Algèbre linéaire propédeutique
— version abrégée —

Ioan Manolescu

December 27, 2020

Avant-Propos

Ces notes accompagnent le cours d'algèbre linéaire propédeutique donné au semestre d'automne des années 2020-2021 à l'Université de Fribourg.

Le cours a était restructuré par rapport aux années précédentes pour accentuer d'avantage les applications concretes. En particulier, les matrices, le déterminant et l'interpretation géométrique sont introduits le plus tôt possible.

Le but du cours est double. Il vise principalement de familiariser les étudiants avec l'algèbre linéaire, en accentuant les aspects pratiques comme le calcul matriciel, la résolution des systèmes linéaires et les notions de valeur et de vecteur propres. Un deuxième but est de faire découvrir aux étudiants le fonctionnement des mathématiques, à savoir la construction d'une théorie à partir d'axiomes et le concept de preuve. Ce processus vise à créer un cadre abstrait qui permet de décrire des problèmes issus de la réalité. Trouver un cadre abstrait, général, plutôt que d'en créer un pour chaque problème rencontré, cela nous permet d'apercevoir des connexions entre différents problèmes et d'en offrir des solutions plus robustes.

Enfin, dans le développement de toute science, arrive un moment où, pour dépasser les approches ad-hoc, il est nécessaire de construire un cadre théorique qui permet une étude systématique. Les sciences plus “mathématiques” comme la physique ou l'informatique travaillent déjà dans un tel cadre. Dans les sciences traditionnellement plus empiriques (biologie, médecine, économie etc.) ce cadre est en train de se former par les travaux de modélisation de plus en plus fréquentes. Ainsi, il est essentiel pour les scientifiques en formation de se confronter à une construction abstraite, comme celle présentée dans ce cours.

Cette version du polycopié résume les définitions et les résultats essentiels du cours; elle correspond à ce qui est présenté en classe. Ainsi, ce sont uniquement les preuves simples et illustratrices qui y sont présentées. Ceux qui sont intéressés par les preuves plus complexes peuvent consulter la version complète du polycopié. Les exercices sont destinés à offrir une compréhension plus profonde du cours et sont facultatifs. Ils sont différents des exercices des séries qui consistent surtout en exemples concrets.

Ce polycopié n'est pas censé remplacer le cours donné en classe; il se veut plutôt un complément qui permet aux étudiants de revoir certains points.

Contents

Chapter 1

Systèmes linéaires, pivot de Gauss, espace $\mathbb{R}^d$

Pour $d \geq 1$, $\mathbb{R}^d$ est l'ensemble des d-uplets d'éléments de $\mathbb{R}$:

$$\mathbb{R}^d = \{x = (x_1, \dots, x_d) : x_1, \dots, x_d \in \mathbb{R}\}.$$

Suivant la situation, on utilisera aussi l'écriture en colonne $x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$. Pour l'instant, on considère ces deux écritures équivalentes.

1.1 Systèmes linéaires; ensemble des solutions

Définition 1.1. *Soient $m, n \in \mathbb{N}$ et $(a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, $(b_i)_{1 \leq i \leq m}$ deux familles de scalaires (c.à d. des éléments de $\mathbb{R}$). Le système linéaire de m équations à n inconnues $x_1, \dots, x_n$ avec les coefficients $(a_{ij})_{i,j}$, $(b_i)_i$ est l'ensemble d'équations*

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \dots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m. \end{cases} \tag{1.1}$$

Une solution du système est une famille $(x_1, \dots, x_n) \in \mathbb{R}^n$ qui satisfait toutes les équations de (1.1). On appelle $\mathcal{S} \subset \mathbb{R}^n$ l'ensemble des solutions de (1.1).

On verra par la suite trois possibilités pour l'ensemble des solutions d'un système linéaire:

- $\mathcal{S} = \emptyset$: le système n'admet pas de solution;
- $\mathcal{S}$ est un singleton: le système admet une unique solution;
- $\mathcal{S}$ est infini: le système admet une infinité de solutions.

Le *système homogène* associé à (1.1) est le suivant

$$\begin{cases} a_{11}y_1 + \cdots + a_{1n}y_n = 0 \\ \dots \\ a_{m1}y_1 + \cdots + a_{mn}y_n = 0. \end{cases}$$

Le système homogène admet toujours au moins une solutions, à savoir $y_1 = \cdots = y_n = 0$, mais peut en général en admettre plusieurs. Notons $\mathcal{S}_{hom}$ l'ensemble des solutions du système homogène.

Supposons qu'on dispose d'une solution $(x_1, \dots, x_n)$ du système (1.1); on appellera celle-ci une *solution particulière*. Alors, si $(x'_1, \dots, x'_n)$ est une autre solution de (1.1), on a

$$\begin{cases} a_{11}(x_1 - x'_1) + \cdots + a_{1n}(x_n - x'_n) = b_1 - b_1 = 0 \\ \dots \\ a_{m1}(x_1 - x'_1) + \cdots + a_{mn}(x_n - x'_n) = b_m - b_m = 0, \end{cases}$$

donc $(x_1 - x'_1, \dots, x_n - x'_n)$ est solution du système homogène. Inversement, si $(y_1, \dots, y_n)$ est solution du système homogène, alors on vérifie aisément que $(x_1 + y_1, \dots, x_n + y_n) \in \mathcal{S}$. On conclut que les solutions du système originel peuvent être obtenues comme la somme d'une solutions particulière avec les solutions du système homogène:

$$\mathcal{S} = (x_1, \dots, x_n) + \mathcal{S}_{hom} = \big\{(x_1 + y_1, \dots, x_n + y_n) \,:\, (y_1, \dots, y_n) \in \mathcal{S}_{hom}\big\}.$$

1.2 Opérations entre les équations; écriture réduite

Notons les équations du système comme suit

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 & (L_1) \\ \dots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m. & (L_m) \end{cases} \tag{1.2}$$

Dans la résolution d'un tel système d'equations, trois types d'opérations vont nous aider à simplifier le système

- pour $i \neq j$ et $\lambda \in \mathbb{R}$, l'opération $\text{Add}_{i;\lambda,j}$ consiste à rajouter à l'équation L_i l'équation L_j multipliée par λ; le système devient alors

$$\left\{\begin{array}{rcllll} a_{11}x_1 + & \dots & +a_{1n}x_n & = b_1 & (L_1) \\ \vdots & \vdots & & \vdots & \vdots \\ (a_{i1} + \lambda a_{j1})x_1 + & \dots & +(a_{in} + \lambda a_{jn})x_n & = b_i + \lambda b_j & (L_i) + \lambda(L_j) \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1}x_1 + & \dots & +a_{mn}x_n & = b_m. & (L_m) \end{array}\right. \tag{1.3}$$

- pour $\lambda \neq 0$, l'opération $\mathrm{Mult}_{\lambda,i}$ consiste à multiplier l'équation i par λ; le système devient alors

$$\left\{\begin{array}{rcrll} a_{11}x_1+ & \dots & +a_{1n}x_n & = b_1 & (L_1) \\ \vdots & \vdots & & \vdots & \vdots \\ \lambda a_{i1}x_1+ & \dots & +\lambda a_{in}x_n & = \lambda b_i & \lambda(L_i) \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1}x_1+ & \dots & +a_{mn}x_n & = b_m. & (L_m) \end{array}\right. \tag{1.4}$$

- l'opération $\mathrm{Ech}_{i;j}$ consiste à échanger les lignes i et j.

Proposition 1.2. *Les opérations* $\mathrm{Add}_{i;\lambda,j}$ *(avec $i \neq j$),* $\mathrm{Mult}_{\lambda,i}$ *(avec $\lambda \neq 0$) et* $\mathrm{Ech}_{i;j}$ *ne changent pas l'ensemble des solutions du système.*

On va dire que tous les systèmes obtenus à partir de (1.2) par des applications de $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ et $\mathrm{Ech}_{i;j}$ sont équivalents.

Pour écrire un système de manière plus compacte on utilise l'écriture réduite, qui consiste à éliminer les inconnues et les opérations. Ainsi, l'écriture réduite de (1.2) est

$$\left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{array}\right) \tag{1.5}$$

Définition 1.3. *Un système linéaire est sous* forme échelonnée[(i)] *(selon les lignes) si son écriture réduite est de la forme*

$$\begin{array}{c} \begin{array}{cccccccccc} & j_1 & & j_2 & & j_3 & & j_k & & \\ & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \end{array} \\ \left(\begin{array}{ccccccccc|c} 0\dots 0 & 1 & *\dots * & 0 & *\dots * & 0 & \dots * & 0 & *\dots * & b_1 \\ 0\ \dots & & \dots\ 0 & 1 & *\dots * & 0 & \dots * & 0 & *\dots * & b_2 \\ 0\ \dots & & \dots\ \dots & \dots & \dots\ 0 & 1 & \dots * & 0 & *\dots * & b_3 \\ \vdots & & & & & & \ddots & \vdots & \vdots & \vdots \\ 0\ \dots & & \dots\ \dots & \dots & \dots & \dots & \dots\ 0 & 1 & *\dots * & b_k \\ 0\ \dots & & \dots\ \dots & \dots & \dots & \dots & \dots & 0 & \dots\ 0 & b_{k+1} \\ \vdots & & & & & & & \vdots & \vdots & \vdots \\ 0\ \dots & & \dots\ \dots & \dots & \dots & \dots & \dots & 0 & \dots\ 0 & b_m \end{array}\right) \end{array}$$

Dans cette expression, les $*$ représentent des nombres quelconques. On appelle un tel système échelonné de rang k.

[(i)] reduzierte Zeilenstufenform

Attention, dans cette définition on peut avoir $j_1 = 1$ (pas de colonnes nulles), $k = m$ (pas de lignes nulles – seulement si $m \leq n$) et des $j_1, \dots, j_k$ consécutifs (pas de colonnes contenant des $*$).

Les systèmes sous forme échelonnée sont particulièrement faciles à étudier, comme on va voir plus bas. Il est donc intéressant de transformer un système quelconque en un système équivalent sous forme échelonnée. Le théorème suivant nous garantie que cela peut se faire.

Théorème 1.4. *Tout système linéaire peut être transformé en un système forme échelonnée en appliquant une série d'opérations sur les lignes du type* $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Ech}_{i;j}$ *et* $\mathrm{Mult}_{\lambda,i}$.

La preuve du théorème nous donne aussi l'algorithme à suivre pour obtenir la forme échelonnée de A. Cet algorithme, appelé le *pivot de Gauss*[(ii)], est présenté dans la partie suivante.

1.3 Pivot de Gauss pour la résolution des systèmes linéaires

Transformer un système dans un système sous forme échelonnée.
Considérons un système linéaire écrit sous forme réduite

$$\left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{array}\right). \tag{1.6}$$

On va procéder de façon itérative, suivant les colonnes.

Supposons qu'est arrivé à un système dont les $j-1$ premières colonnes sont sous forme échelonnée. Soit $k-1 \leq j-1$ le nombre des lignes non-nulles du tableau formée des les $j-1$ premières colonnes. (Au pas initial, on prend $j = 1$ et $k = 1$ et on applique la procédure décrite ci-dessous.)

On s'occupe de la colonne j, plus précisément de l'entrée a_{kj}. On distingue plusieurs cas:

(i) Si $a_{kj} = 1$, alors on soustrait la ligne k multiplié par a_{ij} à la ligne i pour chaque $i = 1, \dots, k-1, k+1, \dots, n$. Plus précisément on applique $\mathrm{Add}_{i;-a_{ij},k}$ pour $i = 1, \dots, k-1, k+1, \dots, m$. Ainsi on obtient un système qui contient un coefficient 1 sur la position k, j et des 0 sur le reste de la colonne j. Les colonnes $1, \dots, j-1$ ne sont pas affectées par ces transformations car la ligne k a ses premières $j-1$ entrées égales à 0. On conclut que les j premières colonnes ainsi obtenues sont échelonnées; on peut donc passer à la colonne suivante.

(ii) Si $a_{kj} \neq 0$ est une valeur quelconque, alors on divise la ligne k par a_{kj} (c.à-d. on applique $\mathrm{Mult}_{k,\frac{1}{a_{kj}}}$) et on obtient ainsi un système comme celui traité au point (i). On continue en appliquant le point (i) à ce système.

(iii) Si $a_{kj} = 0$ mais il existe $\ell > k$ tel que $a_{\ell,j} \neq 0$, alors on échange les lignes k et ℓ (c.à-d. on applique $\mathrm{Ech}_{k,\ell}$) et on se ramène ainsi à un système comme celui traité au point (ii); on continue en appliquant le point (ii).

[(ii)] Gaußsches Eliminationsverfahren

(iv) Si $a_{kj} = a_{k+1,j} = \cdots = a_{mj} = 0$, alors les j premières colonnes sont déjà sous forme échelonnée et on peut passer à la colonne suivante.

L'algorithme finit quand $j = n+1$ ou $k = m+1$. On vérifie facilement que dans les deux cas, le système résultant est bien sous forme échelonnée. Un exemple est donné plus bas.

Attention! Pendant tout l'algorithme, toutes les opérations doivent aussi être appliquées à la colonne b. Pourtant, on n'essaye pas de mettre cette colonne sous forme échelonnée.
L'ordre d'application des transformations est important!

Résoudre un système sous forme échelonnée.
Supposons maintenant qu'on veut résoudre un système déjà écrit sous forme échelonnée:

$$\begin{array}{c}
\begin{array}{ccccccccc} & j_1 & & j_2 & & j_3 & & j_k & \\ & \downarrow & & \downarrow & & \downarrow & & \downarrow & \end{array} \\
\left(\begin{array}{ccccccccccccccc|c}
0 & \ldots & 0 & 1 & * \ldots * & 0 & * \ldots * & 0 & \ldots & * & 0 & * & \ldots & * & & b_1 \\
0 & \ldots & \ldots & & & 0\;1 & * \ldots * & 0 & \ldots & * & 0 & * & \ldots & * & & b_2 \\
0 & \ldots & \ldots & \ldots & \ldots & & & 0\;1 & \ldots & * & 0 & * & \ldots & * & & b_3 \\
\vdots & & & & & & & & \ddots & & \vdots & & & \vdots & & \vdots \\
0 & \ldots & \ldots & \ldots & \ldots & \ldots & & \ldots & & 0 & 1 & * & \ldots & * & & b_k \\
0 & \ldots & \ldots & \ldots & \ldots & \ldots & & \ldots & & & 0 & & \ldots & 0 & & b_{k+1} \\
\vdots & & & & & & & & & & \vdots & & & \vdots & & \vdots \\
0 & \ldots & \ldots & \ldots & \ldots & \ldots & & \ldots & & & 0 & & \ldots & 0 & & b_m
\end{array}\right)
\end{array} \tag{1.7}$$

où $j_1 < \cdots < j_k$ sont les colonnes contenant les premiers 1 de chaque ligne. Alors $(x_1, \ldots, x_n)$ est solution du système si et seulement si

$$\begin{cases} x_{j_1} + \sum_{j>j_1} a_{1j} x_j &= b_1 \\ \vdots & \\ x_{j_k} + \sum_{j>j_k} a_{kj} x_j &= b_k \\ 0 &= b_{k+1} \\ \vdots & \\ 0 &= b_m \end{cases}$$

Ainsi, le système admet des solutions si et seulement si $b_{k+1} = \cdots = b_m = 0$. De plus, si $b_{k+1} = \cdots = b_m = 0$, les solutions sont obtenues comme suit: on choisi arbitrairement les valeurs de x_j pour $j \notin \{j_1, \ldots j_k\}$ (on appelle ces x_j des variables libres), ensuite, pour chaque j_ℓ, on pose

$$x_{j_\ell} = b_\ell - \sum_{j>j_k} a_{kj} x_j.$$

En autres mots, on a

$$\mathcal{S} = \begin{cases} \big\{(x_1, \ldots, x_n) \in \mathbb{R}^n \,:\, x_{j_\ell} = b_\ell - \sum_{j>j_k} a_{kj} x_j \;\forall 1 \le \ell \le k\big\} & \text{si } b_{k+1} = \cdots = b_m = 0; \\ \emptyset & \text{sinon.} \end{cases}$$

On peut également voir la structure des solutions comme solution particulière plus les solutions du système homogène (quand $b_{k+1} = \cdots = b_m = 0$). Une solutions particulière du système est $x_{j_\ell} = b_\ell$ pour tout $1 \le \ell \le k$ et $x_j = 0$ pour $j \notin \{j_1, \dots j_k\}$. Le système homogène associé est

$$\begin{cases} x_{j_1} + \sum_{j>j_1} a_{1j}x_j &= 0 \\ \vdots & \\ x_{j_k} + \sum_{j>j_k} a_{kj}x_j &= 0 \end{cases} \qquad \text{ou encore} \qquad \begin{cases} x_{j_1} &= -\sum_{j>j_1} a_{1j}x_j \\ \vdots & \\ x_{j_k} &= -\sum_{j>j_k} a_{kj}x_j. \end{cases}$$

En fin, identifions les trois cas possibles, déjà discutés dans la partie précédente:

- $\mathcal{S} = \emptyset$: cela arrive quand $k < m$ et il existe un term non-nul parmi $b_{k+1}, \dots, b_m$;
- $\mathcal{S}$ est un singleton: cela arrive quand $k = n$ et $b_{k+1} = \dots = b_m = 0$;
- $\mathcal{S}$ est infini: cela arrive quand $k < n$ et $b_{k+1} = \dots = b_m = 0$.

Le deuxième cas peut seulement arriver quand $n \le m$ (au moins autant d'équations que d'inconnues). De plus, si $m = n$, le deuxième cas arrive seulement quand le système est de la forme

$$\left(\begin{array}{cccc|c} 1 & 0 & \dots & 0 & b_1 \\ 0 & 1 & \dots & 0 & b_2 \\ \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & \dots & 1 & b_m \end{array}\right). \tag{1.8}$$

Dans le troisième cas, on dit que l'ensemble des solutions est de dimension $n - k$; on verra par la suite pourquoi.

Exemple: Considérons le système suivant, aux inconnues $x_1, \dots, x_5 \in \mathbb{R}$,

$$\begin{cases} 4x_1 - 4x_2 - 4x_3 - 8x_4 - 8x_5 = -8 \\ 3x_1 - 3x_2 - 3x_3 - x_4 + 4x_5 = 9 \\ x_1 + 2x_3 + x_4 + 5x_5 = 5 \\ 2x_1 + 4x_3 + 6x_5 = 4 \\ -2x_1 + x_2 - x_3 + 4x_4 + 3x_5 = 6 \end{cases}$$

On l'écrit sous forme réduite comme dans (1.5):

$$\left(\begin{array}{rrrrr|r} 4 & -4 & -4 & -8 & -8 & -8 \\ 3 & -3 & -3 & -1 & 4 & 9 \\ 1 & 0 & 2 & 1 & 5 & 5 \\ 2 & 0 & 4 & 0 & 6 & 4 \\ -2 & 1 & -1 & 4 & 3 & 6 \end{array}\right)$$

On applique l'algorithme pour mettre le système sous forme échelonnée: on commence par $k = j = 1$. Vu que l'entrée en position $1, 1$ est non-nulle, on applique le point (ii), à savoir on divise la première ligne par 4 pour obtenir un 1 en haut à gauche:

$$\left(\begin{array}{rrrrr|r} 1 & -1 & -1 & -2 & -2 & -2 \\ 3 & -3 & -3 & -1 & 4 & 9 \\ 1 & 0 & 2 & 1 & 5 & 5 \\ 2 & 0 & 4 & 0 & 6 & 4 \\ -2 & 1 & -1 & 4 & 3 & 6 \end{array}\right) \qquad L_1/4$$

Ensuite on applique l'étape (i) de l'algorithme: on additionne la première ligne multipliée par -3, -1, -2, 2 aux lignes $2, 3, 4$ et 5, respectivement. Ainsi on obtient des zéros sur le reste de la première colonne.

$$\left(\begin{array}{rrrrr|r} 1 & -1 & -1 & -2 & -2 & -2 \\ 0 & 0 & 0 & 5 & 10 & 15 \\ 0 & 1 & 3 & 3 & 7 & 7 \\ 0 & 2 & 6 & 4 & 10 & 8 \\ 0 & -1 & -3 & 0 & -1 & 2 \end{array}\right) \quad \begin{array}{l} \\ L_2 - 3L_1 \\ L_3 - L_1 \\ L_4 - 2L_1 \\ L_5 + 2L_1 \end{array}$$

La première colonne est maintenant échelonnée; on passe à la suivante, donc à $j = k = 2$. Comme l'entrée en position $2, 2$ est nulle, mais qu'il y a des valeurs non-nulles sous cette entrée, on applique le point (iii) de l'algorithme: on échange les lignes 2 et 3 pour mettre une valeur non-nulle à la position $2, 2$.

$$\left(\begin{array}{rrrrr|r} 1 & -1 & -1 & -2 & -2 & -2 \\ 0 & 1 & 3 & 3 & 7 & 7 \\ 0 & 0 & 0 & 5 & 10 & 15 \\ 0 & 2 & 6 & 4 & 10 & 8 \\ 0 & -1 & -3 & 0 & -1 & 2 \end{array}\right) \quad \begin{array}{l} \\ L_3 \\ L_2 \\ \\ \\ \end{array}$$

L'entrée à la position $2, 2$ est maintenant déjà égale à 1, il n'est donc pas nécessaire d'appliquer le point (ii), on passe directement au point (i). On additionne la ligne 2 multipliée par 1, -2 et 1 aux lignes 1, 3 et 5, respectivement, pour éliminer les autres entrées de la colonne 2:

$$\left(\begin{array}{rrrrr|r} 1 & 0 & 2 & 1 & 5 & 5 \\ 0 & 1 & 3 & 3 & 7 & 7 \\ 0 & 0 & 0 & 5 & 10 & 15 \\ 0 & 0 & 0 & -2 & -4 & -6 \\ 0 & 0 & 0 & 3 & 6 & 9 \end{array}\right) \quad \begin{array}{l} L_1 + L_2 \\ \\ \\ L_4 - 2L_2 \\ L_5 + L_2 \end{array}$$

On passe à la colonne 3 (à savoir à $j = k = 3$). Toutes les entrées sous le niveau déjà traité sont nulles (on est dans le cas (iv) de l'algorithme), on peut donc passer à la colonne 4, à savoir à $j = 4$ et $k = 3$. On applique le point (ii): on divise la ligne 3 par 5 pour obtenir l'entrée 1 en position $3, 4$:

$$\left(\begin{array}{rrrrr|r} 1 & 0 & 2 & 1 & 5 & 5 \\ 0 & 1 & 3 & 3 & 7 & 7 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & -2 & -4 & -6 \\ 0 & 0 & 0 & 3 & 6 & 9 \end{array}\right) \quad \begin{array}{l} \\ \\ \frac{1}{5}L_3 \\ \\ \\ \end{array}$$

Pour éliminer les autres entrées sur la colonne 4 on applique le point (i): on additionne la ligne 3 multipliée par $-1, -3, 2 - 3$ aux lignes $1, 2, 4$ et 5, respectivement.

$$\left(\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & 3 & 2 \\ 0 & 1 & 3 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \qquad \begin{array}{l} L_1 - L_3 \\ L_2 - 3L_3 \\ \\ L_4 + 2L_3 \\ L_5 - 3L_3 \end{array}$$

En fin on passe à la colonne 5 ($j = 5$, $k = 4$). On est à nouveau dans le cas (iv). Comme toutes les colonnes ont été analysées, l'algorithme est fini.

Le système ainsi obtenue est échelonné; les variables libres sont x_3 et x_5. Le système initial est équivalent à

$$\left\{\begin{array}{rl} x_1 + 2x_3 + 3x_5 & = 2 \\ x_2 + 3x_3 + x_5 & = -2 \\ x_4 + 2x_5 & = 3 \end{array}\right.$$

Les solutions sont obtenues en choisissant x_3 et x_5 arbitrairement, puis en posant

$$\left\{\begin{array}{ll} x_1 & = 2 - 2x_3 - 3x_5 \\ x_2 & = -2 - 3x_3 - x_5 \\ x_4 & = 3 - 2x_5 \end{array}\right. \tag{1.9}$$

Ainsi $\mathcal{S} = \left\{ \begin{pmatrix} 2 - 2\lambda - 3\mu \\ -2 - 3\lambda - \mu \\ \lambda \\ 3 - 2\mu \\ \mu \end{pmatrix} : \lambda, \mu \in \mathbb{R} \right\}$, ou λ et μ jouent le rôle des variables libres x_3 et x_5.

Exercice 1.1.
Représenter dans l'espace $\mathbb{R}^3$ les solutions de

$$\left\{\begin{array}{l} 2x + 2y + z = 4 \\ x + y + z = 0 \end{array}\right.$$

Exercice 1.2.
Mettre sous forme échelonnée les systèmes suivantes et calculer leurs solutions:

$$a) \left\{\begin{array}{l} x + 2y + z = 4 \\ 2x + y + z = 3 \\ x + 5y + 2z = 7. \end{array}\right. \qquad b) \left\{\begin{array}{l} 3x + 2z = 5y - 6 \\ z = x + 2 \\ 2x - y - z = 0. \end{array}\right.$$

$$c) \left\{\begin{array}{l} x + 2y + z + t = 4 \\ 2x + 3y + 3z - t = 3 \\ x + 5y + 2z + 3t = 7 \\ 2x + 5y + 4z + 3t = 2. \end{array}\right. \qquad d) \left\{\begin{array}{l} x + y + z + t = 0 \\ x - y + 2z - 2t = 0 \\ x + y + 4z + 4t = 0 \\ x - y + 8z - 8t = 0. \end{array}\right.$$

1.4 L'espace vectoriel $\mathbb{R}^d$

Pour $d \geq 1$, l'ensemble $\mathbb{R}^d$ peut être naturellement muni de deux types d'opérations; on obtient ainsi ce qu'on appelle l'espace vectoriel $\mathbb{R}^d$. On appellera les éléments de $\mathbb{R}^d$ des *vecteurs* et ceux de $\mathbb{R}$ des *scalaires*.

- Pour $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ et $\lambda \in \mathbb{R}$ la *multiplication du vecteur x par le scalaire λ* est le vecteur
$$\lambda \cdot x = (\lambda x_1, \dots, \lambda x_d) \in \mathbb{R}^d$$
(par convention, on écrit toujours le scalaire devant le vecteur);
- pour deux vecteurs $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ et $y = (y_1, \dots, y_d) \in \mathbb{R}^d$ la *somme* du vecteur x et du vecteur y est le vecteur
$$x + y = (x_1 + y_1, \dots, x_d + y_d) \in \mathbb{R}^d.$$

Les opérations sur $\mathbb{R}^d$ suivent des réglés similaires aux opérations usuelles sur $\mathbb{R}$.

Proposition 1.5. *Soit $d \geq 1$. Pour tous $x, y, z \in \mathbb{R}^d$ et $\lambda, \mu \in \mathbb{R}$:*

(i) $x + y = y + x$ *(commutativité de +);*

(ii) $(x + y) + z = x + (y + z)$ *(associativité de +);*

(iii) $\lambda \cdot (\mu \cdot x) = (\lambda\mu) \cdot x$ *(associativité de ·);*

(iv) $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$ *(distributivité pour +);*

(v) $\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$ *(distributivité pour +).*

On écrit $0 = (0, \dots, 0)$ pour le vecteur nul.

Définition 1.6. *Un ensemble $F \subset \mathbb{R}^d$ est appelé un* sous-espace vectoriel *de $\mathbb{R}^d$ si $F \neq \emptyset$ et si,*

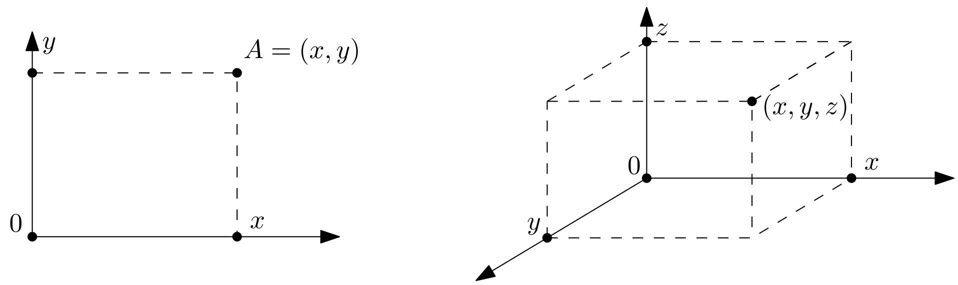
$$\text{pour tous } x, y, \in F \text{ et } \lambda, \mu \in \mathbb{R}, \quad \lambda x + \mu y \in F.$$

Un ensemble $G \subset \mathbb{R}^d$ est appelé un sous-espace affine *de $\mathbb{R}^d$ s'il existe F sous-espace vectoriel de E et $x \in E$ tel que $G = \{x + y \,:\, y \in F\}$.*

Observons que tout sous-espace vectoriel de $\mathbb{R}^d$ contient le vecteur nul $0 = (0, \dots, 0)$. Deux sous-espaces vectoriels de $\mathbb{R}^d$ sont $\{0\}$ (le plus petit) et $\mathbb{R}^d$ (le plus grand). Un autre exemple (quand $d \geq 2$) est $\{(x_1, \dots, x_n) : x_n = 0\}$.

Interprétation géométrique Pour $d = 2$ et $d = 3$, l'espace $\mathbb{R}^d$ admet une interprétation géométrique intuitive. Commençons par le cas $d = 2$. Fixons dans un plan un système d'axes perpendiculaires qui s'intersectent en un point qu'on notera 0. De plus, supposons que les deux axes sont munis d'un système de graduation. Alors tout point A du plan est identifié à l'unique point $(x, y) \in \mathbb{R}^2$ où x et y sont ses coordonnées sur l'axe horizontal et vertical, respectivement.

De même l'espace $\mathbb{R}^3$ s'identifie à l'espace ambient de dimension 3, muni d'un système de coordonnées. Ainsi, tout point $(x, y, z) \in \mathbb{R}^3$ correspond au point de coordonnées x, y et z de l'espace.



Souvent dans ce contexte on identifie un point A de l'espace à la flèche qui pointe de 0 vers A; on la note $\overrightarrow{OA}$. L'addition des vecteurs correspond géométriquement à l'addition par la règle du parallélogramme. La multiplication par un scalaire positif correspond à l'elongation de la flèche et celle par -1 à un changement de sens.

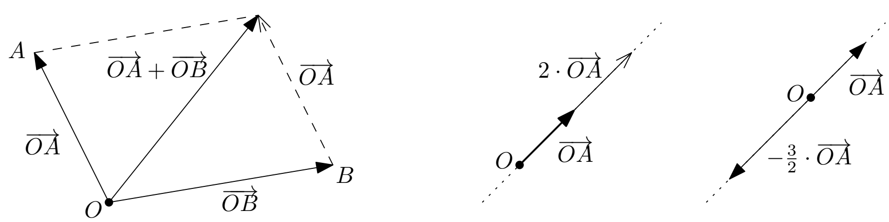


Figure 1.1: L'addition des vecteurs peut se faire par la règle du parallélogramme. Dans l'image de gauche, les côtés opposés du parallélogramme représentent le même vecteur. La multiplication par une constante se fait en gardant la même direction, mais en multipliant la longueur du vecteur. Si la constante est négative, la flèche change de sens.

Les sous-espaces vectoriels de $\mathbb{R}^2$, $\mathbb{R}^3$ (et plus généralement $\mathbb{R}^d$) s'identifient aux droites, plans etc. qui passent par 0. Le point 0 par lui même forme le sous-espace vectoriel trivial $\{0\}$. Les droites et plans qui ne passent pas forcement par 0 sont des espaces affines.

1.5 Indépendance linéaire; espace engendré; bases

Espace engendré Fixons $d \geq 1$. Pour une famille $x_1, \ldots, x_n \in \mathbb{R}^d$, on appelle *combinaison linéaire*[(iii)] de $x_1, \ldots, x_n$ tout vecteur de la forme

$$\lambda_1 x_1 + \cdots + \lambda_n x_n \in \mathbb{R}^d,$$

(iii) Linearkombination

où $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ sont des scalaires quelconques.

Définition 1.7. *Soit $x_1, \dots, x_n \in \mathbb{R}^d$ une famille finie de vecteurs. Le sous-espace vectoriel engendré par $x_1, \dots, x_n$*[(iv)] *est*

$$\mathrm{Vect}(x_1, \dots, x_n) := \{\lambda_1 x_1 + \dots + \lambda_n x_n : \lambda_1, \dots, \lambda_n \in \mathbb{R}\}.$$

Par convention, on pose $\mathrm{Vect}(\emptyset) = \{0\}$.

On dit que la famille $(x_1, \dots, x_n)$ est génératrice[(v)] *(pour $\mathbb{R}^d$) si* $\mathrm{Vect}(x_1, \dots, x_n) = \mathbb{R}^d$.

En d'autres termes, $\mathrm{Vect}(x_1, \dots, x_n)$ est l'ensemble formé de toutes les combinaisons linéaires de $x_1, \dots, x_n$. La famille $(x_1, \dots, x_n)$ est dite génératrice si tout vecteur de $\mathbb{R}^d$ peut s'écrire comme combinaison linéaire de $x_1, \dots, x_n$.

L'ordre des vecteurs $x_1, \dots, x_n$ n'affecte pas l'espace qu'ils engendrent, donc ni le fait que la famille soit génératrice ou pas.

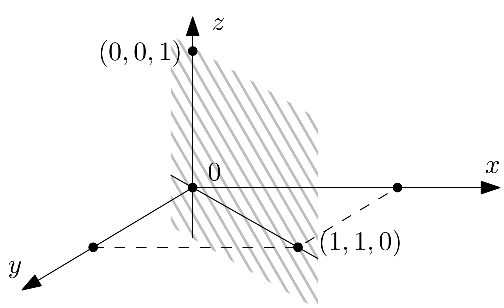
Remarque 1.8. La définition de $\mathrm{Vect}(x_1, \dots, x_n)$ ci-dessus nécessite une preuve. En effet $\mathrm{Vect}(x_1, \dots, x_n)$ est défini comme un *sous-ensemble* de $\mathbb{R}^d$, non-pas comme un sous-espace vectoriel. Il faut donc montrer que $\mathrm{Vect}(x_1, \dots, x_n)$ est en effet un sous-espace vectoriel de $\mathbb{R}^d$. On ne va pas s'occuper pour l'instant de cela; on reviendra à ces questions dans un cadre plus général dans la partie **??**

Exemple: Dans l'espace vectoriel $\mathbb{R}^2$ la famille $((0,1);(1,0))$ est génératrice. En effet, tout vecteur $(a,b) \in \mathbb{R}^2$ s'écrit $(a,b) = b \cdot (0,1) + a \cdot (1,0)$.

Dans l'espace vectoriel $\mathbb{R}^3$, la famille $((1,1,0);(0,0,1))$ n'est pas génératrice. En effet, le vecteur $(0,1,0)$ ne peut pas s'écrire comme $\alpha \cdot (1,0,0) + \beta \cdot (0,1,1)$ avec $\alpha, \beta \in \mathbb{R}$. On peut voir que

$$\mathrm{Vect}((1,1,0);(0,0,1)) = \{(a,b,c) \in \mathbb{R}^3 : \ a = b\}.$$

On peut aussi facilement se convaincre que $\mathrm{Vect}((1,1,0);(0,0,1))$ est un s.e.v. de $\mathbb{R}^3$. Il s'agit du plan vertical qui passe par la diagonale $x = y$ du plan horizontal. Voir l'image.



(iv) lineare Hülle von $(x_1, \dots, x_n)$, aussi notée $\mathrm{span}(x_1, \dots, x_n)$

(v) Erzeugendensystem

Indépendance linéaire

Définition 1.9. *Soit $x_1, \dots, x_n \in \mathbb{R}^d$ une famille de vecteurs. On dit que la famille $(x_1, \dots, x_n)$ est* liée *(ou linéairement dépendante)*[vi] *s'il existe des scalaires $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ non tous nuls, tels que*

$$\lambda_1 x_1 + \cdots + \lambda_n x_n = 0.$$

Une famille qui n'est pas liée est dite libre *(ou linéairement indépendante)*[vii].

En d'autres termes, une famille est liée s'il existe une combinaison linéaire non-triviale de ses vecteurs qui vaut 0. Bien évidemment, le fait qu'une famille de vecteurs est libre ou liée ne dépend pas de l'ordre des vecteurs. Par contre, si un même vecteur apparait deux fois dans la famille, alors la famille est forcement liée.

Si on désire montrer qu'une famille $(x_1, \dots, x_n)$ de vecteurs est libre, on peut considérer des scalaires $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ tels que $\lambda_1 x_1 + \cdots + \lambda_n x_n = 0$ et déduire que $\lambda_1 = \cdots = \lambda_n = 0$. Inversement, si on veut montrer que la famille est liée, le plus simple est d'exhiber une combinaison linéaire non-triviale de $x_1, \dots, x_n$ qui vaut 0. Cela se réduit à la resolution d'un système linéaire (voir la partie suivante).

Exemple: Dans l'espace vectoriel $\mathbb{R}^2$ la famille $((0,1);(1,0))$ est libre. En effet, si $\lambda, \mu \in \mathbb{R}$ sont tels que $\lambda \cdot (0,1) + \mu \cdot (1,0) = (\lambda, \mu) = (0,0)$, alors $\lambda = \mu = 0$.
Par contre, la famille $((0,1);(1,0);(1,1))$ est liée. En effet, on a

$$1 \cdot (0,1) + 1 \cdot (1,0) + (-1) \cdot (1,1) = (0,1) + (1,0) - (1,1) = (0,0).$$

Le lemme suivant offre un critère pratique pour montrer qu'une famille est liée.

Lemme 1.10. *Soit $(x_1, \dots, x_n)$ une famille de vecteurs de $\mathbb{R}^d$. Alors elle est liée si et seulement s'il existe un vecteur x_i qui s'écrit comme combinaison linéaire des autres. Plus précisément si et seulement s'il existe $i \in \{1, \dots, n\}$ et des scalaires $\lambda_1, \dots, \lambda_{i-1}, \lambda_{i+1}, \dots, \lambda_n$ tels que*

$$x_i = \lambda_1 x_1 + \cdots + \lambda_{i-1} x_{i-1} + \lambda_{i+1} x_{i-1} \cdots + \lambda_n x_n = \sum_{j \neq i} \lambda_j x_j.$$

Pour la preuve il suffit d'observer que l'équations du lemme 1.10 se traduit en une combinaison linéaire non triviale nulle.

Exemple: En général le fait qu'une famille est libre ou liée n'implique pas le fait qu'elle est génératrice ou pas. Donnons des exemples pour les quatre situations possibles; on va se placer dans l'espace vectoriel $\mathbb{R}^2$:
La famille $(1,0);(0,1)$ est génératrice et libre.
La famille $(1,0);(1,1);(0,1)$ est génératrice et liée (car $(1,1) = (1,0) + (0,1)$).
La famille formée uniquement du vecteur $(1,0)$ est libre mais pas génératrice.
La famille $(1,0);(0,0);(2,0)$ n'est pas génératrice et est liée.

[vi] linear abhängig
[vii] linear unabhängig

Dimension; bases

Définition 1.11. *Soit F un sous-espace vectoriel de $\mathbb{R}^d$. La* dimension *de F est le plus petit nombre de vecteurs qui génère F:*

$$\dim(F) = \min\{n \geq 0 \,:\, \exists\, x_1, \ldots, x_n \in \mathbb{R}^d \text{ tels que } \mathrm{Vect}(x_1, \ldots, x_n) = F\} \tag{1.10}$$

Par convention $\dim(\{0\}) = 0$.
Une famille libre $x_1, \ldots, x_n$ avec $\mathrm{Vect}(x_1, \ldots, x_n) = F$ *est appelée un* base[viii] *de F.*

La dimension de $\mathbb{R}^d$ est d. En effet, une base est la base canonique

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_d = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

La dimension d'un sous-espace vectoriel de $\mathbb{R}^d$ est toujours comprise entre 0 et d. De plus, les seuls sous-espaces vectoriels de $\mathbb{R}^d$ de dimensions 0 et d sont $\{0\}$ et $\mathbb{R}^d$, respectivement. Pour n strictement entre 0 et d, il existe une infinité de sous-espaces vectoriels de dimensions n.

Une famille contenant $\dim(F)$ vecteurs qui génère F est forcement libre, et donc une base de F. Inversement, toute base de F contient exactement $\dim(F)$ vecteurs. Sauf dans le cas dégénéré ou $\dim(F) = 0$, il existent une infinité de bases différentes de F.

Proposition 1.12. *Si $x_1, \ldots, x_n$ est une base d'un sous-espace vectoriel F, alors tout $y \in F$ s'écrit de manière unique comme combinaison linéaire de $x_1, \ldots, x_n$:*

$$\text{Pour tout } y \in F,\ \exists! \lambda_1, \ldots, \lambda_n \in \mathbb{R} \text{ tels que } y = \lambda_1 x_1 + \cdots + \lambda_n x_n.$$

Preuve: L'existence d'une combinaison linéaire est garantie par le fait que $x_1, \ldots, x_n$ génère F; l'unicité par le fait que $x_1, \ldots, x_n$ est libre. □

Définition 1.13. *Le* rang[ix] *d'une famille de vecteurs $x_1, \ldots, x_n \in \mathbb{R}^d$ est défini par* $\mathrm{rang}(x_1, \ldots, x_n) = \dim(\mathrm{Vect}(x_1, \ldots, x_n))$.

Le rang d'une famille $x_1, \ldots, x_n$ est toujours compris entre 0 et n. D'autre part, comme il s'agit de vecteurs dans $\mathbb{R}^d$, il est aussi plus petit que d. Il s'avère que le rang de $x_1, \ldots, x_n$ coincide avec le nombre de vecteurs *parmi* $x_1, \ldots, x_n$ nécessaires pour engendrer $\mathrm{Vect}(x_1, \ldots, x_n)$.

[viii] Basis
[ix] Rang

Proposition 1.14. *Une famille de vecteurs $x_1, \dots, x_n \in \mathbb{R}^d$ est libre si et seulement si* $\mathrm{rang}(x_1, \dots, x_n) = n$. *En particulier, si $n > d$, la famille est forcement liée. Si $n = d$, la famille est libre si et seulement si elle est génératrice de $\mathbb{R}^d$.*

Les résultats cités ci-dessus vont être discutés plus en détail dans la partie **??**.

1.6 Retour aux systèmes linéaires

Considérons un systèmes linéaires à m équations et n inconnues.

$$\begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = b_1 \\ \dots \\ a_{m1}x_1 + \dots + a_{mn}x_n = b_m. \end{cases} \tag{1.11}$$

Notons $\mathbf{a}_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix}$ pour $1 \leq i \leq n$ et $\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$. Alors $\mathbf{a}_1, \dots, \mathbf{a}_n$ et $\mathbf{b}$ sont des vecteurs de $\mathbb{R}^m$ et (1.12) s'écrit comme une équation vectorielle:

$$x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n = \mathbf{b}.$$

Ainsi, le système admet des solutions si et seulement si $\mathbf{b} \in \mathrm{Vect}(\mathbf{a}_1, \dots, \mathbf{a}_n)$. Si c'est le cas, alors la solution est unique si et seulement si $\mathbf{a}_1, \dots, \mathbf{a}_n$ forment une famille libre.

Rappelons que, en général, les solutions de (1.12) sont la somme d'une solutions particulière de ce système et des solutions $\mathcal{S}_{hom}$ du système homogène $x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n = 0$. Il est facile de vérifier que $\mathcal{S}_{hom}$ est un sous-espace vectoriel de $\mathbb{R}^n$. Ainsi, si (1.12) admet des solutions, alors $\mathcal{S}$ est un sous-espace affine de $\mathbb{R}^n$.

Si on ramène le système sous forme échelonnée et on utilise la notation de (1.7), alors $\mathcal{S}_{hom}$ est un sous-espace vectoriel de dimension $n - k$. Une famille de vecteurs qui engendre $\mathcal{S}_{hom}$ est donnée par les vecteurs $y^{(j)} = (y_1^{(j)}, \dots, y_n^{(j)})$ avec $j \in \{1, \dots, n\} \setminus \{j_1, \dots, j_k\}$, ou $y^{(j)}$ est la solution du système homogène avec $y_j^{(j)} = 1$ et $y_\ell^{(j)} = 0$ pour tout $\ell \notin \{j_1, \dots, j_k\}$ et $\ell \neq j$.

Rappelons que le système admet des solutions si et seulement si $b_{k+1} = \dots = b_m = 0$ (ici et plus bas, b_i sont les scalaires de la dernière colonne du système échelonné). Dans le cas échéant, une solution particulière $x = (x_1, \dots, x_n)$ est donnée par $x_{j_1} = b_{j_1}, \dots, x_{j_k} = b_{j_k}$ et $x_j = 0$ pour tout $j \notin \{j_1, \dots, j_k\}$.

Dans l'exemple de la partie 1.3, la forme initiale et échelonnée du système étaient

$$\left(\begin{array}{rrrrr|r} 4 & -4 & -4 & -8 & -8 & -8 \\ 3 & -3 & -3 & -1 & 4 & 9 \\ 1 & 0 & 2 & 1 & 5 & 5 \\ 2 & 0 & 4 & 0 & 6 & 4 \\ -2 & 1 & -1 & 4 & 3 & 6 \end{array}\right) \quad \text{et} \quad \left(\begin{array}{ccccc|r} 1 & 0 & 2 & 0 & 3 & 2 \\ 0 & 1 & 3 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

On a alors $k = 3$, $j_1 = 1$, $j_2 = 2$ et $j_3 = 4$. Ainsi $\mathcal{S}_{hom}$ est un s.e.v. de dimension $5 - 3 = 2$. Une base pour $\mathcal{S}_{hom}$ est donnée par les deux vecteurs

$$y^{(3)} = \begin{pmatrix} -2 \\ -3 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{et} \quad y^{(5)} = \begin{pmatrix} -3 \\ -1 \\ 0 \\ -2 \\ 1 \end{pmatrix}.$$

De plus, le système admet la solution particulière $x = \begin{pmatrix} 2 \\ -2 \\ 0 \\ 3 \\ 0 \end{pmatrix}$. On conclut donc que $\mathcal{S}_{hom} = \{\lambda y^{(3)} + \mu y^{(5)} : \lambda, \mu \in \mathbb{R}\}$ est un s.e.v. de dimension 2 et $\mathcal{S} = \{x + \lambda y^{(3)} + \mu y^{(5)} : \lambda, \mu \in \mathbb{R}\}$ est une espace affine.

Inversement, si on se donne une famille de vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$, on peut tester si elle est libre ou génératrice en utilisant la résolution des systèmes linéaires par le pivot de Gauss. En effet, on écrit les vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_n$ dans les colonnes d'un tableau comme dans (1.6) (la forme réduite d'un système linéaire); on n'a pas à se soucier de la colonne des $b_1, \dots, b_m$, on peut donc l'ignorer pour ce calcul. On applique ensuite le pivot de Gauss pour réduire le tableau à sa forme échelonnée; ci-dessous on utilise les même notations que dans (1.7):

- La famille $\mathbf{a}_1, \dots, \mathbf{a}_n$ est génératrice si et seulement si $k = m$, c'est à dire que la forme échelonnée ne contient aucune ligne entièrement nulle. En effet, si c'est le cas, le système $x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n = \mathbf{b}$ admet des solutions pour tout $\mathbf{b} \in \mathbb{R}^m$, ce qui veut dire que $\mathrm{Vect}(\mathbf{a}_1, \dots, \mathbf{a}_n) = \mathbb{R}^m$.

  Plus généralement, on peut tester si un vecteur $\mathbf{b} \in \mathbb{R}^m$ est dans $\mathrm{Vect}(\mathbf{a}_1, \dots, \mathbf{a}_n)$ en résolvant le système $x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n = \mathbf{b}$.

- La famille $\mathbf{a}_1, \dots, \mathbf{a}_n$ est libre si et seulement si $k = n$, ce qui veut dire que la forme échelonnée ne contient aucune variable libre (les variable libres correspondent aux colonnes contenant des $*$). En effet, si c'est le cas, le système $x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n = 0$ admet une unique solution, à savoir $(x_1, \dots, x_n) = (0, \dots, 0)$.

  Plus généralement, si on trouve une forme échelonnée avec $k < n$, alors $\mathrm{Vect}(\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_k}) = \mathrm{Vect}(\mathbf{a}_1, \dots, \mathbf{a}_n)$. En d'autres mots, chaque $\mathbf{a}_j$ avec $j \notin \{j_1, \dots, j_k\}$ est redondant car il peut s'écrire comme combinaison linéaire de $\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_k}$.

Description d'un sous-espace vectoriel On a deux options pour décrire un sous-espace vectoriel F de $\mathbb{R}^d$:

- exhiber une base de F – cela est plus simple quand la dimension de F est petite;
- écrire F comme l'ensemble des solutions d'un système linéaire homogène, avec un nombre minimal d'équations (c.à-d. $d - \dim(F)$ équations) – cela est plus simple quand la dimension de F est proche de d.

Revenons à l'exemple du s.e.v. de $\mathbb{R}^3$ $F = \text{Vect}((1,1,0),(0,0,1))$. Cette description est du premier type. Elle n'est pas unique; une autre est $F = \text{Vect}((1,1,1),(-2,-2,1))$. Dans les deux cas, les paires de vecteurs utilisées sont des bases de F.

De manière alternative, on peut écrire $F = \{(x_1,x_2,x_3) \in \mathbb{R}^3 : x_1 = x_2\}$. Le système linéaire ici est formé d'une unique équation, à savoir $x_1 - x_2 = 0$.

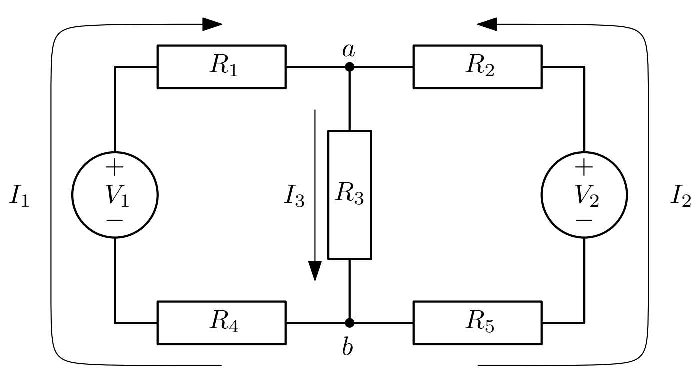
1.7 Application: réseaux électriques

Le flux du courant dans un réseau électrique formé de sources et de résistances est décrit par deux quantités: la tension (entre tous deux points du circuit) et l'intensité du courent (le long de toute arrête du circuit). Trois lois relient ces quantités:

- **Loi de Ohm:** la tension entre deux bornes d'une résistance est proportionnelle au courant qui le traverse $U = IR$; le facteur de proportionnalité est la résistance R.
- **Loi des nœuds de Kirchhoff:** la somme des intensités des courants qui entrent dans un nœud est égale à la somme des intensités des courants qui sortent du même nœud.
- **Loi des mailles de Kirchhoff:** la somme tensions le long de toute maille est nulle.

A l'aide de ces lois et des systèmes linéaires, on peut calculer les intensités et les tensions dans tout circuit. On illustre cela par un exemple.

Considérons le circuit suivant



Les lois de Kirchhoff s'écrivent alors

$$\begin{cases} I_1 + I_2 &= I_3 & \text{pour le nœud } a \text{ ou } b \\ V_1 - I_1R_1 - I_3R_3 - I_1R_4 &= 0 & \text{pour la maille de gauche} \\ V_2 - I_2R_2 - I_3R_3 - I_2R_5 &= 0 & \text{pour la maille de droite} \\ V_2 - I_2R_2 + I_1R_1 - V_1 + I_1R_4 - I_2R_5 &= 0 & \text{pour la grande maille} \end{cases}$$

Dans ce système V_1, V_2 et $R_1, \ldots, R_5$ sont des coefficients (ils nous sont donnés), I_1, I_2, I_3 sont des inconnues. Ainsi, il s'agit d'une système de quatre équations à trois inconnues; c'est une situations ou typiquement le système n'admet pas de solutions. Toutefois, on sait physiquement que le système admet une unique solution. En effet, la quatrième équation est la difference de la troisième et de la deuxième, et le système peut donc s'écrire

$$\begin{cases} I_1 + I_2 - I_3 &= 0 \\ I_1R_1 + I_3R_3 + I_1R_4 &= V_1 \\ I_2R_2 + I_3R_3 + I_2R_5 &= V_2 \end{cases}$$

Un calcul direct en utilisant le pivot de Gauss nous montre que ce système admet une unique solution, notamment

$$I_1 = \frac{V_1(R_1+R_3+R_4) - V_2R_3}{(R_1+R_3+R_4)(R_2+R_3+R_5) - R_3^2}, \qquad I_2 = \frac{V_2(R_2+R_3+R_5) - V_1R_3}{(R_1+R_3+R_4)(R_2+R_3+R_5) - R_3^2},$$
$$I_3 = \frac{V_1(R_1+R_4) + V_2(R_2+R_5)}{(R_1+R_3+R_4)(R_2+R_3+R_5) - R_3^2}.$$

Les differences de potentiels se calculent facilement en utilisant la loi de Ohm.

1.8 Application: optimisation linéaire

On commence par un exemple. Supposons qu'un magasin de meubles vend des tables et des chaises. Le magasin achète les chaises aux prix de 200chf/pièce et les tables au prix de 600chf/pièce; il fait un profit de 50chf par chaise vendue et 200chf par table vendue. On va supposer que le magasin vend tous les meubles mis en vente. La question est de trouver la stratégie du magasin pour maximiser son profit, sachant qu'il est sujet aux limitations suivantes:

(C1) le magasin dispose d'un budget totale de 48000chf,

(C2) le magasin peut contenir au plus 180 pièces de mobilier et

(C3) la livraison est faite par un camion qui peut transporter en tout au plus 2700kg. Une chaise pèse 5kg alors qu'une table pèse $40kg$.

Pour formaliser cela, notons x le nombre de chaises et y le nombre de tables que le magasin décide de mettre en vente. La question devient alors de maximiser $50x + 200y$ sous les conditions suivantes

$$\begin{cases} 200x + 600y &\leq 48000, & \text{(C1)} \\ x + y &\leq 180, & \text{(C2)} \\ 5x + 40y &\leq 2700, & \text{(C3)} \\ x &\geq 0, \\ y &\geq 0. \end{cases}$$

Appelons *l'ensemble admissible* les couples $(x, y) \in \mathbb{R}^2$ qui satisfont les conditions ci-dessus. Il s'agit de l'intérieur d'un polygone. Ses cotés sont des parties des droites

$$\begin{aligned} 200x + 600y &= 48000, && \text{(C1)} \\ x + y &= 180, && \text{(C2)} \\ 5x + 40y &= 2700, && \text{(C3)} \\ x &= 0, && \text{(Ay)} \\ y &= 0. && \text{(Ax)} \end{aligned}$$

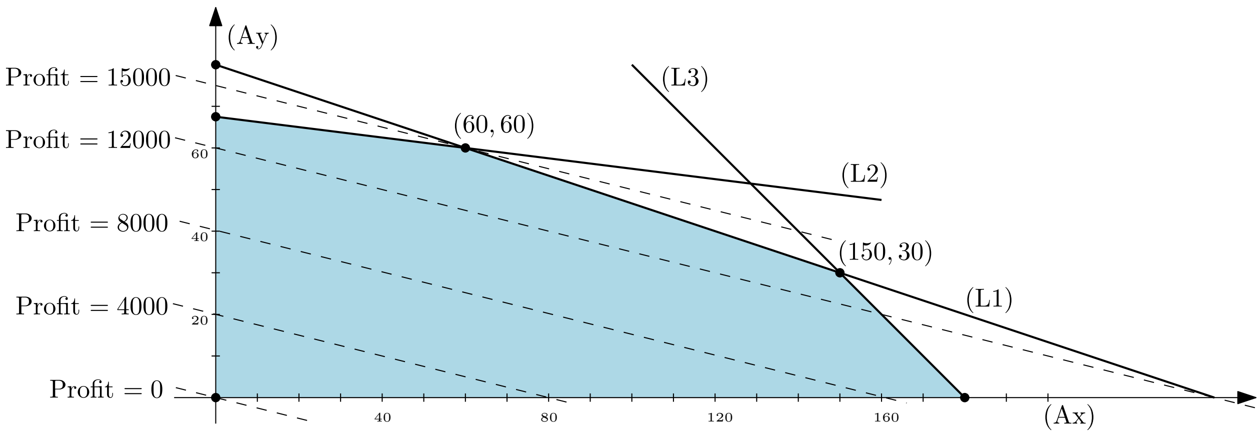


Figure 1.2: Les lignes en gras représentent les solutions aux équations (L1)-(Ax). La zone bleue est l'ensemble des couples admissibles. Les lignes pointillées représentent les couples (x, y) dont le profit est constant.

Ses sommets sont donnés par certaines intersection de ces droites, donc par les systèmes de deux équations avec deux inconnues formés des paires d'équations parmi les cinque ci-dessus. La zone admissible est représentée graphiquement dans la Figure 1.2.

On peut montrer généralement que le profit maximal est atteint à un des sommets du polygone. Dans le cas présent il s'agit du sommet $(60, 60)$ qui est l'intersection de (L1) et (L3). Le profit est alors $50 \cdot 60 + 200 \cdot 60 = 15000$chf.

Plus généralement, dans un problème d'optimisation linéaire, on dispose de n variables $x_1, \dots, x_n$, on vise à maximiser une certaine combinaison linéaire $c_1x_1 + \cdots + c_nx_n$ (qu'on appelle *fonction revenu*) sous m contraintes linéaires

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n \leq b_1 \\ \dots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \leq b_m. \end{cases} \tag{1.12}$$

Souvent, on rajoute les contraintes $x_1 \geq 0, \dots,\ x_n \geq 0$. La zone admissible est un polyèdre dans $\mathbb{R}^n$ et le maximum de la fonction revenu est atteint toujours à un des sommets du polyèdre.

À retenir

- La resolution des systèmes linéaires par le pivot de Gauss. Si le système admet des solutions, elles forment un espace affine.

- $\mathbb{R}^d$ est un espace vectoriel. Les éléments de $\mathbb{R}^d$ sont appelés des *vecteurs*; les nombres sont appelés des *scalaires*.

- Un *sous-espace vectoriel* (s.e.v.) F de $\mathbb{R}^d$ est un sous ensemble de $\mathbb{R}^d$ stable par addition et multiplication par scalaire.
  Un sous-espace vectoriel contient nécessairement le vecteur nul $0 = (0, \dots, 0)$.

- Pour une famille $(x_1, \dots, x_n)$ de vecteurs de $\mathbb{R}^d$, $\mathrm{Vect}(x_1, \dots, x_n)$ est le *sous-espace vectoriel* de $\mathbb{R}^d$ *engendré* par $x_1, \dots, x_n$. Il est composé de toutes les combinaisons linéaires de $x_1, \dots, x_n$.
  Si $\mathrm{Vect}(x_1, \dots, x_n) = \mathbb{R}^d$ on dit que $x_1, \dots, x_n$ est *génératrice* pour $\mathbb{R}^d$.

- Une famille $(x_1, \dots, x_n)$ de vecteurs de $\mathbb{R}^d$ est dite *liée* (ou linéairement dépendante) si un des vecteurs s'écrit comme combinaison linéaire des autres.
  Si elle n'est pas liée, la famille est dite *libre* (ou linéairement indépendante).

- Une *base* d'un s.e.v. F est une famille de vecteurs qui est libre et qui génère F. Le nombre de vecteurs dans la base est la *dimension* de F; on l'écrit $\dim(F)$.

- Dans $\mathbb{R}^d$, une famille libre a au plus d vecteurs; une famille génératrice en a au moins d.

- Pour une famille $x_1, \dots, x_n \in \mathbb{R}^d$ on définit le *rang* de la famille par
  $$\mathrm{rang}(x_1, \dots, x_n) = \dim(\mathrm{Vect}(x_1, \dots, x_n)).$$
  On a $\mathrm{rang}(x_1, \dots, x_n) \leq n$ et $\mathrm{rang}(x_1, \dots, x_n) \leq d$. La famille est libre si et seulement si $\mathrm{rang}(x_1, \dots, x_n) = n$; elle est génératrice si et seulement si $\mathrm{rang}(x_1, \dots, x_n) = d$.

Zu wissen

- Die Auflösung eines linearen Systems mit dem Gauss-Pivot. Wenn das System Lösungen hat, dann bilden diese einen affinen Raum.

- $\mathbb{R}^d$ ist ein Vektorraum. Die Elementen von $\mathbb{R}^d$ werden *Vektoren* genannt; die Werte werden *Skalare* genannt.

- Ein *Untervektorraum* F von $\mathbb{R}^d$ ist eine Teilmenge von $\mathbb{R}^d$, die unter Addition und Multiplikation mit einem Skalar stabil ist.
  Ein Untervektorraum enthält den Nullvektor $0 = (0, \dots, 0)$ sowieso.

- Für eine Familie $(x_1, \dots, x_n)$ von Vektoren aus $\mathbb{R}^d$, wird $\mathrm{Vect}(x_1, \dots, x_n)$ der von $x_1, \dots, x_n$ *aufgespannte Untervektorraum* von $\mathbb{R}^d$ genannt. Es besteht aus alle Linearkombinationen von $x_1, \dots, x_n$.
  Wenn $\mathrm{Vect}(x_1, \dots, x_n) = \mathbb{R}^d$, sagt man, dass $x_1, \dots, x_n$ $\mathbb{R}^d$ *aufspannt.*

- Eine Familie $(x_1, \dots, x_n)$ von Vektoren aus $\mathbb{R}^d$ ist *abhängig* (oder linear abhängig), falls ein Vektor als Linearkombination der anderen Vektoren geschrieben werden kann.
  Falls die Familie nicht abhängig ist, wird sie *frei* (oder linear unabhängig) genannt.

- Eine *Basis* eines Untervektorraums F ist eine Familie von Vektoren, die frei ist und die F erzeugt. Der Anzahl von Basisvektoren ist die *Dimension* von F; man schreibt $\dim(F)$.

- In $\mathbb{R}^d$ hat eine freie Familie höchstens d Vektoren; eine erzeugende Familie hat mindestens d Vektoren.

- Für eine Familie $x_1, \dots, x_n \in \mathbb{R}^d$ definiert man der Rang einer Familie durch

  $$\mathrm{rang}(x_1, \dots, x_n) = \dim(\mathrm{Vect}(x_1, \dots, x_n)).$$

  Es gilt $\mathrm{rang}(x_1, \dots, x_n) \leq n$ und $\mathrm{rang}(x_1, \dots, x_n) \leq d$. Eine Familie ist frei, dann und nur dann, wenn $\mathrm{rang}(x_1, \dots, x_n) = n$; sie ist eine erzeugende Familie, dann und nur dann, wenn $\mathrm{rang}(x_1, \dots, x_n) = d$.

Chapter 2

Matrices

2.1 Définitions, opérations

Les matrices sont des objets omniprésentes dans les différentes domaines des mathématiques. De plus elles sont très souvent utilisées dans les applications des mathématiques. Une de leur utilisations essentielles est la description des applications linéaires, qui va être abordé dans le chapitre 4. Dans ce chapitre on se contente d'une description des matrices et des opérations s'y appliquant.

Définition 2.1. *Soient $m, n \in \mathbb{N}$. Une matrice A de taille $m \times n$ à coefficients dans $\mathbb{R}$ est une famille $(a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ d'éléments de $\mathbb{R}$. Elle est représenté par un tableau rectangulaire*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

L'indice i est l'indice des lignes et l'indice j est celui des colonnes.
L'ensemble de matrices de taille $m \times n$ à coefficients dans $\mathbb{R}$ est noté $\mathcal{M}_{mn}(\mathbb{R})$ (ou $\mathbb{R}^{m \times n}$).

2.1.1 Addition et multiplication des matrices

Les matrices peuvent être additionnées, multipliées par des constantes et dans certains cas multipliées entre elles.

Soient $m, n \in \mathbb{N}$ et deux matrices $A = (a_{ij})$ et $B = (b_{ij})$ de $\mathcal{M}_{mn}(\mathbb{R})$. De plus, soit $\lambda \in \mathbb{R}$. On défini les matrices $A + B \in \mathcal{M}_{mn}(\mathbb{R})$ et $\lambda \cdot A \in \mathcal{M}_{mn}(\mathbb{R})$ par

$$A + B = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix} \quad \text{et} \quad \lambda \cdot A = \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & \ddots & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{pmatrix}.$$

L'ensemble $\mathcal{M}_{mn}(\mathbb{R})$ avec les opérations définies au-dessus est un espace vectoriel: il satisfait les propriétés de la proposition 1.5.

Mais en plus des opérations d'addition et multiplication par un scalaire, les matrices peuvent également être multipliées. Soient $m, n, p \in \mathbb{N}$. Pour $A = (a_{ik}) \in \mathcal{M}_{mn}(\mathbb{R})$ et $B = (b_{kj}) \in \mathcal{M}_{np}(\mathbb{R})$ on définit le *produit des matrices* A et B comme la matrice $AB = (c_{ij}) \in \mathcal{M}_{mp}(\mathbb{R})$ (aussi notée $A \times B$) dont les entrées sont données par

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \qquad \text{pour } 1 \leq i \leq m \text{ et } 1 \leq j \leq p.$$

Attention! Pour des matrices A, B dont les tailles ne sont pas compatibles (comme dans la définition au-dessus) le produit de A et B n'est pas définit.

Les règles de la multiplication matricielle sont décrites dans la proposition suivante.

Proposition 2.2. *Soient $m, n, p, q \in \mathbb{N}$.*
(i) *Pour tout $A \in \mathcal{M}_{mn}(\mathbb{R})$ et $B, C \in \mathcal{M}_{np}(\mathbb{R})$ on a $A(B + C) = AB + AC$.*
(ii) *Pour tout $A, B \in \mathcal{M}_{mn}(\mathbb{R})$ et $C \in \mathcal{M}_{np}(\mathbb{R})$ on a $(A + B)C = AC + BC$.*
(iii) *Pour tout $A \in \mathcal{M}_{mn}(\mathbb{R})$ et $B \in \mathcal{M}_{np}(\mathbb{R})$ et $C \in \mathcal{M}_{pq}(\mathbb{R})$ on a $(AB)C = A(BC)$.*
(iv) *Pour tout $A \in \mathcal{M}_{mn}(\mathbb{R})$ et $B \in \mathcal{M}_{np}(\mathbb{R})$ et $\lambda \in \mathbb{R}$ on a $\lambda(AB) = (\lambda A)B = A(\lambda B)$.*

On va surtout utiliser la multiplication des matrices pour les matrices carrées, à savoir les matrices de $\mathcal{M}_{nn}(\mathbb{R})$ pour un $n \in \mathbb{N}$. On écrira $\mathcal{M}_n(\mathbb{R})$ à la place de $\mathcal{M}_{nn}(\mathbb{R})$ pour raccourcir la notation.

Attention! Généralement la multiplication des matrices n'est pas commutative. En effet, pour $n \geq 2$ ils existent $A, B \in \mathcal{M}_n(\mathbb{R})$ telles que $AB \neq BA$.

Exercice 2.1.
Trouver pour tout $n \geq 2$ des matrices $A, B \in \mathcal{M}_n(\mathbb{R})$ telles que $AB \neq BA$.

Proposition 2.3. *Soit $n \in \mathbb{N}$. Alors la matrice*

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

est l'unique matrice de $\mathcal{M}_n(\mathbb{R})$ telle que, pour tout $A \in \mathcal{M}_n(\mathbb{R})$,

$$AI_n = I_nA = A. \tag{2.1}$$

On appelle I_n la matrice identité *de taille n.*

Preuve: Commençons par montrer que 2.1 est satisfait pour tout $A \in \mathcal{M}_n(\mathbb{R})$. Par la règle de multiplication des matrices, pour une matrice $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$, on a

$$A \cdot I_n = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = A.$$

De même $I_n \cdot A = A$.

Montrons maintenant l'unicité de la matrice I_n. Soit $J \in \mathcal{M}_n(\mathbb{R})$ telle que $A \cdot J = J \cdot A = A$ pour tout $A \in \mathcal{M}_n(\mathbb{R})$. En appliquant cela à $A = I_n$, et en utilisant (2.1) on trouve

$$I_n = I_n \cdot J = J.$$

□

Exercice 2.2.
Montrer que pour tout $n \geq 2$ ils existent des matrices $A, B \in \mathcal{M}_n(\mathbb{R})$ telles que $AB = 0$, mais $A \neq 0$ et $B \neq 0$.

2.1.2 Matrices inversibles

Définition 2.4. *Soit $n \in \mathbb{N}$ et $A \in \mathcal{M}_n(\mathbb{R})$. On dit qu'une matrice $B \in \mathcal{M}_n(\mathbb{R})$ est un inverse de A si $AB = BA = I_n$. Si une telle matrice existe, on dit que A est* inversible(i).

L'étude des matrices inversibles est particulièrement intéressante. On donne une première proposition les concernant; on va y revenir dans les parties **??** et 2.2.

Proposition 2.5.
(i) *Si A est inversible, alors il existe un unique inverse qu'on note A^{-1}.*
(ii) *Soient $A, B \in \mathcal{M}_n(\mathbb{R})$ deux matrices inversibles. Alors AB est une matrice inversible et $(AB)^{-1} = B^{-1}A^{-1}$.*

Exercice 2.3.
Soient $A, B, C \in \mathcal{M}_n(\mathbb{R})$ des matrices telles que $AB = AC$. Sous quelle condition peut-on déduire que $B = C$?

Exemple: Regarder

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

(i) invertierbar

2.1.3 La transposition

Une opération spécifique aux matrices est la transposition. Elle consiste à échanger les lignes et les colonnes d'une matrice. On peut aussi la voir comme une réflexion de la matrice suivant sa diagonale.

Définition 2.6. *Soient $m, n \in \mathbb{N}$ et $A = (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}} \in \mathcal{M}_{mn}(\mathbb{R})$. La* transposée *de A est la matrice $A^T = (a^T_{ij})_{\substack{1\le i\le n\\ 1\le j\le m}} \in \mathcal{M}_{nm}(\mathbb{R})$ définie par $a^T_{ij} = a_{ji}$ pour tout $1 \le i \le n$ et $1 \le j \le m$.*

Donnons quelques exemples pour illustrer la notion.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

Proposition 2.7. *Soit $A, B \in \mathcal{M}_n(\mathbb{R})$. Alors $(AB)^T = B^T A^T$.*

Preuve: Montrons l'égalité élément par élément. Notons $A = (a_{ij})_{1\le i,j\le n}$ et $B = (b_{ij})_{1\le i,j\le n}$. On va aussi noter $a^T_{ij} = a_{ji}$ et $b^T_{ij} = b_{ji}$ les entrées de A^T et B^T.

Soient $k, \ell \in \{1, n\}$. L'entrée de la ligne k et la colonne ℓ de $(AB)^T$ est égale à l'entrée de la ligne ℓ et la colonne k de AB. Il s'agit donc de

$$(AB)^T_{k,\ell} = a_{\ell,1}b_{1k} + \cdots + a_{\ell,n}b_{nk}.$$

D'autre part, l'entrée de la ligne k et la colonne ℓ de $B^T A^T$ est

$$(B^T A^T)_{k,\ell} = b^T_{k1}a^T_{1,\ell} + \cdots + b^T_{kn}a^T_{n,\ell} = b_{1k}a_{\ell,1} + \cdots + b_{nk}a_{\ell,n}.$$

Ainsi $(AB)^T = B^T A^T$. □

Corollaire 2.8. *Soit $A \in \mathcal{M}_n(\mathbb{R})$. Alors A est inversible si et seulement si A^T l'est. Quand les deux sont inversibles, $(A^T)^{-1} = (A^{-1})^T$.*

Preuve: En utilisant la proposition 2.7 on vérifie que si A est inversible

$$(A^{-1})^T A^T = (AA^{-1})^T = I_n = (A^{-1}A)^T = A^T(A^{-1})^T.$$

Donc A^T est inversible et $(A^T)^{-1} = (A^{-1})^T$. □

Exercice 2.4.
Trouver toutes les matrices $A \in \mathcal{M}_n(\mathbb{R})$ telles que
(a) $A = A^T$.
(b) $A = -A^T$.

2.1.4 Formes particulières

Il est souvent commode d'écrire une matrice comme une collection de lignes ou de colonnes. Soit $A = (a_{ij}) \in \mathcal{M}_{mn}(\mathbb{R})$. On peut alors écrire

$$A = \begin{pmatrix} \\ C_1 \dots C_n \\ \\ \end{pmatrix} = \begin{pmatrix} L_1 \\ \vdots \\ L_m \end{pmatrix},$$

où $C_1, \dots, C_n$ sont les colonnes de A et $L_1, \dots, L_m$ sont les lignes. Plus précisément, pour $1 \leq i \leq m$ et $1 \leq j \leq n$,

$$L_i = (a_{i1}, \dots, a_{in}) \in \mathcal{M}_{1n} \quad \text{et} \quad C_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathcal{M}_{m1}.$$

Pour les matrices carrées, certaines formes de matrices sont particulièrement intéressantes.

Définition 2.9. *Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$. On dit que A est une* matrice triangulaire supérieure *si $a_{ij} = 0$ pour tout $i > j$, et* triangulaire inférieure, *si $a_{ij} = 0$ pour tout $i < j$.*

On dit que A est une matrice diagonale si $a_{ij} = 0$ pour tout $i \neq j$.

Les formes générales des matrices triangulaires supérieures, triangulaires inférieures et diagonales, respectivement, sont:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}; \qquad \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}; \qquad \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}.$$

Pour les matrices triangulaires il existe un critère simple pour déterminer l'inversibilité.

Lemme 2.10. *Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$. une matrice triangulaire supérieure, triangulaire inférieure ou diagonale. Alors A est inversible si et seulement si ses éléments diagonaux sont tous non-nuls. De plus, si A est diagonale, alors*

$$A^{-1} = \begin{pmatrix} 1/a_{11} & 0 & \dots & 0 \\ 0 & 1/a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/a_{nn} \end{pmatrix}. \tag{2.2}$$

La preuve va être donnée dans la partie 2.2.

2.1.5 Puissances des matrices

Soit $A \in \mathcal{M}_n(\mathbb{R})$. Comme pour les nombres, on peut multiplier A par elle même plusieurs fois pour obtenir les puissances de A. Ainsi on pose, pour $k \in \mathbb{N}$, $k \geq 1$:

$$A^k = \underbrace{A \dots A}_{k \text{ fois}}.$$

On pose aussi par convention $A^0 = I_n$. De plus, si A est inversible, on pose $A^{-k} = (A^{-1})^k$. Si A n'est pas inversible, les puissances négatives de A ne sont pas définies.

Certaines propriétés des puissances des nombres sont conservées. Pour $k, \ell \in \mathbb{N}$ on a bien

$$A^k A^\ell = A^{k+\ell} \quad \text{et} \quad \left(A^k\right)^\ell = A^{k\ell}.$$

D'autres ne le sont pas...

Exercice 2.5.
Trouver trois solutions distinctes $A \in \mathcal{M}_2(\mathbb{R})$ de l'équation $A^2 = A$.

Puissance des matrices diagonales Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice diagonale:

$$A = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}.$$

On peut alors facilement calculer les puissances de A:

$$A^k = \begin{pmatrix} a_{11}^k & 0 & \dots & 0 \\ 0 & a_{22}^n & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn}^n \end{pmatrix}, \quad \text{pour tout } k \in \mathbb{N}. \tag{2.3}$$

Exercice 2.6.
Montrer la formule (2.3) par récurrence sur n.

Le calcul des puissances d'une matrice est un problème très intéressant d'un point de vue pratique. On a vu quelques exemples où le calcul est possible, mais en général, il s'agit d'un problème compliqué, au quel on reviendra dans la parie 4. Quelques illustrations de l'utilité des puissances d'une matrice sont données dans la partie 2.5.

2.2 Image, noyau et rang d'une matrice

Commençons par définir l'image et le noyau d'une matrice $M \in \mathcal{M}_{mn}(\mathbb{R})$.

Définition 2.11. *On pose*

$$\begin{aligned}\operatorname{Im}(M) &= \big\{Y \in \mathcal{M}_{m1}(\mathbb{R}) : \exists X \in \mathcal{M}_{n1} \ t.q.\ MX = Y\big\} \subset \mathbb{R}^m,\\ \operatorname{Ker}(M) &= \big\{X \in \mathcal{M}_{n1}(\mathbb{R}) : MX = 0\big\} \subset \mathbb{R}^n,\\ \operatorname{rang}(M) &= \dim(\operatorname{Im}(M)).\end{aligned}$$

On peut facilement vérifier que $\operatorname{Im}(M)$ et $\operatorname{Ker}(M)$ sont des sous-espaces vectoriels de $\mathbb{R}^m$ et $\mathbb{R}^n$, respectivement. Leurs dimensions sont reliées par la relation suivante.

Théorème 2.12 (Théorème du rang pour les matrices)**.** *Soit $M \in \mathcal{M}_{mn}(\mathbb{R})$, alors*

$$\operatorname{rang}(M) + \dim(\operatorname{Ker}(M)) = n.$$

Pour les matrices, le rang peut se calculer de plusieurs manières, comme l'illustre la proposition suivante.

Proposition 2.13.

Soit $M \in \mathcal{M}_{mn}(\mathbb{R})$, $M = \begin{pmatrix} C_1 \dots C_n \end{pmatrix} = \begin{pmatrix} L_1 \\ \vdots \\ L_m \end{pmatrix}$. Alors $\operatorname{Im}(M) = \operatorname{Vect}(C_1, \dots, C_n)$ et

$$\operatorname{rang}(M) = \operatorname{rang}(C_1, \dots, C_n) = \operatorname{rang}(L_1, \dots, L_m).$$

Les preuves des deux résultats précédents sont données (partiellement) dans la partie 2.3. Une conséquence immédiate de la proposition 2.13 est le critère suivant pour l'invisibilité des matrices.

Corollaire 2.14. *Soit $M \in \mathcal{M}_n(\mathbb{R})$. Alors on a équivalence de*

(i) *la matrice M est inversible,*
(ii) $\operatorname{rang}(M) = n$,
(iii) *la famille des colonnes $C_1, \dots, C_n$ est libre,*
(iv) *la famille des lignes $L_1, \dots, L_n$ est libre,*
(v) $\operatorname{Ker}(M) = \{0\}$.

Preuve: (i)$\Leftrightarrow$(ii) : Si la matrice M est inversible, alors tout $X \in \mathbb{R}^n$ s'écrit $X = M\,(M^{-1}X)$. Ainsi $\operatorname{Im}(M) = \mathbb{R}^n$, donc $\operatorname{rang}(M) = n$.

Inversement, si $\mathrm{rang}(M) = n$, alors $\mathrm{Im}(M) = \mathbb{R}^n$. Ainsi, pour chaque vecteur de la base canonique $e_1, \dots, e_n$ de $\mathcal{M}_{n1}^n$, il existe $X_i \in \mathcal{M}_{n1}$ tel que $MX_i = e_i$. Alors, si on note M^{-1} la matrices dont les colonnes sont $X_1, \dots, X_n$, on a

$$M \cdot M^{-1} = \begin{pmatrix} MX_1 \dots MX_n \end{pmatrix} = \begin{pmatrix} e_1 \dots e_n \end{pmatrix} = I_n.$$

L'égalité $M^{-1}M = I_n$ s'en suit par une manipulation plus compliquée, qu'on ne va pas expliciter ici.

(ii)⇔(iii) : Par la proposition précédente, $\mathrm{rang}(M) = n$ revient à $\mathrm{rang}(C_1, \dots, C_n) = n$. Mais cela est équivalent au fait que la famille $(C_1, \dots, C_n)$ est libre.

(ii)⇔(iv) : Par la proposition précédente, $\mathrm{rang}(M) = n$ revient à $\mathrm{rang}(L_1, \dots, L_n) = n$. Mais cela est équivalent au fait que la famille $(L_1, \dots, L_n)$ est libre.

(ii)⇔(v) : Par le théorème du rang $\mathrm{rang}(M) = n - \dim(\mathrm{Ker}(M))$. Ainsi $\mathrm{rang}(M) = n$ si et seulement si $\dim(\mathrm{Ker}(M)) = 0$, donc $\mathrm{Ker}(M) = \{0\}$. □

Comme promis, donnons la preuve du Lemme 2.10.

Preuve du Lemme 2.10: Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ triangulaire inferieure. Notons $C_1, \dots, C_n \in \mathcal{M}_{n\,1}$ les colonnes de A.

Supposons que tous les a_{ii} sont non-nuls et montrons que la famille des colonnes est libre, donc que A est inversible. Soient $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ tels que

$$\sum_{i=1}^{n} \lambda_i C_i = \lambda_1 C_1 + \dots + \lambda_n C_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = 0.$$

Observons que la première coordonnée de $\sum_{i=1}^n \lambda_i C_i$ vaut simplement $\lambda_1 a_{11}$. Comme $a_{11} \neq 0$, on conclut que $\lambda_1 = 0$.

Alors la deuxième coordonnée de $\sum_{i=1}^n \lambda_i C_i$ vaut $\lambda_1 a_{21} + \lambda_2 a_{22} = \lambda_2 a_{22}$. Comme $a_{22} \neq 0$, on conclut que $\lambda_2 = 0$. On continue ainsi et on montre que $\lambda_1 = \lambda_2 = \dots = \lambda_n = 0$. Ainsi la famille $C_1, \dots, C_n$ est libre, donc A est inversible par le corollaire 2.14.

Inversement, supposons qu'il existe au moins un 0 sur la diagonale de A: $a_{ii} = 0$ pour un certain i. Alors les colonnes $C_1, \dots, C_i$ sont toutes des vecteurs du sous-espace vectoriel

$$F = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : x_i = x_{i+1} = \dots = x_n = 0 \right\}.$$

Cet espace a dimension $i - 1$, d'ou on déduit que la famille $C_1, \dots, C_i$ est liée. Ainsi $C_1, \dots, C_n$ est également liée et le corollaire 2.14 implique que A n'est pas inversible.

Le résultat pour les matrices triangulaires supérieures s'obtient par transposition. Celui pour les matrices diagonales est un cas particulier des résultats pour les matrices triangulaires. En fin, si A est diagonale et $a_{ii} \neq 0$ pour tout i, on montre par calcul direct que le produit entre A et la matrice donnée dans (2.2) est égal à I_n. □

Lemme 2.15. *Soit $M \in \mathcal{M}_{mn}(\mathbb{R})$, $P \in \mathcal{M}_m(\mathbb{R})$ et $Q \in \mathcal{M}_n(\mathbb{R})$. Alors*

(i) *si* P *est inversible, alors* $\mathrm{Ker}(M) = \mathrm{Ker}(PM)$ *et* $\mathrm{rang}(M) = \mathrm{rang}(PM)$*;*

(ii) *si* Q *est inversible, alors* $\mathrm{Im}(M) = \mathrm{Im}(MQ)$ *et* $\mathrm{rang}(M) = \mathrm{rang}(MQ)$*.*

Ce qu'il faut retenir de ce lemme est que le rang d'une matrice n'est pas modifié si on la multiplie, à gauche ou à droite, par des matrices inversibles. On laisse la preuve en exercice.

Exercice 2.7.
Démonter le lemme 2.15.

Exercice 2.8.
Soient $A, B \in \mathcal{M}_n(\mathbb{R})$ tels que $AB = I_n$. Montrer que A est inversible et que $A^{-1} = B$. Trouver un couple de matrices $A \in \mathcal{M}_{mn}(\mathbb{R})$ et $B \in \mathcal{M}_{nm}(\mathbb{R})$ pour $m < n$, tels que $AB = I_m$. Calculer BA. Que dire des images et noyaux de A et B?

2.3 Systèmes linéaires et matrices

Considérons le système linéaire de m équations à n inconnues

$$\left\{ \begin{array}{l} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \dots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m. \end{array} \right. \tag{2.4}$$

ou $m, n \in \mathbb{N}$ et $(a_{ij})_{\substack{1\leq i\leq m,\\ 1\leq j\leq n}}$, $(b_i)_{1\leq i\leq m}$ deux familles de d'éléments de $\mathbb{R}$. On peut écrire ce système sous forme matricielle de la façon suivante. Soient

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \in \mathcal{M}_{mn}(\mathbb{R}) \quad \text{et} \quad B = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathcal{M}_{m1}(\mathbb{R}).$$

Alors, (2.4) devient

$$AX = B, \tag{2.5}$$

ou $X \in \mathcal{M}_{n1}(\mathbb{R})$ est le vecteur des inconnues $x_1, \dots, x_n$. On peut ainsi écrire

$$\mathcal{S} = \left\{ X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathcal{M}_{n1}(\mathbb{R}) : \ AX = B \right\}.$$

Ainsi, $\mathcal{S} \neq \emptyset$ (c.à d. que le système admet au moins une solution) si et seulement si $B \in \mathrm{Im}(A)$. Supposons que $B \in \mathrm{Im}(A)$, et soit $X_0 \in \mathcal{M}_{n1}(\mathbb{R})$ une solution. Alors, pour toute solution $X \in \mathcal{S}$,

$$A(X - X_0) = AX - AX_0 = B - B = 0,$$

donc $X - X_0 \in \mathrm{Ker}(A)$. Inversement, si $X \in \mathcal{M}_{n1}(\mathbb{R})$ est tel que $X - X_0 \in \mathrm{Ker}(A)$, alors le calcul précédent montre que $X \in \mathcal{S}$. On arrive ainsi à la conclusion suivante.

Théorème 2.16. *Soient* $A \in \mathcal{M}_{mn}(\mathbb{R})$ *et* $B \in \mathcal{M}_{m1}(\mathbb{R})$. *On note* $\mathcal{S}$ *l'ensemble des solutions* $X \in \mathcal{M}_{n1}(\mathbb{R})$ *de* $AX = B$. *Alors*

(i) *si* $B \notin \mathrm{Im}(A)$, *alors* $\mathcal{S} = \emptyset$,

(ii) *si* $B \in \mathrm{Im}(A)$, *alors* $\mathcal{S} \neq \emptyset$. *De plus, si* X_0 *est une solution, on a*

$$\mathcal{S} = X_0 + \mathrm{Ker}(A) = \{X_0 + Y : Y \in \mathrm{Ker}(A)\}.$$

Dans le cas *(ii)* on dit que X_0 est une solution particulière du système et que Y est une solution générale du système homogène. En effet, que $Y \in \mathrm{Ker}(A)$ s'écrit aussi $AY = 0$, ou encore,

$$\begin{cases} a_{11}y_1 + \cdots + a_{1n}y_n = 0 \\ \dots \\ a_{m1}y_1 + \cdots + a_{mn}y_n = 0. \end{cases}$$

On reconnait ici le système homogène associé à (2.4).

Rappelons nous que $\mathrm{Ker}(A)$ est un s.e.v. de $\mathbb{R}^n$. Ainsi, quand le système admet des solutions, $\mathcal{S}$ est le translaté d'un espace vectoriel par un vecteur X_0. On appelle ce type d'espace un *espace affine* et on pose

$$\dim(\mathcal{S}) = \dim(\mathrm{Ker}(A)).$$

Une conséquence du théorème du rang qui peut être utile est que $\dim(\mathrm{Ker}(A)) = n - \mathrm{rang}(A)$.

Preuve du Théorème 2.12: Soit $M \in \mathcal{M}_{mn}(\mathbb{R})$. Alors $\mathrm{Ker}(M)$ est l'ensemble des solutions du système linéaire homogène $MX = 0$. Si on note $\tilde{M}$ la forme échelonnée de M et k le nombre de lignes non-nulles de $\tilde{M}$, on rappelle de la partie 1.6 que $\dim(\mathrm{Ker}(M)) = \dim(\mathcal{S}_{hom}) = n - k$.

D'autre par, si on écrit $C_1, \dots, C_n$ pour les colonnes de M et on considère $X \in \mathcal{M}_{n1}(\mathbb{R})$, alors $MX = x_1C_1 + \dots, x_nC_n$. Ainsi $\mathrm{Im}(M) = \mathrm{Vect}(C_1, \dots, C_n)$. Comme expliqué dans la partie 1.6, en utilisant la notation de (1.7), les colonnes $C_{j_1}, \dots, C_{j_k}$ forment une base de $\mathrm{Vect}(C_1, \dots, C_n) = \mathrm{Im}(M)$, donc $\mathrm{rang}(M) = k$. L'égalité désirée s'en suit. □

Preuve de la Proposition 2.13: Soit $M \in \mathcal{M}_{mn}(\mathbb{R})$ dont les lignes sont $L_1, \dots, L_m$ et les colonnes $C_1, \dots, C_n$. Dans la preuve précédente on a montré que $\mathrm{Im}(M) = \mathrm{Vect}(C_1, \dots, C_n)$, ce qui implique directement que $\mathrm{rang}(M) = \mathrm{rang}(C_1, \dots, C_n)$.

La dernière égalité (à savoir $\mathrm{rang}(C_1, \dots, C_n) = \mathrm{rang}(L_1, \dots, L_m)$) est plus délicate et nécessite une construction qu'on ne traite pas dans ce cours. On va l'admettre.

□

2.4 Inverse, image, noyau par le pivot de Gauss

Il est souvent intéressant de vérifier si A est inversible et de calculer son inverse A^{-1}. Le pivot de Gauss offre une procédure pratique pour faire cela.

Commençons par définir les matrices $\mathrm{Add}_{i;\lambda,j}$ (pour $\lambda \in \mathbb{K}$ et $i \neq j$), $\mathrm{Mult}_{\lambda,i}$ (pour $\lambda \neq 0$) et $\mathrm{Ech}_{i;j}$ (pour $i \neq j$) de $\mathcal{M}_m(\mathbb{R})$ comme ayant les formes suivantes:

$$\begin{array}{cc} \begin{matrix} & & & j & \\ & & & \downarrow & \end{matrix} & \\ \begin{pmatrix} 1 & & 0 & \lambda & 0 \\ & 1 & & & \\ & & \ddots & & 0 \\ & 0 & & 1 & \\ & & & & 1 \end{pmatrix} & \leftarrow i \end{array} \qquad \begin{array}{cc} \begin{matrix} & & i & & \\ & & \downarrow & & \end{matrix} & \\ \begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \\ & & & \lambda & \\ & 0 & & 1 & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} & \leftarrow i \end{array} \qquad \begin{array}{cc} \begin{matrix} & i & & j & \\ & \downarrow & & \downarrow & \end{matrix} & \\ \begin{pmatrix} 1 & & & & \\ & \ddots & 0 & 0 & \\ & 1 & & & \\ & & 0 & 1 & \\ 0 & & 1 & & 0 \\ & & \ddots & & \\ & & & 1 & \\ & & 1 & 0 & \\ & & & 1 & \\ 0 & 0 & & & \ddots \\ & & & & 1 \end{pmatrix} & \begin{matrix} \\ \\ \\ \leftarrow i \\ \\ \\ \\ \leftarrow j \\ \\ \\ \\ \end{matrix} \end{array}$$

‘

Proposition 2.17. *Soient $A \in \mathcal{M}_{mn}(\mathbb{R})$, $1 \leq i, j \leq m$ distincts et $\lambda \in \mathbb{R}^*$. Alors*

- $\mathrm{Add}_{i;\lambda,j}A$ *est la matrice obtenue à partir de A en rajoutant à la ligne i la ligne j multipliée par λ;*
- $\mathrm{Mult}_{\lambda,i}A$ *est la matrice obtenue à partir de A en multipliant la ligne i par λ;*
- $\mathrm{Ech}_{i;j}A$ *est la matrice obtenue à partir de A en échangeant les lignes i et j.*

De plus, les matrices $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ et $\mathrm{Ech}_{i;j}$ sont inversibles.

Preuve: L'effet de la multiplication de M par $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ et $\mathrm{Ech}_{i;j}$ peut être vérifié par calcul direct. Les inverses de $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ et $\mathrm{Ech}_{i;j}$ sont les matrices $\mathrm{Add}_{i;-\lambda,j}$, $\mathrm{Mult}_{1/\lambda,i}$ et $\mathrm{Ech}_{i;j}$, respectivement. □

On dit qu'une matrice $A \in \mathcal{M}_{mn}(\mathbb{R})$ est *échelonnée selon les lignes* si A est de la forme

$$A = \begin{pmatrix} 0 \dots 0\, 1 * \dots * & 0 * \dots * & 0 & \dots * & 0 * \dots * \\ 0 \ \dots \quad \dots & 0\, 1 * \dots * & 0 & \dots * & 0 * \dots * \\ 0 \ \dots \quad \dots \ \dots & \dots & 0\, 1 & \dots * & 0 * \dots * \\ \vdots & & & \ddots & \vdots \quad\ \ \vdots \\ 0 \ \dots \quad \dots \ \dots & \dots \quad \dots & \dots & \dots & 0\, 1 * \dots * \\ 0 \ \dots \quad \dots \ \dots & \dots \quad \dots & \dots & & 0 \ \dots \ 0 \\ \vdots & & & & \vdots \quad\ \ \vdots \\ 0 \ \dots \quad \dots \ \dots & \dots \quad \dots & \dots & & 0 \ \dots \ 0 \end{pmatrix}, \tag{2.6}$$

où les $*$ représentent des nombres quelconques.

L'algorithme du pivot de Gauss peut s'appliquer aux matrices pour les réduire à une matrice échelonnée. Chaque pas de l'algorithme correspond à une multiplication à gauche par une matrice $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ ou $\mathrm{Ech}_{i;j}$. Le résultat suivant est une consequence de l'algorithme du pivot de Gauss.

Lemme 2.18. *Soit $A \in \mathcal{M}_{mn}(\mathbb{R})$. Alors il existe une suite $P_1, \dots, P_k$ de matrices de la forme $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ ou $\mathrm{Ech}_{i;j}$ avec $i \neq j$ et $\lambda \neq 0$ telle que $P_k \dots P_1 A$ est échelonnée.*

Remarque 2.19. Si on reviens au système $AX = B$ avec $A \in\in \mathcal{M}_{mn}(\mathbb{R})$, $X \in \mathcal{M}_{n,1}(\mathbb{R})$ et $B \in \mathcal{M}_{m,1}(\mathbb{R})$ et on note $P_1, \dots, P_k$ les matrices transformant A en une matrice échelonnée, alors

$$AX = B \quad \text{si et seulement si } P_k \dots P_1 AX = P_k \dots P_1 B.$$

L'equivalence entre ces deux égalités est garantie par le fait que $P_1, \dots, P_k$ sont inversibles. L'avantage de l'expression $P_k \dots P_1 AX = P_k \dots P_1 B$ est qu'il s'agit d'un système sous forme échelonnée, donc facilement resoluble.

Cette remarque resume la resolution des systèmes linéaires en utilisant le pivot de Gauss, déjà décrite dans la partie 1.3.

Le pivot de Gauss pour les matrices nous permet aussi de calculer l'inverse d'une matrice. Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice carrée. Appliquons le pivot de Gauss à A et à la matrice identité I_n simultanément. Plus précisément, chaque transformation appliquée à A dans l'algorithme du pivot de Gauss, est aussi appliquée à I_n. Notons $\tilde{A}$ la forme échelonnée de A qui en résulte, et $\tilde{I}$ le résultat pour I_n.

Proposition 2.20.
Si $\tilde{A} \neq I_n$, alors A n'est pas inversible.
Si $\tilde{A} = I_n$, alors A est inversible et $A^{-1} = \tilde{I}$.

Preuve: Pendant le pivot de Gauss, on applique des transformations à A qui correspondent à des multiplications à gauche par des matrices inversibles $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ ou $\mathrm{Ech}_{i;j}$. Notons ces matrices $P_1, \dots, P_k$ dans l'ordre d'application. Ainsi $\tilde{A} = P_k \dots P_1 A$ et $\tilde{I} = P_k \dots P_1 I_n = P_k \dots P_1$.

Si $\tilde{A}$ n'est pas égale à I_n, alors elle contient au moins une ligne nulle, donc n'est pas inversible. Comme $P = P_k \dots P_1$ est inversible, cela implique que A n'est pas inversible non-plus (car le produit de deux matrices inversibles est forcement inversible − voir la proposition 2.5).

Si $\tilde{A} = PA = I_n$ alors $A = P^{-1}$ est inversible et $A^{-1} = P = P_k \dots P_1 = \tilde{I}$. □

Exercice 2.9.
Déterminer si les matrices suivantes sont inversibles. Si c'est le cas, calculer leur inverse. Multiplier la matrice par le résultat pour vérifier.

$$a) \begin{pmatrix} 1 & 2 & 3 \\ 3 & 7 & 6 \\ 2 & 8 & -5 \end{pmatrix} \in \mathcal{M}_3(\mathbb{R}); \qquad b) \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \in \mathcal{M}_5(\mathbb{R});$$

$$c) \begin{pmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \in \mathcal{M}_5(\mathbb{R}); \qquad d) \begin{pmatrix} 1 & -1 & \dots & -1 & -1 \\ 0 & 1 & \dots & -1 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \in \mathcal{M}_n(\mathbb{R});$$

$$e)\begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 2 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ n-1 & n-2 & \dots & 1 & 0 \\ n & n-1 & \dots & 2 & 1 \end{pmatrix} \in \mathcal{M}_n(\mathbb{R}); \qquad f)\begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ 0 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \in \mathcal{M}_n(\mathbb{R}).$$

Rang d'une familles de vecteurs par pivot de Gauss. Soit $C_1, \dots, C_n \in \mathcal{M}_{m1}$ une famille de vecteurs de $\mathbb{R}^m$ écrits en format colonne. A l'aide du pivot de Gauss on peut déterminer si la famille $C_1, \dots, C_n$ est libre; plus généralement on peut calculer son rang.

En effet, posons $A = \Big(C_1, \dots, C_n\Big) \in \mathcal{M}_{mn}$ la matrice formée des colonnes $C_1, \dots, C_n$. Alors $\mathrm{rang}(C_1, \dots, C_n) = \mathrm{rang}(A) = \mathrm{rang}(\tilde{A})$, ou $\tilde{A}$ est la matrice échelonnée obtenue à partir de A par le pivot de Gauss. En particulier, la famille $C_1, \dots, C_n$ est libre si et seulement si $\mathrm{rang}(\tilde{A}) = n$.

Mentionnons qu'on peut aussi déterminer les scalaires $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ tels que

$$\lambda_1 C_1 + \dots + \lambda_n C_n = 0. \tag{2.7}$$

Si on pose $\Lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$, alors (2.7) s'écrit

$$A\Lambda = \lambda_1 C_1 + \dots + \lambda_n C_n = 0,$$

ce qui revient à dire que $\Lambda \in \mathrm{Ker}(A) = \mathrm{Ker}(\tilde{A})$.

En conclusion, une famille $\lambda_1, \dots, \lambda_n$ satisfait (2.7) si et seulement si le vecteur colonne qu'elle forme est dans $\mathrm{Ker}(A) = \mathrm{Ker}(\tilde{A})$. On peut donc, de façon alternative, voir si la famille $C_1, \dots, C_n$ est libre en calculant $\mathrm{Ker}(\tilde{A})$. La famille est libre si et seulement si l'unique famille de scalaires satisfaisant (2.7) est nulle, donc si et seulement si $\mathrm{Ker}(A) = \mathrm{Ker}(\tilde{A}) = \{0\}$.

Image et noyau par pivot de Gauss. Soit $A = (a_{ij})_{ij} \in \mathcal{M}_{mn}(\mathbb{R})$ une matrice dont on veut trouver le noyau et l'image. Rappelons que $\mathrm{Im}(A)$ est un sous-espace vectoriel de $\mathbb{R}^m$ et $\mathrm{Ker}(A)$ un sous-espace vectoriel de $\mathbb{R}^n$.

Si $\dim(\mathrm{Im}(A)) = \mathrm{rang}(A) = m$, alors l'image est simple à décrire: $\mathrm{Im}(A) = \mathbb{R}^m$; de plus, si $\dim(\mathrm{Ker}(A)) = 0$, alors $\mathrm{Ker}(A) = \{0\}$. A part dans ces cas extremes, la meilleure façon de décrire $\mathrm{Im}(A)$ et $\mathrm{Ker}(A)$ est en donnant une base de chaque espace.

Pour faire cela, notons $C_1, \dots, C_n$ les colonnes de A et écrivons $\tilde{A} = (\tilde{a}_{ij})_{i,j}$ pour la forme échelonnée associée à A. Rappelons que les colonnes de $\tilde{A}$ sont de deux types: les colonnes $j_1, \dots, j_k$ contenant un 1 entouré de zeros, et les autres (contenant des $*$ dans (2.6)).

Alors $\dim(\mathrm{Ker}(A)) = n - k$ donc toute base de $\mathrm{Ker}(A)$ contient $n - k$ vecteurs; Une telle base peut être construite comme suit. À chaque $j \notin \{j_1, \dots, j_k\}$ on associe le vecteur

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{où} \quad x_i = \begin{cases} 1 & \text{si } i = j; \\ -\tilde{a}_{kj} & \text{si } i = j_\ell, \text{ pour } \ell = 1, \dots, k; \\ 0 & \text{sinon.} \end{cases}$$

L'image de A a dimension k et une base de $\mathrm{Im}(A)$ est formée des colonnes $C_{j_1}, \dots, C_{j_k}$ de A (voir aussi la preuve du Théorème 2.12).

2.5 Applications: les matrices comme outil de modélisation

Les matrices peuvent être utilisées pour décrire une grande variété de problèmes, comme par exemple des évolutions déterministes ou stochastiques (voir plus bas). Pour plus d'exemples d'utilisations des matrices, voir Chapitre 10 de [1].

2.5.1 Population de deux types de bactéries

Supposons qu'on a une population de bactéries de deux types A et B. A chaque instant, une bactérie de type A donne naissance à une bactérie de type A et quatre de type B, puis meure. Une bactérie de type B donne naissance à une bactérie de type A et une de type B. Quelle est la population après n étapes, sachant qu'on commence par une seule bactérie de type A?

On peut modéliser cela comme suit. Notons a_n et b_n le nombre de bactéries de type A et B, respectivement, à l'étape n. Alors on a la formule suivante:

$$a_{n+1} = a_n + b_n \quad \text{et} \quad b_{n+1} = 4a_n + b_n. \tag{2.8}$$

De plus la condition initiale est $a_0 = 1$ et $b_0 = 0$.

La population totale de bactéries après n est alors $p_n = a_n + b_n$. Remarquons que $p_{n+1} = 5a_n + 2b_n$ (grace aux deux équations précédentes) ne s'écrit pas simplement en fonction de p_n. Par contre, si on note $X_n = \begin{pmatrix} a_n \\ b_n \end{pmatrix}$ pour chaque $n \in \mathbb{N}$, on observe que

$$X_{n+1} = \begin{pmatrix} a_n + b_n \\ 4a_n + b_n \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} X_n.$$

Ainsi

$$X_n = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}^n X_0 = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}^n \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Malheureusement il n'est pas evident de deviner la forme générale de la matrice $\begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}^n$. On verra dans la partie 4 comment le faire pour cette matrice, ainsi que pour (presque) toute autre matrice. Pour l'instant, on va résoudre ce problème par une astuce: si on pose $c_n = 2a_n - b_n$ et $d_n = 2a_n + b_n$ pour tout n, alors on observe que

$$c_{n+1} = -c_n \quad \text{et} \quad d_{n+1} = 3d_n, \quad \forall n \in \mathbb{N}.$$

Ainsi

$$\begin{pmatrix} c_n \\ d_n \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}^n \begin{pmatrix} c_0 \\ d_0 \end{pmatrix} = \begin{pmatrix} (-1)^n c_0 \\ 3^n d_0 \end{pmatrix} = \begin{pmatrix} 2 \cdot (-1)^n \\ 2 \cdot 3^n \end{pmatrix}.$$

Cela suffit pour retrouver a_n et b_n car

$$a_n = \frac{1}{4}(c_n + d_n) = \frac{1}{2}(3^n + (-1)^n) \quad \text{et} \quad b_n = \frac{1}{2}(d_n - c_n) = 3^n - (-1)^n.$$

Interlude Dans l'exemple précédent on observe que les population de types a et b augmentent comme 3^n (les facteurs $1/2$ et $(-1)^n$ sont peu importants). On appelle ce type de croissance (à savoir comme α^n avec $\alpha > 1$) une croissance *géométrique* ou *exponentielle*; elle est souvent rencontrée dans les évolutions des populations. Il s'agit d'une croissance très rapide, comme illustré par l'anecdote suivante.

Un voyageur arriva à la cour de l'empereur de Chine et lui apprit le jeu d'échecs. L'empereur de chine aima tant le jeu qu'il offera à son inventeur de satisfaire ses désirs. Celui-ci ne demanda que du riz. "Je voudrais un grain de riz sur la première case de l'échiquier, deux sur la deuxième, quatre sur la troisième, et ainsi de suite jusqu'à la dernière case." L'empereur s'empressa d'accepter, pensant que tout cela ne représentait que deux ou trois sacs de riz.

Calculons la quantité de riz que l'empereur devait au voyageur. L'échiquier est de taille 8 sur 8, donc contient $8^2 = 64$ cases, qu'on peut numéroter de 1 à 64. Le nombre de grains sur la case k est alors 2^k; c'est en effet $2 \cdot 2^{k-1}$, donc le double de la case précédente. Ainsi, le nombre total de grains de riz dus au voyageur est

$$N = 1 + 2 + 4 + \cdots + 2^{63} = \sum_{k=0}^{63} 2^k.$$

On peut calculer ce nombre à l'aide de la formule qui donne la somme d'une série géométrique ($\sum_{k=0}^{n} \alpha^k = \frac{\alpha^{n+1}-1}{\alpha-1}$). Dans le cas présent, cela donne

$$N = 2^{64} - 1.$$

Une première chose qu'on observe (et qui devrait inquiéter l'empereur) est que le nombre total est le double de ce qui est placé sur la dernière case (à un grain près).

Ca peut être utile de connaitre les premières puissances de 2 et d'avoir une technique pour estimer rapidement les puissances supérieures:

k	0	1	2	3	4	5	6	7	8	9	10
2^k	1	2	4	8	16	32	64	128	256	512	1024

Pour estimer des puissances plus grandes, comme 2^{64}, on utilisera l'approximation $2^{10} = 1024 \approx 1000 = 10^3$. Ainsi on trouve

$$2^{65} = 2^4 \cdot (2^{10})^6 \approx 2^4 \cdot (10^3)^6 = 16 \cdot 10^{18}.$$

Cela va plus nous parler si on l'exprime en grammes, kilos ou tonnes... On considérera qu'un grain de riz pèse 0.025g. Alors

$$\begin{aligned} 16 \cdot 10^{18} \text{ grains} &= 16 \cdot 0.025 \cdot 10^{18}\text{g} = 0.4 \cdot 10^{18}\text{g} \\ &= 4 \cdot 10^{17}\text{g} = 4 \cdot 10^{14}\text{kg} = 4 \cdot 10^{11}\text{t} = 400\ 000\ 000\ 000\text{t}. \end{aligned}$$

Pour comprendre cette somme, mentionnons que la production totale de riz est actuellement de 700 000 000t. Ainsi, ca prendrait plus de 500 ans pour produire suffisamment de riz pour payer le voyageur.

2.5.2 Matrices d'adjacence et matrices stochastiques

Un graphe est un couple d'ensembles $G = (V, E)$ où les éléments de V sont les sommets de G et $E \subset V \times V$ représente les arêtes de G. Il symbolise un réseau avec des arêtes orientées (ou pas, suivant la définition exacte) entre certains de ses sommets. Pour un graphe fini G (c.à-d. avec V fini), on définit sa matrice d'adjacence comme suit.

On commence par noter $V = \{v_1, \dots, v_n\}$. La matrice d'adjacence $A = (a_{ij})_{1\le i,j\le n} \in \mathcal{M}_n(\mathbb{R})$ est définie par

$$a_{ij} = 1 \text{ si et seulement si } (v_i, v_j) \in E,$$

donc si et seulement s'il y a une arête de v_i vers v_j. Sinon, on pose $a_{ij} = 0$. Ainsi la matrice d'adjacence décrit parfaitement le graphe G. Un graphe non-orienté est décrit par une matrice d'adjacence symétrique ($A = A^T$ car $a_{ij} = a_{ji}$).

Si de plus le graphe G a un poids positif w_e associé à chaque arête $e \in E$, on peut rajouter cette information à la matrice A, en posant

$$a_{ij} = \begin{cases} w_{(v_i,v_j)} & \text{si } (v_i, v_j) \in E, \\ 0 & \text{si } (v_i, v_j) \notin E. \end{cases}$$

Un vecteur colonne $X \in \mathcal{M}_{n1}(\mathbb{R})$ (avec des entrées positives) peut être interprété comme une attribution de “masse” à chaque sommet: chaque sommet v_i a une masse x_i.

Imaginons la dynamique suivante: à chaque étape, chaque sommet v_i transfère une proportion a_{ij} de sa masse au sommet v_j. Pour que cela ai un sens, il faut que

$$\sum_{j=1}^{n} a_{ij} = 1, \qquad \forall 1 \le i \le n. \tag{2.9}$$

Alors, si on commence par une attribution de masse X, après une étape, on aura une distribution de masse Y, où

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad \text{avec} \quad y_j = \sum_{i=1}^{n} a_{ij} x_i, \ \forall 1 \le j \le n.$$

On peut écrire cela de façon plus compacte comme $Y = A^T X$. Plus généralement, après k étapes, on aura une distribution $(A^T)^k X$.

La condition (2.9) garantie qu'il y a conservation de la masse totale dans le graphe. Si cette condition est satisfaite, on dit que A est une *matrice stochastique* à droite et A^T une matrice stochastique à gauche. Ainsi, dans une matrice stochastique à droite, la somme des entrées sur chaque ligne vaut 1; dans une matrice stochastique à gauche, c'est la somme des entrées sur chaque colonne qui vaut 1.

Cette dynamique décrit une *marche aléatoire* sur le graphe G (un cas particulier d'une *chaîne de Markov*). Imaginons qu'on dispose d'un kilo de sable. On se donne un vecteur colonne $X = (x_i)_{1\le i\le n} \in \mathcal{M}_{n1}$ avec

$$x_i \ge 0, \text{ pour tout } 1 \le i \le n \quad \text{ et } \quad \sum_{i=1}^{n} x_i = 1.$$

Pour chaque i, on place x_i kilos de sable au sommet v_i de G. On lance ensuite la dynamique décrite avant: le sable de v_i est distribué entre ses voisins, v_j recevant de v_i une proportion a_{ij} de sable.

Supposons qu'un grain de sable est coloré en rouge et qu'on le suit pendent cette évolution. Alors, à chaque étape, si le grain se trouve à v_i, il a une probabilité a_{ij} d'être envoyé au sommet v_j. On va supposer que cette dynamique est markovienne, c.à-d. que l'évolution future du grain de sable dépend uniquement de sa position actuelle, pas de son évolution passée.

Ainsi, la position du grain rouge après k étapes est décrite par le vecteur

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \left(A^T\right)^k X.$$

En effet, il y a une quantité y_i de sable au sommet v_i, donc une probabilité y_i que le grain rouge s'y trouve.

Exercice 2.10.
Soit $A \in \mathcal{M}_n$. Montrer que A est stochastique à droite si et seulement si

$$A \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

De plus, A est stochastique à gauche si et seulement si $\begin{pmatrix} 1 & \dots & 1 \end{pmatrix} A = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}$.

Exemple: la parade nuptiale des bourdons. Comme discuté avant, les matrices peuvent être utilisées pour décrire certaines évolutions aléatoires appelées des *chaines de Markov*. On va illustrer cela par un exemple concret.

L'accouplement des bourdons suit une procédure à plusieurs étapes. Celles-ci peuvent être classées comme suit:

- (App) Approche: un mâle se dirige vers la reine. Il s'approche à courte distance de la reine, et peut continuer la parade, ou se retirer (c'est le plus souvent la cas).
- (IF) Inspection de la femelle: le mâle suit la reine avec ses antennes tendues vers elle. Il inspecte souvent la reine au niveau de la tête (région où se trouvent les glandes produisant les phéromones sexuelles), mais parfois au niveau de l'abdomen.
- (T) Tentative d'accouplement: le mâle s'approche de la reine, il s'accroche à elle. Il frotte de ses pattes antérieures l'extrémité de l'abdomen de la femelle. Il sort ses génitalias (appareil reproducteur) et tente de pénétrer la reine.
- (Acc) Accouplement: lors de l'accouplement, le comportement du mâle se caractérise par des mouvements de battements des pattes sur l'extrémité de l'abdomen de la reine.

Pour observer la séquence, on place 80 bourdons dans un milieu favorable, et on les observe pendant 15 minutes. Les bourdons passent par les différentes phases de la parade nuptiale. On observe chaque bourdon chaque minute pour déterminer les phases par lesquelles il passe. On rajoute quelques états dans notre tableau:

(D) depart: la situation de depart.
(AA) Accouplement accompli: le mâle quitte la séance après accouplement.
(AM) Abandon du mâle: lors de la séquence, le bourdon mâle peut adopter un comportement indifférent vis-à-vis de la reine; il sort de la parade nuptiale et n'y revient jamais.

Les observations sont classées dans le tableau suivant:

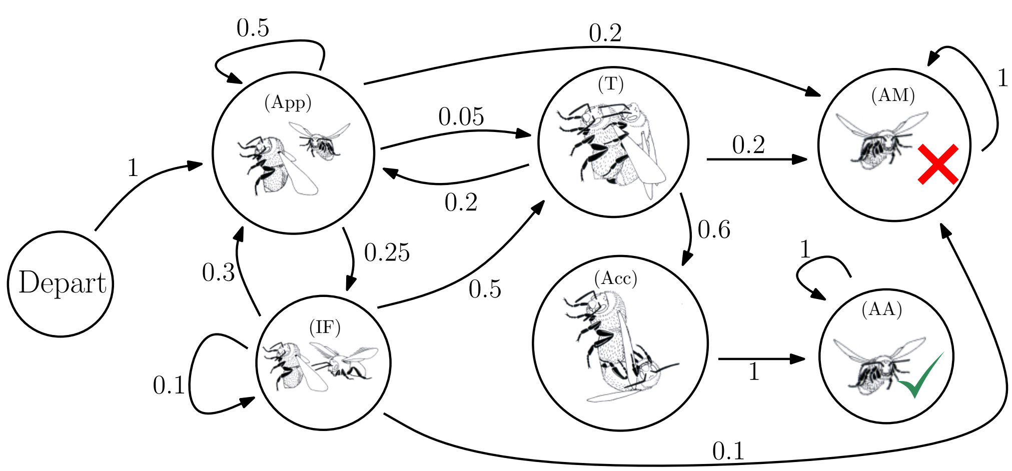
	Vers					
De ↓	App	IF	T	Acc	AM	Total
D	80	0	0	0	0	80
App	102	51	10	0	41	204
IF	16	6	28	0	7	57
T	6	0	0	22	10	38

Si on suppose l'évolution des bourdons markovienne, on peut estimer à partir de ces observations les probabilités de passage d'une phase à une autre. Elles sont obtenues en divisant chaque ligne du tableau précédent par le nombre total d'observations y correspondent. Rajoutons les transitions à partir des états Acc, AM et AA:

	App	IF	T	Acc	AM	AA
D	1	0	0	0	0	0
App	$\frac{102}{204}=0.5$	$\frac{51}{204}=0.25$	$\frac{10}{204}=0.05$	0	$\frac{41}{204}=0.2$	0
IF	$\frac{16}{57}=0.3$	$\frac{6}{57}=0.1$	$\frac{28}{57}=0.5$	0	$\frac{7}{57}=0.1$	0
T	$\frac{6}{38}=0.15$	0	0	$\frac{22}{38}=0.6$	$\frac{10}{38}=0.25$	0
Acc	0	0	0	0	0	1
AM	0	0	0	0	1	0
AA	0	0	0	0	0	1

(2.10)

Cette évolution peut être décrite par le graphe suivant



La matrice associé à ce graphe (en ignorant l'état D) est

$$A = \begin{pmatrix} 0.5 & 0.25 & 0.05 & 0 & 0.2 & 0 \\ 0.3 & 0.1 & 0.5 & 0 & 0.1 & 0 \\ 0.15 & 0 & 0 & 0.6 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{et} \quad A^T = \begin{pmatrix} 0.5 & 0.3 & 0.15 & 0 & 0 & 0 \\ 0.25 & 0.1 & 0 & 0 & 0 & 0 \\ 0.05 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0 & 0 \\ 0.2 & 0.1 & 0.25 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Exercice 2.11.
Vérifier que les matrices A et A^T sont stochastiques (à droite et à gauche, respectivement). Si on suppose que les bourdons se comportent de façon markovienne, donner une distribution, minute par minute, des bourdons entre les différents états de la parade nuptiale. Pour cela, prenons X_0 le vecteur colonne avec 80 en première coordonnée et 0 ailleurs, et posons $X_{k+1} = A^T X_k$ pour $k \geq 0$. Alors X_k contient le nombre moyen de bourdons dans chaque état de la parade après k minutes.
Est-ce que les observations sont cohérentes avec les simulations? Pour cela il faut voir si le nombre total (sur toutes les étapes de l'algorithme) de bourdons observés dans chaque état est le même dans notre simulation que dans le tableau (2.10).

Exemple: Page rank algorithm. Comment décider quelles pages internet sont plus importantes que d'autres lors une recherche? Le contenu individuel de chaque page (par exemple le nombre de fois que le mot recherché y apparait) n'est évidemment pas un bon indicateur. On peut par contre considérer que plus il y a de liens vers une page, plus cette page est importante. Cette observation est à l'origine des algorithmes utilisés par les moteurs de recherche.

Imaginons l'internet comme un graphe $G = (V, E)$ dirigé, chaque sommet étant une page internet, et chaque arête symbolisant un lien. Posons $n = |V|$ et donnons pour commencer une importance égale à chaque page:

$$X_0 = \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix} \in \mathcal{M}_{n1}(\mathbb{R}).$$

Cette attribution d'importance n'est pas réaliste. On va donc considérer en première approximation, que chaque page est aussi importante qu'il y a d'arêtes pointant vers elle. On considère ainsi que chaque page v_i distribue son importance uniformément parmi ses voisins.

Notons $d(v_i)$ le nombre d'arêtes sortantes de v_i et posons

$$A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R}) \text{ avec } a_{ij} = \begin{cases} \frac{1}{d(v_i)} & \text{si } (v_i, v_j) \in E \\ 0 & \text{sinon.} \end{cases}$$

La matrice A est stochastique à droite. L'importance de chaque sommet v_i est alors donné en cette première approximation par $X_1 = A^T X_0$.

Cette façon de classer les pages n'est pas parfaite non-plus: une page devrait être plus importante si d'autres pages importantes y sont reliées. Ainsi on pose

$$X_2 = A^T X_1, \; X_3 = A^T X_2, \; etc.$$

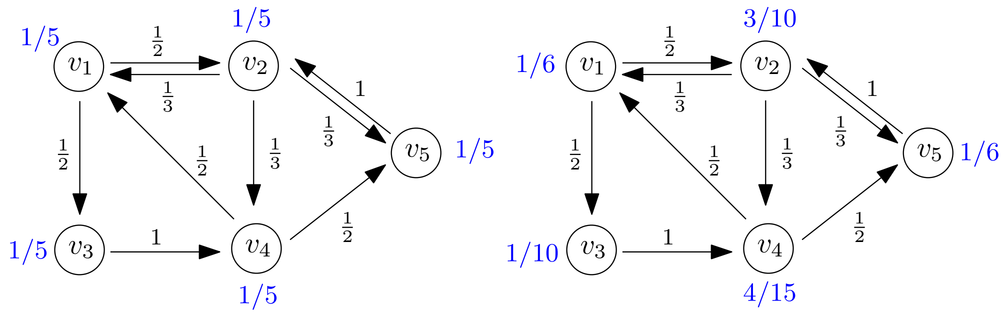


Figure 2.1: Un graphe avec les proportions transmise par chaque arête. A gauche une commence avec chaque sommet ayant une importance égale, à savoir 1/5. Après un pas de redistribution, on obtient X_1, décrit à droite.

Prenons l'exemple du graphe de la figure 2.1. Dans ce cas la matrice A s'écrit

$$A = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad \text{et} \quad A^T = \begin{pmatrix} 0 & \frac{1}{3} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

Si on applique A^T de façon répétée on trouve

	X	A^TX	$(A^T)^2X$	$(A^T)^3X$	$(A^T)^4X$	$(A^T)^5X$	$(A^T)^6X$	$(A^T)^7X$	$(A^T)^8X$
v_1	20%	17%	23%	18%	20%	21%	19%	20%	20%
v_2	20%	30%	25%	35%	28%	30%	31%	29%	31%
v_3	20%	10%	8%	12%	9%	10%	10%	10%	10%
v_4	20%	27%	20%	17%	23%	18%	20%	21%	19%
v_5	20%	17%	23%	18%	20%	21%	19%	20%	20%

À retenir

- Une *matrice* de taille $m \times n$ est un tableau rectangulaire de $m \cdot n$ scalaires notée $A = (a_{ij})_{\substack{1\leq i\leq m \\ 1\leq j\leq n}}$. On écrit $\mathcal{M}_{mn}(\mathbb{R})$ pour l'ensemble des matrices de taille $m \times n$.

- On peut additionner deux matrices de même taille et multiplier une matrice par un scalaire.

- On peut multiplier une matrice $A \in \mathcal{M}_{mn}(\mathbb{R})$ avec $B \in \mathcal{M}_{np}(\mathbb{R})$ et on obtient $AB \in \mathcal{M}_{mp}(\mathbb{R})$. La multiplication des matrices n'est pas commutative!

- L'élément neutre pour la multiplication dans $\mathcal{M}_n(\mathbb{R})$ est la matrice $I_n = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$.

  Une matrice A est *inversible* s'il existe $A^{-1} \in \mathcal{M}_n(\mathbb{R})$ telle que $AA^{-1} = A^{-1}A = I_n$.

- Pour $A \in \mathcal{M}_{mn}(\mathbb{R})$ on définit le *noyau* $\mathrm{Ker}(A)$ et l'*image* $\mathrm{Im}(A)$ par

$$\begin{aligned} &\mathrm{Ker}(A) = \{X \in \mathcal{M}_{n1}(\mathbb{R}) : AX = 0\} \quad \text{et} \\ &\mathrm{Im}(A) = \{Y \in \mathcal{M}_{m1}(\mathbb{R}) : \exists X \in \mathcal{M}_{n1} \text{ avec } AX = Y\}. \end{aligned}$$

  Ce sont des sous-espaces vectoriels de $\mathcal{M}_{n1}(\mathbb{R})$ et $\mathcal{M}_{m1}(\mathbb{R})$, respectivement.

- On définit le *rang* d'une matrice $A \in \mathcal{M}_{mn}(\mathbb{R})$ par $\mathrm{rang}(A) = \dim(\mathrm{Im}(A))$. C'est également le rang de la famille des lignes de A (comme vecteurs de $\mathbb{R}^n$) et celui de la famille des colonnes (comme vecteurs de $\mathbb{R}^m$).

- Pour $A \in \mathcal{M}_{mn}(\mathbb{R})$ on a, $\mathrm{rang}(A) + \dim(\mathrm{Ker}(A)) = n$.

- Une matrice $A \in \mathcal{M}_n(\mathbb{R})$ est inversible si et seulement si $\mathrm{rang}(A) = n$.

- Un *système linéaire* est une équation matricielle $AX = B$ où $A \in \mathcal{M}_{mn}(\mathbb{R})$, $B \in \mathcal{M}_{m1}(\mathbb{R})$ sont les *coefficients* et $X \in \mathcal{M}_{n1}(\mathbb{R})$ est l'*inconnue*.

- L'inverse d'une matrice peut être calculé par l'algorithme du pivot de Gauss.

Zu wissen

- Eine *Matrix* der Grösse $m \times n$ ist eine rechteckige Tabelle von $m \cdot n$ Skalaren, die $A = (a_{ij})_{\substack{1\leq i\leq m\\1\leq j\leq n}}$ geschrieben wird. Man bezeichnet die Menge der $(m \times n)$-Matrizen durch $\mathcal{M}_{mn}(\mathbb{R})$.
- Man kann zwei Matrizen addieren und eine Matrix mit einem Skalar multiplizieren.
- Man kann eine Matrix $A \in \mathcal{M}_{mn}(\mathbb{R})$ mit $B \in \mathcal{M}_{np}(\mathbb{R})$ multiplizieren und erhält die Matrix $AB \in \mathcal{M}_{mp}(\mathbb{R})$. Die Matrixmultiplikation ist nicht kommutativ!
- Das neutrale Element der Multiplikation in $\mathcal{M}_n(\mathbb{R})$ ist durch die Matrix $I_n = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$ gegeben. Eine Matrix A ist *invertierbar*, falls es eine Matix $A^{-1} \in \mathcal{M}_n(\mathbb{R})$ existiert, so dass $AA^{-1} = A^{-1}A = I_n$.
- Für $A \in \mathcal{M}_{mn}(\mathbb{R})$ definiert man den *Kern* $\mathrm{Ker}(A)$ und das *Bild* $\mathrm{Im}(A)$ durch

  $$\begin{aligned} &\mathrm{Ker}(A) = \{X \in \mathcal{M}_{n1}(\mathbb{R}) : AX = 0\} \text{ und} \\ &\mathrm{Im}(A) = \{Y \in \mathcal{M}_{m1}(\mathbb{R}) : \exists X \in \mathcal{M}_{n1} \text{ mit } AX = Y\}. \end{aligned}$$

  Beide sind Untervektorräume von $\mathcal{M}_{n1}(\mathbb{R})$ beziehungsweise $\mathcal{M}_{m1}(\mathbb{R})$.
- Man definiert der *Rang* einer Matrix $A \in \mathcal{M}_{mn}(\mathbb{R})$ durch $\mathrm{rang}(A) = \dim(\mathrm{Im}(A))$. Ebenfalls ist es (gleich) der Rang der Familie der Zeilenvektoren von A (als Vektoren von $\mathbb{R}^n$) und der Rang der Familie der Spaltenvektoren (als Vektoren von $\mathbb{R}^m$).
- Für $A \in \mathcal{M}_{mn}(\mathbb{R})$ gilt, $\mathrm{rang}(A) + \dim(\mathrm{Ker}(A)) = n$.
- Eine Matrix $A \in \mathcal{M}_n(\mathbb{R})$ ist invertierbar dann und nur dann $\mathrm{rang}(A) = n$.
- Ein *Linearsystem* ist eine Matrizengleichung $AX = B$, wobei $A \in \mathcal{M}_{mn}(\mathbb{R})$, $B \in \mathcal{M}_{m1}(\mathbb{R})$ die *Koeffizienten* sind und $X \in \mathcal{M}_{n1}(\mathbb{R})$ die *Unbekannte* ist.
- Die Inverse einer Matrix kann dank dem Gauss-Pivot Algorithmus berechnet werden.

Chapter 3

Le déterminant

On a vu qu'une propriété importante des matrices est l'inversibilité. Le déterminant nous offre un critère pratique pour vérifier si une matrice est inversible.

3.1 Définition et propriétés de base

Le déterminant associe à une matrice carrée $A \in \mathcal{M}_n(\mathbb{R})$ un scalaire qu'on va noter $\det(A) \in \mathbb{R}$. Il y a plusieurs façons de définir le déterminant, on choisit celle par récurrence. On aura besoin de la notation suivante:

Définition 3.1. *Soient $A \in \mathcal{M}_n(\mathbb{R})$ et $1 \leq i, j \leq n$. La* sous-matrice *$A_{i,j}$ de A est la matrice de $\mathcal{M}_{n-1}(\mathbb{R})$ obtenue en éliminant la i^{eme} ligne et j^{eme} colonne de A.*

Définition 3.2. *Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$.*

- *Si $n = 1$, alors $\det(A) = a_{11}$.*
- *Si $n \geq 2$, alors $\det(A) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(A_{1,j})$.*

Donnons quelques conséquences de cette définition.

- Pour les matrices de $\mathcal{M}_2(\mathbb{R})$ le déterminant est donné par :

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc;$$

- Pour les matrices de $\mathcal{M}_3(\mathbb{R})$ le déterminant est donné par :

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \tag{3.1}$$
$$= a_{11}(a_{22}a_{33} - a_{32}a_{23}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{31}a_{22})$$
$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{21}a_{12}a_{33} - a_{11}a_{32}a_{23}.$$

Un moyen de se souvenir de cette formule est par l'image suivante.

$$\begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{array} \quad - \quad \begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{array}$$

- Plus généralement, le déterminant d'une matrice $A \in \mathcal{M}_n(\mathbb{R})$ est une somme de produits, chaque produit contenant n entrées, une par ligne et une par colonne. Ainsi il y a $n!$ produits dans l'expression du déterminant (certains ayant un signe $+$, certains un signe $-$). Formellement

$$\det(A) = \sum_{\sigma} \epsilon(\sigma) a_{1,\sigma(1)} \dots a_{n,\sigma(n)}, \tag{3.2}$$

où la somme porte sur toutes les bijections $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$ et où ϵ vaut $+1$ ou -1 en fonction de σ ($\epsilon(\sigma)$ est déterminé de la manière suivante: une inversion de σ est une paire $1 \le i < j \le n$ avec $\sigma(i) > \sigma(j)$; $\epsilon(\sigma) = 1$ si et seulement si le nombre d'inversions de σ est pair).

Généralement le calcul du déterminant d'une matrice est compliqué. Pour les matrices 3×3 le nombre de termes dans la somme de (3.1) est 6. On peut écrire une formule similaire pour les matrices 4×4; elle va contenir 24 termes ...

Néanmoins, dans certains cas particulier, le déterminant se calcule facilement.

Proposition 3.3. *Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice triangulaire supérieure ou inférieure. Alors,*

$$\det(A) = a_{11}a_{22} \dots a_{nn}.$$

Cette proposition se montre par récurrence sur la taille de A; on la laisse en exercice.

3.2 Déterminant et inversibilité

On va donner ici quelques propriétés essentielles du déterminant.

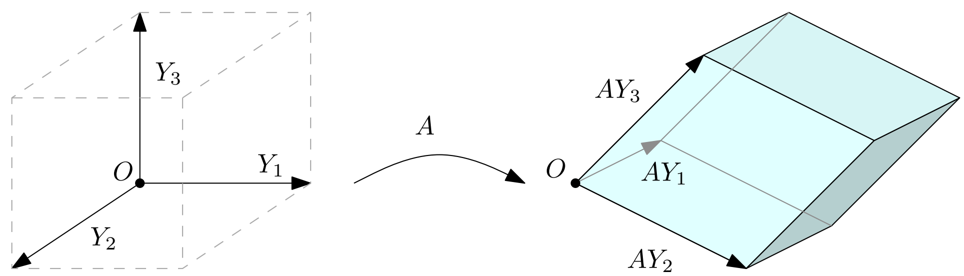


Figure 3.1: Une matrice A transforme les trois vecteurs de la base canonique de $\mathbb{R}^3$. L'image du cube unité par A est marquée en bleu; son volume est (à signe près) le déterminant de A.

Proposition 3.4. *Soient $A, B \in \mathcal{M}_n(\mathbb{R})$. Alors,*

$$\det(AB) = \det A \cdot \det B.$$

On va admettre cette proposition.

Théorème 3.5. *Soit $A \in \mathcal{M}_n(\mathbb{R})$. Alors, A est inversible si et seulement si* $\det A \neq 0$.

Une preuve rapide est basée sur le pivot de Gauss; on la donne dans la partie suivante.

Une façon géométrique de voir le déterminant d'une matrice A est comme le volume du polyèdre déterminé par les vecteurs colonne de A. Notons $C_1, \dots, C_n$ les colonnes de A; alors

$$|\det(A)| = \mathrm{Vol}\Big\{ \sum_{i=1}^{n} \lambda_i C_i : \lambda_1, \dots, \lambda_n \in [0,1] \Big\}.$$

On a déjà vu que A n'est pas inversible si et seulement si $C_1 \dots, C_n$ forment une famille liée. Cela reviens à dire que $\mathrm{rang}(C_1, \dots, C_n) < n$, donc à ce que le polyèdre mentionné fasse partie d'un hyperplan de $\mathbb{R}^n$. On voit bien alors que son volume est nul.

L'inverse d'une matrice de déterminant non-nul peut être explicité en utilisant les déterminants de ses mineurs, comme illustré par le lemme suivant. La preuve est omise dans cette version, mais peut être trouvée dans la version longue.

Lemme 3.6. *Pour $A \in \mathcal{M}_n$ definitions la matrice de cofacteurs $M = (m_{ij})_{1 \le i,j \le n} \in \mathcal{M}_n$ où*

$$m_{ij} = (-1)^{i+j} \det(A_{i,j}) \qquad \text{pour } 1 \le i, j \le n.$$

Si A est telle que $\det(A) \neq 0$, *alors $A^{-1} = \frac{1}{\det(A)} M^T$.*

Exemple: Soit $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathbb{R})$. Alors A est inversible si et seulement si $ad - bc \neq 0$. De plus, dans ce cas

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

3.3 Opérations sur lignes; pivot de Gauss

Rappelons nous des opérations sur les lignes utilisées pour le pivot de Gauss. Ces opérations changent le déterminant comme suit.

Proposition 3.7. *Pour $i \neq j$ et $\lambda \neq 0$,*

- $\mathrm{Add}_{i;\lambda,j}$ *ne change pas le déterminant,*
- $\mathrm{Mult}_{\lambda,i}$ *multiplie le déterminant par λ,*
- Ech_{ij} *change le signe du déterminant (c.à d. le multiplie par -1).*

Rappelons nous que ces opérations correspondent à des multiplications à gauche par des matrices spécifiques, qu'on a notées aussi $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ et Ech_{ij}. Un calcul direct mène au lemme suivant.

Lemme 3.8. *Pour $i \neq j$ et $\lambda \neq 0$,*

$$\det(\mathrm{Add}_{i;\lambda,j}) = 1 \quad \text{et} \quad \det(\mathrm{Mult}_{\lambda,i}) = \lambda \quad \text{et} \quad \det(\mathrm{Ech}_{ij}) = -1.$$

Comme le déterminant est multiplicatif, la proposition suit directement du lemme.

Pour une matrice carrée de taille n, le calcul du déterminant par la formule récursive qui le définit est extrêmement long. On peut voir par récurrence qu'il a une complexité algorithmique d'ordre $n!$. Un moyen beaucoup plus rapide (de complexité n^2) est offert par le pivot de Gauss.

Soit $A \in \mathcal{M}_n(\mathbb{R})$. Rappelons nous du théorème 1.4 qui nous dit qu'on peut transformer la matrice A en une matrice échelonnée $\tilde{A}$ en utilisant les opérations sur les lignes $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ et Ech_{ij}. Vu qu'on connait l'effet de ces opérations sur le déterminant, il suffit de savoir calculer le déterminant d'une matrice carrée échelonnée.

Ce calcul est particulièrement facile vu la proposition 3.3. Une matrice carrée échelonnée $\tilde{A} \in \mathcal{M}_n(\mathbb{R})$ est forcément triangulaire supérieure. Ses coefficients sur la diagonale valent soit 0 soit 1. Ainsi le déterminant de $\tilde{A}$ vaut soit 1 (si tous les coefficient diagonaux valent 1), soit 0. Rappelons également que l'unique matrice carrée échelonnée contenant que des 1 sur la diagonale est la matrice identité.

Ainsi, on arrive à la conclusion suivante.

Corollaire 3.9. *Soit $A \in \mathcal{M}_n(\mathbb{R})$ et $\tilde{A} = P_k \dots P_1 A$ la matrice échelonnée obtenue à partir de A en utilisant les opérations $P_1, \dots, P_k$ de type $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ et Ech_{ij}. Alors,*

$$\det(A) = \begin{cases} 0, & \text{si } \tilde{A} \neq I_n; \\ \frac{1}{\det(P_k)\dots\det(P_1)}, & \text{si } \tilde{A} = I_n \end{cases}$$

Comme promis, on donne maintenant la preuve du théorème 3.5.

Preuve: [Théorème 3.5] Soit $A \in \mathcal{M}_n(\mathbb{R})$ et $\tilde{A} = P_k \dots P_1 A$ la matrice échelonnée obtenue à partir de A en utilisant les opérations $P_1, \dots, P_k$ de type $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ et Ech_{ij}. On a vu dans la partie 2.4 que A est inversible, si et seulement si $\tilde{A} = I_n$. Mais le corollaire 3.9 nous dit que cela est équivalent à $\det(A) \neq 0$. $\square$

Remarque 3.10. En général on n'est pas obligé de ramener A sous forme échelonnée pour calculer son déterminant. Il suffit d'utiliser $\mathrm{Add}_{i;\lambda,j}$, $\mathrm{Mult}_{\lambda,i}$ et Ech_{ij} pour la ramener sous une forme triangulaire supérieure ou inférieure.
Dans certains cas on peut montrer par des moyens plus simple que le rang de A est strictement plus petit que n. Cela implique que A n'est pas inversible, donc que $\det(A) = 0$.

3.4 Compléments

Pour complétude, on mentionne le résultat suivant qui peut être utile en pratique. Il s'applique également à l'écriture par lignes. On va admettre ce résultat, même s'il suit directement de (3.2).

Proposition 3.11. *Soient $C_1, \dots, C_n, C_i' \in \mathcal{M}_{n1}(\mathbb{R})$ une famille de colonnes et $\lambda \in \mathbb{R}$. Alors,*

$$\begin{aligned}
&\det\Big(C_1 \dots C_{i-1} \quad C_i + C_i' \quad C_{i+1} \dots C_n\Big) \\
&\qquad = \det\Big(C_1 \dots C_{i-1} \quad C_i \quad C_{i+1} \dots C_n\Big) + \det\Big(C_1 \dots C_{i-1} \quad C_i' \quad C_{i+1} \dots C_n\Big), \\
&\det\Big(C_1 \dots C_{i-1} \quad \lambda C_i \quad C_{i+1} \dots C_n\Big) = \lambda \det\Big(C_1 \dots C_{i-1} \quad C_i \quad C_{i+1} \dots C_n\Big).
\end{aligned}$$

On dit que le déterminant est une forme multi-linéaire *de la famille des colonnes.*

On viens de voir l'effet sur le déterminant de certaines opérations sur les lignes (proposition 3.7) et sur les colonnes (proposition 3.11). Les deux sont reliées par le lemme suivant. On va admettre ce résultat (il peut être montré en utilisant (3.2)).

Lemme 3.12. *Pour tout $A \in \mathcal{M}_n(\mathbb{R})$, $\det(A^T) = \det(A)$.*

À retenir

- Le déterminant est un nombre associé à une matrice carrée; uniquement les matrices carrées ont des déterminants.
- Le déterminant d'une matrice $n \times n$ représente un volume dans $\mathbb{R}^n$: celui du parallélépipède engendré par les vecteurs colonnes de la matrice.
- Une matrice carrée est inversible si et seulement si son déterminant est non-nul.
- Formules de calcul des déterminants des matrices 2×2 et 3×3.
- Le déterminant d'une matrice triangulaire supérieure, triangulaire inférieure ou diagonale est le produit de ses coefficients diagonaux.
- Calcul du déterminant par le pivot de Gauss.

Chapter 4

Matrices et applications linéaires; diagonalisation

4.1 Applications linéaires

Définition 4.1. *Soit $u : \mathbb{R}^n \to \mathbb{R}^m$ une fonction. On dit que u est une* application linéaire[i] *de $\mathbb{R}^n$ dans $\mathbb{R}^m$ si les conditions suivantes sont satisfaites*

(i) *pour tout $x, y \in \mathbb{R}^n$, $u(x+y) = u(x) + u(y)$.*

(ii) *pour tout $x \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}$, $u(\lambda x) = \lambda u(x)$.*

Une application linéaire de $\mathbb{R}^n$ dans lui-même est appelée un endomorphisme de $\mathbb{R}^n$.

Le lemme suivant nous permet de vérifier la linéarité d'une application en utilisant une seule équation.

Lemme 4.2. *Une application $u : \mathbb{R}^n \to \mathbb{R}^m$ est linéaire si et seulement si*

$$u(\lambda\, x + \mu\, y) = \lambda\, u(x) + \mu\, u(y) \qquad \text{pour tout } x, y \in \mathbb{R}^n \text{ et } \lambda, \mu \in \mathbb{R}. \tag{4.1}$$

Observons que (4.1) implique que si u est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$, alors

$$u(0) = 0 \quad \text{et} \quad u(-x) = -u(x) \text{ pour tout } x \in \mathbb{R}^n.$$

La preuve du lemme 4.2 est simple mais omise ici.

Un premier exemple d'applications linéaires est la multiplication par les matrices. Fixons $m, n \geq 1$ et $A \in \mathcal{M}_{m,n}(\mathbb{R})$. Alors, la matrice A définie une fonction de $\mathcal{M}_{n,1}(\mathbb{R}) = \mathbb{R}^n$ dans $\mathcal{M}_{m,1}(\mathbb{R}) = \mathbb{R}^m$ par $u(X) = AX$. En effet, les règles de multiplication des matrices montre que

$$A(X+Y) = AX + AY \quad \text{et} \quad A(\lambda X) = \lambda(AX) \qquad \text{pour tous } X, Y \in \mathcal{M}_{n,1}(\mathbb{R}) \text{ et } \lambda \in \mathbb{R}.$$

(i) Lineare Abbildung oder Vektorraumhomomorphismus

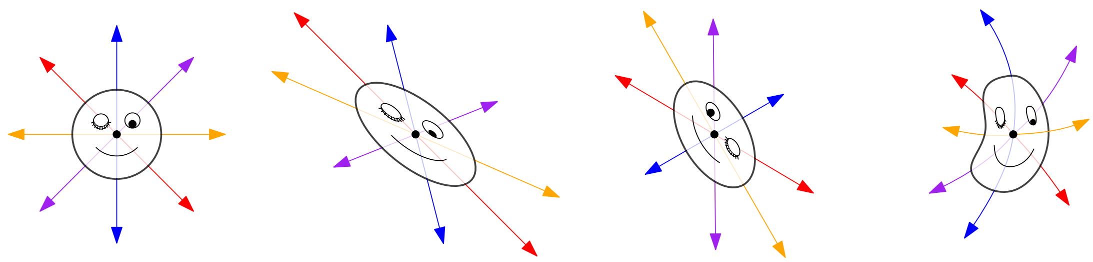


Figure 4.1: Deux transformation linéaires de la figure de gauche et une qui n'est pas linéaire. Observez l'effet de ces transformations sur les différents vecteurs.

De plus, on verra dans la partie 4.2.2 que toutes les applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}^m$ sont essentiellement des multiplications par des matrices $m \times n$.

Une application u de $\mathbb{R}^n$ dans $\mathbb{R}^n$ devrait être perçue comme une transformation de l'espace. Le nom d'application linéaire viens du fait qu'une telle application transforme les droites en droites (ou dans le point 0 quand elle est "dégénérée").

Inversement, toute application de u de $\mathbb{R}^n$ dans $\mathbb{R}^n$ qui fixe 0 et qui transforme les droites en droites est nécessairement linéaire[(ii)]. Si on renonce à la condition que 0 soit envoyé sur 0, alors u est dite affine. Voir la figure 4.2 pour quelques exemples.

En pratique: *Pour montrer qu'une application $u : \mathbb{R}^n \to \mathbb{R}^m$ est linéaire, on considère deux vecteurs génériques $x, y \in \mathbb{R}^n$ et deux nombres $\lambda, \mu \in \mathbb{R}$ et on montre que $u(\lambda\, x + \mu\, y) = \lambda\, u(x) + \mu\, u(y)$.*
Si on souhaite monter qu'une application u n'est pas linéaire, il suffit d'exhiber un contre-exemple à (4.1). Alternativement, on peut montrer que $u(0) \neq 0$.

Exemple: Quelles des applications suivantes de $\mathbb{R}^2 \to \mathbb{R}^2$ sont linéaires

$$u : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} y^2 \\ x+1 \end{pmatrix} \quad v : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ xy \end{pmatrix} \quad w : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2x+y \\ y-x \end{pmatrix}?$$

La réponse pour chaque application doit être structurée comme suit:

- u n'est pas linéaire car $u\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.
- v n'est pas linéaire car $v\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} = v\begin{pmatrix} 1 \\ 0 \end{pmatrix} + v\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.
- w est linéaire car, pour $\lambda, \mu \in \mathbb{R}$ et $\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x' \\ y' \end{pmatrix} \in \mathbb{R}^2$,

$$\begin{aligned} w\Big[\lambda\begin{pmatrix} x \\ y \end{pmatrix} + \mu\begin{pmatrix} x' \\ y' \end{pmatrix}\Big] &= w\Big[\begin{pmatrix} \lambda x + \mu x' \\ \lambda y + \mu y' \end{pmatrix}\Big] = \begin{pmatrix} 2\lambda x + 2\mu x' + \lambda y + \mu y' \\ \lambda y + \mu y' - \lambda x - \mu x' \end{pmatrix} \\ &= \lambda\begin{pmatrix} 2x+y \\ y-x \end{pmatrix} + \mu\begin{pmatrix} 2x'+y' \\ y'-x' \end{pmatrix} = \lambda\, w\Big[\begin{pmatrix} x \\ y \end{pmatrix}\Big] + \mu\, w\Big[\begin{pmatrix} x' \\ y' \end{pmatrix}\Big]. \end{aligned}$$

[(ii)] si on impose en plus que u soit bijective et continue.

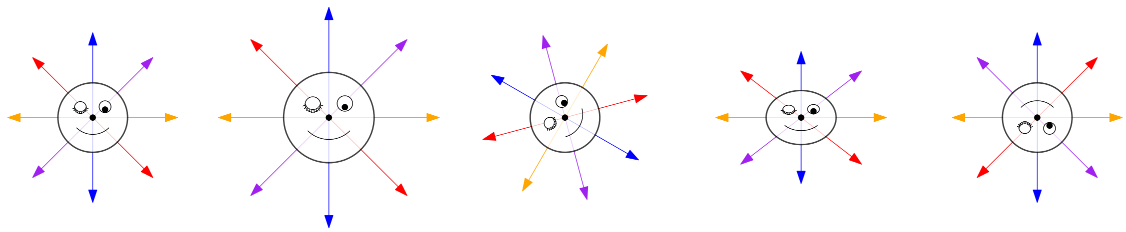


Figure 4.2: Les tranformations u, v, w et τ.

Exemple: Considérons quelques transformations linéaires du plan $\mathbb{R}^2$:

- la dilatation u d'un facteur λ (appelée *homothétie*);
- la rotation v autour de 0 d'angle θ (pour un certain $\theta \in [0, 2\pi]$);
- l'élongation τ dans la direction verticale par un facteur λ;
- la reflection τ par rapport à l'axe horizontal.

Ces transformations s'écrivent

$$u : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}; \quad v : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \cos\theta\, x - \sin\theta\, y \\ \sin\theta\, x + \cos\theta\, y \end{pmatrix},$$
$$w : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ \lambda y \end{pmatrix}, \quad \tau : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ -y \end{pmatrix}.$$

Observons que dans tous ces cas, les applications linéaires s'écrivent comme des multiplication de $\begin{pmatrix} x \\ y \end{pmatrix}$ par des matrices, à savoir $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$, $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$ et $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, respectivement.

En général, une application $u : \mathbb{R}^n \to \mathbb{R}^m$ est linéaire si et seulement si, pour $x = (x_1, \dots, x_n) \in \mathbb{R}^m$, chaque entrée du vecteur $u(x)$ s'écrit comme combinaison linéaire de $x_1, \dots, x_n$.

4.2 Représentations dans des bases

4.2.1 Représentation d'un vecteur dans une base

Une base de $\mathbb{R}^n$ sert essentiellement à *décomposer* les vecteurs de $\mathbb{R}^n$. Le théorème suivant donne un sens précis à la décomposition d'un vecteur dans une base.

Théorème 4.3. *Soit $n \geq 1$ et $\mathcal{B} = (b_1, \dots, b_n)$ une base de $\mathbb{R}^n$. Alors, pour tout $x \in \mathbb{R}^n$,* *il existe une unique famille de scalaires* $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ *telle que*

$$x = \lambda_1 b_1 + \cdots + \lambda_n b_n. \tag{4.2}$$

Preuve: Soit $x \in \mathbb{R}^n$. Comme $\mathcal{B}$ est une base de $\mathbb{R}^n$, $x \in \mathrm{Vect}(\mathcal{B}) = E$. Ainsi, il existe $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ tels que $x = \lambda_1 b_1 + \cdots + \lambda_n b_n$.

Maintenant qu'on a prouvé l'existence de la famille de scalaires désirée, montrons aussi son unicité. Soit $\lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_n \in \mathbb{R}$ tels que

$$x = \lambda_1 b_1 + \cdots + \lambda_n b_n = \mu_1 b_1 + \cdots + \mu_n b_n.$$

Alors, $(\lambda_1 - \mu_1) b_1 + \cdots + (\lambda_n - \mu_n) b_n = 0$. La famille $b_1, \dots, b_n$ étant libre, cela implique $\lambda_1 - \mu_1 = \cdots = \lambda_n - \mu_n = 0$. En d'autres termes, les familles $(\lambda_1, \dots, \lambda_n)$ et $(\mu_1, \dots, \mu_n)$ sont égales, ce qui prouve l'unicité. $\square$

A chaque $x \in \mathbb{R}^n$ on associe le vecteur $V_{\mathcal{B}}(x) := (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$, ou $\lambda_1, \dots, \lambda_n$ sont les scalaires tels que (4.2) soit vérifié. Pour des raisons qu'on va voir plus tard, on va désormais écrire $V_{\mathcal{B}}(x)$ verticalement, à savoir:

$$V_{\mathcal{B}}(x) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

On appellera le vecteur $V_{\mathcal{B}}(x)$ *l'écriture de* x *dans la base* $\mathcal{B}$.

Remarque 4.4. Si $\mathcal{E} = (e_1, \dots, e_n)$ est la base canonique de $\mathbb{R}^n$ et $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$, alors

$$X = \begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{pmatrix} = x_1 e_1 + \cdots + x_n e_n,$$

donc $V_{\mathcal{B}}(X) = X$.

Plus généralement, si $\mathcal{B} = (b_1, \dots, b_n)$ est une base quelconque de $\mathbb{R}^n$, alors $V_{\mathcal{B}}(b_i) = e_i$ pour tout i.

Proposition 4.5. *Soit* $\mathcal{B} = (b_1, \dots, b_n)$ *une base de* $\mathbb{R}^n$. *Alors, pour* $x, y \in \mathbb{R}^n$ *et* $\lambda \in \mathbb{R}$, *notons*

$$V_{\mathcal{B}}(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{et} \quad V_{\mathcal{B}}(y) = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Alors:

$$V_{\mathcal{B}}(x+y) = V_{\mathcal{B}}(x) + V_{\mathcal{B}}(y) := \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} \quad et \quad V_{\mathcal{B}}(\lambda x) = \lambda V_{\mathcal{B}}(x) := \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix},$$

ce qui est équivalent à dire que $V_{\mathcal{B}} : \mathbb{R}^n \to \mathbb{R}^n$ est une application linéaire. De plus, la fonction $V_{\mathcal{B}} : E \to \mathbb{R}^n$ est une bijection.

Rappel: le fait que $V_{\mathcal{B}}$ est bijective revient à dire que, pour tout vecteur $Y \in \mathbb{R}^n$, il existe un unique vecteur $X \in \mathbb{R}^n$ tel que $V_{\mathcal{B}}(X) = Y$.

Preuve: On va admettre ce résultat. $\square$

Exemple: Evidemment, on peut décomposer un même vecteur dans deux bases différentes. Pour $\mathbb{R}^2$, on dispose de la base canonique $\mathcal{E} = (e_1, e_2) = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)$. Mais on peut également considérer la base $\mathcal{F} = (f_1, f_2)$ ou

$$f_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{et} \quad f_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

(On vérifie facilement que $\mathcal{F}$ est en effet une base de $\mathbb{R}^2$.) Le vecteur $x = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$ se décompose alors dans les bases $\mathcal{E}$ et $\mathcal{F}$ comme suit (voir aussi la figure 4.3)

$$\begin{pmatrix} 7 \\ 3 \end{pmatrix} = 7e_1 + 3e_2 = 5f_1 + 2f_2.$$

Ainsi

$$V_{\mathcal{E}}(x) = \begin{pmatrix} 7 \\ 3 \end{pmatrix} \quad \text{et} \quad V_{\mathcal{F}}(x) = \begin{pmatrix} 5 \\ 2 \end{pmatrix}.$$

Pour trouver la representation d'un vecteur dans une base on peut procéder graphiquement, comme dans la figure 4.3, ou en résolvant un système linéaire. En effet, si $\mathcal{F} = (f_1, \ldots, f_n)$ est une base de $\mathbb{R}^n$ et $x \in \mathbb{R}^n$ un vecteur, alors $\lambda_1 f_1 + \cdots + \lambda_n f_n = x$ est un système linéaire de n équation avec n inconnues $\lambda_1, \ldots, \lambda_n$. Le fait que $\mathcal{F}$ est une base garantit que ce système admet une unique solution.

4.2.2 Représentation des applications linéaires par des matrices

Si $\mathcal{E}$ est une base de $\mathbb{R}^n$, on a vu que tout vecteur $x \in \mathbb{R}^n$ admet une représentation dans la base $\mathcal{E}$ qu'on a noté $V_{\mathcal{E}}(x) \in \mathcal{M}_{n,1}$ (voir la partie 4.2.1). De manière similaire, les applications linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$ sont représentées par des matrices $m \times n$, quand on fixe des bases $\mathcal{E}$ et $\mathcal{F}$ de $\mathbb{R}^n$ et $\mathbb{R}^m$, respectivement.

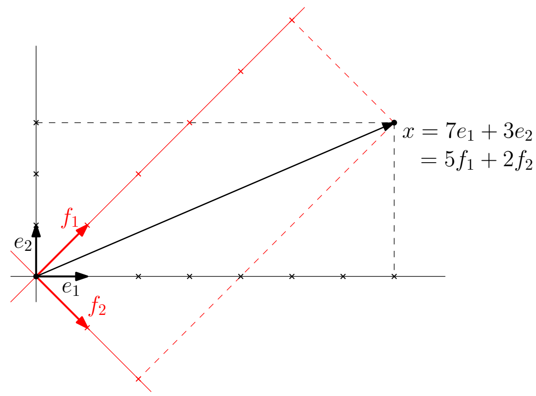


Figure 4.3: Un même vecteur x décomposé dans deux bases $\mathcal{E}$ et $\mathcal{F}$.

Définition 4.6. *Soient u une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^n$ et $\mathcal{E} = (e_1, \ldots, e_n)$ et $\mathcal{F} = (f_1, \ldots, f_m)$ des bases de E et F respectivement. La matrice $\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u) = (a_{ij}) \in \mathcal{M}_{mn}(\mathbb{R})$ définie par*

$$\begin{pmatrix} a_{1j} \\ \vdots \\ a_{m,j} \end{pmatrix} = V_{\mathcal{F}}\big(u(e_j)\big), \qquad \textit{pour tout } j \in \{1, \ldots, n\}, \tag{4.3}$$

est la matrice de u dans les bases $\mathcal{E}$, $\mathcal{F}$.

Le plus souvent on va traiter le cas des endomorphismes de E. Dans ce cadre il est naturel d'utiliser la même base pour l'espace d'arrivée et de départ. Ainsi, pour $u \in \mathcal{L}(E)$ et $\mathcal{E}$ une base de E, on écrit $\mathrm{Mat}_{\mathcal{E}}(u) = \mathrm{Mat}_{\mathcal{E},\mathcal{E}}(u) \in \mathcal{M}_n(\mathbb{R})$.

Théorème 4.7. *Soient u une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^n$, $\mathcal{E}$ et $\mathcal{F}$ des bases de E et F respectivement, et $x \in E$. Alors*

$$V_{\mathcal{F}}\big(u(x)\big) = \mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u) V_{\mathcal{E}}(x). \tag{4.4}$$

De plus $\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u)$ est l'unique matrice qui a cette propriété.

Preuve: On commence par montrer que $\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u)$ satisfait (4.4) pour les vecteurs de la base $\mathcal{E}$.

Soit e_j un vecteur de $\mathcal{E}$. Alors

$$V_{\mathcal{E}}(e_j) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

où l'entrée 1 est sur la j^{eme} ligne. Ainsi,

$$\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u) V_{\mathcal{E}}(e_j) = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = V_{\mathcal{F}}\big(u(e_j)\big). \tag{4.5}$$

La première égalité vient de la règle de multiplication des matrices, la deuxième vient de la definition de $\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u)$

Montrons maintenant (4.4) pour un vecteur quelconque x. Soit $x \in E$ et notons $V_{\mathcal{E}}(x) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$, de sorte que $x = \lambda_1 e_1 + \cdots + \lambda_n e_n$. Alors, par la linéarité de u, $u(x) = \lambda_1 u(e_1) + \cdots + \lambda_n u(e_n)$. Ainsi

$$\begin{aligned}
&\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u) V_{\mathcal{E}}(x) \\
&\quad = \mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u)\big(\lambda_1 V_{\mathcal{E}}(e_1) + \cdots + \lambda_n V_{\mathcal{E}}(e_n)\big) && \text{par linéarité de } V_{\mathcal{E}} \\
&\quad = \lambda_1 \mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u) V_{\mathcal{E}}(e_1) + \cdots + \lambda_n \mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u) V_{\mathcal{E}}(e_n) && \text{par linéarité de multip. matricielle} \\
&\quad = \lambda_1 V_{\mathcal{F}}\big(u(e_1)\big) + \cdots + \lambda_n V_{\mathcal{F}}\big(u(e_n)\big) && \text{par (4.5)} \\
&\quad = V_{\mathcal{F}}\big(\lambda_1 u(e_1) + \cdots + \lambda_n u(e_n)\big) && \text{par linéarité de } V_{\mathcal{F}} \\
&\quad = V_{\mathcal{F}}\big(u(\lambda_1 e_1 + \cdots + \lambda_n e_n)\big) && \text{par linéarité de } u \\
&\quad = V_{\mathcal{F}}\big(u(x)\big).
\end{aligned}$$

En fin, observons que $\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u)$ est l'unique matrice qui satisfait (4.4) pour tout $x \in E$. En effet si on applique (4.4) au vecteur e_i de la base $\mathcal{E}$, on obtient (4.3). Ainsi (4.3) n'est qu'un cas particulier de (4.4). $\square$

Exemple: Considérons les transformations suivantes du plan $\mathbb{R}^2$: u est la rotation autour de 0 d'angle θ (pour un certain $\theta \in [0, 2\pi]$) et soit v la réflexion par rapport à la diagonale principale de $\mathbb{R}^2$ (voir fig. 4.4) Les deux sont des applications linéaires. On peut calculer leur matrices dans la base canonique $\mathcal{E}$ de $\mathbb{R}^2$, à savoir la base formée de $e_1 = (1, 0)$ et $e_2 = (0, 1)$. On a

$$u(e_1) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \cos\theta e_1 + \sin\theta e_2, \qquad u(e_2) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = -\sin\theta e_1 + \cos\theta e_2,$$

$$\text{et} \quad v(e_1) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = e_2, \qquad v(e_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = e_1.$$

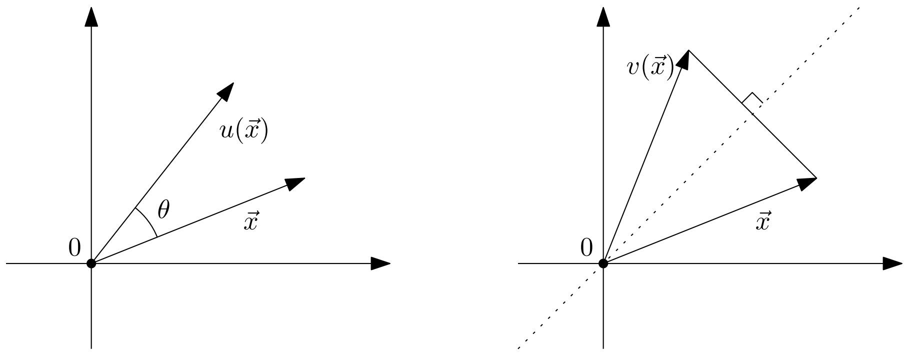


Figure 4.4: L'effet des transformations u et v.

Ainsi

$$\mathrm{Mat}_{\mathcal{E}}(u) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{et} \quad \mathrm{Mat}_{\mathcal{E}}(v) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

On peut également essayer d'écrire les mêmes transformations dans une autre base. Posons $f_1 = e_1 + e_2$ et $f_2 = -e_1 + e_2$. Alors il est facile de voir que $\mathcal{F} = (f_1, f_2)$ est une base de $\mathbb{R}^2$. Un calcul rapide nous montre que

$$\mathrm{Mat}_{\mathcal{F}}(u) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{et} \quad \mathrm{Mat}_{\mathcal{F}}(v) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Exemple: Fixons $m, n \geq 1$ et $A \in \mathcal{M}_{m,n}(\mathbb{R})$. On viens de dire que l'application $u : X \mapsto AX$ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$. Si $\mathcal{E}$ et $\mathcal{F}$ représentent les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$, respectivement, alors $\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u) = A$. Toutefois, si $\mathcal{F}$ et $\mathcal{E}$ désignent d'autres bases que les bases canoniques, $\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u)$ peut être différente de A.
Inversement, si v est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$ et $\mathcal{E}$ et $\mathcal{F}$ représentent les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$, alors

$$v(X) = V_{\mathcal{F}}(v(X)) = \mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u) V_{\mathcal{E}}(X) = \mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u) X.$$

Ainsi toute application linéaire est la multiplication par une matrice.

Exemple: Parmi les endomorphismes de $\mathbb{R}^n$, on en identifie deux particuliers:

$$\begin{aligned} \mathrm{id} : \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto x \end{aligned} \qquad\qquad \begin{aligned} 0 : \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto 0 \end{aligned}$$

Pour toute base $\mathcal{B}$ de $\mathbb{R}^n$ on a $\mathrm{Mat}_{\mathcal{B}}(\mathrm{id}) = I_n$ et $\mathrm{Mat}_{\mathcal{B}}(\mathrm{id}) = I_n$. Ces applications sont les seules dont la représentation ne dépend pas de la base; pour toutes les autres applications linéaires u, $\mathrm{Mat}_{\mathcal{B}}(u)$ dépend de $\mathcal{B}$.

Exemple: Considérons l'application linéaire $u : \mathbb{R}^2 \to \mathbb{R}^2$ donnée par $u\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ 4x+y \end{pmatrix}$. Elle est bien linéaire et sa matrice dans la base canonique de $\mathbb{R}^2$ est $\begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$ Observons toutefois que son effet sur les vecteurs

$$f_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \text{ et } f_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ est } u(f_1) = \begin{pmatrix} 1 \\ -2 \end{pmatrix} = -f_1 \text{ et } u(f_2) = \begin{pmatrix} 3 \\ 6 \end{pmatrix} = 3f_2.$$

Ainsi, dans la base $\mathcal{F} = (f_1, f_2)$, la matrice de u s'écrit

$$\mathrm{Mat}_{\mathcal{F}}(u) = \begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}.$$

4.2.3 Image, noyau, composition, inverse

Les notions d'image, noyau et rang d'une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$ sont similaires à ceux des matrices:

$$\mathrm{Im}(u) = \{u(x) \,:\, x \in \mathbb{R}^n\}, \quad \mathrm{Ker}(u) = \{x \in \mathbb{R}^n \,:\, u(x) = 0\} \quad \text{et} \quad \mathrm{rang}(u) = \dim(\mathrm{Im}(u)).$$

L'image et le noyau de u sont des sous-espaces vectoriels de $\mathbb{R}^m$ et $\mathbb{R}^n$, respectivement.

Proposition 4.8. *Soient $\mathcal{E}$ une base de $\mathbb{R}^n$, $\mathcal{F}$ une base de $\mathbb{R}^m$ et u une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$. Alors*

$$\mathrm{Im}(\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u)) = V_{\mathcal{F}}(\mathrm{Im}(u)), \ \ \mathrm{Ker}(\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u)) = V_{\mathcal{E}}(\mathrm{Ker}(u)), \ \ \mathrm{rang}(u) = \mathrm{rang}(\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u)).$$

Le preuve de cette proposition suit directement de la formule (4.4) et de la bijectivité de $V_{\mathcal{E}}$ et $V_{\mathcal{F}}$.

Proposition 4.9. *Soient u et v des applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}^m$ et de $\mathbb{R}^m$ dans $\mathbb{R}^\ell$, respectivement, alors $v \circ u : x \mapsto v(u(x))$ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^\ell$. De plus, si $\mathcal{E}$, $\mathcal{F}$ et $\mathcal{G}$ sont des bases de $\mathbb{R}^n$, $\mathbb{R}^m$ et $\mathbb{R}^\ell$, respectivement, alors*

$$\mathrm{Mat}_{\mathcal{G},\mathcal{E}}(v \circ u) = \mathrm{Mat}_{\mathcal{G},\mathcal{F}}(v)\mathrm{Mat}_{\mathcal{F},\mathcal{E}}(u)$$

Exemple: Pour $\theta \in \mathbb{R}$, notons $u_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ la rotation d'angle θ et $v : \mathbb{R}^2 \to \mathbb{R}^2$ la reflection par rapport à l'axe horizontal. De plus, notons $\mathcal{E} = (e_1, e_2)$ la base canonique de $\mathbb{R}^2$. Alors:

$$\mathrm{Mat}_{\mathcal{E}}(u_\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{et} \quad \mathrm{Mat}_{\mathcal{E}}(v) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Considerons les application linéaires $u_\theta \circ u_{\theta'}$, $v \circ u_\theta$ et $u_\theta \circ v$. On a

$$\begin{aligned}
\mathrm{Mat}_{\mathcal{E}}(u_\theta \circ u_{\theta'}) &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta' & -\sin\theta' \\ \sin\theta' & \cos\theta' \end{pmatrix} \\
&= \begin{pmatrix} \cos\theta\, \cos\theta' - \sin\theta\, \sin\theta' & -\cos\theta\, \sin\theta' - \sin\theta \cos\theta' \\ \sin\theta\, \cos\theta' + \cos\theta\, \sin\theta' & -\sin\theta\, \sin\theta' + \cos\theta\, \cos\theta' \end{pmatrix} \\
&= \begin{pmatrix} \cos(\theta+\theta') & -\sin(\theta+\theta') \\ \sin(\theta+\theta') & \cos(\theta+\theta') \end{pmatrix} = \mathrm{Mat}_{\mathcal{E}}(u_{\theta+\theta'})
\end{aligned}$$

Cela n'est pas surprenant, en effet la composition d'une rotation d'angle θ avec une rotation d'angle θ' donne bien une rotation d'angle $\theta+\theta'$. De plus, on observe que $\mathrm{Mat}_{\mathcal{E}}(u_{\theta'} \circ u_\theta) = \mathrm{Mat}_{\mathcal{E}}(u_{\theta+\theta'})$ aussi. En d'autre mots, u_θ et $u_{\theta'}$ commutent (ou, de manière équivalente, leur matrices $\mathrm{Mat}_{\mathcal{E}}(u_\theta)$ et $\mathrm{Mat}_{\mathcal{E}}(u_{\theta'})$ commutent).

D'autre part

$$\begin{aligned}
\mathrm{Mat}_{\mathcal{E}}(v \circ u_\theta) &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{pmatrix} \quad \text{et} \\
\mathrm{Mat}_{\mathcal{E}}() &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} = \mathrm{Mat}_{\mathcal{E}}(v \circ u_{-\theta}).
\end{aligned}$$

Ainsi, $u_\theta \circ v = v \circ u_{-\theta} \neq v \circ u_{-\theta}$, ce qui entraine que u_θ et v ne commutent pas.

En fin, on dit qu'une application linéaire $u : \mathbb{R}^n \to \mathbb{R}^n$ est inversible s'il existe une autre application linéaire v qui a l'effet inverse de u, donc telle que

$$u(v(x)) = v(u(x)) = x \qquad \text{pour tout } x \in \mathbb{R}^n.$$

Si une telle application existe, elle est unique, et on pose $u^{-1} = v$. L'équation précédente s'écrit plus simplement $u \circ u^{-1} = u^{-1} \circ u = \mathrm{id}$.

Proposition 4.10. *Soit $u : \mathbb{R}^n \to \mathbb{R}^n$ linéaire et $\mathcal{E}$ une base de $\mathbb{R}^n$. Alors u est inversible si et seulement si $\mathrm{Mat}_{\mathcal{E}}(u)$ est inversible. Dans le cas ou les deux sont inversibles, on a*

$$\mathrm{Mat}_{\mathcal{E}}\big(u^{-1}\big) = \mathrm{Mat}_{\mathcal{E}}(u)^{-1}.$$

De plus, grace au corollaire 2.14 et à la proposition 4.8, on a aussi équivalence de

Application linéaire	Matrice
• u est inversible	• $\mathrm{Mat}_{\mathcal{E}}(u)$ est inversible
• $\mathrm{rang}(u) = n$	• $\mathrm{rang}(\mathrm{Mat}_{\mathcal{E}}(u)) = n$
• $\mathrm{Ker}(u) = \{0\}$	• $\mathrm{Ker}(\mathrm{Mat}_{\mathcal{E}}(u)) = \{0\}$
• u est surjective	• la famille des colonnes de $\mathrm{Mat}_{\mathcal{E}}(u)$ est libre
• u est injective	• la famille des lignes de $\mathrm{Mat}_{\mathcal{E}}(u)$ est libre
• u est bijective	

Exercice 4.1.
Soient $A, B \in \mathcal{M}_{mn}(\mathbb{R})$. Montrer que si $AX = BX$ pour tout $X \in \mathcal{M}_{n1}(\mathbb{R})$, alors $A = B$.

4.3 Changements de base; matrices semblables

4.3.1 Changement de base pour la représentation d'un vecteur

En pratique, il est des fois nécessaire de travailler avec plusieurs bases d'un même espace vectoriel (voir par exemple les suites définies par récurrence traités dans la partie **??**, ou encore la diagonalisation des matrices traitée dans la partie 4). Ainsi, il est important d'avoir un outil qui permet d'obtenir l'écriture d'un vecteur dans une base à partir de son écriture dans une autre base. La solution est donnée par les matrices de changement de base.

Fixons deux bases $\mathcal{B} = (b_1, \ldots, b_n)$ et $\mathcal{E} = (e_1, \ldots, e_n)$ de $\mathbb{R}^n$.

Définition 4.11. *La* matrice de changement de base *de $\mathcal{E}$ à $\mathcal{B}$ est la matrice notée* $\mathrm{P}_{\mathcal{B},\mathcal{E}} = (p_{ij}) \in \mathcal{M}_n(\mathbb{R})$ *dont les entrées sont données colonne par colonne par*

$$C_j = \begin{pmatrix} p_{1j} \\ \vdots \\ p_{nj} \end{pmatrix} = V_{\mathcal{B}}(e_j) \qquad \textit{pour tout } j \in \{1, \ldots, n\}. \tag{4.6}$$

Comme illustré par son nom, la matrice permet d'obtenir l'écriture d'un vecteur X dans la base $\mathcal{B}$ à partir de son écriture dans la base $\mathcal{E}$.

Proposition 4.12. *Pour tout* $X \in \mathcal{R}^n$, $V_{\mathcal{B}}(X) = \mathrm{P}_{\mathcal{B},\mathcal{E}} V_{\mathcal{E}}(X)$.

Preuve: Admis. □

On aimerai dire que le changement de base de $\mathcal{B}$ à $\mathcal{E}$ est l'inverse de celui de $\mathcal{E}$ à $\mathcal{B}$. La proposition suivante en donne le sens précis.

Proposition 4.13. *La matrice* $\mathrm{P}_{\mathcal{B},\mathcal{E}}$ *est inversible et* $\mathrm{P}_{\mathcal{B},\mathcal{E}}^{-1} = \mathrm{P}_{\mathcal{E},\mathcal{B}}$.

Preuve: En appliquant la proposition 4.12 deux fois, on obtient que pour tout $X \in E$,

$$V_{\mathcal{B}}(X) = \mathrm{P}_{\mathcal{B},\mathcal{E}} V_{\mathcal{E}}(X) = \mathrm{P}_{\mathcal{B},\mathcal{E}} \mathrm{P}_{\mathcal{E},\mathcal{B}} V_{\mathcal{B}}(X).$$

Le même calcul s'applique à $\mathrm{P}_{\mathcal{E},\mathcal{B}} \mathrm{P}_{\mathcal{B},\mathcal{E}}$. Comme $V_{\mathcal{B}}$ et $V_{\mathcal{E}}$ sont surjectives, on déduit que

$$\mathrm{P}_{\mathcal{B},\mathcal{E}} \mathrm{P}_{\mathcal{E},\mathcal{B}} X = \mathrm{P}_{\mathcal{E},\mathcal{B}} \mathrm{P}_{\mathcal{B},\mathcal{E}} X = X, \qquad \text{pour tout } X \in \mathcal{M}_{n1}(\mathbb{R}).$$

Il s'en suit (voir par exemple l'exercice 4.1) que $\mathrm{P}_{\mathcal{B},\mathcal{E}} \mathrm{P}_{\mathcal{E},\mathcal{B}} = \mathrm{P}_{\mathcal{E},\mathcal{B}} \mathrm{P}_{\mathcal{B},\mathcal{E}} = I_n$, donc que $\mathrm{P}_{\mathcal{B},\mathcal{E}} = \mathrm{P}_{\mathcal{E},\mathcal{B}}^{-1}$. □

De plus, il est facile de voir que toute matrice inversible est une matrice de changement de base. En effet, si $P \in \mathcal{M}_n(\mathbb{R})$ est inversible et $\mathcal{B}$ est une base de E, il suffit de créer la base $\mathcal{E}$ par la formule (4.6), et on obtient $P = \mathrm{P}_{\mathcal{B},\mathcal{E}}$.

Remarque 4.14. Inversement, toute matrice inversible est un matrice de changement de base. En effet, soit $M = (m_{ij}) \in \mathcal{M}_n(\mathbb{R})$ et $\mathcal{B} = (b_1, \dots, b_n)$ une base de $\mathbb{R}^n$. Posons $\mathcal{E} = (e_1, \dots, e_n)$ avec $e_j = m_{1j}b_1 + \cdots + m_{nj}b_n$. Comme M est inversible, on déduit que $\mathcal{E}$ est une base de $\mathbb{R}^n$.

Alors, un calcul direct nous montre que $M = \mathrm{P}_{\mathcal{B},\mathcal{E}}$. En effet, pour chaque $1 \le j \le n$, la $j$ème colonne de M est $\begin{pmatrix} m_{1j} \\ \vdots \\ m_{nj} \end{pmatrix} = V_{\mathcal{B}}(e_j)$, par définition de e_j.

Exemple: Revenons au bases $\mathcal{E} = (e_1, e_2)$ et $\mathcal{F} = (f_1, f_2)$ de $\mathbb{R}^2$ données par

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{et} \quad f_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Il s'agit des bases utilisés dans l'exemple de la partie 4.2.2. Calculons les matrices de changement de base $\mathrm{P}_{\mathcal{E},\mathcal{F}}$ et $\mathrm{P}_{\mathcal{F},\mathcal{E}}$.

Comme $\mathcal{E}$ est la base canonique de $\mathbb{R}^2$, c'est simple d'observer que

$$V_{\mathcal{E}}(f_1) = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \quad V_{\mathcal{E}}(f_2) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Ainsi $\mathrm{P}_{\mathcal{E},\mathcal{F}} = \Big(V_{\mathcal{E}}(f_1)\, V_{\mathcal{E}}(f_2) \Big) = \begin{pmatrix} -1 & 1 \\ 2 & 2 \end{pmatrix}$. De plus

$$\mathrm{P}_{\mathcal{F},\mathcal{E}} = \mathrm{P}_{\mathcal{E},\mathcal{F}}^{-1} = \frac{1}{\det \mathrm{P}_{\mathcal{E},\mathcal{F}}} \begin{pmatrix} 2 & -1 \\ -2 & -1 \end{pmatrix} = -\frac{1}{4} \begin{pmatrix} 2 & -1 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} -1/2 & 1/4 \\ 1/2 & 1/4 \end{pmatrix}$$

En effet, $V_{\mathcal{F}}(e_1) = \mathrm{P}_{\mathcal{F},\mathcal{E}} V_{\mathcal{E}}(e_1) = \begin{pmatrix} -1/2 \\ 1/2 \end{pmatrix}$ et $V_{\mathcal{F}}(e_2) = \begin{pmatrix} 1/4 \\ 1/4 \end{pmatrix}$.

De plus, pour l'application linéaire $u : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x + y \\ 4x + y \end{pmatrix}$ de l'exemple de la partie 4.2.2, $\mathrm{Mat}_{\mathcal{E}}(u) = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$, donc

$$\mathrm{Mat}_{\mathcal{F}}(u) = \mathrm{P}_{\mathcal{F},\mathcal{E}} \mathrm{Mat}_{\mathcal{E}}(u) \mathrm{P}_{\mathcal{E},\mathcal{F}} = \begin{pmatrix} -1/2 & 1/4 \\ 1/2 & 1/4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}.$$

4.3.2 Matrices semblables

Définition 4.15. *Soit $A, B \in \mathcal{M}_n(\mathbb{R})$ deux matrices. On dit qu'elles sont* semblables *s'il* *existe* $P \in \mathcal{M}(\mathbb{R})$ *inversible telle que*

$$B = P^{-1}AP. \tag{4.7}$$

La notion de matrices semblable est motivée par la représentation des endomorphismes dans différentes bases. Pour illustrer cela, on commence par une proposition qui donne la formule essentielle de changement de base pour des endomorphismes.

Proposition 4.16. *Notons* $\mathcal{B}$ *et* $\mathcal{E}$ *deux bases* $\mathbb{R}^n$ *et* u *un endomorphisme de* $\mathbb{R}^n$.

$$\mathrm{Mat}_{\mathcal{B}}(u) = \mathrm{P}_{\mathcal{E},\mathcal{B}}^{-1}\,\mathrm{Mat}_{\mathcal{E}}(u)\,\mathrm{P}_{\mathcal{E},\mathcal{B}}.$$

Grace à cette proposition et à la remarque 4.14, on conclut que deux matrices sont semblables si et seulement si elles représentent la même application linéaire dans deux bases différentes.

En effet, si deux matrices A et B sont semblables, alors $B = PAP^{-1}$ avec $P \in \mathcal{M}_n(\mathbb{R})$ qui est inversible, donc qui est la matrice de changement de base de la base canonique vers une certaine base $\mathcal{B}$ de $\mathbb{R}^n$. Ainsi $B = \mathrm{P}_{\mathcal{E},\mathcal{B}}^{-1}\,A\,\mathrm{P}_{\mathcal{E},\mathcal{B}}$.

Preuve: Soient u, $\mathcal{B}$ et $\mathcal{E}$ comme dans l'énoncé. Alors, pour $x \in \mathbb{R}^n$, $V_{\mathcal{E}}(x) = \mathrm{P}_{\mathcal{E},\mathcal{B}}V_{\mathcal{B}}(x)$ et donc

$$V_{\mathcal{B}}(u(x)) = \mathrm{P}_{\mathcal{B},\mathcal{E}}V_{\mathcal{E}}(u(x)) = \mathrm{P}_{\mathcal{B},\mathcal{E}}\mathrm{Mat}_{\mathcal{E}}(u)V_{\mathcal{E}}(x) = \mathrm{P}_{\mathcal{B},\mathcal{E}}\mathrm{Mat}_{\mathcal{E}}(u)\mathrm{P}_{\mathcal{E},\mathcal{B}}V_{\mathcal{B}}(x).$$

Par l'unicité de la matrice dans le théorème 4.7,

$$\mathrm{Mat}_{\mathcal{B}}(u) = \mathrm{P}_{\mathcal{B},\mathcal{E}}\mathrm{Mat}_{\mathcal{E}}(u)\mathrm{P}_{\mathcal{E},\mathcal{B}} = \mathrm{P}_{\mathcal{E},\mathcal{B}}^{-1}\,\mathrm{Mat}_{\mathcal{E}}(u)\,\mathrm{P}_{\mathcal{E},\mathcal{B}}\,V_{\mathcal{B}}(X).$$

□

4.4 Valeurs et vecteurs propres

Définition 4.17. *On dit que* $\lambda \in \mathbb{R}$ *est une* valeur propre[(iii)] *de* A *s'il existe un vecteur* $X \in \mathcal{M}_{n1}(\mathbb{R})$, $X \neq 0$, *tel que*

$$AX = \lambda X.$$

Dans ce cas on dit que X *est un* vecteur propre[(iv)] *de* A *pour la valeur propre* λ. *L'ensemble des valeurs propres de* A *s'appelle le* spectre[(v)] *de* A *et est noté* $Sp(A)$. *Pour* $\lambda \in Sp(A)$, *on note*

$$E_\lambda(A) = \big\{X \in \mathcal{M}_{n1}(\mathbb{R}) : AX = \lambda X\big\}.$$

[(iii)] Eigenwert
[(iv)] Eigenvektor
[(v)] Spektrum

Proposition 4.18.

(i) *Pour chaque $\lambda \in Sp(A)$, $E_\lambda(A)$ est un sous-espace vectoriel de $\mathbb{R}^n$ qu'on appelle* l'espace propre[vi] *de A associé à λ.*

(ii) *Si $\lambda_1, \dots, \lambda_k \in Sp(A)$ sont deux à deux distinctes et $X_1 \in E_{\lambda_1}, \dots, X_k \in E_{\lambda_k}$ sont des vecteurs non-nuls, alors la famille $(X_1, \dots, X_k)$ est libre.*

(iii) *Le spectre de A a au plus n éléments.*

Preuve: (i) Soit $\lambda \in Sp(A)$. Alors $X \in E_\lambda(A)$ si et seulement si $AX - \lambda X = (A - \lambda I_n)X = 0$. Ainsi

$$E_\lambda(A) = \mathrm{Ker}(A - \lambda I_n).$$

En particulier on déduit que $E_\lambda(A)$ est un s.e.v. de $\mathbb{R}^n$.

(ii) Ce point est admis.

(iii) Soient $\lambda_1, \dots, \lambda_k \in Sp(A)$ deux à deux distincts. On peut alors choisir des vecteurs non-nuls $X_1 \in E_{\lambda_1}, \dots, X_k \in E_{\lambda_k}$. Par le point précèdent $(X_1, \dots, X_k)$ est une famille libre de $\mathbb{R}^n$, donc $k \leq n$. Ainsi $|Sp(A)| \leq n$. $\square$

Remarque 4.19. On insiste sur l'importance du point (i) et de sa preuve. Vu que

$$E_\lambda(A) = \mathrm{Ker}(A - \lambda I_n),$$

les valeurs propres de A sont exactement les $\lambda \in \mathbb{R}$ pour lesquels $A - \lambda I_n$ n'est pas inversible.

Observons que X est le vecteur propre d'une application linéaire A si et seulement si X est envoyé par A sur un multiple de X. Ainsi, il est naturel que les vecteurs et valeures propres d'une matrice soient invariantes par changement de base.

Proposition 4.20. *Soit $P \in \mathcal{M}_n(\mathbb{R})$ une matrice inversible. Alors*

(i) *$Sp(A) = Sp(P^{-1}AP)$;*

(ii) *pour tout $\lambda \in \mathbb{R}$, $X \in E_\lambda(P^{-1}AP)$ si et seulement si $PX \in E_\lambda(A)$.*

Preuve: On commence par le point (ii). Soit $\lambda \in \mathbb{R}$ et $X \in \mathcal{M}_{n1}(\mathbb{R})$. Supposons que $X \in E_\lambda(P^{-1}AP)$. Alors $P^{-1}APX = \lambda X$. Si on multiplie cette égalité par P à gauche, on obtient $APX = \lambda PX$, donc $PX \in E_\lambda(A)$.

Inversement si $PX \in E_\lambda(A)$, alors $P^{-1}APX = \lambda P^{-1}PX = \lambda X$, donc $X \in E_\lambda(A)$.

Passons au point (i). Soit $\lambda \in Sp(P^{-1}AP)$ et $X \in E_\lambda(P^{-1}AP)$ non-nul. Alors, par le point précédent, $PX \in E_\lambda(A)$. De plus, comme P est inversible, $PX \neq 0$, donc $E_\lambda(A) \neq \{0\}$. On en déduit que $\lambda \in Sp(A)$, donc que $Sp(P^{-1}AP) \subset Sp(A)$.

Inversement, si $\lambda \in Sp(A)$ et $X \in E_\lambda(A)$ non-nul, alors $P^{-1}X \in Sp(P^{-1}AP)$. A nouveau $P^{-1}X \neq 0$ car P^{-1} est inversible, donc $\lambda \in Sp(P^{-1}AP)$. Il s'en suit que $Sp(A) \subset Sp(P^{-1}AP)$. La double inclusion montre que $Sp(A) = Sp(P^{-1}AP)$. $\square$

[vi] Eigenraum

4.5 Matrices diagonalisables

Définition 4.21. *On dit que A est une matrice* diagonalisable[(vii)] *(dans $\mathcal{M}_n(\mathbb{R})$) s'il existe une matrice inversible $P \in \mathcal{M}_n(\mathbb{R})$ telle que $P^{-1}AP$ est une matrice diagonale.*

L'intérêt de cette notion peut s'expliquer par l'observation suivante. Dans beaucoup de situations pratiques il est intéressant de calculer les puissances d'une matrice carrée (voir les exemples de la partie 2.5). Comment peut-on donc calculer A^k pour une grande matrice A, ou k est une très grande valeur, sans faire trop de calculs?

Si A est diagonalisable, le calcul est simple. Soit $P \in \mathcal{M}_n(\mathbb{R})$ inversible, telle que la matrice $D = P^{-1}AP$ soit diagonale. Alors $A = PDP^{-1}$, et

$$A^k = (PDP^{-1})^k = \underbrace{PD\cancel{P^{-1}}\cancel{P}D\cancel{P^{-1}}\ldots\cancel{P}DP^{-1}}_{k \text{ fois}} = PD^kP^{-1}.$$

Mais D est une matrice diagonale de la forme $D = \begin{pmatrix} d_{11} & 0 & \ldots & 0 \\ 0 & d_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_{nn} \end{pmatrix}$. On peut donc facilement calculer

$$D^k = \begin{pmatrix} d_{11}^k & 0 & \ldots & 0 \\ 0 & d_{22}^k & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_{nn}^k \end{pmatrix}.$$

Ainsi, le calcul des puissances des matrices diagonalisables est facile, surtout si on connaît la matrice P.

4.5.1 Lien avec les vecteurs propres

Théorème 4.22. *La matrice A est diagonalisable si et seulement s'il existe une base $\mathcal{E} = (X_1, \ldots, X_n)$ de $\mathbb{R}^n$ formée entièrement de vecteurs propres de A.*

Remarque 4.23. Dans les faits, la matrice P n'est rien d'autre que la matrice de changement de base de la base de vecteurs propres de A à la base canonique de $\mathbb{R}^n$. On aimerait insister sur le fait que la base de vecteurs propres (et donc la matrice P qui diagonalise A) ne sont pas uniques!

Si $A \in \mathcal{M}_n(\mathbb{R})$ est diagonalisable et P est une matrice inversible telle que $P^{-1}AP = D$ est diagonale, alors les valeurs propres de A sont exactement les éléments de la diagonale de D. De plus, pour $\lambda \in Sp(A)$, $\dim(E_\lambda(A))$ est simplement le nombre d'apparitions de λ sur la diagonale de D.

(vii) diagonalisierbar

Exemple: Rappelons nous de la dynamique de population donnée dans la partie 2.5:

$$a_{n+1} = a_n + b_n \quad \text{et} \quad b_{n+1} = 4a_n + b_n, \qquad \text{pour } n \geq 0.$$

Si on pose, $X_n = \begin{pmatrix} a_n \\ b_n \end{pmatrix}$ pour chaque $n \in \mathbb{N}$, on observe que

$$X_n = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} X_{n-1} = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}^n X_0 = A^n X_0,$$

où $A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$. Si on arrive a diagonaliser A, ca devient facile de calculer A^n et par la suite X_n.

Observons que

$$A \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{et} \quad A \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} = (-1) \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

Ainsi, A admet les vecteurs propres $Y_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ et $Y_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$, avec les valeurs propres $\lambda_1 = 3$ et $\lambda_2 = -1$. Les vecteurs Y_1, Y_2 sont évidemment libres, donc forment une base de $\mathcal{M}_{21}$. On est donc dans le cadre du théorème 4.22, et on conclut que

$$P^{-1}AP = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{où} \quad P = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix}.$$

On peut calculer P^{-1} par le pivot de Gauss et on trouve

$$P^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & -\frac{1}{4} \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix}.$$

La forme générale de X_n est donc donnée par

$$\begin{aligned} X_n &= P \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} P^{-1} X_{n-1} = P \begin{pmatrix} 3^n & 0 \\ 0 & (-1)^n \end{pmatrix} P^{-1} X_0 \\ &= \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 3^n & 0 \\ 0 & (-1)^n \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}[3^n + (-1)^n] & \frac{1}{4}[3^n - (-1)^n] \\ 3^n - (-1)^n & \frac{1}{2}[3^n + (-1)^n] \end{pmatrix} \begin{pmatrix} a_0 \\ b_0 \end{pmatrix}. \end{aligned}$$

Cela explique le choix qu'on avait fait dans la partie 2.5 de poser $c_n = 2a_n - b_n$ et $d_n = 2a_n + b_n$, car

$$\begin{aligned} \begin{pmatrix} c_n \\ d_n \end{pmatrix} &= P^{-1} X_n = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} P^{-1} X_{n-1} \\ &= \begin{pmatrix} 3^n & 0 \\ 0 & (-1)^n \end{pmatrix} P^{-1} X_0 = \begin{pmatrix} 3^n & 0 \\ 0 & (-1)^n \end{pmatrix} \begin{pmatrix} c_0 \\ d_0 \end{pmatrix} = \begin{pmatrix} 3^n c_0 \\ (-1)^n d_0 \end{pmatrix}. \end{aligned}$$

Le critère de diagonalisation suivant va nous servir par la suite; il découle du théorème précédent.

Proposition 4.24. *La matrice A est diagonalisable si et seulement si*

$$\sum_{\lambda \in Sp(A)} \dim(E_\lambda(A)) = n.$$

Preuve: On va admettre ce résultat. □

Malheureusement pas toutes les matrices sont diagonalisables.

Lemme 4.25. *La matrice $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ n'est pas diagonalisable.*

Preuve: On va procéder par l'absurde. Supposons que N est diagonalisable. Soit $P \in \mathcal{M}_2(\mathbb{R})$ inversible, telle que $P^{-1}NP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ soit une matrice diagonale. Alors

$$\begin{pmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{pmatrix} = (P^{-1}NP)^2 = P^{-1}N^2P = P^{-1}0P = 0,$$

donc $\lambda_1 = \lambda_2 = 0$. Cela implique que $N = P^{-1}0P = 0$, ce qui est clairement faux. □

Comme les matrices diagonalisables sont souvent plus facile à traiter, il est intéressant d'avoir des critères pour les reconnaître. On commence par un critère suffisant, mais pas nécessaire; il suit facilement des résultats déjà mentionnés.

Corollaire 4.26. *Si A est telle que $Sp(A)$ contient n valeurs propres distinctes, alors A est diagonalisable.*

On insiste sur le fait que le condition que le n valeurs propres sont distinctes est suffisante pour la diagonalisation, mais pas nécessaire.

Preuve: Si $Sp(A)$ contient n valeurs propres distinctes $\lambda_1, \dots, \lambda_n$, alors

$$\dim(E_{\lambda_1}) + \cdots + \dim(E_{\lambda_n}) \geq n.$$

Mais on a vu que cette somme est toujours plus petite que n, donc

$$\dim(E_{\lambda_1}) + \cdots + \dim(E_{\lambda_n}) = n.$$

La proposition 4.24 implique que A est diagonalisable. □

Matrices symétriques

Théorème 4.27. *Si $A \in \mathcal{M}_n(\mathbb{R})$ est symétrique (c.-à-d. $A^T = A$) alors A est diagonalisable. De plus ses valeurs propres sont toutes réelles et il existe une matrice $P \in \mathcal{M}_n(\mathbb{R})$ inversible avec $P^{-1} = P^T$, telle que*

$$P^{-1}AP = D$$

est une matrice diagonale.

Une matrice $P \in \mathcal{M}_n(\mathbb{R})$ telle que $P^{-1} = P^T$ est dite orthogonale. Ce type de matrices a une signification particulière dans l'interprétation géométrique des espaces vectoriels; on va y revenir dans le chapitre 5. On va admettre ce théorème.

Matrices positives: Perron-Frobenius

Théorème 4.28. *Soit $A = (a_{ij})_{1\leq i,j\leq n} \in \mathcal{M}_n(\mathbb{R})$ une matrice avec $a_{ij} > 0$ pour tout $1 \leq i, j \leq n$. Alors il existe $\lambda_1 \in (0, +\infty)$ une valeur propre de A telle que:*

(i) *pour toute autre valeur propre $\lambda \in Sp(A) \setminus \{\lambda_1\}$, on a $|\lambda| < \lambda_1$;*

(ii) $\dim E_{\lambda_1}(A) = 1$ *(on dit que la valeur propre λ_1 est simple);*

(iii) *il existe $X \in E_{\lambda_1}(A)$ avec toutes les entrées de X réelles et strictement positives;*

(iv) *si $X \in \mathcal{M}_{n1}(\mathbb{R})$ est un vecteur propre de A avec toutes les entrées positives ou nulles, alors $X \in E_{\lambda_1}(A)$.*

On va admettre ce théorème (il s'agit d'un théorème difficile).

Mentionnons que ce résultat se généralise à certaines matrices aux entrées positives ou nulles. Ainsi, si $A \in \mathcal{M}_n(\mathbb{R})$ est telle que $a_{ij} \geq 0$ pour tout i, j et s'il existe $k \in \mathbb{N}$ tel que toutes les entrées de A^k sont strictement positives, alors la conclusion du théorème est encore valable pour A.

En pratique on rencontre souvent des matrices dont les entrées sont positives. Les matrices qui décrivent les évolutions de populations sont souvent de ce type (voir exemple dans la partie 4.5.2), ainsi que celles décrivant des chaines de Markov.

Supposons que le théorème de Perron-Frobenius s'applique à une matrice A et notons $\lambda_1 \geq 0$ la plus grande valeur propre de A et $X_1 \in \mathcal{M}_{n1}((0, +\infty))$ un vecteur propre associé. Si on suppose que A est diagonalisable avec valeurs propres $\lambda_1, \ldots, \lambda_n$ et vecteurs propres $(X_1, \ldots, X_n)$, alors pour tout vecteur $X = \sum_i \alpha_i X_i$ avec $\alpha_1 \neq 0$, on a

$$\frac{1}{\lambda_1^n} A^n X = \sum_{i=1}^{d} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^n X_i \xrightarrow[n\to\infty]{} \alpha_1 X_1.$$

Ainsi le "taux" de croissance de $A^n X$ est λ_1^n et on observe une distribution (après normalisation par λ_1^{-n}) proportionnelle au vecteur positif X_1.

Le même résultat asymptotique peut être montré même si A n'est pas diagonalisable. On en parlera plus dans l'exemple donné dans la partie 4.5.2.

Perron-Frobenius pour les matrices stochastiques Supposons que $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice stochastique à gauche. Alors tous les coefficients de A sont positifs ou nuls. Supposons de plus qu'il existe $N \geq 1$ tel que tous les coefficients de A^N soit strictement positifs (on dit alors que A est irréductible et apériodique). On peut alors appliquer le théorème de Perron-Frobenius à A (une version un peu plus générale du théorème quand $N > 1$) et déduire l'existence d'un vecteur propre $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ d'entrées strictement positives et de la valeur propre λ associée, qui est strictement positive et satisfait les points (i) et (ii) du théorème.

Du fait que $AX = \lambda X$ on peut déduire

$$\lambda \sum_{i=1}^n x_i = \sum_{i=1}^n (AX)_i = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_j = \sum_{j=1}^n \Big(x_j \sum_{i=1}^n a_{ij} \Big) = \sum_{j=1}^n x_j.$$

Dans la dernière égalité on a utilisé que $\sum_{i=1}^n a_{ij} = 1$ car A est stochastique à gauche. En fin, les x_i sont tous strictement positifs, donc $\sum_{i=1}^n x_i = \sum_{j=1}^n x_j > 0$. On peut simplifier par cette quantité pour trouver $\lambda = 1$.

On conclut que toutes les autres valeurs propres de A sont de module strictement plus petit que 1 et que tout vecteur propre de valeur propre 1 est proportionnel à X. Si on choisit X de sorte que $\sum_{i=1}^n x_i = 1$ (ce qui le détermine uniquement) on peut voir X comme une probabilité sur $\{1, \dots, n\}$ (avec une probabilité x_i associée à i).

De la discussion qui précède, on conclut que cette probabilité est

- *invariante*: car $AX = X$;
- *asymptotique*: car pour tout vecteur Y d'entrées positives se sommant à 1, on a

$$A^N Y \xrightarrow[N\to\infty]{} X.$$

De plus, chaque propriété détermine X de manière unique.

Exercice 4.2.
Soient $A \in \mathcal{M}_n(\mathbb{R})$ une matrice diagonalisable, avec n valeurs propres distinctes. Soit $B \in \mathcal{M}_n(\mathbb{R})$ une matrice qui commute avec A (c.-à-d. telle que $AB = BA$).

Montrer que B est aussi diagonalisable. De plus, montrer que B admet les mêmes vecteurs propres que A (mais pas forcement les mêmes valeurs propres).
Est-ce que le résultat reste vrai si les valeurs propres de A ne sont pas supposées distinctes?

Exercice 4.3.
Trouver une matrice diagonalisable A, et plusieurs matrices inversibles P telles que $P^{-1}AP$ est diagonale.
Que dire de A si $P^{-1}AP$ est diagonale pour toute matrice inversible $P \in \mathcal{M}_n(\mathbb{R})$.

4.5.2 Application: matrices de Leslie

Les matrices de Leslie sont un modèle d'évolution d'une population classée selon l'âge. On considère une population qui évolue en temps discret (on peut par exemple supposer qu'on la mesure tous les ans/mois) dont les individus peuvent avoir un âge compris entre 1 et $d \in \mathbb{N}$.

Ainsi, la population à tout moment $t \in \mathbb{N}$ est représenté par un vecteur $N(t) = \begin{pmatrix} n_1(t) \\ \vdots \\ n_d(t) \end{pmatrix}$.

Supposons que l'évolution suit la dynamique suivante:

$$\begin{aligned} n_{k+1}(t+1) &= s_k n_k(t), && \text{pour } 1 \leq k < d, \text{ et} \\ n_1(t+1) &= a_1 n_1(t) + \cdots + a_d n_d(t). \end{aligned}$$

où $s_1, \ldots, s_{d-1} \in (0, 1]$ et $a_1, \ldots a_d \in (0, +\infty)$ sont des paramètres fixés. Ainsi $1-s_k$ représente la proportion d'individus d'âge k qui meurent avant d'arriver à l'âge $k+1$; a_k représente le nombre moyen d'enfants d'un individu de la génération k.

On peut écrire l'évolution sous forme matricielle comme suit:

$$N(t+1) = \begin{pmatrix} a_1 & a_2 & \ldots & a_{d-1} & a_d \\ s_1 & 0 & \ldots & 0 & 0 \\ 0 & s_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & s_{d-1} & 0 \end{pmatrix} N(t) = AN(t),$$

où $A \in \mathcal{M}_d(\mathbb{R})$ est la matrice affichée. Même si la matrice A n'est pas strictement positive, A^d l'est, et on peut appliquer le théorème de Perron Frobenius à A.

Cherchons λ (différent de 0) et $N \in \mathcal{M}_{d1}(\mathbb{R}_+)$ tels que $AN = \lambda N$. Cela reviens à

$$\begin{aligned} n_{k+1} &= \frac{1}{\lambda} s_k n_k = \cdots = \frac{1}{\lambda^k} s_1 \ldots s_k n_1, && \text{pour } 1 \leq k < d, \text{ et} \\ n_1 &= \frac{1}{\lambda}(a_1 n_1 + \cdots + a_d n_d) = \Big(\sum_{k=1}^{d} \frac{1}{\lambda^k} s_1 \ldots s_{k-1} a_k \Big) n_1 \end{aligned}$$

Ainsi, λ est valeur propre de A (avec $n_1 \neq 0$) si et seulement si

$$\phi(\lambda) := \sum_{k=1}^{d} \frac{1}{\lambda^k} s_1 \ldots s_{k-1} a_k = 1;$$

de plus pour cette valeur propre, un vecteur propre associé est donné par $n_k = \frac{s_1 \ldots s_{k-1}}{\lambda^{k-1}}$ pour $k \geq 1$ (en particulier $n_1 = 1$).

On voit bien que $\phi(\lambda) = 1$ admet exactement une solution positive, qu'on va noter λ_1.

Supposons que A est diagonalisable, avec $Sp(A) = \{\lambda_1, \ldots, \lambda_d\}$. Soient $X_1, \ldots, X_d$ des vecteurs propres associés; ils forment une base de $\mathbb{R}^d$ qu'on note $\mathcal{X}$. Alors, si on suppose que λ_1 est la valeur propre positive,

$$|\lambda_i| < \lambda_1 \quad \forall i \geq 2.$$

Soit $N(0)$ une distribution de population initiale. On écrit alors $N(0) = \sum_i \alpha_i X_i$ où

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{pmatrix} = V_{\mathcal{X}}(N(0)) \in \mathcal{M}_{d1}(\mathbb{R}).$$

Un calcul immédiat montre que

$$N(t) = A^t N(0) = \sum_{i=1}^{d} \alpha_i \lambda_i^t X_i.$$

Ainsi

$$\lambda_1^{-t} N(t) = \alpha_1 X_1 + \sum_{i=2}^{d} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^t X_i \xrightarrow[t\to\infty]{} \alpha_1 X_1. \tag{4.8}$$

Cela doit être compris comme suit. Si $\alpha_1 > 0$, alors la population a un taux de croissance λ_1 et si on la normalise, elle devient proportionnelle à une distribution suivant l'âge donnée par X_1.

Une analyse plus approfondie des espaces propres des matrices montre que la convergence de (4.8) est vraie même si A n'est pas diagonalisable.

4.6 Polynôme caractéristique

Une façon de déterminer les valeurs propres d'une matrice est donnée par le critère suivant. Fixons $A \in \mathcal{M}_n(\mathbb{R})$.

Proposition 4.29. $\lambda \in \mathbb{R}$ *est valeur propre de* A *si et seulement si* $\det(\lambda I_n - A) = 0$.

On peut vérifier (par exemple par récurrence sur la taille de A) que la fonction $\lambda \mapsto \det(\lambda I_n - A)$ est un polynôme de degré n en λ (et de coefficient dominant 1). Il est appelé le *polynôme caractéristique*(viii) de A. On le note $\chi_A(\lambda)$.

La proposition 4.29 nous dit que les valeurs propres de A sont exactement les racines de χ_A.

Preuve: On a déjà vu (remarque 4.19) que $\lambda \in \mathbb{R}$ est une valeur propre de A si et seulement si $A - \lambda I_n$ n'est pas inversible. Par ailleurs, le théorème 3.5 nous dit que cela est équivalent à $\det(\lambda I_n - A) = 0$. □

Exercice 4.4.
Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ une matrice triangulaire supérieure. Calculer χ_A et montrer que

$$Sp(A) = \{a_{11}, \dots, a_{nn}\}.$$

4.6.1 Application: recherche de vecteurs propres

On est maintenant en mesure de décrire un algorithme pour chercher les valeurs et vecteurs propres d'une matrice $A \in \mathcal{M}_n(\mathbb{R})$.

(viii) Charakteristisches Polynom

On commence par calculer le polynôme caractéristique de A. Ensuite on trouve ses racines, notons les $\lambda_1, \dots, \lambda_k$ et notons $m_1, \dots, m_k \in \mathbb{N}$ leur multiplicité.

Pour chaque $1 \leq i \leq k$, en utilisant le pivot de Gauss, on résout l'équation matricielle $(\lambda_i I_n - A)X = 0$, pour $X \in \mathcal{M}_n(\mathbb{R})$. L'ensemble des solutions est $E_{\lambda_i} = \mathrm{Ker}(\lambda_i I_n - A)$. Vu que $\lambda_i I_n - A$ n'est pas inversible, on va trouver un espace de solutions de dimension au moins 1. On a ainsi trouver les valeurs propres $\lambda_1, \dots, \lambda_k$ et les vecteurs propres associés.

Si on trouve des espaces propres avec $\dim(E_{\lambda_1}) + \dots + \dim(E_{\lambda_k}) = n$, alors A est diagonalisable (voir la proposition 4.24). De plus, la preuve de la proposition 4.24, décrit comment créer une base de $\mathbb{R}^n$ formée de vecteurs propres de A. Le théorème 4.22 (plus précisément sa preuve) nous dit alors comment créer une matrice $P \in \mathcal{M}_n(\mathbb{R})$ qui diagonalise A. Si par contre, $\dim(E_{\lambda_1}) + \dots + \dim(E_{\lambda_k}) < n$, alors A n'est pas diagonalisable (voir à nouveau la proposition 4.24). On distingue trois situations:

(i) χ_A admet n racines distinctes (c.à-d. si $k = n$ et $m_1 = \dots = m_n = 1$). Alors la matrice A est diagonalisable et chaque espace propre est de dimension 1. (Rappelons nous du corollaire 4.26 qui nous dit que si $|Sp(A)| = n$, alors A est diagonalisable.)

Exemple: Posons

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

On trouve par un calcul direct $\chi_A(\lambda) = \lambda^3 - 6\lambda^2 + 11\lambda - 6 = (\lambda - 1)(\lambda - 2)(\lambda - 3)$. Ainsi on déduit que A est bien diagonalisable et que ses valeurs propres sont $1, 2$ et 3. Un calcul rapide nous offre une base de vecteurs propres associés, notamment

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}.$$

(ii) χ_A est *scindé*, c.à-d. $m_1 + \dots + m_k = n$, mais certaines racines ont des multiplicités plus grandes que 1. Alors, pour chaque $i = 1, \dots, k$, on calcule $E_i = \mathrm{Ker}(\lambda_i I_n - A)$ et on trouve

$$1 \leq \dim(E_i) \leq m_i.$$

(Le fait que $\dim(E_i) \leq m_i$ va être admis). La matrice est diagonalisable si et seulement si $\dim(E_i) = m_i$ pour chaque i.

Exemple: Posons

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

Par calcul direct $\chi_B(\lambda) = (\lambda - 1)^2(\lambda - 3)$ et $\chi_C(\lambda) = (\lambda - 3)(\lambda + 1)^2$. En utilisant le pivot de Gauss, on peut calculer les espaces propres de B. On trouve

$$\operatorname{Ker}(I_3-B) = \operatorname{Vect}\left[\begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}\right] \quad \text{et} \quad \operatorname{Ker}(3I_3-B) = \operatorname{Vect}\left[\begin{pmatrix}-1\\-1\\1\end{pmatrix}\right].$$

On en déduit que B est diagonalisable; une base de vecteurs propres est donnée par les bases de $\operatorname{Ker}(I_3 - B)$ et $\operatorname{Ker}(3I_3 - B)$. Plus précisément, si on pose

$$P = \begin{pmatrix}0 & 1 & -1\\1 & 0 & -1\\0 & 1 & 1\end{pmatrix} \quad \text{alors} \quad P^{-1}BP = \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 3\end{pmatrix}.$$

Le même type de calcul pour C nous permet d'obtenir ses espaces propres, à savoir

$$\operatorname{Ker}(3I_3 - C) = \operatorname{Vect}\left[\begin{pmatrix}1/2\\1\\1\end{pmatrix}\right] \quad \text{et} \quad \operatorname{Ker}(-I_3 - C) = \operatorname{Vect}\left[\begin{pmatrix}1\\2\\1\end{pmatrix}\right].$$

Ainsi C n'est pas diagonalisable.

(iii) Si on travaille sur $\mathbb{R}$ il se peut que le polynôme χ_A n'ai pas toutes ses racines réelles. Si c'est le cas (c.à-d. si $m_1 + \cdots + m_k < n$) alors A n'est pas diagonalisable dans $\mathbb{R}$, c.à-d. il n'existe pas $P \in \mathcal{M}_n(\mathbb{R})$ inversible telle que $P^{-1}AP$ est diagonale. (Pour voir si A est diagonalisable dans $\mathbb{C}$, voir les points (i) et (ii).)

Exemple: Posons

$$D = \begin{pmatrix}3 & -1 & 0\\6 & -3 & 2\\8 & -6 & 5\end{pmatrix}.$$

Un calcul direct montre que $\chi_D(\lambda) = \lambda^3 - 5\lambda^2 + 9\lambda - 5 = (\lambda-1)(\lambda^2-4\lambda+5)$. En utilisant la résolution des équations d'ordre 2, on voit que $\lambda^2 - 4\lambda + 5$ n'as pas de racines réelles. Ainsi D n'est pas diagonalisable par une matrice de $\mathcal{M}_n(\mathbb{R})$.

Si on prend en compte les racines complexes, alors presque tout polynôme admet n racines distinctes. Il faut avoir quelques notions d'analyse mathématique pour donner un sens précis à cette affirmation. Informellement, on peut dire que, si $A \in \mathcal{M}_n(\mathbb{C})$ est prise “au hasard”, alors elle est diagonalisable. Ou encore que toute matrice $A \in \mathcal{M}_n(\mathbb{C})$ peut être approchée par des matrices diagonalisables.

En général, quand $n \geq 5$, les racines de χ_A n'admettent pas d'expression exacte. Ainsi, on doit se contenter d'approximations numériques pour trouver $Sp(A)$.

À retenir

- Une application $u : \mathbb{R}^n \to \mathbb{R}^m$ est linéaire si $u(\lambda x + \mu y) = \lambda u(x) + \nu u(y)$ pour tout $x, y \in \mathbb{R}^n$ et $\lambda, \nu \in \mathbb{R}$.

- Si $\mathcal{E} = (e_1, \dots, e_n)$ est une base de $\mathbb{R}^n$ et $x \in \mathbb{R}^n$, alors $V_\mathcal{E}(x)$ est la décomposition de x dans la base $\mathcal{E}$:
$$V_\mathcal{E}(x) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \quad \text{où } x = \lambda_1 e_1 + \cdots + \lambda_n e_n.$$

- Pour des bases $\mathcal{E}$ et $\mathcal{F}$ de $\mathbb{R}^n$ et $\mathbb{R}^m$, respectivement, et $u : \mathbb{R}^n \to \mathbb{R}^m$ linéaire, u est représentée dans les bases $\mathcal{E}$ et $\mathcal{F}$ par la matrice $\mathrm{Mat}_{\mathcal{F},\mathcal{E}} \in \mathcal{M}_{m,n}$; c'est l'unique matrice telle que
$$V_\mathcal{F}(u(x)) = \mathrm{Mat}_{\mathcal{F},\mathcal{E}} V_\mathcal{E}(x) \qquad \forall x \in \mathbb{R}^n.$$
Quand $u : \mathbb{R}^n \to \mathbb{R}^n$, on utilise la même base pour $\mathcal{E}$ et $\mathcal{F}$.

- La multiplication des matrices correspond à la composition des applications. Une application linéaire est inversible si et seulement si la matrice associée l'est (dans une base quelconque).

- La représentation des vecteurs et des applications linéaires par des matrices dépend des bases $\mathcal{E}$ et $\mathcal{F}$. Pour passer d'une base à une autre on utilise la matrice de changement de base. Pour deux bases $\mathcal{E}$ et $\mathcal{B}$ de E et $x \in E$ et $u \in \mathcal{L}(E)$,
$$V_\mathcal{B}(x) = \mathrm{P}_{\mathcal{B},\mathcal{E}} V_\mathcal{E}(x) \quad \text{et} \quad \mathrm{Mat}_\mathcal{B}(u) = \mathrm{P}_{\mathcal{E},\mathcal{B}}^{-1} \mathrm{Mat}_\mathcal{E}(u) \mathrm{P}_{\mathcal{E},\mathcal{B}}.$$

- Pour une matrice carrée $A \in \mathcal{M}_n(\mathbb{R})$ on dit qu'un vecteur non-nul $X \in \mathbb{R}^n$ est *vecteur propre* de A si $AX = \lambda X$ pour un $\lambda \in \mathbb{R}$. Le scalaire λ s'appelle *valeur propre* de A. L'ensemble des valeurs propres est le *spectre* de A, noté $Sp(A)$.

- L'ensemble des vecteurs propres pour une valeur propre λ est le s.e.v.
$$E_\lambda(A) = \mathrm{Ker}(A - \lambda I_n).$$

- $\lambda \in Sp(A)$ si et seulement si $A - \lambda I_n$ n'est pas inversible.

- La matrice A est dite *diagonalisable* si elle est semblable à une matrice diagonale, c.-à-d. s'il existe $P \in \mathcal{M}_n(\mathbb{R})$ inversible, telle que $P^{-1}AP$ est diagonale.
Pas toutes les matrices sont diagonalisables!

- A est diagonalisable si et seulement s'il existe une base de $\mathbb{R}^n$ formée entièrement de vecteurs propres de A.

- Si A est diagonalisable par une matrice inversible P:

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix}, \quad \text{alors} \quad A^k = P \begin{pmatrix} \lambda_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^k \end{pmatrix} P^{-1}.$$

- Les matrices symétriques sont diagonalisables.

- Le *polynôme caractéristique* d'une matrice $A \in \mathcal{M}_n(\mathbb{R})$ est $\chi_A(\lambda) = \det(\lambda I_n - A)$. Ses racines sont exactement les valeurs propres de A.

- La procédure de la partie 4.6.1 pour vérifier si une matrice est diagonalisable.

Zu wissen

- Für eine quadratische Matrix $A \in \mathcal{M}_n(\mathbb{R})$ nennt man einen Vektor $X \in \mathbb{R}^n$ $X \neq 0$, einen *Eigenvektor* von A, wenn $AX = \lambda X$ für ein $\lambda \in \mathbb{R}$. Der Skalar λ heisst *Eigenwert* von A. Die Menge der Eigenvektoren ist das *Spektrum* von A, bezeichnet mit $Sp(A)$.

- Die Menge der Eigenvektoren für einen Eigenwert λ ist der Untervektorraum
$$E_\lambda(A) = \operatorname{Ker}(A - \lambda I_n).$$

- $\lambda \in Sp(A)$ genau dann, wenn $A - \lambda I_n$ nicht invertierbar ist.

- Die Matrix A heisst *diagonalisierbar*, wenn sie zu einer Diagonalmatrix ähnlich ist, d.h. wenn es eine invertierbare Matrix $P \in \mathcal{M}_n(\mathbb{R})$ gibt, so dass $P^{-1}AP$ diagonal ist.
Nicht alle Matrizen sind diagonalisierbar!

- A ist diagonalisierbar genau dann, wenn es für $\mathbb{R}^n$ eine Basis bestehend aus Eigenvektoren von A gibt.

- A ist diagonalisierbar genau dann, wenn $\sum_{\lambda \in Sp(A)} \dim(E_\lambda) = n$.

- Ist A diagonalisierbar mit einer invertierbaren Matrix P,
$$P^{-1}AP = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix}, \quad \text{dann gilt} \quad A^k = P \begin{pmatrix} \lambda_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^k \end{pmatrix} P^{-1}.$$

- Jede symmetrische Matrix in $\mathcal{M}_n(\mathbb{R})$ ist diagonalisierbar.

- Um die Umkehrbarkeit einer Matrix zu überprüfen, benutzt man die *Determinante*: $A \in \mathcal{M}_n(\mathbb{R})$ ist invertierbar genau dann, wenn $\det(A) \neq 0$.

- Für $A, B \in \mathcal{M}_n(\mathbb{R})$ gilt $\det(AB) = \det(A)\det(B)$.

- Die Wirkung der elementaren Zeilenumformungen:
  - Multiplikation einer Zeile der Matrix mit λ multipliziert die Determinante mit λ;
  - Vertauschen von zwei Zeilen multipliziert die Determinante mit -1;
  - Addition einer Zeile zu einer anderen verändert die Determinante nicht.

  Die Determinante kann mit Hilfe des Gauß-Algorithmus berechnet werden.

- Die Determinante einer Diagonalmatrix ist gleich dem Produkt der Diagonaleinträge.

- Das *charakteristische Polynom* einer Matrix $A \in \mathcal{M}_n(\mathbb{R})$ ist $\chi_A(\lambda) = \det(\lambda I_n - A)$. Seine Nullstellen sind genau die Eigenwerte von A.

- Das Vorgehen in Abschnitt 4.6.1, um zu überprüfen, ob eine Matrix diagonalisierbar ist.

Chapter 5

Géométrie des espaces vectoriels $\mathbb{R}^d$

On a déjà parlé de la représentation des vecteurs de $\mathbb{R}^d$ dans l'espace de dimension d; pour tous les exemples on prendra $d = 2$ ou $d = 3$. En plus de la structure d'espace vectoriel, l'espace ambient a des notions de longueur, aire, volume, angles etc. Dans cette partie, on explique la relation entre ces notions et la structure d'espace vectoriel de $\mathbb{R}^d$.

En mathématiques, les longueurs ne sont pas mesurés en unités; elles sont toutes données en utilisant comme unité les vecteurs de la base canonique $e_i = (0, \dots, 0, 1, 0, \dots, 0)$, dont la longueur est prise, par convention, comme égale à 1.

5.1 Produit scalaire, norme, angles

Définition 5.1. *Pour $x = (x_1, \dots, x_d)$ et $y = (y_1, \dots, y_d)$ des vecteurs de $\mathbb{R}^d$, le* produit scalaire *de x et y est*

$$\langle x, y\rangle = \sum_{i=1}^{d} x_i y_i = x_1 y_1 + \cdots + x_d y_d.$$

La norme (euclidienne) de x est

$$\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + \cdots + x_d^2}.$$

Observons que $\langle x, x\rangle \geq 0$ pour tout $x \in \mathbb{R}^d$, donc que la racine carrée dans la définition de $\|x\|$ existe.

Proposition 5.2. (a) *Pour $x, y, z \in \mathbb{R}^d$ et $\lambda \in \mathbb{R}$*

$$\begin{aligned} &\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle \qquad \langle \lambda x, z\rangle = \lambda\langle x, z\rangle, \\ &\langle x, z + y\rangle = \langle x, y\rangle + \langle x, z\rangle, \qquad \langle x, \lambda z\rangle = \lambda\langle x, z\rangle, \\ &\langle x, y\rangle = \langle y, x\rangle. \end{aligned}$$

On dit que le produit scalaire est une forme bi-linéaire symmetrique.

(b) *Pour* $x, y, z \in \mathbb{R}^d$ *et* $\lambda \in \mathbb{R}$

$$\|\lambda x\| = |\lambda|\,\|x\|,$$
$$\|x + y\| \leq \|x\| + \|y\|, \qquad \text{(inégalité triangulaire)}$$
$$\|x\| \geq 0, \ \textit{avec égalité ssi } x = 0.$$

On dit que $\|.\|$ *est une norme sur* $\mathbb{R}^d$*.*

(c) *Pour* $x, y \in \mathbb{R}^d$,

$$\langle x, y\rangle \leq \|x\|\,\|y\|, \qquad \text{(Cauchy-Schwarz)}$$

de plus l'égalité se produit si et seulement si $y = 0$ *ou s'il existe* $\lambda \geq 0$ *tel que* $x = \lambda y$ *(c.à-d. les vecteurs* x *et* y *sont colinéaires et de même sens).*

Preuve: (a) Suit directement de la formule définissant le produit scalaire.

(c) Fixons $x, y \in \mathbb{R}^d$. Considérons la fonction quadratique $f : \mathbb{R} \to \mathbb{R}$ donnée par

$$f(\lambda) = \lambda^2\langle y, y\rangle - 2\lambda\langle x, y\rangle + \langle x, x\rangle^2 = \langle x - \lambda y, x - \lambda y\rangle.$$

Par la définition du produit scalaire, on a $f(\lambda) \geq 0$ pour tout $\lambda \in \mathbb{R}$. Ainsi le discriminant de cette fonction quadratique est positif, donc

$$\Delta = 4\langle x, y\rangle^2 - 4\langle y, y\rangle\langle x, x\rangle \geq 0,$$

ce qui implique (Cauchy-Schwarz).

Analysons le cas d'égalité dans (Cauchy-Schwarz). Si $y = 0$, alors on obtient bien $0 = 0$ dans (Cauchy-Schwarz). Supposons par la suite que $y \neq 0$. Alors $\Delta = 0$ si et seulement si f admet une racine double réelle, donc si et seulement si $x = \lambda y$ pour un certain $\lambda \in \mathbb{R}$.

(b) Fixons $\lambda \in \mathbb{R}$ et $x, y \in \mathbb{R}^d$. Alors

$$\|\lambda x\| = \sqrt{\lambda^2 x_1^2 + \dots + \lambda^2 x_d^2} = \sqrt{\lambda^2}\sqrt{x_1^2 + \dots + x_d^2} = |\lambda|\,\|x\|.$$

De plus, par le point (a),

$$\begin{aligned}\|x + y\|^2 = \langle x + y, x + y\rangle &= \|x\|^2 + 2\langle x, y\rangle + \|y\|^2\\ &\leq \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,\end{aligned}$$

ou l'inégalité suit de (Cauchy-Schwarz). En prenant la racine carrée de cette équation, on obtient $\|x + y\| \leq \|x\| + \|y\|$.

En fin, si $\|x\| = 0$, alors $\sum_{i=1}^d x_i^2 = 0$, donc $x_i = 0$ pour chaque $i = 1, \dots, d$, ce qui revient à $x = 0$. Inversement $\|0\| = 0$ évidement. $\square$

Remarque 5.3. Si on écrit $x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$ et $y = \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}$ sous forme colonne, alors le produit scalaire s'écrit comme produit matriciel:

$$\langle x, y\rangle = x^T y = \begin{pmatrix} x_1, \dots, x_d \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} = x_1 y_1 + \cdots + x_d y_d.$$

Interpretation géométrique La norme d'un vecteur définie ci-dessus représente la longueur du vecteur (d'ou le nom de norme euclidienne). Ainsi, pour $x, y \in \mathbb{R}^d$, la distance entre les points x et y est la longueur du vecteur $x - y$, donc est égale à $\|x - y\|$.

Le produit scalaire de deux vecteurs x y peut s'écrire

$$\langle x, y\rangle = \|x\|\,\|y\| \cos\theta, \tag{5.1}$$

où $\theta \in (-\pi, \pi]$ représente l'angle entre les vecteurs x et y. Observons que l'inégalité de Cauchy-Schwarz nous dit que $\frac{\|x\|\,\|y\|}{\langle x,y\rangle} \in [-1, 1]$, ce qui nous permet d'écrire ce nombre comme le cosinus d'un angle.

Observons également que $\langle x, y\rangle = \|x\|\,\|y\|$ si et seulement si $\cos\theta = 0$, ce qui revient à dire que les vecteurs sont colinéaires et de même sens, ou encore que $x = \alpha y$ pour un $\alpha > 0$. Inversement $\langle x, y\rangle = -\|x\|\,\|y\|$ si et seulement si $\cos\theta = \pi$, ce qui revient à dire que les vecteurs sont colinéaires et de sens opposé, ou encore que $x = \alpha y$ pour un $\alpha < 0$.

5.2 Orthogonalité

On a vu apparaitre la notion de produit scalaire, qui induit la notion de longueur. Elle induit également la notion d'angle entre deux vecteurs.

Définition 5.4. *Soit $\langle \cdot, \cdot\rangle$ un produit scalaire sur $\mathbb{R}^d$. Deux vecteurs $x, y \in \mathbb{R}^d$ sont dit* orthogonaux *si $\langle x, y\rangle = 0$. On note ça $x \perp y$.*

Cette définition est cohérente avec (5.1), vu que $\cos\theta = 0$ si et seulement si $\theta = \pm\pi/2$. De plus, elle correspond à la notion d'orthogonalité géométrique qu'on a en dimensions deux et trois. En fin, observons que si $x \perp y$ alors

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2. \tag{Pythagore}$$

Définition 5.5. *Soient $\langle \cdot, \cdot\rangle$ un produit scalaire sur $\mathbb{R}^d$ et $\mathcal{F} = (f_1, \dots, f_n)$ une famille de vecteurs de $\mathbb{R}^d$.*

(i) *On dit que $\mathcal{F}$ est* orthogonale *si, pour tout $1 \le i < j \le n$, $f_i \perp f_j$.*

(ii) *On dit que $\mathcal{F}$ est* orthonormée *si elle est orthogonale et $\|f_i\| = 1$ pour tout $1 \le i \le n$.*

(iii) *On dit que $\mathcal{F}$ est une* base orthonormée *d'un sous-espace F de $\mathbb{R}^d$ si $\mathcal{F}$ est orthonormée et est une base de F.*

Proposition 5.6. *Soit $\mathcal{F} = (f_1, \ldots, f_n)$ une famille orthogonale de vecteurs non-nuls de $\mathbb{R}^d$. Alors $\mathcal{F}$ est libre.*

Preuve: Soit $\mathcal{F} = (f_1, \ldots, f_n)$ une famille orthogonale de vecteurs non-nuls de $\mathbb{R}^d$. Soient $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ une famille de scalaires tels que $\sum_{i=1}^n \lambda_i f_i = 0$. Alors, pour chaque $1 \leq j \leq n$, on peut écrire

$$0 = \langle f_j, 0\rangle = \langle f_j, \sum_{i=1}^n \lambda_i f_i\rangle = \sum_{i=1}^n \lambda_i \langle f_j, f_i\rangle = \lambda_j \|f_j\|^2. \tag{5.2}$$

Ainsi, $\lambda_1 = \cdots = \lambda_n = 0$, donc la famille est libre. □

Une consequence de cette proposition est que, si $\mathcal{F} = (f_1, \ldots, f_d)$ est une famille orthonormée dans $\mathbb{R}^d$ contenant d vecteur, alors $\mathcal{F}$ est une base. En effet, par la proposition précédente, $\mathcal{F}$ est libre; comme le nombre de vecteurs dans $\mathcal{F}$ est égal à la dimension de l'espace $\mathbb{R}^d$, $\mathcal{F}$ est également génératrice.

La proposition suivante est utile pour décomposer un vecteur dans une base orthonormée.

Proposition 5.7. *Soit $\mathcal{F} = (f_1, \ldots, f_d)$ une base orthonormée de $\mathbb{R}^d$. Alors, pour tout $x \in \mathbb{R}^d$,*

$$x = \langle x, f_1\rangle\, f_1 + \cdots + \langle x, f_d\rangle\, f_d.$$

Observons donc que les coordonnées de x dans la base $\mathcal{F}$ sont particulièrement faciles à calculer étant donné l'orthonormalité de $\mathcal{F}$.

Preuve: Pour $1 \leq j \leq d$ calculons le produit scalaire

$$\Big\langle x - \sum_{i=1}^d \langle x, f_i\rangle\, f_i,\ f_j\Big\rangle = \langle x, f_j\rangle - \sum_{i=1}^d \langle x, f_i\rangle\, \langle f_i, f_j\rangle = \langle x, f_j\rangle - \langle x, f_j\rangle = 0.$$

Par linéarité on trouve $(x - \sum_{i=1}^d \langle x, f_i\rangle\, f_i) \perp y$ pour tout $y \in \mathrm{Vect}(f_1, \ldots, f_d) = \mathbb{R}^d$. Ainsi $(x - \sum_{i=1}^d \langle x, f_i\rangle\, f_i) \perp (x - \sum_{i=1}^d \langle x, f_i\rangle\, f_i)$, d'ou $\|x - \sum_{i=1}^d \langle x, f_i\rangle\, f_i\| = 0$, ce qui implique $x = \sum_{i=1}^d \langle x, f_i\rangle\, f_i$. □

5.3 Matrices et applications linéaires orthogonales

5.3.1 Définitions

Définition 5.8. *Une application linéaire $u : \mathbb{R}^d \to \mathbb{R}^d$ est dite orthogonale si et seulement si*

$$\langle u(x), u(y)\rangle = \langle x, y\rangle \qquad \text{pour tous } x, y \in \mathbb{R}^d.$$

On dit qu'une matrice $M \in \mathcal{M}_d(\mathbb{R})$ est orthogonale si c'est la matrice d'une application linéaire orthogonale dans la base canonique.

Notons $\mathcal{E}$ la base canonique de $\mathbb{R}^d$. Si $M = \mathrm{Mat}_{\mathcal{E}}(u)$ est la matrice d'une application linéaire orthogonale u dans la base $\mathcal{E}$, alors $\langle Me_i, Me_j\rangle = 0$ si $i \neq j$ et 1 si $i = j$. Mais Me_i est la $i^{\text{ème}}$ colonne de M, notons la C_i. On conclut donc que $C_1, \dots, C_d$ forment une famille orthornormée.

En fin, observons aussi que $\langle Me_i, Me_j\rangle = C_i^T C_j$ est l'entree de la ligne i et la colonne j de la matrice $M^T M$. Ainsi, on arrive au résultat suivant.

Lemme 5.9. *Pour $M \in \mathcal{M}_d(\mathbb{R})$ on a équivalence de*

(i) *M est orthogonale;*

(ii) *les colonnes de M forment une famille orthonormée;*

(iii) *$M^T M = I_d$;*

(iv) *M est inversible et $M^{-1} = M^T$.*

Notons $O(n)$ l'ensemble des matrices orthogonales de dimension n. Alors, si $M \in O(n)$,

$$1 = \det M^T M = \det M^T \, \det M = (\det M)^2,$$

donc $\det M \in \{\pm 1\}$. Notons $SO(d)$ les matrices de $O(n)$ de déterminant $+1$.

Il est évident que l'application successive de deux transformations orthogonales est une transformation orthogonale. La même chose s'applique alors aux matrices orthogonales. On obtient donc le résultat suivant.

Lemme 5.10. *Soient $u, v : \mathbb{R}^d \to \mathbb{R}^d$ deux applications linéaire orthogonales. Alors $u \circ v$ est une application linéaire orthogonale.*

Soient $A, B \in \mathcal{M}_d$ deux matrices orthogonales. Alors AB est une matrice orthogonale.

En fin, supposons que $\mathcal{E} = (e_1, \dots, e_d)$ et $\mathcal{F} = (f_1, \dots, f_d)$ sont deux bases orthonormées de $\mathbb{R}^d$. Alors la matrice de changement de base $\mathrm{P}_{\mathcal{F},\mathcal{E}}$ de $\mathcal{E}$ à $\mathcal{F}$ est orthogonale et donc $\mathrm{P}_{\mathcal{E},\mathcal{F}} = \mathrm{P}_{\mathcal{F},\mathcal{E}}^{-1} = \mathrm{P}_{\mathcal{F},\mathcal{E}}^T$. En particulier, si u est une application linéaire,

$$\mathrm{Mat}_{\mathcal{F}}(u) = \mathrm{P}_{\mathcal{E},\mathcal{F}}^T \mathrm{Mat}_{\mathcal{E}}(u) \mathrm{P}_{\mathcal{E},\mathcal{F}}.$$

Cette formule explique l'intérêt du théorème 4.27.

5.3.2 Dimensions 2 et 3

Dimension 2. Les transformations linéaires orthogonales de $\mathbb{R}^2$ dont le déterminant vaut $+1$ (c.à-d. les éléments de $SO(2)$) sont les rotations. La matrice d'une rotation d'angle θ dans la base canonique s'écrit $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

Un exemple de transformation linéaire orthogonale de $\mathbb{R}^2$ dont le déterminant vaut -1 est la symétrie par rapport à l'axe horizontale. Sa matrice dans la base canonique s'écrit $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Toute transformation orthogonale de $\mathbb{R}^2$ de déterminant -1 est la composition de cette réflexion et d'une transformation orthogonale de déterminant $+1$. Ainsi, les transformations linéaires orthogonales dans $\mathbb{R}^2$ de déterminant -1 correspondent aux matrices de la forme $\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$ avec $\theta \in [0, 2\pi)$.

Dimension 3. Les transformations linéaires orthogonales dans $\mathbb{R}^3$ dont le déterminant vaut $+1$ (c.à-d. les éléments de $SO(3)$) sont les rotations dans $\mathbb{R}^3$. Chaque telle rotation est déterminé une axe et un angle de rotation. Si l'axe de rotation est colinéaire au vecteur $(0, 0, 1)$ et l'angle est θ, alors la matrice de la rotation dans la base canonique s'écrit

$$\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ \sin\theta & -\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Si M est la matrice d'une rotation dans $\mathbb{R}^3$ d'axe arbitraire, alors il existe une base orthonormée dont le troisième vecteur est l'axe de la rotation. Ainsi, si on change de base, M s'écrit comme avant:

$$M = P^T \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ \sin\theta & -\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} P,$$

pour une certaine matrice orthogonale P et $\theta \in [0, 2\pi)$.

Passons aux matrices orthogonales de dimension 3 et de déterminant -1. Vu que pour les matrices A de dimension 3, $\det(-A) = -\det A$, une matrice M est orthogonale de déterminant -1 si et seulement si $-M \in SO(3)$.

5.4 Projection orthogonale

Soit F un sous-espace vectoriel de $\mathbb{R}^d$. On pose

$$F^\perp = \{x \in \mathbb{R}^d \,:\, x \perp y \text{ pour tout } y \in F\}.$$

Proposition 5.11. *$F^\perp$ est un sous-espace vectoriel de $\mathbb{R}^d$ de dimension* $\dim F^\perp = d - \dim F$.

Preuve: Admis. □

Proposition 5.12. (a) *Pour tout $x \in \mathbb{R}^d$, il existe une unique décomposition*

$$x = \mathrm{proj}_F(x) + \mathrm{proj}_{F^\perp}(x),$$

avec $\mathrm{proj}_F(x) \in F$ *et* $\mathrm{proj}_{F^\perp}(x) \in F^\perp$.

(b) *L'application* $x \mapsto \mathrm{proj}_F(x)$ *est linéaire.*

(c) *On a*

$$\|x - \mathrm{proj}_F(x)\| = \min_{y \in F} \|x - y\| =: \mathrm{dist}(x, F),$$

et $\mathrm{proj}_F(x)$ *est le seul* $y \in F$ *qui réalise le minimum dans l'équation précédente.*

Preuve: Admis. □

Calcul de la projection orthogonale quand $\dim(F) = 1$. Si $\dim(F) = 1$, alors $F = \mathrm{Vect}(y)$ pour $y \in \mathbb{R}^d$ avec $y \neq 0$. Dans ce cas,

$$\mathrm{proj}_{\mathrm{Vect}(y)}(x) = \frac{\langle x, y\rangle}{\|y\|^2} y.$$

Calcul de la projection orthogonale en général. Soit $\mathcal{E} = (e_1, \dots, e_d)$ est une base orthonormée et supposons pour commencer que $F = \mathrm{Vect}(e_1, \dots, e_k)$ pour un certain $1 \leq k \leq d$. Alors

$$\mathrm{proj}_{\mathrm{Vect}(e_1,\dots,e_k)}(x) = \sum_{i=1}^{k} \langle x, e_i\rangle\, e_i.$$

Dans le cas ou F est un sous-espace vectoriel de dimension générale k, on peut construire une base orthonormée $(e_1, \dots, e_d)$ de $\mathbb{R}^d$ de telle sorte que $e_1, \dots, e_k \in F$. On utilise pour cela l'algorithme de Gram-Schmidt. Pour calculer $\mathrm{proj}_F(x)$ on utilise alors la formule précédente.

Produit vectoriel. On se place ici dans le cas de $\mathbb{R}^3$. Soient $a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ et $b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ deux vecteurs de $\mathbb{R}^3$. Alors le produit vectoriel de a et b est défini par

$$a \times b = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \qquad \text{avec } c_1 = \det\begin{pmatrix} a_2 & b_2 \\ a_3 & b_3 \end{pmatrix},\ c_2 = -\det\begin{pmatrix} a_1 & b_1 \\ a_3 & b_3 \end{pmatrix},\ c_3 = \det\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}.$$

Le vecteur $a \times b$ obtenu ainsi est

- orthogonal à $\mathrm{Vect}(a, b)$;

- $\|a \times b\| = \|a\|\,\|b\| \sin\theta$ ou $\theta \in [0, \pi]$ est l'angle entre les deux vecteurs a et b;
- le sens est donnée par la règle de la main droite: si l'indexe de la main droite pointe dans la direction de a et le majeur dans celle de b, alors le pouce pointe dans la direction de $a \times b$.

Observons que $a \times b$ est de norme 0 si a et b sont colinéaires. Plus généralement, $\|a \times b\|$ représente l'aire du parallélogramme de cotés a et b. De plus, si a, b engendrent un espace de dimension 2, alors

$$\mathrm{Vect}(a,b)^\perp = \mathrm{Vect}(a \times b) \quad \text{ et } \quad \mathrm{proj}_{\mathrm{Vect}(a,b)^\perp}(x) = \frac{\langle x, a \times b\rangle}{\|a \times b\|^2}\, a \times b.$$

Le produit vectoriel a les mêmes propriétés que le produit sur $\mathbb{R}$ (associativité, distributivité) avec une exception importante: il est anti-commutatif. En effet, $a \times b = -b \times a$.

Régression linéaire. Supposons qu'on dispose d'un grand nombre d'observations, chacune formée de $d+1$ nombres: $(a_{i1}, \dots, a_{i\,d}, b_i) \in \mathbb{R}^{d+1}$ pour $i = 1, \dots, n$. Les données ont une structure particulière: les variables $a_{i1}, \dots, a_{i\,d}$ (appelées variables explicatives, variables indépendantes, variables exogènes ou encore prédicteurs) sont essentiellement aléatoires, mais elles donnent une information sur les variables b_i (qu'on appelle des variables expliquées, variables dépendantes, variables endogènes ou encore réponses). La dépendance des variables expliquées en termes des variables explicatives est supposée (approximativement) linéaire:

$$b_i = a_{i1}x_1 + \dots + a_{i\,d}x_d + \epsilon_i \qquad \text{pour chaque } 1 \le i \le n, \tag{5.3}$$

ou $x_1, \dots, x_d \in \mathbb{R}$ sont des paramètres et $\epsilon_1, \dots, \epsilon_n \in \mathbb{R}$ sont des erreurs. Observons que (5.3) s'écrit sous forme matricielle comme

$$B = AX + \epsilon$$

avec $B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in \mathcal{M}_{n,1}$, $A = (a_{ij})_{\substack{1 \le i \le n \\ 1 \le j \le d}} \in \mathcal{M}_{n,d}$, $X = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \in \mathcal{M}_{d,1}$ et $\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \in \mathcal{M}_{n,1}$.

On dit que les paramètres $x_1, \dots, x_d$ explique bien la dépendance linéaire si les erreurs dans (5.3) sont petites. Evidement, le mieux est d'avoir $\epsilon_1 = \dots = \epsilon_n = 0$, ce qui revient à $AX = B$. Si cela n'est pas possible, on va choisir X de manière à minimiser l'erreur totale:

$$\|\epsilon\|^2 = \epsilon_1^2 + \dots + \epsilon_n^2.$$

En pratique, le nombre d de variables explicatives est petit (par exemple $d = 1, 2$ ou 3) alors que le nombre d'observations est grand. Il est alors peu probable que le système linéaire $AX = B$ de n équations à d inconnues admettent une solution.

On va dire que X est solution du problème des moindres carrées si

$$\|AX - B\| = \min\{\|AY - B\| \,:\, Y \in \mathcal{M}_{d,1}\}.$$

Rappelons que $\{AY : Y \in \mathcal{M}_{d,1}\} = \mathrm{Im}(A)$ et donc que

$$\min\{\|AY - B\| : Y \in \mathcal{M}_{d,1}\} = \min\{\|B - Z\| : Z \in \mathrm{Im}(A)\} = \mathrm{dist}(B, \mathrm{Im}(A)).$$

La projection orthogonale nous dit alors qu'il existe exactement un $Z \in \mathrm{Im}(A)$ qui réalise ce minimum, notamment $\hat{B} = \mathrm{proj}_{\mathrm{Im}(A)}(B)$. Ainsi, X est solution du problème des moindres carrées si et seulement si $AX = \hat{B}$.

Supposons par la suite que les colonnes de A forment une famille libre de $\mathbb{R}^n$ (cela est généralement le cas, car il s'agit d'un petit nombre de vecteurs dans un espace de très grande dimension). Alors le système linéaire de n équations à d inconnues $AX = \hat{B}$ admet une unique solution. De plus, la matrice $A^T A \in \mathcal{M}_{d,d}$ est inversible. En fin, dans cette situation, on a équivalence de:

(a) X est solution du problème des moindres carrées;
(b) $AX = \hat{B}$;
(c) $A^T AX = A^T B$;
(d) $X = (A^T A)^{-1} A^T B$.

Observons que la résolution de (c) est relativement simple, il s'agit de résoudre un système de d équations à d inconnues.

Preuve: On viens de voir que (a) et (b) sont équivalents. D'autre part, (c) et (d) sont trivialement équivalents car $A^T A$ est inversible.

Pour voir que (b) implique (c), notons $C_1, \dots, C_d$ les colonnes de A et rappelons que $\mathrm{Im}(A) = \mathrm{Vect}(C_1, \dots, C_d)$. Alors, par la définition de $\hat{B}$, $\langle C_j, B - \hat{B}\rangle = C_j^T (B - \hat{B}) = 0$ pour tout $1 \le j \le d$. Ainsi $A^T(B - \hat{B}) = 0$, ce qui s'écrit encore $A^T B = A^T \hat{B}$. Il s'en suit que si $AX = \hat{B}$, alors $A^T AX = A^T \hat{B} = A^T B$.

En fin, montrons que (c) implique (b). Les deux systèmes linéaires $AX = \hat{B}$ et $A^T AX = A^T B$ ont une seule solution chacun. D'autre part, on vient de montrer qu'une solution du premier est aussi solution du deuxième. Cela implique qu'ils ont exactement la même solution. $\square$

À retenir

- La définition du produit scalaire $\langle x, y\rangle = \sum_{i=1}^d x_i y_i$ sur $\mathbb{R}^d$; les propriétés de bi-linéarité et symétrie.
- Que $\|x\| := \sqrt{\langle x, x\rangle}$ est la norme euclidienne sur $\mathbb{R}^d$; ses propriétés; l'inégalité de Cauchy-Schwarz. La norme représente la *longueur* des vecteurs; $\|x-y\|$ est la distance entre les points $x, y \in \mathbb{R}^d$.
- On dit que x est orthogonal à y (noté $x \perp y$) si $\langle x, y\rangle = 0$. Si $x \perp y$, alors $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.
- Une base $(e_i)_{1\le i\le d}$ est dite orthonormée si $e_i \perp e_j$ pour $i \neq j$ et $\|e_i\| = 1$ pour chaque i. La décomposition d'un vecteur x dans une base orthonormée est simple à calculer:

$$x = \sum_{i=1}^d \langle x, e_i\rangle\, e_i.$$

- Une matrice $M \in \mathcal{M}_d$ est orthogonale si $M^{-1} = M^T$. Une telle matrice correspond à une transformation de l'espace qui préserve le produit scalaire (et par conséquent les distances et les angles). Elle correspond également à un changement de base, d'une base orthonormée vers une autre.
- Pour F sous-espace vectoriel de $\mathbb{R}^d$, la projection orthogonale $\mathrm{proj}_F(x)$ sur F d'un point $x \in \mathbb{R}^d$ est le point de F le plus proche de x.

Références

[1] Anton, H. and Rorres C., *Elementary Linear Algebra: Applications Version*, Wiley 11^{th} ed. (2014).

[2] Lay, D. and Lay, S. and McDonald., *Linear algebra and its applications*, Pearson 5^{th} ed. (2016). Existe également en français.

[3] Loser, F., *Un premier cours de logique*. http://www.math.ens.fr/~loeser/cours_26_01_2010.pdf.

[4] Strang, G., *Linear Algebra and Its Applications*, Academic Press (1976).

[5] https://fr.wikipedia.org/wiki/Table_de_symboles_mathématiques
https://de.wikipedia.org/wiki/Liste_mathematischer_Symbole

[6] 3Blue1Brown, chaine YouTube https://www.youtube.com/playlist?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab