



Algèbre Linéaire et Analyse de Données

Licence 2 MIASHS (2021-2022)

Guillaume Metzler

Institut de Communication (ICOM)

Université de Lyon, Université Lumière Lyon 2


Laboratoire ERIC UR 3083, Lyon, France

guillaume.metzler@univ-lyon2.fr

Résumé

Ce cours est à destination des étudiants de la Licence 2 MIASHS de l'Université Lumière, Lyon 2. Il se présente en deux parties : une première partie présente les outils **d'algèbre linéaire** nécessaires à l'étude des techniques **d'analyses de données** présentées dans la seconde partie de ce document.

La première partie introduit essentiellement les notions d'espaces vectoriels, d'applications linéaires. Nous verrons ensuite comment étudier de tels objets ainsi que les propriétés de ces derniers. Nous accorderons beaucoup d'importances à l'étude des matrices dans ce cours car ce sont les principaux objets utilisés pour représenter les données, mais c'est aussi avec cette représentation que nous pourrons analyser et "faire davantage de calculs" pour illustrer les différentes notions. La géométrie sera également mise à l'honneur car elle s'avère nécessaire pour l'étude des données afin de voir si des groupes de données se forment et éventuellement étudier les distances entre les groupes.

Cette première partie est essentiellement théorique mais très formatrice. Elle permettra au lecteur de se plonger dans un formalisme abstrait qu'il pourra ensuite mettre en application. Pour l'aider, des exercices lui permettant de vérifier ses connaissances et de s'entraîner à effectuer des démonstrations sont disponibles. Des exercices plus calculatoires sont également disponibles. Tout cela à la fin de chaque section. Enfin, quand le contexte le permet, chaque fin de section sera illustrée par des exemples sous  afin d'introduire progressivement les commandes qui nous serviront à faire de l'Analyse de Données dans la deuxième partie.

La deuxième partie se concentrera plus particulièrement sur les techniques d'analyses de données comme l'analyse en composante principale, la décomposition en valeurs singulières, l'analyse factorielle ou encore l'analyse en composantes multiples. Nous verrons comment toutes ces techniques nous permettront de synthétiser les informations, de les localiser mais aussi de voir si des groupes d'individus se trouvent dans nos données et quelles sont les variables qui les caractérisent.

Les différentes méthodes d'analyses seront illustrées avec des exemples sous .

Remerciements Un grand merci à Julien Ah-Pine dont les ressources ont été d'une aide précieuse pour la préparation de ce cours tant sur le contenu que sur la création des fiches d'exercices.

Un grand merci également aux différents chargés de TD qui ont permis d'améliorer les exercices et les corrections de ces derniers à travers les différentes remarques constructives :

- 2022 : Mickaël Lallouche et Martial Amovin dont les remarques ont été très pertinentes et dont bons nombres d'entre elles doivent encore être prises en compte pour l'amélioration de ce document.
- 2023 : ...

Table des matières

I	Introduction	6
II	Algèbre Linéaire	12
1	Espaces vectoriels et applications linéaires	14
1.1	Généralités sur les espaces vectoriels	14
1.2	Propriétés des sous-espaces vectoriels	19
1.3	Applications linéaires	23
1.4	Projecteurs et symétries	30
2	Espaces vectoriels de dimension finie	36
2.1	Familles libres, familles génératrices et bases	36
2.2	Espace vectoriel de dimension finie	43
2.3	Représentation d'une application linéaire dans une base	47
2.4	Retour sur sous-espaces vectoriels	51
3	Matrices et calcul matriciel	59
3.1	Rappels sur les matrices	59
3.2	Représentation des applications linéaires	64
3.3	Familles de vecteurs et changement de bases	66
3.4	Opérations élémentaires sur les matrices	71
3.5	Rang, déterminant, et inverse d'une matrice	73
3.5.1	Rang d'une matrice	73
3.5.2	Déterminant d'une matrice	74
3.5.3	Inverse d'une matrice	78
3.6	Calculs explicites	79
3.6.1	Matrice échelonnée (réduite)	81
3.6.2	Calcul du rang	82
3.6.3	Calcul du déterminant	83
3.6.4	Inverse d'une matrice	86

4	Systèmes linéaires	98
4.1	Définition et interprétation	98
4.2	Résolution d'un système linéaire	99
5	Réduction des endomorphismes	110
5.1	Vecteurs propres et valeurs propres	111
5.2	Diagonalisation	111
5.2.1	Recherche des valeurs propres	112
5.2.2	Recherche des vecteurs propres	115
5.3	Caractérisation des endomorphismes diagonalisables	117
6	Formes quadratiques et espaces euclidiens	127
6.1	Formes bilinéaires	127
6.2	Formes quadratiques	130
6.3	Espaces euclidiens	135
III	Analyse de données	158
7	Généralités et Décomposition en Valeurs Singulières (SVD)	161
7.1	Généralités	161
7.2	Vers la recherche de valeurs propres et vecteurs propres	165
7.3	Décomposition en valeurs singulières	170
8	Analyse en Composantes Principales (ACP)	178
8.1	Principe	178
8.2	Transformations	179
8.3	Graphes et commentaires	186
9	Généralisation des méthodes	194
10	Analyse Factorielle des Correspondances (AFC)	199
10.1	Préliminaires	199
10.2	Analyse Factorielle des Correspondances	203
10.2.1	Application aux profils lignes	204
10.2.2	Application aux profils colonnes	208
10.3	Remarques et interprétations	211

11 Analyse factorielle des Correspondances Multiples (ACM)	220
11.1 Préliminaires	220
11.2 Analyse factorielle des Correspondances Multiples	224
11.2.1 Applications aux profils lignes.	227
11.2.2 Applications aux profils colonnes.	227
11.3 Remarques et interprétations	229

Première partie

Introduction

Jusqu'à présent, le lecteur aura uniquement eu des cours sur la statistique descriptive où les concepts de base de statistique sont présentés comme la **moyenne** et la **variance**.

Ces cours auront eu pour objectif de décrire les propriétés de variables aléatoires ou encore l'étude d'un ou de plusieurs échantillons issu(s) d'une population. La plupart de ces analyses sont dites *unidimensionnelles* et ne portent que sur l'étude d'une variable aléatoire seule voire l'étude d'un lien entre deux variables aléatoires. Mais l'analyse de données ne se limite pas à l'étude d'une seule ou de deux variables aléatoires, on cherchera davantage à prendre en compte l'ensemble des descripteurs des individus pour les étudier, c'est ce que l'on appelle l'**analyse multidimensionnelle**.

Vers l'analyse multidimensionnelle

La réalité est parfois beaucoup plus *complexe* et *riche* que le contexte d'étude imposé par la simple statistique descriptive. Elle nécessite de pouvoir étudier un échantillon en prenant en compte l'ensemble de l'information et non pas un ou deux critères isolés que l'on étudierait sans prendre en compte les autres. L'objectif de ce cours est d'explorer et de trouver la richesse des informations présente dans nos données mais en prenant en compte **toute l'information**. Mais cela ne se fait pas sans difficultés, tout d'abord conceptuelles et graphiques.

En effet, là où l'analyse univariée ou bivariée se voulait rassurante via l'étude de graphiques en deux dimensions (histogrammes, droite de régression, ...) nous serons vite plongés dans des espaces de dimension supérieure à 3 que nous ne sommes plus capables de représenter et de visualiser en tant qu'humain.

Un exemple pour voir l'intérêt

Considérons un exemple¹ qui consiste à étudier un questionnaire de satisfaction.

Nous demandons aux étudiants qui suivent ce cours s'ils sont satisfaits, à travers 5 critères sur lesquels ils doivent se positionner sur une échelle de 1 à 5, 1 correspondant à "très insatisfait" et 5 à "très satisfait". Voici ces 5 critères :

1. Cet exemple est issu de l'adresse suivante : <https://openclassrooms.com/fr/courses/4525281-realisez-une-analyse-exploratoire-de-donnees/5280363-decouvrez-linteret-de-l-analyse-multidimensionnelle>.

	Q1	Q2	Q3	Q4	Q5
Individu 1	3	3	3	5	5
Individu 2	2	3	1	5	4
Individu 3	2	3	3	4	5
Individu 4	1	1	1	1	1
Individu 5	5	5	4	3	3
Individu 6	4	5	5	2	3
Individu 7	5	5	5	3	3
Individu 8	1	1	1	1	1

TABLE 1 – Résultats du questionnaire sur un ensemble de 8 individus.

	Q1	Q2	Q3	Q4	Q5
Moyenne	2.875	3.25	3	3	3.125

TABLE 2 – Réponse moyenne sur les différents critères vis à vis des réponses fournies au questionnaire et présentées dans la Table 1.

1. Clarté du cours écrit
2. Fluidité de la lecture du cours écrit
3. Facilité à comprendre les exemples du cours
4. Clarté des vidéos
5. Satisfaction vis à vis de l'enseignant dans la vidéo

On peut alors étudier les réponses fournies par 8 étudiants qui se trouvent dans la Table 1. On fera ici abstraction de la significativité de cette étude mais c'est simplement pour illustrer l'intérêt de l'analyse multidimensionnelle.

Si nous reprenons simplement les outils statistiques vus jusqu'à présent, nous serons simplement tenté de calculer la moyenne de chacune des réponses sur l'ensemble des individus. Cette moyenne nous donnerait un individu *moyen* dont les caractéristiques présentées en Table 2 laissent à penser que l'échantillon est globalement moyennement satisfait vis à vis de l'ensemble des critères étudiés.

Connaître le "comportement moyen" est une chose importante lorsque l'on cherche à résumer l'information de façon très synthétique. Mais ce qui nous intéresse ici, c'est aussi le profil des étudiants, c'est-à-dire l'ensemble des 5 valeurs qui décrivent cet étudiant répondant (on parlera plus tard de vecteur). On va analyser ces 5 valeurs pour voir s'il est possible de regrouper ces étudiants par "paquets" pour ainsi déterminer les différents profils des répondants.

Comme le montre la Table 3, nous identifions trois profils différents :

- le premier profil en blanc, formé des individus 4 et 8, se présente comme l'ensemble des individus qui ne sont pas satisfaits par la formation ou le contenu

	Q1	Q2	Q3	Q4	Q5
Individu 1	3	3	3	5	5
Individu 2	2	3	2	5	4
Individu 3	2	3	3	4	5
Individu 4	1	1	1	1	1
Individu 5	5	5	4	4	3
Individu 6	4	5	5	2	3
Individu 7	5	5	5	3	3
Individu 8	1	1	1	1	1

TABLE 3 – Création de groupes en fonction des réponses des individus. On identifie trois profils différents en fonction de la nature des réponses, identifiés en marron, gris et blanc.

de cette dernière. Ce type de réponse peut parfois être très "classique" dans l'étude d'un questionnaire mais peut être interprété de façon différente : (i) elle peut montrer une réelle insatisfaction de la part du répondant ou (ii) cela peut aussi être caractéristique de réponses fournies sans avoir pris le soin de lire la question. Le questionnaire est perçu comme chronophage pour l'individu qui souhaite alors y passer le moins de temps possible.

- un deuxième profil en marron, formé des individus 1 à 3, qui représente l'ensemble des individus qui sont satisfaits du contenu vidéo mais qui le sont moins par le contenu dit "textuel"
- un dernier profil en gris, formé des individus 5 à 7, qui représente l'ensemble des individus qui sont satisfaits du contenu textuel mais qui le sont moins par le contenu vidéo.

Différentes méthodes d'analyse

Dans l'analyse de données, on peut regrouper les méthodes dans trois grandes familles :

- les méthodes **supervisées** qui sont notamment employées dans les modèles de régression qui vous seront présentés l'année prochaine. Ces modèles consistent par exemple à chercher à prédire une valeur d'un individu en fonction des valeurs des descripteurs de cet individu. Ces méthodes peuvent aussi constituer une bonne transition vers les modèles dits de *Machine Learning*. Mais on ne va pas aborder cette partie là ici.
- les méthodes **non supervisées**. Par opposition aux méthode supervisée, on va ici chercher à partitionner nos données, en les rangeant dans des "classes" sans que ces dernières ne soient définies à l'avance. C'est ce que nous avons fait lorsque nous avons précédemment étudié les profils des répondants à notre questionnaire. Sans le savoir, nous avons fait ce que l'on appelle du **clustering**,

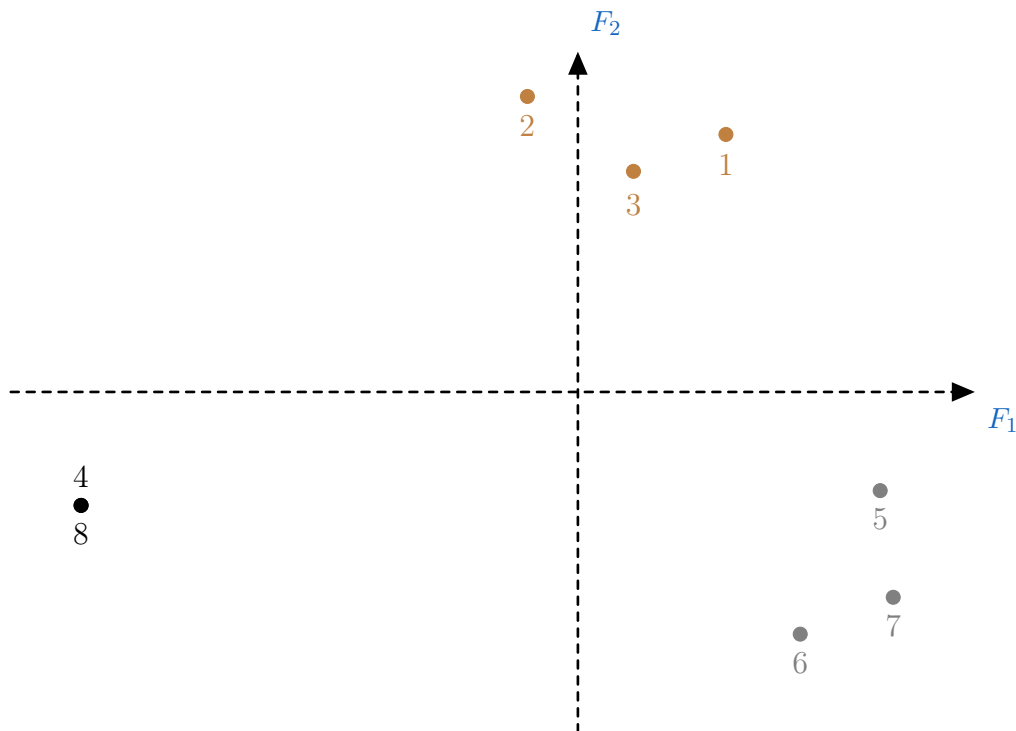


FIGURE 1 – Représentation des individus, *i.e.* des réponses fournies au questionnaire, dans un espace à deux dimensions obtenu par l'ACP. La couleur des individus correspond à celle utilisée dans la Table 3 pour identifier les différents profils. Les individus 4 et 8 étant identiques, ils ont la même représentation.

i.e. du partitionnement de nos données. Ces méthodes là ne seront pas étudiées dans le cadre de ce cours, faute de temps.

- les méthodes **factorielles** qui vont résumer l'information présente dans le jeu de données, en projetant les données dans des espaces de dimensions plus faibles. Ces espaces permettent d'avoir une représentation *visuelle* de l'information visible et qui préserve au mieux cette information. Une méthode très connue qui permet de faire cela est l'**analyse en composantes principales**.

Concentrons-nous sur les méthodes factorielles, par exemple l'ACP (pour analyse en composantes principales) permet de réduire le nombre de variables en trouvant de nouvelles variables qui en synthétisent plusieurs. Trouver une variable synthétique permet de remplacer plusieurs colonnes du tableau par une seule. Malheureusement, cette transformation nous fera perdre un peu d'information, mais on sera capable de quantifier la part d'informations qui sera contenue sur une ou plusieurs variables synthétiques.

Par exemple, si on projette notre précédent jeu de données dans un espace à deux dimensions via une ACP, nous aurons la représentation donnée en Figure 1.

Cette ACP nous a permis de projeter les données dans un espace à 2 dimensions

	Q1	Q2	Q3	Q4	Q5
Individu 1	bleu	France	chat	brun clair	grande
Individu 2	bleu	Allemagne	chat	blond	moyenne
Individu 3	vert	Belgique	souris	gris	grande
Individu 4	marron	Pologne	lion	noir	petite
Individu 5	vert	Autriche	hamster	blond	grande
Individu 6	marron	Ukraine	tigre	noir	petite
Individu 7	marron	Hongrie	chien	brun	petite
Individu 8	marron	Roumanie	chien	brun	moyenne

TABLE 4 – Résultats du deuxième questionnaire sur un ensemble de 8 individus.

ce qui permet de visualiser graphiquement les données initialement décrites dans un espace de dimension 5. Cette représentation permet également de reconnaître les différents profils observés dans notre tableau initial. C'est donc une technique très puissante pour synthétiser les données ! Il reste à voir comment obtenir une telle représentation ... mais c'est tout l'objet de ce cours !

L'ACP nous permettra également d'étudier :

- la variabilité entre les individus, on pourra ainsi détecter d'éventuelles différences ou ressemblances entre les individus.
- les liaisons entre les variables : est-ce que certaines variables sont très corrélées entre elles et pourraient être regroupées pour former des variables synthétiques en nombre moins important ?

L'ACP n'est bien sûr pas la seule méthode d'analyse factorielle, et heureusement car elle présente un inconvénient majeur, elle ne permet d'étudier que des variables dites *quantitatives*.

Imaginons que l'on dispose des résultats d'une autre enquête où l'on a posé les questions suivantes :

1. Quelle est la couleur de vos yeux ?
2. De quel pays êtes vous originaire ?
3. Quel est votre animal préféré ?
4. Quelle est la couleur de vos cheveux ?
5. Comment qualifieriez vous votre taille ? (petite, moyenne, grande)

Les réponses à ces questions sont synthétisées dans la Table 4. On ne pourra pas utiliser une ACP pour étudier ce tableau. Heureusement l'ACP n'est pas la seule méthode d'analyse factorielle qui existe et nous étudierons deux autres méthodes pour synthétiser des variables *qualitatives* : l'**analyse factorielle des correspondances** et sa généralisation l'**analyse des correspondances multiples**.

Mais toutes ces méthodes nécessitent, des notions en géométrie et indirectement en algèbre linéaire nous permettant de manipuler ces tableaux et la projection de ces données dans un espace de dimension réduite. La présentation de cet outil fait l'objet de la prochaine partie de ce cours !

Deuxième partie

Algèbre Linéaire

Cette première partie permet d'introduire tous les outils d'algèbre linéaire qui nous serviront à l'étude des techniques d'analyse de données.

Nous reprenons les fondamentaux en Algèbre Linéaire depuis la présentation des espaces vectoriels jusqu'à l'étude des formes quadratiques et des fondamentaux en géométrie comme l'analyse des espaces euclidiens. L'idée de cette première partie est de présenter les outils pas à pas afin que le lecteur puisse suivre aisément le cheminement. Bien qu'elle soit abstraite, cette partie va se révéler nécessaire pour la présentation des techniques d'analyse de données qui seront présentées par la suite et nous essaierons d'appuyer tous les résultats fondamentaux et nécessaires à l'étude des données. En outre, pour rendre les notions plus abordables mais surtout moins abstraites pour le lecteur, nous illustrons les résultats par des graphiques ou encore des exemples simples.

Une première section sera dédiée aux généralités sur les espaces vectoriels et les applications linéaires. Elle introduit ce concept de base qu'est la notion d'espace des vecteurs qui sont des éléments essentiels et qui représenteront une donnée ou encore l'ensemble des valeurs d'une caractéristique, *i.e.* une *feature* par un ensemble d'individus, *i.e.* un échantillon. Nous présenterons aussi les notions de sous-espaces vectoriels ainsi que quelques sous-espaces particuliers et leurs propriétés. Enfin nous introduirons également les projecteurs qui joueront un rôle majeur dans la deuxième partie.

La deuxième section se consacre à l'étude des objets précédemment étudiés mais dans le cadre de la dimension finie, ce qui sera toujours le cas lorsque nous étudierons des données. Les notions de bases d'espace vectoriels et les propriétés des applications linéaires y seront alors présentées.

La troisième section aura un rôle plus important car elle abordera la représentation des applications linéaires dans une base, *i.e.* elle introduit la notion de matrice. Elle sera l'occasion de se familiariser avec des objets qui constituent une représentation naturelle et conventionnelle de nos données dans une forme dite *structurée*. On rappellera comment sont définis *trace*, *rang*, *déterminant* et *inverse* d'une matrice. On montrera comment représenter un vecteur ou une application linéaire dans une base mais aussi comment d'une représentation d'une base à une autre. Enfin, on terminera par quelques rappels sur la façon de calculer les différentes caractéristiques d'une matrice.

La quatrième section portera sur la résolution des systèmes linéaires de façon générale en présentant la méthode dite du *Pivot de Gauss* ainsi que les systèmes de *Cramer*.

La cinquième cherchera à savoir s'il est possible d'obtenir une représentation plus simple des applications linéaires, *i.e.* est-ce qu'il existe une base dans laquelle les matrices sont diagonales? Nous verrons que la réponse n'est pas si triviale et qu'obtenir une telle base n'est pas forcément évident. Nous verrons cependant, quand cela est le cas, comment obtenir une telle base et quelles conditions il est possible d'avoir une représentation simple d'un endomorphisme mais aussi dans quelles conditions. L'étude de la réduction des endomorphismes nous permettra, par exemple, de déterminer les directions de l'espace selon lesquelles nos données présentent une grande richesse en terme d'informations, *i.e.* une grande **variance**.

La sixième et dernière section consistera à introduire quelques éléments de nature géométrique. En effet, il est tout fait naturel de voir un vecteur ou une donnée comme un point dans un certain espace. De plus, l'objectif de l'analyse de données est de chercher à synthétiser l'information, c'est-à-dire de passer d'une représentation complexe des données (dans un espace à grande dimension) à une représentation dans un espace "simple", *i.e.* dans un espace de dimension 2 ou 3 que l'humain pourra alors facilement analyser. Pour ces besoins, nous devons présenter quelques concepts géométriques comme la notion de produit scalaire, de normes et surtout comment définir la projection d'un ensemble de points sur une droite ou sur un sous-espace.

Pour s'assurer que le lecteur a bien assimilé les différentes notions d'algèbre linéaire, nous proposons un petit questionnaire de vérification des connaissances à chaque fin de section.

1 Espaces vectoriels et applications linéaires

Dans cette première section, nous commençons par introduire les définitions d'espace vectoriel et de sous-espace vectoriels pour en étudier quelques propriétés (stabilité par intersection ou encore somme de sous-espaces vectoriels). Dans un deuxième temps nous introduisons les morphismes d'espaces vectoriels (applications linéaires, projecteurs, symétries) et regardons les propriétés des *noyaux* et *images* de telles applications.

1.1 Généralités sur les espaces vectoriels

Définition 1.1: Espace vectoriel

Soit E un ensemble muni d'une loi de composition interne notée "+" et d'une loi externe notée "." définie sur $\mathbb{K} \times E$ par

$$\begin{aligned}\mathbb{K} \times E &\rightarrow E \\ (\alpha, \mathbf{x}) &\rightarrow \alpha \cdot \mathbf{x}.\end{aligned}$$

On dit que E est un \mathbb{K} -espace vectoriel (*e.v.*) si les propriétés suivantes sont vérifiées

- i) $(E, +)$ est un groupe abélien (*i.e.* commutatif)
- ii) $\forall \mathbf{x} \in E, 1 \cdot \mathbf{x} = \mathbf{x}$.
- iii) $\forall (\alpha, \beta) \in \mathbb{K}^2, \forall \mathbf{x} \in E, (\alpha + \beta) \cdot \mathbf{x} = \alpha \cdot \mathbf{x} + \beta \cdot \mathbf{x}$.
- iv) $\forall \alpha \in \mathbb{K}, \forall \mathbf{x}, \mathbf{x}' \in E, \alpha \cdot (\mathbf{x} + \mathbf{x}') = \alpha \cdot \mathbf{x} + \alpha \cdot \mathbf{x}'$.
- v) $\forall (\alpha, \beta) \in \mathbb{K}^2, \forall \mathbf{x} \in E, \alpha \cdot (\beta \cdot \mathbf{x}) = (\alpha\beta) \cdot \mathbf{x}$.

Un groupe abélien (i) est défini par les points suivants :

- a) $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in E, (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$, c'est l'*associativité*.
- b) $\forall \mathbf{x}, \mathbf{y} \in E, \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$, c'est la *commutativité*.
- c) il existe un élément, noté $\mathbf{0}$ qui vérifie, pour tout $\mathbf{x} \in E : \mathbf{0} + \mathbf{x} = \mathbf{x}$.
- d) $\forall \mathbf{x} \in E$, il existe un élément $\mathbf{x}' \in E$ qui vérifie $\mathbf{x} + \mathbf{x}' = \mathbf{0}$. C'est l'existence d'un *inverse*, en général noté $-\mathbf{x}$.

Le plus souvent, dans la définition précédente, $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} . Dans la deuxième partie de ce cours (relative à l'analyse de données) on considérera uniquement $\mathbb{K} = \mathbb{R}$ donc des \mathbb{R} -espaces vectoriels, notés *r.e.v.*.

Le point *iii*) traduit la distributivité à droite de la loi multiplicative \cdot par rapport à la loi additive $+$ de \mathbb{K} . Le point *iv*) traduit la distributivité à gauche de la loi multiplicative \cdot par rapport à la loi additive $+$ de E . Enfin le point *v*) énonce l'associativité par rapport aux éléments de \mathbb{K} .

Les éléments d'un espace vectoriel sont appelés des **vecteurs**.

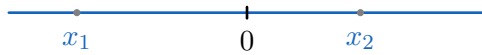
Dans la suite, et pour simplifier l'écriture, nous mettrons de côté le \cdot pour la multiplication externe.

Exemple 1.1. Nous avons tous déjà manipulé des espaces vectoriels sans s'en rendre compte. En effet, l'ensemble \mathbb{R} est un espace vectoriel sur \mathbb{R} . De même que l'ensemble des nombres complexes \mathbb{C} qui peut être vu comme un \mathbb{C} -espace vectoriel ou encore un \mathbb{R} -espace vectoriel. C'est-à-dire :

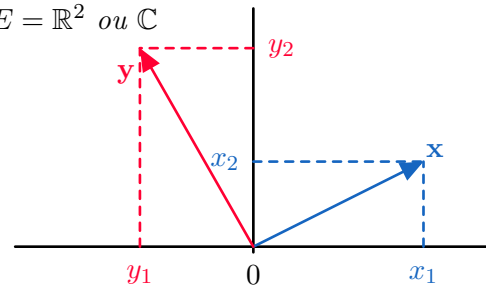
- i) $\forall a, a', b, b' \in \mathbb{R}, (a + ib) + (a' + ib') = (a + a') + i(b + b')$
- ii) $\forall \lambda, a, b \in \mathbb{R}, \lambda(a + ib) = \lambda a + i\lambda b$

Cela peut se voir facilement car un vecteur (ou un élément) de \mathbb{C} est totalement identifié en connaissant sa partie réelle et sa partie imaginaire. Ainsi les éléments de \mathbb{C} peuvent se voir comme des éléments de \mathbb{R}^2 .

$E = \mathbb{R}$



$E = \mathbb{R}^2$ ou \mathbb{C}



Plus généralement, nous avons le résultat suivant :

Proposition 1.1: \mathbb{K} -espace vectoriel

Pour tout $n \in \mathbb{N}$, l'espace \mathbb{K}^n est un \mathbb{K} -espace vectoriel pour les lois $+$ et \cdot définies par :

- $\forall ((\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n), (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)) \in (\mathbb{K}^n)^2 :$

$$(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) + (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) = (\mathbf{x}_1 + \mathbf{y}_1, \mathbf{x}_2 + \mathbf{y}_2, \dots, \mathbf{x}_n + \mathbf{y}_n),$$

- $\forall \alpha \in \mathbb{K}, \forall (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \in \mathbb{K}^n :$

$$(\alpha \mathbf{x}_1, \alpha \mathbf{x}_2, \dots, \alpha \mathbf{x}_n) = \alpha (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n).$$

Démonstration. La démonstration est laissée à titre d'exercice, il s'agit de vérifier les différents axiomes d'espace vectoriel. \square

Les espaces vectoriels sont des objets dans lesquels nous pouvons additionner des **vecteurs** ou encore les multiplier par un scalaire. La combinaison de ces éléments forme ce que l'on appelle une *combinaison linéaire*. Cette notion reviendra souvent lorsque l'on cherchera à réduire la dimension de l'espace de représentation de nos données. Nous serons alors amenés à exprimer une nouvelle information (typiquement un axe) comme une combinaison linéaire de vecteurs (axes existants).

Définition 1.2: Combinaison linéaire

Soit E un \mathbb{K} -espace vectoriel et soit $(\mathbf{x}_1, \dots, \mathbf{x}_n) \in E^n$. On appelle *combinaison linéaire* des vecteurs de la famille $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ toute somme du type :

$$\sum_{k=1}^n \alpha_k \mathbf{x}_k \in E, \quad \text{où } \alpha_1, \dots, \alpha_n \in \mathbb{K}.$$

Définition 1.3: Sous-espace vectoriel

Soit E un \mathbb{K} -espace vectoriel soit F une partie de E (on peut aussi dire, un sous-ensemble de E). F est un **sous-espace vectoriel** de E si F est lui-même doté d'une structure d'espace vectoriel pour les lois induites par les lois définies par E .

A priori, pour montrer que F est un sous-espace vectoriel de E , il nous faudrait vérifier la totalité des axiomes de la définition d'espace vectoriel, ce qui peut être relativement fastidieux.

En réalité il y a beaucoup plus simple pour montrer cela. Les lois (additive et multiplicative) sur F sont celles induites par les lois sur E , comme elles ont les bonnes propriétés sur E , elles les ont aussi sur F qui est un sous-ensemble de E .

Il reste donc à vérifier, comme le montre la proposition suivante, que le sous-ensemble F est non vide (sinon l'étude n'est pas intéressante ...) et qu'il est stable par combinaison linéaire.

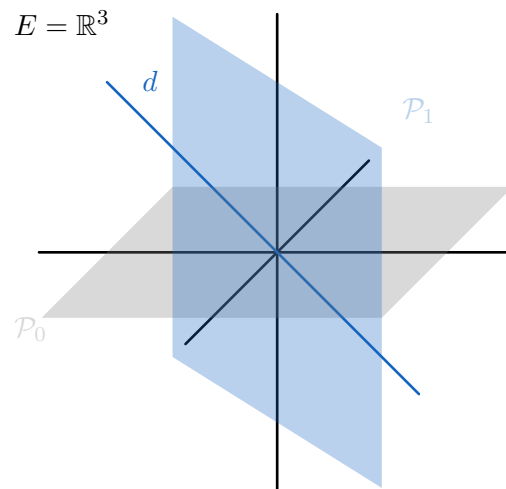
Proposition 1.2: Caractérisation sous-espace vectoriel

Soit E un \mathbb{K} -espace vectoriel et soit F un sous-ensemble de E . F est un **sous-espace** vectoriel de E si et seulement si les deux propriétés suivantes sont vraies :

- i) $F \neq \emptyset$, i.e. F est non vide,
- ii) $\forall(\mathbf{x}, \mathbf{x}') \in F^2, \forall(\alpha, \beta) \in \mathbb{R}^2, \alpha\mathbf{x} + \beta\mathbf{x}' \in F$, i.e. F est stable par combinaison linéaire.

Il est donc plus simple, pour montrer qu'un objet est un espace vectoriel, qu'il est un sous-espace vectoriel d'un espace vectoriel connu.

Exemple 1.2. Les exemples de sous-espaces vectoriels sont à nouveau nombreux, on peut par exemple prendre n'importe quel espace vectoriel qui se trouve dans un espace vectoriel donné. Prenons un exemple concret avec $E = \mathbb{R}^3$, alors toutes les droites et tous les plans qui passent par l'origine sont des sous-espaces vectoriels de E .



Dans cet exemple, les plans vectoriels $\mathcal{P}_0, \mathcal{P}_1$ et la droite vectorielle d sont des sous-espaces de $E = \mathbb{R}^3$.

Exemple 1.3. Considérons l'espace vectoriel $E = \mathbb{R}^2$. Regardons un autre exemple et montrons que l'ensemble F ci-dessous définit bien un sous-espace vectoriel :

$$F = \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 - x_2 = 0\}.$$

F est un sous-ensemble de \mathbb{R}^2 , ainsi, plutôt que de vérifier tous les axiomes d'espace vectoriel, on va plutôt montrer que c'est un sous-espace vectoriel de \mathbb{R}^2 . On va donc vérifier les deux points de la Proposition 1.2.

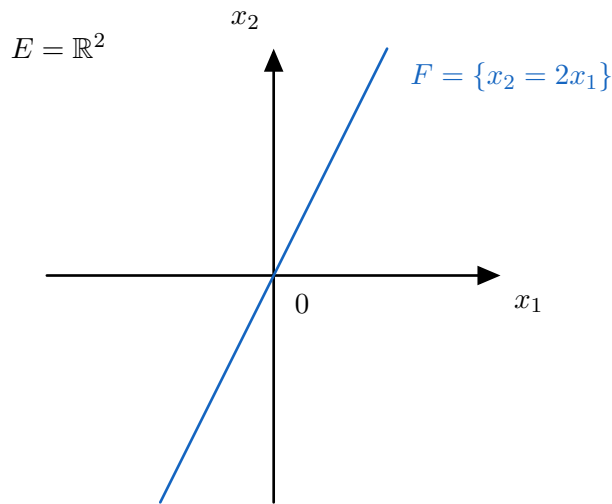
- i) On vérifie que le sous ensemble est non vide. Ce qui est le cas car le vecteur $\mathbf{x} = \mathbf{0} = (0, 0)$ appartient bien à F . En effet $2 \cdot 0 - 0 = 0$.
- ii) Il reste à vérifier la stabilité par combinaison linéaire, c'est-à-dire que si on considère deux vecteurs \mathbf{x}, \mathbf{x}' quelconques de F alors pour tout $\alpha, \beta \in \mathbb{R}$, le vecteur $\alpha\mathbf{x} + \beta\mathbf{x}' \in F$. Alors allons-y !

$$\alpha\mathbf{x} + \beta\mathbf{x}' = (\alpha x_1 + \beta x'_1, \alpha x_2 + \beta x'_2).$$

Regardons si les propriétés des éléments de F sont bien vérifiées pour ce vecteur, i.e. montrons que $2(\alpha x_1 + \beta x'_1) - \alpha x_2 - \beta x'_2 = 0$.

$$\begin{aligned} 2(\alpha x_1 + \beta x'_1) - \alpha x_2 - \beta x'_2 &= 2\alpha x_1 + 2\beta x'_1 - \alpha x_2 - \beta x'_2, \\ &= \alpha \left(\underbrace{2x_1 - x_2} \right) + \beta \left(\underbrace{2x'_1 - x'_2} \right), \\ &\quad \downarrow \mathbf{x} \in F \implies 2x_1 - x_2 = 0 \text{ et } \mathbf{x}' \in F \implies 2x'_1 - x'_2 = 0. \\ &= \alpha \cdot 0 + \beta \cdot 0. \\ &= 0. \end{aligned}$$

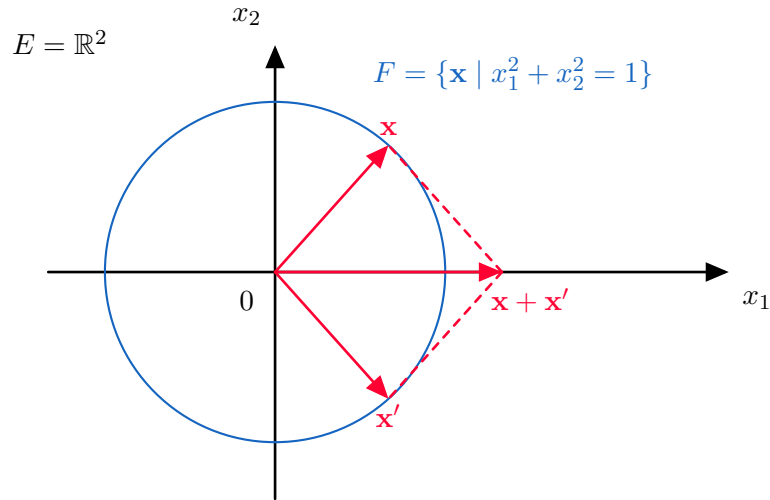
On peut représenter ce sous-espace F dans \mathbb{R}^2 comme suit :



Exemple 1.4. Considérons l'espace vectoriel $E = \mathbb{R}^2$. Regardons si l'ensemble F ci-dessous défini est un sous-espace vectoriel de \mathbb{R}^2

$$F = \{ \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1 \}.$$

F est un sous-ensemble de \mathbb{R}^2 . En revanche, on peut montrer qu'il ne s'agit pas d'un sous-espace vectoriel de \mathbb{R}^2 .



En effet, considérons deux vecteurs $\mathbf{x} = (x_1, x_2)$ et $\mathbf{x}' = (x'_1, x'_2)$ de F et $\alpha, \beta \in \mathbb{R}$ et montrons que le vecteur $\mathbf{z} = \alpha\mathbf{x} + \beta\mathbf{x}' \notin F$, i.e. il faut montrer que ce nouveau vecteur ne se trouve pas sur le cercle unité.

$$\begin{aligned} z_1^2 + z_2^2 &= (\alpha x_1 + \beta x'_1)^2 + (\alpha x_2 + \beta x'_2)^2, \\ &= \underbrace{\alpha^2(x_1^2 + x_2^2)} + \underbrace{\beta^2(x_1'^2 + x_2'^2)} + 2\alpha\beta(x_1x'_1 + x_2x'_2), \\ &\quad \downarrow \mathbf{x} \in F \text{ et } \mathbf{x}' \in F. \\ &= \alpha^2 + \beta^2 + 2\alpha\beta(x_1x'_1 + x_2x'_2) \end{aligned}$$

Si nous prenons par exemple $\alpha = \beta = 1$, comme illustré par le graphe ci-dessus, on a alors

$$z_1^2 + z_2^2 = 1 + 1 + 2(x_1x'_1 + x_2x'_2)$$

qui est en général différent de 1 comme on peut le voir.

Nous aurions également pu voir cela plus simplement en montrant que le vecteur nul n'appartient pas à cet ensemble F . En effet, un espace vectoriel contient forcément le vecteur nul.

1.2 Propriétés des sous-espaces vectoriels

Au cours de la précédente section, nous avons introduit la définition de sous-espace vectoriel. Dans cette section, nous nous concentrons sur les propriétés (intersection et

somme) des sous-espaces vectoriels d'un espace vectoriel E .

Le premier résultat ci-dessous montre la stabilité par intersection de la définition de sous-espace vectoriel.

Proposition 1.3: Sous-espace vectoriel et intersection

Soit E un \mathbb{K} espace vectoriel, et soient F et G deux sous-espaces vectoriels de E . Alors $F \cap G$ est un sous-espace vectoriel de E .

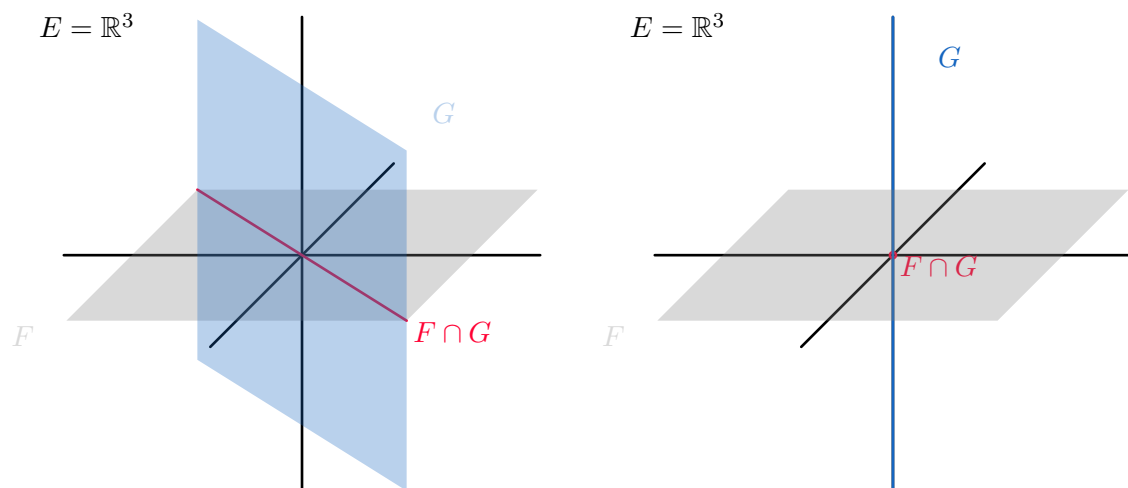
Démonstration. Il nous faut simplement vérifier que (i) l'intersection est non vide et (ii) la stabilité par combinaison linéaire.

- i) $\mathbf{0}_E \in F$ et $\mathbf{0}_E \in G$ car F et G sont des sous-espaces de E donc $\mathbf{0}_E \in F \cap G$, donc $F \cap G$ est non vide
- ii) Soit $(\mathbf{x}, \mathbf{y}) \in (F \cap G)^2$ et soit $(\alpha, \beta) \in \mathbb{K}^2$:
 $\mathbf{x} \in F$ et $\mathbf{y} \in F$ donc $\alpha\mathbf{x} + \beta\mathbf{y} \in F$ car F est un sous-espace de E .
 De même, $\mathbf{x} \in G$ et $\mathbf{y} \in G$ donc $\alpha\mathbf{x} + \beta\mathbf{y} \in G$ car G est un sous-espace de E
 On a donc bien $\alpha\mathbf{x} + \beta\mathbf{y} \in F \cap G$

□

Cette proposition se généralise très facilement à une famille de sous-espace de E . Ainsi si $(F_i)_{i \in I}$ est une famille de sous-espaces de E alors $\bigcap_{i \in I} F_i$ est un sous-espace vectoriel de E .

Exemple 1.5. Soit $E = \mathbb{R}^3$ un \mathbb{R} -espace vectoriel et soient F et G des sous-espaces de E . Considérons les deux graphes ci-dessous :



Sur la figure de gauche, on voit que l'intersection des sous-espaces F et G forment une droite vectorielle dans E qui est un sous-espace de E . De même sur la figure de droite, l'intersection de ces deux sous-espaces est réduite à $\mathbf{0}_E$, i.e. le sous-espace trivial.

On peut également générer un espace vectoriel à partir de la donnée d'un sous-ensemble d'un espace vectoriel, i.e. la donnée de quelques vecteurs d'un espace vectoriel E permet d'engendrer un sous-espace vectoriel. Plus formellement :

Définition 1.4: Espace engendré

Soit E un \mathbb{K} -espace vectoriel et soit A une partie non vide de E . On appelle **sous-espace vectoriel engendré par A** , noté $Vect(A)$ (ou encore $\langle A \rangle$) le plus petit sous-espace vectoriel (au sens de l'inclusion) de E qui contienne A .

On en déduit le résultat suivant :

Proposition 1.4: Sous-espace engendré

Soit E un \mathbb{K} -espace vectoriel et soit A une partie non vide de E . Posons $\mathcal{E} = \{F \text{ sous-espace de } E \mid A \subset F\}$ alors $Vect(A) = \bigcap_{F \in \mathcal{E}} F$.

Démonstration. On démontre le résultat en montrant les deux inclusions :

- i) par définition, tout sous-espace vectoriel F contenant A , on doit avoir $Vect(A) \subset F$. On a donc $Vect(A) \subset \bigcap_{F \in \mathcal{E}} F$.
- ii) réciproquement, $Vect(A)$ est un sous-espace vectoriel qui contient A , donc $Vect(A) \in \mathcal{E}$ ainsi $\bigcap_{F \in \mathcal{E}} F \subset Vect(A)$. On a donc bien $Vect(A) = \bigcap_{F \in \mathcal{E}} F$.

□

Ce résultat peut également être vu d'un point de vue un peu plus "pratique". En effet, l'espace engendré par A , $Vect(A)$ peut également se voir comme l'ensemble des vecteurs qui peuvent s'écrire comme une combinaison linéaire des éléments de A .

Nous avons étudié l'intersection entre des sous-espaces vectoriels, on peut également voir ce que représenterait la somme de deux sous-espaces vectoriels.

Définition 1.5: Somme de sous-espaces

Soit E un \mathbb{K} -espace vectoriel, et soient F et G deux sous-espaces de E . On appelle **somme de F et G** , notée $F + G$, le sous-espace vectoriel engendré par $F \cup G$. Par définition, nous avons donc $F + G = \text{Vect}(F \cup G)$.

Proposition 1.5: Décomposition dans une somme de sous-espaces

Soit E un \mathbb{K} -espace vectoriel, et soient F et G deux sous-espaces de E , alors $F + G = \{\mathbf{x} \in E \mid \exists(\mathbf{x}_1, \mathbf{x}_2) \in F \times G, \text{ tel que } \mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2\}$.

Ce résultat devient particulièrement intéressant si les espaces sont dits *supplémentaires*. Une notion que l'on retrouvera lorsque nous présenterons l'*Analyse en Composantes Principales* à la fin de ce document mais surtout lorsque l'on parlera de *projection*. Elle est beaucoup plus forte que la simple somme de deux sous-espaces.

Définition 1.6: Espaces supplémentaires

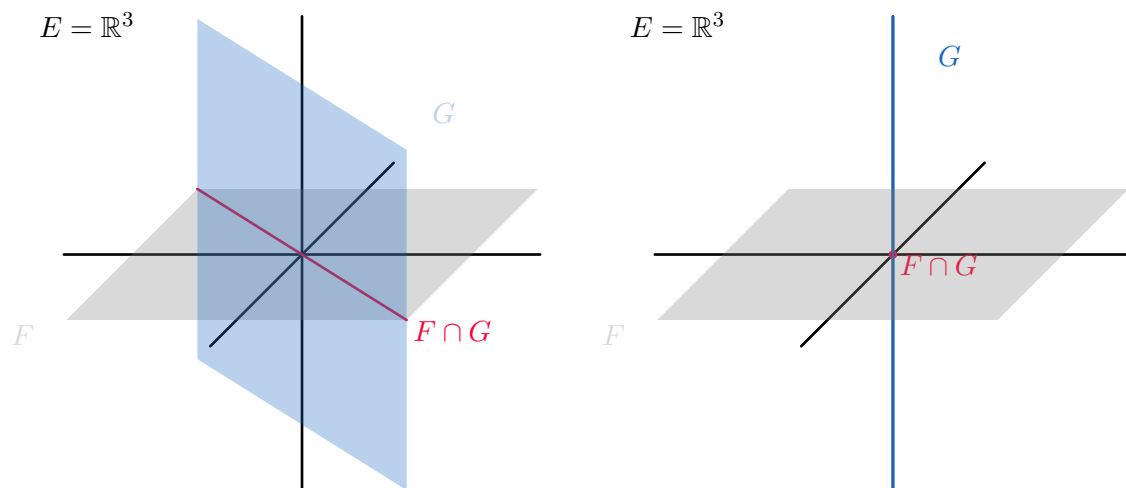
Soit E un \mathbb{K} -espace vectoriel, et soient F et G deux sous-espaces de E . F et G sont dits **supplémentaires** si les deux propriétés suivantes sont vérifiées :

i) $F \cap G = \{0_E\}$,

ii) $F + G = E$.

On notera alors $E = F \oplus G$.

Exemple 1.6. Soit $E = \mathbb{R}^3$ un \mathbb{R} -espace vectoriel et soient F et G des sous-espaces de E . Considérons les deux graphes ci-dessous :



Sur la figure de gauche, on a bien deux sous-espaces vectoriels de E mais ces derniers ne sont pas supplémentaires car l'intersection (représentée en rouge) n'est pas réduite à $\{0_E\}$ mais à une droite. La figure de droite montre deux sous-espaces supplémentaires dans \mathbb{R}^3 (une droite et un plan).

La notion de supplémentaire a une conséquence importante sur la décomposition d'éléments \mathbf{x} comme le montre le résultat suivant

Proposition 1.6: Somme Supplémentaire

Soit E un \mathbb{K} -espace vectoriel, et soient F et G deux sous-espaces de E , alors $E = F \oplus G$ si et seulement si $\forall \mathbf{x} \in E \mid \exists! (\mathbf{x}_1, \mathbf{x}_2) \in F \times G$, tel que $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$.

Démonstration. La démonstration est laissée en exercice. Il s'agit de montrer les deux implications :

- i) $E = F \oplus G \implies \forall \mathbf{x} \in E \mid \exists! (\mathbf{x}_1, \mathbf{x}_2) \in F \times G$, tel que $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$.
Il faut donc partir d'un élément de $\mathbf{x} \in E$ et montrer que sa décomposition est unique.
- ii) $\forall \mathbf{x} \in E \mid \exists! (\mathbf{x}_1, \mathbf{x}_2) \in F \times G$, tel que $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 \implies E = F \oplus G$.
Il s'agira ici de vérifier les deux points de la définition de sous-espaces supplémentaires.

□

1.3 Applications linéaires

Définition 1.7: Application linéaire

Soient E et E' deux \mathbb{K} -espaces vectoriels, et soit f une application de E dans E' . On dit que f est une application linéaire si :

$$\forall (\alpha, \beta) \in \mathbb{K}^2, \forall (\mathbf{x}, \mathbf{x}') \in E^2, f(\alpha \mathbf{x} + \beta \mathbf{x}') = \alpha f(\mathbf{x}) + \beta f(\mathbf{x}').$$

On peut résumer la définition d'application linéaire de la façon suivante : *l'image d'une combinaison linéaire par cette application est la combinaison linéaire des images de cette application*. Les applications linéaires sont également appelées **morphismes d'espaces vectoriels** (ou encore **homomorphisme** d'espaces vectoriels). Cette définition peut très facilement se généraliser (par récurrence) pour une combinaison linéaire de n vecteurs.

Proposition 1.7: Application linéaire

Soient E et E' deux \mathbb{K} -espaces vectoriels et f une application linéaire de E dans E' . Alors :

$$\forall (\mathbf{x}_1, \dots, \mathbf{x}_n) \in E^n, \forall \alpha_1, \dots, \alpha_n \in \mathbb{K}, \text{ on a } f\left(\sum_{k=1}^n \alpha_k \mathbf{x}_k\right) = \sum_{k=1}^n \alpha_k f(\mathbf{x}_k).$$

En dépit de cette définition abstraite, mais très simple ! Les applications linéaires vont se retrouver au coeur de toutes les analyses que nous effectuerons, notamment lorsque nous chercherons à **projeter** nos données dans un certain espace. La projection étant elle-même une application linéaire.

Exemple 1.7. On considère l'application f définie de \mathbb{R}^3 dans \mathbb{R}^2 par

$$f(x_1, x_2, x_3) = (x_1 + x_2, x_1 + x_3).$$

Cette application est bien linéaire. En effet, pour tout $\alpha, \beta \in \mathbb{R}$, nous avons :

$$\begin{aligned} f(\alpha \mathbf{x} + \beta \mathbf{y}) &= f(\alpha(x_1, x_2, x_3) + \beta(y_1, y_2, y_3)), \\ &\quad \downarrow \text{on applique la définition de } f \\ &= (\alpha(x_1 + x_2) + \beta(y_1 + y_2), \alpha(x_1 + x_3) + \beta(y_1 + y_3)), \\ &\quad \downarrow \text{on développe} \\ &= (\alpha x_1 + \alpha x_2 + \beta y_1 + \beta y_2, \alpha x_1 + \alpha x_3 + \beta y_1 + \beta y_3), \\ &\quad \downarrow \text{on décompose les éléments} \\ &= (\underbrace{\alpha x_1 + \alpha x_2, \alpha x_1 + \alpha x_3}_{\text{bleu}}) + (\underbrace{\beta y_1 + \beta y_2, \beta y_1 + \beta y_3}_{\text{rouge}}), \\ &\quad \downarrow \text{on factorise par } \alpha \text{ et } \beta \\ &= \alpha(x_1 + x_2, x_1 + x_3) + \beta(y_1 + y_2, y_1 + y_3), \\ &\quad \downarrow \text{on retrouve la définition de } f \\ &= \alpha f(\mathbf{x}) + \beta f(\mathbf{y}). \end{aligned}$$

Exemple 1.8. Considérons l'application f définie de \mathbb{R}^2 dans \mathbb{R} par

$$f(x_1, x_2) = x_1^2 + x_2^2.$$

On vérifie facilement que cette application n'est pas linéaire. En effet, pour tout $\alpha \in \mathbb{R}$ nous avons

$$f(\alpha \mathbf{x}) = f(\alpha(x_1, x_2)) = \alpha^2 x_1^2 + \alpha^2 x_2^2 = \alpha^2(x_1^2 + x_2^2) = \alpha^2 f(\mathbf{x}) \neq \alpha f(\mathbf{x}).$$

On peut aussi reconnaître facilement une application linéaire d'un simple coup d'oeil ! Il suffit que cette dernière s'écrive comme la somme de monômes de degré 1.

Une autre propriété intéressante des applications linéaires est qu'elles permettent de conserver la structure d'espace vectoriel, *i.e.* les images **directes** et **réci-proques** d'un espace vectoriel par une application linéaire sont des sous-espaces vectoriels. Plus précisément :

Proposition 1.8: Image application linéaire

Soient E et E' deux \mathbb{K} -espaces vectoriels et soit f une application linéaire de E dans E' .

- a) Soit F un sous espace vectoriel de E , alors $f(F)$ est un sous espace vectoriel de E' .
- b) Soit F' un sous-espace vectoriel de E' , alors $f^{-1}(F')$ est un sous-espace vectoriel de E .

Démonstration. On va démontrer les deux points de la proposition :

- a) F est un sous-espace vectoriel de E , il est donc non vide, de même que son image par f . On peut alors considérer \mathbf{y} et \mathbf{y}' des éléments de $f(F)$ ainsi que $\alpha \in \mathbb{K}$. Il existe des éléments \mathbf{x} et \mathbf{x}' de F tels que $\mathbf{y} = f(\mathbf{x})$ et $\mathbf{y}' = f(\mathbf{x}')$ et :

$$\begin{aligned}\mathbf{y} + \mathbf{y}' &= f(\mathbf{x}) + f(\mathbf{x}') = f(\mathbf{x} + \mathbf{x}'), \\ \alpha \mathbf{y} &= \alpha f(\mathbf{x}) = f(\alpha \mathbf{x}).\end{aligned}$$

Or F est un sous-espace vectoriel de E , donc $\mathbf{x} + \mathbf{x}'$ et $\alpha \mathbf{x}$ sont des éléments de F et $\mathbf{y} + \mathbf{y}'$ ainsi que $\alpha \mathbf{y}$ appartiennent à $f(F)$. Donc $f(F)$ est un sous-espace vectoriel de E' .

- b) Le vecteur nul de E' appartient au sous-espace F' et $f(0) = 0$, donc le vecteur nul de E est un élément de l'image réciproque de F' par f . Cette image réciproque est donc non vide.

Soient \mathbf{x}, \mathbf{x}' des vecteurs de $f^{-1}(F')$ et $\alpha \in \mathbb{K}$. Alors $f(\mathbf{x})$ et $f(\mathbf{x}')$ sont des éléments de F' , ainsi que $f(\mathbf{x}) + f(\mathbf{x}')$ et $\alpha f(\mathbf{x})$ puisque F' est un sous-espace vectoriel de E' . Les relations :

$$\begin{aligned}f(\mathbf{x} + \mathbf{x}') &= f(\mathbf{x}) + f(\mathbf{x}') \\ f(\alpha \mathbf{x}) &= \alpha f(\mathbf{x})\end{aligned}$$

entraînent que $\mathbf{x} + \mathbf{x}'$ et $\alpha \mathbf{x}$ sont des éléments de $f^{-1}(F')$.

□

Cette propriété nous assurera, plus tard, que les projections de nos données conservent la structure d'espaces vectoriels et toutes les propriétés qui en découlent !

Cependant, cette proposition interroge sur la nature des images des applications linéaires. Lors de l'étude d'une fonction réelle (de \mathbb{R} dans \mathbb{R}) il n'est pas rare d'étudier les *valeurs que peut prendre* une telle fonction mais aussi ses *zéros*, *i.e.* les valeurs pour lesquelles elle s'annule. On peut faire une étude semblable pour les applications linéaires où l'ensemble des valeurs prises par la fonction est appelé *image* et l'ensemble de ses zéros est appelé *noyau*.

Définition 1.8: Noyau et Image

Soient E et E' deux \mathbb{K} -espaces vectoriels et soit f une application linéaire de E dans E' . On appelle :

- *noyau de f* , noté $\text{Ker}(f)$, le sous-espace vectoriel de E défini par :

$$\text{Ker}(f) = \{\mathbf{x} \in E \mid f(\mathbf{x}) = \mathbf{0}_{E'}\}.$$

- *image de f* , notée $\text{Im}(f)$, le sous espace vectoriel de E' défini par :

$$\text{Im}(f) = \{\mathbf{y} \in E' \mid \exists \mathbf{x} \in E \text{ tel que } f(\mathbf{x}) = \mathbf{y}\}.$$

Cette définition comporte également un résultat dans son énoncé et qu'il nous faudrait montrer : $\text{Ker}(f)$ et $\text{Im}(f)$ sont des sous-espaces vectoriels de E et E' . Heureusement, c'est une conséquence immédiate de la Proposition 1.8.

On peut maintenant caractériser la nature des applications linéaires en fonction des espaces de "départ" et "d'arrivée" mais aussi en fonction de la nature du noyau et de l'image de ces applications. Mais avant cela faisons quelques rappels.

Définition 1.9: Nature d'une application

Soient X et Y deux ensembles non vides et f une application de X dans Y . Alors :

- f est dite **injective** si tout élément de Y admet **au plus** un antécédent dans X . Ce que l'on peut aussi formuler :

$$\forall x, x' \in X, x \neq x' \implies f(x) \neq f(x')$$

ou encore

$$\forall x, x' \in X, f(x) = f(x') \implies x = x'.$$

- f est dite **surjective** si tout élément Y admet *au moins* un antécédent dans X , *i.e.* :

$$\forall y \in Y, \exists x \in X \text{ tel que } f(x) = y.$$

- f est dite **bijjective** si elle est à la fois *injective et surjective*, ce que l'on peut écrire :

$$\forall y \in Y, \exists! x \in X \text{ tel que } f(x) = y.$$

Ce premier rappel nous permet d'introduire la définition suivante qui se présente plus comme un point de vocabulaire pour la suite de la présentation.

Définition 1.10: Morphismes et propriétés

Soient E et E' des \mathbb{K} -espaces vectoriels et soit f une application linéaire de E dans E' , i.e. $f \in \mathcal{L}(E, E')$: ensemble des applications linéaires de E dans E' . De plus :

- dans le cas où $E = E'$, f est appelée *endomorphisme de E* et on notera $\mathcal{L}(E)$ l'ensemble des endomorphismes de E ,
- dans le cas où $E' = \mathbb{K}$ ($= \mathbb{R}$ ou \mathbb{C}), f est appelée *forme linéaire sur E* , i.e. il s'agit d'une application qui à un vecteur associe un nombre.

De plus :

- si f est bijective, alors f est appelée *isomorphisme de E dans E'* ,
- enfin, si f est un isomorphisme de E dans lui même (donc un endomorphisme bijectif), alors f est appelée *automorphisme de E* et on note $\mathcal{GL}(E)$ l'ensemble des automorphismes de E . Dans ce cas, l'application f admet un inverse notée f^{-1} qui, pour tout $\mathbf{x} \in E$ vérifie

$$f(f^{-1}(\mathbf{x})) = f^{-1}(f(\mathbf{x})).$$

Des définitions 1.8 et 1.9 découle le résultat suivant sur le caractère injectif d'une application linéaire.

Proposition 1.9: Caractérisation de l'injectivité

Soient E et E' deux \mathbb{K} -espaces vectoriels et soit f une application linéaire de E dans E' , alors f est injective si et seulement si $\text{Ker}(f) = \{0\}$, i.e. si son noyau est réduit au vecteur nul.

Démonstration. Il s'agit de démontrer le résultat en montrant les deux implications :

- on commence par supposer que f est injective, i.e. si $f(\mathbf{x}) = f(\mathbf{x}')$ alors $\mathbf{x} = \mathbf{x}'$. On sait déjà que $\text{Ker}(f)$ est un sous-espace vectoriel de E il contient donc $\mathbf{0}$, donc $0 \subset \text{Ker}(f)$. Il nous reste à montrer $\text{Ker}(f) \subset \mathbf{0}$. Pour cela, on se donne un élément du noyau et on montre que celui est nécessairement nul :

$$\begin{aligned} \mathbf{x} - \mathbf{x}' \in \text{Ker}(f) &\implies f(\mathbf{x} - \mathbf{x}') = 0, \\ &\downarrow f \text{ est une application linéaire} \\ &\implies f(\mathbf{x}) - f(\mathbf{x}') = \mathbf{0}, \\ &\implies f(\mathbf{x}) = f(\mathbf{x}'), \end{aligned}$$

$$\begin{aligned}
&\downarrow f \text{ est injective par hypothèse} \\
&\implies \mathbf{x} = \mathbf{x}', \\
&\implies \mathbf{x} - \mathbf{x}' = \mathbf{0}.
\end{aligned}$$

Donc $\text{Ker}(f) \subset \mathbf{0}$.

- on suppose maintenant que $\text{Ker}(f) = \{\mathbf{0}\}$ et on doit montrer que f est injective. On se donne \mathbf{x}, \mathbf{x}' tels que $f(\mathbf{x}) = f(\mathbf{x}')$ et on doit montrer que $\mathbf{x} = \mathbf{x}'$.

$$\begin{aligned}
f(\mathbf{x}) &= f(\mathbf{x}'), \\
f(\mathbf{x}) - f(\mathbf{x}') &= \mathbf{0}, \\
&\downarrow \text{ car } f \text{ est une application linéaire.} \\
f(\mathbf{x} - \mathbf{x}') &= \mathbf{0}.
\end{aligned}$$

Cette dernière égalité implique que $\mathbf{x} - \mathbf{x}'$ est un élément du noyau de f qui est lui même réduit à $\mathbf{0}$, *i.e.* $\mathbf{x} - \mathbf{x}' = \mathbf{0}$. Nous avons donc $\mathbf{x} = \mathbf{x}'$ donc f est injective.

□

Cela constitue un premier résultat sur les applications linéaires. Nous verrons plus tard, lorsque nous étudierons *la dimension* des espaces vectoriels qu'il est également possible de caractériser les applications *surjectives* et donc les applications *bijectives*.

Exemple 1.9. Considérons l'application f de \mathbb{R}^2 dans \mathbb{R} définie par

$$f(\mathbf{x}) = f(x_1, x_2) = x_1.$$

L'application n'est pas injective, car les vecteurs $(0, 1)$ et $(0, 2)$ ont la même image par l'application f .

On peut en revanche montrer que l'application est surjective. En effet, quelque soit x_1 , le vecteur (x_1, t) est bien un antécédent de x_1 par l'application f , pour tout $t \in \mathbb{R}$.

Cette application n'est pas bijective.

Exemple 1.10. Considérons l'application f de \mathbb{R}^2 dans \mathbb{R} définie par

$$f(\mathbf{x}) = f(x_1, x_2) = (x_1, 2x_2).$$

Pour montrer que l'implication est injective, on considère deux vecteurs \mathbf{x} et \mathbf{y} tels que $f(\mathbf{x}) = f(\mathbf{y})$ et montrons que $\mathbf{x} = \mathbf{y}$.

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

↓ définition de f

$$(x_1, 2x_2) = (y_1, 2y_2),$$
$$(x_1 - y_1, 2(x_2 - y_2)) = 0.$$

Cette dernière égalité implique que $x_1 = y_1$ et $x_2 = y_2$, donc $\mathbf{x} = \mathbf{y}$.

Pour la surjectivité, on considère un vecteur $\mathbf{y} \in \mathbb{R}^2$ et on montre que le vecteur $\mathbf{x} = (y_1, \frac{1}{2}y_2)$ vérifie bien $f(\mathbf{x}) = \mathbf{y}$.

L'application linéaire étant surjective et injective, elle est donc bijective.

Nous allons maintenant étudier un type particulier d'applications linéaires : les *projecteurs* mais de façon "abstraite" pour le moment. Nous verrons cela de façon plus explicite lorsque nous aborderons les notions de base d'un espace vectoriel, le calcul matriciel et la géométrie euclidienne.

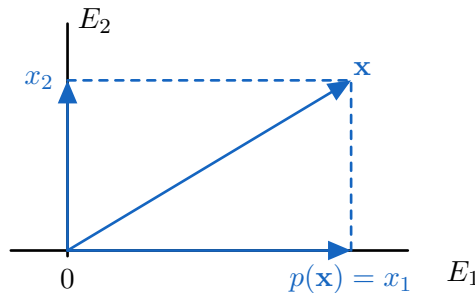
1.4 Projecteurs et symétries

Définition 1.11: Projecteur

Soit E un \mathbb{K} -espace vectoriel, et soient E_1 et E_2 des sous-espaces vectoriels supplémentaires de E . On appelle **projecteur sur E_1 parallèlement E_2** , l'application $p : E \rightarrow E$ qui, à un vecteur $\mathbf{x} \in E$ se décomposant comme $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ avec $(\mathbf{x}_1, \mathbf{x}_2) \in E_1 \times E_2$, associe le vecteur \mathbf{x}_1 .

Si la définition paraît abstraite, elle est en fait très simple à comprendre avec un exemple.

Exemple 1.11. Soit E un espace vectoriel, disons $E = \mathbb{R}^2$ et considérons le vecteur $\mathbf{x} = (x_1, x_2) = (5, 3) = (5, 0) + (0, 3) \in E$. Soit la projection p qui consiste à conserver uniquement la première composante de ce vecteur, alors $p(\mathbf{x}) = (5, 0)$.



On peut résumer cette définition par le schéma suivant

$$p : \quad E = E_1 \oplus E_2 \quad \rightarrow E.$$

$$\mathbf{x} = \underbrace{\mathbf{x}_1}_{\in E_1} + \underbrace{\mathbf{x}_2}_{\in E_2} \quad \mapsto \mathbf{x}_1$$

La proposition suivante montre que les projecteurs ont des images et noyaux qui sont faciles à déterminer à partir de leur définition.

Proposition 1.10: Propriété projecteur

Soit E un \mathbb{K} -espace vectoriel, et soient E_1 et E_2 deux sous-espaces supplémentaires de E . Soit p le projecteur sur E_1 parallèlement à E_2 . Alors p est un endomorphisme de E dont le noyau et l'image sont :

$$\text{Ker}(p) = E_2 \text{ et } \text{Im}(p) = E_1.$$

Démonstration. La démonstration de ce résultat n'est pas compliquée, il faut d'abord montrer que p est un endomorphisme E et ensuite montrer les égalités entre les différents ensembles. \square

Cette proposition énonce simplement qu'un projecteur projette sur son image parallèlement à son noyau. Dans ce cas image et noyau d'un projecteur sont des espaces supplémentaires et on peut écrire :

$$E = \text{Ker}(f) \oplus \text{Im}(f).$$

La proposition ci-dessous permet de caractériser les projecteurs parmi l'ensemble des applications linéaires.

Proposition 1.11: Idempotence

Soit E un \mathbb{K} -espace vectoriel et soit p un endomorphisme de E , alors p est un projecteur de E si et seulement si $p \circ p = p$.

La démonstration de ce résultat n'est pas bien compliquée mais est un peu fastidieuse, on va donc la laisser de côté.

Concentrons nous sur un dernier type d'applications : *les symétries* dont nous donnerons la définition et deux résultats. En effet, nous verrons qu'une symétrie peut se définir à l'aide d'une projection !

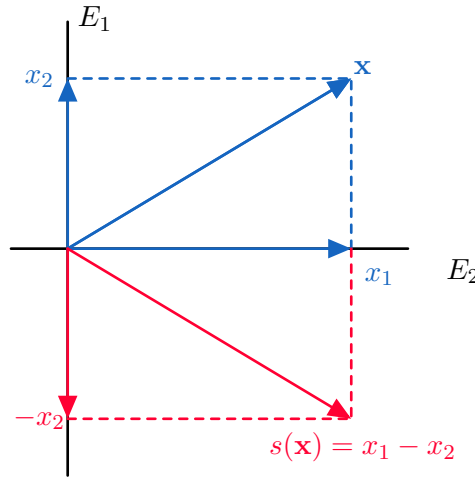
Définition 1.12: Symétrie

Soit E un \mathbb{K} -espace vectoriel, et soient E_1 et E_2 deux sous-espaces supplémentaires de E . On appelle alors **symétrie par rapport à E_1 parallèlement à E_2** l'application s de E dans E qui, à un vecteur $\mathbf{x} \in E$ se décomposant comme $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ où $(\mathbf{x}_1, \mathbf{x}_2) \in E_1 \times E_2$, associe le vecteur $\mathbf{x}_1 - \mathbf{x}_2$.

Exemple 1.12. Reprenons l'exemple précédent où E est l'espace vectoriel \mathbb{R}^2 et \mathbf{x} un vecteur quelconque de E . Prenons $\mathbf{x} = (5, 3)$, alors la symétrie s par rapport à E_1 parallèlement à E_2 du vecteur \mathbf{x} nous donne

$$s(\mathbf{x}) = x_1 - x_2 = (5, -3)$$

Dans cet exemple, cela revient à déterminer le symétrique de \mathbf{x} par rapport à l'axe des abscisses.



De la même façon que pour les projecteurs, on peut résumer la définition par le schéma suivant :

$$\begin{aligned} s : E = E_1 \oplus E_2 &\rightarrow E. \\ \mathbf{x} = \underbrace{\mathbf{x}_1}_{\in E_1} + \underbrace{\mathbf{x}_2}_{\in E_2} &\mapsto \mathbf{x}_1 - \mathbf{x}_2 \end{aligned}$$

Proposition 1.12: Propriété symétrie

Soit E un \mathbb{K} -espace vectoriel, et soient E_1 et E_2 des sous-espaces supplémentaires de E . Soit s la symétrie par rapport E_1 parallèlement à E_2 .

Alors s est un automorphisme de E et, de plus,

$$E_1 = \text{Ker}(s - \text{Id}_E) = E_2 \text{ et } E_2 = \text{Ker}(s + \text{Id}_E),$$

où Id_E est l'application *identité* de E , i.e. $\text{Id}_E : \mathbf{x} \mapsto \mathbf{x}$.

On peut à nouveau caractériser les symétries par rapport à l'ensemble des applications linéaires. Cette caractérisation ne doit pas surprendre car elle est très intuitive.

Proposition 1.13: Caractérisation symétrie

Soit E un \mathbb{K} -espace vectoriel et soit p un endomorphisme de E , alors s est une symétrie si et seulement si $s \circ s = \text{Id}_E$.

Un dernier résultat permet de faire le lien entre un projecteur et une symétrie, c'est pourquoi nous nous concentrerons par la suite sur l'étude des projecteurs.

Proposition 1.14: Lien projecteur - symétrie

Soit E un \mathbb{K} -espace vectoriel, et soient E_1 et E_2 deux sous-espaces supplémentaires de E . Soit p le projecteur sur E_1 parallèlement à E_2 et s la symétrie par rapport à E_1 parallèlement à E_2 .

Alors $s = 2p - \text{Id}_E$.

Démonstration. Il suffit de vérifier l'identité à l'aide des définitions de p et s .

Soit $\mathbf{x} \in E$, alors il se décompose de façon unique comme la somme d'un élément de E_1 avec un élément de E_2 (car les deux sous-espaces sont supplémentaires).

Les définitions de la symétrie s et de la projection p nous donnent $s(\mathbf{x}) = x_1 - x_2$ et $p(\mathbf{x}) = x_1$. On vérifie dans ce cas que

$$2p(\mathbf{x}) - \text{Id}_E(\mathbf{x}) = 2x_1 - (x_1 + x_2) = x_1 - x_2 = s(\mathbf{x}).$$

Nous avons bien $s = 2p - \text{Id}_E$. □

Vérifier ses connaissances

1. Soit un E un \mathbb{K} -espace vectoriel. Les propositions suivantes sont-elles vraies ou fausses ? Justifiez.
 - (a) E est un sous-espace vectoriel de E .
 - (b) \emptyset est un sous-espace vectoriel de E ?
 - (c) $\{0_E\}$ est un sous-espace vectoriel de E .
 - (d) Un sous-espace vectoriel de E non réduit à $\{0_E\}$ peut avoir un nombre fini d'éléments.
2. Soit un E un \mathbb{K} -espace vectoriel et F un sous-espace vectoriel de E . Les propositions suivantes sont-elles vraies ou fausses ? Justifiez.
 - (a) $0_E \in F$.
 - (b) Si $x \in F$, alors $2x \in F$.
 - (c) Si $x \in F$ et $y \in E \setminus F$ alors $x + y \in E \setminus F$.
 - (d) Si $x \in E \setminus F$ et $y \in E \setminus F$ alors $x + y \in E \setminus F$.
 - (e) Si $x \in E \setminus F$ alors $2x \in E \setminus F$.
 - (f) $E \setminus F$ est un sous-espace vectoriel de E .
 - (g) $F = E \iff E \subset F$.
 - (h) $F = \{0_E\} \iff \{0_E\} \subset F$
 - (i) $F = \{0_E\} \iff F \subset \{0_E\}$
 - (j) Si G est un sous-espace vectoriel de E inclus dans F , alors G est un sous-espace vectoriel de F .
3. Soient E et E' deux \mathbb{K} -espaces vectoriels et soit f une application linéaire de E dans E' . Montrer que $f(0_E) = 0_{E'}$.
4. Les espaces suivants sont-ils des \mathbb{R} -espaces vectoriels ?
 - (a) \mathbb{R}^2
 - (b) \mathbb{R}
 - (c) \mathbb{C}
 - (d) \mathbb{Q}
5. Soit E un \mathbb{K} -espace vectoriel et soient F et G deux sous-espaces supplémentaires de E .
 - (a) Soit $x \in E \setminus F$, peut-on affirmer que $x \in G$?

- (b) Soit p le projecteur sur F parallèlement à G .
Montrer que :

$$\forall \mathbf{x} \in E, \mathbf{x} \in F \iff p(\mathbf{x}) = \mathbf{x},$$

et montrer que

$$\forall \mathbf{x} \in E, \mathbf{x} \in G \iff p(\mathbf{x}) = \mathbf{0}.$$

2 Espaces vectoriels de dimension finie

La section précédente a permis d'introduire des notions fondamentales sur les espaces vectoriels et les applications linéaires. Cette section est dédiée plus spécifiquement aux espaces de dimension finie dont \mathbb{R}^2 et \mathbb{R}^3 sont les plus connus. Nous allons introduire la notion de base d'espace vectoriel et verrons que tout élément d'un espace vectoriel peut se représenter de façon unique par des éléments de cette base (pensez aux coordonnées lorsque vous faisiez de la géométrie dans le plan ou dans l'espace).

Dans toute cette section E désignera un \mathbb{K} -espace vectoriel.

2.1 Familles libres, familles génératrices et bases

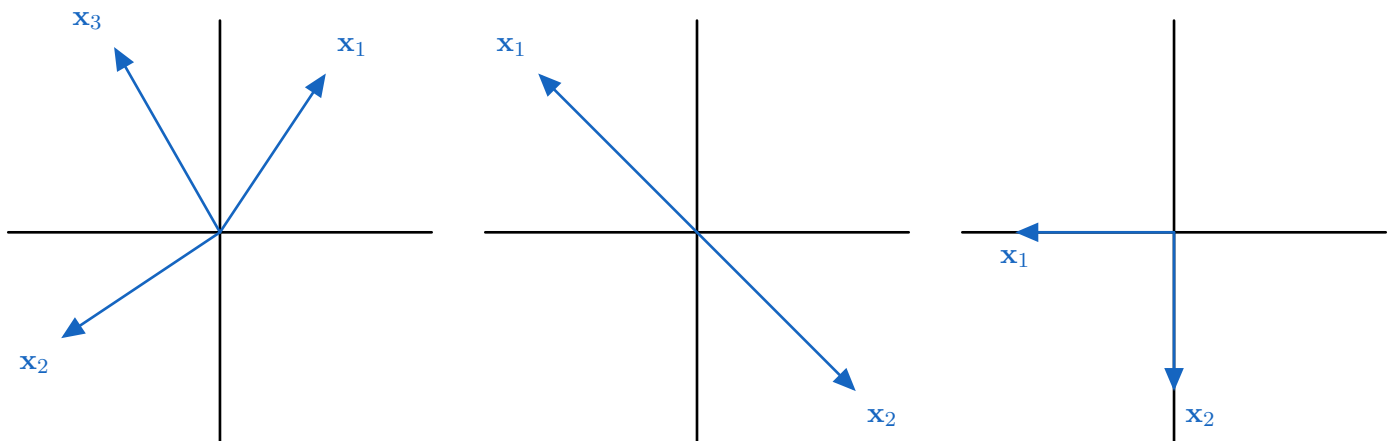
Définition 2.1: Familles libres, liées

Soit $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ une famille de vecteurs de E . Cette famille est dite :

- **libre** si $\forall (\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n, \sum_{k=1}^n \lambda_k \mathbf{x}_k = 0 \implies \lambda_k = 0 \forall k \in \llbracket 1, n \rrbracket$.
- **liée** si elle n'est pas libre, c'est-à-dire s'il existe $(\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n \setminus \{0\}^n$ tel que :

$$(\lambda_1, \dots, \lambda_n) \neq (0, \dots, 0) \quad \text{et} \quad \sum_{k=1}^n \lambda_k \mathbf{x}_k = 0_E.$$

Exemple 2.1. Soit $E = \mathbb{R}^2$ un espace vectoriel et considérons les graphes ci-dessous avec des familles de vecteurs $(\mathbf{x}_1, \mathbf{x}_2)$ ou $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$.



Dans le premier cas, nous avons trois vecteurs distincts dans un espace à deux dimensions, on peut donc écrire l'un des vecteurs comme combinaison linéaire des deux autres, la **famille n'est donc pas libre, elle est liée**. Dans le deuxième cas, les vecteurs

sont colinéaires et on a clairement $\mathbf{x}_1 = -\mathbf{x}_2$, à nouveau, **la famille est liée**. Dans le dernier cas, **la famille est bien libre** car les deux vecteurs sont orthogonaux. On verra plus tard que le troisième exemple montre l'exemple d'une famille de vecteurs que l'on appellera **base** de l'espace vectoriel.

Exemple 2.2. On considère la famille de vecteurs $\mathbf{v}_1 = (1, 2, 1, 0)$ et $\mathbf{v}_2 = (-1, 0, -1, 1)$ de \mathbb{R}^4 .

On montre que cette famille est bien une famille libre. En effet, considérons $\lambda_1, \lambda_2 \in \mathbb{R}$, alors

$$\lambda_1 \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Si on concentre sur la deuxième et la dernière équation, on trouve directement que $\lambda_1 = \lambda_2 = 0$.

Exemple 2.3. On considère la famille de vecteurs $\mathbf{v}_1 = (1, -1)$, $\mathbf{v}_2 = (3, -2)$ et $\mathbf{v}_3 = (0, 1)$ de \mathbb{R}^2 .

Cette famille de vecteurs forme une famille liée. En effet, on vérifie aisément que l'on a

$$\mathbf{v}_2 = 3\mathbf{v}_1 + \mathbf{v}_3.$$

Nous verrons plus tard que nous aurions pu directement écrire que cette famille de vecteurs est liée sans effectuer de calculs. Pourquoi selon vous ?

Exemple 2.4. On considère la famille de vecteurs $\mathbf{v}_1 = (1, 1)$ et $\mathbf{v}_2 = (1, -1)$. On montre qu'il s'agit d'une famille libre. En effet, considérons $\lambda_1, \lambda_2 \in \mathbb{R}$, alors

$$\lambda_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

On doit alors résoudre le système formé par les deux équations $\lambda_1 + \lambda_2 = 0$ et $\lambda_1 - \lambda_2 = 0$.

La somme de ces deux équations implique $\lambda_1 = 0$ et la différence de ces deux équations implique $\lambda_2 = 0$.

Proposition 2.1: Caractérisation famille liée

Soit $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ une famille d'au moins deux vecteurs de E , cette famille est *liée* si et seulement si l'un de ses vecteurs peut s'écrire comme combinaison linéaire de ses autres vecteurs.

Démonstration. Il s'agit de démontrer l'équivalence par implications réciproques.

- Supposons que la famille $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ est liée. Dans ce cas, il existe $(\lambda_1, \dots, \lambda_n) \in (\mathbb{K} \setminus 0)^n$ tel que $\sum_{k=1}^n \lambda_k \mathbf{x}_k = \mathbf{0}_E$.
En particulier, il existe un indice $j \in \llbracket 1, n \rrbracket$ tel que $\lambda_j \neq 0$, d'où

$$\lambda_j \mathbf{x}_j + \sum_{k \neq j=1}^n \lambda_k \mathbf{x}_k = \mathbf{0}_E \implies \mathbf{x}_j = - \sum_{k \neq j=1}^n \frac{\lambda_k}{\lambda_j} \mathbf{x}_k.$$

On a donc bien exprimé un vecteur de la famille comme combinaison linéaire de ses autres vecteurs.

- Supposons maintenant qu'un vecteur de $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ puisse s'écrire comme combinaison linéaire des autres vecteurs de cette même famille. Soit alors $j \in \llbracket 1, n \rrbracket$ tel que \mathbf{x}_j puisse s'écrire comme combinaison linéaire des autres vecteurs :

$$\exists (\lambda_1, \dots, \lambda_{j-1}, \lambda_{j+1}, \dots, \lambda_n) \in \mathbb{K}^{n-1} \quad \text{tel que} \quad \mathbf{x}_j = \sum_{k \neq j=1}^n \lambda_k \mathbf{x}_k.$$

En posant alors $\lambda_j = -1$, on aura l'existence de $(\lambda_1, \dots, \lambda_n) \in (\mathbb{K} \setminus 0)^n$ et $\sum_{k=1}^n \lambda_k \mathbf{x}_k = \mathbf{0}_E$.

La famille $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ est donc liée.

□

Remarquons, par contraposée, qu'une famille est dite *libre* si aucun de ses vecteurs ne peut s'écrire comme combinaison linéaire de ses autres vecteurs (on parle d'*indépendance linéaire*).

Le résultat précédent implique directement le fait que :

- toute sur-famille d'une famille liée est liée,
- toute sous-famille d'une famille libre est libre.

Lors de la présentation générale des espaces vectoriels, nous avons également abordé la notion d'espace engendré par des sous-espaces vectoriels. La notion de famille *génératrice* présentée ci-dessous va permettre d'introduire la notion plus générale d'espaces vectoriels générés par des *vecteurs*.

Définition 2.2: Famille génératrice

Soit $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ une famille de vecteurs de E , cette famille est *génératrice* si $\text{Vect}(\mathbf{x}_1, \dots, \mathbf{x}_n) = E$. C'est-à-dire, si tout élément \mathbf{x} de E peut s'écrire comme une combinaison linéaire des éléments de cette famille :

$$\forall \mathbf{x} \in E, \exists (\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n, \text{ tel que } \mathbf{x} = \sum_{k=1}^n \lambda_k \mathbf{x}_k.$$

On en déduit que toute sur-famille d'une famille génératrice de E est une famille génératrice de E . Les graphes de gauche et de droite de l'Exemple 2.1 constituent deux exemples de famille génératrice de $E = \mathbb{R}^2$.

Exemple 2.5. Reprenons la famille de vecteurs $\mathbf{v}_1 = (1, 1)$ et $\mathbf{v}_2 = (1, -1)$ et montrons qu'il s'agit d'une famille génératrice de \mathbb{R}^2 .

Pour cela, on considère un vecteur quelconque $\mathbf{x} = (x_1, x_2)$ de \mathbb{R}^2 et on va montrer que ce vecteur peut s'écrire comme une combinaison des vecteurs de \mathbf{v}_1 et \mathbf{v}_2 , i.e.

$$\mathbf{x} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2.$$

On cherche donc les valeurs α_1, α_2 telles que

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \alpha_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

On montre qu'une solution est donnée par

$$\alpha_1 = \frac{x_1 + x_2}{2} \quad \text{et} \quad \alpha_2 = \frac{x_1 - x_2}{2}.$$

Exemple 2.6. Considérons la famille de vecteurs $\mathbf{v}_1 = (1, 1)$, $\mathbf{v}_2 = (1, 0)$ et $\mathbf{v}_3 = (-1, 1)$.

On peut facilement voir qu'il s'agit d'une famille de vecteurs liée. En effet, on peut remarquer que l'on a le lien suivant entre les éléments de cette famille de vecteurs

$$\mathbf{v}_1 = 2\mathbf{v}_2 + \mathbf{v}_3.$$

On peut aussi montrer qu'il s'agit d'une famille génératrice de \mathbb{R}^2 . Considérons à nouveau un vecteur quelconque $\mathbf{x} = (x_1, x_2)$ de \mathbb{R}^2 et montrons qu'il peut

s'exprimer comme une combinaison linéaire des vecteurs de notre famille.

Par exemple :

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (x_1 - x_2) \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

ou encore

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1 + x_2) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Nous aurions également pu écrire ce vecteur de la façon suivante :

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{x_2 + x_1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{x_2 - x_1}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

On a donc trois façons d'exprimer notre vecteur \mathbf{x} à l'aide de vecteurs de cette famille !

Ce dernier exemple introduit quelques ambiguïtés sur la notion de représentation d'un vecteur.

Le fait de pouvoir représenter un vecteur par des combinaisons différentes d'une même famille se révèle peu intéressant dans les faits, en plus d'être source de problèmes lorsque l'on cherchera à *représenter* une application linéaire.

C'est pourquoi il est nécessaire de demander une propriété supplémentaire à nos familles de vecteurs génératrices : l'objet de notre définition suivante, qui porte sur la notion d'une base d'un espace vectoriel. On va voir qu'elle regroupe à la fois les caractéristiques d'une famille libre et génératrice.

Définition 2.3: Base

On appelle base de E toute famille d'éléments de E à la fois libre et génératrice.

Exemple 2.7. Dans \mathbb{R}^2 , les vecteurs $\mathbf{e}_1 = (1, 0)$ et $\mathbf{e}_2 = (0, 1)$ forment une famille libre et génératrice de \mathbb{R}^2 . En effet :

$$\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 = \mathbf{0} \iff (\lambda_1, \lambda_2) = \mathbf{0},$$

i.e. si seulement si $\lambda_1 = \lambda_2 = 0$.

De plus tout vecteur $\mathbf{x} = (x_1, x_2)$ de \mathbb{R}^2 peut s'écrire comme combinaison linéaire des vecteurs \mathbf{e}_1 et \mathbf{e}_2

$$\mathbf{x} = (x_1, x_2) = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2.$$

On verra un peu plus loin que cette famille de vecteurs forme ce que l'on appelle la base canonique de \mathbb{R}^2 .

Exemple 2.8. La famille de vecteurs $\mathbf{v}_1 = (-1, 1)$ et $\mathbf{v}_2 = (1, 1)$ forme une base de \mathbb{R}^2 . En effet, elle forme une famille libre

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 = \mathbf{0} \iff (-\lambda_1 + \lambda_2, \lambda_1 + \lambda_2) = \mathbf{0}.$$

En résolvant le système linéaire :

$$\begin{cases} -\lambda_1 + \lambda_2 &= 0 \\ \lambda_1 + \lambda_2 &= 0 \end{cases} \iff \begin{cases} -\lambda_1 + \lambda_2 &= 0 \\ 2\lambda_2 &= 0 \end{cases}$$

On en déduit $\lambda_1 = \lambda_2 = 0$.

De plus, on montre que cette famille est bien une famille génératrice de \mathbb{R}^2 . Considérons $\mathbf{x} = (x_1, x_2)$ est un vecteur de \mathbb{R}^2 , alors il peut s'écrire :

$$\mathbf{x} = \frac{x_2 - x_1}{2} \mathbf{v}_1 + \frac{x_1 + x_2}{2} \mathbf{v}_2.$$

Tout vecteur de \mathbb{R}^2 peut donc s'écrire comme une combinaison linéaire des vecteurs \mathbf{v}_1 et \mathbf{v}_2 .

Proposition 2.2: Caractérisation base

Soit $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ une famille de vecteurs de E , cette famille est une base de E si et seulement si

$$\forall \mathbf{x} \in E, \exists! (\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n, \text{ tel que } \mathbf{x} = \sum_{k=1}^n \lambda_k \mathbf{x}_k.$$

Remarquez bien la différence avec la définition de famille génératrice ! L'écriture de tout élément de \mathbf{x} de E s'exprime de façon **unique** comme élément d'une base de E , alors que cette décomposition n'est pas unique pour une famille génératrice.

Démonstration. On démontre à nouveau l'équivalence par implications réciproques.

- Supposons que $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ est une base de E et soit \mathbf{x} un élément de E . Comme la famille est génératrice :

$$\exists(\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n, \quad \text{tel que} \quad \mathbf{x} = \sum_{k=1}^n \lambda_k \mathbf{x}_k.$$

Il nous faut montrer que cette décomposition est unique, on suppose donc qu'il en existe une deuxième, *i.e.* que $\mathbf{x} = \sum_{k=1}^n \mu_k \mathbf{x}_k$ où $(\mu_1, \dots, \mu_n) \in \mathbb{K}^n$. On a donc

$$\sum_{k=1}^n (\mu_k - \lambda_k) \mathbf{x}_k = \mathbf{0}.$$

Or la famille est libre donc $\mu_k - \lambda_k = 0 \forall k \in \llbracket 1, n \rrbracket$.

Cela montre l'unicité de la décomposition.

- Supposons que tout élément \mathbf{x} de E admette une unique décomposition, *i.e.* $\forall \mathbf{x} \in E, \exists!(\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n$ tel que $\mathbf{x} = \sum_{k=1}^n \lambda_k \mathbf{x}_k$, ce qui montre que la famille est bien génératrice. Il reste à montrer qu'il s'agit d'une famille libre. Pour cela, soit $(\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n$ tel que $\sum_{k=1}^n \lambda_k \mathbf{x}_k = \mathbf{0}_E$. En particulier, nous avons $\mathbf{0}_E = \sum_{k=1}^n 0 \cdot \mathbf{x}_k$. L'unicité de la décomposition assure que $\forall k \in \llbracket 1, n \rrbracket, \lambda_k = 0$.

La famille est donc libre et génératrice, c'est donc une base de E .

□

Un autre point à souligner au regard de cette démonstration est que l'existence de la décomposition des éléments de E provient du caractère *générateur* de la famille et l'unicité vient du caractère *libre*. A partir de cette remarque on comprend que seule la famille du graphe de droite de l'Exemple 2.1 constitue une base de $E = \mathbb{R}^2$.

Exemple 2.9 (Base canonique de \mathbb{R}^n). Soit $n \in \mathbb{N}^*$, l'espace \mathbb{R}^n est l'ensemble des n -uplets (x_1, \dots, x_n) où pour tout $j \in \llbracket 1, n \rrbracket, \mathbf{x}_j \in \mathbb{R}$. On définit alors la famille $(\mathbf{e}_1, \dots, \mathbf{e}_n) \in (\mathbb{R}^n)^n$, par :

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, 0, \dots, 0, 0), \\ \mathbf{e}_2 &= (0, 1, 0, \dots, 0, 0), \\ &\vdots \\ \mathbf{e}_n &= (0, 0, 0, \dots, 0, 1). \end{aligned}$$

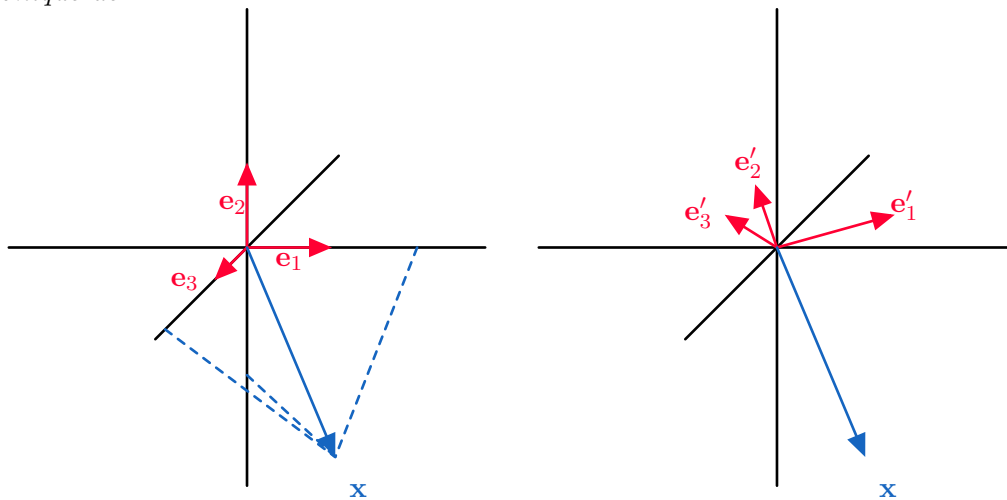
Tout vecteur $\mathbf{x} = (x_1, \dots, x_n)$ se décompose alors de manière unique comme combinaison linéaire des vecteurs de la base canonique :

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \dots + x_n\mathbf{e}_n.$$

Les éléments (x_1, \dots, x_n) sont donc appelés *coordonnées* du vecteur \mathbf{x} dans la base $(\mathbf{e}_1, \dots, \mathbf{e}_n)$, cette base est appelée *base canonique* de \mathbb{R}^n .

Définir un vecteur dans \mathbb{R}^n revient donc à déterminer exactement ses coordonnées dans la base $(\mathbf{e}_1, \dots, \mathbf{e}_n)$.

Dans l'exemple de gauche, nous avons $\mathbf{x} = 2\mathbf{e}_1 - 1.5\mathbf{e}_2 + 2.5\mathbf{e}_3$ et dans l'exemple de droite, les coordonnées sont plus complexes à déterminer car nous n'utilisons pas la base canonique de \mathbb{R}^n .



Il convient aussi de noter qu'un espace vectoriel E peut admettre plusieurs bases, voire un nombre infini de bases. Elles auront cependant toutes une caractéristique commune comme nous allons le voir juste après.

2.2 Espace vectoriel de dimension finie

L'exemple présenté dans la section précédente permet de mettre en avant une propriété importante de certains espaces vectoriels comme le fait d'admettre une **base finie**, *i.e.* un nombre fini de vecteurs de base. Ce nombre de vecteurs permet de définir ce que l'on appelle la **dimension d'un espace vectoriel**.

Avant cela, nous pouvons commencer par caractériser les espaces vectoriels de dimension finie avec la définition ci-dessous :

Définition 2.4: Dimension finie

Un espace vectoriel E est dit de *dimension finie* s'il admet une famille génératrice finie.

Un point très important mais que nous ne démontrerons pas ici, est que **tout espace vectoriel de dimension finie admet une base (qui est finie)**.

Théorème 2.1: Définition de la dimension

Soit E un espace vectoriel sur \mathbb{K} de dimension finie, alors toutes les bases de E ont le même nombre fini d'éléments, ce nombre d'éléments est appelé *dimension* de l'espace vectoriel E , il est noté $\dim(E)$.

Exemple 2.10. L'exemple présenté précédemment permet de montrer que les espaces vectoriels \mathbb{R}^n , $n \in \mathbb{N}^*$ sont des espaces vectoriels de dimension n . Nous avons également exhibé une base pour de tels espaces vectoriels comme la base canonique.

Dans le cas où E est réduit à 0_E , on dira que E est un espace vectoriel de dimension 0 (réduit à un point). Cela revient à considérer que la famille vide \emptyset est la seule base de E .

Un autre résultat important concerne la donnée d'une famille libre d'un espace vectoriel E .

Théorème 2.2: Théorème de la base incomplète

Soit E un espace vectoriel sur \mathbb{K} de dimension finie, alors toute famille libre de E peut être complétée en une base de E .

Démonstration. Pour démontrer ce théorème, nous aurons besoin du lemme suivant :

Lemme 2.1. Soit E un espace vectoriel de dimension finie et soit $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ une famille génératrice de E , alors toute sous-famille libre de $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ peut être complétée en une base de E à l'aide d'éléments de $(\mathbf{x}_1, \dots, \mathbf{x}_n)$.

Soit $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ une famille libre de E . L'espace E étant de dimension finie, il admet une base $(\mathbf{e}_1, \dots, \mathbf{e}_d)$.

La famille $(\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{e}_1, \dots, \mathbf{e}_d)$ est donc une sur-famille de $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ donc elle est génératrice. La famille $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ est alors une sous-famille de $(\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{e}_1, \dots, \mathbf{e}_d)$. Ainsi, d'après le lemme, il est possible de compléter la famille $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ en une base de E à l'aide d'éléments de $(\mathbf{e}_1, \dots, \mathbf{e}_d)$. \square

Exemple 2.11. Considérons la famille de vecteurs $\mathbf{v}_1 = (1, -1, 0)$ et $\mathbf{v}_2 = (2, 5, 0)$.

On montre que ces deux vecteurs forment une famille libre de \mathbb{R}^3 mais ne forment pas une famille génératrice de \mathbb{R}^3 .

On peut cependant lui ajouter le vecteur $\mathbf{v}_3 = (0, 0, 1)$ afin que cette famille libre devienne une base de \mathbb{R}^3 .

Le résultat suivant est très utile en pratique lorsque l'on souhaite montrer qu'une famille constitue une base d'un espace vectoriel.

Théorème 2.3: Propriétés familles

Soit E un espace vectoriel de dimension finie $n \in \mathbb{N}^*$. Soit $(\mathbf{x}_1, \dots, \mathbf{x}_n) \in E^n$, alors les propositions suivantes sont équivalentes :

- i) $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ est une famille libre de E ,
- ii) $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ est une famille génératrice de E ,
- iii) $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ est une base de E .

Tout comme nous venons de le faire pour les espaces vectoriels, il est également possible de caractériser la dimension d'un sous-espace vectoriel.

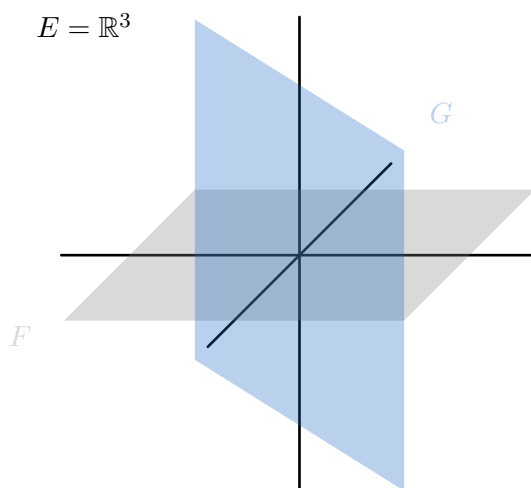
Proposition 2.3: Dimension sous-espace vectoriel

Soit E un espace vectoriel de dimension finie, alors tout sous espace vectoriel F de E est aussi de dimension finie et $\dim(F) \leq \dim(E)$.

De plus, on a l'équivalence $E = F \iff \dim(E) = \dim(F)$

Ce résultat peut aussi se résumer de la façon suivante : les sous-espaces vectoriels stricts de E ont une dimension strictement inférieure à celle de E .

Exemple 2.12. Considérons l'espace $E = \mathbb{R}^3$, alors les sous-espaces F et G représentés ci-dessous sont des sous espaces vectoriels de \mathbb{R}^3 de dimension 2.



On verra plus tard que ces espaces F et G sont aussi appelés hyperplans de E .

Exemple 2.13. Dans le cas où E est réduit à 0_E , on dira que E est un espace vectoriel de dimension 0 (réduit à un point). Cela revient à considérer que la famille vide \emptyset est la seule base de E .

Exemple 2.14. L'espace $\mathcal{M}_2(\mathbb{R})$ des matrices carrées d'ordre 2 est un espace de dimension 4 dont une base est donnée par les matrices suivantes :

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \text{ et } \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Cette base est la base canonique de $\mathcal{M}_2(\mathbb{R})$

Proposition 2.4: Sous-espace supplémentaire

Soit E un espace vectoriel de dimension finie, tout sous-espace F de E admet un supplémentaire dans E .

Nous avons déjà vu la notion d'espaces supplémentaires dans la section précédente.

Proposition 2.5: Recollement de bases

Soit E un espace vectoriel de dimension finie, et soient F et G des sous-espaces supplémentaires de E . Soit également $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ et $(\mathbf{y}_1, \dots, \mathbf{y}_q)$ des bases F et G respectivement.

Alors la famille $(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}_1, \dots, \mathbf{y}_q)$ est une base de E .

Cette proposition admet un corollaire immédiat liant la dimension de E à celles de deux sous-espaces supplémentaires de E .

Corollaire 2.1: Dimension espaces supplémentaires

Soit E un espace vectoriel de dimension finie et soient F et G des sous-espaces supplémentaires de E , alors $\dim(E) = \dim(F) + \dim(G)$.

C'est un résultat particulier pour des espaces supplémentaires d'un espace vectoriel E . Le résultat suivant généralise le corollaire précédent à deux sous-espaces vectoriels quelconques de E .

Proposition 2.6: Dimension somme de sous-espaces quelconques

Soit E un espace vectoriel de dimension finie et soient F et G des sous-espaces vectoriels de E , alors

$$\dim(F + G) = \dim(F) + \dim(G) - \dim(F \cap G).$$

Retournons maintenant sur nos familles de vecteurs et regardons une caractéristique de ces dernières, appelée **rang**, qui permet de faire le lien entre l'espace engendré par la famille de vecteurs et la dimension de l'espace engendré par cette même famille.

Définition 2.5: Rang d'une famille

Soit E un espace vectoriel et $\mathcal{F} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ une famille de vecteurs de E . On appelle *rang* de la famille \mathcal{F} , noté $\text{rang}(\mathcal{F})$ ou encore $\text{rg}(\mathcal{F})$, la dimension de l'espace vectoriel engendré par \mathcal{F} , *i.e.*

$$\text{rg}(\mathbf{x}_1, \dots, \mathbf{x}_n) = \dim(\text{Vect}(\mathbf{x}_1, \dots, \mathbf{x}_n)).$$

Cette notion de rang est fondamentale et reviendra lorsque nous reviendrons sur les applications linéaires et leurs représentations avec le *théorème du rang* mais aussi lorsque nous introduirons les matrices dans la prochaine section.

Proposition 2.7: Propriétés rang d'une famille

Soit E un espace vectoriel et soit $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ une famille de vecteurs de E , alors :

- $\text{rg}(\mathbf{x}_1, \dots, \mathbf{x}_n) \leq n$,
- $\text{rg}(\mathbf{x}_1, \dots, \mathbf{x}_n) = n$ si et seulement si $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ est libre.

2.3 Représentation d'une application linéaire dans une base

On considère un espace vectoriel E de dimension finie qui admet une base $(\mathbf{x}_1, \dots, \mathbf{x}_n)$, F un autre espace vectoriel et considérons $f \in \mathcal{L}(E, F)$ une application linéaire. Soit \mathbf{x} un élément de E alors \mathbf{x} peut s'écrire de façon unique $\mathbf{x} = \sum_{k=1}^n \lambda_k \mathbf{x}_k$. Si on applique la fonction f au vecteur \mathbf{x} , on a

$$f(\mathbf{x}) = \sum_{k=1}^n \lambda_k f(\mathbf{x}_k).$$

L'application linéaire est donc entièrement déterminée par les images $(f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))$ des vecteurs de bases $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ de E .

Réciproquement, pour $(\mathbf{y}_1, \dots, \mathbf{y}_n)$ une famille de vecteurs de F , on définit l'application f de E dans F par : $\forall \mathbf{x} \in E$ tel que $\mathbf{x} = \sum_{k=1}^n \lambda_k \mathbf{x}_k$, $f(\mathbf{x}) = \sum_{k=1}^n \lambda_k \mathbf{y}_k$. On a alors, $\forall k \in \llbracket 1, n \rrbracket$, $f(\mathbf{x}_k) = \mathbf{y}_k$. De plus f est une application linéaire de E dans F .

Ainsi, on peut définir une application linéaire f en fixant les images $\mathbf{y}_1, \dots, \mathbf{y}_n$ des vecteurs de la base $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ de E .

On peut résumer tout cela par la proposition suivante :

Proposition 2.8: Application et image d'une base

Soit E un espace vectoriel qui admet une base $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ et soit F un espace vectoriel.

Alors $\forall (\mathbf{y}_1, \dots, \mathbf{y}_n) \in F^n$, $\exists ! f \in \mathcal{L}(E, F)$ tel que $\forall k \in \llbracket 1, n \rrbracket$, $f(\mathbf{x}_k) = \mathbf{y}_k$.

Cette proposition, ainsi que ce qui la précède, permet

- étant donnée une application linéaire $f \in \mathcal{L}(E, F)$, il suffit de connaître l'image des vecteurs de bases de E par cette application pour la déterminer entièrement.
- ou encore, si l'on connaît n vecteurs de F , il est possible de construire une (unique) application linéaire f de E dans F telle que pour tout $k \in \llbracket 1, n \rrbracket$, $f(\mathbf{x}_k) = \mathbf{y}_k$.

Après avoir vu comment construire une telle application, on va voir que l'on est aussi capable de caractériser les propriétés des morphismes en fonction de l'image des vecteurs de bases.

Proposition 2.9: Caractérisation morphisme et famille

Soit E un espace vectoriel de dimension finie $n \in \mathbb{N}^*$ muni d'une base $(\mathbf{e}_1, \dots, \mathbf{e}_n)$. Soit E' un espace vectoriel de f une application linéaire de E dans E' , alors :

- f est injective si et seulement si la famille $(f(\mathbf{e}_1), \dots, f(\mathbf{e}_n))$ est une famille libre de E' .
- f est surjective si et seulement si $(f(\mathbf{e}_1), \dots, f(\mathbf{e}_n))$ est une famille génératrice de E' .
- f est bijective si et seulement si $(f(\mathbf{e}_1), \dots, f(\mathbf{e}_n))$ est une base de E'

En particulier, si f est bijective on dit que les espaces E et E' sont isomorphes. Nous avons, en outre, le résultat suivant :

Proposition 2.10: Caractérisation isomorphismes

Soient E et E' deux espaces vectoriels de dimensions finies, alors E et E' sont isomorphes si et seulement si $\dim(E) = \dim(E')$.

On en déduit le résultat suivant :

Corollaire 2.2: Isomorphismes de \mathbb{K}^n

Soit $n \in \mathbb{N}^*$, tout \mathbb{K} -espace vectoriel de dimension n est isomorphe à \mathbb{K}^n .

Revenons maintenant à la dimension de rang. Nous avons précédemment introduit la notion de rang d'une famille de vecteurs. On va voir qu'il est également possible de définir le rang d'une application linéaire en regardant le rang de l'image des vecteurs d'une base de E par cette application linéaire.

Définition 2.6: Rang application linéaire

Soient E et E' deux \mathbb{K} -espaces vectoriels de dimensions finies, et $f \in \mathcal{L}(E, E')$. On appelle *rang de f* , noté $\text{rang}(f)$ ou $\text{rg}(f)$, la dimension de l'espace $\text{Im}(f)$.

Nous pouvons même relier le rang d'une application linéaire $f \in \mathcal{L}(E, E')$ à la dimension de l'espace de départ E et la dimension du noyau, c'est ce que l'on verra avec le **théorème du rang**. Mais avant cela, nous avons besoin du résultat suivant :

Lemme 2.1: Supplémentaire et isomorphisme

Soient E et E' deux \mathbb{K} -espaces vectoriels de dimensions finies et soit $f \in \mathcal{L}(E, E')$. Soit F un supplémentaire de $\text{Ker}(f)$ dans E (rappelons que $\text{Ker}(f)$ est un sous-espace de E), alors la restriction de f à F , notée $f|_F$, réalise un isomorphisme de F sur $\text{Im}(f)$.

Démonstration. Il nous faut montrer que l'application $f|_F \in \mathcal{L}(F, \text{Im}(f))$ est bijective vu que la linéarité de l'application est déjà acquise par hypothèse.

Pour cela, on rappelle que l'application est définie par :

$$\begin{aligned} f|_F : \quad F &\rightarrow \text{Im}(f), \\ \mathbf{x} &\mapsto f|_F(\mathbf{x}) = f(\mathbf{x}). \end{aligned}$$

- **Injectivité :**

Soit $\mathbf{x} \in F$ alors $\mathbf{x} \in \text{Ker}(f|_F)$ si et seulement si \mathbf{x} appartient à F et au noyau

de l'application f , *i.e.* si et seulement si \mathbf{x} appartient à l'intersection de ces deux ensembles. Or F et $\text{Ker}(f)$ sont supplémentaires dans E , donc $\mathbf{x} = 0$.

On a donc $\text{Ker}(f|_F) = \{0\}$, donc $f|_F$ est injective.

• **Surjectivité :**

Soit $\mathbf{y} \in \text{Im}(f)$, alors il existe $\mathbf{x} \in E$ tel que $f(\mathbf{x}) = \mathbf{y}$. Or nous avons $E = F \oplus \text{Ker}(f)$ donc on il existe $(\mathbf{x}_1, \mathbf{x}_2) \in F \times \text{Ker}(f)$ tel que $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$. Or $\mathbf{y} = f(\mathbf{x})$, ce qui entraîne

$$\mathbf{y} = f(\mathbf{x}_1) + \underbrace{f(\mathbf{x}_2)}_{=0} = f(\mathbf{x}_1) = f|_F(\mathbf{x}_1).$$

La dernière égalité vient du fait que $\mathbf{x}_1 \in F$, donc $\mathbf{y} \in \text{Im}(f|_F)$.

Ce qui montre que $\text{Im}(f) = \text{Im}(f|_F)$ et que $f|_F$ est surjective.

$f|_F$ est donc injective et surjective, par conséquent elle est bijective et est donc bien un isomorphisme de F dans $f|_F$. □

Théorème 2.4: Théorème du rang

Soient E et E' deux \mathbb{K} -espaces vectoriels de dimensions finies et soit $f \in \mathcal{L}(E, E')$, alors $\dim(\text{Ker}(f)) + \dim(\text{Im}(f)) = \dim(E)$.

Démonstration. Soit F un supplémentaire de $\text{Ker}(f)$ dans E . Nous avons vu plus tôt que pour deux sous-espaces vectoriels supplémentaires dans un espace vectoriel donné voient la somme de leur dimension égale à la somme de l'espace totale, *i.e.*

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

De plus, le lemme précédent nous montre que la restriction $f|_F$ réalise un isomorphisme de F dans $\text{Im}(f)$, les deux espaces sont donc isomorphes et, par conséquent, ont la même dimension :

$$\dim(\text{Im}(f)) = \dim(F).$$

On a donc bien $\dim(\text{Im}(f)) + \dim(F) = \dim(E)$. □

Remarque : Il est également possible d'écrire le résultat de ce théorème de la façon suivante :

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

Ce résultat va permettre de très facilement caractériser les morphismes définies entre deux espaces de même dimension.

Proposition 2.11: Morphismes et dimension

Soient E et E' deux espaces vectoriels de dimensions finies tels que $\dim(E) = \dim(E')$. Soit $f \in \mathcal{L}(E, E')$, il y a alors équivalence entre les trois propriétés suivantes :

- i) f est injective
- ii) f est surjective
- iii) f est bijective

Démonstration. Ce résultat découle immédiatement du théorème du rang. \square

Ce résultat qui n'a rien de surprenant une fois que l'on connaît le théorème du rang, se révèle cependant très utile lorsque l'on cherche à montrer qu'une application linéaire entre deux espaces de même dimension réalise un isomorphisme. En effet, il suffira simplement de montrer qu'elle est injective **ou** surjective.

2.4 Retour sur sous-espaces vectoriels

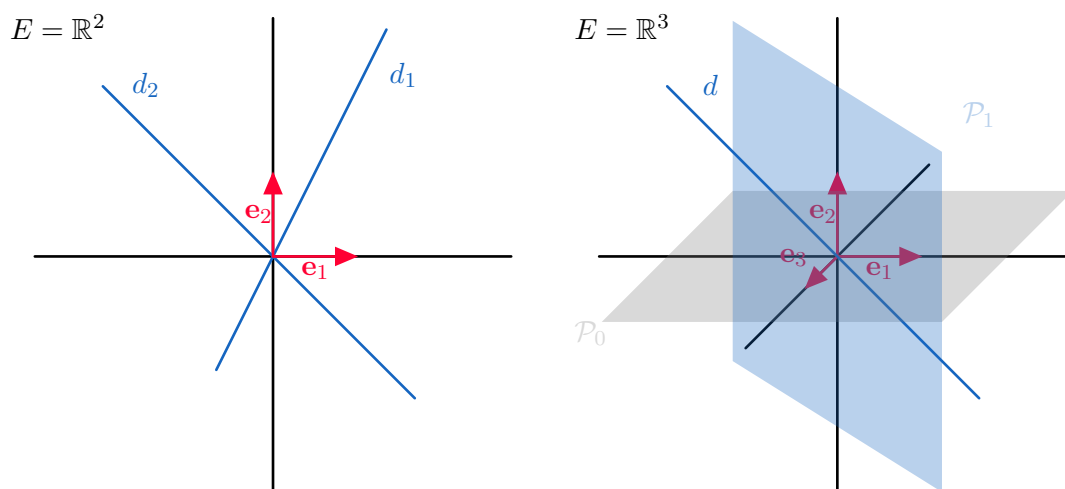
Cette dernière partie introduit la notion **d'hyperplan** qui est un cas particulier de sous-espaces vectoriels. Plus précisément :

Définition 2.7: Hyperplan

Soit E un espace vectoriel de dimension finie $n \geq 2$. On appelle *hyperplan* de E tout sous-espace vectoriel de E de dimension $n - 1$.

En tant que sous-espace vectoriel, les hyperplans doivent donc nécessairement contenir le vecteur nul, *i.e.* le vecteur 0_E .

Exemple 2.15. *L'exemple le plus simple de sous-espace vectoriel que l'on puisse imaginer est une droite dans un plan. En effet, un plan est un espace de dimension 2 et la droite est un espace de dimension 1. Les droites d_1 et d_2 sont des sous espaces vectoriels de \mathbb{R}^2 .*



Même chose avec les plans dans $E = \mathbb{R}^3$, les plans \mathcal{P}_0 et \mathcal{P}_1 sont des hyperplans de \mathbb{R}^3 ; en revanche la droite d n'est pas un hyperplan de \mathbb{R}^3 .

Nous verrons également qu'il existe aussi, dans un contexte géométrique, ce que l'on appelle des *hyperplans affines*.

Proposition 2.12: Supplémentaire hyperplan

Soit E un espace vectoriel de dimension finie $n \geq 2$ et soit H un hyperplan de E . Alors pour tout vecteur a de E , si $a \notin H$, on a $H \oplus \text{Vect}(a) = E$, i.e. l'espace vectoriel engendré par a (qui est une droite) est un supplémentaire de H dans E .

Proposition 2.13: Forme linéaire et noyau

Soit E un espace vectoriel de dimension finie $n \geq 2$ et soit H un sous-espace vectoriel de E , alors H est un hyperplan de E si et seulement si H est le noyau d'une forme linéaire non nulle sur E .

Démonstration. Il s'agit de démontrer que $\dim(H) = n-1 \iff (\exists \varphi \in \mathcal{L}(E, K) \text{ telle que } \varphi \neq 0 \text{ et } H = \text{Ker}(\varphi))$. On effectue la démonstration par implications réciproques à nouveau.

- **Supposons que $\dim(H) = n-1$:**

on peut donc définir une base $(\mathbf{e}_1, \dots, \mathbf{e}_{n-1})$ de H que l'on peut compléter pour obtenir une base $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ de E . Soit $\varphi \in \mathcal{L}(E, \mathbb{K})$ définie par $\forall k \in \llbracket 1, n-1 \rrbracket, \varphi(\mathbf{e}_k) = 0$ et $\varphi(\mathbf{e}_n) = 1$.

On a donc montré l'existence d'une forme linéaire non nulle sur E , il reste à vérifier que $\text{Ker}(\varphi) = H$.

Soit $\mathbf{x} \in E$ on a alors :

$$\mathbf{x} = \sum_{k=1}^n \lambda_k \mathbf{e}_k, \quad \text{où } (\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n.$$

Alors $\mathbf{x} \in \text{Ker}(\varphi) \iff \sum_{k=1}^n \lambda_k \mathbf{e}_k = 0 \iff \lambda_n = 0 \iff \mathbf{x} \in H$.

- **On suppose maintenant qu'il existe une forme linéaire φ non nulle telle que son noyau est égal à H .**

En appliquant le théorème du rang, on a

$$\dim(\text{Ker}(\varphi)) + \text{rg}(\varphi) = n.$$

Or l'image de φ est un sous-espace vectoriel de \mathbb{K} , donc $\text{Im}(\varphi) = \text{rg}(\varphi) \leq \dim(\mathbb{K}) = 1$. Or φ est une forme linéaire non nulle, donc l'image de φ est nulle, donc de dimension strictement positive. Par conséquent, $\dim(\text{Im}(\varphi)) = 1$, on en conclut que $\dim(\text{Ker}(\varphi)) = \dim(H) = n - 1$.

□

Vérifier ses connaissances


1. Soit E un \mathbb{K} -espace vectoriel. Les propositions suivantes sont-elles vraies ou fausses ? Justifiez.
 - (a) Soit \mathbf{x} un élément non nul de E , alors la famille (\mathbf{x}) est libre.
 - (b) Soient $\mathbf{x} \neq \mathbf{y} \in E$, alors la famille (\mathbf{x}, \mathbf{y}) est libre.
 - (c) Si une famille de vecteurs de E contient le vecteur nul, elle est liée.
 - (d) Une famille libre de vecteurs $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ est une base de $F = Vect((\mathbf{x}_1, \dots, \mathbf{x}_n))$.
2. Soit E un \mathbb{K} -espace vectoriel non nul. Les propositions A et B suivantes sont-elles équivalentes ? Est-ce que l'un implique l'autre ?
 - (a) A : (\mathbf{x}, \mathbf{y}) est une famille liée de E .
B : $\exists \lambda \in \mathbb{K}$ tel que $\mathbf{y} = \lambda \mathbf{x}$.
 - (b) A : $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ est une famille libre de E .
B : si $(\lambda_1, \dots, \lambda_n) = (0, \dots, 0)$ alors :

$$\sum_{k=1}^n \lambda_k \mathbf{x}_k = \mathbf{0}_E$$

- (c) A : E est un espace vectoriel de dimension 1.
B : $\forall \mathbf{a} \in E$ tel que $\mathbf{a} \neq \mathbf{0}$, alors $E = Vect(\mathbf{a})$.
3. Soit E un \mathbb{K} -espace vectoriel de dimension finie $n \in \mathbb{N}^*$. Les propositions suivantes sont-elles vraies ou fausses ?
 - (a) Toute famille libre de E compte au plus n vecteurs.
 - (b) Toute famille génératrice de E compte au plus n vecteurs.
 - (c) Une famille qui compte moins de n vecteurs est libre.
 - (d) Une famille qui compte plus de n vecteurs est liée.
 - (e) Une famille qui compte plus de n vecteurs est génératrice.
 - (f) Si F est un sous-espace vectoriel de E , une famille libre de vecteurs de F est une famille libre de vecteurs de E .
 - (g) Si F est un sous-espace vectoriel de E , une famille génératrice de F est une famille génératrice de E .
 4. Soit E un \mathbb{K} -espace vectoriel de dimension $n \in \mathbb{N}^*$ et soient F et G des sous-espaces vectoriels de E . Les propositions suivantes sont-elles équivalentes ? L'une implique-t-elle l'autre ?
A : $F \oplus G = E$
B : $F \cap G = \{\mathbf{0}\}$ et $\dim F + \dim G = n$.

5. Soit E un \mathbb{K} -espace vectoriel de dimension $n \in \mathbb{N}^*$ muni d'une base $(\mathbf{e}_1, \dots, \mathbf{e}_n)$. Soit E' un \mathbb{K} -espace vectoriel de dimension $n' \in \mathbb{N}^*$ et $f \in \mathcal{L}(E, E')$. Les propositions suivantes sont-elles vraies ou fausses ?
- (a) $(f(\mathbf{e}_1), \dots, f(\mathbf{e}_n))$ est une base de $Im(f)$
 - (b) $rg(f) = rg((f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)))$
 - (c) f est injective si et seulement si $dim(Ker(f)) = 0$
 - (d) f est surjective si et seulement si $dim(Im(f)) = n$
 - (e) f est surjective si et seulement si $dim(Im(f)) = n'$
6. Soit E un \mathbb{K} -espace vectoriel et soit φ une forme linéaire sur E . Montrer que φ est nulle ou surjective.
7. Soit E un \mathbb{K} -espace vectoriel de dimension finie $n \in \mathbb{N}^*$ et soit $f \in \mathcal{L}(E)$. Les propositions suivantes sont-elles vraies ou fausses ?
- (a) Si $rg(f) = n$, alors $f \in \mathcal{GL}(E)$.
 - (b) $Ker(f)$ et $Im(f)$ sont supplémentaires dans E .
 - (c) Si $Ker(f) \cap Im(f) = \{\mathbf{0}\}$, alors $Ker(f)$ et $Im(f)$ sont supplémentaires dans E .

Les variables et vecteurs sur

Cette deuxième section, bien que très théorique et introductive, va nous permettre d'introduire les éléments de bases pour l'Analyse de Données en regardant comment créer un vecteur sur  ainsi que quelques calculs élémentaires (vu que l'on a abordé les espaces vectoriels de dimension finie!)

Regardons déjà comment créer une variable **x** qui contient la valeur 3.

```
# Assignment d'une valeur à une variable
x = 3
# ou
x <- 3
```

Si l'on souhaite afficher la valeur de **x** (mais cela reste vrai pour tout objet défini), il suffit de taper le nom de la variable dans la console.

Les opérations élémentaires sur les nombres réels se font de la façon suivante

```
x = 3
y = 2
# Addition, soustraction, multiplication, division
x + y

## [1] 5

x - y

## [1] 1

x*y

## [1] 6

x/y

## [1] 1.5
```

Nous pourrions également stocker ces résultats dans des variables afin de les garder en mémoire, ce qui peut se révéler très utile lorsque l'on souhaite réutiliser les résultats plus tard.

Il est également possible de créer une variable qui contient plusieurs valeurs, par exemple un vecteur de \mathbb{R}^3 , $\mathbf{x} = (1, 2, 3)$.

```
# Création d'un vecteur de dimension 3
x = c(1,2,3)
# ou encore
x = c(1:3)
```

Remarquez la deuxième notation qui permet de créer facilement un vecteur de dimension 3 en indiquant la première et dernière valeur prise par le vecteur. Par exemple la commande ci-dessous :

```
# Création d'un vecteur de dimension 6
x = c(5:10)
```

va créer un vecteur de taille 6 contenant les valeurs 5, 6, 7, 8, 9 et 10. Nous avons tout ce dont nous avons besoin pour illustrer les opérations internes et externes sur un espace vectoriel. Pour cela, on va considérer deux vecteurs \mathbf{x} et \mathbf{y} ainsi qu'un scalaire (*i.e.* un nombre réel) α .

```
# Vecteurs et scalaire
x = c(1,2,3)
y = c(1,2,1)
alpha = 3

# Addition interne
x + y

## [1] 2 4 4

# produit externe
alpha*x

## [1] 3 6 9
```

Il est également possible de définir la différence entre deux vecteurs ou encore la division d'un vecteur par un scalaire (si ce dernier est non nul!). Notons que l'addition interne revient à faire une addition composante par composante, le produit externe consiste quant à lui à multiplier toutes les composantes du vecteur par le scalaire α . Nous pouvons également définir un produit entre deux vecteurs (ce que l'on appelle un *produit de Hadamard*²) sur \mathbb{R} ou l'addition entre un scalaire et un vecteur.

2. Du nom du mathématicien français **Jacques Salomon Hadamard** (1865-1963) ayant contribué dans différents domaines des mathématiques comme la théorie des nombres ou la géométrie différentielle. Ce produit est normalement défini pour des matrices dont les vecteurs sont un cas particulier.

```

# Produit entre deux vecteurs
x * y

## [1] 1 4 3

# produit externe
x + alpha

## [1] 4 5 6

```

Dans le premier cas, cela revient à faire le produit composante par composante. Nous verrons que cela sera pareil lorsque l'on manipulera des matrices, *i.e.*

$$x * y = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} * \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 \\ 2 \cdot 2 \\ 3 \cdot 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix}.$$

Ensuite, lorsque l'on ajoute un nombre réel à un vecteur, cela revient à ajouter ce nombre réel à l'ensemble des composantes du vecteur, *i.e.*

$$x + \alpha = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + 3 = \begin{pmatrix} 1 + 3 \\ 2 + 3 \\ 3 + 3 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}.$$

3 Matrices et calcul matriciel

Dans la section précédente nous avons largement étudié les espaces vectoriels de dimension finie et nous avons vu que nous pouvions munir de tels espaces d'une **base** et que tout vecteur pouvait se **représenter de façon unique** dans cette base. Nous avons également vu que l'on pouvait définir une application linéaire à partir de la connaissance de l'image des vecteurs de bases par cette application linéaire. Plus précisément, une application linéaire f entre deux espaces E et F de dimensions finies p et n est entièrement définie par la donnée des images des p vecteurs de la base de E et ces images peuvent chacune être décomposées dans la base de F .

On va donc obtenir p n -uplets de coordonnées que l'on va écrire dans un tableau appelé **matrice**.

Cette section sera donc principalement dédiée à l'étude des matrices et à leurs propriétés. Nous verrons que de ces propriétés sur les matrices nous pourrions en déduire des propriétés sur les applications linéaires qu'elles représentent. Beaucoup de ces points constituent des rappels, mais ces propriétés seront importantes pour la suite de ce cours et nous ne pouvons donc pas faire l'impasse dessus.

Dans toute cette section, on supposera que les espaces vectoriels étudiés sont des \mathbb{K} -espaces vectoriels avec $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} .

3.1 Rappels sur les matrices

Définition 3.1: Matrice

On appelle *matrice* à n lignes et p colonnes à coefficients dans \mathbb{K} , toute application de $\llbracket 1, n \rrbracket \times \llbracket 1, p \rrbracket$ dans \mathbb{K} .

Une telle matrice, notée A , se note alors

$$A = (a_{ij})_{i,j=1}^{n,p} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix}.$$

La matrice A est également appelée matrice de type (n, p) pour dire qu'elle comporte n lignes et p colonnes.

Dans cette définition, notez bien que le premier indice, l'indice i , fait référence à la i -ème ligne du tableau. Le deuxième indice, l'indice j , désigne la j -ème colonne de notre tableau.

La matrice est donc à voir comme un tableau de valeurs. Cette vision est très importante pour la suite car elle correspond à la représentation la plus couramment utilisée lorsque l'on représente des données avec les conventions suivantes :

- **en ligne** : on représente les individus (au nombre total de n , chaque individu étant décrit par p attributs).
- **en colonne** : on représente les valeurs prises par les différents attributs (au nombre total de p) sur les différents éléments de l'échantillon (que sont nos individus et au nombre de n).

Un autre cas particulier sont les matrices de type $(1, n)$ et $(n, 1)$ qui désignent des **vecteurs** ou **matrices** lignes et colonnes respectivement.

Définition 3.2: Vecteurs lignes et colonnes

Soit A une matrice de type (n, p) .

Pour tout $i \in \llbracket 1, n \rrbracket$, on appelle i -ème *vecteur ligne* de A le vecteur de \mathbb{K}^p défini par

$$L_i = (a_{i1}, a_{i2}, \dots, a_{ip}).$$

Pour tout $j \in \llbracket 1, p \rrbracket$, on appelle j -ème *vecteur colonne* de A le vecteur de \mathbb{K}^n défini par

$$C_j = (a_{1j}, a_{2j}, \dots, a_{pj}).$$

Définition 3.3: Matrices carrées

On appelle **matrice carrée** toute matrice de type (n, n) . Ce type de matrice est dit matrice d'ordre n .

Une matrice carrée A est dite **diagonale** si pour tout $i \neq j$, $a_{ij} = 0$.

Une matrice carrée A est dite **triangulaire supérieure** (respectivement **triangulaire inférieure**) si pour tout $i > j$ (respectivement pour tout $i < j$) $a_{ij} = 0$.

Exemple 3.1. Les matrices A, B, C et D suivantes sont respectivement des matrices carrées de type $(3, 3)$ quelconque, diagonale, triangulaire supérieure et triangulaire inférieure.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \pi \end{pmatrix}, \quad C = \begin{pmatrix} e & 3 & \ln(2) \\ 0 & 1 & 0 \\ 0 & 0 & \pi \end{pmatrix}, \quad D = \begin{pmatrix} e & 0 & 0 \\ 1 & 2 & 0 \\ 0 & \gamma & 0.12 \end{pmatrix}.$$

La littérature regorge de matrices carrées de taille n avec des propriétés très intéressantes. L'une d'entre elle, que l'on présentera plus tard, reviendra très souvent dans

le contexte d'analyse de données.

Une dernière matrice importante mais très simple à manipuler est la matrice identité. C'est une matrice qui représente l'application identité d'un espace vectoriel dans lui-même (si on souhaite faire référence à la section précédente) et elle s'écrit très simplement comme le montre la définition suivante.

Définition 3.4: Matrice identité

On appelle **matrice identité d'ordre n** , notée I_n , la matrice carrée diagonale d'ordre n dont tous les coefficients diagonaux sont égaux à 1, on a donc :

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}.$$

La définition suivante montre que l'espace des matrices a une structure bien particulière que l'on a déjà rencontré.

Définition 3.5: Structure de l'espace des matrices

On note $\mathcal{M}_{n,p}(\mathbb{K})$ l'ensemble des matrices de type (n,p) à coefficients dans \mathbb{K} . L'ensemble des matrices carrées d'ordre n à coefficients dans \mathbb{K} est quant à lui noté $\mathcal{M}_N(\mathbb{K})$.

On peut munir ces espaces de l'addition et de la multiplication par un scalaire $\lambda \in \mathbb{K}$ définies comme suit, pour tout $A, B \in \mathcal{M}_{n,p}(\mathbb{K})$

$$A + B = \begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{np} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1p} \\ \vdots & & \vdots \\ b_{n1} & \dots & b_{np} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1p} + b_{1p} \\ \vdots & & \vdots \\ a_{n1} + b_{n1} & \dots & a_{np} + b_{np} \end{pmatrix},$$

$$\lambda A = \lambda \begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{np} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \dots & \lambda a_{1p} \\ \vdots & & \vdots \\ \lambda a_{n1} & \dots & \lambda a_{np} \end{pmatrix}.$$

Proposition 3.1: Espace vectoriel des matrices

Muni de ces lois internes (addition) et externe (multiplication par un scalaire), l'espace $\mathcal{M}_{n,p}(\mathbb{K})$ est alors un espace vectoriel.

Comme tout espace vectoriel, il est possible de définir une base de cet espace et ainsi de définir sa dimension.

Proposition 3.2: Dimension et base de $\mathcal{M}_{n,p}(\mathbb{K})$

L'espace vectoriel $\mathcal{M}_{n,p}(\mathbb{K})$ est de dimension $n \times p$. Une base de cet espace vectoriel est donnée par la famille des matrices $(E_{ij})_{i,j=1}^{n,p}$ définies par :

$$E_{ij} = \begin{pmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{pmatrix} \begin{matrix} \leftarrow i\text{-ème ligne} \\ \\ \uparrow \\ j\text{-ème colonne} \end{matrix}$$

Démonstration. On vérifie facilement que les éléments E_{ij} constituent une base de cet espace vectoriel en remarquant que

$$A = (a)_{i,j=1}^{n,p} = \sum_{i=1}^n \sum_{j=1}^p a_{ij} E_{ij}.$$

Cette base est une famille de $n \times p$ éléments donc $\mathcal{M}_{n,p}(\mathbb{K})$ est un espace de dimension $n \times p$. \square

Cette base formée des éléments $(E_{ij})_{i,j=1}^{n,p}$ constitue même la base canonique de $\mathcal{M}_{n,p}(\mathbb{K})$.

On rappelle rapidement que le produit de deux matrices $A \in \mathcal{M}_{m,n}(\mathbb{K})$ et $B \in \mathcal{M}_{n,p}(\mathbb{K})$ résulte en une matrice $C \in \mathcal{M}_{m,p}(\mathbb{K})$ dont les éléments c_{ij} sont définies, pour tout $i \in \llbracket 1, m \rrbracket$ et $j \in \llbracket 1, p \rrbracket$, par :

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}.$$

Notez bien le format de la matrice C , on pourrait établir la règle suivante sur le format des matrices :

$$(m, p) = (m, n) \times (n, p).$$

La proposition suivante rappelle que le produit matriciel est *associatif*.

Proposition 3.3: Associativité produit matriciel

Soient A, B et C telles que $A \in \mathcal{M}_{m,n}(\mathbb{K})$, $B \in \mathcal{M}_{n,p}(\mathbb{K})$ et $C \in \mathcal{M}_{p,q}(\mathbb{K})$, alors

$$(AB)C = A(BC).$$

Démonstration. Elle est laissée à titre d'exercice car uniquement calculatoire. Il faut employer le rappel sur le produit matriciel. \square

Définition 3.6: Transposition d'une matrice

Soit $A \in \mathcal{M}_{n,p}(\mathbb{K})$, on appelle **transposée de la matrice** A , notée A^T , la matrice A' de $\mathcal{M}_{p,n}(\mathbb{K})$ définie pour tout $(i, j) \in \llbracket 1, p \rrbracket \times \llbracket 1, n \rrbracket$ par $a'_{ij} = a_{ji}$. Ecrit autrement

$$\text{si } A = \begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{np} \end{pmatrix} \quad \text{alors} \quad A^T = \begin{pmatrix} a_{11} & \dots & a_{n1} \\ \vdots & & \vdots \\ a_{1p} & \dots & a_{np} \end{pmatrix}$$

Notez bien que si la matrice A est de type (n, p) , sa transposée est elle de type (p, n) . La transposition est une transformation qui admet les propriétés suivantes

Proposition 3.4: Propriétés transposition

Soient A, B deux matrices de $\mathcal{M}_{m,n}(\mathbb{K})$, C une matrice de $\mathcal{M}_{n,p}(\mathbb{K})$, et soit $\lambda \in \mathbb{K}$ un scalaire, alors :

- i) $(\lambda A)^T = \lambda A^T$,
- ii) $(A + B)^T = A^T + B^T$,
- iii) $(AC)^T = C^T A^T$.

Démonstration. La démonstration est purement calculatoire, elle est donc laissée à titre d'exercice. \square

Le cas des matrices carrées est plus intéressant à étudier et permet de mettre en avant une propriété importante de certaines matrices : **la symétrie**.

Définition 3.7: Matrices symétriques et anti-symétriques

Soit $A \in \mathcal{M}_n(\mathbb{K})$, alors :

- i) A est dite symétrique si $A^T = A$,
- ii) A est dite anti-symétrique si $A^T = -A$

En général on note $S_n(\mathbb{K})$ l'ensemble des matrices symétriques d'ordre n sur le corps K et $A_n(\mathbb{K})$ l'ensemble des matrices anti-symétriques d'ordre n sur le corps \mathbb{K} . On pourra même montrer que ces deux ensembles forment des sous-espaces vectoriels supplémentaires de $\mathcal{M}_n(\mathbb{K})$, *i.e.* toute matrice carrée peut s'écrire comme la somme d'une matrice symétrique et d'une matrice anti-symétrique.

Exemple 3.2. Les matrices S et A suivantes sont respectivement des matrices symétriques et anti-symétriques d'ordre 4

$$S = \begin{pmatrix} a & b & c & d \\ b & e & f & g \\ c & f & h & i \\ d & g & i & j \end{pmatrix} \quad \text{et} \quad A = \begin{pmatrix} 0 & -b & -c & -d \\ b & 0 & -f & -g \\ c & f & 0 & -i \\ d & g & i & 0 \end{pmatrix}.$$

Quid de la matrice nulle à votre avis (celle ne comportant que des 0) ? Notez que le fait d'être anti-symétrique impose nécessairement que la diagonale de la matrice soit nulle.

3.2 Représentation des applications linéaires

Après ces quelques rappels sur les matrices, nous allons maintenant pouvoir faire le lien entre les applications linéaires présentées aux sections précédentes et leur représentation matricielle.

Pour cela, nous allons considérer deux espaces vectoriels E et F tous deux de dimension finie p et n respectivement. Les espaces E et F seront également munis des bases $(\mathbf{e}_1, \dots, \mathbf{e}_p)$ et $(\mathbf{f}_1, \dots, \mathbf{f}_n)$ respectivement. Enfin on désignera par u (non plus f comme dans les sections précédentes pour éviter les confusions) une application linéaire de E dans F .

Nous avons précédemment vu que u est entièrement déterminée par l'image des vecteurs de la base de E dans la base de F . Formalisons cela.

Définition 3.8: Représentation matricielle

Pour tout $j \in \llbracket 1, p \rrbracket$, notons $(a_{ij})_{i=1}^n$ les coordonnées de $u(\mathbf{e}_j)$ dans la base de F , on a donc :

$$u(\mathbf{e}_j) = \sum_{i=1}^n a_{ij} \mathbf{f}_i.$$

La matrice $A = (a_{ij})_{i,j=1}^{n,p}$ obtenue est alors appelée **matrice de u** relativement aux bases de E et F . On la note en générale $A = \underset{\mathcal{B}_E, \mathcal{B}_F}{Mat}(u)$ ou encore $Mat(u)$ lorsque le contexte n'est pas ambigu.

On peut représenter cette matrice de la façon suivante :

$$\underset{\mathcal{B}_E, \mathcal{B}_F}{Mat}(u) = \begin{pmatrix} u(\mathbf{e}_1) & \cdots & u(\mathbf{e}_p) \\ a_{11} & \cdots & a_{1p} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{np} \end{pmatrix} \begin{pmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_n \end{pmatrix}.$$

Ainsi la j -ème colonne représente l'image du vecteur \mathbf{e}_j par u dans la base de F .

Exemple 3.3. Considérons l'application linéaire u , de \mathbb{R}^3 dans \mathbb{R}^3 munis de leur base canonique, définie par

$$f(\mathbf{x}) = (2x_1 + x_3, x_2 - x_3, x_1 - x_2 + 3x_3).$$

Pour déterminer sa représentation matricielle, on calcule l'image des vecteurs de base par l'application u

$$\begin{aligned} u(\mathbf{e}_1) &= (2, 0, 1) = 2\mathbf{e}_1 + 0\mathbf{e}_2 + 1\mathbf{e}_3, \\ u(\mathbf{e}_2) &= (0, 1, -1) = 0\mathbf{e}_1 + 1\mathbf{e}_2 + -1\mathbf{e}_3, \\ u(\mathbf{e}_3) &= (1, -1, 3) = 1\mathbf{e}_1 + -1\mathbf{e}_2 + 3\mathbf{e}_3 \end{aligned}$$

Ainsi la matrice de l'application linéaire u est $\begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 3 \end{pmatrix}.$

On peut maintenant utiliser cette définition pour définir l'image \mathbf{y} d'un vecteur \mathbf{x} par l'application linéaire.

Pour tout vecteur $\mathbf{x} \in E$, on peut décomposer \mathbf{x} dans la base de E et ainsi $\mathbf{x} = \sum_{j=1}^p x_j \mathbf{e}_j$, on a alors :

$$u(\mathbf{x}) = \sum_{j=1}^p x_j u(\mathbf{e}_j) = \sum_{j=1}^p x_j \sum_{i=1}^n a_{ij} \mathbf{f}_i = \sum_{j=1}^p \sum_{i=1}^n a_{ij} x_j \mathbf{f}_i.$$

Cette dernière égalité donne directement une décomposition de l'image de \mathbf{x} par u dans la base de F .

Alors pour tout $i \in \llbracket 1, n \rrbracket$, $y_i = \sum_{j=1}^p a_{ij} x_j$. Ce que l'on peut écrire, de façon matricielle : $\mathbf{y} = A\mathbf{x}$

Cette définition montre que derrière chaque matrice se cache en fait une application linéaire. En effet, nous pouvons même montrer qu'il existe un isomorphisme entre $\mathcal{L}(E, F)$ et $\mathcal{M}_{n,p}(\mathbb{K})$.

Ainsi lorsque l'on multiplie des matrices, cela revient en fait à composer des applications linéaires !

Cette remarque reste valable pour les matrices carrées d'ordre n . On peut à nouveau montrer que les espaces $\mathcal{L}(E)$ et $\mathcal{M}_n(\mathbb{K})$ sont isomorphes.

Cependant, les matrices carrées ont une propriété de plus que les matrices quelconques ... certaines d'entre elles sont **inversibles** ! Souvenez-vous, nous avons déjà vu que certaines applications linéaires sont inversibles, c'est ce que l'on appelé **les automorphismes**.

Donc si une application linéaire est inversible, il en est de même pour la matrice associée.

Proposition 3.5: Matrice et inverse d'une application linéaire

Soit E un espace vectoriel sur le corps \mathbb{K} de dimension finie $n \in \mathbb{N}^*$ muni d'une base et soit $u \in \mathcal{L}(E)$. Notons A la représentation matricielle de u , alors u est inversible si et seulement si la matrice A associée est inversible. De plus, si u est inversible on note $A^{-1} = \text{Mat}(u^{-1})$ la matrice inverse de A .

On conserve également les mêmes propriétés que pour les endomorphismes inversibles de E , à savoir l'équivalence entre les propriétés suivantes :

- i) A est inversible à droite,
- ii) A est inversible à gauche,
- iii) A est inversible.

3.3 Familles de vecteurs et changement de bases

La représentation matricielle ne sert pas uniquement à représenter des applications linéaires, elle peut aussi être utilisée pour représenter une famille de vecteurs dans une

base. C'est d'ailleurs cette vision là que nous adoptons lorsque l'on souhaite représenter nos données sous forme de tableaux.

Considérons un espace vectoriel E de dimension finie $n \in \mathbb{N}^*$ muni d'une base \mathcal{B}_E et considérons une famille de vecteurs $(\mathbf{x}_1, \dots, \mathbf{x}_p)$, alors pour tout $j \in \llbracket 1, p \rrbracket$ on peut décomposer le vecteur \mathbf{x}_j dans la base de E

$$\mathbf{x}_j = \sum_{i=1}^n a_{ij} \mathbf{e}_i,$$

où les coefficients a_{ij} représentent les coordonnées des vecteurs dans la base de E . On note alors $A = (a_{ij})_{i,j=1}^{n,p}$ la représentation matricielle de cette famille de vecteurs. Ce que l'on peut encore écrire

$$Mat_{\mathcal{B}_E}(\mathbf{x}_1, \dots, \mathbf{x}_p) = \begin{pmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_p \\ a_{11} & \cdots & a_{1p} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{np} \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{pmatrix}.$$

Proposition 3.6: Base et inversibilité

Soit E un espace vectoriel de dimension $n \in \mathbb{N}^*$ muni d'une base \mathcal{B}_E et considérons une famille de vecteurs $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ de E , alors cette famille est une base de E si et seulement si la matrice associée à cette famille est inversible.

Ce résultat est important pour introduire la notion de changement de bases. En effet il est possible que l'on ne souhaite pas forcément travailler avec la base canonique, ce qui est très souvent le cas en analyse de données où l'on préfère regarder les données "sous un autre angle". Il faut alors voir comment faire pour passer d'une base à une autre.

Définition 3.9: Changement de bases

Soit E un espace vectoriel sur \mathbb{K} de dimension finie $n \in \mathbb{N}^*$, et soient $\mathcal{B} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ et $\mathcal{B}' = (\mathbf{e}'_1, \dots, \mathbf{e}'_n)$ deux bases de E . On appelle **matrice de passage de \mathcal{B} dans \mathcal{B}' ou de \mathcal{B} vers \mathcal{B}'** la matrice $P_{\mathcal{B}, \mathcal{B}'}$ définie par :

$$P_{\mathcal{B}, \mathcal{B}'} = Mat_{\mathcal{B}}(\mathcal{B}') = Mat_{\mathcal{B}}(\mathbf{e}'_1, \dots, \mathbf{e}'_n).$$

Pour dire les choses plus simplement, les colonnes de la matrice de changement de bases de \mathcal{B} vers \mathcal{B}' sont formées par les coordonnées des vecteurs de \mathcal{B}' dans la base \mathcal{B} .

Exemple 3.4. On va considérer l'espace vectoriel $E = \mathbb{R}^2$ muni de deux bases différentes :

$$\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2) = ((1, 1), (-1, 1)) \quad \text{et} \quad \mathcal{B}' = (\mathbf{e}'_1, \mathbf{e}'_2) = ((1, 0), (3, 1))$$

Pour déterminer la matrice de changement de base, il faut alors exprimer les vecteurs de la base \mathcal{B}' dans \mathcal{B} .

On montre alors que la matrice de changement de base est définie par :

$$P_{\mathcal{B}, \mathcal{B}'} = \begin{pmatrix} \mathbf{e}'_1 & \mathbf{e}'_2 \\ 0.5 & 2 \\ -0.5 & -1 \end{pmatrix} \begin{matrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{matrix}.$$

Pour \mathbf{e}'_2 on a bien $\mathbf{e}'_2 = 2\mathbf{e}_1 - \mathbf{e}_2$. On peut ici trouver les coefficients de tête, mais nous verrons plus tard comment faire cela en résolvant ce que l'on appelle des systèmes linéaires.

Le résultat suivant montre comment on peut "rebrousser chemin" i.e. passer de la base \mathcal{B}' vers la base \mathcal{B} , il est une conséquence de la Proposition 3.6.

Proposition 3.7: Bases et inversibilité

Soit E un espace vectoriel sur \mathbb{K} de dimension finie $n \in \mathbb{N}^*$ et soient \mathcal{B} et \mathcal{B}' deux bases de E . Alors $P_{\mathcal{B}, \mathcal{B}'}$ est inversible et $(P_{\mathcal{B}, \mathcal{B}'})^{-1} = P_{\mathcal{B}', \mathcal{B}}$.

Maintenant que l'on dispose de ces résultats on va pouvoir comment représenter un vecteur \mathbf{x} ou une matrice A dans des bases différentes.

Proposition 3.8: Changement de base pour un vecteur

Soit E un espace vectoriel sur \mathbb{K} de dimension finie $n \in \mathbb{N}^*$. Soient $\mathcal{B} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ et $\mathcal{B}' = (\mathbf{e}'_1, \dots, \mathbf{e}'_n)$ deux bases de E , et soit $P = P_{\mathcal{B}, \mathcal{B}'}$.

Considérons un vecteur $\mathbf{x} \in E$, on peut alors écrire

$$\mathbf{x} = \sum_{k=1}^n x_k \mathbf{e}_k \quad \text{et} \quad \mathbf{x} = \sum_{k=1}^n x'_k \mathbf{e}'_k.$$

Notons alors $\mathbf{x}_{\mathcal{B}} = (x_1, \dots, x_n)$ et $\mathbf{x}_{\mathcal{B}'} = (x'_1, \dots, x'_n)$, alors

$$\mathbf{x}_{\mathcal{B}} = P \mathbf{x}_{\mathcal{B}'}.$$

Nous ne cherchons pas à démontrer ce résultat mais simplement à l'illustrer dans le cas particulier d'un espace de dimension 2.

Exemple 3.5. Considérons deux bases $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2)$ et $\text{dom}\mathcal{B}' = (\mathbf{e}'_1, \mathbf{e}'_2)$ d'un espace vectoriel E de dimension 2 et un élément \mathbf{x} de E dont les coordonnées sont respectivement notées $\mathbf{x}_{\mathcal{B}} = (x_1, x_2)$ et $\mathbf{x}_{\mathcal{B}'} = (x'_1, x'_2)$ les coordonnées du vecteur \mathbf{x} dans les bases \mathcal{B} et \mathcal{B}' .

Supposons que l'on a également les relations suivantes entre les vecteurs des deux bases

$$\begin{aligned}\mathbf{e}'_1 &= a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2, \\ \mathbf{e}'_2 &= a_{12}\mathbf{e}_1 + a_{22}\mathbf{e}_2.\end{aligned}$$

On va maintenant chercher à trouver notre matrice de passage P de \mathcal{B} vers \mathcal{B}' . Pour cela, on prend l'expression de notre vecteur \mathbf{x} exprimée dans la nouvelle base \mathcal{B}' et on va chercher ces coordonnées dans l'ancienne base \mathcal{B} .

$$\begin{aligned}\mathbf{x} &= x'_1 \underbrace{\mathbf{e}'_1}_{\text{définition de } \mathbf{e}'_1 \text{ et } \mathbf{e}'_2} + x'_2 \underbrace{\mathbf{e}'_2}_{\text{définition de } \mathbf{e}'_1 \text{ et } \mathbf{e}'_2}, \\ &\downarrow \text{on factorise} \\ &= x'_1 a_{11} \mathbf{e}_1 + a_{21} \mathbf{e}_2 + x'_2 a_{12} \mathbf{e}_1 + a_{22} \mathbf{e}_2, \\ &\downarrow \text{on factorise} \\ &= (a_{11}x'_1 + a_{12}x'_2)\mathbf{e}_1 + (a_{21}x'_1 + a_{22}x'_2)\mathbf{e}_2, \\ &\downarrow \text{définition de } \mathbf{x} \text{ dans la base } (\mathbf{e}_1, \mathbf{e}_2) \\ &= x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2.\end{aligned}$$

En étudiant les deux dernières égalités, nous aboutissons aux relations suivantes :

$$\begin{aligned}x_1 &= a_{11}x'_1 + a_{12}x'_2, \\ x_2 &= a_{21}x'_1 + a_{22}x'_2.\end{aligned}$$

Ainsi, nous avons la relation $\mathbf{x}_{\mathcal{B}} = P\mathbf{x}_{\mathcal{B}'}$ où $P = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

Proposition 3.9: Changement de base pour une application linéaire

Soient E et F des espaces vectoriels sur \mathbb{K} de dimensions finies p et n . Soient $\mathcal{B}_E = (\mathbf{e}_1, \dots, \mathbf{e}_p)$ et $\mathcal{B}'_E = (\mathbf{e}'_1, \dots, \mathbf{e}'_p)$ deux bases de E et soit P la matrice de passage de \mathcal{B}_E dans \mathcal{B}'_E . De même, soient $\mathcal{B}_F = (\mathbf{f}_1, \dots, \mathbf{f}_n)$ et $\mathcal{B}'_F = (\mathbf{f}'_1, \dots, \mathbf{f}'_n)$ deux bases de F et soit Q la matrice de passage de \mathcal{B}_F dans \mathcal{B}'_F . Soit $u \in \mathcal{L}(E, F)$ et notons $A = \text{Mat}_{\mathcal{B}_E, \mathcal{B}_F}(u)$ et $A' = \text{Mat}_{\mathcal{B}'_E, \mathcal{B}'_F}(u)$. Alors

$$A' = Q^{-1}AP.$$

La démonstration devrait permettre de lever l'ombre qui plane sur cette relation.

Démonstration. Soient \mathbf{x} et \mathbf{x}' les représentations d'un vecteur de E dans les bases \mathcal{B}_E et \mathcal{B}'_E et soient \mathbf{y} et \mathbf{y}' les représentations d'un vecteur de E dans les bases \mathcal{B}_F et \mathcal{B}'_F . A est l'unique matrice telle que pour tout $(\mathbf{x}, \mathbf{y}) \in E$, $u(\mathbf{x}) = \mathbf{y}$ si et seulement si $A\mathbf{x} = \mathbf{y}$.

A est l'unique matrice telle que pour tout $(\mathbf{x}', \mathbf{y}') \in E$, $u(\mathbf{x}') = \mathbf{y}'$ si et seulement si $A\mathbf{x}' = \mathbf{y}'$.

De plus, nous avons les relations $\mathbf{x}' = P\mathbf{x}$ et $\mathbf{y}' = Q\mathbf{y}$.

Donc

$$A\mathbf{x} = \mathbf{y} \iff AP\mathbf{x}' = Q\mathbf{y}' \iff Q^{-1}AP\mathbf{x}' = \mathbf{y}',$$

ainsi $A' = Q^{-1}AP$. □

Corollaire 3.1: Changement de base d'un endomorphisme

Soit E un espace vectoriel sur \mathbb{K} de dimension finie n . Soient $\mathcal{B} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ et $\mathcal{B}' = (\mathbf{e}'_1, \dots, \mathbf{e}'_n)$ deux bases de E et soit P la matrice de passage de \mathcal{B} vers \mathcal{B}' . Considérons un endomorphisme u de E et notons A sa représentation matricielle dans la base \mathcal{B} et A' sa représentation matricielle dans la base \mathcal{B}' . Alors $A' = P^{-1}AP$.

Démonstration. Il suffit d'appliquer la proposition précédente dans le cas où $E = F$, $\mathcal{B}_E = \mathcal{B}_F = \mathcal{B}$ et $\mathcal{B}'_E = \mathcal{B}'_F = \mathcal{B}'$. Dans ce cas : $P = Q$. □

Exemple 3.6. Soit u un endomorphisme de $E = \mathbb{R}^2$ dont la matrice associée, notée A , dans la base canonique est donnée par

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

On reconnaît ici l'endomorphisme u qui pour tout $\mathbf{x} \in \mathbb{R}^2$ est défini par

$$u(\mathbf{x}) = (2x_1 + x_2, -x_1 + 3x_2).$$

On souhaite maintenant exprimer cet endomorphisme dans la base formée par les vecteurs $\mathbf{v}_1 = (-1, 1)$ et $\mathbf{v}_2 = (1, 1)$. Dans un premier temps, il nous faut déterminer la matrice de passage P de la base canonique vers la nouvelle base définie par \mathbf{v}_1 et \mathbf{v}_2 .

On rappelle que la matrice de passage P est obtenue en exprimant les vecteurs de la nouvelle base \mathcal{B}' en fonction des vecteurs de l'ancienne base \mathcal{B} , cette représentation se faisant en colonne :

$$P_{\mathcal{B}, \mathcal{B}'} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \\ -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix}.$$

Cette matrice est inversible et son inverse P^{-1} est donnée par

$$P^{-1} = \begin{pmatrix} -0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}.$$

Il reste alors à calculer le produit $P^{-1}AP$, ce qui nous donne

$$P^{-1}AP = \frac{1}{2} \begin{pmatrix} -1 & 11 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 5 & -1 \\ 3 & 5 \end{pmatrix}.$$

Le cas des endomorphismes sera particulièrement intéressant lorsque nous traiterons de la réduction des endomorphismes plus tard dans ce document.

3.4 Opérations élémentaires sur les matrices

On présente ici quelques opérations élémentaires sur les matrices qui nous serviront lorsque l'on effectuera du calcul matriciel.

Définition 3.10: Opérations élémentaires

On appelle **opération élémentaire sur les lignes (ou sur les colonnes)** d'une matrice l'une des trois opérations suivantes :

- i) multiplier une ligne (ou une colonne) par un scalaire λ , *i.e.* faire l'opération $L_i \leftarrow \lambda L_i$ (ou $C_i \leftarrow \lambda C_i$),
- ii) ajouter à une ligne (ou à une colonne) donnée le produit d'une autre ligne (ou d'une autre colonne) par un scalaire λ , *i.e.* $L_i \leftarrow L_i + \lambda L_j$ (ou $C_i \leftarrow C_i + \lambda C_j$),
- iii) intervertir deux lignes (ou deux colonnes), *i.e.* $L_i \leftrightarrow L_j$ (ou $C_i \leftrightarrow C_j$).

Ces transformations peuvent s'écrire matriciellement. Pour cela, considérons une matrice $A \in \mathcal{M}_{n,p}(\mathbb{K})$ sur laquelle on va effectuer les opérations élémentaires précédentes sur les lignes de la matrice A .

- i) **Opération** $L_i \leftarrow \lambda L_i$, cela revient à faire le produit

$$i \rightarrow \begin{pmatrix} & & i \downarrow & & \\ & & \downarrow & & \\ 1 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \\ & & & \lambda & \\ & 0 & & \ddots & \\ & & & & 1 \end{pmatrix} A$$

ii) **Opération** $L_i \leftarrow L_i + \lambda L_j$, cela revient à faire le produit

$$i \rightarrow \begin{pmatrix} & & & j \downarrow & & \\ & & & \downarrow & & \\ 1 & & & & & 0 \\ & \ddots & & & & \\ & & 1 & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & \lambda & \cdots & 0 \\ & & & & 1 & & & & \\ & & & & & \ddots & & & \\ & 0 & & & & & \ddots & & \\ & & & & & & & 1 & \\ & & & & & & & & 1 \end{pmatrix} A$$

iii) **L'opération** $L_i \leftrightarrow L_j$, cela revient à faire :

$$\begin{matrix} & i \downarrow & & j \downarrow & \\ i \rightarrow & \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 0 & \cdots & \cdots & \cdots & 1 \\ & & \vdots & 1 & & & \vdots \\ & & \vdots & & \ddots & & \vdots \\ & & \vdots & & & 1 & \vdots \\ & & 1 & \cdots & \cdots & \cdots & 0 \end{pmatrix} & A \\ j \rightarrow & \begin{pmatrix} & & & & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} \end{matrix}$$

Remarque. On voit qu'effectuer une opération élémentaire sur les lignes revient à pré-multiplier la matrice A par l'opération élémentaire. Si on souhaite travailler sur les colonnes, cela revient alors à post-multiplier la matrice A par la matrice d'opération élémentaire.

On prendra cependant garde à la dimension des matrices représentant les opérations élémentaires selon que l'on agisse sur les lignes ou les colonnes !

3.5 Rang, déterminant, et inverse d'une matrice

3.5.1 Rang d'une matrice

Tout comme nous avons défini le rang d'une application linéaire, on peut également définir le rang d'une matrice.

Définition 3.11: Rang d'une matrice

Soit A une matrice de $\mathcal{M}_{n,p}(\mathbb{K})$, on appelle **rang de A** , noté $rg(A)$, le rang de la famille de vecteurs colonnes de A qui sont des vecteurs de \mathbb{K}^n .

Pourquoi on s'intéresse aux vecteurs colonnes, car ces derniers représentent les images des vecteurs de bases de l'espace vectoriel E , le rang de la famille de vecteurs colonne de la matrice A va donc déterminer la dimension de l'espace image de l'application linéaire étudiée.

Proposition 3.10: Rang et base

Soit E un \mathbb{K} -espace vectoriel de dimension finie n muni d'une base $\mathcal{B} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ et soit $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ une famille de vecteurs de E , alors :

$$rg(\mathbf{x}_1, \dots, \mathbf{x}_p) = rg(Mat_{\mathcal{B}}(\mathbf{x}_1, \dots, \mathbf{x}_p)).$$

De ce résultat on en déduit immédiatement que toute matrice carrée A d'ordre n est inversible si seulement son rang est égale à n . De plus, ce que l'on a vu sur les familles des vecteurs reste valable sur les matrices d'applications linéaires.

Proposition 3.11: Rang matrice d'une application linéaire

Soient E et F deux espaces vectoriels sur \mathbb{K} de dimensions finies p et n munis de bases \mathcal{B}_E et \mathcal{B}_F . Soit u une application linéaire de E dans F dont $A = Mat_{\mathcal{B}_E, \mathcal{B}_F}(u)$ est la représentation matricielle.
Alors $rg(u) = rg(A)$.

On terminera cette introduction sur le rang d'une matrice par le résultat suivant

Proposition 3.12: Propriété du rang

Soit $A \in \mathcal{M}_{n,p}(\mathbb{K})$, alors $rg(A^T) = rg(A)$. Le rang d'une matrice est donc invariant par transposition.

3.5.2 Déterminant d'une matrice

Présenter proprement ce qu'est le déterminant d'une matrice nécessiterait de présenter différents concepts plus ou moins abstraits qui ne seront pas utiles pour la suite de ce cours. Nous ne présenterons donc même pas une définition propre du déterminant et nous nous contenterons de donner quelques propriétés le concernant.

Le lecteur désirant en savoir plus sur la construction du déterminant est invité à étudier le *groupe symétrique* ainsi que les *formes n -linéaires alternées*. On se contentera uniquement de donner la définition du déterminant pour une matrice avec un minimum d'explications.

Définition 3.12: Déterminant

Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$ dont les éléments sont notés $(a_{ij})_{i,j=1}^n$. Soit \mathfrak{S}_n le groupe des permutations de $\llbracket 1, n \rrbracket$, c'est-à-dire l'ensemble des bijections de $\llbracket 1, n \rrbracket$ dans $\llbracket 1, n \rrbracket$ (comme une transposition qui échange de place deux éléments i et j). On définit le **déterminant** de la matrice A par la relation

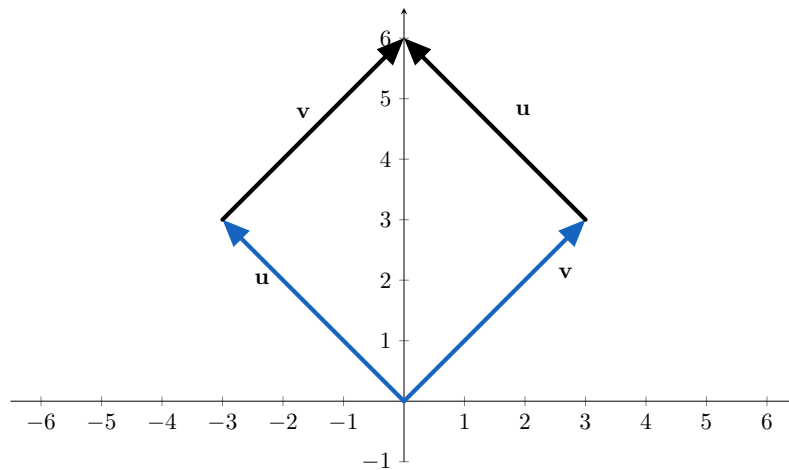
$$\det(A) = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma) \prod_{i=1}^n a_{\sigma(i), i},$$

où $\varepsilon(\sigma)$ est appelé **signature de la permutation** et qui vaut ± 1 .

Le déterminant est une valeur que l'on peut associer à une famille de vecteurs, aux applications linéaires en dimension finie ou plus généralement à des matrices.

Dans un contexte algébrique, il va permettre de déterminer si une famille de vecteurs constitue **une base** d'un espace vectoriel.

Remarque On peut également interpréter ce nombre d'un point de vue géométrique. En effet, dans des espaces de dimension 2 ou 3, il représentera l'aire du parallélogramme ou le volume d'un parallélépipède dans l'espace générés par des familles deux ou trois vecteurs respectivement.



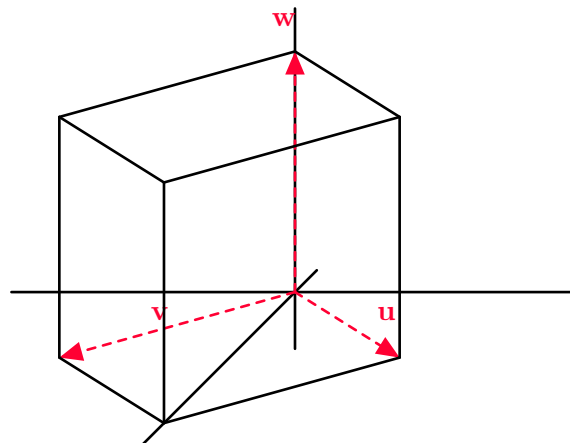
Dans ce premier exemple les vecteurs (\mathbf{u}, \mathbf{v}) sont définis par $\mathbf{u} = (-3, 3)$ et $\mathbf{v} = (3, 3)$ ce qui génère un carré avec une surface de 18. Nous aurions donc :

$$\det(\mathbf{u}, \mathbf{v}) = \begin{vmatrix} 3 & -3 \\ 3 & 3 \end{vmatrix} = 18.$$

Considérons maintenant les vecteurs $\mathbf{u}, \mathbf{v}, \mathbf{w}$ qui sont définis par

$$\mathbf{u} = \begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} -3 \\ 3 \\ 0 \end{pmatrix} \quad \text{et} \quad \mathbf{w} = \begin{pmatrix} 0 \\ 0 \\ \sqrt{18} \end{pmatrix}.$$

Ce qui permet de générer le parallélépipède suivant dans un espace de dimension 3, dont le volume qui n'est rien d'autre que le déterminant de la familles des 3 vecteurs, est égal à $18\sqrt{18} = 54\sqrt{2}$.



En effet :

$$\det(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \begin{vmatrix} 3 & -3 & 0 \\ 3 & 3 & 0 \\ 0 & 0 & \sqrt{18} \end{vmatrix} = 18\sqrt{18}.$$

Nous verrons un peu plus loin comment calculer les déterminants de telles familles de vecteurs. Mais avant cela, regardons quelques propriétés du déterminant.

Proposition 3.13: Base et déterminant

Soit $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ une famille de vecteurs d'un espace vectoriel E de dimension finie n . Alors cette famille constitue une base de cet espace vectoriel si et seulement si le déterminant associé à cette famille de vecteurs est non nul, *i.e.* $\det(\mathbf{x}_1, \dots, \mathbf{x}_n) \neq 0$.

De la même façon, on a vu que l'on pouvait représenter les endomorphismes u sur un espace vectoriel E en déterminant les images des vecteurs de la base de E par u , les images forment les vecteurs colonnes de la matrice. On peut donc voir si les images forment une base de l'espace E et donc voir si notre application est bijective en étudiant le déterminant de cette application linéaire ou de sa matrice associée.

Proposition 3.14: Caractérisation automorphisme

Soit $u \in \mathcal{L}(E)$, alors u est un automorphisme, *i.e.* un endomorphisme bijectif si et seulement si $\det(u) \neq 0$.

Proposition 3.15: Inversibilité d'une matrice

Soit $A \in \mathcal{M}_n(\mathbb{K})$ alors A est inversible si et seulement si $\det(A) \neq 0$.

Remarque. Le déterminant n'a de sens que pour les matrices carrées *i.e.* pour les endomorphismes d'espaces vectoriels.

On termine cette présentation du déterminant par quelques propriétés importantes concernant ce dernier.

Proposition 3.16: Propriétés du déterminant

Soient A et B deux matrices de $\mathcal{M}_n(\mathbb{K})$ alors

$$\det(A^T) = \det(A) \quad \text{et} \quad \det(AB) = \det(A)\det(B).$$

Si de plus A est inversible, *i.e.* si son déterminant est non nul alors

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

En revanche il n'y a pas de propriétés particulières concernant le déterminant de la somme de deux matrices ! En général

$$\det(A + B) \neq \det(A) + \det(B).$$

Ces propriétés seront illustrées sur des exemples simples dans la section suivante consacrée aux calculs sur les matrices.

Lien avec les opérations élémentaires. Le déterminant possède encore quelques propriétés qui seront utiles lorsque l'on cherchera à le calculer :

- il ne change pas si l'on ajoute à une colonne une combinaison linéaire des autres colonnes
- il change de signe si l'on échange les positions de deux colonnes
- si on multiplie une colonne par un scalaire λ alors le déterminant est lui-même multiplié par ce scalaire λ (car le déterminant est linéaire par rapport à chaque colonne).

Toutes ces remarques formulées sur les lignes restent valables sur les colonnes vu que $\det(A^T) = \det(A)$.

Mineures d'une matrice. Les résultats de cette section seront très utiles pour le calcul pratique du déterminant, ils nous serviront également à expliciter l'inverse d'une matrice.

Définition 3.13: Mineure et Cofacteur

Soit $A = (a_{ij})_{i,j=1}^n \in \mathcal{M}_n(\mathbb{K})$ alors pour tout couple (i, j)

- on appelle **mineur de A relatif à la i -ème ligne et à la j -ème colonne**, notée Δ_{ij} , le déterminant de la matrice obtenue en supprimant la i -ème ligne et la j -ème colonne de A ,
- on appelle **cofacteur de A**, notée A_{ij} , le scalaire défini par $A_{ij} = (-1)^{i+j} \Delta_{ij}$.

Proposition 3.17: Déterminant et mineurs

Soit $A = (a_{ij})_{i,j=1}^n \in \mathcal{M}_n(\mathbb{K})$ alors pour tout $1 \leq j \leq n$, $\det(A) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \Delta_{ij} = \sum_{i=1}^n a_{ij} A_{ij}$.

3.5.3 Inverse d'une matrice

Il nous reste à définir ce que l'on appelle la comatrice afin de pouvoir fournir une expression de l'inverse d'une matrice $A \in \mathcal{M}_n(\mathbb{K})$.

Définition 3.14: Comatrice

Soit $A = (a_{ij})_{i,j=1}^n \in \mathcal{M}_n(\mathbb{K})$, on appelle **comatrice de A**, notée $Com(A)$, la matrice des cofacteurs de A , c'est-à-dire $Com(A) = (a_{ij})_{i,j=1}^n$.

Un cofacteur A_{ij} est défini comme le déterminant de la matrice A à laquelle on aura supprimé la i -ème ligne et la j -ème colonne.

Proposition 3.18: Inverse d'une matrice

Soit $A \in \mathcal{M}_n(\mathbb{K})$ et supposons que A est inversible, alors $Com(A)^T A = A^T Com(A) = \det(A) I_n$.

On peut également réécrire le résultat de cette proposition comme suit : si A est inversible alors

$$A^{-1} = \frac{1}{\det(A)} Com(A)^T.$$

3.6 Calculs explicites

C'est le moment de mettre un peu de concret sur ce qui é été développé tout au long de cette section. Nous allons voir comment déterminer le rang, calculer le déterminant et l'inverse d'une matrice en appliquant les définitions, mais aussi en utilisant quelques petites "*astuces*" pratiques.

Avant de faire cela, on rappelle rapidement comment effectuer un produit matriciel.

On commence par le produit entre une matrice $A = (a_{ij})_{i,j=1}^{n,p} \in \mathcal{M}_{n,p}(\mathbb{R})$ et un vecteur de $\mathbf{x} = (x_1, x_2, \dots, x_p)^T \in \mathbb{R}^p$.

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{np} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^p a_{1k}x_k \\ \sum_{k=1}^p a_{2k}x_k \\ \vdots \\ \sum_{k=1}^p a_{nk}x_k \end{pmatrix}$$

On peut également représenter d'une façon plus graphique ce produit entre la matrice A et le vecteur \mathbf{x}

$$L_i \begin{pmatrix} a_{11} & \cdots & a_{ik} & \cdots & a_{1p} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ik} & \cdots & a_{ip} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & \cdots & \cdots & a_{np} \end{pmatrix} \cdots \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ \vdots \\ x_p \end{pmatrix} = y_i$$

Rappelons également comment effectuer le produit entre deux matrices de dimensions quelconques. On considère $A = (a_{ij})_{i,j=1}^{n,p} \in \mathcal{M}_{n,p}(\mathbb{R})$ et $B = (b_{jk})_{j,k=1}^{p,q} \in \mathcal{M}_{p,q}(\mathbb{R})$. Si on note C le produit matriciel AB , alors l'élément d'indice (i, j) de cette matrice C est égal à

$$c_{ij} = (AB)_{ij} = \sum_{k=1}^p a_{ik}b_{kj}.$$

$$C_{ij}$$

$$L_i \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & & \vdots \\ a_{i1} & \cdots & a_{ip} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{np} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1q} \\ \vdots & & \vdots & & \vdots \\ \vdots & & b_{kj} & & \vdots \\ \vdots & & \vdots & & \vdots \\ b_{n1} & \cdots & b_{pj} & \cdots & b_{pq} \end{pmatrix} = \begin{pmatrix} \cdots & c_{ij} & \cdots \end{pmatrix}$$

Exemple 3.7. Soient $A \in \mathcal{M}_{3,4}(\mathbb{R})$, $B \in \mathcal{M}_{4,2}(\mathbb{R})$ et $\mathbf{x} \in \mathbb{R}^4$ définis par

$$A = \begin{pmatrix} 2 & -1 & 0 & 4 \\ -1 & 0 & 3 & -2 \\ 5 & 0 & 0 & -4 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 4 & 3 \\ 0 & -2 \\ 1 & -1 \end{pmatrix} \quad et \quad \mathbf{x} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix}.$$

Alors le produit de la matrice A avec le vecteur \mathbf{x} (qui revient donc à calculer l'image du vecteur \mathbf{x} par l'application linéaire associée à la matrice A) est

$$A\mathbf{x} = \begin{pmatrix} 2 & -1 & 0 & 4 \\ -1 & 0 & 3 & -2 \\ 5 & 0 & 0 & -4 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \times 1 - 1 \times (-1) + 0 \times 0 + 4 \times 2 \\ \dots \\ \dots \end{pmatrix} = \begin{pmatrix} 10 \\ -5 \\ -3 \end{pmatrix}$$

Alors le produit $C = AB$ est

$$C = AB = \begin{pmatrix} 2 & -1 & 0 & 4 \\ -1 & 0 & 3 & -2 \\ 5 & 0 & 0 & -4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 4 & 3 \\ 0 & -2 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 \times 0 - 1 \times 4 + 0 \times 0 + 4 \times 1 & \dots \\ \dots & \dots \\ \dots & \dots \end{pmatrix} = \begin{pmatrix} 0 & -5 \\ -2 & -5 \\ -4 & 9 \end{pmatrix}.$$

Lorsque l'on cherche à effectuer un produit matriciel, on prendra garde à faire en sorte que le produit soit **licite**! Un produit AB est licite si *la matrice A possède autant*

de colonne que la matrice B possède de ligne.

Enfin, noter que le produit de deux matrices A et B donne une matrice C dont le nombre ligne est égal au nombre de ligne de la matrice A et dont le nombre de colonne est égal au nombre de colonne de la matrice B . Dit autrement, le produit d'une matrice de $\mathcal{M}_{n,p}(\mathbb{R})$ avec une matrice de $\mathcal{M}_{p,q}(\mathbb{R})$ donne une matrice de $\mathcal{M}_{n,q}(\mathbb{R})$.

3.6.1 Matrice échelonnée (réduite)

Définition 3.15: Matrice échelonnée

Une matrice est dite **échelonnée** en lignes si son nombre de zéros précédant la première valeur non d'une ligne augmente ligne par ligne jusqu'à l'éventuelle obtention de lignes ne comportant que des zéros.

Définition 3.16: Matrice échelonnée réduite

Une matrice échelonnée est dite **réduite** si elle est échelonnée et si les premières valeurs non nulles de chaque ligne sont égales à 1. Ces valeurs là sont appelées des **pivots**.

Remarque : on peut également définir une matrice échelonnée (réduite) en raisonnant sur les colonnes.

Les matrices échelonnées, et surtout les matrices échelonnées réduites, vont jouer un rôle important dans la méthode du *Pivot de Gauss* aussi appelée *élimination de Gauss-Jordan* que l'on verra dans la prochaine section.

C'est également avec les opérations effectuées pour obtenir de telles matrices que l'on va pouvoir directement déterminer le rang ou l'inverse d'une matrice donnée (si elle est inversible!).

Exemple 3.8. Les matrices A, B et C suivantes sont respectivement échelonnée, échelonnée réduite et non échelonnée.

$$A = \begin{pmatrix} 0 & 4 & 3 & 2 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 4 & -3 & 5 \\ 0 & 1 & -2 & 7 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{et} \quad C = \begin{pmatrix} 2 & 4 & -3 & 5 \\ 4 & 1 & -2 & 7 \\ 6 & 2 & 4 & 1 \\ 8 & -5 & -2 & 1 \end{pmatrix}.$$

3.6.2 Calcul du rang

Pour déterminer le rang d'une matrice, il n'y a rien de plus simple. Il suffit d'effectuer des opérations élémentaires sur les lignes d'une matrice (ou sur les colonnes) afin d'obtenir une matrice échelonnée ou échelonnée réduite. Le rang de la matrice est alors directement égal au nombre de lignes non nulles de la matrice.

Exemple 3.9. Reprenons les matrices de l'exemple précédent

$$A = \begin{pmatrix} 0 & 4 & 3 & 2 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 4 & -3 & 5 \\ 0 & 1 & -2 & 7 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{et} \quad C = \begin{pmatrix} 2 & 4 & -3 & 5 \\ 4 & 1 & -2 & 7 \\ 6 & 2 & 4 & 1 \\ 8 & -5 & -2 & 1 \end{pmatrix}.$$

On voit directement que les matrices A et B sont de rang 3 et 4. Le rang est cependant moins évident en ce qui concerne la matrice C , on va donc effectuer des opérations élémentaires sur les lignes pour déterminer son rang.

$$C = \begin{pmatrix} 2 & 4 & -3 & 5 \\ 4 & 1 & -2 & 7 \\ 6 & 2 & 4 & 1 \\ 8 & -5 & -2 & 1 \end{pmatrix}$$

↓ On va se servir de la valeur 2 en position (1,1) pour annuler les autres lignes

$$\begin{pmatrix} 2 & 4 & -3 & 5 \\ 0 & -7 & 4 & -3 \\ 0 & -10 & 13 & -18 \\ 0 & -21 & 10 & -19 \end{pmatrix} \begin{array}{l} \\ L_2 \leftarrow L_2 - 2L_1 \\ L_3 \leftarrow L_3 - 3L_1 \\ L_4 \leftarrow L_4 - 4L_1 \end{array}$$

↓ on se sert ensuite du -7 en position (2,2) pour faire apparaître des 0 dans les deux dernières lignes

$$\begin{pmatrix} 2 & 4 & -3 & 5 \\ 0 & -7 & 4 & -3 \\ 0 & 0 & 51 & -96 \\ 0 & 0 & -2 & -10 \end{pmatrix} \begin{array}{l} \\ \\ L_3 \leftarrow 7L_3 - 10L_2 \\ L_4 \leftarrow L_4 - 3L_2 \end{array}$$

On peut remarquer que les deux dernières lignes sont indépendantes, la matrice C est donc de rang 4.

Cet exemple montre bien l'intérêt d'utiliser des matrices échelonnées réduites qui permettent de simplifier grandement les calculs.

Remarque. On peut également calculer le rang d'une matrice non carrée car il s'agit simplement de la dimension de l'espace image de l'application linéaire associée. On peut essayer de calculer le rang de la matrice suivante par exemple :

$$A = \begin{pmatrix} 1 & 4 & 7 & -2 & 0 \\ 3 & -1 & 2 & -9 & 6 \\ 7 & 2 & 4 & 5 & 2 \end{pmatrix}.$$

On se rappelle que le rang d'une matrice est toujours inférieur ou égal au minimum entre le nombre de colonnes et le nombre de lignes de la matrice. Dans cet exemple, le rang ne peut donc pas excéder 3. De façon générale, si A est une matrice avec n lignes et p colonnes, alors :

$$rg(A) \leq \inf(n, p).$$

3.6.3 Calcul du déterminant

Le calcul du déterminant est en général très complexe sauf pour des matrices très particulières. En revanche, ce calcul est très simple dans des espaces à faibles dimensions, comme en dimension 2 et 3 où des formules nous permettent de calculer très facilement le déterminant.

- **En dimension 2 :** le déterminant d'une matrice $A = (a_{ij})_{i,j=1}^2$ est donné par :

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

- **En dimension 3 :** le déterminant d'une matrice $A = (a_{ij})_{i,j=1}^3$ est donné par :

$$\begin{aligned} \det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11}a_{22}a_{33} + a_{12}a_{13}a_{31} + a_{21}a_{32}a_{13} \\ &\quad - a_{11}a_{13}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{aligned}$$

Cette règle s'appelle la *règle de Sarrus*.

Sur ces deux exemples là, on peut retenir que le déterminant est égal à *la somme des produits des éléments se trouvant sur une même diagonale étendue moins la somme des produits des éléments anti-diagonaux*. Ces deux relations vont nous servir à calculer le déterminant de matrices dont la taille est plus importante, il est donc important de bien les mémoriser.

Quelques cas particuliers. Le déterminant reste extrêmement simple à calculer pour certaines matrices, peu importe leurs tailles. Ces cas particuliers seront des cas on essaiera de se ramener à l'aide de transformation sur notre matrice afin de pouvoir en calculer aisément le déterminant

- **Matrice diagonale :** soit $A \in \mathcal{M}_n(\mathbb{K})$ une matrice diagonale, *i.e.* $a_{ij} = 0$ pour tout $i \neq j$, alors $\det(A) = \prod_{i=1}^n a_{ii}$.
On peut aussi le voir comme :

$$\det(A) = \begin{vmatrix} a_{11} & 0 & \cdots & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & a_{nn} \end{vmatrix} = a_{11}a_{22} \dots a_{nn} = \prod_{i=1}^n a_{ii}.$$

De ce résultat, on voit tout de suite qu'une condition nécessaire et suffisante pour qu'une matrice diagonale soit inversible est que ses éléments diagonaux soient non nuls.

- **Matrice triangulaire** soit $A \in \mathcal{M}_n(\mathbb{K})$ une matrice triangulaire supérieure (le résultat reste identique dans le cas des matrices triangulaires supérieures) , *i.e.* $a_{ij} = 0$ pour tout $i > j$, alors $\det(A) = \prod_{i=1}^n a_{ii}$.
On peut aussi le voir comme :

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & a_{(n-1)n} \\ 0 & 0 & \cdots & 0 & a_{nn} \end{vmatrix} = a_{11}a_{22} \dots a_{nn} = \prod_{i=1}^n a_{ii}.$$

A nouveau, une matrice triangulaire est inversible si et seulement si ses éléments diagonaux sont non nuls.

Exemple 3.10. Dans le cas des matrices de dimension 2 et 3, on peut reprendre les exemples utilisées pour illustrer géométriquement ce qu'est le déterminant.

Considérons maintenant deux matrices A et B respectivement diagonales et triangulaires supérieures de type 4, on peut alors vérifier que :

$$\det(A) = \begin{vmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{vmatrix} = 24 \quad \text{et} \quad B = \begin{vmatrix} 2 & 0 & 3 & 0 \\ 0 & 6 & 2 & 1 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{vmatrix} = 240.$$

Développement selon une colonne. Ce qui est développé dans ce paragraphe s'applique également aux lignes d'une matrice. Il s'agit de reprendre et mettre en application la Proposition 3.17.

On se propose de traiter cela une matrice quelconque $A \in \mathcal{M}_3(\mathbb{K})$ et de voir si on retrouve la règle de Sarrus. Pour cela on considère la matrice A définie par

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Et on se propose pour cela de développer selon la première colonne (nous pourrions aussi faire le choix de développer selon la première ligne). Nous avons donc :

$$\begin{aligned} \det(A) &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= a_{11}(-1)^{1+1} \begin{vmatrix} \cancel{a_{11}} & \cancel{a_{12}} & \cancel{a_{13}} \\ \cancel{a_{21}} & a_{22} & a_{23} \\ \cancel{a_{31}} & a_{32} & a_{33} \end{vmatrix} + a_{21}(-1)^{2+1} \begin{vmatrix} \cancel{a_{11}} & a_{12} & a_{13} \\ \cancel{a_{21}} & \cancel{a_{22}} & \cancel{a_{23}} \\ \cancel{a_{31}} & a_{32} & a_{33} \end{vmatrix} + a_{31}(-1)^{3+1} \begin{vmatrix} \cancel{a_{11}} & a_{12} & a_{13} \\ \cancel{a_{21}} & a_{22} & a_{23} \\ \cancel{a_{31}} & \cancel{a_{32}} & \cancel{a_{33}} \end{vmatrix}, \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}, \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{31}(a_{12}a_{23} - a_{13}a_{22}), \\ &= a_{11}a_{22}a_{33} + a_{12}a_{13}a_{31} + a_{21}a_{32}a_{13} - a_{11}a_{13}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{aligned}$$

Exemple 3.11. On se propose de calculer le déterminant de la matrice A définie par

$$A = \begin{pmatrix} 3 & 1 & -2 & 4 \\ 0 & 0 & 5 & 1 \\ 0 & 2 & 0 & 0 \\ 3 & 6 & 0 & 1 \end{pmatrix}$$

à l'aide de transformation sur les colonnes et en développant sur les colonnes (ou les lignes). Nous avons donc

$$\begin{aligned} \det(A) &= \begin{vmatrix} 3 & 1 & -2 & 4 \\ 0 & 0 & 5 & 1 \\ 0 & 2 & 0 & 0 \\ 3 & 6 & 0 & 1 \end{vmatrix}, \\ &\quad \downarrow \text{ le déterminant reste inchangé par combinaison linéaire de lignes} \\ &= \begin{vmatrix} \color{red}{3} & 1 & -2 & 4 \\ \color{red}{0} & 0 & 5 & 1 \\ \color{red}{0} & 2 & 0 & 0 \\ \color{red}{0} & 5 & 2 & -3 \end{vmatrix} \begin{matrix} \\ \\ \\ L_4 \leftarrow L_4 - L_1 \end{matrix}, \end{aligned}$$

$$\begin{aligned}
& \downarrow \text{ on développe selon la première colonne} \\
& = 3 \begin{vmatrix} 0 & 5 & 1 \\ 2 & 0 & 0 \\ 5 & 2 & -3 \end{vmatrix} \\
& \downarrow \text{ on développe selon la deuxième ligne} \\
& = 3 \times (-2) \begin{vmatrix} 5 & 1 \\ 2 & -3 \end{vmatrix}, \\
& \downarrow \text{ on développe le déterminant de taille 2} \\
& = 3 \times (-2) \times (5 \times (-3) - 1 \times 2) = 102.
\end{aligned}$$

3.6.4 Inverse d'une matrice

Une fois que l'on a calculé le déterminant, il reste à déterminer l'expression de la comatrice, *i.e.* la matrice des cofacteurs. Pour cela il faut déterminer toutes les mineures de notre matrice de départ. Pour une matrice $A \in \mathcal{M}_2(\mathbb{K})$ inversible nous avons directement :

$$\text{si } A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \text{alors} \quad A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

Pour les matrices carrées de taille 2, la transposée de la matrice des cofacteurs consiste simplement à échanger de place les éléments diagonaux et à changer le signe des éléments anti-diagonaux.

Exemple 3.12. Reprenons notre matrice A précédente et déterminons son inverse. Comme nous avons déjà calculé son déterminant ($\det(A) = 102$), il nous reste simplement à déterminer la matrice des cofacteurs A_{ij} . Une fois cette matrice obtenue, il ne faudra pas oublier de la transposer afin de déterminer l'inverse de la matrice A . On se contentera de calculer les valeurs de la première ligne de la comatrice.

$$\bullet A_{11} = (-1)^2 \Delta_{11} = \begin{vmatrix} 0 & 5 & 1 \\ 2 & 0 & 0 \\ 6 & 0 & 1 \end{vmatrix} = -2 \times \begin{vmatrix} 5 & 1 \\ 0 & 1 \end{vmatrix} = 10.$$

On aura développé selon la deuxième ligne pour calculer le déterminant de notre matrice carrée de taille 3.

$$\bullet A_{12} = (-1)^3 \Delta_{12} = - \begin{vmatrix} 0 & 5 & 1 \\ 0 & 0 & 0 \\ 3 & 0 & 1 \end{vmatrix} = 0.$$

C'est immédiat car notre matrice comprend une ligne avec des zéros uniquement, son déterminant est donc nul.

$$\bullet A_{13} = (-1)^4 \Delta_{13} = \begin{vmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 3 & 6 & 1 \end{vmatrix} = \begin{vmatrix} 0 & 2 \\ 3 & 6 \end{vmatrix} = -6.$$

On aura développé selon la première ligne pour obtenir notre nouveau déterminant de taille 2.

$$\bullet A_{14} = (-1)^5 \Delta_{14} = - \begin{vmatrix} 0 & 0 & 5 \\ 0 & 2 & 0 \\ 3 & 6 & 0 \end{vmatrix} = -5 \times \begin{vmatrix} 0 & 2 \\ 3 & 6 \end{vmatrix} = 30.$$

On aura développé selon la première ligne à nouveau.

En calculant tous les coefficients, on montre alors que la matrice des cofacteurs est donnée par :

$$\text{com}(A) = \begin{pmatrix} -10 & 0 & -6 & 30 \\ -4 & 0 & 18 & 12 \\ -127 & 51 & -15 & 75 \\ 44 & 0 & 6 & -30 \end{pmatrix}.$$

Il existe aussi une méthode plus pratique pour calculer l'inverse d'une matrice qui repose sur le fait que *les opérations élémentaires sur les lignes et les colonnes sont des automorphismes, ils ne changent donc pas le caractère inversible d'une matrice donnée*. On peut se servir de cela pour trouver une suite d'opérations élémentaires qui va permettre de transformer une matrice A inversible en la matrice identité. Si on applique, en parallèle, les mêmes transformations à la matrice identité, on sera mesure de déterminer la matrice inverse de la matrice A .

Deux choix s'offrent à nous en terme d'écriture, mais ce choix va conditionner le travail à effectuer et inversement :

- si on souhaite effectuer des opérations élémentaires sur **les lignes de la matrice de A** , on travaillera sur la *matrice étendue* suivante

$$(A \mid I) = \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1n} & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} & 0 & \cdots & 1 \end{array} \right),$$

afin d'aboutir, via une succession de manipulations sur les lignes à une matrice étendue de la forme

$$(I \mid A') = \left(\begin{array}{ccc|ccc} 1 & \cdots & 0 & a'_{11} & \cdots & a'_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & a'_{n1} & \cdots & a'_{nn} \end{array} \right),$$

où $A' = A^{-1}$ désigne l'inverse de la matrice A .

- si on souhaite effectuer des opérations élémentaires sur **les colonnes de la matrice de A** , on travaillera sur la *matrice étendue* suivante

$$\left(\frac{A}{I}\right) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \\ 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{pmatrix}.$$

Regardons cela sur un exemple à nouveau.

Exemple 3.13. Soit $A \in \mathcal{M}_3(\mathbb{R})$ une matrice que l'on admettra inversible (vous pourrez calculer son déterminant afin de vous en convaincre), définie par :

$$A = \begin{pmatrix} 5 & 2 & 2 \\ 0 & 3 & 5 \\ 0 & 0 & 1 \end{pmatrix}.$$

La matrice A est triangulaire supérieure dans cet exemple, on se propose de déterminer son inverse en effectuant des opérations élémentaires sur les lignes. On considère la matrice étendue suivante :

$$\left(\begin{array}{ccc|ccc} 5 & 2 & 2 & 1 & 0 & 0 \\ 0 & 3 & 5 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right).$$

On commence par se servir de la valeur 1 en bas à droite de notre matrice à inverser.

On oubliera pas de répercuter les mêmes transformations sur la matrice identité !

$$\left(\begin{array}{ccc|ccc} 5 & 2 & 0 & 1 & 0 & -2 \\ 0 & 3 & 0 & 0 & 1 & -5 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right) \begin{array}{l} L_1 \leftarrow L_1 - 2L_3 \\ L_2 \leftarrow L_2 - 5L_3. \end{array}$$

On va ensuite faire apparaître la valeur 1 dans la deuxième ligne et deuxième colonne. On garde à l'esprit que l'on doit faire apparaître la matrice identité à gauche et on itère ce même type d'opérations.

$$\left(\begin{array}{ccc|ccc} 5 & 2 & 0 & 1 & 0 & -2 \\ 0 & 1 & 0 & 0 & 1/3 & -5/3 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right) L_2 \leftarrow L_2/3,$$

$$\left(\begin{array}{ccc|ccc} 5 & 0 & 0 & 1 & -2/3 & 4/3 \\ 0 & 1 & 0 & 0 & 1/3 & -5/3 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right) L_1 \leftarrow L_1 - 2L_2,$$

$$\left(\begin{array}{ccc|ccc} \textcolor{red}{1} & 0 & 0 & \textcolor{red}{1/5} & \textcolor{red}{-2/15} & \textcolor{red}{4/15} \\ 0 & 1 & 0 & 0 & 1/3 & -5/3 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right) \textcolor{blue}{L_1 \leftarrow L_1/5}.$$

On a donc

$$A^{-1} = \begin{pmatrix} 1/5 & -2/15 & 4/15 \\ 0 & 1/3 & -5/3 \\ 0 & 0 & 1 \end{pmatrix}.$$

Vérifier ses connaissances

1. Les propositions suivantes sont-elles vraies ou fausses ?
 - (a) Une colonne d'une matrice de $\mathcal{M}_{n,p}(\mathbb{R})$ est un vecteur de \mathbb{R}^p .
 - (b) Une ligne d'une matrice de $\mathcal{M}_{n,p}(\mathbb{R})$ est un vecteur de \mathbb{R}^n .
 - (c) Une matrice non nulle n'a pas de coefficient nul.
 - (d) Toute matrice carrée admet un inverse
 - (e) Les matrices diagonales sont les matrices carrées à la fois triangulaires inférieures et triangulaires supérieures.
 - (f) Dans $\mathcal{M}_n(\mathbb{R})$, la multiplication matricielle est commutative.
 - (g) Le produit d'une matrice par une matrice nulle est nul.
 - (h) Si A et B sont des matrices telles que le produit AB ait un sens alors

$$AB = 0 \iff A = 0 \quad \text{ou} \quad B = 0.$$

- (i) Si A est une matrice inversible de $\mathcal{M}_n(\mathbb{R})$ alors pour tout B

$$AB = 0 \iff B = 0.$$

- (j) Le rang d'une matrice de $\mathcal{M}_{n,p}(\mathbb{R})$ est inférieure à $\inf(p, n)$.
 - (k) La transposée d'un produit de matrices est le produit des transposées de ces matrices.
 - (l) L'inverse d'un produit de matrices inversibles est le produit des inverses de ces matrices.
2. Préciser la base canonique de $\mathcal{M}_{3,2}(\mathbb{R})$

3. Calculer le produit AB où

$$A = \begin{pmatrix} 1 & 3 & 0 \\ 0 & -2 & -1 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} 5 & -1 & -3 \\ 1 & 1 & 4 \\ -2 & 3 & 2 \end{pmatrix}.$$

4. Soit E un espace vectoriel de dimension finie $n \in \mathbb{N}^*$. Les propositions suivantes sont-elles vraies ou fausses ?
 - (a) La matrice de Id_E relativement à un couple de bases de E est I_n .
 - (b) La matrice de Id_E relativement à une base de E est I_n .
 - (c) La matrice d'un automorphisme de E relativement à un couple de bases de E est inversible.

- (d) La matrice d'un automorphisme de E relativement à une base de E est inversible.
5. Soit $P \in \mathcal{M}_n(\mathbb{R})$. Les propositions suivantes sont-elles équivalentes ? Est-ce que l'une implique l'autre ?
- (a) A : P est inversible, B : P est de rang n .
 - (b) A : P est inversible,
B : Pour tout $\mathbf{x} = (x_1, \dots, x_n)^T$, $P\mathbf{x} = \mathbf{0} \implies \mathbf{x} = \mathbf{0}$
 - (c) A : P est inversible,
B : aucune colonne de P n'est nulle
 - (d) A : P est inversible,
B : L'endomorphisme de \mathbb{R}^n canoniquement associé à P est un automorphisme.
6. Soient A et B deux matrices diagonales, que vaut le produit AB ?
7. Si D est une matrice diagonale, que vaut D^p pour $p \in \mathbb{N}$?
8. Justifier que deux matrices A et B de $\mathcal{M}_{n,p}(\mathbb{R})$ sont égales si et seulement si pour tout $\mathbf{x} \in \mathbb{R}^p$, on a $A\mathbf{x} = B\mathbf{x}$.
9. Soit E un espace vectoriel de dimension finie $n \geq 1$. Les propositions suivantes sont-elles vraies ou fausses ?
- (a) Si \mathcal{B} et \mathcal{B}' sont deux bases de E alors $\det_{\mathcal{B}}(\mathcal{B}') = 1$.
 - (b) Si \mathcal{B} est une base de E alors $\det_{\mathcal{B}}(\mathcal{B}) = 1$.
 - (c) Le déterminant de Id_E est égal à 1.
 - (d) Si $u \in \mathcal{L}(E)$ est tel que $\det(u) = 1$ alors $u = Id_E$.
 - (e) Si $u \in \mathcal{L}(E)$ est tel que $\det(u) = 1$ alors $u \in \mathcal{GL}(E)$.
 - (f) Si $u \in \mathcal{L}(E)$ alors $\forall \lambda \in \mathbb{K}$, $\det(\lambda u) = \lambda \det(u)$.
10. On considère $M \in \mathcal{M}_n(\mathbb{K})$. Les propositions A et B sont-elles équivalentes, l'une implique-t-elle l'autre ?
- (a) A : $\det(M) = 0$,
B : une des colonnes de M est nulle.
 - (b) A : $\det(M) = 0$,
B : deux colonnes de M sont proportionnelles.

- (c) A : $\det(M) = 0$,
B : une des colonnes de M est combinaison linéaire des autres colonnes.

11. Déterminer le rang des matrices suivantes :

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 8 & 5 & 0 \\ -12 & 3 & 14 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 3 \\ 1 & 1 & 3 \end{pmatrix} \quad \text{et} \quad C = \begin{pmatrix} 1 & 1 & 1-s \\ 1+s & -1 & 2 \\ 2 & -s & 3 \end{pmatrix}, \quad \text{où } s \in \mathbb{R}.$$

12. Montrer que la matrice suivante est inversible et déterminer son inverse

$$\begin{pmatrix} 0 & 1 & 2 \\ 2 & 2 & 3 \\ 4 & 0 & 1 \end{pmatrix}.$$

13. Soient $(a, b) \in \mathbb{R}^2$. Calculer le déterminant des matrices suivantes et préciser si elles sont inversibles et à quelle(s) condition(s).

$$A = \begin{pmatrix} 1 & a \\ 1 & b \end{pmatrix}, \quad B = \begin{pmatrix} 1 & a & b \\ a & 1 & b \\ b & a & 1 \end{pmatrix} \quad \text{et} \quad C = \begin{pmatrix} a & 0 & 0 & b \\ b & 1 & 1 & a \\ a & 1 & 1 & b \\ 0 & 2 & 1 & a \end{pmatrix}.$$

Les matrices et opérations sur

Matrices et opérations

Nous complétons notre première partie pratique en introduisant cette fois-ci les matrices, qui représentent des applications linéaires ou que l'on peut aussi voir comme une famille de vecteurs.

On regardera aussi comment faire des opérations sur les matrices : *somme et produits de matrices, calcul du rang, calcul de déterminant* ou encore le *calcul de l'inverse d'une matrice*.

Commençons par regarder comment définir une matrice. Nous avons vu que cela pouvait être la concaténation de plusieurs vecteurs. Prenons l'exemple de la matrice suivante :

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

On peut faire cela de plusieurs façons :

```
# Concaténation en ligne ...
A = rbind(c(1,2,3),c(4,5,6),c(7,8,9))
# ou en colonne.
A = cbind(c(1,4,7),c(2,5,8),c(3,6,9))
# On peut aussi écrire :
A = matrix(c(1:9), nrow = 3, ncol = 3, byrow = TRUE)
# Affichage
A

##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    4    5    6
## [3,]    7    8    9
```

L'objet ainsi défini est bien une matrice de taille 3×3 . Regardons chaque définition :

- la fonction **rbind** permet de concaténer des vecteurs en **ligne** (r pour *row*).
- la fonction **cbind** permet de concaténer des vecteurs en **colonne** (c pour *column*).

Attention, pour faire cela, il faut que tous les vecteurs aient la même dimension. Il est également possible de définir directement la matrice avec la fonction **matrix** en précisant : *les éléments de la matrice*, son nombre de ligne et de colonne et la façon dont

elle est remplie.

Il est également possible d'accéder à un élément particulier d'une matrice, que cela soit une ligne (resp. une colonne), un ensemble de lignes (resp. de colonnes) ou encore une entrée spécifique de la matrice. Pour cela reprenons notre matrice précédente

```
# Première ligne
A[1,]

## [1] 1 2 3

# Deuxième colonne
A[,2]

## [1] 2 5 8

# Lignes 1 et 3
A[c(1,3),]

##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    7    8    9

# Colonnes 2 et 3
A[,c(2,3)]

##      [,1] [,2]
## [1,]    2    3
## [2,]    5    6
## [3,]    8    9

# a_{23} : deuxième ligne, troisième colonne
A[2,3]

## [1] 6
```

On peut maintenant regarder quelques opérations sur les matrices :

- la multiplication par un scalaire $\alpha = 3$ soit αA

```
# Multiplication d'une matrice par un scalaire
alpha = 3
alpha * A
##      [,1] [,2] [,3]
## [1,]    3    6    9
## [2,]   12   15   18
## [3,]   21   24   27
```

A nouveau, comme pour un vecteur, il s'agit de multiplier chaque élément de la matrice a_{ij} par la valeur de α .

- le produit *matriciel* : on considère la matrice A définie précédemment et la matrice B définie par

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

cela s'écrit :

```
# Produit de deux matrices
B = rbind(c(1,0,0),c(0,3,0),c(0,0,2))
A%%B
##      [,1] [,2] [,3]
## [1,]    1    6    6
## [2,]    4   15   12
## [3,]    7   24   18
```

On retrouve le produit matriciel tel que défini plus tôt dans cette section.

- le produit *d'Hadamard*

```
# Produit de deux matrices
A*B
##      [,1] [,2] [,3]
## [1,]    1    0    0
## [2,]    0   15    0
## [3,]    0    0   18
```


On prendra garde à faire la distinction entre les deux opérateurs utilisés. En pratique, nous utiliserons uniquement le produit matriciel "standard" avec l'opérateur `%*%`.

On pourra enfin calculer la *transposée* d'une matrice avec la fonction `t`.

```
# Transposée
t(A)
```

```
##      [,1] [,2] [,3]
## [1,]    1    4    7
## [2,]    2    5    8
## [3,]    3    6    9
```

Rang, trace, déterminant et inverse

Il existe des fonctions simples qui permettent d'effectuer différentes opérations simple sur les matrices. Malheureusement, certaines d'entre elles nécessitent l'installation d'un package spécifique sur . C'est par exemple le cas pour le calcul du rang d'une matrice qui nécessite le package *Matrix*.

```
# Installation du package
install.packages("Matrix")
# Chargement de la librairie
library(Matrix)
```

Notez qu'une fois que le package est installé sur votre ordinateur, il n'est plus nécessaire de le réinstaller ! Il suffira simplement de charger la librairie.

Pour le calcul du rang, reprenons l'Exemple 3.9, il suffira alors d'appliquer la fonction *rankMatrix* à la matrice.

```
C = rbind(c(2,4,-3,5),
          c(4,1,-2,7),
          c(6,2,4,1),
          c(8,-5,-2,1))
# Calcul du rang
rankMatrix(C)
```

Regardons maintenant comment déterminer la trace et le déterminant d'une matrice. On va reprendre la matrice *A* définie dans l'Exemple 3.11.

```
A = rbind(c(3,1,-2,4),
          c(0,0,5,1),
          c(0,2,0,0),
          c(3,6,0,1))
# Calcul de la trace
sum(diag(A))

## [1] 4
```



```
# Calcul du déterminant
det(A)
```

```
## [1] 102
```

La fonction *diag* permet d'extraire les éléments diagonaux d'une matrice carrée et la fonction *sum* permet de faire la somme des éléments d'un objet.
Par exemple :

```
# Extraction de la diagonale de A
diag(A)
```

```
## [1] 3 0 0 1
```

```
# Trace de A
sum(diag(A))
```

```
## [1] 4
```

```
# Somme des éléments de A
sum(A)
```


```
## [1] 24
```

la fonction *diag* renvoie donc un élément qui est un *vecteur*. La fonction *sum* va faire la somme de ces éléments. Sur un vecteur, elle va donc faire la somme des éléments du vecteur. Sur une matrice, elle fera la somme de toutes les entrées de la matrice.

Enfin, reprenons maintenant l'Exemple 3.13 pour regarder comment calculer l'inverse d'une matrice.

```
A = rbind(c(5,2,2),c(0,3,5),c(0,0,1))
# Calcul de l'inverse
solve(A)
```

```
##      [,1]      [,2]      [,3]
## [1,] 0.2 -0.1333333 0.2666667
## [2,] 0.0 0.3333333 -1.6666667
## [3,] 0.0 0.0000000 1.0000000
```

On retrouve bien le résultat que nous avons trouvé à la main dans cet exemple. Bien évidemment  calculera l'inverse d'une matrice carrée inversible uniquement.

4 Systèmes linéaires

4.1 Définition et interprétation

Définition 4.1: Système linéaire

On appelle *système d'équations linéaires* de n équations à p inconnues tout système S de la forme

$$S : \begin{cases} a_{11}x_1 + \dots + a_{1p}x_p &= b_1 \\ a_{21}x_1 + \dots + a_{2p}x_p &= b_2 \\ \vdots & \\ a_{n1}x_1 + \dots + a_{np}x_p &= b_n \end{cases}$$

où

- les scalaires x_1, \dots, x_p sont les inconnues du système d'équations,
- les coefficients $(a_{ij})_{i,j=1}^{n,p}$ sont les coefficients du système,
- le vecteur $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{K}^n$ est le second membre du système.

L'objectif de cette section est de voir comment nous pourrions résoudre ce type de système mais aussi comment vont s'interpréter les solutions.

Pour cela, commençons par réécrire le système sous une forme en lien avec la notion développée dans la précédente section, *i.e.* sous forme matricielle. En effet, si l'on note $A \in \mathcal{M}_{n,p}(\mathbb{K})$ la matrice associée aux coefficients a_{ij} , \mathbf{x} le vecteur des inconnues x_j et \mathbf{b} le vecteur des scalaires qui constituent le second membre de notre système, nous pouvons réécrire S sous la forme

$$A\mathbf{x} = \mathbf{b}.$$

Tout comme on a introduit la notion de rang pour une matrice, on peut aussi parler de rang pour un système d'équations linéaires. Ainsi, le rang de la matrice A est égal au rang du système S .

Dans le cas où le second membre est nul ($\mathbf{b} = \mathbf{0}$), le système S est dit homogène et s'écrit

$$A\mathbf{x} = \mathbf{0}.$$

Si le second membre n'est pas nul ($\mathbf{b} \neq \mathbf{0}$), le système S_0 défini par $A\mathbf{x} = \mathbf{0}$ est appelé système homogène associé à S . On va voir qu'étudier un tel système va permettre de définir l'ensemble des solutions d'un système d'équations linéaires.

Interprétation. On se rappelle qu'à toute matrice A est associée une application linéaire, disons u . Ainsi le système S peut aussi s'écrire sous la forme $u(\mathbf{x}) = \mathbf{b}$ et le

système d'équation homogène associé S_0 s'écrira $u(\mathbf{x}) = \mathbf{0}$.

Cela permet de faire un premier lien avec l'algèbre linéaire et les solutions d'un système homogène :

Les solutions d'un système homogène S_0 sont exactement les éléments de $\text{Ker}(u)$.

De plus, la nouvelle écriture du système S nous indique que ce dernier admet une solution si et seulement si $\mathbf{b} \in \text{Im}(u)$. Si tel est le cas, notons \mathbf{x}_0 une solution *particulière* du système S , nous avons donc

$$u(\mathbf{x}) = u(\mathbf{x}_0) \iff u(\mathbf{x} - \mathbf{x}_0) = \mathbf{0} \iff \mathbf{x} - \mathbf{x}_0 \in \text{Ker}(u) \iff \exists K \in \text{Ker}(u) \mid \mathbf{x} = \mathbf{x}_0 + K.$$

On peut donc faire la remarque suivante :

L'ensemble des solutions du système S est donc $\{\mathbf{x}_0 + K \mid K \in \text{Ker}(u)\}$. On peut dire que les solutions de S sont obtenues en trouvant une solution particulière à laquelle on ajoute une solution du système homogène associée.

Remarque : cette observation est également valable dans d'autres contextes en mathématiques :

- lorsque l'on cherche à déterminer les solutions d'une équation séquentielle de la forme

$$av_{n+1} + bv_n + cv_{n-1} = d(n),$$

- ou encore lorsque l'on cherche à déterminer les solutions d'une équation différentielle (ce qui arrive très souvent en physique) :

$$ay''(t) + by'(t) + cy(t) = f(t).$$

4.2 Résolution d'un système linéaire

On va maintenant regarder comment déterminer les solutions d'un système linéaire pour des systèmes dits de *Cramer* et pour des systèmes d'équations linéaires quelconques. Pour de tels systèmes, trois cas se présentent à nous en ce qui concerne l'espace des solutions :

- Le système peut ne pas avoir de solutions (sauf si c'est un système *homogène*, auquel cas le vecteur nul est toujours solution !)
- le système admet une unique solution
- le système admet une infinité de solutions

Sans informations supplémentaire sur la nature du système, il est difficile de savoir si le système admet ou non des solutions.

On peut cependant être plus précis pour les systèmes homogènes d'**équations linéaires indépendantes**. Dans ce cas :

- le système admet une seule et unique solution s'il y a autant d'inconnues que d'équations
- le système admet une infinité de solutions lorsque le nombre d'équations est strictement inférieur au nombre d'inconnues.

Cette dernière remarque est très intéressante car elle permet de déterminer si un système d'équations linéaires homogènes admet des solutions ou non uniquement en regardant le *rang de la matrice A associée au système d'équations linéaires homogènes*.

Il est également possible de caractériser la dimension de l'espace des solutions pour un système linéaire non homogène, c'est le résultat du **Théorème de Rouché-Fontené**

Théorème 4.1: Théorème de Rouché-Fontené

Soit S un système d'équations linéaires à n variables, de la forme $A\mathbf{x} = \mathbf{b}$. Ce système possède une solution, si et seulement si le rang de la matrice A est égal à celui de la matrice augmentée $(A \mid \mathbf{b})$. Si des solutions existent, elles forment un sous-espace **affine** de \mathbb{R}^n de dimension $n - \text{rg}(A)$.

De plus :

- si $n = \text{rg}(A)$, alors S admet une unique solution
- sinon, il existe une infinité de solutions.

Résolution d'un système de Cramer

Définition 4.2: Système de Cramer

On appelle **système de Cramer** tout système de n équations à n inconnues de rang n , $n \in \mathbb{N}^*$.

Un système de Cramer est donc un système S de la forme $A\mathbf{x} = \mathbf{b}$ où $A \in \mathcal{M}_n(\mathbb{K})$ est inversible. Comme la matrice A est inversible, on peut définir son inverse A^{-1} qui vérifie

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

Cette relation montre que tout système de Cramer admet une seule et unique solution définie par la relation précédente. De plus, la matrice A étant inversible, le système homogène associé n'admet pour solution que la solution triviale $\mathbf{x} = \mathbf{0}$. En effet A inversible implique que le noyau de l'application linéaire associé est réduit à $\mathbf{0}$.

Nous avons vu comment déterminer A^{-1} dans la section traitant des matrices, ce qui nous suffirait pour résoudre un tel système. Mais il existe un résultat qui permet aussi de donner immédiatement les solutions d'un tel système. Plus précisément :

Proposition 4.1: Solutions(s) d'un Système de Cramer

Considérons le système de Cramer S suivant :

$$S : \begin{cases} a_{11}x_1 + \dots + a_{1n}x_n &= b_1 \\ \vdots & \\ a_{n1}x_1 + \dots + a_{nn}x_n &= b_n \end{cases}$$

dont l'écriture matricielle est $A\mathbf{x} = \mathbf{b}$ où $A \in \mathcal{GL}_n(\mathbb{K})$. Alors l'unique solution de S est le n -uplet (x_1, \dots, x_n) défini par :

$$x_k = \frac{\det(A_k)}{\det(A)}, \quad \forall k \in \llbracket 1, n \rrbracket,$$

où A_k est la matrice A dans laquelle on a remplacé la k -ème colonne par le vecteur \mathbf{b} .

Exemple 4.1. On considère le système d'équations linéaires suivant que l'on se propose de résoudre.

$$S : \begin{cases} 2x_1 + 3x_2 = 2 \\ x_1 - 5x_2 = 1 \end{cases}$$

On commence par montrer que ce système $A\mathbf{x} = \mathbf{b}$ est un système de Cramer donc que la matrice A définie par

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

est inversible. Cela se vérifie facilement en calculant le déterminant qui est ici égal à -13 . On peut donc utiliser le résultat de la proposition précédente pour déterminer la solution de ce système.

$$\begin{aligned}
\bullet x_1 &= \frac{\det(A_1)}{\det(A)} = \frac{\begin{vmatrix} 2 & 3 \\ 1 & -5 \end{vmatrix}}{\begin{vmatrix} 2 & 3 \\ 1 & -5 \end{vmatrix}} = \frac{-13}{-13} = 1, \\
\bullet x_2 &= \frac{\det(A_2)}{\det(A)} = \frac{\begin{vmatrix} 2 & 2 \\ 1 & 1 \end{vmatrix}}{\begin{vmatrix} 2 & 3 \\ 1 & -5 \end{vmatrix}} = \frac{0}{-13} = 0.
\end{aligned}$$

Résolution d'un système de Cramer triangulaire Il s'agit là d'un cas particulier de Système de Cramer où la matrice A associée est triangulaire (triangulaire supérieure par exemple). Notre système s'écrit alors

$$S : \begin{cases} a_{11}x_1 + \dots + a_{1n}x_n &= b_1 \\ a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \vdots & \\ a_{n-1,1}x_{n-1} + a_{n-1,n}x_n &= b_{n-1} \\ a_{nn}x_n &= b_n \end{cases}$$

C'est un système particulier qui peut se résoudre "de bas en haut". En effet, il s'agit d'un système de Cramer, la matrice A est donc inversible. Or elle est triangulaire supérieure, cela implique que tous éléments diagonaux de la matrice sont non nuls. La résolution donne :

$$\begin{aligned}
\bullet x_n &= \frac{b_n}{a_{nn}} \\
\bullet x_{n-1} &= \frac{1}{a_{n-1,n-1}}(b_{n-1} - a_{n-1,n}x_n) \\
\bullet \dots & \\
\bullet x_1 &= \frac{1}{a_{11}}(b_1 - a_{12}x_2 - \dots - a_{1n}x_n).
\end{aligned}$$

Ce type de système est donc très simple à résoudre car il suffit d'utiliser les lignes du "bas vers le haut" pour déterminer les différentes valeurs de x_i .

Exemple 4.2. Résoudre le système d'équations linéaires suivant :

$$S : \begin{cases} 2x_1 + 3x_2 - x_3 &= 7 \\ -2x_2 + 3x_3 &= -1 \\ 7x_3 &= -7 \end{cases}$$

On va donc déterminer x_3 à l'aide de la troisième équation puis x_2 et enfin x_1 , ce qui nous donne :

$$x_3 = -1,$$

$$x_2 = -\frac{1}{2}(-1 - 3x_3) = -\frac{1}{2}(-1 + 3) = -1,$$

$$x_1 = \frac{1}{2}(7 - 3x_2 + x_3) = \frac{1}{2}(7 + 3 - 1) = \frac{9}{2}.$$

Comme on peut le voir, résoudre un système triangulaire de Cramer est beaucoup plus simple que de résoudre un système de Cramer "classique". En effet, nous n'avons plus besoin de calculer plusieurs déterminants, ce qui, à haute dimension, peut se révéler très fastidieux. On va donc voir dans la prochaine section comment se ramener à un tel système de triangulaire, pas forcément de Cramer, nous allons donc traiter le cas général

Méthode du Pivot de Gauss Considérons un système d'équations linéaires indépendantes suivant :

$$S : \begin{cases} a_{11}x_1 + \dots + a_{1p}x_p = b_1 \\ a_{21}x_1 + \dots + a_{2p}x_p = b_2 \\ \vdots \\ a_{n1}x_1 + \dots + a_{np}x_p = b_n \end{cases},$$

où $(b_1, \dots, b_n) \neq 0$. D'après la remarque en début de section nous avons vu que :

- si $p < n$, alors le système n'admet pas de solutions
- si $p = n$, alors le système admet une seule et unique solution
- si $p > n$, alors le système admet une infinité de solutions

On va se placer dans le cas où $p = n$ et on va donc essayer de se ramener à un système triangulaire supérieur à l'aide de la **méthode du pivot de Gauss**. On va donc effectuer des opérations élémentaires sur les lignes de la matrice A associée au système :

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \longrightarrow \begin{pmatrix} a'_{11} & a'_{12} & \dots & a'_{1n} \\ 0 & a'_{22} & \dots & a'_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a'_{nn} \end{pmatrix}.$$

Pour obtenir un tel système triangulaire, on va servir de a_{11} comme pivot afin d'annuler tous les éléments se trouvant dans la même colonne que a_{11} , *i.e.* tous les éléments de la forme a_{k1} . Concrètement on va effectuer les opérations :

$$A \rightarrow \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} - \frac{a_{21}}{a_{11}}a_{12} & \dots & a_{2n} - \frac{a_{21}}{a_{11}}a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} - \frac{a_{21}}{a_{11}}a_{12} & \dots & a_{nn} - \frac{a_{21}}{a_{11}}a_{1n} \end{pmatrix} \begin{matrix} \\ L_2 \leftarrow L_2 - \frac{a_{21}}{a_{11}}L_1 \\ \\ L_n \leftarrow L_n - \frac{a_{n1}}{a_{11}}L_1 \end{matrix}.$$

Notre objectif est faire apparaître des 0 sur tous les éléments de la colonne se trouvant sous le pivot afin de se ramener à une matrice échelonnée. On va ensuite appliquer ce principe colonne par colonne mais en n'oubliant de descendre d'une ligne à chaque fois. En effet, si nous ne faisons pas cela, nous risquons de refaire apparaître des valeurs non nulles !

Ensuite, on va donc se servir de la première valeur non nulle, *i.e.* $a_{22} - \frac{a_{21}}{a_{11}}a_{12}$, comme nouveau pivot pour annuler toutes les autres valeurs se trouvant dans la même colonne mais pour des indices de lignes supérieurs à la ligne du pivot. Et on continue ainsi de suite jusqu'à l'obtention d'un système triangulaire, ce qui nous donne :

$$A \rightarrow \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} - \frac{a_{21}}{a_{11}}a_{12} & \dots & a_{2n} - \frac{a_{21}}{a_{11}}a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} - \frac{a_{21}}{a_{11}}a_{1n} - \gamma \end{pmatrix},$$

$$\text{où } \gamma = \frac{a_{n2} - \frac{a_{21}}{a_{11}}a_{12}}{a_{22} - \frac{a_{21}}{a_{11}}a_{12}} \left(a_{2n} - \frac{a_{21}}{a_{11}}a_{1n} \right).$$

Dans le cas où le système n'est pas homogène, les opérations effectuées sur les lignes doivent aussi s'effectuer sur les lignes du vecteur **b** !

Remarque : on remarque que dans la présentation effectuée de la méthode du pivot de Gauss, on essaie de faire intervenir des matrices échelonnées. Il est parfois d'usage de faire apparaître des matrices échelonnées réduites, *i.e.* la première valeur non nulle d'une ligne donnée est un 1. Cela permet de simplifier les calculs dans certaines situations. Cela donnerait :

$$\begin{aligned}
A &\longrightarrow \begin{pmatrix} 1 & \frac{a_{12}}{a_{11}} & \dots & \frac{a_{1n}}{a_{11}} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} L_1 \leftarrow L_1/a_{11} \\
&\quad \downarrow \text{ on sert maintenant de la valeur 1 comme pivot} \\
&\longrightarrow \begin{pmatrix} 1 & \frac{a_{12}}{a_{11}} & \dots & \frac{a_{1n}}{a_{11}} \\ 0 & a_{22} - \frac{a_{21}a_{12}}{a_{11}} & \dots & a_{2n} - \frac{a_{21}a_{1n}}{a_{11}} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2} - \frac{a_{n1}a_{12}}{a_{11}} & \dots & a_{nn} - \frac{a_{n1}a_{1n}}{a_{11}} \end{pmatrix} \begin{aligned} L_2 &\leftarrow L_2 - a_{21}L_1 \\ & \\ L_n &\leftarrow L_n - a_{n1}L_1 \end{aligned}
\end{aligned}$$

Et ainsi de suite ... regardons cela sur quelques exemples.

Exemple 4.3. On se propose de résoudre le système S défini par :

$$S : \begin{cases} 2x_1 - 4x_2 - 2x_3 = 8 \\ 5x_1 - 2x_2 + 3x_3 = -1 \\ -x_1 + 3x_2 + 7x_3 = -2 \end{cases}$$

On peut réécrire notre système sous la forme :

$$(A \mid \mathbf{b}) = \left(\begin{array}{ccc|c} 2 & -4 & -2 & 8 \\ 5 & -2 & 3 & -1 \\ -1 & 3 & 7 & -2 \end{array} \right).$$

On peut vérifier qu'il s'agit d'un système d'équations indépendantes de trois équations à trois inconnus, c'est donc bien un système de Cramer. On va appliquer la méthode du pivot de Gauss afin de transformer la matrice A en une matrice triangulaire supérieure

$$\begin{aligned}
A &\longrightarrow \left(\begin{array}{ccc|c} 2 & -4 & -2 & 8 \\ 0 & 8 & 8 & -21 \\ 0 & 1 & 6 & 2 \end{array} \right) \begin{aligned} L_2 &\leftarrow L_2 - 2.5L_1 \\ L_3 &\leftarrow L_3 + 0.5L_1 \end{aligned} \\
&\quad \downarrow \text{ on sert ensuite de la valeur 8 comme pivot} \\
&\longrightarrow \left(\begin{array}{ccc|c} 2 & -4 & -2 & 8 \\ 0 & 8 & 8 & -21 \\ 0 & 0 & 5 & \frac{37}{8} \end{array} \right) L_3 \leftarrow L_3 - \frac{1}{8}L_2
\end{aligned}$$

Notre système S est donc équivalent au système

$$S : \begin{cases} 2x_1 - 4x_2 - 2x_3 &= 8 \\ 8x_2 + 8x_3 &= -21 \\ 5x_3 &= \frac{37}{8} \end{cases}$$

que l'on peut alors résoudre "de bas en haut".

Ce qui nous donne le triplet solution suivant

$$S : \begin{cases} x_1 &= -\frac{435}{200} \\ x_2 &= -\frac{20}{37} \\ x_3 &= \frac{37}{40} \end{cases}$$

Dans cet exemple l'espace des solutions est un **point** dans un espace de dimension 3.

On procédera exactement de la même façon dans le cas où $p > n$, seule la nature des solutions va changer. En effet, nous aboutirons à un système "pseudo-triangular" dont la matrice A associée sera de la forme

$$A = \begin{pmatrix} a'_{11} & a'_{12} & \cdots & a'_{1n} & \cdots & a_{1p} \\ 0 & a'_{22} & \cdots & a'_{2n} & \cdots & a_{2p} \\ \vdots & \ddots & \ddots & \vdots & & \vdots \\ 0 & \cdots & 0 & a'_{nn} & \cdots & a_{np} \end{pmatrix}$$

Un tel système n'admet cependant pas une unique solution, mais des solutions paramétrées. On pourra par exemple poser $(x_{n+1}, \dots, x_p) = (t_{n+1}, \dots, t_p) \in \mathbb{R}^{p-n}$, qui deviennent des paramètres quelconques du système.

On pourra ensuite résoudre notre système comme un système triangulaire classique.

Remarque : dans ce cas, l'espace des solutions ne se limite pas à un point, mais cela sera un sous-espace dit *affine* de dimension $p - n$. La dimension est égale au nombre de paramètres, *i.e.* à la différence entre le nombre d'inconnus du système et le nombre d'équations indépendantes, soit $p - n$.

Exemple 4.4. On se propose de résoudre le système S défini par :

$$S : \begin{cases} x_1 - x_2 - 2x_3 &= -1 \\ 3x_1 - 2x_2 + 3x_3 &= 2 \end{cases}$$

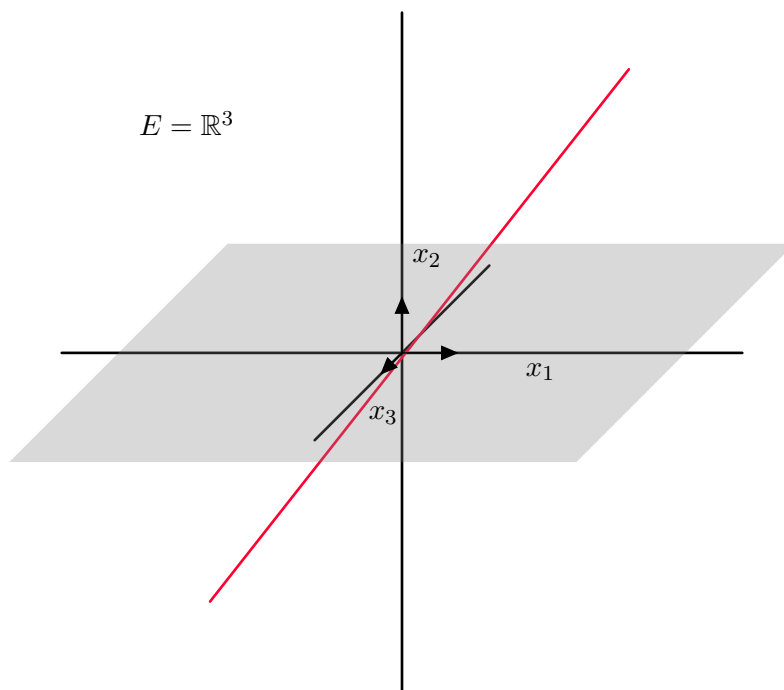
On peut facilement voir que les deux équations sont linéairement indépendantes. On va se servir de la valeur 1 comme pivot et on obtient le système équivalent

$$S : \begin{cases} x_1 - x_2 - 2x_3 &= -1 \\ x_2 + 9x_3 &= 5 \end{cases} : L_2 \leftarrow L_2 - 3L_1$$

La deuxième équation comporte deux inconnues, on va fixer $x_3 = t \in \mathbb{R}$ qui sera donc un paramètre. On aura donc :

$$S : \begin{cases} x_1 &= -1 + 2t + x_2 \\ x_2 &= 5 - 9t \end{cases}, \longrightarrow \begin{cases} x_1 &= 4 - 7t \\ x_2 &= 5 - 9t \end{cases}$$

L'espace des solutions est ici un espace de dimension 1 dont une représentation est donnée ci-dessous :



Attention, contrairement à ce que le graphique laisse croire, la droite des solutions ne passent pas par l'origine du repère.

En effet, on rappelle que dans ce cas, les solutions sont données par la somme d'une solution particulière à laquelle on ajoute les vecteurs de la forme $\lambda \mathbf{v}$ où \mathbf{v} est un élément du noyau de A associé au système S .

Une solution particulière est donnée par $x_0 = (4, 5, 0)$ et un élément du noyau de A est donné par $\mathbf{v} = (-7, -9, 1)$.

Vérifier ses connaissances

1. On considère le système suivant

$$\begin{cases} 2x - 3y &= 4 \\ 3x + 7y &= -1 \end{cases}$$

- (a) Le système est-il un système de Cramer ?
- (b) Résoudre le système S à l'aide des formules de Cramer puis en utilisant la méthode du Pivot

2. On considère le système suivant

$$\begin{cases} 2x - 3y + 6z - 3u &= 2 \\ z + u &= -5 \\ 4u &= -8 \\ 7y - 3z + 2u &= -1 \end{cases}$$

- (a) Le système est-il un système de Cramer ?
- (b) Résoudre le système S par la méthode de votre choix.

3. On considère le système suivant

$$\begin{cases} x - 2y + 2z &= 3 \\ 3x - 2z &= -7 \\ -x + y + z &= 6 \end{cases}$$


- (a) Le système est-il un système de Cramer ?
- (b) Résoudre le système S par la méthode de votre choix.


4. On considère le système suivant

$$\begin{cases} x - 2y + 2z + 2u &= 1 \\ 3x - 2z - 5u &= -4 \\ -x + y + z + 3u &= 2 \end{cases}$$

- (a) Le système est-il un système de Cramer ?
- (b) Résoudre le système S à l'aide de la méthode du Pivot de Gauss.
- (c) Que peut-on dire quant à la dimension de l'espace des solutions ?

Résolution d'un système linéaire sous

Les outils nécessaires sous  ont été introduit à la section précédente, il suffit simplement de les mettre en application.

Remarque :  ne nous permettra de résoudre des systèmes linéaires de n équations à n inconnues !

Regardons cela avec l'Exemple 4.3 qui se présente sous la forme $A\mathbf{x} = \mathbf{b}$ où A est une matrice carrée inversible.

Alors la solution est donnée par $\mathbf{x} = A^{-1}\mathbf{b}$, ce qui s'écrit :

```
A = rbind(c(2,-4,-2),c(5,-2,3),c(-1,3,7))
b = c(8,-1,-2)
# Résolution système linéaire
solve(A,b)

## [1] -2.175 -3.550  0.925

#ou encore
solve(A)%*%b

##           [,1]
## [1,] -2.175
## [2,] -3.550
## [3,]  0.925
```

En réalité, il existe des fonctions permettant de résoudre des systèmes linéaires de n équations avec p inconnues mais cela nécessite l'utilisation du package *matlib*³.

3. Une description avec des exemple est disponible à l'adresse suivante pour le lecteur curieux : <https://cran.r-project.org/web/packages/matlib/vignettes/linear-equations.html>

5 Réduction des endomorphismes

Dans cette section, nous allons concentrer sur l'étude des endomorphismes u d'espaces vectoriels E de dimension finie $n \in \mathbb{N}^*$. Nous avons vu que si E est muni d'une base $\mathcal{B} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$, typiquement la base canonique de \mathbb{R}^n , alors u peut-être représentée matriciellement en déterminant les images des vecteurs de bases par l'application u . La matrice obtenue, notée A , est alors de la forme :

$$A = \begin{pmatrix} u(\mathbf{e}_1) & u(\mathbf{e}_2) & \dots & u(\mathbf{e}_n) \\ a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{pmatrix}.$$

Nous avons également vu comment représenter ce même endomorphisme dans des bases différentes \mathcal{B} et \mathcal{B}' dont les représentations sont notées A et A' respectivement et on note P la matrice de passage de \mathcal{B} vers \mathcal{B}' . Alors

$$A' = P^{-1}AP.$$

Dans cette section, nous allons chercher une base \mathcal{B}' dans laquelle la matrice A' se présente sous la forme d'une matrice diagonale, comme suit :

$$A' = \begin{pmatrix} a'_{11} & 0 & \dots & 0 \\ 0 & a'_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a'_{nn} \end{pmatrix}.$$

Un endomorphisme pour lequel il existe une base dans laquelle sa représentation est une matrice diagonale est dit **diagonalisable**. Il est plus simple d'étudier les matrices diagonales et ces dernières viendront "naturellement" en analyses de données. Par exemple, on cherchera souvent à diagonaliser des matrices dans le cadre de l'Analyse en Composantes Principales afin de déterminer quelles sont les directions de l'espace dans lesquelles les données présentent une variance élevée.

En revanche, nous verrons que toutes les matrices ne sont pas nécessairement diagonalisables. On va donc chercher à caractériser les endomorphismes qui sont diagonalisables mais aussi voir comment diagonaliser de telles applications. Enfin, nous verrons comment obtenir les matrices de passage permettant d'avoir une représentation sous forme de matrice diagonale de notre endomorphisme.

Dans la suite on considérera uniquement le cas $\mathbb{K} = \mathbb{R}$. Le cas $\mathbb{K} = \mathbb{C}$ fera l'objet d'une remarque en fin de section.

5.1 Vecteurs propres et valeurs propres

Définition 5.1: Vecteur et valeur propres

Soit E un espace vectoriel sur \mathbb{K} de dimension finie n et soit u un endomorphisme de E . On dit qu'un vecteur $\mathbf{x} \in E$ est un **vecteur propre** de u s'il existe $\lambda \in \mathbb{K}$ tel que

$$u(\mathbf{x}) = \lambda \mathbf{x}.$$

Dans ce cas, le scalaire λ est appelé **valeur propre** de l'endomorphisme u .

Si \mathbf{x} est un vecteur propre, il existe une et une seule valeur propre λ associée telle que $u(\mathbf{x}) = \lambda \mathbf{x}$. Si λ est une valeur propre, l'ensemble des vecteurs propres associés est l'ensemble des vecteurs non nuls de l'espace $\text{Ker}(u - \lambda \text{Id})$. En effet :

$$\lambda \text{ valeur propre} \iff \exists \mathbf{x} \neq 0 \text{ t.q. } u(\mathbf{x}) = \lambda \mathbf{x} \iff \exists \mathbf{x} \neq 0 \text{ t.q. } (u - \lambda \text{Id})(\mathbf{x}) = 0 \iff \mathbf{x} \in \text{Ker}(u - \lambda \text{Id}).$$

Si les vecteurs propres sont non nuls par définition, rien n'empêche la valeur propre d'être nulle, dans ce cas les vecteurs propres associés à la valeur propre $\lambda = 0$ sont exactement les vecteurs non nuls de $\text{Ker}(u)$. On peut aussi dire que 0 est une valeur propre de u si et seulement si u n'est pas injectif.

Remarquons également que si \mathbf{x} est un vecteur propre de u associé à la valeur propre λ , alors $\alpha \mathbf{x}$, où $\alpha \in \mathbb{K}$ est aussi un vecteur propre u .

5.2 Diagonalisation

Définition 5.2: Endomorphisme diagonalisable

Un endomorphisme u d'un espace vectoriel E sur \mathbb{K} de dimension finie n est **diagonalisable** si et seulement s'il existe une base de E formée des vecteurs propres de u .

Cette définition est plutôt intuitive. En effet, nous avons que si \mathbf{x}_i est un vecteur propre associé à la valeur propre λ_i alors $u(\mathbf{x}_i) = \lambda_i \mathbf{x}_i$. Si on note A la représentation matricielle de u dans la base de vecteurs propres, nous avons alors :

$$A = \begin{pmatrix} u(\mathbf{x}_1) & u(\mathbf{x}_2) & \dots & u(\mathbf{x}_n) \\ \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}.$$

Réciproquement s'il existe une base $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ dans laquelle l'endomorphisme u est représenté par une matrice diagonale alors cette base est formée des vecteurs propres de u .

Toute la question est maintenant de savoir comment, étant donné un endomorphisme u , déterminer ses valeurs propres et ses vecteurs propres. En répondant à cette question, nous mettrons à nouveau une condition nécessaire et suffisante pour qu'un endomorphisme soit diagonalisable.

5.2.1 Recherche des valeurs propres

Rappelons que λ est une valeur propre d'un endomorphisme u si l'application $u - \lambda Id$ n'est pas injective. On se rappelle que si une application n'est pas injective, elle n'est donc pas bijective et ... son déterminant est donc nul! C'est ce que nous allons utiliser pour déterminer les valeurs propres de notre application. Ainsi λ est une valeur propre de u si et seulement si

$$\det(u - \lambda Id) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1n} \\ a_{n1} & \dots & a_{nn-1} & a_{nn} - \lambda \end{vmatrix} = 0.$$

Si on développe ce déterminant, on obtient un polynôme de degré n dont les racines sont les valeurs propres de u .

Proposition 5.1: Valeurs propres d'un endomorphisme

Soient E un espace vectoriel de dimension finie n et u un endomorphisme de E . Alors les valeurs propres de u sont exactement les racines du polynôme P_u défini par :

$$P_u(\lambda) = \det(u - \lambda Id).$$

Ce polynôme est appelé **polynôme caractéristique** de u . Il est de degré n en λ .

Remarque : on peut également définir le polynôme caractéristique d'un endomorphisme en utilisant sa représentation matricielle A . Ainsi

$$P_u(\lambda) = \det(u - \lambda Id) = \det(A - \lambda I_n) = P_A(\lambda).$$

Certains auteurs définissent le polynôme caractéristique par :

$$P_u(\lambda) = \det(\lambda Id - u) = \det(\lambda I_n - A) = P_A(\lambda).$$

Ce qui revient simplement, pour passer de l'une à l'autre des expressions, à multiplier par $(-1)^n$. La seconde définition est parfois plus simple à manipuler car le coefficient du terme de degré n devant le polynôme est égal à 1 au lieu de -1 (si n est impair). Le polynôme est alors dit *unitaire*.

Exemple 5.1. *Considérons l'endomorphisme u dont la représentation dans la base canonique est donnée par :*

$$A = \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix}.$$

Le polynôme caractéristique $P_A(\lambda)$ est donné par :

$$P_A(\lambda) = \det(A - \lambda Id) = \begin{vmatrix} 1 - \lambda & -1 \\ 3 & 5 - \lambda \end{vmatrix} = (1 - \lambda)(5 - \lambda) + 3 = (\lambda - 2)(\lambda - 4).$$

Ainsi les valeurs propres de A sont $\lambda_1 = 2$ et $\lambda_2 = 4$.

Exemple 5.2. *Considérons l'endomorphisme u dont la représentation dans la base canonique est donnée par :*

$$A = \begin{pmatrix} 2 & -1 \\ 0 & 2 \end{pmatrix}.$$

La matrice est triangulaire supérieure, on peut donc directement lire les valeurs propres sur la diagonale de matrice. En effet, on se souvient que le déterminant d'une matrice triangulaire est égale au produit des éléments diagonaux. Ainsi les valeurs propres de A sont $\lambda_1 = 2$ et $\lambda_2 = 2$. On dira aussi que l'endomorphisme n'admet qu'une seule valeur propre mais de multiplicité 2.

Définition 5.3: Spectre d'une matrice ou d'un endomorphisme

L'ensemble des valeurs propres d'un endomorphisme u est appelé **spectre de u** , noté $\text{Spec}(u)$ ou encore $\text{Spec}(A)$, si A est la représentation matricielle de u .

La proposition suivante est très pratique lorsque l'on manipule des matrices de taille raisonnable, surtout des matrices de $\mathcal{M}_2(\mathbb{K})$ afin de vérifier que l'on ne s'est pas trompé dans le calcul des valeurs propres :

Proposition 5.2: Somme et produit des valeurs propres

Soit $A \in \mathcal{M}_n(\mathbb{K})$ et notons $\lambda_1, \dots, \lambda_n$ ses valeurs propres, c'est-à-dire les racines du polynôme caractéristique associé à A . Alors :

- $\text{tr}(A) = \sum_{i=1}^n \lambda_i$ et
- $\det(A) = \prod_{i=1}^n \lambda_i$.

Démonstration. Notons $(\lambda_1, \dots, \lambda_n)$ les valeurs propres de la matrice A et rappelons que le polynôme caractéristique de la matrice A s'écrit :

$$P_A(\lambda) = \prod_{i=1}^n (\lambda - \lambda_i) = \lambda^n - \lambda^{n-1} \left(\sum_{i=1}^n \lambda_i \right) + \dots + (-1)^n \prod_{i=1}^n \lambda_i.$$

Il est important d'avoir cela en tête pour la suite, nous allons donner une expression différente de ce polynôme et faire de l'identification terme à terme.

On rappelle que l'on a également, par définition du polynôme caractéristique :

$$P_A(\lambda) = \det(\lambda I_n - A) = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma) \prod_{i=1}^n (\lambda \delta_{\sigma(i), i} - a_{\sigma(i) i}),$$

où $\delta_{i,j} = 1$ si et seulement si $i = j$, dans le cas contraire $\delta_{i,j} = 0$ (c'est ce que l'on appelle le *symbole de kronecker*).

Lorsque σ n'est pas l'identité, il existe au moins deux éléments distincts k et l de $\llbracket 1, n \rrbracket$ tels que $\sigma(l) = k$ et $\sigma(k) = l$, donc les termes $\delta_{\sigma(k), k}$ et $\delta_{\sigma(l), l}$ sont nuls. Ainsi les termes de degrés n et $n-1$ sont ceux liés à $\sigma = \text{Id}$ (i.e. on ne permute pas les indices) et ils sont donnés par les termes de degrés n et $n-1$ du produit :

$$\prod_{i=1}^n (\lambda - a_{ii}).$$

Les termes de degrés n et $n-1$ sont donc λ^n et $-(a_{11} + \dots + a_{nn})\lambda^{n-1}$. Par identification entre les deux polynômes, plus précisément en comparant les termes de degré

$n - 1$, on a bien $\text{tr}(A) = \sum_{i=1}^n \lambda_i$.

Enfin, on posant $\lambda = 0$ dans les deux expressions de P_A , on trouve bien que :

$$P_A(0) = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma) \prod_{i=1}^n (-a_{\sigma(i)i}) = (-1)^n \det(A) = (-1)^n \prod_{i=1}^n \lambda_i.$$

□

Exemple 5.3. *Considérons l'endomorphisme u dont la représentation dans la base canonique est donnée par :*

$$A = \begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix}.$$

On pourrait alors utiliser le résultat précédent pour calculer les deux valeurs propres λ_1 et λ_2 ! En effet, d'après la proposition précédente, le polynôme caractéristique de A s'écrit :

$$P_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2 = \lambda - \text{tr}(A)\lambda + \det(A).$$

On a donc le système suivant :

$$\begin{cases} \lambda_1\lambda_2 = \det(A) = 4 \\ \lambda_1 + \lambda_2 = \text{tr}(A) = 5 \end{cases},$$

dont l'unique solution est donnée par $\lambda_1 = 1$ et $\lambda_2 = 4$.

5.2.2 Recherche des vecteurs propres

Il s'agit maintenant de déterminer *les vecteurs propres associés à chaque valeur propre*. Cette étape s'effectue donc *après avoir déterminé les valeurs propres d'un endomorphisme*.

Pour cela nous devons déterminer les éléments \mathbf{x} tels que $\mathbf{x} \in \text{Ker}(u - \lambda \text{Id})$. Cela peut se faire en résolvant le système d'équations linéaires homogène :

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

pour toutes les valeurs de λ appartenant au spectre de A . La connaissance de ces vecteurs propres va permettre de définir une base dans laquelle la nouvelle matrice représentant u sera diagonale mais aussi la matrice de passage associée de la base courante vers la base de vecteurs propres.

Exemple 5.4. Considérons l'endomorphisme u dont la représentation dans la base canonique est donnée par :

$$A = \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix}.$$

Nous avons vu que cet endomorphisme possède deux valeurs propres qui sont $\lambda_1 = 2$ et $\lambda_2 = 4$. Déterminons maintenant les vecteurs propres associés à ces deux valeurs propres en cherchant éléments de $\text{Ker}(A - \lambda I)$.

- **Pour $\lambda = \lambda_1 = 2$:**

Nous devons résoudre le système linéaire $(A - 2I)\mathbf{x} = 0$

$$\begin{cases} -x_1 - x_2 &= 0 \\ 3x_1 + 3x_2 &= 0 \end{cases}$$

Les deux équations sont dépendantes, on peut donc se concentrer sur la première équation dont une solution est donnée par $(x_1, x_2) = (-1, 1)$.

- **Pour $\lambda = \lambda_2 = 4$:**

Nous devons résoudre le système linéaire $(A - 4I)\mathbf{x} = 0$

$$\begin{cases} -3x_1 - x_2 &= 0 \\ 3x_1 + x_2 &= 0 \end{cases}$$

A nouveau, les deux équations sont dépendantes, on peut donc se concentrer sur la première équation dont une solution est $(x_1, x_2) = (-1, 3)$.

On a donc deux vecteurs propres et on peut vérifier qu'ils forment une base de $E = \mathbb{R}^2$ en calculant le déterminant :

$$\begin{vmatrix} -1 & -1 \\ 1 & 3 \end{vmatrix} = -2.$$

Ainsi la matrice A est diagonalisable et elle est semblable à la matrice $A' = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$ et la matrice de passage P vers la vers base des vecteurs propres est

défini par $P = \begin{pmatrix} -1 & -1 \\ 1 & 3 \end{pmatrix}$.

On peut ensuite facilement vérifier que $A' = P^{-1}AP$

A retenir : la matrice de passage est composée des vecteurs propres en colonne !

Exemple 5.5. Considérons l'endomorphisme u dont la représentation dans la base canonique est donnée par :

$$A = \begin{pmatrix} 2 & -1 \\ 0 & 2 \end{pmatrix}.$$

Nous avons vu que cette matrice admet une seule valeur propre $\lambda = 2$ dont la multiplicité est égale à 2. Déterminons les vecteurs propres associés en cherchant les solutions du système $(A - 2I)\mathbf{x} = 0$:

$$\begin{cases} -x_2 &= 0 \\ 0 &= 0 \end{cases}$$

Ce système admet pour solution le vecteur de la forme $(x_1, x_2) = (1, 0)$ et ... c'est tout ! On ne trouve ici qu'un seul vecteur propre ... la matrice A n'est donc pas diagonalisable a priori car on ne trouve pas une base de vecteurs propres de \mathbb{R}^2 .

Ce dernier exemple montre qu'il semble exister un lien entre le fait d'être diagonalisable et le nombre de vecteurs propres associés à une valeur propre de multiplicité supérieure ou égale à 2.

5.3 Caractérisation des endomorphismes diagonalisables

Définition 5.4: Espace propre

On appelle **sous-espace vectoriel propre** associé à une valeur propre λ , le sous-espace vectoriel

$$E_\lambda(u) = \text{Ker}(u - \lambda Id).$$

Un tel sous espace vectoriel est au moins de dimension 1. Si λ n'est pas une valeur propre de u , alors cet espace est de dimension 0 (il est réduit au vecteur nul).

Cette définition va nous permettre de donner une première caractérisation des endomorphismes (ou matrices) diagonalisables.

Proposition 5.3: Propriétés des sous-espaces propres

Soient $\lambda_1, \dots, \lambda_n$ les valeurs propres, deux à deux distincts, d'un endomorphisme u . Alors la famille des sous-espaces propres $\{E_{\lambda_i}(u)\}_{i=1}^n$ est en somme directe.

Démonstration. On ne démontre pas ce résultat car il nécessiterait de faire une grande parenthèse sur les polynômes ainsi que d'introduire un résultat préliminaire que l'on appelle le *lemme des noyaux*. □

Ce premier résultat est important car si on dispose de n sous-espaces de dimension au moins égale à 1 qui sont en somme directe dans un espace vectoriel de dimension n . Cela signifie que les vecteurs propres forment une base de cet espace vectoriel, donc l'endomorphisme est diagonalisable.

C'est en fait une conséquence du résultat, plus général, suivant :

Théorème 5.1: Diagonalisation et sous-espaces propres

Soit u un endomorphisme d'un espace vectoriel E de dimension n et notons $\lambda_1, \dots, \lambda_n$ ses valeurs propres (pas nécessairement distinctes). Alors les propriétés suivantes sont équivalentes :

- i) u est diagonalisable
- ii) E est la somme directe de ses sous-espaces vectoriels propres.

Ce théorème a deux conséquences immédiates :

Corollaire 5.1: Diagonalisation et valeurs propres

Soit E un espace vectoriel de dimension n et u un endomorphisme de E qui admet $r \leq n$ valeurs propres distinctes $\lambda_1, \dots, \lambda_r$. Alors :

- l'endomorphisme u est diagonalisable si et seulement si la somme $\sum_{k=1}^r \dim(E_{\lambda_k}(u))$ est égale à n .
- l'endomorphisme u est diagonalisable si et seulement si, pour tout $k \in \llbracket 1, r \rrbracket$, nous avons

$$\dim(E_{\lambda_k}(u)) = m(\lambda_k),$$

où $m(\lambda_k)$ est la multiplicité de la valeur propre λ_k .

Ainsi, d'après ce corollaire, si un endomorphisme admet des valeurs propres deux à deux distinctes, alors ce dernier est automatiquement diagonalisable. Plus précisément :

Corollaire 5.2: Diagonalisation et polynôme caractéristique

Soit E un espace vectoriel de dimension finie n et u un endomorphisme de E . On notera $A \in \mathcal{M}_n(\mathbb{R})$ la représentation matricielle de cet endomorphisme. Si le polynôme caractéristique de A est scindé sur \mathbb{R} (*i.e.* il peut s'écrire comme le produit de monômes à coefficients dans \mathbb{R}) et possède toutes ses racines simples, alors A est diagonalisable dans $\mathcal{M}_n(\mathbb{R})$.

L'hypothèse *polynôme caractéristique scindé sur \mathbb{R}* est une condition importante pour la diagonalisation. En effet considérons la matrice A définie par

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

dont le polynôme caractéristique $P_A(\lambda)$ est donné par $P_A(\lambda) = \lambda^2 + 1$. Ce polynôme n'est pas scindé dans \mathbb{R} et n'est donc pas diagonalisable car ces deux valeurs propres sont complexes i et $-i$.

Reprenons aussi l'Exemple 5.5, nous avons remarqué que cette matrice n'était pas diagonalisable. En effet, l'espace vectoriel E est de dimension 2 et le sous-espace-propre engendré par la seule et uniquement valeur propre 2 n'était que de dimension 1.

Cet exemple permet aussi d'illustrer le résultat suivant.

Proposition 5.4: Dimension des sous-espaces propres

Soit u un endomorphisme d'un espace vectoriel E de dimension n , soit λ une valeur propre de E de multiplicité $m(\lambda)$ et $E_\lambda(u)$ le sous-espace propre associé. Alors :

$$\dim(E_\lambda(u)) \leq m(\lambda).$$

Finissons par un exemple où la matrice admet une valeur propre multiple mais qui est malgré tout diagonalisable.

Exemple 5.6. Considérons l'endomorphisme u dont la représentation dans la base canonique est donnée par :

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

et montrons que ce dernier est diagonalisable. Son polynôme caractéristique est donné par le déterminant $\det(A - \lambda I_3)$:

$$\begin{aligned} P_A(\lambda) &= \begin{vmatrix} -\lambda & 3 & 2 \\ -2 & 5-\lambda & 2 \\ 2 & -3 & -\lambda \end{vmatrix}, \\ &\downarrow \text{on va développer selon la première ligne} \\ &= -\lambda \begin{vmatrix} 5-\lambda & 2 \\ -3 & -\lambda \end{vmatrix} - 3 \begin{vmatrix} -2 & 2 \\ 2 & -\lambda \end{vmatrix} + 2 \begin{vmatrix} -2 & 5-\lambda \\ 2 & -3 \end{vmatrix}, \\ &\downarrow \text{on développe chaque déterminant d'ordre 2} \\ &= \underbrace{-\lambda(-\lambda(5-\lambda) + 6)}_{\text{bleu}} - \underbrace{3(2\lambda - 4)}_{\text{rouge}} + \underbrace{2(6 - 2(5-\lambda))}_{\text{vert}}, \end{aligned}$$

$$\begin{aligned}
&= -\lambda^3 + 5\lambda^2 - 6\lambda - 6\lambda + 12 + 12 - 20 + 4\lambda, \\
&\quad \downarrow \text{factorisation} \\
&= -\lambda^3 + 5\lambda^2 - 8\lambda + 4, \\
&\quad \downarrow \text{on observe que 2 est racine du polynôme} \\
&= -(\lambda - 2)(\lambda^2 - 3\lambda + 2), \\
&\quad \downarrow \text{on observe à nouveau que 2 et -1 sont racines} \\
&= -(\lambda - 2)^2(\lambda - 1).
\end{aligned}$$

Le polynôme caractéristique admet donc deux valeurs propres distinctes $\lambda_1 = 1$ et $\lambda_2 = 2$ où $m(\lambda_1) = 1$ et $m(\lambda_2) = 2$.

Pour vérifier si notre endomorphisme est diagonalisable, il faut montrer que sous-espace propre associé à la valeur propre λ_2 est bien de dimension 2. Cependant, on va également chercher à déterminer le sous-espace propre E_{λ_1} afin de définir notre matrice de passage.

On rappelle que les sous-espaces propres sont définis par :

$$E_\lambda = \text{Ker}(u - \lambda \text{Id}).$$

• **Espace propre E_{λ_1} :**

On cherche à résoudre le système suivant en utilisant la méthode du pivot :

$$\begin{cases} -x_1 + 3x_2 + 2x_3 &= 0 \\ -2x_1 + 4x_2 + 2x_3 &= 0 : L_2 \leftarrow L_2 - 2L_1 \\ 2x_1 - 3x_2 - x_3 &= 0 : L_3 \leftarrow L_3 + 2L_1 \end{cases}$$

$$\begin{cases} -x_1 + 3x_2 + 2x_3 &= 0 \\ -2x_2 - 2x_3 &= 0 \\ 3x_2 + 3x_3 &= 0 \end{cases}$$

Le système peut donc être réduit au système à deux équations suivants :

$$\begin{cases} -x_1 + 3x_2 + 2x_3 &= 0 \\ x_2 &= -x_3 \end{cases}$$

soit

$$\begin{cases} x_1 &= -x_3 \\ x_2 &= -x_3 \end{cases}$$

Le sous espace propre E_{λ_1} est alors généré par le vecteur propre $\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$

• **Espace propre E_{λ_2} :**

On cherche à résoudre le système suivant en utilisant la méthode du pivot :

$$\begin{cases} -2x_1 + 3x_2 + 2x_3 &= 0 \\ -2x_1 + 3x_2 + 2x_3 &= 0 \\ 2x_1 - 3x_2 - 2x_3 &= 0 \end{cases}$$

On remarque que le système peut-être réduit à l'équation suivante :

$$-2x_1 + 3x_2 + 2x_3 = 0$$

dont les solutions sont paramétrés par des paramètres $x_1 = t_1$ et $x_2 = t_2$. Une première solution est alors donnée par le vecteur $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ en posant $t_2 = 0$ et $t_1 = 1$

et une deuxième solution, indépendante est donnée par le vecteur $\begin{pmatrix} 0 \\ -1/3 \\ 1/2 \end{pmatrix}$ en posant $t_1 = 0$ et $t_2 = -1/3$.

Ainsi le sous-espace propre E_{λ_2} est engendré par les vecteurs $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ et $\begin{pmatrix} 0 \\ -1/3 \\ 1/2 \end{pmatrix}$.

Pour chaque valeur propre, la dimension du sous-espace propre est égale à la multiplicité de la valeur propre associée, notre endomorphisme est donc diagonalisable. Dans la base de vecteurs propres, la matrice diagonale est définie par $\text{Diag}(1, 2, 2)$ et la matrice de passage est donnée par :

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1/3 \\ -1 & 1 & 1/2 \end{pmatrix}.$$

On pourra vérifier que la matrice $P^{-1}AP$ est bien diagonale.

Remarque : dans cette section, nous avons supposé que le corps sur lequel nous avons effectué notre étude est le corps des nombres réels \mathbb{R} .

Sur le corps des nombres complexes, la théorie reste inchangée mais on peut cependant remarquer que si des matrices ne sont pas diagonalisables dans \mathbb{R} elle peuvent l'être dans \mathbb{C} ! Encore mieux, il se peut qu'une matrice n'admette aucune valeur propre réelle mais qu'elle soit diagonalisable (dans \mathbb{C} mais pas dans \mathbb{R} bien évidemment).

Un exemple classique qui permet d'illustrer cela est la matrice de rotation $R(\theta)$ où $\theta \in \mathbb{R}$ est définie par :

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Le polynôme caractéristique de cette matrice est défini par

$$\begin{aligned} P_{R(\theta)}(\lambda) &= (\cos(\theta) - \lambda)^2 + \sin(\theta)^2, \\ &= \lambda^2 - 2\cos(\theta)\lambda + 1. \end{aligned}$$

Si on cherche les racines de ce polynôme, en utilisant le fait que $\cos(\theta) = \frac{1}{2}(e^{i\theta} + e^{-i\theta})$, on trouve $\lambda_{\pm} = \{e^{-i\theta}, e^{i\theta}\}$ où i est tel que $i^2 = -1$. Donc $P_{R(\theta)}(\lambda) = (\lambda - e^{-i\theta})(\lambda - e^{i\theta})$. Ce polynôme est donc scindé sur \mathbb{C} mais ne l'est pas sur \mathbb{R} sauf dans le cas $\theta = 0$. Comme le polynôme caractéristique est scindé à racines simples dans \mathbb{C} , $R(\theta)$ est diagonalisable dans \mathbb{C} . De plus $R(\theta)$ est diagonalisable dans \mathbb{R} si et seulement si $\theta \equiv 0 \pmod{\pi}$.

Vérifier ses connaissances

1. Les propositions suivantes sont-elles vraies ou fausses ?
 - (a) Tout endomorphisme est diagonalisable.
 - (b) Tout automorphisme d'espace vectoriel est diagonalisable.
 - (c) Un endomorphisme d'espace vectoriel injectif admet nécessairement 0 comme valeur propre.
 - (d) Un automorphisme d'espace vectoriel n'admet que des valeurs propres non nulles.
 - (e) Les racines du polynôme caractéristique sont exactement les valeurs propres de l'endomorphisme étudié.
 - (f) Toutes les matrices triangulaires (inférieures ou supérieures) ne sont pas diagonalisables.

2. Déterminer le polynôme caractéristique des matrices suivantes

$$A = \begin{pmatrix} 1 & 2 & -2 \\ 2 & 1 & -2 \\ 2 & 2 & -3 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 0 & -\alpha_0 \\ 1 & 0 & -\alpha_1 \\ 0 & 1 & -\alpha_2 \end{pmatrix} \quad \text{et} \quad C = \begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & 5 \end{pmatrix}.$$

3. Les matrices suivantes sont-elles diagonalisables. Si oui, déterminez les vecteurs propres.

$$A = \begin{pmatrix} 2 & 5 & -6 \\ 4 & 6 & -9 \\ 3 & 6 & -8 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 3 & 3 \\ -1 & 8 & 6 \\ 2 & -14 & -10 \end{pmatrix} \quad \text{et} \quad C = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$


Valeurs propres et vecteurs propres sous

Nous allons simplement regarder comment calculer les valeurs propres et les vecteurs propres d'un endomorphisme. Pour cela reprenons la matrice A de l'Exemple 5.6 et vérifions que ce que nous avons fait est correct.

Pour déterminer les valeurs propres et les vecteurs propres d'un endomorphisme, on utilisera la fonction *eigen*.

```
A = rbind(c(0,3,2),c(-2,5,2),c(2,-3,0))
# Réduction d'un endomorphisme
eigen(A)

## eigen() decomposition
## $values
## [1] 2 2 1
##
## $vectors
##           [,1]      [,2]      [,3]
## [1,] -0.1908801  0.3333333 -0.5773503
## [2,] -0.6294228  0.6666667 -0.5773503
## [3,]  0.7532541 -0.6666667  0.5773503
```

On remarque que  nous renvoie un objet avec deux *listes*. Une première liste *\$values* avec les différentes valeurs propres et une deuxième liste *\$vectors* avec les vecteurs propres. associés.

```
# Valeurs propres
eigen(A)$values

## [1] 2 2 1

# Vecteurs propres
eigen(A)$vectors

##           [,1]      [,2]      [,3]
## [1,] -0.1908801  0.3333333 -0.5773503
## [2,] -0.6294228  0.6666667 -0.5773503
## [3,]  0.7532541 -0.6666667  0.5773503
```

La valeur propre 2 apparaît deux fois, c'est donc une racine double du polynôme caractéristique associé. Les vecteurs propres associés à cette valeur propre sont représentés par les deux premières **colonnes** de notre matrice $eigen(A)$vectors$. Enfin la valeur

propre 1 est associée au vecteur propre représenté par la dernière colonne de matrice.

On peut alors vérifier la définition de vecteurs/valeurs propres

```
# Vecteurs propres
v1 = eigen(A)$vectors[,1]
v2 = eigen(A)$vectors[,2]
v3 = eigen(A)$vectors[,3]
# Valeurs propres
lambda1 = eigen(A)$values[1]
lambda2 = eigen(A)$values[2]
lambda3 = eigen(A)$values[3]
# Matrice identité
I3 = diag(c(1,1,1))
# Calcul de (A-lambda I3)v
(A-lambda1*I3)%*%v1


##           [,1]
## [1,] -2.220446e-16
## [2,] -4.440892e-16
## [3,]  6.661338e-16

(A-lambda2*I3)%*%v2


##           [,1]
## [1,]  0.000000e+00
## [2,]  2.220446e-16
## [3,] -4.440892e-16

(A-lambda3*I3)%*%v3

##           [,1]
## [1,]  0.000000e+00
## [2,] -2.220446e-16
## [3,]  3.330669e-16
```

Les valeurs sont nulles bien que  indique le contraire. Mais cette imprécision est purement numérique.

On remarque que les valeurs des vecteurs propres sont différentes de celles obtenues dans l'exemple. Mais pas d'inquiétude, ils définissent la même droite. La seule différence

réside dans la norme de ces vecteurs.  cherche des vecteurs propres de norme égale à 1, ce qui n'est pas le cas lorsque nous avons fait les calculs à la main.

```
# Vecteurs propres dans eigen(A)$vectors
v1 = eigen(A)$vectors[,1]
v2 = eigen(A)$vectors[,2]
v3 = eigen(A)$vectors[,3]
# Calcul de la norme
sqrt(sum(v1^2))

## [1] 1

sqrt(sum(v2^2))

## [1] 1

sqrt(sum(v3^2))

## [1] 1
```

Nous n'avons pas encore parlé de norme mais on anticipe un peu ici, on rappelle que la norme d'un vecteur \mathbf{x} est donnée par

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{j=1}^d x_j^2}.$$

6 Formes quadratiques et espaces euclidiens

6.1 Formes bilinéaires

Définition 6.1: Forme bilinéaire

Soit E un espace vectoriel sur \mathbb{K} et ϕ une application de $E \times E$ dans \mathbb{K} . L'application ϕ est dite **bilinéaire** si elle est linéaire en ses deux arguments, *i.e.* si pour tout $\mathbf{x}, \mathbf{x}', \mathbf{x}'' \in E$ et $\alpha \in \mathbb{K}$ elle est vérifiée

- $\phi(\mathbf{x} + \mathbf{x}'', \mathbf{x}') = \phi(\mathbf{x}, \mathbf{x}') + \phi(\mathbf{x}'', \mathbf{x}')$ (linéarité à gauche)
- $\phi(\mathbf{x}, \mathbf{x}' + \mathbf{x}'') = \phi(\mathbf{x}, \mathbf{x}') + \phi(\mathbf{x}, \mathbf{x}'')$ (linéarité à droite)
- $\phi(\lambda \mathbf{x}, \mathbf{x}') = \phi(\mathbf{x}, \lambda \mathbf{x}') = \lambda \phi(\mathbf{x}, \mathbf{x}')$

Regardons un peu comment se caractérise une forme bilinéaire si on se place dans un espace E de dimension finie n muni d'une base $\mathcal{B} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ (par exemple, la base canonique de \mathbb{R}^n). Pour cela considérons $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$ et $\mathbf{y} = \sum_{i=1}^n y_i \mathbf{e}_i$. Alors

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

Ainsi, la forme bilinéaire est entièrement définie par la connaissance des images des vecteurs de base par l'application ϕ . Cette première écriture permet également de montrer que toutes les formes bilinéaires peuvent s'écrire de la forme suivante :

$$\phi(\mathbf{x}, \mathbf{y}) = x_1 y_1 \phi_{11} + x_2 y_1 \phi_{21} + x_1 y_2 \phi_{12} + \dots + x_i y_j \phi_{ij} + \dots + x_n y_n \phi_{nn},$$

où $\phi_{ij} = \phi(\mathbf{e}_i, \mathbf{e}_j)$. Une forme bilinéaire se présente donc comme la somme de monômes en $x_i y_j$ où chacun des termes se retrouvent à la puissance 1. Si l'on prend le cas particulier de la base canonique de \mathbb{R}^n , nous aurions

$$\phi(\mathbf{x}, \mathbf{y}) = x_1 y_1 \phi_{11} + x_2 y_2 \phi_{22} + \dots + x_n y_n \phi_{nn},$$

Exemple 6.1. Les applications suivantes sont des formes bilinéaires

- $3x_1 y_1 + 2x_2 y_1 - 3.14x_1 y_3$
- $(x_1 - x_2)^2 - (x_1 + x_2)^2$

Les applications suivantes ne sont pas des formes bilinéaires

- $3x_1^2 y_2 - x_2 y_1 + 5x_3 y_3$ (présence d'un terme de degré 3)
- $3x_1 - x_2 y_1 + 5x_3 y_3$ (présence d'un terme de degré un uniquement)

Dans la suite, nous allons nous focaliser sur un type particulier de formes bilinéaires que nous serons très souvent amenés à étudier par la suite : *les formes bilinéaires*

symétriques.

Définition 6.2: Forme bilinéaire symétrique

Soit ϕ une forme bilinéaire sur un espace vectoriel E . Cette application ϕ est dite bilinéaire **symétrique** si et seulement si pour tout $\mathbf{x}, \mathbf{y} \in E$

$$\phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{y}, \mathbf{x}).$$

Une forme bilinéaire symétrique qui nous intéressera plus particulièrement un petit peu plus tard sera *le produit scalaire*. On verra même que cette forme bilinéaire dispose de propriétés supplémentaires.

Pour le moment, remarquons qu'une forme bilinéaire est dite symétrique si elle est invariante sous l'action de transposition des vecteurs \mathbf{x} et \mathbf{y} .

Exemple 6.2. L'application bilinéaire définie par :

$$\phi(\mathbf{x}, \mathbf{y}) = x_1y_1 + x_2y_1 + x_1y_2 - 3x_2y_3 + 3x_3y_2$$

n'est pas symétrique car la transposition $\mathbf{x} \leftrightarrow \mathbf{y}$ ne laisse pas le résultat invariant comme le montre les deux derniers termes. En revanche l'application

$$\phi(\mathbf{x}, \mathbf{y}) = x_1y_1 + x_2y_1 + x_1y_2 - 3x_2y_3 - 3x_3y_2$$

est bien symétrique.

Représentation matricielle. Comme pour les applications linéaires, il est possible de représenter les applications bilinéaires dès lors que l'on connaît l'image des vecteurs de base par notre application. La différence réside dans le fait que nous devons déterminer l'image par de tous les couples $(\mathbf{e}_i, \mathbf{e}_j)$ par notre application ϕ . En effet, on se rappelle que pour tout $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$ et $\mathbf{y} = \sum_{j=1}^n y_j \mathbf{e}_j$, on a

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

Définition 6.3: Matrice d'une forme bilinéaire

Soit ϕ une forme bilinéaire d'un espace vectoriel E de dimension finie n et muni d'une base $\mathcal{B} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$. La représentation matricielle de la forme bilinéaire s est alors donnée par

$$A_\phi = \begin{pmatrix} \phi(\mathbf{e}_1, \mathbf{e}_1) & \phi(\mathbf{e}_1, \mathbf{e}_2) & \dots & \phi(\mathbf{e}_1, \mathbf{e}_n) \\ \phi(\mathbf{e}_2, \mathbf{e}_1) & \phi(\mathbf{e}_2, \mathbf{e}_2) & \dots & \phi(\mathbf{e}_2, \mathbf{e}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(\mathbf{e}_n, \mathbf{e}_1) & \phi(\mathbf{e}_n, \mathbf{e}_2) & \dots & \phi(\mathbf{e}_n, \mathbf{e}_n) \end{pmatrix}$$

On prendra cependant à ne pas confondre cela avec la représentation matricielle d'une application linéaire ! Bien que les objets se ressemblent, ils n'ont pas la même signification !

Une fois que cette représentation est adoptée, il est possible de réécrire de façon totalement matricielle notre produit scalaire.

En effet, si on considère un espace vectoriel muni d'une base et que l'on note A la représentation matricielle d'une forme bilinéaire ϕ , alors pour tout vecteur \mathbf{x} et \mathbf{y} de E nous avons :

$$\phi(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A \mathbf{y}.$$

Exemple 6.3. Considérons l'application bilinéaire $\phi : E \times E \rightarrow \mathbb{R}$, où E est un espace vectoriel sur \mathbb{R} muni de la base canonique de \mathbb{R}^3 , définie par

$$\phi(\mathbf{x}, \mathbf{y}) = 3x_1y_1 + 6x_2y_1 - 3x_1y_2 - 6x_3y_1 + 2x_1y_3 + 2x_3y_3.$$

La matrice associée à cette forme bilinéaire est alors donnée par

$$A = \begin{pmatrix} 3 & -3 & 2 \\ 6 & 0 & 0 \\ -6 & 0 & 2 \end{pmatrix}$$

La connaissance de cette représentation nous permet par exemple de pouvoir déterminer simplement l'image des vecteurs $\mathbf{x} = (-1, 0, 1)$ et $\mathbf{y} = (0, 1, 1)$

$$\phi(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A \mathbf{y} = (-1, 0, 1) \begin{pmatrix} 3 & -3 & 2 \\ 6 & 0 & 0 \\ -6 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 3.$$

Exemple 6.4. Considérons l'application bilinéaire $\phi : E \times E \rightarrow \mathbb{R}$, où E est un espace vectoriel sur \mathbb{R} muni de la base canonique de \mathbb{R}^3 , définie par

$$\phi(\mathbf{x}, \mathbf{y}) = 3x_1y_1 + 2x_2y_1 + 2x_1y_2 - 3x_3y_1 - 3x_1y_3 + 2x_3y_3 + 5y_2x_2.$$

La matrice associée à cette forme bilinéaire est alors donnée par

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

On reconnaît ici la matrice d'une forme bilinéaire symétrique. La connaissance de cette représentation nous permet par exemple de pouvoir déterminer simplement l'image des vecteurs $\mathbf{x} = (-1, 0, 1)$ et $\mathbf{y} = (0, 1, 1)$

$$\phi(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A \mathbf{y} = (-1, 0, 1) \begin{pmatrix} 3 & 2 & -3 \\ 2 & 5 & 0 \\ -3 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 3.$$

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

La prochaine section va se focaliser sur le cas particulier des formes *bilinéaires symétriques*. Nous verrons que ces matrices là possèdent des propriétés intéressantes en terme de valeurs propres! Et, plus tard, que leur matrice de passage associée possède également une structure particulière.

6.2 Formes quadratiques

Généralités. Dans ce qui précède, nous avons supposé que les vecteurs \mathbf{x} et \mathbf{y} sont des vecteurs quelconques de \mathbb{R}^n , mais si l'on s'intéresse au cas particulier où $\mathbf{x} = \mathbf{y}$ on associe alors à notre forme bilinéaire ϕ ce que l'on appelle une forme quadratique.

Définition 6.4: Forme quadratique

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

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

est appelée **forme quadratique** sur E associée à la forme bilinéaire ϕ .

En tant que forme **quadratique**, l'application q précédemment définie n'est pas linéaire ! Ainsi, pour tout $\lambda \in \mathbb{K}$ l'égalité $q(\lambda \mathbf{x}) = \lambda q(\mathbf{x})$ est généralement fausse ! En effet, nous avons

$$q(\lambda \mathbf{x}) = \lambda^2 q(\mathbf{x}).$$

Si on note $A = (a_{ij})_{i,j=1}^n$ la représentation matricielle de l'application bilinéaire symétrique, alors $A \in S_n(\mathbb{K})$, i.e. l'ensemble des matrices symétriques à coefficients dans \mathbb{K} , alors la forme quadratique est définie par :

$$q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \sum_{i=1}^n a_{ii} x_i^2 + 2 \sum_{i>j}^n a_{ij} x_i x_j.$$

Cette expression est très pratique lorsque l'on souhaite reconnaître des formes quadratiques. En effet, elle se compose de monômes de degré 2 en les composantes x_i de \mathbf{x} .

Exemple 6.5. Considérons l'application bilinéaire $\phi : E \times E \rightarrow \mathbb{R}$, où E est un espace vectoriel sur \mathbb{R} muni de la base canonique de \mathbb{R}^3 , définie par

$$\phi(\mathbf{x}, \mathbf{y}) = 3x_1y_1 + 2x_2y_1 + 2x_1y_2 - 3x_3y_1 - 3x_1y_3 + 2x_3y_3 + 5y_2x_2.$$

La matrice associée à cette forme bilinéaire symétrique est alors donnée par

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

La forme quadratique est donnée par $q(\mathbf{x}) = \phi(\mathbf{x}, \mathbf{x}) = 3x_1^2 + 5x_2^2 + 2x_3^2 + 4x_1x_2 - 6x_1x_3$.

Si a toute forme bilinéaire symétrique il est possible d'associer une forme quadratique, la réciproque est également vraie.

Proposition 6.1: Identités de polarisation

Soit $q : E \rightarrow \mathbb{K}$ une forme quadratique sur E . Il existe alors une et une seule forme bilinéaire symétrique $\phi : E \times E \rightarrow \mathbb{K}$ telle que pour tout $\mathbf{x} \in E$: $q(\mathbf{x}) = \phi(\mathbf{x}, \mathbf{x})$. Elle est donnée par l'une des relations suivantes :

$$\begin{aligned}\phi(\mathbf{x}, \mathbf{y}) &= \frac{1}{2} (q(\mathbf{x} + \mathbf{y}) - q(\mathbf{x}) - q(\mathbf{y})) \\ \phi(\mathbf{x}, \mathbf{y}) &= \frac{1}{2} (q(\mathbf{x}) + q(\mathbf{y}) - q(\mathbf{x} - \mathbf{y})) \\ \phi(\mathbf{x}, \mathbf{y}) &= \frac{1}{4} (q(\mathbf{x} + \mathbf{y}) - q(\mathbf{x} - \mathbf{y})) .\end{aligned}$$

L'application ϕ est appelée **forme polaire** de q et les relations précédentes sont appelées **identités de polarisation**.

Démonstration. Nous démontrons uniquement la première relation, le fonctionnement est le même pour les autres identités. On se propose de faire cela de deux façons différentes.

- **A l'aide de la représentation matricielle :**

Pour cela on va considérer la matrice $A \in S_n(\mathbb{K})$ associée à q et deux vecteurs \mathbf{x} et \mathbf{y} de E . Nous avons :

$$\begin{aligned}q(\mathbf{x} + \mathbf{y}) - q(\mathbf{x}) - q(\mathbf{y}) &= \sum_{i=1}^n a_{ii}(x_i + y_i)^2 + 2 \sum_{i>j}^n a_{ij}(x_i + y_i)(x_j + y_j) \\ &\quad - \sum_{i=1}^n a_{ii}x_i^2 - 2 \sum_{i>j}^n a_{ij}x_ix_j - \sum_{i=1}^n a_{ii}y_i^2 - 2 \sum_{i>j}^n a_{ij}y_iy_j, \\ &\quad \downarrow \text{ en développant le premier terme et en simplifiant les termes en bleu} \\ &= 2 \sum_{i=1}^n a_{ii}x_iy_i + 2 \sum_{i>j}^n a_{ij}(x_i + y_i)(x_j + y_j) - 2 \sum_{i>j}^n a_{ij}x_ix_j - 2 \sum_{i>j}^n a_{ij}y_iy_j, \\ &\quad \downarrow \text{ en développant le premier terme en rouge puis en simplifiant} \\ &= 2 \sum_{i=1}^n a_{ii}x_iy_i + 4 \sum_{i>j}^n a_{ij}x_iy_j, \\ &= 2 \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij}x_iy_j \right), \\ &= 2\phi(\mathbf{x}, \mathbf{y}).\end{aligned}$$

- **A l'aide de la définition :**

On se rappelle que ϕ est une application bilinéaire symétrique donc

$$\begin{aligned} q(\mathbf{x} + \mathbf{y}) &= \phi(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) \\ &= \phi(\mathbf{x}, \mathbf{x}) + 2\phi(\mathbf{x}, \mathbf{y}) + \phi(\mathbf{y}, \mathbf{y}) \\ &= q(\mathbf{x}) + 2\phi(\mathbf{x}, \mathbf{y}) + q(\mathbf{y}) \end{aligned}$$

On obtient la deuxième relation en changeant \mathbf{y} en $-\mathbf{y}$ dans la démonstration précédente.

Enfin, la troisième relation s'obtient en calculant

$$q(\mathbf{x} + \mathbf{y}) - q(\mathbf{x} - \mathbf{y}) = \phi(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) - \phi(\mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y}).$$

□

Propriétés des formes quadratiques Les formes bilinéaires peuvent être représentées par des matrices, il en est donc de même pour les formes quadratiques qui admettront exactement la même représentation que la forme bilinéaire associée, comme nous avons pu le voir dans les exemples précédentes. On continuera d'appeler cette matrice A et cette dernière sera toujours symétrique !

Comme il s'agit d'une matrice, il est possible de diagonaliser cette matrice et de déterminer ses valeurs propres. Cela se révèle d'un grand intérêt dans le domaine de l'optimisation afin de dériver des propriétés d'un algorithme lorsque ces formes quadratiques sont étudiées en tant que fonction. On va donc rapidement lister quelques résultats intéressants sur ces dernières.

Définition 6.5: Caractérisation des formes quadratiques

Soit E un espace vectoriel sur \mathbb{K} et q une forme quadratique sur E . Notons A sa représentation matricielle. Alors q est dite :

- *positive* (resp. *définie positive*) si pour tout $\mathbf{x} \in E$ non nul :

$$q(\mathbf{x}) \geq 0 \quad \text{ou} \quad \mathbf{x}^T A \mathbf{x} \geq 0 \quad (\text{resp.} \quad q(\mathbf{x}) > 0 \quad \text{ou} \quad \mathbf{x}^T A \mathbf{x} > 0).$$

On dit parfois que la matrice associée à q est *SDP* : Semi-Définie Positive (resp. *DP* : Définie Positive).

- *négative* (resp. *définie négative*) si pour tout $\mathbf{x} \in E$ non nul :

$$q(\mathbf{x}) \leq 0 \quad \text{ou} \quad \mathbf{x}^T A \mathbf{x} \leq 0 \quad (\text{resp.} \quad q(\mathbf{x}) < 0 \quad \text{ou} \quad \mathbf{x}^T A \mathbf{x} < 0).$$

On dit parfois que la matrice associée à q est *SDN* : Semi-Définie Négative (resp. *DN* : Définie Négative).

Regardons un cas particulier que sont les formes quadratiques en dimension 2 pour illustrer ce résultat. Ces dernières se présentent sous la forme

$$q(\mathbf{x}) = \alpha x_1^2 + 2\beta x_1 x_2 + \gamma x_2^2.$$

Regardons à quelles conditions sur les paramètres α, β et γ , cette forme quadratique est positive ou négative. L'étude consiste à étudier un trinôme du second degré que l'on peut réécrire :

$$q(\mathbf{x}) = 2\beta x_1 x_2 + \gamma x_2^2 \quad \text{si } \alpha = 0, \quad (1)$$

$$q(\mathbf{x}) = \alpha \left(x_1 + \frac{\beta}{\alpha} x_2 \right)^2 + \frac{\alpha\gamma - \beta^2}{\alpha} x_2^2 \quad \text{si } \alpha \neq 0. \quad (2)$$

Considérons maintenant trois cas

- **Si $\alpha\gamma - \beta^2$ est strictement négatif :**

Si α est nul, alors β est non nul. Dans ce cas, la forme quadratique peut prendre n'importe quelle valeur réelle d'après l'expression 1, ce qui peut se voir en considérant la valeur $q(t, 1)$ où $t \in \mathbb{R}$.

Si α est strictement positif alors $q(1, 0) = \alpha$ et $q(\beta, -\alpha) = \alpha(\alpha\gamma - \beta^2)$ est alors strictement négatif.

De même lorsque α est négatif en considérant $-q$.

- **Si $\alpha\gamma - \beta^2$ est strictement positif :**

Dans ce cas le produit $\alpha\gamma > 0$ et α et γ sont nuls et de même signe. Si α est strictement positif, l'expression 2 montre que la forme quadratique est toujours positive. Elle s'annule uniquement si $x_1 + \frac{\beta}{\alpha} x_2$ et x_2 sont nuls simultanément, donc si $x_1 = x_2 = 0$.

Dans le cas où α est strictement négatif, on montre, de façon analogue, que la forme quadratique est strictement négative sauf si $x_1 = x_2 = 0$.

- **Si $\alpha\gamma - \beta^2$ est nul :**

Si $\alpha = 0$ alors $\beta = 0$, dans ce cas $q(\mathbf{x}) = \gamma x_2^2$ est du signe de γ et s'annule en $x_2 = 0$.

Si $\alpha > 0$, on a alors $q(\mathbf{x}) = \alpha \left(x_1 + \frac{\beta}{\alpha} x_2 \right)^2$ qui est toujours strictement positif sauf en $(-\beta, \alpha)$.

A nouveau le cas $\alpha < 0$ s'en déduit en considérant $-q$.

La forme quadratique est donc définie positive si et seulement si $\alpha\gamma - \beta^2 > 0$ et α (ou γ) est strictement positif.

Bien évidemment, il ne s'agit pas, pour montrer qu'une forme quadratique est semi-définie positive, de considérer tous les vecteurs \mathbf{x} possibles et de vérifier que la forme quadratique renvoie une valeur positive. Nous verrons plus tard comment faire cela en regardant uniquement les valeurs propres de l'endomorphisme.

6.3 Espaces euclidiens

A partir de maintenant on se place uniquement dans des espaces vectoriels réels. Ainsi, E désignera toujours un espace vectoriel sur \mathbb{R} .

Définition 6.6: Produit scalaire

Soit E un espace vectoriel sur \mathbb{K} . On appelle **produit scalaire** sur E , tout forme bilinéaire ϕ *symétrique et définie positive* sur E :

- **Bilinéaire** : voir Définition 6.1
- **Symétrique** : $\phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{y}, \mathbf{x})$
- **Définie positive** : $\phi(\mathbf{x}, \mathbf{x}) > 0$ et $\phi(\mathbf{x}, \mathbf{x}) = 0 \iff \mathbf{x} = 0$.

Le produit scalaire le plus classique que l'on connaît est celui défini relativement à la canonique de \mathbb{R}^n , appelé aussi **produit scalaire euclidien**. Ce produit scalaire est en général noté $\langle \mathbf{x}, \mathbf{y} \rangle$. Cependant la littérature est très riche en notation pour le produit scalaire, on peut également rencontrer $\langle \mathbf{x} | \mathbf{y} \rangle$ en physique notamment.

Exemple 6.6. Soit E un espace vectoriel et montrons que l'application $\phi : E \times E \rightarrow \mathbb{R}$ définie par

$$\phi(\mathbf{x}, \mathbf{y}) = x_1y_1 + 3x_1y_2 + 3x_2y_1 + 10x_2y_2,$$

définie un produit scalaire. On va donc vérifier les trois points de la définition.

- L'application est bien symétrique. En effet :

$$\phi(\mathbf{x}, \mathbf{y}) = x_1y_1 + 3x_1y_2 + 3x_2y_1 + 10x_2y_2 = \phi(\mathbf{y}, \mathbf{x}).$$

- Elle est bien définie positive. En effet

$$\phi(\mathbf{x}, \mathbf{x}) = x_1^2 + 6x_1x_2 + 10x_2^2 = (x_1 + 3x_2)^2 + x_2^2.$$

- **Bilinéarité** : Soient \mathbf{x}, \mathbf{y} et \mathbf{z} des éléments de E et $\lambda \in \mathbb{K}$ alors :

$$\begin{aligned} \phi(\mathbf{x} + \lambda\mathbf{z}, \mathbf{y}) &= (x_1 + \lambda z_1)y_1 + 3(x_1 + \lambda z_1)y_2 + 3(x_2 + \lambda z_2)y_1 + 10(x_2 + \lambda z_2)y_2 \\ &\quad \downarrow \text{on développe} \\ &= x_1y_1 + 3x_1y_2 + 3x_2y_1 + 10x_2y_2 \end{aligned}$$

$$\begin{aligned}
& + \lambda(z_1 y_1 + 3z_1 y_2 + 3z_2 y_1 + 10z_2 y_2), \\
& \downarrow \text{définition de } \phi \\
& = \phi(\mathbf{x}, \mathbf{y}) + \lambda \phi(\mathbf{z}, \mathbf{y}).
\end{aligned}$$

Proposition 6.2: Inégalité de Cauchy-Schwartz

Soit E un espace vectoriel. Alors pour tout vecteur \mathbf{x} et \mathbf{y} de E nous avons

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

De plus, l'égalité est atteinte lorsque les vecteurs \mathbf{x} et \mathbf{y} sont colinéaires.

Démonstration. Soit $(\mathbf{x}, \mathbf{y}) \in E^2$ et soit $\lambda \in \mathbb{R}$. Comme le produit scalaire est une forme bilinéaire symétrique définie positive, nous avons

$$\langle \mathbf{x} + \lambda \mathbf{y}, \mathbf{x} + \lambda \mathbf{y} \rangle \geq 0.$$

Or

$$\begin{aligned}
\langle \mathbf{x} + \lambda \mathbf{y}, \mathbf{x} + \lambda \mathbf{y} \rangle &= \langle \mathbf{x}, \mathbf{x} \rangle + \lambda \langle \mathbf{x}, \mathbf{y} \rangle + \lambda \langle \mathbf{x}, \mathbf{y} \rangle + \lambda^2 \langle \mathbf{y}, \mathbf{y} \rangle. \\
&= \langle \mathbf{x}, \mathbf{x} \rangle + 2\lambda \langle \mathbf{x}, \mathbf{y} \rangle + \lambda^2 \langle \mathbf{y}, \mathbf{y} \rangle.
\end{aligned}$$

Le cas \mathbf{x} ou \mathbf{y} est trivial, on va donc supposer que les deux vecteurs sont nuls dans la suite. L'expression précédente peut donc se voir comme un polynôme de degré 2 en λ . Or $\langle \mathbf{x}, \mathbf{x} \rangle + 2\lambda \langle \mathbf{x}, \mathbf{y} \rangle + \lambda^2 \langle \mathbf{y}, \mathbf{y} \rangle \geq 0$ et $\langle \mathbf{y}, \mathbf{y} \rangle > 0$, donc le discriminant de ce polynôme

$$4\langle \mathbf{x}, \mathbf{y} \rangle^2 - 4\langle \mathbf{y}, \mathbf{y} \rangle \langle \mathbf{x}, \mathbf{x} \rangle$$

est négatif ou nul. On en déduit directement le résultat. \square

Définition 6.7: Espace euclidien

On appelle **espace euclidien** tout espace vectoriel réel de dimension finie muni d'un produit scalaire.

Un produit scalaire étant par définition une forme bilinéaire symétrique définie positive, il est donc possible de définir plusieurs produits scalaires sur un même espace vectoriel (une infinité même) vu que l'on peut définir une infinité de forme bilinéaire

symétrique définie positive.

Définition 6.8: Norme Euclidienne

On appelle **norme euclidienne** associée au produit scalaire $\langle \cdot, \cdot \rangle$, notée $\| \cdot \|$, l'application définie de E dans \mathbb{R}_+ pour tout $\mathbf{x} \in E$ par

$$\|\mathbf{x}\|_2^2 = \langle \mathbf{x}, \mathbf{x} \rangle.$$

Remarque. La définition précédente ne définit qu'un exemple particulier, mais il existe une définition plus générale de normes que vous serez amenés à rencontrer lorsque l'on fait du calcul matriciel ou encore de la modélisation statistique (*ex* : norme matricielle ou encore les normes L^p pour des méthodes pénalisées).

Définition 6.9: Norme

On appelle **norme** sur un espace vectoriel E toute application N définie de E dans \mathbb{R}_+ vérifiant les axiomes suivants :

- (i) $\forall (\alpha, \mathbf{x}) \in \mathbb{R} \times E, N(\alpha \mathbf{x}) = |\alpha|N(\mathbf{x})$ (homogénéité),
- (ii) $\forall \mathbf{x} \in E, N(\mathbf{x}) = 0 \iff \mathbf{x} = 0$ (séparation),
- (iii) $\forall (\mathbf{x}, \mathbf{x}') \in E \times E, N(\mathbf{x} + \mathbf{x}') \leq N(\mathbf{x}) + N(\mathbf{x}')$ (inégalité triangulaire).

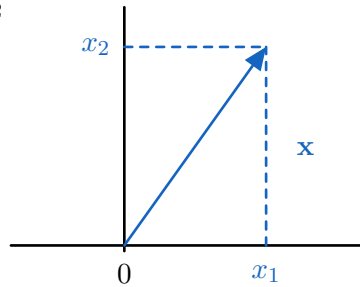
Exemple 6.7. On donne ici quelques exemples de normes :

- L'espace \mathbb{R} muni de la valeur absolue est un espace euclidien, ce qui veut dire dire que l'application $N(\cdot) = |\cdot|$
- La norme euclidienne définie précédemment est une norme, plus précisément, nous avons

$$\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^n x_i^2}.$$

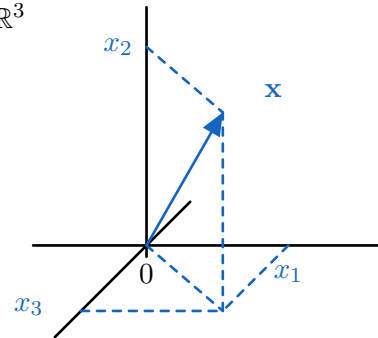
Cette norme correspond à un résultat que l'on connaît très bien en dimension 2 et le généralise à la dimension n ... **le théorème de Pythagore !**

$E = \mathbb{R}^2$



$$\|x\| = \sqrt{x_1^2 + x_2^2}$$

$E = \mathbb{R}^3$



$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

Vous pouvez essayer de montrer que la norme euclidienne est effectivement une norme en montrant que les trois axiomes sont vérifiés.

- Plus généralement, toutes les applications $N : \mathbf{x} \mapsto \|\mathbf{x}\|_p$, également appelées p -norme, sont des normes.

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}$$

L'exemple de la norme euclidienne et plus précisément le théorème de Pythagore permet de revenir sur un point important que l'on a déjà abordé précédemment : la notion d'*orthogonalité*.

Définition 6.10: Orthogonalité

Soit E un espace euclidien muni du produit scalaire $\langle \cdot, \cdot \rangle$ et \mathbf{v}, \mathbf{w} deux vecteurs de E . On dit que \mathbf{v} et \mathbf{w} sont orthogonaux si

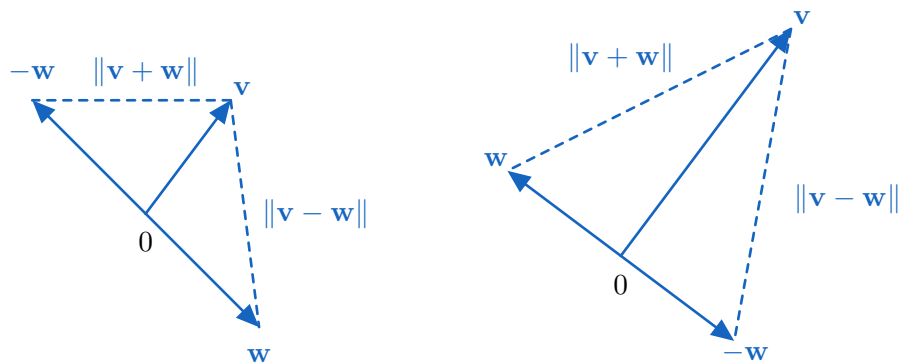
$$\langle \mathbf{v}, \mathbf{w} \rangle = 0.$$

Le résultat suivant permet également de montrer que deux vecteurs sont orthogonaux uniquement en regardant des normes.

Proposition 6.3: Caractérisation de l'Orthogonalité

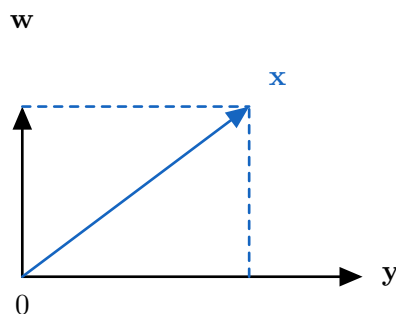
Soit E un espace euclidien muni du produit scalaire $\langle \cdot, \cdot \rangle$, $\|\cdot\|$ la norme associée et \mathbf{v}, \mathbf{w} deux vecteurs de E . Les vecteurs \mathbf{v} et \mathbf{w} sont orthogonaux si et seulement si

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



Démonstration. La démonstration est laissée à titre d'exercice. \square

A partir de cette notion d'orthogonalité, il est également possible de définir l'angle entre deux vecteurs d'un espace euclidien E . Pour cela regardons un petit exemple pour mieux comprendre la construction. Considérons deux vecteurs \mathbf{x} et \mathbf{y} définis dans une certaine base \mathcal{B} . On cherche à déterminer l'angle que fait le vecteur \mathbf{x} avec le vecteur \mathbf{y} en utilisant la notion d'orthogonalité vu précédemment.



Le vecteur \mathbf{x} peut se décomposer comme la somme de vecteurs orthogonaux, *i.e.* il existe un vecteur \mathbf{w} orthogonal à \mathbf{y} ainsi qu'un scalaire λ tels que :

$$\mathbf{x} = \lambda \mathbf{y} + \mathbf{w}.$$

Rappelez-vous, c'est une conséquence des résultats vus précédemment dans cette partie! Le vecteur $\lambda \mathbf{y}$ est appelé projeté orthogonal de \mathbf{x} sur \mathbf{y} . Il nous faut alors déterminer la valeur du scalaire λ . Pour cela, on se rappelle que \mathbf{y} est orthogonal à \mathbf{w} , donc $\langle \mathbf{x} - \lambda \mathbf{y}, \mathbf{y} \rangle = 0$, ce qui donne

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

et nous permet de définir indirectement la projection d'un vecteur \mathbf{x} sur un vecteur \mathbf{y}

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

Cette valeur λ qui représente le coefficient de la projection de \mathbf{x} sur \mathbf{y} est fortement liée à l'angle entre les deux vecteurs.

Définition 6.11: Angles entre deux vecteurs

Soient \mathbf{x} et \mathbf{y} deux vecteurs d'un espace euclidien E et notons \mathbf{x}' et \mathbf{y}' les normalisations de ces deux vecteurs, *i.e.*

$$\mathbf{x}' = \frac{\mathbf{x}}{\|\mathbf{x}\|} \quad \text{et} \quad \mathbf{y}' = \frac{\mathbf{y}}{\|\mathbf{y}\|}.$$

Alors le cosinus de l'angle formé entre les deux vecteurs \mathbf{x} et \mathbf{y} , noté $\lambda = \cos(\theta)$, est donné par

$$\lambda = \langle \mathbf{x}', \mathbf{y}' \rangle = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}.$$

C'est donc une valeur comprise entre -1 et 1 . En effet, cela est une conséquence de l'inégalité de Cauchy-Schwarz

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

On parle ici d'angle non orienté car deux valeurs sont possibles pour θ dans cette définition dans l'intervalle $[0, 2\pi]$. En revanche, il n'existe qu'une seule valeur θ dans l'intervalle $[0, \pi]$ telle que

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

Retour sur le produit scalaire. La définition 6.6 a permis d'introduire le produit scalaire comme une forme **bilinéaire, symétrique et définie positive**. Nous pouvons donc représenter ce produit scalaire à l'aide d'une matrice, si notre espace E est muni d'une base (voir définition 6.3).

Proposition 6.4: Représentation produit scalaire

Soit s un produit scalaire sur un espace euclidien E muni d'une base $(\mathbf{e}_i)_{i \in \mathbb{N}}$ et notons S la représentation de s dans cette base. Considérons deux vecteurs \mathbf{x} et \mathbf{x}' de E alors

$$s(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle_s = \mathbf{x}^T S \mathbf{y}.$$

En tant que matrice représentante d'un produit scalaire, cette matrice S doit vérifier les mêmes points que la définition de produit scalaire, *i.e.* elle doit être :

- **bilinéaire** ,
- **symétrique** : $S = S^T$,
- **définie positive** : $\forall \mathbf{x} \in E \quad \mathbf{x}^T S \mathbf{x} \geq 0$ et $\mathbf{x}^T S \mathbf{x} = 0 \iff \mathbf{x} = 0$

Remarque. La matrice associée à un produit scalaire est diagonalisable en tant que matrice symétrique réelle.

On voit donc qu'à toute forme bilinéaire, comme le produit scalaire, il est possible d'associer une matrice S . A cette matrice S est en fait associée un endomorphisme de E , que l'on notera g , et qui possède les mêmes propriétés que S . On peut ainsi écrire notre produit scalaire $s(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T S \mathbf{y} = \langle \mathbf{x}, g(\mathbf{y}) \rangle$

Nous serons amenés cette notation dans le cadre de la démonstration de certains résultats

On peut même montrer mieux que cela et sans grandes difficultés.

Proposition 6.5: Isomorphisme entre g et ϕ_g

Considérons l'application qui à un endomorphisme g associe la forme bilinéaire ϕ_g (notée aussi ϕ) est un isomorphisme de l'espace des endomorphismes symétriques $S(E)$ sur l'espace des forme bilinéaires symétriques sur E , $\mathcal{BL}_{sym}(E)$.

Démonstration. On vérifie facilement que l'application $g \mapsto \phi_g$ est linéaire, il suffit simplement d'écrire la définition.

Montrons maintenant que cette application est injective. Supposons que ϕ_g est nul, cela signifie que pour tout $\mathbf{x}, \mathbf{y} \in E$: $\langle \mathbf{x}, g(\mathbf{y}) \rangle$, ce qui signifie que $g(\mathbf{y}) = 0$ pour tout vecteur \mathbf{y} donc $g = 0$.

Observons maintenant que les espaces $S(E)$ et $\mathcal{BL}_{sym}(E)$ ont la même dimension.

L'application est donc bien un isomorphisme. \square

Exemple 6.8. Un premier exemple bien connu de produit scalaire est celui employé jusqu'à présent, c'est-à-dire le produit scalaire canonique de \mathbb{R}^n pour lequel le produit scalaire est communément appelé produit scalaire euclidien et qui peut se représenter sous la forme

$$\langle \mathbf{x}, \mathbf{y} \rangle_s = \mathbf{x}^T I \mathbf{x} = \mathbf{x}^T \mathbf{y}, \quad \text{dans ce cas } S = I.$$

On peut également définir un produit scalaire sur les matrices, par exemple sur l'ensemble des matrices carrées d'ordre n , $\mathcal{M}_n(\mathbb{R})$. Le produit scalaire ainsi défini s'appelle le produit scalaire de Frobenius, noté $\langle \cdot, \cdot \rangle_F$ (auquel on pourra également rattacher une norme, la norme de Frobenius).

Etant données deux matrices carrées A et B d'ordre n , dont les coefficients sont respectivement notés a_{ij} et b_{ij} , alors le produit scalaire de Frobenius est défini par :

$$\langle A, B \rangle_F = \text{Tr}(A^T B) = \sum_{i,j=1}^n a_{ij} b_{ij}.$$

Vérifions qu'il s'agit bien d'un produit scalaire :

- **Symétrie** : on doit vérifier que $\langle A, B \rangle_F = \langle B, A \rangle_F$ or

$$\begin{aligned} \langle A, B \rangle_F &= \sum_{i,j=1}^n a_{ij} b_{ij}, \\ &= \sum_{i,j=1}^n b_{ij} a_{ij}, \\ \langle A, B \rangle_F &= \langle A, B \rangle_F. \end{aligned}$$

- **Bilinéarité** : soit de plus $\lambda \in \mathbb{R}$ et $C \in \mathcal{M}_n(\mathbb{R})$

$$\begin{aligned} \langle A, B + \lambda C \rangle_F &= \sum_{i,j=1}^n a_{ij} b_{ij} + \lambda c_{ij}, \\ &= \sum_{i,j=1}^n a_{ij} b_{ij} + \lambda \sum_{i,j=1}^n a_{ij} c_{ij}, \\ \langle A, B + \lambda C \rangle_F &= \langle A, B \rangle_F + \lambda \langle A, C \rangle_F. \end{aligned}$$

L'application est donc bien linéaire à droite, elle l'est aussi à gauche par symétrie, elle est donc bilinéaire.

- **Définie positive** : on aisément $\langle A, A \rangle_F = \sum_{i,j=1}^n a_{ij}^2 \geq 0$ et cette somme est égale à 0 si et seulement si tous les termes sont égaux à 0 (en tant que somme de termes positifs), donc $a_{ij}^2 = 0 \iff a_{ij} = 0$.

Changement de base et bases orthonormées. Regardons comment le changement de base influence la représentation matricielle de notre produit scalaire (donc indirectement d'une forme bilinéaire symétrique définie positive).

Proposition 6.6: Changement de base produit scalaire

Soit \mathcal{B} et \mathcal{B}' deux bases d'une espace euclidien E muni du produit scalaire s et notons P la matrice de passage de \mathcal{B} vers \mathcal{B}' . Notons également S et S' les représentations matricielles de S et S' respectivement dans les bases \mathcal{B} et \mathcal{B}' , alors

$$S' = P^T S P.$$

Démonstration. La démonstration est semblable à celle effectuée en Proposition 3.9.

Soient \mathbf{x} et \mathbf{x}' les représentations d'un vecteur de E dans les bases \mathcal{B} et \mathcal{B}' et soient \mathbf{y} et \mathbf{y}' les représentations d'un vecteur de E dans ces mêmes bases. Nous avons vu comment opérer un changement de base sur les vecteurs et nous avons les relations

$$\mathbf{x}' = P\mathbf{x} \quad \text{et} \quad \mathbf{y}' = P\mathbf{y}.$$

Par conséquent

$$s(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T S \mathbf{y} = (P\mathbf{x}')^T S P \mathbf{y}' = \mathbf{x}'^T (P^T S P) \mathbf{y}' = \mathbf{x}'^T S' \mathbf{y}' = s'(\mathbf{x}', \mathbf{y}').$$

□

Exemple 6.9. Considérons \mathbf{x} et \mathbf{y} deux vecteurs de \mathbb{R}^2 et soient \mathcal{B} la base canonique de \mathbb{R}^2 et $(\mathbf{e}'_1, \mathbf{e}'_2)$ une autre base de \mathbb{R}^2 telle que

$$\mathbf{e}'_1 = 2\mathbf{e}_1 - 4\mathbf{e}_2 \quad \text{et} \quad \mathbf{e}'_2 = -\mathbf{e}_1 + 2\mathbf{e}_2.$$

On cherche à écrire le produit scalaire canonique de \mathbb{R}^2 dans notre nouvelle base $(\mathbf{e}'_1, \mathbf{e}'_2)$. Remarquons que $S = I$ dans le cas présent et que la matrice de passage P s'écrit $P = \begin{pmatrix} 3 & -1 \\ -4 & 2 \end{pmatrix}$.

On aura donc

$$S' = P^T I P = P^T P = \begin{pmatrix} 3 & -1 \\ -4 & 2 \end{pmatrix} \begin{pmatrix} 3 & -4 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 10 & -14 \\ -14 & 20 \end{pmatrix}.$$

Définition 6.12: Base orthogonale/orthonormée

Soit E un espace euclidien muni d'une base $\mathcal{B} = (\mathbf{e}_i)_{i=1}^n$. Cette base est dite orthogonale si pour tout i, j tels que $i \neq j$ on a

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

De plus, cette base est dite orthonormée si elle vérifie

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{si } i=j, \\ 0 & \text{sinon.} \end{cases}$$

De cette définition on peut tout de suite voir que de toute base orthogonale, il est possible de construire une base orthonormée. Pour cela, si $(\mathbf{e}_i)_{i=1}^n$ forme une base **orthogonale**, alors $\left(\frac{\mathbf{e}_i}{\|\mathbf{e}_i\|^2} \right)_{i=1}^n$ forme une base **orthonormale**.

Dans une base orthonormale, le produit scalaire s'exprime de façon très simple et semblable au produit scalaire canonique. En effet, considérons $(\mathbf{e}_i)_{i=1}^n$ une base orthonormale de E et \mathbf{x} et \mathbf{y} deux vecteurs de E exprimés dans cette base. Alors

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i,j=1}^n \langle x_i \mathbf{e}_i, y_j \mathbf{e}_j \rangle = \sum_{i,j=1}^n x_i y_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \sum_{i=1}^n x_i y_i.$$

Il existe un moyen, partant d'une base quelconque d'un espace vectoriel euclidien, de construire une base orthogonale (voire orthonormale). Cela peut se faire à l'aide du **procédé d'orthogonalisation de Gram-Schmidt**, nous ne l'étudierons pas ici mais on en présente rapidement l'idée.

Soit $(\mathbf{e}_i)_{i=1}^n$ une base a priori quelconque de E et définissons la **projection du vecteur \mathbf{e} sur le vecteur \mathbf{u}** :

$$proj_{\mathbf{u}}(\mathbf{e}) = \frac{\langle \mathbf{u}, \mathbf{e} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}.$$

Alors la famille $(\mathbf{u}_i)_{i=1}^n$ définie par :

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{e}_1 \\ \mathbf{u}_2 &= \mathbf{e}_2 - proj_{\mathbf{u}_1}(\mathbf{e}_2) = \mathbf{e}_2 - \frac{\langle \mathbf{u}_1, \mathbf{e}_2 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 \\ \mathbf{u}_3 &= \mathbf{e}_3 - proj_{\mathbf{u}_2}(\mathbf{e}_3) - proj_{\mathbf{u}_1}(\mathbf{e}_3) = \mathbf{e}_3 - \frac{\langle \mathbf{u}_2, \mathbf{e}_3 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle} \mathbf{u}_2 - \frac{\langle \mathbf{u}_1, \mathbf{e}_3 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 \\ &\vdots \\ \mathbf{u}_k &= \mathbf{e}_k - \sum_{l=1}^{k-1} proj_{\mathbf{u}_l}(\mathbf{e}_k) = \mathbf{e}_k - \sum_{l=1}^{k-1} \frac{\langle \mathbf{u}_l, \mathbf{e}_k \rangle}{\langle \mathbf{u}_l, \mathbf{u}_l \rangle} \mathbf{u}_l \\ &\vdots \\ \mathbf{u}_n &= \mathbf{e}_n - \sum_{l=1}^{n-1} proj_{\mathbf{u}_l}(\mathbf{e}_n) = \mathbf{e}_n - \sum_{l=1}^{n-1} \frac{\langle \mathbf{u}_l, \mathbf{e}_n \rangle}{\langle \mathbf{u}_l, \mathbf{u}_l \rangle} \mathbf{u}_l \end{aligned}$$

est une base orthogonale de E .

Si l'on dispose d'une famille de vecteurs $(\mathbf{e}_i)_{i=1}^n$ d'un espace euclidien E qui est orthonormale, relativement à la base canonique de E et que l'on note P la matrice de passage de la base canonique vers $(\mathbf{e}_i)_{i=1}^n$ alors $P^T P = I$.
Les matrices ayant une telle propriété sont dites **orthogonales** et vérifient $P^{-1} = P^T$.

Le procédé d'orthogonalisation de Gram-Schmidt évoqué précédemment permet de monter le résultat suivant :

Proposition 6.7: Existence d'une base orthonormée

Tout espace euclidien E possède une base orthonormée.

On peut maintenant retourner à l'étude de nos endomorphismes symétriques réelles dont le résultat suivant va jouer un rôle majeur dans certaines méthodes d'analyses de données.

Proposition 6.8: Endomorphismes symétriques et diagonalisation

Soit ϕ une forme bilinéaire symétrique sur un espace euclidien E , alors il existe une base orthonormée de E dans laquelle la matrice de ϕ est diagonale.

Bien que nous ne montrons pas ce résultat cela nous permet de justifier l'existence du résultat suivant permettant de caractériser les formes quadratiques relativement au signe de ses valeurs propres.

Proposition 6.9: Caractérisation des formes quadratiques

Soient E un espace vectoriel sur \mathbb{K} et q une forme quadratique sur E dont la représentation matricielle est notée A . Alors

- q est semi-définie positive (*resp.* définie positive) si et seulement si ses valeurs propres sont positives (*resp.* strictement positives)
- q est semi-définie négative (*resp.* définie négative) si et seulement si ses valeurs propres sont négatives (*resp.* strictement négatives).

Démonstration. Soit \mathcal{B} une base orthonormée dans laquelle l'endomorphisme g associée à la matrice A est diagonale et notons $\lambda_1, \dots, \lambda_n$ ses valeurs propres. Pour tout vecteur \mathbf{x} de E , on a :

$$\langle \mathbf{x}, g(\mathbf{x}) \rangle = \sum_{i=1}^n \lambda_i x_i^2,$$

où les x_i représentent les coordonnées du vecteur \mathbf{x} dans la base \mathcal{B} .

Remarquons que si g est positif, alors $\lambda_i = \langle \mathbf{e}_i, g(\mathbf{e}_i) \rangle$ est positif pour tout i . Les vecteurs \mathbf{e}_i désignant les vecteurs de la base orthonormée \mathcal{B} . Réciproquement, si λ_i est positif pour tout i alors, pour tout \mathbf{x} :

$$\langle \mathbf{x}, g(\mathbf{x}) \rangle = \sum_{i=1}^n \lambda_i x_i^2 \geq 0.$$

L'endomorphisme g est donc bien positif.

On démontre le côté défini positif en remarquant que g est défini positif, si et seulement si il est positif et inversible. La démonstration est analogue pour le deuxième point.

□

Espaces Orthogonaux

Définition 6.13: Sous espace orthogonal

Soit E un espace vectoriel euclidien et F un sous-espace quelconque de E strictement inclus dans E . On appelle sous espace orthogonal de F dans E , noté F^\perp , l'ensemble des vecteurs \mathbf{x} de E qui sont orthogonaux à tout vecteur de F

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

On peut montrer que F^\perp est un sous-espace vectoriel de E .

Proposition 6.10: Propriétés espaces orthogonaux

Soit E un espace vectoriel euclidien et F un sous espace de E . Alors

- $\dim(E) = \dim(F) + \dim(F^\perp)$,
- $E = F \oplus F^\perp$,
- $(F^\perp)^\perp = F$.

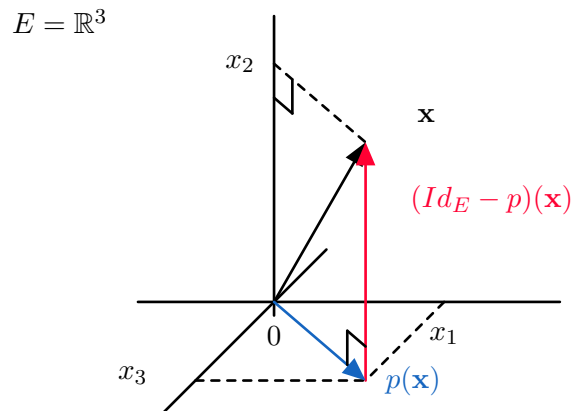
Nous avons déjà construits de tels sous-espaces précédemment à l'aide de certaines applications. Souvenez vous des projecteurs orthogonaux étudiés au tout début de cette partie (voir la Définition 1.11 et les Propositions 1.10 et 1.11).

Pour rappel, ces résultats ont montré que le noyau et l'image d'un projecteur sont supplémentaires dans l'espace de départ. De plus, nous avons

$$\text{Ker}(p) = \text{Im}(Id_E - p) = (\text{Im}(p))^\perp.$$

On rappelle également que le projecteur est une application idempotente (i.e. $p \circ p = p$).

Illustrons cela à l'aide d'un petit schéma où l'on considère un vecteur \mathbf{x} de \mathbb{R}^3



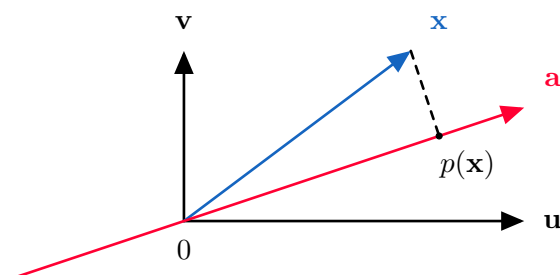
On considère un projecteur p sur l'espace engendré par les vecteurs de base $(\mathbf{e}_1, \mathbf{e}_3)$ parallèlement à la droite vectorielle engendrée par le vecteur \mathbf{e}_2 . Nous avons donc

$$\text{Ker}(p) = \text{Vect}(\mathbf{e}_2) = \text{Im}(Id_E - p) \quad \text{et} \quad \text{Im}(p) = \text{Ker}(Id_E - p) = \text{Vect}(\mathbf{e}_1, \mathbf{e}_3).$$

Projections sur un sous espace Nous terminons cette section en regardant comment déterminer le projeté orthogonal d'un vecteur sur un sous espace vectoriel d'un espace vectoriel euclidien.

Pour cela, considérons un espace euclidien E de dimension n et F un sous-espace de E de dimension $p < n$. Soit \mathbf{x} un vecteur de E . Nous allons regarder comment déterminer le projeté de \mathbf{x} sur une droite vectorielle (*i.e.* le cas où F est un espace de dimension 1) puis on traitera le cas plus général où la dimension de F est $2 \leq p < n$.

- **Projection sur une droite :** on considère la droite vectorielle F engendrée par un vecteur \mathbf{a} et on cherche à déterminer le projeté de \mathbf{x} sur \mathbf{a} .



$p(\mathbf{x})$ correspond au projeté du vecteur \mathbf{x} sur la droite vectorielle engendrée par \mathbf{a} . En ce sens $p(\mathbf{x})$ peut s'interpréter comme le point de $\text{Vect}(\mathbf{a})$ le plus proche de \mathbf{x} , *i.e.* cela reviendrait à chercher les coordonnées α sur le vecteur \mathbf{a} qui soit solution du problème suivant :

$$\min_{\alpha \in \mathbb{R}} \|\mathbf{x} - \alpha \mathbf{a}\|_2^2,$$

où $\alpha \mathbf{a} = p(\mathbf{x})$.

Plutôt que de chercher à résoudre ce problème d'optimisation dont la résolution sera traitée dans un cours dédié, nous allons simplement nous contenter des outils de l'algèbre linéaire.

Pour cela, utilisons simplement le fait que $p(\mathbf{x}) = \alpha \mathbf{a}$ comme énoncé plus tôt (*i.e.* le projeté appartient à la droite vectorielle engendrée par \mathbf{a}). Dans ce cas, les vecteurs \mathbf{a} et $\mathbf{x} - \alpha \mathbf{a}$ sont orthogonaux, d'où

$$\begin{aligned} \langle \mathbf{a}, \mathbf{x} - \alpha \mathbf{a} \rangle &= 0, \\ \alpha \langle \mathbf{a}, \mathbf{a} \rangle &= \langle \mathbf{a}, \mathbf{x} \rangle, \\ \alpha &= \frac{\langle \mathbf{a}, \mathbf{x} \rangle}{\langle \mathbf{a}, \mathbf{a} \rangle}, \\ \alpha &= \frac{\langle \mathbf{a}, \mathbf{x} \rangle}{\|\mathbf{a}\|^2}. \end{aligned}$$

On aura simplement développer les expressions.

Ainsi $p(\mathbf{x}) = \alpha \mathbf{a} = \frac{\langle \mathbf{a}, \mathbf{x} \rangle}{\|\mathbf{a}\|^2} \mathbf{a}$. Remarquons que seul le vecteur \mathbf{x} a une influence sur l'expression $p(\mathbf{x})$ mais pas le vecteur \mathbf{a} qui ne sert qu'à donner la direction, *i.e.* multiplier le vecteur \mathbf{a} par λ ne change pas l'expression du projeté, mais multiplier \mathbf{x} par λ implique de multiplier l'expression de $p(\mathbf{x})$ par λ .

Dernière remarque, on peut réécrire l'expression de $p(\mathbf{x})$ sous forme matricielle, en remarquant que $p(\mathbf{x}) = \frac{\mathbf{a}\mathbf{a}^T}{\|\mathbf{a}\|^2} \mathbf{x} = P\mathbf{x}$ où $P = \frac{\mathbf{a}\mathbf{a}^T}{\|\mathbf{a}\|^2}$.

Dans ce cas, la matrice P définit bien une matrice projection (de rang 1) car

$$P^2 = P \quad \text{et} \quad P^T = P.$$

Généralisons cela pour un sous espace de dimension $2 \leq p < n$.

- **Projection sur un hyperplan :**

Cette fois-ci F est un sous-espace vectoriel de E de dimension $2 \leq p < n$ généré par les vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_p$ de \mathbb{R}^n qui forment une base de F .

Dans ce cas, l'espace F est l'espace générée par les colonnes de la matrice $A = (\mathbf{a}_1, \dots, \mathbf{a}_p)$. Nous avons précédemment vu que le projeté de \mathbf{x} sur le sous espace F pouvait s'écrire $p(\mathbf{x}) = P\mathbf{x}$ où P est une certaine matrice, mais surtout, on sait que le projeté va s'écrire comme une combinaison linéaire des vecteurs colonnes de A , *i.e.* $p(\mathbf{x}) = \sum_{j=1}^p \alpha_j \mathbf{a}_j = A\boldsymbol{\alpha}$.

On peut alors utiliser le même raisonnement que précédemment, on sait que le vecteur $\mathbf{x} - p(\mathbf{x})$ est orthogonal à tout vecteur de F , *i.e.*

$$\begin{aligned}\langle \mathbf{a}_1, \mathbf{x} - A\boldsymbol{\alpha} \rangle &= 0, \\ &\vdots \\ \langle \mathbf{a}_p, \mathbf{x} - A\boldsymbol{\alpha} \rangle &= 0.\end{aligned}$$

Ce qui peut s'écrire de façon matricielle :

$$\langle A, \mathbf{x} - A\boldsymbol{\alpha} \rangle = A^T(\mathbf{x} - A\boldsymbol{\alpha}) = 0.$$

Notre objectif est de trouver le vecteur $\boldsymbol{\alpha}$, on va donc l'isoler.

$$\begin{aligned}A^T(\mathbf{x} - A\boldsymbol{\alpha}) &= 0, \\ A^T A\boldsymbol{\alpha} &= A^T \mathbf{x}\end{aligned}$$

Comme les colonnes de A forment une base de F la matrice $A^T A \in \mathcal{M}_p(\mathbb{R})$ est inversible. On a donc

$$\boldsymbol{\alpha} = (A^T A)^{-1} A^T \mathbf{x}.$$

Donc le projeté orthogonal de \mathbf{x} sur F est $p(\mathbf{x}) = A (A^T A)^{-1} A^T \mathbf{x}$.

On peut résumer tout cela par la proposition suivante.

Proposition 6.11: Projection orthogonal sur un sous-espace

Soit E un espace euclidien de dimension n et F un sous espace de E de dimension $p < n$ dont les vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_p$ forment une base. On note A la matrice dont les colonnes sont formées des vecteurs de cette base.

Alors, pour tout vecteur \mathbf{x} de E , le projeté orthogonal $p(\mathbf{x})$ de \mathbf{x} sur F est donné par

$$p(\mathbf{x}) = A (A^T A)^{-1} A^T \mathbf{x}$$

et $P = (A^T A)^{-1} A^T$ est la matrice de projection sur le sous-espace F .

Vérifier ses connaissances

1. Les propositions suivantes sont-elles vraies ou fausses ?

(a) L'application $\varphi : \mathbb{R}^2 \times \mathbb{R}$ définie par

$$\phi(\mathbf{x}, \mathbf{y}) = 3x_1y_2 + 6x_2y_2 + 3x_1y_1 + 4x_2y_1$$

est-elle une forme bilinéaire ?

(b) L'application $\varphi : \mathbb{R}^2 \times \mathbb{R}$ définie par

$$\phi(\mathbf{x}, \mathbf{y}) = -2x_1y_2^2 + 5x_2y_2 - 7x_1y_1$$

est-elle une forme bilinéaire ?

(c) L'application $\varphi : \mathbb{R}^2 \times \mathbb{R}$ définie par

$$\phi(\mathbf{x}, \mathbf{y}) = 2x_1y_1 - 3x_1y_2 - 3x_2y_1 + 6x_2y_2$$

est-elle une forme bilinéaire symétrique ?

(d) L'application $\varphi : \mathbb{R}^2 \times \mathbb{R}$ définie par

$$\phi(\mathbf{x}, \mathbf{y}) = 2x_1 - 3x_1y_2 - 3x_2y_1 + 2y_2$$

est-elle une forme bilinéaire symétrique ?

(e) Les matrices associées aux formes bilinéaires symétriques sont nécessairement symétriques.

(f) A toute forme bilinéaire symétrique, il est possible d'associer une forme quadratique.

(g) Une forme quadratique est dite positive si et seulement si toutes ses valeurs propres sont positives.

(h) L'application $\varphi : \mathbb{R}^2 \times \mathbb{R}^2$ définie par

$$\phi(\mathbf{x}, \mathbf{y}) = 2x_1y_1 + 3x_1y_2 + 3x_2y_1 + 6x_2y_2$$

est-elle un produit scalaire ?

(i) L'application $\varphi : \mathbb{R}^2 \times \mathbb{R}$ définie par

$$\phi(\mathbf{x}, \mathbf{y}) = 2x_1y_1 + 3x_1y_2 + 3x_2y_1 - 6x_2y_2$$

est-elle un produit scalaire ?

(j) On peut associer une norme à tout produit scalaire $\langle \cdot, \cdot \rangle$.

(k) Tout endomorphisme d'un espace euclidien est diagonalisable.

(l) De toute base $\mathcal{B} = (\mathbf{e}_i)_{i=1}^n$ d'un espace euclidien E de dimension n , il est possible de construire une base orthonormale.

- (m) L'inégalité de Cauchy-Schwarz devient une égalité lorsque les deux vecteurs sont orthogonaux.
 - (n) Si une matrice P est orthogonale, alors son inverse est égale à sa transposée.
 - (o) L'application définie comme la combinaison linéaire de deux normes est une norme.
 - (p) Une application définie comme la somme positive de deux normes est encore une norme.
 - (q) Toute matrice orthogonale P est inversible.
2. Déterminer si les colonnes des matrices suivantes forment une base orthogonale de \mathbb{R}^3 :

$$A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad B = \begin{pmatrix} -1 & 2 & 0 \\ 1 & 1 & -2 \\ 1 & 1 & 2 \end{pmatrix} \quad \text{et} \quad C = \begin{pmatrix} 2 & -2 & 1 \\ -1 & -1 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

3. Déterminer le signe des formes quadratiques suivantes

$$A = \begin{pmatrix} 2 & 5 & -6 \\ 5 & 6 & 3 \\ -6 & 3 & -1 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 3 & 3 \\ 3 & 2 & -1 \\ 3 & -1 & -2 \end{pmatrix} \quad \text{et} \quad C = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$


4. On se donne une droite générée par le vecteur $\mathbf{a} = (-2, 3, 4)$. Déterminer le projeté orthogonal des vecteurs suivant sur la droite vectorielle engendrée par \mathbf{a} .
- (a) $\mathbf{x} = (-2, 4, 1)$,
 - (b) $\mathbf{y} = (0, 3, 0)$,
 - (c) $\mathbf{z} = (-1, -3, 0.5)$.

Code R

Toutes les commandes nécessaires à la manipulation des objets dans cette section ont déjà été présentées dans les sections précédentes, mise à part le produit scalaire entre deux vecteurs.

```
# Définition des vecteurs
x = c(1,2,3)
y = c(0,3,4)
# Produit scalaire de x et y
x%*%y

##          [,1]
## [1,]      18
```

On va surtout profiter de cette section pour étudier les options graphiques sous  via l'étude des formes quadratiques et linéaire dans un espace de dimension 2.

On va, pour simplifier, considérer que la forme quadratique q a déjà été réduite et qu'elle se trouve dans une base telle que la matrice associée soit diagonale, *i.e.* la matrice de notre forme quadratique est de la forme

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

où $\lambda_1, \lambda_2 \in \mathbb{R}$. On souhaite maintenant regarder l'allure de la forme quadratique en fonction des valeurs de λ_1 et λ_2 , plus précisément, en fonction de leur signe (positif, nul ou négatif). On va s'intéresser aux courbes de la forme $q(\mathbf{x}) + l(\mathbf{x}) = k$, $k \in \mathbb{R}$, c'est à dire aux courbes d'équation

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + ax_1 + bx_2 = k$$

Cas 1 Imaginons que les deux valeurs propres sont strictement positives et prenons $k = 1$ et $a = b = 0$. On va essayer de représenter la courbe. Pour cela on va essayer d'exprimer x_2 (que l'on assimilera à l'ordonnée) en fonction de x_1 , ainsi

$$x_2 = \pm \sqrt{\frac{1}{\lambda_2} (-\lambda_1 x_1^2 + 1)},$$

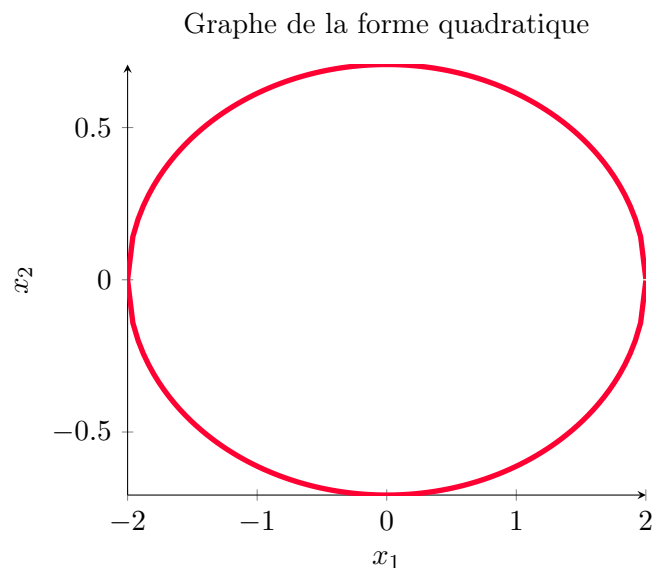
où x_1 est tel que $1 - \lambda_1 x_1^2$ soit positif, on va fixer $\lambda_2 = 1$ et $\lambda_1 = 1/4$ dans la suite. Ainsi $x_1 \in [-2, 2]$.

```
# On va definir notre courbe en deux équations differentes
# Cela donnera alors deux courbes que l'on superposera.


# On definit les objets
lambda1 = 1/4
lambda2 = 2
x1 = seq(-2,2,by=0.01)
# On calcule nos deux images
x2_plus = sqrt((1/lambda2)*(1-lambda1*x1^2))
x2_moins = -sqrt((1/lambda2)*(1-lambda1*x1^2))
```

Il ne nous reste plus qu'à représenter notre courbe, qui aura donc la forme d'une ellipse.

```
# On trace notre courbe en deux morceaux.
plot(x1,x2_plus, type = 'l', lwd = 2 ,
     xlim = c(-2.5,2.5) ,
     ylim = c(-1.5,1.5), xlab = "" , ylab = "",
     main = "Les valeurs propres sont positives")
lines(x1,x2_moins, type = 'l', lwd = 2)
```



Ainsi, pour représenter graphiquement un nuage de points, on peut utiliser la fonction *plot*, qui va prendre différents arguments, mais nous ne présentons que ceux utilisés ci-dessus :

- le premier argument "x" spécifie les valeurs en abscisse : ici x_1
- le deuxième argument "y" spécifie les valeurs en abscisse : ici x_2
- *type* indique le type de tracé que nous souhaitons. Avec *type* = 'l', nous demandons à  de relier les points, sinon nous aurions par défaut un nuage de points.
- *lwd* permet de régler l'épaisseur du trait.
- *main* : permet de donner un nom au graphique, ce nom sera mis en entre "".
- *xlim* et *ylim* permettent de régler la fenêtre graphique en indiquant les valeurs minimales et maximales sur chaque axe.
- *xlab* et *ylab* (lab pour *label*, i.e. étiquette, permet de donner un nom aux axes. Ici l'utilisation des "" permet de supprimer le nom des axes.

La commande *lines* utilisée pour tracer le deuxième graphe permet de superposer un graphique à un graphique existant.

Cas 2 On suppose que l'une des valeurs propres est nulle, disons $\lambda_2 = 0$ et que $k = a = 0$ également. Dans ce cas, on se retrouve avec une équation de la forme

$$\lambda_2 x_1^2 = b x_2 \iff x_2 = \frac{\lambda_1}{b} x_1^2.$$

On reconnaît ici l'équation d'une parabole, vérifions cela de suite en représentant graphiquement cette fonction comme nous l'avons fait précédemment. On posera $\lambda_1 = 2$ et $b = 3$.

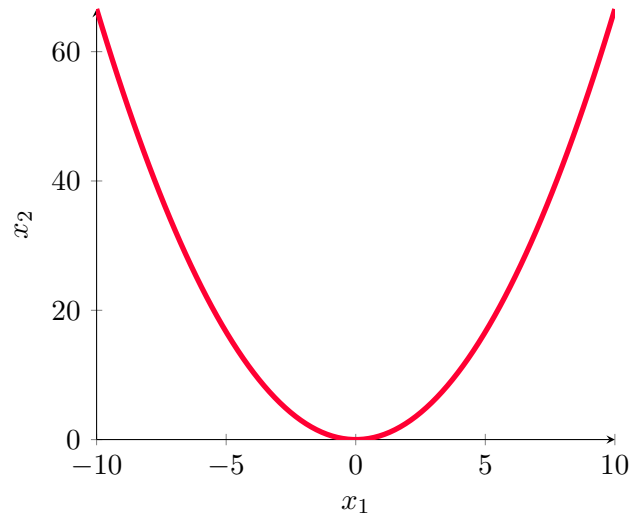
```
# Definition des variables
lambda1 = 2
b = 3
x1 = seq(-5,5,by=0.01)
# On calcule l'image de x1 par notre fonction
x2 = (lambda1/b)*x1^2
```

Il ne nous reste plus qu'à représenter notre courbe, qui aura donc la forme d'une parabole

```
# On trace notre courbe.
plot(x1,x2, type = 'l', lwd = 2, col = "red",
main = "Une valeur propre positive")
```

L'option *col* permet de définir la couleur du graphique.

Graphe de la forme quadratique



Cas 3 On suppose que les deux valeurs propres sont de signes opposés $\lambda_1 < 0 < \lambda_2$ et que $a = b = 0$. On va également supposer que $k = 1$ dans le cas présent. Dans ce cas, on se retrouve avec une équation de la forme

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 = 1 \iff x_2 = \pm \sqrt{\frac{1}{\lambda_2} (1 - \lambda_1 x_1^2)}.$$

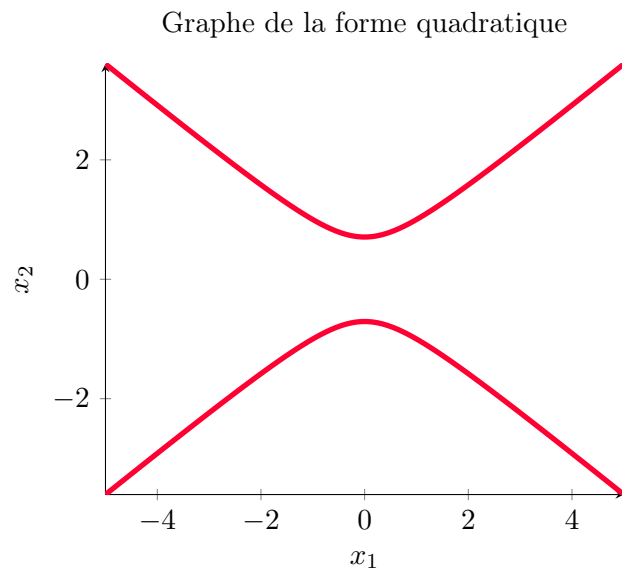
Comme dans le Cas 1, on va à nouveau tracer notre graphique en plusieurs étapes, on va donc superposer plusieurs graphes. Pour cela, on va commencer par supposer que $\lambda_1 = -1$ et $\lambda_2 = 2$ et réécrire l'équation de notre forme quadratique.

```
# Definition des variables
lambda1 = -1
lambda2 = 2
x1 = seq(-5,5,by=0.01)
# On calcule nos deux images
x2_plus = sqrt((1/lambda2)*(1-lambda1*x1^2))
x2_moins = -sqrt((1/lambda2)*(1-lambda1*x1^2))
```

Il ne nous reste plus qu'à représenter notre courbe en deux phases.

```
# On trace notre courbe en deux morceaux.
plot(x1,x2_plus, type = 'l', lwd = 2 ,
      xlim = c(-2.5,2.5) ,
      ylim = c(-1.5,1.5), xlab = "" , ylab = "",
```

```
main = "Valeurs propres de signes opposés")
lines(x1,x2_moins, type = 'l', lwd = 2)
```



La figure ainsi obtenue représente les deux branches d'une hyperbole qui ne se coupent pas et qui sont séparées par l'axe des abscisses. Nous pourrions obtenir un graphe analogue mais avec deux branches d'hyperbole "verticales", *i.e.* séparées par l'axe des ordonnées si nous avions échangé le signe des valeurs propres (ce que l'on comprendrait aussi si on échangeait l'axe des abscisses et celui des ordonnées sur la figure actuelle).

Troisième partie

Analyse de données

La première partie a permis d'introduire toutes les définitions et outils algébriques et géométriques pour l'analyse de forme quadratiques et notamment la recherche et le calcul de valeurs propres d'un endomorphisme. Nous allons maintenant voir comment mobiliser ces outils là pour analyser des données de différents types : réelles (température, salaire, taille, nombre d'enfants, ...) - catégorielles (profession, sexe, ...) ou encore ordinales (score attribué à un produit ou film, ...).

Cette deuxième partie traite des méthodes réduction de la dimension, nous verrons comment synthétiser des informations issues de données plongées dans des espaces de dimensions n ou p . Plus concrètement, cette synthèse se fera en adoptant une représentation des données dans un espace de dimension 2 voire 3, permettant de **visualiser** les informations contenues dans les données. Les outils utilisés combinent l'usage des statistiques et de l'algèbre linéaire, nous utilisons surtout les notions suivantes :

- la notion de distances entre des points, les projections orthogonales et la notion de métrique,
- la recherche de valeurs propres d'un endomorphisme et ses vecteurs propres.

Dans toute cette partie, la notation n désignera le nombre d'individus dans notre échantillon (ou le nombre d'exemples) et p le nombre de descripteurs pour un exemple donné (*i.e.* le nombre de variables).

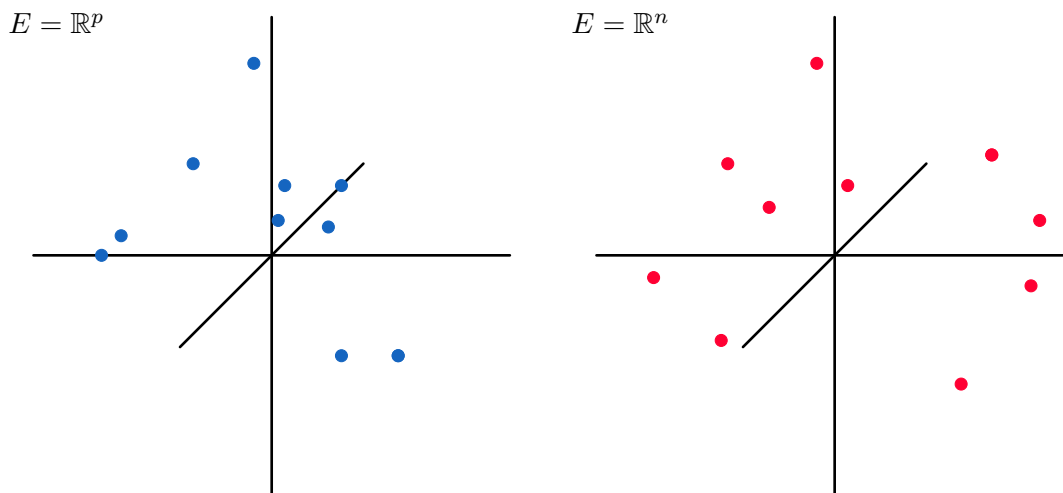
En analyse de données, il est d'usage de représenter ces individus dans un *grand tableau de données* que l'on pourra alors considérer comme une matrice avec certaines dimensions. Pour cela on considère

$$X = \begin{matrix} & \mathbf{v}_1 & \cdots & \mathbf{v}_k & \cdots & \mathbf{v}_p \\ \mathbf{x}_1 & \left(\begin{matrix} x_{11} & \cdots & x_{1k} & \cdots & x_{1p} \\ \vdots & & \vdots & & \vdots \\ \mathbf{x}_i & \begin{pmatrix} x_{i1} & \cdots & x_{ik} & \cdots & x_{ip} \end{pmatrix} \\ \vdots & & \vdots & & \vdots \\ \mathbf{x}_n & \begin{pmatrix} x_{n1} & \cdots & x_{nk} & \cdots & x_{np} \end{pmatrix} \end{matrix} \right. \end{matrix},$$

où \mathbf{x}_i représente l'individu i avec les valeurs x_{ik} prises par les différents descripteurs \mathbf{v}_k .

A partir de ce simple tableaux de données, il est possible d'adopter deux représentations

- On peut choisir de représenter les **individus** \mathbf{x}_i dans l'espaces des **variables** \mathbf{v}_k , une première représentation qui est sûrement la plus utilisée. Dans ce cas chaque point \mathbf{x}_i a pour coordonnées $(\mathbf{x}_{i1}, \dots, \mathbf{x}_{ip})$ dans l'espace \mathbb{R}^p .
On obtient un premier nuage de points que l'on notera **nuage des individus**.
- On peut également faire le choix de représenter les **variables** dans l'espace des **individus**. Dans ce cas chaque point \mathbf{v}_k a pour coordonnées $(\mathbf{v}_{1k}, \dots, \mathbf{v}_{nk})$ dans l'espace \mathbb{R}^n .
Ce deuxième nuage de points est appelé **nuage des variables**.



On montrera qu'il existe un lien très fort entre ces deux représentations et que ce dernier repose sur la décomposition en valeurs singulières de cette matrice de données.

L'objectif de cette partie est de fournir des réponses à des questions comme

- Est-ce que des variables sont corrélées entre elles ?
- Quelles directions de l'espace permettent d'expliquer au mieux la variabilité observée au sein des données ?
- Est-ce qu'il est possible d'obtenir une représentation fiable de nos données dans un espace de dimension faible afin de visualiser les informations ? Quel serait le sens de cette nouvelle présentation ?
- Est-ce qu'il existe des groupes d'individus dont le comportement pourrait être expliqué par des variables particulières ?

Bien évidemment, ces questions de représentations vont se limiter aux espaces de dimension 2 et 3 pour les aspects visualisation. Il serait ridicule de chercher une représentation dans un espace de dimension supérieure.

Bien évidemment, passer d'un espace de dimension p (ou n) à un espace de dimension 2

ou 3 entraîne souvent une perte d'informations, mais on va chercher les représentations les plus "fidèles". D'autre part, nous verrons également comment quantifier cette perte d'informations.

Les techniques de réduction de dimension étudiées dans cette partie sont :

- l'Analyse en Composantes Principales (ACP),
- l'Analyse Factorielle des Correspondance (AFC),
- l'Analyse factorielle des Correspondances Multiples (ACM).

7 Généralités et Décomposition en Valeurs Singulières (SVD)

Cette première section se focalise sur les prémices de l'analyse de données en énonçant les principales quantités utilisées pour les analyses ainsi que la méthode générale employée pour toutes les analyses présentées dans les sections suivantes.

7.1 Généralités

Lorsque l'on étudie des données, nous sommes amenés, le plus souvent, à nous intéresser à deux choses :

- l'analyse des corrélations entre les variables
- l'analyse des distances entre les individus

Ces deux critères recherches nous permettent de voir si notre jeu de données est riche en information. Pour quantifier cette information dans un nuage de points composé de n individus dans un espace de dimension p , on va mesurer la **variance**, notée Var_{tot} qui se trouve dans ce nuage (encore appelée **inertie totale**) définie par :

$$Var_{tot} = \frac{1}{n^2} \sum_{\mathbf{x} \in X} \sum_{\mathbf{x}' \in X} d^2(\mathbf{x}, \mathbf{x}') = \frac{1}{n} \sum_{i=1}^n d^2(\mathbf{x}_i, \bar{\mathbf{x}}),$$

où $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ est appelé **barycentre du nuage de points**. Il représente un individu *moyen* qui représente le nuage de points.

Plus la variance dans le nuage de points, *i.e.* plus les individus sont *dispersés* autour du barycentre, plus ce nuage de points est riche en information. A l'inverse, on comprend que si cette quantité est petite, cela signifie que tous les points sont concentrés autour du barycentre, et, potentiellement, ce nuage de points sera moins riche en information. Nous reviendrons sur ce point plus tard.

La notion de *variance* employée ici fait appel à la notion de distance que nous n'avons pas encore définie.

Définition 7.1: Distance

Soit E un ensemble (par exemple un espace vectoriel, mais ce n'est pas une obligation). On appelle **distance** sur l'espace E , tout application d de $E \times E$ à valeurs dans \mathbb{R}^+ qui vérifient les propriétés suivantes :

- **symétrie** : $\forall \mathbf{x}, \mathbf{x}' \in E, d(\mathbf{x}, \mathbf{x}') = d(\mathbf{x}', \mathbf{x})$
- **séparation** : $\forall \mathbf{x}, \mathbf{x}' \in E, d(\mathbf{x}, \mathbf{x}') = 0 \iff \mathbf{x} = \mathbf{x}'$
- **inégalité triangulaire** : $\forall \mathbf{x}, \mathbf{x}', \mathbf{x}'' \in E, d(\mathbf{x}, \mathbf{x}'') \leq d(\mathbf{x}, \mathbf{x}') + d(\mathbf{x}', \mathbf{x}'')$.

Exemple 7.1. Nous avons déjà rencontré des distances dans la première partie de ce document. Ce sont les distances induites par les normes, lorsque l'ensemble E est un espace vectoriel. On a alors

$$\forall \mathbf{x}, \mathbf{x}' \in E, \quad d(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|.$$

Ainsi on peut définir des distances pour tout entier $p \geq 1$ comme dans le cas des normes

$$\text{Distance de Manhattan} \quad d(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^n |\mathbf{x} - \mathbf{x}'|.$$

$$\text{Distance Euclidienne} \quad d(\mathbf{x}, \mathbf{x}') = \sqrt{\sum_{i=1}^n (\mathbf{x} - \mathbf{x}')^2}.$$

$$\text{Distance de Minkowski} \quad d(\mathbf{x}, \mathbf{x}') = \sqrt[p]{\sum_{i=1}^n |\mathbf{x} - \mathbf{x}'|^p}.$$

$$\text{Distance de Tchebychev} \quad d(\mathbf{x}, \mathbf{x}') = \lim_{p \rightarrow \infty} \sqrt[p]{\sum_{i=1}^n |\mathbf{x} - \mathbf{x}'|^p} = \sup_{1 \leq i \leq n} |\mathbf{x}_i - \mathbf{x}'_i|.$$

Cette première distance est sûrement la plus employée pour les analyses classiques où l'on étudie les distances entre individus décrits par des **variables quantitatives réelles**.

On peut également définir une distance discrète sur un ensemble quelconque E par

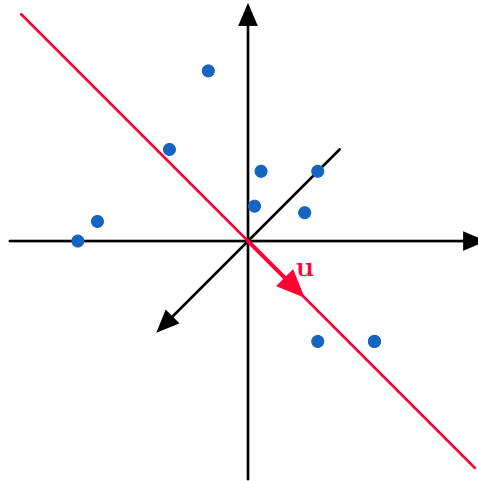
$$\forall \mathbf{x}, \mathbf{x}' \in E, \quad d(\mathbf{x}, \mathbf{x}') = \begin{cases} 0 & \text{si } \mathbf{x} = \mathbf{x}', \\ 1 & \text{si } \mathbf{x} \neq \mathbf{x}'. \end{cases}$$

Cela peut servir par exemple à étudier la distance entre individus décrits par des **variables qualitatives**.

Un autre cas intéressant que l'on peut aussi rencontrer est la notion de distance d'édition aussi appelée distance de Levenshtein qui est une distance définie entre des chaînes de caractères. Ce qui est très intéressant lorsque l'on souhaite étudier la distance entre deux mots ou des textes par exemple.

Cette notion de distance est importante car elle est à la base des méthodes d'analyse de données (ou plus généralement de réduction de la dimension, mais nous ferons une parenthèse à ce sujet en conclusion).

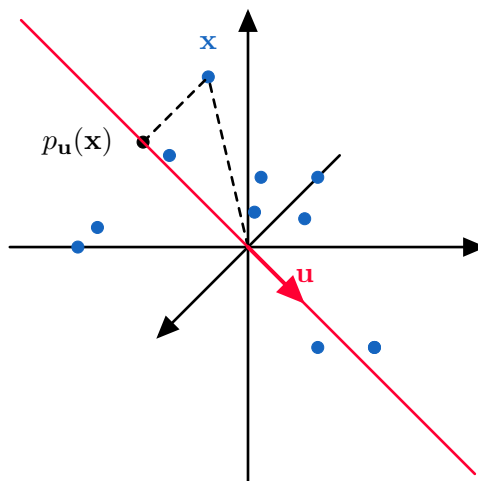
Un des premier objectif des méthodes que nous étudierons est de pouvoir fournir une représentation fidèle des données dans un espace de dimension inférieure.



Dans l'exemple ci-dessous, on va chercher à projeter les données de notre espace de dimension 3 sur une droite vectorielle de façon à conserver une grande partie de l'information initialement présente. On va donc chercher la droite \mathbf{u} qui maximise la variance de cette nouvelle représentation. Cela conduit à résoudre un problème d'optimisation suivant :

$$\min_{\mathbf{u} \in \mathbb{R}^p} \sum_{i=1}^n \|\mathbf{x}_i - p_{\mathbf{u}}(\mathbf{x}_i)\|^2,$$

où $p_{\mathbf{u}}(\mathbf{x})$ désigne le projeté orthogonal de \mathbf{x} sur la droite vectorielle engendrée par \mathbf{u} .



Remarque. En pratique, nous ne chercherons que très rarement à résoudre ce type de problème analytiquement, nous utiliserons des logiciels qui nous permettront de résoudre ce type de problème, c'est ce que nous verrons lors des séances pratiques. Cela ne nous empêche cependant pas d'étudier ce type de problèmes.

Le problème d'optimisation précédent est aussi appelé problème des *moindres carrés ordinaire (MCO)* car il consiste à minimiser la somme de carrés.

On se rappelle que \mathbf{x} est un vecteur et $p_{\mathbf{u}}(\mathbf{x})$ est son projeté, alors le vecteur $\mathbf{x} - p_{\mathbf{u}}(\mathbf{x})$ est orthogonal à \mathbf{u} et il représente un terme d'erreur due à la projection. Ainsi le problème des MCO est à voir comme un problème de projection qui cherche à minimiser les erreurs dues à la projection, *i.e.* on cherche le vecteur qui déforme le moins l'espace !.

En utilisant les propriétés d'orthogonalité entre les différents vecteurs et le théorème de Pythagore, nous avons la relation suivante

$$\|\mathbf{x} - p_{\mathbf{u}}(\mathbf{x})\|^2 = \|\mathbf{x}\|^2 - \|p_{\mathbf{u}}(\mathbf{x})\|^2.$$

Ainsi, notre problème de minimisation initial est équivalent à

$$\max_{\mathbf{u} \in \mathbb{R}^p} \sum_{i=1}^n \|p_{\mathbf{u}}(\mathbf{x}_i)\|^2,$$

car $\|\mathbf{x}\|^2$ est indépendant de \mathbf{u} . Ainsi, notre problème d'optimisation peut aussi se voir comme la recherche d'un axe selon lequel la variance des données projetées est maximisée ! C'est une notion fondamentale que l'on va retrouver lorsque l'on parlera de l'Analyse en Composantes Principales.

7.2 Vers la recherche de valeurs propres et vecteurs propres

On va maintenant se concentrer sur la deuxième formulation de notre problème d'optimisation que nous allons décortiquer pour essayer de faire le lien avec la recherche de valeurs propres.

On se rappelle que le projeté $p_{\mathbf{u}}(\mathbf{x})$ d'un vecteur \mathbf{x} sur la droite vectorielle engendrée par \mathbf{u} est donnée par

$$p_{\mathbf{u}}(\mathbf{x}) = \frac{\langle \mathbf{x}, \mathbf{u} \rangle}{\|\mathbf{u}\|} \mathbf{u}.$$

Dans la suite on supposera, sans perte de généralité, que le vecteur \mathbf{u} est un vecteur de norme égale à 1. Nous aurons alors : $p_{\mathbf{u}}(\mathbf{x}) = \langle \mathbf{x}, \mathbf{u} \rangle \mathbf{u}$. Dans ce cas

$$\begin{aligned} \|p_{\mathbf{u}}(\mathbf{x})\|^2 &= \langle p_{\mathbf{u}}(\mathbf{x}), p_{\mathbf{u}}(\mathbf{x}) \rangle, \\ &= \langle \langle \mathbf{x}, \mathbf{u} \rangle \mathbf{u}, \langle \mathbf{x}, \mathbf{u} \rangle \mathbf{u} \rangle, \\ &\quad \downarrow \text{linéarité du produit scalaire} \\ &= \langle \mathbf{x}, \mathbf{u} \rangle^2 \langle \mathbf{u}, \mathbf{u} \rangle, \\ &\quad \downarrow \text{car } \mathbf{u} \text{ est un vecteur unitaire} \\ &= \langle \mathbf{x}, \mathbf{u} \rangle^2, \\ &\quad \downarrow \text{simple réécriture} \\ &= (\mathbf{x}^T \mathbf{u})^T (\mathbf{x}^T \mathbf{u}), \\ &= \mathbf{u}^T \mathbf{x} \mathbf{x}^T \mathbf{u}. \end{aligned}$$

Notons maintenant que $X = (\mathbf{x}_1^T, \dots, \mathbf{x}_n^T)$ est une matrice à n lignes et p colonnes dont les lignes sont formées par les \mathbf{x}_i . On peut alors écrire :

$$\sum_{i=1}^n \|p_{\mathbf{u}}(\mathbf{x}_i)\|^2 = \sum_{i=1}^n \mathbf{u}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{u} = \sum_{i=1}^n \mathbf{u}^T (\mathbf{x}_i \mathbf{x}_i^T) \mathbf{u} = \mathbf{u}^T (X^T X) \mathbf{u}.$$

Dans cette relation, la matrice $X^T X$ est une matrice de $\mathcal{M}_p(\mathbb{R})$. Nous avons alors l'équivalence suivante :

$$\max_{\mathbf{u} \in \mathbb{R}^p} \sum_{i=1}^n \|p_{\mathbf{u}}(\mathbf{x}_i)\|^2 = \max_{\mathbf{u} \in \mathbb{R}^p, \|\mathbf{u}\|^2=1} \mathbf{u}^T X^T X \mathbf{u}.$$

On commence à voir des matrices qui apparaissent, ce qui est un premier élément de la recherche de valeurs propres et de vecteurs propres. Mais il n'est pas encore clair

de savoir qu'elle est la signification de ces éléments dans ce contexte, d'autant plus que la matrice étudiée ici est $X^T X$, qui est une matrice **des produits scalaires entre les variables !**

Avant de répondre à cette question, on va déjà étudier cette fameuse matrice $X^T X \in \mathcal{M}_p(\mathbb{R})$ qui est **la matrice des produits scalaires entre les variables**.

Définition 7.2: Matrice de Gram

Soit E un espace euclidien de dimension p et $\mathbf{x}_1, \dots, \mathbf{x}_n$ des vecteurs de E . On appelle **matrice de Gram**, notée G , la matrice carrée des produits scalaires scalaires entre les individus, dont la matrice d'ordre n telle que :

$$G_{ij} = \langle \mathbf{x}_i, \mathbf{x}_j \rangle, \forall i, j = 1, \dots, n.$$

Nous pourrions définir une matrice similaire pour définir le produit scalaire entre les variables.
Il s'agit également d'une matrice de Gram.

Ces deux matrices sont définies par des produits scalaires qui a l'avantage d'être symétrique ! Ce qui signifie que cette matrice G (ou son homologue dans l'espace dual) sont diagonalisables (même orthogonalement semblable à une matrice diagonale, voir Proposition 6.8).

Remarque : bien que sa définition reste simple, cette matrice est très importante et vous la retrouverez notamment lors de vos (futurs) cours d'Apprentissage Machine.

Mais sa définition permet d'en dire un peu plus sur ses valeurs propres.

Proposition 7.1: Valeurs propres de la matrice de Gram

Soit E un espace euclidien de dimension p et $\mathbf{x}_1, \dots, \mathbf{x}_n$ des vecteurs de E et considérons la matrice G carrée d'ordre n définie par

$$G_{ij} = \langle \mathbf{x}_i, \mathbf{x}_j \rangle, \forall i, j = 1, \dots, n.$$

Alors G est symétrique et semi-définie positive, *i.e.* elle admet n valeurs propres positives ou nulles.

Démonstration. Rappelons que si \mathbf{u} est un vecteur propre (non nul) G associée à la valeur propre λ , alors

$$G\mathbf{u} = \lambda\mathbf{u}.$$

Or $G = XX^T$, donc $G\mathbf{u} = XX^T\mathbf{u} = \lambda\mathbf{u}$. Si on pré-multiplie chaque membre de l'égalité par \mathbf{u}^T , nous avons

$$\begin{aligned}\mathbf{u}^T XX^T\mathbf{u} &= \lambda\mathbf{u}^T\mathbf{u}, \\ &\downarrow \text{définition transposée et norme} \\ (X^T\mathbf{u})^T (X^T\mathbf{u}) &= \lambda\|\mathbf{u}\|^2, \\ &\downarrow \text{définition de norme} \\ \|X^T\mathbf{u}\|^2 &= \lambda\|\mathbf{u}\|^2, \\ &\downarrow \mathbf{u} \text{ est un vecteur non nul} \\ \lambda &= \frac{\|X^T\mathbf{u}\|^2}{\|\mathbf{u}\|^2} \geq 0.\end{aligned}$$

On peut rédiger une démonstration analogue dans le cas où $G = X^T X$. Ce qu'on laisse à titre d'exercice. \square

Finalement, en posant $G' = X^T X$ (qui possède les mêmes propriétés que G mais dans un espace différent), notre problème d'optimisation peut s'écrire sous la forme :

$$\max_{\mathbf{u} \in \mathbb{R}^p, \|\mathbf{u}\|^2=1} \mathbf{u}^T G' \mathbf{u}.$$

Cette dernière expression doit normalement vous faire penser à quelque chose que nous avons également vu plus tôt ... en effet, cela ressemble à une forme quadratique ! Finalement, ce problème d'optimisation ne consiste en rien d'autre qu'en la recherche d'un vecteur propre associé à la plus grande valeur propre de la forme quadratique étudiée !

On fait une petite parenthèse ci-dessous pour justifier cela.

Remarque. Ces problèmes d'optimisation, sont appelés des *problèmes d'optimisation sous contrainte(s)*, ici la contrainte est la recherche d'un vecteur propre unitaire \mathbf{u} : $\|\mathbf{u}\|^2 = 1$.

Pour résoudre ce type de problème, on fait intervenir une nouvelle quantité que l'on appelle le **lagrangien** L ainsi qu'une nouvelle variable, disons α que l'on appelle **variable lagrangienne associée à la contrainte**. La fonction qui nous intéresse est alors définie par :

$$L(\mathbf{u}, \alpha) = \mathbf{u}^T G' \mathbf{u} - \alpha(\mathbf{u}^T \mathbf{u} - 1).$$

Le premier terme de cette fonction correspond à la quantité à optimiser initialement, le deuxième terme est la contrainte qui est cette fois-ci **retranchée**.

Notre objectif est de trouver les extrema de cette fonction, ces derniers sont données en cherchant les valeurs pour lesquels les dérivées par rapport à \mathbf{u} et α **s'annulent simultanément**, *i.e.* on cherche à résoudre :

$$\nabla L(\mathbf{u}, \alpha) = \mathbf{0} = \begin{pmatrix} \frac{\partial L}{\partial \mathbf{u}}(\mathbf{u}, \alpha) \\ \frac{\partial L}{\partial \alpha}(\mathbf{u}, \alpha) \end{pmatrix}.$$

La dérivée partielle par rapport à la variable α est très peu intéressante et elle n'impose rien de plus que le fait que la contrainte soit vérifiée. En revanche, si on dérive par rapport au vecteur \mathbf{u} , nous avons

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{u}}(\mathbf{u}, \alpha) = \mathbf{0} &\iff 2G' \mathbf{u} - 2\alpha \mathbf{u} = \mathbf{0}, \\ &\iff G' \mathbf{u} = \alpha \mathbf{u}, \\ &\quad \downarrow \text{définition de } G' \\ &\iff X^T X \mathbf{u} = \alpha \mathbf{u}. \end{aligned}$$

On retrouve donc bien un problème aux valeurs propres pour la matrice $G' = X^T X$.

Le vecteur \mathbf{u} , qui est le vecteur propre associé à la plus grande valeur propre de G' est donc celui qui permet d'avoir la meilleure représentation possible dans un espace de dimension 1.

Généralisation. Mais pourquoi se limiter à une projection sur un espace de dimension 1 ? En effet, on se doute bien que l'on risque de perdre énormément d'informations en projetant nos données d'un espace de dimension quelconque dans un espace de dimension 1.

On se doute bien que la représentation sera plus fidèle si on augmente la dimension de l'espace de projection. On va donc regarder comment généraliser cela.

Supposons que l'on souhaite maintenant projeter les données sur un espace de dimension $1 < p' < p$. Cela se fera toujours en étudiant les valeurs propres et vecteurs

propres de la matrice G' . Vu que cette matrice est symétrique, on rappelle que ses vecteurs propres forment une base orthonormée de l'espace de départ. Ainsi, si l'on cherche la représentation dans l'espace de dimension p' qui maximise la variance, il suffit de déterminer les p' vecteurs propres associés aux p' ($\mathbf{u}_1, \dots, \mathbf{u}_{p'}$) plus grandes valeurs propres ($\lambda_1, \dots, \lambda_{p'}$) associées à la matrice G' .

Les coordonnées d'une donnée \mathbf{x}_i sur la droite vectorielle engendrée par \mathbf{u}_s est donnée par :

$$\langle \mathbf{x}_i, \mathbf{u}_s \rangle = \mathbf{x}_i^T \mathbf{u}_s.$$

Pour l'ensemble des vecteurs \mathbf{x}_i , les coordonnées sur la droite vectorielle engendrée par \mathbf{u}_s sont données par le vecteur :

$$\langle X^T, \mathbf{u}_s \rangle = X \mathbf{u}_s.$$

On peut déterminer les coordonnées de cette façon pour l'ensemble des vecteurs \mathbf{x}_i dans la base des vecteurs propres \mathbf{u}_s . Ces coordonnées sont données par la matrice

$$XU_{p'},$$

où $U_{p'}$ est une matrice de dimension $p \times p'$.

Projection des variables dans un espace dimension réduit Pour l'instant nous avons simplement regarder comment projeter les données dans un espace de dimension inférieure, mais nous avons dit que nous pouvions faire de même pour les variables ! (Nous en avons parlé lorsqu'il a été fait mention de produit scalaire entre les variables). C'est le moment de montrer la dualité qui existe entre la représentation dans l'espace des variables et l'espace des individus ainsi que les liens entre les matrices G et G' .

Regardons déjà ce qui se passe si l'on cherche à réduire l'espace de représentation des variables en passant d'un espace de dimension n à un espace de dimension $n' < n$. On peut reprendre tout ce que l'on a vu précédemment et l'appliquer à la matrice $G = XX^T$ (qui donne donc la matrice des produits scalaires entre les variables) et noter $\mathbf{u}'_1, \dots, \mathbf{u}'_n$ les vecteurs propres de cette matrice.

On peut donc représenter les vecteurs des variables (nous les avons appelés \mathbf{v} en introduction) dans le sous-espace de dimension n' formé par les vecteurs propres $\mathbf{u}'_1, \dots, \mathbf{u}'_n$, à l'aide des coordonnées données par la matrice

$$X^T U'_{n'}.$$

Dualité entre les deux représentations Nous venons de voir que chercher à obtenir une représentation d'un ensemble d'individus revient à diagonaliser la matrice $X^T X$ de $\mathcal{M}_p(\mathbb{R})$. De même, si on souhaite projeter les variables, représentés dans l'espace des individus, dans un espace de dimension inférieure, nous devons diagonaliser la matrice XX^T de $\mathcal{M}_n(\mathbb{R})$.

Il y a cependant un fait plutôt marquant entre ces deux matrices ... toutes les valeurs propres non nulles sont égales ! En effet, notons λ_k la k -ème plus grande valeur propre de la matrice XX^T et \mathbf{u}'_k le vecteur propre associé. Par définition, nous avons alors :

$$XX^T \mathbf{u}'_k = \lambda_k \mathbf{u}'_k \quad \text{d'où} \quad X^T XX^T \mathbf{u}'_k = \lambda_k X^T \mathbf{u}'_k.$$

Ainsi si \mathbf{u}'_k est le vecteur propre de la matrice XX^T associé à la valeur propre λ_k , alors le vecteur $X^T \mathbf{u}'_k$ est un vecteur propre de la matrice $X^T X$ associé à la même valeur propre λ_k , or cette matrice $X^T X$ admet \mathbf{u}_k comme vecteur propre associé à la même valeur propre λ_k .

On en déduit les relations suivantes entre les deux vecteurs propres \mathbf{u}'_k et $X\mathbf{u}'_k = \mathbf{u}_k$:

$$\mathbf{u}'_k = \frac{X^T \mathbf{u}_k}{\|X^T \mathbf{u}_k\|} = \frac{1}{\sqrt{\lambda_k}} X^T \mathbf{u}'_k.$$

Nous avons également la relation inverse

$$\mathbf{u}_k = \frac{X \mathbf{u}'_k}{\|X \mathbf{u}'_k\|} = \frac{1}{\sqrt{\lambda_k}} X \mathbf{u}'_k.$$

La normalisation qui intervient dans les relations de dualité est due au fait que les vecteurs propres \mathbf{u}_k et \mathbf{u}'_k sont des vecteurs normés, *i.e.* $\|\mathbf{u}_k\| = \|\mathbf{u}'_k\| = 1$.

Pour l'instant nous sommes passés par le biais de matrices carrées annexes pour étudier les valeurs propres associées à une matrice de données. Mais comment étudier directement les "valeurs propres" associées à une matrice de données qui n'est pas nécessairement carrée ? C'est l'objectif de la prochaine section.

7.3 Décomposition en valeurs singulières

On commence par fournir le résultat important de cette section et nous passerons le reste de cette section à l'expliquer.

Proposition 7.2: Décomposition en valeurs singulières

Soit X une matrice réelle d'ordre $n \times p$ et de rang $r \leq \min(n, p)$. Alors la matrice X peut être factorisée de la façon suivante

$$X = U \Sigma U'^T,$$

où $U \in \mathcal{M}_n(\mathbb{R})$ et $U' \in \mathcal{M}_p(\mathbb{R})$ sont des matrices orthogonales et $\Sigma \in \mathcal{M}_{n,p}(\mathbb{R})$ est une matrice remplie de 0 sauf sur la diagonale principale où, pour tout $i = 1, \dots, r$, on a :

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

où les σ_i sont les valeurs propres non nulles indifféremment de la matrice XX^T ou X^TX . On supposera de plus que l'on a $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$.

Dans cette proposition :

- les valeurs σ_i sont appelées *valeurs singulières* de la matrice X .
- la matrice U est composée des *vecteurs singuliers* à gauche de la matrice X . Ils correspondent aux vecteurs propres de la matrice XX^T .
- la matrice U' est composée des *vecteurs singuliers* à droite de la matrice X . Ils correspondent aux vecteurs propres de la matrice X^TX .

De plus, les colonnes des matrices U et U' forment une base orthonormée des espaces \mathbb{R}^n et \mathbb{R}^p respectivement.

En effet, nous avons vu que toute matrice symétrique est orthogonalement semblable à une matrice diagonale et que les vecteurs propres forment une base de l'espace de départ. D'où

$$U^T U = I_n \quad \text{et} \quad U'^T U' = I_p.$$

Analyse de la proposition. Après avoir présenté les différents éléments de ce résultat, essayons de le justifier les définitions des valeurs singulières et des vecteurs singuliers.

Pour cela remarquons que si $X = U \Sigma U'^T$ alors $X^T = U' \Sigma^T U^T$ ce qui implique :

$$\begin{aligned} X^T X &= X^T U \Sigma U'^T, \\ &\quad \downarrow \text{définition de } X^T \\ &= U' \Sigma^T U^T U \Sigma U'^T, \\ &\quad \downarrow \text{orthogonalité de } U \\ &= U' \Sigma^T \Sigma U'^T, \\ &= U' \Sigma_p^2 U'^T, \end{aligned}$$

où $\Sigma^T \Sigma = \Sigma_p^2 \in \mathcal{M}_p(\mathbb{R})$. De la même façon, nous avons :

$$\begin{aligned}
 XX^T &= XU'\Sigma^T U^T, \\
 &\quad \downarrow \text{définition de } X \\
 &= U\Sigma U'^T U'\Sigma^T U^T, \\
 &\quad \downarrow \text{orthogonalité de } U' \\
 &= U\Sigma\Sigma^T U^T, \\
 &= U\Sigma_n^2 U^T,
 \end{aligned}$$

où $\Sigma\Sigma^T = \Sigma_n^2 \in \mathcal{M}_n(\mathbb{R})$. On rappelle que les matrices Σ_n^2 et Σ_p^2 sont des matrices diagonales dont tous les éléments diagonaux non nuls sont égaux.

Cette première observation permet de montrer que les *valeurs singulières* de la matrice X ne sont rien d'autres que la racine carrée des valeurs propres de Σ_n^2 (ou indifféremment de Σ_p^2).

De plus le premier développement permet bien de montrer que les vecteurs singuliers à droite sont bien les vecteurs propres de la matrice $X^T X$. De la même façon, les vecteurs singuliers à gauche sont bien les vecteurs propres de la matrice XX^T .

Enfin nous pouvons, comme dans la section précédente, exhiber une relation entre les vecteurs singuliers à gauche et à droite de la matrice X en fonction de la matrice elle-même et des valeurs singulières σ_k de cette dernière.

Les relations $XX^T \mathbf{u}_k = \lambda_k \mathbf{u}_k = \sigma_k^2 \mathbf{u}_k$ et $X^T X \mathbf{u}'_k = \lambda_k \mathbf{u}'_k = \sigma_k^2 \mathbf{u}'_k$ permettent d'écrire

$$\frac{X^T \mathbf{u}_k}{\sigma_k} = \mathbf{u}'_k \quad \text{et} \quad \frac{X \mathbf{u}'_k}{\sigma_k} = \mathbf{u}_k.$$

On montre cela de la même façon que précédemment (voir section précédente).

Qualité de l'approximation et illustration Pour analyser les données, nous avons dit qu'il est parfois intéressant de projeter ces dernières sur des sous-espaces de dimensions inférieures (à des fins de visualisation par exemple). Nous avons également énoncé le fait que ces projections ne sont pas sans conséquences sur la quantité d'information initialement présente dans les données. En effet, il paraît légitime de penser que de chercher à synthétiser l'information implique une perte d'information.

La question est maintenant de savoir comment nous pouvons quantifier cette perte en fonction de la projection effectuée et des données.

Pour cela considérons notre matrice $X \in \mathcal{M}_{n,p}(\mathbb{R})$, où $\mathcal{M}_{n,p}(\mathbb{R})$ est l'espace vectoriel des matrices muni du produit scalaire de Frobenius $\langle \cdot, \cdot \rangle_F$ et de la norme induite $\|\cdot\|_F$ qui nous servira à définir la distance entre deux matrices. Pour rappel, pour deux matrices A et B de $\mathcal{M}_{n,p}(\mathbb{R})$ nous avons :

$$\langle A, B \rangle = \text{Tr}(A^T B) = \sum_{i=1}^n \sum_{j=1}^p a_{ij} b_{ij} \quad \text{et} \quad \|A - B\|_F^2 = \sum_{i=1}^n \sum_{j=1}^p (a_{ij} - b_{ij})^2.$$

La norme de Frobenius de la différence de deux matrices peut ainsi se voir comme la distance euclidienne entre deux matrices.

On souhaite maintenant approximer, au sens de la norme de Frobenius, une matrice X par une matrice \tilde{X} telle que \tilde{X} soit de rang inférieur à un rang donné s . On souhaite donc résoudre le problème suivant :

$$\min_{\tilde{X} \in \mathcal{M}_{n,p}(\mathbb{R}), \text{rg}(\tilde{X}) \leq s} \|X - \tilde{X}\|_F^2.$$

La solution à ce problème d'optimisation est donnée par le *Théorème d'Eckart-Young*.

Théorème 7.1: Théorème d'Eckart-Young

Soit X une matrice de $\mathcal{M}_{n,p}(\mathbb{R})$ et considérons le problème d'optimisation suivant :

$$\min_{\tilde{X} \in \mathcal{M}_{n,p}(\mathbb{R}), \text{rg}(\tilde{X}) \leq s} \|X - \tilde{X}\|_F^2.$$

Alors la solution de ce problème d'optimisation est donnée par la décomposition en valeurs singulières de la matrice X .

On ne démontre pas ce résultat car la démonstration est technique et elle n'apporte rien dans le cadre de ce cours.

Ce résultat reste valable lorsque l'on remplace la norme de Frobenius par la *norme spectrale* aussi appelée norme subordonnée à la norme euclidienne.

Quelques remarques à propos de ce théorème :

- si on cherche une approximation de rang s et que la matrice de design (autre nom donnée à la matrice des données) X est de rang $< s$, on a alors $X = \tilde{X}$.
- si la matrice de design est de rang $r > s$, alors elle possède r valeurs singulières ainsi que r vecteurs singuliers à gauche $(\mathbf{u}_1, \dots, \mathbf{u}_r)$ et à droite $(\mathbf{u}'_1, \dots, \mathbf{u}'_r)$

qui sont normés.

Ainsi la meilleure approximation de rang s est donnée par

$$\tilde{X} = \sigma_1 \mathbf{u}_1 (\mathbf{u}'_1)^T + \cdots + \sigma_s \mathbf{u}_s (\mathbf{u}'_s)^T,$$

où les valeurs singulières sont rangées par ordre décroissant : $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_s$.
Chaque élément de cette somme représente des matrices de rang 1 mais comme pour tout $i \neq j$, $\mathbf{u}_i \perp \mathbf{u}_j$ et $\mathbf{u}'_i \perp \mathbf{u}'_j$, alors la somme des matrices $\mathbf{u}_i (\mathbf{u}'_i)^T$ avec $\mathbf{u}_j (\mathbf{u}'_j)^T$ donnent bien une matrice de rang 2.

Ainsi notre approximation \tilde{X} peut se représenter comme

$$\tilde{X} = \sigma_1 \begin{matrix} \mathbf{u}'_1 \\ \mathbf{u}_1 \end{matrix} + \sigma_2 \begin{matrix} \mathbf{u}'_2 \\ \mathbf{u}_2 \end{matrix} + \dots + \sigma_s \begin{matrix} \mathbf{u}'_s \\ \mathbf{u}_s \end{matrix}$$

- enfin, si l'on souhaite mesurer la qualité τ de l'information, on peut regarder un premier indicateur naïf qu'est le quotient de la somme des valeurs singulières associées à l'approximation sur la somme des valeurs singulières de la matrice X :

$$\tau = \frac{\sum_{k=1}^s \sigma_k(X)}{\sum_{k=1}^r \sigma_k(X)}.$$

Plus la valeur de s est grande, meilleure sera l'approximation.

Cette valeur est un bon indicateur pour savoir si l'information contenue dans la matrice de design peut être synthétisée dans un espace de dimension faible.

En revanche, ce critère ne repose pas sur la même norme que celle employée dans le problème d'optimisation, il faudrait donc introduire un moyen de mesurer qualité de l'approximation qui se fonde sur la norme de Frobenius. Pour cela on mesure plutôt :

$$\tau = \frac{\|\tilde{X}\|_F^2}{\|X\|_F^2},$$

où la norme de Frobenius d'une matrice X est donnée par

$$\|X\|_F^2 = \sum_{i=1}^n \sum_{j=1}^p x_{ij}^2.$$

Mais plutôt que de calculer cette somme là, on va voir que l'on peut à nouveau faire intervenir les valeurs singulières de la matrice X (ou les valeurs propres de la matrice XX^T !). Pour cela, on se rappelle que si X est une matrice de rang r , alors $X = \sum_{k=1}^r \sigma_k \mathbf{u}_k (\mathbf{u}'_k)^T$

$$\begin{aligned} \|X\|_F^2 &= \sum_{i=1}^n \sum_{j=1}^p x_{ij}^2, \\ &\downarrow \text{ en utilisant la SVD} \\ &= \sum_{k=1}^r \sigma_k^2 \sum_{i=1}^n \sum_{j=1}^p (\mathbf{u}_k)_i^2 (\mathbf{u}'_k)_j^2, \\ &\downarrow \text{ on arrange les termes} \\ &= \sum_{k=1}^r \sigma_k^2 \left(\sum_{i=1}^n (\mathbf{u}_k)_i^2 \right) \left(\sum_{j=1}^p (\mathbf{u}'_k)_j^2 \right), \\ &\downarrow \text{ les vecteurs } \mathbf{u}_k \text{ et } \mathbf{u}'_k \text{ sont normés} \\ &= \sum_{k=1}^r \sigma_k^2. \end{aligned}$$

Ainsi

$$\tau = \frac{\sum_{k=1}^s \sigma_k^2(X)}{\sum_{k=1}^r \sigma_k^2(X)} = \frac{\sum_{k=1}^s \lambda_k(X)}{\sum_{k=1}^r \lambda_k(X)}.$$

Mettons cela en oeuvre sur un exemple.

Exemple 7.2. Soit $X = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ une matrice de $\mathcal{M}_{3,2}(\mathbb{R})$. On va essayer de déterminer la décomposition en valeurs singulières de cette matrice, ses vecteurs singuliers contenus dans les matrices U et U' ainsi que la qualité d'une approximation de rang 1.

On commence par déterminer les valeurs propres de la matrice $X^T X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ en déterminant les racines du polynôme caractéristique $\mathcal{X}_{X^T X}(\lambda)$ défini par :

$$\mathcal{X}_{X^T X}(\lambda) = \det(X^T X - \lambda I_3),$$

$$= \begin{vmatrix} 1-\lambda & 1 & 0 \\ 1 & 2-\lambda & 1 \\ 0 & 1 & 1-\lambda \end{vmatrix},$$

↓ on développe selon la ligne

$$(1-\lambda) \begin{vmatrix} 2-\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} - \begin{vmatrix} 1 & 1 \\ 0 & 1-\lambda \end{vmatrix},$$

↓ on développe les deux déterminants d'ordre 2 et on factorise
 $-\lambda(\lambda-3)(\lambda-1).$

Ainsi les racines de $\mathcal{X}_{X^T X}(\lambda)$ et donc les valeurs propres de $X^T X$ sont 3, 1 et 0. Les valeurs singulières de X sont donc $\sigma_1 = \sqrt{3}$ et $\sigma_2 = 1$.

On cherche maintenant les vecteurs propres associés à chaque valeur propre de la matrice $X^T X$. Ces vecteurs propres vont définir les vecteurs singuliers à droite de notre matrice X (i.e. ils vont définir la matrice U' .)

- **Vecteur propre associé à $\lambda = 3$** : on a $X^T X - 3I_3 = \begin{pmatrix} -2 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & -2 \end{pmatrix}$ dont

la forme échelonnée réduite est donnée par la matrice $\begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$. Un élément du noyau de cette matrice est donnée par le vecteur $(1, 2, 1)$, donc une base orthonormée de ce sous-espace propre est donné par le vecteur $\mathbf{u}'_1 = \frac{1}{\sqrt{6}}(1, 2, 1)$.

On en déduit de suite $\mathbf{u}_1 = \frac{X\mathbf{u}'_1}{\sigma_1} = \frac{1}{\sqrt{2}}(1, 1)$.

- **Vecteur propre associé à $\lambda = 1$** : on effectue le même processus que précédemment. On a $X^T X - I_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ dont la forme échelonnée réduite est

donnée par la matrice $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Un élément du noyau de cette matrice est donnée par le vecteur $(1, 0, -1)$, donc une base orthonormée de ce sous-espace propre est donné par le vecteur $\mathbf{u}'_2 = \frac{1}{\sqrt{2}}(1, 0, -1)$.

On en déduit de suite $\mathbf{u}_2 = \frac{X\mathbf{u}'_2}{\sigma_2} = \frac{1}{\sqrt{2}}(1, -1)$.

- **Vecteur propre associé à $\lambda = 0$** : On a $X^T X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ dont la forme

échelonnée réduite est donnée par la matrice $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Un élément du noyau de cette matrice est donné par le vecteur $(1, -1, 1)$, donc une base orthonormée de ce sous-espace propre est donné par le vecteur $\mathbf{u}'_3 = \frac{1}{\sqrt{3}}(1, -1, 1)$.

Finalement, en posant $U = (\mathbf{u}_1, \mathbf{u}_2)$, $\Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ et $U' = (\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3)$, on obtient la décomposition en valeurs singulières de la matrice X .

Regardons maintenant la qualité de l'approximation de rang 1. La matrice \tilde{X} associée est définie par :

$$\tilde{X} = \sigma_1 \mathbf{u}_1 (\mathbf{u}'_1)^T = \frac{\sqrt{3}}{2} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix}.$$

La qualité de cette approximation de rang 1 est donnée par

$$\tau = \frac{\|\tilde{X}\|_F^2}{\|X\|_F^2} = \frac{3}{3+1} = \frac{3}{4}.$$

Ainsi, l'approximation de rang 1 permet de restituer 75% de l'information initialement présente.

Maintenant que nous avons posé les bases de la réduction de dimension, nous allons pouvoir en regarder quelques techniques. Une première que l'on utilisera pour analyser des données quantitatives réelles, à savoir l'*Analyse en Composantes Principales* puis une deuxième que l'on utilisera pour analyser des données qualitatives, à savoir l'*Analyse Factorielle des Correspondances*.

8 Analyse en Composantes Principales (ACP)

8.1 Principe

L'analyse en composantes principale est une méthode réduction de la dimension qui s'emploie lorsque les variables étudiées sont **quantitatifs réelles**, *i.e.* les variables prennent des valeurs réelles.

A nouveau, on supposera que l'on disposera d'un ensemble de n exemples, *i.e.* n vecteurs décrits par p attributs, on a donc $\mathbf{x}_i \in \mathbb{R}^p$. Ainsi chaque \mathbf{x}_i est un vecteur *individu*.

Ces descripteurs sont notés \mathbf{v}_k et prennent autant de valeurs qu'il y a des exemples dans notre échantillon d'étude, *i.e.* $\mathbf{v}_k \in \mathbb{R}^n$. Un tel vecteur est appelé vecteur *variable*.

Toutes ces données sont synthétisées dans la matrice $X \in \mathcal{M}_{np}(\mathbb{R})$ suivante

$$X = \begin{matrix} & \mathbf{v}_1 & \cdots & \mathbf{v}_k & \cdots & \mathbf{v}_p \\ \mathbf{x}_1 & x_{11} & \cdots & x_{1k} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathbf{x}_i & x_{i1} & \cdots & x_{ik} & \cdots & x_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathbf{x}_n & x_{n1} & \cdots & x_{nk} & \cdots & x_{np} \end{matrix},$$

Comme nous l'avons mentionné dans la section précédente, la mesure que nous utiliserons pour mesurer la distance entre les individus est la distance euclidienne :

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^p ((\mathbf{x}_i)_k - (\mathbf{x}_j)_k)^2.$$

Cette distance sera donc la base de l'étude du nuage des individus.

De la même façon nous pouvons nous intéresser au nuage des variables. En revanche, nous utiliserons des grandeurs différentes pour décrire et étudier ce nuage. Premièrement, nous n'utiliserons pas de notion de distance entre les variables mais nous allons plutôt chercher à mesurer la **liaison** entre ces dernières en étudiant le **coefficient de corrélation** noté $r(\mathbf{v}_k, \mathbf{v}_l)$. Pour cela, nous aurons besoin d'étudier les grandeurs suivantes :

La moyenne m_k d'une variable \mathbf{v}_k

$$m_k = \frac{1}{n} \sum_{i=1}^n (\mathbf{v}_k)_i,$$

ainsi que la variance s_k^2 de ces mêmes variables

$$s_k^2 = \frac{1}{n} \sum_{l=1}^n ((\mathbf{v}_k)_l - m_k)^2.$$

Ces quantités nous serviront à étudier notre matrice de design X avec notre ACP.

Objectif de l'ACP : il s'agit d'une technique de réduction de dimension qui va nous permettre de projeter les données sur des sous-espaces afin de synthétiser l'information. Pour faire cela, l'ACP va transformer des variables **corrélées** en un nouvel ensemble de variables **décorrélées** qui vont se présenter comme une combinaison linéaire des anciennes variables. Ces nouvelles variables seront appelées **axes principaux**. Ces axes correspondent à des directions de l'espace selon lesquelles la variance est maximale.

L'ACP combine donc des outils à la fois statistiques (moyenne, variance, corrélation) avec des outils de géométries et d'algèbre linéaire pour obtenir une nouvelle représentation des données dans laquelle l'information est moins redondante.

C'est une technique très utile lorsque le jeu de données présente un très grand nombre de variables et où l'on est presque sûr que la plupart d'entre elles sont corrélées. Elle permet de réduire les ressources de calculs nécessaires après l'application d'une ACP.

En tant que technique cherchant à préserver au maximum la variance dans les données, l'ACP se présente comme un problème aux valeurs propres sur une certaine matrice de variance : *la matrice de variance-covariance* des données X . On pourra également (et c'est le choix que l'on fera par la suite) de travailler sur *la matrice de corrélation* des variables. Nous verrons d'ailleurs que cette matrice de corrélation va nous permettre de représenter graphiquement les liens entre les variables.

8.2 Transformations

Avant de s'attaquer à la recherche des axes principaux, nous devons déjà effectuer quelques transformations sur notre échantillon.

Transformation du jeu de données. La première étape consiste à **centrer** notre jeu de données, *i.e.* faire en sorte que les moyennes de chaque variables \mathbf{v}_k soient égales à 0. On obtient alors une nouvelle matrice X_{centre} définie par

$$X_{centre} = \begin{pmatrix} x_{11} - m_1 & \cdots & x_{1k} - m_k & \cdots & x_{1p} - m_p \\ \vdots & & \vdots & & \vdots \\ x_{i1} - m_1 & \cdots & x_{ik} - m_k & \cdots & x_{ip} - m_p \\ \vdots & & \vdots & & \vdots \\ x_{n1} - m_1 & \cdots & x_{nk} - m_k & \cdots & x_{np} - m_p \end{pmatrix} = X - \mathbf{1}\mathbf{m}^T,$$

où $\mathbf{m} = (m_1, \dots, m_p)$ est le vecteur des moyennes des variables. C'est aussi le barycentre du nuage de points et $\mathbf{1}$ est un vecteur colonne de taille n ne comprenant que des 1.

Cette première transformation sur la variance ou l'inertie présente initialement dans le nuage de points car on effectue simplement une transformation sur ce dernier, les distances sont invariantes par translation. On peut également vérifier que pour chaque colonne de cette nouvelle matrice, nous avons bien

$$\sum_{i=1}^n (x_{ik} - m_k) = 0, \quad \forall k \in \llbracket 1, p \rrbracket.$$

On peut ensuite réduire notre jeu de données en divisant chaque variable par son écart-type, c'est le choix que nous ferons ici mais il n'est pas obligatoire. On obtient alors une nouvelle matrice $X_{cen-red}$ définie par

$$X_{cen-red} = \begin{pmatrix} \frac{x_{11} - m_1}{s_1} & \cdots & \frac{x_{1k} - m_k}{s_k} & \cdots & \frac{x_{1p} - m_p}{s_p} \\ \vdots & & \vdots & & \vdots \\ \frac{x_{i1} - m_1}{s_1} & \cdots & \frac{x_{ik} - m_k}{s_k} & \cdots & \frac{x_{ip} - m_p}{s_p} \\ \vdots & & \vdots & & \vdots \\ \frac{x_{n1} - m_1}{s_1} & \cdots & \frac{x_{nk} - m_k}{s_k} & \cdots & \frac{x_{np} - m_p}{s_p} \end{pmatrix} = \text{diag} \left(\frac{1}{s_1}, \dots, \frac{1}{s_p} \right) (X - \mathbf{1}\mathbf{m}^T).$$

On vérifie facilement que, dans ce cas, toutes les variables ont une variance égale à 1

$$\frac{1}{n} \sum_{i=1}^n (x_{ik} - m_k)^2 = 1, \quad \forall k \in \llbracket 1, p \rrbracket.$$

Réduire ses données n'est pas sans conséquence sur l'ACP qui repose sur la variance des données. En effet en faisant ce choix :

- toutes les variables vont avoir la même variance (égale à 1) ce qui va éviter de tirer l'ACP vers les variables dont la variance est élevée simplement parce que les valeurs prises par cette dernière sont plus grandes.
- en revanche, si les données associées à une variable présentent un bruit important (mauvaise collecte des données, problème avec l'outil de mesure, ...) alors cette dernière aura une variance semblable à une variable qui serait elle plus informative.

Il est donc important de s'assurer que toutes les variables sont informatives avant de faire cela. Un argument plus physique peut motiver le choix d'une telle transformation, l'étude sera plus pertinente, par la suite, si on étudie des variables qui sont adimensionnées.

Enfin, cette transformation a une conséquence sur l'inertie totale du nuage du point car les distances sont modifiées cette fois-ci. En effet, lorsque les données sont standardisées (*i.e.* lorsque l'on divise les valeurs par l'écart-type) la distance entre deux points \mathbf{x}_i et \mathbf{x}_j devient

$$d^2(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^p \frac{1}{s_k^2} (x_{ik} - x_{jk})^2.$$

Remarque importante : il y a deux écoles pour réaliser l'ACP : l'ACP *classique* et une ACP dite *normée*. Cette deuxième version nécessite de diviser l'ensemble des colonnes de notre matrice $X_{cen-red}$ par \sqrt{n} .

On suppose maintenant que nos données sont centrées et réduites et les valeurs sont divisées par \sqrt{n} comme décrit précédemment et on cherche maintenant à projeter nos individus dans un espace de dimension plus faible.

Projection des individus. On rappelle que dans ce cas, on cherche à préserver les distances entre les individus, *i.e.* à préserver l'inertie totale. suite aux transformations effectuées précédemment, notamment l'opération de recentrage des données, l'inertie du nuage de points peut s'écrire

$$\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p \left(\frac{x_{ik} - m_k}{s_k} \right)^2.$$

L'inertie totale du nuage de points n'est donc rien d'autre que **la moyenne pondérée des nouveaux vecteurs des individus qui sont centrés et réduits**.

On va procéder de façon semblable à la SVD. Si on cherche à projeter les données sur un sous-espace de dimension s , on va commencer par chercher des directions $\mathbf{u}'_1, \dots, \mathbf{u}'_s$ sur lesquelles on va maximiser la variance (ou l'inertie) du nuage de points. Ces directions seront appelées **axes principaux** avec cette même convention que pour la SVD :

- l'axe défini par le vecteur \mathbf{u}'_1 est le sous-espace de dimension 1 qui maximise l'inertie du nuage du point après projection,
- l'axe défini par le vecteur \mathbf{u}'_2 , *orthogonal* à \mathbf{u}'_1 est le deuxième sous-espace de dimension 1 qui maximise l'inertie du nuage du point après projection,
- on continue avec \mathbf{u}'_3 qui est *orthogonal* à \mathbf{u}'_1 et \mathbf{u}'_2 , et ainsi de suite.

Mais comment obtenir ces vecteurs formellement ? Lorsque nous avons étudié la décomposition en valeurs singulières, nous nous sommes intéressés aux matrices $X^T X$ et XX^T dont nous avons cherché les valeurs propres et les vecteurs propres. Et bien nous allons faire la même chose ici, mais nous ne travaillerons pas directement sur la matrice de design X mais plutôt sur sa version centrée-réduite normée $X_{cen-red}$.

En effet, on rappelle que le but de l'ACP est aussi de décorréler les variables, il faut donc faire intervenir cette matrices de corrélation. Pour rappel, il s'agit de la matrice $C \in \mathcal{M}_p(\mathbb{R})$ définie par

$$C = \frac{1}{n} X_{cen-red}^T X_{cen-red},$$

où les termes c_{ij} sont données par $c_{ij} = \frac{1}{n} \sum_{l=1}^n \left(\frac{x_{li} - m_i}{s_i} \right) \left(\frac{x_{lj} - m_j}{s_j} \right)$ et sont tous compris dans l'intervalle $[-1, 1]$.

Dans la suite, nous noterons z_{ij} les éléments de la matrice $X_{cen-red}/\sqrt{n}$ qui correspondent aux données centrées réduites normées de la matrice de design X afin d'éviter toute confusion. Ainsi, on définit

$$Z = \begin{pmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_k \\ \vdots \\ \mathbf{z}_n \end{pmatrix} \begin{pmatrix} \frac{\mathbf{v}'_1}{s_1\sqrt{n}} & \dots & \frac{\mathbf{v}'_k}{s_k\sqrt{n}} & \dots & \frac{\mathbf{v}'_p}{s_p\sqrt{n}} \\ \frac{x_{11} - m_1}{s_1\sqrt{n}} & \dots & \frac{x_{1k} - m_k}{s_k\sqrt{n}} & \dots & \frac{x_{1p} - m_p}{s_p\sqrt{n}} \\ \vdots & & \vdots & & \vdots \\ \frac{x_{i1} - m_1}{s_1\sqrt{n}} & \dots & \frac{x_{ik} - m_k}{s_k\sqrt{n}} & \dots & \frac{x_{ip} - m_p}{s_p\sqrt{n}} \\ \vdots & & \vdots & & \vdots \\ \frac{x_{n1} - m_1}{s_1\sqrt{n}} & \dots & \frac{x_{nk} - m_k}{s_k\sqrt{n}} & \dots & \frac{x_{np} - m_p}{s_p\sqrt{n}} \end{pmatrix}.$$

Notons que cette nouvelle matrice Z est une matrice dont toutes les colonnes ont une norme égale à 1.

En définissant $C = Z^T Z$, on définit une matrice symétrique réelle, elle est donc orthogonalement semblable à une matrice diagonale, *i.e.* il existe donc une matrice orthogonale $U' \in \mathcal{M}_p(\mathbb{R})$ et une matrice diagonale $\Sigma_p = \text{diag}(\lambda_1, \dots, \lambda_p)$ telle que $\lambda_1 \geq \dots \geq \lambda_p$. On a

$$C = U' \Sigma_p U'^T \quad \text{et pour tout } m \leq p, \quad C \mathbf{u}'_m = \lambda_m \mathbf{u}'_m.$$

où U' est formée des vecteurs propres de C . Donc le vecteur propre \mathbf{u}'_m est associé à la valeur propre λ_m qui est la m -ème plus grande valeur propre.

Ainsi, les nouvelles coordonnées d'un individu \mathbf{x}_i sur un axe principal \mathbf{u}' sont données par :

$$p_{\mathbf{u}'}(\mathbf{z}_i) = \langle \mathbf{z}_i, \mathbf{u}' \rangle = \sum_{k=1}^p z_{ik} u'_k.$$

On peut montrer que le nuage projeté sur un axe principal \mathbf{u} est également centré, *i.e.*

$$\frac{1}{n} \sum_{i=1}^n p_{\mathbf{u}'}(\mathbf{z}_i) = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p z_{ik} u'_k = \sum_{k=1}^p u'_k \underbrace{\frac{1}{n} \sum_{i=1}^n z_{ik}}_{=0} = 0.$$

Enfin, on peut également calculer la variance associée au m -ème axe principal \mathbf{u}_m :

$$\frac{1}{n} \sum_{i=1}^n p_{\mathbf{u}'_m}(\mathbf{z}_i)^2 = \frac{1}{n} \sum_{i=1}^n \langle \mathbf{z}_i, \mathbf{u}'_m \rangle^2 = \frac{1}{n} \sum_{i=1}^n \mathbf{u}'_m{}^T \mathbf{z}_i \mathbf{z}_i^T \mathbf{u}'_m = \frac{1}{n} \mathbf{u}'_m{}^T C \mathbf{u}'_m = \frac{1}{n} \lambda_m.$$

Cette dernière égalité nous indique que la variance du projeté sur l'axe engendré par le vecteur \mathbf{u}'_m est égal, au facteur $1/n$ près, à la valeur propre associée au vecteur \mathbf{u}'_m .

Plus généralement, on obtient alors les composantes principales des n individus sur l'axe factoriel \mathbf{u}'_m en déterminant

$$\mathbf{p}_{\mathbf{u}'_m} = \mathbf{p}_{\mathbf{u}'_m}(Z) = Z \mathbf{u}'_m.$$

Projection des variables. On peut également procéder au même type d'analyse en travaillant sur le nuage des variables (un peu moins commun il faut bien l'avouer). En revanche, la distance utilisée sera un peu différente vu que l'on ne s'intéresse plus aux individus mais plutôt aux variables. On va continuer de travailler sur les variables centrée

réduites et normalisées définies précédemment.

La distance euclidienne employée fera alors intervenir le coefficient de corrélation :

$$d(\mathbf{v}'_k, \mathbf{v}'_l)^2 = \|\mathbf{v}'_k - \mathbf{v}'_l\|^2 = \langle \mathbf{v}'_k, \mathbf{v}'_k \rangle + \langle \mathbf{v}'_l, \mathbf{v}'_l \rangle - 2\langle \mathbf{v}'_k, \mathbf{v}'_l \rangle = 2(1 - c_{kl}),$$

$\langle \mathbf{v}'_j, \mathbf{v}'_j \rangle$ désigne la variance du vecteur \mathbf{v}'_j qui est égale à 1 par définition et c_{kl} est le coefficient de corrélation linéaire entre les variables \mathbf{v}'_k et \mathbf{v}'_l définies dans le paragraphe précédent.

Bien que la distance se présente différemment que dans le cas des variables, l'objectif reste identique : **représenter les variables dans un espace de dimension inférieure** en préservant cette notion de distance. On procédera à nouveau par une décomposition en valeurs singulières. En revanche, on va cette fois-ci s'intéresser à la matrice des produits scalaires entre les individus. On retrouve donc notre matrice de Gram $G \in \mathcal{M}_n(\mathbb{R})$ sur les individus

$$G = ZZ^T.$$

Or G est une matrice symétrique réelle, elle est donc orthogonalement semblable à une matrice diagonale, *i.e.* il existe donc une matrice orthogonale $U \in \mathcal{M}_n(\mathbb{R})$ et une matrice diagonale $\Sigma_n = \text{diag}(\lambda_1, \dots, \lambda_n)$ telle que $\lambda_1 \geq \dots \geq \lambda_n$. On a

$$G = U\Sigma_n U^T \quad \text{et pour tout } m \leq n, \quad G\mathbf{u}_m = \lambda_m \mathbf{u}_m.$$

où U est formée des vecteurs propres de G . Donc le vecteur propre \mathbf{u}_m est associé à la valeur propre λ_m qui est la m -ème plus grande valeur propre.

Ainsi, les nouvelles coordonnées d'une variable \mathbf{v}'_j sur un axe principal \mathbf{u} sont données par :

$$p_{\mathbf{u}}(\mathbf{v}'_j) = \langle \mathbf{v}'_j, \mathbf{u} \rangle = \sum_{i=1}^n z_{ij} u_i.$$

Regardons maintenant d'un peu plus près les propriétés du vecteur $\mathbf{p}_{\mathbf{u}} = \mathbf{p}_{\mathbf{u}}(Z^T) \in \mathbb{R}^p$ (on regarde les composantes des p variables sur \mathbf{u}) et remarquons le lien suivant entre les vecteurs variables et les coordonnées des projections

$$\mathbf{u}_m = \frac{1}{\sqrt{\lambda_m}} Z \mathbf{u}'_m = \frac{1}{\sqrt{\lambda_m}} \mathbf{p}_{\mathbf{u}'_m}(Z).$$

Ce dernier élément correspond à la projection des n individus sur l'axe factoriel \mathbf{u}'_m . Les composantes de l'axe factoriel \mathbf{u}_m sont donc proportionnelles aux composantes

principales des individus sur l'axe factoriel \mathbf{u}'_m .

On peut également montrer que $p_{\mathbf{u}_m}(\mathbf{v}'_j)$ est égal au coefficient de corrélation linéaire entre les vecteurs \mathbf{u}_m et \mathbf{v}'_j (les deux étant des vecteurs de norme égale à 1), c'est donc une valeur comprise entre -1 et 1 . Mais c'est aussi égal au coefficient de corrélation linéaire entre les vecteurs \mathbf{v}'_j et $\mathbf{p}_{\mathbf{u}'_m}(Z)$.

Enfin, on retiendra que les projections des variables sur les différents sous-espaces engendrés par les vecteurs \mathbf{u} se trouvent dans une hypersphère de rayon 1 (les coordonnées étant des coefficients de corrélation).

Dualité. Regardons maintenant les liens qui existent l'espace des individus et l'espace des variables. Nous avons déjà donné un petit aperçu avec qui ce qui précède mais nous pouvons également rajouter que la variance (ou l'inertie) est la même quelque soit la représentation choisie.

En effet l'inertie du nuage des individus est

$$\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^p \left(\frac{x_{ik} - m_k}{s_k} \right)^2 = \sum_{i=1}^n \sum_{k=1}^p (z_{ik})^2 = p.$$

car toutes les variables ont une variance de 1 et que les vecteurs sont normés.

En projetant les nuages de points sur les axes successifs, on décompose la variance, donc la variance totale n'est rien d'autre que la somme des valeurs propres de notre problème

$$Var = \sum_{k=1}^p \lambda_k = p.$$

Comme nous l'avons vu dans le cas de la SVD, il est également possible de faire le lien entre les deux espaces de représentations à l'aide des relations suivantes :

$$\mathbf{u}'_m = \frac{1}{\sqrt{\lambda_m}} \mathbf{p}_{\mathbf{u}_m}(Z) = \frac{1}{\sqrt{\lambda_m}} Z \mathbf{u}_m.$$

déjà présentée précédemment, ainsi que

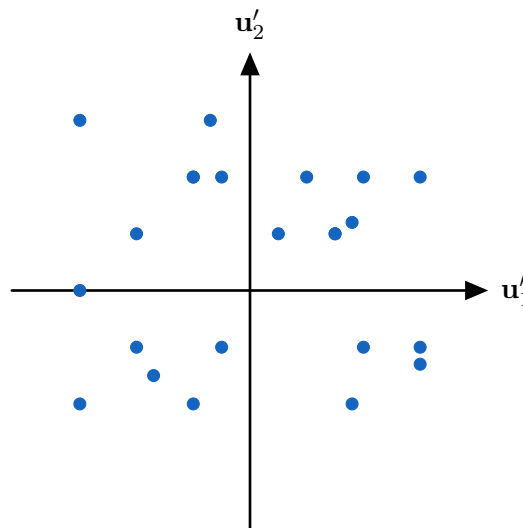
$$\mathbf{u}_m = \frac{1}{\sqrt{\lambda_m}} \mathbf{p}_{\mathbf{u}'_m}(Z) = \frac{1}{\sqrt{\lambda_m}} Z^T \mathbf{u}'_m.$$

Ces relations montrent qu'il est suffisant de procéder à une seule diagonalisation (par exemple celle de la matrice des corrélations linéaires) afin d'obtenir toutes les informations nécessaires à la synthèse des données dans les deux espaces de représentations.

8.3 Graphes et commentaires

Représentation des individus Si on cherche à représenter les individus dans un espace de dimension $s < p$, on va considérer un nouveau repère affine qui sera centré en le barycentre des données $\mathbf{m} = (m_1, \dots, m_p) \in \mathbb{R}^p$ et on prendra comme repère les axes définis par les vecteurs propres de la matrice de corrélations linéaires C .

Dans ce nouvel espace, les individus auront alors pour coordonnées les composantes principales des axes $\mathbf{p}_{\mathbf{u}'_1}(Z), \mathbf{p}_{\mathbf{u}'_2}(Z), \dots, \mathbf{p}_{\mathbf{u}'_s}(Z)$.

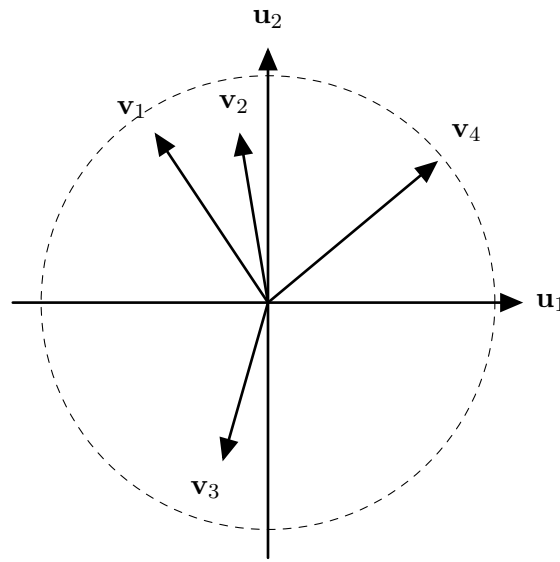


Projection des individus dans l'espace engendré par $(\mathbf{u}'_1, \mathbf{u}'_2)$.

Ce graphique peut s'interpréter de la façon suivante. Tout individu possédant une grande valeur positive (ou négative) pour un axe \mathbf{u}'_k donné est alors fortement corrélé à cet axe.

Or, rappelons qu'un axe principal est en fait une combinaison linéaire des variables initiales. Ainsi, si un axe \mathbf{u}'_k est fortement corrélé à un groupe de variables (*i.e.* les variables pour lesquelles le coefficient de corrélation avec l'axe \mathbf{u}'_k est élevé), alors l'ensemble des individus possédant une grande valeur positive (ou négative) à cet axe \mathbf{u}'_k , peuvent être caractérisés par ce groupe de variables.

Représentation des variables En ce qui concerne le variable, on considère le repère affine centré en l'origine de l'espace \mathbb{R}^n (et non plus centré en le barycentre!) dont une base est formée par les axes principaux $\mathbf{u}_1, \dots, \mathbf{u}_s$ qui forment à nouveau une base orthonormale. On rappelle que, dans cet espace, les points ont, sur chaque axe principal, comme coordonnées les valeurs $\mathbf{p}_{\mathbf{u}_1}(Z^T), \mathbf{p}_{\mathbf{u}_2}(Z^T), \dots, \mathbf{p}_{\mathbf{u}_s}(Z^T)$. Pour rappel, ces coordonnées ne sont rien d'autres que les coefficients de corrélations linéaires entre les variables \mathbf{v}_k et l'axe principale \mathbf{u}_k . Les coordonnées sont donc toutes comprises dans l'intervalle $[-1, 1]$



Cercle des corrélations dans l'espace engendré par $(\mathbf{u}_1, \mathbf{u}_2)$.

Comment interpréter ce cercle, pour cela il suffit de souvenir du sens des coordonnées de chacun des vecteurs \mathbf{v}'_k . Ainsi, sur cet exemple, on remarque que les vecteurs \mathbf{v}'_2 et \mathbf{v}'_3 sont corrélés (respectivement positivement et négativement) au deuxième axe principal alors que \mathbf{v}'_4 est la variable la plus corrélée au premier axe principal.

On est également en droit de demander quel est le sens de ces axes principaux. Sans signification des variables \mathbf{v}'_1 à \mathbf{v}'_4 , il est difficile de leur donner un sens. On peut cependant trouver de nombreux exemples⁴ pour lesquels il est facile de leur donner un sens.

Choix du nombre d'axes Il y a plusieurs façon afin de déterminer un choix optimal du nombre d'axes à conserver pour la réduction de dimensions par ACP. Cela peut se

4. On peut par exemple regarder l'exemple présenté au lien <https://openclassrooms.com/fr/courses/4525281-realisez-une-analyse-exploratoire-de-donnees/5278723-interpretez-le-cercle-des-corrélations>. Il permettra également de donner une autre approche de l'ACP.

faire par le biais de l'interprétation des axes et par une bonne connaissance des données et du domaine dont elles sont extraites.

Le plus souvent, on emploie des règles empiriques qui sont fondées sur les valeurs propres, car elles seules nous permettent de savoir à quel point l'information est bien synthétisée dans le nouveau sous-espace. Nous présenterons deux règles souvent utilisées qui supposent que les valeurs propres sont rangées par ordre décroissant $\lambda_1 \geq \dots \geq \lambda_p$

- **La règle de Cattell** : cette première représentation se base sur l'utilisation d'un histogramme sur lequel on représente les valeurs propres par ordre décroissant.

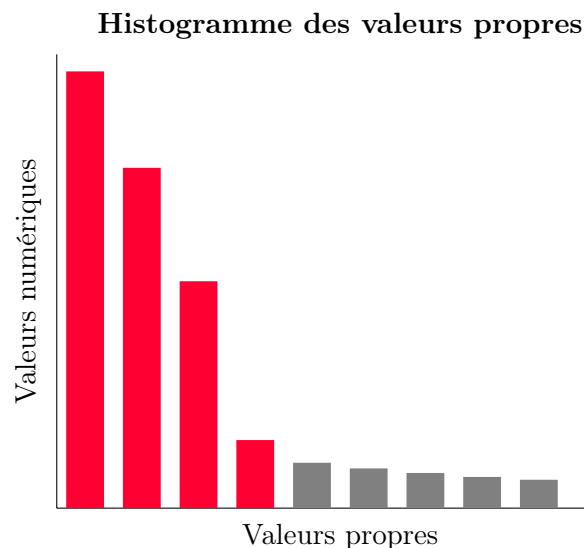
Le but est de chercher une rupture dans la courbe de décroissance des valeurs propres et de ne conserver que les espaces propres associés aux valeurs propres se trouvant avant cette rupture.

Cette rupture se trouve en regardant à partir de quelle valeur propre la décroissance faiblie.

- **La règle de Kayser** : cette règle est un peu plus simple que la précédente, elle consiste simplement à considérer les espaces propres dont les valeurs propres associées sont supérieures à 1.

Pourquoi cette valeur ? Tout simplement car la moyenne des valeurs propres est égale à 1 (on se rappelle que la somme des valeurs propres est égale à p et que l'on a exactement p valeurs propres).

Cela revient donc à conserver les axes dont la variance est supérieure à la moyenne.



Sur cet exemple, on note une rupture (ou un coude) après la quatrième valeur propre et on fait figurer en rouge les valeurs propres dont la valeur est supérieure à 1. Les deux règles présentées précédemment nous incitent donc à choisir une représentation

dans un sous-espace de dimension 4.

Interprétation de la représentation L'Analyse en Composantes Principales nous permet d'obtenir une représentation des données dans un sous-espace dans lequel on conserve l'essentielle de l'information. On pourra cependant donner quelques critères quantitatifs pour juger de la qualité de la représentation que ce soit **à l'échelle des individus** ou encore **à l'échelle des variables**.

- On peut évaluer la qualité de la représentation de **l'individu i sur l'axe \mathbf{u}'_m** en évaluant l'inertie de la projection de cet individu sur l'axe \mathbf{u}'_m et en divisant cela par l'inertie totale de l'individu i . Cette valeur est égale au \cos^2 de l'angle formé entre le vecteur de la projection de \mathbf{z}_i sur \mathbf{u}'_m et \mathbf{z}_i :

$$\frac{(p_{\mathbf{u}'_m}(\mathbf{z}_i))^2}{\sum_{k=1}^p (z_{ik})^2}.$$

On peut également mesurer la qualité de la représentation du nuage complet sur l'axe \mathbf{u}'_m par la qualité suivante

$$\frac{\lambda_m}{p}.$$

Enfin, la contribution de **l'individu i à l'axe \mathbf{u}_m** est l'inertie de la projection de i sur \mathbf{u}'_m mais divisée cette fois-ci par l'inertie totale du nuage projeté sur ce même axe

$$\frac{(p_{\mathbf{u}'_m}(\mathbf{z}_i))^2}{\lambda_m}.$$

- Nous pouvons faire la même chose en ce qui concerne les variables avec les éléments présentés dans les sections précédentes. On rappelle que le vecteur $\mathbf{p}_{\mathbf{u}_m}(Z^T)$ contient les coordonnées des individus \mathbf{z}_i sur l'axe principal \mathbf{u}_m qui sont des coefficients de corrélation. D'où

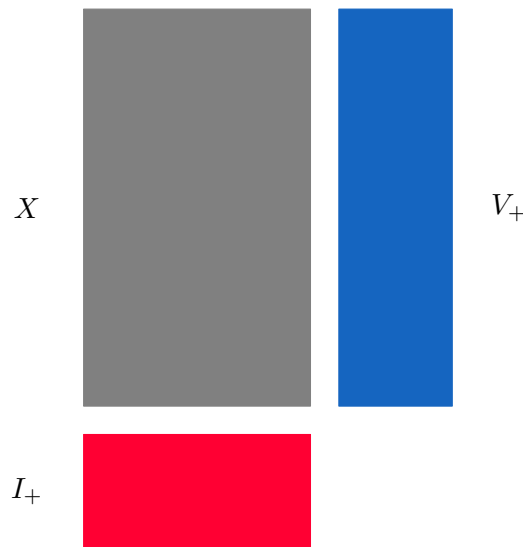
$$(\mathbf{p}_{\mathbf{u}'_m}(Z))_k = \text{cor}(\mathbf{v}'_k, \mathbf{u}_m).$$

Les variables fortement corrélées à un axe sont donc celles qui contribuent le plus à la définition de cet axe. Etant donnée un axe \mathbf{u}_m , on s'intéresse donc aux variables ayant les plus fortes coordonnées sur cet axe.

Sachant que $\mathbf{p}_{\mathbf{u}'_m}(Z) = \sqrt{\lambda_m} \mathbf{u}_m$ mais aussi que $(\mathbf{p}_{\mathbf{u}_m}(Z^T))_k = \text{cor}(\mathbf{v}'_k, \mathbf{u}_m)$, on pourra interpréter l'axe principal $\mathbf{p}_{\mathbf{u}'_m}(Z)$ en fonction des groupements des variables ayant une coordonnées forte sur \mathbf{u}_m .

Ajout de nouvelles informations Après avoir réalisé une ACP (que ce soit pour de la visualisation ou pour simplement pour des tâches de pré-traitements des données) nous sommes souvent amenés à considérer de nouvelles données : aussi bien des **individus** (ce qui arrive surtout dans un contexte de Machine Learning) que des **variables** à prendre en compte pour une nouvelle analyse.

Notre objectif est de savoir comment se servir des traitements effectués pour réaliser notre ACP et les utiliser sur ces nouveaux **individus/variables**. Pour cela on considère notre matrice $X \in \mathcal{M}_{n,p}(\mathbb{R})$ avec nos données initiales et on considère deux matrices $V_+ \in \mathcal{M}_{n,p_+}(\mathbb{R})$ et $I_+ \in \mathcal{M}_{n_+,p}(\mathbb{R})$ qui désignent respectivement la matrice des variables et des individus supplémentaires.



Il est alors facile de représenter les **individus supplémentaires** I_+ dans le sous-espace engendré par les s premiers axes principaux $\mathbf{u}'_1, \dots, \mathbf{u}'_s \in \mathbb{R}^p$ en calculant simplement leur projection sur ces axes

$$\mathbf{p}_U(I_+) = Z_+ U',$$

où $U' = (\mathbf{u}'_1, \dots, \mathbf{u}'_s)$ est le sous-espace engendré par les s premiers axes principaux dans \mathbb{R}^p et Z_+ représente une certaine transformation des individus I_+

On procède exactement de la même façon avec les **variables supplémentaires** V_+ , la représentation sur les s premiers axes principaux $\mathbf{u}_1, \dots, \mathbf{u}_s \in \mathbb{R}^n$ est donnée par

$$\mathbf{p}_U(V_+) = Y_+ U,$$

où $U' = (\mathbf{u}'_1, \dots, \mathbf{u}'_s)$ est le sous-espace engendré par les s premiers axes principaux et Y_+ représente une certaine transformation des variables V_+

Que cela soit pour les individus où les variables, nous avons effectués des transformations sur les données avant de procéder au traitement de ces dernières : rappelez vous le passage de la matrice X à la matrice Z . Regardons comment transformer les nouvelles données afin de pouvoir appliquer l'ACP sur ces dernières que cela soit du point de vue des **individus** ou encore des **variables**.

- **A l'échelle des individus** : de multiples raisons font que de nouveaux individus ne sont pas utilisés ou pris en compte pour la réalisation de l'ACP et ne sont utilisés qu'a posteriori : *données tests en Machine Learning - individus qui auraient un impact trop important sur les résultats ou sur lesquels planent une incertitude sur l'extraction des caractéristiques et que l'on préfère écarter pour ne pas fausser l'analyse,*

Nous avons procédé au **centrage** (on retranchait le vecteur moyenne $\mathbf{m} = (m_1, \dots, m_p)$) et à la **réduction** (multiplication par $\text{diag}(s_1^{-1}, \dots, s_p^{-1})$) puis à la normalisation des données initiales X avant d'effectuer l'ACP. Pour ces nouvelles données I_+ on effectue exactement la même chose et on obtient une nouvelle matrice Z_+ dont les éléments sont définis par

$$(Z_+)_{ij} = \frac{(I_+)_{ij} - m_j}{s_j \sqrt{n}}.$$

On fera attention au fait que les moyennes m_j et écart-types s_j employés sont bien ceux calculés sur les données initiales X et non sur la concaténation des anciens individus avec les individus supplémentaires.

La projection de ces nouveaux individus sur un axe principal \mathbf{u}'_m se déduit alors directement de ce qui précède

$$\mathbf{p}_{\mathbf{u}'_m}(Z_+) = Z_+ \mathbf{u}'_m.$$

- Qu'en est-il de l'ajout de nouvelles variables ? Le fonctionnement est un peu différent car ce sont des caractéristiques que l'on n'a pas encore rencontré. On va cette fois-ci obtenir notre matrice transformée en utilisant comme moyennes et écart-types ceux **des nouvelles variables** !. On définit alors

$$m_{+,j} = \frac{1}{n} \sum_{i=1}^n (V_+)_{ij} \text{ et } s_{+,j} = \frac{1}{n} \sum_{i=1}^n ((V_+)_{ij} - m_{+,j})^2.$$

On peut alors définir notre matrice Y_+ dont le terme général en (i, j) est donné par

$$(Y_+)_{ij} = \frac{(V_+)_{ij} - m_{+,j}}{s_{+,j} \sqrt{n}}.$$

La projection des nouvelles variables sur le sous-espace engendré par \mathbf{u}_m se déduit alors directement de ce qui précède

$$\mathbf{p}_{\mathbf{u}_m}(Y_+) = Y_+^T \mathbf{u}_m.$$

On termine cette section par une dernière remarque concernant les individus. Il n'est pas rare que ces derniers ne présentent pas que des caractéristiques numériques mais ils peuvent aussi être décrits par une (ou plusieurs variables) nominale : *sexe - catégorie socio-professionnelle - ...* dont la prise en compte peut se révéler intéressante dans l'analyse de l'ACP et de ses différents groupes.

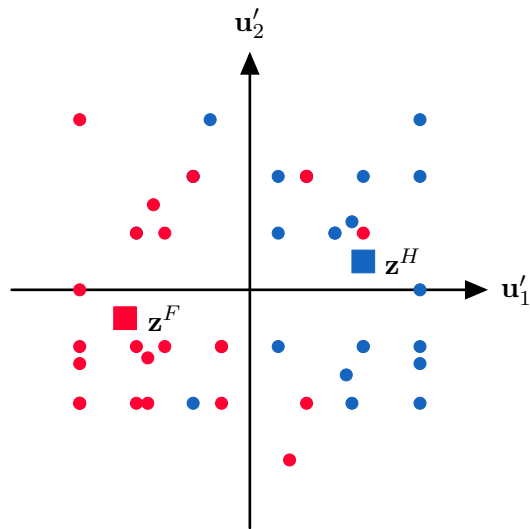
En pratique, la méthodologie présentée précédemment reste inchangée et s'applique à l'ensemble des données indépendamment de la modalité de la variable nominale. Cette variable nominale interviendra uniquement sur la représentation graphique et sur la représentation des différents groupes par un individu moyen.

Considérons une variable nominale \mathbf{q} possédant q modalités, prenons l'exemple d'une variable avec deux modalités pour simplifier la présentation (Homme H et Femme F). On peut calculer les barycentres \mathbf{z}^H et \mathbf{z}^F des groupes :

$$\mathbf{z}^H = \frac{1}{|H|} \sum_{i:q_i=H} z_i \quad \text{et} \quad \mathbf{z}^F = \frac{1}{|F|} \sum_{i:q_i=F} z_i.$$

Ces individus moyens peuvent ainsi être projetés sur le sous-espace engendré par un axe principal \mathbf{u}'_m comme nous l'avons plus tôt pour la projection de nouveaux individus, *i.e.* à l'aide de la relation :

$$p_{\mathbf{u}'_m}(\mathbf{z}^H) = (\mathbf{z}^H)^T \mathbf{u}'_m = \langle \mathbf{z}^H, \mathbf{u}'_m \rangle \quad \text{et} \quad p_{\mathbf{u}'_m}(\mathbf{z}^F) = (\mathbf{z}^F)^T \mathbf{u}'_m = \langle \mathbf{z}^F, \mathbf{u}'_m \rangle.$$



9 Généralisation des méthodes

Cette courte section nous permet de présenter rapidement comment généraliser la décomposition en valeurs singulières mais aussi l'ACP lorsque nous travaillons avec des métriques différentes (*i.e.* en changeant la notion de distance ou encore de produit scalaire) mais aussi lorsque les exemples d'un jeu de données n'ont pas tous la même importance.

On conserve les mêmes notations que précédemment et on appelle toujours $X \in \mathcal{M}_{n,p}(\mathbb{R})$ notre matrice de design à n individus décrits par p variables.

Métrique générale. Jusqu'à présent, nous avons toujours considéré la distance euclidienne, *i.e.* la distance induite par la métrique I_p ou encore le produit scalaire canonique. Mais rien de nous empêche de considérer un autre produit scalaire et donc une autre métrique M ⁵.

Soit $M \in \mathcal{M}_p(\mathbb{R})$ une matrice carrée d'ordre p symétrique et définie positive. Nous avons vu que ce type de métrique représente une forme bilinéaire symétrique, *i.e.* un produit scalaire. Ainsi, étant donné deux vecteurs \mathbf{x} et \mathbf{x}' , on peut définir le produit scalaire relativement à M par

$$\langle \mathbf{x}, \mathbf{x}' \rangle_M = \mathbf{x}^T M \mathbf{x}' = \sum_{k=1}^p \sum_{l=1}^p x_k m_{kl} x'_l.$$

On rappelle également que, relativement à cette métrique M , un vecteur \mathbf{x} est dit normé si

$$\|\mathbf{x}\|_M^2 = \mathbf{x}^T M \mathbf{x} = 1.$$

Enfin, deux vecteurs \mathbf{x} et \mathbf{x}' sont dits orthogonaux si

$$\langle \mathbf{x}, \mathbf{x}' \rangle_M = \mathbf{x}^T M \mathbf{x}' = 0.$$

Poids quelconques. Dans les présentations effectuées précédemment, nous avons supposé que tous les exemples (individus) avaient le même poids, *i.e.* qu'ils avaient la même la importance l'un par rapport à l'autre. Il n'est cependant pas rare d'accorder un poids

5. Cette notion de métrique fait également l'objet d'un pan de recherche entier en Apprentissage Machine, que l'on appelle le *Metric Learning* et qui consiste à apprendre une nouvelle représentation des données dans laquelle les tâches d'apprentissages sont plus simples et plus performantes

plus important à certains individus ou exemples qui sont plus représentatifs de la population ou dont l'acquisition de la donnée est plus fiable. Cela permet de leur accorder plus d'importance dans les méthodes d'analyse. A l'inverse, une donnée "bruitée" se verra souvent accorder un poids plus faible afin de minimiser son impact dans les méthodes d'analyse.

Nous faisons maintenant l'hypothèse que les individus \mathbf{x}_i ont des poids associés $w_i \geq 0$ et on note $\mathbf{w} \in \mathbb{R}^n$ ce vecteur de poids et $W = \text{diag}(\mathbf{w})$ la matrice correspondante. On supposera que ces poids somment à n , le nombre d'exemples.

L'introduction d'un vecteur de poids va impacter les valeurs des moyennes et écart-types associées aux données qui deviennent alors :

$$m_k = \frac{1}{n} \sum_{i=1}^n w_i x_{ik} \quad \text{et} \quad s_k^2 = \frac{1}{n} \sum_{i=1}^n w_i (x_{ik} - m_k)^2, \quad \text{où} \quad \sum_{i=1}^n w_i = n.$$

Cette attribution de poids va donc modifier les axes principaux de l'ACP car certains individus vont d'avantage contribuer à la définition de ces axes.

Regardons maintenant ce que devient la notion d'inertie (de variance) relativement à l'introduction de poids sur les individus mais aussi à la nouvelle métrique M . Rappelons que l'inertie est définie comme un terme de variance, donc comme la somme du carré des distances au barycentre du nuage de points sauf que cette fois-ci :

- les individus sont pondérés par des poids w_i
- la métrique utilisée (et donc la distance) n'est plus euclidienne, mais dépend d'une matrice M .

L'inertie totale du nuage de points est donc :

$$\frac{1}{n} \sum_{i=1}^n w_i d_M^2(\mathbf{x}_i, \bar{\mathbf{x}}) = \sum_{i=1}^n w_i \|\mathbf{x}_i - \bar{\mathbf{x}}\|_M^2 = \sum_{i=1}^n w_i \langle \mathbf{x}_i - \bar{\mathbf{x}}, \mathbf{x}_i - \bar{\mathbf{x}} \rangle_M,$$

où $\bar{\mathbf{x}} = \sum_{i=1}^n \mathbf{x}_i = \mathbf{m}$ représente le barycentre du nuage de points.

Il s'agit maintenant de voir les impacts sur la définition des axes principaux. Considérons maintenant \mathbf{u}_1 le premier axe principal, *i.e.* le vecteur de \mathbb{R}^p selon lequel la variance du jeu de données est maximale (on rappelle que c'est aussi l'axe selon lequel on préserve au maximum l'information présente dans le jeu de données). Ainsi

$$\max_{\mathbf{u}_1 \in \mathbb{R}^p} \text{Var}(X) = \max_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{i=1}^n w_i \left\| \frac{\langle \mathbf{u}_1, \mathbf{x}_i \rangle_M}{\|\mathbf{u}_1\|_M} \mathbf{u}_1 \right\|_M^2.$$

On supposera, sans perte de généralités, que le vecteur \mathbf{u}_1 est normé par rapport à la métrique M , donc $\|\mathbf{u}_1\|_M^2 = 1$. En développant le terme à maximiser on trouve :

$$\begin{aligned}
\sum_{i=1}^n w_i \left\| \frac{\langle \mathbf{u}_1, \mathbf{x}_i \rangle_M}{\|\mathbf{u}_1\|_M} \mathbf{u}_1 \right\|_M^2 &= \sum_{i=1}^n w_i \langle \mathbf{x}_i, \mathbf{u}_1 \rangle_M \langle \mathbf{x}_i, \mathbf{u}_1 \rangle_M \\
&\downarrow \text{ bilinéarité du produit scalaire} \\
&= \sum_{i=1}^n w_i \langle \mathbf{x}_i, \mathbf{u}_1 \rangle_M^2 \\
&\downarrow \text{ car } \|\mathbf{u}_1\|_M^2 = \langle \mathbf{u}_1, \mathbf{u}_1 \rangle_M = 1 \\
&= \sum_{i=1}^n w_i \langle \mathbf{x}_i, \mathbf{u}_1 \rangle_M^2 \\
&= \sum_{i=1}^n w_i (\mathbf{x}_i^T M \mathbf{u}_1)^T (\mathbf{x}_i^T M \mathbf{u}_1) \\
&\downarrow \text{ on développe la transposée} \\
&= \sum_{i=1}^n w_i \mathbf{u}_1^T M \mathbf{x}_i \mathbf{x}_i^T M \mathbf{u}_1 \\
&= \mathbf{u}_1^T M \left(\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^T \right) M \mathbf{u}_1 \\
&= \mathbf{u}_1^T M X^T W X M \mathbf{u}_1,
\end{aligned}$$

On aboutit donc au problème d'optimisation sous contrainte suivant :

$$\max_{\mathbf{u}_1 \in \mathbb{R}^p} \sum_{i=1}^n w_i \left\| \frac{\langle \mathbf{u}_1, \mathbf{x}_i \rangle_M}{\|\mathbf{u}_1\|_M} \mathbf{u}_1 \right\|_M^2 = \max_{\mathbf{u}_1 \in \mathbb{R}^p, \|\mathbf{u}_1\|_M^2 = 1} \mathbf{u}_1^T M X^T W X M \mathbf{u}_1,$$

où $M X^T W X M$ est une matrice de Gram, elle est donc symétrique définie positive et le problème d'optimisation consiste à donc à maximiser une forme quadratique.

Le problème peut donc se résoudre par méthode spectrale. En effet, en écrivant le Lagrangien de notre problème d'optimisation, on a

$$L(\mathbf{u}_1, \alpha) = \mathbf{u}_1^T M X^T W X M \mathbf{u}_1 - \alpha (\mathbf{u}_1^T M \mathbf{u}_1 - 1).$$

On cherche à nouveau les extrema de cette fonction, on va donc chercher les valeurs de \mathbf{u}_1 et α pour lesquelles le gradient du Lagrangien s'annule :

$$\nabla L(\mathbf{u}_1, \alpha) = \mathbf{0} = \begin{pmatrix} \frac{\partial L}{\partial \mathbf{u}_1}(\mathbf{u}, \alpha) \\ \frac{\partial L}{\partial \alpha}(\mathbf{u}, \alpha) \end{pmatrix}.$$

En particulier, la dérivée par rapport à \mathbf{u}_1 nous donne :

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{u}_1}(\mathbf{u}_1, \alpha) = \mathbf{0} &\iff 2K\mathbf{u} - 2\alpha M\mathbf{u}_1 = \mathbf{0}, \\ &\iff K\mathbf{u}_1 = \alpha M\mathbf{u}_1, \\ &\quad \downarrow \text{définition de } K \\ &\iff MX^T WXM\mathbf{u}_1 = \alpha M\mathbf{u}_1, \\ &\quad \downarrow M \text{ est définie positive donc inversible} \\ &\iff X^T WXM\mathbf{u}_1 = \alpha \mathbf{u}_1. \end{aligned}$$

Il nous faut donc chercher les valeurs propres et vecteurs propres de la matrice $X^T WXM$.

Quid des variables ? Dans le cas précédent, nous avons décidé d'employer une métrique M qui pondère l'importance de chaque variable (mais aussi les corrélations) et d'attribuer des poids non uniformes W aux individus.

Nous pourrions très bien décider d'appliquer une métrique non plus dans l'espace \mathbb{R}^p mais dans l'espace \mathbb{R}^n (que l'on pourrait noter M') et d'attribuer des poids non uniformes aux variables W' .

On applique alors tout ce qui précède non plus à la matrice X mais à la matrice X^T . Dans ce cas, chercher les axes principaux revient à décomposer la matrice $XW'X^T M'$.

Il faut cependant avouer que cette analyse (duale) se rencontre très rarement en pratique.

Enfin, si l'on souhaite projeter nos données sur le sous-espace engendré par les s premiers vecteurs propres $\mathbf{u}_1, \dots, \mathbf{u}_s$, alors la matrice des coordonnées des vecteurs dans cette base est donnée par :

$$\mathbf{p}_U(X) = XMU,$$

$$\text{où } U = (\mathbf{u}_1, \dots, \mathbf{u}_m) \text{ et } (\mathbf{p}_U(X))_{im} = p_{\mathbf{u}_m}(\mathbf{x}_i) = \langle \mathbf{x}_i, \mathbf{u}_m \rangle_M.$$

A nouveau, le principe reste le même lorsque l'on cherche à projeter de nouveaux individus et/ou de nouvelles variables, nous ne réécrivons donc pas les détails.

Remarque Noter que dans cette section, il n'a pas été fait mention d'une éventuelle transformation des données (centrage, réduction et division par \sqrt{n}).

10 Analyse Factorielle des Correspondances (AFC)

Dans les trois première sections, nous nous sommes concentrés sur l'analyse de données quantitatives réelles. L'étude de l'ACP a cependant laisser apercevoir qu'il pourrait aussi être intéressant d'étudier les **variables qualitatives nominales**. C'est l'objet de la présente section à travers la présentation de l'Analyse Factorielle des correspondances *AFC*.

L'AFC consiste à analyser des tables de contingence, tables qui croisent deux variables qualitatives, disons P et Q , ayant chacune un certain nombre de modalités p et q respectivement. Ainsi, une table de contingence, notée N , sera un tableau de taille $p \times q$ dont chaque entrée n_{ij} sera égal au nombre d'individus possédant **à la fois la caractéristique p_i et la caractéristique q_j** .

$$N = \begin{matrix} & \mathbf{q}_1 & \cdots & \mathbf{q}_j & \cdots & \mathbf{q}_p \\ \mathbf{p}_1 & \left(\begin{matrix} n_{11} & \cdots & n_{1j} & \cdots & n_{1q} \\ \vdots & & \vdots & & \vdots \\ \mathbf{p}_i & n_{i1} & \cdots & n_{ij} & \cdots & n_{iq} \\ \vdots & & \vdots & & \vdots \\ \mathbf{p}_n & n_{p1} & \cdots & n_{pj} & \cdots & n_{pq} \end{matrix} \right) \end{matrix},$$

où \mathbf{p}_i désigne le vecteur des répartitions des modalités de Q parmi tous les individus ayant la modalité p_i . De la même façon, \mathbf{q}_j est le vecteur des répartitions des modalités P parmi tous les individus ayant la modalité q_j .

Exemple 10.1. *Un exemple (purement fictif) de matrice de contingence est donné en considérant deux modalités :*

- **animal de compagnie** (P) dont les modalités sont chat - chien - souris - lézard,
- **couleur des yeux** (Q) dont les modalités sont bleu - vert - marron

Pour cela on dispose d'un échantillon de 248 personnes.

$$N = \begin{matrix} & \text{bleu} & \text{vert} & \text{marron} \\ \begin{matrix} \text{chat} \\ \text{chien} \\ \text{souris} \\ \text{lézard} \end{matrix} & \left(\begin{matrix} 45 & 34 & 12 \\ 3 & 21 & 90 \\ 2 & 6 & 8 \\ 12 & 8 & 7 \end{matrix} \right) \end{matrix},$$

10.1 Préliminaires

L'objectif de cette partie est de commencer par quelques analyses simples de la table de contingence.

$$N = \begin{matrix} & \mathbf{q}_1 & \cdots & \mathbf{q}_j & \cdots & \mathbf{q}_p \\ \mathbf{p}_1 & \left(\begin{matrix} n_{11} & \cdots & n_{1j} & \cdots & n_{1p} \end{matrix} \right) & n_{1\cdot} \\ \vdots & \left(\begin{matrix} \vdots & & \vdots & & \vdots \end{matrix} \right) & \vdots \\ \mathbf{p}_i & \left(\begin{matrix} n_{i1} & \cdots & n_{ij} & \cdots & n_{ip} \end{matrix} \right) & n_{i\cdot} \\ \vdots & \left(\begin{matrix} \vdots & & \vdots & & \vdots \end{matrix} \right) & \vdots \\ \mathbf{p}_n & \left(\begin{matrix} n_{n1} & \cdots & n_{nj} & \cdots & n_{np} \end{matrix} \right) & n_{n\cdot} \\ & n_{\cdot 1} & \cdots & n_{\cdot j} & \cdots & n_{\cdot p} & n \end{matrix}$$

On peut d'abord définir ce que l'on appelle des **marges** relatives à cette table :

- $n_{i\cdot} = \sum_{j:q_j \in Q} n_{ij}$ qui correspond au nombre d'individus ayant la modalité p_i pour la variable P .
- $n_{\cdot j} = \sum_{i:p_i \in P} n_{ij}$ qui correspond au nombre d'individus ayant la modalité q_j pour la variable Q .

On peut alors avoir accès au nombre total d'individus n par les relations

$$n = \sum_{j:q_j \in Q} n_{\cdot j} = \sum_{i:p_i \in P} n_{i\cdot} = \sum_{i=1}^p \sum_{j=1}^q n_{ij}.$$

Exemple 10.2. Reprenons notre exemple précédent avec nos 248 personnes et faisons cette fois-ci figurer les marges :

$$N = \begin{matrix} & \text{bleu} & \text{vert} & \text{marron} \\ \text{chat} & \left(\begin{matrix} 45 & 34 & 12 \end{matrix} \right) & 91 \\ \text{chien} & \left(\begin{matrix} 3 & 21 & 90 \end{matrix} \right) & 114 \\ \text{souris} & \left(\begin{matrix} 2 & 6 & 8 \end{matrix} \right) & 16 \\ \text{lézard} & \left(\begin{matrix} 12 & 8 & 7 \end{matrix} \right) & 27 \\ & 62 & 69 & 117 & 248 \end{matrix}$$

Cette première table N est appelée **table des effectifs bruts**, mais il est parfois d'usage de travailler non pas avec les effectifs mais plutôt avec les fréquences. Pour cela on définit une table des fréquences F ayant la même dimension que N et de terme général

$$f_{ij} = \frac{n_{ij}}{n},$$

i.e. on regarde simplement la proportion d'individus qui possèdent à la fois les caractéristiques p_i et q_j parmi les n individus considérés.

$$F = \begin{matrix} & \mathbf{q}_1 & \cdots & \mathbf{q}_j & \cdots & \mathbf{q}_p \\ \mathbf{p}_1 & \left(\begin{array}{ccccc} f_{11} & \cdots & f_{1j} & \cdots & f_{1p} \end{array} \right) & f_{1\cdot} \\ \vdots & \left(\begin{array}{ccccc} \vdots & & \vdots & & \vdots \end{array} \right) & \vdots \\ \mathbf{p}_i & \left(\begin{array}{ccccc} f_{i1} & \cdots & f_{ij} & \cdots & f_{ip} \end{array} \right) & f_{i\cdot} \\ \vdots & \left(\begin{array}{ccccc} \vdots & & \vdots & & \vdots \end{array} \right) & \vdots \\ \mathbf{p}_n & \left(\begin{array}{ccccc} f_{n1} & \cdots & f_{nj} & \cdots & f_{np} \end{array} \right) & f_{n\cdot} \\ n_{\cdot 1} & \cdots & f_{\cdot j} & \cdots & f_{\cdot p} & 1 \end{matrix}$$

De la même façon on pourra également définir des **marges** sur la matrice de fréquences

- $f_{i\cdot} = \sum_{j:q_j \in Q} f_{ij}$ qui correspond à la proportion d'individus ayant la modalité p_i .
- $f_{\cdot j} = \sum_{i:p_i \in P} f_{ij}$ qui correspond à la proportion d'individus ayant la modalité q_j .

La somme totale des éléments de F (qui doit valoir 1 !) peut se définir par

$$1 = \sum_{j:q_j \in Q} f_{\cdot j} = \sum_{i:p_i \in P} f_{i\cdot} = \sum_{i=1}^p \sum_{j=1}^q f_{ij}.$$

Pour la suite de la présentation, nous noterons \mathbf{p}_i le **vecteur des fréquences des modalités de Q des individus ayant la modalité p_i** et \mathbf{q}_j le **vecteur des fréquences des modalités de P des individus ayant la modalité q_j** . On ne résonnera donc qu'avec les fréquence à partir de maintenant.

Exemple 10.3. On reprend à nouveau notre exemple précédent avec notre table de contingence N et on calcule la table des fréquences correspondante

$$F = \begin{matrix} & \text{bleu} & \text{vert} & \text{marron} \\ \text{chat} & \left(\begin{array}{ccc} 0.18 & 0.14 & 0.05 \end{array} \right) & 0.37 \\ \text{chien} & \left(\begin{array}{ccc} 0.01 & 0.09 & 0.36 \end{array} \right) & 0.46 \\ \text{souris} & \left(\begin{array}{ccc} 0.01 & 0.02 & 0.03 \end{array} \right) & 0.06 \\ \text{lézard} & \left(\begin{array}{ccc} 0.05 & 0.03 & 0.03 \end{array} \right) & 0.11 \\ & 0.25 & 0.28 & 0.47 & 1 \end{matrix}$$

Cette table des fréquences va nous permettre d'étudier les liaisons ou corrélations entre les variables P et Q .⁶

Plus précisément, on va plutôt chercher à savoir si **des variables sont indépendantes ou non**. D'un point de vue statistique, deux variables P et Q sont indépendantes,

6. Vous verrez, l'année prochaine en cours de statistiques, comment étudier si deux variables qualitatives sont corrélées ou non entre elles à travers des tests statistiques et la loi de probabilité du χ^2 . On verra un petit peu plus tard dans ce document que l'on peut également définir une distance du même nom que cette loi, la distance du χ^2 .

notée $P \perp Q$ si les fréquences observées pour une variable donnée ne dépendent pas des fréquences des modalités observées pour l'autre variable. On peut aussi traduire cela par :

$$\forall(i, j) : (p_i, q_j) \in P \times Q \quad f_{ij} = f_{i.} f_{.j}.$$

A travers cette relation on peut aussi dire que la probabilité qu'un individu possède en même temps les attributs p_i et q_j est égale à la probabilité qu'un individu possède l'attribut p_i , $f_{i.}$ multipliée par la probabilité qu'il possède l'attribut q_j , $f_{.j}$. On dit aussi que **la probabilité jointe est égale au produit des probabilités marginales**.

Ce que l'on va chercher à faire ici c'est de mesurer l'écart à cette indépendance en comparant les valeurs de f_{ij} aux produits $f_{i.} f_{.j}$. Une telle comparaison offre deux cas possibles :

- $f_{ij} > f_{i.} f_{.j}$, la probabilité jointe est supérieure au produit des probabilités marginales, cela signifie que les modalités p_i et q_j **s'attirent**,
- au contraire si $f_{ij} < f_{i.} f_{.j}$, la probabilité jointe est inférieure au produit des probabilités marginales, cela signifie que les modalités p_i et q_j **se repoussent**.

Pour cela on va comparer une table des fréquences dite empirique , qui contient les fréquences observées, *i.e.* il s'agit de notre table F , et on va construire une deuxième table F' qui contiendra les effectifs théoriques dans le cas où les deux variables étudiées sont indépendantes.

Exemple 10.4. Reprenons la table des fréquences observées F de notre exemple

$$F = \begin{array}{c} \begin{array}{cc} & \begin{array}{ccc} \text{bleu} & \text{vert} & \text{marron} \end{array} \\ \begin{array}{c} \text{chat} \\ \text{chien} \\ \text{souris} \\ \text{lézard} \end{array} & \left(\begin{array}{ccc} 0.18 & 0.14 & 0.05 \\ 0.01 & 0.09 & 0.36 \\ 0.01 & 0.02 & 0.03 \\ 0.05 & 0.03 & 0.03 \end{array} \right) \end{array} \begin{array}{c} 0.37 \\ 0.46 \\ 0.06 \\ 0.11 \end{array} \\ \begin{array}{ccc} 0.25 & 0.28 & 0.47 \end{array} \quad 1 \end{array}$$

et construisons la table des fréquences théoriques F' dans le cas où nos deux variables sont indépendantes, pour rappel, il suffit simplement de faire le produit des marginales qui sont données ci-dessus

$$F' = \begin{array}{c} \begin{array}{cc} & \begin{array}{ccc} \text{bleu} & \text{vert} & \text{marron} \end{array} \\ \begin{array}{c} \text{chat} \\ \text{chien} \\ \text{souris} \\ \text{lézard} \end{array} & \left(\begin{array}{ccc} 0.10 & 0.10 & 0.17 \\ 0.11 & 0.13 & 0.22 \\ 0.01 & 0.02 & 0.03 \\ 0.03 & 0.03 & 0.05 \end{array} \right) \end{array} \begin{array}{c} 0.37 \\ 0.46 \\ 0.06 \\ 0.11 \end{array} \\ \begin{array}{ccc} 0.25 & 0.28 & 0.47 \end{array} \quad 1 \end{array}$$

Les entrées mises en avant en **rouges** représentent des cas où les modalités ont tendance à **s'attirer**, c'est-à-dire que $f_{ij} > f_{i.}f_{.j}$. À l'inverse, celles en **bleues** représentent des cas où les modalités ont plutôt tendance à **se repousser** c'est-à-dire que $f_{ij} < f_{i.}f_{.j}$.

En cas d'indépendances entre les deux variables, nous avons les propriétés suivantes sur les lignes et les colonnes

- les lignes i de terme général $\frac{f_{ij}}{f_{i.}}$ sont **proportionnelles à la marge** :

$$f_{ij} = f_{i.}f_{.j} \iff \frac{f_{ij}}{f_{i.}} = f_{.j}$$

- les colonnes j de terme général $\frac{f_{ij}}{f_{.j}}$ sont **proportionnelles à la marge** :

$$f_{ij} = f_{i.}f_{.j} \iff \frac{f_{ij}}{f_{.j}} = f_{i.}$$

On va donc naturellement introduire les concepts de **profils lignes** et **profils colonnes** qui sont les répartitions (en proportion) à l'intérieur d'une ligne et d'une colonne respectivement.

Les **profils lignes** i sont définis par $\frac{f_{ij}}{f_{i.}}$, i.e. le pourcentage des individus ayant la modalité q_j de Q parmi ceux ayant la modalité p_i de P .

Les **profils colonnes** j sont définis par $\frac{f_{ij}}{f_{.j}}$, i.e. le pourcentage des individus ayant la modalité p_i de P parmi ceux ayant la modalité q_j de Q .

10.2 Analyse Factorielle des Correspondances

L'Analyse Factorielle des Correspondances a les mêmes objectifs que l'ACP, il s'agit à nouveau d'une méthode de réduction de la dimension mais qui ne traite bien évidemment pas du même type de données, ici on traite plutôt des données qualitatives. Contrairement à l'ACP, on ne parlera pas ici d'espace des *variables* ou des *individus*, la structure des données étant très différente. On va plutôt parler de profils :

- **les profils lignes** : qui peuvent être vus comme des vecteurs de l'espace \mathbb{R}^q . Ils correspondent aux *individus* dans le cadre de l'ACP.
- **les profils colonnes** : qui peuvent être vus comme des vecteurs de l'espace \mathbb{R}^p . Ils correspondent aux *variables* dans le cadre de l'ACP.

L'objectif reste le même, trouver des sous-espaces de faibles dimensions dans lesquels on conserve le maximum d'information et dans lesquels l'interprétation des données se retrouve facilitée. Précédemment, nous avons vu que conserver l'information initialement présente revenait à conserver l'inertie du nuage de points. Cette inertie étant basée

sur une distance entre les individus par exemple. Cependant, dans le cadre de l'AFC, nous n'avons pas d'individus, nous avons uniquement des informations en terme de fréquences, nous ne pourrions donc pas utiliser une distance euclidienne dans ce contexte mais nous allons plutôt devoir définir une distance plus adaptée⁷.

10.2.1 Application aux profils lignes

Commençons par représenter nos profils lignes. Nous en avons déjà donné une définition à la section précédente mais on rappelle qu'il s'agit d'une matrice L de taille $p \times q$ dont chaque entrée l_{ij} est définie par

$$l_{ij} = \frac{f_{ij}}{f_{i\cdot}}.$$

$$L = \begin{matrix} & \mathbf{l}_1 & \begin{pmatrix} \frac{f_{11}}{f_{1\cdot}} & \dots & \frac{f_{1j}}{f_{1\cdot}} & \dots & \frac{f_{1q}}{f_{1\cdot}} \end{pmatrix} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{l}_i & \begin{pmatrix} \frac{f_{i1}}{f_{i\cdot}} & \dots & \frac{f_{ij}}{f_{i\cdot}} & \dots & \frac{f_{iq}}{f_{i\cdot}} \end{pmatrix} & 1, \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{l}_n & \begin{pmatrix} \frac{f_{p1}}{f_{p\cdot}} & \dots & \frac{f_{pj}}{f_{p\cdot}} & \dots & \frac{f_{pq}}{f_{p\cdot}} \end{pmatrix} & 1 \end{matrix}$$

où on rappelle que $f_{i\cdot} = \sum_{j=1}^q f_{ij}$ et \mathbf{l}_i est un vecteur de \mathbb{R}^q tel que $\mathbf{l}_i = \left(\frac{f_{i1}}{f_{i\cdot}}, \dots, \frac{f_{ij}}{f_{i\cdot}}, \dots, \frac{f_{iq}}{f_{i\cdot}} \right)$. De plus tous les éléments des \mathbf{l}_i somment à 1.

La question est de savoir quelle distance employer pour analyser ces profils lignes. La distance euclidienne n'est pas forcément adaptée ici, on va plutôt employer la **distance du χ^2** qui est définie par

$$d_{\chi^2}^2(\mathbf{l}_a, \mathbf{l}_b) = \sum_{j=1}^q \frac{1}{f_{\cdot j}} \left(\frac{f_{aj}}{f_{a\cdot}} - \frac{f_{bj}}{f_{b\cdot}} \right)^2.$$

Cette distance compare aussi deux profils lignes en calculant le carré de la différence terme à terme des deux profils. On prendra cependant garde au fait que cet écart est pondéré par un facteur $\frac{1}{f_{\cdot j}}$ correspondant à la j -ème dimension (ou entrée) du vecteur. Ainsi, plus la fréquence de la modalité $j : q_j$ de la variable Q est faible, plus la distance associée sera élevée.

7. La distance qui sera présentée dans cette section sera également étudiée en Statistiques dans le cadre de tests statistiques pour évaluer si deux variables qualitatives sont ou non corrélées entre elles.

La distance du \mathcal{X} va donc privilégier, au sens : augmenter la distance, les écarts vis-à-vis d'une modalité peu fréquente voire "rare".

Ecriture matricielle Contrairement à l'ACP, les lignes et les colonnes ont naturellement **une pondération non uniforme** qui est imposée par la matrice de contingence en terme de fréquences, *i.e.* la matrice F . Cette pondération, sur les lignes par exemple, est la proportion d'individus dans l'échantillon qui possède la modalité $i : p_i$ de la variable P .

On pourra alors également définir **un barycentre $\bar{\mathbf{l}}$ des profils lignes** par la relation :

$$\bar{\mathbf{l}} = \sum_{i=1}^p w_i \mathbf{l}_i,$$

où $w_i = f_{i.}$. Plus précisément, une coordonnée \bar{l}_j de ce vecteur est donnée par

$$\begin{aligned} \bar{l}_j &= \sum_{i=1}^p \cancel{f_{i.}} \frac{f_{ij}}{\cancel{f_{i.}}}, \\ &= f_{.j}. \end{aligned}$$

Finalement, on voit que les composantes du barycentre des profils lignes ne sont rien d'autres que marges des modalités de la variable Q .

Exemple 10.5. Reprenons notre exemple habituel qui nous suivra tout au long de cette section. On rappelle que la matrice F est donnée par

$$F = \begin{array}{ccccc} & \text{bleu} & \text{vert} & \text{marron} & \\ \text{chat} & 0.18 & 0.14 & 0.05 & 0.37 \\ \text{chien} & 0.01 & 0.09 & 0.36 & 0.46 \\ \text{souris} & 0.01 & 0.02 & 0.03 & 0.06 \\ \text{lézard} & 0.05 & 0.03 & 0.03 & 0.11 \\ & 0.25 & 0.28 & 0.47 & 1 \end{array}.$$

A l'aide de cette matrice F , on peut déterminer la matrice L ainsi que le barycentre des profils lignes $\bar{\mathbf{l}}$

$$L = \begin{array}{ccccc} & \text{chat} & \text{chien} & \text{souris} & \text{lézard} \\ & 0.49 & 0.38 & 0.13 & 1 \\ & 0.02 & 0.20 & 0.78 & 1 \\ & 0.17 & 0.33 & 0.50 & 1 \\ & 0.46 & 0.27 & 0.27 & 1 \\ \bar{\mathbf{l}} & 0.25 & 0.28 & 0.47 & 1 \end{array}.$$

Cette matrice L peut être entièrement déterminée à l'aide du calcul matriciel. Pour cela introduisons les matrices diagonales $D_P \in \mathcal{M}_p(\mathbb{R}^p)$ et $D_Q \in \mathcal{M}_q(\mathbb{R}^q)$ dont les éléments diagonaux sont les marges des variables P et Q , *i.e.* les probabilités d'avoir une caractéristique p_i ou q_j respectivement. Ce qui nous donne

$$D_P = \begin{pmatrix} f_{1\cdot} & & \\ & \ddots & \\ & & f_{p\cdot} \end{pmatrix} \quad D_Q = \begin{pmatrix} f_{\cdot 1} & & \\ & \ddots & \\ & & f_{\cdot q} \end{pmatrix}.$$

Nous sommes alors capable d'exprimer la matrice L en fonction des matrices D_P et F . En effet nous avons vu que L se composait des éléments de la matrice F divisés par les marges de chaque lignes $f_{i\cdot}$. On a donc

$$L = D_P^{-1} F.$$

On peut également faire de même avec la distance du \mathcal{X}^2 entre deux **profils lignes**. On rappelle que cette dernière est la somme pondérée des différences au carrées entre deux profils lignes \mathbf{l}_a et \mathbf{l}_b , où la pondération est donnée par **les marges des colonnes** sur chaque éléments des profils lignes :

$$d_{\mathcal{X}^2}^2(\mathbf{l}_a, \mathbf{l}_b) = \langle \mathbf{l}_a - \mathbf{l}_b, \mathbf{l}_a - \mathbf{l}_b \rangle_{D_Q^{-1}} = (\mathbf{l}_a - \mathbf{l}_b) D_Q^{-1} (\mathbf{l}_a - \mathbf{l}_b).$$

Réduction de la dimension. Dans la Section 9, nous avons vu comment effectuer la recherche de valeurs propres et de vecteurs propres lorsque l'on pondère les exemples à l'aide d'une matrice W mais aussi lorsque l'on emploie une métrique différente M . Ces deux éléments servant à définir la notion d'inertie du nuage de points étudié. Dans ce cas, nous devons chercher les valeurs propres et les vecteurs propres de la matrice

$$X^T W X M.$$

Dans le cadre de l'AFC, c'est la matrice D_Q^{-1} qui va jouer le rôle de métrique et la matrice D_P qui joue le rôle de la matrice de pondération.

Dans ce cas, on peut montrer⁸ que les vecteurs propres $\mathbf{u}_1, \dots, \mathbf{u}_s$ peuvent être obtenus en diagonalisant la matrice $S \in \mathcal{M}_q(\mathbb{R})$ définie par

$$S = F^T D_P^{-1} F D_Q^{-1},$$

dont le terme général est donné par

8. Il faudrait reprendre tout le processus effectué à la Section 9 pour aboutir au résultat annoncé.

$$\forall i, j, \quad s_{ij} = \sum_{k=1}^p = \frac{f_{ki}f_{kj}}{f_{k\cdot}f_{\cdot j}}.$$

Une fois que l'on a trouvé un vecteur propre \mathbf{u}_m (appelé aussi *axe factoriel* dans ce contexte), on projette les profils lignes sur celui-ci pour obtenir les coordonnées (ou composantes) des profils dans l'espace engendré par \mathbf{u}_m . On prendra bien évidemment garde à effectuer cette projection au sens la métrique D_Q^{-1} employée !

Ainsi, si l'on note $\mathbf{p}_{\mathbf{u}_m}(L) \in \mathbb{R}^p$ la projection des profils lignes sur l'axe \mathbf{u}_m , nous avons⁹

$$\begin{aligned} \mathbf{p}_{\mathbf{u}_m}(F) &= \underbrace{D_P^{-1} F D_Q^{-1}}_{\downarrow \text{ par définition de } L} \mathbf{u}_m, \\ &= L D_Q^{-1} \mathbf{u}_m, \end{aligned}$$

dont une composante de ce vecteur est donnée par

$$p_{\mathbf{u}_m}(F)_k = \langle \mathbf{l}_k, \mathbf{u}_m \rangle_{D_Q^{-1}} = \mathbf{l}_k^T D_Q^{-1} \mathbf{u}_m = \sum_{j=1}^q \frac{f_{kj}}{f_{k\cdot}} \frac{1}{f_{\cdot j}} (u_m)_j.$$

Remarques. Dans cette présentation, nous avons fait le choix de ne pas centrer le nuage des profils lignes, mais nous aurions pu ! Cela ne change pas les résultats en pratique.

Lorsque l'on pratique l'AFC sur le *nuage non centré*, le premier axe factoriel possède les propriétés suivantes :

- Il relie l'origine du repère $\mathbf{0}$ au barycentre $\bar{\mathbf{I}}$
- Les projection des profils lignes sur cet axe se retrouvent toutes en $\bar{\mathbf{I}}$
- L'inertie des profils lignes projetés sur cet axe vaut 1.

Les autres axes factoriels sont ensuite identiques que l'on centre ou non le nuage des profils lignes. **On gardera à l'esprit que le premier axe factoriel est donc un axe trivial qui n'apporte pas d'information, on aura donc tendance à l'enlever en pratique.**

9. Le lecteur est invité à se référer à la toute fin de la Section 9 en cas de doute.

10.2.2 Application aux profils colonnes

Tout ce que nous avons introduit précédemment pour les profils lignes s'applique également aux profils colonnes. En fait, les lignes et colonnes de F jouent cette fois-ci un rôle symétrique (ce sont juste deux variables différentes), ce qui n'était pas le cas avec l'ACP.

On peut donc définir q profils colonnes $\mathbf{c}_1, \dots, \mathbf{c}_q$ au sein d'une matrice de données C de dimension $p \times q$ dont le terme général est défini par

$$c_{ij} = \frac{f_{ij}}{f_{\cdot j}}.$$

Ainsi

$$C = \begin{pmatrix} \mathbf{c}_1 & \dots & \mathbf{c}_j & \dots & \mathbf{c}_q \\ \frac{f_{11}}{f_{\cdot 1}} & \dots & \frac{f_{1j}}{f_{\cdot j}} & \dots & \frac{f_{1q}}{f_{\cdot q}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{i1}}{f_{\cdot 1}} & \dots & \frac{f_{ij}}{f_{\cdot j}} & \dots & \frac{f_{iq}}{f_{\cdot q}} \\ \vdots & & \vdots & & \vdots \\ \frac{f_{p1}}{f_{\cdot 1}} & \dots & \frac{f_{pj}}{f_{\cdot j}} & \dots & \frac{f_{pq}}{f_{\cdot q}} \\ 1 & \dots & 1 & \dots & 1 \end{pmatrix}$$

où on rappelle que $f_{\cdot j} = \sum_{i=1}^p f_{ij}$ et \mathbf{c}_j est un vecteur de \mathbb{R}^p tel que $\mathbf{c}_j = \left(\frac{f_{1j}}{f_{\cdot j}}, \dots, \frac{f_{ij}}{f_{\cdot j}}, \dots, \frac{f_{pj}}{f_{\cdot j}} \right)$. De plus tous les éléments des \mathbf{c}_j somment à 1.

On pourra également définir une distance du χ^2 entre deux profils colonnes \mathbf{c}_a et \mathbf{c}_b

$$d_{\chi^2}^2(\mathbf{c}_a, \mathbf{c}_b) = \sum_{i=1}^p \frac{1}{f_{i\cdot}} \left(\frac{f_{ia}}{f_{\cdot a}} - \frac{f_{ib}}{f_{\cdot b}} \right)^2.$$

Ici, chaque entrée d'un vecteur \mathbf{c} est pondérée par un facteur $\frac{1}{f_{i\cdot}}$. Ce poids dépend donc directement de la valeur de la marge de chaque ligne de la matrice F .

Ecriture matricielle. Procédons comme pour les profils lignes et regardons comment écrire tout cela à l'aide de matrices.

Tout comme les profils lignes, on peut noter que chaque profil colonne est pondéré par sa marge $f_{\cdot j}$, ce qui conduit à la définition suivante du barycentre $\bar{\mathbf{c}} \in \mathbb{R}^p$ des profils colonnes

$$\begin{aligned}\bar{\mathbf{c}} &= \sum_{j=1}^q f_{\cdot j} \mathbf{c}_j, \\ &= \left(\sum_{j=1}^q f_{\cdot j} \frac{f_{1j}}{f_{\cdot j}}, \dots, \sum_{j=1}^q f_{\cdot j} \frac{f_{ij}}{f_{\cdot j}}, \dots, \sum_{j=1}^q f_{\cdot j} \frac{f_{pj}}{f_{\cdot j}} \right), \\ &= (f_{1\cdot}, \dots, f_{i\cdot}, \dots, f_{p\cdot}).\end{aligned}$$

Le barycentre est cette fois égale à la marge des modalités de la variable P .

Exemple 10.6. *Considérons toujours notre même exemple dont on rappelle la matrice F*

$$F = \begin{array}{ccccc} & \text{bleu} & \text{vert} & \text{marron} & \\ \text{chat} & 0.18 & 0.14 & 0.05 & 0.37 \\ \text{chien} & 0.01 & 0.09 & 0.36 & 0.46 \\ \text{souris} & 0.01 & 0.02 & 0.03 & 0.06 \\ \text{lézard} & 0.05 & 0.03 & 0.03 & 0.11 \\ & 0.25 & 0.28 & 0.47 & 1 \end{array}.$$

A l'aide de cette matrice F , on peut déterminer la matrice C ainsi que le barycentre des profils colonnes $\bar{\mathbf{c}}$

$$L = \begin{array}{ccccc} & \text{bleu} & \text{vert} & \text{marron} & \bar{\mathbf{c}} \\ & 0.72 & 0.50 & 0.11 & 0.37 \\ & 0.04 & 0.32 & 0.77 & 0.46 \\ & 0.04 & 0.07 & 0.06 & 0.06 \\ & 0.20 & 0.11 & 0.06 & 0.11 \\ & 1 & 1 & 1 & \end{array}.$$

Cette matrice C peut être entièrement déterminée à l'aide du calcul matriciel. Pour cela introduisons les matrices diagonales $D_P \in \mathcal{M}_p(\mathbb{R}^p)$ et $D_Q \in \mathcal{M}_p(\mathbb{R}^q)$ dont les éléments diagonaux sont les marges des variables P et Q , *i.e.* les probabilités d'avoir une caractéristique p_i ou q_j respectivement. Ce qui nous donne

$$D_P = \begin{pmatrix} f_{1\cdot} & & \\ & \ddots & \\ & & f_{p\cdot} \end{pmatrix} \quad D_Q = \begin{pmatrix} f_{\cdot 1} & & \\ & \ddots & \\ & & f_{\cdot q} \end{pmatrix}.$$

Nous sommes alors capable d'exprimer la matrice C en fonction des matrices D_Q et F . En effet nous avons vu que C se composait des éléments de la matrice F divisés par les marges de chaque colonne $f_{\cdot j}$. On a donc

$$C^T = D_Q^{-1} F^T.$$

On peut également faire de même avec la distance du χ^2 entre deux **profils colonnes**. On rappelle que cette dernière est la somme pondérée des différences au carrées entre deux profils lignes \mathbf{c}_a et \mathbf{c}_b , où la pondération est donnée par **les marges des lignes** sur chaque éléments des profils colonnes :

$$d_{\chi^2}^2(\mathbf{c}_a, \mathbf{c}_b) = \langle \mathbf{c}_a - \mathbf{c}_b, \mathbf{c}_a - \mathbf{c}_b \rangle_{D_P^{-1}} = (\mathbf{c}_a - \mathbf{c}_b) D_P^{-1} (\mathbf{c}_a - \mathbf{c}_b).$$

Réduction de la dimension. A nouveau, on se rappelle le travail effectué à la Section 9 et on peut montrer que chercher les axes factoriels $\mathbf{u}'_1, \dots, \mathbf{u}'_s$, revient à diagonaliser¹⁰ matrice carrée T de taille p définie par :

$$T = F D_Q^{-1} F^T D_P^{-1},$$

dont le terme général est donné par :

$$\forall i, j, \quad t_{ij} = \sum_{k=1}^q \frac{f_{ik} f_{kj}}{f_{\cdot k} f_{j \cdot}}.$$

En effet, dans le contexte de l'AFC appliquée sur les **profils colonnes**, c'est la matrice D_P^{-1} qui fait office de métrique et la matrice D_Q^{-1} joue le rôle de matrice de pondération.

Si on compare la matrice à diagonaliser dans le cas de l'étude des profils colonnes comparée à l'étude des profils lignes. On voit que l'on a simplement échangé le rôle des matrices D_Q et D_P et que l'on considère la matrice transposée de F pour passer d'une situation à une autre.

Une fois que l'on a trouvé un vecteur propre ou *axe factoriel* \mathbf{u}'_m , on projette les profils colonnes sur celui-ci pour obtenir les coordonnées (ou composantes) des profils

10. En cas de doute, il suffit d'appliquer la démarche présentée à la Section 9 pour obtenir le résultat. On prendra garde à la pondération des données D_Q^{-1} mais aussi à la métrique D_P^{-1} employée.

dans l'espace engendré par \mathbf{u}'_m . On prendra bien évidemment garde à effectuer cette projection au sens la métrique D_P^{-1} employée !

Ainsi, si l'on note $\mathbf{p}_{\mathbf{u}'_m}(C) \in \mathbb{R}^q$ la projection des profils colonnes sur l'axe \mathbf{u}'_m , nous avons ¹¹

$$\begin{aligned}\mathbf{p}_{\mathbf{u}'_m}(F) &= \underbrace{D_Q^{-1} F^T D_P^{-1}}_{\downarrow \text{ par définition de } C} \mathbf{u}'_m, \\ &= C^T D_P^{-1} \mathbf{u}'_m,\end{aligned}$$

dont une composante de ce vecteur est donnée par

$$p_{\mathbf{u}'_m}(F)_k = \langle \mathbf{c}_k, \mathbf{u}'_m \rangle_{D_P^{-1}} = \mathbf{c}_k^T D_P^{-1} \mathbf{u}'_m = \sum_{i=1}^p \frac{f_{ik}}{f_{\cdot k}} \frac{1}{f_{\cdot i}} (u_m)_i.$$

Enfin, toutes les remarques effectuées à propos du centrage du nuage des profils lignes restent valables en ce qui concerne le nuage des profils colonnes.

10.3 Remarques et interprétations

On va finir cette étude l'AFC par quelques analyses, remarques et illustrations.

Dualité profils lignes et profils colonnes. Comme pour l'ACP, il est possible de faire le lien entre l'analyse des profils lignes et des profils colonnes dans le cadre de l'AFC.

Tout d'abord, on peut remarquer que les axes factoriels \mathbf{u}_m et \mathbf{u}'_m sont des vecteurs propres associés à la même valeur propre λ_m . On peut donc écrire des relations entre les axes factoriels \mathbf{u}_m et \mathbf{u}'_m à partir de cette relation. Plus précisément, nous avons

$$\mathbf{u}_m = \frac{1}{\sqrt{\lambda_m}} F^T D_P^{-1} \mathbf{u}'_m \quad \text{et} \quad \mathbf{u}'_m = \frac{1}{\sqrt{\lambda_m}} F D_Q^{-1} \mathbf{u}_m.$$

On notera l'intervention des métriques associées dans chacune des relations.

En effet, nous avons les relations suivantes

11. Le lecteur est invité à se référer à la toute fin de la Section 9 en cas de doute.

$$T\mathbf{u}'_m = FD_Q^{-1}F^TD_P^{-1}\mathbf{u}'_m = \lambda_m\mathbf{u}'_m \quad \text{et} \quad S\mathbf{u}_m = F^TD_P^{-1}FD_Q^{-1}\mathbf{u}_m = \lambda_m\mathbf{u}_m,$$

où \mathbf{u}_m et \mathbf{u}'_m sont des vecteurs de norme égale à 1 (par rapport aux métriques D_Q^{-1} et D_P^{-1} respectivement). En pré-multipliant la première équation par $F^TD_P^{-1}$ on a

$$F^TD_P^{-1}FD_Q^{-1} \underbrace{F^TD_P^{-1}\mathbf{u}'_m}_{\text{vecteur propre de } S} = \lambda_m \underbrace{F^TD_P^{-1}\mathbf{u}'_m}_{\text{vecteur propre de } S},$$

or \mathbf{u}_m est un vecteur propre S associé à la même valeur propre λ_m , donc $F^TD_P^{-1}\mathbf{u}'_m$ est aussi un vecteur propre de S mais celui-ci n'est pas de norme égale à 1. Sa norme¹² est égale à

$$\begin{aligned} \|F^TD_P^{-1}\mathbf{u}'_m\|_{D_Q^{-1}}^2 &= (F^TD_P^{-1}\mathbf{u}'_m)^T D_Q^{-1} (F^TD_P^{-1}\mathbf{u}'_m), \\ &\quad \downarrow \text{on développe} \\ &= \mathbf{u}'_m D_P^{-1} \underbrace{FD_Q^{-1}F^TD_P^{-1}}_{\lambda_m} \mathbf{u}'_m, \\ &\quad \downarrow \mathbf{u}'_m \text{ est une valeur propre de } T \\ &= \mathbf{u}'_m D_P^{-1} \lambda_m \mathbf{u}'_m. \\ &= \lambda_m \underbrace{\mathbf{u}'_m D_P^{-1} \mathbf{u}'_m}_{1} \\ &\quad \downarrow \text{définition de la } D_P^{-1} \text{ - norme} \\ &= \lambda_m \|\mathbf{u}'_m\|, \\ &\quad \downarrow \mathbf{u}'_m \text{ est un vecteur unitaire pour cette norme} \\ &= \lambda_m. \end{aligned}$$

On a donc bien

$$\mathbf{u}_m = \frac{1}{\sqrt{\lambda_m}} F^TD_P^{-1}\mathbf{u}'_m.$$

On obtient la deuxième relation en procédant de façon analogue.

On pourra également faire le lien entre les coordonnées sur les différents axes principaux (correspondance entre les composantes principales). Pour cela, rappelons que $\mathbf{p}_{\mathbf{u}_m}(F) = D_P^{-1}FD_Q^{-1}\mathbf{u}_m$ et $\mathbf{p}_{\mathbf{u}'_m}(F) = D_Q^{-1}FD_P^{-1}\mathbf{u}'_m$. On en déduit la relation de dualité

12. on calcule la norme d'un vecteur propre associé aux profils lignes, la norme utilisée dépend donc de la métrique associée, c'est-à-dire D_Q^{-1}

suivante entre les composantes des profils lignes et colonnes sur les axes factoriels associés à la même valeur propre.

$$\mathbf{p}_{\mathbf{u}_m}(F) = \sqrt{\lambda_m} D_P^{-1} \mathbf{u}'_m \quad \text{et} \quad \mathbf{p}_{\mathbf{u}'_m}(F) = \sqrt{\lambda} D_Q^{-1} \mathbf{u}_m,$$

dont les i -ème et j -ème composantes sont respectivement données par

$$p_{\mathbf{u}_m}(F)_i = \sqrt{\lambda_m} \sum_{j=1}^q \frac{f_{ij}}{f_{i\cdot}} p_{\mathbf{u}'_m}(F)_j = \frac{1}{\sqrt{\lambda_m}} \mathbf{l}_i^T \mathbf{p}_{\mathbf{u}'_m}(F),$$

$$p_{\mathbf{u}'_m}(F)_j = \sqrt{\lambda_m} \sum_{i=1}^p \frac{f_{ij}}{f_{\cdot j}} p_{\mathbf{u}_m}(F)_i = \frac{1}{\sqrt{\lambda_m}} \mathbf{c}_j^T \mathbf{p}_{\mathbf{u}_m}(F).$$

Ces relations sont appelées, **relations de transitions ou barycentriques**.

Regardons maintenant l'expression $p_{\mathbf{u}_m}(F)_i = \sqrt{\lambda_m} \sum_{j=1}^q \frac{f_{ij}}{f_{i\cdot}} p_{\mathbf{u}'_m}(F)_j$ d'un peu plus près.

On remarque que la composante de d'un *profil ligne* $p_{\mathbf{u}_m}(F)_i$ sur l'axe \mathbf{u}_m s'exprime comme *une combinaison linéaire des composantes des profils colonnes* $p_{\mathbf{u}'_m}(F)_j$ où les coefficients de la combinaison linéaire sont les coordonnées du profil ligne \mathbf{l}_i .

On peut faire une interprétation analogue quant à $p_{\mathbf{u}'_m}(F)_j$. On remarque donc qu'il est très facile de passer de la composante des profils lignes aux composantes des profils colonnes sur un axe associé à une même valeur propre !

Propriétés. Regardons un peu les propriétés des composantes principales associées à une valeur propre λ_m . On pourra tout d'abord remarquer que les coordonnées sur un axe principale sont centrées ! Plus précisément, la somme pondérée des composantes principales par leurs fréquences relatives est nulle

$$\sum_{i=1}^p f_{i\cdot} p_{\mathbf{u}_m}(F)_i = 0 \quad \text{et} \quad \sum_{j=1}^q f_{\cdot j} p_{\mathbf{u}'_m}(F)_j = 0$$

(chercher pourquoi)

Comme les composantes sont centrées l'inertie ou la variance des nuages de points sur les axes \mathbf{u}_m et \mathbf{u}'_m sont respectivement

$$\sum_{i=1}^p f_{i\cdot} p_{\mathbf{u}_m}(F)_i^2 = \lambda_m \quad \text{et} \quad \sum_{j=1}^q f_{\cdot j} p_{\mathbf{u}'_m}(F)_j^2 = \lambda_m.$$

Nos nuages étant initialement non centrés, la première valeur propre est $\lambda_1 = 1$ et donc toutes les valeurs propres sont inférieures à 1. (pareil à expliquer pourquoi $\lambda_1 = 1$)

On rappelle également que l'inertie ou la variance peut être vue comme la moyenne pondérée des distances au carré au barycentre du nuage. On va donc faire le lien entre distance du \mathcal{X}^2 et variance totale (des profils lignes par exemple)

$$\begin{aligned} Var_{\text{tot}} &= \sum_{i=1}^p f_{i\cdot} d_{\mathcal{X}^2}^2(\mathbf{l}_i, \bar{\mathbf{l}}), \\ &\downarrow \text{définition de la distance du } \mathcal{X}^2 \\ &= \sum_{i=1}^p f_{i\cdot} \sum_{j=1}^q \frac{1}{f_{\cdot j}} \left(\frac{f_{ij}}{f_{i\cdot}} - f_{\cdot j} \right)^2, \\ &= \sum_{i=1}^p \sum_{j=1}^q \frac{1}{f_{\cdot j} f_{i\cdot}} (f_{ij} - f_{i\cdot} f_{\cdot j})^2. \end{aligned}$$

Cette inertie totale, notée parfois φ^2 , est une grandeur qui permet de mesurer l'intensité de la liaison entre deux variables qualitatives.¹³ Cette grandeur peut également être directement reliée aux valeurs propres λ_m

$$\varphi^2 = \sum_{m=2}^{\min(p,q)} \lambda_m.$$

Concernant les modalités. Il arrive parfois que certaines modalités soient peu représentées dans les données et qu'il soit parfois nécessaire de les agréger, c'est en en général ce que l'on fait lorsque deux modalités ont une sémantique très proche.

On peut montrer que si l'on agrège deux modalités d'une même variable **ayant des profils identiques** (la contrainte reste forte !) en une nouvelle modalité affectée de la somme de leurs masses, alors les distances entre les profils de cette variable restent inchangées ! C'est ce que l'on appelle **l'équivalence distributionnelle**.

Par exemple, si on considère deux profils lignes \mathbf{l}_a et \mathbf{l}_b avec des fréquences relatives f_a et f_b potentiellement différentes.

- \mathbf{l}_a et \mathbf{l}_b étant confondus dans l'espace \mathbb{R}^q , l'agrégation $\frac{\mathbf{l}_a + \mathbf{l}_b}{f_a + f_b}$ est également confondu en ces deux points et elle n'altère donc en rien le nuage de points.

13. On peut également relier cette quantité à la statistique du \mathcal{X}^2 qui permet de tester la liaison entre deux variables qualitatives. Mais nous ne détaillerons pas tout cela car cela fera l'objet d'études plus poussées en cours de Statistiques en L3.

- cette même modification ne change pas non plus les distances (en terme de χ^2) entre les profils colonnes.

Exemple 10.7. *Considérons deux variables P et Q désignant respectivement les catégories socioprofessionnelles (CSP) et les départements.*

Parmi les différentes modalités de P on se concentre sur les ouvriers qualifiés et les ouvriers non qualifiés. Supposons que ces deux profils lignes sont très similaires, donc avec des composantes proches, ce qui signifie que la répartition au sein des différents départements est quasi-identique.

Il est alors possible d'agréger les deux modalités en une seule, que l'on pourrait nommer ouvrier, sans changer drastiquement les résultats de l'AFC.

Interprétation. On continue la présentation de l'AFC par quelques interprétation sur les valeurs propres et quelques illustrations.

Comme pour l'ACP, les valeurs propres nous renseignent sur l'inertie du nuage projeté sur les axes. Nous avons déjà vu que l'inertie du nuage des profils lignes sur l'axe \mathbf{u}_m est égale à λ_m (de même pour l'inertie du nuage des profils colonnes sur l'axe \mathbf{u}'_m).

On peut alors définir le taux d'inertie d'un axe \mathbf{u}_m par

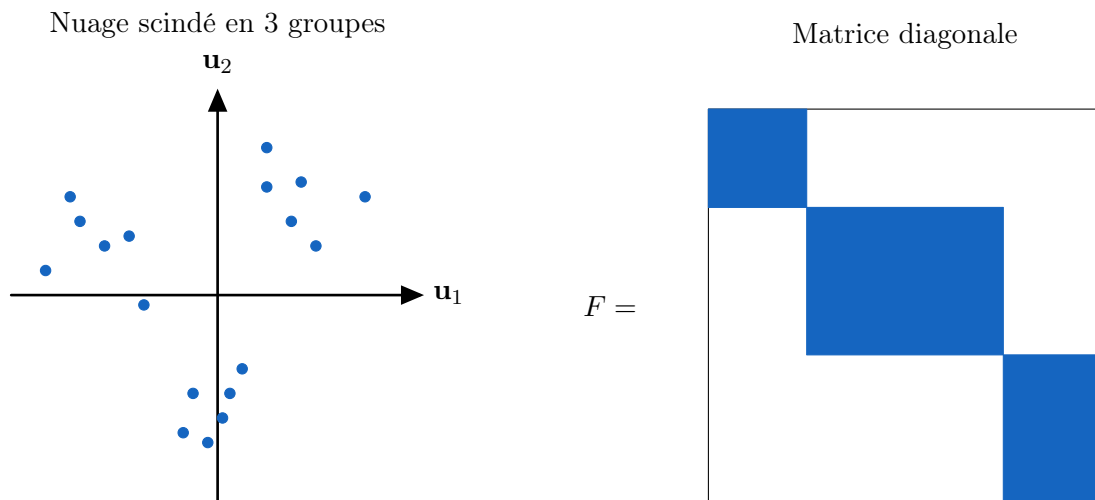
$$\frac{\lambda_m}{\varphi^2} = \frac{\lambda_m}{\sum_{k=2}^{\min(p,q)} \lambda_k}.$$

L'inertie d'un axe mesure l'association entre des modalités des deux variables qu'il met en évidence. Cette valeur ne peut pas dépasser 1. Si elle est proche 1 cela signifie que l'axe permet de séparer l'ensemble des modalités d'une variable en deux sous-groupes. En revanche, si toutes les valeurs propres sont proches de 1, alors chaque modalité d'une variable est en correspondance presque exclusive avec une modalité d'une autre variable.

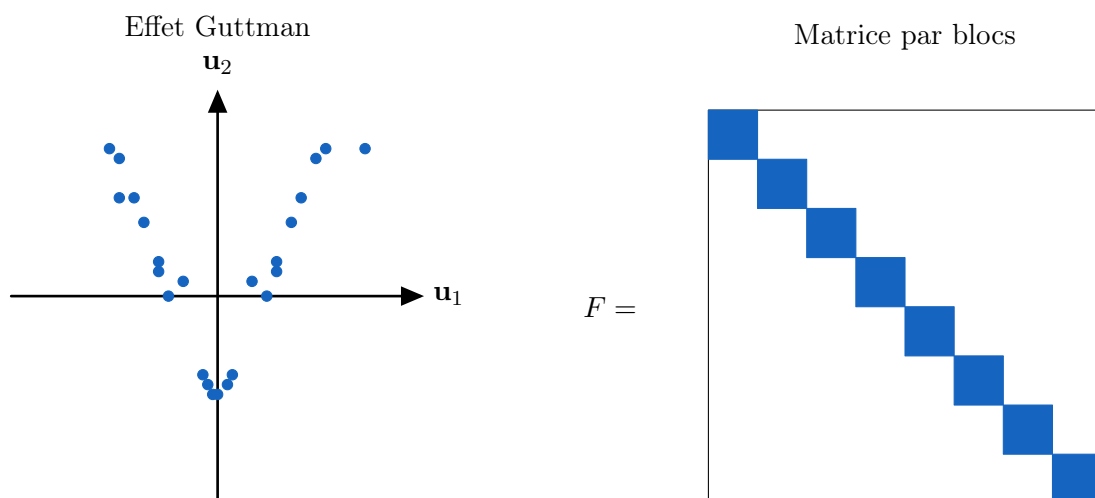
Cela montre que les valeurs propres peuvent nous renseigner sur la forme du nuage projeté ! On dénombre quelques cas ci-dessous :

- imaginons que le nuage est **scindé en deux sous-groupes**, la matrice F peut être réorganisée en diagonale par blocs en ordonnant les coordonnées des lignes vis-à-vis des colonnes de leurs premiers facteurs respectifs. C'est le cas correspondant aux valeurs propres très proche de 1.
- On peut généraliser cela en **nuage scindé en k sous-groupes**, la matrice F peut alors également être réorganisée en diagonales par blocs en permutant les

lignes et les colonnes. Cela signifie que k valeurs propres sont proches de 1.



- **L'effet Guttman** (à revérifier), le nuage de points a une **forme parabolique**. La matrice F peut être réorganisée sous forme **diagonale** qui est l'illustration typique de cet effet. Dans ce cas, le premier axe oppose des modalités opposées tandis que le deuxième permet de séparer les valeurs intermédiaires des valeurs extrêmes.



Qualité de la représentation. On finit cette présentation de l'AFC par quelques remarques sur les qualités de représentation des individus, leurs contributions et en présentant comment ajouter des individus supplémentaires.

On peut, comme pour l'ACP, juger de la qualité de la représentation d'un profil

ligne ou colonne sur un axe factoriel donné par les quantités suivantes :

- **Pour un profil ligne l_i sur l'axe u_m** , la qualité de la représentation du profil sur cet axe est donnée par

$$\frac{(p_{u_m}(F)_i)^2}{\sum_{j=1}^q \frac{1}{f_{\cdot j}} \left(\frac{f_{ij}}{f_{i\cdot}} - f_{\cdot j} \right)^2}.$$

- **Pour un profil colonne c_j sur l'axe u'_m** , la qualité de la représentation du profil sur cet axe est donnée par

$$\frac{(p_{u'_m}(F)_j)^2}{\sum_{i=1}^p \frac{1}{f_{i\cdot}} \left(\frac{f_{ij}}{f_{\cdot j}} - f_{i\cdot} \right)^2}.$$

Ainsi, pour une analyse des profils, on va privilégier ceux dont la représentation est de bonne qualité!

On pourra également évaluer la contribution des différents profils aux différents axes factoriels :

- **Pour un profil ligne l_i à l'axe u_m** , la contribution du profil ligne à cet axe est donnée par

$$f_{i\cdot} \frac{(p_{u_m}(F)_i)^2}{\lambda_m}.$$

- **Pour un profil colonne c_j à l'axe u'_m** , la contribution du profil colonne à cet axe est donnée par

$$f_{\cdot j} \frac{(p_{u'_m}(F)_j)^2}{\lambda_m}.$$

On en déduit, puisque la somme des contribution de l'ensemble des profils lignes est égale à 1 (même chose pour les profils colonnes), que

$$\sum_{i=1}^p f_{i\cdot} (p_{u_m}(F)_i)^2 = \lambda_m \quad \text{et} \quad \sum_{j=1}^q f_{\cdot j} (p_{u'_m}(F)_j)^2 = \lambda_m.$$

Ainsi, pour interpréter un axe, on va privilégier les profils qui contribuent fortement à la définition de cet axe.

Enfin, quid des données supplémentaires. En effet, il n'est pas rare de vouloir projeter des éléments supplémentaires dans les espaces réduits afin d'enrichir notre analyse.

Pour cela, supposons que l'on dispose de q_+ colonnes supplémentaires qui correspondent à q_+ modalités en plus de la variable Q que l'on souhaite projeter. Ces q colonnes

supplémentaires sont aussi décrites par des vecteurs de taille p . Ce tableau supplémentaire sera appelé C_+ .



Nouvelle table des fréquences relatives : ajout de profils colonnes

Les éléments de la table C_+ se calculent de la même façon que les éléments de la table C , elle a donc pour terme général

$$c_{+,ij} = \frac{n_{+,ij}}{n_{+,j}},$$

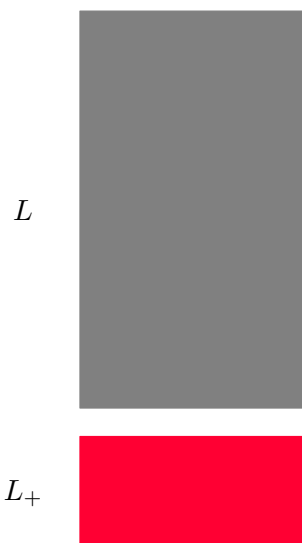
où $n_{+,ij}$ représente le nombre d'individus possédant la modalité p_i et la nouvelle modalité $q_{+,j}$ et $n_{+,j}$ représente le nombre d'individus possédant la nouvelle modalité $q_{+,j}$.

Pour projeter ces nouveaux profils colonnes sur l'axe \mathbf{u}'_m , nous pouvons utiliser la relation décrivant le lien entre les composantes principales d'un profil colonne sur \mathbf{u}'_m et celle des profils lignes sur l'axe \mathbf{u}_m , *i.e.*

$$p_{\mathbf{u}'_m}(C_+)_j = \frac{1}{\sqrt{\lambda_m}} \sum_{i=1}^p c_{+,ij} p_{\mathbf{u}_m}(F)_i,$$

où $p_{\mathbf{u}_m}(F)_i$ correspond à la projection des profils lignes sur l'axe \mathbf{u}_m .

De même si l'on dispose de modalités supplémentaires pour la variable P , nous aurons donc de nouvelles nouveaux profils lignes à calculer qui se trouveront dans une table L_+ .



Nouvelle table des fréquences relatives : ajout de profils lignes

Nous pouvons calculer les fréquences relatives, et donc les coordonnées des nouveaux profils lignes comme pour les profils colonnes

$$l_{+,ij} = \frac{n_{+,ij}}{n_{+,i}},$$

où $n_{+,ij}$ représente le nombre d'individus possédant la modalité q_j et la nouvelle modalité $p_{+,i}$ et $n_{+,i}$ représente le nombre d'individus possédant la nouvelle modalité $p_{+,i}$.

Ainsi, la projection de ces nouveaux profils lignes sur l'axe \mathbf{u}_m pourra se faire en utilisant la relation décrivant le lien entre les composantes principales d'un profil ligne sur l'axe \mathbf{u}_m et celle des profils colonnes sur l'axe \mathbf{u}'_m , *i.e.*

$$p_{\mathbf{u}_m}(L_+)_i = \frac{1}{\sqrt{\lambda_m}} \sum_{j=1}^q l_{+,ij} p_{\mathbf{u}'_m}(F)_j,$$

où $p_{\mathbf{u}'_m}(F)_j$ correspond à la projection des profils lignes sur l'axe \mathbf{u}'_m .

11 Analyse factorielle des Correspondances Multiples (ACM)

La section précédente a permis d'exposer les outils d'analyses nécessaires à l'étude de lien entre deux variables qualitatives via l'Analyse Factorielle des Correspondances. La dernière partie de ce document a pour objectif de généraliser l'AFC à l'étude de plusieurs variables qualitatives, c'est ce que l'on appelle l'**Analyse factorielle en Composantes Multiples** (ACM).

11.1 Préliminaires

L'ACM va nous permettre d'étudier une population de n individus décrits par un ensemble de p variables qualitatives nominales (ou catégorielles). Bien évidemment, chacune de ces variables qualitatives p_j présentera un certain nombre de modalités.

L'ACM est souvent employée pour analyser des réponses issues d'une enquête. Les questions de cette enquête se présentant comme des questions fermées avec une réponse possible parmi plusieurs. Dans ce cas, **une question correspond à une variable** tandis qu'**une réponse à une question donnée correspondra à une modalité de cette variable**.

Structure des données. En terme de présentation des données, nous disposons de p variables notée $\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_p$ qui sont des éléments de \mathbb{R}^n (ces variables correspondent aux questions) et enfin de n individus $\mathbf{o}_1, \dots, \mathbf{o}_i, \dots, \mathbf{o}_n$ qui sont des éléments de \mathbb{R}^p (ces individus correspondent aux personnes sondées au cours d'une enquête). On rappelle aussi que chaque variable \mathbf{v}_j prend ses valeurs dans un ensemble discret (la réponse fournie à une question).

On peut synthétiser tout cela dans une matrice de design N de taille $n \times p$

$$N = \begin{matrix} & \mathbf{v}_1 & \cdots & \mathbf{v}_j & \cdots & \mathbf{v}_p \\ \mathbf{o}_1 & \left(\begin{matrix} n_{11} & \cdots & n_{1j} & \cdots & n_{1p} \end{matrix} \right. \\ \vdots & \left. \begin{matrix} \vdots & & \vdots & & \vdots \end{matrix} \right. \\ \mathbf{o}_i & \left(\begin{matrix} n_{i1} & \cdots & n_{ij} & \cdots & n_{ip} \end{matrix} \right. \\ \vdots & \left. \begin{matrix} \vdots & & \vdots & & \vdots \end{matrix} \right. \\ \mathbf{o}_n & \left(\begin{matrix} n_{n1} & \cdots & n_{nj} & \cdots & n_{np} \end{matrix} \right. \end{matrix}$$

où n_{ij} correspond à la réponse fournie par l'individu i à la question j .

Par exemple, si l'individu i a coché la première réponse à la question j , alors $n_{ij} = 1$. S'il a coché la réponse 3, alors $n_{ij} = 3$.

Les valeurs n'ont pas de sens à proprement parlé, il serait donc absurde de chercher un sens à la valeur numérique. Ces valeurs correspondent à un pré-traitement des données (ou des questionnaires) : on parle plus **d'encodage des réponses**

Exemple 11.1. *Intéressons nous à trois variables qualitatives*

- le **genre** $\mathbf{v}_1 \in \{H, F\}$,
- la **tranche d'âge** \mathbf{v}_2 dont les réponses possibles sont : moins de 25 ans - entre 25 et 44 ans - entre 45 et 64 ans ou 65 ans et plus,
- le **statut civil** \mathbf{v}_3 dont les modalités sont : célibataire - marié(e) ou divorcé(e).

Il est d'usage d'encoder les réponses avec des valeurs numériques. Ainsi, pour une question ayant 4 réponses possibles, on numérotera les choix de 1 à 4.

On aboutit à une matrice de données du type :

$$N = \begin{matrix} & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \\ \mathbf{o}_1 & \begin{pmatrix} 1 & 4 & 3 \end{pmatrix} \\ \mathbf{o}_2 & \begin{pmatrix} 2 & 4 & 3 \end{pmatrix} \\ \mathbf{o}_3 & \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \\ \mathbf{o}_4 & \begin{pmatrix} 1 & 2 & 1 \end{pmatrix} \\ \mathbf{o}_5 & \begin{pmatrix} 2 & 3 & 2 \end{pmatrix} \\ \mathbf{o}_6 & \begin{pmatrix} 2 & 2 & 2 \end{pmatrix} \\ \mathbf{o}_7 & \begin{pmatrix} 1 & 3 & 1 \end{pmatrix} \end{matrix}$$

Les valeurs sur ces tables n'ayant pas de sens en soit, nous ne pouvons donc pas calculer de grandeurs statistiques dessus comme une moyenne ou une variance. Cela ne veut pour autant pas dire que l'on ne peut pas traiter ces données.

Codage. Comme les valeurs numériques n'ont pas de sens, il est d'usage de passer par une phase dite de **codage** des réponses¹⁴. Cette phase de codage consiste à projeter chaque variable dans un espace de dimension égale au nombre de modalités de cette variable, *i.e.* au nombre de réponses possibles à la question.

Une variable \mathbf{v}_j admettant k_j modalités peut ainsi être plongé dans un espace de dimension \mathbb{R}^k . Ainsi, pour chaque variable \mathbf{v}_j admettant k_j modalités et dont le nombre, il est possible de transformer notre vecteur de \mathbb{R}^n des réponses en matrice Z_j de taille $n \times k_j$ par une binarisation des réponses, *i.e.*

$$(Z_j)_{ik} = \begin{cases} 1 & \text{si l'individu } i \text{ a coché la } k\text{-ème réponse à la question } j, \\ 0 & \text{sinon.} \end{cases}$$

Exemple 11.2. *Reprenons notre exemple précédent et concentrons nous sur la deuxième variable \mathbf{v}_2 correspondant à la tranche d'âge des répondants. Il y avait 4 réponses possibles numérotées de 1 à 4.*

La binarisation des réponses se fait de la façon suivante

14. Attention, cette étape n'est pas à confondre avec la phase d'**encodage** des réponses qui revient simplement à attribuer une valeur numérique à la réponse d'une question donnée !

$$\begin{aligned}
1 &\rightarrow (1, 0, 0, 0), \\
2 &\rightarrow (0, 1, 0, 0), \\
3 &\rightarrow (0, 0, 1, 0), \\
4 &\rightarrow (0, 0, 0, 1).
\end{aligned}$$

Ainsi, on passe d'un vecteur de réponses de \mathbb{R}^7 à la matrice binaire de taille 7×4 suivante :

$$\mathbf{v}_2 = \begin{pmatrix} 4 \\ 4 \\ 1 \\ 2 \\ 3 \\ 2 \\ 3 \end{pmatrix} \rightarrow Z_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

On peut également voir cette grande matrice comme une juxtaposition de vecteurs binaires $(Z_j)_k$, k_j vecteurs exactement.

$$Z_j = ((Z_j)_1, \dots, (Z_j)_{k_1}, \dots, (Z_j)_{k_j}).$$

Exemple 11.3. Sur notre matrice Z_2 précédente, cela nous donne

$$Z_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \rightarrow (Z_2)_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, (Z_2)_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, (Z_2)_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ et } (Z_2)_4 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Finalement, pour représenter l'ensemble des modalités de l'ensemble des variables a sein d'un même tableau ou d'une même matrice, on peut faire le choix de concaténer l'ensemble des vecteurs binaires $(Z_j)_k$ ou encore l'ensemble des matrices Z_j .

Cela permet de définir une seule et même matrice que l'on nomme **tableau disjonctif complet** qui sera alors de taille $n \times q$ où $q = \sum_{j=1}^p k_j$.

$$Z = (Z_1, \dots, Z_p) = ((Z_1)_1, \dots, (Z_1)_{k_1}, \dots, (Z_p)_1, \dots, (Z_p)_{k_p}).$$

Exemple 11.4. *Toujours sur notre même exemple, aboutissons au tableau disjonctif suivant :*

$$N = \begin{matrix} & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \\ \mathbf{o}_1 & \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 2 \\ 2 \\ 1 \end{pmatrix} & \begin{pmatrix} 4 \\ 4 \\ 1 \\ 2 \\ 3 \\ 2 \\ 3 \end{pmatrix} & \begin{pmatrix} 3 \\ 3 \\ 1 \\ 1 \\ 2 \\ 2 \\ 1 \end{pmatrix} \end{matrix} \rightarrow Z = \begin{matrix} & (Z_1)_1 & (Z_1)_2 & (Z_2)_1 & (Z_2)_2 & (Z_2)_3 & (Z_2)_4 & (Z_3)_1 & (Z_3)_2 & (Z_3)_3 \\ \mathbf{o}_1 & \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \end{matrix}$$

Avec un tel codage, un individu \mathbf{o}_i est finalement représenté par un vecteur binaire de taille q , plus précisément $\mathbf{o}_i \in \{0, 1\}^q$. De plus, pour chaque variable \mathbf{v}_j , un individu ne pourra prendre qu'une et une seule modalité. Pour chaque ligne i , de Z_j , nous avons alors

$$\sum_{l=1}^{k_j} (Z_j)_{l,i} = 1,$$

où $(Z_j)_{l,i}$ désigne la l -ème modalité de la variable j pour l'individu i . Chaque modalité k d'une variable \mathbf{v}_j est aussi représentée par un vecteur binaire $(Z_j)_k$ dont la somme des éléments représente le nombre d'individus ayant la modalité k pour la variable j . Enfin, pour chaque variable \mathbf{v}_j , la somme de tous les termes vérifie :

$$\sum_{i=1}^n \sum_{l=1}^{k_j} (Z_j)_{l,i} = n.$$

De ces propriétés précédentes sur les tableaux Z_j nous pouvons également déduire quelques propriétés sur le tableau disjonctif Z . Par exemple, les marges des lignes \mathbf{o}_i de Z , sont telles que pour tout $i \in 1, \dots, n$:

$$\sum_{l=1}^q (o_i)_l = o_{i\cdot} = p,$$

c'est-à-dire que les marges de chaque ligne sont égales au nombre de variables ou encore au nombre de questions.

Les marges sur les colonnes, notée $Z_{\cdot j}$ de la matrice Z sont égales à la fréquence de la modalité k d'une variable donnée. Enfin, la somme de tous les éléments de Z est égale à np , *i.e.* le nombre de d'individus multiplié par le nombre de questions.

Exemple 11.5. Notre tableau disjonctif auquel on ajoute les marges devient alors

	$(Z_1)_1$	$(Z_1)_2$	$(Z_2)_1$	$(Z_2)_2$	$(Z_2)_3$	$(Z_2)_4$	$(Z_3)_1$	$(Z_3)_2$	$(Z_3)_3$	
\mathbf{o}_1	1	0	0	0	0	1	0	0	1	$o_{1.} = 3$
\mathbf{o}_2	0	1	0	0	0	1	0	0	1	$o_{2.} = 3$
\mathbf{o}_3	1	0	1	0	0	0	1	0	0	$o_{3.} = 3$
\mathbf{o}_4	1	0	0	1	0	0	1	0	0	$o_{4.} = 3$
\mathbf{o}_5	0	1	0	0	1	0	0	1	0	$o_{5.} = 3$
\mathbf{o}_6	0	1	0	1	0	0	0	1	0	$o_{6.} = 3$
\mathbf{o}_7	1	0	0	0	1	0	1	0	0	$o_{7.} = 3$
Z	$Z_{.1} = 4$	$Z_{.2} = 3$	$Z_{.3} = 1$	$Z_{.4} = 2$	$Z_{.5} = 2$	$Z_{.6} = 2$	$Z_{.7} = 3$	$Z_{.8} = 2$	$Z_{.9} = 2$	

Dans la suite, on ne va plus regarder les variables en soit mais on va plutôt se placer dans l'espace dans modalités.

11.2 Analyse factorielle des Correspondances Multiples

L'objectif de l'ACM est le même que celui de l'AFC ou de l'ACP, il s'agit toujours d'une méthode réduction de la dimension.

On pourra à nouveau se concentrer sur deux types de nuages de points : le **nuage des individus** où chaque ligne de la table Z est alors représenté par un vecteur de \mathbb{R}^q et le **nuage des modalités** (et non plus celui des variables) où chaque colonne de la table Z est alors représenté par un vecteur de \mathbb{R}^n .

On cherche des sous espaces de dimension réduite qui permettent de préserver au mieux l'inertie des nuages de points et dans lesquels nous sommes capables de tirer d'éventuelles des conclusions quant aux données observées. Comme pour l'AFC, la notion d'inertie sera définie au sens de la distance du \mathcal{X}^2

Nous verrons que l'ACM se présente comme une AFC mais sur un tableau disjonctif complet ! En ce sens, les tableaux étudiées dans les deux cas ayant des propriétés différentes, nous verrons que l'ACM possèdent ses propres particularités que nous comparerons à celles de l'AFC.

Préparation des données. Tout comme pour l'AFC, il est possible de voir la matrice Z comme une grande table de contingence où les individus sont représentés en lignes et les modalités en colonnes.

On pourra donc calculer une fréquence pour chacune des entrées de la matrice, cette dernière valant 0 ou $1/(np)$

$$F = \begin{matrix} & (Z_1)_1 & \cdots & (Z_p)_{k_p} \\ \mathbf{o}_1 & \left(\begin{array}{ccc} f_{11} & \cdots & f_{1q} \\ \vdots & & \vdots \\ f_{i1} & \cdots & f_{iq} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{nq} \end{array} \right) \end{matrix}$$

où $f_{ik} = \frac{n_{ik}}{np}$, $\forall (i, k) \in \llbracket 1, n \rrbracket \times \llbracket 1, q \rrbracket$. L'indice k parcourt donc l'ensemble des modalités de l'ensemble des variables et $n_{ik} = 0$ ou 1 selon que le i -ème individu ait sélectionné la k -ème modalité en guise de réponse. On va donc conserver ces notations pour la suite.

Exemple 11.6. Sur notre même exemple, le tableau des fréquences serait

$$F = \begin{matrix} & \mathbf{f}_1 & \mathbf{f}_2 & \mathbf{f}_3 & \mathbf{f}_4 & \mathbf{f}_5 & \mathbf{f}_6 & \mathbf{f}_7 & \mathbf{f}_8 & \mathbf{f}_9 \\ \mathbf{o}_1 & \left(\begin{array}{c|ccc|cc|cc} \frac{1}{21} & 0 & 0 & 0 & 0 & \frac{1}{21} & 0 & 0 & \frac{1}{21} \end{array} \right) & f_{1\cdot} = \frac{3}{21} \\ \mathbf{o}_2 & \left(\begin{array}{c|ccc|cc|cc} 0 & \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} & 0 & 0 & \frac{1}{21} \end{array} \right) & f_{2\cdot} = \frac{2}{21} \\ \mathbf{o}_3 & \left(\begin{array}{c|ccc|cc|cc} \frac{1}{21} & 0 & \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} & 0 & 0 \end{array} \right) & f_{3\cdot} = \frac{2}{21} \\ \mathbf{o}_4 & \left(\begin{array}{c|ccc|cc|cc} \frac{1}{21} & 0 & 0 & \frac{1}{21} & 0 & 0 & \frac{1}{21} & 0 & 0 \end{array} \right) & f_{4\cdot} = \frac{2}{21} \\ \mathbf{o}_5 & \left(\begin{array}{c|ccc|cc|cc} 0 & \frac{1}{21} & 0 & 0 & \frac{1}{21} & 0 & 0 & \frac{1}{21} & 0 \end{array} \right) & f_{5\cdot} = \frac{2}{21} \\ \mathbf{o}_6 & \left(\begin{array}{c|ccc|cc|cc} 0 & \frac{1}{21} & 0 & \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} & 0 \end{array} \right) & f_{6\cdot} = \frac{2}{21} \\ \mathbf{o}_7 & \left(\begin{array}{c|ccc|cc|cc} \frac{1}{21} & 0 & 0 & 0 & \frac{1}{21} & 0 & \frac{1}{21} & 0 & 0 \end{array} \right) & f_{7\cdot} = \frac{3}{21} \end{matrix}$$

où l'on renomme chaque colonne \mathbf{f}_j pour désigner le vecteur des fréquences de la modalité j . On pourra également définir la somme des fréquences d'un profil ligne \mathbf{o}_i qui sera toujours égale à $1/n$ pour toutes les lignes, cette somme est notée $f_{i\cdot}$. On peut aussi définir la somme des fréquences d'un profil colonne \mathbf{f}_j qui correspond au nombre de fois où la modalité est sélectionnée sur le nombre total de réponses $n \times p$, elle est notée $f_{\cdot j}$.

L'exemple nous motive à introduire deux nouvelles matrices D_O (faisant référence aux observations) et D_M (faisant références aux modalités), toutes deux diagonales et définies par

$$D_O = \begin{pmatrix} f_{1\cdot} & & & \\ & \ddots & & \\ & & f_{i\cdot} & \\ & & & \ddots \\ & & & & f_{p\cdot} \end{pmatrix} \quad \text{et} \quad D_M = \begin{pmatrix} f_{1\cdot} & & & \\ & \ddots & & \\ & & f_{i\cdot} & \\ & & & \ddots \\ & & & & f_{q\cdot} \end{pmatrix}$$

Etant donnée la matrice F précédemment définie, les deux matrices sont plus précisément égales à

$$D_O = \begin{pmatrix} \frac{1}{n} & & & \\ & \ddots & & \\ & & \frac{1}{n} & \\ & & & \ddots \\ & & & & \frac{1}{n} \end{pmatrix} \quad \text{et} \quad D_M = \begin{pmatrix} \frac{Z_{\cdot 1}}{np} & & & \\ & \ddots & & \\ & & \frac{Z_{\cdot j}}{np} & \\ & & & \ddots \\ & & & & \frac{Z_{\cdot q}}{np} \end{pmatrix}$$

On définit ensuite les profils \mathbf{l}_i comme étant le vecteur des fréquences des modalités divisé par le nombre de variables, *i.e.*

$$\mathbf{l}_i = \left(\frac{n_{i1}}{p}, \dots, \frac{n_{ik}}{p}, \dots, \frac{n_{iq}}{p} \right).$$

Si on note L l'ensemble de ces profils lignes, nous avons alors

$$L = D_O^{-1} F = \frac{Z}{p}.$$

Nous avons vu que les marges des lignes de Z sont toutes égales, cela revient donc à attribuer le même poids à l'ensemble des individus qui est égal $1/n$. La matrice D_O^{-1} joue donc le rôle de pondération des profils lignes. Ce qui signifie que la matrice D_M , et plus précisément D_M^{-1} va jouer le rôle de métrique dans le calcul de la distance du \mathcal{X}^2 .¹⁵ La distance du \mathcal{X}^2 entre deux profils lignes (ou deux individus) \mathbf{l}_a et \mathbf{l}_b s'écrit :

$$d_{\mathcal{X}^2}^2 = (\mathbf{l}_a - \mathbf{l}_b)^T D_M^{-1} (\mathbf{l}_a - \mathbf{l}_b) = \sum_{k=1}^q \frac{np}{Z_{\cdot k}} \left(\frac{n_{ak}}{p} - \frac{n_{bk}}{p} \right)^2.$$

15. On peut se référer au travail effectué dans le cadre de l'AFC pour faire l'analogie entre les deux.

11.2.1 Applications aux profils lignes.

On rappelle que notre objectif est de déterminer des sous-espaces de telle que à ce que la projection des profils lignes qui conservent au maximum l'inertie du nuage de points.

Comme précédemment, on va donc chercher une suite d'axes $\mathbf{u}_1, \dots, \mathbf{u}_s$, également appelés axes factoriels, de dimension q qui permettent de préserver l'inertie du nuage projeté.

On montre¹⁶ que cela se fait en diagonalisant la matrice $S \in \mathcal{M}_q(\mathbb{R})$ définie par :

$$S = F^T D_0^{-1} F D_M^{-1} = \frac{1}{np^2} Z^T Z D_M^{-1}.$$

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 S . Comme pour les méthodes précédentes, une fois que l'on a on trouvé l'axe factoriel \mathbf{u}_m , il ne reste qu'à projeter les individus sur ce dernier pour obtenir les composantes (*i.e.* les coordonnées) des individus dans le sous-espace engendré par \mathbf{u}_m . On prendra attention au fait que cette projection *orthogonale* doit se faire au sens de la métrique employée, *i.e.* D_M^{-1} .

Si on note $\mathbf{p}_{\mathbf{u}_m}(Z)$ la projection des individus sur l'axe \mathbf{u}_m , on a :

$$\begin{aligned} \mathbf{p}_{\mathbf{u}_m}(Z) &= \underbrace{D_0^{-1} F D_M^{-1}}_{\downarrow \text{lien entre les matrices } F \text{ et } Z} \mathbf{u}_m, \\ &= \underbrace{\frac{1}{p} Z D_M^{-1}}_{\downarrow \text{définition de } L} \mathbf{u}_m, \\ &= L D_M^{-1} \mathbf{u}_m. \end{aligned}$$

Ainsi, pour tout profil ligne \mathbf{l}_i , nous avons

$$(p_{\mathbf{u}_m})(\mathbf{l}_i) = \langle \mathbf{l}_i, \mathbf{u}_m \rangle_{D_M^{-1}} = \mathbf{l}_i^T D_M^{-1} \mathbf{u}_m = \sum_{k=1}^q \frac{n_{ik}}{p} \frac{np}{Z_{.k}} (u_m)_k.$$

11.2.2 Applications aux profils colonnes.

On peut appliquer le même développement que précédemment aux nuages des modalités. On prendra cependant garde au fait que la métrique ainsi que la matrice de

16. Il suffit de se rappeler le travail effectuée sur la généralisation de la SVD en Section 9.

pondération ne sont pas les mêmes que précédemment dans ce cas.

Les profils colonnes (*i.e.* les profils des modalités) \mathbf{c}_k sont définis comme étant les vecteurs de \mathbb{R}^n (au nombre de q , le nombre de modalités) dont les éléments sont définis par

$$\mathbf{c}_k = \left(\frac{n_{1k}}{Z_{\cdot k}}, \dots, \frac{n_{ik}}{Z_{\cdot k}}, \dots, \frac{n_{nk}}{Z_{\cdot k}} \right).$$

Les points colonnes sont obtenus par :

$$C^T = D_M^{-1} F^T = \frac{1}{np} D_M^{-1} Z^T.$$

La matrice des poids des modalités est alors donnée par la matrice D_M et la métrique associée est alors celle induite par la matrice $D_O^{-1} = nI_{n..}$. Connaissant cette métrique, on peut alors définir la distance du χ^2 entre deux profils colonnes \mathbf{c}_a et \mathbf{c}_b par

$$d_{\chi^2}(\mathbf{c}_a, \mathbf{c}_b) = (\mathbf{c}_a - \mathbf{c}_b) D_O^{-1} (\mathbf{c}_a - \mathbf{c}_b) = \sum_{i=1}^n \left(\frac{n_{ia}}{Z_{\cdot a}} - \frac{n_{ib}}{Z_{\cdot b}} \right)^2.$$

La suite des sous-espaces propres engendrés par les vecteurs $\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_s$ s'obtient cette fois-ci en diagonalisant la matrice $T \in \mathcal{M}_n(\mathbb{R})$ définie par

$$T = F D_M^{-1} F^T D_O^{-1} = \frac{1}{np^2} Z D_M^{-1} Z^T.$$

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 S . Comme pour les méthodes précédentes, une fois que l'on a on trouvé l'axe factoriel \mathbf{u}'_m , il ne reste qu'à projeter les profils colonnes sur ce dernier pour obtenir les composantes (*i.e.* les coordonnées) des modalités dans le sous-espace engendré par \mathbf{u}'_m . On prendra attention au fait que cette projection *orthogonale* doit se faire au sens de la métrique employée, *i.e.* D_O^{-1} .

Si on note $\mathbf{p}_{\mathbf{u}'_m}(Z)$ la projection des individus sur l'axe \mathbf{u}'_m , on a :

$$\begin{aligned} \mathbf{p}_{\mathbf{u}'_m}(Z) &= \underbrace{D_M^{-1} F^T D_O^{-1}}_{\downarrow \text{lien entre les matrices } F \text{ et } Z} \mathbf{u}'_m, \\ &= \frac{1}{p} Z D_M^{-1} \mathbf{u}'_m, \\ &\quad \downarrow \text{définition de } C \\ &= C^T D_O^{-1} \mathbf{u}'_m. \end{aligned}$$

Ainsi, pour tout profil colonne, ou toute modalité \mathbf{c}_k ,

$$(p_{\mathbf{u}'_m})(\mathbf{c}_k) = \langle \mathbf{c}_k, \mathbf{u}'_m \rangle_{D_O^{-1}} = \mathbf{c}_k^T D_O^{-1} \mathbf{u}'_m = n \sum_{i=1}^n \frac{n_{ik}}{Z_{\cdot k}} (u'_m)_i.$$

11.3 Remarques et interprétations

Nous finissons cette section par quelques remarques et propriétés de l'ACM.

Dualité. Comme pour les autres méthodes, il existe un lien entre l'analyse faite sur les profils lignes et celle faite sur les profils colonnes.

Rappelons tout d'abord que les axes factoriels \mathbf{u}_m et \mathbf{u}'_m sont associés à la même valeur propre λ_m et que

$$L = \frac{Z}{p} \quad \text{et} \quad C^T = \frac{1}{np} D_M^{-1} Z^T.$$

Il est donc facile d'obtenir les relations entre les coordonnées sur l'axe \mathbf{u}_m des profils lignes \mathbf{l}_i et les coordonnées des profils colonnes \mathbf{c}_k sur l'axe \mathbf{u}'_m :

$$\mathbf{p}_{\mathbf{u}_m}(Z) = \frac{1}{\sqrt{\lambda_m}} L \mathbf{p}_{\mathbf{u}'_m}(Z) = \frac{1}{p\sqrt{\lambda_m}} Z \mathbf{p}_{\mathbf{u}'_m}(Z).$$

De la même façon ¹⁷ nous avons aussi

$$\mathbf{p}_{\mathbf{u}'_m}(Z) = \frac{1}{\sqrt{\lambda_m}} C^T \mathbf{p}_{\mathbf{u}_m}(Z) = \frac{1}{np\sqrt{\lambda_m}} D_M^{-1} Z^T \mathbf{p}_{\mathbf{u}_m}(Z).$$

Les termes généraux (les coordonnées du vecteur) de ces deux relations sont donnés par

$$p_{\mathbf{u}_m}(\mathbf{l}_i) = \frac{1}{\sqrt{\lambda_m}} \sum_{k=1}^q \frac{n_{ik}}{Z_{i\cdot}} p_{\mathbf{u}'_m}(\mathbf{c}_k) = \frac{1}{\sqrt{\lambda_m}} \left(\frac{1}{p} \sum_{k:n_{ik}=1} p_{\mathbf{u}'_m}(\mathbf{c}_k) \right).$$

Première relation dans laquelle on peut voir, à un facteur $1/\sqrt{\lambda_m}$, que la coordonnée d'un profil ligne \mathbf{l}_i sur un axe factorielle \mathbf{u}_m est la **moyenne arithmétique** des coordonnées des modalités des profils \mathbf{c}_k sur l'axe \mathbf{u}'_m

17. Ces deux relations peuvent s'obtenir de la même façon que celles obtenues après analyse de l'AFC, on ne justifie donc pas les relations. Le lecteur peut se référer à la section précédente.

De la même façon, nous avons

$$p_{\mathbf{u}'_m}(\mathbf{c}_k) = \frac{1}{\sqrt{\lambda_m}} \sum_{i=1}^n \frac{n_{ik}}{Z_{\cdot k}} p_{\mathbf{u}_m}(\mathbf{l}_i) = \frac{1}{\sqrt{\lambda_m}} \left(\frac{1}{Z_{\cdot k}} \sum_{i:n_{ik}=1} p_{\mathbf{u}_m}(\mathbf{l}_i) \right),$$

où, à un facteur $1/\sqrt{\lambda_m}$, la coordonnée d'un profil ligne \mathbf{c}_k sur un axe factorielle \mathbf{u}'_m est la **moyenne arithmétique** des coordonnées des modalités des profils \mathbf{l}_i sur l'axe \mathbf{u}_m .

Propriétés. En ACM, il existe naturellement des dépendances linéaires entre les colonnes de la matrice Z . Ces colonnes représentent des modalités et sont définies au sein d'une matrice $C = \frac{1}{np} Z D_M^{-1}$ dont le rang est au plus $q - p + 1$, où q représente le nombre de modalités et p le nombre de variables. Par conséquent, la matrice S définie plus tôt possède cette même propriété.

Comme pour l'AFC, si on enlève la première valeur propre associées au centre de gravité du nuage de points, nous aurons en pratique, au plus $q - p$ valeurs propres non nulles.

On rappelle que le barycentre du nuage des modalités, est le vecteur $\bar{\mathbf{c}}$ des marges des lignes de la matrice F :

$$\bar{\mathbf{c}} = \left(\frac{1}{n}, \dots, \frac{1}{n} \right).$$

La distance du \mathcal{X}^2 entre un profil colonne \mathbf{c}_k et le barycentre $\bar{\mathbf{c}}$ est donnée par

$$\begin{aligned} d_{\mathcal{X}^2}^2(\mathbf{c}_k, \bar{\mathbf{c}}) &= \sum_{i=1}^n n \left(\frac{n_{ik}}{Z_{\cdot k}} - \frac{1}{n} \right)^2, \\ &= \frac{n}{Z_{\cdot k}} - 1. \end{aligned}$$

L'inertie totale du nuage du points est alors donnée par la somme pondérée des distances du \mathcal{X}^2 , où la pondération dépend des marges sur les colonnes $Z_{\cdot k}$

$$\sum_{k=1}^q \frac{Z_{\cdot k}}{np} d_{\mathcal{X}^2}^2(\mathbf{c}_k, \bar{\mathbf{c}}) = \frac{q}{p} - 1.$$

Donc l'inertie totale n'est rien d'autre que le rapport *nombre de modalités* sur *nombre de variables* auquel on retranche la valeur 1.

Interprétations. Finalement, on peut aussi chercher à mesurer la qualité de la représentation d'un profil ligne \mathbf{l}_i sur l'axe \mathbf{u}_m par la relation ¹⁸

$$q_{\mathbf{u}_m}(\mathbf{l}_i) \frac{p_{\mathbf{u}_m}(\mathbf{l}_i)^2}{\sum_{k=1}^q \frac{1}{f_{\cdot k} \left(\frac{f_{ik}}{f_{i\cdot}} - f_{\cdot k} \right)^2}} = \frac{p_{\mathbf{u}_m}(\mathbf{l}_i)^2}{\sum_{k=1}^q \frac{np}{Z_{\cdot k} \left(\frac{n_{ik}}{p} - \frac{Z_{\cdot k}}{p} \right)^2}}.$$

et de la représentation d'un profil colonne \mathbf{c}_k sur l'axe \mathbf{u}'_m

$$q_{\mathbf{u}'_m}(\mathbf{c}_k) \frac{p_{\mathbf{u}'_m}(\mathbf{c}_k)^2}{\sum_{i=1}^n \frac{1}{f_{i\cdot} \left(\frac{f_{ik}}{f_{\cdot k}} - f_{i\cdot} \right)^2}} = \frac{p_{\mathbf{u}'_m}(\mathbf{c}_k)^2}{\sum_{i=1}^n n \left(\frac{n_{ik}}{Z_{\cdot k}} - \frac{1}{n} \right)^2}.$$

Ces mesures de qualité sont des mesures aussi appelées \cos^2 , c'est une mesure de l'angle formé entre le point initial dans l'espace \mathbb{R}^q et celui résultant de la projection sur l'axe factoriel. Plus cet angle est fermé (proche de zéro), plus les vecteurs sont proches et meilleure sera la qualité de la représentation (ce qui veut dire que la valeur \cos^2 sera proche de 1).

Rappelons que **l'ACM représente des modalités qui peuvent être issues d'une même variable \mathbf{v} !** On peut donc chercher à évaluer **la qualité de la représentation de la variable \mathbf{v}_j sur un axe \mathbf{u}'_m** en sommant la qualité des représentations des différents profils colonnes

$$q_{\mathbf{u}'_m}(\mathbf{v}_j) = \sum_{j:\mathbf{c}_j \in \mathbf{v}_k} q_{\mathbf{u}'_m}(\mathbf{c}_k).$$

Tout comme nous avons mesuré la qualité de la représentation, on peut aussi mesurer la contribution d'un profil ligne \mathbf{l}_i sur l'axe \mathbf{u}_m par la relation

$$c_{\mathbf{u}_m}(\mathbf{l}_i) = f_{i\cdot} \frac{p_{\mathbf{u}_m}(\mathbf{l}_i)^2}{\lambda_m} = \frac{1}{n} \frac{p_{\mathbf{u}_m}(\mathbf{l}_i)^2}{\lambda_m}.$$

De même pour la contribution d'un profil colonne \mathbf{c}_k sur l'axe \mathbf{u}'_m par la relation

$$c_{\mathbf{u}'_m}(\mathbf{c}_k) = f_{\cdot k} \frac{p_{\mathbf{u}'_m}(\mathbf{c}_k)^2}{\lambda_m} = \frac{Z_{\cdot k}}{np} \frac{p_{\mathbf{u}'_m}(\mathbf{c}_k)^2}{\lambda_m}.$$

Enfin, il est également possible d'ajouter des individus supplémentaires ou encore des modalités supplémentaires. Prenons par exemple un individu supplémentaire que l'on

18. ces relations sont analogues à celles présentées dans le cadre de l'AFC

notera $\mathbf{l}_{+,i}$ qui est vecteur binaire de \mathbb{R}^q . C'est donc une ligne supplémentaire dans ntre tableau disjonctif complet. Alors la représentation de ce nouvel individu sur l'axe factoriel \mathbf{u}_m (*i.e.* la coordonnée de ce nouvelle individu sur cet axe) est donnée par

$$p_{\mathbf{u}_m}(\mathbf{l}_{+,i}) = \frac{1}{\sqrt{\lambda_m}} \left(\frac{1}{p} \sum_{k:(z_{+,i})_k=1} p_{\mathbf{u}'_m}(\mathbf{c}_k) \right).$$

De même la représentation d'une nouvelle modalité $\mathbf{c}_{+,k}$ dans le tableau disjonctif est donnée par

$$p_{\mathbf{u}'_m}(\mathbf{c}_{+,k}) = \frac{1}{\sqrt{\lambda_m}} \left(\frac{1}{Z_{\cdot k}} p_{\mathbf{u}_m}(\mathbf{l}_i) \right).$$