf{A}' = M(f)_{\mathbf{e}'_i}$ et $\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i}$ est la matrice de passage de la base $\{\mathbf{e}_i\}$ à la base $\{\mathbf{e}'_i\}$, on a : $\mathbf{A}' = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$.
- ⇒ Objectif : chercher des bases de \mathbb{E} dans lesquelles la matrice associée à f est la plus simple possible càd **diagonale**. Plus précisément, on dira que f est **diagonalisable** s'il existe une base $\{\mathbf{v}_i\}$ telle que :

$$M(f)_{\mathbf{v}_i} = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_{nn} \end{pmatrix}$$

Rappel du Sommaire

2 Algèbre Linéaire

- Espaces vectoriels
- La méthode du pivot de Gauss
- Applications linéaires et matrices
- Changement de base
- Déterminants
- Réduction des endomorphismes
 - Vecteurs propres
 - Recherche des valeurs propres
 - Recherche des vecteurs propres
 - Caractérisation des endomorphismes diagonalisables
- Formes bilinéaires et formes quadratiques
- Espace euclidien

Introduction (suite)

- En particulier, ce qui nous intéresse se décline en deux points :
 - ▶ Caractériser les endomorphismes diagonalisables.
 - ▶ Déterminer, si elles existent, les bases dans lesquelles la matrice est diagonale.
- En d'autres termes, puisque deux matrices \mathbf{A} et \mathbf{A}' liées par la relation $\mathbf{A}' = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ représentent le même endomorphisme en des bases différentes, alors ce qui nous intéresse peut s'énoncer de la manière suivante :
 - ▶ Caractériser les matrices $\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$ pour lesquelles il existe $\mathbf{P} \in \mathcal{M}_n(\mathbb{K})$ inversible, telles que $\mathbf{A}' = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ soit diagonale.
 - ▶ Déterminer effectivement \mathbf{P} et \mathbf{A}' .
- La clé de la diagonalisation est la notion de **vecteur propre**.

Définition d'un vecteur propre

Définition.

Soit $f \in \text{End}_{\mathbb{K}}(\mathbb{E})$. Un vecteur $\mathbf{v} \in \mathbb{E}$ est dit **vecteur propre** de f si :

- $\mathbf{v} \neq \mathbf{0}$.
- $\exists \lambda \in \mathbb{K} : f(\mathbf{v}) = \lambda \mathbf{v}$.

Le scalaire λ est dit **valeur propre** correspondante à \mathbf{v} .

- Les vecteurs propres sont par définition non nuls. Par contre la valeur propre peut être nulle : les vecteurs (non nuls) de $\text{Ker}(f)$ sont justement les vecteurs propres correspondant à $\lambda = 0$.
- Si \mathbf{v} est un vecteur propre correspondant à la valeur propre λ , alors pour tout $\mu \in \mathbb{K}, \mu \neq 0$, $\mu \mathbf{v}$ est aussi vecteur propre correspondant à la même valeur propre λ car on a : $f(\mu \mathbf{v}) = \mu f(\mathbf{v}) = \mu(\lambda \mathbf{v}) = \lambda(\mu \mathbf{v})$.
- Les vecteurs propres sont soit les vecteurs du noyau, soit les vecteurs qui ne changent pas de direction sous l'action de f .

Théorème

Théorème.

$f \in \text{End}_{\mathbb{K}}(\mathbb{E})$ est **diagonalisable** s.s.i. il existe une base de \mathbb{E} formée de vecteurs propres.

- Si $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ est une base formée de vecteurs propres de valeurs propres $\lambda_1, \dots, \lambda_n$, alors on a : $f(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$;

$$M(f)_{\mathbf{v}_i} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \text{ et donc } f \text{ est diagonalisable.}$$

- Réciproquement, s'il existe une base $\{\mathbf{e}_i\}$ telle que $M(f)_{\mathbf{e}_i}$ est

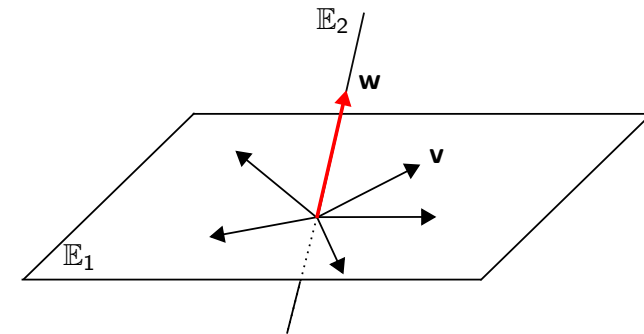
$$\text{diagonale } M(f)_{\mathbf{e}_i} = \begin{pmatrix} a_{11} & & \\ & \ddots & \\ & & a_{nn} \end{pmatrix} \text{ alors on a } f(\mathbf{e}_i) = a_{ii} \mathbf{e}_i \text{ ce qui}$$

signifie que les \mathbf{e}_i sont des vecteurs propres.

▷ **Diagonaliser** f c'est rechercher une **base de vecteurs propres**.

Exemple

- Soit $\mathbb{E} = \mathbb{R}^3$ et f la projection sur un plan \mathbb{E}_1 parallèlement à une droite \mathbb{E}_2 .
 - ▶ pour tout vecteur $\mathbf{v} \neq \mathbf{0}$ du plan, on a $f(\mathbf{v}) = \mathbf{v}$
 - ▶ pour tout vecteur $\mathbf{w} \in \mathbb{E}_2$, on a $f(\mathbf{w}) = \mathbf{0}$.



- On a donc deux types de vecteurs propres car on peut voir que tout autre vecteur change de direction sous l'action de f .

Préambule

- Soit λ une valeur propre de f ; il existe donc un vecteur $\mathbf{v} \in \mathbb{E}, \mathbf{v} \neq \mathbf{0}$ tel que $f(\mathbf{v}) = \lambda \mathbf{v}$, c'à d $(f - \lambda \text{id})\mathbf{v} = \mathbf{0}$. Comme $\mathbf{v} \neq \mathbf{0}$, cela signifie que l'endomorphisme $(f - \lambda \text{id})$ n'est pas injectif, ce qui équivaut à :

$$\det(f - \lambda \text{id}) = 0$$

- Aussi, si $\{\mathbf{e}_i\}$ est une base de \mathbb{E} et :

$$M(f)_{\mathbf{e}_i} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

- La condition pour que λ soit une valeur propre s'écrit :

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \dots & \ddots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix} = 0$$

Polynôme caractéristique

- En développant ce déterminant, on aboutit à une équation du type :

$$(-1)^n \lambda^n + b_{n-1} \lambda^{n-1} + \dots + b_1 \lambda + b_0 = 0$$

dont les racines¹ (dans \mathbb{K}) sont les valeurs propres de f .

- Cette équation est appelée **équation caractéristique** et le membre de gauche **polynôme caractéristique** de f noté $P_f(\lambda)$.

Propriété.

Soit $f \in \text{End}_{\mathbb{K}}(\mathbb{E})$ où \mathbb{E} est un ev de dimension finie n . Les valeurs propres de f sont les racines du polynôme :

$$P_f(\lambda) = \det(f - \lambda \text{id})$$

$P_f(\lambda)$ est un polynôme de degré n en λ appelé **polynôme caractéristique** de f .

1. Valeurs de λ pour lesquelles l'équation est vérifiée.

Spectre d'une application linéaire

Définition.

L'ensemble des valeurs propres de f est dit **spectre** de f et est noté $\text{Sp}_{\mathbb{K}}(f)$ (ou $\text{Sp}_{\mathbb{K}}(\mathbf{A})$, si \mathbf{A} est la matrice qui représente f dans une base).

- Remarque : $\lambda = 0$ est valeur propre de f s.s.i. $\det(f) = 0$, car $P_f(0) = \det(f)$.

Exemple

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'endomorphisme qui, dans la base canonique, est représentée par la matrice :

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$$

- On a :

$$P_f(\lambda) = \det(f - \lambda \text{id}) = \begin{vmatrix} 1-\lambda & 2 \\ -1 & 4-\lambda \end{vmatrix} = \lambda^2 - 5\lambda + 6$$

- On remarque que 2 est une racine $P_f(2) = 0$, on en déduit la factorisation suivante : $P_f(\lambda) = (\lambda - 2)(\lambda - 3)$.
- Donc les valeurs propres de f sont $\lambda_1 = 2$ et $\lambda_2 = 3$.
- Remarque : si $\mathbf{A} = M(f)_{e_i}$, $P_f(\lambda) = \det(\mathbf{A} - \lambda \text{id})$ sera noté aussi $P_{\mathbf{A}}(\lambda)$.

Rappels sur quelques propriétés

- Pour déterminer les racines d'un polynôme P , on pourra s'aider des propriétés remarquables sur les polynômes ou la propriété de distributivité de la multiplication sur l'addition qui permettent de factoriser les polynômes. Par exemple :
 - $ax^2 + bx = ax(x + \frac{b}{a})$
 - $x^2 + 2ax + a^2 = (x + a)^2$
 - $x^2 - 2ax + a^2 = (x - a)^2$
 - $x^2 - a^2 = (x + a)(x - a)$
- Par ailleurs on rappelle que les solutions d'une équation du second degré $ax^2 + bx + c = 0$ sont réelles si $\Delta = b^2 - 4ac \geq 0$ et dans ce cas :
 - Si $\Delta > 0$:
 - ★ $x_1 = \frac{-b - \sqrt{\Delta}}{2a}$
 - ★ $x_2 = \frac{-b + \sqrt{\Delta}}{2a}$
 - Si $\Delta = 0$:
 - ★ $x_1 = x_2 = \frac{-b}{2a}$

Exemple

- Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'endomorphisme qui, dans la base canonique, est représentée par la matrice :

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & -4 \end{pmatrix}$$

- On a :

$$P_{\mathbf{A}}(\lambda) = \det(f - \lambda \text{id}) = \begin{vmatrix} 1-\lambda & 2 \\ 3 & -4-\lambda \end{vmatrix} = \lambda^2 + 3\lambda - 10$$

- Résolution de l'équation du second degré : $\Delta = 3^2 - 4(-10) = 49$,
 $\lambda_1 = \frac{-3-\sqrt{49}}{2} = -5$, $\lambda_2 = \frac{-3+\sqrt{49}}{2} = 2$.
- Donc les valeurs propres de f sont $\lambda_1 = -5$ et $\lambda_2 = 2$.
- On peut écrire : $P_{\mathbf{A}}(\lambda) = (\lambda + 5)(\lambda - 2)$.

Exemple

- Reprenons l'exemple de $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'endomorphisme qui, dans la base canonique, est représentée par la matrice :

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$$

- On a vu que $Sp_{\mathbb{K}}(\mathbf{A}) = \{2, 3\}$. Il existe donc deux vecteurs propres \mathbf{v}_1 et \mathbf{v}_2 tels que : $f(\mathbf{v}_1) = \lambda_1 \mathbf{v}_1 = 2\mathbf{v}_1$ et $f(\mathbf{v}_2) = \lambda_2 \mathbf{v}_2 = 3\mathbf{v}_2$.
 - ▶ Calcul de \mathbf{v}_1 : il faut résoudre $f(\mathbf{v}_1) = 2\mathbf{v}_1$ càd $(f - 2\text{id})\mathbf{v}_1 = \mathbf{0}$ ou encore $(\mathbf{A} - 2\mathbf{I})\mathbf{v}_1 = \mathbf{0}$.
 - ▶ On a :

$$\begin{aligned} (\mathbf{A} - 2\mathbf{I})\mathbf{v}_1 = \mathbf{0} &\Leftrightarrow \begin{pmatrix} -1 & 2 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\Leftrightarrow \begin{cases} -x_1 + 2x_2 = 0 \\ -x_1 + 2x_2 = 0 \end{cases} \end{aligned}$$

- ▶ Les solutions sont donc engendrés par le vecteur $\mathbf{v}_1 = (2, 1)$.

Recherche des vecteurs propres par résolution de systèmes d'équations linéaires homogènes

- ▶ Une fois calculée les valeurs propres de \mathbf{A} (càd son spectre $Sp_{\mathbb{K}}(\mathbf{A})$), **on détermine les vecteurs propres en résolvant les systèmes linéaires homogènes** suivants :

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$$

où $\lambda \in Sp_{\mathbb{K}}(\mathbf{A})$.

- Si \mathbf{A} est diagonalisable alors les vecteurs propres $\{\mathbf{v}_i\}$ définissent la matrice de passage $\mathbf{P} = \mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{v}_i}$ telle que :

$$\mathbf{A}' = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$$

Exemple (suite)

- ▶ Calcul de \mathbf{v}_2 : il faut résoudre $f(\mathbf{v}_2) = 3\mathbf{v}_2$ càd $(f - 3\text{id})\mathbf{v}_2 = \mathbf{0}$ ou encore $(\mathbf{A} - 3\mathbf{I})\mathbf{v}_2 = \mathbf{0}$.
- ▶ On a :

$$\begin{aligned} (\mathbf{A} - 3\mathbf{I})\mathbf{v}_2 = \mathbf{0} &\Leftrightarrow \begin{pmatrix} -2 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\Leftrightarrow \begin{cases} -2x_1 + 2x_2 = 0 \\ -x_1 + x_2 = 0 \end{cases} \end{aligned}$$

- ▶ Les solutions sont donc engendrés par le vecteur $\mathbf{v}_2 = (1, 1)$.

- On a donc deux vecteurs propres $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ et $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Forment-ils une base de \mathbb{R}^2 ?
- Oui car $\det(\mathbf{v}_1, \mathbf{v}_2) = \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = 1 \neq 0$.

Exemple (suite)

- $\{\mathbf{v}_1, \mathbf{v}_2\}$ formant une base, par conséquent, f est diagonalisable et :

$$M(f)_{\mathbf{v}_i} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

- La matrice de passage $\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{v}_i}$ est telle que :

$$\mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{v}_i} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

- On vérifie facilement que : $\mathbf{A}' = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ avec $\mathbf{A}' = M(f)_{\mathbf{v}_i}$.

Espaces propres

Définition.

Soit $\lambda \in \mathbb{K}$. On note :

$$\mathbb{E}_\lambda = \{\mathbf{v} \in \mathbb{E} : f(\mathbf{v}) = \lambda \mathbf{v}\}$$

\mathbb{E}_λ est un sous-espace vectoriel de \mathbb{E} dit **espace propre** correspondant à λ .

- Si λ n'est pas valeur propre, $\mathbb{E}_\lambda = \{\mathbf{0}\}$.
- Si λ est valeur propre, alors :
 $\mathbb{E}_\lambda = \{\text{vecteurs propres associés à } \lambda\} \cup \{\mathbf{0}\}$ et $\dim(\mathbb{E}_\lambda) \geq 1$.

Propriétés des valeurs propres

Propriété.

Soit $\mathbf{A} \in \mathcal{M}_n(\mathbb{K})$ et $Sp(\mathbf{A}) = \{\lambda_1, \dots, \lambda_n\}$ (càd l'ensemble des racines dans \mathbb{K} égales ou distinctes du polynôme caractéristique). Alors :

- $tr(\mathbf{A})^a = \lambda_1 + \dots + \lambda_n$
- $det(\mathbf{A}) = \lambda_1 \dots \lambda_n$

a. tr est l'opérateur trace correspondant à la somme des termes de la diagonale

Preuve de la propriété sur le déterminant.

Comme $\{\lambda_1, \dots, \lambda_n\}$ sont les valeurs propres de f , ce sont les racines du polynôme caractéristique $P_f(x)$. On peut donc factoriser le polynôme caractéristique de la manière suivante :

$P_f(x) = (-1)^n(x - \lambda_1) \dots (x - \lambda_n)$. Pour $x = 0$: $P_f(0) = \lambda_1 \dots \lambda_n$. Or $P_f(0) = det(f - 0id) = det(f)$. \square

Espaces propres

Propriété.

Soient $\lambda_1, \dots, \lambda_p$ des scalaires deux à deux distincts. Alors les espaces propres $\mathbb{E}_{\lambda_1}, \dots, \mathbb{E}_{\lambda_p}$ sont en somme directe.

- Les espaces propres sont toujours en somme directe mais il se peut que leur somme "ne remplisse pas" totalement \mathbb{E} càd :
 $\mathbb{E}_{\lambda_1} \oplus \dots \oplus \mathbb{E}_{\lambda_p} \subset \mathbb{E}$.
- C'est en particulier le cas si $\dim(\mathbb{E}_{\lambda_1}) + \dots + \dim(\mathbb{E}_{\lambda_p}) < \dim(\mathbb{E})$. Dans ce cas f n'est pas diagonalisable.

Caractérisations des endomorphismes diagonalisables

Théorème.

Soit $f \in \text{End}_{\mathbb{K}}(\mathbb{E})$ et $\lambda_1, \dots, \lambda_p$ les valeurs propres de f . Alors les propriétés suivantes sont équivalentes :

- f est diagonalisable.
- \mathbb{E} est somme directe des espaces propres : $\mathbb{E} = \mathbb{E}_{\lambda_1} \oplus \dots \oplus \mathbb{E}_{\lambda_p}$
- $\dim(\mathbb{E}_{\lambda_1}) + \dots + \dim(\mathbb{E}_{\lambda_p}) = \dim(\mathbb{E})$.

Corollaire.

Si f admet n valeurs propres deux à deux distinctes alors f est diagonalisable.

Exemple

- Soit $\mathbf{A} = \begin{pmatrix} 2 & 0 & 4 \\ 3 & -4 & 12 \\ 1 & -2 & 5 \end{pmatrix}$.
- Calculez les valeurs propres de \mathbf{A} . \mathbf{A} est-elle diagonalisable ?
- Calcul des valeurs propres :

$$\begin{aligned} \begin{vmatrix} 2-\lambda & 0 & 4 \\ 3 & -4-\lambda & 12 \\ 1 & -2 & 5-\lambda \end{vmatrix} &= (2-\lambda) \begin{vmatrix} -4-\lambda & 12 \\ -2 & 5-\lambda \end{vmatrix} + 4 \begin{vmatrix} 3 & -4-\lambda \\ 1 & -2 \end{vmatrix} \\ &= (2-\lambda)((-4-\lambda)(5-\lambda) + 24) + 4(-6 + (4+\lambda)) \\ &= (2-\lambda)((-4-\lambda)(5-\lambda) + 20) \\ &= (2-\lambda)(\lambda^2 - \lambda) \\ &= -\lambda(\lambda-1)(\lambda-2) \end{aligned}$$

- $P_{\mathbf{A}}$ a trois racines distinctes. $\text{Sp}(\mathbf{A}) = \{0, 1, 2\}$. \mathbf{A} est diagonalisable.
- Déterminez les espaces propres associés aux trois valeurs propres.

Propriétés des espaces propres et diagonalisation

Propriété.

Soit $f \in \text{End}_{\mathbb{K}}(\mathbb{E})$ et λ une valeur propre d'ordre (multiplicité) α . Alors :

$$\dim(\mathbb{E}_{\lambda}) \leq \alpha$$

Théorème.

Soit $f \in \text{End}_{\mathbb{K}}(\mathbb{E})$. Alors f est **diagonalisable** ssi :

- 1 $P_f(x)$ s'écrit :

$$P_f(x) = (-1)^n (x - \lambda_1)^{\alpha_1} \dots (x - \lambda_p)^{\alpha_p}$$

avec $\lambda_1, \dots, \lambda_p \in \mathbb{K}$ et $\alpha_1 + \dots + \alpha_p = n$.

- 2 Les dimensions des espaces propres sont **maximales** càd : pour chaque valeur propre λ_i de multiplicité α_i , on a :

$$\dim(\mathbb{E}_{\lambda_i}) = \alpha_i$$

Exemple (suite)

- Pour $\lambda_1 = 0$:

$$\begin{aligned} (f - 0\text{id})\mathbf{v}_1 &= \mathbf{0} \Leftrightarrow \begin{cases} 2x_1 + 4x_3 &= 0 \\ 3x_1 - 4x_2 + 12x_3 &= 0 \\ x_1 - 2x_2 + 5x_3 &= 0 \end{cases} \\ &\Leftrightarrow \begin{cases} x_1 - 2x_2 + 5x_3 &= 0 \\ 2x_1 + 4x_3 &= 0 \\ 3x_1 - 4x_2 + 12x_3 &= 0 \end{cases} \\ &\Leftrightarrow \begin{cases} x_1 - 2x_2 + 5x_3 &= 0 \\ 4x_2 - 6x_3 &= 0 \\ 2x_2 - 3x_3 &= 0 \end{cases} \\ &\Leftrightarrow \begin{cases} x_1 - 2x_2 + 5x_3 &= 0 \\ 4x_2 - 6x_3 &= 0 \end{cases} \\ &\Leftrightarrow x_3 \in \mathbb{R}; x_2 = \frac{3}{2}x_3; x_1 = -2x_3 \end{aligned}$$

Exemple (suite)

- Pour $\lambda_1 = 0$: la solution du système homogène s'écrit $\begin{pmatrix} -2x_3 \\ \frac{3}{2}x_3 \\ x_3 \end{pmatrix}$. On

peut prendre par exemple $x_3 = 2$ et on a $\mathbf{v}_1 = \begin{pmatrix} -4 \\ 3 \\ 2 \end{pmatrix}$ d'où

$$\mathbb{E}_{\lambda_1} = \text{Vec}\left\{\begin{pmatrix} -4 \\ 3 \\ 2 \end{pmatrix}\right\}.$$

Exemple (suite)

- Pour $\lambda_2 = 1$: la solution du système homogène s'écrit $\begin{pmatrix} -4x_3 \\ 0 \\ x_3 \end{pmatrix}$. On

peut prendre par exemple $x_3 = 1$ et on a $\mathbf{v}_2 = \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}$ d'où

$$\mathbb{E}_{\lambda_2} = \text{Vec}\left\{\begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}\right\}.$$

Exemple (suite)

- Pour $\lambda_2 = 1$:

$$\begin{aligned} (f - 1id)\mathbf{v} = \mathbf{0} &\Leftrightarrow \begin{cases} 2x_1 + 4x_3 - x_1 &= 0 \\ 3x_1 - 4x_2 + 12x_3 - x_2 &= 0 \\ x_1 - 2x_2 + 5x_3 - x_3 &= 0 \end{cases} \\ &\Leftrightarrow \begin{cases} x_1 + 4x_3 &= 0 \\ 3x_1 - 5x_2 + 12x_3 &= 0 \\ x_1 - 2x_2 + 4x_3 &= 0 \end{cases} \\ &\Leftrightarrow \begin{cases} x_1 + 4x_3 &= 0 \\ -5x_2 &= 0 \\ -2x_2 &= 0 \end{cases} \\ &\Leftrightarrow \begin{cases} x_2 = 0 \\ x_3 \in \mathbb{R} \\ x_1 = -4x_3 \end{cases} \end{aligned}$$

Exemple (suite)

- Pour $\lambda_3 = 2$:

$$\begin{aligned} (f - 2id)\mathbf{v}_3 = \mathbf{0} &\Leftrightarrow \begin{cases} 2x_1 + 4x_3 - 2x_1 &= 0 \\ 3x_1 - 4x_2 + 12x_3 - 2x_2 &= 0 \\ x_1 - 2x_2 + 5x_3 - 2x_3 &= 0 \end{cases} \\ &\Leftrightarrow \begin{cases} 4x_3 &= 0 \\ 3x_1 - 6x_2 + 12x_3 &= 0 \\ x_1 - 2x_2 + 3x_3 &= 0 \end{cases} \\ &\Leftrightarrow \begin{cases} x_3 &= 0 \\ 3x_1 - 6x_2 &= 0 \\ x_1 - 2x_2 &= 0 \end{cases} \\ &\Leftrightarrow \begin{cases} x_3 = 0 \\ x_2 \in \mathbb{R} \\ x_1 = 2x_2 \end{cases} \end{aligned}$$

Exemple (suite)

- Pour $\lambda_3 = 2$: la solution du système homogène s'écrit $\begin{pmatrix} 2x_2 \\ 1x_2 \\ 0 \end{pmatrix}$. On

peut prendre par exemple $x_2 = 1$ et on a $\mathbf{v}_3 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$ d'où

$$\mathbb{E}_{\lambda_3} = \text{Vec}\left\{\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}\right\}.$$

Introduction

- Les formes bilinéaires sont des outils essentiels pour définir les espaces euclidiens. En effet, les notions d'angle, d'orthogonalité, de distances et de normes sont définies à partir du produit scalaire qui est une forme bilinéaire symétrique définie positive.
- Les formes quadratiques sont également importantes. Elle permettent d'étudier le signe d'une forme bilinéaire et de savoir notamment si elle est définie positive.
- Par ailleurs, nous étudions implicitement les formes quadratiques en ADD. En particulier, la matrice de variance-covariance ou des coefficients de corrélation définit une forme quadratique définie positive. La décomposition spectrale de ces matrices aboutissent aux axes factoriels (cf plus loin la partie ADD du cours).

Rappel du Sommaire

2 Algèbre Linéaire

- Espaces vectoriels
- La méthode du pivot de Gauss
- Applications linéaires et matrices
- Changement de base
- Déterminants
- Réduction des endomorphismes
- Formes bilinéaires et formes quadratiques
- Espace euclidien

Forme bilinéaire

Définition. (Forme bilinéaire)

Soit \mathbb{E} un ev sur \mathbb{K} . On appelle une **forme bilinéaire**, une application $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{K}$ qui est linéaire en les deux arguments, càd qui vérifie :

- $s(\mathbf{x} + \mathbf{y}, \mathbf{z}) = s(\mathbf{x}, \mathbf{z}) + s(\mathbf{y}, \mathbf{z})$
- $s(\mathbf{x}, \mathbf{y} + \mathbf{z}) = s(\mathbf{x}, \mathbf{y}) + s(\mathbf{x}, \mathbf{z})$
- $s(\lambda \mathbf{x}, \mathbf{y}) = s(\mathbf{x}, \lambda \mathbf{y}) = \lambda s(\mathbf{x}, \mathbf{y})$

Caractéristique d'une forme bilinéaire

- Soit \mathbb{E} un ev de dimension n , $\{\mathbf{e}_i\}$ une base de \mathbb{E} et s une forme bilinéaire de \mathbb{E} . Si $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$ et $\mathbf{y} = \sum_{j=1}^n y_j \mathbf{e}_j$, on a :

$$s(\mathbf{x}, \mathbf{y}) = s\left(\sum_{i=1}^n x_i \mathbf{e}_i, \sum_{j=1}^n y_j \mathbf{e}_j\right) = \sum_{i,j=1}^n x_i y_j s(\mathbf{e}_i, \mathbf{e}_j)$$

- Les $s(\mathbf{e}_i, \mathbf{e}_j)$ sont des éléments de \mathbb{K} . Si on pose $s_{ij} = s(\mathbf{e}_i, \mathbf{e}_j)$ alors l'expression de s dans la base $\{\mathbf{e}_i\}$ est :

$$s(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^n s_{ij} x_i y_j$$

càd s est du type :

$$s(\mathbf{x}, \mathbf{y}) = s_{11}x_1y_1 + s_{12}x_1y_2 + \dots + s_{ij}x_iy_j + \dots + s_{nn}x_ny_n$$

Caractéristique d'une forme bilinéaire symétrique

- Soit s une forme bilinéaire, elle est **symétrique** ssi :

$$s(\mathbf{x}, \mathbf{y}) = s(\mathbf{y}, \mathbf{x})$$

- Ainsi, on reconnaît qu'une forme bilinéaire est symétrique par le fait que, en l'écrivant dans une base, on a $\forall i, j = 1, \dots, n : s_{ij} = s_{ji}$.
- Exemples :
 - ▶ $s(\mathbf{x}, \mathbf{y}) = x_1y_1 + 2x_2y_2 - x_3y_1 + x_1y_3$ n'est pas une forme bilinéaire symétrique.
 - ▶ $s(\mathbf{x}, \mathbf{y}) = 2x_1y_1 - 3x_1y_2 - 3x_2y_1 + x_1y_3 + x_3y_1$ est une forme bilinéaire symétrique.

Caractéristique d'une forme bilinéaire

- Ainsi, si s est une forme bilinéaire elle peut être mise sous la forme :

$$s(\mathbf{x}, \mathbf{y}) = s_{11}x_1y_1 + s_{12}x_1y_2 + \dots + s_{ij}x_iy_j + \dots + s_{nn}x_ny_n$$

- Par conséquent, on reconnaît une forme bilinéaire par le fait que, en l'écrivant dans une base, on obtient une **somme de monômes dans lesquels x_i et y_j apparaissent à la puissance 1**.
- Exemples :
 - ▶ $s(\mathbf{x}, \mathbf{y}) = x_1y_1 + 2x_2y_2 - x_3y_1 + x_1y_3$ est une forme bilinéaire.
 - ▶ $s(\mathbf{x}, \mathbf{y}) = 2x_1y_1 + x_1^2y_2 - x_3y_2^2 + x_1y_3$ n'est pas une forme bilinéaire.
 - ▶ $s(\mathbf{x}, \mathbf{y}) = 2x_1y_1 + x_1 - x_3y_2$ n'est pas une forme bilinéaire.

Représentation matricielle d'une forme bilinéaire

- Pour les calculs, il est plus pratique d'utiliser l'**expression matricielle de la forme bilinéaire**.
- Soit \mathbb{E} un ev de dimension finie n sur \mathbb{K} et $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{K}$ une forme bilinéaire.
- Si $\{\mathbf{e}_i\}$ est une base de \mathbb{E} et $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$, $\mathbf{y} = \sum_{j=1}^n y_j \mathbf{e}_j$, on a :

$$s(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^n x_i y_j s(\mathbf{e}_i, \mathbf{e}_j)$$

- s est donc déterminée par la **connaissance des valeurs $s(\mathbf{e}_i, \mathbf{e}_j)$ sur une base**.

Matrice d'une forme bilinéaire

Définition.

Soient s une forme bilinéaire sur \mathbb{E} et $\{\mathbf{e}_i\}$ une base de \mathbb{E} . On appelle **matrice de s dans la base $\{\mathbf{e}_i\}$** la matrice :

$$M_b(s)_{\mathbf{e}_i} = \begin{pmatrix} s(\mathbf{e}_1, \mathbf{e}_1) & s(\mathbf{e}_1, \mathbf{e}_2) & \dots & s(\mathbf{e}_1, \mathbf{e}_n) \\ s(\mathbf{e}_2, \mathbf{e}_1) & s(\mathbf{e}_2, \mathbf{e}_2) & \dots & s(\mathbf{e}_2, \mathbf{e}_n) \\ \vdots & \vdots & \dots & \vdots \\ s(\mathbf{e}_n, \mathbf{e}_1) & s(\mathbf{e}_n, \mathbf{e}_2) & \dots & s(\mathbf{e}_n, \mathbf{e}_n) \end{pmatrix}$$

On écrira aussi $M_b(s)_{\mathbf{e}_i} = (s(\mathbf{e}_i, \mathbf{e}_j))$.

- Ainsi $s(\mathbf{e}_i, \mathbf{e}_j)$ correspond au coefficient de $x_i y_j$.
- L'application M_b qui à toute forme bilinéaire associe une matrice, ne doit pas être confondue avec l'isomorphisme M qui à toute application linéaire associe une matrice.

Exemples

- Soit $s : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ la forme bilinéaire qui dans la base canonique $\{\mathbf{e}_i\}$ s'écrit :
 $s(\mathbf{x}, \mathbf{y}) = 5x_1 y_1 - 2x_2 y_2 + 4x_3 y_3 + 7x_1 y_2 + 6x_1 y_3 - 4x_3 y_1 + 2x_2 y_3 + 8x_3 y_2$.

- On a alors :

$$M_b(s)_{\mathbf{e}_i} = \begin{pmatrix} 5 & 7 & 6 \\ 0 & -2 & 2 \\ -4 & 8 & 4 \end{pmatrix}$$

- L'image des vecteurs $\mathbf{x} = (1, 1, 0)$ et $\mathbf{y} = (0, 2, 4)$ est :

$$\begin{aligned} s(\mathbf{x}, \mathbf{y}) &= (1 \ 1 \ 0) \begin{pmatrix} 5 & 7 & 6 \\ 0 & -2 & 2 \\ -4 & 8 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 4 \end{pmatrix} \\ &= 42 \end{aligned}$$

Détermination des images par calcul matriciel

Propriété.

Soient \mathbb{E} un ev sur \mathbb{K} de base $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. Pour toute forme bilinéaire $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{K}$ représentée par la matrice $M_b(s)_{\mathbf{e}_i} = \mathbf{S}$ et pour tout vecteur $\mathbf{x}, \mathbf{y} \in \mathbb{E}$, on a :

$$\begin{aligned} s(\mathbf{x}, \mathbf{y}) &= M(\mathbf{x})_{\mathbf{e}_i}^\top M_b(s)_{\mathbf{e}_i} M(\mathbf{y})_{\mathbf{e}_i} \\ &= \mathbf{x}^\top \mathbf{S} \mathbf{y} \end{aligned}$$

Exemples (suite)

- Soit $s : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ la forme bilinéaire symétrique qui dans la base canonique $\{\mathbf{e}_i\}$ s'écrit : $s(\mathbf{x}, \mathbf{y}) = 2x_1 y_1 - 3x_1 y_2 - 3x_2 y_1 + x_1 y_3 + x_3 y_1$.
- On a alors :

$$M_b(s)_{\mathbf{e}_i} = \begin{pmatrix} 2 & -3 & 1 \\ -3 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

- L'image des vecteurs $\mathbf{x} = (1, 1, 0)$ et $\mathbf{y} = (0, 2, 4)$ est :

$$\begin{aligned} s(\mathbf{x}, \mathbf{y}) &= (1 \ 1 \ 0) \begin{pmatrix} 2 & -3 & 1 \\ -3 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 4 \end{pmatrix} \\ &= -2 \end{aligned}$$

- On pourra vérifier que $s(\mathbf{x}, \mathbf{y}) = s(\mathbf{y}, \mathbf{x})$.

Forme quadratique

- A toute forme bilinéaire symétrique $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}$, on peut lui associer une application $q : \mathbb{E} \rightarrow \mathbb{R}$.

Définition. (Forme quadratique)

Soit \mathbb{E} un ev sur \mathbb{K} et $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{K}$ une forme bilinéaire symétrique. Alors l'application $q : \mathbb{E} \rightarrow \mathbb{K}$ définie par :

$$q(\mathbf{x}) = s(\mathbf{x}, \mathbf{x})$$

est appelée **forme quadratique** sur \mathbb{E} associée à s .

- q n'est pas une forme possédant la propriété de linéarité.

Caractéristique d'une forme quadratique

- Soit s une forme bilinéaire symétrique et q la forme quadratique associée.
- On a vu comment reconnaître algébriquement une forme bilinéaire symétrique.
- Comme $q(\mathbf{x}) = s(\mathbf{x}, \mathbf{x})$, il est facile de voir que l'écriture algébrique de q dans une base de \mathbb{E} est un **polynôme homogène de degré 2** en les composantes x_i de \mathbf{x} .
- Exemple :
 - ▶ Si $s(\mathbf{x}, \mathbf{y}) = 2x_1y_1 - 3x_1y_2 - 3x_2y_1 + x_1y_3 + x_3y_1$ alors la forme quadratique associée est $q(\mathbf{x}) = s(\mathbf{x}, \mathbf{x}) = 2x_1^2 - 6x_1x_2 + 2x_1x_3$.
 - ▶ Si $s(\mathbf{x}, \mathbf{y}) = 2x_1y_1 + x_1 - x_3y_2$ alors $s(\mathbf{x}, \mathbf{x}) = 2x_1^2 + x_1 - x_2x_3$ n'est pas une forme quadratique (s n'étant pas une forme bilinéaire).

Forme polaire d'une forme quadratique

- Il existe en fait une correspondance bijective entre l'ensemble des formes bilinéaires s sur $\mathbb{E} \times \mathbb{E}$ et les formes quadratiques sur \mathbb{E} .

Propriété. (Forme polaire de q)

Soit $q : \mathbb{E} \rightarrow \mathbb{K}$ une forme quadratique sur \mathbb{E} . Il existe alors une et une seule forme bilinéaire symétrique $s : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{K}$ telle que $\forall \mathbf{x} \in \mathbb{E}, s(\mathbf{x}, \mathbf{x}) = q(\mathbf{x})$. s est donnée par :

$$s(\mathbf{x}, \mathbf{y}) = \frac{1}{2} (q(\mathbf{x} + \mathbf{y}) - q(\mathbf{x}) - q(\mathbf{y}))$$

s est dite **forme polaire** de q .

Détermination des images par calcul matriciel

- Comme pour les formes bilinéaires, nous pouvons représenter les formes quadratiques par des matrices (cf slide 156).
- Et donc calculer les images $q(\mathbf{x})$ par calcul matriciel.

Propriété.

Soient \mathbb{E} un ev sur \mathbb{K} de base $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. Pour toute forme quadratique $q : \mathbb{E} \rightarrow \mathbb{K}$ représentée par la matrice $M_b(q)_{\mathbf{e}_i} = \mathbf{Q}$ et pour tout vecteur $\mathbf{x} \in \mathbb{E}$, on a :

$$\begin{aligned} q(\mathbf{x}) &= M(\mathbf{x})_{\mathbf{e}_i}^\top M_b(q)_{\mathbf{e}_i} M(\mathbf{x})_{\mathbf{e}_i} \\ &= \mathbf{x}^\top \mathbf{Q} \mathbf{x} \end{aligned}$$

- Pour simplifier la lecture on notera la matrice de la forme quadratique q associée à s par \mathbf{Q} et non \mathbf{S} .
- On remarquera que \mathbf{Q} est forcément **symétrique** ($\mathbf{Q} = \mathbf{Q}^\top$).

Signe d'une forme quadratique

Définition.

Soit $q : \mathbb{E} \rightarrow \mathbb{K}$ une forme quadratique sur \mathbb{E} et soit \mathbf{Q} sa représentation matricielle. On dit que q est :

- *définie* si : $q(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
- *définie positive (dp)* si elle est définie et si :

$$\forall \mathbf{x} \in \mathbb{E}, \mathbf{x} \neq \mathbf{0}, \mathbf{x}^\top \mathbf{Q} \mathbf{x} > 0$$
- *définie négative (dn)* si elle est définie et si :

$$\forall \mathbf{x} \in \mathbb{E}, \mathbf{x} \neq \mathbf{0}, \mathbf{x}^\top \mathbf{Q} \mathbf{x} < 0$$
- *semi-définie positive (sdp)* si :

$$\forall \mathbf{x} \in \mathbb{E}, \mathbf{x} \neq \mathbf{0}, \mathbf{x}^\top \mathbf{Q} \mathbf{x} \geq 0$$
- *semi-définie négative (sdn)* si :

$$\forall \mathbf{x} \in \mathbb{E}, \mathbf{x} \neq \mathbf{0}, \mathbf{x}^\top \mathbf{Q} \mathbf{x} \leq 0$$
- *indéfinie* si elle est ni sdp ni sdn.

Valeurs propres et signe d'une forme quadratique

Théorème.

Les valeurs propres d'une matrice carrée symétrique réelle sont réelles.

Théorème.

Soit $q : \mathbb{E} \rightarrow \mathbb{K}$ une forme quadratique sur \mathbb{E} et \mathbf{Q} sa matrice carré symétrique associée.

- q est dp ssi ses valeurs propres sont strictement positives.
- q est dn ssi ses valeurs propres sont strictement négatives.
- q est sdp ssi ses valeurs propres sont positives ou nulles.
- q est sdn ssi ses valeurs propres sont négatives ou nulles.
- q est indéfinie ssi ses valeurs propres sont de signes distincts.

Comment déterminer le signe d'une forme quadratique

- Il existe plusieurs méthodes pour étudier le signe d'une forme quadratique :
 - ▶ Le critère de Sylvester reposant sur le calcul de déterminants de sous-matrices de \mathbf{Q} .
 - ▶ La réduction en carrés de Gauss qui est un algorithme utilisant la représentation algébrique.
 - ▶ La détermination du spectre de \mathbf{Q} où le signe des valeurs propres permet de conclure sur le signe de la forme quadratique.
- Dans la suite par abus de langage, nous noterons une forme quadratique soit par q soit par sa matrice représentation \mathbf{Q} . Les notations $\mathbf{Q} > 0$, $\mathbf{Q} < 0$, $\mathbf{Q} \geq 0$ et $\mathbf{Q} \leq 0$ désigneront respectivement une forme quadratique dp, dn, sdp, sdn.

Rappel du Sommaire

2 Algèbre Linéaire

- Espaces vectoriels
- La méthode du pivot de Gauss
- Applications linéaires et matrices
- Changement de base
- Déterminants
- Réduction des endomorphismes
- Formes bilinéaires et formes quadratiques
- Espace euclidien
 - Produit scalaire dans un espace vectoriel. Espaces euclidiens
 - Base orthonormée
 - Norme d'un vecteur. Angle non orienté (cas général)
 - Sous-espaces orthogonaux et matrices de projection

Introduction

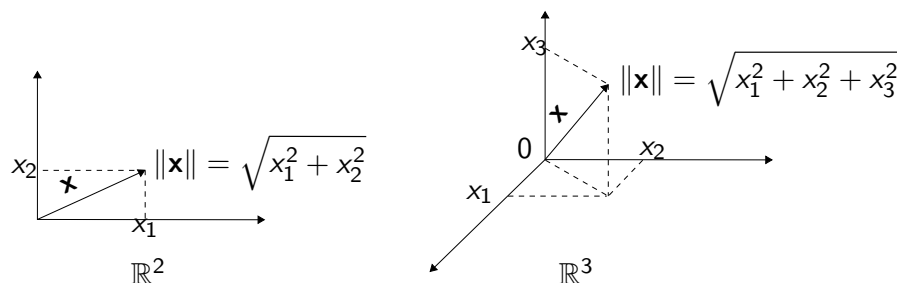
- Les espaces vectoriels constituent le cadre général de l'algèbre linéaire : il s'agit de la structure minimale permettant de traiter les problèmes linéaires.
 - On peut ajouter des concepts et structures supplémentaires aux espaces vectoriels permettant ainsi d'enrichir encore plus ces objets mathématiques.
- ⇒ Objectif : Ajouts des concepts de **métriques** aux ev permettant de définir les notions de normes, angles et distances ... La notion de **produit scalaire** est fondamentale dans cette perspective.
- ⇒ Nous voyons également les opérateurs de **projection orthogonale** qui sont utilisées implicitement dans de nombreuses méthodes statistiques.

Norme dans \mathbb{R}^2 et \mathbb{R}^3

- On appelle **norme** (associée au produit scalaire $\langle \cdot, \cdot \rangle$), du vecteur \mathbf{x} , notée $\|\mathbf{x}\|$, le scalaire :

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

- Cette notion correspond à celle déduite du théorème de Pythagore.



Produit scalaire canonique dans \mathbb{R}^2 ou \mathbb{R}^3

- Soient dans \mathbb{R}^3 , $\mathbf{x} = (x_1, x_2, x_3)$ et $\mathbf{y} = (y_1, y_2, y_3)$. On appelle **produit scalaire canonique** de \mathbf{x} par \mathbf{y} , noté $\langle \mathbf{x}, \mathbf{y} \rangle$ le scalaire défini par :

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + x_2 y_2 + x_3 y_3$$

- Propriétés du produit scalaire canonique dans \mathbb{R}^3 (ou \mathbb{R}^2) :

- Il est **bilinéaire**, $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3, \forall \lambda \in \mathbb{R}$:

- ★ $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
- ★ $\langle \lambda \mathbf{x}, \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{z} \rangle$
- ★ $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$
- ★ $\langle \mathbf{x}, \lambda \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{z} \rangle$

- Il est **symétrique**, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^3 : \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$

- Il est **défini positif**, $\forall \mathbf{x} \in \mathbb{R}^3$:

- ★ $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$
- ★ $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$

- $\langle \cdot, \cdot \rangle$ est une forme bilinéaire symétrique définie positive.

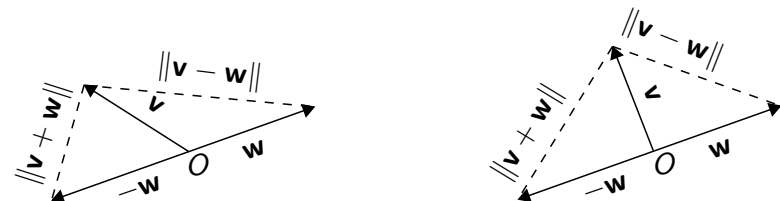
Orthogonalité dans \mathbb{R}^2 et \mathbb{R}^3

- On dit que deux vecteurs \mathbf{v} et \mathbf{w} sont **orthogonaux** (pour le produit scalaire $\langle \cdot, \cdot \rangle$), si :

$$\langle \mathbf{v}, \mathbf{w} \rangle = 0$$

- On peut exprimer le fait que ces vecteurs soient orthogonaux par la condition suivante :

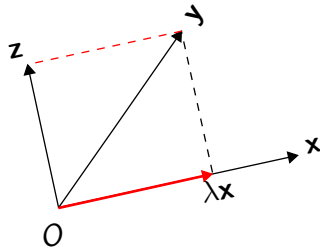
$$\|\mathbf{v} + \mathbf{w}\| = \|\mathbf{v} - \mathbf{w}\|$$



Angle non orienté dans \mathbb{R}^2 et \mathbb{R}^3

- Soient \mathbf{x} et \mathbf{y} deux vecteurs non nuls. Le coefficient de la **projection orthogonale (par rapport au produit scalaire $\langle \cdot, \cdot \rangle$)** de \mathbf{y} sur \mathbf{x} , noté $Proj_{\mathbf{x}}(\mathbf{y})$, est le scalaire λ tel que le vecteur $\mathbf{z} = \mathbf{y} - \lambda\mathbf{x}$ soit orthogonal à \mathbf{x} :

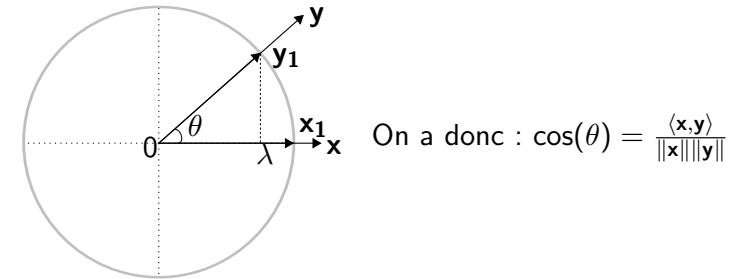
$$\langle \mathbf{y} - \lambda\mathbf{x}, \mathbf{x} \rangle = 0 \Rightarrow \lambda = Proj_{\mathbf{x}}(\mathbf{y}) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|^2}$$



Angle non orienté dans \mathbb{R}^2 et \mathbb{R}^3 (suite)

- Si \mathbf{x} et \mathbf{y} sont deux vecteurs non nuls, l'**angle non orienté** entre ces vecteurs est l'angle compris entre 0 et π que forment les vecteurs de longueur 1 : $\mathbf{x}_1 = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ et $\mathbf{y}_1 = \frac{\mathbf{y}}{\|\mathbf{y}\|}$.
- En trigonométrie, le **cosinus** de θ est défini par : $\cos(\theta) = \lambda$ où λ est le coefficient de la projection orthogonale de \mathbf{y}_1 sur \mathbf{x}_1 :

$$\lambda = \frac{\langle \mathbf{x}_1, \mathbf{y}_1 \rangle}{\|\mathbf{x}_1\|^2} = \langle \mathbf{x}_1, \mathbf{y}_1 \rangle = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$$



Espaces euclidiens

- Les définitions précédentes dans $\mathbb{R}^2, \mathbb{R}^3$ se généralisent à tout ev \mathbb{E} .
- D'une part nous pouvons définir des produits scalaires dans tout e.v. \mathbb{E} de dimension quelconque.
- D'autre part, il existe une infinité de produits scalaires dans \mathbb{E} puisqu'il existe une infinité de façon de définir des applications bilinéaires symétriques et définie positive.

Définition.

Soit \mathbb{E} un espace vectoriel réel, on appelle **produit scalaire** sur \mathbb{E} , une application $\langle \cdot, \cdot \rangle : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}$ qui est :

- Bilinéaire.
- Symétrique.
- Définie positive.

Un espace vectoriel réel de dimension finie muni d'un produit scalaire est dit **espace euclidien**.

Représentation matricielle du produit scalaire

Propriété.

Soit s un produit scalaire sur \mathbb{E} , $\{\mathbf{e}_i\}$ une base, et $\mathbf{S} = M_b(s)_{\mathbf{e}_i}$, $\mathbf{x} = M(\mathbf{x})_{\mathbf{e}_i}$, $\mathbf{y} = M(\mathbf{y})_{\mathbf{e}_i}$ où $\mathbf{x}, \mathbf{y} \in \mathbb{E}$. On a :

$$s(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{S} \mathbf{y}$$

On note également $s(\mathbf{x}, \mathbf{y})$ par $\langle \mathbf{x}, \mathbf{y} \rangle_s$.

- Exemple : $\mathbf{S} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ et $\mathbf{y} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$, on a :

$$s(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = 6$$

Représentation matricielle du produit scalaire (suite)

- Le produit scalaire canonique est associée à la matrice identité \mathbf{I} :

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{I}} = \mathbf{x}^{\top} \mathbf{I} \mathbf{y} = \mathbf{x}^{\top} \mathbf{y}$$

- Un produit scalaire s est tel que sa matrice $\mathbf{S} = M_b(s)_{\mathbf{e}_i}$ vérifie :

- ▶ $\mathbf{S} = \mathbf{S}^{\top}$
- ▶ $\forall \mathbf{x} \in \mathbb{E} : \mathbf{x}^{\top} \mathbf{S} \mathbf{x} \geq 0$
- ▶ $\mathbf{x}^{\top} \mathbf{S} \mathbf{x} = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$

Propriété.

La matrice associée à un produit scalaire est telle que ses valeurs propres sont réelles et strictement positives.

Autre exemple (suite)

- Il faut vérifier que $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j=1}^n a_{ij} b_{ij}$ est :

- ▶ Bilinéaire :

- ★ $\langle \mathbf{A} + \mathbf{B}, \mathbf{C} \rangle = \sum_{i,j=1}^n (a_{ij} + b_{ij}) c_{ij} = \sum_{i,j=1}^n a_{ij} c_{ij} + \sum_{i,j=1}^n b_{ij} c_{ij} = \langle \mathbf{A}, \mathbf{C} \rangle + \langle \mathbf{B}, \mathbf{C} \rangle$
- ★ $\langle \lambda \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j=1}^n (\lambda a_{ij}) b_{ij} = \lambda \sum_{i,j=1}^n a_{ij} b_{ij} = \lambda \langle \mathbf{A}, \mathbf{B} \rangle$
- ★ $\langle \mathbf{A}, \mathbf{B} + \mathbf{C} \rangle = \langle \mathbf{A}, \mathbf{B} \rangle + \langle \mathbf{A}, \mathbf{C} \rangle$
- ★ $\langle \mathbf{A}, \lambda \mathbf{B} \rangle = \lambda \langle \mathbf{A}, \mathbf{B} \rangle$

- ▶ Symétrique : $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j=1}^n a_{ij} b_{ij} = \sum_{i,j=1}^n b_{ij} a_{ij} = \langle \mathbf{B}, \mathbf{A} \rangle$

- ▶ Définie positive :

- ★ $\langle \mathbf{A}, \mathbf{A} \rangle = \sum_{i,j=1}^n a_{ij}^2 \geq 0$
- ★ $\langle \mathbf{A}, \mathbf{A} \rangle = 0 \Leftrightarrow \sum_{i,j=1}^n a_{ij}^2 = 0 \Leftrightarrow \forall i, j = 1, \dots, n : a_{ij} = 0 \Leftrightarrow \mathbf{A} = \mathbf{0}$

Autre exemple

- Soit $\mathbb{E} = \mathcal{M}_2(\mathbb{R})$ l'ev des matrices carrées d'ordre 2. On définit le produit scalaire de Froebenius $\langle \mathbf{A}, \mathbf{B} \rangle_F = \text{tr}(\mathbf{A}^{\top} \mathbf{B})$. Plus précisément si $\mathbf{A} = (a_{ij})$ et $\mathbf{B} = (b_{ij})$ on a :

$$\langle \mathbf{A}, \mathbf{B} \rangle_F = \text{tr}(\mathbf{A}^{\top} \mathbf{B}) = \sum_{i,j=1}^n a_{ij} b_{ij}$$

- Vérifions que l'opérateur ainsi défini est un produit scalaire.

Action du changement de base sur la représentation matricielle d'un produit scalaire

- Soient $\{\mathbf{e}_i\}$ et $\{\mathbf{e}'_i\}$ deux bases de \mathbb{E} , $\mathbf{P} = \mathbf{P}_{\mathbf{e}_i \rightarrow \mathbf{e}'_i}$ la matrice de passage et $\mathbf{x}, \mathbf{y} \in \mathbb{E}$ et s un produit scalaire.
- Si $\mathbf{S} = M_b(s)_{\mathbf{e}_i}$, $\mathbf{S}' = M_b(s)_{\mathbf{e}'_i}$, $\mathbf{x} = M(\mathbf{x})_{\mathbf{e}_i}$, $\mathbf{x}' = M(\mathbf{x})_{\mathbf{e}'_i}$, $\mathbf{y} = M(\mathbf{y})_{\mathbf{e}_i}$, $\mathbf{y}' = M(\mathbf{y})_{\mathbf{e}'_i}$ alors on a :
- $\mathbf{x}' = \mathbf{P}^{-1} \mathbf{x}$, $\mathbf{x} = \mathbf{P} \mathbf{x}'$, $\mathbf{y}' = \mathbf{P}^{-1} \mathbf{y}$ et $\mathbf{y} = \mathbf{P} \mathbf{y}'$.
- Par conséquent on a :

$$s(\mathbf{x}, \mathbf{y}) = \mathbf{x}^{\top} \mathbf{S} \mathbf{y} = (\mathbf{P} \mathbf{x}')^{\top} \mathbf{S} (\mathbf{P} \mathbf{y}') = (\mathbf{x}')^{\top} (\mathbf{P}^{\top} \mathbf{S} \mathbf{P}) \mathbf{y}'$$

- Par ailleurs on a par définition dans la base $\{\mathbf{e}'_i\}$:

$$s(\mathbf{x}, \mathbf{y}) = (\mathbf{x}')^{\top} \mathbf{S}' \mathbf{y}'$$

- On en conclut que :

$$\mathbf{S}' = \mathbf{P}^{\top} \mathbf{S} \mathbf{P}$$

Exemple

- Écrivez le produit scalaire canonique de \mathbb{R}^2 dans la base $\{\mathbf{e}'_1, \mathbf{e}'_2\}$ où :

$$\mathbf{e}'_1 = 2\mathbf{e}_1 + 3\mathbf{e}_2; \quad \mathbf{e}'_2 = -8\mathbf{e}_1 + 5\mathbf{e}_2$$

- On a $\mathbf{S} = \mathbf{I}$ et $\mathbf{P} = \begin{pmatrix} 2 & -8 \\ 3 & 5 \end{pmatrix}$.
- On a vu que $\mathbf{S}' = \mathbf{P}^\top \mathbf{S} \mathbf{P}$ et donc :

$$\mathbf{S}' = \mathbf{P}^\top \mathbf{I} \mathbf{P} = \mathbf{P}^\top \mathbf{P} = \begin{pmatrix} 2 & 3 \\ -8 & 5 \end{pmatrix} \begin{pmatrix} 2 & -8 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 13 & -1 \\ -1 & 89 \end{pmatrix}$$

- En d'autres termes on a $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{e}'_i} = 13x'_1y'_1 + 89x'_2y'_2 - x'_1y'_2 - x'_2y'_1$.

Remarque sur les bases orthonormées

- Si $\{\epsilon_i\}$ est une base orthogonale alors les vecteurs $\frac{\epsilon_i}{\|\epsilon_i\|}$ forment une base orthonormée.
- Toute famille de vecteurs non nuls $\{\epsilon_1, \dots, \epsilon_n\}$ deux à deux orthogonaux est libre.
- Soit $\{\mathbf{e}_i\}$ une base orthonormée de \mathbb{E} et $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$, $\mathbf{y} = \sum_{i=1}^n y_i \mathbf{e}_i$ deux vecteurs de \mathbb{E} . On a :

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle &= \left\langle \sum_{i=1}^n x_i \mathbf{e}_i, \sum_{j=1}^n y_j \mathbf{e}_j \right\rangle = \sum_{i,j=1}^n x_i y_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \sum_{i,j=1}^n x_i y_j \delta_{ij} \\ &= x_1 y_1 + x_2 y_2 + \dots + x_n y_n \end{aligned}$$

Ainsi, **dans une base orthonormée** le produit scalaire prend la forme du **produit scalaire canonique**.

Base orthonormée

- Rappels de définitions :
 - Norme du vecteur \mathbf{x} associé au produit scalaire $\langle \cdot, \cdot \rangle$: $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$.
 - \mathbf{x} et \mathbf{y} sont orthogonaux par rapport au produit scalaire $\langle \cdot, \cdot \rangle$ si : $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ (ce qu'on notera également par $\mathbf{x} \perp \mathbf{y}$)

Définition.

Soit \mathbb{E} un espace euclidien. Une base $\{\mathbf{e}_i\}$ est dite **orthogonale** si, $\forall i, j$ tels que $i \neq j$:

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$$

Elle est dite **orthonormée** si de plus chaque vecteur est de norme 1 :

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$$

où δ_{ij} est le symbole de Kronecker défini par : $\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$

Théorème fondamental des espaces euclidiens

Théorème.

Dans un espace euclidien il existe toujours des bases orthonormées.

Corollaire.

Soit \mathbb{E} un espace euclidien. Le choix d'une base orthonormée $\mathbb{B} = \{\mathbf{e}_i\}$ dans un espace euclidien, permet d'identifier \mathbb{E} à \mathbb{R}^n muni du produit scalaire canonique par l'**isomorphisme** :

$$\phi_{\mathbb{B}} = \begin{matrix} \mathbb{E} & \rightarrow & \mathbb{R}^n \\ \mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i & \rightarrow & (x_1, \dots, x_n) \end{matrix}$$

- Remarque : le procédé d'orthonormalisation de Gram-Schmidt permet de transformer une base quelconque en une base orthonormée.

Diagonalisation de matrices réelles symétriques

Théorème.

Toute matrice réelle symétrique \mathbf{A} est diagonalisable et les espaces propres sont deux à deux orthogonaux (pour le produit scalaire canonique de \mathbb{R}^n).

- Exemple :

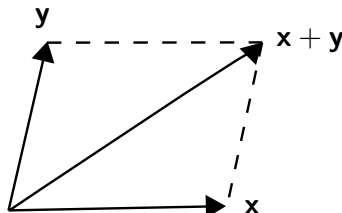
$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- On a 3 valeurs propres : $Sp_{\mathbb{K}}(\mathbf{A}) = \{0, 1, 2\}$.
- On a 3 vecteurs propres : $\mathbf{v}_1 = (1, -1, 0)$, $\mathbf{v}_2 = (0, 0, 1)$, $\mathbf{v}_3 = (1, 1, 0)$.
- Les 3 vecteurs propres sont mutuellement orthogonaux.
- Les 3 vecteurs propres normalisés sont : $\mathbf{e}_1 = \frac{1}{\sqrt{2}}(1, -1, 0)$, $\mathbf{e}_2 = \mathbf{v}_2 = (0, 0, 1)$, $\mathbf{e}_3 = \frac{1}{\sqrt{2}}(1, 1, 0)$.

Inégalité triangulaire

- Inégalité triangulaire :

$$n(\mathbf{x} + \mathbf{y}) \leq n(\mathbf{x}) + n(\mathbf{y})$$



- L'inégalité triangulaire exprime le fait que dans un triangle la longueur d'un côté est inférieure ou égale à la somme des longueurs des deux autres côtés.

Norme

Définition.

Une norme est une application $n : \mathbb{E} \rightarrow \mathbb{R}^+$ qui vérifie les propriétés suivantes, $\forall \mathbf{x} \in \mathbb{E}; \forall \lambda \in \mathbb{R}$:

- $n(\lambda \mathbf{x}) = |\lambda| n(\mathbf{x})$
- $n(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
- Inégalité triangulaire : $n(\mathbf{x} + \mathbf{y}) \leq n(\mathbf{x}) + n(\mathbf{y})$
L'égalité a lieu s.s.i. il existe $\lambda \geq 0$ tel que $\mathbf{y} = \lambda \mathbf{x}$.

- Soit \mathbb{E} un ev réel muni d'un produit scalaire $\langle \cdot, \cdot \rangle$. L'application $\|\cdot\| : \mathbb{E} \rightarrow \mathbb{R}^+$ définie $\forall \mathbf{x} \in \mathbb{E}$ par :

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

est la **norme dite euclidienne** (càd associée à $\langle \cdot, \cdot \rangle$) de \mathbb{E} .

- En revanche, toute norme définie sur \mathbb{E} n'est pas forcément euclidienne (exemple : la norme ℓ_1).

Inégalités de Cauchy-Schwarz

Propriété. (Inégalité de Cauchy-Schwarz)

Pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{E}$ un espace euclidien, on a :

$$\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2$$

ou encore :

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$$

Angle non orienté

- Soient \mathbf{x}, \mathbf{y} deux vecteurs non nuls. D'après l'inégalité de Cauchy-Schwarz on a la propriété suivante :

$$\frac{|\langle \mathbf{x}, \mathbf{y} \rangle|}{\|\mathbf{x}\| \|\mathbf{y}\|} \leq 1 \Leftrightarrow -1 \leq \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \leq 1$$

- Il existe un et un seul $\theta \in [0, \pi]$ tel que :

$$\cos(\theta) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

θ est dit **angle** (non orienté) entre les vecteurs \mathbf{x} et \mathbf{y} .

- Notons enfin la relation qui exprime le **produit scalaire en fonction de la norme** :

$$\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2} (\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2)$$

Sous-espaces orthogonaux

Définition.

Soient \mathbb{E} un espace euclidien, $\mathbb{F} \subseteq \mathbb{E}$ un sous-espace quelconque. On définit \mathbb{F}^\perp , le sous-espace orthogonal de \mathbb{F} , de la façon suivante :

$$\mathbb{F}^\perp = \{\mathbf{x} \in \mathbb{E} \mid \langle \mathbf{x}, \mathbf{u} \rangle = 0, \forall \mathbf{u} \in \mathbb{F}\}$$

- En raison de la symétrie, on a également $\mathbb{F}^\perp = \{\mathbf{x} \in \mathbb{E} \mid \langle \mathbf{u}, \mathbf{x} \rangle = 0, \forall \mathbf{u} \in \mathbb{F}\}.$
- \mathbb{F}^\perp est un sev de \mathbb{E} même si \mathbb{F} ne l'est pas initialement.
- $\{\mathbf{0}\}^\perp = \mathbb{E}.$
- $\mathbb{E}^\perp = \{\mathbf{0}\}.$

Distance

Définition.

Une distance est une application $d : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}^+$ qui vérifie les propriétés suivantes, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{E}; \forall \lambda \in \mathbb{R}$:

- Symétrie : $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x}).$
- $d(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}.$
- Inégalité triangulaire : $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}).$

- Soit \mathbb{E} un ev réel muni d'un produit scalaire $\langle \cdot, \cdot \rangle$. L'application $\|\cdot\| : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}^+$ définie $\forall \mathbf{x}, \mathbf{y} \in \mathbb{E}$ par :

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle}$$

est la **distance dite euclidienne de \mathbb{E}** et vérifie donc les propriétés énoncées précédemment.

Sous-espaces orthogonaux (suite)

Propriété.

Pour tout sev \mathbb{F} d'un espace euclidien \mathbb{E} , on a :

- $\dim(\mathbb{E}) = \dim(\mathbb{F}) + \dim(\mathbb{F}^\perp).$
- $\mathbb{E} = \mathbb{F} \oplus \mathbb{F}^\perp.$
- $(\mathbb{F}^\perp)^\perp = \mathbb{F}.$

Projecteur

- Nous avons déjà présenté précédemment des applications linéaires qui sont des projecteurs.
- La particularité de ces applications linéaires est la suivante :

Définition.

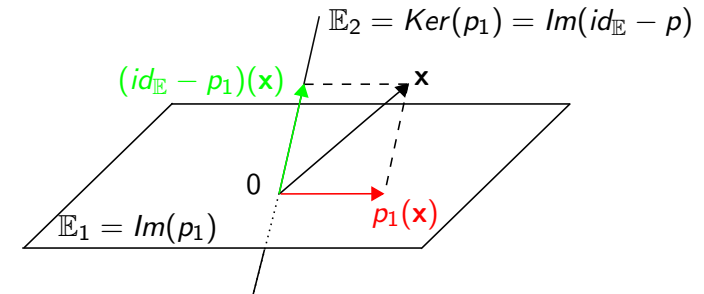
Soit \mathbb{E} un espace vectoriel. On appelle **projecteur** un endomorphisme p de \mathbb{E} tel que $p^2 = p \circ p = p$.

Propriété.

On a pour un projecteur p les propriétés suivantes :

- $\mathbb{E} = \text{Ker}(p) \oplus \text{Im}(p)$.
- $\text{Ker}(p) = \text{Im}(id_{\mathbb{E}} - p)$ où $id_{\mathbb{E}}$ est l'application identité.

Exemple de projecteur



- p_1 est le projecteur sur \mathbb{E}_1 parallèlement à \mathbb{E}_2 .
- $p_1^2(x) = p_1(p_1(x)) = p_1(x)$.
- $\text{Im}(p_1) = \mathbb{E}_1$ et $\text{Ker}(p_1) = \mathbb{E}_2$.
- $\text{Im}(id_{\mathbb{E}} - p_1) = \mathbb{E}_2$.
- $\mathbb{E} = \mathbb{E}_1 \oplus \mathbb{E}_2$.

Projecteur orthogonale

Définition.

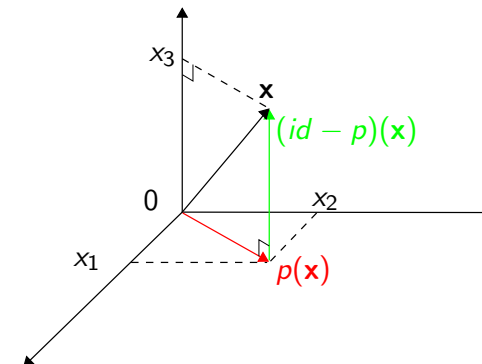
Soit \mathbb{E} un espace euclidien et \mathbb{F} un sous-espace de \mathbb{E} . On appelle **projecteur orthogonal** sur \mathbb{F} , le projecteur sur \mathbb{F} parallèlement à \mathbb{F}^\perp .

Propriété.

Nous avons les propriétés équivalentes suivantes :

- p est un projecteur orthogonal.
- $\text{Im}(id_{\mathbb{E}} - p) = (\text{Im}(p))^\perp$.
- $\forall \mathbf{x}, \mathbf{y} : \langle p(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, p(\mathbf{y}) \rangle$ (on dit que p est **auto-adjoint**).
- Autres propriétés si p est un projecteur orthogonal sur \mathbb{F} :
 - $\text{Im}(p) = \mathbb{F}$ et $\text{Ker}(p) = \mathbb{F}^\perp$.
 - $\mathbf{P} = M(p)_{\mathbf{e}_i}$ est tel que $\boxed{\mathbf{P}^2 = \mathbf{P}}$ (on dit que \mathbf{P} est **idempotent**).
 - $id - p$ est (également) un projecteur orthogonal sur $(\text{Im}(p))^\perp$ et donc $\boxed{(\mathbf{I} - \mathbf{P})^2 = \mathbf{I} - \mathbf{P}}$.

Exemple de projecteur orthogonale



- $\mathbb{E} = \mathbb{R}^3$ et p est le projecteur orthogonal sur $\mathbb{F} = \text{Vec}\{\mathbf{e}_1, \mathbf{e}_2\}$ (plan vectoriel engendré par \mathbf{e}_1 et \mathbf{e}_2).
- $\mathbb{F}^\perp = \text{Vec}\{\mathbf{e}_3\}$ (droite vectorielle engendrée par \mathbf{e}_3).
- $\text{Im}(id_{\mathbb{E}} - p) = \text{Vec}\{\mathbf{e}_3\}$.

Projection orthogonale et moindres carrés

Théorème.

Soit \mathbb{E} un ev muni d'un produit scalaire et soit \mathbb{F} un sev de \mathbb{E} . Soit $\mathbf{u} \in \mathbb{E}$ alors :

$$\forall \mathbf{v} \in \mathbb{F} : \|\mathbf{u} - \text{Proj}_{\mathbb{F}}(\mathbf{u})\| \leq \|\mathbf{u} - \mathbf{v}\|$$

- On dit que $\text{Proj}_{\mathbb{F}}(\mathbf{u}) \in \mathbb{F}$ est la meilleur approximation de \mathbf{u} par des vecteurs de \mathbb{F} .

Démonstration.

On a $\mathbf{u} - \mathbf{v} = \mathbf{u} - \text{Proj}_{\mathbb{F}}(\mathbf{u}) + \text{Proj}_{\mathbb{F}}(\mathbf{u}) - \mathbf{v}$. Or $\text{Proj}_{\mathbb{F}}(\mathbf{u}) - \mathbf{v} \in \mathbb{F}$ et $\mathbf{u} - \text{Proj}_{\mathbb{F}}(\mathbf{u}) \in \mathbb{F}^{\perp}$. En raison de l'orthogonalité et du théorème de Pythagore, on en conclut que

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u} - \text{Proj}_{\mathbb{F}}(\mathbf{u})\|^2 + \|\text{Proj}_{\mathbb{F}}(\mathbf{u}) - \mathbf{v}\|^2. \text{ Comme } \|\text{Proj}_{\mathbb{F}}(\mathbf{u}) - \mathbf{v}\|^2 \geq 0 \text{ on a finalement } \|\mathbf{u} - \mathbf{v}\|^2 \geq \|\mathbf{u} - \text{Proj}_{\mathbb{F}}(\mathbf{u})\|^2 \text{ et donc } \|\mathbf{u} - \mathbf{v}\| \geq \|\mathbf{u} - \text{Proj}_{\mathbb{F}}(\mathbf{u})\|. \quad \square$$

Projection orthogonale et moindres carrés (suite)

- Reprenons un système de p équations linéaires à n inconnues (x_1, \dots, x_n) s'écrivant $\mathbf{Ax} = \mathbf{b}$ où \mathbf{A} est de taille $(p \times n)$ et \mathbf{b} est de taille $(p \times 1)$.
- Supposons que ce système n'a pas de solution. Cela veut dire que \mathbf{b} ne peut s'écrire comme une combinaison linéaire des colonnes de \mathbf{A} ou encore que $\mathbf{b} \notin \text{Vec}\{\mathbf{a}^1, \dots, \mathbf{a}^n\}$ où \mathbf{a}^i est la i -ème colonne de \mathbf{A} .
- Dans ce cas, on peut chercher un vecteur de $\mathbb{F} = \text{Vec}\{\mathbf{a}^1, \dots, \mathbf{a}^n\}$ qui puisse approximer au mieux \mathbf{b} .
- Ce problème peut donc être formaliser de la façon suivante :

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|^2$$

- La solution \mathbf{x} est appelée solution des **Moindres Carrés Ordinaires** (MCO).

Projection orthogonale et moindres carrés (suite)

- Cette appellation est due au fait qu'on essaie d'approximer \mathbf{b} par un vecteur engendré par les colonnes de \mathbf{A} et que nous cherchons pour cela à faire le minimum d'erreur :

$$\|\mathbf{Ax} - \mathbf{b}\|^2 = \|\epsilon\|^2$$

où $\epsilon \in \mathbb{R}^p$ de terme général $\epsilon_i = \sum_{j=1}^n a_{ij}x_j - b_i$.

- On cherche donc à minimiser la somme des erreurs aux carrés $(\epsilon_i)^2$.
- Or nous avons vu que la meilleure approximation de \mathbf{b} en fonction de vecteurs de \mathbb{F} est donnée par la projection orthogonale $\text{Proj}_{\mathbb{F}}(\mathbf{b})$.
- On propose donc de résoudre le système :

$$\mathbf{Ax} = \text{Proj}_{\mathbb{F}}(\mathbf{b})$$

Projection orthogonale et moindres carrés (suite)

- Pour résoudre $\mathbf{Ax} = \text{Proj}_{\mathbb{F}}(\mathbf{b})$ on peut raisonner géométriquement comme suit.
- Si \mathbf{x} est tel que $\mathbf{Ax} = \text{Proj}_{\mathbb{F}}(\mathbf{b})$, alors le vecteur $\mathbf{b} - \mathbf{Ax} = \mathbf{b} - \text{Proj}_{\mathbb{F}}(\mathbf{b})$ est un vecteur appartenant à \mathbb{F}^{\perp} .
- \mathbb{F} est l'espace engendré par les colonnes de \mathbf{A} et tout vecteur de \mathbb{F} est orthogonal à $\mathbf{b} - \mathbf{Ax}$. On a en particulier :

$$\mathbf{A}^{\top}(\mathbf{b} - \mathbf{Ax}) = 0$$

- Cette équation conduit au **système normal** suivant (associé au système $\mathbf{Ax} = \mathbf{b}$) :

$$\mathbf{A}^{\top} \mathbf{b} = \mathbf{A}^{\top} \mathbf{Ax}$$

- Les solutions du système normal sont les solutions au sens des MCO (donc approximative) du système $\mathbf{Ax} = \mathbf{b}$.

Projection orthogonale et moindres carrés (suite)

Théorème.

Pour tout système d'équations linéaires $\mathbf{Ax} = \mathbf{b}$ le système normal associé $\mathbf{A}^\top \mathbf{Ax} = \mathbf{A}^\top \mathbf{b}$ a au moins une solution. Toutes les solutions du système normal sont des solutions au sens des MCO de $\mathbf{Ax} = \mathbf{b}$. De plus, si \mathbb{F} est l'espace engendré par les colonnes de \mathbf{A} et si \mathbf{x} est une solution des MCO de $\mathbf{Ax} = \mathbf{b}$ alors : $\text{Proj}_{\mathbb{F}}(\mathbf{b}) = \mathbf{Ax}$.

Théorème.

Soit \mathbf{A} une matrice dont les vecteurs colonnes sont linéairement indépendants. Alors la matrice $\mathbf{A}^\top \mathbf{A}$ est inversible.

Théorème.

Soit \mathbf{A} une matrice dont les vecteurs colonnes sont linéairement indépendants. Alors tout système linéaire $\mathbf{Ax} = \mathbf{b}$ a une solution unique au sens des MCO qui est donnée par $\mathbf{x} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$.

Projection orthogonale et moindres carrés (suite)

- Du système normal nous avons d'une part : $\mathbf{x} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$.
- D'autre part, nous avons par définition : $\mathbf{Ax} = \text{Proj}_{\mathbb{F}}(\mathbf{b})$.
- $\text{Proj}_{\mathbb{F}}$ est un endomorphisme. Notons par $\mathbf{P}_{\mathbb{F}}$ sa matrice associée.
- Nous avons alors matriciellement :

$$\mathbf{Ax} = \mathbf{P}_{\mathbb{F}} \mathbf{b}$$

- En remplaçant \mathbf{x} par sa solution des MCO, on obtient :

$$\mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} = \mathbf{P}_{\mathbb{F}} \mathbf{b}$$

- Enfin par identification, nous avons la matrice de la projection orthogonale sur \mathbb{F} (le sev engendré par les colonnes de \mathbf{A}) qui s'écrit :

$$\mathbf{P}_{\mathbb{F}} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$$

- Cette matrice de projection intervient dans de nombreux contextes notamment en régression linéaire et en analyse de données.

Applications en statistique multidimensionnelle

- Les MCO sont des techniques courantes en statistique.
- La régression linéaire multiple :
 - ▶ Modèle qui fait l'hypothèse qu'une variable à expliquer Y peut être approximée par une combinaison linéaire de p variables explicatives :

$$Y = \sum_{k=1}^p a_k X^k + \epsilon$$

où ϵ est l'erreur entre la vraie valeur Y et celle prédite par le modèle linéaire $\sum_{k=1}^p a_k X^k$.

- ▶ Remarque : Y est appelée variable dépendante et les $\{X^k\}_{k=1}^p$ variables indépendantes.
- ▶ Pour inférer les valeurs des coefficients on minimise la somme des carrés des erreurs.
- ▶ Soit $\mathbf{X} = (\mathbf{x}^1 \dots \mathbf{x}^p)$ la matrice des variables explicatives et \mathbf{y} le vecteur de la variable et $\hat{\mathbf{a}}_{mco}$ le vecteur des coefficients estimé selon les MCO. On a donc :

$$\hat{\mathbf{a}}_{mco} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

Applications en statistique multidimensionnelle (suite)

- Méthodes de réduction de dimension :
 - ▶ Etant donné un nuage de points dans un espace euclidien de p dimensions, on cherche à déterminer un sous-espace euclidien de dimension plus petite que p de sorte à ce que le nuage de points projetés sur ce sous-espace soit le "moins déformé" possible.
 - ▶ Une façon de poser le problème est par MCO : on minimise les distances euclidiennes entre chaque point et son projeté.
 - ▶ La résolution du problème fait appel à une généralisation de la décomposition spectrale à des matrices rectangulaires.
 - ▶ C'est ce que nous voyons dans la suite du cours.

Rappel du Sommaire

- 1 Introduction
- 2 Algèbre Linéaire
- 3 Méthodes de réduction de dimension
- 4 Méthodes classiques de classification automatique (introduction)

Rappel du Sommaire

- 3 Méthodes de réduction de dimension
 - Analyse générale : la décomposition en valeurs singulières (SVD)
 - Analyse en composantes principales (ACP)
 - Analyse générale : extension à des métriques et poids quelconques
 - Analyse factorielle des correspondances (AFC)
 - Analyse (factorielle) des correspondances multiples (ACM)

Introduction

- Nous abordons maintenant les applications en statistiques exploratoires multidimensionnelles :
 - ▶ les méthodes de réduction de dimension,
 - ▶ les méthodes de classification automatique.
- Les fondements mathématiques de ces techniques reposent sur les concepts développés en espace euclidien que nous avons revus précédemment.
- La suite du cours reste formalisée mais avec une perspective plus ancrée en statistiques appliquées.
- Nous voyons notamment comment des concepts en analyse de données comme information, approximation... se traduisent en termes métriques dans le cadre des espaces euclidiens.

Introduction

- Les techniques de réduction de dimension permettent de représenter de façon synthétique et graphique l'information contenue dans une table de données composées de nombreuses variables ($p > 2$).
- Il s'agit de méthodes statistiques descriptives mais qui sont puissantes puisqu'elles permettent de découvrir des connaissances².
- Les outils algébriques fondamentaux sont :
 - ▶ distances entre points, projections orthogonales et métriques ;
 - ▶ décomposition en valeurs propres de matrices carrées.
- Il existe en fait une dualité (càd des relations fortes) entre l'analyse du nuage des individus et celle du nuage des variables.
- Sous l'angle algébrique, cette dualité est fondée sur la décomposition en valeurs singulières d'une matrice rectangulaire (SVD).
- Nous faisons le lien entre les principes forts en réduction de dimension et la SVD dans ce qui suit.

2. Ces approches sont très employées en fouille de données (data mining)/découverte de connaissances (knowledge discovery).

Le point de départ en ADD : une table de données (rappel)

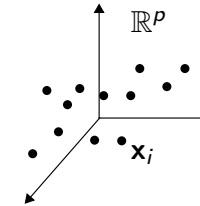
- Notons \mathbf{X} une matrice de données de taille $(n \times p)$:
$$\mathbf{X} = \begin{matrix} & \mathbf{x}^1 & \dots & \mathbf{x}^k & \dots & \mathbf{x}^p \\ \mathbf{x}_1 & x_{11} & & x_{1k} & & x_{1p} \\ \vdots & & & \vdots & & \\ \mathbf{x}_i & x_{i1} & \dots & x_{ik} & \dots & x_{ip} \\ \vdots & & & \vdots & & \\ \mathbf{x}_n & x_{n1} & & x_{nk} & & x_{np} \end{matrix}$$
- x_{ik} est le terme général de \mathbf{X} et $\forall i = 1, \dots, n; \forall k = 1, \dots, p : x_{ik} \in \mathbb{R}$.
- Chaque ligne i de \mathbf{X} représente un vecteur \mathbf{x}_i de taille $(p \times 1)$.
- Chaque colonne k de \mathbf{X} représente un vecteur \mathbf{x}^k de taille $(n \times 1)$.

$$\forall i : \mathbf{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix} \quad \text{et} \quad \forall k : \mathbf{x}^k = \begin{pmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{pmatrix}$$

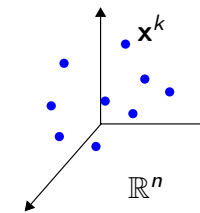
- On notera également un vecteur comme ceci : $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$.

Une table de données mais deux nuages de points (rappel)

- L'ensemble des vecteurs $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ peut être représenté dans \mathbb{R}^p .
On notera ce nuage $\mathbb{N}\mathbb{O}$.



- L'ensemble des vecteurs $\{\mathbf{x}^1, \dots, \mathbf{x}^p\}$ peut être représenté dans \mathbb{R}^n .
On notera ce nuage $\mathbb{N}\mathbb{A}$.



Objectifs des méthodes de ce cours (rappel)

- Pour fixer les idées supposons que les lignes représentent des individus et les colonnes des variables de différentes natures (par exemple sexe, âge, nombre d'enfants, salaire...)
- Pour appréhender l'information contenue dans la table de données, on s'intéresse aux nuages de points formés par les individus d'une part et par les variables d'autre part.
- Le point de vue géométrique de ces données permettrait d'apporter des éléments de réponses aux questions suivantes :
 - Quels sont les individus qui sont proches et qui présentent donc des similitudes ?
 - Quelles sont les variables corrélées entre elles ?
 - Peut-on représenter visuellement ces informations dans un espace dont les axes porteraient une sémantique particulière ?
 - Peut-on dégager des groupes homogènes d'individus et expliquer ces groupes par un comportement particulier de plusieurs variables ?

Objectifs des méthodes de ce cours (suite)

- Toutefois, il n'est pas possible d'appréhender ces données dans des espaces de dimension supérieure à 3 puisqu'on ne sait pas représenter visuellement ce type d'espace.
- On cherche donc à représenter ces données dans des espaces réduits de dimension 2 ou 3. Mais passer d'un espace de dimension p (ou n) à un espace de dimension 2 implique une perte d'information. On cherche donc des espaces réduits qui conservent au mieux l'information initiale.
- Les concepts de données, espace, projection et information sont abordés du point de vue géométrique. Pour cela on a recours **revu** des outils mathématiques issus principalement de l'AL.
- On verra également par la suite quelques éléments en optimisation pour une meilleure compréhension des méthodes en ADD.

Variance ou inertie d'un nuage de points

- La quantité d'information contenue dans un nuage de points \mathbb{NO} constitué de n vecteurs est mesurée par la **variance** ou l'**inertie totale**. Elle sera notée $int(\mathbb{NO})$ et définie par :

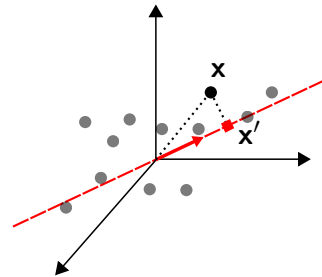
$$int(\mathbb{NO}) = \frac{1}{2n^2} \sum_{\mathbf{x} \in \mathbb{NO}} \sum_{\mathbf{y} \in \mathbb{NO}} d^2(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n d^2(\mathbf{x}_i, \bar{\mathbf{x}})$$

où $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ est le vecteur moyen appelé **barycentre** de \mathbb{NO} .

- De façon brève : il s'agit de la distance quadratique moyenne entre chaque point du nuage et le barycentre.
- Interprétation : plus les points sont dispersés autour du barycentre plus il y a de l'information. A contrario, un nuage de points concentré autour du barycentre est peu d'informatif.

Approximer \mathbb{NO} par un nuage dans un espace réduit (suite)

- Retour sur l'exemple :



- Le théorème de Pythagore implique la relation suivante :

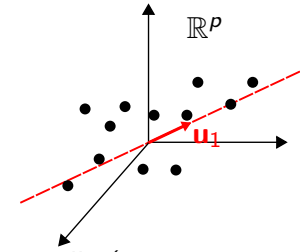
$$\|\mathbf{x} - \mathbf{x}'\|^2 = \|\mathbf{x}\|^2 - \|\mathbf{x}'\|^2$$

- Seuls les \mathbf{x}' dépendent de \mathbf{u}_1 (la variable inconnue) on a donc :

$$\min_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{\mathbf{x} \in \mathbb{NO}} \|\mathbf{x} - \mathbf{x}'\|^2 \Leftrightarrow \max_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{\mathbf{x} \in \mathbb{NO}} \|\mathbf{x}'\|^2$$

Approximer \mathbb{NO} par un nuage dans un espace réduit

- Exemple illustratif :



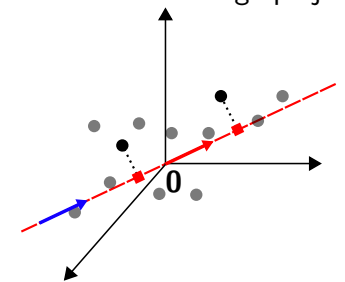
- On cherche dans \mathbb{R}^p une droite vectorielle (un sous-espace de dimension 1) portée par un vecteur \mathbf{u}_1 tel que lorsqu'on projette \mathbb{NO} sur cette droite, la variance est la plus grande possible.
- Pour maximiser la variance du nuage projeté on voit qu'il faut "déformer" au minimum le nuage de point.
- Soit \mathbf{x}' le vecteur projeté de $\mathbf{x} \in \mathbb{NO}$ sur la droite portée par \mathbf{u}_1 . On peut formuler le problème par moindres carrés ordinaires :

$$\min_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{\mathbf{x} \in \mathbb{NO}} \|\mathbf{x} - \mathbf{x}'\|^2$$

Approximer \mathbb{NO} par un nuage dans un espace réduit (suite)

- Les MCO permettent ici de maximiser la variance du nuage projeté.

- En effet, remarquons que dans un espace euclidien les distances entre points ne sont pas modifiées si l'on change l'origine du repère :



- On peut donc supposer que le repère est centré en le barycentre $\bar{\mathbf{x}}$. On a donc $\bar{\mathbf{x}} = \mathbf{0}$. On voit également que $\bar{\mathbf{x}}'$ la projection de $\bar{\mathbf{x}}$ sur \mathbf{u}_1 est en $\mathbf{0}$. Notons $int_{\mathbf{u}_1}(\mathbb{NO})$ l'inertie du nuage projeté sur \mathbf{u}_1 , on a donc :

$$\max_{\mathbf{u}_1 \in \mathbb{R}^p} int_{\mathbf{u}_1}(\mathbb{NO}) \Leftrightarrow \max_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{\mathbf{x}' \in \mathbb{NO}} \|\mathbf{x}' - \bar{\mathbf{x}}'\|^2 \Leftrightarrow \max_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{\mathbf{x}' \in \mathbb{NO}} \|\mathbf{x}'\|^2$$

Approximer \mathbb{NO} par un nuage dans un espace réduit (suite)

- Par ailleurs, \mathbf{x}' le vecteur \mathbf{x} projeté sur \mathbf{u}_1 , est défini comme suit :

$$\mathbf{x}' = \frac{\langle \mathbf{x}, \mathbf{u}_1 \rangle}{\|\mathbf{u}_1\|} \mathbf{u}_1$$

- On remarquera que si on prend \mathbf{u}_1 normée, càd $\|\mathbf{u}_1\| = 1$, on a $\mathbf{x}' = \langle \mathbf{x}, \mathbf{u}_1 \rangle \mathbf{u}_1$ et alors :

$$\begin{aligned} \|\mathbf{x}'\|^2 &= \langle \mathbf{x}', \mathbf{x}' \rangle \\ &= \langle \langle \mathbf{x}, \mathbf{u}_1 \rangle \mathbf{u}_1, \langle \mathbf{x}, \mathbf{u}_1 \rangle \mathbf{u}_1 \rangle \\ &= \langle \mathbf{x}, \mathbf{u}_1 \rangle^2 \langle \mathbf{u}_1, \mathbf{u}_1 \rangle \text{ (car linéarité de } \langle \cdot, \cdot \rangle \text{)} \\ &= \langle \mathbf{x}, \mathbf{u}_1 \rangle^2 \text{ (car } \langle \mathbf{u}_1, \mathbf{u}_1 \rangle = 1 \text{)} \\ &= (\mathbf{x}^\top \mathbf{u}_1)^2 \\ &= (\mathbf{x}^\top \mathbf{u}_1)^\top (\mathbf{x}^\top \mathbf{u}_1) \text{ (car scalaires)} \\ &= \mathbf{u}_1^\top \mathbf{x} \mathbf{x}^\top \mathbf{u}_1 \end{aligned}$$

Approximer \mathbb{NO} par un nuage dans un espace réduit (suite)

- On remarquera que $\mathbf{X}^\top \mathbf{X}$ est également une matrice de produits scalaires. En effet, $\mathbf{X} = (\mathbf{x}^1 \dots \mathbf{x}^p)$ et donc :

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} (\mathbf{x}^1)^\top \\ \vdots \\ (\mathbf{x}^p)^\top \end{pmatrix} (\mathbf{x}^1 \dots \mathbf{x}^p) = \underbrace{((\mathbf{x}^k)^\top \mathbf{x}^l)}_{\langle \mathbf{x}^k, \mathbf{x}^l \rangle}_{k,l=1,\dots,p}$$

Définition. (Matrice de Gram)

Soit $\mathbf{x}_1, \dots, \mathbf{x}_n$ une famille de vecteurs d'un espace euclidien \mathbb{E} . On appelle matrice de Gram de $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, notée \mathbf{G} , la matrice carrée d'ordre n , de terme général $\mathbf{G}_{ij} = \langle \mathbf{x}_i, \mathbf{x}_j \rangle, \forall i, j = 1, \dots, n$.

Propriété. (Matrice de Gram)

Toute matrice de Gram est symétrique et semi-définie positive.

Approximer \mathbb{NO} par un nuage dans un espace réduit (suite)

- Prenons tous les vecteurs de $\mathbb{NO} = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n\}$. Les projetés sur la droite engendrée par \mathbf{u}_1 sont notés $\{\mathbf{x}'_1, \dots, \mathbf{x}'_i, \dots, \mathbf{x}'_n\}$. On a :

$$\sum_{i=1}^n \|\mathbf{x}'_i\|^2 = \sum_{i=1}^n \mathbf{u}_1^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{u}_1 = \mathbf{u}_1^\top \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top \right) \mathbf{u}_1 = \mathbf{u}_1^\top (\mathbf{X}^\top \mathbf{X}) \mathbf{u}_1$$

- En effet $\mathbf{X} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_n^\top)$ et donc :

$$\mathbf{X}^\top \mathbf{X} = (\mathbf{x}_1 \dots \mathbf{x}_n) \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top$$

- On obtient le problème d'optimisation contraint suivant :

$$\max_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{i=1}^n \|\mathbf{x}'_i\|^2 \Leftrightarrow \max_{\mathbf{u}_1 \in \mathbb{R}^p : \|\mathbf{u}_1\|^2=1} \mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X} \mathbf{u}_1$$

Approximer \mathbb{NO} par un nuage dans un espace réduit (suite)

- Posons $\mathbf{Q} = \mathbf{X}^\top \mathbf{X}$, le problème d'optimisation contraint précédent, $\max_{\mathbf{u}_1 \in \mathbb{R}^p} \mathbf{u}_1^\top \mathbf{Q} \mathbf{u}_1$ sous la contrainte (s.l.c.) $\|\mathbf{u}_1\|^2 = 1$, consiste à trouver le vecteur maximiseur de la forme quadratique \mathbf{Q} s.l.c. que la solution soit normée.
- En optimisation contrainte, pour déterminer une telle solution \mathbf{u}_1^* on définit le **lagrangien** (ou fonction de Lagrange) noté $lag(\mathbf{u}_1, \lambda)$:

$$lag(\mathbf{u}_1, \lambda) = \mathbf{u}_1^\top \mathbf{Q} \mathbf{u}_1 - \lambda(\mathbf{u}_1^\top \mathbf{u}_1 - 1)$$

- On cherche ensuite les solutions des conditions nécessaires du premier ordre de lag (point où le gradient de lag s'annule) :

$$\nabla lag(\mathbf{u}_1, \lambda) = \begin{pmatrix} \frac{\partial lag}{\partial \mathbf{u}_1} \\ \frac{\partial lag}{\partial \lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix} \text{ où } \frac{\partial lag}{\partial \mathbf{u}_1} = \begin{pmatrix} \frac{\partial lag}{\partial u_{1,1}} \\ \vdots \\ \frac{\partial lag}{\partial u_{1,p}} \end{pmatrix}$$

Approximer \mathbb{NO} par un nuage dans un espace réduit (suite)

- Le gradient par rapport à \mathbf{u}_1 donne la condition suivante :

$$\begin{aligned}\frac{\partial \text{lag}}{\partial \mathbf{u}_1}(\mathbf{u}_1, \lambda) = \mathbf{0} &\Leftrightarrow 2\mathbf{Q}\mathbf{u}_1 - 2\lambda\mathbf{u}_1 = \mathbf{0} \\ &\Leftrightarrow \mathbf{Q}\mathbf{u}_1 = \lambda\mathbf{u}_1 \\ &\Leftrightarrow \mathbf{X}^\top \mathbf{X}\mathbf{u}_1 = \lambda\mathbf{u}_1\end{aligned}$$

- ▷ La solution du problème est un **vecteur propre** de $\mathbf{X}^\top \mathbf{X}$!

- Lequel ? A partir des conditions précédentes on voit que :

$$\begin{aligned}\mathbf{X}^\top \mathbf{X}\mathbf{u}_1 = \lambda\mathbf{u}_1 &\Rightarrow \mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1 = \lambda\mathbf{u}_1^\top \mathbf{u}_1 \\ &\Rightarrow \mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}_1 = \lambda \text{ (car le vecteur est choisi normé)}\end{aligned}$$

- ▷ La valeur de l'inertie du nuage projeté est égale à la valeur propre λ !
Donc \mathbf{u}_1^* est le vect. propre normé de $\mathbf{X}^\top \mathbf{X}$ associé à la plus grande valeur propre.

Approximer \mathbb{NO} par un nuage dans un espace réduit (suite)

- ▷ Ce résultat pour $m = 1, 2$ se généralise jusqu'à s et pour déterminer un sous-espace de dimension $s < p$ qui maximise l'inertie du nuage projeté on prendra comme base les s vecteurs propres normés de $\mathbf{X}^\top \mathbf{X}$ associés aux s plus grande valeurs propres : $\{\mathbf{u}_1, \dots, \mathbf{u}_s\}$.
- ▷ Les s premières valeurs propres $\{\lambda_1, \dots, \lambda_s\}$ sont resp. les mesures de l'inertie du nuage projeté sur chacun des axes $\{\mathbf{u}_1, \dots, \mathbf{u}_s\}$.
- ▷ Les coordonnées de \mathbf{x}_i sur la droite vectorielle \mathbf{u}_m est donnée par :

$$\langle \mathbf{x}_i, \mathbf{u}_m \rangle = \mathbf{x}_i^\top \mathbf{u}_m$$

- ▷ Plus généralement la représentation de tous les vecteurs dans l'espace réduit de dimension s est donnée par :

$$\mathbf{F} = \mathbf{X}\mathbf{U}$$

$$\text{où } \mathbf{U} = (\mathbf{u}_1 \quad \dots \quad \mathbf{u}_s).$$

Approximer \mathbb{NO} par un nuage dans un espace réduit (suite)

- On a trouvé un sous-espace de dimension 1 qui permet de représenter le nuage \mathbb{NO} en cherchant à conserver au mieux l'information.
- Comment trouver un sous-espace de dimension 2 ?
- On construit incrémentalement : on cherche un deuxième vecteur $\mathbf{u}_2 \in \mathbb{R}^p$ tel que $\|\mathbf{u}_2\| = 1$ et $\mathbf{u}_2 \perp \mathbf{u}_1$ et qui permet de conserver au mieux l'information (càd l'inertie).
- Comme précédemment, on montre que \mathbf{u}_2^* est le vecteur propre normé de $\mathbf{X}^\top \mathbf{X}$ associé à la **seconde**³ plus grande valeur propre.

3. On a vu slide 185 que les espaces propres d'une matrice carrée symétrique réelle sont orthogonaux entre eux.

Approximer \mathbb{NA} par un nuage dans un espace réduit

- Rappel : à la table \mathbf{X} on lui associe deux nuages de points \mathbb{NO} et $\mathbb{NA} = \{\mathbf{x}^1, \dots, \mathbf{x}^p\}$ dont les vecteurs sont représentés dans \mathbb{R}^n .
 - Les développements précédents permettent de déterminer de façon similaire un sous-espace de dimension réduite ($s < n$) pour représenter \mathbb{NA} tout en déformant au minimum ce dernier (au sens de l'inertie).
- ▷ On montre que la base de sous-espace est donné par $\mathbf{v}^1, \dots, \mathbf{v}^s$ les s vecteurs propres normés de $\mathbf{X}\mathbf{X}^\top$ associés aux plus grandes valeurs propres.

Dualité entre l'analyse de \mathbb{NO} et celle de \mathbb{NA}

- L'analyse de \mathbb{NO} revient à diagonaliser $\mathbf{X}^\top \mathbf{X}$ d'ordre p tandis que l'analyse de \mathbb{NA} revient à diagonaliser $\mathbf{X}\mathbf{X}^\top$ d'ordre n . La symétrie de ces deux analyses fait que nous avons des propriétés particulières.
- Les valeurs propres non nulles de $\mathbf{X}^\top \mathbf{X}$ et $\mathbf{X}\mathbf{X}^\top$ sont les mêmes !
- Les bases des sous-espaces réduits sont liées entre elles par les formules de dualité. Notons λ_m la m -ème plus grande valeur propre de $\mathbf{X}\mathbf{X}^\top$ et \mathbf{v}^m le vecteur propre associé, on a :

$$\mathbf{X}\mathbf{X}^\top \mathbf{v}^m = \lambda_m \mathbf{v}^m \Rightarrow \mathbf{X}^\top \mathbf{X} \mathbf{X}^\top \mathbf{v}^m = \lambda_m \mathbf{X}^\top \mathbf{v}^m$$

- ▷ $\mathbf{X}^\top \mathbf{v}^m$ et \mathbf{u}_m , le vecteur propre normé de $\mathbf{X}^\top \mathbf{X}$ associé à λ_m , sont proportionnels. Pour satisfaire aux contraintes de normes unitaires on voit que (**relations de dualité**) :

$$\mathbf{u}_m = \frac{\mathbf{X}^\top \mathbf{v}^m}{\|\mathbf{X}^\top \mathbf{v}^m\|} = \frac{1}{\sqrt{\lambda_m}} \mathbf{X}^\top \mathbf{v}^m \quad \text{et} \quad \mathbf{v}^m = \frac{\mathbf{X} \mathbf{u}_m}{\|\mathbf{X} \mathbf{u}_m\|} = \frac{1}{\sqrt{\lambda_m}} \mathbf{X} \mathbf{u}_m$$

Factorisation de matrices : la décomposition en valeurs singulières (suite)

- Les valeurs $\{\sigma_1, \dots, \sigma_r\}$ sont appelées **valeurs singulières** de \mathbf{A} .
- Les colonnes de \mathbf{U} sont appelés **vecteurs singuliers à gauche** de \mathbf{A} et les colonnes de \mathbf{V} sont appelés **vecteurs singuliers à droite** de \mathbf{A} .
- ▷ Les colonnes de \mathbf{U} et les colonnes de \mathbf{V} forment respectivement une base orthonormée de \mathbb{R}^n et \mathbb{R}^p ce qui implique que :

$$\mathbf{U}^\top \mathbf{U} = \mathbf{I}_n \quad \text{et} \quad \mathbf{V}^\top \mathbf{V} = \mathbf{I}_p$$

- ▷ ou encore (définition équivalente de matrices orthogonales) :

$$\mathbf{U}^\top = \mathbf{U}^{-1} \quad \text{et} \quad \mathbf{V}^\top = \mathbf{V}^{-1}$$

Factorisation de matrices : la SVD

Théorème. (Décomposition en valeurs singulières (SVD))

Soit $\mathbf{A} \in \mathcal{M}_{n,p}(\mathbb{R})$ une matrice de réels de taille $(n \times p)$ et de rang^a $\text{rg}(\mathbf{A}) = r \leq \min(n, p)$. Alors \mathbf{A} peut être factorisée de la façon suivante :

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$$

où $\mathbf{U} \in \mathcal{M}_n$ et $\mathbf{V} \in \mathcal{M}_p$ sont des matrices orthogonales^b, et $\mathbf{\Sigma} \in \mathcal{M}_{n,p}$ est une matrice remplie de 0 sauf sur sa diagonale principale où, $\forall i = 1, \dots, r$, on a :

$$\Sigma_{ii} = \sigma_i$$

Les $\{\sigma_i\}_{i=1, \dots, r}$ sont uniques et on supposera que $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$ (tri dans l'ordre décroissant).

a. cf slide 87

b. $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_n$ et $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_p$

Déterminations des éléments singuliers

- Remarquons que si $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$ alors $\mathbf{A}^\top = \mathbf{V} \mathbf{\Sigma} \mathbf{U}^\top$ et on infère :

$$\begin{aligned} \mathbf{A}^\top \mathbf{A} &= \mathbf{A}^\top \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top \\ &= \mathbf{V} \mathbf{\Sigma} \mathbf{\Sigma} \mathbf{V}^\top \\ &= \mathbf{V} \mathbf{\Sigma}^2 \mathbf{V}^\top \end{aligned}$$

- ▷ Les valeurs singulières de \mathbf{A} et les vecteurs singuliers à droite (càd les colonnes de \mathbf{V}) sont resp. les racines carrées des valeurs propres et les vecteurs propres de $\mathbf{A}^\top \mathbf{A}$.
- ▷ Similairement, on montre que les valeurs singulières de \mathbf{A} et les vecteurs singuliers à gauche (càd les colonnes de \mathbf{U}) sont resp. les racines carrées des valeurs propres et les vecteurs propres de $\mathbf{A} \mathbf{A}^\top$.
- Comme nous le savons déjà, les valeurs propres de $\mathbf{A} \mathbf{A}^\top$ et $\mathbf{A}^\top \mathbf{A}$ sont les mêmes mais, leurs vecteurs propres respectifs sont distincts en général.

Liens entre vecteurs singuliers à gauche et à droite

- Les développements précédents, indiquent des relations fortes entre la décomposition spectrale de $\mathbf{A}^\top \mathbf{A}$ et celle de $\mathbf{A} \mathbf{A}^\top$.
- Les vect. propres de $\mathbf{A}^\top \mathbf{A}$ sont \mathbf{V} , les vect. singuliers à droite de \mathbf{A} .
- Les vect. propres de $\mathbf{A} \mathbf{A}^\top$ sont \mathbf{U} , les vect. singuliers à gauche de \mathbf{A} .
- Nous avons la **relation de dualité** suivante, $\forall i = 1, \dots, r$:

$$(\mathbf{A} \mathbf{A}^\top \mathbf{u}_i = \sigma_i \mathbf{u}_i) \wedge (\mathbf{A}^\top \mathbf{A} \mathbf{v}^i = \sigma_i \mathbf{v}^i) \Rightarrow \left(\frac{\mathbf{A}^\top \mathbf{u}_i}{\sqrt{\sigma_i}} = \mathbf{v}^i \right) \wedge \left(\frac{\mathbf{A} \mathbf{v}^i}{\sqrt{\sigma_i}} = \mathbf{u}_i \right)$$

- En effet, nous avons la même démonstration que précédemment :

$$\begin{aligned} \mathbf{A} \mathbf{A}^\top \mathbf{u}_i &= \sigma_i \mathbf{u}_i \Rightarrow \mathbf{A}^\top \mathbf{A} \mathbf{A}^\top \mathbf{u}_i = \sigma_i \mathbf{A}^\top \mathbf{u}_i \\ &\Rightarrow \mathbf{A}^\top \mathbf{u}_i \text{ est vecteur propre de } \mathbf{A}^\top \mathbf{A} \text{ associé à } \sigma_i \end{aligned}$$

- Ensuite, \mathbf{v}^i doit être normé (par définition). On a $\|\mathbf{A}^\top \mathbf{u}_i\|^2 = \langle \mathbf{A}^\top \mathbf{u}_i, \mathbf{A}^\top \mathbf{u}_i \rangle = \mathbf{u}_i^\top \mathbf{A} \mathbf{A}^\top \mathbf{u}_i = \sigma_i$ d'où $\frac{1}{\sqrt{\sigma_i}} \mathbf{A}^\top \mathbf{u}_i = \mathbf{v}^i$.

Théorème d'Eckart-Young

- On souhaite approximer au sens de $\|\cdot\|_F$, une matrice \mathbf{A} par une matrice $\hat{\mathbf{A}}$, s.l.c. que $\hat{\mathbf{A}}$ soit de rang plus petit qu'un entier donné s . Formellement on a :

$$\min_{\hat{\mathbf{A}} \in \mathcal{M}_{n,p} : \text{rg}(\hat{\mathbf{A}}) \leq s} \|\mathbf{A} - \hat{\mathbf{A}}\|_F$$

Théorème. (Théorème d'Eckart-Young)

La solution du problème précédent est donnée par la SVD de \mathbf{A} .

- Plus précisément définissons $\hat{\mathbf{A}}$ comme suit :

$$\hat{\mathbf{A}} = \sum_{m=1}^s \sigma_m \mathbf{u}_m (\mathbf{v}^m)^\top$$

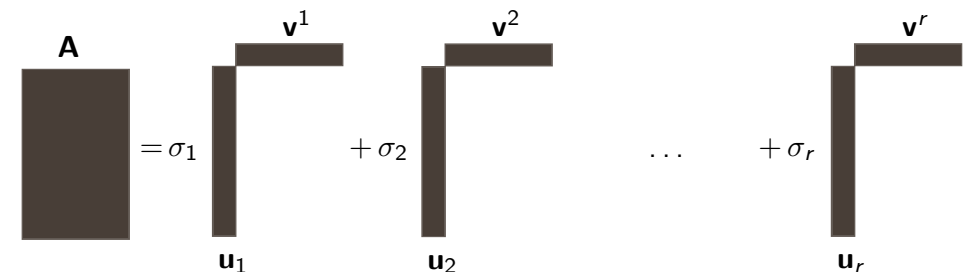
- ▷ $\hat{\mathbf{A}}$ ainsi définie est la matrice de rang s qui soit la **meilleure approximation** de \mathbf{A} au sens de $\|\cdot\|_F$.

Approximation matricielle de faible rang

- Nous venons de faire le parallélisme entre la dualité en ADD et les résultats en AL concernant la SVD.
- Nous avons vu que les valeurs singulières de la matrice \mathbf{X} pouvait s'interpréter en terme de maximisation de la forme quadratique $\mathbf{X}^\top \mathbf{X}$ s.l.c. de norme unitaire (cf slide 218).
- Nous voyons un autre résultat important qui nous éclaire encore davantage sur le bien fondé des méthodes de réduction de dimension.
- On raisonne dans l'espace vectoriel $\mathcal{M}_{n,p}$ (çad l'ensemble des matrices de taille $(n \times p)$). On munit de cet espace le produit scalaire de Froebenius (cf slide 178) qui donne la distance euclidienne suivante :

$$\|\mathbf{A} - \mathbf{B}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^p (a_{ij} - b_{ij})^2}$$

Illustration



- Rappelons que :
 - $r = \text{rg}(\mathbf{A})$, on a donc r valeurs singulières non nulles.
 - $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$
 - les vecteurs singuliers $\mathbf{u}_1, \dots, \mathbf{u}_r$ et $\mathbf{v}^1, \dots, \mathbf{v}^r$ sont normés.
- $\sigma_1 \mathbf{u}_1 (\mathbf{v}^1)^\top$ = meilleure approximation de rang 1,
- $\sigma_1 \mathbf{u}_1 (\mathbf{v}^1)^\top + \sigma_2 \mathbf{u}_2 (\mathbf{v}^2)^\top$ = meilleure approximation de rang 2,
- ...
- ▷ $\sum_{m=1}^r \sigma_m \mathbf{u}_m (\mathbf{v}^m)^\top = \mathbf{A}$!

Exemple

- Prenons l'exemple suivant :

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \Rightarrow \mathbf{A}\mathbf{A}^\top = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

- $rg(\mathbf{A}) = r = 2$.
- $\sigma_1 = 2, \sigma_2 = \sqrt{2}, \sigma_3 = 0$.
- $\mathbf{u}_1 = (1, 1, 0)/\sqrt{2}, \mathbf{u}_2 = (0, 0, 1), \mathbf{u}_3 = (-1, 1, 0)/\sqrt{2}$.
- $\mathbf{v}^1 = (1, 1, 0, 0)/\sqrt{2}, \mathbf{v}^2 = (0, 0, 1, 1)/\sqrt{2}, \mathbf{v}^3 = (-1, 1, 0, 0)/\sqrt{2}$.
- Calculez la meilleure approximation de rang 1 :

$$2\mathbf{u}_1(\mathbf{v}^1)^\top = \frac{2}{\sqrt{2}\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} (1 \ 1 \ 0 \ 0) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Qualité de l'approximation (suite)

- ▷ Comme de plus $\mathbf{A} = \sum_{m=1}^r \sigma_m \mathbf{u}_m (\mathbf{v}^m)^\top$, on a donc :

$$\tau_s = \frac{\sum_{m=1}^s \sigma_m^2}{\sum_{m=1}^r \sigma_m^2}$$

- Autrement dit, la qualité de l'approximation de \mathbf{A} par $\hat{\mathbf{A}}$ peut être mesurée par les valeurs singulières de \mathbf{A} .
- ▷ Clairement $\tau_s \in [0, 1]$ et elle peut être interprétée tel un **pourcentage de reconstitution**.
- ▷ En ADD, σ_m^2 le carré de la m -ème plus grande valeur singulière de \mathbf{X} :
 - ▶ correspond à λ_m , la m -ème plus grande valeur propre de $\mathbf{X}^\top \mathbf{X}$ et $\mathbf{X}\mathbf{X}^\top$,
 - ▶ est l'**inertie des nuages projetés** sur les vect. propres de $\mathbf{X}^\top \mathbf{X}$ et $\mathbf{X}\mathbf{X}^\top$ associés à λ_i .
- Calculez τ_1 pour l'exemple précédent. On trouve $\tau_1 = 4/6 = 2/3$.

Qualité de l'approximation

- Supposons que $\hat{\mathbf{A}}$ soit une approximation de rang s de \mathbf{A} .
- On peut mesurer la **qualité de l'approximation** par la quantité :

$$\tau_s = \frac{\|\hat{\mathbf{A}}\|_F^2}{\|\mathbf{A}\|_F^2}$$

- En utilisant la SVD on a $\hat{\mathbf{A}} = \sum_{m=1}^s \sigma_m \mathbf{u}_m (\mathbf{v}^m)^\top$ et il vient :

$$\begin{aligned} \|\hat{\mathbf{A}}\|_F^2 &= \sum_{i=1}^n \sum_{j=1}^p \hat{a}_{ij}^2 = \sum_{i=1}^n \sum_{j=1}^p \sum_{m=1}^s \sigma_i^2 u_{m,i}^2 (v_j^m)^2 \\ &= \sum_{m=1}^s \sigma_m^2 \left(\sum_{i=1}^n u_{m,i}^2 \right) \left(\sum_{j=1}^p (v_j^m)^2 \right) \\ &= \sum_{m=1}^s \sigma_m^2 \end{aligned}$$

Rappel du Sommaire

3 Méthodes de réduction de dimension

- Analyse générale : la décomposition en valeurs singulières (SVD)
- Analyse en composantes principales (ACP)
- Analyse générale : extension à des métriques et poids quelconques
- Analyse factorielle des correspondances (AFC)
- Analyse (factorielle) des correspondances multiples (ACM)

Données traitées par l'ACP et notations

- L'ACP traite des tables de données dont les **variables sont quantitatives continues**.
- On suppose n individus $\{\mathbf{t}_1, \dots, \mathbf{t}_i, \dots, \mathbf{t}_n\}$ où $\mathbf{t}_i \in \mathbb{R}^p$.
- On suppose p variables $\{\mathbf{t}^1, \dots, \mathbf{t}^k, \dots, \mathbf{t}^p\}$ où $\mathbf{t}^k \in \mathbb{R}^n$.
- Ces données se retrouvent dans la matrice \mathbf{T} de taille $(n \times p)$:

$$\mathbf{T} = \begin{matrix} & \mathbf{t}^1 & \dots & \mathbf{t}^k & \dots & \mathbf{t}^p \\ \begin{matrix} \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_i \\ \vdots \\ \mathbf{t}_n \end{matrix} & \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ \dots & \dots & t_{ik} & \dots & \dots \\ & & & & \\ & & & & \end{pmatrix} \end{matrix}$$

où $\forall i = 1, \dots, n; \forall k = 1, \dots, p : t_{ik} \in \mathbb{R}$.

Données traitées par l'ACP et notations

- Chaque ligne de \mathbf{T} est un ensemble de p nombres réels. C'est ce qu'on appellera un vecteur individu :

$$\mathbf{t}_i = (t_{i1}, \dots, t_{ik}, \dots, t_{ip})$$

- Chaque colonne de \mathbf{T} est un ensemble de n nombres réels. C'est ce qu'on appellera un vecteur variable :

$$\mathbf{t}^k = \begin{pmatrix} t_{1k} \\ \vdots \\ t_{ik} \\ \vdots \\ t_{nk} \end{pmatrix}$$

- Dans la suite, par abus de langage, on parlera du vecteur \mathbf{t}_i et de l'individu i de façon équivalente. Il en va de même pour le vecteur \mathbf{t}^k et la variable k .

Proximité et nuage des individus

- L'ensemble des individus (appelés également objets) sera noté \mathbb{O} .
- Pour mesurer la proximité entre deux individus i et j on utilise (le carré de) la **distance euclidienne** notée $d^2(\mathbf{t}_i, \mathbf{t}_j)$:

$$d^2(\mathbf{t}_i, \mathbf{t}_j) = \sum_{k=1}^p (t_{ik} - t_{jk})^2$$

- Les individus \mathbb{O} sont représentés dans l'espace \mathbb{R}^p dont les dimensions sont les variables $(\mathbf{t}^1, \dots, \mathbf{t}^p)$.
- L'ensemble de ces n points forme le **nuage des individus** noté \mathbb{NO} .
- On s'intéresse dans cet espace aux **relations de "proximité" entre individus** (qui se ressemblent (càd proches) ? qui sont différents (càd éloignés) ?).

Proximité et nuage des variables

- L'ensemble des variables (appelés également attributs) sera noté \mathbb{A} .
- L'ensemble des p points $\mathbf{t}^1, \dots, \mathbf{t}^p$ forme le **nuage des variables** noté \mathbb{NA} appartenant à l'espace \mathbb{R}^n dont les dimensions sont cette fois-ci les individus $(\mathbf{t}_1, \dots, \mathbf{t}_n)$.
- Pour mesurer la **liaison** entre deux variables k et l on utilise le **coefficient de corrélation** noté $r(\mathbf{t}^k, \mathbf{t}^l)$.
- Etant donné une variable \mathbf{t}^k , on introduit au préalable,
 - ▶ sa **moyenne empirique** notée m_k (càd $\bar{\mathbf{t}}^k$) :

$$m_k = \frac{1}{n} \sum_{i=1}^n t_{ik}$$

- ▶ sa **variance empirique** notée s_k^2 :

$$s_k^2 = \frac{1}{n} \sum_{i=1}^n (t_{ik} - m_k)^2$$

Proximité et nuage des variables (suite)

- On **centre et on réduit** la variable \mathbf{t}^k puis on **divise par** \sqrt{n} et on obtient la variable \mathbf{x}^k dont le terme général est donné par :

$$x_{ik} = \frac{t_{ik} - m_k}{s_k \sqrt{n}}$$

- On s'intéresse aux **relations de "proximité" entre variables** et on utilise ici le **coefficient de corrélation** (empirique) donné par :

$$r(\mathbf{t}^k, \mathbf{t}^l) = \sum_{i=1}^n \underbrace{\left(\frac{t_{ik} - m_k}{s_k \sqrt{n}} \right)}_{x_{ik}} \underbrace{\left(\frac{t_{il} - m_l}{s_l \sqrt{n}} \right)}_{x_{il}}$$

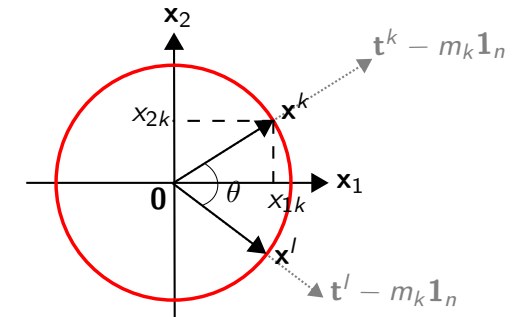
- Remarques sur le coefficient de corrélation :
 - $r(\mathbf{t}^k, \mathbf{t}^l)$ est le produit scalaire entre les vecteurs \mathbf{x}^k et \mathbf{x}^l :
 $r(\mathbf{t}^k, \mathbf{t}^l) = \langle \mathbf{x}^k, \mathbf{x}^l \rangle = \sum_{i=1}^n x_{ik} x_{il}$.
 - $\forall k : r(\mathbf{t}^k, \mathbf{t}^k) = \langle \mathbf{x}^k, \mathbf{x}^k \rangle = \|\mathbf{x}^k\|^2 = 1$, donc tous les vecteurs \mathbf{x}^k sont de norme unitaire et ils appartiennent tous à une hypersphère.

Objectifs

- ▶ Etant donné la matrice \mathbf{T} de dimension p (on suppose $p > 3$), on s'intéresse aux relations de proximité entre **individus d'une part** et entre **variables d'autre part**. En d'autres termes on cherche à définir une **typologie** des ces entités respectives.
- Cette mise en lumière est réalisée par le biais de **graphiques** représentant respectivement les nuages d'individus et de variables dans des **sous-espaces à 2 dimensions** et dans lesquels les relations de proximités initiales sont respectées au mieux.
- ▶ Raisonnons par exemple sur le nuage des individus. Le but est de représenter le nuage de points appartenant à un grand espace à p dimensions dans un espace plus petit, de sorte à ce que les individus proches dans l'espace initiale le reste dans l'espace réduit autant se faire que peut.
- ▶ Pour définir cet espace de faible dimension, on va déterminer des **nouvelles variables qui "synthétise" les variables initiales** et qui préservent au mieux la variabilité du nuage des individus.

Hypersphère et coefficient de corrélation

- Illustrations dans \mathbb{R}^2 du coefficient de corrélation entre \mathbf{t}^k et \mathbf{t}^l :



- Une fois les variables réduites et divisées par \sqrt{n} , on obtient les \mathbf{x}^k qui appartiennent au cercle de centre \mathbb{R}^n et de rayon 1.
- $r(\mathbf{t}^k, \mathbf{t}^l)$ est égale au **cosinus de θ , l'angle formé par $(\mathbf{x}^k, \mathbf{x}^l)$ (ou $(\mathbf{t}^k, \mathbf{t}^l)$)**. Il est aussi égale au produit scalaire entre ces vecteurs.
- Remarquons aussi que $r(\mathbf{t}^k, \mathbf{t}^l) = r(\mathbf{t}^k - m_k \mathbf{1}_n, \mathbf{t}^l - m_l \mathbf{1}_n) = r(\mathbf{x}^k, \mathbf{x}^l)$ où $\mathbf{1}_n$ est le vecteur rempli de 1 et de dimension n .

Idées principales de la méthode

- ▶ Ces nouvelles variables sont des vecteurs de \mathbb{R}^n appelées **composantes principales**. Une composante principale est en fait les coordonnées de la projection du nuage de points qui appartient à \mathbb{R}^p , sur un vecteur de ce même espace appelé **axe factoriel**.
- ▶ Une **composante principale est un vecteur de \mathbb{R}^n obtenu par une combinaison linéaire des variables initiales** \mathbf{A} . Les coefficients de ces combinaisons linéaires sont les coefficients de corrélation entre la composante principale et les différentes variables initiales.
- ▶ L'étude de ces coefficients permet d'associer à toute composante principale un groupe de variables initiales qui lui est associé.
- ▶ Préserver au mieux la variabilité du nuage des individus revient à déterminer dans \mathbb{R}^p un **axe factoriel** qui est tel que lorsqu'on projette les vecteurs individus sur celui-ci la variance reste forte.

Transformation des données et ACP normée

- Nous supposons désormais que la matrice des données est celle des variables centrées, réduites et divisée par \sqrt{n} (comme décrit précédemment).
- Nous noterons \mathbf{X} cette matrice que nous utiliserons donc désormais :

$$\mathbf{X} = \begin{matrix} & \mathbf{x}^1 & \dots & \mathbf{x}^k & \dots & \mathbf{x}^p \\ \begin{matrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_i \\ \vdots \\ \mathbf{x}_n \end{matrix} & \left(\begin{array}{cccccc} & & & & & \\ & & & & & \\ & & & & & \\ \dots & \dots & x_{ik} & \dots & \dots & \\ & & & & & \\ & & & & & \end{array} \right) \end{matrix}$$

- Autrement dit, dans l'approche classique de l'ACP normée, la première étape consiste à transformer \mathbf{T} en \mathbf{X} .

Variance ou inertie du nuage des individus

- La notion de variance de nuage de points est centrale puisque c'est la quantité que nous cherchons à préserver.
- Dans le repère initial des variables \mathbb{A} , le nuage \mathbb{NO} a une **inertie totale** définie par :

$$int_{\mathbb{A}}(\mathbb{NO}) = \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n d^2(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{n} \sum_{i=1}^n d^2(\mathbf{x}_i, \bar{\mathbf{x}})$$

- L'information contenue dans \mathbb{NO} est mesurée par l'inertie qui indique, en moyenne, de combien s'éloignent les points du barycentre.
- Si l'inertie est faible alors les points sont en moyenne très proche du barycentre ce qui est peu informatif. Si l'inertie est grande, c'est donc qu'il y a plein de disparités entre les individus. Le but alors est d'appréhender cette information de façon synthétique.

Transformation des données et ACP normée (suite)

- Dans la suite, les vecteurs représentant les individus seront donc notés \mathbf{x}_i et les vecteurs représentant les variables seront notés \mathbf{x}^k .
- Les propriétés de \mathbf{X} sont alors les suivantes, $\forall k = 1, \dots, p$:

$$\sum_{i=1}^n x_{ik} = 0 \quad \text{et} \quad \sum_{i=1}^n (x_{ik})^2 = 1$$

- Remarques :

- ▶ \mathbf{m} le barycentre de \mathbb{NO} calculé dans le repère initial devient l'origine du repère du nouvel espace affine. L'opération de centrage agit telle une **translation** de \mathbb{NO} de l'origine initiale au barycentre.
- ▶ Comme déjà fait remarquer, il est intéressant de noter que le centrage ne change pas les distances euclidiennes entre individus.
- ▶ Après réduction les variables appartiennent à une hypersphère.
- ▶ En pratique, la réduction permet aux variables de s'affranchir de leurs **unités de mesure** ce qui rend l'analyse plus robuste face aux biais associés aux différences d'échelles. On parle d'**ACP normée**.

Variance ou inertie du nuage des individus

- Du fait du centrage, le barycentre devient l'origine du nouveau repère on montre que dans ce repère affine :

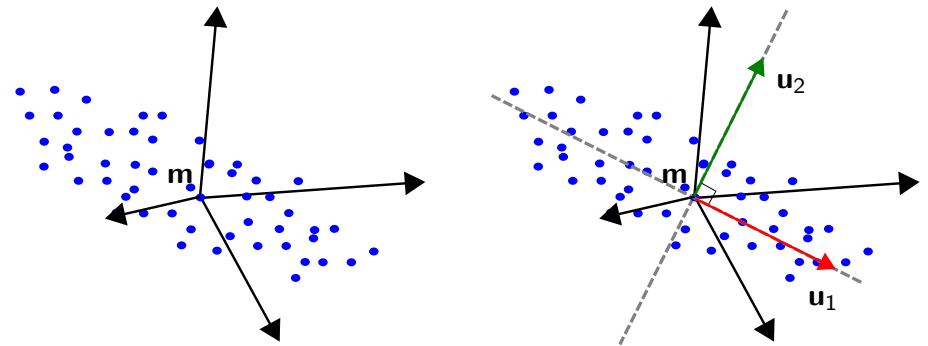
$$int_{\mathbb{A}}(\mathbb{NO}) = \frac{1}{n} \sum_{i=1}^n d^2(\mathbf{x}_i, \mathbf{0}) = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2$$

- Donc dans l'espace affine de \mathbb{R}^p dont l'origine est le barycentre, l'inertie est la moyenne pondérée des normes des vecteurs des individus.
- ▶ L'objectif est de déterminer des **sous-espaces** de sorte à ce que la projection des \mathbf{x}_i dans ceux-ci **conservernt au mieux l'inertie**. Autrement dit, on souhaite que l'image de \mathbb{NO} dans ce sous-espace soit **le moins déformé possible**.

Ajustement du nuage des individus

- Pour déterminer en pratique ces sous-espaces, on cherche une **suite de s directions privilégiées** ($s < p$) dans \mathbb{R}^p notées $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_s$ qui permettent de **maximiser l'inertie du nuage projeté**.
- Ce sont en fait s vecteurs de \mathbb{R}^p appelés **axes factoriels (ou principaux)** qui ont les propriétés suivantes :
 - ▶ \mathbf{u}_1 est le sous-espace de dimension 1 qui maximise l'inertie du nuage projeté.
 - ▶ \mathbf{u}_2 est orthogonal à \mathbf{u}_1 et le plan engendré par $\{\mathbf{u}_1, \mathbf{u}_2\}$ est l'espace de dimension 2 qui maximise l'inertie du nuage projeté.
 - ▶ \mathbf{u}_3 est orthogonal à \mathbf{u}_1 et \mathbf{u}_2 et le sous-espace engendré par $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ est l'espace de dimension 3 qui maximise l'inertie du nuage projeté.
 - ▶ ...
- Les vecteurs $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_s$ sont donc orthogonaux entre eux et permettent de maximiser l'inertie des points images lorsque l'on projette le nuage $\mathbb{N}\mathbb{O}$.

Illustration du plan factoriel



Détermination des axes factoriels (ou principaux)

- ▶ On montre que (cf partie sur la SVD) les vecteurs $\mathbf{u}_1, \dots, \mathbf{u}_s$ peuvent être obtenus en **diagonalisant la matrice des coefficients de corrélation** que l'on notera par \mathbf{C} et qui est de taille $(p \times p)$.
- ▶ En utilisant la matrice de données centrées-réduites \mathbf{X} , on a :

$$\mathbf{C} = \mathbf{X}^T \mathbf{X}$$

où \mathbf{X}^T est la transposée de \mathbf{X}

- De façon explicite, nous avons le terme général de \mathbf{C} , $\forall k, l$:

$$C_{kl} = \sum_{i=1}^n x_{ik} x_{il} \quad (\text{cf slide 241})$$

- ▶ Pour tout $m = 1, \dots, s$, \mathbf{u}_m est le vecteur propre associé à λ_m la m -ème plus grande valeur propre de \mathbf{C} .

Coordonnées des individus sur les axes factoriels

- ▶ L'axe factoriel \mathbf{u}_m est un vecteur de \mathbb{R}^p et on supposera par la suite qu'il est normé (càd $\langle \mathbf{u}_m, \mathbf{u}_m \rangle = 1$).
- Projeter $\mathbb{N}\mathbb{O}$ sur \mathbf{u}_m , c'est déterminer les coordonnées des individus \mathbf{x}_i sur ce vecteur.
- On obtient ces coordonnées en projetant orthogonalement les \mathbf{x}_i sur \mathbf{u}_m . Notons \mathbf{f}^m le vecteur de taille n comportant les composantes des individus sur l'axe \mathbf{u}_m . On montre que :

$$\mathbf{f}^m = \mathbf{X} \mathbf{u}_m$$

- De façon explicite on a pour tout individu i :

$$f_i^m = \langle \mathbf{x}_i, \mathbf{u}_m \rangle = \sum_{k=1}^p x_{ik} u_{m,k}$$

où f_i^m est le i -ème terme de \mathbf{f}^m et $u_{m,k}$ est le k -ème terme de \mathbf{u}_m .

Propriétés du vecteur des composantes des individus \mathbf{f}^m

- ▶ Les facteurs \mathbf{f}^m sont des vecteurs de \mathbb{R}^n obtenus à partir de combinaisons linéaires des variables initiales \mathbf{x}^k . On les appelle les **composantes principales**.
- Puisque $\mathbb{N}\mathbb{O}$ est centré selon son propre barycentre, les moyennes des \mathbf{f}^m sont nulles, $\forall m = 1, \dots, s$:

$$\frac{1}{n} \sum_{i=1}^n f_i^m = 0$$

- On montre également que les variances de \mathbf{f}^m ont des valeurs bien particulières, $\forall m = 1, \dots, s$:

$$\frac{1}{n} \sum_{i=1}^n (f_i^m)^2 = \frac{\lambda_m}{n}$$

- ▶ Autrement dit, la **variance du nuage projeté sur \mathbf{u}_m est égal (à un facteur $1/n$ près) à la valeur propre associée à \mathbf{u}_m !**

Ajustement du nuage des variables

- Pour ajuster le nuage des variables $\mathbb{N}\mathbb{A}$, la démarche est la même à cela près que les points sont les \mathbf{x}^k et que l'espace considéré est \mathbb{R}^n .
- La **distance euclidienne entre les variables** fait intervenir les coefficients de corrélation :

$$d^2(\mathbf{x}^k, \mathbf{x}^l) = \langle \mathbf{x}^k, \mathbf{x}^k \rangle + \langle \mathbf{x}^l, \mathbf{x}^l \rangle - 2\langle \mathbf{x}^k, \mathbf{x}^l \rangle = 2(1 - r(\mathbf{x}^k, \mathbf{x}^l))$$

- Remarques :

- ▶ Cette distance ne se fait pas par rapport au barycentre dans \mathbb{R}^n de $\mathbb{N}\mathbb{A}$ puisqu'on a centré les données par rapport au barycentre dans \mathbb{R}^p de $\mathbb{N}\mathbb{O}$. Cette opération correspond à une projection (et non à une translation) pour le nuage $\mathbb{N}\mathbb{A}$.
- ▶ Ainsi même si la notion d'inertie du nuage $\mathbb{N}\mathbb{A}$ n'est pas strictement la même que pour $\mathbb{N}\mathbb{O}$ l'objectif reste le même : on cherche des sous-espaces qui déforment le moins possible le nuage projeté de $\mathbb{N}\mathbb{A}$.

Détermination des axes factoriels

- ▶ La méthode est similaire : on détermine une suite de s directions privilégiées ($s < n$) dans \mathbb{R}^n notées $\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^s$ telles que la projection du nuage $\mathbb{N}\mathbb{A}$ conserve au mieux l'inertie de $\mathbb{N}\mathbb{A}$.
- ▶ On montre que (cf partie sur la SVD) les $\mathbf{v}^1, \dots, \mathbf{v}^s$ peuvent être obtenues en **diagonalisant la matrice des produits scalaires entre individus** que l'on notera par \mathbf{K} et qui est de taille $(n \times n)$:

$$\mathbf{K} = \mathbf{X}\mathbf{X}^\top$$

- Le terme général de \mathbf{K} est, $\forall i, j$:

$$K_{ij} = \sum_{k=1}^p x_{ik} x_{jk}$$

- ▶ Pour tout $m = 1, \dots, s$, \mathbf{v}^m est le vecteur propre associé à la m -ème plus grande valeur propre de \mathbf{K} .

Coordonnées des variables sur les axes factoriels

- \mathbf{v}^m est un vecteur de \mathbb{R}^n et on supposera par la suite qu'il est normé (càd $\langle \mathbf{v}^m, \mathbf{v}^m \rangle = 1$).
- Projeter $\mathbb{N}\mathbb{A}$ sur \mathbf{v}^m , c'est déterminer les coordonnées des variables \mathbf{x}^k sur ce vecteur.
- On obtient ces coordonnées en projetant orthogonalement les \mathbf{x}^k sur \mathbf{v}^m . Notons \mathbf{g}_m le vecteur de taille p comportant les composantes des variables sur l'axe \mathbf{v}^m . On montre que :

$$\mathbf{g}_m = \mathbf{X}^\top \mathbf{v}^m$$

- De façon explicite on a pour toute variable k :

$$g_{m,k} = \langle \mathbf{x}^k, \mathbf{v}^m \rangle = \sum_{i=1}^n x_{ik} v_i^m$$

où $g_{m,k}$ est le k -ème terme de \mathbf{g}_m et v_i^m est le i -ème terme de \mathbf{v}^m .

Propriétés du vecteur des composantes des variables \mathbf{g}_m

- ▷ On a la relation suivante qui fait le lien entre les axes factoriels issus de l'analyse de \mathbb{NO} et ceux issus de l'analyse de \mathbb{NA} , $\forall m = 1, \dots, s$:

$$\mathbf{v}^m = \frac{1}{\sqrt{\lambda_m}} \mathbf{X} \mathbf{u}_m = \frac{1}{\sqrt{\lambda_m}} \mathbf{f}^m$$

- ▷ \mathbf{v}^m est donc proportionnel à la composante principale \mathbf{f}^m .
 ▷ On montre par ailleurs que :

$$g_{m,k} = r(\mathbf{x}^k, \mathbf{v}^m)$$

- ▷ Comme $\forall \mathbf{a}, \mathbf{b} \in \mathbb{R}^n; \forall \alpha > 0 : r(\mathbf{a}, \mathbf{b}) = r(\mathbf{a}, \alpha \mathbf{b})$, on en déduit que :

$$g_{m,k} = r(\mathbf{x}^k, \mathbf{v}^m) = r(\mathbf{x}^k, \frac{1}{\sqrt{\lambda_m}} \mathbf{f}^m) = r(\mathbf{x}^k, \mathbf{f}^m)$$

- ▷ La coordonnée de la projection de la variable \mathbf{x}^k sur l'axe factoriel \mathbf{v}^m est donc aussi le coefficient de corrélation entre \mathbf{x}^k et le facteur \mathbf{f}^m .

Propriétés du vecteur des composantes des variables \mathbf{g}_m (suite)

- ▷ En fait, les vecteurs $\mathbf{v}^1, \dots, \mathbf{v}^s$ sont des vecteurs de \mathbb{R}^n qui sont **mutuellement orthogonaux et qui permettent de maximiser les sommes suivantes**, $\forall m = 1, \dots, s$:

$$\sum_{k=1}^p r^2(\mathbf{x}^k, \mathbf{v}^m)$$

- Les éléments de \mathbf{g}_m étant des coefficients de corrélations, nous avons, $\forall k = 1, \dots, m$:

$$|g_{m,k}| \leq 1$$

- Autrement dit, la projection d'une variable \mathbf{x}^k sur un vecteur \mathbf{v}^m est bornée entre -1 et 1 .
 • Ainsi, Les projections des variables dans les sous-espaces engendrés par les vecteurs \mathbf{v}^m sont à l'intérieur des **hypersphères de rayon 1**.

Dualité individus-variables

- Précédemment nous avons donné des **relations faisant le lien entre l'analyse de l'espace des individus et celle de l'espace des variables**. Nous donnons ici quelques autres points importants.
 • Tout d'abord, l'inertie de \mathbb{NO} et celle de \mathbb{NA} sont les mêmes !

$$int_{\mathbb{A}}(\mathbb{NO}) = int_{\mathbb{O}}(\mathbb{NA}) = \sum_{i=1}^n \sum_{k=1}^p (x_{ik})^2 = p$$

- La projection de chacun des deux nuages sur une suite d'axes factoriels correspond à une **décomposition de cette inertie totale**.
 ▷ L'inertie du nuage \mathbb{NO} projeté sur \mathbf{u}_m est noté $int_{\mathbf{u}_m}(\mathbb{NO})$ et on a :

$$int_{\mathbf{u}_m}(\mathbb{NO}) = int_{\mathbf{v}^m}(\mathbb{NA}) = \lambda_m \text{ (càd la valeur propre associée!)}$$

- ▷ La somme des valeurs propres donne l'inertie totale !

$$int_{\mathbb{A}}(\mathbb{NO}) = int_{\mathbb{O}}(\mathbb{NA}) = \sum_{k=1}^p \lambda_k$$

Dualité individus-variables (suite)

- Nous avons exposé précédemment la relation suivante :

$$\mathbf{v}^m = \frac{1}{\sqrt{\lambda_m}} \mathbf{f}^m$$

- Mais nous avons également la relation symétrique :

$$\mathbf{u}_m = \frac{1}{\sqrt{\lambda_m}} \mathbf{g}_m$$

- On voit que (cf également partie sur la SVD) l'analyse de \mathbb{NO} et celle de \mathbb{NA} sont **symétriques**. Les relations précédentes sont des **formules de transition** permettant de passer des résultats de l'analyse d'un espace à ceux de l'autre (à un facteur $1/\sqrt{\lambda_m}$ multiplicatif près).
 • Cette **dualité** a un premier intérêt pratique important : il n'est pas nécessaire de faire deux diagonalisations mais une seule ! Si $p < n$ (cas le plus courant) on choisira donc de diagonaliser \mathbf{C} .

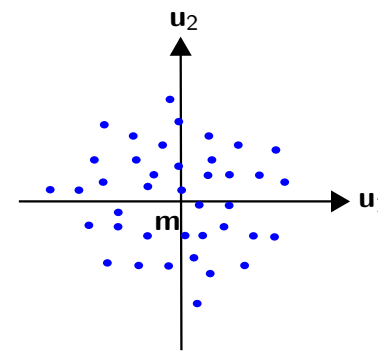
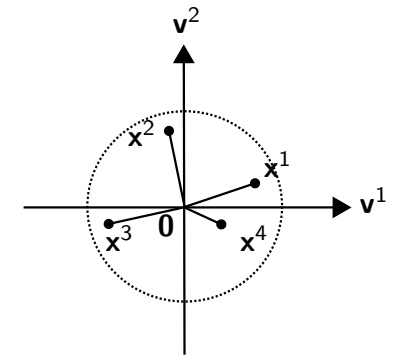
Représentations graphiques

- Pour les individus, nous prenons le repère affine centré en le barycentre \mathbf{m} (qui est un point de \mathbb{R}^p) et les axes factoriels $\mathbf{u}_1, \dots, \mathbf{u}_s$ qui forment une base orthonormale de \mathbb{R}^s (vu ici comme un sous-espace de \mathbb{R}^p). Dans cet espace affine, les individus ont comme coordonnées les composantes principales $\mathbf{f}^1, \dots, \mathbf{f}^s$.
- Pour les variables, nous prenons le repère affine centré en l'origine de \mathbb{R}^n (contrairement aux individus) et les axes factoriels $\mathbf{v}^1, \dots, \mathbf{v}^s$ qui forment aussi une base orthonormale de \mathbb{R}^s (vu ici comme un sous-espace de \mathbb{R}^n). Les variables dans cet espace sont les points ayant pour coordonnées les composantes des facteurs $\mathbf{g}_1, \dots, \mathbf{g}_s$.
- Comme l'analyse de NA est basée sur les angles et les corrélations entre vecteurs variables, on représente généralement ces points par des vecteurs (icàd des flèches). Aussi, comme ces points sont dans une hypersphère, on représente sur un plan factoriel le cercle de rayon unitaire. Ce graphique est appelé **cercle des corrélations**.

Représentations graphiques

- Les formules de transition permettent une **interprétation conjointe** des résultats des analyses des deux nuages. Cet aspect est un atout des méthodes factorielles souvent exploité dans divers domaines.
- Exemple d'interprétation : si un individu i a une forte valeur positive f_i^m et que la composante principale \mathbf{f}^m est fortement corrélée à un groupe de variables \mathbf{x}^k qui sont celles ayant des valeurs $g_{m,k}$ (càd $r(\mathbf{x}^k, \mathbf{f}^m)$) proches de 1, alors on peut dire que l'individu i est caractérisé par ce groupe de variables. De plus, les individus j proches de i sur l'axe \mathbf{u}_m (càd $f_j^m \approx f_i^m$) sont donc de type similaire.
 - Il faut cependant faire attention à la superposition des deux graphiques présentés précédemment :
 - Ces deux représentations n'ont pas la même origine et on ne peut donc pas faire correspondre \mathbf{m} (dans \mathbb{R}^p) et $\mathbf{0}$ (l'origine initiale de \mathbb{R}^n).
 - De plus, les distances entre les points individus d'un côté et les points variables de l'autre n'auraient pas de sens.

Représentations graphiques

Plan factoriel $(\mathbf{u}_1, \mathbf{u}_2)$ de NO.Plan factoriel (ou **cercle des corrélations**) $(\mathbf{v}^1, \mathbf{v}^2)$ de NA.

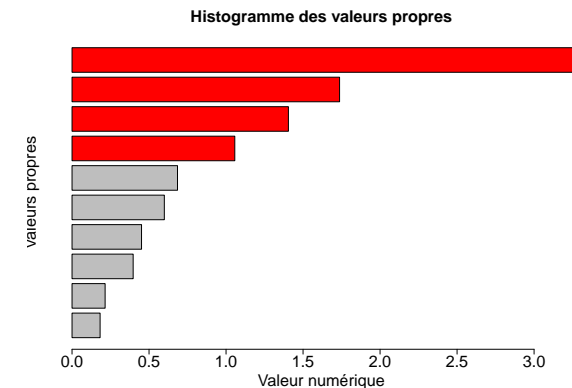
Choix du nombre d'axes

- Une question pratique importante est le choix du nombre d'axes à retenir càd la valeur de s .
- Il existe plusieurs approches :
 - des règles empiriques (critères de Catell et de Kayser, ...),
 - des procédures de validation externe (connaissances expertes et externes aux données, ...),
 - des critères statistiques basés sur les propriétés des valeurs propres (intervalles de confiance d'Anderson, ...),
 - des méthodes statistiques pour l'étude de la stabilité des axes (rééchantillonnage, bootstrap, ...)
- Nous discuterons brièvement des règles empiriques.

Choix du nombre d'axes (suite)

- Les **règles empiriques** consistent à observer l'allure de la séquence des valeurs propres λ_m rangées dans l'ordre décroissant.
- La **règle de Cattell** : on cherche des **ruptures** (ou discontinuités ou "coudes") dans l'"histogramme" de décroissance des valeurs propres. On interprète une rupture comme un "changement structurel" (càd non aléatoire) dans l'espace réduit qui justifierait le choix de s . A contrario, l'absence de rupture et d'une décroissance lente et régulière des valeurs propres indiquent un nuage de points relativement sphérique et laisse présager un faible intérêt des axes factoriels.
- La **règle de Kayser** : on retient uniquement les axes dont les valeurs propres sont supérieures à 1. En effet, on montre que la moyenne des valeurs propres vaut 1, ainsi on sélectionne les axes dont l'inertie est au-dessus de la moyenne (donc les plus informatifs). Attention ! c'est une règle simple mais qui peut-être mis en défaut dans plusieurs cas notamment lorsqu'on observe un **effet taille** important (cf plus loin).

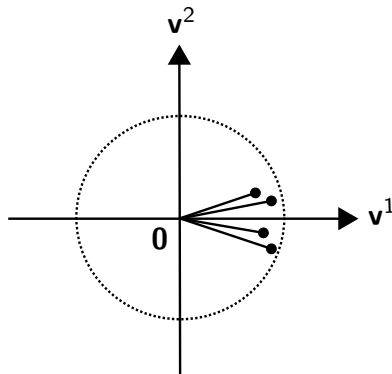
Exemple d'histogramme de décroissance des valeurs propres



- Il y a une rupture ou un "coude" après la 4ème valeur propre.
- En **rouge** sont les valeurs propres supérieures à 1.

L'effet taille

- Si les variables sont toutes corrélées positivement alors le nuage \mathcal{NA} est loin de l'origine de \mathbb{R}^n et le premier axe factoriel traduit cette information.
- Dans ce cas, les variables sont mal représentées dans l'espace réduit : elles se projettent à peu près toutes au même endroit sur le premier axe factoriel.



L'effet taille (suite)

- C'est l'**effet taille** et il se produit lorsqu'une partie des individus a de faibles valeurs pour toutes les variables tandis qu'un autre groupe a de fortes valeurs pour toutes les variables.
- On a alors un premier axe principal qui oppose ces deux groupes et dont la valeur propre associée est très forte en comparaison des autres.
- Avec un effet taille important, il est probable que les autres axes factoriels aient des valeurs propres associées en dessous de 1. Dans ce cas la règle de Kayser n'est pas bonne. En effet, malgré l'effet taille il est intéressant d'aller observer les autres axes factoriels qui traduisent d'autres types d'information sur la typologie des données.

Aide à l'interprétation à partir des données actives

- L'ACP aboutit à des approximations de nuages de points dans des espaces réduits. Nous avons des outils permettant d'apprécier la qualité de ces approximations.
- Ces outils sont aussi utilisés pour interpréter les résultats d'une ACP.
- Nous avons les quantités suivantes :
 - ▶ **Qualité** de la représentation **d'un élément** sur un axe.
 - ▶ **Qualité** de la représentation **d'un nuage** sur un axe.
 - ▶ **Contribution d'un élément** à la définition d'un axe.

Aide à l'interprétation de la représentation de $\mathbb{N}\mathbb{A}$

- En ce qui concerne les variables, nous avons déjà vu des éléments d'interprétations.
- Le vecteur \mathbf{g}_m contient les coordonnées des \mathbf{x}^k sur \mathbf{v}^m qui sont des coefficients de corrélation. En effet, rappelons que :

$$g_{m,k} = r(\mathbf{x}^k, \mathbf{v}^m)$$
- Ainsi, graphiquement, on voit que les variables fortement corrélés (positivement ou négativement) à un axe sont celles qui contribuent le plus à la définition de cet axe. Etant donné \mathbf{v}^m , on s'intéresse donc aux variables ayant les plus fortes coordonnées sur cet axe.
- Sachant que $\mathbf{f}^m = \sqrt{\lambda_m} \mathbf{v}^m$ et que $g_{m,k} = r(\mathbf{x}^k, \mathbf{f}^m)$ également, on interprétera la composante principale \mathbf{f}^m en fonction des groupements des variables ayant une coordonnée forte sur \mathbf{v}^m .
- Attention ! Il faut en revanche éviter d'interpréter les proximités entre variables qui ne sont pas proches du cercle des corrélations (càd celles qui sont proches du centre $\mathbf{0}$).

Aide à l'interprétation de la représentation de $\mathbb{N}\mathbb{O}$

- La **qualité de la représentation de l'individu i sur l'axe \mathbf{u}_m** est l'inertie de la projection de i sur \mathbf{u}_m divisée par l'inertie totale de i . Cette valeur égale aussi le (carré du) cosinus de l'angle formé entre le vecteur de la projection de \mathbf{x}_i sur \mathbf{u}_m et le vecteur \mathbf{x}_i :

$$qlt_{\mathbf{u}_m}(\mathbf{x}_i) = (f_i^m)^2 / \sum_{k=1}^p (x_{ik})^2$$

- La **qualité de la représentation de $\mathbb{N}\mathbb{O}$ sur l'axe \mathbf{u}_m** est l'inertie du nuage projeté (qui vaut λ_m) sur l'inertie totale du nuage (qui vaut p) :

$$qlt_{\mathbf{u}_m}(\mathbb{N}\mathbb{O}) = \lambda_m / p$$

- La **contribution de l'individu i à l'axe \mathbf{u}_m** est l'inertie de la projection de i sur \mathbf{u}_m divisée par l'inertie du nuage projeté sur \mathbf{u}_m :

$$ctr_{\mathbf{u}_m}(\mathbf{x}_i) = (f_i^m)^2 / \lambda_m$$

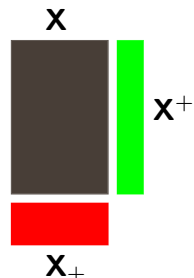
Ajouts de nouveaux éléments

- La matrice de données \mathbf{X} de taille $(n \times p)$ à partir de laquelle on extrait le sous-espace de dimension réduite contient les éléments (individus et variables) **actifs**.
- Au-delà de ces éléments, on peut projeter *a posteriori* d'autres éléments dans ces sous-espaces réduits. Ces éléments (individus ou variables) sont dits **supplémentaires ou illustratifs**.
- Ces éléments supplémentaires peuvent avoir un intérêt en termes d'interprétations lorsqu'il s'agira de caractériser les vecteurs des bases⁴ des sous-espaces réduits (cf ci-après).
- Pour formaliser la projection des éléments supplémentaires dans les sous-espaces réduits, on introduit les matrices suivantes :
 - ▶ \mathbf{X}_+ de taille $(n_+ \times p)$ des individus supplémentaires.
 - ▶ \mathbf{X}^+ de taille $(n \times p^+)$ des variables supplémentaires.

4. On parlera plutôt d'axes.

Ajouts de nouveaux éléments (suite)

- Illustration des matrices \mathbf{X} , \mathbf{X}_+ et \mathbf{X}^+ :



- La représentation des ind. supplémentaires dans l'espace réduit de \mathbb{R}^p engendré par (les colonnes de) $\mathbf{U} = (\mathbf{u}_1 \dots \mathbf{u}_s)$ est donnée par :

$$\mathbf{F}^+ = \mathbf{X}_+ \mathbf{U}$$

- La représentation des var. supplémentaires dans l'espace réduit de \mathbb{R}^n engendré par (les colonnes de) $\mathbf{V} = (\mathbf{v}^1 \dots \mathbf{v}^s)$ est donnée par :

$$\mathbf{G}_+ = (\mathbf{X}^+)^T \mathbf{V}$$

Aide à l'interprétation à partir de données supplémentaires

- Précédemment, nous avons exploité les données de la matrice \mathbf{X} afin d'interpréter les axes et de définir des indicateurs de qualité des nuages projetés. Il est possible d'utiliser des données supplémentaires afin d'affiner l'interprétation des sorties d'une ACP.
- On peut vouloir que certains individus ne soient pas pris en compte pour la détermination des axes principaux. On souhaite plutôt projeter ces **individus supplémentaires (ou illustratifs)** a posteriori (càd une fois les axes déterminés). Leurs positionnements par rapport aux **individus actifs** (càd ceux ayant contribué à la détermination des axes) peuvent avoir un intérêt pour l'analyse.
- De même, on peut projeter a posteriori des **variables continues supplémentaires (ou illustratives)** sur le cercle des corrélations. On peut alors visualiser les corrélations entre d'un côté les **variables actives**, axes principaux (ou composantes principales) et de l'autre côté, les variables supplémentaires.

Individus supplémentaires

- Les individus supplémentaires peuvent être des objets présentant un caractère exceptionnel (dont la prise en compte pèserait trop lourdement pour la détermination des axes en comparaison des autres), des objets comportant des erreurs de mesures ou des objets extérieurs à l'étude mais appartenant à un domaine voisin, ...
- Soit \mathbf{T}_+ une matrice de taille $(n_+ \times p)$ de terme général $t_{+,ik}$, dont les lignes sont n_+ vecteurs de \mathbb{R}^p représentant des individus supplémentaires dans l'espace engendré par \mathbf{A} .
- Pour positionner ces individus dans l'espace réduit, comme pour les individus actifs, on commence par centrer, réduire et diviser par \sqrt{n} la matrice \mathbf{T}_+ . On obtient \mathbf{X}_+ de terme général :

$$x_{+,ik} = \frac{t_{+,ik} - m_k}{s_k \sqrt{n}}$$

où m_k et s_k sont les moyennes et écart-types de \mathbf{x}^k (càd calculés à partir des n individus actifs).

Individus supplémentaires (suite)

- Afin de déterminer les coordonnées des individus supplémentaires on les projette sur les différents axes factoriels $\mathbf{u}_1, \dots, \mathbf{u}_s$.
- Ainsi, les composantes principales des n_+ individus supplémentaires sur l'axe \mathbf{u}_m sont données par :

$$\mathbf{f}^{+,m} = \mathbf{X}_+ \mathbf{u}_m$$

Variables continues supplémentaires

- Soit \mathbf{T}^+ une matrice de taille $(n \times p^+)$ de terme général t_{ik}^+ , dont les colonnes sont p^+ vecteurs de \mathbb{R}^n représentant des variables supplémentaires dans l'espace engendré par \mathbb{O} .
- On centre, réduit et divise par \sqrt{n} les éléments de \mathbf{T}^+ mais **contrairement aux individus, on utilise comme moyennes et écart-types ceux des nouvelles variables**. Soit $\mathbf{t}^{+,k}$ un vecteur colonne de \mathbf{T}^+ on a donc :

$$m_k^+ = \frac{1}{n} \sum_{i=1}^n t_{ik}^+ \quad \text{et} \quad (s_k^+)^2 = \frac{1}{n} \sum_{i=1}^n (t_{ik}^+ - m_k^+)^2$$

- On calcule ensuite la matrice \mathbf{X}^+ de terme général :

$$x_{ik}^+ = \frac{t_{ik}^+ - m_k^+}{s_k^+ \sqrt{n}}$$

Variables qualitatives nominales supplémentaires

- En pratique, il est souvent utile de **projeter a posteriori des variables nominales illustratives**.
- Ceci permet d'avoir une analyse plus riche et une interprétation plus fine des résultats.
- En revanche, il s'agit d'un type de variable différent de celui utilisé pour la détermination des axes donc les transformations précédentes ne sont pas valables.
- Soit \mathbf{q} une variable nominale ayant p_q modalités. Par exemple si la variable nominale est le genre alors il a $p_q = 2$ modalités et \mathbf{q} est un vecteur discret de taille n dont les éléments sont soit "F" soit "M".
- ▶ Pour représenter une variable nominale, on raisonne dans \mathbb{R}^p :
 - ▶ On va calculer les barycentres de chaque groupe d'individus correspondant à une modalité.
 - ▶ On projette ces p_q barycentres sur l'espace réduit engendré par $\mathbf{u}_1, \dots, \mathbf{u}_s$ (càd comme si on projetait des individus supplémentaires).

Variables continues supplémentaires (suite)

- Afin de déterminer les coordonnées des variables supplémentaires on les projette sur les différents axes factoriels $\mathbf{v}^1, \dots, \mathbf{v}^s$.
- ▶ Ainsi, les coordonnées des p^+ variables continues supplémentaires sur l'axe \mathbf{v}^m sont données par :

$$\mathbf{g}_{+,m} = (\mathbf{X}^+)^T \mathbf{v}^m$$

- $g_{+,m,k}$ est le coefficient de corrélation entre $\mathbf{x}^{+,k}$ et l'axe factoriel \mathbf{v}^m . Ainsi, la projection des variables illustratives sont également à l'intérieur d'hypersphères de rayon 1.

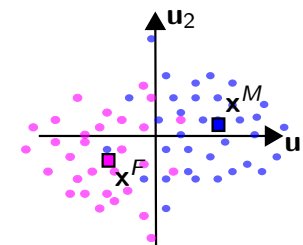
Variables qualitatives nominales supplémentaires (suite)

- Reprenons l'exemple précédent avec la variable du genre :
 - ▶ On calcule donc deux vecteurs \mathbf{x}^F et \mathbf{x}^M de \mathbb{R}^p défini par :

$$x_k^F = \frac{1}{|F|} \sum_{i:q_i=F} x_{ik} \quad \text{et} \quad x_k^M = \frac{1}{|M|} \sum_{i:q_i=M} x_{ik}$$
 où $|F|$ et $|M|$ sont resp. le nombre de femmes et d'hommes dans \mathbb{O} .
 - ▶ On projette \mathbf{x}^F et \mathbf{x}^M sur \mathbf{u}_m en calculant les coordonnées suivantes :

$$(\mathbf{x}^F)^T \mathbf{u}_m = \langle \mathbf{x}^F, \mathbf{u}_m \rangle \quad \text{et} \quad (\mathbf{x}^M)^T \mathbf{u}_m = \langle \mathbf{x}^M, \mathbf{u}_m \rangle$$

- Illustration :



Rappel du Sommaire

3 Méthodes de réduction de dimension

- Analyse générale : la décomposition en valeurs singulières (SVD)
- Analyse en composantes principales (ACP)
- Analyse générale : extension à des métriques et poids quelconques
- Analyse factorielle des correspondances (AFC)
- Analyse (factorielle) des correspondances multiples (ACM)

Métrique non canonique dans \mathbb{R}^p

- Nous raisonnons sans perte de généralité avec les vecteurs $\{\mathbf{x}_i\}_{i=1,\dots,n}$ qui constituent le nuage des objets ou individus \mathbb{NO} .
- L'espace de description est celui engendré par les p variables. On est donc dans \mathbb{R}^p .
- Nous avons supposé jusqu'ici que dans \mathbb{R}^p , la métrique était \mathbf{I}_p , c-à-d le produit scalaire canonique.
- De façon plus générale, soit \mathbf{M} de terme général $(m_{kl})_{k,l=1,\dots,p}$, une matrice carrée d'ordre p qui soit symétrique et définie positive.
- \mathbf{M} représente un **produit scalaire dans \mathbb{R}^p** et on note :

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle_{\mathbf{M}} = \underbrace{\mathbf{x}_i^{\top} \mathbf{M} \mathbf{x}_j}_{\text{écrit. mat.}} = \underbrace{\sum_{k=1}^p \sum_{l=1}^p x_{ik} x_{jl} m_{kl}}_{\text{écrit. alg.}}$$

- Rappel : en supposons une métrique \mathbf{M} , \mathbf{x} est normé si $\mathbf{x}^{\top} \mathbf{M} \mathbf{x} = 1$ et \mathbf{x} et \mathbf{y} sont orthogonaux si $\mathbf{x}^{\top} \mathbf{M} \mathbf{y} = 0$.

Généralisation à des poids et métriques non uniformes

- Nous avons vu les liens entre la recherche d'un sous-espace optimal en ADD et la SVD d'une matrice rectangulaire en AL et nous avons appliqué ceci dans le cadre de l'ACP normée.
- Nous avons jusqu'ici fait implicitement les hypothèses suivantes :
 - ▶ un *produit scalaire canonique* dans \mathbb{R}^p et \mathbb{R}^n ,
 - ▶ un *poids uniforme* sur les individus et les variables.
- Or, nous avons vu dans la 1ère partie du cours que nous pouvions définir dans un espace euclidien une infinité de produits scalaires.
- De plus en statistique, il est souvent utile de donner des poids différents aux individus et aux variables.
- ▶ Nous généralisons les résultats précédents :
 - ▶ à **tout type de métrique** (produit scalaire) dans \mathbb{R}^p et \mathbb{R}^n ,
 - ▶ à **tout type de pondération** sur les individus et ou sur les variables.

Poids non uniformes

- Nous faisons également l'hypothèse qu'à tout vecteur individu \mathbf{x}_i est associé un poids $n_i \geq 0$.
- Soit \mathbf{n} le vecteur de taille $(n \times 1)$ des poids des individus.
- On introduit \mathbf{N} la matrice des poids qui est carrée d'ordre n , diagonale de terme général :

$$\mathbf{N}_{ii} = n_i$$

- Rappel : $\bar{\mathbf{x}}$, le barycentre des $\{\mathbf{x}_i\}_{i=1,\dots,n}$ est alors défini comme suit :

$$\bar{\mathbf{x}} = \sum_{i=1}^n n_i \mathbf{x}_i$$

Inertie et approximation de NO dans le cas général

- L'inertie définie en slide 213 avec \mathbf{M} et \mathbf{N} est maintenant **pondérée** :

$$\begin{aligned} \text{int}(\text{NO}) &= \sum_{i=1}^n n_i d_{\mathbf{M}}^2(\mathbf{x}_i - \bar{\mathbf{x}}) = \sum_{i=1}^n n_i \|\mathbf{x}_i - \bar{\mathbf{x}}\|_{\mathbf{M}}^2 \\ &= \sum_{i=1}^n n_i \langle \mathbf{x}_i - \bar{\mathbf{x}}, \mathbf{x}_i - \bar{\mathbf{x}} \rangle_{\mathbf{M}} \end{aligned}$$

- Notons $\mathbf{x}' = \frac{\langle \mathbf{x}, \mathbf{u}_1 \rangle_{\mathbf{M}}}{\|\mathbf{u}_1\|_{\mathbf{M}}} \mathbf{u}_1$, le **vecteur projeté** de \mathbf{x} sur $\mathbf{u}_1 \in \mathbb{R}^p$.
- Les développements des slides 214 et 215 restent valables et la recherche de la droite vectorielle dans \mathbb{R}^p maximisant la variance du nuage projeté au sens de \mathbf{M} et tenant compte des poids \mathbf{N} se réduit à :

$$\max_{\mathbf{u}_1 \in \mathbb{R}^p} \text{int}_{\mathbf{u}_1}(\text{NO}) = \max_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{i=1}^n n_i \|\mathbf{x}'_i - \bar{\mathbf{x}}'\|_{\mathbf{M}}^2 = \max_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{i=1}^n n_i \|\mathbf{x}'_i\|_{\mathbf{M}}^2$$

$$\text{où } \mathbf{x}'_i = \frac{\langle \mathbf{x}_i, \mathbf{u}_1 \rangle_{\mathbf{M}}}{\|\mathbf{u}_1\|_{\mathbf{M}}} \mathbf{u}_1 \text{ et } \bar{\mathbf{x}}' = \sum_{i=1}^n n_i \mathbf{x}'_i.$$

Inertie et approximation de NO dans le cas général (suite)

- Similairement aux slides 220, 221, on pose $\mathbf{Q} = \mathbf{M}\mathbf{X}^T \mathbf{N} \mathbf{X} \mathbf{M}$ et on cherche le vecteur de \mathbb{R}^p qui maximise l'inertie (pondérée) du nuage projeté au sens de \mathbf{M} s.l.c. que ce vecteur soit \mathbf{M} -normé.
- On pose le lagrangien :

$$\text{lag}(\mathbf{u}_1, \lambda) = \mathbf{u}_1^T \mathbf{Q} \mathbf{u}_1 - \lambda (\mathbf{u}_1^T \mathbf{M} \mathbf{u}_1 - 1)$$

- Les conditions nécessaires du premier ordre donnent :

$$\begin{aligned} \frac{\partial \text{lag}}{\partial \mathbf{u}_1}(\mathbf{u}_1, \lambda) = \mathbf{0} &\Leftrightarrow 2\mathbf{Q}\mathbf{u}_1 - 2\lambda\mathbf{M}\mathbf{u}_1 = \mathbf{0} \\ &\Leftrightarrow \mathbf{Q}\mathbf{u}_1 = \lambda\mathbf{M}\mathbf{u}_1 \\ &\Leftrightarrow \mathbf{M}\mathbf{X}^T \mathbf{N} \mathbf{X} \mathbf{M} \mathbf{u}_1 = \lambda\mathbf{M}\mathbf{u}_1 \\ &\Leftrightarrow \mathbf{X}^T \mathbf{N} \mathbf{X} \mathbf{M} \mathbf{u}_1 = \lambda \mathbf{u}_1 \end{aligned}$$

- Les solutions du problème correspondent à la décomposition spectrale de la matrice $\mathbf{X}^T \mathbf{N} \mathbf{X} \mathbf{M}$.

Inertie et approximation de NO dans le cas général (suite)

- On peut faire les mêmes développements que dans les slides 217 et 218 et en supposant que \mathbf{u}_1 est \mathbf{M} -normé, on obtient :

$$\begin{aligned} \sum_{i=1}^n n_i \|\mathbf{x}'_i\|_{\mathbf{M}}^2 &= \sum_{i=1}^n n_i \mathbf{u}_1^T \mathbf{M} \mathbf{x}_i \mathbf{x}_i^T \mathbf{M} \mathbf{u}_1 \\ &= \mathbf{u}_1^T \mathbf{M} \left(\sum_{i=1}^n n_i \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{M} \mathbf{u}_1 \\ &= \mathbf{u}_1^T \mathbf{M} \mathbf{X}^T \mathbf{N} \mathbf{X} \mathbf{M} \mathbf{u}_1 \end{aligned}$$

- On obtient le problème d'optimisation contraint suivant :

$$\max_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{i=1}^n \|\mathbf{x}'_i\|_{\mathbf{M}}^2 \Leftrightarrow \max_{\mathbf{u}_1 \in \mathbb{R}^p : \|\mathbf{u}_1\|_{\mathbf{M}}^2 = 1} \mathbf{u}_1^T \underbrace{\mathbf{M} \mathbf{X}^T \mathbf{N} \mathbf{X} \mathbf{M}}_{\mathbf{Q}} \mathbf{u}_1$$

où $\mathbf{M} \mathbf{X}^T \mathbf{N} \mathbf{X} \mathbf{M}$ est une matrice de Gram (et donc carrée symétrique et sdp) qui est vue telle une forme quadratique.

Remarque sur la dualité dans le cas général

- Précédemment l'hypothèse du produit scalaire canonique valait pour NO dans \mathbb{R}^p mais également pour NA dans \mathbb{R}^n .
- Comme nous venons de voir, nous pouvons définir une métrique \mathbf{M} dans \mathbb{R}^p et une pondération non uniforme \mathbf{N} pour les individus.
- Mais nous pouvons faire de même et de façon indépendante, pour \mathbb{R}^n où sont représentés les vecteurs $\{\mathbf{x}^k\}_{k=1, \dots, p}$ de NA. Pour cela les développements des derniers slides peuvent être appliqués à la matrice \mathbf{X}^T au lieu de \mathbf{X} et nous prenons \mathbf{P} la métrique dans \mathbb{R}^n ainsi qu'une pondération non uniforme pour les variables représentées par la matrice diagonale \mathbf{Q} . C'est alors la décomposition spectrale de $\mathbf{X} \mathbf{Q} \mathbf{X}^T \mathbf{P}$ qui donne l'espace réduit pour NA.
- Si les couples (\mathbf{M}, \mathbf{N}) pour $(\mathbb{R}^p, \text{NO})$ et (\mathbf{P}, \mathbf{Q}) pour $(\mathbb{R}^n, \text{NA})$ n'ont pas de relations particulières alors la dualité vue en slides 225 et 229 n'est en général pas valide.
- Néanmoins, en ADD, ces couples sont en fait liés et nous verrons comment tirer profit de la dualité dans chacun des cas particuliers.

Sous-espace réduits et projections dans le cas général

- ▶ Pour déterminer un sous-espace de dimension $s < p$ qui maximise l'inertie du nuage projeté compte tenu de la métrique \mathbf{M} et de la pondération \mathbf{N} , on prendra comme base les s vecteurs propres normés de $\mathbf{X}^\top \mathbf{N} \mathbf{X} \mathbf{M}$ associés aux s plus grande valeurs propres : $\{\mathbf{u}_1, \dots, \mathbf{u}_s\}$.
- ▶ Les s premières valeurs propres $\{\lambda_1, \dots, \lambda_s\}$ sont resp. les mesures de l'inertie du nuage projeté sur chacun des axes $\{\mathbf{u}_1, \dots, \mathbf{u}_s\}$.
- ▶ Les coordonnées de \mathbf{x}_i sur la droite vectorielle \mathbf{u}_m est donnée par :

$$\langle \mathbf{x}_i, \mathbf{u}_m \rangle_{\mathbf{M}} = \mathbf{x}_i^\top \mathbf{M} \mathbf{u}_m$$

- ▶ Plus généralement la représentation de tous les vecteurs dans l'espace réduit de dimension s est donnée par :

$$\mathbf{F} = \mathbf{X} \mathbf{M} \mathbf{U} = (f_{im})_{i=1, \dots, n; m=1, \dots, s} \text{ où } f_{im} = \langle \mathbf{x}_i, \mathbf{u}_m \rangle_{\mathbf{M}}$$

$$\text{où } \mathbf{U} = (\mathbf{u}_1 \quad \dots \quad \mathbf{u}_s).$$

Ajouts de nouveaux éléments dans le cas général (suite)

- Ce cas change par rapport au slide 272, c'est que nous avons des métriques non canoniques dans le cas général dont il faut tenir compte.
- La représentation des ind. supplémentaires dans l'espace réduit de \mathbb{R}^p engendré par (les colonnes de) $\mathbf{U} = (\mathbf{u}_1 \quad \dots \quad \mathbf{u}_s)$ est donnée par :

$$\mathbf{F}^+ = \mathbf{X}_+ \mathbf{M} \mathbf{U}$$

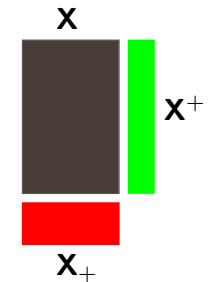
- La représentation des var. supplémentaires dans l'espace réduit de \mathbb{R}^n engendré par (les colonnes de) $\mathbf{V} = (\mathbf{v}^1 \quad \dots \quad \mathbf{v}^s)$ est donnée par :

$$\mathbf{G}_+ = (\mathbf{X}^+)^{\top} \mathbf{P} \mathbf{V}$$

Ajouts de nouveaux éléments dans le cas général

- Les motivations sont les mêmes que celles exposées au slide 272.
- Dans la matrice de données \mathbf{X} de taille $(n \times p)$, les individus et variables sont dits **actifs**.
- On dispose d'individus et/ou de variables **supplémentaires** ou **illustratifs**. Les matrices de données supplémentaires sont notées :
 - ▶ \mathbf{X}_+ de taille $(n_+ \times p)$ des individus supplémentaires.
 - ▶ \mathbf{X}^+ de taille $(n \times p^+)$ des variables supplémentaires.

- Illustration des matrices \mathbf{X} , \mathbf{X}_+ et \mathbf{X}^+ :



Utilité de la généralisation

- L'ACP normée applique de façon quasi-systématique les éléments d'AL développés du slide 208 à 235 en supposant des métriques canoniques et des poids uniformes.
- Nous voyons dans ce qui suit le traitement de table de données issues de variables qualitatives nominales. Les techniques de réduction de dimension s'appellent alors AFC et ACM.
- Ces méthodes reposent en fait sur la SVD généralisée que nous venons de voir avec des métriques et poids particuliers.
- Dans le cadre de l'AFC et de l'ACM, les différents concepts géométriques ont également des interprétations statistiques (test et distance du χ^2) que nous préciserons au moment venu.

Rappel du Sommaire

3 Méthodes de réduction de dimension

- Analyse générale : la décomposition en valeurs singulières (SVD)
- Analyse en composantes principales (ACP)
- Analyse générale : extension à des métriques et poids quelconques
- Analyse factorielle des correspondances (AFC)
- Analyse (factorielle) des correspondances multiples (ACM)

Données traitées par l'AFC et notations (suite)

- On a donc :

$$\mathbf{N} = \begin{matrix} & \mathbf{n}_{1.} & \dots & \mathbf{n}_{j.} & \dots & \mathbf{n}_{q.} \\ \begin{matrix} \mathbf{n}_{1.} \\ \vdots \\ \mathbf{n}_{i.} \\ \vdots \\ \mathbf{n}_{p.} \end{matrix} & \left(\begin{array}{cccc} & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \end{array} \right) \end{matrix}$$

- Le vecteur $\mathbf{n}_{i.} = (n_{i1}, \dots, n_{iq})$ donne la répartition des effectifs selon les modalités j de Q au sein de la population des individus ayant la modalité i de P .
- Le vecteur $\mathbf{n}_{.j} = (n_{1j}, \dots, n_{pj})$ donne la répartition des effectifs selon les modalités i de P au sein de la population des individus ayant la modalité j de Q .

Données traitées par l'AFC et notations

- L'Analyse Factorielle des Correspondances (AFC) traite des **tableaux de contingence (ou tableaux croisés)**.
- Le tableau de contingence croise **deux variables qualitatives nominales** et compte le nb d'effectifs pour chaque paire de modalités.
- Formellement, soient P et Q deux variables qualitatives possédant p et q modalités. On note \mathbf{N} le tableau de contingence qui est une matrice de taille $(p \times q)$ de terme général :

n_{ij} = Nombre d'individus ayant les modalités i de P et j de Q

- Notons par $\mathbf{n}_{i.}$ le vecteur de taille q donnant pour l'ensemble des individus possédant la modalité i de P , la répartition des effectifs selon les modalités de Q .
- De façon symétrique $\mathbf{n}_{.j}$ est le vecteur de taille p donnant pour l'ensemble des individus possédant la modalité j de Q , la répartition des effectifs selon les modalités de P .

Exemple

- Exemple tiré de L. Lebart, M. Piron, A. Morineau. *Statistique exploratoire multidimensionnelle*. Dunod. 2006 (4ème Ed.).
- P =Couleur des yeux.
- Q =Couleur des cheveux.
- Sur un ensemble de $n=592$ femmes.

	<i>brun</i>	<i>chatain</i>	<i>roux</i>	<i>blond</i>
<i>marron</i>	68	119	26	7
<i>noisette</i>	15	54	14	10
<i>vert</i>	5	29	14	16
<i>bleu</i>	20	84	17	94

Données traitées par l'AFC et notations (suite)

$$\mathbf{N} = \begin{matrix} & \begin{matrix} n_{1.} & \dots & n_{.j} & \dots & n_{.q} \end{matrix} \\ \begin{matrix} n_{1.} \\ \vdots \\ n_{i.} \\ \vdots \\ n_{p.} \end{matrix} & \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ \dots & \dots & n_{ij} & \dots & \dots \\ & & & & \\ & & & & \end{pmatrix} & \begin{matrix} n_{1.} \\ \vdots \\ n_{i.} \\ \vdots \\ n_{p.} \end{matrix} \\ \begin{matrix} n_{1.} & \dots & n_{.j} & \dots & n_{.q} \end{matrix} & & n \end{matrix}$$

- On définit également les **marges de N** comme suit :
 - $n_{i.} = \sum_{j \in Q} n_{ij}$ = nombre total d'individus ayant la modalité i de P .
 - $n_{.j} = \sum_{i \in P} n_{ij}$ = nombre total d'individus ayant la modalité j de Q .
- Le nombre total d'individus n est tel que $n = \sum_{i \in P} n_{i.} = \sum_{j \in Q} n_{.j}$.

Exemple

$$\mathbf{N} = \begin{matrix} & \begin{matrix} brun & chatain & roux & blond \end{matrix} \\ \begin{matrix} marron \\ noisette \\ vert \\ bleu \end{matrix} & \begin{pmatrix} 68 & 119 & 26 & 7 \\ 15 & 54 & 14 & 10 \\ 5 & 29 & 14 & 16 \\ 20 & 84 & 17 & 94 \end{pmatrix} & \begin{matrix} 220 \\ 93 \\ 64 \\ 215 \end{matrix} \\ \begin{matrix} 108 & 286 & 71 & 127 \end{matrix} & & 592 \end{matrix}$$

Tableau des fréquences relatives

- \mathbf{N} est le tableau d'effectifs bruts. En AFC, on transforme \mathbf{N} afin d'obtenir des **tableaux de fréquences relatives**. Ce tableau est une matrice notée \mathbf{F} de taille $(p \times q)$ et de terme général :

$$f_{ij} = \frac{n_{ij}}{n}$$

= **Pourcentage** d'individus ayant les modalités (i, j) de $P \times Q$

- Dans ce cas, nous avons les définitions suivantes :

$$f_{i.} = \sum_{j \in Q} f_{ij}$$

= Pourcentage des individus ayant la modalité i de P

$$f_{.j} = \sum_{i \in P} f_{ij}$$

= Pourcentage des individus ayant la modalité j de Q

Tableau des fréquences relatives (suite)

$$\mathbf{F} = \begin{matrix} & \begin{matrix} f_{1.} & \dots & f_{.j} & \dots & f_{.q} \end{matrix} \\ \begin{matrix} f_{1.} \\ \vdots \\ f_{i.} \\ \vdots \\ f_{p.} \end{matrix} & \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ \dots & \dots & \frac{n_{ij}}{n} & \dots & \dots \\ & & & & \\ & & & & \end{pmatrix} \end{matrix}$$

- Le vecteur $\mathbf{f}_{i.} = (f_{i1}, \dots, f_{iq})$ donne les fréquences d'observation des modalités j de Q au sein de la population des individus possédant la modalité i de P .
- Le vecteur $\mathbf{f}_{.j} = (f_{1j}, \dots, f_{pj})$ donne les fréquences d'observation des modalités i de P au sein de la population des individus possédant la modalité j de Q .

Tableau des fréquences relatives (suite)

$$\mathbf{F} = \begin{matrix} & \begin{matrix} f_{.1} & \dots & f_{.j} & \dots & f_{.q} \end{matrix} \\ \begin{matrix} f_{1.} \\ \vdots \\ f_{i.} \\ \vdots \\ f_{p.} \end{matrix} & \left(\begin{array}{ccccc} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{array} \right) & \begin{matrix} f_{1.} \\ \vdots \\ f_{i.} \\ \vdots \\ f_{p.} \end{matrix} \\ \begin{matrix} f_{.1} & \dots & f_{.j} & \dots & f_{.q} \end{matrix} & & 1 \end{matrix}$$

- Les marges de \mathbf{F} sont donc définies comme suit :
 - $f_{i.} = \sum_{j \in Q} f_{ij}$ = Pourcentage des individus ayant la modalité i de P .
 - $f_{.j} = \sum_{i \in P} f_{ij}$ = Pourcentage des individus ayant la modalité j de Q .
- La somme totale des éléments de \mathbf{F} est telle que

$$\sum_{i,j} f_{ij} = \sum_i f_{i.} = \sum_j f_{.j} = 1.$$

Exemple

$$\mathbf{F} = \begin{matrix} & \begin{matrix} brun & chatain & roux & blond \end{matrix} \\ \begin{matrix} marron \\ noisette \\ vert \\ bleu \end{matrix} & \left(\begin{array}{cccc} & & & \\ & & & \\ & & & \\ & & & \\ & & & \end{array} \right) & \begin{matrix} .37 \\ .16 \\ .11 \\ .36 \end{matrix} \\ \begin{matrix} .12 & .20 & .04 & .01 \\ .03 & .09 & .02 & .02 \\ .01 & .05 & .02 & .03 \\ .03 & .14 & .03 & .16 \end{matrix} & & \begin{matrix} .19 & .48 & .11 & .22 & 1.0 \end{matrix} \end{matrix}$$

Indépendance et liaisons entre variables qualitatives

- Le tableau \mathbf{F} permet d'étudier la liaison entre P et Q .
- Un concept central dans ce cas est l'**indépendance statistique** entre deux variables qualitatives.
- On dit que P et Q sont **indépendantes**, ce qu'on note aussi par $P \perp Q$, si leur tableau \mathbf{F} est tel que :

$$\forall (i,j) \in P \times Q : f_{ij} = f_{i.} f_{.j}$$

- La situation d'indépendance exprime l'absence de liaison.** En effet $f_{ij} = f_{i.} f_{.j}$ indique que la probabilité d'observer de façon jointe la paire de modalités (i,j) de $P \times Q$ est égale au produit de la probabilité d'observer la modalité i de P et de la probabilité d'observer la modalité j de Q .
- On dit également qu'il y a indépendance lorsque **la probabilité jointe est égale au produit des probabilités marginales**.

Indépendance et liaisons entre variables qualitatives (suite)

- L'indépendance s'interprète aussi en termes de probabilités conditionnelles $P \perp Q$ si pour toute paire de modalités (i,j) , le fait d'observer la modalité j de Q n'impacte aucunement la probabilité d'observer la modalité i de P (et vice versa).
- L'écart à l'indépendance permet donc de mesurer la liaison entre P et Q .** Ainsi, plus les valeurs f_{ij} sont distinctes des produits $f_{i.} f_{.j}$ plus la liaison entre P et Q est avérée. Dans ce cas, la probabilité d'observer la modalité i de P dépend de la modalité j de Q qui est observée (et vice versa). Nous pouvons alors considérer deux sous-cas :
 - Si $f_{ij} > f_{i.} f_{.j}$, la probabilité jointe est supérieure au produit des probabilités marginales : " i et j s'**attirent**".
 - Si $f_{ij} < f_{i.} f_{.j}$, la probabilité jointe est inférieure au produit des probabilités marginales : " i et j se **repoussent**".

Exemple

- Table de fréquences empiriques :

$$F = \begin{array}{c} \text{marron} \\ \text{noisette} \\ \text{vert} \\ \text{bleu} \end{array} \begin{pmatrix} \text{brun} & \text{chatain} & \text{roux} & \text{blond} \\ \begin{pmatrix} .12 & .20 & .04 & .01 \\ .03 & .09 & .02 & .02 \\ .01 & .05 & .02 & .03 \\ .03 & .14 & .03 & .16 \end{pmatrix} \end{pmatrix} \begin{array}{c} .37 \\ .16 \\ .11 \\ .36 \end{array}$$

.19 .48 .11 .22 1.0

- Table de fréquences théorique s'il y avait indépendance :

$$F = \begin{array}{c} \text{marron} \\ \text{noisette} \\ \text{vert} \\ \text{bleu} \end{array} \begin{pmatrix} \text{brun} & \text{chatain} & \text{roux} & \text{blond} \\ \begin{pmatrix} .07 & .18 & .04 & .08 \\ .03 & .08 & .02 & .03 \\ .02 & .05 & .01 & .02 \\ .07 & .18 & .04 & .08 \end{pmatrix} \end{pmatrix} \begin{array}{c} .37 \\ .16 \\ .11 \\ .36 \end{array}$$

.19 .48 .11 .22 1.0

Indépendance et liaisons entre variables qualitatives (suite)

$$F = \begin{pmatrix} f_{11} & \dots & f_{1j} & \dots & f_{1q} \\ \vdots & & \vdots & & \vdots \\ f_{i1} & \dots & f_{ij} & \dots & f_{iq} \\ \vdots & & \vdots & & \vdots \\ f_{p1} & \dots & f_{pj} & \dots & f_{pq} \\ f_{\cdot 1} & \dots & f_{\cdot j} & \dots & f_{\cdot q} \end{pmatrix} \begin{array}{c} f_{1\cdot} \\ \vdots \\ f_{i\cdot} \\ \vdots \\ f_{p\cdot} \end{array}$$

- En cas d'**indépendance**, nous avons aussi les propriétés suivantes

- Les **lignes** i de terme général $f_{ij}/f_{i\cdot}$ sont **proportionnelles à la marge** :

$$f_{ij} = f_{i\cdot} f_{\cdot j} \Leftrightarrow \frac{f_{ij}}{f_{i\cdot}} = f_{\cdot j}$$

- Les **colonnes** j de terme général $f_{ij}/f_{\cdot j}$ sont **proportionnelles à la marge** :

$$f_{ij} = f_{i\cdot} f_{\cdot j} \Leftrightarrow \frac{f_{ij}}{f_{\cdot j}} = f_{i\cdot}$$

Profils lignes et profils colonnes

- On introduit alors les concepts de **profils lignes** et **profils colonnes** qui sont des répartitions en pourcentage à l'intérieur d'une ligne et d'une colonne respectivement.
- Les **profils lignes** i sont définis comme suit :

$$\frac{f_{ij}}{f_{i\cdot}} = \text{Pourcentage des individus ayant la modalité } j \text{ de } Q \text{ parmi ceux ayant la modalité } i \text{ de } P$$

On remarquera que $\frac{f_{ij}}{f_{i\cdot}} = \frac{n_{ij}}{n_{i\cdot}}$.

- Les **profils colonnes** j sont définis comme suit :

$$\frac{f_{ij}}{f_{\cdot j}} = \text{Pourcentage des individus ayant la modalité } i \text{ de } P \text{ parmi ceux ayant la modalité } j \text{ de } Q$$

On remarquera que $\frac{f_{ij}}{f_{\cdot j}} = \frac{n_{ij}}{n_{\cdot j}}$.

Objectifs de l'AFC

- L'objectif général est identique à celui poursuivi en ACP même si la nature des données est différente.
- On a **deux types de nuage de points** :
 - NL le **nuage des profils lignes** : chaque profil ligne est vu tel un vecteur appartenant à un espace à q dimensions.
 - NC le **nuage des profils colonnes** : chaque profil colonne est vu tel un vecteur appartenant à un espace à p dimensions.
- On cherche des sous-espaces de dimension réduite qui permettent de conserver au mieux l'inertie des nuages de points.
- On cherche également à déterminer des typologies des éléments de chaque nuage et d'interpréter celles-ci conjointement.
- L'inertie est, comme pour l'ACP, basée sur la notion de distance entre points mais **en AFC on utilise une distance particulière** qui s'interprète également en termes d'indépendance statistique.

Nuage des profils lignes NL

- Les p profils lignes sont représentés par p vecteurs $\mathbf{l}_1, \dots, \mathbf{l}_p$ au sein d'une matrice de données \mathbf{L} de taille $(p \times q)$:

$$\mathbf{L} = \begin{matrix} & \mathbf{l}_1 & \dots & \mathbf{l}_j & \dots & \mathbf{l}_p \\ \begin{matrix} \vdots \\ \vdots \\ \vdots \end{matrix} & \begin{pmatrix} \frac{f_{11}}{f_{1.}} & \dots & \frac{f_{1j}}{f_{1.}} & \dots & \frac{f_{1q}}{f_{1.}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{i1}}{f_{i.}} & \dots & \frac{f_{ij}}{f_{i.}} & \dots & \frac{f_{iq}}{f_{i.}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{p1}}{f_{p.}} & \dots & \frac{f_{pj}}{f_{p.}} & \dots & \frac{f_{pq}}{f_{p.}} \end{pmatrix} & \begin{matrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{matrix} \end{matrix}$$

- Le vecteur \mathbf{l}_i est de composantes $(\frac{f_{i1}}{f_{i.}}, \dots, \frac{f_{ij}}{f_{i.}}, \dots, \frac{f_{iq}}{f_{i.}})$.
- NL est donc la représentation des $\mathbf{l}_1, \dots, \mathbf{l}_p$ dans \mathbb{R}^q .
- Comme tous les profils lignes sont de somme unitaire, les \mathbf{l}_i appartiennent à un hyperplan de \mathbb{R}^q .

Barycentre de NL

- En AFC, les vecteurs lignes et colonnes ont également une **pondération non uniforme** qui est imposée contrairement à l'ACP.
- Chaque profil ligne \mathbf{l}_i a un poids p_i qui est égal à $f_{i.}$ (la fréquence de la population ayant la modalité i de P).
- Le barycentre de NL dénoté $\bar{\mathbf{l}}$ est par conséquent défini par :

$$\bar{\mathbf{l}} = \sum_{i=1}^p p_i \mathbf{l}_i$$

- Plus précisément son terme général \bar{l}_j est tel que $\forall i = 1, \dots, q$:

$$\begin{aligned} \bar{l}_j &= \sum_{i=1}^p f_{i.} \frac{f_{ij}}{f_{i.}} \\ &= f_{.j} \end{aligned}$$

► Ainsi, les composantes de $\bar{\mathbf{l}}$ sont les marges des modalités de Q .

Distance du χ^2 entre profils lignes

- Contrairement à l'ACP où on utilise la distance euclidienne classique pour mesurer la proximité entre individus, on utilise en AFC la **distance du χ^2** définie comme suit :

$$d_{\chi^2}^2(\mathbf{l}_i, \mathbf{l}_{i'}) = \sum_{j=1}^q \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2$$

- Interprétation de la distance du χ^2 :
 - Comme pour la distance euclidienne classique, on compare pour chaque dimension l'écart entre les deux vecteurs et on prend son carré.
 - Mais ici, il y a une **pondération $1/f_j$ associée à la dimension j** de l'espace. Ainsi, plus la fréquence de la modalité j est forte, moins le poids de cette dimension est important. La distance du χ^2 avantage donc les écarts vis à vis d'une modalité qui est peu fréquente ("rare").

Exemple

$$\mathbf{L} = \begin{matrix} & \begin{matrix} marron \\ noisette \\ vert \\ bleu \\ \bar{\mathbf{l}} \end{matrix} & \begin{pmatrix} .31 & .54 & .12 & .03 \\ .16 & .58 & .15 & .11 \\ .08 & .45 & .22 & .25 \\ .09 & .39 & .08 & .44 \\ .19 & .48 & .11 & .22 \end{pmatrix} & \begin{matrix} 1.0 \\ 1.0 \\ 1.0 \\ 1.0 \\ 1.0 \end{matrix} \end{matrix}$$

- On remarquera par exemple que le dernier terme de $\bar{\mathbf{l}}$ est tel que :

$$\begin{aligned} \bar{l}_4 &= 0.37 * 0.03 + 0.16 * 0.11 + 0.11 * 0.25 + 0.36 * 0.44 \\ &\approx 0.22 \end{aligned}$$

Formulations matricielles

- Nous pouvons formuler les quantités précédentes en termes matriciels.
- Notons d'abord par \mathbf{D}_P et \mathbf{D}_Q les matrices diagonales de taille $(p \times p)$ et $(q \times q)$ dont les éléments diagonaux sont les marges de P et Q respectivement

$$\mathbf{D}_P = \begin{pmatrix} f_{1.} & & \\ & \ddots & \\ & & f_{p.} \end{pmatrix} ; \quad \mathbf{D}_Q = \begin{pmatrix} f_{.1} & & \\ & \ddots & \\ & & f_{.q} \end{pmatrix}$$

- Remarquons par ailleurs que :

$$\mathbf{D}_P^{-1} = \begin{pmatrix} \frac{1}{f_{1.}} & & \\ & \ddots & \\ & & \frac{1}{f_{p.}} \end{pmatrix} ; \quad \mathbf{D}_Q^{-1} = \begin{pmatrix} \frac{1}{f_{.1}} & & \\ & \ddots & \\ & & \frac{1}{f_{.q}} \end{pmatrix}$$

Ajustement du nuage des profils lignes

- Dans le cadre de \mathbf{NL} , on cherche à représenter dans un espace de dimension réduite la proximité des modalités i de P .
- Chaque modalité i est représentée par un profil ligne \mathbf{l}_i qui est considéré comme un point de \mathbb{R}^q .
- L'information contenue dans \mathbf{NL} est mesurée par l'inertie c'est à dire la distance moyenne des profils lignes par rapport au barycentre de \mathbf{NL} (comme en ACP mais en utilisant la distance $d_{\chi^2}^2$ -cf slide 247-).

- ▷ L'objectif est donc de déterminer des **sous-espaces** de sorte à ce que la projection des \mathbf{l}_i dans ceux-ci **conservent au mieux l'inertie**. Autrement dit, on souhaite que l'image de \mathbf{NL} dans ce sous-espace soit **le moins déformé possible**.
- ▷ Comme en ACP, on cherche alors une suite de s ($s < q$) directions privilégiées appelés **axes factoriels** dans \mathbb{R}^q notées $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_s$ qui permettent de **maximiser l'inertie du nuage projeté**.

Formulations matricielles (suite)

- Nous avons alors l'expression suivante de \mathbf{L} , la matrice dont les lignes sont les profils lignes :

$$\mathbf{L} = \mathbf{D}_P^{-1} \mathbf{F}$$

On rappelle que \mathbf{D}_P^{-1} est la matrice diagonale dont les termes diagonaux sont $(1/f_{1.}, \dots, 1/f_{p.})$.

- La distance du χ^2 donne un poids à chaque dimension et dans \mathbb{R}^q la dimension j a un poids $1/f_{.j}$.
- Pondérer de façon non uniforme chaque dimension c'est utiliser une **métrique diagonale** qui, dans \mathbb{R}^q , est donnée par la matrice symétrique définie positive \mathbf{D}_Q^{-1} .
- La distance du χ^2 entre deux profils lignes \mathbf{l}_i et $\mathbf{l}_{i'}$ vaut :

$$d_{\chi^2}^2(\mathbf{l}_i, \mathbf{l}_{i'}) = \langle \mathbf{l}_i - \mathbf{l}_{i'}, \mathbf{l}_i - \mathbf{l}_{i'} \rangle_{\mathbf{D}_Q^{-1}} = (\mathbf{l}_i - \mathbf{l}_{i'})^\top \mathbf{D}_Q^{-1} (\mathbf{l}_i - \mathbf{l}_{i'})$$

Détermination des axes factoriels

- Nous avons vu (cas général) en slide 289 qu'avec une métrique \mathbf{M} et une matrice de poids \mathbf{N} , l'espace réduit des lignes d'une matrice de données \mathbf{X} était donné par les vecteurs propres de $\mathbf{X}^\top \mathbf{N} \mathbf{X} \mathbf{M}$.
- En AFC pour les profils lignes la métrique est \mathbf{D}_Q^{-1} , la matrice de poids est \mathbf{D}_P et la matrice de données est $\mathbf{L} = \mathbf{D}_P^{-1} \mathbf{F}$.
- On montre donc que les vecteurs $\mathbf{u}_1, \dots, \mathbf{u}_s$ peuvent être obtenus en **diagonalisant la matrice \mathbf{S}** de taille $(q \times q)$ définie comme suit :

$$\mathbf{S} = \mathbf{F}^\top \mathbf{D}_P^{-1} \mathbf{F} \mathbf{D}_Q^{-1}$$

- De façon explicite, nous avons le terme général de \mathbf{S} , $\forall j, j'$:

$$s_{jj'} = \sum_{i=1}^p \frac{f_{ij} f_{ij'}}{f_{i.} f_{.j'}}$$

- ▷ Pour tout $m = 1, \dots, s$, \mathbf{u}_m est le vecteur propre associé à λ_m la m -ème plus grande valeur propre de \mathbf{S} .

Coordonnées des profils lignes sur les axes factoriels

- Une fois trouvé l'axe factoriel \mathbf{u}_m , on projette les profils lignes sur celui-ci et on obtient les coordonnées (ou composantes) factorielles (ou facteurs) des profils lignes dans l'espace engendré par \mathbf{u}_m .
- Attention ! La projection orthogonale est au sens de la métrique utilisée dans \mathbb{R}^q càd \mathbf{D}_Q^{-1} .
- Nous noterons par \mathbf{f}^m le vecteur de taille p comprenant les coordonnées des p profils lignes sur l'axe \mathbf{u}_m . On a

$$\mathbf{f}^m = \underbrace{\mathbf{D}_P^{-1} \mathbf{F} \mathbf{D}_Q^{-1}}_{\mathbf{L}} \mathbf{u}_m$$

- De façon explicite, on a pour tout profil ligne \mathbf{l}_i :

$$f_i^m = \langle \mathbf{l}_i, \mathbf{u}_m \rangle_{\mathbf{D}_Q^{-1}} = \mathbf{l}_i^\top \mathbf{D}_Q^{-1} \mathbf{u}_m = \sum_{j=1}^q \frac{f_{ij}}{f_{i.}} \frac{1}{f_{.j}} u_{m,j}$$

Nuage des profils colonnes NC

- Tout ce qui a été introduit précédemment pour les profils lignes est valable pour les profils colonnes. En fait, lignes et colonnes de \mathbf{F} jouent un rôle symétrique contrairement à l'ACP (où on a clairement distinguer les distances entre individus et celles entre variables).
- Les q profils colonnes sont représentés par q vecteurs $\mathbf{c}^1, \dots, \mathbf{c}^q$ au sein d'une matrice de données \mathbf{C} de taille $(p \times q)$:

$$\mathbf{C} = \begin{pmatrix} \mathbf{c}^1 & \dots & \mathbf{c}^j & \dots & \mathbf{c}^q \\ \frac{f_{11}}{f_{.1}} & & \frac{f_{1j}}{f_{.j}} & & \frac{f_{1q}}{f_{.q}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{i1}}{f_{.1}} & & \frac{f_{ij}}{f_{.j}} & & \frac{f_{iq}}{f_{.q}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{p1}}{f_{.1}} & & \frac{f_{pj}}{f_{.j}} & & \frac{f_{pq}}{f_{.q}} \\ 1 & \dots & 1 & \dots & 1 \end{pmatrix}$$

A propos du centrage de NL

- On peut montrer qu'il n'est pas nécessaire de centrer NL avant de procéder à l'AFC.
- En effet, l'AFC du nuage non centré et celle du nuage centré conduisent au même résultat !
- Si l'on pratique l'AFC sur le nuage non centré le premier axe factoriel possède les propriétés suivantes :
 - Il relie l'origine du repère initial $\mathbf{0}$ au barycentre $\bar{\mathbf{l}}$.
 - Les projections des profils sur cet axe se retrouvent toutes en $\bar{\mathbf{l}}$.
 - Cet axe est orthogonal au sens de la métrique \mathbf{D}_q^{-1} à l'hyperplan contenant NL.
 - L'inertie projetée de NL sur cet axe vaut 1.
- Les autres axes factoriels sont ensuite identiques entre l'AFC de NL non centré et l'AFC de NL centré.
- Dans la suite, **nous indiquons les méthodes de détermination des axes factoriels du nuage non centré**. On gardera donc en mémoire que le premier axe est **trivial** et qu'on l'enlève en pratique.

Distance du χ^2 et barycentre dans NC

- Le vecteur \mathbf{c}^j est de composantes $(\frac{f_{1j}}{f_{.j}}, \dots, \frac{f_{ij}}{f_{.j}}, \dots, \frac{f_{pj}}{f_{.j}})$.
- NC est donc la représentation des $\mathbf{c}^1, \dots, \mathbf{c}^q$ dans \mathbb{R}^p .
- Les \mathbf{c}^j appartiennent à un hyperplan de \mathbb{R}^p .
- La distance du χ^2 entre profils colonnes est comme suit :

$$d_{\chi^2}^2(\mathbf{c}^j, \mathbf{c}^{j'}) = \sum_{i=1}^p \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2$$

- Les \mathbf{c}^j ont un poids imposé $f_{.j}$ et $\bar{\mathbf{c}}$, le barycentre de NC, est tel que, $\forall i = 1, \dots, p$:

$$\begin{aligned} \bar{c}_i &= \sum_{j=1}^q f_{.j} \frac{f_{ij}}{f_{.j}} \\ &= f_{i.} \end{aligned}$$

- ▷ Comme précédemment, le barycentre est égale à la marge mais des modalités de P dans ce cas.

Exemple

$$\mathbf{C} = \begin{array}{ccccc} & \text{brun} & \text{chatain} & \text{roux} & \text{blond} & \bar{c} \\ \left(\begin{array}{cccc} .63 & .42 & .37 & .06 \\ .14 & .19 & .20 & .08 \\ .05 & .10 & .20 & .13 \\ .19 & .29 & .24 & .74 \end{array} \right) & \begin{array}{c} .37 \\ .16 \\ .11 \\ .36 \end{array} \\ 1.0 & 1.0 & 1.0 & 1.0 & \end{array}$$

- On remarquera par exemple que le dernier terme de \bar{c} est tel que :

$$\begin{aligned} \bar{c}_4 &= 0.19 * 0.19 + 0.48 * 0.29 + 0.11 * 0.24 + 0.22 * 0.74 \\ &\approx 0.36 \end{aligned}$$

Ajustement du nuage des profils colonnes et axes factoriels

- Contrairement à l'ACP, en AFC, l'ajustement de \mathbf{NC} est totalement symétrique vis à vis de l'ajustement de \mathbf{NL} .
- Dans ce cas, chaque modalité j de Q est considérée comme un point de \mathbb{R}^p dont les composantes sont données par les profils colonnes \mathbf{c}^j .
- On note $\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^s$ les vecteurs de \mathbb{R}^p permettant de **maximiser l'inertie du nuage projeté**. Ces axes dits factoriels peuvent être obtenus en **diagonalisant la matrice \mathbf{T}** de taille $(p \times p)$ définie comme suit :

$$\mathbf{T} = \mathbf{F} \mathbf{D}_Q^{-1} \mathbf{F}^\top \mathbf{D}_P^{-1}$$

- Le terme général de \mathbf{T} est défini comme suit, $\forall i, i'$:

$$t_{ii'} = \sum_{j=1}^q \frac{f_{ij} f_{i'j}}{f_{.j} f_{.i'}}$$

- Pour tout $m = 1, \dots, s$, \mathbf{v}^m est le vecteur propre associé à λ_m la m -ème plus grande valeur propre de \mathbf{T} .

Formulations matricielles

- La matrice des profils colonnes est noté \mathbf{C} et vérifie :

$$\mathbf{C}^\top = \mathbf{D}_Q^{-1} \mathbf{F}^\top$$

où \mathbf{C}^\top est la transposée de \mathbf{C} . On rappelle que \mathbf{D}_Q^{-1} est la matrice diagonale dont les termes diagonaux sont $(1/f_{.1}, \dots, 1/f_{.q})$.

- La distance du χ^2 donne un poids à chaque dimension. Dans \mathbb{R}^p la dimension i a un poids $1/f_{.i}$.
- La **métrique diagonale** dans \mathbb{R}^p est donc \mathbf{D}_P^{-1} et la distance du χ^2 entre deux profils colonnes \mathbf{c}^j et $\mathbf{c}^{j'}$ est donnée par :

$$d_{\chi^2}^2(\mathbf{c}^j, \mathbf{c}^{j'}) = \langle \mathbf{c}^j - \mathbf{c}^{j'}, \mathbf{c}^j - \mathbf{c}^{j'} \rangle_{\mathbf{D}_P^{-1}} = (\mathbf{c}^j - \mathbf{c}^{j'})^\top \mathbf{D}_P^{-1} (\mathbf{c}^j - \mathbf{c}^{j'})$$

Coordonnées des profils colonnes sur les axes factoriels

- Une fois trouvé l'axe factoriel \mathbf{v}^m , on projette les profils colonnes sur celui-ci pour obtenir les coordonnées factorielles des \mathbf{c}^j sur cet axe.
- La projection orthogonale est au sens de la métrique utilisée dans \mathbb{R}^p c'à d \mathbf{D}_P^{-1} .
- Nous noterons par \mathbf{g}_m le vecteur de taille q comprenant les coordonnées des q profils colonnes sur l'axe \mathbf{v}^m . On a

$$\mathbf{g}_m = \underbrace{\mathbf{D}_Q^{-1} \mathbf{F}^\top}_{\mathbf{C}^\top} \mathbf{D}_P^{-1} \mathbf{v}^m$$

- De façon explicite, on a pour tout profil colonne \mathbf{c}^j :

$$g_{m,j} = \langle \mathbf{c}^j, \mathbf{v}^m \rangle_{\mathbf{D}_P^{-1}} = (\mathbf{c}^j)^\top \mathbf{D}_P^{-1} \mathbf{v}^m = \sum_{i=1}^p \frac{f_{ij}}{f_{.j}} \frac{1}{f_{.i}} v_i^m$$

A propos du centrage de NC

- Comme pour NL, il n'est pas nécessaire de centrer NC avant de procéder à l'AFC. Nous avons exactement les mêmes propriétés citées précédemment.
- Ainsi, dans ce qui suit, nous indiquons aussi les méthodes de détermination des axes factoriels du nuage NC non centré et on gardera également en mémoire que le premier axe est celui rejoignant l'origine et le barycentre de NC et qu'on le retire en pratique.

Dualité en AFC - Relations barycentriques

- Rappelons que $f_i^m = \sum_{j=1}^q \frac{f_{ij}}{f_{i.}} \frac{1}{f_{i.}} u_{m,j}$ (cf slide 317). Précédemment nous avons établi $\mathbf{g}_m = \sqrt{\lambda_m} \mathbf{D}_Q^{-1} \mathbf{u}_m$ ce qui explicitement s'écrit $g_{m,j} = \sqrt{\lambda_m} \frac{1}{f_{i.}} u_{m,j}$. Nous en déduisons la relation suivante :

$$f_i^m = \frac{1}{\sqrt{\lambda_m}} \sum_{j=1}^q \frac{f_{ij}}{f_{i.}} g_{m,j} = \frac{1}{\sqrt{\lambda_m}} \mathbf{l}_i^\top \mathbf{g}_m$$

- Un même développement conduit à la relation suivante :

$$g_{m,j} = \frac{1}{\sqrt{\lambda_m}} \sum_{i=1}^p \frac{f_{ij}}{f_{i.}} f_i^m = \frac{1}{\sqrt{\lambda_m}} (\mathbf{c}^j)^\top \mathbf{f}^m$$

- Ces relations entre les coordonnées factorielles des profils des deux nuages sont appelées **relation de transition ou barycentriques**.

Dualité en AFC - Relations entre espaces réduits

- La relation de dualité entre l'ajustement de NL et celui de NC est plus forte en AFC qu'en ACP.
- Les axes factoriels \mathbf{u}_m et \mathbf{v}^m ont même valeur propre λ_m .
- Il existe les relations de transition entre les deux espaces réduits :

$$\mathbf{u}_m = \frac{1}{\sqrt{\lambda_m}} \mathbf{F}^\top \mathbf{D}_P^{-1} \mathbf{v}^m \text{ et } \mathbf{v}^m = \frac{1}{\sqrt{\lambda_m}} \mathbf{F} \mathbf{D}_Q^{-1} \mathbf{u}_m$$

On remarquera l'usage des métriques associées aux deux espaces.

- Rappelons que $\mathbf{f}^m = \mathbf{D}_P^{-1} \mathbf{F} \mathbf{D}_Q^{-1} \mathbf{u}_m$ et que $\mathbf{g}_m = \mathbf{D}_Q^{-1} \mathbf{F}^\top \mathbf{D}_P^{-1} \mathbf{v}^m$. Nous avons alors la relation de dualité suivante entre d'une part les composantes des profils d'un nuage et les axes factoriels de l'espace réduit dans lequel est représenté l'autre nuage :

$$\mathbf{f}^m = \sqrt{\lambda_m} \mathbf{D}_P^{-1} \mathbf{v}^m \text{ et } \mathbf{g}_m = \sqrt{\lambda_m} \mathbf{D}_Q^{-1} \mathbf{u}_m$$

Dualité en AFC - Relations barycentriques (suite)

- Interprétation de

$$f_i^m = \frac{1}{\sqrt{\lambda_m}} \sum_{j=1}^q \frac{f_{ij}}{f_{i.}} g_{m,j}$$

- ▶ Au facteur $1/\sqrt{\lambda_m}$ près, la coordonnée factorielle sur l'axe \mathbf{u}_m d'une modalité i de NL (f_i^m) est le **barycentre** des coordonnées factorielles sur l'axe \mathbf{v}^m des modalités j de Q ($\{g_{m,j}\}_j$) avec comme système de pondération le profil ligne de i ($\{f_{ij}/f_{i.}\}_j$) !
- ▶ De la même manière, au facteur $1/\sqrt{\lambda_m}$ près, la coordonnée factorielle sur l'axe \mathbf{v}^m d'une modalité j de NC ($g_{m,j}$) est le **barycentre** des coordonnées factorielles sur l'axe \mathbf{u}_m des modalités i de P ($\{f_i^m\}_i$) avec comme système de pondération le profil ligne de j ($\{f_{ij}/f_{i.}\}_i$) !
- ▶ Etant donnés les facteurs d'un nuage de points sur un axe, on peut donc calculer les facteurs de l'autre nuage de points sur l'axe associé à la même valeur propre.

Dualité en AFC - Représentation simultanée

- ▶ En raison des relations de dualité en AFC, il est naturellement possible de faire une représentation simultanée des deux nuages \mathbb{NL} et \mathbb{NC} au sein d'un même plan réduit (contrairement à l'ACP).
- En AFC, à l'aide des liaisons barycentriques on peut positionner et interpréter un profil d'un nuage en fonction de l'ensemble des profils de l'autre nuage (même si les deux nuages appartiennent à deux espaces réduits distincts).
- Par exemple, si l'on est dans un plan factoriel de \mathbb{NL} on pourra donc représenter dans celui-ci les modalités de \mathbb{NC} selon la formule barycentrique suivante :

$$g_{m,j} = \frac{1}{\sqrt{\lambda_m}} \sum_{i=1}^p \frac{f_{ij}}{f_{.j}} f_i^m$$

Liens avec la statistique du χ^2

- En cas d'indépendance, les profils sont identiques aux marges (cf slide 306). Par ailleurs, les composantes d'un barycentre d'un nuage sont les marges de l'autre variable (cf slide 311). Indépendance et distance du χ^2 sont donc liées.
- L'inertie d'un nuage étant égale à la distance (du χ^2) moyenne entre les profils et leur barycentre, nous pouvons faire le lien entre ce concept et la statistique du χ^2 :

$$\begin{aligned} \text{int}(\mathbb{NL}) &= \sum_{i=1}^p f_i d_{\chi^2}^2(\mathbf{l}_i, \bar{\mathbf{l}}) \\ &= \sum_{i=1}^p f_i \sum_{j=1}^q \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_i} - f_{.j} \right)^2 \\ &= \sum_{i=1}^p \sum_{j=1}^q \frac{(f_{ij} - f_i f_{.j})^2}{f_i f_{.j}} \end{aligned}$$

Propriétés des facteurs

- Comme précédemment nous considérons que les profils sont affectés de poids égaux à leurs fréquences relatives.
- Les **coordonnées factorielles sont alors centrées** :

$$\sum_{i=1}^p f_i f_i^m = 0 \text{ et } \sum_{j=1}^q f_j g_{m,j} = 0$$

- ▶ L'**inertie** (ou variance) des nuages de points \mathbb{NL} et \mathbb{NC} sur \mathbf{u}_m et \mathbf{v}^m respectivement est égale à λ_m :

$$\sum_{i=1}^p f_i (f_i^m)^2 = \lambda_m \text{ et } \sum_{j=1}^q f_j (g_{m,j})^2 = \lambda_m$$

- Les nuages étant non centrés, la première valeur propre λ_1 vaut 1 et permet de retrouver les barycentres des nuages (cf slide 318). Cela implique que **toutes les valeurs propres sont inférieures à 1**.

Liens avec la statistique du χ^2 (suite)

- L'**inertie totale** de \mathbb{NL} est égale à une grandeur statistique appelée le ϕ^2 ("phi 2") qui permet de mesurer l'intensité de l'association entre deux variables qualitatives en statistique des contingences. Or $n\phi^2 = \chi^2$ (n étant la taille de l'échantillon).
- Nous avons donc :

$$\text{int}(\mathbb{NL}) = \chi^2 / n$$

- ▶ On montre de la même manière que $\text{int}(\mathbb{NC}) = \chi^2 / n$ et donc :

$$\text{int}(\mathbb{NC}) = \text{int}(\mathbb{NL}) = \phi^2$$

- ▶ Ainsi, en AFC, l'inertie s'interprète également du point de vue statistique puisqu'elle est directement liée à la statistique du χ^2 .
- Nous avons aussi la propriété suivante (l'axe trivial ayant été enlevé) :

$$\phi^2 = \sum_{m=1}^{\min(p,q)-1} \lambda_m$$

Liens avec la statistique du χ^2 (suite)

- Rappelons également que la statistique du χ^2 mesurée sur un tableau de contingence croisant deux variables P et Q avec respectivement p et q modalités suit asymptotiquement une loi du χ^2 à $(p-1)(q-1)$ degrés de libertés.
- En statistique inférentielle, on rejette l'hypothèse nulle d'indépendance entre P et Q si la mesure du χ^2 dépasse une valeur critique définissant une région de rejet étant donné un risque de premier espèce α (de 5 ou 10%).
- Une statistique de χ^2 forte cela veut dire, dans notre cas, une inertie forte et donc beaucoup de dispersion au sein des profils.
- ▶ Lorsqu'un test d'indépendance du χ^2 sera rejeté, il est donc très utile de compléter l'analyse par une AFC qui rend compte des différentes tendances de la dispersion au sein des modalités.

Distance du χ^2 et équivalence distributionnelle (suite)

- La propriété d'équivalence distributionnelle apporte un intérêt pratique dans de nombreux cas.
- Prenons l'exemple $P=\text{CSP}$ et $Q=\text{Départements}$:
 - ▶ Considérons les deux modalités de P suivantes, "ouvrier qualifié" et "ouvrier non qualifié".
 - ▶ Supposons que les profils lignes de ces deux modalités soient très similaires (càd leurs répartition au sein des différents départements sont quasi-identiques).
 - ▶ Il est alors possible d'agréger ces deux modalités en une seule, "ouvrier", sans changer drastiquement les résultats de l'AFC.
- La propriété permet donc en pratique de regrouper des modalités d'une variable de profils voisins afin de définir une nomenclature (càd liste de modalités de la variable) plus ou moins détaillée.

Distance du χ^2 et équivalence distributionnelle

- En AFC, on analyse une table de contingence et on utilise la distance du χ^2 entre profils. Ceci permet d'obtenir des propriétés remarquables de cette méthode notamment l'**équivalence distributionnelle**.
- Selon cette propriété, **on peut agréger deux modalités d'une même variable (P ou Q) ayant des profils identiques** en une nouvelle modalité affectée de la somme de leurs masses (càd leurs fréquences relatives), **sans rien changer**, ni aux distances entre les profils de cette variable, ni aux distances entre les profils de l'autre variable.
- Par exemple, considérons deux profils lignes i et i' identiques mais de fréquences relatives $f_{i.}$ et $f_{i'.$ pouvant être différentes :
 - ▶ \mathbf{l}_i et $\mathbf{l}_{i'}$ étant confondus dans \mathbb{R}^q , l'agrégation $(\mathbf{l}_i + \mathbf{l}_{i'})/(f_{i.} + f_{i'.$) est également confondu en ces deux points et elle n'altère donc en rien le nuage NL.
 - ▶ Plus étonnamment remplacer \mathbf{l}_i et $\mathbf{l}_{i'}$ par $(\mathbf{l}_i + \mathbf{l}_{i'})/(f_{i.} + f_{i'.$), ne change pas non plus les distances de χ^2 entre les profils colonnes !

Valeurs propres et inertie sur les axes factoriels

- Comme en ACP, les valeurs propres indiquent l'**inertie du nuage projeté** sur les axes :

$$\text{int}_{\mathbf{u}_m}(\text{NL}) = \text{int}_{\mathbf{v}^m}(\text{NC}) = \lambda_m$$

- ▶ On définit alors le **taux d'inertie** d'un axe \mathbf{u}_m par :

$$\frac{\lambda_m}{\phi^2}$$

- L'inertie d'un axe mesure l'association entre des modalités des deux variables qu'il met en évidence. Nous avons vu qu'elle ne peut pas dépasser 1.
- Si elle est proche de 1, l'axe permet de séparer en deux sous-groupes l'ensemble des modalités d'une variable.
- Si toutes les vp sont proches de 1, alors chaque modalité d'une variable est en correspondance presque exclusive avec une modalité de l'autre variable.

Quelques formes caractéristiques de nuage

- Ainsi, les valeurs propres permettent d'avoir une idée sur la forme du nuage projeté. Il existe plusieurs cas spécifiques :
 - ▶ Le nuage est **scindé en deux sous-groupes** la matrice **F** peut être réorganisée en diagonale par blocs en ordonnant les coordonnées des lignes et des colonnes vis à vis de leurs premiers facteurs respectifs. Dans ce cas, la valeur propre vaut 1 (outre la valeur propre triviale).
 - ▶ Le nuage est **scindé en k sous-groupes** la matrice **F** peut être aussi réorganisée en diagonale par blocs en permutant lignes et colonnes. Dans ce cas, k valeurs propres valent 1 (outre la valeur propre triviale).
 - ▶ L'**effet Guttman** le nuage de points a une **forme parabolique**. Le tableau **F** peut être réorganisée suivant une diagonale (on parle de sériation) qui est typique d'un effet Guttman. Dans ce cas, le premier axe oppose des modalités opposées tandis que le deuxième permet de séparer les valeurs intermédiaires des valeurs extrêmes. Exemple dans une étude de cas en sociologie politique, on peut retrouver sur le premier axe une opposition entre droite et gauche tandis que sur le deuxième axe oppose les modérés et les extrêmes.

Aide à l'interprétation à partir des données actives

- Les indices d'interprétation définis en ACP reposent sur des concepts géométriques et sont aussi valables pour l'AFC. Il s'agit de mesures de **qualité** et de **contribution** (cf également slide 270).
- Il faut cependant garder à l'esprit, qu'en AFC, les éléments ont des poids imposés qui interviennent dans les mesures de contribution.
- **Qualité de la représentation du profil ligne l_i sur l'axe u_m :**

$$qlt_{u_m}(l_i) = \frac{(f_i^m)^2}{\sum_{j=1}^q \frac{1}{f_j} (f_{ij} - f_j)^2}$$

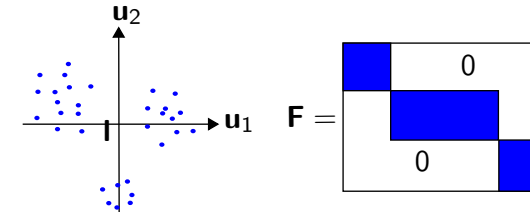
- **Qualité de la représentation du profil colonne c^j sur l'axe v^m :**

$$qlt_{v^m}(c^j) = \frac{(g_{m,j})^2}{\sum_{i=1}^p \frac{1}{f_i} (f_{ij} - f_i)^2}$$

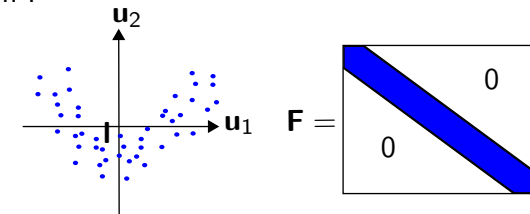
- Pour analyser les points, on privilégiera ceux dont la qualité est élevée.

Illustrations

- Nuage scindé en 3 groupes :



- Effet Guttman :



Aide à l'interprétation à partir des données actives (suite)

- La **contribution du profil ligne l_i à l'axe u_m :**

$$ctr_{u_m}(l_i) = f_i \frac{(f_i^m)^2}{\lambda_m}$$

- La **contribution du profil colonne c^j à l'axe v^m :**

$$ctr_{v^m}(c^j) = f_j \frac{(g_{m,j})^2}{\lambda_m}$$

- Nous remarquerons que comme $\sum_i f_i (f_i^m)^2 = \lambda_m$, nous avons donc :

$$\sum_{i=1}^p ctr_{u_m}(l_i) = 1$$

Il en va de même pour les profils colonnes c^j de NC.

- Pour interpréter un axe, on privilégiera les profils dont les contributions sont élevées.

Aide à l'interprétation à partir de données supplémentaires

- Comme en ACP, il est souvent utile de projeter des éléments supplémentaires dans les espaces réduits afin d'affiner les interprétations (cf également slide 274).
- Supposons que nous avons q_+ colonnes supplémentaires portant à $q + q_+$ le nombre total de modalités de Q .
- L'idée est de projeter ces nouveaux q_+ profils colonnes dans l'espace réduit dans lequel est déjà représenté le nuage actif \mathbb{NC} .
- Soit alors n_{ij}^+ la i ème coordonnée de la j ème colonne supplémentaire dans le tableau de contingence. Son profil \mathbf{c}_j^+ est donné par, $\forall i = 1, \dots, p$:

$$c_{ij}^+ = \frac{n_{ij}^+}{n_{+j}^+}$$

où $n_{+j}^+ = \sum_{i=1}^p n_{ij}^+$ et on rappelle que de façon générale $\frac{f_{ij}}{f_{+j}} = \frac{n_{ij}}{n_{+j}}$.

Aide à l'interprétation à partir de données supplémentaires (suite)

- Remarque : les éléments supplémentaires n'intervenant pas dans la construction des axes factoriels, leurs contributions sont nulles. En revanche, leur qualité vis à vis d'un axe étant exactement le calcul d'un cosinus (au carré), cet indicateur reste valide dans ce cas pour l'aide à l'interprétation de ces éléments nouveaux.

Aide à l'interprétation à partir de données supplémentaires (suite)

- On projette \mathbf{c}_j^+ sur \mathbf{v}^m en utilisant la formule de transition vue au slide 327 :

$$g_{m,j}^+ = \frac{1}{\sqrt{\lambda_m}} \sum_{i=1}^p \frac{n_{ij}^+}{n_{+j}^+} f_i^m$$

Ici, les f_i^m sont les facteurs des profils lignes actifs sur l'axe \mathbf{u}_m .

- Si on suppose que nous disposons de lignes supplémentaires au sein des modalités de P , alors on calcule le profil ligne de ces nouveaux éléments pour obtenir les vecteurs $\mathbf{l}_{+,i}$ et on projette ces derniers sur \mathbf{u}_m par le biais de la formule suivante :

$$f_i^{+,m} = \frac{1}{\sqrt{\lambda_m}} \sum_{j=1}^q \frac{n_{+,ij}}{n_{+,i}} g_{m,j}$$

Rappel du Sommaire

3 Méthodes de réduction de dimension

- Analyse générale : la décomposition en valeurs singulières (SVD)
- Analyse en composantes principales (ACP)
- Analyse générale : extension à des métriques et poids quelconques
- Analyse factorielle des correspondances (AFC)
- Analyse (factorielle) des correspondances multiples (ACM)

Données traitées par l'ACM et notations

- L'**Analyse des Correspondances Multiples (ACM)** permet de généraliser l'AFC dans le cas où nous étudions plus de 2 variables qualitatives.
- L'ACM permet d'étudier une population de n **individus décrits par un ensemble de p variables qualitatives nominales**.
- L'application la plus courante de l'ACM est l'analyse des réponses à une enquête où chaque question est une variable et chaque réponse proposée à chaque variable est une modalité (questionnaire à choix fermé).
- Les données brutes se présentent sous la forme d'une matrice \mathbf{T} de taille $(n \times p)$ comme en ACP sauf que les variables sont à valeurs discrètes.

Exemple

- 3 variables qualitatives $\mathbf{t}^1 = \text{genre}$, $\mathbf{t}^2 = \text{tranche d'âge}$, $\mathbf{t}^3 = \text{statut marital}$ avec $\mathbb{T}^1 = \{H, F\}$, $\mathbb{T}^2 = \{<25, 25-44, 45-64, >65\}$ et $\mathbb{T}^3 = \{\text{célibataire, marié(e), divorcé(e)}\}$.
- Souvent les modalités sont codées numériquement $\mathbb{T}^1 = \{1, 2\}$, $\mathbb{T}^2 = \{1, 2, 3, 4\}$ et $\mathbb{T}^3 = \{1, 2, 3\}$.
- Attention ! ces codages sont symboliques et aucunement numériques donc faire une moyenne par exemple n'a pas de sens.

$$\mathbf{T} = \begin{matrix} & \mathbf{t}^1 & \mathbf{t}^2 & \mathbf{t}^3 \\ \mathbf{t}_1 & \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \\ \mathbf{t}_2 & \begin{pmatrix} 2 & 2 & 2 \end{pmatrix} \\ \mathbf{t}_3 & \begin{pmatrix} 1 & 3 & 3 \end{pmatrix} \\ \mathbf{t}_4 & \begin{pmatrix} 2 & 4 & 2 \end{pmatrix} \\ \mathbf{t}_5 & \begin{pmatrix} 1 & 2 & 2 \end{pmatrix} \\ \mathbf{t}_6 & \begin{pmatrix} 2 & 1 & 1 \end{pmatrix} \\ \mathbf{t}_7 & \begin{pmatrix} 2 & 2 & 3 \end{pmatrix} \end{matrix}$$

Données traitées par l'ACM et notations (suite)

- On suppose p variables $\{\mathbf{t}^1, \dots, \mathbf{t}^k, \dots, \mathbf{t}^p\}$.
- Chaque variable \mathbf{t}^k prend ses valeurs dans un ensemble discret \mathbb{T}^k (ensemble des modalités) de dimension p_k (nombre de modalités).
- On suppose n individus $\{\mathbf{t}_1, \dots, \mathbf{t}_i, \dots, \mathbf{t}_n\}$ où $\mathbf{t}_i \in \prod_{k=1}^p \mathbb{T}^k$.
- Ces données se retrouvent dans la matrice \mathbf{T} de taille $(n \times p)$:

$$\mathbf{T} = \begin{matrix} & \mathbf{t}^1 & \dots & \mathbf{t}^k & \dots & \mathbf{t}^p \\ \mathbf{t}_1 & \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{pmatrix} \\ \vdots & & & & & \\ \mathbf{t}_i & \begin{pmatrix} \dots & \dots & t_{ik} & \dots & \dots \end{pmatrix} \\ \vdots & & & & & \\ \mathbf{t}_n & \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{pmatrix} \end{matrix}$$

où $\forall i = 1, \dots, n; \forall k = 1, \dots, p : t_{ik} \in \mathbb{T}^k$.

Codage par tableau disjonctif complet

- En ACM, on utilise une autre représentation symbolique de ces données et qui consiste en la représentation explicite des modalités de chaque variable comme dimension de l'espace de description.
- Tout d'abord, remarquons que chaque variable j peut être représentée par une matrice binaire \mathbf{Z}^j de taille $(n \times p_j)$ et de terme général :

$$z_{ik}^j = \begin{cases} 1 & \text{si l'individu } i \text{ a la modalité } k \text{ de la variable } j \\ 0 & \text{sinon} \end{cases}$$

- Exemple :

$$\mathbf{t}^1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 1 \\ 2 \\ 2 \end{pmatrix} \rightarrow \mathbf{Z}^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

Codage par tableau disjonctif complet (suite)

- Ensuite, remarquons que la matrice \mathbf{Z}^j peut aussi être vue telle une juxtaposition de p_j vecteurs binaires \mathbf{z}_k^j avec $k = 1, \dots, p_j$.
- Exemple :

$$\mathbf{Z}^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \rightarrow \mathbf{z}_1^1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \text{ et } \mathbf{z}_2^1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

- Nous avons donc pour chaque variable j :

$$\mathbf{Z}^j = (\mathbf{z}_1^j \quad \dots \quad \mathbf{z}_k^j \quad \dots \quad \mathbf{z}_{p_j}^j)$$

Exemple

$$\underbrace{\begin{matrix} & \mathbf{t}^1 & \mathbf{t}^2 & \mathbf{t}^3 \\ \mathbf{t}_1 & \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \\ \mathbf{t}_2 & \begin{pmatrix} 2 & 2 & 2 \end{pmatrix} \\ \mathbf{t}_3 & \begin{pmatrix} 1 & 3 & 3 \end{pmatrix} \\ \mathbf{t}_4 & \begin{pmatrix} 2 & 4 & 2 \end{pmatrix} \\ \mathbf{t}_5 & \begin{pmatrix} 1 & 2 & 2 \end{pmatrix} \\ \mathbf{t}_6 & \begin{pmatrix} 2 & 1 & 1 \end{pmatrix} \\ \mathbf{t}_6 & \begin{pmatrix} 2 & 2 & 3 \end{pmatrix} \end{matrix}}_{\mathbf{T}} \rightarrow \underbrace{\begin{matrix} & \mathbf{Z}^1 & & \mathbf{Z}^2 & & \mathbf{Z}^3 \\ & \mathbf{z}_1^1 & \mathbf{z}_2^1 & \mathbf{z}_1^2 & \mathbf{z}_2^2 & \mathbf{z}_3^2 & \mathbf{z}_4^2 & \mathbf{z}_1^3 & \mathbf{z}_2^3 & \mathbf{z}_3^3 \\ \mathbf{z}_1 & \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \\ \mathbf{z}_2 & \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \\ \mathbf{z}_3 & \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\ \mathbf{z}_4 & \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix} \\ \mathbf{z}_5 & \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \\ \mathbf{z}_5 & \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix} \\ \mathbf{z}_6 & \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}}_{\mathbf{Z}}$$

- Avec ce codage, un individu i est donc représenté par un vecteur binaire \mathbf{z}_i de taille q .
- Autrement dit les individus appartiennent à un espace binaire de dimension q que l'on notera $\{0, 1\}^q$.

Codage par tableau disjonctif complet (suite)

- Pour représenter l'ensemble des variables au sein d'une même matrice, nous pouvons de la même manière concaténer l'ensemble des vecteurs binaires \mathbf{z}^j (ou de façon équivalente l'ensemble des matrices \mathbf{Z}^j) en une seule.
- Nous définissons ainsi la matrice \mathbf{Z} que l'on nomme **tableau disjonctif complet** qui est de taille $(n \times q)$ où $q = \sum_{j=1}^p p_j$:

$$\begin{aligned} \mathbf{Z} &= (\mathbf{Z}^1 \quad \dots \quad \mathbf{Z}^p) \\ &= (\mathbf{z}_1^1 \quad \dots \quad \mathbf{z}_{p_1}^1 \quad \mathbf{z}_1^2 \quad \dots \quad \dots \quad \mathbf{z}_{p_p}^p) \end{aligned}$$

- Pour chaque variable j , chaque individu i ne peut prendre qu'une et une seule modalité. Donc chaque ligne i de \mathbf{Z}^j est telle que :

$$\sum_{k=1}^{p_j} z_{ik}^j = 1$$

- Chaque modalité k de chaque variable j est représentée par le vecteur binaire de présence/absence \mathbf{z}_k^j . Nous avons donc :

$$\sum_{i=1}^n z_{ik}^j = \text{Fréquence de la modalité } k \text{ de la variable } j$$

- Pour chaque variable j , la somme de tous les termes de \mathbf{Z}^j vérifie :

$$\sum_{i=1}^n \sum_{k=1}^q z_{ik}^j = n$$

Codage par tableau disjonctif complet (suite)

- Les propriétés précédentes des matrices \mathbf{Z}^j , impliquent les propriétés suivantes pour le tableau $\mathbf{Z} = (\mathbf{Z}^1 \dots \mathbf{Z}^p)$.
- Les marges des lignes de \mathbf{Z} sont telles que, $\forall i = 1, \dots, n$:

$$\sum_{k=1}^q z_{ik} = z_{i.} = p$$

- Les marges des colonnes de \mathbf{Z} sont telles que, $\forall k = 1, \dots, q$:

$$\sum_{i=1}^n z_{ik} = z_{.k} = \text{Fréquence de la modalité } k$$

- La somme de tous les termes de \mathbf{Z} vérifie :

$$\sum_{i=1}^n \sum_{k=1}^q z_{ik} = z_{..} = np$$

Objectifs de l'ACM

- L'objectif général est identique aux méthodes précédentes et la matrice de données est ici \mathbf{Z} .
- On a **deux types de nuage de points** :
 - NO le **nuage des individus** : chaque ligne de \mathbf{Z} est vu tel un vecteur appartenant à un espace à q dimensions.
 - NM le **nuage des modalités** : chaque colonne de \mathbf{Z} est vu tel un vecteur appartenant à un espace à n dimensions.
- On cherche des sous-espaces de dimension réduite qui permettent de conserver au mieux l'inertie des nuages de points et qui contribuent à déterminer des typologies des éléments de chaque nuage.
- L'inertie est basée sur la notion de distance entre points mais comme **en AFC on utilise la distance du χ^2** .
- ▷ L'ACM est en fait une **AFC du tableau disjonctif complet \mathbf{Z}** . Mais, \mathbf{Z} possédant des propriétés différentes des tables de contingence, l'ACM présentent donc des spécificités par rapport à l'AFC.

Exemple

$$\mathbf{Z} = \begin{array}{c} \begin{matrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \\ \mathbf{z}_4 \\ \mathbf{z}_5 \\ \mathbf{z}_6 \end{matrix} \begin{pmatrix} \begin{matrix} z_1^1 & z_2^1 & z_3^1 & z_4^1 & z_5^1 & z_6^1 \end{matrix} \\ \begin{matrix} z_1^2 & z_2^2 & z_3^2 & z_4^2 & z_5^2 & z_6^2 \end{matrix} \\ \begin{matrix} z_1^3 & z_2^3 & z_3^3 & z_4^3 & z_5^3 & z_6^3 \end{matrix} \end{pmatrix} \begin{matrix} z_{1.} = 3 \\ z_{2.} = 3 \\ z_{3.} = 3 \\ z_{4.} = 3 \\ z_{5.} = 3 \\ z_{6.} = 3 \end{matrix} \end{array}$$

$$\begin{array}{ccccccccc} z_{.1} & z_{.2} & z_{.3} & z_{.4} & z_{.5} & z_{.6} & z_{.7} & z_{.8} & z_{.9} \\ \parallel & \parallel & \parallel & \parallel & \parallel & \parallel & \parallel & \parallel & \parallel \\ 3 & 4 & 2 & 3 & 1 & 1 & 2 & 3 & 2 \end{array}$$

- Dans la suite, les modalités ne sont plus nécessairement rattachées à leur variable d'origine mais représentent les dimensions de l'espace.

ACM et AFC

- Le tableau \mathbf{Z} est en fait vu tel une "table de contingence croisant deux variables qualitatives" : les individus d'une part notés O et les modalités d'autre part notées M .
- Comme en AFC, on calcule la matrice des fréquences \mathbf{F} de taille $(n \times q)$ donné par :

$$\mathbf{F} = \frac{\mathbf{Z}}{np} \text{ de terme général } f_{ik} = \frac{z_{ik}}{np}$$

$$\bullet \text{ On a donc : } \mathbf{F} = \begin{array}{c} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_i \\ \vdots \\ \mathbf{f}_n \end{array} \begin{pmatrix} \mathbf{f}^1 & \dots & \mathbf{f}^k & \dots & \mathbf{f}^p \\ \vdots & & \vdots & & \vdots \\ \dots & \dots & \frac{z_{ik}}{np} & \dots & \dots \\ \vdots & & \vdots & & \vdots \end{pmatrix}$$

Exemple

$$\mathbf{F} = \begin{array}{c} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \\ \mathbf{f}_4 \\ \mathbf{f}_5 \\ \mathbf{f}_5 \\ \mathbf{f}_6 \end{array} \begin{array}{c} \mathbf{f}^1 \\ \mathbf{f}^2 \\ \mathbf{f}^1 \\ \mathbf{f}^2 \\ \mathbf{f}^3 \\ \mathbf{f}^4 \\ \mathbf{f}^1 \\ \mathbf{f}^2 \\ \mathbf{f}^3 \end{array} \begin{array}{c} \left(\begin{array}{cc|cc|cc|cc} \frac{1}{21} & 0 & \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} & 0 & 0 \\ 0 & \frac{1}{21} & 0 & \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} & 0 \\ \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} \\ 0 & \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} & 0 & \frac{1}{21} & 0 \\ \frac{1}{21} & 0 & 0 & \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} & 0 \\ 0 & \frac{1}{21} & 1 & 0 & 0 & 0 & \frac{1}{21} & 0 & 0 \\ 0 & \frac{1}{21} & 0 & \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} & 0 \end{array} \right) \end{array} \begin{array}{c} f_{1.} = \frac{1}{7} \\ f_{2.} = \frac{1}{7} \\ f_{3.} = \frac{1}{7} \\ f_{4.} = \frac{1}{7} \\ f_{5.} = \frac{1}{7} \\ f_{6.} = \frac{1}{7} \\ f_{7.} = \frac{1}{7} \end{array}$$

$$\begin{array}{c} f_{1.} \\ f_{2.} \\ f_{3.} \\ f_{4.} \\ f_{5.} \\ f_{6.} \\ f_{7.} \end{array} \begin{array}{c} f_{.1} \\ f_{.2} \\ f_{.3} \\ f_{.4} \\ f_{.5} \\ f_{.6} \\ f_{.7} \\ f_{.8} \\ f_{.9} \end{array} \begin{array}{c} \parallel \\ \parallel \\ \parallel \\ \parallel \\ \parallel \\ \parallel \\ \parallel \\ \parallel \\ \parallel \end{array} \begin{array}{c} \frac{3}{21} \\ \frac{4}{21} \\ \frac{2}{21} \\ \frac{3}{21} \\ \frac{1}{21} \\ \frac{1}{21} \\ \frac{2}{21} \\ \frac{3}{21} \\ \frac{2}{21} \end{array}$$

- Dans l'exemple $n = 7$, $p = 3$ et $q = 9$.

Proximité et nuage des individus

- Les points lignes sont dans ce cas obtenus par :

$$\mathbf{L} = \mathbf{D}_O^{-1} \mathbf{F} = \frac{\mathbf{Z}}{p}$$

- Les lignes de \mathbf{L} forment le nuage NO qui est l'ensemble des vecteurs $\{\mathbf{l}_1, \dots, \mathbf{l}_n\}$ appartenant à \mathbb{R}^q . On a :

$$\mathbf{l}_i = \left(\frac{z_{i1}}{p}, \dots, \frac{z_{iq}}{p} \right)$$

- Les marges des lignes de \mathbf{Z} étant constantes, cela revient à attribuer à chaque individu un poids uniforme $1/n$.
- Comme en AFC, la métrique est celle associée à la matrice \mathbf{D}_M^{-1} et la distance du χ^2 entre deux individus \mathbf{l}_i et $\mathbf{l}_{i'}$ s'écrit :

$$d_{\chi^2}^2(\mathbf{l}_i, \mathbf{l}_{i'}) = (\mathbf{l}_i - \mathbf{l}_{i'})^\top \mathbf{D}_M^{-1} (\mathbf{l}_i - \mathbf{l}_{i'}) = \sum_{k=1}^q \frac{np}{z_{.k}} \left(\frac{z_{ik}}{p} - \frac{z_{i'k}}{p} \right)^2$$

Matrices des poids et des métriques

- Puis, on détermine les matrices diagonales suivantes :

$$\mathbf{D}_O = \begin{pmatrix} f_{1.} & & \\ & \ddots & \\ & & f_{p.} \end{pmatrix} \text{ et } \mathbf{D}_M = \begin{pmatrix} f_{.1} & & \\ & \ddots & \\ & & f_{.q} \end{pmatrix}$$

- Les propriétés de \mathbf{Z} impliquent les relations suivantes :

$$\mathbf{D}_O = \begin{pmatrix} \frac{1}{n} & & \\ & \ddots & \\ & & \frac{1}{n} \end{pmatrix} \text{ et } \mathbf{D}_M = \begin{pmatrix} \frac{z_{.1}}{np} & & \\ & \ddots & \\ & & \frac{z_{.q}}{np} \end{pmatrix}$$

- ▷ On en déduit les matrices inverses suivantes :

$$\mathbf{D}_O^{-1} = \begin{pmatrix} n & & \\ & \ddots & \\ & & n \end{pmatrix} \text{ et } \mathbf{D}_M^{-1} = \begin{pmatrix} \frac{np}{z_{.1}} & & \\ & \ddots & \\ & & \frac{np}{z_{.q}} \end{pmatrix}$$

Ajustement du nuage des individus et détermination des axes factoriels

- L'objectif est donc de déterminer des **sous-espaces** de sorte à ce que la projection des \mathbf{l}_i dans ceux-ci **conservent au mieux l'inertie**.
- ▷ Comme précédemment, on cherche alors une suite de s ($s < q$) directions privilégiées appelés **axes factoriels** dans \mathbb{R}^q notées $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_s$ qui permettent de **maximiser l'inertie du nuage projeté**.
- Les résultats sur la SVD généralisée (cf slide 289) montrent que les vecteurs $\mathbf{u}_1, \dots, \mathbf{u}_s$ peuvent être obtenus en **diagonalisant la matrice \mathbf{S}** de taille $(q \times q)$ définie comme suit :

$$\mathbf{S} = \mathbf{F}^\top \mathbf{D}_O^{-1} \mathbf{F} \mathbf{D}_M^{-1} = \frac{1}{np^2} \mathbf{Z}^\top \mathbf{Z} \mathbf{D}_M^{-1}$$

- ▷ Pour tout $m = 1, \dots, s$, \mathbf{u}_m est le vecteur propre associé à λ_m la m -ème plus grande valeur propre de \mathbf{S} .

Coordonnées des individus sur les axes factoriels

- Une fois trouvé l'axe factoriel \mathbf{u}_m , on projette les individus sur celui-ci et on obtient les coordonnées (ou composantes) factorielles (ou facteurs) des individus dans l'espace engendré par \mathbf{u}_m .
- ▷ Attention ! La projection orthogonale est au sens de la métrique utilisée dans \mathbb{R}^q càd \mathbf{D}_M^{-1} .
- Nous noterons par \mathbf{f}^m le vecteur de taille n comprenant les coordonnées des n points lignes sur l'axe \mathbf{u}_m . On a

$$\mathbf{f}^m = \underbrace{\frac{1}{p} \mathbf{Z} \mathbf{D}_M^{-1} \mathbf{u}_m}_{\mathbf{L}}$$

- De façon explicite, on a pour tout point ligne \mathbf{l}_i :

$$f_i^m = \langle \mathbf{l}_i, \mathbf{u}_m \rangle_{\mathbf{D}_M^{-1}} = \mathbf{l}_i^\top \mathbf{D}_M^{-1} \mathbf{u}_m = \sum_{k=1}^q \frac{z_{ik}}{p} \frac{np}{z_{.k}} u_{m,k}$$

Ajustement du nuage des modalités et détermination des axes factoriels

- ▷ Comme précédemment, on cherche une suite de s ($s < q$) directions privilégiées appelés **axes factoriels** dans \mathbb{R}^n notées $\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^s$ qui permettent de **maximiser l'inertie du nuage projeté**.
- Les résultats sur la SVD généralisée (cf slide 289) montrent que les vecteurs $\mathbf{v}^1, \dots, \mathbf{v}^s$ peuvent être obtenus en **diagonalisant la matrice \mathbf{T}** de taille $(n \times n)$ définie comme suit :

$$\mathbf{T} = \mathbf{F} \mathbf{D}_M^{-1} \mathbf{F}^\top \mathbf{D}_O^{-1} = \frac{1}{np^2} \mathbf{Z} \mathbf{D}_M^{-1} \mathbf{Z}^\top$$

- ▷ Pour tout $m = 1, \dots, s$, \mathbf{v}^m est le vecteur propre associé à λ_m la m -ème plus grande valeur propre de \mathbf{S} .

Proximité et nuage des modalités

- ▷ Les développements précédents s'appliquent quasi-identiquement au deuxième nuage mais attention aux poids et métrique qui changent !
- Les points colonnes sont obtenus par :

$$\mathbf{c}^\top = \mathbf{D}_M^{-1} \mathbf{F}^\top = \frac{1}{np} \mathbf{D}_M^{-1} \mathbf{Z}^\top$$

- Les colonnes de \mathbf{C} forment le nuage NM qui est l'ensemble des vecteurs $\{\mathbf{c}^1, \dots, \mathbf{c}^q\}$ appartenant à \mathbb{R}^n . On a :

$$\mathbf{c}^k = \left(\frac{z_{1k}}{z_{.k}}, \dots, \frac{z_{nk}}{z_{.k}} \right)$$

- La matrice des poids des modalités est donnée par \mathbf{D}_M .
- La métrique est celle associée à la matrice $\mathbf{D}_O^{-1} = n\mathbf{I}_n$ et la distance du χ^2 entre deux modalités \mathbf{c}^k et $\mathbf{c}^{k'}$ s'écrit :

$$d_{\chi^2}^2(\mathbf{c}^i, \mathbf{c}^{k'}) = (\mathbf{c}^k - \mathbf{c}^{k'})^\top \mathbf{D}_O^{-1} (\mathbf{c}^k - \mathbf{c}^{k'}) = \sum_{i=1}^n n \left(\frac{z_{ik}}{z_{.k}} - \frac{z_{ik'}}{z_{.k'}} \right)^2$$

Coordonnées des modalités sur les axes factoriels

- Une fois trouvé l'axe factoriel \mathbf{v}^m , on projette les modalités sur celui-ci et on obtient les coordonnées (ou composantes) factorielles (ou facteurs) des modalités dans l'espace engendré par \mathbf{v}^m .
- ▷ Attention ! La projection orthogonale est au sens de la métrique utilisée dans \mathbb{R}^n càd $\mathbf{D}_O^{-1} = n\mathbf{I}_n$.
- Nous noterons par \mathbf{g}_m le vecteur de taille n comprenant les coordonnées des q points colonnes sur l'axe \mathbf{v}^m . On a :

$$\mathbf{g}_m = \underbrace{\mathbf{D}_M^{-1} \mathbf{F}^\top}_{\mathbf{c}^\top} \mathbf{D}_O^{-1} \mathbf{v}^m = \frac{1}{p} \mathbf{D}_M^{-1} \mathbf{Z}^\top \mathbf{v}^m$$

- De façon explicite, on a pour tout point colonne \mathbf{c}^k :

$$g_{m,k} = \langle \mathbf{c}^k, \mathbf{v}^m \rangle_{\mathbf{D}_O^{-1}} = (\mathbf{c}^k)^\top \mathbf{D}_O^{-1} \mathbf{v}^m = n \sum_{i=1}^n \frac{z_{ik}}{z_{.k}} v_i^m$$

A propos du centrage des nuages

- L'ACM étant en pratique une AFC, les remarques relatives au centrage des nuages de points faits précédemment aux slides 318 et 325 restent valides.
- Plus particulièrement, en ACM, le vecteur propre trivial associé à la valeur propre 1 est lié au barycentre qui est proportionnel au vecteur rempli de 1 (première bissectrice).
- Ainsi, dans la suite, **nous indiquons les méthodes de détermination des axes factoriels du nuage non centré**. On gardera donc en mémoire que le premier axe est **trivial** et qu'on l'enlève en pratique.

Dualité en ACM - Relations barycentriques (suite)

- Le terme général de $\mathbf{f}^m = \frac{1}{p\sqrt{\lambda_m}} \mathbf{Z} \mathbf{g}_m$ s'écrit :

$$f_i^m = \frac{1}{\sqrt{\lambda_m}} \sum_{k=1}^q \underbrace{\frac{z_{ik}}{z_{i.}}}_{\frac{1}{p}} g_{m,k}$$

- On peut réécrire f_i^m comme suit :

$$f_i^m = \frac{1}{\sqrt{\lambda_m}} \left(\frac{1}{p} \sum_{k: z_{ik}=1} g_{m,k} \right)$$

- ▷ Au facteur $1/\sqrt{\lambda_m}$ près, la coordonnée factorielle sur l'axe \mathbf{u}_m d'un individu i est la **moyenne arithmétique** des coordonnées factorielles sur l'axe \mathbf{v}^m des modalités k prises par i !

Dualité en ACM - Relations barycentriques

- Comme pour l'AFC, en ACM la dualité entre l'analyse de NO et NM est forte.
- Les axes factoriels ont même valeur propre λ_m .
- Rappelons que $\mathbf{L} = \mathbf{Z}/p$ et $\mathbf{C}^\top = \frac{1}{np} \mathbf{D}_M^{-1} \mathbf{Z}^\top$.
- ▷ Les **relations barycentriques** données ci-dessous permettent d'obtenir les coordonnées des éléments d'un nuage en fonction des coordonnées des éléments de l'autre nuage :

$$\mathbf{f}^m = \frac{1}{\sqrt{\lambda_m}} \mathbf{L} \mathbf{g}_m = \frac{1}{p\sqrt{\lambda_m}} \mathbf{Z} \mathbf{g}_m$$

$$\mathbf{g}_m = \frac{1}{\sqrt{\lambda_m}} \mathbf{C}^\top \mathbf{f}^m = \frac{1}{np\sqrt{\lambda_m}} \mathbf{D}_M^{-1} \mathbf{Z}^\top \mathbf{f}^m$$

Dualité en ACM - Relations barycentriques (suite)

- Le terme général de $\mathbf{g}_m = \frac{1}{np\sqrt{\lambda_m}} \mathbf{D}_M^{-1} \mathbf{Z}^\top \mathbf{f}^m$ s'écrit :

$$g_{m,k} = \frac{1}{\sqrt{\lambda_m}} \sum_{i=1}^n \frac{z_{ik}}{z_{.k}} f_i^m$$

- On peut réécrire $g_{m,k}$ comme suit :

$$g_{m,k} = \frac{1}{\sqrt{\lambda_m}} \left(\frac{1}{z_{.k}} \sum_{i: z_{ik}=1} f_i^m \right)$$

- ▷ Au facteur $1/\sqrt{\lambda_m}$ près, la coordonnée factorielle sur l'axe \mathbf{v}^m d'une modalité k est la **moyenne arithmétique** des coordonnées factorielles sur l'axe \mathbf{u}_m des individus i ayant la modalité k !

Dualité en ACM Représentation simultanée

- Contrairement à l'AFC les points lignes sont de poids uniformes. Les relations barycentriques se réduisent donc à des moyennes arithmétiques.
- Pour rappel, la dualité nous invite à ne pas faire deux analyses mais l'analyse d'un seul nuage de points suffit. On obtient la représentation graphique de l'autre nuage à partir des relations barycentriques.
- Par ailleurs, comme en AFC, ces propriétés indiquent que les nuages NO et NM peuvent être représentés sur un même graphique.
- Néanmoins, en pratique, cette représentation simultanée n'est pas souvent utilisée car :
 - ▶ Lorsque n et q sont grands la représentation simultanée projetant $n + p$ points, elle devient rapidement surchargée et donc illisible.
 - ▶ Souvent, l'ACM est utilisée dans le cas d'enquêtes où les individus sont anonymisés. Dans ce cas, la représentation des points lignes ne permet pas d'interpréter des éléments particuliers. En revanche, il est utile malgré tout d'analyser la forme du nuage projeté afin d'identifier les régions de fortes ou faibles densités.

Propriétés de l'ACM

- Comme évoqué précédemment, en ACM, il existe par nature des dépendances linéaires parmi les colonnes de \mathbb{Z} .
- Les modalités sont représentés par les colonnes de la matrice $\mathbf{C} = \frac{1}{np} \mathbf{Z} \mathbf{D}_M^{-1}$ (cf slide 362).
- On montre que le rang de $\mathbf{Z} \mathbf{D}_M^{-1}$ est au plus $q - p + 1$.
- Par conséquent, le rang de la matrice \mathbf{S} (cf slide 360) qui croise les modalités est donc également au plus de rang $q - p + 1$.
- Ainsi, si on enlève la 1ère valeur propre 1 associée au centre de gravité (cf slide 365), on aura en pratique $q - p$ valeurs propres non nulles.

Sous-nuage des modalités d'une même variable

- Le tableau disjonctif complet \mathbf{Z} , même s'il est interprété tel une table de contingence pour les besoins de la réduction de dimension, possède des propriétés spécifiques ce qui fait de l'ACM une méthode à part, vue comme l'équivalent de l'ACP pour des variables qualitatives.
- Une particularité propre à l'ACM est due au fait que les modalités sont associées par groupe à plusieurs variables qualitatives. Dans notre exemple précédent, nous avons $q = 9$ modalités mais $p = 3$ variables.
- On peut alors s'intéresser aux sous-nuages des points colonnes relatifs à la même variable. Nous avons donc p sous-nuages distincts.
- Les dimensions de \mathbb{R}^q représentant les modalités d'une même variable sont linéairement dépendantes. On peut montrer que les centres de gravité des p sous-nuages sont les mêmes et qu'ils correspondent au centre de gravité globale de NM .
- ▶ D'un point de vue pratique et graphique, cela veut dire que les axes factoriels permettent d'opposer les modalités d'une même variable.

Propriétés de l'ACM (suite)

- Le barycentre de NM est le vecteur $\bar{\mathbf{c}} = (\frac{1}{n}, \dots, \frac{1}{n})$ (marge des lignes de \mathbf{F} -cf l'exemple slide 357-).
- La distance du χ^2 entre un point colonne \mathbf{c}^k et $\bar{\mathbf{c}}$ vaut donc :

$$d_{\chi^2}^2(\mathbf{c}^k, \bar{\mathbf{c}}) = \sum_{i=1}^n n \left(\frac{z_{ik}}{z_{.k}} - \frac{1}{n} \right)^2$$

$$= \frac{n}{z_{.k}} - 1$$

- Ensuite, l'inertie totale de NM est donnée par :

$$\text{int}(\text{NM}) = \sum_{k=1}^q \frac{z_{.k}}{np} d_{\chi^2}^2(\mathbf{c}^k, \bar{\mathbf{c}}) = \frac{q}{p} - 1$$

- Remarque : contrairement à l'AFC pour le coup, l'inertie totale en ACM n'a pas d'interprétation statistique.

Aide à l'interprétation à partir des données actives

- On adapte les formules de l'AFC vues précédemment au cas de l'ACM. On peut alors exprimer les mesures en fonction de \mathbf{Z} .
- Qualité de la représentation du point ligne \mathbf{l}_i sur l'axe \mathbf{u}_m :**

$$qlt_{\mathbf{u}_m}(\mathbf{l}_i) = \frac{(f_i^m)^2}{\sum_{k=1}^q \frac{1}{f_{i.}} \left(\frac{f_{ik}}{f_{i.}} - f_{.k} \right)^2} = \frac{(f_i^m)^2}{\sum_{k=1}^q \frac{np}{z_{.k}} \left(\frac{z_{ik}}{p} - \frac{z_{.k}}{p} \right)^2}$$

- Qualité de la représentation du point colonne \mathbf{c}^k sur l'axe \mathbf{v}^m :**

$$qlt_{\mathbf{v}^m}(\mathbf{c}^k) = \frac{(g_{m,k})^2}{\sum_{i=1}^p \frac{1}{f_{i.}} \left(\frac{f_{ik}}{f_{i.}} - f_{i.} \right)^2} = \frac{(g_{m,k})^2}{\sum_{i=1}^n n \left(\frac{z_{ik}}{z_{.k}} - \frac{1}{n} \right)^2}$$

Aide à l'interprétation à partir des données actives (suite)

- La contribution du point ligne \mathbf{l}_i à l'axe \mathbf{u}_m :**

$$ctr_{\mathbf{u}_m}(\mathbf{l}_i) = f_{i.} \frac{(f_i^m)^2}{\lambda_m} = \frac{1}{n} \frac{(f_i^m)^2}{\lambda_m}$$

- La contribution du point colonne \mathbf{c}^k à l'axe \mathbf{v}^m :**

$$ctr_{\mathbf{v}^m}(\mathbf{c}^k) = f_{.k} \frac{(g_{m,k})^2}{\lambda_m} = \frac{z_{.k}}{np} \frac{(g_{m,k})^2}{\lambda_m}$$

- Rappelons que les contributions sont des pourcentages et qu'elles somment à 1.

Aide à l'interprétation à partir des données actives (suite)

- Rappelons que les mesures de qualité sont des mesures cosinus au carré. Plus précisément, l'angle qui nous intéresse est celui formé entre le point initial dans \mathbb{R}^q et celui résultant de la projection sur l'axe.
- Plus l'angle est fermé, plus les vecteurs sont proches et meilleure est la qualité (valeur proche de 1).
- Nous avons à nouveau une spécificité de l'ACM qui est due au fait que plusieurs modalités (dimensions de \mathbb{R}^q) sont issues d'une même variable.
- Nous pouvons calculer la **qualité de la représentation de la variable \mathbf{t}^j sur l'axe \mathbf{v}^m :**

$$qlt_{\mathbf{v}^m}(\mathbf{t}^j) = \sum_{k \in \mathbf{t}^j} qlt_{\mathbf{v}^m}(\mathbf{c}^k)$$

- On montre que $qlt_{\mathbf{v}^m}(\mathbf{t}^j) = \frac{1}{p} \eta(\mathbf{g}_m, \mathbf{t}^j)$ où $\eta(\mathbf{g}_m, \mathbf{t}^j)$ est le **rapport de corrélation** entre \mathbf{g}_m et \mathbf{t}^j .

Aide à l'interprétation à partir de données supplémentaires

- Nous pouvons ajouter également des individus ou variables/modalités supplémentaires.
- En raison des formules barycentriques propres à l'ACM données en slide 366 (et ceux qui suivent), il est encore plus simple dans ce cas de déterminer les coordonnées des éléments supplémentaires en ACM.
- Soit $\mathbf{z}_{+,i} \in 0,1^q$ la ligne du tableau disjonctif complet instanciant la représentation d'un individu supplémentaire $(z_{+,i1}, \dots, z_{+,iq})$. Alors :

$$f_i^{+,m} = \frac{1}{\sqrt{\lambda_m}} \left(\frac{1}{p} \sum_{k: \mathbf{z}_{+,ik}=1} g_{m,k} \right)$$

Aide à l'interprétation à partir de données supplémentaires (suite)

- Soit $\mathbf{z}_k^+ \in 0, 1^n$ la colonne du tableau disjonctif complet instanciant la représentation d'une modalité supplémentaire $(z_{1k}^+, \dots, z_{nk}^+)$. Alors :

$$g_{m,j}^+ = \frac{1}{\sqrt{\lambda_m}} \sum_{i=1}^p \frac{n_{ij}^+}{n_j^+} f_i^m = \frac{1}{\sqrt{\lambda_m}} \left(\frac{1}{z_{\cdot k}^+} \sum_{i: z_{ik}^+=1} f_i^m \right)$$

Rappel du Sommaire

- 4 Méthodes classiques de classification automatique (introduction)
 - Introduction
 - Les k -moyennes (k -means)
 - La classification ascendante hiérarchique (CAH) de Ward
 - Stabilisation de la CAH de Ward par les k -moyennes

Rappel du Sommaire

- 1 Introduction
- 2 Algèbre Linéaire
- 3 Méthodes de réduction de dimension
- 4 Méthodes classiques de classification automatique (introduction)

Introduction

- L'un des objectifs des méthodes de réduction est de mettre en lumière une typologie des individus et/ou des variables d'une table de données.
- Les axes factoriels permettent d'avoir une esquisse de cette typologie en opposant les éléments de part et d'autre du barycentre (repère du plan factoriel).
- La représentation des éléments se fait dans un espace euclidien de dimension réduite où les proximités entre éléments contribuent à la mise en place de cette typologie.
- Pour aller plus loin et compléter la visualisation par une véritable partition des éléments, nous pouvons associer aux méthodes de réduction de dimension des méthodes de classification automatique.
- Dans cette dernière partie du cours nous voyons des méthodes classiques de classification automatique.

La classification automatique

- La classification automatique (également dénommée typologie, segmentation, apprentissage non-supervisé, clustering (en anglais)...) consiste à déterminer automatiquement des **groupes homogènes d'éléments**.
- Homogénéité : les éléments appartenants au même groupe doivent être plus **proches** entre eux qu'avec des éléments appartenants à d'autres groupes.
- Les notions de **proximités** et donc de **distances** sont ici également des ingrédients fondamentaux du problème.
- Il faut ensuite définir un critère permettant de définir le degré d'homogénéité des groupes. On parle de **critère de partitionnement**.

La classification automatique (suite)

- La classification automatique (comme les méthodes de réduction) est une tâche courante en fouille de données⁵ /science des données⁶.
- Quelques exemples classiques :
 - ▶ En marketing : segmentation de la clientèle, campagne publicitaire, gestion de la relation clients...
 - ▶ En bio-informatique : analyse de puces à ADN pour déterminer par exemple des gènes ayant les même types de fonction...
- En SHS, comme expliqué précédemment, on l'utilise souvent en complément d'une méthode de réduction de dimension afin de mieux identifier les différents groupes d'un ensemble de données :
 - ▶ Typologie des sondés d'une enquête,
 - ▶ Typologie de données de recensement en démographie,
 - ▶ Détection de classes sémantiques entre termes en linguistique...
- Plus récemment, en sociologie/communication, la classification automatique de données de réseaux/graphes permet de faire la détection de communautés.

5. https://fr.wikipedia.org/wiki/Exploration_de_donn%C3%A9es

6. https://fr.wikipedia.org/wiki/Science_des_donn%C3%A9es

La classification automatique (suite)

- Rassembler des éléments en des groupes homogènes est une activité humaine et scientifique de base.
- Quelques exemples :
 - ▶ Exemple : les noms du langage désignent en ensemble d'éléments partageant des caractéristiques communes. Le mot table regroupe tout meuble avec une surface plane et reposant sur un ou plusieurs supports.
 - ▶ Exemple : en science du vivant, on classifie traditionnellement les être vivants en différentes catégories : Animaux, Plantes, Champignons...
 - ▶ Exemple : en chimie, le tableau périodique des éléments de Mendeleïev sont ordonnés par numéro atomique et organisés en fonction de leur configuration électronique.
 - ▶ Exemple : en sociologie, les professions et catégories socioprofessionnelles forment une nomenclature permettant de classer les métiers.

Plusieurs structures de classification

- Il existe plusieurs façons de classifier des éléments.
- Plusieurs types de structure de classification :
 - ▶ Partition à k classes : ensemble de k sous-ensembles mutuellement disjoints et recouvrant l'ensemble des éléments.
 - ▶ Classification hiérarchique représenté par un arbre binaire : ensemble de partitions emboîtées allant de la partition triviale à n classes à la partition triviale à 1 classe.
- Plusieurs types de degré d'appartenance aux classes :
 - ▶ Un élément appartient à une et une seule classe.
 - ▶ Un élément peut appartenir à plusieurs classes (classes recouvrantes).
 - ▶ Un élément à une probabilité d'appartenir à une classe.

Partitions et relations d'équivalence

- L'ensemble des partitions d'un ensemble \mathbb{O} est en bijection avec l'ensemble des relations d'équivalence sur \mathbb{O} .

Définition. (Relation binaire sur \mathbb{O})

Une relation binaire R sur \mathbb{O} est un sous-ensemble du produit cartésien $\mathbb{O} \times \mathbb{O}$, c'àd un **ensemble de paires ordonnées** : $R \subset \mathbb{O} \times \mathbb{O}$.
Si $(a, b) \in R$ (noté aussi aRb) on dit que " a est **en relation avec** b ".

Définition. (Relation d'équivalence \mathbb{O})

Une relation binaire R est une relation d'équivalence si elle satisfait les conditions suivantes :

- **Réflexivité** : $\forall x (xRx)$
- **Symétrie** : $\forall x, y (xRy \Rightarrow yRx)$
- **Transitivité** : $\forall x, y, z ((xRy \wedge yRz) \Rightarrow xRz)$

Classification hiérarchique et arbre binaire

Définition. (Classification hiérarchique et dendrogramme)

Une classification hiérarchique de \mathbb{O} est un ensemble de partitions emboîtées de \mathbb{O} . Elle peut être représenté par un **arbre binaire** où :

- le noeud racine contient tous les éléments de \mathbb{O} ,
- chaque noeud parent est une classe qui est l'union des classes de ses noeuds fils,
- chaque feuille est une classe singleton c'àd ne comportant qu'un seul élément de \mathbb{O} .

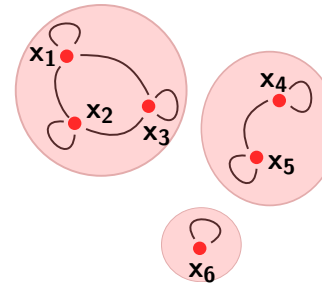
Plus formellement, si n, n' sont deux noeuds de l'arbre binaire alors :

$$(n \cap n' = \emptyset) \vee (n \subset n') \vee (n' \subset n)$$

Un **dendrogramme** est un arbre binaire représentant une classification hiérarchique, pour lequel on attribue également à chaque noeud n , une valeur numérique appelée la hauteur du noeud.

Illustration

- On peut représenter une relation d'équivalence par un graphe.
- Illustration d'une relation d'équivalence sur $\mathbb{O} = \{x_1, \dots, x_6\}$

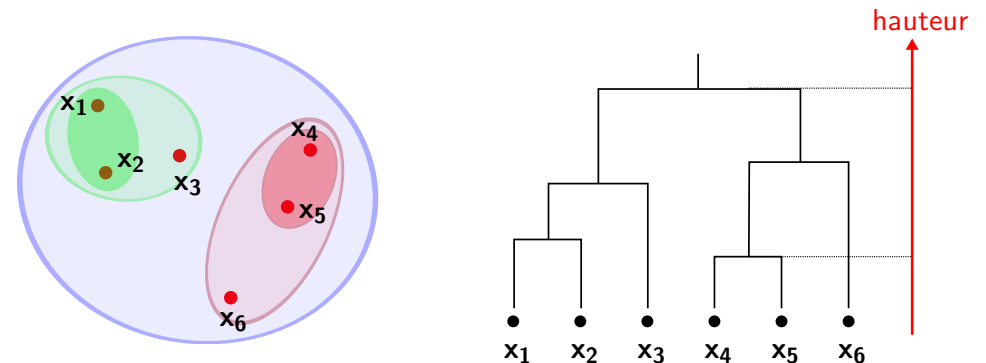


- R signifie "est dans la même classe que".
- **Réflexivité** :
 x_1Rx_1, \dots, x_6Rx_6
- **Symétrie** :
 $(x_1Rx_2 \wedge x_2Rx_1), \dots,$
 $(x_4Rx_5 \wedge x_5Rx_4)$
- **Transitivité** :
 $x_1Rx_2 \wedge x_2Rx_3 \Rightarrow x_1Rx_3,$
 $x_1Rx_3 \wedge x_3Rx_2 \Rightarrow x_1Rx_2,$
 $x_3Rx_2 \wedge x_2Rx_1 \Rightarrow x_3Rx_1, \dots$

- **Partition** :
 $\{x_1, x_2, x_3\}, \{x_4, x_5\}, \{x_6\}$

Illustration

- Illustration d'une classification hiérarchique de $\mathbb{O} = \{x_1, \dots, x_6\}$.



Un problème d'optimisation combinatoire

- Notons $\mathbb{P}(\mathbb{O})$ l'ensemble des partitions possibles de \mathbb{O} et notons $C \in \mathbb{P}(\mathbb{O})$ une partition quelconque.
- Supposons que nous avons une fonction $f : \mathbb{P}(\mathbb{O}) \rightarrow \mathbb{R}$ associant à toute partition C une mesure de sa qualité. f est le critère de partitionnement que nous cherchons à optimiser.
- Une approche simple du problème de partitionnement consisterait à :
 - ▶ Enumérer tous les éléments $C \in \mathbb{P}(\mathbb{O})$.
 - ▶ Mesurer pour chacun de ces éléments $f(C)$.
 - ▶ Une fois parcouru $\mathbb{P}(\mathbb{O})$ tout entier, retenir la partition $C^* = \arg \max_{C \in \mathbb{P}(\mathbb{O})} f(C)$.
- Cette approche est naïve car $|\mathbb{P}(\mathbb{O})|$ croît de façon exponentielle en fonction de $|\mathbb{O}|$!

Un problème d'optimisation combinatoire (suite)

- Voici les premiers nbs de Stirling de 2ème espèce et de Bell :

$n \backslash k$	0	1	2	3	4	5	6	$B(n)$
0	1	0	0	0	0	0	0	1
1	0	1	0	0	0	0	0	1
2	0	1	1	0	0	0	0	2
3	0	1	3	1	0	0	0	5
4	0	1	7	6	1	0	0	15
5	0	1	15	25	10	1	0	52
6	0	1	31	90	65	15	1	203

- Autre exemple : $B(71) \simeq 4 \times 10^{74}$!
- En bref, il est impossible d'énumérer toutes les partitions d'un ensemble de taille n dès que n dépasse des dizaines !
- On a alors recours à des **heuristiques** pour trouver des solutions (approchées) à la tâche de classification automatique.

Un problème d'optimisation combinatoire (suite)

- Le nb de partitions en k classes d'un ensemble \mathbb{O} de taille n est donné par le nb de Stirling de 2ème espèce dénoté $S(n, k)$:

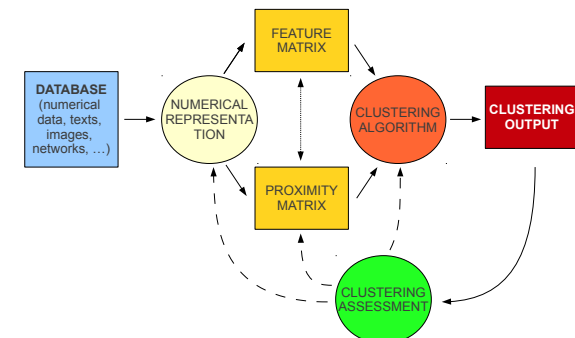
$$S(n, k) = \frac{1}{k!} \sum_{j=0}^k (-1)^j \binom{n}{j} j^n$$

- Le nb total de partitions d'un ensemble de taille n est donné par le nb de Bell dénoté $B(n)$ et qui vaut donc :

$$B(n) = \sum_{k=0}^n S(n, k)$$

La classification automatique en pratique

- Il existe plusieurs façons de classifier, plusieurs critères de partitionnement et le problème est à la base exponentiel (NP-complet). Il existe donc plusieurs algorithmes de classification automatique !
- En général, il s'agit d'une procédure itérative :



La classification automatique dans le cadre de ce cours

- Nous nous concentrons sur les méthodes de bases.
- En particulier, nous étudions celles utilisant des distances (euclidiennes) pour la mesure de proximité.
- De plus, l'objectif premier est d'étudier des méthodes complémentaires aux techniques de réduction de dimension.
- Il existe dans ce cas deux méthodes classiques :
 - ▶ la méthode des **k-moyennes**,
 - ▶ la classification hiérarchique ascendante de **Ward**,
 - ▶ la **méthode hybride** combinant les deux méthodes précédentes.

Représentation euclidienne des données

- Nous supposons que les individus sont les éléments de l'ensemble \mathbb{O} avec $|\mathbb{O}| = n$.
- Ils sont représentés par des vecteurs $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ dans un espace euclidien.
- L'espace euclidien est engendré par p variables quantitatives. Nous avons donc $\mathbf{x}_i \in \mathbb{R}^p$ pour tout $i = 1, \dots, n$.
- Nous disposons donc comme précédemment d'une matrice de données :

$$\mathbf{X} = \begin{matrix} & \mathbf{x}_1 & & \mathbf{x}_i & & \mathbf{x}_n \end{matrix} \begin{pmatrix} x_{11} & & x_{1k} & & x_{1p} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{ik} & \dots & x_{ip} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & & x_{nk} & & x_{np} \end{pmatrix}$$

Définition des distances

Définition.

Soit \mathbb{E} un ensemble d'éléments. Une distance est une application $d : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}^+$ qui vérifie les propriétés suivantes, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{E}$:

- **non-négativité** : $d(\mathbf{x}, \mathbf{y}) \geq 0$,
- **symétrie** : $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$,
- **identité des indiscernables** : $\mathbf{x} = \mathbf{y} \Leftrightarrow d(\mathbf{x}, \mathbf{y}) = 0$,
- **inégalité triangulaire** : $\forall \mathbf{z} (d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}))$.

Représentation euclidienne des données (suite)

- Rappelons que la distance euclidienne (au carré) entre \mathbf{x}_i et \mathbf{x}_j vaut :

$$\begin{aligned} d^2(\mathbf{x}_i, \mathbf{x}_j) &= \|\mathbf{x}_i - \mathbf{x}_j\|^2 \\ &= \langle \mathbf{x}_i - \mathbf{x}_j, \mathbf{x}_i - \mathbf{x}_j \rangle \\ &= \sum_{k=1}^p (x_{ik} - x_{jk})^2 \end{aligned}$$

- La distance euclidienne vérifie donc les propriétés précédentes.
- Remarque : il existe des distances qui sont autre qu'euclidienne. Exemple la distance de "Manhattan" ou "city-block" ou ℓ_1 :

$$d_{manh}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

Variance intra-classe

- Soit C une partition. Notons $|C|$ le nombre de classe de C . Notons également C_l une classe de C avec $l = 1, \dots, |C|$.
- Par exemple : $C = \{\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}, \{\mathbf{x}_4, \mathbf{x}_5\}, \{\mathbf{x}_6\}\}$, $|C| = 3$, $C_1 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, $C_2 = \{\mathbf{x}_4, \mathbf{x}_5\}$ et $C_3 = \{\mathbf{x}_6\}$.
- Le critère de partitionnement que nous allons étudier principalement est la **variance (ou inertie) intra-classe** d'une partition C :

$$int_w(C) = \frac{1}{n} \sum_{l=1}^{|C|} \sum_{\mathbf{x} \in C_l} \underbrace{\|\mathbf{x} - \mathbf{m}_l\|^2}_{d^2(\mathbf{x}, \mathbf{m}_l)}$$

où $\mathbf{m}_l = \frac{1}{|C_l|} \sum_{\mathbf{x} \in C_l} \mathbf{x}$ est le barycentre (vecteur moyen) de C_l .

- ▶ On veut minimiser int_w donc on cherche C tel que les éléments d'un groupe sont proches de leur barycentre. On a alors des classes homogènes.

Fondements de la méthode des k -moyennes

- Méthode proposée initialement par Forgry en 1965 puis MacQueen en 1967. L'algorithme usuel est celui décrit par Hartigan 1975.
- k est un paramètre correspond au nombre de classes souhaités et qui doit donc être donné par l'utilisateur.
- La méthode cherche donc à minimiser int_w sous la contrainte que C soit une partition en k classes.
- L'algorithme est une heuristique (le problème est NP-complet) et consiste à transférer un individu de sa classe actuelle à une autre si cela permet de diminuer int_w .

Rappel du Sommaire

4 Méthodes classiques de classification automatique (introduction)

- Introduction
- Les k -moyennes (k -means)
- La classification ascendante hiérarchique (CAH) de Ward
- Stabilisation de la CAH de Ward par les k -moyennes

Algorithme des k -moyennes

- Pseudo-code de l'algorithme des k -moyennes.

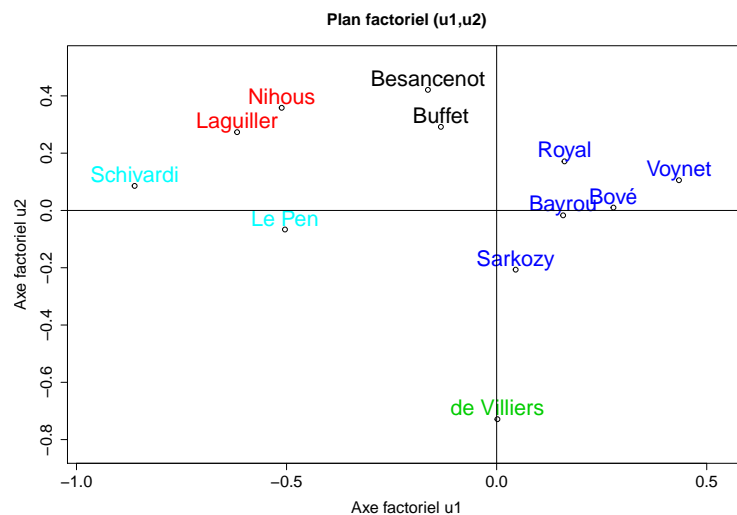
```

1  Input :  $\mathbf{X}$  and  $k$ 
2  Initialize  $C$  with  $k$  different clusters
3  While a stopping criterion is not reached do
4      For all  $\mathbf{x} \in \mathbf{O}$  do
5          For all  $C_l \in C$  do
6              Compute  $d^2(\mathbf{x}, \mathbf{m}_l)$ 
7          End For
8          Find  $C_{l^*} = \operatorname{argmin}_{C_l \in C} d^2(\mathbf{x}, \mathbf{m}_l)$ 
9          Move  $\mathbf{x}$  from its current cluster to  $C_{l^*}$ 
10         Update the mean vectors accordingly
11     End For
12 End While
13 Ouput :  $C$ 
```

Algorithme des k -moyennes

- On distingue donc 2 phases :
 - Une phase d'initialisation : l'approche de base consiste à affecter aléatoirement les éléments dans k classes.
 - Une phase d'itérations : l'algo calcule les distances euclidiennes (au carré) de tous les éléments avec les barycentres de toutes les classes puis transfère les éléments vers les classes dont ils sont les plus proches. Chaque barycentre est alors mis à jour selon les éléments qui constituent la classe après les transferts.
- Le critère d'arrêt peut être :
 - un nb maximal d'itérations ou,
 - la convergence du critère int_w (quand la valeur ne change plus bcp) ou,
 - la convergence de la partition C (quand il n'y a plus de transfert).

Code R (suite)



Code R

- Application aux données du TP5 sur les candidats à l'élection présidentielle de 2007 représentés dans le premier plan factoriel.

```
> res_kmeans=kmeans(cbind(f1,f2),centers = 5)
> res_kmeans$cluster
[1] 5 2 1 1 4 4 4 2 4 4 3 5
> res_kmeans$centers
      [,1]      [,2]
1 -0.148141998  0.356495820
2 -0.564683222  0.315860312
3  0.001860403 -0.728753980
4  0.215364045  0.012627602
5 -0.682717391  0.009615943
> res_kmeans$tot.withinss
[1] 0.2632774
```

Avantages/Inconvénients (en bref)

- Avantages :
 - Un des algorithmes des plus utilisés (si ce n'est pas le plus utilisé).
 - Algorithme de complexité linéaire par rapport à n (si k et p sont petits). On peut donc l'appliquer sur de grandes bases de données.
- Inconvénients :
 - On doit fixer le nombre k et si on n'a pas d'a priori cela peut poser un problème important.
 - La phase d'initialisation est aléatoire et conditionne le résultat obtenu.
 - Les classes trouvées sont de forme sphérique.

Rappel du Sommaire

4 Méthodes classiques de classification automatique (introduction)

- Introduction
- Les k -moyennes (k -means)
- La classification ascendante hiérarchique (CAH) de Ward
- Stabilisation de la CAH de Ward par les k -moyennes

Algorithme de la CAH de Ward

- Pseudo-code de la classification hiérarchique ascendante :

- 1 **Input** : \mathbf{D} (squared euclidean distance matrix)
- 2 Initialize the tree representation with n leaves
- 3 **While** not all data points are grouped together **do**
- 4 Merge the two closest clusters according to some distance measure
- 5 Add a parent node in the tree representation accordingly
- 6 **End While**
- 7 **Output** : tree representation

Fondements de la méthode

- Proposé par Ward en 1963.
- La matrice donnée en entrée est celle des distances euclidiennes au carré pour chaque paire d'éléments de $\mathbb{O} \times \mathbb{O}$.
- C'est un algorithme de classification hiérarchique ascendante :
 - ▶ on commence par n classes (singletons/feuilles de l'arbre),
 - ▶ à chaque itération on regroupe deux classes en une seule (ajout d'un noeud dans l'arbre binaire),
 - ▶ jusqu'à obtenir une classe regroupant tous les éléments (racine de l'arbre).
- A chaque itération les classes regroupées sont celles permettant de minimiser int_w .
- Le critère est donc le même que pour les k -moyennes mais on utilise ici une heuristique gloutonne conduisant à la détermination d'une classification hiérarchique.

Mise à jour des distances dans le cas de Ward

- A l'étape 4 précédente, supposons que nous fusionnons C_l^* et $C_{l'}^*$:

$$(C_l^*, C_{l'}^*) = \arg \min_{C_l, C_{l'} \in \mathbb{P}^t} d_{ward}(C_l, C_{l'})$$

où \mathbb{P}^t est la partition de \mathbf{O} à l'itération t .

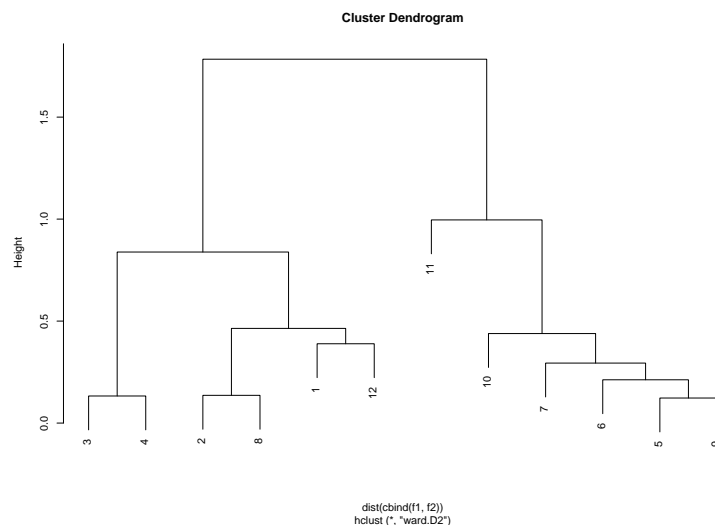
- Il nous faut déterminer les distances de Ward entre la nouvelle classe $C_l^* \cup C_{l'}^*$ et toutes les classes existantes $C_k \in \mathbb{P}^t$.
- Dans le cas de Ward, ces nouvelles distances sont calculées comme suit :

$$\begin{aligned} d_{ward}(C_k, C_l^* \cup C_{l'}^*) &= \frac{|C_k| + |C_l^*|}{|C_k| + |C_l^*| + |C_{l'}^*|} d_{ward}(C_k, C_l^*) \\ &\quad + \frac{|C_k| + |C_{l'}^*|}{|C_k| + |C_l^*| + |C_{l'}^*|} d_{ward}(C_k, C_{l'}^*) \\ &\quad - \frac{|C_k|}{|C_k| + |C_l^*| + |C_{l'}^*|} d_{ward}(C_l^*, C_{l'}^*) \end{aligned}$$

Cas général : la formule de Lance-Williams

- On montre que fusionner à chaque itération les classes dont la distance de Ward est la plus petite conduit à diminuer au mieux int_w .
- La formule de mise à jour de Ward est un cas particulier qui est spécifiquement lié au critère int_w .
- Il existe bien d'autres distances dans le cadre de la classification hiérarchique ascendante.
- La formule de Lance-Williams donnée en 1967 est une équation paramétrique permettant de généraliser 7 distances proposées dans la littérature.

Code R (suite)



Code R

- Application aux données du TP5 sur les candidats à l'élection présidentielle de 2007 représentés dans le premier plan factoriel.

```
> res_ward=hclust(dist(cbind(f1,f2)),method = "ward.D2")
> str(as.dendrogram(res_ward))
--[dendrogram w/ 2 branches and 12 members at h = 1.78]
|  |--[dendrogram w/ 2 branches and 6 members at h = 0.839]
|    |  |--[dendrogram w/ 2 branches and 2 members at h = 0.133]
|    |    |  |--leaf 3
|    |    |  '--leaf 4
|    |  '--[dendrogram w/ 2 branches and 4 members at h = 0.464]
|    |    |  |--[dendrogram w/ 2 branches and 2 members at h = 0.136]
|    |    |    |  |--leaf 2
|    |    |    |  '--leaf 8
|    |    '--[dendrogram w/ 2 branches and 2 members at h = 0.389]
|    |      |  |--leaf 1
|    |      '--leaf 12
|  '--[dendrogram w/ 2 branches and 6 members at h = 0.996]
```

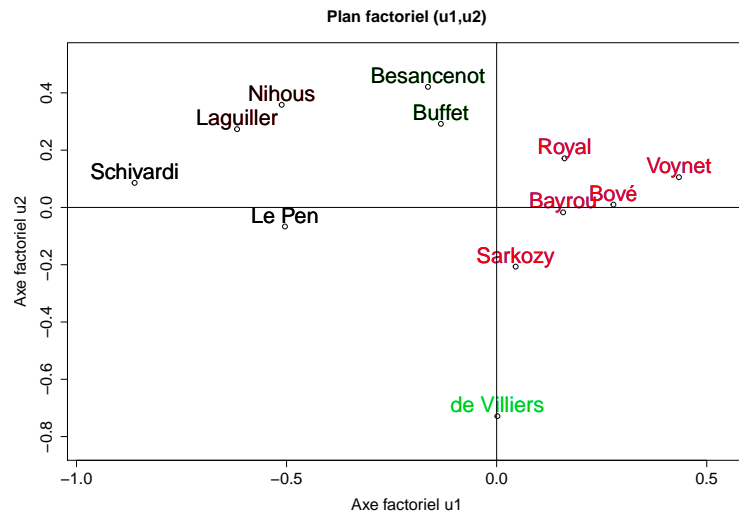
Code R (suite)

- Nous pouvons couper l'arbre à tout niveau souhaité!

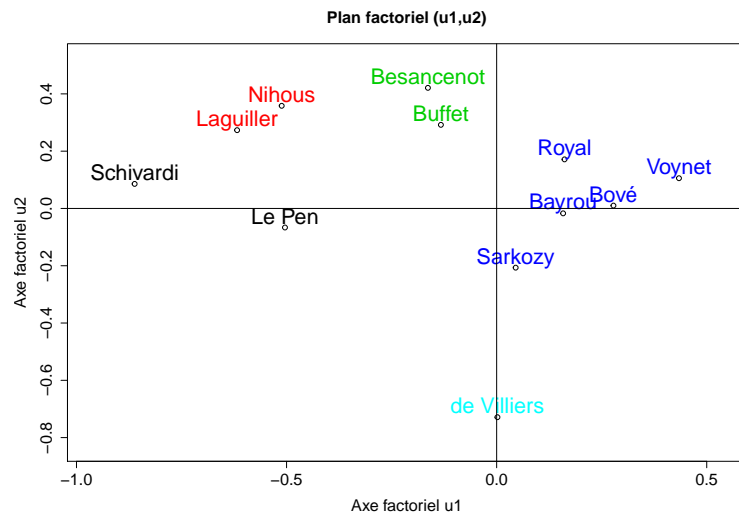
```
> cutree(res_ward,k = 3)
[1] 1 1 1 1 2 2 2 1 2 2 3 1
> cutree(res_ward,k = 4)
[1] 1 1 2 2 3 3 3 1 3 3 4 1
> cutree(res_ward,k = 5)
[1] 1 2 3 3 4 4 4 2 4 4 5 1
```

- Dans cet exemple coupé à 5 classes donne le même résultat que les k -moyennes.

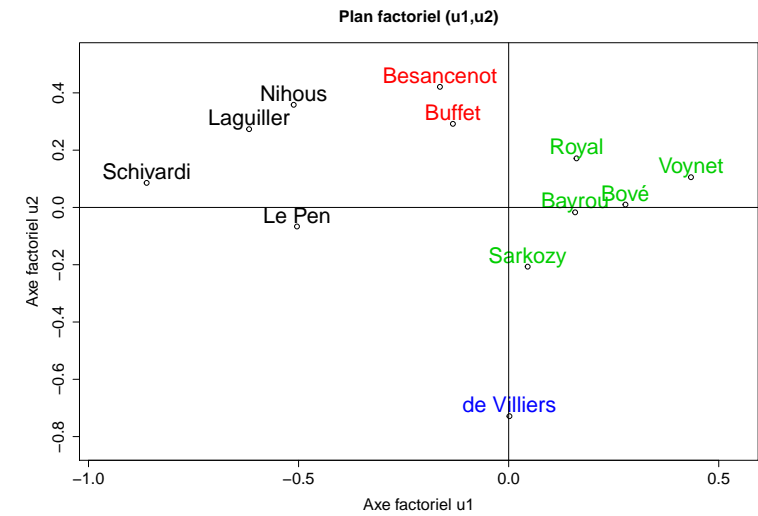
Code R (suite)



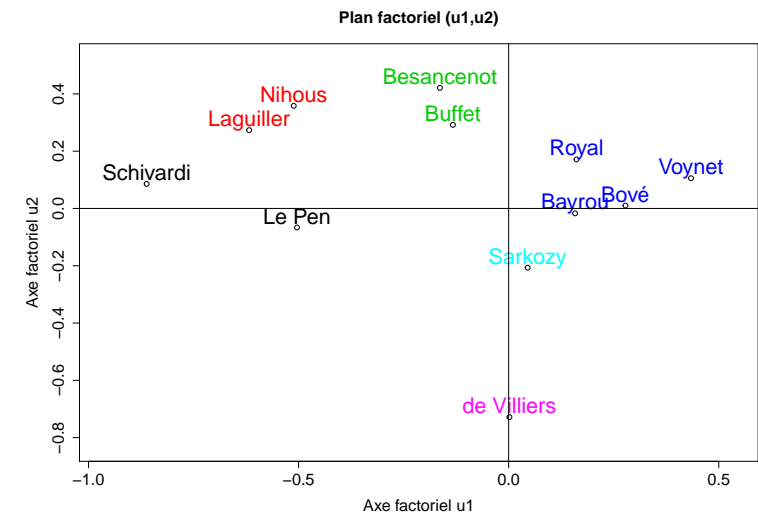
Code R (suite)



Code R (suite)



Code R (suite)



Avantages/Inconvénients

- Avantages :
 - ▶ On ne fixe pas a priori le nombre de classes.
 - ▶ La hiérarchie est plus informative que le partitionnement car elle permet de voir quelles classes ont été fusionnées à chaque étape. On a donc plus d'information.
- Inconvénients :
 - ▶ La stratégie est dite gloutonne : ce qui a été regroupé précédemment ne peut être séparé. Donc l'algorithme est d'une certaine façon sous-optimal.
 - ▶ La complexité en mémoire est quadratique (matrice de distances donnée en entrée) et surtout la complexité du calcul est par défaut cubique. Il existe des méthodes permettant de réduire la complexité mais elle reste trop grande pour permettre à la méthode de traiter de très grandes bases de données.

Idée de l'approche hybride

- On part d'une classification hiérarchique de Ward.
- On coupe à un niveau donné.
- La méthode de Ward est (par construction) sous-optimale pour chaque niveau intermédiaire.
- L'idée est donc d'améliorer le résultat de la partition obtenue en utilisant les transferts de k -moyennes.
- Cette approche peut également être vue comme une solution pour l'initialisation des k -moyennes (mais elle est coûteuse d'un point de vue computationnelle).
- Dans de nombreux logiciels d'ADD cette approche hybride est implémentée par défaut.

Rappel du Sommaire

- 4 Méthodes classiques de classification automatique (introduction)
 - Introduction
 - Les k -moyennes (k -means)
 - La classification ascendante hiérarchique (CAH) de Ward
 - Stabilisation de la CAH de Ward par les k -moyennes