Algèbre Linéaire

S. Friedli (CMS, EPFL)

2024-08-12

Table des matières

Chapitre I

Préface

I.1 À propos de ce cours

"For deep learning, it is linear algebra that matters the most. "

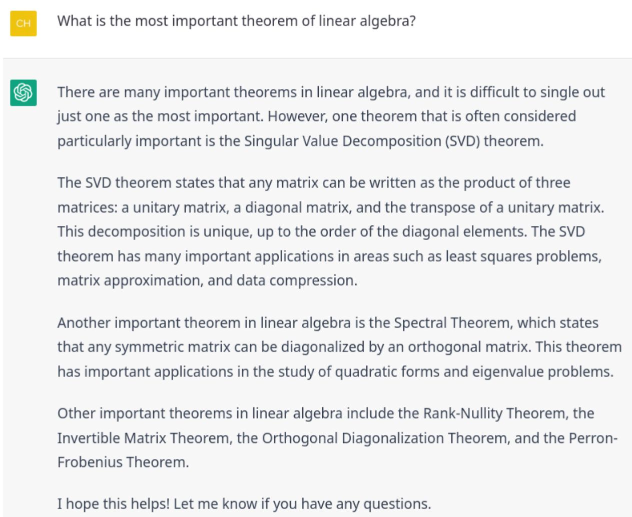
Gilbert Strang, Linear algebra and learning from data

Ce polycopié, qui est en construction, contient l'essentiel de mon cours d'algèbre linéaire, donné aux étudiant.e.s des sections de Génie Civil, Génie Électrique, et Sciences de l'Environnement, à l'EPFL.

Initialement, la structure de ce cours était empruntée (à l'exception du dernier chapitre sur les chaînes de Markov) au cours donné par Simone Deparis (SMA, EPFL), lui-même basé sur celui du Professeur Assyr Abdulle (SMA, EPFL). Je remercie Simone et Assyr de m'avoir fourni ce matériel, qui a grandement facilité la préparation de mes cours et de mes séries d'exercices à mon arrivée à l'EPFL.

Bien-sûr, la rédaction d'un texte engendre l'utilisation d'un style, des changements, des réarrangements, des ajouts de matériel supplémentaire, etc. et donc le contenu de ce cours va progressivement s'écarter de ce qu'était le cours d'Assyr.

Le format online de ce polycopié est emprunté de celui de mon cours d'**Analyse 1** (lien web). On trouvera sur ce dernier plusieurs information additionnelles que je n'ai pas jugé nécessaire de reproduire ici.



I.2 Références bibliographiques

- ⋆ **D.C. Lay**, *Algèbre linéaire et applications*, $4^{ème}$ édition, Parson. Ce texte est celui adopté par l'ensemble des enseignants d'algèbre linéaire. Il est volumineux mais très facile à suivre. Contient de nombreux exercices.
- ⋆ Le **MOOC (Massive Online Open Course) d'Algèbre linéaire** (lien web) de Donna Testerman (SMA, EPFL) contient tout ce qui est dit ici, et bien plus encore.
- ⋆ Le livre de **S. Balac et F. Sturm**, *Algèbre et analyse* ($2^{ème}$ édition, Presses polytechniques et universitaires romandes), est aussi une référence intéressante. http ://scherk.pbwiki.com/
- ⋆ Le livre de **Jacques Rappaz** et **Marco Picasso**, *Introduction à l'analyse numérique* (Presses polytechniques et universitaires romandes), montre également quelques applications de l'algèbre linéaire à la résolution numérique de certains problèmes classiques de physique.
- ⋆ **Hamilton Prado Bueno**, *Álgebra Linear, Um segundo curso*. Textos Universitários, Sociedade Brasileira de Matemática.
- ⋆ Pour d'autres quiz (sur l'algèbre linéaire et d'autres chapitres des mathématiques), créés par **Terence Tao** (lien web), cliquer **ici** (lien web).

Chapitre 1

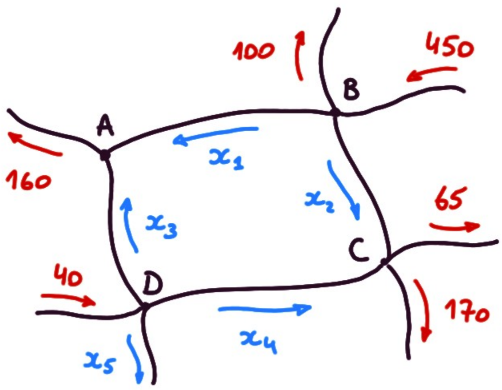
Systèmes linéaires

1.1 Motivation

Ce premier chapitre présente les systèmes d'équations qui seront étudiés dans ce cours. Voyons comment ce type de système peut apparaître, dans des situations très pratiques.

Traffic routier

Dans une petite ville ne possédant que 4 croisements, on a mesuré les flux de voitures sur quelques axes routiers entrants et sortants de la ville, dans le but de prévoir les flux résultant sur le réseau interne, et de préparer les aménagements nécessaires :



Ces mesures indiquent, par exemple, que le flux de voitures entrant au croisement B, venant de l'est de la ville, est de 450 voitures par heure.

Étant données ces contraintes, se pose la question de savoir s'il est possible de calculer les flux résultants sur les autres axes, indiqués par les lettres x_1 à x_5 sur la figure.

Le principe régissant les flux à un croisement est le même que celui utilisé dans les réseaux électriques (Loi de Kirchoff) : en chaque noeud du réseau, la somme des flux entrants doit être égal à la somme des flux sortants, ce qui donne, en chacun des points du réseau,

$$\begin{aligned}
A:&\quad +x_1+x_3=160\\
B:&\quad 450=100+x_1+x_2\\
C:&\quad x_2+x_4=65+170\\
D:&\quad 40=x_3+x_4+x_5
\end{aligned}$$

On peut récrire ces relations comme suit :

$$\begin{cases} x_1 & & & + & x_3 & & & & & = & 160 \\ x_1 & + & x_2 & & & & & & & = & 350 \\ & & x_2 & & & + & x_4 & & & = & 235 \\ & & & & x_3 & + & x_4 & + & x_5 & = & 40 \end{cases}$$

Plusieurs questions se posent :

- ⋆ Existe-t-il des nombres x_1, x_2, x_3, x_4, x_5 satisfaisant simultanément à ces 4 conditions ?
- ⋆ Si oui, ces nombres sont-ils tous positifs, pour respecter les sens imposés sur la figure, ou alors certains de ces nombres sont-ils négatifs (auquel cas on devra inverser le sens du traffic sur les axes concernés) ?
- ⋆ Si oui toujours, est-ce que ces nombres sont uniques ? Existe-t-il plusieurs solutions ? Et si ils ne sont pas uniques, quelles contraintes y a-t-il sur les choix que l'on peut faire ?
- ⋆ Et si la ville contenait des milliers de croisements, avec des centaines de flux entrants/sortants ?

Le système de 4 équations à 5 inconnues ci-dessus est un exemple de ce qu'on appelle un *système linéaire*. Ce type de système forme une part importante de ce cours, et on commencera leur étude dans la section suivante.

Informel 1.1. La dernière question ("et si la ville était beaucoup plus grande ?") montre qu'il est important d'aborder l'étude de ces systèmes de façon rigoureuse, en acceptant qu'ils peuvent être de taille arbitrairement grande.

Discrétisation d'équations différentielles

(faire)

1.2 Définition et exemples

Définition 1.2. Un **système linéaire en les variables** $x_1, x_2, \dots, x_n$ est une famille d'équations du type suivant :

$$(*) \begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & & & & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases}$$

Un tel système est dit "$m \times n$" : il contient m équations, et n variables. Les nombres a_{kj} ($1 \leqslant k \leqslant m$, $1 \leqslant j \leqslant n$) sont appelés les **coefficients** du système, les b_k ($1 \leqslant k \leqslant m$) forment le **second membre**.

Remarque 1.3. Insistons sur le fait que les coefficients a_{ij}, ainsi que le second membre, sont des *nombres fixés* qui ne dépendent pas des x_i ; en général ils sont donnés par une situation pratique. ⋄

On peut voir un système $(*)$ comme une famille de m **contraintes** que les variables $x_1, \dots, x_n$ doivent satisfaire, où la $k$ème contrainte est

$$a_{k1}x_1 \quad + \quad a_{k2}x_2 \quad + \quad \cdots \quad + \quad a_{kn}x_n \quad = \quad b_k$$

(On appelle cette contrainte une **équation linéaire**.)

Résoudre un système linéaire

Considérons un système $m \times n$ donné, comme $(*)$.

Définition 1.4. Une famille de nombres $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ est **solution** de $(*)$ si elle satisfait simultanément aux m contraintes spécifiées par $(*)$. L'ensemble des solutions de $(*)$ est noté $S_{(*)}$.

- ⋆ S'il existe au moins une solution, $S_{(*)} \neq \varnothing$, on dit que $(*)$ est **compatible**.
- ⋆ S'il n'existe aucune solution, $S_{(*)} = \varnothing$, on dit que $(*)$ est **incompatible** (ou **singulier**).

Lorsqu'un système est compatible, le **résoudre** signifiera trouver *toutes* ses solutions. Dans ce cas, on devra aussi savoir décrire précisément $S_{(*)}$. Voyons deux exemples simples.

Exemple 1.5. Le système 2×2

$$(*) \begin{cases} x_1 & + & x_2 & = & 1 \\ x_1 & + & x_2 & = & 0 \end{cases}$$

est incompatible. En effet, quelle que soit la valeur de x_1 et x_2, la somme $x_1 + x_2$ ne peut pas être à la fois égale à 1 et à 0. Donc $S_{(*)} = \varnothing$. ⋄

Exemple 1.6. Considérons le système 1×2 suivant :

$$\begin{cases} x_1 & + & x_2 & = & 1 \end{cases}$$

On trouve facilement des solutions : $(1, 0)$, $(2, -1)$, $(3, -2)$, etc. Donc ce système est compatible, et semble même posséder une infinité de solutions. Pour décrire son ensemble de solutions précisément (pour n'en oublier aucune !), il suffit de remarquer que l'on peut toujours *choisir* une des variables, et prendre l'autre en fonction de façon à ce que la relation soit satisfaite. Par exemple, en choisissant x_1, on garantit que la contrainte est satisfaite en prenant

$$x_2 = 1 - x_1 \,.$$

Lorsqu'on peut ainsi choisir une variable, appelée **variable libre**, on a avantage à y penser comme à un *paramètre*, et à utiliser une autre lettre pour la décrire. Si on utilise la lettre t pour ce paramètre, on a

$$\begin{aligned} x_1 &= t \,, \\ x_2 &= 1 - t \,. \end{aligned}$$

Les variables x_1 et x_2 étant exprimées en fonction des variables libres, on les appelle **variables liées** (ou **variables de base**). On peut finalement exprimer l'ensemble des solutions comme suit :

$$S = \{(t, 1-t) \,:\, t \in \mathbb{R}\} \,.$$

⋄

1.3 Sur le nombre de solutions d'un système linéaire

Un de nos objectifs dans ce qui suit sera de trouver des conditions suffisantes pour déterminer si un système est compatible ou incompatible.

Mais avant cela, nous allons énoncer une propriété générale, satisfaite par n'importe quel système linéaire, concernant le *nombre* de solutions qu'il peut posséder.

Théorème 1.7. *Si un système est compatible, alors soit il possède exactement une solution, soit il en possède une infinité.*

Preuve: **Remarque :** Un peu plus loin dans le cours, nous redonnerons la preuve de ce théorème, mais en utilisant le langage vectoriel, ce qui la rendra plus transparente.

Considérons un système $m \times n$ de la forme $(*)$, que l'on suppose être compatible. Si sa solution n'est pas unique, c'est qu'il existe au moins deux familles distinctes, que l'on notera $(\bar{x}_1, \dots, \bar{x}_n)$ et $(\bar{y}_1, \dots, \bar{y}_n)$, qui sont toutes deux solutions de $(*)$; cela signifie qu'elles satisfont toutes les contraintes : pour tout $k = 1, 2, \dots, m$, on a d'une part que

$$a_{k1}\bar{x}_1 \quad + \quad \cdots \quad + \quad a_{kn}\bar{x_n} \quad = \quad b_k$$

et d'autre part que

$$a_{k1}\bar{y}_1 \quad + \quad \cdots \quad + \quad a_{kn}\bar{y_n} \quad = \quad b_k$$

Prenons alors un réel λ quelconque, différent de 0 et de 1, et définissons la famille $(\bar{z}_1, \dots, \bar{z}_n)$, où

$$z_j := \lambda\bar{x}_j + (1-\lambda)\bar{y}_j \,, \quad j = 1, 2, \dots, n \,.$$

Montrons alors que $(\bar{z}_1, \dots, \bar{z}_n)$ est aussi solution, en montrant qu'elle satisfait chacune des m contraintes du système. En effet,

$$\begin{aligned}
&a_{k1}\bar{z}_1 + \cdots + a_{kn}\bar{z_n} \\
&\quad = a_{k1}(\lambda\bar{x}_1 + (1-\lambda)\bar{y}_1) + \cdots + a_{kn}(\lambda\bar{x}_n + (1-\lambda)\bar{y}_n) \\
&\quad = \lambda\big(\underbrace{a_{k1}\bar{x}_1 + \cdots + a_{kn}\bar{x_n}}_{=b_k}\big) + (1-\lambda)\big(\underbrace{a_{k1}\bar{y}_1 + \cdots + a_{kn}\bar{y_n}}_{=b_k}\big) \\
&\quad = \lambda b_k + (1-\lambda) b_k \\
&\quad = b_k \,,
\end{aligned}$$

ce qui signifie que la $k$ème contrainte est satisfaite.
Puisque $(\bar{x}_1, \dots, \bar{x}_n)$ et $(\bar{y}_1, \dots, \bar{y}_n)$ sont distinctes, il existe au moins un k tel que $\bar{x}_k \neq \bar{y}_k$. Cela signifie que si λ n'est égal ni à 0 ni à 1, le nombre $z_k = \lambda\bar{x}_k + (1-\lambda)\bar{y}_k$ est différent à la fois de $\bar{x}_k$ et de $\bar{y}_k$. On peut donc, en choisissant λ, construire autant de nouvelles solutions. Ceci signifie que le système possède une infinité de solutions. □

Informel 1.8. En d'autres termes, le nombre de solutions de n'importe quel système linéaire ne peut être que 0 (s'il est incompatible), 1 ou ∞. Plus tard on se référera à ce résultat comme **le Théorème "$0, 1, \infty$"**.

Pour des petites valeurs de n, l'affirmation du Théorème "$0, 1, \infty$" peut s'interpréter géométriquement.

Interprétation géométrique dans le cas $n = 2$

Fixons $n = 2$, et considérons un système $m \times 2$:

$$(*) \left\{ \begin{array}{ccccc}
a_{11}x_1 & + & a_{12}x_2 & = & b_1 \\
a_{21}x_1 & + & a_{22}x_2 & = & b_2 \\
 & & & \vdots & \\
a_{m1}x_1 & + & a_{m2}x_2 & = & b_m
\end{array} \right.$$

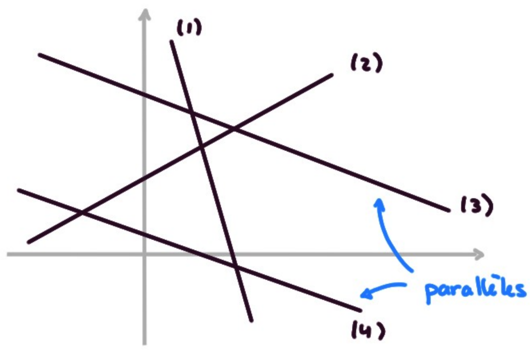
Ici, une paire (x_1, x_2) peut s'interpréter comme les coordonnées d'un point dans le plan, relativement à un repère orthonormé fixé. Aussi, on sait (voir cours de géométrie analytique) qu'une contrainte de la forme

$$a_{k1}x_1 \quad + \quad a_{k2}x_2 \quad = \quad b_k$$

signifie que le point de coordonnées (x_1, x_2) est sur une **droite**.

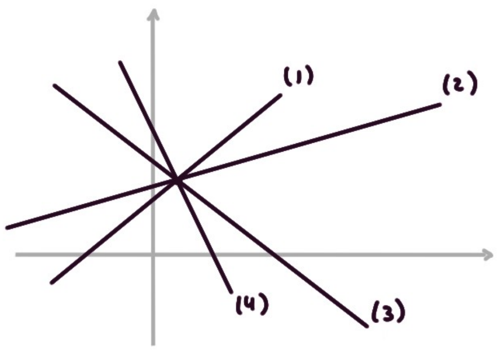
Donc une paire (x_1, x_2) sera solution de $(*)$, $(x_1, x_2) \in S_{(*)}$, si et seulement si le point (x_1, x_2) appartient en même temps à chacune des m droites spécifiées dans $(*)$. Or appartenir à m droites en même temps est une contrainte en général difficile à satisfaire, surtout si m est grand.

⋆ $S_{(*)}$ est en général vide, surtout si on parle de plus de deux droites, ou dès que deux de ces droites sont parallèles et distinctes :

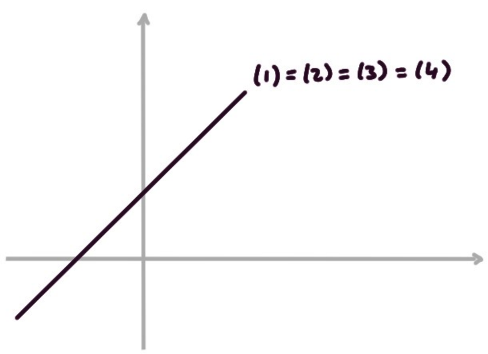


Sur ce dessin, on voit qu'il n'existe aucun point (x_1, x_2) qui appartient aux quatre droites à la fois.

⋆ $S_{(*)}$ contient seulement un élément si les droites s'intersectent en exactement un point :



⋆ $S_{(*)}$ contient une infinité d'éléments si les droites sont confondues :



On comprend que géométriquement, il est impossible de créer m droites dans le plan qui s'intersectent, par exemple, en exactement 4 points.

Interprétation géométrique dans le cas $n = 3$

Fixons $n = 3$, et considérons un système $m \times 3$:

$$(*) \begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \\ \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 = b_m \end{cases}$$

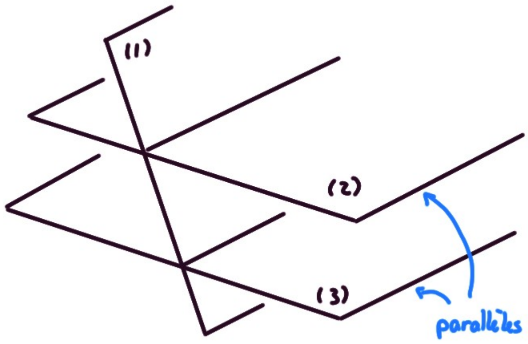
Ici, un triplet (x_1, x_2, x_3) peut s'interpréter comme les coordonnées d'un point dans l'espace, relativement à un repère orthonormé fixé. Aussi, on sait (voir cours de géométrie analytique) qu'une contrainte de la forme

$$a_{k1}x_1 + a_{k2}x_2 + a_{k3}x_3 = b_k$$

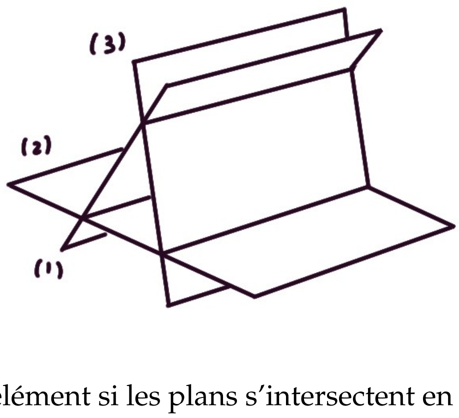
signifie que le point de coordonnées (x_1, x_2, x_3) est sur un **plan**.

Donc un triplet (x_1, x_2, x_3) sera solution de $(*)$ si et seulement si le point (x_1, x_2, x_3) appartient en même temps à chacun des m plans spécifiés. Or ici aussi il est géométriquement clair que $S_{(*)}$ ne peut contenir que 0, 1 ou une infinité de points.

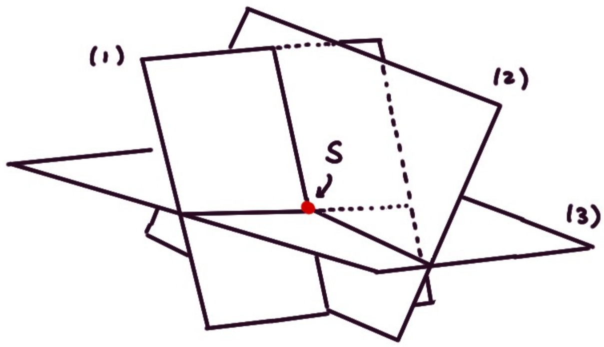
- ⋆ $S_{(*)}$ est vide dès que 2 de ces plans sont parallèles, distincts :



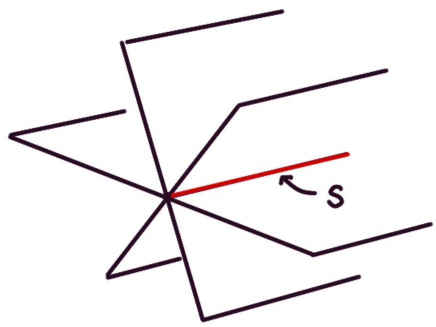
Ou alors ils peuvent aussi n'avoir aucun point en commun mais s'intersecter 2 à 2, sans que certains soient parallèles :



- ⋆ $S_{(*)}$ contient exactement un élément si les plans s'intersectent en seulement un point :



* $\star$ $S_{(*)}$ contient une infinité d'éléments si les plans sont confondus, ou s'intersectent selon une droite :



1.4 Transformer un système en un autre

Notre but pour la suite du chapitre est de présenter une méthode qui permet de savoir si un système est compatible ou incompatilbe et qui, lorsqu'il est compatible, permet en plus de décrire précisément l'ensemble de toutes ses solutions.

Cette méthode est utile non-seulement parce qu'elle mène à un algorithme (l'*algorithme de Gauss*) que l'on peut facilement implémenter sur une machine à l'aide d'un programme de quelques lignes, mais aussi parce qu'elle fournit un résultat théorique qui sera utilisé souvent dans la suite du cours.

Informel 1.9. Attention : Ce que nous présentons ci-dessous est une *méthode de calcul*. Elle sera utilisée souvent dans la suite du cours, mais ne constitue pas, en soi, "l'essence de l'algèbre linéaire" !

Un idéal : les systèmes triangulaires

Pour comprendre l'idée derrière la méthode générale qui va suivre, commençons par considérer un type de système dont la structure suggère elle-même une méthode de résolution.

Exemple 1.10. Considérons le système 3×3 suivant :

$$(*)\left\{\begin{array}{ccccccc} x_1 & - & x_2 & + & 2x_3 & = & 4 \\ & & x_2 & - & 3x_3 & = & -5 \\ & & & & 5x_3 & = & 10 \end{array}\right. ,$$

que l'on peut comprendre comme étant en fait

$$(*)\left\{\begin{array}{ccccccc} x_1 & - & x_2 & + & 2x_3 & = & 4 \\ \color{red}{0x_1} & + & x_2 & - & 3x_3 & = & -5 \\ \color{red}{0x_1} & + & \color{red}{0x_2} & + & 5x_3 & = & 10 \end{array}\right.$$

La présence des zéros en bas à gauche donne à ce système une structure triangulaire, qui suggère une résolution simple, “du bas vers le haut” :

1) Dans la troisième équation, on calcule $x_3 = \frac{10}{5} = 2$.
2) On injecte x_3 dans la deuxième équation, pour trouver
$$x_2 = -5 + 3x_3 = -5 + 6 = 1\,.$$
3) On injecte x_3 et x_2 dans la première équation, pour trouver
$$x_1 = 4 + x_2 - 2x_3 = 4 + 1 - 4 = 1\,.$$

Donc la solution est unique, donnée par $(x_1, x_2, x_3) = (1, 1, 2)$. $\diamond$

Le système de ce dernier exemple était très particulier, puisque les coefficients $a_{21} = a_{31} = a_{32} = 0$, lui conférant une structure simple à traiter. Mais même si le système était très grand, toujours avec la même structure triangulaire,

$$(*)\left\{\begin{array}{ccccccccccccc}
x_1 & + & 2x_2 & - & 4x_3 & - & 8x_4 & + & x_5 & + & 9x_6 & = & -3 \\
 & & x_2 & + & 4x_3 & - & 5x_4 & + & x_5 & - & x_6 & = & 6 \\
 & & & & x_3 & + & 6x_4 & + & x_5 & - & 6x_6 & = & -2 \\
 & & & & & & -x_4 & + & x_5 & + & x_6 & = & 3 \\
 & & & & & & & & 4x_5 & + & x_6 & = & 0 \\
 & & & & & & & & & & 7x_6 & = & 11
\end{array}\right.$$

on traiterait le problème de la même façon, “du bas vers le haut”...

Voyons un autre exemple dans lequel on profite de la présence de coefficients nuls dans la partie inférieure gauche, mais où le nombre de variables est supérieur au nombre d’équations :

Exemple 1.11. Considérons
$$\left\{\begin{array}{ccccccc}
2x_1 & + & x_2 & - & x_3 & = & 0 \\
{\color{red}0x_1} & + & x_2 & + & 2x_3 & = & 3
\end{array}\right.$$

Ce système 2×3 n’est pas “aussi triangulaire” que l’on voudrait. Pourtant, une opération naturelle est de déplacer les termes contenant x_3 du côté droit, pour récrire ce système comme

$$\left\{\begin{array}{ccccc}
2x_1 & + & x_2 & = & x_3 \\
{\color{red}0x_1} & + & x_2 & = & 3 - 2x_3
\end{array}\right.$$

Écrit sous cette forme, on voit qu’on peut choisir la valeur de x_3, ce qui fixe le côté droit, et qu’on se retrouve ensuite avec un système 2×2 triangulaire dont second membre dépend de x_3. Puisque la valeur de x_3 peut être choisie, on dit que c’est une variable **libre**, elle joue le rôle de paramètre ; on la notera plutôt t :

$$\left\{\begin{array}{ccccc}
2x_1 & + & x_2 & = & t \\
 & & x_2 & = & 3 - 2t
\end{array}\right.$$

Procédant “du bas vers le haut”, on obtient $x_2 = 3 - 2t$, puis
$$x_1 = \tfrac{1}{2}(t - x_2) = \tfrac{1}{2}(t - (3 - 2t)) = \tfrac{3}{2}(t - 1)\,.$$

On a donc, pour tout choix de t, une solution (x_1, x_2, x_3) donnée par
$$\begin{aligned}
x_1 &= \tfrac{3}{2}(t - 1) \\
x_2 &= 3 - 2t \\
x_3 &= t\,.
\end{aligned}$$

On peut donc écrire l'ensemble de toutes les solutions de la façon suivante :

$$S = \{(\tfrac{3}{2}(t-1), 3-2t, t) \,:\, t \in \mathbb{R}\}\,.$$

$\diamond$

Ainsi, la stratégie générale, pour résoudre un système quelconque, sera d'arriver à le *transformer en un système aussi triangulaire que possible*. Pour que ce nouveau système soit utile, il faudra être sûr que son ensemble de solutions soit *exactement le même* que le système de départ.

Opérations élémentaires

Considérons un système $m \times n$, noté $(*)$. Dans la suite, on utilisera le symbole L_i pour représenter la $i$ème ligne de $(*)$:

$$L_i : \qquad a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i\,.$$

Définissons plusieurs façons d'agir sur les lignes d'un système :

Définition 1.12. 1) L'opération consistant à échanger les lignes i et j est appelée **opération élémentaire de Type I**, et sera notée

$$\boxed{L_i \leftrightarrow L_j}$$

2) L'opération consistant à multiplier la $i$ème ligne par un scalaire non-nul $\lambda \in \mathbb{R}^*$ est appelée **opération élémentaire de Type II**, et sera notée

$$\boxed{L_i \leftarrow \lambda L_i}$$

3) L'opération consistant à rajouter à la $i$ème ligne un multiple (par un scalaire λ) de la $j$ème est appelée **opération élémentaire de Type III**, et sera notée

$$\boxed{L_i \leftarrow L_i + \lambda L_j}$$

Une opération élémentaire a pour effet de transformer un système linéaire $(*)$ en un autre système linéaire $(*)'$, de même dimension ; ces deux systèmes sont alors dits **équivalents selon les lignes** (ou **ligne-équivalents**).

Exemple 1.13. Considérons le système

$$(*)\left\{\begin{array}{ccccccccc} x_1 & + & 3x_2 & + & x_3 & + & 4x_4 & = & 3 \\ 2x_1 & + & 5x_2 & + & x_3 & - & x_4 & = & -1 \\ -x_1 & - & 2x_2 & - & 3x_3 & - & 9x_4 & = & 0 \end{array}\right.$$

⋆ En appliquant à $(*)$ l'opération de Type I donnée par $L_1 \leftrightarrow L_2$, on obtient

$$(*)'\left\{\begin{array}{ccccccccc} 2x_1 & + & 5x_2 & + & x_3 & - & x_4 & = & -1 \\ x_1 & + & 3x_2 & + & x_3 & + & 4x_4 & = & 3 \\ -x_1 & - & 2x_2 & - & 3x_3 & - & 9x_4 & = & 0 \end{array}\right.$$

⋆ En appliquant à $(*)$ l'opération de Type II donnée par $L_2 \leftarrow \frac{1}{2}L_2$, on obtient

$$(*)'\left\{\begin{array}{ccccccccc} x_1 & + & 3x_2 & + & x_3 & + & 4x_4 & = & 3 \\ x_1 & + & \frac{5}{2}x_2 & + & \frac{1}{2}x_3 & - & \frac{1}{2}x_4 & = & -\frac{1}{2} \\ -x_1 & - & 2x_2 & - & 3x_3 & - & 9x_4 & = & 0 \end{array}\right.$$

⋆ En appliquant à $(*)$ l'opération de Type III donnée par $L_3 \leftarrow L_3 + L_1$, on obtient

$$(*)' \left\{ \begin{array}{rcrcrcrcr} x_1 & + & 3x_2 & + & x_3 & + & 4x_4 & = & 3 \\ 2x_1 & + & 5x_2 & + & x_3 & - & x_4 & = & -1 \\ & & x_2 & - & 2x_3 & - & 5x_4 & = & 3 \end{array} \right.$$

◇

Un propriété remarquable des opérations définies ci-dessus est qu'elles sont toutes *inversibles*, dans le sens suivant : si $(*)'$ s'obtient en appliquant une opération élémentaire (de Type I, II ou III) à $(*)$, alors il existe une opération élémentaire **réciproque** qui permet, si on l'applique à $(*)'$, de revenir au système $(*)$ de départ.

1) La réciproque de $L_i \leftrightarrow L_j$ est $L_i \leftrightarrow L_j$ (évident).
2) La réciproque de $L_i \leftarrow \lambda L_i$ ($\lambda \in \mathbb{R}^*$) est $L_i \leftarrow \frac{1}{\lambda} L_i$ (en effet, on peut "défaire" la multiplication de la ligne i par λ en... divisant la ligne i par λ).
3) La réciproque de $L_i \leftarrow L_i + \lambda L_j$ est $L_i \leftarrow L_i - \lambda L_j$ (évident).

Comme conséquence de l'inversibilité :

Théorème 1.14. *Soient deux systèmes de mêmes dimensions, $(*)$ et $(*)'$. Si $(*)'$ est obtenu à partir de $(*)$ par une d'opération élémentaire, alors $(*)$ et $(*)'$ ont le même ensemble de solutions :*

$$S_{(*)} = S_{(*)'} .$$

L'usage du théorème ci-dessus se fera comme suit :

Corollaire 1. *Soient deux systèmes de mêmes dimensions, $(*)_1$ et $(*)_2$. Si $(*)_2$ peut être obtenu à partir de $(*)_1$ par une suite finie d'opérations élémentaires,*

$$(*)_1 \rightsquigarrow \cdots \rightsquigarrow (*)_2 ,$$

alors $(*)_1$ et $(*)_2$ ont le même ensemble de solutions :

$$S_{(*)_1} = S_{(*)_2} .$$

Preuve: Par le théorème, on sait que lors de chaque étape, l'ensemble de solutions est préservé. Puisqu'il y un nombre fini d'étapes, ceci implique $S_{(*)_1} = S_{(*)_2}$. □

Ce dernier corollaire suggère une méthode de résolution d'un système $m \times n$ quelconque : puisque les opérations élémentaires ne changent pas l'ensemble des solutions, on pourra appliquer au système de départ des opérations qui font peu à peu apparaître des zéros dans la partie inférieure gauche. Une fois le système "triangularisé", on procédera "du bas vers le haut", comme vu précédemment.

Plutôt que d'écrire explicitement un algorithme très général, considérons un premier exemple, qui contient déjà l'idée de la méthode, et montre dans quel ordre les opérations élémentaires sont choisies :

Exemple 1.15. Considérons le système

$$(*)_1 \left\{ \begin{array}{rcrcrcr} 3x_1 & + & 5x_2 & + & 4x_3 & = & 1 \\ 6x_1 & + & 12x_2 & + & 6x_3 & = & 0 \\ -2x_1 & - & 2x_2 & - & 7x_3 & = & 5 \end{array} \right.$$

D'abord, simplifions la deuxième ligne en la divisant par 6, ce qui revient à faire $L_2 \leftarrow \frac{1}{6}L_2$:

$$(*)_1' \left\{ \begin{array}{rcrcrcl} 3x_1 & + & 5x_2 & + & 4x_3 & = & 1 \\ x_1 & + & 2x_2 & + & x_3 & = & 0 \\ -2x_1 & - & 2x_2 & - & 7x_3 & = & 5 \end{array} \right.$$

Regardons maintenant la première colonne, et les coefficients présents devant x_1. Pour ce qui va suivre, on a avantage à faire $L_1 \leftrightarrow L_2$:

$$(*)_1'' \left\{ \begin{array}{rcrcrcl} x_1 & + & 2x_2 & + & x_3 & = & 0 \\ 3x_1 & + & 5x_2 & + & 4x_3 & = & 1 \\ -2x_1 & - & 2x_2 & - & 7x_3 & = & 5 \end{array} \right.$$

On va maintenant profiter du "x_1" en haut à gauche, appelé **pivot**, pour faire apparaître des zéros dans la partie inférieure de la première colonne. Par exemple, pour faire disparaître le "$3x_1$" de la deuxième ligne, on a besoin de lui soustraire 3 fois la première ligne. Donc en faisant $L_2 \leftarrow L_2 - 3L_1$, on obtient

$$(*)_1''' \left\{ \begin{array}{rcrcrcl} x_1 & + & 2x_2 & + & x_3 & = & 0 \\ 0x_1 & - & x_2 & + & x_3 & = & 1 \\ -2x_1 & - & 2x_2 & - & 7x_3 & = & 5 \end{array} \right. .$$

Ensuite, en faisant $L_3 \leftarrow L_3 + 2L_1$,

$$(*)_1'''' \left\{ \begin{array}{rcrcrcl} x_1 & + & 2x_2 & + & x_3 & = & 0 \\ 0x_1 & - & x_2 & + & x_3 & = & 1 \\ 0x_1 & + & 2x_2 & - & 5x_3 & = & 5 \end{array} \right. .$$

Ensuite, on passe à la deuxième colonne. C'est maintenant le "$-x_2$", dans L_2, qui joue le rôle de pivot et dicte le choix de l'opération suivante, qui fait apparaître encore un zéro au bas de la deuxième colonne : $L_3 \leftarrow L_3 + 2L_2$, qui donne

$$(*)_2 = (*)_1''''' \left\{ \begin{array}{rcrcrcl} x_1 & + & 2x_2 & + & x_3 & = & 0 \\ 0x_1 & - & x_2 & + & x_3 & = & 1 \\ 0x_1 & + & 0x_2 & - & 3x_3 & = & 7 \end{array} \right.$$

Remarquons que dans cette dernière étape, l'opération élémentaire n'a pas affecté les zéros de la première colonne !

En transformant $(*)_1$ en $(*)_2$, nous dirons plus loin que nous avons *échelonné* le système.

En procédant "du bas vers le haut" dans $(*)_2$, on obtient

$$S_{(*)_2} = \{(\tfrac{27}{3}, -\tfrac{10}{3}, -\tfrac{7}{3})\},$$

et le corollaire permet de conclure que $S_{(*)_1} = S_{(*)_2}$. ◇

1.5 Matrices et algorithme de Gauss

On comprend que la méthode illustrée sur l'exemple précédent, appelée **algorithme de Gauss**, s'applique à des systèmes de tailles quelconques. Il consiste à utiliser des opérations élémentaires qui font progressivement apparaître des zéros dans la partie inférieure gauche de la matrice.

Nous verrons d'autres exemples plus loin, mais arrêtons-nous un instant pour introduire une certaine simplification d'écriture.

Matrices associées à un système

Les opérations élémentaires que l'on effectue sur un système général du type

$$(*)\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}\right. ,$$

agissent sur les coefficients a_{ij}, et sur les termes du second membre, les b_i. Dans ces opérations, le nom que l'on donne aux variables (jusqu'à présent : $x_1, \dots, x_n$) n'a pas d'importance. Il est donc utile de simplifier la manipulation des systèmes en ne gardant que la structure numérique des coefficients et du second membre.

Définition 1.16. Soit $(*)$ le système ci-dessus.

1) La **matrice associée à** $(*)$ est le tableau de $m \times n$ nombres défini par

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

2) La **matrice augmentée associée à** $(*)$ est le tableau de $m \times (n+1)$ nombres défini par

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right)$$

(La ligne verticale, qui sépare les a_{kn} des b_k, est là pour rappeler que la dernière colonne doit être interprétée comme le second membre dans $(*)$.)

Exemple 1.17. La matrice augmentée du système

$$\left\{\begin{array}{ccccccccc} -x_1 & + & x_2 & & & + & x_4 & = & \alpha \\ x_1 & & & + & x_3 & + & 7x_4 & = & \beta \\ & & x_2 & - & x_3 & + & x_4 & = & \gamma \end{array}\right.$$

est donnée par

$$\left(\begin{array}{cccc|c} -1 & 1 & 0 & 1 & \alpha \\ 1 & 0 & 1 & 7 & \beta \\ 0 & 1 & -1 & 1 & \gamma \end{array}\right)$$

◇

Opérations élémentaires sur les matrices

Il est clair que les opérations élémentaires (de Type I,II,III) que l'on effectue sur un système se traduisent naturellement en des opérations sur les lignes de la matrice et de la matrice augmentée associée.

De manière générale, on peut effectuer des opérations élémentaires sur une matrice sans se référer au système dont elle est issue. On peut donc définir deux matrices comme étant **équivalentes selon les lignes** (ou **ligne-équivalentes**) si l'une peut s'obtenir à partir de l'autre par un nombre fini d'opérations élémentaires.

Matrices échelonnées

L'algorithme de Gauss mène, en général, à une matrice qui est ce que nous appelions "aussi triangulaire que possible" ; définissons ce terme précisément.

Définition 1.18. Pour une matrice $m \times n$ quelconque,

1) Le **coefficient principal de la i-ème ligne** est, s'il y en a un, le premier coefficient non-nul trouvé, en partant de la gauche.
2) Une matrice est **échelonnée** si les deux conditions suivantes sont satisfaites :
   - ⋆ Si elle possède des lignes nulles (c'est-à-dire ne contenant que des coefficients nuls), alors celles-ci sont toutes dans la partie inférieure de la matrice.
   - ⋆ Si elle possède des lignes non-nulles, alors le coefficient principal de chacune de ces lignes se trouve strictement à droite du coefficient principal de la ligne du dessus.

Exemple 1.19. La matrice

$$\left(\begin{array}{cccc|c} 3 & 1 & 2 & 0 & 7 \\ 0 & 0 & 4 & 3 & -1 \\ 0 & 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

est échelonnée. En effet, la seule ligne ne contenant que des zéros est tout en bas, et sur chacune des autres lignes, le coefficient principal est strictement à droite du coefficient principal de la ligne du dessus : $-\frac{1}{2}$ (coefficient principal de la troisième ligne) est à droite de 4 (coefficient principal de la deuxième ligne), qui est lui-même à droite de 3 (coefficient principal de la première ligne). ◇

Exemple 1.20. La matrice

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

n'est pas échelonnée, car le coefficient principal de la troisième ligne est juste sous le (et non *strictement à droite du*) coefficient principal de la deuxième ligne. ◇

Théorème 1.21. *Toute matrice peut être transformée, à l'aide d'un nombre fini de transformations élémentaires, en une matrice échelonnée. (En d'autres termes : toute matrice est équivalente selon les lignes à une matrice échelonnée.)*

Preuve: La méthode utilisée dans les exemples traités plus haut se généralise sans difficulté à n'importe quelle matrice. □

Exemple 1.22. Considérons la matrice augmentée du système vu plus haut :

$$\left(\begin{array}{ccc|c} 3 & 5 & 4 & 1 \\ 6 & 12 & 6 & 0 \\ -2 & -2 & -7 & 5 \end{array}\right).$$

En appliquant successivement les opérations élémentaires

$$L_2 \leftarrow \tfrac{1}{6}L_2\,, \quad L_1 \leftrightarrow L_2\,, \quad L_2 \leftarrow L_2 - 3L_1\,, L_3 \leftarrow L_3 + 2L_1\,, \quad L_3 \leftarrow L_3 + 2L_2\,,$$

on obtient comme on a vu la matrice

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -3 & 7 \end{array}\right),$$

qui est échelonnée. ◇

Remarque 1.23. La matrice échelonnée obtenue dépend du choix des opérations élémentaires faites en chemin ! Par exemple, en partant de

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 3 \end{pmatrix},$$

alors en faisant $L_2 \leftarrow L_2 - L_1$ on obtient une première échelonnée

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

mais en faisant $L_1 \leftrightarrow L_2$ suivie de $L_2 \leftarrow L_2 - L_1$ on obtient

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

qui est une autre forme échelonnée. Donc une matrice peut posséder plusieurs formes échelonnées. ◇

Concrètement, dans la transformation d'un système à l'aide d'opérations élémentaires, c'est une fois que la matrice obtenue est échelonnée que l'on peut déjà savoir si le système est compatible, si oui identifier les variables libres, et exprimer l'ensemble des solutions. Voyons un exemple assez complet :

Exemple 1.24. Supposons que l'on parte du système

$$(*)\left\{\begin{array}{ccccccccc} 3x_1 & + & x_2 & + & 2x_3 & & & = & 7 \\ 6x_1 & + & 2x_2 & + & 12x_3 & + & 7x_4 & = & 12 \\ 6x_1 & + & 2x_2 & + & 8x_3 & + & 4x_4 & = & 13 \\ 3x_1 & + & x_2 & + & 6x_3 & + & 4x_4 & = & 6 \end{array}\right.,$$

dont la matrice augmentée est

$$\left(\begin{array}{cccc|c} 3 & 1 & 2 & 0 & 7 \\ 6 & 2 & 12 & 7 & 12 \\ 6 & 2 & 8 & 4 & 13 \\ 3 & 1 & 6 & 4 & 6 \end{array}\right)$$

Faisons d'abord apparaître des zéros dans la première colonne, en appliquant successivement $L_2 \leftarrow L_2 - 2L_1$, $L_3 \leftarrow L_3 - 2L_1$, $L_4 \leftarrow L_4 - L_1$:

$$\left(\begin{array}{cccc|c} 3 & 1 & 2 & 0 & 7 \\ 0 & 0 & 8 & 7 & -2 \\ 0 & 0 & 4 & 4 & -1 \\ 0 & 0 & 4 & 4 & -1 \end{array}\right)$$

Comme les deux dernières lignes sont égales, on peut faire $L_4 \leftarrow L_4 - L_3$:

$$\left(\begin{array}{cccc|c} 3 & 1 & 2 & 0 & 7 \\ 0 & 0 & 8 & 7 & -2 \\ 0 & 0 & 4 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Ensuite, $L_2 \leftarrow L_2 - 2L_3$, suivie de $L_2 \leftrightarrow L_3$, nous donne une version échelonnée :

$$\left(\begin{array}{cccc|c} 3 & 1 & 2 & 0 & 7 \\ 0 & 0 & 4 & 4 & -1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Comment poursuivre, pour obtenir $S_{(*)}$? Pour y voir plus clair, revenons au système correspondant à cette dernière matrice augmentée :

$$(*)' \left\{ \begin{array}{ccccccccc} 3x_1 & + & x_2 & + & 2x_3 & + & 0x_4 & = & 7 \\ 0x_1 & + & 0x_2 & + & 4x_3 & + & 4x_4 & = & -1 \\ 0x_1 & + & 0x_2 & + & 0x_3 & - & x_4 & = & 0 \\ 0x_1 & + & 0x_2 & + & 0x_3 & + & 0x_4 & = & 0 \end{array} \right. ,$$

que l'on peut écrire plus simplement comme

$$(*)' \left\{ \begin{array}{ccccccccc} 3x_1 & + & x_2 & + & 2x_3 & & & = & 7 \\ & & & & 4x_3 & + & 4x_4 & = & -1 \\ & & & & & - & x_4 & = & 0 \end{array} \right. .$$

On a supprimé la dernière ligne "$0x_1 + 0x_2 + 0x_3 + 0x_4 = 0$". En effet, cette contrainte n'en est pas une, puisqu'elle est satisfaite par n'importe quel quadruplet (x_1, x_2, x_3, x_4).

On observe maintenant que la variable $x_2 = t$ est **libre**, puisqu'on peut la passer du côté droit pour obtenir un système triangulaire en (x_1, x_3, x_4), qui sont les variables **de base** :

$$(*)' \left\{ \begin{array}{ccccccc} 3x_1 & + & 2x_3 & & & = & 7-t \\ & & 4x_3 & + & 4x_4 & = & -1 \\ & & & - & x_4 & = & 0 \end{array} \right. .$$

Maintenant, en procédant "de bas en haut", on obtient

$$x_4 = 0 , \qquad x_3 = -\tfrac{1}{4} , \qquad x_1 = \frac{15-2t}{6} ,$$

et donc :

$$S_{(*)} = S_{(*)'} = \left\{ \left(\tfrac{15-2t}{6}, t, -\tfrac{1}{4}, 0\right) \, : \, t \in \mathbb{R} \right\}$$

Le système est donc compatible, et possède une infinité de solutions. ◇

Exemple 1.25. Considérons le système

$$(*) \left\{ \begin{array}{ccccccc} x_1 & + & x_2 & - & x_3 & = & 5 \\ 3x_1 & + & 4x_2 & & & = & 4 \\ 4x_1 & + & 5x_2 & - & x_3 & = & 7 \end{array} \right.$$

Après échelonnage, sa matrice devient

$$\left(\begin{array}{ccc|c} 1 & 1 & -1 & 5 \\ 0 & 1 & 3 & -11 \\ 0 & 0 & 0 & -3 \end{array} \right) .$$

Le système correspondant est

$$(*)' \left\{ \begin{array}{ccccccc} x_1 & + & x_2 & - & x_3 & = & 5 \\ 0x_1 & + & x_2 & + & 3x_3 & = & -11 \\ 0x_1 & + & 0x_2 & + & 0x_3 & = & -3 \end{array} \right. ,$$

qui est incompatible, puisque la dernière contrainte ne peut jamais être satisfaite, quel que soit (x_1, x_2, x_3). Donc $(*)$ est aussi incompatible. ◇

La réduite de Gauss

Définition 1.26. Une matrice est **échelonnée réduite** si elle est échelonnée et si, de plus,

- ⋆ tous ses coefficients principaux sont égaux à 1, et si
- ⋆ chaque coefficient principal est l'unique élément non-nul de sa colonne.

Les coefficients principaux d'une matrice échelonnée réduite sont appelés **pivots**.

Exemple 1.27. La matrice

$$\begin{pmatrix} 0 & 1 & 0 & 3 & 8 & 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 5 & 9 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 3 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

est échelonnée réduite (les pivots sont indiqués en bleu). ⋄

Voyons comment obtenir une échelonnée réduite d'une matrice. Une fois encore, la méthode présentée dans cet exemple particulier montre comment procéder en général :

Exemple 1.28. Considérons encore une fois la matrice du premier exemple de cette section :

$$\left(\begin{array}{ccc|c} 3 & 5 & 4 & 1 \\ 6 & 12 & 6 & 0 \\ -2 & -2 & -7 & 5 \end{array}\right).$$

Pour obtenir sa réduite, on commence par l'échelonner, comme vu plus haut :

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -3 & 7 \end{array}\right).$$

On poursuit en rendant tous les coefficients principaux égaux à 1, à l'aide de $L_2 \leftarrow (-1)L_2$ et $L_3 \leftarrow \frac{1}{-3}L_3$:

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & -7/3 \end{array}\right),$$

Remarquons que maintenant, lorsqu'on additionne un multiple d'une ligne L_i à une ligne L_j située *au-dessus*, la présence de zéros fait *qu'on ne modifie pas les coefficients de L_j situés à droite du coefficient principal de L_i*.

On procède donc "du bas vers le haut" pour faire apparaître des zéros au-dessus des "1". D'abord, on utilise $L_2 \leftarrow L_2 + L_3$ pour faire partir le "-1" de la troisième colonne, ce qui donne

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & 1 & 0 & -10/3 \\ 0 & 0 & 1 & -7/3 \end{array}\right).$$

Ensuite, on peut faire disparaître le "2" de la deuxième colonne en faisant $L_1 \leftarrow L_1 - 2L_2$,

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 20/3 \\ 0 & 1 & 0 & -10/3 \\ 0 & 0 & 1 & -7/3 \end{array}\right),$$

puis le "1" de la troisième colonne en faisant $L_1 \leftarrow L_1 - L_3$:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 27/3 \\ 0 & 1 & 0 & -10/3 \\ 0 & 0 & 1 & -7/3 \end{array}\right),$$

qui est *la* forme échelonnée réduite de la matrice de départ. On remarque que cette matrice correspond au système

$$\left\{\begin{array}{ccccc} x_1 & & & = & 27/3 \\ & x_2 & & = & -10/3 \\ & & x_3 & = & -7/3 \end{array}\right.,$$

pour lequel on a la solution "sous les yeux" : $(x_1, x_2, x_3) = (\frac{27}{3}, \frac{-10}{3}, \frac{-7}{3})$. ◇

On l'a mentionné plus haut, une matrice peut être échelonnée d'une infinité de façons différentes. Pour la réduite, c'est différent :

Théorème 1.29. *Toute matrice A peut être transformée, à l'aide d'un nombre fini de transformations élémentaires, en une matrice échelonnée réduite. De plus, cette échelonnée réduite est unique et ne dépend pas des opérations élémentaires avec lesquelles elle a été obtenue ; on l'appelle* ***la réduite de Gauss de*** *A.*

Preuve: (Omise pour l'instant.) □

Dans la résolution d'un système, on n'a pas besoin d'aller jusqu'à la réduite pour obtenir la solution ; n'importe quelle échelonnée suffit. Mais puisque la réduite est *unique*, comme l'affirme le théorème ci-dessus, elle fournit des informations importantes sur la matrice de départ, que nous exploiterons dans les chapitres suivants, en particulier dans l'étude des applications linéaires. D'un point de vue pratique, elle nous permet toujours d'identifier les variables de base et les variables libres, et donc de résoudre complètement le système.

Chapitre 2

Vecteurs de $\mathbb{R}^n$

2.1 Définitions

Dans ce chapitre, nous laissons les systèmes de côté un instant, pour introduire le langage de base nécessaire au développement de l'algèbre linéaire dans les espaces $\mathbb{R}^n$, $n \geqslant 1$. Pour commencer, nous introduirons la notion de *vecteur*, centrale en algèbre linéaire, et particulièrement utile pour décrire les systèmes.

Vecteurs

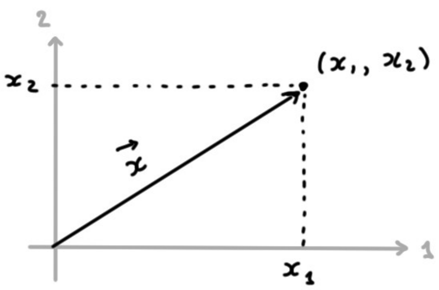
On l'a déjà mentionné plus haut : toute liste de nombres réels $(x_1, x_2, \dots, x_n)$ peut être identifiée avec un point de l'espace $\mathbb{R}^n$. Or les points de $\mathbb{R}^n$ sont plus facilement manipulables lorsqu'on les interprète comme des objets appelés *vecteurs*.

On identifiera donc $(x_1, \dots, x_n)$ avec le **vecteur** (dit aussi **vecteur-colonne**), noté

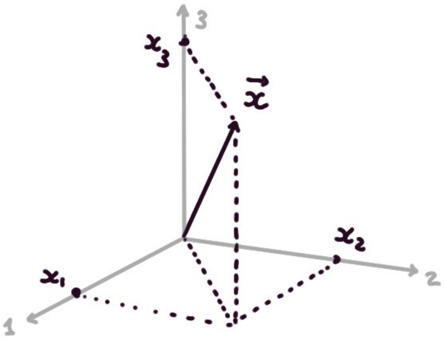
$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} .$$

Informel 2.1. En analyse, $\mathbb{R}^n$ est considéré comme un ensemble de *points*. En algèbre linéaire, $\mathbb{R}^n$ est considéré comme un ensemble de *vecteurs*.

Il est clair que la liste $(x_1, \dots, x_n)$ et le vecteur $\mathbf{x}$ contiennent la même information, mais ici il faut interpréter $\mathbf{x}$ comme le *déplacement* depuis l'origine jusqu'au point $(x_1, x_2, \dots, x_n)$. Par exemple, dans le plan $\mathbb{R}^2$,



ou dans l'espace $\mathbb{R}^3$:



L'avantage d'identifier des points avec des vecteurs est que le langage vectoriel permet d'introduire des opérations qui rendent les vecteurs propices au *calcul*.

Addition et multiplication par des scalaires

On munit l'ensemble des vecteurs de $\mathbb{R}^n$ de deux opérations :

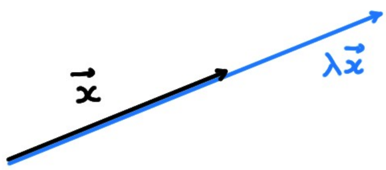
1) (Multiplication par un scalaire) La **multiplication d'un vecteur**

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

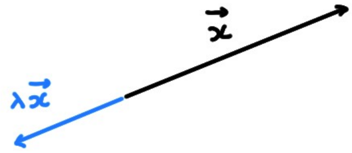
par un scalaire $\lambda \in \mathbb{R}$ est le vecteur $\lambda\mathbf{x}$ défini par

$$\lambda\mathbf{x} := \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

Du point de vue géométrique, la multiplication par un scalaire $\lambda > 0$ correspond à étirer (si $\lambda \geqslant 1$) ou comprimer (si $0 < \lambda < 1$) :



Lorsque $\lambda < 0$, cette transformation est en plus accompagnée d'un changement de sens :



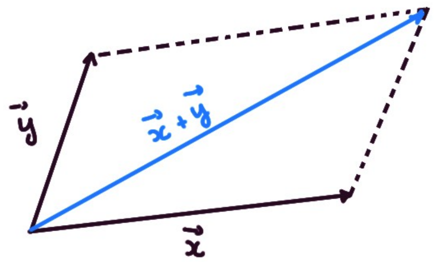
2) (Addition vectorielle) Si deux vecteurs

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \text{et} \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

sont donnés, leur **somme** est définie par

$$\mathbf{x}+\mathbf{y}:=\begin{pmatrix}x_1+y_1\\x_2+y_2\\\vdots\\x_n+y_n\end{pmatrix}.$$

Remarque 2.2. En petites dimensions, $n=2$ et $n=3$, l'addition vectorielle peut s'interpréter géométriquement comme la **règle du parallélogramme** :



On comprend ici que c'est l'interprétation d'un vecteur comme à un *déplacement* qui rend cette opération d'addition naturelle. ◇

Le **vecteur nul** est celui dont toutes les composantes sont égales à zéro. On le notera

$$\mathbf{0}=\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}.$$

Par définition, on a $0\mathbf{x}=\mathbf{0}$, et

$$\mathbf{x}+\mathbf{0}=\mathbf{x}.$$

Pour tout vecteur $\mathbf{x}$, on appelle **opposé de** $\mathbf{x}$ le vecteur $-\mathbf{x}:=(-1)\mathbf{x}$. Par définition,

$$\mathbf{x}+(-\mathbf{x})=(-\mathbf{x})+\mathbf{x}=\mathbf{0}.$$

Les opérations d'addition et de multiplication par un scalaire permettent de manipuler les vecteurs à l'aide de calculs. Ces calculs obéissent aux règles standard de l'arithmétique :

Proposition 1. *Pour tous vecteurs $\mathbf{x}, \mathbf{y}, \mathbf{z}$ de $\mathbb{R}^n$, et pour tous scalaires $\lambda, \mu \in \mathbb{R}$,*

1) $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$ *(commutativité)*
2) $\mathbf{x}+(\mathbf{y}+\mathbf{z})=(\mathbf{x}+\mathbf{y})+\mathbf{z}$ *(associativité)*
3) $\lambda(\mathbf{x}+\mathbf{y})=\lambda\mathbf{x}+\lambda\mathbf{y}$ *(distributivité)*
4) $(\lambda+\mu)\mathbf{x}=\lambda\mathbf{x}+\mu\mathbf{x}$ *(distributivité)*
5) $(\lambda\mu)\mathbf{x}=\lambda(\mu\mathbf{x})=\mu(\lambda\mathbf{x})$

Preuve: Ces propriétés ne font que refléter une propriété élémentaire semblable, valide dans le corps des nombres réels. Ci-dessous, on indiquera par le symbole $\overset{!}{=}$ une identité obtenue en utilisant une propriété de base dans les réels, pour chacune des composantes :

1)

$$\mathbf{x}+\mathbf{y}=\begin{pmatrix}x_1+y_1\\x_2+y_2\\\vdots\\x_n+y_n\end{pmatrix}\overset{!}{=}\begin{pmatrix}y_1+x_1\\y_2+x_2\\\vdots\\y_n+x_n\end{pmatrix}=\mathbf{y}+\mathbf{x}$$

2)

$$
\begin{aligned}
\mathbf{x}+(\mathbf{y}+\mathbf{z}) &= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1+z_1 \\ y_2+z_2 \\ \vdots \\ y_n+z_n \end{pmatrix} \\
&= \begin{pmatrix} x_1+(y_1+z_1) \\ x_2+(y_2+z_2) \\ \vdots \\ x_n+(y_n+z_n) \end{pmatrix} \\
&\overset{!}{=} \begin{pmatrix} (x_1+y_1)+z_1 \\ (x_2+y_2)+z_2 \\ \vdots \\ (x_n+y_n)+z_n \end{pmatrix} \\
&= \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \\
&= (\mathbf{x}+\mathbf{y})+\mathbf{z}
\end{aligned}
$$

3)

$$
\begin{aligned}
\lambda(\mathbf{x}+\mathbf{y}) &= \lambda \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{pmatrix} \\
&= \begin{pmatrix} \lambda(x_1+y_1) \\ \lambda(x_2+y_2) \\ \vdots \\ \lambda(x_n+y_n) \end{pmatrix} \\
&\overset{!}{=} \begin{pmatrix} \lambda x_1+\lambda y_1 \\ \lambda x_2+\lambda y_2 \\ \vdots \\ \lambda x_n+\lambda y_n \end{pmatrix} \\
&= \lambda\mathbf{x}+\lambda\mathbf{y}
\end{aligned}
$$

4)

$$
\begin{aligned}
(\lambda+\mu)\mathbf{x} &= \begin{pmatrix} (\lambda+\mu)x_1 \\ (\lambda+\mu)x_2 \\ \vdots \\ (\lambda+\mu)x_n \end{pmatrix} \\
&\overset{!}{=} \begin{pmatrix} \lambda x_1+\mu x_1 \\ \lambda x_2+\mu x_2 \\ \vdots \\ \lambda x_n+\mu x_n \end{pmatrix} = \lambda\mathbf{x}+\mu\mathbf{x}
\end{aligned}
$$

5)

$$(\lambda\mu)\mathbf{x} = \begin{pmatrix} (\lambda\mu)x_1 \\ (\lambda\mu)x_2 \\ \vdots \\ (\lambda\mu)x_n \end{pmatrix}$$

$$\stackrel{!}{=} \begin{pmatrix} \lambda(\mu x_1) \\ \lambda(\mu x_2) \\ \vdots \\ \lambda(\mu x_n) \end{pmatrix} = \lambda(\mu\mathbf{x})$$

$$\stackrel{!}{=} \begin{pmatrix} \mu(\lambda x_1) \\ \mu(\lambda x_2) \\ \vdots \\ \mu(\lambda x_n) \end{pmatrix} = \mu(\lambda\mathbf{x})$$

□

Avec ces propriétés, on peut résoudre des équations vectorielles, dont l'inconnue est un vecteur $\mathbf{x}$, de la même façon qu'on résout des équations élémentaires où l'inconnue est un réel x.

Exemple 2.3. Considérons deux vecteurs $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ fixés, et étudions l'équation vectorielle

$$2\mathbf{a} - 3\mathbf{x} = 5\mathbf{x} + 7\mathbf{b}\,.$$

Utilisons les propriétés démontrées ci-dessus pour *isoler* $\mathbf{x}$, comme on le fait quand on résout une équation en arithmétique élémentaire.

Rajoutons $+3\mathbf{x}$ des deux côtés. Du côté gauche, détaillons l'utilisation des propriétés :

$$2\mathbf{a} - 3\mathbf{x} + 3\mathbf{x} = 2\mathbf{a} + (-3+3)\mathbf{x} = 2\mathbf{a} + 0\mathbf{x} = 2\mathbf{a} + \mathbf{0} = 2\mathbf{a}\,.$$

En procédant de même du côté droit, on obtient

$$2\mathbf{a} = 8\mathbf{x} + 7\mathbf{b}\,.$$

En soustrayant $7\mathbf{b}$ des deux côtés,

$$8\mathbf{x} = 2\mathbf{a} - 7\mathbf{b}\,,$$

puis en multipliant par $\frac{1}{8}$,

$$\mathbf{x} = \tfrac{1}{4}\mathbf{a} - \tfrac{7}{8}\mathbf{b}\,.$$

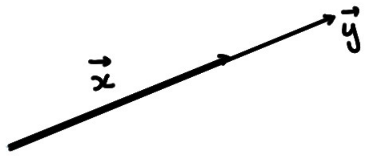
◇

2.2 Colinéarité

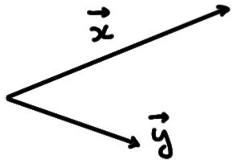
La *colinéarité* est une généralisation du *parallélisme*.

Définition 2.4. Deux vecteurs $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ sont **colinéaires** si il existe un scalaire $\lambda \in \mathbb{R}$ tel que $\mathbf{y} = \lambda\mathbf{x}$ ou $\mathbf{x} = \lambda\mathbf{y}$.

Deux vecteurs sont colinéaires si ils sont supportés par la même droite,



et non-colinéaires (on dira bientôt *indépendants*) si ils pointent dans des directions différentes :



Exemple 2.5. Parmi les trois vecteurs de $\mathbb{R}^4$ suivants,

$$\mathbf{x} = \begin{pmatrix} 3 \\ 4 \\ 0 \\ -1 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} -3 \\ 1/2 \\ 0 \\ 1 \end{pmatrix}, \qquad \mathbf{z} = \begin{pmatrix} 6 \\ -1 \\ 0 \\ -2 \end{pmatrix},$$

seuls $\mathbf{y}$ et $\mathbf{z}$ sont colinéaires, puisque $\mathbf{z} = -2\mathbf{y}$. ◇

Remarque 2.6. Le vecteur nul, $\mathbf{0}$, est colinéaire à n'importe quel autre vecteur. En effet, quel que soit $\mathbf{y} \in \mathbb{R}^n$, on peut toujours écrire $\mathbf{0} = \lambda\mathbf{y}$, où $\lambda = 0$. ◇

2.3 Combinaisons linéaires et parties engendrés

En algèbre linéaire, une façon standard et non-triviale d'obtenir de nouveaux vecteurs à partir d'une famille donnée est de former des *combinaisons linéaires*.

Définition 2.7. Soient $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ des vecteurs de $\mathbb{R}^n$. Une somme du type

$$\lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k\,,$$

où les $\lambda_1, \lambda_2, \dots, \lambda_k$ sont des scalaires fixés, est appelée **combinaison linéaire** des vecteurs $\mathbf{v}_1, \dots, \mathbf{v}_k$. Les scalaires λ_j sont les **coefficients** de la combinaison linéaire.

Exemple 2.8. Dans $\mathbb{R}^2$, considérons $\mathbf{v}_1 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$. En prenant $\lambda_1 = -2$, $\lambda_2 = 3$,

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 = -2\begin{pmatrix} 3 \\ 5 \end{pmatrix} + 3\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} -9 \\ -4 \end{pmatrix}.$$

Choisissons maintenant un troisième vecteur : $\mathbf{w} = \begin{pmatrix} 5 \\ -2 \end{pmatrix}$, et posons la question : est-il possible d'écrire $\mathbf{w}$ comme une combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$? Il s'agit donc de voir si il existe des scalaires λ_1, λ_2 tels que

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 = \mathbf{w}\,.$$

Lorsqu'on exprime cette relation en composantes,

$$\begin{pmatrix} 3\lambda_1 - \lambda_2 \\ 5\lambda_1 + 2\lambda_2 \end{pmatrix} = \begin{pmatrix} 5 \\ -2 \end{pmatrix}$$

Puisque deux vecteurs sont égaux si et seulement si leurs composantes sont égales deux-à-deux, on en déduit que λ_1 et λ_2 doivent être solution du système

$$\begin{cases} 3\lambda_1 - \lambda_2 = 5 \\ 5\lambda_1 + 2\lambda_2 = -2 \end{cases}$$

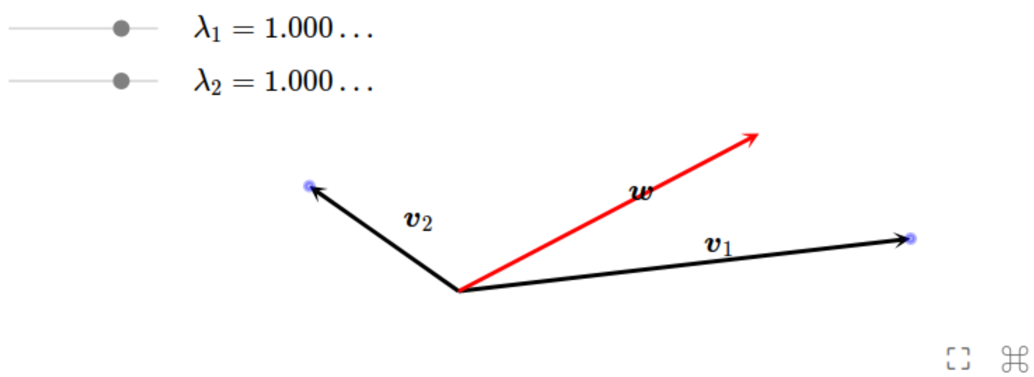
La solution de ce système est unique, donnée par $\lambda_1 = \frac{8}{11}$, $\lambda_2 = \frac{-31}{11}$. On en déduit que $\mathbf{w}$ est bien combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$:

$$\mathbf{w} = \tfrac{8}{11}\mathbf{v}_1 - \tfrac{31}{11}\mathbf{v}_2 \,.$$

◇

Plus généralement, fixons deux vecteurs $\mathbf{v}_1$ et $\mathbf{v}_2$ dans le plan, et considérons toutes les combinaisons linéaires de la forme

$$\mathbf{w} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 \,, \qquad \lambda_1, \lambda_2 \in \mathbb{R} \,.$$



On remarque que

- ⋆ Si $\mathbf{v}_1$ et $\mathbf{v}_2$ ne sont pas colinéaires, alors *toutes les combinaisons linéaires possibles de* $\mathbf{v}_1$ et $\mathbf{v}_2$ *remplissent le plan*, dans le sens suivant : n'importe quel vecteur $\mathbf{w}$ peut s'écrire comme combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$.
- ⋆ Si $\mathbf{v}_1$ et $\mathbf{v}_2$ sont colinéaires, alors seulement certains vecteurs $\mathbf{w}$ du plan peuvent s'écrire comme combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$ (essentiellement ceux qui sont sur la droite portée par $\mathbf{v}_1$ et $\mathbf{v}_2$).

Exemple 2.9. Dans $\mathbb{R}^3$, considérons

$$\mathbf{v}_1 = \begin{pmatrix} -3 \\ 2 \\ 0 \end{pmatrix} , \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} , \quad \mathbf{w} = \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix}$$

Est-ce que $\mathbf{w}$ est combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$? Pour le savoir, cherchons λ_1, λ_2 tels que

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 = \mathbf{w} \,,$$

qui mène au système

$$\begin{cases} -3\lambda_1 + \lambda_2 = 4 \\ 2\lambda_1 + \lambda_2 = 0 \\ - \lambda_2 = 3 \end{cases}$$

Ce système est incompatible, donc $\mathbf{w}$ ne peut *pas* s'écrire comme combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$. ◇

Informel 2.10. Dans ce dernier exemple, on a résolu un problème de combinaison linéaire en l'exprimant sous la forme d'un système linéaire. Dans la section suivante nous ferons l'inverse, en montrant qu'un système linéaire peut se traduire en un problème de combinaison linéaire.

Parties engendrées

Définition 2.11. Soient $\mathbf{v}_1, \ldots, \mathbf{v}_p$ des vecteurs de $\mathbb{R}^n$ donnés. La **partie de $\mathbb{R}^n$ engendrée par la famille** $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, notée

$$\mathrm{Vect}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}\,,$$

est définie comme l'ensemble des vecteurs de $\mathbb{R}^n$ qui peuvent s'écrire comme combinaison linéaire des vecteurs $\mathbf{v}_1, \ldots, \mathbf{v}_p$:

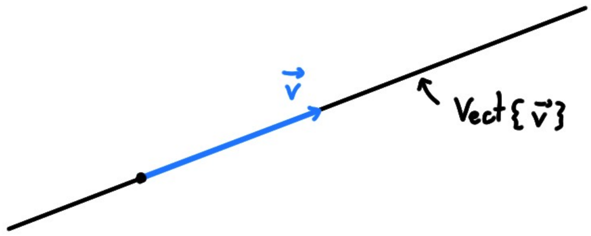
$$\mathbf{w} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_p \mathbf{v}_p\,.$$

Informel 2.12. La partie engendrée par une famille de vecteurs, c'est l'ensemble de toutes les combinaisons linéaires possibles de ces vecteurs.

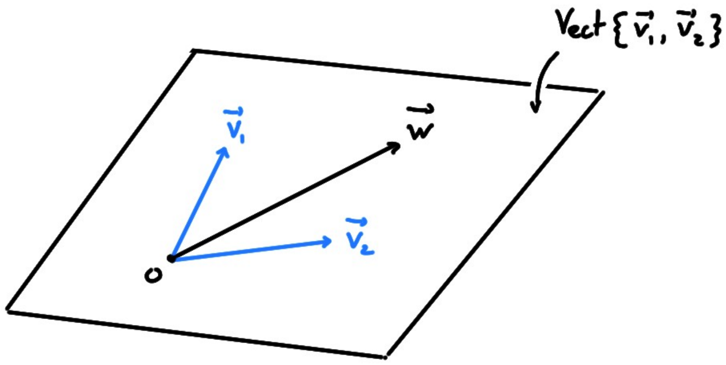
Remarque 2.13. En anglais, $\mathrm{Vect}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ se note $\mathrm{Span}\,\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$. ◇

Pour les familles contenant un ou deux vecteurs :

⋆ Lorsqu'on considère une famille $\{\mathbf{v}\}$ contenant un seul vecteur non-nul $\mathbf{v}$, $\mathrm{Vect}\{\mathbf{v}\}$ est constitué de tous les vecteurs colinéaires à $\mathbf{v}$, c'est-à-dire de la forme $\mathbf{w} = \lambda\mathbf{v}$. Il est donc naturel d'interpréter $\mathrm{Vect}\{\mathbf{v}\}$ comme **la droite de $\mathbb{R}^n$ engendrée par $\mathbf{v}$**, passant par l'origine :



⋆ Lorsqu'on considère une famille $\{\mathbf{v}_1, \mathbf{v}_2\}$ contenant deux vecteurs non-colinéaires, on interprète $\mathrm{Vect}\{\mathbf{v}_1, \mathbf{v}_2\}$ comme **le plan de $\mathbb{R}^n$ engendrée par** $\mathbf{v}_1$ et $\mathbf{v}_2$, passant par l'origine :

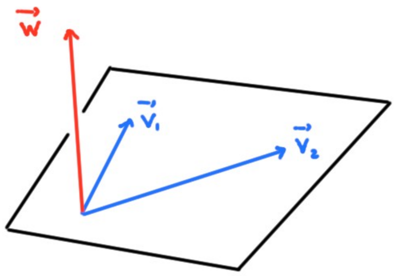


Même si cette terminologie ("droite", "plan") est empruntée à la géométrie du plan $(n = 2)$ et de l'espace ($n = 3$), nous l'utiliserons aussi dans les dimensions supérieures ($n > 3$).

Exemple 2.14. Plus haut, nous avions défini

$$\mathbf{v}_1 = \begin{pmatrix} -3 \\ 2 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix},$$

et montré que $\mathbf{w}$ n'est pas combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$. Nous pouvons maintenant interpréter ceci en disant que $\mathbf{w}$ n'est pas dans le plan engendré par $\mathbf{v}_1$ et $\mathbf{v}_2$:



◇

La base canonique de $\mathbb{R}^n$

Définissons, pour tout $k = 1, \ldots, n$, le vecteur $\mathbf{e}_k \in \mathbb{R}^n$ comme étant le vecteur dont toutes les composantes sont nulles, sauf la k-ème, qui vaut 1.

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \cdots \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Cette famille de vecteur peut être utilisée pour décomposer n'importe quel vecteur, comme suit :

$$\begin{aligned} \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} &= \begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{pmatrix} \\ &= x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \\ &= x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \cdots + x_n \mathbf{e}_n . \end{aligned}$$

En d'autres termes,

$$\mathbb{R}^n = \text{Vect}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\} .$$

Plus tard, on appellera $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ la **base canonique de $\mathbb{R}^n$**.

2.4 In-dépendance linéaire

La notion *d'indépendance linéaire* (et celle qui lui est associée, la *dépendance linéaire*) est une des plus importantes de l'algèbre linéaire.

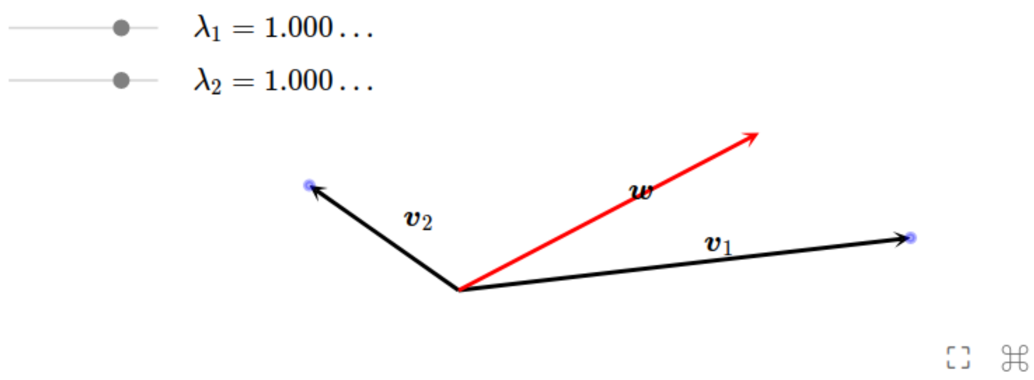
Informel 2.15. En effet, il sera important de comprendre comment des vecteurs peuvent être utilisés pour "remplir l'espace", à l'aide de combinaisons linéaires. Pour ce faire, il faudra pouvoir décrire dans quelle mesure ces vecteurs pointent dans des *dimensions différentes* de $\mathbb{R}^n$. Et pour avoir en main une notion qui permette de travailler (et faire des calculs !), il faut introduire une définition abstraite, qui s'utilise en toute dimension. Cette notion, c'est l'*indépendance linéaire*.

Motivation : une caractérisation de la non-colinéarité

En guise de motivation, considérons deux vecteurs dans le plan. Clairement, si ces vecteurs ne sont pas colinéaires, c'est qu'ils pointent dans des directions différentes. Or on peut reformuler ce que signifie être *non-colinéaire* un peu différemment.

Fixons deux vecteurs du plan, $\mathbf{v}_1$ et $\mathbf{v}_2$, et étudions toutes les combinaisons linéaires de la forme

$$\mathbf{w} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 \,, \qquad \lambda_1, \lambda_2 \in \mathbb{R} \,.$$



Bien-sûr, quels que soient $\mathbf{v}_1$ et $\mathbf{v}_2$, on a $\mathbf{w} = \mathbf{0}$ dès que $\lambda_1 = \lambda_2 = 0$, puisque

$$0\mathbf{v}_1 + 0\mathbf{v}_2 = \mathbf{0} \,.$$

Mais posons-nous la question de savoir s'il existe d'autres paires (λ_1, λ_2) telles que $\mathbf{w} = \mathbf{0}$. Par un simple calcul, ou en utilisant l'animation ci-dessus , on se convainc facilement des deux faits suivants :

1) Si $\mathbf{v}_1$ et $\mathbf{v}_2$ ne sont pas colinéaires, alors l'unique façon d'avoir $\mathbf{w} = \mathbf{0}$ est de prendre $\lambda_1 = \lambda_2 = 0$.
2) Si $\mathbf{v}_1$ et $\mathbf{v}_2$ sont colinéaires, alors il existe une infinité de choix possibles pour λ_1 et λ_2 qui garantissent $\mathbf{w} = \mathbf{0}$.
3) La même chose fonctionne en toute dimension.

On conclut de cette simple discussion que la non-colinéarité, pour deux vecteurs, peut s'exprimer de façon plus abstraite, par cette condition à propos de leurs combinaisons linéaires nulles :

Lemme 1. *Deux vecteurs non-nuls $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ sont non-colinéaires si et seulement si l'unique combinaison linéaire nulle,*

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 = \mathbf{0} \,,$$

est celle pour laquelle $\lambda_1 = \lambda_2 = 0$.

L'avantage de cette caractérisation de la non-colinéarité de deux vecteurs, proposée dans le lemme précédent, est qu'elle se généralise naturellement à des familles contenant plus que deux vecteurs (de $\mathbb{R}^n$). Voyons comment, dans la section suivante.

Définition

Définition 2.16. Soient $\mathbf{v}_1, \dots, \mathbf{v}_p$ des vecteurs de $\mathbb{R}^n$ donnés. La famille $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est dite

- ⋆ **linéairement indépendante** (ou **libre**) si l'unique combinaison linéaire nulle,

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_p \mathbf{v}_p = \mathbf{0}$$

  est celle pour laquelle $\lambda_1 = \lambda_2 = \cdots = \lambda_p = 0$.

- ⋆ **dépendante** (ou **liée**) si il existe des coefficients $\lambda_1, \lambda_2, \dots, \lambda_p$, dont au moins un n'est pas nul, tels que

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_p \mathbf{v}_p = \mathbf{0}\,.$$

Remarque 2.17. Dès qu'un des vecteurs de la famille $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est nul, cette famille est dépendante. En effet, supposons que $\mathbf{v}_k = \mathbf{0}$. On peut alors écrire l'identité suivante, toujours vraie,

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_{k-1} + 1 \underbrace{\mathbf{v}_k}_{=\mathbf{0}} + 0\mathbf{v}_{k+1} + \cdots + 0\mathbf{v}_p = \mathbf{0}\,,$$

qui implique bien que $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ est dépendante. ◇

Exemple 2.18. Considérons la famille de vecteurs de $\mathbb{R}^4$ contenant les trois vecteurs

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \\ -3 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 0 \\ -2 \\ 1 \\ 5 \end{pmatrix}, \qquad \mathbf{v}_3 = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \end{pmatrix}.$$

Cette famille est-elle libre ou liée ? Pour répondre, considérons la relation linéaire

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \lambda_3 \mathbf{v}_3 = \mathbf{0}\,.$$

Lorsqu'on écrit explicitement cette relation en composantes, on obtient le système 4×3 suivant :

$$\left\{ \begin{array}{rcrcrcl} \lambda_1 & & & + & 3\lambda_3 & = & 0 \\ & - & 2\lambda_2 & + & 2\lambda_3 & = & 0 \\ 2\lambda_1 & + & \lambda_2 & + & \lambda_3 & = & 0 \\ -3\lambda_1 & + & 5\lambda_2 & & & = & 0 \end{array} \right.$$

La matrice augmentée de ce système devient, après échelonnage,

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

La solution du système correspondant est $\lambda_1 = \lambda_2 = \lambda_3 = 0$. On conclut que $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ est une famille libre. ◇

Exemple 2.19. Montrons que la famille formée des vecteurs de la base canonique de $\mathbb{R}^n$, $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, est libre. Pour ce faire, considérons la relation linéaire

$$\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 + \cdots + \lambda_n \mathbf{e}_n = \mathbf{0}\,.$$

Lorsqu'on l'exprime en composantes, cette dernière devient

$$\left\{ \begin{array}{ccc} \lambda_1 & = & 0 \\ \lambda_2 & = & 0 \\ & \vdots & \\ \lambda_n & = & 0 \end{array} \right.,$$

qui montre bien que la famille est libre. ◇

Théorème 2.20. *Une famille* $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ *est liée si et seulement si un de ses vecteurs peut s'écrire comme combinaison linéaire des autres, plus précisément : si il existe* $k \in \{1, \dots, p\}$ *tel que* $\mathbf{v}_k$ *peut s'écrire comme combinaison linéaire des* $\mathbf{v}_j$, $j \neq k$.

Preuve: Si la famille est liée, alors il existe des nombres $\lambda_1, \dots, \lambda_p$, non tous nuls, tels que

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k + \cdots + \lambda_p \mathbf{v}_p = \mathbf{0}\,.$$

Si on supposons que le coefficient $\lambda_k \neq 0$, on peut isoler $\mathbf{v}_k$ dans cette dernière :

$$\mathbf{v}_k = \sum_{\substack{j=1\\ j\neq k}}^{p} \frac{(-\lambda_j)}{\lambda_k} \mathbf{v}_j\,.$$

On a donc bien exprimé $\mathbf{v}_k$ comme combinaison linéaire des autres. Inversément, si un $\mathbf{v}_k$ peut s'écrire comme combinaison linéaire des autres,

$$\mathbf{v}_k = \sum_{\substack{j=1\\ j\neq k}}^{p} \alpha_j \mathbf{v}_j\,,$$

et on peut récrire cette dernière comme

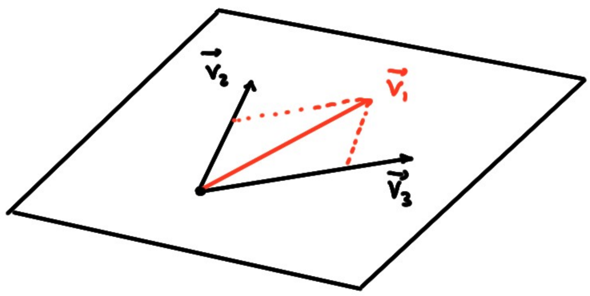
$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_{k-1} \mathbf{v}_{k-1} + (-1)\mathbf{v}_k + \alpha_{k+1} \mathbf{v}_{k+1} + \cdots + \alpha_p \mathbf{v}_p = \mathbf{0}\,,$$

qui montre bien que la famille est liée. □

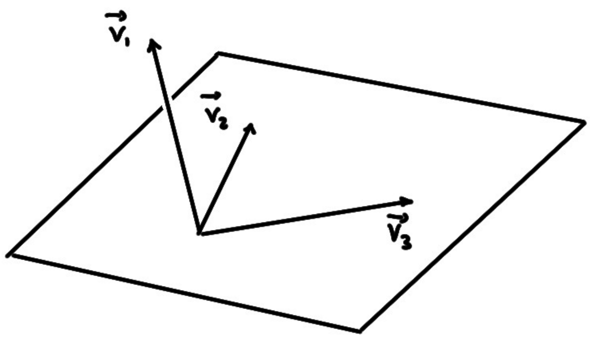
À la lumière de ce dernier théorème, illustrons encore la différence libre/liée dans le cas simple de trois vecteurs dans $\mathbb{R}^3$.

Exemple 2.21. Considérons une famille de trois vecteurs de $\mathbb{R}^3$, $\mathcal{F} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

⋆ Si $\mathcal{F}$ est liée, alors le théorème précédent implique que l'un des vecteurs, disons $\mathbf{v}_1$, peut s'écrire comme combinaison linéaire des deux autres. En d'autres termes, cela signifie que $\mathbf{v}_1$ est dans le plan engendré par $\mathbf{v}_2, \mathbf{v}_3$:



⋆ Par contre, si $\mathcal{F}$ est libre, alors le théorème implique qu'aucun des vecteurs ne peut s'écrire comme combinaison linéaire des autres (aucun n'est dans le plan engendré par les deux autres), ce qui exprime bien le fait que ces trois vecteurs *pointent tous dans des dimensions différentes* :



◇

Chapitre 3

Systèmes : formulation vectorielle

3.1 Systèmes : formulation matricielle

Dans ce chapitre, nous allons reformuler ce qui a été dit à propos des systèmes en utilisant le langage vectoriel de l'algèbre linéaire. Ceci aura plusieurs avantages, et mènera en particulier à une compréhension plus profonde des divers aspects liés à la recherche des solutions d'un système linéaire.

On peut voir un système $m \times n$ général, de la forme

$$(*)\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}\right. ,$$

comme une égalité entre deux vecteurs de $\mathbb{R}^m$:

$$\begin{pmatrix} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n \\ & & & \vdots & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} .$$

Or on peut récrire cette dernière comme suit :

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} ,$$

dans laquelle on reconnaît maintenant, dans le membre de gauche, une combinaison linéaire *des colonnes de la matrice associée au système*, qui est donnée, rappelons-le, par

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

Récrivons la même chose de façon plus compacte, en commençant par définir le vecteur associé au second membre,

$$\mathbf{b} := \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} ,$$

et récrivons la matrice du système comme une famille de colonnes,

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} ,$$

où, la k-ème colonne est le vecteur de $\mathbb{R}^m$ donné par

$$\mathbf{a}_k := \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{pmatrix} .$$

Donc la recherche de solutions $(x_1, x_2, \dots, x_n)$ au système $(*)$ est équivalente à demander si le membre de droite $\mathbf{b}$ appartient à la partie de $\mathbb{R}^m$ engendrée par les colonnes de A, c'est-à-dire si on peut écrire $\mathbf{b}$ comme une combinaison linéaire des colonnes de A :

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b} .$$

Dans cette formulation, les inconnues $x_1, \dots, x_n$ jouent le rôle de coefficients de la combinaison linéaire.

Une dernière définition permettra de faire encore un pas dans la description du système $(*)$.

Définition 3.1. Soit A une matrice $m \times n$, dont la k-ème colonne est notée $\mathbf{a}_k \in \mathbb{R}^m$,

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} ,$$

et soit $\mathbf{x}$ un vecteur de $\mathbb{R}^n$,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} .$$

Le **produit de A par $\mathbf{x}$** est le vecteur $A\mathbf{x} \in \mathbb{R}^m$ défini par la combinaison linéaire

$$A\mathbf{x} := x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n .$$

Le produit d'une matrice A $(m \times n)$ par un vecteur $\mathbf{x} \in \mathbb{R}^n$ crée donc un vecteur $A\mathbf{x} \in \mathbb{R}^m$. Cette transformation est l'exemple standard de ce que l'on appellera plus tard une *transformation linéaire*, puisqu'elle satisfait aux deux propriétés suivantes :

Lemme 2. *Soit A une matrice $m \times n$. Alors pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ et pour tout scalaire $\lambda \in \mathbb{R}$,*

1) $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$,
2) $A(\lambda\mathbf{x}) = \lambda\, A\mathbf{x}$.

*Ensemble, ces deux propriétés constituent la **linéarité**.*

Preuve: Notons la matrice $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$, et les vecteurs

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} , \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} .$$

Pour la première propriété,

$$\begin{aligned}
A(\mathbf{x}+\mathbf{y}) &= \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix}\begin{pmatrix}x_1+y_1\\ x_2+y_2\\ \vdots\\ x_n+y_n\end{pmatrix}\\
&= (x_1+y_1)\mathbf{a}_1+\cdots+(x_n+y_n)\mathbf{a}_n\\
&= \big(x_1\mathbf{a}_1+\cdots+x_n\mathbf{a}_n\big)+\big(y_1\mathbf{a}_1+\cdots+y_n\mathbf{a}_n\big)\\
&= A\mathbf{x}+A\mathbf{y}\,.
\end{aligned}$$

Pour la deuxième,

$$\begin{aligned}
A(\lambda\mathbf{x}) &= \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix}\begin{pmatrix}\lambda x_1\\ \lambda x_2\\ \vdots\\ \lambda x_n\end{pmatrix}\\
&= (\lambda x_1)\mathbf{a}_1+\cdots+(\lambda x_n)\mathbf{a}_n\\
&= \lambda(x_1\mathbf{a}_1)+\cdots+\lambda(x_n\mathbf{a}_n)\\
&= \lambda\big(x_1\mathbf{a}_1+\cdots+x_n\mathbf{a}_n\big)\\
&= \lambda\,A\mathbf{x}\,.
\end{aligned}$$

□

Avec les notations introduites, on peut maintenant écrire le système $(*)$ sous une forme purement vectorielle :

$$(*):\quad A\mathbf{x}=\mathbf{b}$$

L'existence d'une solution $\mathbf{x}$, pour ce système, signifie donc qu'il existe au moins une façon d'écrire le membre de droite $\mathbf{b}$ comme combinaison linéaire des colonnes de A.

3.2 Sur le nombre de solutions d'un système linéaire (BIS)

Comme première application de la formulation vectorielle d'un système linéaire $m\times n$, revisitons le Théorème "$0,1,\infty$", en donnant une preuve plus transparente que celle vue précédemment :

Théorème 3.2. *Soit A une matrice $m\times n$, $\mathbf{b}\in\mathbb{R}^m$ un second membre, et soit $S_{(*)}$ l'ensemble des vecteurs $\mathbf{x}\in\mathbb{R}^n$ solutions de*

$$(*):\quad A\mathbf{x}=\mathbf{b}\,,$$

Si $S_{(*)}$ n'est pas vide, alors soit il contient exactement un vecteur, soit il en contient une infinité.

Preuve: (La preuve est la même que dans la première version, mais formulée dans un langage vectoriel.) Supposons que $S_{(*)}$ n'est pas vide, et qu'il contient plus d'un élément. On a donc deux vecteurs $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$ distincts, tels que

$$A\mathbf{x}=\mathbf{b}\,,\qquad A\mathbf{y}=\mathbf{b}\,.$$

Considérons un scalaire λ quelconque, et définissons

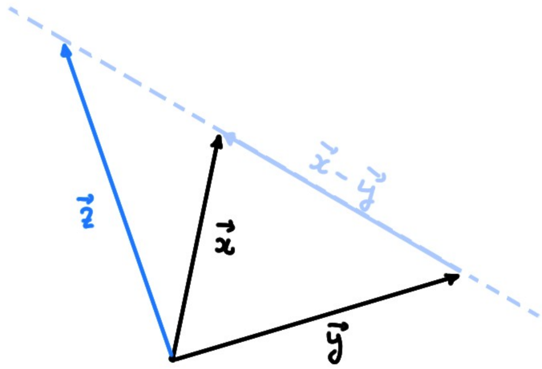
$$\mathbf{z}:=\mathbf{y}+\lambda(\mathbf{x}-\mathbf{y})\,.$$

Si λ est différent de 0 et 1, alors $\mathbf{z}$ est différent de $\mathbf{x}$ et de $\mathbf{y}$. Vérifions que $\mathbf{z}$ est aussi solution de $(*)$. En effet, par la linéarité démontrée dans le lemme,

$$A\mathbf{z}=A\big(\mathbf{y}+\lambda(\mathbf{x}-\mathbf{y})\big)=\underbrace{A\mathbf{y}}_{=\mathbf{b}}+\lambda\underbrace{\big(A\mathbf{x}-A\mathbf{y}\big)}_{=\mathbf{b}-\mathbf{b}=\mathbf{0}}=\mathbf{b}\,.$$

On peut donc, en choisissant λ, créer une infinité de nouvelles solutions.

La formulation vectorielle permet d'interpréter géométriquement la preuve donnée ci-dessus. En effet, on sait de la géométrie analytique que le vecteur $\mathbf{z} = \mathbf{y} + \lambda(\mathbf{x} - \mathbf{y})$ a son extrémité située sur la droite passant $\mathbf{y}$, dirigée par $\mathbf{x} - \mathbf{y}$:



L'infinité de solutions vient du fait qu'il existe une infinité de vecteurs ayant tous leur extrémité sur cette droite. □

3.3 Systèmes homogènes et inhomogènes

Dans l'étude des systèmes $m \times n$ du type

$$(*) : \quad A\mathbf{x} = \mathbf{b}\,,$$

il sera important de distinguer ceux dont le second membre $\mathbf{b}$ est nul.

Définition 3.3. Soit $\mathbf{b} \in \mathbb{R}^m$.

- ⋆ Si $\mathbf{b} = \mathbf{0}$, le système $(*)$ est dit **homogène** :

$$A\mathbf{x} = \mathbf{0}\,.$$

- ⋆ Si $\mathbf{b} \neq \mathbf{0}$, le système $(*)$ est dit **inhomogène**.

Solutions des systèmes homogènes

Commençons par une remarque importante : *tout système homogène est compatible*. En effet, il possède toujours la **solution triviale** , donnée par le vecteur nul $\mathbf{0} \in \mathbb{R}^n$, puisque le produit d'une matrice par le vecteur nul donne toujours le vecteur nul :

$$A\mathbf{0} = \mathbf{0}\,.$$

(Attention, le "$\mathbf{0}$" du membre de gauche est le vecteur nul de $\mathbb{R}^n$, alors que le "$\mathbf{0}$" du membre de droite est le vecteur nul de $\mathbb{R}^m$!)

Exemple 3.4. Étudions le système 3×3 homogène donné par

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 3 & 0 \\ 4 & 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\,.$$

L'opération $L_3 \leftarrow L_3 - 4L_1$ donne

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 3 & 0 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} ,$$

qui correspond au système triangulaire

$$\begin{cases} x_1 & & + & 2x_3 & = & 0 \\ & 3x_2 & & & = & 0 \\ & & - & 3x_3 & = & 0 \end{cases} ,$$

dont l'unique solution est $(x_1, x_2, x_3) = (0, 0, 0)$. On conclut que dans ce cas, il n'y a pas d'autre solution que la solution triviale. $\diamond$

Mais un système homogène peut posséder des solutions autres que la solution triviale. En fait, dès qu'il possède une solution autre que la triviale, on sait qu'il doit en posséder une infinité.

Exemple 3.5. Le système 3×3 homogène

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 8 & -2 \\ 2 & 5 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

est équivalent à

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 3 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} .$$

On voit que x_3 est libre, et donc que le système possède une infinité de solutions, décrites par

$$S = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = t \begin{pmatrix} -1 \\ 1 \\ 3 \end{pmatrix} \middle| t \in \mathbb{R} \right\}$$

On retrouve bien-sûr la solution triviale en prenant $t = 0$, mais toute valeur $t \neq 0$ donne une solution non-triviale.

Remarquons encore que l'ensemble S ci-dessus n'est autre que la partie de $\mathbb{R}^3$ engendrée par le vecteur non-nul $\mathbf{v} = \left(\begin{smallmatrix} -1 \\ 1 \\ 3 \end{smallmatrix}\right)$: $S = \text{Vect}\{\mathbf{v}\}$. On peut donc interpréter S comme l'ensemble de tous les vecteurs situés sur la droite dirigée par $\mathbf{v}$, passant par l'origine de $\mathbb{R}^3$. $\diamond$

Informel 3.6. Remarquons que quand on travaille dans les réels, l'équation (avec $a \neq 0$)

$$ax = 0$$

ne possède que "$x = 0$" comme solution. Ici, un système homogène

$$A\mathbf{x} = \mathbf{0}$$

peut posséder une infinité de solutions *non-nulles* (même si A contient des coefficients différents de zéro).

Exemple 3.7. Considérons le système homogène 2×4 suivant :

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

Deux opérations élémentaires ($L_1 \leftarrow L_1 - L_2$, suivie de $L_1 \leftrightarrow L_2$) mènent à la forme réduite

$$\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} ,$$

qui correspond à

$$(*) \left\{ \begin{array}{ccccccccc} x_1 & & & - & x_3 & & & = & 0 \\ & & x_2 & + & 2x_3 & + & x_4 & = & 0 \end{array} \right. ,$$

dans lequel $x_3 = s$ et $x_4 = t$ sont libres. On a donc

$$S = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} s \\ -2s - t \\ s \\ t \end{pmatrix} \Big| s, t \in \mathbb{R} \right\}$$

$\diamond$

Systèmes homogènes et indépendance linéaire

Par définition, le problème de savoir si une famille $\{\mathbf{v}_1, \dots, \mathbf{v}_p\} \subset \mathbb{R}^n$ est libre ou liée revient à étudier les familles de coefficients $\alpha_1, \dots, \alpha_p$ pour lesquelles la condition

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_p \mathbf{v}_p = \mathbf{0}$$

peut être satisfaite. D'un point de vue calculatoire, ce problème peut être reformulé comme suit.

Définissons la matrice $n \times p$,

$$A := \left[\mathbf{v}_1 \cdots \mathbf{v}_p \right] ,$$

et introduisons le vecteur

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix} \in \mathbb{R}^p .$$

Alors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ est libre si et seulement si le système homogène

$$A\alpha = \mathbf{0}$$

ne possède que la **solution triviale** : $\alpha = \mathbf{0}$.

Exemple 3.8. Soient

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} , \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ -4 \end{pmatrix} , \quad \mathbf{v}_3 = \begin{pmatrix} 7 \\ -2 \\ 3 \end{pmatrix} .$$

La famille $\mathcal{F} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \subset \mathbb{R}^3$ est-elle libre ou liée ?

Le système $A\alpha = \mathbf{0}$ correspondant est

$$\begin{pmatrix} 1 & 0 & 7 \\ 2 & 1 & -2 \\ 3 & -4 & 3 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} ,$$

qui est équivalent à

$$\begin{pmatrix} 1 & 0 & 7 \\ 0 & 1 & -16 \\ 0 & 0 & 41 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} ,$$

qui ne possède que la solution triviale. Donc $\mathcal{F}$ est libre. ◇

Cette façon de traiter l'indépendance linéaire permet d'énoncer un résultat général sur l'indépendance linéaire :

Théorème 3.9. *Dans $\mathbb{R}^n$, toute famille de plus de n vecteurs est liée.*

Preuve: Soit $\mathcal{F} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\} \subset \mathbb{R}^n$, avec $p > n$. On peut ranger ces vecteurs dans une matrice $n \times p$, qui a plus de colonnes que de lignes :

$$A := [\mathbf{v}_1 \dots \mathbf{v}_p] = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix} ,$$

où a_{ij}, $i = 1, \dots, n$, sont les composantes du vecteur $\mathbf{v}_j$.

Maintenant, étudions la dépendance en considérant la relation linéaire

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_p \mathbf{v}_p = \mathbf{0} .$$

Celle-ci correspond au système $A\alpha = \mathbf{0}$, qui a pour matrice augmentée

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1p} & 0 \\ a_{21} & a_{22} & \cdots & a_{2p} & 0 \\ \vdots & \vdots & & \vdots & 0 \\ a_{n1} & a_{n2} & \cdots & a_{np} & 0 \end{array} \right)$$

Puisque $p > n$, sa forme réduite doit contenir au moins une colonne ne contenant pas de pivot, et donc le système possède au moins une variable libre. Ceci implique, comme le second membre est nul, qu'il existe une infinité de solutions non-triviales, et donc que $\mathcal{F}$ est liée. □

Solutions des systèmes inhomogènes

Fixons maintenant un $\mathbf{b} \in \mathbb{R}^m$ *non-nul*, et considérons le système inhomogène

$$(*) : \quad A\mathbf{x} = \mathbf{b} .$$

À l'opposé des systèmes homogènes (qui ont toujours au moins la solution triviale), il n'y a aucune garantie concernant l'existence d'une solution. Mais pour que la discussion ci-dessous ne soit pas vide, supposons que ce système est compatible : $S_{(*)} \neq \emptyset$. Notre but ci-dessous sera décrire une propriété générale de l'ensemble $S_{(*)}$.

Le résultat suivant va nous montrer que les solutions de ce système sont intimement liées à celle du **système homogène associé**, qui est celui avec la même matrice A, mais dans lequel on remplace $\mathbf{b}$ par $\mathbf{0}$:

$$(*)_h : \quad A\mathbf{x} = \mathbf{0} .$$

Théorème 3.10. *Supposons déjà connue une solution de* $(*)$, *que l'on nommera* ***particulière***, *et que l'on notera* $\mathbf{v}_p$. *Alors toute autre solution de* $(*)$, $\mathbf{v} \in S_{(*)}$, *peut s'écrire comme*

$$\mathbf{v} = \mathbf{v}_p + \mathbf{v}_h \,,$$

où $\mathbf{v}_h$ *est une certaine solution du problème homogène* $(*)_h$ *associé.*

Preuve: Si $\mathbf{v} \in S_{(*)}$, alors

$$A\mathbf{v} = \mathbf{b}\,.$$

Mais puisque $\mathbf{v}_p \in S_{(*)}$, on a aussi

$$A\mathbf{v}_p = \mathbf{b}\,.$$

En soustrayant ces deux dernières expressions, on obtient

$$A\mathbf{v} - A\mathbf{v}_p = \mathbf{b} - \mathbf{b} = \mathbf{0}\,.$$

Par linéarité, ceci implique que

$$A(\mathbf{v} - \mathbf{v}_p) = \mathbf{0}\,.$$

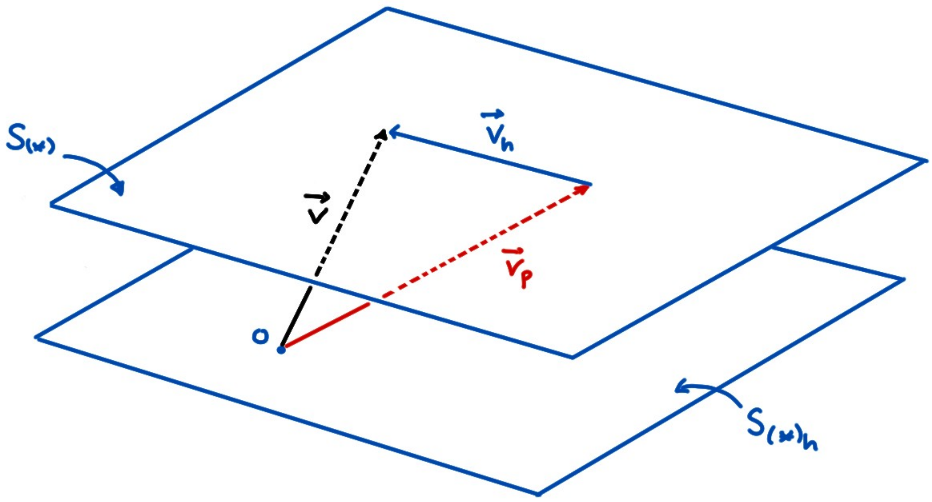
Ainsi, en définissant $\mathbf{v}_h := \mathbf{v} - \mathbf{v}_p$, cette dernière dit bien que $\mathbf{v}_h$ est solution du problème homogène associé : $A\mathbf{v}_h = \mathbf{0}$. Puisque $\mathbf{v} = \mathbf{v}_p + \mathbf{v}_h$, ceci démontre le résultat. □

Ce théorème peut être résumé comme suit : si on connaît seulement *une* solution du système $(*)$, et si on sait complètement résoudre le système homogène associé $(*)_h$, alors on connaît *toutes* les solutions du système $(*)$. Plus concrètement, pour résoudre $(*)$, on pourra procéder comme suit :

1) Chercher une solution particulière $\mathbf{v}_p$ de $(*)$.
2) Résoudre le système homogène associé $(*)_h$, c'est-à-dire trouver l'ensemble $S_{(*)_h}$.
3) Combiner les deux, pour produire

$$S_{(*)} = \left\{\mathbf{v} = \mathbf{v}_p + \mathbf{v}_h \,|\, \mathbf{v}_h \in S_{(*)_h}\right\}.$$

Interprétation géométrique :



Voyons comment cette structure peut s'observer sur un exemple concret.

Exemple 3.11. Considérons le système 3×3 suivant :

$$(*)\left\{\begin{array}{ccccccc} x_1 & + & 3x_2 & + & x_3 & = & 1 \\ 2x_1 & + & x_2 & - & 3x_3 & = & 7 \\ -x_1 & + & 5x_2 & + & 7x_3 & = & -9 \end{array}\right.$$

qui correspond à

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 1 & -3 \\ -1 & 5 & 7 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 1 \\ 7 \\ -9 \end{pmatrix}.$$

Remarquons que le vecteur

$$\mathbf{v}_p := \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}$$

est solution du système. En effet,

$$A\mathbf{v}_p = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 1 & -3 \\ -1 & 5 & 7 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 7 \\ -9 \end{pmatrix} = \mathbf{b}\,.$$

On a donc une solution particulière $\mathbf{v}_p$. On sait maintenant, par le théorème, que l'on aura toutes les autres solutions en résolvant le système homogène associé, c'est-à-dire

$$(*)_h : \qquad \begin{pmatrix} 1 & 3 & 1 \\ 2 & 1 & -3 \\ -1 & 5 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

En procédant comme d'habitude, on obtient

$$S_{(*)_h} = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = t \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\}.$$

On sait donc, par le théorème, que toutes les solutions du problème inhomogène sont données par

$$\begin{aligned} S_{(*)} &= \left\{ \mathbf{v} = \mathbf{v}_p + \mathbf{v}_h \,|\, \mathbf{v}_h \in S_{(*)_h} \right\} \\ &= \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix} + t \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\}. \end{aligned}$$

Bien-sûr, on observe cette structure aussi si on résout le système avec la technique habituelle. En partant de la matrice augmentée,

$$\left(\begin{array}{ccc|c} 1 & 3 & 1 & 1 \\ 2 & 1 & -3 & 7 \\ -1 & 5 & 7 & -9 \end{array}\right),$$

dont l'échelonnage mène à identifier la variable libre $x_3 = t$, on trouve

$$\begin{aligned} x_1 &= 4 + 2t\,, \\ x_2 &= -1 - t\,, \\ x_3 &= t\,. \end{aligned}$$

Vectoriellement, on peut écrire l'ensemble des solutions comme

$$S_{(*)} = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\}$$

On voit donc encore une fois la structure "solution particulière + toutes les solutions du problème homogène associé". ◇

3.4 Applications linéaires : introduction

Plus haut, nous avons défini le membre de droite d'un système $m \times n$, à savoir "$A\mathbf{x}$", comme le produit d'une matrice (de taille $m \times n$) A par le vecteur $\mathbf{x} \in \mathbb{R}^n$. Ce produit étant défini comme une combinaison linéaire des colonnes de A, $A\mathbf{x}$ est un vecteur de $\mathbb{R}^m$.

La multiplication par une matrice $m \times n$ est donc une opération qui transforme les vecteurs de $\mathbb{R}^n$ en des vecteurs de $\mathbb{R}^m$. C'est un cas particulier d'une **application (ou fonction) de $\mathbb{R}^n$ dans $\mathbb{R}^m$** :

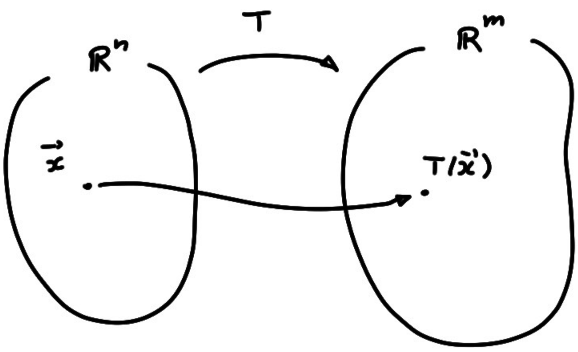
$$\begin{aligned} \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{x} &\mapsto A\mathbf{x}\,. \end{aligned}$$

Si nécessaire, quelques notions générales sur les fonctions sont rappelées **ici** (lien web).

Applications : le point de vue général

Plus généralement, une application n'est pas forcément définie à l'aide d'une matrice. On utilisera souvent la lettre "T" pour représenter une application générique :

$$\begin{aligned} T : \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{x} &\mapsto T(\mathbf{x})\,. \end{aligned}$$



Le vecteur $\mathbf{y} = T(\mathbf{x})$ est appelé **image** de $\mathbf{x}$ (par T), et $\mathbf{x}$ est une **préimage** de $\mathbf{y}$.

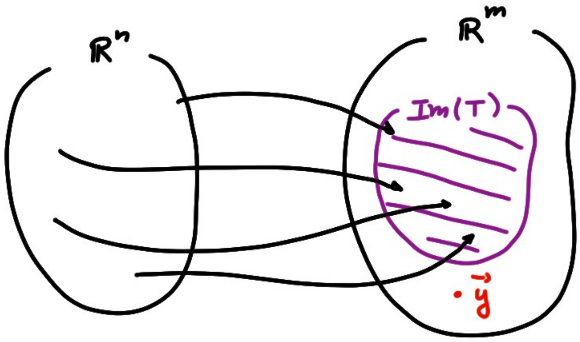
Considérons un instant une équation du type suivant

$$(*) \;:\; T(\mathbf{x}) = \mathbf{b}\,,$$

où le membre de droite $\mathbf{b} \in \mathbb{R}^m$ est fixé. L'existence d'au moins une solution $\mathbf{x} \in \mathbb{R}^n$, pour cette équation, revient à demander si $\mathbf{b}$ fait partie des éléments de l'ensemble d'arrivée qui sont "atteints" par l'application, c'est-à-dire pour lesquels il existe au moins une préimage. Ceci mène à la définition suivante :

Définition 3.12. L'**ensemble image** de $T : \mathbb{R}^n \to \mathbb{R}^m$ est défini par

$$\mathrm{Im}(T) := \left\{ \mathbf{y} \in \mathbb{R}^m \ : \ \exists \mathbf{x} \in \mathbb{R}^n \text{ t.q. } T(\mathbf{x}) = \mathbf{y} \right\}.$$



On a donc, pour l'équation $(*)$ ci-dessus :

- ⋆ Si $\mathbf{b} \notin \mathrm{Im}(T)$, alors $(*)$ ne possède aucune solution.
- ⋆ Si $\mathbf{b} \in \mathrm{Im}(T)$, alors $(*)$ possède au moins une solution.

Définition de la linéarité

Définition 3.13. Une application

$$\begin{aligned} T : \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{x} &\mapsto T(\mathbf{x}) \end{aligned}$$

est dite **linéaire** si elle satisfait aux deux propriétés suivantes :

1) $T(\mathbf{x} + \mathbf{x}') = T(\mathbf{x}) + T(\mathbf{x}')$ pour tous $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$.
2) $T(\lambda \mathbf{x}) = \lambda T(\mathbf{x})$ pour tous $\lambda \in \mathbb{R}$ et $\mathbf{x} \in \mathbb{R}^n$.

Exemple 3.14. Considérons l'application $T : \mathbb{R}^3 \to \mathbb{R}^2$ définie ainsi :

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{x} \mapsto T(\mathbf{x}) := \begin{pmatrix} -x_1 + 3x_2 + 5x_3 \\ x_3 + 7x_1 \end{pmatrix}$$

Montrons, "à la main", uniquement à l'aide de la définition de linéarité, que T est linéaire. Pour ce faire, prenons un $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ et un scalaire λ. On utilise la définition de T pour calculer

$$\begin{aligned} T(\lambda \mathbf{x}) &= T\left(\begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ x_3 \end{pmatrix} \right) \\ &= \begin{pmatrix} (-\lambda x_1) + 3(\lambda x_2) + 5(\lambda x_3) \\ (\lambda x_3) + 7(\lambda x_1) \end{pmatrix} \\ &= \lambda \begin{pmatrix} -x_1 + 3x_2 + 5\lambda x_3 \\ x_3 + 7x_1 \end{pmatrix} \\ &= \lambda T(\mathbf{x}). \end{aligned}$$

Ensuite, pour toute paire $\mathbf{x}, \mathbf{y}$,

$$\begin{aligned}
T(\mathbf{x}+\mathbf{y}) &= T\Big(\begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ x_3+y_3 \end{pmatrix}\Big) \\
&= \begin{pmatrix} -(x_1+y_1)+3(x_2+y_2)+5(x_3+y_3) \\ (x_3+y_3)+7(x_1+y_1) \end{pmatrix} \\
&= \begin{pmatrix} -x_1+3x_2+5x_3 \\ x_3+7x_1 \end{pmatrix} + \begin{pmatrix} -y_1+3y_2+5y_3 \\ y_3+7y_1 \end{pmatrix} \\
&= T(\mathbf{x}) + T(\mathbf{y}) \,.
\end{aligned}$$

On a donc bien montré que T est linéaire. ◇

Nous avons déjà vu que si $T : \mathbb{R}^n \to \mathbb{R}^m$, et si il existe une matrice $m \times n$ telle que

$$T(\mathbf{x}) = A\mathbf{x} \qquad \forall \mathbf{x} \in \mathbb{R}^n \,,$$

alors T est linéaire . Mais a priori, une application peut être linéaire sans forcément être associée à une matrice.

Exemple 3.15. Si on reprend l'application T de l'exemple précédent, on peut remarquer que

$$T(\mathbf{x}) = \begin{pmatrix} -x_1+3x_2+5x_3 \\ x_3+7x_1 \end{pmatrix} = \underbrace{\begin{pmatrix} -1 & 3 & 5 \\ 7 & 0 & 1 \end{pmatrix}}_{=:A} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = A\mathbf{x} \,.$$

Ainsi, T est linéaire. ◇

Pour montrer qu'une application n'est *pas* linéaire, il suffit de montrer qu'une des deux conditions qui définit la linéarité n'est pas satisfaite, en exhibant un *contre-exemple*. On pourra donc

- ⋆ soit trouver un $\mathbf{x}_*$ et un scalaire λ tel que $T(\lambda\mathbf{x}_*) \neq \lambda T(\mathbf{x}_*)$,
- ⋆ soit trouver deux vecteurs $\mathbf{x}_*, \mathbf{y}_*$ tels que $T(\mathbf{x}_* + \mathbf{y}_*) \neq T(\mathbf{x}_*) + T(\mathbf{y}_*)$.

Exemple 3.16. Considérons l'application $T : \mathbb{R}^2 \to \mathbb{R}^3$ définie par

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x} \mapsto T(\mathbf{x}) := \begin{pmatrix} x_1x_2 \\ -x_2 \\ x_1 \end{pmatrix}$$

L'apparition de la multiplication "x_1x_2" indique que cette application n'est probablement pas linéaire. Comme contre-exemple, prenons $\lambda = 2$, et $\mathbf{x}_* = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$. On a

$$T(\lambda\mathbf{x}_*) = T\Big(2\begin{pmatrix} 1 \\ 3 \end{pmatrix}\Big) = T\Big(\begin{pmatrix} 2 \\ 6 \end{pmatrix}\Big) = \begin{pmatrix} 12 \\ -6 \\ 2 \end{pmatrix} \,,$$

alors que

$$\lambda T(\mathbf{x}_*) = 2T\Big(\begin{pmatrix} 1 \\ 3 \end{pmatrix}\Big) = 2\begin{pmatrix} 3 \\ -3 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ -6 \\ 2 \end{pmatrix} \,.$$

On a donc $T(\lambda\mathbf{x}_*) \neq \lambda T(\mathbf{x}_*)$, ce qui implique que T n'est pas linéaire. ◇

Pour la suite...

Nous aurons encore beaucoup à dire sur les applications linéaires, qui sont les vraies “fonctions” étudiées en algèbre linéaire (un peu comme les fonctions continues sont les fonctions les plus étudiées en analyse).

Mais avant d’en dire plus, nous allons faire un pause, dans le chapitre suivant, et reprendre tout ce que nous avons fait jusqu’ici, en adoptant un point de vue beaucoup plus général, celui des *espaces vectoriels* abstraits. Nous introduirons plus de choses dans ce cadre, en particulier à propos des applications linéaires d’un espace vectoriel dans un autre. Plus tard, nous appliquerons alors ces notions lorsque nous reviendrons plus en profondeur sur les applications linéaires du type $\mathbf{x} \mapsto T(\mathbf{x}) = A\mathbf{x}$.

Chapitre 4

Espaces vectoriels

4.1 Motivation

On a pu apprécier, dans les dernières sections, à quel point l'introduction de la notion abstraite de *vecteur* s'est avérée utile, non seulement dans la description des systèmes linéaires, mais aussi dans l'avantage qu'ils représentent d'un point de vue calculatoire : on peut les manipuler un peu comme de simples *nombres* , sans se soucier du fait qu'ils représentent, a priori, des objets de grandes dimensions.

Les vecteurs nous ont également permis de développer le début de la théorie des applications linéaires $T : \mathbb{R}^n \to \mathbb{R}^m$, qui nous occuperont pour la plupart de ce que nous allons faire jusqu'à la fin de ce cours.

Mais avant de poursuivre cette étude, nous allons généraliser tout ce que nous avons fait jusqu'ici, pour l'utiliser dans d'autres situations.

En effet, il est profitable, dans beaucoup de situations qui vont bien au-delà de ce que nous avons vu jusqu'à maintenant, d'avoir une structure vectorielle abstraite qui permette de manipuler des objets à l'aide d'une *addition vectorielle* et d'une *multiplication par un scalaire*, telle que les propriétés classiques de l'arithmétique (commutativité, distributivité, etc) soient satisfaites. Cette structure, qui généralise la notion de vecteur dans $\mathbb{R}^n$, est ce qu'on appelle un *espace vectoriel*, et constitue le sujet de ce chapitre.

Les espaces vectoriels offrent un cadre de travail sur lequel nous redéfinirons naturellement tout ce que nous avons fait dans le cas de $\mathbb{R}^n$. Nous introduirons également de nouvelles notions, qui seront après utilisées dans le cas particulier des espaces $\mathbb{R}^n$.

Informel 4.1. Attention : le contenu de ce chapitre est *abstrait* ! La difficulté principale, pour le novice, est d'accepter le fait que l'on va parler de “vecteurs” sans dire exactement ce qu'ils sont ; il faudra s'habituer à travailler avec ces objets en utilisant uniquement les propriétés qui les définissent, et qui sont décrites dans la définition de la section suivante.

4.2 Définition et exemples

Commençons par introduire la généralisation abstraite de la notion de vecteur rencontrée dans les chapitres précédents :

Définition 4.2. Un **espace vectoriel** est un ensemble non-vide, noté souvent V, dont les éléments sont appelés **vecteurs**, notés souvent $u, v, w, \dots$, muni d'une **addition** et d'une **multiplication par un scalaire**, satisfaisant aux propriétés suivantes :

1) $u + v = v + u$ pour tous $u, v \in V$ (commutativité)
2) $u + (v + w) = (u + v) + w$ pour tous $u, v, w \in V$ (associativité)
3) Il existe un **vecteur nul**, noté "0", tel que pour tout $v \in V$,
$$0 + v = v + 0 = v\,.$$
4) Pour tout $v \in V$, il existe un vecteur **opposé**, noté $-v$, tel que
$$v + (-v) = (-v) + v = 0\,.$$
5) $\lambda(u + v) = \lambda u + \lambda v$ pour tous $\lambda \in \mathbb{R}$, $u, v \in V$
6) $(\lambda + \mu)v = \lambda v + \mu v$ pour tous $\lambda, \mu \in \mathbb{R}$, $v \in V$
7) $\lambda(\mu v) = (\lambda\mu)v = \mu(\lambda v)$ pour tous $\lambda, \mu \in \mathbb{R}$, $v \in V$
8) $1v = v$ pour tout $v \in V$

Remarque 4.3. Ce que l'on vient de définir est généralement appelé espace vectoriel *réel*, car les scalaires utilisés pour multiplier les vecteurs sont des nombres *réels*. ◇

Donc un espace vectoriel est simplement un ensemble d'objets abstraits appelés vecteurs, dans lequel un "+" permet d'additionner ces vecteurs, et dans lequel on peut multiplier les vecteurs par des scalaires.

Informel 4.4. Cela peut prendre du temps de s'habituer à ce niveau d'abstraction, et d'imaginer que ce genre de structure existe ailleurs que dans le cadre des "flèches de $\mathbb{R}^n$". C'est surtout à la fin du cours qu'on se rendra compte de l'utilité de cette généralisation, lorsqu'on pourra résoudre des problèmes concrets en appliquant des méthodes algébriques/géométriques (par exemple : la méthode des moindres carrés) dans un espace vectoriel abstrait.

Voyons quelques-uns des principaux exemples d'espaces vectoriels.

Espaces $\mathbb{R}^n$

Le premier exemple d'espace vectoriel que nous avons rencontré est bien-sûr celui où V est formé de tous les vecteurs de $\mathbb{R}^n$. Dans ce cas l'addition et la multiplication par un scalaire avaient été définis de façon naturelle, à savoir *composante par composante*. C'est souvent le même procédé qui est utilisé dans des cas plus généraux.

Espaces de fonctions

Dans ce premier exemple, nous allons voir comment des ensembles de *fonctions* peuvent aussi être vus comme des espaces vectoriels.

Soit $I \subset \mathbb{R}$ un intervalle (borné ou non, I peut même être la droite toute entière), et soit V l'ensemble de toutes les fonctions définies sur I, à valeurs réelles :
$$V = \big\{\text{fonctions } f : I \to \mathbb{R}\big\}\,.$$

Remarque 4.5. Une fonction $f \in V$ est définie une fois que l'on a défini la valeur du réel $f(t)$ pour chaque $t \in I$. Ainsi, deux fonctions $f, g \in V$ sont **égales**, ce qu'on écrit $f = g$, si et seulement si elles prennent la même valeur en tout point, c'est-à-dire si

$$f(t) = g(t)\,, \quad \forall t \in I\,.$$

$\diamond$

- $\star$ Définissons une **addition** sur V. Pour ce faire, nous devons associer à toute paire $f, g \in V$ une nouvelle fonction $f + g \in V$. On doit donc définir le réel $(f + g)(t)$ pour tout $t \in I$, ce que l'on fait naturellement en posant

$$(f + g)(t) := f(t) + g(t)\,, \qquad \forall t \in I\,.$$

- $\star$ Définissons la **multiplication par un scalaire** : si $f \in V$ et $\lambda \in \mathbb{R}$, alors $\lambda f \in V$ est la fonction $\lambda f : I \to \mathbb{R}$ définie comme suit :

$$(\lambda f)(t) := \lambda f(t)\,, \qquad \forall t \in I\,.$$

Nous devons maintenant vérifier que V est bien un espace vectoriel. Pour cela, nous aurons besoin de la **fonction nulle** $0 \in V$, comme étant la fonction qui vaut zéro en tout point,

$$0(t) := 0\,, \quad \forall t \in I\,,$$

et l'**opposé** d'une fonction $f \in V$, notée $-f \in V$, est la fonction

$$(-f)(t) := -f(t)\,, \quad \forall t \in I\,.$$

Théorème 4.6. *Muni de l'addition et de la multiplication par un scalaire (définies ci-dessus), V est un espace vectoriel.*

Preuve: On vérifie une à une chacune des propriétés qui définissent un espace vectoriel. (On remarquera qu'à chaque fois, c'est une propriété des *réels* qui fait le travail !)

1) Soient $f, g \in V$. Si on fixe $t \in I$, on peut écrire

$$(f + g)(t) = f(t) + g(t) = g(t) + f(t) = (g + f)(t)\,.$$

   Comme cette identité est vraie pour tout $t \in I$, cela implique bien que $f + g = g + f$.

2) Soient $f, g, h \in V$. Si on fixe $t \in I$, alors

$$\begin{aligned}(f + (g + h))(t) &= f(t) + (g + h)(t)\\ &= f(t) + (g(t) + h(t))\\ &= (f(t) + g(t)) + h(t)\\ &= (f + g)(t) + h(t) = ((f + g) + h)(t)\,.\end{aligned}$$

   Comme cette identité est vraie pour tout $t \in I$, cela implique bien que $f + (g + h) = (f + g) + h$.

3) Par la définition de la fonction nulle, on a bien-sûr que $f + 0 = f$ pour toute $f \in V$, puisque

$$(f + 0)(t) = f(t) + 0(t) = f(t)\,, \qquad \forall t \in I\,.$$

4) Avec l'opposé $-f$ défini plus haut, pour tout $t \in I$,

$$(f + (-f))(t) = f(t) + (-f)(t) = f(t) - f(t) = 0 = 0(t)\,,$$

   ce qui implique que $f + (-f) = 0$.

5) Soient $f, g \in V$, et soit $\lambda \in \mathbb{R}$. Pour tout $t \in I$, on a

$$\begin{aligned}(\lambda(f+g))(t) &= \lambda((f+g)(t)) \\ &= \lambda(f(t)+g(t)) \\ &= \lambda f(t) + \lambda g(t) \\ &= (\lambda f)(t) + (\lambda g)(t) \\ &= (\lambda f + \lambda g)(t)\,,\end{aligned}$$

ce qui implique $\lambda(f+g) = \lambda f + \lambda g$.

6) Soient $\lambda, \mu \in \mathbb{R}$, et $f \in V$. On a, pour tout $t \in I$,

$$\begin{aligned}((\lambda+\mu)f)(t) &= (\lambda+\mu)f(t) \\ &= \lambda f(t) + \mu f(t) \\ &= (\lambda f)(t) + (\mu f)(t) \\ &= (\lambda f + \mu f)(t)\,,\end{aligned}$$

ce qui implique bien que $(\lambda+\mu)f = \lambda f + \mu f$.

7) Soient $\lambda, \mu \in \mathbb{R}$, $f \in V$. On a, pour tout $t \in I$,

$$\begin{aligned}(\lambda(\mu f))(t) &= \lambda((\mu f)(t)) \\ &= \lambda(\mu f(t)) \\ &= (\lambda\mu) f(t) \\ &= (\mu\lambda) f(t) \\ &= \mu(\lambda f(t)) \\ &= \mu((\lambda f)(t)) \\ &= (\mu(\lambda f))(t)\,,\end{aligned}$$

ce qui implique bien que $\lambda(\mu f) = (\lambda\mu)f = \mu(\lambda f)$.

8) Soit $f \in V$. On a, pour tout $t \in I$,

$$(1f)(t) = 1 \cdot f(t) = f(t)\,,$$

ce qui implique bien $1f = f$.

□

Informel 4.7. La preuve est étonnamment longue, mais ne présente *aucune* subtilité ! (La seule difficulté, peut-être, est de comprendre pourquoi il est nécessaire de faire tout ça !)

Espaces de polynômes

Les polynômes sont des fonctions très particulières mais fournissent un cas important d'espace vectoriel, jouant un rôle important dans de nombreuses applications.

Soit $n \geqslant 1$ un entier, et soit $\mathbb{P}_n$ l'**ensemble de tous les polynômes réels de degré au plus égal à** n. Cela signifie qu'un élément $p \in \mathbb{P}_n$ est un polynôme de la forme

$$p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n\,, \qquad t \in \mathbb{R}\,,$$

où les coefficients $a_0, a_1, \dots, a_n$ sont des réels. On additionne et multiplie (par des scalaires) comme on l'a fait pour les fonctions.

Théorème 4.8. *Muni de l'addition et de la multiplication par un scalaire, $\mathbb{P}_n$ est un espace vectoriel.*

Preuve: (voir exercices) □

Espace des matrices

Considérons l'**ensemble des matrices** $m \times n$ **à coefficients réels**, noté $\mathbb{M}_{m\times n}(\mathbb{R})$. Si une matrice $A \in \mathbb{M}_{m\times n}(\mathbb{R})$ a des coefficients a_{ij} ($i = 1, \dots, m$, $j = 1, \dots, n$), on écrira simplement

$$A = (a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}}$$

On rappelle les définitions d'addition et de multiplication par un scalaire, introduites précédemment :

- ⋆ Si $A, B \in \mathbb{M}_{m\times n}(\mathbb{R})$,
$$A = (a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}}\,, \qquad B = (b_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}}\,,$$
on définit $A + B \in \mathbb{M}_{m\times n}(\mathbb{R})$ comme la matrice dont les éléments sont les nombres $a_{ij} + b_{ij}$:
$$A + B := (a_{ij} + b_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}}\,.$$
- ⋆ Pour un scalaire $\lambda \in \mathbb{R}$, on définit $\lambda A \in \mathbb{M}_{m\times n}(\mathbb{R})$ comme la matrice dont les éléments sont les nombres λa_{ij} :
$$\lambda A := (\lambda a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}}\,.$$

Théorème 4.9. *Muni de l'addition et de la multiplication par un scalaire (définies ci-dessus), $\mathbb{M}_{m\times n}(\mathbb{R})$ est un espace vectoriel.*

Preuve: En exercice ! L'élément nul "0" est la matrice $m \times n$ dont tous les éléments sont égaux à zéro, et l'opposé d'une matrice A est la matrice dont tous les éléments sont les opposés de ceux de A. □

Autres exemples

La structure d'espace vectoriel apparaît dans de nombreuses situations.

Exemple 4.10. Soit V l'ensemble des **suites de réels**, dans lequel une suite est notée simplement $\mathbf{x} = (x_n)_{n\geqslant 0}$. En définissant une multiplication par un scalaire $\lambda \in \mathbb{R}$,

$$\lambda\mathbf{x} := (\lambda x_n)_{n\geqslant 0}\,,$$

et l'addition

$$\mathbf{x} + \mathbf{y} := (x_n + y_n)_{n\geqslant 0}\,,$$

on peut vérifier (en exercice) que V a une structure d'espace vectoriel. ⋄

4.3 Colinéarité et in-dépendance linéaire

Une fois que l'on est dans un espace vectoriel V bien défini, on peut *importer* n'importe quelle notion vectorielle, rencontrée dans $\mathbb{R}^n$, dans V. Ceci permettra de profiter de ces notions pour résoudre des problèmes dans un cadre abstrait, ayant parfois des conséquences pratiques surprenantes.

Arrêtons-nous sur quelques-unes de ces notions, qui seront empruntées directement de ce que nous avons fait dans $\mathbb{R}^n$.

Colinéarité

Définition 4.11. Soit V un espace vectoriel. Deux vecteurs $u, v \in V$ sont **colinéaires** si l'un d'eux peut s'écrire comme un multiple de l'autre.

Exemple 4.12. Les matrices $A, B \in \mathbb{M}_{2\times 3}(\mathbb{R})$ définies par

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & -3 & \frac{1}{2} \end{pmatrix}, \qquad B = \begin{pmatrix} -2 & 0 & -4 \\ 0 & 6 & -1 \end{pmatrix}$$

sont colinéaires, puisque $B = -2A$. Par contre,

$$A' = \begin{pmatrix} 1 & 1 & 2 \\ 1 & -3 & \frac{1}{2} \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 7 & 2 \\ 0 & -2 & -1 \end{pmatrix}$$

ne sont pas colinéaires, parce qu'il n'existe aucun $\lambda \in \mathbb{R}$ tel que $A = \lambda B$ ou tel que $B = \lambda A$. ◇

Exemple 4.13. Soient $f, g : \mathbb{R} \to \mathbb{R}$ les fonctions

$$f(t) := \sin(t) \qquad g(t) := \cos(t) \qquad t \in \mathbb{R}\,.$$

Montrons que f et g ne sont pas colinéaires. On le démontre par l'absurde : supposons qu'il existe $\lambda \in \mathbb{R}$ tel que $g = \lambda f$, c'est-à-dire tel que

$$\cos(t) = \lambda \sin(t)\,, \qquad \forall t \in \mathbb{R}\,.$$

En écrivant cette relation pour le choix particulier $t = \frac{\pi}{4}$, on obtient

$$\frac{\sqrt{2}}{2} = \lambda \cdot \frac{\sqrt{2}}{2}\,,$$

qui implique $\lambda = 1$. Mais, pour le choix $t = \frac{\pi}{2}$, on obtient

$$0 = \lambda \cdot 1\,,$$

qui implique $\lambda = 0$. On conclut qu'il ne peut pas exister de scalaire λ qui fonctionne pour tous les $t \in \mathbb{R}$. On conclut que f et g ne sont pas colinéaires. ◇

Combinaisons linéaires et in-dépendance

Si $v_1, \dots, v_p$ est une famille de vecteurs d'un espace vectoriel V, et si $\lambda_1, \dots, \lambda_p$ sont des scalaires, on peut considérer la **combinaison linéaire**

$$\lambda_1 v_1 + \cdots + \lambda_p v_p\,.$$

On peut alors généraliser la notion d'indépendance linéaire :

Définition 4.14. Soient $v_1, \dots, v_p$ des vecteurs d'un espace vectoriel V. La famille $\{v_1, \dots, v_p\}$ est dite

⋆ **linéairement indépendante** (ou **libre**) si l'unique combinaison linéaire nulle,

$$\lambda_1 v_1 + \cdots + \lambda_p v_p = 0$$

est celle pour laquelle $\lambda_1 = \lambda_2 = \cdots = \lambda_p = 0$.

⋆ **dépendante** (ou **liée**) si il existe des coefficients $\lambda_1, \lambda_2, \dots, \lambda_p$, dont au moins un n'est pas nul, tels que

$$\lambda_1 v_1 + \cdots + \lambda_p v_p = 0\,.$$

Exemple 4.15. Considérons les matrices

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

et montrons que la famille $\{A, B, C\}$ est libre. Pour ce faire, considérons la relation linéaire

$$\lambda_1 A + \lambda_2 B + \lambda_3 C = 0\,,$$

qui signifie en fait

$$\lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

c'est-à-dire

$$\begin{pmatrix} \lambda_1 + \lambda_3 & \lambda_2 - \lambda_3 \\ \lambda_2 & -\lambda_1 + \lambda_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Deux matrices sont égales si et seulement si tous leurs coefficients sont égaux, donc cette dernière égalité entre matrices 2×2 est équivalente à

$$(*) \left\{ \begin{array}{rcrcrcl} \lambda_1 & & & + & \lambda_3 & = & 0 \\ & & \lambda_2 & - & \lambda_3 & = & 0 \\ & & \lambda_2 & & & = & 0 \\ -\lambda_1 & & & + & \lambda_3 & = & 0 \end{array} \right.$$

Ce système ne possédant que la solution triviale, $\lambda_1 = \lambda_2 = \lambda_3 = 0$, on en conclut que $\{A, B, C\}$ est libre ou, en d'autres termes, qu'aucune de ces matrices ne peut s'écrire comme combinaison linéaire des deux autres. ◇

Exemple 4.16. Dans l'espace V de toutes les fonctions de $\mathbb{R}$ dans $\mathbb{R}$, considérons pour tout $k = 0, 1, \ldots, p$, le polynôme $f_k(t) := t^k$, c'est-à-dire que

$$f_0(t) := 1\,, \quad f_1(t) := t\,, \quad f_2(t) := t^2\,, \cdots, f_p(t) := t^p\,.$$

Lemme 3. *$\{f_0, f_1, \ldots, f_p\}$ est libre.*

Preuve: Soient $\lambda_0, \lambda_1, \ldots, \lambda_p$ des scalaires. On a

$$\lambda_0 f_0 + \lambda_1 f_1 + \cdots + \lambda_p f_p = 0$$

si et seulement si

$$\lambda_0 + \lambda_1 t + \lambda_2 t^2 + \cdots + \lambda_p t^p = 0\,, \qquad \forall t \in \mathbb{R}\,.$$

On va maintenant utiliser le résultat suivant :

Théorème 4.17. *Soit $p(t)$ un polynôme à coefficients réels :*

$$p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_p t^p\,.$$

Si $p(t) = 0$ pour tout $t \in I$ (où I est un intervalle ouvert), alors p est le polynôme nul, c'est-à-dire que $a_0 = a_1 = \cdots = a_p = 0$.

Preuve: Voir par exemple **ici** (lien web). ◇

Ce résultat implique, en prenant $I = \mathbb{R}$, que $\lambda_0 = \lambda_1 = \cdots = \lambda_p = 0$, et donc que la famille $\{f_0, f_1, \ldots, f_p\}$ est libre. ◇

◇

Exemple 4.18. Dans l'espace V de toutes les fonctions de $\mathbb{R}$ dans $\mathbb{R}$, considérons la famille $\{f, g, h\}$, où pour tout $t \in \mathbb{R}$,

$$f(t) := 7\,, \qquad g(t) := \cos(2t)\,, \qquad h(t) := \cos^2(t)\,.$$

Pour savoir $\{f, g, h\}$ est libre ou liée, on considère la relation linéaire

$$\lambda_1 f + \lambda_2 g + \lambda_3 h = 0\,,$$

qui signifie

$$7\lambda_1 + \lambda_2 \cos(2t) + \lambda_3 \cos^2(t) = 0 \qquad \forall t \in \mathbb{R}\,.$$

Or si on se souvient de la relation trigonométrique

$$\cos^2(\alpha) = \frac{1 + \cos(2\alpha)}{2}\,,$$

on peut écrire

$$h(t) = \cos^2(t) = \frac{1}{2} + \frac{1}{2}\cos(2t) = \frac{1}{14} f(t) + \frac{1}{2} g(t)\,.$$

Donc $h = \frac{1}{14} f + \frac{1}{2} g$, ce qui montre que la famille $\{f, g, h\}$ est liée. $\diamond$

4.4 Sous-espaces vectoriels

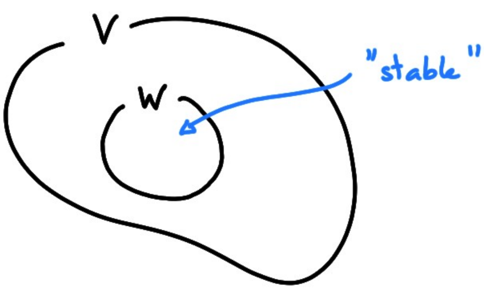
Définition 4.19. Soit V un espace vectoriel. Un sous-ensemble non vide $W \subset V$ est un **sous-espace vectoriel de** V si, lorsqu'on le munit de l'addition de V et de la multiplication par les scalaires, devient un espace vectoriel.

Proposition 2. *Un sous-ensemble non vide $W \subset V$ est un sous-espace vectoriel de V si et seulement si les trois conditions ci-dessous sont satisfaites :*

1) $0 \in W$.
2) *Si* $w \in W$ *et* $\lambda \in \mathbb{R}$, *alors* $\lambda w \in W$.
3) *Si* $w, w' \in W$, *alors* $w + w' \in W$.

Preuve: (exercice) $\square$

On dit aussi qu'un sous-espace vectoriel est un sous-ensemble de V qui est *stable* par addition et par multiplication par des scalaires. Schématiquement :



Remarque 4.20. ⋆ Soit W un sous-espace vectoriel de V. Si on fixe un $w \in W$, alors $(-1)w = -w \in W$, et donc $w + (-w) \in W$, ce qui implique $0 \in W$. (Ceci montre que la première condition de la définition est en fait superflue.)

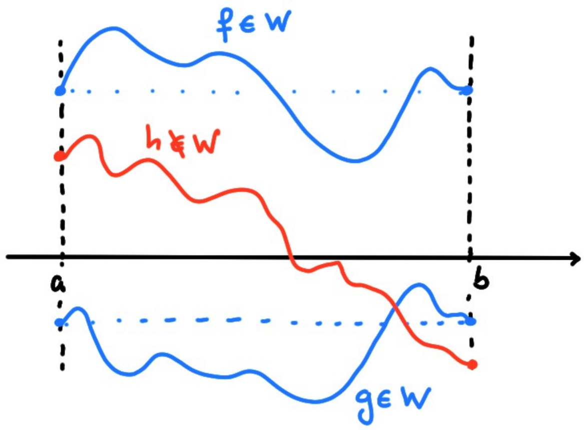
- ⋆ V, vu comme sous-ensemble de lui-même, peut être considéré comme un sous-espace vectoriel. Donc tout ce que nous dirons dans la suite peut aussi être appliqué au cas où $W = V$.

◇

Exemple 4.21. Soit V l'espace vectoriel des fonctions réelles sur l'intervalle $I = [a, b]$. Considérons

$$W := \{f \in V \mid f(a) = f(b)\}.$$

Donc les éléments de W sont les fonctions sur $[a, b]$ dont le graphe a un point initial à même hauteur que le point final :



Montrons que W est un sous-espace vectoriel de V.

1) D'abord, la fonction nulle 0 est évidemment dans W, puisque $0(a) = 0(b) = 0$.
2) Ensuite, si $f \in W$ et $\lambda \in \mathbb{R}$, alors

$$(\lambda f)(a) = \lambda f(a) = \lambda f(b) = (\lambda f)(b),$$

et donc $\lambda f \in W$.
3) Finalement, si $f, g \in W$, alors

$$(f + g)(a) = f(a) + g(a) = f(b) + g(b) = (f + g)(b),$$

et donc $f + g \in W$.

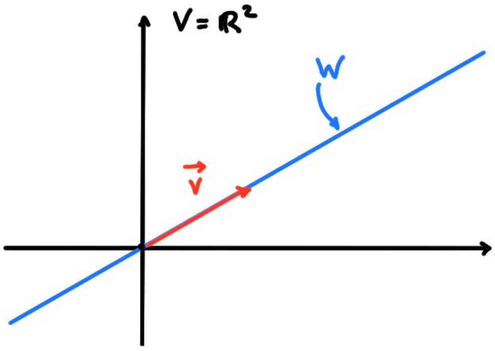
Sur le même espace vectoriel V (des fonctions réelles définies sur $[a, b]$), les sous-ensembles suivants sont aussi des sous-espaces vectoriels :

- ⋆ Les fonctions paires (si $[a, b]$ est symétrique).
- ⋆ Les fonctions impaires (si $[a, b]$ est symétrique).
- ⋆ Les fonctions continues sur $[a, b]$.
- ⋆ Les fonctions continues sur $[a, b]$, dérivables sur $]a, b[$.

◇

Exemple 4.22. Si V est l'espace de toutes les fonctions réelles définies sur $\mathbb{R}$, et si $\mathbb{P}_n$ est l'ensemble de tous les polynômes de degré au plus égal à n, alors $\mathbb{P}_n$ est un sous-espace vectoriel de V. (Voir exercices.) ◇

Exemple 4.23. Dans $V = \mathbb{R}^2$, considérons le sous-ensemble W des vecteurs situés sur la droite dirigée par $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, passant par l'origine :



Intuitivement, il est clair que cet ensemble W est "stable" : si on multiplie un vecteur de W par un scalaire, on obtient un vecteur qui est aussi dans W, et si on additionne deux vecteurs de W, alors on obtient un vecteur qui est aussi dans W : "on ne sort pas de W" en additionnant ou en multipliant par des scalaires.

Plus rigoureusement, montrons que W est un sous-espace vectoriel de $V = \mathbb{R}^2$. *Preuve*: Par définition, $W = \mathrm{Vect}\{\mathbf{v}\}$: $\mathbf{w} \in W$ si et seulement si il existe un scalaire λ tel que $\mathbf{w} = \lambda\mathbf{v}$.

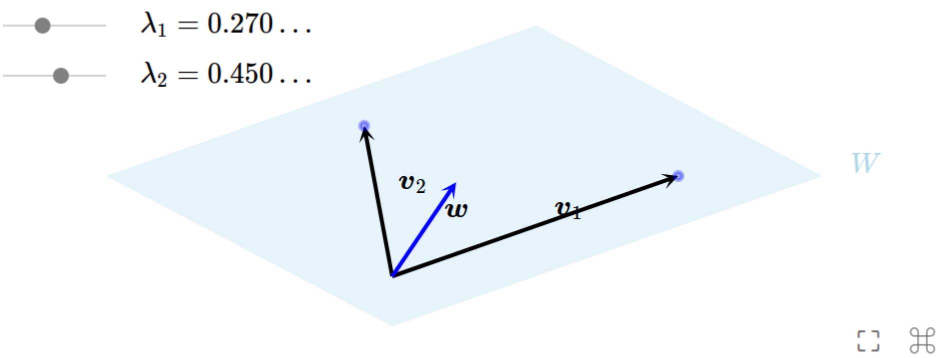
1) Le vecteur nul $\mathbf{0}$ est évidemment dans W puisque $\mathbf{0} = 0\mathbf{v}$.
2) Soit $\mathbf{w} \in W$, c'est-à-dire qu'il existe $\lambda \in \mathbb{R}$ tel que $\mathbf{w} = \lambda\mathbf{v}$, et soit $\mu \in \mathbb{R}$. Alors clairement $\mu\mathbf{w} \in W$ puisque $\mu\mathbf{w} = \mu(\lambda\mathbf{v}) = (\mu\lambda)\mathbf{v} \in W$.
3) Si $\mathbf{w}, \mathbf{w}' \in W$, alors il existe $\lambda, \lambda' \in \mathbb{R}$ tels que $\mathbf{w} = \lambda\mathbf{v}$ et $\mathbf{w}' = \lambda'\mathbf{v}$, et donc
$$\mathbf{w} + \mathbf{w}' = \lambda\mathbf{v} + \lambda'\mathbf{v} = (\lambda + \lambda')\mathbf{v}\,,$$
ce qui entraîne $\mathbf{w} + \mathbf{w}' \in W$.

◇

◇

Exemple 4.24. Dans $V = \mathbb{R}^3$, considérons le plan $W = \mathrm{Vect}\{\mathbf{v}_1, \mathbf{v}_2\}$ dirigé par les vecteurs
$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} -5 \\ 3 \\ 7 \end{pmatrix}.$$

Par définition, tout vecteur $\mathbf{w} \in W$ est de la forme
$$\mathbf{w} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2\,, \qquad \lambda_1, \lambda_2 \in \mathbb{R}\,.$$



On affirme : W est un sous-espace vectoriel de $\mathbb{R}^3$. *Preuve*: Clairement, W est formé de tous les vecteurs qui sont combinaisons linéaires de $\mathbf{v}_1, \mathbf{v}_2$, donc $\mathbf{w} \in W$ si et seulement si il existe des scalaires λ_1, λ_2 tels que
$$\mathbf{w} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2\,.$$

En d'autres termes : $W = \mathrm{Vect}\{\mathbf{v}_1, \mathbf{v}_2\}$.

1) Clairement, $\mathbf{0} \in W$ (prendre $\lambda_1 = \lambda_2 = 0$).
2) Soit $\mathbf{w} \in W$, de la forme $\mathbf{w} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2$, et soit $\mu \in \mathbb{R}$. Alors $\mu \mathbf{w} \in W$, car

$$\mu \mathbf{w} = \mu(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) = (\mu\lambda_1)\mathbf{v}_1 + (\mu\lambda_2)\mathbf{v}_2 .$$

3) Si $\mathbf{w}, \mathbf{w}' \in W$, de la forme $\mathbf{w} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2$, $\mathbf{w}' = \lambda_1' \mathbf{v}_1 + \lambda_2' \mathbf{v}_2$, alors

$$\begin{aligned} \mathbf{w} + \mathbf{w}' &= (\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) + (\lambda_1' \mathbf{v}_1 + \lambda_2' \mathbf{v}_2) \\ &= (\lambda_1 + \lambda_1')\mathbf{v}_1 + (\lambda_2 + \lambda_2')\mathbf{v}_2 , \end{aligned}$$

et donc $\mathbf{w} + \mathbf{w}' \in W$.

◇

◇

Les deux derniers exemples sont des cas particuliers d'un procédé très général permettant de construire des sous-espaces vectoriels.

Définition 4.25. Soit V un espace vectoriel, et soient $v_1, \dots, v_p$ des vecteurs de V. Le sous-ensemble de V **engendré par** $v_1, \dots, v_p$, noté $\mathrm{Vect}\{v_1, \dots, v_p\}$, est défini comme l'ensemble de toutes les combinaisons linéaires possibles des vecteurs $v_1, \dots, v_p$.

Lemme 4. $W = \mathrm{Vect}\{v_1, \dots, v_p\}$ *est un sous-espace vectoriel de* V.

Preuve: Fonctionne exactement comme les deux preuves dans les exemples ci-dessus.

1) La combinaison linéaire dont tous les coefficients sont nuls, donne l'élément nul :

$$0 = 0v_1 + \cdots + 0v_p \in W .$$

2) Si $w \in W$, $w = \lambda_1 v_p + \cdots + \lambda_p v_p$, alors pour tout scalaire $\lambda \in \mathbb{R}$,

$$\lambda w = \lambda(\lambda_1 v_1 + \cdots + \lambda_p v_p) = (\lambda\lambda_1)v_p + \cdots + (\lambda\lambda_p)v_p \in W$$

3) Si $w \in W$ et $w' \in W$,

$$\begin{aligned} w &= \lambda_1 v_1 + \cdots + \lambda_p v_p , \\ w' &= \lambda_1' v_1 + \cdots + \lambda_p' v_p , \end{aligned}$$

alors

$$w + w' = (\lambda_1 + \lambda_1')v_1 + \cdots + (\lambda_p + \lambda_p')v_p \in W .$$

□

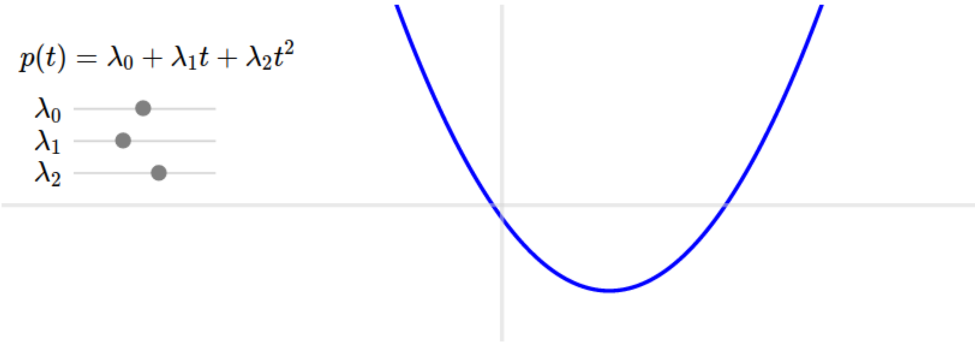
Exemple 4.26. Si V est l'espace de toutes les fonctions de $\mathbb{R}$ dans $\mathbb{R}$, et si $f_0, f_1, f_2 \in V$ sont définies par $f_k(t) := t^k$ pour tout $t \in \mathbb{R}$, alors

$$W = \mathrm{Vect}\{f_0, f_1, f_2\} = \mathbb{P}_2$$

est le sous-espace vectoriel de V contenant toutes les combinaisons linéaires

$$p = \lambda_0 f_0 + \lambda_1 f_1 + \lambda_2 f_2 ,$$

c'est-à-dire tous les polynômes p de degré plus petit ou égal à 2 :

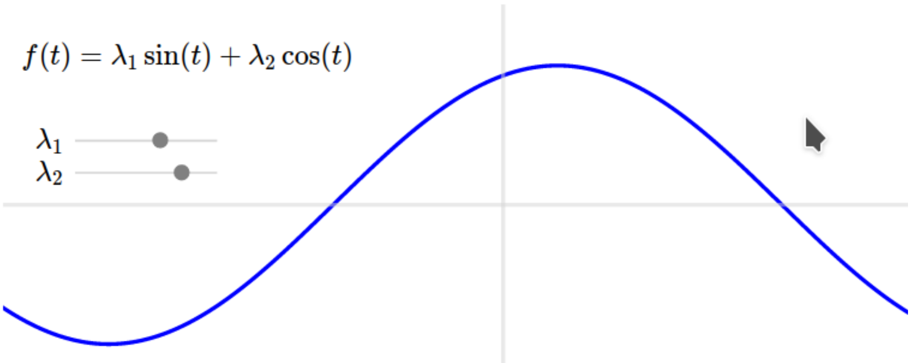


◇

Exemple 4.27. Si V est l'espace de toutes les fonctions de $\mathbb{R}$ dans $\mathbb{R}$, et si $f_1, f_2 \in V$ sont définies par

$$f_1(t) = \sin(t)\,, \qquad f_2(t) = \cos(t)\,, \qquad t \in \mathbb{R}\,,$$

alors $W = \text{Vect}\{f_1, f_2\}$ est le sous-espace vectoriel de V contenant toutes les combinaisons linéaires :



◇

4.5 Bases

Dans toute cette section, V est un espace vectoriel fixé.

Définition 4.28. Soit W un sous-espace vectoriel de V. Une famille finie de vecteurs $\mathcal{B} = \{v_1, v_2, \ldots, v_p\} \subset W$ est une **base de** W si

1) $\mathcal{B}$ est libre, et si
2) $\mathcal{B}$ engendre W, c'est-à-dire que $W = \text{Vect}\{v_1, v_2, \ldots, v_p\}$.

L'avantage d'une base $\mathcal{B} = \{v_1, v_2, \ldots, v_p\}$ est qu'elle fournit une manière *simple et univoque de représenter les vecteurs de* W.

En effet, fixons un vecteur quelconque $w \in W$. Puisque $\mathcal{B}$ engendre W, w peut s'écrire comme une combinaison linéaire des éléments de $\mathcal{B}$: il existe des scalaires $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$ tels que

$$w = \alpha_1 v_1 + \cdots + \alpha_p v_p\,.$$

Or il se trouve que ces coefficients sont uniques. En effet, supposons qu'il existe une autre famille de scalaires $\alpha'_1, \ldots, \alpha'_p \in \mathbb{R}$, telle que

$$w = \alpha'_1 v_1 + \cdots + \alpha'_p v_p\,.$$

En soustrayant ces deux dernières expressions, on obtient que

$$0 = (\alpha_1 - \alpha_1')v_1 + \cdots + (\alpha_p - \alpha_p')v_p\,.$$

Mais puisque par hypothèse, $\mathcal{B}$ est libre, ceci entraîne

$$\alpha_1 - \alpha_1' = \cdots = \alpha_p - \alpha_p' = 0\,,$$

et donc $\alpha_1 = \alpha_1'$, …, $\alpha_p = \alpha_p'$. Il n'existe donc qu'une seule manière d'exprimer w comme combinaison linéaire des vecteurs de $\mathcal{B}$.

Définition 4.29. Les scalaires $\alpha_1, \dots, \alpha_p$ définis ci-dessus sont les **composantes de** w **relativement à la base** $\mathcal{B}$.

Informel 4.30. Attention : les composantes sont des nombres que l'on peut utiliser pour décrire un vecteur, mais le vecteur **existait**, avant qu'on ne connaisse ses composantes, avant même qu'on ne parle de base !

Représentation en composantes

D'un côté, un vecteur $w \in W$ et un objet abstrait. De l'autre, sa représentation dans la base $\mathcal{B}$, à l'aide des nombres $\alpha_1, \dots, \alpha_p$, en fait un objet avec lequel on peut *faire des calculs*. En effet, on peut stocker ces nombres dans un vecteur de $\mathbb{R}^p$, en définissant

$$[w]_{\mathcal{B}} := \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix}$$

Remarque 4.31. L'*ordre* dans lequel on stocke les $\alpha_1, \dots, \alpha_p$ est important. En effet, la $k$ème composante α_k est associée au $k$ème vecteur de la base, v_k. Il est donc important, quand on introduit une base, de *fixer l'ordre de ses vecteurs*. Donc pour indiquer que les vecteurs de $\mathcal{B}$ sont ordonnés, on écrira dorénavant

$$\mathcal{B} = (v_1, \dots, v_p)\,,$$

qui est une famille ordonnée , au lieu de

$$\mathcal{B} = \{v_1, \dots, v_p\}\,.$$

◇

Insistons sur le fait que le vecteur $[w]_{\mathcal{B}} \in \mathbb{R}^p$ contient *exactement la même information que* w (il **représente** w), puisque w peut toujours être reconstruit exactement à l'aide des composantes de $[w]_{\mathcal{B}}$:

$$\alpha_1 v_1 + \cdots + \alpha_p v_p = w\,.$$

Ceci implique que finalement, dès qu'on est en possession d'une base dans un sous-espace vectoriel, aussi abstrait soit-il, ses vecteurs peuvent être traités comme des vecteurs de $\mathbb{R}^p$!

Dans le cas où $W = V$, la définition ci-dessus revient donc à définir une **base de** V comme étant une famille libre qui engendre V.

Exemple 4.32. Considérons $V = \mathbb{R}^n$. Rappelons que l'on peut écrire tout vecteur $\mathbf{x} \in \mathbb{R}^n$ comme

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 \cdots + x_n\mathbf{e}_n \,,$$

où

$$\mathbf{e}_1 := \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \quad \mathbf{e}_2 := \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, \quad \dots \quad \mathbf{e}_n := \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}.$$

Puisque cette famille de vecteurs est libre , on conclut que $\mathcal{B}_{\text{can}} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ est bien une base, la **base canonique de** $\mathbb{R}^n$. ◇

Exemple 4.33. Dans $V = \mathbb{R}^2$, considérons les vecteurs

$$\mathbf{v}_1 = \begin{pmatrix}2\\1\end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix}-7\\3\end{pmatrix},$$

et montrons que $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2)$ est une base de V. D'abord, on voit que $\mathbf{v}_1$ et $\mathbf{v}_2$ sont ne sont pas colinéaires, et donc que $\mathcal{B}$ est libre. Ensuite, pour montrer qu'elle engendre bien tout V, fixons un $\mathbf{x} \in V$ quelconque, et montrons qu'il peut s'écrire comme combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$, c'est-à-dire qu'il existe des scalaires λ_1 et λ_2 tels que

$$\mathbf{x} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 \,.$$

Si on nome x_1, x_2 les composantes de $\mathbf{x}$, alors cette dernière devient

$$\begin{pmatrix}x_1\\x_2\end{pmatrix} = \lambda_1\begin{pmatrix}2\\1\end{pmatrix} + \lambda_2\begin{pmatrix}-7\\3\end{pmatrix},$$

qui n'est autre que

$$(*)\left\{\begin{array}{rcrcl} 2\lambda_1 & - & 7\lambda_2 & = & x_1 \\ \lambda_1 & + & 3\lambda_2 & = & x_2 \end{array}\right.$$

Après $L_2 \leftarrow L_2 - \frac{1}{2}L_1$,

$$(*)\left\{\begin{array}{rcrcl} 2\lambda_1 & - & 7\lambda_2 & = & x_1 \\ & & \frac{13}{2}\lambda_2 & = & x_2 - \frac{1}{2}x_1 \end{array}\right.$$

En procédant "du bas vers le haut", on trouve

$$\lambda_1 = \tfrac{3}{13}x_1 + \tfrac{7}{13}x_2 \,, \qquad \lambda_2 = -\tfrac{1}{13}x_1 + \tfrac{2}{13}x_2 \,.$$

Ceci montre que $\mathbf{x}$ peut effectivement s'écrire comme combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$. Comme ceci vaut pour tout $\mathbf{x} \in V$, $\mathcal{B}$ engendre bien V.

On a donc montré que $\mathcal{B}$ est une base de V. ◇

Exemple 4.34. Considérons $V = \mathbb{P}_n$, l'ensemble des polynômes à coefficients réels, de degré au plus égal à n. Rappelons que tout élément $p \in \mathbb{P}_n$ est de la forme

$$p(t) = a_0 + a_1t + a_2t^2 + \cdots + a_nt^n \,, \qquad t \in \mathbb{R}\,.$$

Considérons les polynômes $e_0, e_1, \dots, e_n \in \mathbb{P}_n$ définis ainsi : pour tout $t \in \mathbb{R}$,

$$\begin{aligned} e_0(t) &:= 1\,,\\ e_1(t) &:= t\,,\\ e_2(t) &:= t^2\,,\\ &\vdots\\ e_n(t) &:= t^n\,. \end{aligned}$$

Pour le polynôme écrit au-dessus,

$$p = a_0 e_0 + a_1 e_1 + \cdots + a_n e_n \,.$$

Donc la famille $\{e_0, e_1, \ldots, e_p\}$ engendre $\mathbb{P}_n$. Mais on a aussi montré dans une section précédente que cette famille est libre. Ainsi, la famille $\mathcal{B}_{\text{can}} = (e_0, e_1, \ldots, e_n)$ forme une base, appelée **base canonique de** $\mathbb{P}_n$.
Avec la base canonique $\mathcal{B}_{\text{can}}$, l'application $[\cdot]_\mathcal{B}$ associe au polynôme p du dessus le vecteur de $\mathbb{R}^{n+1}$ défini par

$$[p]_\mathcal{B} = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} \,.$$

On peut alors manipuler le polynôme p à l'aide de sa représentation sous la forme $[p]_\mathcal{B}$, exactement comme si c'était un vecteur de $\mathbb{R}^{n+1}$! ◇

Exemple 4.35. Considérons, dans $V = \mathbb{P}_2$, la famille $\mathcal{B} = (q_1, q_2, q_2)$, où

$$q_1(t) = 3\,, \qquad q_2(t) = 1 - 2t\,, \qquad q_3(t) = t^2 + t$$

Montrons que $\mathcal{B}$ est une base de V. Pour commencer, montrons que $\mathcal{B}$ est libre, en posant

$$\lambda_1 q_1 + \lambda_2 q_2 + \lambda_3 q_3 = 0\,,$$

qui signifie, après avoir regroupé les termes,

$$(3\lambda_1 + \lambda_2) + (-2\lambda_2 + \lambda_3)t + \lambda_3 t^2 = 0 \qquad \forall t \in \mathbb{R}\,.$$

On sait qu'un polynôme s'annule en tout $t \in \mathbb{R}$ si et seulement si tous ses coefficients sont nuls. On en déduit que $\lambda_3 = 0$, puis que $\lambda_2 = \frac{1}{2}\lambda_3 = 0$, puis que $\lambda_1 = -\frac{1}{3}\lambda_2 = 0$. Ceci montre que $\mathcal{B}$ est libre.

Montrons ensuite que $\mathcal{B}$ engendre $\mathbb{P}_2$. Pour ce faire, fixons un $p \in \mathbb{P}_2$ quelconque, et montrons qu'on peut trouver des scalaires α_1, α_2 tels que

$$p = \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_3 \,.$$

Si $p(t) = a + bt + ct^2$, cela signifie que

$$a + bt + ct^2 = \alpha_1 3 + \alpha_2(1 - 2t) + \alpha_3(t^2 + t) \qquad \forall t \in \mathbb{R}\,,$$

qui devient, après avoir regroupé les termes,

$$(3\alpha_1 + \alpha_2 - a) + (-2\alpha_2 + \alpha_3 - b)t + (\alpha_3 - c)t^2 = 0 \qquad \forall t \in \mathbb{R}\,.$$

On voit donc que $\alpha_1, \alpha_2, \alpha_3$ doivent satisfaire

$$(*) \left\{ \begin{array}{ccccccc} 3\alpha_1 & + & \alpha_2 & & & = & a \\ & - & 2\alpha_2 & + & \alpha_3 & = & b \\ & & & & \alpha_3 & = & c \end{array} \right.$$

On trouve

$$\alpha_3 = c\,, \qquad \alpha_2 = \tfrac{1}{2}(c - b)\,, \qquad \alpha_3 = \tfrac{1}{3}a + \tfrac{1}{6}b - \tfrac{1}{6}c\,.$$

Ceci montre que $\mathcal{B}$ engendre $\mathbb{P}_2$.

Donc on a bien montré que $\mathcal{B}$ est une base de $\mathbb{P}_2$. ◇

Extraire une base d'une famille génératrice

Supposons qu'un sous-espace $W \subset V$ soit engendré par une famille :

$$W = \text{Vect}\{v_1, \dots, v_p\} .$$

Par définition, tout vecteur $w \in W$ peut s'écrire comme combinaison linéaire des $v_1, \dots, v_p$, mais cela ne signifie pas que ces vecteurs forment une base pour W : il se peut que certains ne soient pas nécessaires dans la description de W ; en d'autres termes, cette famille peut contenir "trop" de vecteurs, certains de ses vecteurs peuvent être superflus.

Lemme 5. *Soit W un sous-espace vectoriel de V, et soit*

$$\mathcal{F} = \{w_1, \dots, w_r\}$$

une famille qui engendre W. Si un des vecteurs de $\mathcal{F}$, disons w_j, peut s'écrire comme combinaison linéaire des autres w_k ($k \neq j$), alors en retirant w_j, la famille

$$\mathcal{F} \setminus \{w_j\} = \{w_1, \dots, w_{j-1}, w_{j+1}, \dots, w_r\}$$

engendre toujours W.

Preuve: Puisque $\mathcal{F}$ engendre W, tout vecteur $w \in W$ peut s'écrire

$$w = a_1 w_1 + \cdots + a_r w_r .$$

Si $w_j = \sum_{i \neq j} \alpha_i w_i$, alors

$$\begin{aligned} w = \sum_{i=1}^{r} a_i w_i = \Big(\sum_{\substack{i=1 \\ i \neq j}}^{r} a_i w_i \Big) + a_j w_j &= \sum_{\substack{i=1 \\ i \neq j}}^{r} a_i w_i + a_j \sum_{\substack{i=1 \\ i \neq j}}^{r} \alpha_i w_i \\ &= \sum_{\substack{i=1 \\ i \neq j}}^{r} (a_i + a_j \alpha_i) w_i , \end{aligned}$$

donc w peut s'écrire comme combinaison linéaire des éléments de $\mathcal{F} \setminus \{w_j\}$. Ceci signifie que la famille $\mathcal{F} \setminus \{w_j\}$ engendre aussi W. □

Ce dernier résultat fournit un algorithme pour construire une base d'un sous-espace W, du moment que l'on possède une famille génératrice.

En effet, supposons qu'une famille finie $\mathcal{F} = \{v_1, \dots, v_r\}$ engendre W. On peut "nettoyer" cette famille à l'aide du lemme précédent, de la façon suivante :

1) Chercher un vecteur $v_j \in \mathcal{F}$ qui peut s'exprimer comme combinaison linéaire des autres.
2) S'il y en a un, retirer v_j de la famille, et recommencer. S'il n'y en a pas, s'arrêter.

Une fois que cet algorithme s'arrête, on obtient une famille $\mathcal{F}' \subset \mathcal{F}$ qui engendre toujours W, et dans laquelle aucun vecteur ne peut s'exprimer comme combinaison linéaire des autres; c'est donc une base de W.

Exemple 4.36. Soit $V = \mathbb{R}^4$, et soit $W \subset V$ le sous-espace défini par

$$W = \text{Vect}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\} ,$$

où

$$\mathbf{w}_1 = \begin{pmatrix} 3 \\ -1 \\ 0 \\ 2 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} 5 \\ -4 \\ -4 \\ 3 \end{pmatrix}, \qquad \mathbf{w}_3 = \begin{pmatrix} 1 \\ 2 \\ 4 \\ 1 \end{pmatrix}.$$

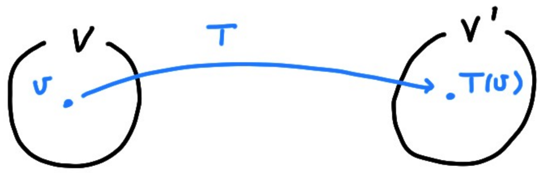
Remarquons que $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ n'est pas libre puisque $\mathbf{w}_2 = 2\mathbf{w}_1 - \mathbf{w}_3$. Donc $\mathbf{w}_2$ est "superflu", et on peut le retirer, sans changer W :

$$W = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_3\}.$$

Maintenant, $\mathbf{w}_1$ et $\mathbf{w}_3$ n'étant pas colinéaires, $\mathcal{B} := (\mathbf{w}_1, \mathbf{w}_3)$ est une base de W. ◇

4.6 Applications linéaires

Dans cette section, nous généralisons la notion d'application linéaire, au cas d'une application d'un espace vectoriel V (de départ) dans un espace vectoriel V' (d'arrivée) :



Rappelons que pour $v \in V$, l'élément $v' = T(v) \in V'$ est appelé **l'image de** v, et v est une **préimage de** v'.

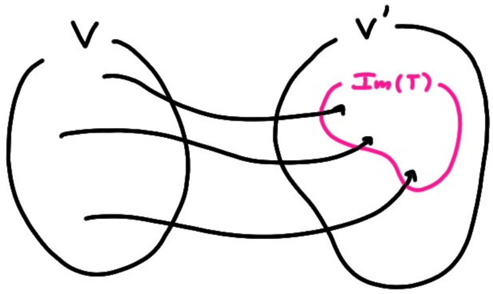
Étant des espaces vectoriels, V et V' possèdent chacun un zéro ; on les notera $0 \in V$ et $0' \in V'$ pour les distinguer. Par contre, l'addition dans ces espaces sera toujours notée "+" pour ne pas trop allourdir les notations .

Généralités

Rappelons rapidement, dans ce cadre général, quelques notions élémentaires de la théorie des fonctions $T : V \to V'$. (Si nécessaire, on pourra aller voir **ici** (lien web), pour d'autres exemples à propos de ces notions.)

Définition 4.37. L'**ensemble image** d'une application $T : V \to V'$ est défini par l'ensemble des éléments de l'ensemble d'arrivée qui possèdent au moins une préimage :

$$\mathrm{Im}(T) := \{v' \in V' \,|\, \exists\, v \in V \text{ tel que } T(v) = v'\}.$$

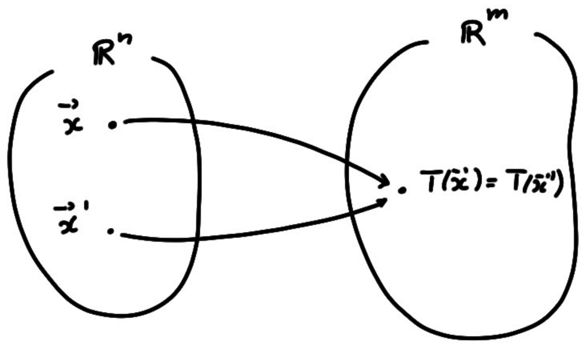


Définition 4.38. Une application $T : V \to V'$ est

- **surjective** si tout élément de l'ensemble d'arrivée V' possède au moins une préimage, c'est-à-dire si $\mathrm{Im}(T) = V'$.
- **injective** si des éléments distincts ont des images distinctes, c'est-à-dire si $v_1 \neq v_2$ implique $T(v_1) \neq T(v_2)$.
- **bijective** si elle est à la fois injective et surjective.

Remarquons que :

- Lorsque $T : V \to V'$ est surjective, alors pour tout $b \in V'$, l'équation
$$T(v) = b$$
possède au moins une solution $v \in V$.
- Une application n'est pas injective si il existe au moins une paire de vecteurs, distincts ($v \neq v'$), tels que $T(v) = T(v')$:



Lorsque $T : V \to V'$ est injective, alors pour tout $b \in V'$, si l'équation
$$T(v) = b$$
possède une solution $v \in V$, cette solution est unique. (En effet, s'il y avait deux solutions, $v_1, v_2 \in V$, alors $T(v_1) = T(v_2) = b$, qui par l'injectivité implique $v_1 = v_2$.)

- L'intérêt d'une application $T : V \to V'$ bijective est que l'on peut l'*inverser*. En effet, fixons un $v' \in V'$. Puisque T est surjective, v' possède au moins une préimage : il existe un $v_* \in V$ tel que
$$T(v_*) = v' .$$
Mais comme T est aussi injective, il ne peut pas exister, à part v_*, d'autre vecteur dont l'image soit égale à v'.

Par ce procédé, on associe à tout $v' \in V'$ un unique $v_* \in V$ tel que $T(v_*) = v'$. L'application qui à chaque v' associe son unique préimage v_* est appelée la **réciproque** de T. On la notera
$$\begin{aligned} T^{-1} : V' &\to V \\ v' &\mapsto T^{-1}(v') := v_* . \end{aligned}$$

Applications linéaires ; définition

Généralisons maintenant la notion d'application linéaire, que nous avions précédemment définie seulement dans le cas $T : \mathbb{R}^n \to \mathbb{R}^m$:

Définition 4.39. Soient V et V' des espaces vectoriels. Une application $T : V \to V'$ est **linéaire** si les deux conditions suivantes sont satisfaites :

1) $T(v_1 + v_2) = T(v_1) + T(v_2)$ pour tous $v_1, v_2 \in V$,
2) $T(\lambda v) = \lambda T(v)$ pour tout $v \in V$ et tout scalaire $\lambda \in \mathbb{R}$.

Remarque 4.40. On peut mettre les deux conditions de la définition en une seule : une application $T : V \to V'$ est **linéaire** si pour tous $v_1, v_2 \in V$, et pour tous scalaires $\alpha, \beta, \in \mathbb{R}$,

$$T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2)\,.$$

◇

Nous avons déjà vu plusieurs exemples d'applications linéaires dans le cas $T : \mathbb{R}^n \to \mathbb{R}^m$. Rappelons le plus important :

Exemple 4.41. Si A est une matrice réelle $m \times n$, alors l'application

$$\begin{aligned} T : \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{x} &\mapsto T(\mathbf{x}) := A\mathbf{x} \end{aligned}$$

est linéaire. ◇

Exemple 4.42. Soit V l'espace des fonctions $f : [a, b] \to \mathbb{R}$, soit $V' = \mathbb{R}^2$, et soit $T : V \to V'$ définie ainsi : pour tout $f \in V$,

$$T(f) := \begin{pmatrix} f(a) \\ f(b) \end{pmatrix} .$$

Alors T est linéaire. En effet, si $f, g \in V$, $\alpha, \beta \in \mathbb{R}$, alors

$$\begin{aligned} T(\alpha f + \beta g) &= \begin{pmatrix} (\alpha f + \beta g)(a) \\ (\alpha f + \beta g)(b) \end{pmatrix} \\ &= \begin{pmatrix} \alpha f(a) + \beta g(a) \\ \alpha f(b) + \beta g(b) \end{pmatrix} \\ &= \alpha \begin{pmatrix} f(a) \\ f(b) \end{pmatrix} + \beta \begin{pmatrix} g(a) \\ g(b) \end{pmatrix} \\ &= \alpha T(f) + \beta T(g)\,. \end{aligned}$$

◇

Exemple 4.43. Soit $V = C([a, b])$ l'espace des fonctions (à valeurs réelles) continues sur $[a, b]$, et soit $V' = \mathbb{R}^2$. Soit $c \in]a, b[$ un point fixé et soit $T : V \to V'$ définie ainsi : pour tout $f \in V$,

$$T(f) := \begin{pmatrix} \displaystyle\int_a^c f(t)\,dt \\ \displaystyle\int_c^b f(t)\,dt \end{pmatrix} .$$

Alors T est linéaire et surjective. (voir exercices) ◇

Linéarité de l'application "composantes"

On peut maintenant en dire plus sur l'application fondamentale :

Lemme 6. *Soit $\mathcal{B} = (v_1, \dots, v_p)$ une base d'un sous-espace vectoriel W. L'application $[\cdot]_{\mathcal{B}}$, qui associe à w le vecteur de $\mathbb{R}^p$ formé des composantes de w relativement à $\mathcal{B}$,*

$$\begin{aligned} [\cdot]_{\mathcal{B}} : W &\to \mathbb{R}^p \\ w &\mapsto [w]_{\mathcal{B}} \end{aligned}$$

est linéaire et bijective.

Preuve: En exercice. □

Noyau

Lorsqu'une application $T : V \to V'$ est linéaire, plusieurs choses peuvent être dites à son sujet.

Par exemple, la linéarité implique que le zéro est toujours envoyé sur le zéro :

$$T(0) = 0' .$$

En effet, en écrivant $0 = 0 + 0$ et en utilisant la linéarité,

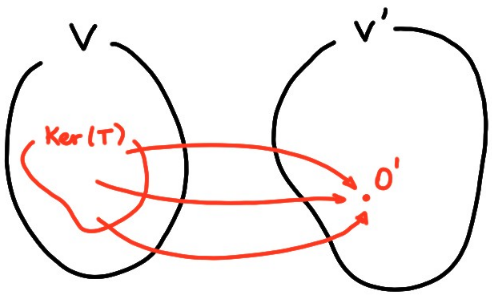
$$T(0) = T(0 + 0) = T(0) + T(0) = (1 + 1)T(0) = 2T(0) ,$$

ce qui implique bien que $T(0) = 0'$.

Si 0 est toujours envoyé sur $0'$, il se pourrait aussi que d'autres éléments de V soient aussi envoyés sur $0'$:

Définition 4.44. Le **noyau** d'une application $T : V \to V'$ est l'ensemble de toutes les préimages de $0'$:

$$\ker(T) := \{v \in V \,|\, T(v) = 0'\} .$$



On a vu plus haut que le noyau contient toujours le zéro de V. On peut en dire un peu plus :

Lemme 7. *Une application linéaire $T : V \to V'$ est injective si et seulement si son noyau ne contient que le zéro :* $\ker(T) = \{0\}$.

Preuve: Supposons d'abord que T est injective. Considérons un $v \in \ker(T)$, c'est-à-dire tel que $T(v) = 0'$. Comme on sait que $T(0) = 0'$, on a donc $T(v) = T(0)$, et l'injectivité implique que $v = 0$. Donc $\ker(T) = \{0\}$.

Supposons maintenant que $\ker(T) = \{0\}$. Considérons $v_1, v_2 \in V$ tels que $T(v_1) = T(v_2)$. Par linéarité, ceci implique $T(v_1 - v_2) = 0'$, et donc $v_1 - v_2 \in \ker(T)$, et donc $v_1 - v_2 = 0$, ce qui implique $v_1 = v_2$. Donc T est injective. □

Exemple 4.45. Soit $T : V \to \mathbb{R}^2$ l'application linéaire définie plus haut ; pour $f : [a,b] \to \mathbb{R}$,

$$T(f) := \begin{pmatrix} f(a) \\ f(b) \end{pmatrix} .$$

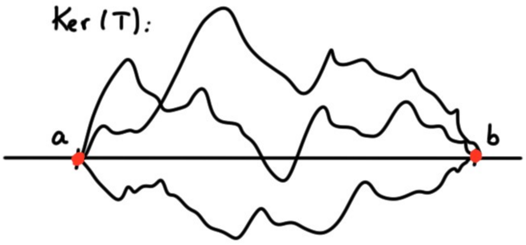
Le noyau de cette application est formé de toutes les fonctions f pour lesquelles

$$T(f) = \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} ,$$

c'est-à-dire

$$\ker(T) = \{f : [a,b] \to \mathbb{R} \,:\, f(a) = f(b) = 0\} .$$

Ce noyau contient en particulier la fonction identiquement nulle bien-sûr, mais aussi une infinité de fonctions non-nulles :



Donc T n'est pas injective. ◇

Finalement, notons que le noyau et l'image sont des sous-ensembles stables de V et V', respectivement :

Lemme 8. *Si $T : V \to V'$ est linéaire, alors*

- ⋆ $\ker(T)$ *est un sous-espace vectoriel de* V.
- ⋆ $\mathrm{Im}(T)$ *est un sous-espace vectoriel de* V'.

Preuve: Commençons par le noyau :

- ⋆ On a vu que $T(0) = 0'$, ce qui signifie que $0 \in \ker(T)$.
- ⋆ Si $v_1, v_2 \in \ker(T)$, et $\alpha, \beta \in \mathbb{R}$, alors la linéarité de T implique

$$T(\alpha v_1 + \beta v_2) = \alpha \underbrace{T(v_1)}_{=0'} + \beta \underbrace{T(v_2)}_{=0'} = 0' ,$$

et donc $\alpha v_1 + \beta v_2 \in \ker(T)$.

Pour $\mathrm{Im}(T)$, voir les exercices. □

Lemme 9. *Si $T : V \to V'$ est linéaire et bijective, alors sa réciproque $T^{-1} : V' \to V$ est aussi linéaire.*

Preuve: (Voir exercices.) □

Images de bases et bijections

Théorème 4.46. *Soit $T : V \to V'$ une application linéaire. Alors T est bijective si et seulement si l'image d'une base $\mathcal{B} = (v_1, \dots, v_p)$ de V, notée $T(\mathcal{B}) = (T(v_1), \dots, T(v_p))$, est une base de V'.*

Preuve: Supposons que T est bijective, et considérons une base de V, $\mathcal{B} = (v_1, \dots, v_p)$. Montrons que $T(\mathcal{B}) := (T(v_1), \dots, T(v_p))$ est une base.

- $\star$ Soit v' un vecteur quelconque de V'. Puisque T est surjective, il existe $v \in V$ tel que $v' = T(v)$. Si on décompose $v = \alpha_1 v_1 + \cdots + \alpha_p v_p$, on a
$$v' = T(v) = \alpha_1 T(v_1) + \cdots + \alpha_p T(v_p)\,.$$
Ceci signifie que $T(\mathcal{B})$ engendre V'.
- $\star$ Considérons une combinaison linéaire nulle,
$$0 = \lambda_1 T(v_1) + \cdots + \lambda_p T(v_p) = T(\alpha_1 v_1 + \cdots + \alpha_p v_p)\,.$$
Puisque T est injective, on a $\alpha_1 v_1 + \cdots + \alpha_p v_p = 0$, et comme $\mathcal{B}$ est une base, on en déduit que $\alpha_1 = \cdots = \alpha_p = 0$. Donc $T(\mathcal{B})$ est libre.

Supposons ensuite que $\mathcal{B} = (v_1, \dots, v_p)$ est une base de V, et que son image $T(\mathcal{B}) := (T(v_1), \dots, T(v_p))$ est une base de V'.

- $\star$ Soit $v \in V$ tel que $T(v) = 0$. Puisqu'on peut décomposer v sur la base $\mathcal{B}$, $v = \alpha_1 v_1 + \cdots + \alpha_p v_p$, on a
$$0 = T(v) = \alpha_1 T(v_1) + \cdots + \alpha_p T(v_p)\,.$$
Puisque $T(\mathcal{B})$ est une base, $\alpha_1 = \cdots = \alpha_p = 0$. Donc $v = 0$, ce qui implique que T est injective.
- $\star$ Soit $v' \in V'$, que l'on peut décomposer sur $T(\mathcal{B})$:
$$v' = \lambda_1 T(v_1) + \cdots + \lambda_p T(v_p) = T(\lambda_1 v_1 + \cdots + \lambda_p v_p)\,,$$
qui implique que $v' \in \mathrm{Im}(T)$. Donc T eset surjective.

Donc T est bijective. □

4.7 Dimension

C'est à l'aide de la notion de *base* que l'on définit naturellement celle de *dimension*.

Dans cette section, nous définirons la dimension pour un sous-espace vectoriel $W \subset V$, en gardant à l'esprit que tout ce qui est dit est aussi valable dans le cas où $W = V$.

Commençons par voir une première conséquence de l'existence d'une base :

Lemme 10. *Si $\mathcal{B} = (v_1, \dots, v_p)$ est une base d'un sous-espace vectoriel W, et si $\mathcal{F} \subset W$ est une famille contenant plus de vecteurs que $\mathcal{B}$ (c'est-à-dire plus de p vecteurs), alors $\mathcal{F}$ est liée.*

Preuve: Le résultat va suivre de ce que nous avons vu dans un chapitre précédent : dans $\mathbb{R}^p$, toute famille de plus de p vecteurs est liée.

Écrivons $\mathcal{F} = \{w_1, \dots, w_k\} \subset W$, avec $k > p$. Considérons la relation linéaire
$$(*)_1 : \qquad \alpha_1 w_1 + \cdots + \alpha_k w_k = 0\,.$$
Appliquons $[\cdot]_{\mathcal{B}}$ des deux côtés de cette relation. Par linéarité, et comme $[0]_{\mathcal{B}} = \mathbf{0}$, on a
$$(*)_2 : \qquad \alpha_1 [w_1]_{\mathcal{B}} + \cdots + \alpha_k [w_k]_{\mathcal{B}} = \mathbf{0}\,.$$
Comme $\{[w_1]_{\mathcal{B}}, \dots, [w_k]_{\mathcal{B}}\}$ est une famille de vecteurs de $\mathbb{R}^p$, on sait qu'elle est liée puisque $k > p$. On conclut qu'il existe une famille de coefficients $\alpha_1, \dots, \alpha_k$, non tous nuls, tels que $(*)_2$ soit vérifiée. Puisque $[\cdot]_{\mathcal{B}}$ est linéaire et inversible, sa réciproque $[\cdot]_{\mathcal{B}}^{-1}$ est aussi linéaire (voir lemme de la section précédente). Donc en appliquant $[\cdot]_{\mathcal{B}}^{-1}$ des deux côtés de $(*)_2$, on récupère $(*)_1$, qui est donc vérifiée pour les mêmes coefficients α_j, ce qui implique que $\mathcal{F}$ est liée. □

Ainsi, si $\mathcal{B}$ est une base de W, on sait qu'une famille libre dans W ne peut pas contenir plus de vecteurs que le nombre de vecteurs contenus dans $\mathcal{B}$. Ceci implique aussi :

Corollaire 2. *Toutes les bases d'un même sous-espace vectoriel W contiennent le même nombre d'éléments.*

Preuve: Soient $\mathcal{B} = (w_1, \dots, w_p)$ et $\mathcal{B}' = (w_1', \dots, w_n')$ deux bases de W. Si on suppose que $p > n$, alors le lemme précédent implique que $\mathcal{B}$ est liée, ce qui n'est pas possible puisque $\mathcal{B}$ est une base ; on conclut que $p \leqslant n$. De même, si on suppose que $n > p$, alors le lemme précédent implique que $\mathcal{B}'$ est liée, ce qui n'est pas possible puisque $\mathcal{B}'$ est une base ; on conclut que $n \leqslant p$. On a donc $p = n$. □

Puisque toutes les bases d'un espace ont le même nombre d'éléments, ce nombre décrit une propriété intrinsèque de cet espace :

Définition 4.47. Si un W possède une base contenant un nombre fini n de vecteurs, on dit que W est **de dimension finie**, et que sa **dimension** est égale à n, ce que l'on note comme suit : $\dim(W) = n$.

Dans le cas d'un espace vectoriel V, la **dimension de** V est donc le nombre d'éléments de n'importe quelle base de V.

Exemple 4.48. Dans $\mathbb{R}^3$, considérons le sous-espace $W = \mathrm{Vect}\{\mathbf{v}_1, \mathbf{v}_2\}$, où

$$\mathbf{v}_1 = \begin{pmatrix} 4 \\ 1 \\ -3 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix}.$$

Puisque $\mathbf{v}_1$ et $\mathbf{v}_2$ ne sont pas colinéaires, et qu'ils engendrent W, on en déduit que $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2)$ est une base de W. Ainsi, $\dim(W) = 2$, c'est un **plan**. ◇

Exemple 4.49. Considérons $V = \mathbb{R}^n$. Comme la base canonique $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ est formée de n vecteurs, n'importe quelle autre base doit aussi avoir n vecteurs, et donc

$$\dim(\mathbb{R}^n) = n\,.$$

◇

Exemple 4.50. Considérons $V = \mathbb{P}^n$. Comme base la base canonique $(e_0, \dots, e_n)$ est formée de $n+1$ vecteurs, n'importe quelle autre base doit aussi avoir $n+1$ vecteurs, et donc

$$\dim(\mathbb{P}^n) = n + 1\,.$$

◇

Remarque 4.51. Il existe des espaces vectoriels, comme par exemple l'espace de toutes les fonctions $f : [a, b] \to \mathbb{R}$, qui ne sont *pas* de dimension finie : il n'existe aucune famille finie $(f_1, \dots, f_n)$ telle que toute fonction puisse s'écrire comme combinaison linéaire de $f_1, \dots, f_n$. On dit que cet espace est **de dimension infinie**. ◇

Théorème 4.52. *Dans un sous-espace vectoriel W de dimension n, toute famille libre contenant n vecteurs est une base de W.*

Preuve: Supposons que $\mathcal{F} \subset W$, $\mathcal{F} = (v_1, \dots, v_n)$, est libre. Prenons un $w \in W$, et définissons $\mathcal{F}' := \mathcal{F} \cup \{w\}$. Par le résultat du dessus, $\mathcal{F}'$ contenant $n+1$ vecteur, elle est liée : il existe $\lambda_1, \dots, \lambda_{n+1}$, pas tous nuls, tels que

$$\lambda_1 v_1 + \cdots + \lambda_n v_n + \lambda_{n+1} w = 0\,.$$

Si $\lambda_{n+1} = 0$, cela signifie qu'au moins un des $\lambda_1, \dots, \lambda_n$ est non-nul, et que

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0\,,$$

et donc que $\mathcal{F}$ est liée, une contradiction. On en conclut que $\lambda_{n+1} \neq 0$, ce qui permet d'écrire w comme combinaison linéaire des éléments de $\mathcal{F}$:

$$w = -\frac{\lambda_1}{\lambda_{n+1}} v_1 - \cdots - \frac{\lambda_n}{\lambda_{n+1}} v_n\,.$$

Donc $\mathcal{F}$ est bien une base de W. □

Construire une base par complétion

Théorème 4.53. *Soit W un sous-espace vectoriel de dimension finie n, et soit*

$$\mathcal{F} = (w_1, \dots, w_r) \subset W$$

*une famille libre, $r < n$. Alors $\mathcal{F}$ peut être **complétée** en une base de W. Plus précisément, il existe des vecteurs $w'_{r+1}, \dots, w'_n$ de W tels que*

$$\mathcal{F}' = (w_1, \dots, w_r, w'_{r+1}, \dots, w'_n)$$

soit une base de W.

Preuve: Puisque par hypothèse ($r < n$) $\mathcal{F}$ n'engendre pas W, il existe au moins un $w'_{r+1} \in W$ qui ne peut pas s'écrire comme combinaison linéaire d'éléments de $\mathcal{F}$. On conclut que

$$(w_1, \dots, w_r, w'_{r+1})$$

est libre. Si cette famille n'engendre toujours pas W, on recommence : il doit exister un $w'_{r+2} \in W$ qui ne peut pas s'écrire comme combinaison linéaire de ses éléments, et donc

$$(w_1, \dots, w_r, w'_{r+1}, w'_{r+2})$$

est libre, etc. Puisque la dimension de W est finie et vaut n, ce procédé continue jusqu'à obtenir une famille libre qui contient exactement n éléments, et qui forme donc une base de W. □

Exemple 4.54. Considérons la famille libre $\mathcal{F} = (\mathbf{v}_1, \mathbf{v}_2) \subset \mathbb{R}^3$, où

$$\mathbf{v}_1 = \begin{pmatrix} 4 \\ 0 \\ -3 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix}.$$

Clairement, $\mathcal{F}$ est libre, mais elle n'engendre pas $\mathbb{R}^3$ (car $2 < 3$!). Par le théorème ci-dessus, on peut compléter $\mathcal{F}$ en une base de $\mathbb{R}^3$, en lui rajoutant un vecteur qui n'est pas une combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$. Comment choisir ce vecteur ?

Remarquons que toute combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$ est de la forme

$$\alpha_1 \begin{pmatrix} 4 \\ 0 \\ -3 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 4\alpha_1 \\ 3\alpha_2 \\ -3\alpha_1 + 2\alpha_2 \end{pmatrix}$$

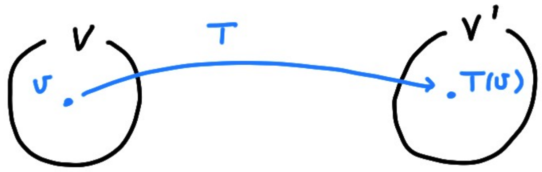
On peut donc prendre n'importe quel vecteur qui n'est pas de cette forme. Par exemple

$$\mathbf{v}_3 = \begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix}.$$

Maintenant, $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ est une base de $\mathbb{R}^3$. ◇

4.8 Matrice d'une application

Considérons deux espaces vectoriels, V et V', ainsi qu'une application linéaire $T : V \to V'$.



Supposons maintenant que ces deux espaces vectoriels sont tous deux de dimension finie, chacun muni d'une base :

- ⋆ $\mathcal{B} = (v_1, \dots, v_p)$ est une base de V,
- ⋆ $\mathcal{B}' = (v'_1, \dots, v'_m)$ est une base de V'.

Nous allons voir maintenant comment l'utilisation de ces bases va permettre de ramener l'étude de T à l'étude d'une application linéaire de $\mathbb{R}^p$ dans $\mathbb{R}^m$.

En effet, prenons un vecteur dans l'ensemble de départ, $v \in V$, et décomposons-le sur $\mathcal{B}$:

$$v = a_1 v_1 + \cdots + a_p v_p \,,$$

ce qui permet de décrire v univoquement à l'aide du vecteur de $\mathbb{R}^p$ qui lui est associé :

$$[v]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} .$$

Ensuite, regardons l'image de v par T. Puisque T est linéaire,

$$\begin{aligned} T(v) &= T(a_1 v_1 + \cdots + a_p v_p) \\ &= a_1 T(v_1) + \cdots + a_p T(v_p) \,. \end{aligned}$$

En utilisant ensuite la linéarité de $[\cdot]_{\mathcal{B}'}$,

$$\begin{aligned} [T(v)]_{\mathcal{B}'} &= \big[a_1 T(v_1) + \cdots + a_p T(v_p)\big]_{\mathcal{B}'} \\ &= a_1 [T(v_1)]_{\mathcal{B}'} + \cdots + a_p [T(v_p)]_{\mathcal{B}'} \,. \end{aligned}$$

Cette dernière ligne est une combinaison linéaire des vecteurs $[T(v_1)]_{\mathcal{B}'}, \cdots, [T(v_p)]_{\mathcal{B}'}$ de $\mathbb{R}^m$, on peut donc l'interpréter comme un produit d'une matrice par le vecteur $[v]_{\mathcal{B}}$:

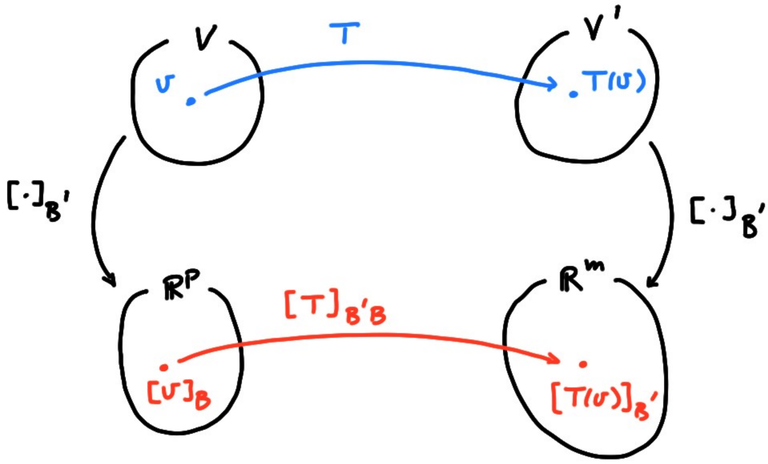
$$\begin{aligned} a_1 [T(v_1)]_{\mathcal{B}'} + \cdots + a_p [T(v_p)]_{\mathcal{B}'} &= \underbrace{\big[[T(v_1)]_{\mathcal{B}'} \cdots [T(v_p)]_{\mathcal{B}'}\big]}_{m \times n} \underbrace{\begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}}_{=[v]_{\mathcal{B}}} \\ &= \big[[T(v_1)]_{\mathcal{B}'} \cdots [T(v_p)]_{\mathcal{B}'}\big] [v]_{\mathcal{B}} \end{aligned}$$

Définition 4.55. La matrice $m \times p$ définie par

$$[T]_{\mathcal{B}'\mathcal{B}} := \big[[T(v_1)]_{\mathcal{B}'} \cdots [T(v_p)]_{\mathcal{B}'}\big]$$

est la matrice qui **représente T relativement aux bases $\mathcal{B}$ (départ)** et **$\mathcal{B}'$ (arrivée)**.

Ce que nous avons fait ci-dessus peut se résumer dans le shéma suivant :



En utilisant les bases $\mathcal{B}$ et $\mathcal{B}'$, ainsi que les applications $[\cdot]_{\mathcal{B}}$ et $[\cdot]_{\mathcal{B}'}$ qui leur sont associées, nous avons pu prendre l'application

$$v \mapsto T(v)$$

qui est abstraite, et nous l'avons rendue plus concrète, en la représentant à l'aide d'une matrice : on peut maintenant la voir comme une application linéaire de $\mathbb{R}^p$ dans $\mathbb{R}^m$, dont la matrice est $[T]_{\mathcal{B}'\mathcal{B}}$:

$$\underbrace{[v]_{\mathcal{B}}}_{\in\mathbb{R}^p} \mapsto \underbrace{[T(v)]_{\mathcal{B}'}}_{\in\mathbb{R}^m} = [T]_{\mathcal{B}'\mathcal{B}}[v]_{\mathcal{B}}$$

Maintenant, l'étude de T peut se réduire à celle de la matrice $[T]_{\mathcal{B}'\mathcal{B}}$.

Exemple 4.56. Considérons l'application $T : \mathbb{P}_3 \to \mathbb{P}_2$ définie ainsi : pour $p \in \mathbb{P}_3$,

$$T(p) = p'\,,$$

qui signifie que $T(p)(t) := p'(t)$ (dérivée de p par rapport à t).

Cette application est clairement linéaire puisque

$$T(\alpha p + \beta q) = (\alpha p + \beta q)' = \alpha p' + \beta q' = \alpha T(p) + \beta T(q)\,.$$

Calculons maintenant la matrice associée à cette application, relativement

- ⋆ à la base canonique $\mathcal{B}_{\text{can}} = (e_0, e_1, e_2, e_3)$ dans $\mathbb{P}_3$, et
- ⋆ à la base canonique $\mathcal{B}'_{\text{can}} = (e_0, e_1, e_2)$ dans $\mathbb{P}_2$.

Par ce qu'on a dit plus haut, cette matrice sera

$$[T]_{\mathcal{B}'_{\text{can}}\mathcal{B}_{\text{can}}} = \big[[T(e_0)]_{\mathcal{B}'_{\text{can}}}\ [T(e_1)]_{\mathcal{B}'_{\text{can}}}\ [T(e_2)]_{\mathcal{B}'_{\text{can}}}\ [T(e_3)]_{\mathcal{B}'_{\text{can}}}\big]\,.$$

Comme

$$e_0(t) = 1 \quad e_1(t) = t \quad e_2(t) = t^2 \quad e_3(t) = t^3\,,$$

on a

$$e_0'(t) = 0 \quad e_1'(t) = 1 \quad e_2'(t) = 2t \quad e_3'(t) = 3t^2\,,$$

et donc

$$T(e_0) = 0\,, \qquad T(e_1) = e_0\,, \qquad T(e_2) = 2e_1\,, \qquad T(e_3) = 3e_2\,,$$

c'est-à-dire

$$\begin{aligned} T(e_0) &= 0e_0 + 0e_1 + 0e_2 \\ T(e_1) &= 1e_0 + 0e_1 + 0e_2 \\ T(e_2) &= 0e_0 + 2e_1 + 0e_2 \\ T(e_3) &= 0e_0 + 0e_1 + 3e_2 \,. \end{aligned}$$

On peut donc écrire

$$[T(e_0)]_{\mathcal{B}'_{\text{can}}} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad [T(e_1)]_{\mathcal{B}'_{\text{can}}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$[T(e_2)]_{\mathcal{B}'_{\text{can}}} = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \qquad [T(e_3)]_{\mathcal{B}'_{\text{can}}} = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$$

La matrice qui représente T est donc, relativement à ce choix de bases,

$$[T]_{\mathcal{B}'_{\text{can}}\mathcal{B}_{\text{can}}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Prenons par exemple le polynôme $p \in \mathbb{P}_3$ défini par

$$p(t) = 2 + t^2 - 5t^3 \,,$$

pour lequel

$$[p]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 2 \\ 0 \\ 1 \\ -5 \end{pmatrix}.$$

Son image par T est $T(p) \in \mathbb{P}_2$, qui relativement à $\mathcal{B}'_{\text{can}}$ est donnée par

$$[T(p)]_{\mathcal{B}'_{\text{can}}} = [T]_{\mathcal{B}'_{\text{can}}\mathcal{B}_{\text{can}}}[p]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ 1 \\ -5 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ -15 \end{pmatrix},$$

qui est bien la décomposition de

$$p'(t) = (2 + t^2 - 5t^3)' = 2t - 15t^2$$

relativement à $\mathcal{B}'_{\text{can}}$:

$$[p']_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 0 \\ 2 \\ -15 \end{pmatrix}.$$

◇

Exemple 4.57. Considérons l'application

$$\begin{aligned} T : \mathbb{P}_2 &\to \mathbb{R}^2 \\ p &\mapsto T(p) := \begin{pmatrix} p(0) \\ p'(1) \end{pmatrix}. \end{aligned}$$

($p'(t)$ est la dérivée de $p(t)$ par rapport à t.) Remarquons que T est linéaire, puisque pour tous $p, q \in \mathbb{P}_2$ et tout scalaires α, β,

$$\begin{aligned} T(\alpha p + \beta q) &= \begin{pmatrix} \alpha p(0) + \beta q(0) \\ \alpha p'(1) + \beta q'(1) \end{pmatrix} \\ &= \alpha \begin{pmatrix} p(0) \\ p'(1) \end{pmatrix} + \beta \begin{pmatrix} q(0) \\ q'(1) \end{pmatrix} = \alpha T(p) + \beta T(q)\,. \end{aligned}$$

Puisqu'on connaît la base canonique $\mathcal{B}_{\text{can}} = (e_0, e_1, e_2)$ dans $\mathbb{P}_2$ et la base canonique $\mathcal{B}'_{\text{can}} = (\mathbf{e}_1, \mathbf{e}_2)$ dans $\mathbb{R}^2$ (on écrit $\mathcal{B}'_{\text{can}}$ juste pour la distinguer de l'autre, mais c'est bien la base canonique de $\mathbb{R}^2$), on peut calculer la matrice 2×3 qui représente T relativement à ces bases :

$$[T]_{\mathcal{B}'_{\text{can}}\mathcal{B}_{\text{can}}} = \big[[T(e_0)]_{\mathcal{B}'_{\text{can}}}\ [T(e_1)]_{\mathcal{B}'_{\text{can}}}\ [T(e_2)]_{\mathcal{B}'_{\text{can}}} \big]\,.$$

Comme $e_0(t) = 1$, $e_1(t) = t$, $e_2(t) = t^2$, on a

$$\begin{aligned} T(e_0) &= \begin{pmatrix} e_0(0) \\ e_0'(1) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ T(e_1) &= \begin{pmatrix} e_1(0) \\ e_1'(1) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ T(e_2) &= \begin{pmatrix} e_2(0) \\ e_2'(1) \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}\,, \end{aligned}$$

et donc

$$[T]_{\mathcal{B}'_{\text{can}}\mathcal{B}_{\text{can}}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \end{pmatrix}\,.$$

Par exemple, prenons le polynôme $p(t) = 9 - 2t + 7t^2$, et calculons son image. Alors

$$\begin{aligned} [T(p)]_{\mathcal{B}'_{\text{can}}} &= [T]_{\mathcal{B}'_{\text{can}}\mathcal{B}_{\text{can}}} [p]_{\mathcal{B}_{\text{can}}} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 9 \\ -2 \\ 7 \end{pmatrix} = \begin{pmatrix} 9 \\ 12 \end{pmatrix}\,, \end{aligned}$$

qui est bien $\begin{pmatrix} p(0) \\ p'(1) \end{pmatrix}$. ◇

Nous reviendrons plus en profondeur sur la représentation d'une application linéaire à l'aide d'une matrice, en particulier dans le cas $T : \mathbb{R}^n \to \mathbb{R}^m$.

4.9 Théorème du Rang

Théorème 4.58. *Soient V, V' deux espaces vectoriels de dimensions finies, et soit $T : V \to V'$ une application linéaire. Alors*

$$\dim(\ker(T)) + \dim(\mathrm{Im}(T)) = \dim(V)\,.$$

Preuve: Soit $n = \dim(V)$, et soit $p = \dim(\ker(T))$. Puisque $\ker(T)$ est un sous-espace vectoriel de V, on a forcément que $p \leqslant n$. Ce que l'on doit donc montrer, c'est que $\dim(\mathrm{Im}(T)) = n - p$.

Si $p = n$, on a $\mathrm{Im}(T) = \{0'\}$ (l'élément nul de V'), et donc $\dim(\mathrm{Im}(T)) = 0$, et le théorème est démontré.

Si $p < n$, posons $r := n - p$, qui est par définition plus grand ou égal à 1. Nous allons montrer que $\dim(\mathrm{Im}(T)) = r$.

Pour ce faire, commençons par considérer une base $\mathcal{B}_{\ker(T)}$ de $\ker(T)$:

$$\mathcal{B}_{\ker(T)} = \big(v_1, \dots, v_p\big) .$$

Puisque $p < n$, $\mathcal{B}_{\ker(T)}$ n'est pas une base de V. Mais on peut malgré tout la compléter en rajoutant $n - p = r$ vecteurs, afin d'obtenir une base de V :

$$\mathcal{B}_V = \big(v_1, \dots, v_p, w_1, \dots, w_r\big) .$$

Montrons maintenant que la famille

$$\mathcal{B}' = \big(T(w_1), \dots, T(w_r)\big)$$

est une base de $\mathrm{Im}(T)$.

1) $\mathcal{B}'$ est libre. En effet, considérons une combinaison linéaire nulle,

$$\alpha_1 T(w_1) + \cdots + \alpha_r T(w_r) = 0 .$$

(On va montrer que $\alpha_1 = \cdots = \alpha_r = 0$.) Par la linéarité de T, on peut écrire cette dernière comme

$$T(\alpha_1 w_1 + \cdots + \alpha_r w_r) = 0 ,$$

qui indique que le vecteur $\alpha_1 w_1 + \cdots + \alpha_r w_r$ est dans $\ker(T)$. On peut donc le décomposer dans la base $\mathcal{B}_{\ker(T)}$:

$$\alpha_1 w_1 + \cdots + \alpha_r w_r = \lambda_1 v_1 + \dots \lambda_p v_p .$$

Or on peut récrire cette dernière comme

$$\lambda_1 v_1 + \dots \lambda_p v_p - \alpha_1 w_1 - \cdots - \alpha_r w_r = 0 .$$

Comme $\mathcal{B}_V = (v_1, \dots, v_p, w_1, \dots, w_r)$ est une base de V, on a donc que

$$\lambda_1 = \cdots = \lambda_p = -\alpha_1 = \cdots = -\alpha_r = 0 .$$

Ainsi, $\alpha_1 = \cdots = \alpha_r = 0$, ce qui démontre l'affirmation.

2) $\mathcal{B}'$ engendre $\mathrm{Im}(T)$. En effet, considérons un $v' \in \mathrm{Im}(T)$, c'est-à-dire un élément $v' \in V'$ pour lequel il existe un $v \in V$ tel que $v' = T(v)$. Puisque l'on peut décomposer v dans la base $\mathcal{B}_V$,

$$v = \lambda_1 v_1 + \cdots + \lambda_r v_r + \lambda_{r+1} w_1 + \cdots + \lambda_n w_p ,$$

on a donc que

$$\begin{aligned} v' &= T(v) \\ &= T(\lambda_1 v_1 + \cdots + \lambda_r v_r + \lambda_{r+1} w_1 + \cdots + \lambda_n w_p) \\ &= \lambda_1 T(v_1) + \cdots + \lambda_r T(v_r) + \lambda_{r+1} T(w_1) + \cdots + \lambda_n T(w_p) \\ &= \lambda_{r+1} T(w_1) + \cdots + \lambda_n T(w_p) . \end{aligned}$$

(Dans la dernière ligne, on a utilisé le fait que chaque $v_k \in \ker(T)$, et donc chaque $T(v_k) = 0$.) Mais cette dernière relation implique que $\mathcal{B}'$ engendre bien $\mathrm{Im}(T)$.

Ainsi, $\mathcal{B}'$ est une base de $\mathrm{Im}(T)$, et comme elle contient r éléments, on a que $\dim(\mathrm{Im}(T)) = r$. On a donc bien que

$$\begin{aligned} \dim(\ker(T)) + \dim(\mathrm{Im}(T)) &= p + r \\ &= p + (n - p) = n = \dim(V) . \end{aligned}$$

□

Chapitre 5

Applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}^m$

5.1 Matrice d'une application

Dans ce chapitre, nous allons appliquer quelques-unes des notions relatives aux applications linéaires $T : V \to V'$, vues dans le chapitre précédent, au cas particulier $T : \mathbb{R}^n \to \mathbb{R}^m$.

Nous avions vu que toute application $T : \mathbb{R}^n \to \mathbb{R}^m$ de la forme $T(\mathbf{x}) = A\mathbf{x}$ est linéaire, et nous savons depuis la dernière section du dernier chapitre que la réciproque est vraie :

Théorème 5.1. *Si $T : \mathbb{R}^n \to \mathbb{R}^m$ est linéaire, alors il existe une unique matrice A ($m \times n$) telle que*

$$T(\mathbf{x}) = A\mathbf{x}\,, \qquad \forall \mathbf{x} \in \mathbb{R}^n\,.$$

De plus, la matrice A est celle dont les colonnes sont les images par T des vecteurs de la base canonique :

$$A = \big[T(\mathbf{e}_1) \cdots T(\mathbf{e}_n)\big]\,.$$

Par ce que nous savons de la section précédente, A n'est autre que $[T]_{\mathcal{B}'_{\text{can}}\mathcal{B}_{\text{can}}}$, où $\mathcal{B}_{\text{can}}$ est la base canonique de $\mathbb{R}^n$, et $\mathcal{B}'_{\text{can}}$ est la base canonique de $\mathbb{R}^m$.

Exemple 5.2. Considérons l'application linéaire $T : \mathbb{R}^3 \to \mathbb{R}^2$ déjà considérée précédemment :

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{x} \mapsto T(\mathbf{x}) := \begin{pmatrix} -x_1 + 3x_2 + 5x_3 \\ x_3 + 7x_1 \end{pmatrix}$$

En calculant les images des vecteurs de base,

$$T(\mathbf{e}_1) = T\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} -1 + 3 \cdot 0 + 5 \cdot 0 \\ 0 + 7 \cdot 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 7 \end{pmatrix}$$

$$T(\mathbf{e}_2) = T\left(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 0 + 3 \cdot 1 + 5 \cdot 0 \\ 0 + 7 \cdot 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

$$T(\mathbf{e}_3) = T\left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 + 3 \cdot 0 + 5 \cdot 1 \\ 1 + 7 \cdot 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}\,,$$

ce qui donne la matrice associée à T :

$$A = \begin{pmatrix} -1 & 3 & 5 \\ 7 & 0 & 1 \end{pmatrix}$$

$\diamond$

Exemple 5.3. Considérons l'application $T : \mathbb{R}^3 \to \mathbb{R}$:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{x} \mapsto T(\mathbf{x}) := x_2 - 3x_1 \,.$$

(On montre facilement que cette application est linéaire.) En calculant les images des vecteurs de base,

$$\begin{aligned} T(\mathbf{e}_1) &= T\Big(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\Big) = 0 - 3 \cdot 1 = -3 \,, \\ T(\mathbf{e}_2) &= T\Big(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\Big) = 1 - 3 \cdot 0 = 1 \,, \\ T(\mathbf{e}_3) &= T\Big(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\Big) = 0 - 3 \cdot 0 = 0 \,, \end{aligned}$$

ce qui donne la matrice 1×3 associée à T :

$$A = \big[T(\mathbf{e}_1)\, T(\mathbf{e}_2)\, T(\mathbf{e}_3)\big] = \begin{pmatrix} -3 & 1 & 0 \end{pmatrix}$$

En effet,

$$T(\mathbf{x}) = A\mathbf{x} = \begin{pmatrix} -3 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = -3x_1 + x_2 \,.$$

$\diamond$

Remarque 5.4. Les applications linéaires $T : \mathbb{R}^n \to \mathbb{R}^m$ définies jusqu'ici ont toujours été définies *en composantes*, c'est-à-dire en définissant les composantes de $T(\mathbf{x}) \in \mathbb{R}^m$ à l'aide des composantes de $\mathbf{x} \in \mathbb{R}^n$, comme dans les deux exemples précédents.

Il faut garder à l'esprit que pour l'instant, ces composantes sont toujours des composantes *associées à la base canonique*.

En général, une application n'a pas besoin d'être définie à l'aide de composantes (voir les exemples de transformations géométriques plus loin dans le chapitre), et on pourra effectivement lui associer une matrice, mais qui dépendra fortement de la base choisie. $\diamond$

5.2 Surjection, injection, bijection

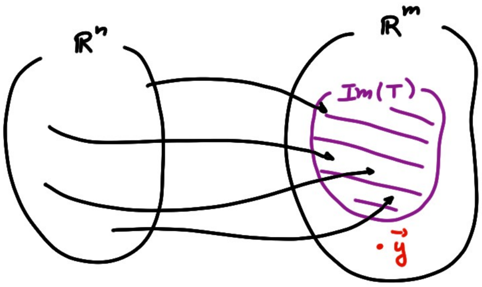
Dans cette section, on illustre les notions de base de la théorie des fonctions, dans le cadre des applications linéaires $T : \mathbb{R}^n \to \mathbb{R}^m$. En particulier, on regardera comment ces notions se traduisent en termes de la matrice qui représente T.

Rappelons que les notions de bases de la théorie des fonctions (injectivité, surjectivité, bijectivité) que nous utiliserons sont résumées **ici** (lien web).

Ensemble image, surjection

Rappelons la définition de l'ensemble image d'une application : c'est l'ensemble des points de l'ensemble d'arrivée qui possèdent au moins une préimage,

$$\mathrm{Im}(T) := \left\{ \mathbf{y} \in \mathbb{R}^m \ : \ \exists \mathbf{x} \in \mathbb{R}^n \text{ t.q. } T(\mathbf{x}) = \mathbf{y} \right\}$$



Rappelons aussi que $T : \mathbb{R}^n \to \mathbb{R}^m$ est **surjective** si $\mathrm{Im}(T) = \mathbb{R}^m$.

Lorsque T est linéaire, on sait qu'il existe une matrice A $(m \times n)$ telle que

$$T(\mathbf{x}) = A\mathbf{x}\,.$$

Puisque T est entièrement déterminée par sa matrice A, on écrira souvent $\mathrm{Im}(A)$ au lieu de $\mathrm{Im}(T)$.

Écrivons maintenant A à l'aide de ses colonnes :

$$A = [\mathbf{a}_1 \cdots \mathbf{a}_n]\,.$$

Puisque $A\mathbf{x}$ est une combinaison linéaire des colonnes de A, $\mathrm{Im}(A)$ représente tous les vecteurs de $\mathbb{R}^m$ que l'on peut obtenir à l'aide de combinaisons linéaires des colonnes de A :

$$\mathrm{Im}(A) = \mathrm{Vect}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}\,.$$

On peut donc se souvenir de l'ensemble image de $T(\mathbf{x}) = A\mathbf{x}$ comme le sous-ensemble de $\mathbb{R}^m$ engendré par les *colonnes* de A ; pour cette raison, il est souvent noté comme suit :

$$\mathrm{Im}(T) = \mathrm{Im}(A) = \mathrm{Col}(A)\,.$$

On a donc une formulation équivalente de la surjectivité :

Théorème 5.5. *Soit $T : \mathbb{R}^n \to \mathbb{R}^m$ une application linéaire et soit A la matrice $m \times n$ qui lui est associée. Alors T est surjective si et seulement si les colonnes de A engendrent tout $\mathbb{R}^m$.*

Exemple 5.6. Montrons que l'application $T : \mathbb{R}^3 \to \mathbb{R}^3$ associée à la matrice

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

est surjective. Pour ce faire, on doit fixer n'importe quel $\mathbf{y} \in \mathbb{R}^3$ et montrer qu'il existe au moins un $\mathbf{x} \in \mathbb{R}^3$ tel que $A\mathbf{x} = \mathbf{y}$. Fixons donc $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$, et étudions le système $A\mathbf{x} = \mathbf{y}$:

$$\begin{pmatrix} 2 & 0 & -1 \\ 1 & -1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

On voit facilement que A est équivalente à

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1/2 \\ 0 & 0 & 7/2 \end{pmatrix},$$

ce qui implique que la solution existe toujours (et est unique), quel que soit $\mathbf{y}$. On conclut que T est surjective : $\mathrm{Im}(T) = \mathbb{R}^3$.

On aurait aussi pu simplement montrer que les colonnes de A forme une famille libre, ce qui implique qu'elle engendre tout $\mathbb{R}^3$ (c'est parce que dans un espace vectoriel de dimension n, toute famille libre de de n vecteurs forme une base). ◇

Exemple 5.7. L'application $T : \mathbb{R}^2 \to \mathbb{R}^4$ associée à une matrice

$$A = \begin{pmatrix} a & \alpha \\ b & \beta \\ c & \gamma \\ d & \delta \end{pmatrix}$$

ne *peut pas* être surjective, puisque deux vecteurs de $\mathbb{R}^4$ ne suffisent jamais pour engendrer $\mathbb{R}^4$. ◇

Injection

Rappelons qu'une application $T : \mathbb{R}^n \to \mathbb{R}^m$ est **injective** si des éléments de $\mathbb{R}^n$ distincts ont des images distinctes :

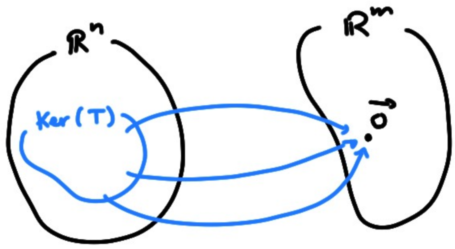
$$\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{x}' \Rightarrow T(\mathbf{x}) \neq T(\mathbf{x}'),$$

ou alors, ce qui est équivalent, si

$$T(\mathbf{x}) = T(\mathbf{x}') \quad \Longrightarrow \quad \mathbf{x} = \mathbf{x}'.$$

Pour les applications qui sont *linéaires*, $T(\mathbf{x}) = A\mathbf{x}$, on sait que l'injectivité peut se caractériser à l'aide du **noyau** :

$$\ker(T) := \{\mathbf{x} \in \mathbb{R}^n \ : \ A\mathbf{x} = \mathbf{0}\}.$$



On écrira souvent le noyau $\ker(A)$ au lieu de $\ker(T)$. Puisque $\ker(A)$ n'est autre que l'ensemble des solutions du système homogène $A\mathbf{x} = \mathbf{0}$, et comme on sait qu'il y a toujours la solution triviale, le noyau n'est jamais vide : $\ker(A) \ni \mathbf{0}$.

Nous avons vu qu'une application linéaire est injective si et seulement si son noyau ne contient *que* le vecteur nul :

$$\ker(A) = \{\mathbf{0}\}.$$

Et comme on sait que l'unicité de la solution du problème homogène caractérise l'indépendance des colonnes de A, l'injectivité peut se formuler en termes de l'indépendance des colonnes de la matrice :

Théorème 5.8. *Soit $T : \mathbb{R}^n \to \mathbb{R}^m$ une application linéaire et soit A la matrice $m \times n$ qui lui est associée. Alors T est injective si et seulement si les colonnes de A sont linéairement indépendantes.*

Exemple 5.9. Soit $T : \mathbb{R}^3 \to \mathbb{R}^3$ l'application linéaire décrite par la matrice (déjà rencontrée plus tôt)

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

Puisque le noyau correspond à l'ensemble des solutions de $A\mathbf{x} = \mathbf{0}$, on l'a déjà calculé :

$$\ker(T) = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = t \begin{pmatrix} -1 \\ 1 \\ 3 \end{pmatrix} \middle| t \in \mathbb{R} \right\}$$

Comme ce noyau contient des vecteurs non-nuls (tout choix de $t \neq 0$ donne une solution non-triviale), T n'est pas injective. Ceci signifie aussi que les colonnes de A sont linéairement dépendantes. En effet, en prenant par exemple la solution correspondant à $t = 1$, on peut écrire

$$(-1) \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ 8 \\ 5 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} .$$

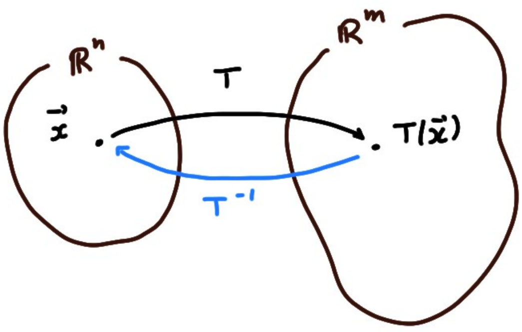
◇

Bijection

Rappelons qu'une application $T : \mathbb{R}^n \to \mathbb{R}^m$ est **bijective** (ou que c'est une **bijection**) si elle est à la fois surjective et injective.

La **réciproque** d'une application bijective : $\mathbb{R}^n \to \mathbb{R}^m$ se note

$$\begin{aligned} T^{-1} : \mathbb{R}^m &\to \mathbb{R}^n \\ \mathbf{y} &\mapsto T^{-1}(\mathbf{y}) = \mathbf{x}_* . \end{aligned}$$



Par construction, on a

$$\begin{aligned} T(T^{-1}(\mathbf{y})) &= \mathbf{y} \qquad \forall \mathbf{y} \in \mathbb{R}^m , \\ T^{-1}(T(\mathbf{x})) &= \mathbf{x} \qquad \forall \mathbf{x} \in \mathbb{R}^n . \end{aligned}$$

On sait que si une application $T : \mathbb{R}^n \to \mathbb{R}^m$ est bijective et linéaire, alors sa réciproque est aussi linéaire. Mais on a aussi :

Théorème 5.10. *Si $T : \mathbb{R}^n \to \mathbb{R}^m$ est linéaire et bijective, alors les ensembles de départ et d'arrivée doivent être de même dimension : $m = n$.*

Preuve: On utilise la représentation de T à l'aide de sa matrice $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$, où chaque $\mathbf{a}_k$ est un vecteur de $\mathbb{R}^m$: $T(\mathbf{x}) = A\mathbf{x}$.

D'une part, T étant injective, on sait que les colonnes de A sont linéairement indépendantes. Ceci implique que $n \leqslant m$ (en effet, on sait que dans $\mathbb{R}^m$, une famille de plus de m vecteurs est toujours liée). D'autre part, T étant surjective, les colonnes de A engendrent $\mathbb{R}^m$, ce qui signifie que $n \geqslant m$.

De là, on conclut que $n = m$. □

Ceci implique que les applications linéaire bijectives ne peuvent exister qu'entre des espaces de même dimensions :

$$T : \mathbb{R}^n \to \mathbb{R}^n$$

De telles applications sont donc représentées par des matrices carrées. Nous y reviendrons plus tard et étudierons leurs propriétés.

5.3 Une base pour $\text{Im}(A)$

Si $T : \mathbb{R}^n \to \mathbb{R}^m$ est une application linéaire définie par une matrice A ($m \times n$). On sait que l'ensemble image $\text{Im}(A) = \text{Col}(A)$ est un sous-espace vectoriel de $\mathbb{R}^m$ (si T n'est pas surjective, c'est un sous-espace strict), et dans cette section nous allons voir un moyen de trouver une base pour le décrire.

Extraire une base des colonnes

D'un point de vue calculatoire, l'ensemble image $\text{Im}(A)$ se calcule en trouvant tous les $\mathbf{y} \in \mathbb{R}^m$ pour lesquels le système

$$A\mathbf{x} = \mathbf{y}$$

possède au moins une solution. Ensuite, chercher une base pour $\text{Im}(A)$ présente a priori une seconde étape.

Mais comme on sait, l'ensemble image est l'ensemble de toutes les combinaisons linéaires des colonnes de A :

$$\text{Im}(T) = \text{Im}(A) = \text{Col}(A) = \text{Vect}\{\mathbf{a}_1 \cdots \mathbf{a}_n\}\,.$$

Puisque les colonnes engendrent $\text{Im}(A)$, il y a tout lieu de penser que certaines d'entre elles peuvent être utilisées pour former une base de $\text{Im}(A)$. Voyons ça sur un exemple (un peu trop) simple.

Exemple 5.11. Considérons l'application $T : \mathbb{R}^3 \to \mathbb{R}^3$ dont la matrice est donnée par

$$A = [\mathbf{a}_1\, \mathbf{a}_2\, \mathbf{a}_3\, \mathbf{a}_4] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Comme $\mathbf{a}_2$ et $\mathbf{a}_4$ sont identiquement nulles, elles ne participent pas à $\text{Col}(A)$:

$$\text{Col}(A) = \text{Vect}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\} = \text{Vect}\{\mathbf{a}_1, \mathbf{a}_3\}\,.$$

De plus, $\mathbf{a}_1$ et $\mathbf{a}_3$ sont linéairement indépendantes, et donc elles forment une base de l'espace qu'elles engendrent. Donc $\mathcal{B} = (\mathbf{a}_1, \mathbf{a}_3)$ forme une base de $\text{Col}(A)$. ◇

Dans ce dernier exemple, on a pu simplement retirer des colonnes nulles, sachant qu'elles ne contribuent pas à l'espace $\mathrm{Col}(A)$.

Nous avons déjà décrit, dans le cadre abstrait des espaces vectoriels (voir **ici** (lien web)), le processus qui permet de retirer les vecteurs "superflus" dans une famille qui engendre un sous-espace W, donnant un algorithme menant à une base de W : on retire un à un les vecteurs qui peuvent être exprimés comme combinaisons linéaires des autres, et quand on ne peut plus en retirer, c'est qu'on est en possession d'une base. Appliquons ce résultat pour calculer l'ensemble image d'une matrice, $W = \mathrm{Col}(A)$:

Exemple 5.12. Considérons l'application linéaire $T : \mathbb{R}^5 \to \mathbb{R}^3$ dont la matrice est

$$A = [\mathbf{a}_1\,\mathbf{a}_2\,\mathbf{a}_3\,\mathbf{a}_4, \mathbf{a}_5] = \begin{pmatrix} 1 & 2 & 0 & 3 & -4 \\ 0 & -2 & 2 & 1 & 1 \\ 1 & 5 & -3 & 1 & -5 \end{pmatrix}.$$

A priori,

$$\mathrm{Col}(A) = \mathrm{Vect}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5\}.$$

Or on remarque que $\mathbf{a}_2 = 2\mathbf{a}_1 - \mathbf{a}_3$, et donc le lemme ci-dessus garantit qu'on peut retirer $\mathbf{a}_2$ sans changer l'espace engendré :

$$\mathrm{Col}(A) = \mathrm{Vect}\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5\}.$$

On remarque aussi que $\mathbf{a}_5 = -\mathbf{a}_1 + \mathbf{a}_3 - \mathbf{a}_4$, et donc

$$\mathrm{Col}(A) = \mathrm{Vect}\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4\}.$$

Maintenant, on peut remarquer que $\mathbf{a}_1$, $\mathbf{a}_3$ et $\mathbf{a}_4$ sont linéairement indépendantes. Comme elles engendrent $\mathrm{Col}(A)$, elles forment donc une base de $\mathrm{Col}(A)$.

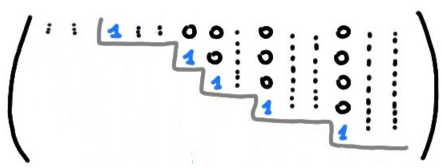
Remarquons en passant que puisque cette base contient trois vecteurs, $\dim(\mathrm{Col}(A)) = 3$, qui est aussi la dimension de l'espace d'arrivée. Ceci a pour conséquence que l'application $T(\mathbf{x}) := A\mathbf{x}$ est surjective. $\diamond$

Une méthode pour identifier les colonnes retirables

Dans le dernier exemple, on a pu trouver des colonnes qui étaient combinaisons linéaires des autres, mais n'y a-t-il pas un moyen plus méthodique de trouver facilement les colonnes "superflues", pour ne garder que celles qui forment une base de $\mathrm{Col}(A)$? La réponse est "oui", et pour le comprendre il faut reprendre le procédé de réduction vu au début du cours.

Définition 5.13. Soit A une matrice $m \times n$ et soit $\widetilde{A}$ sa réduite. Si la k-ème colonne de $\widetilde{A}$ contient un pivot, on dit que la k-ème colonne de A est une **colonne-pivot**

Rappelons que les pivots, dans $\widetilde{A}$, sont les coefficients principaux égaux à 1, seuls coefficients non-nuls de leur colonne :



L'unicité de la réduite implique que la notion de colonne-pivot, pour A, est bien définie.

Exemple 5.14. Considérons
$$A = \begin{pmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{pmatrix}.$$
Alors les colonnes 1 et 2 de A sont des colonnes-pivot, car après réduction, les colonnes 1 et 2 de $\widetilde{A}$ sont celles qui contiennent des pivots :
$$\widetilde{A} = \begin{pmatrix} 1 & 0 & -4/3 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$
◇

Théorème 5.15. *Les colonnes-pivot d'une matrice A forment une base de $\mathrm{Im}(A)$. En particulier, $\dim(\mathrm{Im}(A))$ est égale au nombre de colonnes-pivot de A.*

Pour démontrer le théorème, nous aurons besoin du résultat suivant, qui dit que les dépendances linéaires existant entre des colonnes d'une matrice sont les mêmes que celles existant entre les colonnes correspondantes de sa réduite :

Proposition 3. *Soit $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$, et soit $\mathcal{F}$ un sous-ensemble de colonnes de A :*
$$\mathcal{F} = \{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_l}\} \subset \{\mathbf{a}_1 \cdots \mathbf{a}_n\}.$$
Si $\widetilde{A} = [\mathbf{r}_1 \cdots \mathbf{r}_n]$ est la réduite de A, et si $\widetilde{\mathcal{F}}$ est le sous-ensemble de colonnes de $\widetilde{A}$ correspondant à $\mathcal{F}$,
$$\widetilde{\mathcal{F}} = \{\mathbf{r}_{i_1}, \dots, \mathbf{r}_{i_l}\} \subset \{\mathbf{r}_1 \cdots \mathbf{r}_n\},$$
alors

- ⋆ *$\mathcal{F}$ est liée si et seulement si $\widetilde{\mathcal{F}}$ est liée,*
- ⋆ *$\mathcal{F}$ est libre si et seulement si $\widetilde{\mathcal{F}}$ est libre.*

Preuve: Si les colonnes considérées sont $i_1, \dots, i_l$, alors
$$\mathcal{F} = \{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_l}\}$$
$$\widetilde{\mathcal{F}} = \{\mathbf{r}_{i_1}, \dots, \mathbf{r}_{i_l}\}.$$
Supposons que les colonnes de $\mathcal{F}$ satisfont à une relation linéaire du type
$$\alpha_{i_1}\mathbf{a}_{i_1} + \cdots \alpha_{i_l}\mathbf{a}_{i_l} = \mathbf{0}.$$
Cette relation représente un système homogène, et si on applique sur ce système les mêmes opérations élémentaires qui transforment A en $\widetilde{A}$, il se transforme en le système homogène associé à la relation linéaire
$$\alpha_{i_1}\mathbf{r}_{i_1} + \cdots \alpha_{i_l}\mathbf{r}_{i_l} = \mathbf{0}.$$
Puisque les opérations élémentaires ne changent pas l'ensemble des solutions d'un système, des nombres $\alpha_{i_1}, \dots, \alpha_{i_l}$ sont solutions de la première relation si et seulement ils sont solutions de la deuxième relation. Ceci démontre la proposition. □

Exemple 5.16. Les colonnes 1, 3 et 8 de A sont dépendantes si et seulement si les colonnes 1, 3 et 8 de $\widetilde{A}$ sont dépendantes. ◇

Une conséquence directe du résultat ci-dessus :

Corollaire 3. *Un sous-ensemble des colonnes de A*

$$\mathcal{B} = (\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_l}) \subset \{\mathbf{a}_1 \cdots \mathbf{a}_n\}$$

forme une base de $\mathrm{Col}(A)$ *si et seulement si le sous-ensemble de colonnes de* $\widetilde{A}$ *correspondant,*

$$\widetilde{\mathcal{B}} = (\mathbf{r}_{i_1}, \ldots, \mathbf{r}_{i_l}) \subset \{\mathbf{r}_1 \cdots \mathbf{r}_n\},$$

forme une base de $\mathrm{Col}(\widetilde{A})$.

Exemple 5.17. Soit A une matrice 7×11. Les colonnes 2, 5 et 8 de $\widetilde{A}$ forment une base de $\mathrm{Col}(\widetilde{A})$ si et seulement si les colonnes 2, 5 et 8 de A forment une base de $\mathrm{Col}(A)$. $\diamond$

Nous pouvons maintenant prouver le théorème :
Preuve: Commençons par deux remarques concernant la réduite $\widetilde{A}$.

- ⋆ Dans $\widetilde{A}$, les colonnes contenant des pivots sont linéairement indépendantes, puisqu'elles ont toutes un seul coefficient non nul (le pivot "1"), chaque fois situé à une hauteur différente.
- ⋆ Dans $\widetilde{A}$, toute colonne qui ne contient pas de pivot peut s'écrire comme combinaison linéaire des colonnes qui contiennent un pivot, et qui sont situées à sa gauche.

Par conséquent, le lemme énoncé plus haut garantit que les colonnes de $\widetilde{A}$ ne contenant pas de pivot ne contribuent pas à $\mathrm{Col}(\widetilde{A})$, et les colonnes contenant un pivot forment une base de $\mathrm{Col}(\widetilde{A})$. Par le corollaire, ceci implique que les colonnes-pivot de A forment une base de $\mathrm{Col}(A)$. $\square$

Voyons comment utiliser le théorème pour obtenir plus facilement une base de $\mathrm{Col}(A)$:
Exemple 5.18. Considérons la même matrice que celle du début de cette section :

$$A = [\mathbf{a}_1\, \mathbf{a}_2\, \mathbf{a}_3\, \mathbf{a}_4, \mathbf{a}_5] = \begin{pmatrix} 1 & 2 & 0 & 3 & -4 \\ 0 & -2 & 2 & 1 & 1 \\ 1 & 5 & -3 & 1 & -5 \end{pmatrix}.$$

Après réduction,

$$\widetilde{A} = \begin{pmatrix} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

Comme les colonnes $1, 2$ et 4 de $\widetilde{A}$ sont celles contenant des pivots, on conclut que $\mathcal{B} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4)$ est une base de $\mathrm{Col}(A)$. En particulier, $\dim(\mathrm{Im}(A)) = 3$. $\diamond$

5.4 Une base pour $\ker(A)$

Rappelons que le noyau d'une application $T : \mathbb{R}^n \to \mathbb{R}^m$ associée à une matrice A de taille $m \times n$ est

$$\ker(T) = \ker(A) = \left\{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\right\}.$$

Le noyau étant un sous-espace vectoriel de $\mathbb{R}^n$ (l'ensemble de départ), il est important de trouver une base pour le décrire.

Voyons comment le calcul mène en général directement à une base du noyau, sur un exemple concret. Il est bien important de comprendre la méthode utilisée dans ce cas particulier, car elle sera exploitée dans la preuve du théorème énoncé plus bas :

Exemple 5.19. Calculons le noyau $\ker(A)$ de la matrice d'avant,

$$A = [\mathbf{a}_1\,\mathbf{a}_2\,\mathbf{a}_3\,\mathbf{a}_4,\mathbf{a}_5] = \begin{pmatrix} 1 & 2 & 0 & 3 & -4 \\ 0 & -2 & 2 & 1 & 1 \\ 1 & 5 & -3 & 1 & -5 \end{pmatrix}.$$

Comme on cherche les $\mathbf{x}$ tels que $A\mathbf{x} = \mathbf{0}$, qui est équivalent à $\widetilde{A}\mathbf{x} = \mathbf{0}$, on utilise la réduite déjà calculée,

$$\widetilde{A} = \begin{pmatrix} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

On en déduit la présence de deux variables libres dans le système $\widetilde{A}\mathbf{x} = \mathbf{0}$, x_3 et x_5. Les autres composantes s'expriment en fonction de $x_3 = s$ et $x_5 = t$:

$$\begin{cases} x_1 &= -2s - t \\ x_2 &= s + t \\ x_4 &= t \end{cases}$$

Maintenant, écrivons explicitement la dépendance en s et t, en mettant ces variables *en évidence* :

$$\begin{aligned} \ker(A) &= \left\{ \mathbf{x} \in \mathbb{R}^5 \,\middle|\, A\mathbf{x} = \mathbf{0} \right\} \\ &= \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -2s - t \\ s + t \\ s \\ t \\ t \end{pmatrix} \,\middle|\, s, t \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = s \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \,\middle|\, s, t \in \mathbb{R} \right\} \end{aligned}$$

Comme les vecteurs

$$\mathbf{v}_1 = \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

sont indépendants et engendrent le noyau, ils forment une base de $\ker(A)$. En particulier, $\dim(\ker(A)) = 2$. ◇

Dans ce dernier exemple, nous avons vu apparaître deux variables libres, qui ont donné lieu à deux vecteurs qui formaient directement une base pour le noyau. Il se trouve que **ce procédé mène toujours directement à une base du noyau**.

Théorème 5.20. *Pour toute matrice A, la dimension du noyau $\ker(A)$ est égale au nombre de variables libres apparaissant dans le système $A\mathbf{x} = \mathbf{0}$. (De plus, la méthode directe utilisée dans l'exemple précédent mène toujours à une base du noyau.)*

Preuve: Supposons que A est $m \times n$. Supposons que les variables $x_{i_1}, \dots, x_{i_k}$ soient libres. Lorsqu'on met ces variables en évidence, comme dans l'exemple ci-dessus, à chacune de ces variables x_{i_j} sera associée un vecteur

$\mathbf{v}_j \in \mathbb{R}^n$. Or ces vecteurs possèdent la propriété suivante : pour tout $1 \leqslant j \leqslant k$, la i_j-ème composante de $\mathbf{v}_j$ est un 1, alors que ses composantes $i_{j'}$, pour $j' \neq j$, sont nulles. Ceci implique que la famille $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ est libre. Puisqu'elle engendre $\ker(A)$, elle forme une base du noyau. Ceci implique que $\dim(\ker(A)) = k$, le nombre de variables libres. □

5.5 Transposition

L'opération de **transposition**, pour une matrice, est une opération qui consiste à transformer ses colonnes en lignes. Elle ne sera utilisée que plus tard dans le cours, mais nous la définissons déjà ici, et présentons ses propriétés.

Définition 5.21. Soit A une matrice $m \times n$. La **transposée de** A, notée A^T, est la matrice $n \times m$ dont les éléments sont définis par

$$(A^T)_{ij} := A_{ji}\,.$$

Exemple 5.22. Si A est 2×3, donnée par

$$A = \begin{pmatrix} \alpha & \beta & \gamma \\ \delta & \mu & \varepsilon \end{pmatrix}\,,$$

alors A^T est 3×2, donnée par

$$A^T = \begin{pmatrix} \alpha & \delta \\ \beta & \mu \\ \gamma & \varepsilon \end{pmatrix}\,.$$

◇

Exemple 5.23. Pour une matrice carrée, la transposition revient à refléter ses coefficients à travers la diagonale :

$$A = \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{pmatrix} \quad \Longrightarrow \quad A^T = \begin{pmatrix} a & e & i & m \\ b & f & j & n \\ c & g & k & o \\ d & h & l & p \end{pmatrix}\,.$$

◇

Proposition 4.
1) $(A^T)^T = A$
2) $(\lambda A)^T = \lambda A^T$ *pour tout scalaire* λ
3) *Pour toute paire* A, B *(de mêmes dimensions),*

$$(A + B)^T = A^T + B^T\,.$$

Preuve: Suivent de la définition. □

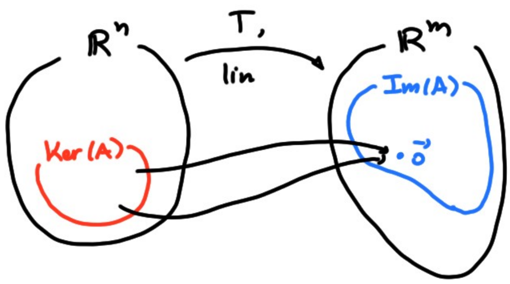
Transposition de vecteurs

Pour des raisons de commodité, on utilisera souvent le fait suivant : si un vecteur de $\mathbb{R}^n$ est vu comme une matrice $n \times 1$ (un "vecteur colonne"), on peut également lui appliquer l'opération de transposition, et le transformer en une matrice $1 \times n$ (un "vecteur ligne") :

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \Longrightarrow \quad \mathbf{x}^T = [x_1\, x_2 \cdots x_n]\,.$$

5.6 Le Théorème du Rang

Considérons une matrice $m \times n$, A, et l'application linéaire associée, $T(\mathbf{x}) = A\mathbf{x}$:



On a déjà dit que

- $\star$ $\ker(A)$ est un sous-espace vectoriel de $\mathbb{R}^n$,
- $\star$ $\mathrm{Im}(A)$ est un sous-espace vectoriel de $\mathbb{R}^m$.

Dans les sections précédentes, nous avons vu comment obtenir des bases pour ces sous-espaces. Ici, nous allons compléter cette analyse en faisant quelques remarques sur les *dimensions* de ces espaces.

Commençons par faire une remarque sur un cas particulier :

Exemple 5.24. Considérons l'application linéaire $T : \mathbb{R}^5 \to \mathbb{R}^3$ rencontrée dans les sections précédentes, dont la matrice est

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

et dont la réduite est

$$\widetilde{A} = \begin{pmatrix} {\color{blue}1} & 0 & 2 & 0 & 1 \\ 0 & {\color{blue}1} & -1 & 0 & -1 \\ 0 & 0 & 0 & {\color{blue}1} & -1 \end{pmatrix}$$

Rappelons ce que nous avons déjà dit :

- $\star$ Les colonnes $1, 2, 4$ de $\widetilde{A}$ contiennent des pivots, ce qui implique que les colonnes $1, 2, 4$ de A sont des colonnes-pivot et forment une base de $\mathrm{Im}(A)$, ce qui implique que
$$\dim(\mathrm{Im}(A)) = 3\,.$$
- $\star$ Les variables x_3, x_5 sont libres, ce qui implique (voir théorème de la section précédente) que
$$\dim(\ker(A)) = 2\,.$$

Par conséquent,

$$\dim(\ker(A)) + \dim(\mathrm{Im}(A)) = 2 + 3 = 5\,.$$

Ici, "5" est également la dimension de l'espace de départ ($\mathbb{R}^5$), qui est également égal au nombre de colonnes de A. $\diamond$

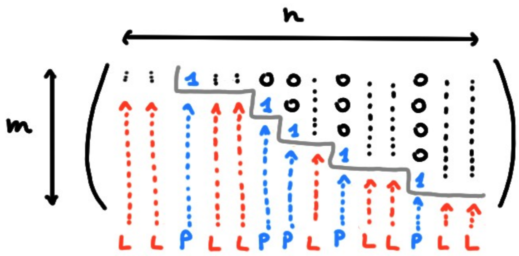
Ce que nous venons d'observer est en fait vrai pour toute matrice : la somme des dimensions de l'ensemble image et du noyau est toujours égale à la dimension de l'espace de départ. C'est le **Théorème du rang** :

Théorème 5.25. *Soit A une matrice $m \times n$. Alors*

$$\dim(\ker(A)) + \dim(\mathrm{Im}(A)) = n\,.$$

Pour le même résultat, mais démontré plus généralement dans le cadre des espaces vectoriels, voir **ici** (lien web).

Preuve: La structure générale d'une matrice réduite sera toujours du type suivant :



- ⋆ Le nombre de colonnes contenant un pivot (au nombre de 5 en bleu sur l'image) donne le nombre d'éléments contenus dans une base de $\mathrm{Im}(A)$, et donc est égal à $\dim(\mathrm{Im}(A))$.
- ⋆ Ensuite toutes les autres colonnes (au nombre de 9 en rouge sur l'image) représentent des variables libres, et donnent donc la dimension du noyau, $\dim(\ker(A))$. Comme il y a en tout n colonnes ($n = 14$ sur l'image), on a bien
$$\dim(\mathrm{Im}(A)) + \dim(\ker(A)) = n\,.$$

□

Le terme "rang" doit encore être défini :

Définition 5.26. Soit A une matrice $m \times n$. Le **rang** de A est défini comme la dimension de son ensemble image :
$$\mathrm{rang}(A) := \dim(\mathrm{Im}(A))\,.$$

Parfois, le rang est aussi noté $\mathrm{rg}(A)$. (En anglais : $\mathrm{rank}(A)$.)

Si A est $m \times n$, alors

1) $\mathrm{rang}(A) \leqslant m$. En effet, l'ensemble image de A étant un sous-espace vectoriel de $\mathbb{R}^m$, sa dimension est au plus égale à m.
2) $\mathrm{rang}(A) \leqslant n$. En effet, la dimension de l'ensemble image est au plus égale au nombre de colonnes de A.

Par conséquent,
$$\mathrm{rang}(A) \leqslant \min\{m, n\}\,.$$

Informel 5.27. Plus le rang d'une matrice $m \times n$ est grand, plus cette matrice définit une application qui "remplit" son ensemble d'arrivée. En particulier, si l'application est surjective, alors son rang vaut m.

Voyons quelques exemples d'utilisation simple du théorème du rang.

Exemple 5.28. Soit A une matrice 6×9. Alors $\ker(A)$ a dimension au moins égale à 3. En effet, $\mathrm{rang}(A) \leqslant \min\{6, 9\} = 6$, et donc par le théorème du rang,

$$\dim(\ker(A)) = 9 - \mathrm{rang}(A) \geqslant 9 - 6 = 3\,.$$

◇

L'espace engendré par les lignes

Nous avons déjà souvent décrit un matrice $m \times n$ à l'aide de ses colonnes $\mathbf{a}_k \in \mathbb{R}^m$:

$$A = [\mathbf{a}_1 \cdots \mathbf{a}_n] .$$

Mais on peut aussi aussi la décrire à l'aide de ses lignes,

$$A = \begin{pmatrix} \ell_1^T \\ \vdots \\ \ell_m^T \end{pmatrix} ,$$

où $\ell_1, \dots, \ell_m$ sont des vecteurs de $\mathbb{R}^n$. En d'autres termes, les lignes de A sont les colonnes de A^T :

$$A^T = [\ell_1 \cdots \ell_m]$$

Exemple 5.29. $A = \begin{pmatrix} 1 & 0 & 2 \\ -4 & 3 & 5 \end{pmatrix}$ peut s'écrire $A = \begin{pmatrix} \ell_1^T \\ \ell_2^T \end{pmatrix}$, où

$$\ell_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} , \qquad \ell_2 = \begin{pmatrix} -4 \\ 3 \\ 5 \end{pmatrix} .$$

◇

Définition 5.30. Soit A une matrice $m \times n$, dont les lignes sont $\ell_1^T, \dots, \ell_m^T$. Alors l'**espace-ligne de** A est le sous-espace vectoriel de $\mathbb{R}^n$ engendré par ses lignes :

$$\operatorname{Lgn}(A) := \operatorname{Vect}\{\ell_1, \dots, \ell_m\} .$$

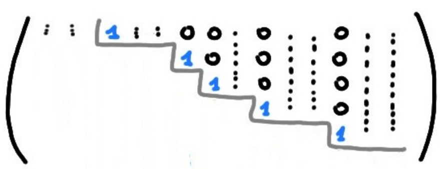
Lemme 11. *Si A et B sont deux matrices équivalentes selon les lignes (on peut passer de l'une à l'autre à l'aide d'un nombre fini d'opérations élémentaires sur les lignes), alors*

$$\operatorname{Lgn}(A) = \operatorname{Lgn}(B) .$$

Preuve: Supposons que B peut s'obtenir par une suite d'opérations élémentaires sur les lignes. Alors toute combinaison linéaire des lignes de B est aussi une combinaison linéaire des lignes de A. Ceci implique $\operatorname{Lgn}(B) \subset \operatorname{Lgn}(A)$. Le même argument montre que $\operatorname{Lgn}(A) \subset \operatorname{Lgn}(B)$, ce qui entraîne $\operatorname{Lgn}(A) = \operatorname{Lgn}(B)$. □

Corollaire 4. *Si $\widetilde{A}$ est la réduite de A, alors les lignes de $\widetilde{A}$ contenant un pivot (s'il y en a) forment une base de* $\operatorname{Lgn}(\widetilde{A})$ ***et de*** $\operatorname{Lgn}(A)$.

Preuve: Regardons $\widetilde{A}$:



Les lignes contenant un pivot possèdent des "1" à des emplacements différents, précédés de "0" : elles sont donc clairement indépendantes. Puisqu'elles engendrent évidemment $\operatorname{Lgn}(\widetilde{A})$, elles forment une base de $\operatorname{Lgn}(\widetilde{A})$.

Par le lemme précédent, toute famille de vecteurs qui forme une base de $\operatorname{Lgn}(\widetilde{A})$ forme aussi une base de $\operatorname{Lgn}(A)$. □

Intéressons-nous maintenant à la dimension de l'espace engendré par les lignes. Par définition,

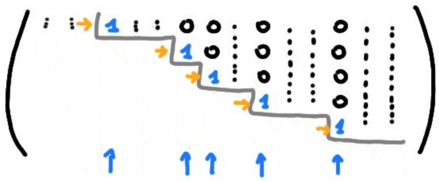
$$\dim(\mathrm{Lgn}(A)) = \mathrm{rang}(A^T).$$

Le résultat suivant montre que les espaces engendrés par les colonnes et les lignes d'une matrice quelconque ont toujours même dimension :

Théorème 5.31. *Si A est une matrice quelconque,*

$$\mathrm{rang}(A) = \mathrm{rang}(A^T)\,.$$

Preuve: Soit $\widetilde{A}$ la réduite de A. La chose importante à remarquer est que dans $\widetilde{A}$, *le nombre de colonnes contenant un pivot est égal au nombre de lignes non nulles*. C'est évident sur un dessin :



On peut donc écrire

$$\begin{aligned}\mathrm{rang}(A) &= \text{nombre de colonnes-pivot de } A\\ &= \text{nombre de colonnes contenant un pivot dans } \widetilde{A}\\ &= \text{nombre de lignes non-nulles dans } \widetilde{A}\\ &= \dim(\mathrm{Lgn}(\widetilde{A}))\\ &= \dim(\mathrm{Lgn}(A))\\ &= \mathrm{rang}(A^T)\,.\end{aligned}$$

Dans la quatrième ligne, on a utilisé le corollaire ci-dessus. Dans la cinquième ligne, on a utilisé le lemme du dessus. □

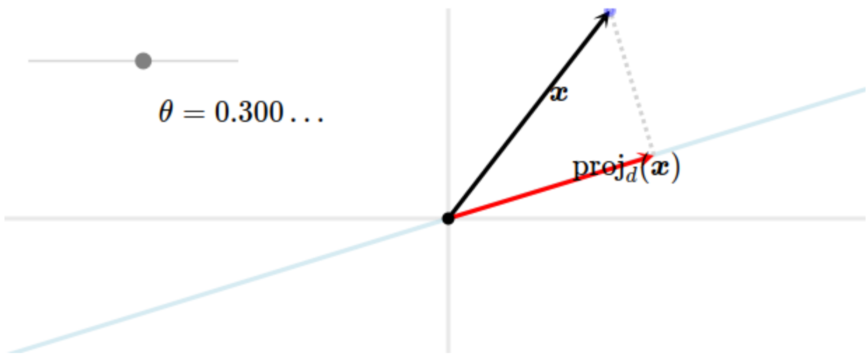
5.7 Transformations géométriques

Dans cette section, on laisse de côté la théorie générale pour considérer quelques exemples importants d'applications linéaires $T : \mathbb{R}^2 \to \mathbb{R}^2$, tous de nature *géométrique*.

Sur ces exemples, on illustrera certaines des notions vues dans les sections précédentes (ensemble image, noyau, etc.), en leur donnant un sens géométrique. On considérera aussi le matrices associées à ces applications relativement à la base canonique, et plus loin relativement à d'autres bases.

Projection sur un axe de $\mathbb{R}^2$

Fixons une droite d dans le plan, passant par l'origine, et considérons la transformation consistant à *projeter un vecteur* $\mathbf{x} \in \mathbb{R}^2$ *orthogonalement sur* d :



Cette opération définit une application

$$\begin{aligned}\mathrm{proj}_d : \mathbb{R}^2 &\mapsto \mathbb{R}^2\\ \mathbf{x} &\mapsto \mathrm{proj}_d(\mathbf{x})\,.\end{aligned}$$

Quelques remarques à propos de cette application :

Par définition de la projection, tout vecteur $\mathbf{v}$ appartenant à d (ou plutôt : colinéaire à un vecteur directeur quelconque de d) ne change pas lorsqu'il est projeté :

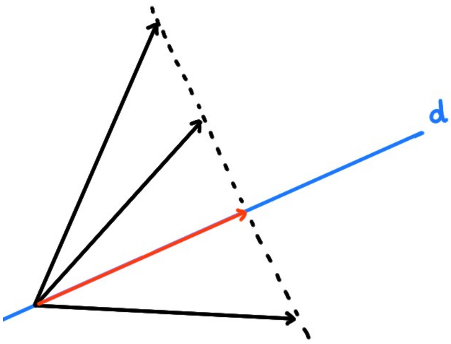
$$\mathrm{proj}_d(\mathbf{v}) = \mathbf{v}\,.$$

Ceci implique en particulier que $\mathrm{Im}(\mathrm{proj}_d) \supset d$. Mais par définition, $\mathrm{Im}(\mathrm{proj}_d) \subset d$, et donc

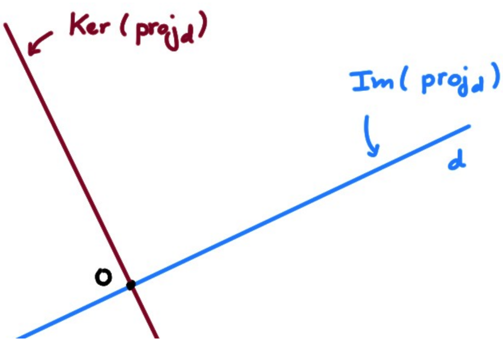
$$\mathrm{Im}(\mathrm{proj}_d) = d\,.$$

Puisque d est un sous-ensemble stricte de $\mathbb{R}^2$, ceci implique que proj_d *n'est pas surjective*.

Ensuite, proj_d *n'est pas injective*, puisqu'il existe une infinité de vecteurs différents dont la projection sur d est la même :



Effectivement le noyau contient (en plus du vecteur nul) une infinité de vecteurs, tous sur la droite d' perpendiculaire à d, passant par l'origine :

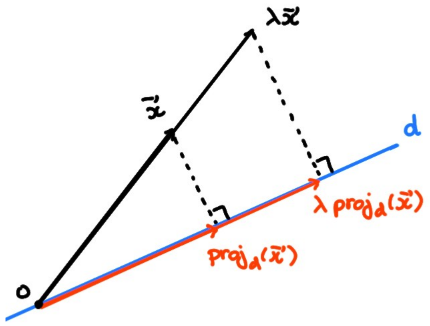


Ceci illustre aussi le théorème du rang :

$$\dim(\ker(\mathrm{proj}_d)) + \dim(\mathrm{Im}(\mathrm{proj}_d)) = 1 + 1 = 2 = \dim(\mathbb{R}^2)\,.$$

Insistons sur le fait que les propriétés décrites ci-dessus ont toutes été obtenues *sans calculs*.

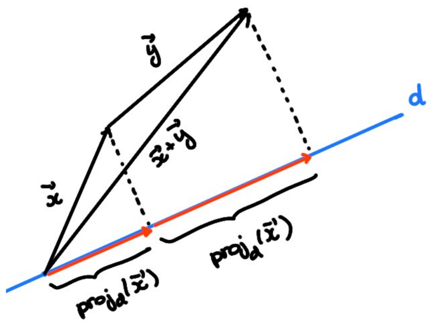
Maintenant, la nature géométrique de la projection permet de montrer sans peine qu'elle est *linéaire*. En effet, si on multiplie $\mathbf{x}$ par un scalaire λ, sa projection est multipliée par le même λ :



En d'autres termes :

$$\mathrm{proj}_d(\lambda\mathbf{x}) = \lambda\mathrm{proj}_d(\mathbf{x})\,.$$

Ensuite, si on additionne deux vecteurs et qu'ensuite on projette leur somme, on obtient le même résultat que si on les avait d'abord projetés séparément pour ensuite les additionner :

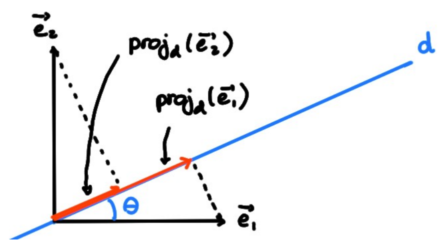


Plus précisément :

$$\mathrm{proj}_d(\mathbf{x}+\mathbf{y}) = \mathrm{proj}_d(\mathbf{x}) + \mathrm{proj}_d(\mathbf{y})\,.$$

Maintenant, puisque proj_d est linéaire, elle peut être représentée à l'aide d'une matrice. Celle-ci est donnée par

$$A = [\mathrm{proj}_d]_{\mathcal{B}_{\mathrm{can}}} = \big[\mathrm{proj}_d(\mathbf{e}_1)\,\mathrm{proj}_d(\mathbf{e}_2)\big]\,.$$



Si on suppose que d fait un angle θ avec $\mathbf{e}_1$ (dans le sens anti-horaire), on trouve

$$\begin{aligned}\mathrm{proj}_d(\mathbf{e}_1) &= \cos(\theta)\begin{pmatrix}\cos(\theta)\\ \sin(\theta)\end{pmatrix}\\ \mathrm{proj}_d(\mathbf{e}_2) &= \sin(\theta)\begin{pmatrix}\cos(\theta)\\ \sin(\theta)\end{pmatrix},\end{aligned}$$

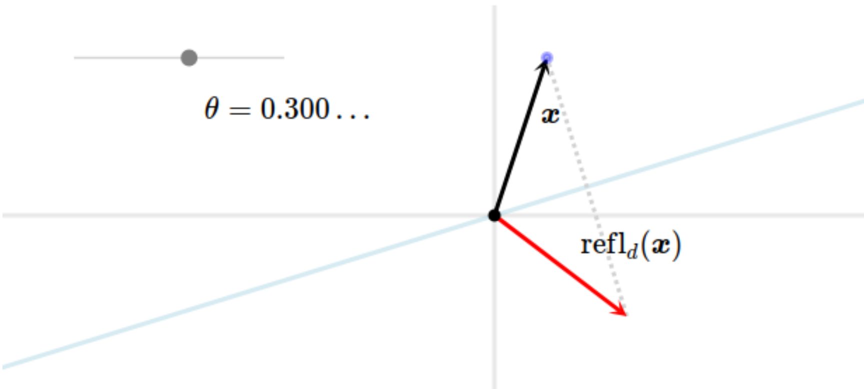
et donc la matrice associée à proj_d est

$$[\mathrm{proj}_d]_{\mathcal{B}_{can}} = \begin{pmatrix}\cos^2(\theta) & \cos(\theta)\sin(\theta)\\ \cos(\theta)\sin(\theta) & \sin^2(\theta)\end{pmatrix}$$

Comme les colonnes sont toutes deux colinéaires au vecteur directeur de d, elles n'engendrent pas $\mathbb{R}^2$, ce qui reflète le fait que proj_d n'est ni injective, ni surjective.

Réflexion à travers un axe de $\mathbb{R}^2$

Reprenons encore une droite d passant par l'origine, et considérons cette fois la transformation consistant à *réfléchir un vecteur* $\mathbf{x} \in \mathbb{R}^2$ *à travers* d. La **réflexion de x à travers** d sera notée $\mathrm{refl}_d(\mathbf{x})$:



Cette opération définit une application

$$\begin{aligned}\mathrm{refl}_d : \mathbb{R}^2 &\mapsto \mathbb{R}^2\\ \mathbf{x} &\mapsto \mathrm{refl}_d(\mathbf{x}).\end{aligned}$$

Quelques remarques :

- ⋆ Par construction, tout vecteur $\mathbf{v}$ appartenant à d (ou plutôt : colinéaire à un vecteur directeur quelconque de d) est invariant sous l'action de la réflexion :

$$\mathrm{refl}_d(\mathbf{v}) = \mathbf{v}.$$

- ⋆ Clairement, le réfléchi du réfléchi de $\mathbf{x}$ est $\mathbf{x}$ lui-même :

$$\mathrm{refl}_d(\mathrm{refl}_d(\mathbf{x})) = \mathbf{x},$$

ce qui implique que refl_d est bijective et qu'elle est égale à sa réciproque :

$${\mathrm{refl}_d}^{-1} = \mathrm{refl}_d.$$

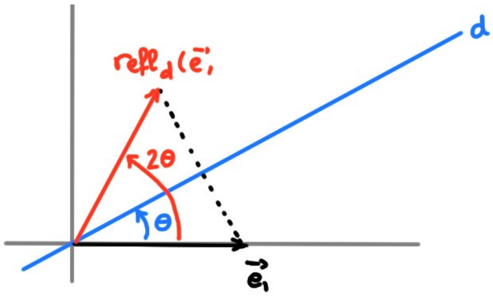
Étant injective, son noyau $\ker(\mathrm{refl}_d) = \{\mathbf{0}\}$. Étant surjective, $\mathrm{Im}(\mathrm{refl}_d) = \mathbb{R}^2$.

$$\dim(\ker(\mathrm{refl}_d)) + \dim(\mathrm{Im}(\mathrm{refl}_d)) = 0 + 2 = 2 = \dim(\mathbb{R}^2).$$

Comme pour la projection, on montre sans peine que refl_d est une application linéaire. Calculons sa matrice relativement à la base canonique :

$$[\mathrm{refl}_d]_{\mathcal{B}_{\mathrm{can}}} = \big[\mathrm{refl}_d(\mathbf{e}_1)\,\mathrm{refl}_d(\mathbf{e}_2)\big]\,.$$

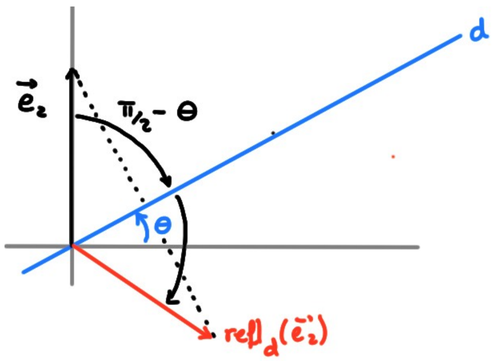
Si encore une fois on suppose que d fait un angle θ avec la direction $\mathbf{e}_1$, alors on remarque que la réflexion de $\mathbf{e}_1$ à travers d le transforme en un vecteur unitaire faisant un angle de 2θ avec l'horizontale :



On a donc

$$\mathrm{refl}_d(\mathbf{e}_1) = \begin{pmatrix}\cos(2\theta)\\ \sin(2\theta)\end{pmatrix}\,.$$

Ensuite, la réflexion de $\mathbf{e}_2$ à travers d le transforme en un vecteur unitaire faisant un angle de $\theta - (\frac{\pi}{2} - \theta) = 2\theta - \frac{\pi}{2}$ avec l'horizontale :



On a donc

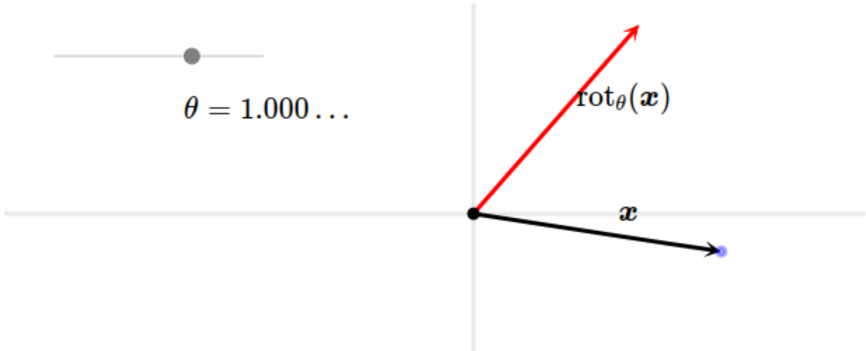
$$\mathrm{refl}_d(\mathbf{e}_2) = \begin{pmatrix}\cos(2\theta - \frac{\pi}{2})\\ \sin(2\theta - \frac{\pi}{2})\end{pmatrix}\,, = \begin{pmatrix}\sin(2\theta)\\ -\cos(2\theta)\end{pmatrix}\,.$$

Ainsi, la matrice qui représente refl_d relativement à la base canonique est donnée par

$$[\mathrm{refl}_d]_{\mathcal{B}_{\mathrm{can}}} = \begin{pmatrix}\cos(2\theta) & \sin(2\theta)\\ \sin(2\theta) & -\cos(2\theta)\end{pmatrix}\,.$$

Rotation d'angle θ autour de l'origine dans $\mathbb{R}^2$

Considérons une **rotation d'angle θ autour de l'origine** (dans le sens trigonométrique) :



Cette opération définit une application

$$\begin{aligned}\mathrm{rot}_\theta : \mathbb{R}^2 &\mapsto \mathbb{R}^2\\ \mathbf{x} &\mapsto \mathrm{rot}_\theta(\mathbf{x})\,.\end{aligned}$$

Quelques remarques :

- ⋆ Si $\theta = 0$ ou un multiple de 2π, la rotation correspond à l'identité.
- ⋆ Puisque
$$\mathrm{rot}_{-\theta}(\mathrm{rot}_\theta(\mathbf{x})) = \mathbf{x}\,,$$
la rotation d'angle θ est bijective, et sa réciproque est la rotation d'angle $-\theta$:
$$\mathrm{rot}_\theta{}^{-1} = \mathrm{rot}_{-\theta}\,.$$
Étant injective, son noyau $\ker(\mathrm{rot}_\theta) = \{\mathbf{0}\}$. Étant surjective, $\mathrm{Im}(\mathrm{rot}_\theta) = \mathbb{R}^2$.
$$\dim(\ker(\mathrm{rot}_\theta)) + \dim(\mathrm{Im}(\mathrm{rot}_\theta)) = 0 + 2 = 2 = \dim(\mathbb{R}^2)\,.$$

Une rotation (autour de l'origine) est clairement une transformation linéaire, et puisque

$$[\mathrm{rot}_\theta(\mathbf{e}_1)]_{\mathcal{B}_{\mathrm{can}}} = \begin{pmatrix}\cos(\theta)\\ \sin(\theta)\end{pmatrix}, \qquad [\mathrm{rot}_\theta(\mathbf{e}_2)]_{\mathcal{B}_{\mathrm{can}}} = \begin{pmatrix}-\sin(\theta)\\ \cos(\theta)\end{pmatrix},$$

sa matrice relativement à la base canonique est donnée par

$$[\mathrm{rot}_\theta]_{\mathcal{B}_{\mathrm{can}}} = \begin{pmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{pmatrix}.$$

Chapitre 6

Le produit matriciel

6.1 Définition

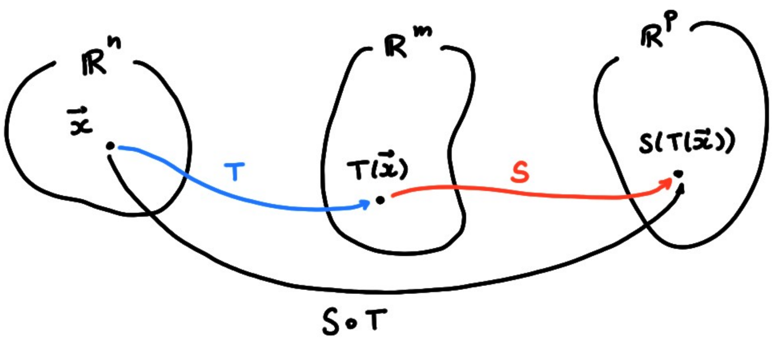
Le *produit* de deux matrices est motivé par la *composition* d'applications linéaires.

Or lorsqu'on veut *composer* deux applications, il faut que les ensembles qui apparaissent dans leurs définitions soient compatibles.

⋆ Soit donc
$$T : \mathbb{R}^n \to \mathbb{R}^m$$
une application linéaire, dont la matrice $m \times n$ est notée A. Si $\mathbf{x} \in \mathbb{R}^n$, la k-ème composante $(1 \leqslant k \leqslant m)$ de $T(\mathbf{x})$ est donnée par
$$(T(\mathbf{x}))_k = (A\mathbf{x})_k = \sum_{j=1}^{n} a_{kj} x_j \,.$$

⋆ Soit ensuite
$$S : \mathbb{R}^m \to \mathbb{R}^p$$
une autre application linéaire, dont la matrice $p \times m$ est notée B. Si $\mathbf{x} \in \mathbb{R}^m$, la k-ème composante $(1 \leqslant k \leqslant p)$ de $S(\mathbf{x})$ est donnée par
$$(S(\mathbf{x}))_k = (B\mathbf{x})_k = \sum_{j=1}^{m} b_{kj} x_j \,.$$

Puisque l'ensemble d'arrivée de T est l'ensemble de départ de S, on peut les composer :



La **composition** est définie par

$$
\begin{aligned}
S \circ T : \mathbb{R}^n &\to \mathbb{R}^p \\
\mathbf{x} &\mapsto (S \circ T)(\mathbf{x}) := S(T(\mathbf{x}))\,.
\end{aligned}
$$

Informel 6.1. Attention, même si on lit le symbole "$S \circ T$" de gauche à droite, en disant "*S composée avec T*", c'est pourtant T que l'on applique en premier, suivie de S !

On peut vérifier (exercice !) que la composée $S \circ T$ est linéaire ; elle peut donc être représentée par une matrice. Quelle est cette matrice ?

Calculons la k-ème composante de $(S \circ T)(\mathbf{x})$:

$$
\begin{aligned}
(S \circ T)(\mathbf{x})_k = (S(T(\mathbf{x})))_k &= (B(A\mathbf{x}))_k \\
&= \sum_{j=1}^{m} b_{kj} (A\mathbf{x})_j \\
&= \sum_{j=1}^{m} b_{kj} \sum_{l=1}^{n} a_{jl} x_l \\
&= \sum_{l=1}^{n} \underbrace{\Big(\sum_{j=1}^{m} b_{kj} a_{jl} \Big)}_{=:c_{kl}} x_l\,.
\end{aligned}
$$

On voit qu'après avoir interverti les sommes sur j et l, on a pu définir des coefficients c_{kl}, qui sont les coefficients d'une matrice $p \times n$, notée C, qui permet d'écrire

$$
(S \circ T)(\mathbf{x})_k = \sum_{l=1}^{n} c_{kl} x_l = (C\mathbf{x})_k\,.
$$

On a donc trouvé la matrice associée à $S \circ T$, et on sait calculer ses coefficients en fonction de ceux de A et B.

Définition 6.2. Soient $B = (b_{ij})$ une matrice $p \times m$, et $A = (a_{ij})$ une matrice $m \times n$. Le **produit matriciel de B par A** est la matrice $p \times n$, notée $C = BA$, dont les coefficients sont définis par

$$
c_{kl} := \sum_{j=1}^{m} b_{kj} a_{jl}\,.
$$

L'expression ci-dessus pour le coefficient c_{kl} montre que ce dernier se calcule en parcourant la k-ème ligne de A et la l-ème colonne de B.

Exemple 6.3. Calculons un produit $BA = C$, pour des matrices 4×4 :

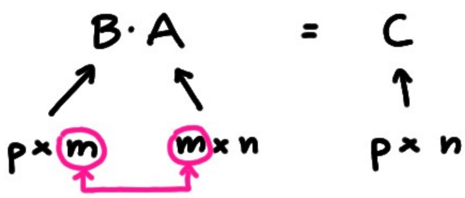
$$
\underbrace{\begin{pmatrix} 1 & 2 & -2 & 0 \\ 3 & 2 & -2 & 1 \\ -1 & 1 & 0 & 1 \\ 5 & 2 & -1 & 6 \end{pmatrix}}_{B}
\underbrace{\begin{pmatrix} 0 & 2 & -1 & 5 \\ 3 & 1 & 5 & -3 \\ 2 & 2 & -1 & 1 \\ 0 & 7 & 1 & 2 \end{pmatrix}}_{A}
=
\underbrace{\begin{pmatrix} 2 & 0 & 11 & -3 \\ 2 & 11 & \mathbf{10} & 9 \\ 3 & 6 & 7 & -6 \\ 4 & 52 & 12 & 30 \end{pmatrix}}_{C}
$$

Comme exemple, on a indiqué le calcul de

$$\begin{aligned} c_{23} &= \sum_{j=1}^{4} b_{2j} a_{j3} \\ &= b_{21}a_{13} + b_{22}a_{23} + b_{23}a_{33} + b_{24}a_{43} \\ &= 3 \cdot (-1) + 2 \cdot 5 + (-2) \cdot (-1) + 1 \cdot 1 = \mathbf{10} \end{aligned}$$

◇

Informel 6.4. On peut multiplier deux matrices de dimensions différentes, BA, mais ces dimensions doivent être compatibles : le nombre de colonnes de B doit être égal au nombre de lignes de A.



Exemple 6.5. Le produit d'une 3×2 par une 2×4 est bien défini :

$$\underbrace{\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}}_{3\times 2} \underbrace{\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix}}_{2\times 4} = \underbrace{\begin{pmatrix} a+5b & 2a+6b & 3a+7b & 4a+8b \\ c+5d & 2c+6d & 3c+7d & 4c+8d \\ e+5f & 2e+6f & 3e+7f & 4e+8f \end{pmatrix}}_{3\times 4}$$

Par contre, dans l'ordre inverse, le produit

$$\underbrace{\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix}}_{2\times 4} \underbrace{\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}}_{3\times 2} \qquad \text{n'est pas défini !}$$

◇

Quelques remarques :

- ⋆ Un vecteur $\mathbf{x} \in \mathbb{R}^n$ peut s'interpréter comme une matrice $n \times 1$. Donc la multiplication d'une matrice $m \times n$ par $\mathbf{x} \in \mathbb{R}^n$, peut s'interpréter comme le produit matriciel d'une $m \times n$ par une $n \times 1$, qui donne une $A\mathbf{x}$ qui est $m \times 1$, c'est-à-dire un vecteur de $\mathbb{R}^m$.
- ⋆ Pour le produit d'une $1 \times n$ par une $n \times 1$, on obtient une matrice 1×1, qui n'est autre qu'un réel :

$$\underbrace{\begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}}_{1\times n} \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{n\times 1} = \underbrace{x_1y_1 + \cdots + x_ny_n}_{\in\mathbb{R}} .$$

- ⋆ Par contre, le produit d'une $m \times 1$ par une $1 \times n$ donne évidemment une $m \times n$. Par exemple,

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \begin{pmatrix} x & y & z & t \end{pmatrix} = \begin{pmatrix} ax & ay & az & at \\ bx & by & bz & bt \\ cx & cy & cz & ct \end{pmatrix}$$

⋆ Considérons le produit de A ($m \times n$) par B ($n \times p$). Si on exprime B à l'aide de ses p colonnes, qui sont des vecteurs de $\mathbb{R}^n$

$$B = [\mathbf{b}_1 \cdots \mathbf{b}_p]\,,$$

alors le produit AB peut s'écrire à l'aide de ses p colonnes :

$$AB = [A\mathbf{b}_1 \cdots A\mathbf{b}_p]\,,$$

où chaque colonne $A\mathbf{b}_k \in \mathbb{R}^m$

6.2 Propriétés

Proposition 5. *Le produit matriciel satisfait aux propriétés suivantes. (Ci-dessous, on suppose que les dimensions des matrices sont toujours compatibles.)*

1) $A(BC) = (AB)C$ *(⚠ **associativité**)*
2) $A(B + C) = AB + AC$ *(**distributivité**)*
3) $(A + B)C = AC + BC$ *(**distributivité**)*
4) $A(\lambda B) = \lambda(AB) = (\lambda A)B$
5) $(AB)^T = B^T A^T$

Preuve: Les premières propriétés seront vérifiées en exercices. Pour la dernière, considérons A de dimensions $m \times n$, et B de dimensions $n \times p$, et calculons l'élément ij de $(AB)^T$:

$$\begin{aligned}
((AB)^T)_{ij} &= (AB)_{ji} \\
&= \sum_{k=1}^{n} A_{jk} B_{ki} \\
&= \sum_{k=1}^{n} (A^T)_{kj} (B^T)_{ik} \\
&= \sum_{k=1}^{n} (B^T)_{ik} (A^T)_{kj} \\
&= (B^T A^T)_{ij}\,.
\end{aligned}$$

□

L'associativité signifie que l'on n'a pas besoin d'utiliser de parenthèses lorsqu'on multiplie plusieurs matrices : les produits peuvent s'effecteur dans n'importe quel ordre. Donc au lieu de $A(BC)$ ou $(AB)C$, on peut simplement écrire ABC.

Ce qu'on n'a pas le droit de faire, par contre, c'est de changer l'ordre des matrices dans un produit : **le produit matriciel n'est pas commutatif**. De fait, en général, même pour des matrices A, B de dimensions compatibles,

$$AB \neq BA\,.$$

En effet, commençons par remarquer que si A est $m \times n$, alors AB et BA sont toutes deux bien définies seulement si B est $n \times m$. Mais alors AB est $m \times m$ et BA est $n \times n$, donc AB et BA sont de dimensions différentes dès que $m \neq n$. Donc pour que les deux matrices AB et BA soient toutes les deux définies et égales, il faut déjà que A et B soient **carrées**, de même dimension $n \times n$.

Or même si A et B sont carrées et de mêmes dimensions, en général $AB \neq BA$.

Exemple 6.6. Avec $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, on a

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix},$$
$$BA = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

et donc $AB \neq BA$. ◇

Définition 6.7. Si A, B sont telles que $AB = BA$, on dit qu'elles **commutent**.

Mentionnons encore une différence importante qui distingue le calcul matriciel du calcul réel. On sait que dans les réels, un produit nul

$$ab = 0$$

implique qu'au moins un des nombres a, b est nul. Par contre, on peut avoir un produit matriciel nul,

$$AB = 0,$$

sans qu'aucune des matrices A, B ne soit identiquement nulle (voir exercices).

Matrice identité

Définition 6.8. La **matrice identité** I_n est la matrice $n \times n$ définie par

$$I_n := \mathrm{diag}(1, 1, \dots, 1) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

L'action de I_n sur un vecteur n'a aucun effet :

$$I_n \mathbf{x} = \mathbf{x}, \qquad \forall \mathbf{x} \in \mathbb{R}^n.$$

De plus, la matrice identité est **l'élément neutre** pour la multiplication des matrices carrées $n \times n$, puisqu'on a, pour toute matrice A ($n \times n$),

$$AI_n = I_n A = A.$$

I_n est d'ailleurs la seule matrice qui commute avec toute matrice $n \times n$ (voir exercices).

Chapitre 7

Inversion

7.1 Motivation

Un des axiomes qui définit le corps des nombres réels est qu'il existe pour tout réel $a \neq 0$ un *inverse*, à savoir un nombre noté a^{-1} tel que

$$aa^{-1} = a^{-1}a = 1\,,$$

ou le nombre "1" est l'élément neutre pour la multiplication dans les réels (c'est-à-dire que $x \cdot 1 = 1 \cdot x = x$ pour tout $x \in \mathbb{R}$). C'est à l'aide de la notion d'inverse que l'on résout une équation du genre

$$ax = b\,,$$

où $a \neq 0$. En effet, en multipliant des deux côtés de l'équation par a^{-1}, on trouve

$$\underbrace{a^{-1}a}_{=1} x = a^{-1}b\,,$$

et donc $x = a^{-1}b$.

Pour les matrices, on aimerait idéalement pouvoir résoudre un système linéaire

$$A\mathbf{x} = \mathbf{b}$$

de la même façon. En effet, si on sait qu'il existe une matrice A^{-1} telle que $A^{-1}A = I_n$, alors en multipliant à gauche des deux côtés de l'équation vectorielle ci-dessus,

$$\underbrace{A^{-1}A}_{=I_n} \mathbf{x} = A^{-1}\mathbf{b}\,,$$

qui donne $\mathbf{x} = A^{-1}\mathbf{b}$.

Cette approche peut sembler élégante, mais elle présuppose qu'il existe une matrice A^{-1} telle que $A^{-1}A = I_n$. Or une telle matrice n'existe pas toujours, comme nous verrons. En effet, pouvoir isoler $\mathbf{x}$, dans l'équation "$A\mathbf{x} = \mathbf{b}$", en multipliant juste par une matrice bien choisie, mène à une solution unique $\mathbf{x} = A^{-1}\mathbf{b}$, et implique en particulier que la solution du système $A\mathbf{x} = \mathbf{b}$ est unique, ce qui n'arrive que dans certains cas (Théorème "$0, 1, \infty$").

Dans ce chapitre, on se propose donc de chercher des conditions sur A qui garantissent l'existence de A^{-1} ; c'est le problème de *l'inversibilité*. Nous verrons aussi plusieurs façons d'obtenir une expression explicite pour A^{-1}.

7.2 Définition et propriétés de base

Voyons le problème d'un point de vue un peu plus général.

Pouvoir isoler $\mathbf{x}$ dans $T(\mathbf{x}) = A\mathbf{x} = \mathbf{b}$ signifie, en termes d'application linéaire, que l'on cherche à récupérer la préimage de $\mathbf{b}$. Pour que cette préimage soit bien définie et unique pour tout $\mathbf{b} \in \mathbb{R}^m$, il faut que T soit bijective.

Or nous avons vu qu'une application linéaire ne peut être bijective que si elle a des espaces de départ et d'arrivée de mêmes dimensions :

$$T : \mathbb{R}^n \to \mathbb{R}^n$$

On peut aussi montrer que dans ce cas, la réciproque $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ est également linéaire (exercice) ; on peut donc lui associer une matrice :

$$T^{-1}(\mathbf{y}) = B\mathbf{y}\,.$$

Les relations

$$\begin{aligned} T(T^{-1}(\mathbf{y})) &= \mathbf{y} \qquad \forall \mathbf{y} \in \mathbb{R}^m\,,\\ T^{-1}(T(\mathbf{x})) &= \mathbf{x} \qquad \forall \mathbf{x} \in \mathbb{R}^n\,. \end{aligned}$$

s'expriment donc, en termes des matrices et du produit matriciel,

$$\begin{aligned} AB\mathbf{y} &= \mathbf{y} \qquad \forall \mathbf{y} \in \mathbb{R}^m\,,\\ BA\mathbf{x} &= \mathbf{x} \qquad \forall \mathbf{x} \in \mathbb{R}^n\,. \end{aligned}$$

Or ces deux dernières conditions peuvent s'exprimer simplement par

$$AB = I_n\,, \qquad BA = I_n\,.$$

La matrice B sera appelée *inverse de* A.

Par la discussion précédente, on ne peut parler d'*inverse* que pour des matrices **carrées**, c'est à dire ayant autant de lignes que de colonnes.

Définition 7.1. Soit A une matrice $n \times n$.

- ⋆ Si il existe une matrice B, $n \times n$, telle que

$$AB = BA = I_n\,,$$

  on dit que A est **inversible**. La matrice B est alors appelée **inverse de** A ; on la note A^{-1} (au lieu de B).
- ⋆ Si A n'est pas inversible, elle est dite **singulière**.

Remarque 7.2. Puisque deux matrices A et B ne commutent a priori pas, la condition "$AB = BA = I_n$" représente en fait *deux* conditions, à savoir $AB = I_n$ et $BA = I_n$. ◇

Exemple 7.3. La matrice $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ est inversible. En effet, en définissant

$$B := \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}\,,$$

on remarque que

$$AB = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2\,,$$

et que

$$BA = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2 \,.$$

Donc A est inversible, et son inverse est $A^{-1} = B$. ◇

Dans cet exemple, on a juste *vérifié* que A était inversible en vérifiant que le produit de A avec B donnait bien la matrice identité. Mais en général, on aimerait des *critères* qui nous permettent d'étudier une matrice donnée A, de savoir si elle est inversible ou pas, et si oui de calculer son inverse.

Informel 7.4. Par exemple, la matrice

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

est-elle inversible ? Si oui, quel est son inverse ?

Exemple 7.5. La matrice $A = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$ est singulière. En effet, quelle que soit B (2×2), le coefficient $(AB)_{11}$ est toujours égal à 0, et donc AB ne peut pas être égale à I_2. Cet exemple montre qu'il ne suffit pas de ne pas être identiquement nulle pour ne pas être inversible. ◇

Listons encore quelques propriétés de base de la matrice inverse.

Proposition 6. *Soit A une matrice $n \times n$ inversible. Alors*

1) *L'inverse A^{-1} est unique,*
2) *A^{-1} est aussi inversible, et $(A^{-1})^{-1} = A$,*
3) *Pour tout scalaire $\lambda \neq 0$, λA est aussi inversible, et $(\lambda A)^{-1} = \frac{1}{\lambda} A^{-1}$,*
4) *A^T est aussi inversible, et $(A^T)^{-1} = (A^{-1})^T$.*

De plus, si M est une autre matrice $n \times n$ inversible, alors AM est inversible, et

$$(AM)^{-1} = M^{-1} A^{-1} \,.$$

Preuve:

1) Supposons qu'il existe deux matrices C, B telles que $AC = CA = I_n$, $AB = BA = I_n$. Alors
$$B = BI_n = B(AC) = (BA)C = I_n C = C \,.$$
2) En considérant A comme la "matrice de départ", l'inversibilité signifie que $AA^{-1} = A^{-1}A = I_n$. Or ces deux conditions peuvent aussi se lire en considérant A^{-1} comme la "matrice de départ", et elles nous disent bien que A^{-1} est inversible et que son inverse est égal à A.
3) Par simple vérification, en utilisant les propriétés de la multiplication d'une matrice par un scalaire,
$$(\lambda A)(\tfrac{1}{\lambda} A^{-1}) = (\lambda \tfrac{1}{\lambda})(AA^{-1}) = I_n$$
De même, $(\frac{1}{\lambda} A^{-1})(\lambda A) = I_n$.
4) Par simple vérification,
$$A^T (A^{-1})^T = (A^{-1} A)^T = I_n^T = I_n \,.$$
De même, $(A^{-1})^T A^T = I_n$.

Finalement, si M est aussi inversible, alors

$$(AM)(M^{-1}A^{-1}) = A(\underbrace{MM^{-1}}_{=I_n})A^{-1} = AA^{-1} = I_n\,,$$
$$(M^{-1}A^{-1})(AM) = M^{-1}(\underbrace{AA^{-1}}_{=I_n})M = MM^{-1} = I_n\,,$$

et donc AM est inversible et son inverse est $M^{-1}A^{-1}$. □

Inverse et résolution de systèmes $n \times n$

Considérons un système $n \times n$,

$$A\mathbf{x} = \mathbf{b}\,,$$

dans lequel la matrice A est inversible. On peut alors résoudre cette équation em multipliant les deux côtés de l'inégalité ci-dessus par A^{-1},

$$\underbrace{A^{-1}A}_{=I_n}\mathbf{x} = A^{-1}\mathbf{b}\,,$$

qui donne directement la solution

$$\mathbf{x} = A^{-1}\mathbf{b}\,.$$

Si cette méthode peut paraître élégante, elle a le désavantage (en plus de ne pouvoir être appliquée que lorsque A est inversible) d'être plus coûteuse en termes de calcul, puisqu'elle

7.3 Le cas 2×2

Avant de nous attaquer au problème général d'une matrice $n \times n$, attardons-nous sur le cas 2×2. Même si ce cas est le plus simple, il va nous permettre de présenter quelques notions qui seront réutilisées dans d'autres chapitres.

Considérons une matrice 2×2 quelconque :

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\,.$$

L'inversibilité de A va dépendre des valeurs des coefficients a, b, c, d bien-sûr, et l'avantage du cas 2×2 est qu'il y a une condition facilement exprimable en fonction de ces coefficients.

Théorème 7.6. $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *est inversible si et seulement si son* ***déterminant****, c'est-à-dire le nombre réel défini par*

$$\det(A) := ad - bc\,,$$

est différent de zéro. De plus, lorsque $\det(A) \neq 0$*, l'inverse de* A *est donné par*

$$A^{-1} = \frac{1}{\det(A)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\,.$$

Preuve: Supposons pour commencer que $\det(A) \neq 0$. Dans ce cas, la matrice A^{-1} de l'énoncé est bien définie, et on vérifie par un calcul direct que $AA^{-1} = A^{-1}A = I_2$. Comme l'inverse est unique, A^{-1} est bien l'inverse de A.

Supposons maintenant que $\det(A) = ad - bc = 0$. Si tous les coefficients de A sont nuls, elle ne peut pas être inversible (on peut la multiplier par n'importe quoi, on n'obtiendra jamais I_n !). Supposons donc qu'au moins un de ses coefficients est non-nul, par exemple $a \neq 0$. Dans ce cas, $ad - bc = 0$ implique $d = \frac{bc}{a}$, et donc

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & \frac{b}{a}a \\ c & \frac{b}{a}c \end{pmatrix},$$

ce qui implique que les colonnes de A sont colinéaires. Comme on sait, ceci implique que A n'est pas inversible. $\square$

Exemple 7.7. À titre d'illustration, considérons la matrice 2×2 déjà mentionnée au début du chapitre :

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

Son déterminant vaut $\det(A) = 1\cdot 4 - 2\cdot 3 = -2 \neq 0$, et donc A est inversible, et son inverse est donné par la formule du théorème :

$$A^{-1} = \frac{1}{-2}\begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix},$$

comme nous avions déjà vérifié.

Cette expression permet maintenant de résoudre n'importe quelle équation vectorielle impliquant A. En effet, le système

$$(*) \left\{ \begin{array}{rcrcl} x_1 & + & 2x_2 & = & b_1 \\ 3x_1 & + & 4x_2 & = & b_2 \end{array} \right.$$

se formule comme $A\mathbf{x} = \mathbf{b}$,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

En multipliant des deux côtés par A^{-1}, on obtient $\mathbf{x} = A^{-1}\mathbf{b}$, qui donne

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 3/2 & 1/2 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} -2b_1 + b_2 \\ \frac{3}{2}b_1 + \frac{1}{2}b_2 \end{pmatrix}.$$

$\diamond$

Exemple 7.8. Considérons quelques transformations linéaires dans le plan.

⋆ Nous avions remarqué que la **projection orthogonale** sur une droite d (passant par l'origine) est une transformation qui n'est ni injective ni surjective, donc pas bijective. On voit maintenant que ceci se reflète dans sa matrice relativement à la base canonique, puisque

$$\det([\mathrm{proj}_d]_{\mathcal{B}_{can}}) = \det\begin{pmatrix} \cos^2(\theta) & \cos(\theta)\sin(\theta) \\ \cos(\theta)\sin(\theta) & \sin^2(\theta) \end{pmatrix} = 0\,.$$

⋆ La **réflexion** d'axe d était inversible, ce que nous voyons maintenant au niveau de sa matrice, puisque

$$\det([\mathrm{refl}_d]_{\mathcal{B}_{\mathrm{can}}}) = \det\begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix} = -1 \neq 0\,.$$

De plus, on sait que $\text{refl}_d{}^{-1} = \text{refl}_d$, ce que l'on vérifie au niveau de la matrice :

$$\begin{aligned}
[\text{refl}_d]_{\mathcal{B}_{\text{can}}}{}^{-1} &= \frac{1}{-1}\begin{pmatrix} -\cos(2\theta) & -\sin(2\theta) \\ -\sin(2\theta) & \cos(2\theta) \end{pmatrix} \\
&= \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix} \\
&= [\text{refl}_d]_{\mathcal{B}_{\text{can}}} .
\end{aligned}$$

⋆ Finalement, nous avions remarqué que la rotation d'angle α est inversible, ce qui au niveau de la matrice se traduit par

$$\det([\text{rot}_\alpha]_{\mathcal{B}_{\text{can}}}) = \det \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} = 1 \neq 0 .$$

En utilisant la formule ci-dessus, on peut vérifier que son inverse correspond, comme on sait, à une rotation de $-\alpha$. En effet, à l'aide des propriétés de parité des fonctions trigonométriques,

$$\begin{aligned}
[\text{rot}_\alpha]_{\mathcal{B}_{\text{can}}}{}^{-1} &= \frac{1}{1}\begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix} \\
&= \begin{pmatrix} \cos(-\alpha) & -\sin(-\alpha) \\ \sin(-\alpha) & \cos(-\alpha) \end{pmatrix} \\
&= [\text{rot}_{-\alpha}]_{\mathcal{B}_{\text{can}}} .
\end{aligned}$$

◇

Le déterminant pour des matrices plus grandes ?

Plus tard, nous verrons comment la notion de déterminant peut se généraliser à des matrices carrées de tailles arbitraires, et comment celui-ci renseigne sur l'inversibilité d'une matrice. Pour l'instant, restons-en à l'étude de l'inversibilité, sans déterminant, en nous tournant vers le cas $n \times n$.

7.4 Le cas $n \times n$: matrices élémentaires et algorithme de Gauss-Jordan

Introduction

D'un point de vue très concret, le problème de l'inversibilité d'une matrice A ($n \times n$) peut se formuler de la façon suivante.

Puisqu'on cherche donc une matrice B ($n \times n$) telle que

$$AB = BA = I_n ,$$

si on écrit l'inconnue B en nommant ses colonnes,

$$B = [\mathbf{b}_1 \cdots \mathbf{b}_n] ,$$

alors le produit devient $AB = [A\mathbf{b}_1 \cdots A\mathbf{b}_n]$, et comme la matrice identité peut aussi s'écrire $I_n = [\mathbf{e}_1 \cdots \mathbf{e}_n]$, la contrainte $AB = I_n$ s'écrit

$$[A\mathbf{b}_1 \cdots A\mathbf{b}_n] = [\mathbf{e}_1 \cdots \mathbf{e}_n] .$$

Les colonnes de B doivent donc être solutions des n systèmes suivants :

$$\begin{cases} (*)_1 & : A\mathbf{b}_1 = \mathbf{e}_1 \,, \\ (*)_2 & : A\mathbf{b}_2 = \mathbf{e}_2 \,, \\ \quad\vdots & \\ (*)_n & : A\mathbf{b}_n = \mathbf{e}_n \,. \end{cases}$$

Si on met en route l'algorithme de Gauss pour résoudre chacun de ces systèmes, on se rend compte que les opérations élémentaires faites pour résoudre le premier système $A\mathbf{b}_1 = \mathbf{e}_1$ pourront être réutilisées dans tous les systèmes suivants . On conclut que l'on peut en fait étudier la résolution de ces n systèmes *en parallèle*, en se concentrant uniquement sur les coefficients de la matrice A.

Si on trouve des vecteurs $\mathbf{b}_1, \dots, \mathbf{b}_n$ solutions, respectivement, de $(*)_1, \dots, (*)_n$, alors on aura déjà une matrice $B = [\mathbf{b}_1 \cdots \mathbf{b}_n]$ satisfaisant $AB = I_n$.

Puisqu'on aimerait maintenir l'interprétation matricielle du résultat final, nous allons garder la trace des opérations élémentaires effectuées successivement, afin d'obtenir notre premier critère d'inversibilité.

Matrices élémentaires

Nous avons précédemment introduit des opérations élémentaires de Type I, II et III, qui agissaient sur un système $m \times n$ ou, de façon équivalente, sur sa matrice augmentée. Il se trouve que chaque opération élémentaire, prise individuellement, peut se formuler à l'aide d'un produit matriciel.

Rappelons que I_n est la matrice identité $n \times n$.

Définition 7.9. Une matrice $n \times n$ est dite **élémentaire** si elle s'obtient en effectuant une (et une seule) opération élémentaire sur I_n.

Exemple 7.10.

- $\star$ $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ est élémentaire, puisqu'on l'obtient à partir de I_3 par l'opération de Type I $L_2 \leftrightarrow L_3$.
- $\star$ $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ est élémentaire, puisqu'on l'obtient à partir de I_3 par l'opération de Type II $L_2 \leftrightarrow -2L_2$.
- $\star$ $\begin{pmatrix} 1 & 3 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ n'est pas élémentaire. (On peut l'obtenir à partir de I_3, mais avec pas moins de *deux* opérations élémentaires.)

$\diamond$

Maintenant, pour effectuer une transformation $\mathcal{E}$ sur une matrice A ($n \times p$), on pourra simplement considérer la matrice élémentaire E ($n \times n$) obtenue en effectuant $\mathcal{E}$ sur I_n, puis l'utiliser pour multiplier A **à gauche** par E : le résultat EA est alors la matrice A sur laquelle on a effectué $\mathcal{E}$.

Exemple 7.11. Considérons une matrice 3×4

$$A = \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \end{pmatrix} .$$

Supposons que l'on veuille effectuer sur A l'opération élémentaire $\mathcal{E}$ donnée par "$L_1 \leftrightarrow L_2$". Pour ce faire, on commence par appliquer $\mathcal{E}$ à I_3, qui donne

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

Puis, on multiplie A par E, à gauche :

$$EA = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \end{pmatrix} = \begin{pmatrix} e & f & g & h \\ a & b & c & d \\ i & j & k & l \end{pmatrix},$$

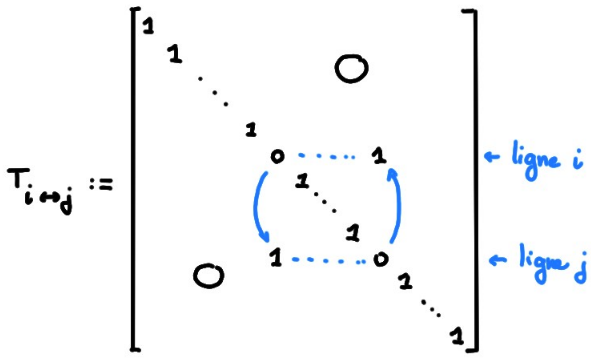
qui est bien ce qu'on voulait. ◇

Nous avions vu qu'une transformation élémentaire effectuée sur un système ne change pas son ensemble de solutions, puisqu'on pouvait toujours revenir au système de départ en appliquant une transformation réciproque. Une traduction de cette affirmation, dans le langage matriciel, est la suivante :

Lemme 12. *Toute matrice élémentaire est inversible.*

Pour le vérifier, écrivons explicitement les matrices élémentaires $n \times n$, ainsi que leurs inverses.

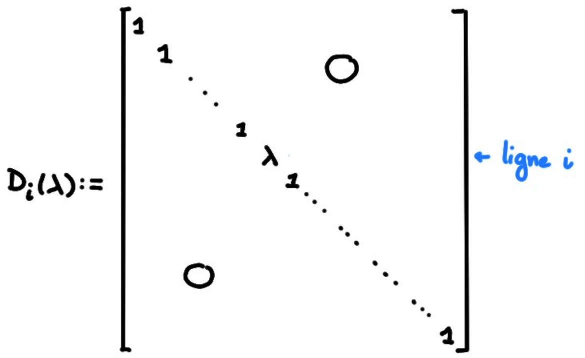
- ⋆ Type I : La matrice élémentaire associée à l'opération $L_i \leftrightarrow L_j$ est



On remarque que $T_{i\leftrightarrow j}T_{i\leftrightarrow j} = I_n$, et donc

$$T_{i\leftrightarrow j}{}^{-1} = T_{i\leftrightarrow j}$$

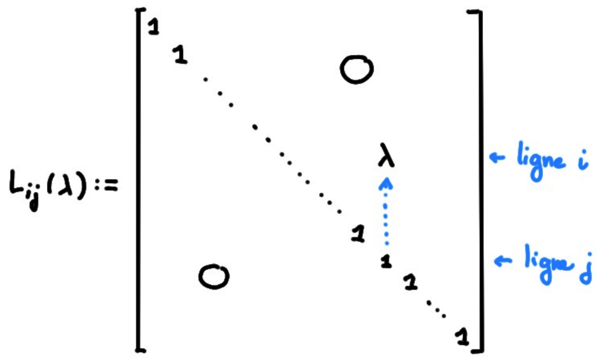
- ⋆ Type II : La matrice élémentaire associée à l'opération $L_i \leftarrow \lambda L_i$, où $\lambda \neq 0$, est



On remarque que $D_i(\lambda)D_i(\lambda^{-1}) = I_n$, et donc

$$D_i(\lambda)^{-1} = D_i(\lambda^{-1})$$

⋆ Type III : La matrice élémentaire associée à l'opération $L_i \leftarrow L_i + \lambda L_j$ est



On remarque que $L_{ij}(\lambda)L_{ij}(-\lambda) = I_n$, et donc

$$L_{ij}(\lambda)^{-1} = L_{ij}(-\lambda)$$

Lorsqu'on résout un système, on choisit une *suite* d'opérations élémentaires, dans le but d'arriver à une forme échelonnée du système. Notons ces opérations $\mathcal{E}^{(1)}, \mathcal{E}^{(2)}, \dots, \mathcal{E}^{(k)}$. L'application de ces opérations (d'abord $\mathcal{E}^{(1)}$, puis $\mathcal{E}^{(2)}$, etc.) correspondent à des multiplications successives de A *à gauche* par les matrices élémentaires qui leur correspondent :

$$\begin{aligned}
\text{Opération } \mathcal{E}^{(1)} : \quad & E^{(1)}A\,,\\
\text{Opération } \mathcal{E}^{(1)} : \quad & E^{(2)}E^{(1)}A\,,\\
& \vdots\\
\text{Opération } \mathcal{E}^{(k)} : \quad & E^{(k)}\cdots E^{(2)}E^{(1)}A\,.
\end{aligned}$$

Comme on sait, une matrice possède une *unique* version échelonnée réduite, et donc il est toujours possible de bien choisir les matrices $E^{(j)}$, de façon à ce que la matrice finale,

$$\widetilde{A} = E^{(k)}\cdots E^{(2)}E^{(1)}A\,,$$

soit la réduite de A.

Théorème 7.12. *Soit A une matrice $n \times n$. Alors A est inversible si et seulement si on peut réduire A à l'identité I_n à l'aide d'un produit de matrices élémentaires, c'est-à-dire s'il existe des matrices élémentaires $E^{(1)}, \dots, E^{(k)}$ telles que la réduite*

$$\widetilde{A} = E^{(k)}\cdots E^{(1)}A$$

soit la matrice identité : $\widetilde{A} = I_n$.

Preuve: Supposons que A est inversible. Alors $T(\mathbf{x}) := A\mathbf{x}$ est bijective, et par conséquent pour tout $\mathbf{b} \in \mathbb{R}^n$, le système $A\mathbf{x} = \mathbf{b}$ possède une unique solution. Ceci implique que sa forme échelonnée réduite ne présente aucune variable libre (chaque pivot est situé immédiatement à droite du pivot de la ligne supérieure). Comme A est $n \times n$, ceci implique que la réduite de A est I_n.

Supposons ensuite qu'il existe des matrices élémentaires $E^{(1)}, \dots, E^{(k)}$ telles que

$$E^{(k)}E^{(k-1)}\cdots E^{(1)}A = I_n\,.$$

Rappelons que chaque $E^{(j)}$ est inversible. En multipliant la relation ci-dessus à gauche par $(E^{(k)})^{-1}$,

$$\underbrace{(E^{(k)})^{-1}E^{(k)}}_{=I_n}E^{(k-1)}\cdots E^{(1)}A = (E^{(k)})^{-1}I_n = (E^{(k)})^{-1}\,.$$

En multipliant successivement, à gauche, par $(E^{(k-1)})^{-1}, \dots, (E^{(1)})^{-1}$, on arrive à

$$A = (E^{(1)})^{-1}\cdots(E^{(k)})^{-1}\,.$$

Puisque chacune des matrices $(E^{(j)})^{-1}$ est inversible, A est un produit de matrices inversibles, et donc elle aussi est inversible. Son inverse est donné par

$$\begin{aligned}A^{-1} &= \big((E^{(1)})^{-1}\cdots(E^{(k)})^{-1}\big)^{-1}\\ &= \big((E^{(k)})^{-1}\big)^{-1}\cdots\big((E^{(1)})^{-1}\big)^{-1}\\ &= E^{(k)}\cdots E^{(1)}\,.\end{aligned}$$

□

Comme conséquence de ce qui a été fait dans la preuve :

Corollaire 5. *Une matrice A ($n \times n$) est inversible si et seulement si elle peut s'écrire comme un produit de matrices élémentaires.*

L'algorithme

L'argument développé dans la preuve du précédent théorème fournit un algorithme pour déterminer si une matrice est inversible et, dans le cas où elle est inversible, de calculer son inverse.

Reprenons l'expression

$$[A\mathbf{b}_1\cdots A\mathbf{b}_n] = [\mathbf{e}_1\cdots\mathbf{e}_n]\,.$$

Pour résoudre n'importe lequel de ces systèmes, $A\mathbf{b}_j = \mathbf{e}_j$, on applique successivement des opérations élémentaires en multipliant à gauche par les matrices correspondantes $E^{(1)}, \dots, E^{(k)}$, jusqu'à obtenir, du côté gauche, la réduite de A :

$$\begin{aligned}A\mathbf{b}_j &= \mathbf{e}_j\\ E^{(1)}A\mathbf{b}_j &= E^{(1)}\mathbf{e}_j\\ E^{(2)}E^{(1)}A\mathbf{b}_j &= E^{(2)}E^{(1)}\mathbf{e}_j\\ &\vdots\\ \underbrace{E^{(k)}\cdots E^{(2)}E^{(1)}A}_{=\widetilde{A}}\mathbf{b}_j &= E^{(k)}\cdots E^{(2)}E^{(1)}\mathbf{e}_j\end{aligned}$$

Si $\widetilde{A} = I_n$, le théorème ci-dessus dit que A est inversible ; de plus la j-ème colonne de son inverse est donnée par

$$\mathbf{b}_j = E^{(k)}\cdots E^{(1)}\mathbf{e}_j\,,$$

qui n'est autre que la j-ème colonne de la matrice $E^{(k)}\cdots E^{(1)}$.

On peut résumer ce procédé dans l'algorithme suivant, appelé **algorithme de Gauss-Jordan** (pour l'étude de l'inversibilité d'une matrice) :

1) Commencer par considérer la matrice $n \times 2n$

$$\big(A\,|\,I_n\big)$$

2) Considérer des opérations élémentaires permettant de transformer A en sa réduite $\widetilde{A}$, et les appliquer *simultanément sur chaque moitié* :

$$\begin{gathered}(A \,|\, I_n)\\ (E^{(1)}A \,|\, E^{(1)}I_n)\\ (E^{(2)}E^{(1)}A \,|\, E^{(2)}E^{(1)}I_n)\\ \vdots \\ (\underbrace{E^{(k)}\cdots E^{(2)}E^{(1)}A}_{=\widetilde{A}} \,|\, \underbrace{E^{(k)}\cdots E^{(2)}E^{(1)}}_{=:C})\,.\end{gathered}$$

3) Regarder le résultat obtenu, $(\widetilde{A} \,|\, C)$, et conclure :
   - ⋆ Si $\widetilde{A} = I_n$, alors A est inversible, et $A^{-1} = C$.
   - ⋆ Si $\widetilde{A} \neq I_n$, alors A est singulière.

Exemple 7.13. Considérons $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, et étudions son inversibilité à l'aide de l'algorithme ci-dessus. On pose donc

$$(A \,|\, I_2) = \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array}\right)$$

Comme il suffit d'une seule transformation pour réduire A, $\mathcal{E}^{(1)} = (L_2 \to L_2 - 2L_1)$, on a

$$(\widetilde{A} \,|\, C) = \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{array}\right)$$

Comme $\widetilde{A} \neq I_2$, A est singulière. (On voit aussi que $\det(A) = 0$, qui montre également que A n'est pas inversible.) ◇

Exemple 7.14. Étudions l'inversibilité de $A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{pmatrix}$. On pose

$$(A \,|\, I_3) = \left(\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right).$$

Appliquons successivement des opérations élémentaires afin de réduire A du côté gauche ; *on applique chaque opération sur toute la matrice, y compris sur le côté droit.*

Commençons par $L_3 \leftarrow L_3 - 4L_2$, suivie de $L_3 \leftarrow L_3 + 3L_1$:

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right).$$

Ensuite, $L_1 \leftrightarrow L_2$,

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

puis $L_2 \leftarrow L_2 - L_3$,

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

$L_3 \leftarrow \frac{1}{2}L_3$,

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

et enfin $L_1 \leftarrow L_1 - 3L_3$:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{array}\right).$$

On a donc obtenu

$$(\widetilde{A}\,|\,C) = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{array}\right).$$

Comme $\widetilde{A} = I_3$, A est inversible, et son inverse est ce qu'on voit du côté droit :

$$A^{-1} = \begin{pmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{pmatrix}.$$

(Exercice : Pourquoi pas vérifier, à la main, que $AA^{-1} = A^{-1}A = I_3$!) ◇

7.5 Critères d'inversibilité

Dans la section précédente, nous avons vu un premier critère d'inversibilité général pour une matrice A, caractérisé par la possibilité de réduire (ou non) A à l'identité. Relions maintenant l'inversibilité à d'autres propriétés algébriques. (Dans la suite du cours, d'autres critères viendront s'ajouter à cette liste.)

Théorème 7.15. *(Critères d'inversibilité) Soit A une matrice $n \times n$. Alors les propriétés suivantes sont toutes équivalentes.*

1) *A est inversible.*
2) *A est un produit fini de matrices élémentaires.*
3) *La réduite de A est I_n.*
4) *Pour tout $\mathbf{b} \in \mathbb{R}^n$, le système $A\mathbf{x} = \mathbf{b}$ possède une unique solution.*
5) *Le système $A\mathbf{x} = \mathbf{0}$ ne possède que la sol. triviale ($\ker(A) = \{\mathbf{0}\}$).*
6) *Les colonnes de A sont linéairement indépendantes.*
7) *Les colonnes de A engendrent $\mathbb{R}^n$ ($\mathrm{Im}(A) = \mathbb{R}^n$).*
8) *$\mathrm{rang}(A) = n$.*
9) *Les colonnes de A forment une base de $\mathbb{R}^n$.*

Preuve: Les équivalences 1.⇔2.⇔3. ont été démontrées dans la section précédente.

1.⇒4. Si A est inversible, alors pour tout $\mathbf{b} \in \mathbb{R}^n$, la solution de $A\mathbf{x} = \mathbf{b}$ est donnée par $\mathbf{x} = A^{-1}\mathbf{b}$, et est donc unique.

4.⇒5. On sait que $A\mathbf{x} = \mathbf{0}$ possède toujours la solution triviale. Donc en prenant $\mathbf{b} = \mathbf{0}$, on assure que la solution triviale est unique.

5.⇔6. Suit du fait que les colonnes d'une matrice A (de taille quelconque) sont linéairement indépendantes si et seulement si le systèmes homogène associé $A\mathbf{x} = \mathbf{0}$ ne possède que la solution triviale.

6.⇒9. En effet, nous savons que dans un espace vectoriel de dimension n, toute famille libre contenant n vecteurs forme une base.

9.⇒1. Supposons que les colonnes de A forment une base de $\mathbb{R}^n$. Pour montrer que A est inversible, il suffit de montrer que l'application $T(\mathbf{x}) = A\mathbf{x}$ est bijective. Montrons d'abord qu'elle est injective. Supposons que $T(\mathbf{x}) = T(\mathbf{x}')$, c'est-à-dire $A(\mathbf{x} - \mathbf{x}') = \mathbf{0}$. Puisque les colonnes de A sont indépendantes, l'unique solution de ce système est $\mathbf{x} - \mathbf{x}' = \mathbf{0}$, et donc $\mathbf{x} = \mathbf{x}'$. Ensuite, T est clairement surjective puisque pour tout $\mathbf{y} \in \mathbb{R}^n$, le système $A\mathbf{x} = \mathbf{y}$ possède une solution (ici on utilise le fait que les colonnes de A engendrent $\mathbb{R}^n$).

9.⇒7⇔8. (Évident)

7.⇒9. Si les colonnes de A engendrent $\mathbb{R}^n$, elles forment nécessairement une famille libre (sinon, la dimension de l'espace qu'elles engendrent serait strictement inférieur à n), et donc une base comme on a déjà dit. □

Conséquence : une simplification de la définition d'inversibilité

Nous avons insisté plusieurs fois sur le fait que la définition d'inversibilité implique *deux* conditions : il doit exister B telle que $AB = I_n$ **et** $BA = I_n$. Or nous avons maintenant les outils pour prouver qu'il suffit qu'*une seule* de ces conditions soit vérifiée :

Proposition 7. *Soit A une matrice $n \times n$.*

1) *Si il existe une matrice C ($n \times n$) telle que $CA = I_n$, alors A est inversible et $A^{-1} = C$.*
2) *Si il existe une matrice B ($n \times n$) telle que $AB = I_n$, alors A est inversible et $A^{-1} = B$.*

Preuve: 1. Supposons que $CA = I_n$. Si $\mathbf{x}$ est solution du système homogène $A\mathbf{x} = \mathbf{0}$, alors

$$\mathbf{x} = I_n\mathbf{x} = CA\mathbf{x} = C(A\mathbf{x}) = C\mathbf{0} = \mathbf{0}\,.$$

Donc le système homogène ne possède que la solution triviale. Par le théorème (critère 5), A est inversible : son inverse A^{-1} existe. En multipliant l'identité $CA = I_n$ à droite par A^{-1}, on obtient $C = A^{-1}$.

2. Supposons que $AB = I_n$. Fixons un $\mathbf{y} \in \mathbb{R}^n$ quelconque. On a alors $AB\mathbf{y} = I_n\mathbf{y} = \mathbf{y}$, que l'on peut récrire $A\mathbf{x}_* = \mathbf{y}$ (où $\mathbf{x}_* = B\mathbf{y}$). Ceci implique bien que $\mathbf{y} \in \mathrm{Im}(A)$. Comme ceci est vrai pour tout $\mathbf{y}$, on a que $\mathrm{Im}(A) = \mathbb{R}^n$. Par le théorème (critère 7.) ce qui implique que A est inversible. En multipliant l'identité $AB = I_n$ à gauche par A^{-1}, on obtient $A^{-1} = B$. □

Chapitre 8

Le changement de base

8.1 Introduction

Dans la pratique, l'étude d'un problème impliquant un espace vectoriel se fait en choisissant une base, et en travaillant sur les composantes des vecteurs relatives à cette base.

Bien-sûr, un problème peut s'énoncer naturellement dans une base $\mathcal{B}$, mais être plus facilement soluble dans une autre base $\mathcal{B}'$, mieux adaptée à la résolution du problème. On aura donc souvent recours à un *changement de base*.

Nous aborderons le changement de base en deux étapes.

1) D'abord, nous considérerons le problème de savoir comment les *composantes* d'un vecteur se transforment quand on change de base dans un espace vectoriel.
2) Ensuite, nous verrons comment la *matrice d'une application linéaire* se transforme lorsqu'on change de base dans les espaces vectoriels de départ et d'arrivée.

Nous présenterons chaque méthode dans un espace vectoriel quelconque, puis l'utiliserons dans diverses situations, en particulier pour les applications $T : \mathbb{R}^n \to \mathbb{R}^m$.

8.2 Effet sur les composantes des vecteurs

Pour commencer, étudions les relations existant entre les composantes d'un même vecteur, exprimé relativement à une base ou à une autre.

Avant de voir l'approche dans le cas général, commençons par un exemple simple.

Exemple 8.1. Dans le plan, considérons le vecteur

$$\mathbf{x} = \begin{pmatrix} 5 \\ 1 \end{pmatrix} .$$

Considérons maintenant la base $\mathcal{B} = (\mathbf{b}_1, \mathbf{b}_2)$, dont les vecteurs sont disons

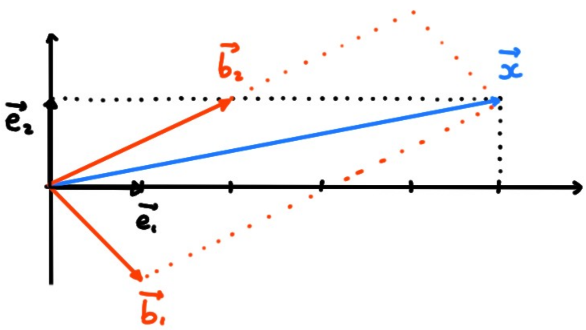
$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} , \qquad \mathbf{b}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} .$$

Quelles sont les composantes de $\mathbf{x}$ relativement à $\mathcal{B}$? Ce qu'on cherche ici est

$$[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} ,$$

qui ne signifie rien d'autre que
$$\mathbf{x} = \beta_1 \mathbf{b}_1 + \beta_2 \mathbf{b}_2 \,.$$
Or cette dernière s'exprime comme
$$\begin{pmatrix} 5 \\ 1 \end{pmatrix} = \beta_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \beta_2 \begin{pmatrix} 2 \\ 1 \end{pmatrix} ,$$
qui est équivalent au système
$$\begin{cases} \beta_1 & + & 2\beta_2 & = & 5 \\ -\beta_1 & + & \beta_2 & = & 1 \end{cases} ,$$
dont la solution est $\beta_1 = 1$, $\beta_2 = 2$. Ainsi,
$$[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} ,$$
qui signifie $\mathbf{x} = \mathbf{b}_1 + 2\mathbf{b}_2$.

Remarque : Il est plus utile de penser que $\mathbf{x}$ est un vecteur dans le plan, et que ce vecteur peut être représenté en composantes, relativement à la base canonique $\mathcal{B}_{\text{can}}$ ou à la base $\mathcal{B}$:
$$[\mathbf{x}]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 5 \\ 1 \end{pmatrix} , \qquad [\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} .$$



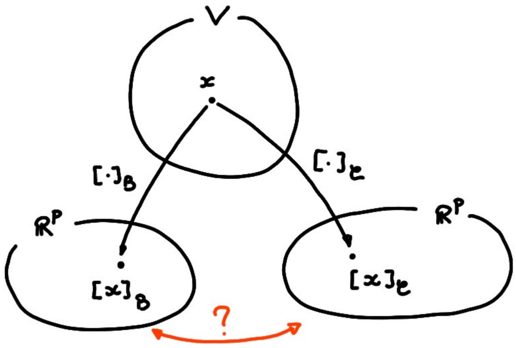
Bien-sûr, il serait intéressant d'avoir un procédé permettant d'obtenir directement les composantes d'un vecteur quelconque dans une base, en fonction des composantes dans l'autre base :
$$[\mathbf{x}]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \overset{?}{\longleftrightarrow} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = [\mathbf{x}]_{\mathcal{B}}$$

◇

La matrice de changement de base

Abordons le problème d'un point de vue général.

Soit V un espace vectoriel de dimension p. Supposons que l'on ait deux bases dans V :
$$\mathcal{B} = (b_1, \dots, b_p) , \qquad \mathcal{C} = (c_1, \dots, c_p) .$$
Si $x \in V$ est un vecteur quelconque, il peut être décomposé dans une base ou dans l'autre, et les composantes relativement à ces bases seront a priori différentes :
$$[x]_{\mathcal{B}} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} , \qquad [x]_{\mathcal{C}} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_p \end{pmatrix} .$$
Nous aimerions savoir comment les composantes relativement à une base, par exemple les $\beta_1, \dots, \beta_p$, peuvent se calculer à partir des composantes dans l'autre base, c'est-à-dire les $\gamma_1, \dots, \gamma_p$.



Nous allons voir que cette relation est linéaire, et peut donc s'exprimer à l'aide d'une matrice :

Théorème 8.2. *Il existe une matrice $p \times p$, notée $P_{\mathcal{CB}}$ (ou parfois : $P_{\mathcal{C}\leftarrow\mathcal{B}}$), telle que*

$$[x]_{\mathcal{C}} = P_{\mathcal{CB}}[x]_{\mathcal{B}} \,.$$

De plus,

- ⋆ $P_{\mathcal{CB}} = \big[[b_1]_{\mathcal{C}} \cdots [b_p]_{\mathcal{C}}\big]$,
- ⋆ $P_{\mathcal{CB}}$ *est inversible, et* ${P_{\mathcal{CB}}}^{-1} = P_{\mathcal{BC}}$.

*On appelle $P_{\mathcal{CB}}$ la **matrice de changement de base de $\mathcal{B}$ vers $\mathcal{C}$**.*

Preuve:

$$\begin{aligned}
[x]_{\mathcal{C}} &= \big[\beta_1 b_1 + \cdots + \beta_p b_p\big]_{\mathcal{C}} \\
&= \beta_1 [b_1]_{\mathcal{C}} + \cdots + \beta_p [b_p]_{\mathcal{C}} \\
&= \big[[b_1]_{\mathcal{C}} \cdots [b_p]_{\mathcal{C}}\big] \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \\
&= P_{\mathcal{CB}}[x]_{\mathcal{B}}
\end{aligned}$$

En procédant dans l'autre sens, on obtient

$$\begin{aligned}
[x]_{\mathcal{B}} &= \big[\gamma_1 c_1 + \cdots + \gamma_p c_p\big]_{\mathcal{B}} \\
&= \gamma_1 [c_1]_{\mathcal{C}} + \cdots + \gamma_p [c_p]_{\mathcal{B}} \\
&= \big[[c_1]_{\mathcal{B}} \cdots [c_p]_{\mathcal{B}}\big] \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_p \end{pmatrix} \\
&= P_{\mathcal{BC}}[x]_{\mathcal{C}}
\end{aligned}$$

Si on réinjecte cette dernière dans celle du dessus,

$$\begin{aligned}
[x]_{\mathcal{C}} &= P_{\mathcal{CB}}[x]_{\mathcal{B}} \\
&= P_{\mathcal{CB}}P_{\mathcal{BC}}[x]_{\mathcal{C}} \,.
\end{aligned}$$

Comme cette dernière est vraie pour tout $x \in V$, on a

$$P_{\mathcal{CB}}P_{\mathcal{BC}} = I_p \,.$$

Donc $P_{\mathcal{CB}}$ est inversible, et son inverse est ${P_{\mathcal{CB}}}^{-1} = P_{\mathcal{BC}}$. □

Remarque 8.3. En bas de page, on explique pourquoi la matrice de changement de base n'est que la représentation de l'application *linéaire identité*, de V dans lui-même. ◇

Exemple 8.4. Dans le plan, considérons comme tout à l'heure le vecteur

$$\mathbf{x} = \begin{pmatrix} 5 \\ 1 \end{pmatrix} .$$

Pour être plus précis, notons $\mathcal{B}_{\text{can}} = (\mathbf{e}_1, \mathbf{e}_2)$ la base canonique, et récrivons

$$[\mathbf{x}]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 5 \\ 1 \end{pmatrix} .$$

Considérons maintenant la base $\mathcal{B} = (\mathbf{b}_1, \mathbf{b}_2)$ définie par :

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} , \qquad \mathbf{b}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} .$$

Calculons $[\mathbf{x}]_{\mathcal{B}}$, en fonction de $[\mathbf{x}]_{\mathcal{B}_{\text{can}}}$, en utilisant le théorème :

$$[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}\mathcal{B}_{\text{can}}} [\mathbf{x}]_{\mathcal{B}_{\text{can}}} ,$$

où

$$P_{\mathcal{B}\mathcal{B}_{\text{can}}} = \big[[\mathbf{e}_1]_{\mathcal{B}} \, [\mathbf{e}_2]_{\mathcal{B}} \big] .$$

On doit donc trouver les composantes de $\mathbf{e}_1$ et $\mathbf{e}_2$ relativement à $\mathcal{B}$. Mais comme

$$[\mathbf{b}_1]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} , \qquad [\mathbf{b}_2]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

signifie en fait

$$\begin{cases} \mathbf{b}_1 & = & \mathbf{e}_1 & - & \mathbf{e}_2 \\ \mathbf{b}_2 & = & 2\mathbf{e}_1 & + & \mathbf{e}_2 \, , \end{cases}$$

on a

$$\begin{cases} \mathbf{e}_1 & = & \frac{1}{3}\mathbf{b}_1 & + & \frac{1}{3}\mathbf{b}_2 \\ \mathbf{e}_2 & = & -\frac{2}{3}\mathbf{b}_1 & + & \frac{1}{3}\mathbf{b}_2 \, , \end{cases}$$

Ainsi,

$$[\mathbf{e}_1]_{\mathcal{B}} = \begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix} , \qquad [\mathbf{e}_2]_{\mathcal{B}} = \begin{pmatrix} -2/3 \\ 1/3 \end{pmatrix} ,$$

et donc

$$P_{\mathcal{B}\mathcal{B}_{\text{can}}} = \big[[\mathbf{e}_1]_{\mathcal{B}} \, [\mathbf{e}_2]_{\mathcal{B}} \big] = \begin{pmatrix} 1/3 & -2/3 \\ 1/3 & 1/3 \end{pmatrix}$$

Donc les coordonnées de $\mathbf{x}$ relativement à $\mathcal{B}$ sont

$$[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}\mathcal{B}_{\text{can}}} [\mathbf{x}]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 1/3 & -2/3 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} ,$$

comme nous avions trouvé plus haut. Si maintenant on souhaite plutôt transformer des composantes relativement à $\mathcal{B}$ en des composantes relativement à $\mathcal{B}_{\text{can}}$, on calcule

$$P_{\mathcal{B}_{\text{can}}\mathcal{B}} = P_{\mathcal{B}\mathcal{B}_{\text{can}}}{}^{-1} = \frac{1}{1/3} \begin{pmatrix} 1/3 & 2/3 \\ -1/3 & 1/3 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} .$$

Donc si par exemple on prend $\mathbf{x}$ tel que

$$[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} ,$$

alors ses composantes relativement à $\mathcal{B}_{\text{can}}$ sont, comme on sait déjà,

$$[\mathbf{x}]_{\mathcal{B}_{\text{can}}} = P_{\mathcal{B}_{\text{can}}\mathcal{B}} [\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix} .$$

◇

Exemple 8.5. Supposons que l'on considère, dans $\mathbb{R}^3$, le vecteur

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} .$$

Considérons la base de $\mathbb{R}^3$, $\mathcal{B} = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$, dont les vecteurs sont (on laisse au lecteur le soin de vérifier que $\mathcal{B}$ est effectivement une base) :

$$\mathbf{b}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} , \qquad \mathbf{b}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} , \qquad \mathbf{b}_3 = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} .$$

Ensuite, cherchons les composantes de $\mathbf{x}$ relativement à $\mathcal{B}$, en utilisant le formalisme présenté plus haut.

Pour bien faire, récrivons explicitement ce que nous savons :

$$[\mathbf{x}]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} ,$$

ainsi que

$$[\mathbf{b}_1]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} , \qquad [\mathbf{b}_2]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} , \qquad [\mathbf{b}_3]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} .$$

Pour exprimer les composantes de $\mathbf{x}$ relativement à $\mathcal{B}$, nous allons utiliser la formule

$$[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}\mathcal{B}_{\text{can}}}[\mathbf{x}]_{\mathcal{B}_{\text{can}}} ,$$

où la matrice de changement de base est donnée par

$$P_{\mathcal{B}\mathcal{B}_{\text{can}}} = \left[[\mathbf{e}_1]_{\mathcal{B}} \, [\mathbf{e}_1]_{\mathcal{B}} \, [\mathbf{e}_1]_{\mathcal{B}} \right] .$$

Or si on écrit explicitement les définitions des vecteurs de la base $\mathcal{B}$,

$$\begin{cases} \mathbf{b}_1 & = & & & \mathbf{e}_3 \\ \mathbf{b}_2 & = & \mathbf{e}_1 & & -\mathbf{e}_3 \\ \mathbf{b}_3 & = & & 2\mathbf{e}_2 & \end{cases}$$

Comme on doit exprimer les composantes des vecteurs de la base canonique par rapport à $\mathcal{B}$, il faut inverser ces relations. On trouve facilement

$$\begin{cases} \mathbf{e}_1 & = & \mathbf{b}_1 & +\mathbf{b}_2 & \\ \mathbf{e}_2 & = & & & \frac{1}{2}\mathbf{b}_3 \\ \mathbf{e}_3 & = & \mathbf{b}_1 & & \end{cases} ,$$

c'est-à-dire

$$[\mathbf{e}_1]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} , \qquad [\mathbf{e}_2]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \end{pmatrix} , \qquad [\mathbf{e}_3]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} ,$$

qui donne

$$\begin{aligned}
[\mathbf{x}]_{\mathcal{B}} &= P_{\mathcal{B}\mathcal{B}_{\text{can}}}[\mathbf{x}]_{\mathcal{B}_{\text{can}}} \\
&= \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\
&= \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix} .
\end{aligned}$$

Remarque : Pour le calcul de $P_{\mathcal{B}\mathcal{B}_{\text{can}}}$, une façon tout à fait équivalente de faire mais écrite différemment aurait été de commencer par calculer

$$P_{\mathcal{B}_{\text{can}}\mathcal{B}} = \big[[\mathbf{b}_1]_{\mathcal{B}_{\text{can}}}\,[\mathbf{b}_2]_{\mathcal{B}_{\text{can}}}\,[\mathbf{b}_3]_{\mathcal{B}_{\text{can}}}\big] = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 1 & -1 & 0 \end{pmatrix} ,$$

puis de calculer son inverse (par exemple avec l'algorithme de Gauss-Jordan) :

$$P_{\mathcal{B}\mathcal{B}_{\text{can}}} = {P_{\mathcal{B}_{\text{can}}\mathcal{B}}}^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix} .$$

◇

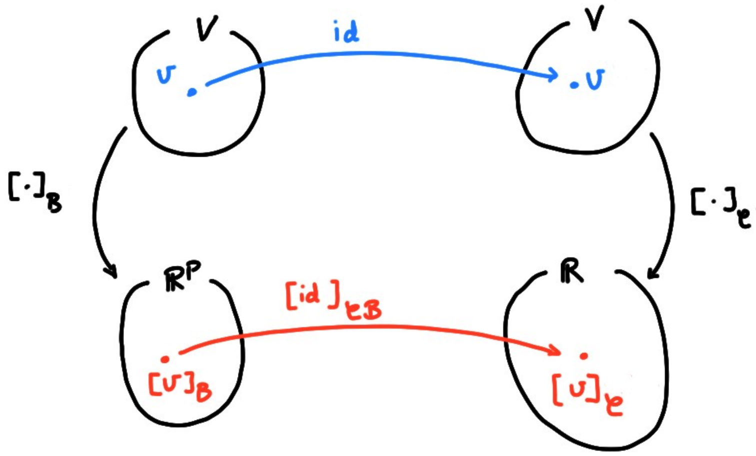
La matrice $P_{\mathcal{C}\mathcal{B}}$ comme matrice d'une application

La matrice de changement de base peut être vue comme un cas particulier de matrice associée à une application linéaire, introduite dans le chapitre sur les espaces vectoriels.

En effet, considérons l'**application identité** $\text{id} : V \to V$, définie par

$$\text{id}(v) := v \qquad \forall v \in V .$$

Cette application ne porte en elle rien de vraiment intéressant. Mais considérons comme avant deux bases pour décrire V, notées $\mathcal{C}$ et $\mathcal{B}$.



Par définition, la matrice qui représente id relativement aux bases $\mathcal{B}$ et $\mathcal{C}$, est celle qui permet d'obtenir $[v]_{\mathcal{C}}$ à partir de $[v]_{\mathcal{B}}$:

$$[v]_{\mathcal{C}} = [\text{id}]_{\mathcal{C}\mathcal{B}}[v]_{\mathcal{B}}.$$

Cette matrice est donc précisément notre matrice de changement de base :

$$P_{\mathcal{C}\mathcal{B}} = [\mathrm{id}]_{\mathcal{C}\mathcal{B}} .$$

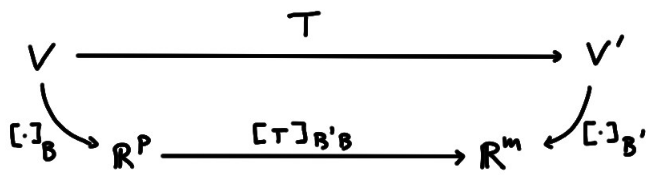
8.3 Effet sur la matrice d’une application

On a vu dans le chapitre sur les espaces vectoriels comment exprimer une application linéaire

$$T : V \to V' ,$$

lorsqu’on possède une base $\mathcal{B}$ dans V, et une base $\mathcal{B}'$ dans V' :

- ⋆ $\mathcal{B} = (v_1, \dots, v_p)$ est une base de V,
- ⋆ $\mathcal{B}' = (v_1', \dots, v_m')$ est une base de V'.



Rappelons que si $v \in V$, et si $v' = T(v) \in V'$ est son image, alors la représentation matricielle de cette correspondance s’exprime par

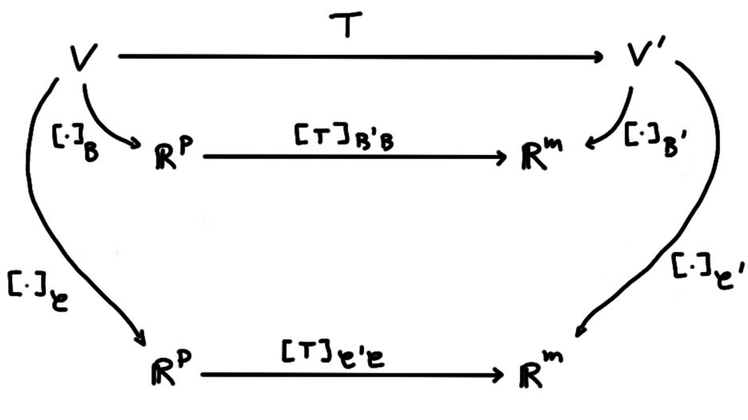
$$\underbrace{[T(v)]_{\mathcal{B}'}}_{\in \mathbb{R}^m} = \underbrace{[T]_{\mathcal{B}'\mathcal{B}}}_{m \times p} \underbrace{[v]_{\mathcal{B}}}_{\in \mathbb{R}^p} ,$$

où les colonnes de la matrice $[T]_{\mathcal{B}'\mathcal{B}}$ s’obtiennent en décomposant les images des vecteurs de $\mathcal{B}$ dans $\mathcal{B}'$:

$$[T]_{\mathcal{B}'\mathcal{B}} = \big[[T(v_1)]_{\mathcal{B}'} \cdots [T(v_p)]_{\mathcal{B}'} \big]$$

Changement de base dans le cas général $T : V \to V'$

Supposons maintenant que l’on ait d’autres bases qui décrivent les espaces de départ et d’arrivée, disons $\mathcal{C}$ dans V et $\mathcal{C}'$ dans V' :



On a donc deux façons de représenter la même application linéaire T :

⋆ En utilisant les bases $\mathcal{B}$ et $\mathcal{B}'$:

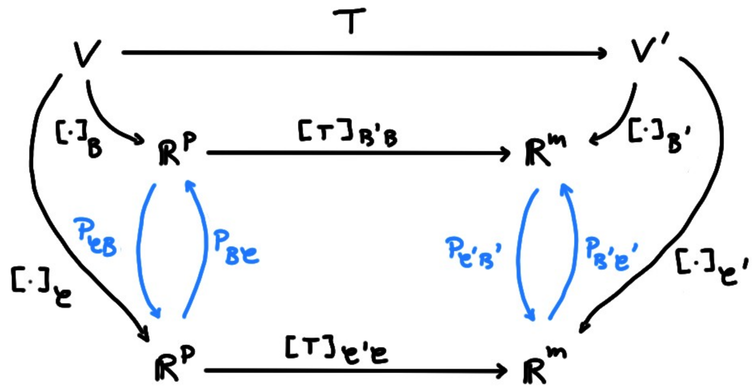
$$[v]_{\mathcal{B}} \mapsto [T(v)]_{\mathcal{B}'} = [T]_{\mathcal{B}'\mathcal{B}}[v]_{\mathcal{B}}\,.$$

⋆ En utilisant les bases $\mathcal{C}$ et $\mathcal{C}'$:

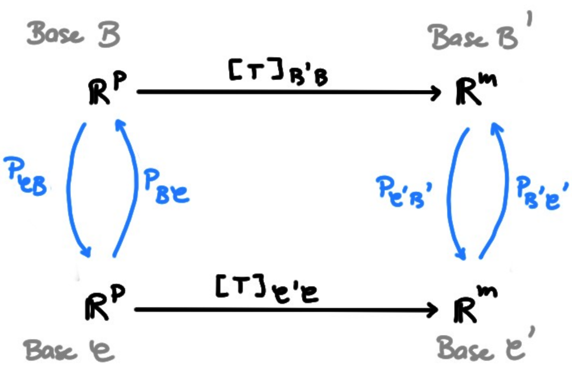
$$[v]_{\mathcal{C}} \mapsto [T(v)]_{\mathcal{C}'} = [T]_{\mathcal{C}'\mathcal{C}}[v]_{\mathcal{C}}\,.$$

Notre but ici est de comprendre la relation entre les matrices $[T]_{\mathcal{B}'\mathcal{B}}$ et $[T]_{\mathcal{C}'\mathcal{C}}$.

Considérons les matrices de changement de base construites dans la section précédente, $P_{\mathcal{C}\mathcal{B}}$ et $P_{\mathcal{C}'\mathcal{B}'}$, et leurs inverses :



Pour simplifier un peu le schéma, gardons uniquement les espaces de départ et d'arrivée, les bases relativement auxquelles ils sont associés, ainsi que les matrices associées à T relativement à ces bases :



Dans ce diagramme, on peut monter ou descendre librement à l'aide des matrices de changement de base, puisqu'elles sont inversibles . On a donc les **formules de changement de base** :

$$\boxed{[T]_{\mathcal{B}'\mathcal{B}} = P_{\mathcal{B}'\mathcal{C}'}[T]_{\mathcal{C}'\mathcal{C}}P_{\mathcal{C}\mathcal{B}}}\,,$$

et

$$\boxed{[T]_{\mathcal{C}'\mathcal{C}} = P_{\mathcal{C}'\mathcal{B}'}[T]_{\mathcal{B}'\mathcal{B}}P_{\mathcal{B}\mathcal{C}}}\,.$$

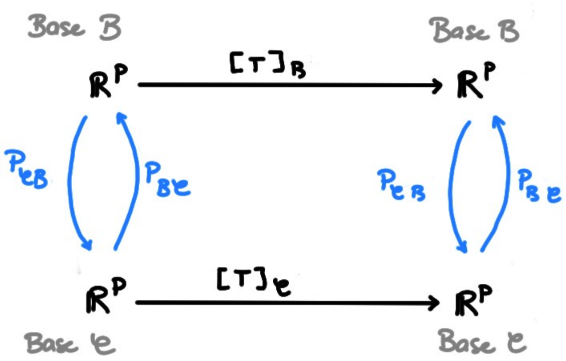
Ces deux formules sont équivalentes : on obtient la deuxième en multipliant la première à droite par $P_{\mathcal{B}\mathcal{C}}$, puis à gauche par $P_{\mathcal{C}'\mathcal{B}'}$.

Changement de base dans le cas $T : V \to V$

Le cas que nous utiliserons le plus est lorsque T applique V dans lui-même, c'est-à-dire où $V' = V$:

$$T : V \to V\,.$$

Dans ce cas, on a $p = m$. Si on suppose aussi que l'on a deux bases pour décrire V, $\mathcal{B}$ et $\mathcal{C}$, et qu'on on prend $\mathcal{C}' = \mathcal{C}$, $\mathcal{B}' = \mathcal{B}$, le shéma devient plus simple :



Maintenant, comme $P_{\mathcal{BC}} = {P_{\mathcal{CB}}}^{-1}$, les formules de changement de base prennent la forme plus connue :

$$\boxed{[T]_{\mathcal{B}} = {P_{\mathcal{CB}}}^{-1}[T]_{\mathcal{C}}P_{\mathcal{CB}}\,,}$$

on encore

$$\boxed{[T]_{\mathcal{B}} = P_{\mathcal{BC}}[T]_{\mathcal{C}}{P_{\mathcal{BC}}}^{-1}\,.}$$

8.4 Exemples

Exploitons les diverses formules de changement de base vues dans les sections précédentes, sur quelques exemples concrets.

Toutes les applications linéaires que nous avons considérées jusqu'à présent ont généralement été définies relativement à la base canonique : leur matrice s'obtenait en calculant les images des vecteurs de la base canonique.

Mais on sait maintenant exprimer la matrice d'une application relativement à n'importe quelle base. Nous allons donc repasser par certaines applications rencontrées précédemment, et étudier leur matrice relativement à des bases qui ne sont pas canoniques.

Exemple 8.6. Considérons l'application linéaire $T : \mathbb{R}^3 \to \mathbb{R}^2$ définie par

$$T\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) := \begin{pmatrix} 2x_2 - 5x_3 \\ x_1 + 3x_2 \end{pmatrix} = \begin{pmatrix} 0 & 2 & -5 \\ 1 & 3 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Remarquons que lorsqu'une application est définie de cette façon, **il est implicitement admis que les coordonnées** (ici x_1, x_2, x_3) **sont relativement aux bases canoniques des ensembles de départ et d'arrivée**. Ici, pour les distinguer, nous noterons temporairement

- $\star$ $\mathcal{B}_{\text{can}} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ la base canonique de $\mathbb{R}^3$,
- $\star$ $\mathcal{B}_{\text{can}} = (\mathbf{e}_1, \mathbf{e}_2)$ la base canonique de $\mathbb{R}^2$.

Donc la matrice ci-dessus est en fait

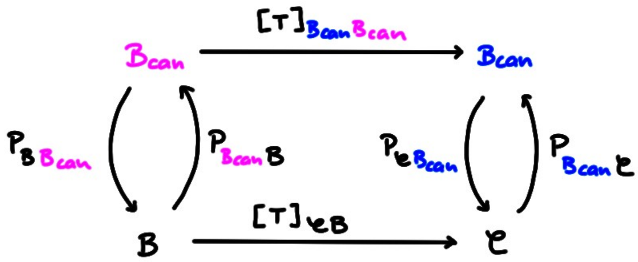
$$\begin{aligned}
[T]_{\mathcal{B}_{\text{can}}\mathcal{B}_{\text{can}}} &= \left[[T(\mathbf{e}_1)]_{\mathcal{B}_{\text{can}}} [T(\mathbf{e}_2)]_{\mathcal{B}_{\text{can}}} [T(\mathbf{e}_3)]_{\mathcal{B}_{\text{can}}} \right] \\
&= \begin{pmatrix} 0 & 2 & -5 \\ 1 & 3 & 0 \end{pmatrix}
\end{aligned}$$

Considérons maintenant les bases $\mathcal{B} = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ de $\mathbb{R}^3$ et $\mathcal{C} = (\mathbf{c}_1, \mathbf{c}_2)$ de $\mathbb{R}^2$, où

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad \mathbf{b}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

$$\mathbf{c}_1 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \mathbf{c}_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Calculons la matrice de T relativement à ces deux nouvelles bases, $[T]_{\mathcal{B}\mathcal{C}}$. On peut s'aider du schéma



pour retrouver la formule :

$$[T]_{\mathcal{C}\mathcal{B}} = P_{\mathcal{C}\mathcal{B}_{\text{can}}} [T]_{\mathcal{B}_{\text{can}}\mathcal{B}_{\text{can}}} P_{\mathcal{B}_{\text{can}}\mathcal{B}}$$

Or puisque les vecteurs de $\mathcal{B}$ ont été donnés en composantes relativement à la base canonique,

$$[\mathbf{b}_1]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad [\mathbf{b}_2]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad [\mathbf{b}_3]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

on a déjà

$$P_{\mathcal{B}_{\text{can}}\mathcal{B}} = \left[[\mathbf{b}_1]_{\mathcal{B}_{\text{can}}} \, [\mathbf{b}_2]_{\mathcal{B}_{\text{can}}} \, [\mathbf{b}_3]_{\mathcal{B}_{\text{can}}} \right] = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}.$$

Ensuite,

$$[\mathbf{c}_1]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad [\mathbf{c}_2]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 2 \\ 3 \end{pmatrix},$$

et donc

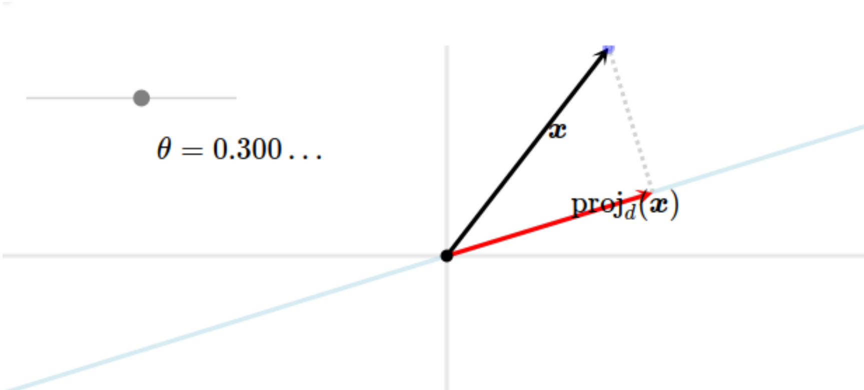
$$\begin{aligned}
P_{\mathcal{C}\mathcal{B}_{\text{can}}} &= P_{\mathcal{B}_{\text{can}}\mathcal{C}}^{-1} \\
&= \left[[\mathbf{c}_1]_{\mathcal{B}_{\text{can}}} \, [\mathbf{c}_2]_{\mathcal{B}_{\text{can}}} \right]^{-1} \\
&= \begin{pmatrix} 0 & 2 \\ -1 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} 3/2 & -1 \\ 1/2 & 0 \end{pmatrix}
\end{aligned}$$

On a donc

$$\begin{aligned}
[T]_{\mathcal{C}\mathcal{B}} &= P_{\mathcal{C}\mathcal{B}_{\text{can}}}[T]_{\mathcal{B}_{\text{can}}\mathcal{B}_{\text{can}}}P_{\mathcal{B}_{\text{can}}\mathcal{B}} \\
&= \begin{pmatrix} 3/2 & -1 \\ 1/2 & 0 \end{pmatrix} \begin{pmatrix} 0 & 2 & -5 \\ 1 & 3 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} -1 & 13/2 & 0 \\ 1 & 5/2 & 1 \end{pmatrix}
\end{aligned}$$

◇

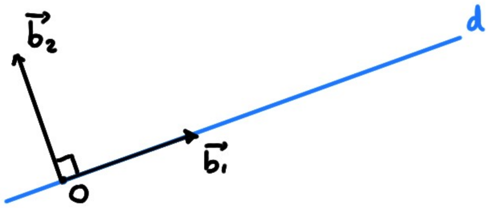
Exemple 8.7. Considérons la projection proj_d sur une droite d passant par l'origine et faisant un angle de θ avec $\mathbf{e}_1$:



Rappelons que sa matrice relativement à la base canonique est donnée par

$$[\text{proj}_d]_{\mathcal{B}_{can}} = \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}$$

Plus naturelle, pour décrire cette projection, serait une base dans laquelle les vecteurs sont orientés dans des directions qui tiennent compte de la position de l'axe d. Par exemple, une base $\mathcal{B} = (\mathbf{b}_1, \mathbf{b}_2)$ où $\mathbf{b}_1$ dirige d, et $\mathbf{b}_2$ est perpendiculaire à d :



Par définition de la projection,

$$\text{proj}_d(\mathbf{b}_1) = \mathbf{b}_1\,, \qquad \text{proj}_d(\mathbf{b}_2) = \mathbf{0}\,,$$

ce qui donne

$$[\text{proj}_d(\mathbf{b}_1)]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\,, \qquad [\text{proj}_d(\mathbf{b}_2)]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\,.$$

Par conséquent, la matrice relativement à cette base prend une forme particulièrement simple :

$$[\text{proj}_d]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Vérifions que c'est bien ce que l'on obtient en faisant le changement de base, de $\mathcal{B}_{\text{can}}$ vers $\mathcal{B}$.

Tout d'abord, on écrit explicitement les vecteurs de la nouvelle base en fonction de ceux de l'ancienne. Puisque d fait un angle θ avec l'horizontale, en les prenant orientés comme sur la figure ci-dessus, et unitaires,

$$[\mathbf{b}_1]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \qquad [\mathbf{b}_2]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

Ainsi, la matrice de changement de base est

$$P_{\mathcal{B}_{\text{can}}\mathcal{B}} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

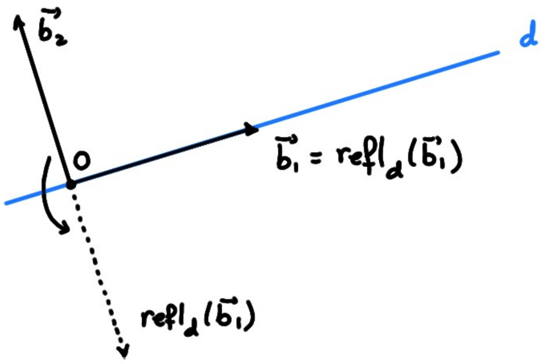
La formule du changement de base donne donc

$$\begin{aligned}
[\text{proj}_d]_{\mathcal{B}} &= P_{\mathcal{B}_{\text{can}}\mathcal{B}}{}^{-1}[\text{proj}_d]_{\mathcal{B}_{\text{can}}} P_{\mathcal{B}_{\text{can}}\mathcal{B}} \\
&= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.
\end{aligned}$$

◇

Dans dernier exemple, on a observé qu'une application (la projection) prenait une forme plus simple quand on la regardait dans une base qui était adaptée à la géométrie du problème. Faisons maintenant l'inverse : prenons une transformation, définie dans une base naturelle, et voyons quelle forme elle prend dans une autre base :

Exemple 8.8. Considérons la **réflexion** par rapport à une droite d qui passe par l'origine, que nous avions notée refl_d. Utilisons à nouveau la base $\mathcal{B} = (\mathbf{b}_1, \mathbf{b}_2)$ du dernier exemple, où $\mathbf{b}_1$ dirige d, et $\mathbf{b}_2$ est perpendiculaire à d. On remarque que l'application de la réflexion sur ces vecteurs prend une forme très simple :



$$\text{refl}_d(\mathbf{b}_1) = \mathbf{b}_1 \qquad \text{refl}_d(\mathbf{b}_2) = -\mathbf{b}_2\,.$$

Par conséquent,

$$[\text{refl}_d]_{\mathcal{B}} = \big[[\text{refl}_d(\mathbf{b}_1)]_{\mathcal{B}}\,[\text{refl}_d(\mathbf{b}_2)]_{\mathcal{B}}\big] = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Maintenant, exprimons la matrice de refl_d relativement à la base canonique : Puisque la matrice de changement de base est la même qu'avant,

$$\begin{aligned}
[\mathrm{refl}_d]_{\mathcal{B}_{\mathrm{can}}} &= P_{\mathcal{B}_{\mathrm{can}}\mathcal{B}}[\mathrm{refl}_d]_{\mathcal{B}} P^{-1}_{\mathcal{B}_{\mathrm{can}}\mathcal{B}} \\
&= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \\
&= \begin{pmatrix} \cos^2\theta - \sin^2\theta & 2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \sin^2\theta - \cos^2\theta \end{pmatrix} \\
&= \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix} .
\end{aligned}$$

Cette expression est bien celle que nous avions obtenue précédemment. ◇

Chapitre 9

Déterminant

9.1 Motivation : le cas 2×2 revisited

La théorie du déterminant, que nous allons aborder dans ce chapitre, est un sujet central en algèbre linéaire. Nous ne le présenterons pas dans sa forme la plus générale, et ne démontrerons pas tous les résultats. Notre but sera de présenter les propriétés de base du déterminant, et de voir leurs conséquences.

Nous avons déjà rencontré le déterminant lorsque nous avons étudié l'inversibilité pour les matrices 2×2. En effet, nous avons vu qu'une matrice

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

est inversible si et seulement si son **déterminant**, qui est le nombre défini par

$$\det(A) := ad - bc\,,$$

est non-nul.

Pouvoir savoir si une matrice 2×2 est inversible ou pas, simplement en calculant un nombre et en vérifiant s'il est nul ou pas, représente certainement un résultat intéressant du point de vue théorique, mais l'étendre au cas $n \times n$ ne sera pas sans difficulté.

En effet, dans le cas $n \times n$, nous avons vu quelques caractérisations équivalentes de l'inversibilité, mais toutes étaient de nature plus calculatoire, et toutes impliquaient plus ou moins l'étude d'un système linéaire.

Pour motiver ce que nous allons faire dans le cas $n \times n$, nous allons revenir sur le cas 2×2, et regarder de plus près cette fonction $A \mapsto \det(A)$, pour nous rendre compte de certaines caractéristiques, et sans du tout nous préoccuper de l'inversibilité.

Plutôt que de voir une matrice 2×2 comme un tableau de 4 nombres rangés dans une grille, voyons la comme la donnée de deux colonnes :

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = [\mathbf{a}_1\, \mathbf{a}_2]\,,$$

où

$$\mathbf{a}_1 = \begin{pmatrix} a \\ c \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} b \\ d \end{pmatrix}\,.$$

Ainsi, le déterminant peut être vu comme une fonction sur les paires de vecteurs de $\mathbb{R}^2$, définie ainsi :

$$\det\left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}\right) := ad - bc\,.$$

Pendant un instant, nous n'utiliserons plus la notation "det", et considérerons seulement la fonction

$$\begin{aligned} \varphi : \mathbb{R}^2 \times \mathbb{R}^2 &\to \mathbb{R} \\ (\mathbf{a}, \mathbf{b}) &\mapsto \varphi(\mathbf{a}\,\mathbf{b}) := a_1 b_2 - a_2 b_1 \,. \end{aligned}$$

Les propriétés suivantes découlent entièrement de sa définition :

Proposition 8. *φ définie ci-dessus jouit des propriétés suivantes :*

1) $\varphi(\mathbf{a}, \mathbf{b}) = -\varphi(\mathbf{b}, \mathbf{a})$ ***(antisymétrie)***
2) $\varphi(\mathbf{a}, \mathbf{a}) = 0$ ***(alternance)***
3) *Pour tout $\mathbf{b} \in \mathbb{R}^2$ fixé, l'application $\mathbf{a} \mapsto \varphi(\mathbf{a}, \mathbf{b})$ est linéaire :*
   - ⋆ $\varphi(\lambda\mathbf{a}, \mathbf{b}) = \lambda\varphi(\mathbf{a}, \mathbf{b})$ *pour tout* $\lambda \in \mathbb{R}$, $\mathbf{a} \in \mathbb{R}^2$
   - ⋆ $\varphi(\mathbf{a} + \mathbf{a}', \mathbf{b}) = \varphi(\mathbf{a}, \mathbf{b}) + \varphi(\mathbf{a}', \mathbf{b})$
4) *Pour tout $\mathbf{a} \in \mathbb{R}^2$ fixé, l'application $\mathbf{b} \mapsto \varphi(\mathbf{a}, \mathbf{b})$ est linéaire :*
   - ⋆ $\varphi(\mathbf{a}, \lambda\mathbf{b}) = \lambda\varphi(\mathbf{a}, \mathbf{b})$ *pour tout* $\lambda \in \mathbb{R}$, $\mathbf{b} \in \mathbb{R}^2$
   - ⋆ $\varphi(\mathbf{a}, \mathbf{b} + \mathbf{b}') = \varphi(\mathbf{a}, \mathbf{b}) + \varphi(\mathbf{a}, \mathbf{b}')$
5) $\varphi(\mathbf{e}_1, \mathbf{e}_2)$*=1* ***(normalisation)***

Preuve: Toutes les propriétés sont vérifiées directement par le calcul :

1) $\varphi(\mathbf{a}, \mathbf{b}) = a_1 b_2 - a_2 b_1 = -(a_2 b_1 - a_1 b_2) = -\varphi(\mathbf{b}, \mathbf{a})$
2) $\varphi(\mathbf{a}, \mathbf{a}) = a_1 a_2 - a_2 a_1 = 0$
3) $\varphi(\lambda\mathbf{a}, \mathbf{b}) = (\lambda a_1) b_2 - (\lambda a_2) b_1 = \lambda(a_1 b_2 - a_2 b_1) = \lambda\varphi(\mathbf{a}, \mathbf{b})$, et

$$\begin{aligned} \varphi(\mathbf{a} + \mathbf{a}', \mathbf{b}) &= (a_1 + a_1')b_2 - (a_2 + a_2')b_1 \\ &= (a_1 b_2 - a_2 b_1) + (a_1' b_2 - a_2' b_1) \\ &= \varphi(\mathbf{a}, \mathbf{b}) + \varphi(\mathbf{a}', \mathbf{b})\,. \end{aligned}$$

4) (pareil)
5) $\varphi(\mathbf{e}_1, \mathbf{e}_2) = 1 \cdot 1 - 0 \cdot 0 = 1.$

□

Remarque 9.1. ⋆ Du moment que la linéarité est vérifiée, les propriétés 1 et 2 sont équivalentes : elles peuvent être déduites l'une à partir de l'autre. En effet, si 1 est vraie, alors $\varphi(\mathbf{a}, \mathbf{a}) = -\varphi(\mathbf{a}, \mathbf{a})$, et donc $\varphi(\mathbf{a}, \mathbf{a}) = 0$. Inversément, si 2 est vraie, alors $\varphi(\mathbf{a} + \mathbf{b}, \mathbf{a} + \mathbf{b}) = 0$, mais la linéarité permet implique

$$\varphi(\mathbf{a} + \mathbf{b}, \mathbf{a} + \mathbf{b}) = \underbrace{\varphi(\mathbf{a}, \mathbf{a})}_{=0} + \varphi(\mathbf{a}, \mathbf{b}) + \varphi(\mathbf{b}, \mathbf{a}) + \underbrace{\varphi(\mathbf{b}, \mathbf{b})}_{=0},$$

ce qui implique $\varphi(\mathbf{b}, \mathbf{a}) = -\varphi(\mathbf{a}, \mathbf{b})$.

⋆ Les propriétés 3 et 4 impliquent que φ est linéaire en chacune de ses variables ; on dit que c'est une application **bilinéaire**.

◇

Il se trouve que les propriétés 1 à 5 énoncées dans la proposition *caractérisent entièrement* la fonction φ ! En d'autres termes, nous allons bientôt voir que si une autre fonction $\psi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ satisfait aux cinq propriétés ci-dessus, alors ceci entraîne que $\psi(\mathbf{a}, \mathbf{b}) = \varphi(\mathbf{a}, \mathbf{b}) = a_1 b_2 - a_2 b_1$.

Nous utiliserons cette caractéristique pour définir le déterminant en dimensions supérieures : nous introduirons une fonction sur les familles de n vecteurs de $\mathbb{R}^n$, en imposant quelques propriétés semblables à celles énoncées ci-dessus, et énoncerons un résultat qui garantit qu'il existe une seule fonction ayant ces propriétés ; c'est ce que nous appellerons le *déterminant*.

9.2 Cas général $n \times n$

Pour définir une notion générale de déterminant, nous allons devoir considérer une *famille* de fonctions.

Nous commencerons avec les fonctions définies sur les paires de vecteurs de $\mathbb{R}^2$ (qui peuvent être vues comme des matrices 2×2), comme dans la sections précédente :

$$\varphi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}\,.$$

Puis, nous considérerons des fonctions sur des triplets de vecteurs de $\mathbb{R}^3$ (qui peuvent être vus comme des matrices 3×3),

$$\varphi : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}\,,$$

puis sur des quadruplets de vecteurs de $\mathbb{R}^4$ (qui peuvent être vus comme des matrices 4×4),

$$\varphi : \mathbb{R}^4 \times \mathbb{R}^4 \times \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}\,,$$

etc.

Nous allons imposer que toutes ces fonctions satisfassent à des propriétés semblables à celles du cas $n = 2$. Nous utiliserons alors un résultat général qui dit qu'il n'y a qu'une seule famille de fonctions vérifiant toutes ces propriétés ; pour un n particulier, φ sera alors la fonction utilisée pour calculer le déterminant des matrices $n \times n$.

Commençons par définir les propriétés, qui généralisent celles du cas 2×2.

Définition 9.2. Une fonction définie sur les familles ordonnées de n vecteurs de $\mathbb{R}^n$

$$\begin{aligned} \varphi : \mathbb{R}^n \times \cdots \times \mathbb{R}^n &\to \mathbb{R} \\ (\mathbf{a}_1, \dots, \mathbf{a}_n) &\mapsto \varphi(\mathbf{a}_1, \dots, \mathbf{a}_n) \end{aligned}$$

est dite

- ⋆ **multilinéaire** si elle est linéaire en chacun de ses vecteurs. (Plus précisément, si pour tout $k = 1, 2, \dots, n$, et pour tous vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_{k-1}, \mathbf{a}_{k+1}, \dots, \mathbf{a}_n$ fixés, l'application

$$\begin{aligned} \mathbb{R}^n &\to \mathbb{R} \\ \mathbf{x} &\mapsto \varphi(\mathbf{a}_1, \dots, \mathbf{a}_{k-1}, \mathbf{x}, \mathbf{a}_{k+1}, \dots, \mathbf{a}_n) \end{aligned}$$

  est linéaire.)
- ⋆ **alternée** si $\varphi(\mathbf{a}_1, \dots, \mathbf{a}_n) = 0$ dès que deux des vecteurs $\mathbf{a}_i$ sont égaux.
- ⋆ **normalisée** si $\varphi(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1$.

Remarque 9.3. On peut montrer, exactement comme on l'a fait dans le cas $n = 2$ (section précédente), que la première propriété d'alternance, conjuguée à celle de multilinéarité, est en fait équivalente à celle d'**antisymétrie** : si on échange deux vecteurs, on change le signe de la fonction :

$$\varphi(\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n) = -\varphi(\mathbf{a}_1, \dots, \mathbf{a}_j, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n)\,.$$

⋄

Puisqu'une famille $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^n$ permet de définir une matrice $n \times n$

$$A := [\mathbf{a}_1 \cdots \mathbf{a}_n]\,,$$

on peut voir $\varphi : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ comme une fonction $\varphi : \mathbb{M}_{n \times n}(\mathbb{R}) \to \mathbb{R}$, en posant

$$\varphi(A) := \varphi(\mathbf{a}_1, \dots, \mathbf{a}_n)\,.$$

Définition 9.4. Si A est une matrice $m \times n$ et si (i,j) est une paire d'indice ($1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$), alors A_{ij} représente la matrice $(m-1) \times (n-1)$ obtenue à partir de A en traçant la i-ème ligne et la j-ème colonne.

Exemple 9.5. Si $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$, alors

$$A_{12} = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}, \quad A_{31} = \begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix}, \quad A_{22} = \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix}.$$

◇

Théorème 9.6. *Pour chaque $n \geqslant 2$, il existe une unique fonction*

$$\begin{aligned} \widetilde{\varphi}_n : \mathbb{R}^n \times \cdots \times \mathbb{R}^n &\to \mathbb{R} \\ (\mathbf{a}_1, \dots, \mathbf{a}_n) &\mapsto \widetilde{\varphi}_n(\mathbf{a}_1, \dots, \mathbf{a}_n) \end{aligned}$$

qui soit en même temps multilinéaire, alternée, et normalisée.

De plus, ces fonctions obéissent au shéma récursif suivant :

- ⋆ *pour $n = 2$,*

$$\widetilde{\varphi}_2(\mathbf{a}, \mathbf{b}) = a_1 b_2 - a_2 b_1 ,$$

- ⋆ *pour $n > 2$, $\widetilde{\varphi}_n$ peut se calculer à l'aide de $\widetilde{\varphi}_{n-1}$ par l'une quelconque des relations suivantes : si A est $n \times n$,*
  - — ***développement selon la i-ème ligne de A :***

$$\widetilde{\varphi}_n(A) = \sum_{k=1}^{n} (-1)^{k+i} a_{ik} \widetilde{\varphi}_{n-1}(A_{ik}) ,$$

  - — ***développement selon la j-ème colonne de A***

$$\widetilde{\varphi}_n(A) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \widetilde{\varphi}_{n-1}(A_{kj}) .$$

Preuve: (omise) □

Le théorème dit en particulier, comme nous avions annoncé précédemment, que la fonction $\varphi(\mathbf{a}, \mathbf{b})$ introduite dans le cas 2×2 est la seule qui possède les trois propriétés simultanément.

Mais le théorème dit aussi comment calculer explicitement ces fonctions *de manière récursive*, en calculant $\widetilde{\varphi}_n$ à l'aide de $\widetilde{\varphi}_{n-1}$.

Exemple 9.7. Considérons la matrice 3×3

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

Pour calculer $\widetilde{\varphi}_3(A)$, le théorème dit que nous avons pas moins de 6 façons équivalentes de procéder.

Par exemple, en développant selon la première ligne,

$$\begin{aligned}\widetilde{\varphi}_3(A) = &\sum_{k=1}^{3}(-1)^{k+1}a_{1k}\widetilde{\varphi}_2(A_{1k})\\ =&(-1)^{1+1}a_{11}\widetilde{\varphi}_2(A_{11}) + (-1)^{2+1}a_{12}\widetilde{\varphi}_2(A_{12}) + (-1)^{3+1}a_{13}\widetilde{\varphi}_2(A_{13})\\ =&\widetilde{\varphi}_2(\begin{pmatrix}5\\8\end{pmatrix}, \begin{pmatrix}6\\9\end{pmatrix}) - 2\widetilde{\varphi}_2(\begin{pmatrix}4\\-7\end{pmatrix}, \begin{pmatrix}6\\9\end{pmatrix}) - 3\widetilde{\varphi}_2(\begin{pmatrix}4\\-7\end{pmatrix}, \begin{pmatrix}5\\8\end{pmatrix})\\ =&(5\cdot 9 - 8\cdot 6) - 2(4\cdot 9 - (-7)\cdot 6) - 3(4\cdot 8 - (-7)\cdot 5)\\ =&-360\,.\end{aligned}$$

Ou alors, en développant selon la 3-ème colonne,

$$\begin{aligned}\widetilde{\varphi}_3(A) = &\sum_{k=1}^{3}(-1)^{k+3}a_{k3}\widetilde{\varphi}_2(A_{k3})\\ =&(-1)^{1+3}a_{13}\widetilde{\varphi}_2(A_{13}) + (-1)^{2+3}a_{23}\widetilde{\varphi}_2(A_{23}) + (-1)^{3+3}a_{33}\widetilde{\varphi}_2(A_{33})\\ =&(-3)\widetilde{\varphi}_2(\begin{pmatrix}4\\-7\end{pmatrix}, \begin{pmatrix}5\\8\end{pmatrix}) - 6\widetilde{\varphi}_2(\begin{pmatrix}1\\-7\end{pmatrix}, \begin{pmatrix}2\\8\end{pmatrix}) + 9\widetilde{\varphi}_2(\begin{pmatrix}1\\4\end{pmatrix}, \begin{pmatrix}2\\5\end{pmatrix})\\ =&-3(4\cdot 8 - (-7)\cdot 5) - 6(1\cdot 8 - (-7)\cdot 2) + 9(1\cdot 5 - 4\cdot 2)\\ =&-360\,.\end{aligned}$$

◇

9.3 Propriétés

Le *déterminant* d'une matrice $n \times n$ est donc un nombre réel, défini à partir de la fonction $\widetilde{\varphi}_n$ dont l'existence a été démontrée dans la section précédente :

Définition 9.8. Pour une matrice $n \times n$, définie par ses colonnes,

$$A = [\mathbf{a}_1 \cdots \mathbf{a}_n]\,,$$

le **déterminant** est défini par

$$\det(A) := \widetilde{\varphi}_n(\mathbf{a}_1, \dots, \mathbf{a}_n)\,.$$

Parfois, pour indiquer la dépendence du déterminant en ses colonnes, on écrira

$$\det(A) := \det(\mathbf{a}_1, \dots, \mathbf{a}_n)\,.$$

Informel 9.9. Dans la littérature, le déterminant de A est parfois noté $|A|$. Nous utiliserons rarement cette notation, car elle rappelle la *valeur absolue*, et donne donc l'impression qu'un déterminant doit toujours être une quantité positive, ce qui n'est pas du tout le cas bien-sûr.

Propriétés

Les propriétés énoncées dans le théorème de la section précédente se résument comme suit :

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

⋆ Développement selon la i-ème ligne d'une matrice $n \times n$:

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+i} a_{ik} \det(A_{ik})$$

⋆ Développement selon la j-ème colonne d'une matrice $n \times n$:

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(A_{kj})$$

Par ces relations de récurrence, le déterminant d'une matrice $n \times n$ peut toujours se calculer en passant par le calcul de n déterminants de sous-matrices $(n-1) \times (n-1)$. Mais à leur tour, le déterminant de chacune de ces matrices $(n-1) \times (n-1)$ passe par le calcul de $n-1$ matrices $(n-2) \times (n-2)$, etc. Ainsi, si N_n représente le nombre d'opérations nécessaires pour calculer le déterminant d'une matrice $n \times n$, on a

$$\begin{aligned} N_n &= nN_{n-1} \\ &= n(n-1)N_{n-2} \\ &= n(n-1)(n-2)N_{n-3} \\ &\vdots \\ &= n(n-1)(n-3)\cdots 4 \cdot 3 \cdot N_2 = \frac{n!}{2} N_2 \,. \end{aligned}$$

Ainsi, le nombre d'opérations augmente **factoriellement** avec n, ce qui rend un calcul de déterminant, a priori, très couteux pour des grandes matrices.

Exemple 9.10. Prenons une matrice 10×10, par exemple

$$A = \begin{pmatrix} 8 & 8 & 1 & 0 & 6 & 1 & 9 & 7 & 5 & 5 \\ 1 & 3 & 3 & 9 & 7 & 4 & 7 & 1 & 5 & 5 \\ 8 & 5 & 6 & 7 & 8 & 3 & 0 & 8 & 9 & 7 \\ 5 & 5 & 4 & 7 & 3 & 8 & 3 & 4 & 5 & 8 \\ 9 & 6 & 8 & 5 & 5 & 3 & 6 & 4 & 3 & 6 \\ 1 & 8 & 9 & 5 & 3 & 7 & 4 & 3 & 5 & 3 \\ 5 & 4 & 1 & 3 & 9 & 6 & 8 & 9 & 3 & 2 \\ 5 & 2 & 1 & 9 & 9 & 5 & 6 & 0 & 4 & 5 \\ 7 & 3 & 6 & 8 & 6 & 0 & 6 & 7 & 3 & 6 \\ 3 & 8 & 2 & 2 & 6 & 6 & 3 & 5 & 9 & 6 \end{pmatrix}$$

Par ce que nous avons dit plus haut, le calcul de $\det(A)$ requiert environ $10!$ (factorielle) opérations, ce qui est de l'ordre de $3'628'800$. Avec une matrice 100×100, le nombre d'opérations est de l'ordre de 10^{158} ; il faudrait à n'importe quel ordinateur, même très puissant, un temps bien supérieur à l'âge de l'univers pour effectuer ce calcul (source : Rappaz-Picasso.)

Remarque : La matrice ci-dessus a été **générée aléatoirement** (lien web). ⋄

Donc en général, on ne calcule pas un déterminant en utilisant les relations de récursion....

En revanche, ce qu'on peut faire est d'utiliser les relations de récurrence, ainsi que les propriétés de base des fonctions $\widetilde{\varphi}_n$, pour dériver d'autres propriétés générales du déterminant. Celles-ci fourniront des méthodes permettant d'économiser autant que possible sur le nombre d'opérations à effectuer pour calculer un déterminant, en *simplifiant* la matrice.

NumChap: chap-determinant, Dernière compilation: 2024-08-12 07:19:48+02:00. (Version Web: botafogo.saitis.net)

Propriétés du déterminant

D'abord, un résultat préliminaire :

Lemme 13. $\det(A^T) = \det(A)$

Preuve: Par récurrence sur n. Lorsque $n = 2$, on a simplement

$$\det(A) = \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$
$$= ad - cb = \det\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \det(A^T)\,.$$

Supposons maintenant que la formule soit correcte pour toute matrice $n \times n$, et considérons une matrice $(n+1)\times(n+1)$, notée A. En développant selon la première colonne, puisque le coefficient $j1$ de A^T est le coefficient $1j$ de A, à savoir a_{1j}, on a que

$$\det(A^T) = \sum_{j=1}^{n+1} (-1)^{j+1} a_{1j} \det((A^T)_{j1})\,.$$

Mais par la définition de la transposition, $(A^T)_{j1} = (A_{1j})^T$. De plus, par l'hypothèse d'induction, A_{1j} étant une matrice $(n-1)\times(n-1)$,

$$\det((A_{1j})^T) = \det(A_{1j})\,,$$

ce qui donne

$$\det(A^T) = \sum_{j=1}^{n+1} (-1)^{j+1} a_{1j} \det(A_{1j}) = \det(A)\,.$$

En effet, cette dernière somme est le déterminant de A, développé selon la première ligne. □

Ensuite, les propriétés qui permettront de simplifier le calcul du déterminant :

Proposition 9. *(Propriétés du déterminant)*

1) *Si A possède deux colonnes (ou deux lignes) égales, alors* $\det(A) = 0$.
2) *Le signe du déterminant change lorsqu'on échange deux colonnes :*

$$\det(\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n)$$
$$= -\det(\mathbf{a}_1, \dots, \mathbf{a}_j, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n)$$

   (Pareil si on échange deux lignes.)
3) *Lorsqu'on multiplie une colonne par λ, le déterminant est multiplié par λ :*

$$\det(\mathbf{a}_1, \dots, \lambda\mathbf{a}_i, \dots, \mathbf{a}_n) = \lambda \det(\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n)$$

   (Pareil si on multiplie une ligne par λ.)
4) *Lorsqu'on rajoute un multiple d'une colonne à une autre colonne, on ne change pas le déterminant :*

$$\det(\mathbf{a}_1, \dots, \mathbf{a}_i + \lambda\mathbf{a}_j, \dots, \mathbf{a}_n) = \det(\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n)$$

   (Pareil si on rajoute un multiple d'une ligne à une autre ligne.)

Preuve: Ces propriétés suivent directement du fait que le déterminant est une fonction des colonnes ($\widetilde{\varphi}_n$), qui est alternée et multilinéaire.

Par exemple, si deux colonnes de A sont égales, $\det(A) = 0$ suit immédiatement du fait que $\det(A) = \widetilde{\varphi}_n(\mathbf{a}_1, \dots, \mathbf{a}_n)$, et que $\widetilde{\varphi}_n$ est alternée. Puis, si deux lignes de A sont égales, alors deux colonnes de A^T sont égales, et donc $\det(A^T) = 0$. Par le lemme précédent, $\det(A) = 0$. □

Exemple 9.11. 1) Deux colonnes égales :

$$\det\begin{pmatrix}-2 & 0 & 8 & 0\\ 5 & 1 & -8 & 1\\ -2 & -3 & 0 & -3\\ 4 & 7 & 1 & 7\end{pmatrix} = 0$$

2) Échange de deux colonnes :

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

3) Extraction d'une constante sur une colonne :

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

4) Rajouter un multiple d'une ligne à une autre :

$$\det\begin{pmatrix}-1 & 2 & 0 & 4\\ 3 & 4 & -7 & 3\\ 5 & -2 & 2 & 5\\ 1 & 6 & -3 & 0\end{pmatrix} = \det\begin{pmatrix}9 & -2 & 4 & 14\\ 3 & 4 & -7 & 3\\ 5 & -2 & 2 & 5\\ 1 & 6 & -3 & 0\end{pmatrix}$$

(On a rajouté 2 fois la troisième ligne à la première.)

◇

Ensuite, il existe des matrices dont le calcul du déterminant ne requiert aucune opération particulière :

Définition 9.12. Une matrice carrée A est **triangulaire supérieure** (resp. **inférieure**) si $a_{ij} = 0$ dès que $i > j$ (resp. $i < j$).

Exemple 9.13. Une matrice 4×4 triangulaire supérieure :

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

◇

Lemme 14. *Si A est une matrice triangulaire supérieure ou inférieure, alors son déterminant est le produit de ses termes diagonaux :*

$$\det(A) = a_{11}\cdots a_{nn} =: \prod_{j=1}^{n} a_{jj}\,.$$

Preuve: Montrons la première affirmation pour les matrices triangulaires supérieures, par récurrence sur n. Pour $n = 2$, on a

$$\det \begin{pmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{pmatrix} = a_{11}a_{22} - 0 \cdot a_{12} = \prod_{j=1}^{2} a_{jj} \,.$$

Supposons que le résultat est prouvé pour un certain entier n, et considérons une matrice A, $(n+1) \times (n+1)$, triangulaire supérieure. En développant selon la première colonne, et en utilisant le fait que tous les $a_{j1} = 0$ lorsque $j = 2, \dots, n+1$, il ne reste que le terme $j = 1$. De plus, puisque A_{11} est une matrice $n \times n$ triangulaire supérieure, on peut utiliser l'hypothèse d'induction :

$$\begin{aligned} \det(A) &= \sum_{j=1}^{n+1} (-1)^{j+1} a_{j1} \det(A_{j1}) \\ &= (-1)^{1+1} a_{11} \det(A_{11}) \\ &= a_{11} \prod_{j=2}^{n+1} a_{jj} \\ &= \prod_{j=1}^{n+1} a_{jj} \,. \end{aligned}$$

Si A est triangulaire inférieure, alors A^T est triangulaire supérieure et ses éléments diagonaux sont les mêmes, donc le résultat est aussi vrai. □

En particulier, si A est **diagonale**,

$$A = \mathrm{diag}(d_1, \dots, d_n) := \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{pmatrix}$$

alors

$$\det(A) = d_1 \cdots d_n = \prod_{j=1}^{n} d_j \,.$$

Exemple 9.14.

$$\det \begin{pmatrix} 1 & 76 & -21 & 98 & -5 & 99 \\ 0 & \sqrt{2} & 0 & -6 & 98 & \\ 0 & 0 & 32 & 53 & 75 & 97 \\ 0 & 0 & 0 & 0 & 42 & 14 \\ 0 & 0 & 0 & 0 & 21 & 32 \\ 0 & 0 & 0 & 0 & 0 & 95 \end{pmatrix} = 1 \cdot \sqrt{2} \cdot 32 \cdot 0 \cdot 21 \cdot 95 = 0 \,.$$

◇

Exemple 9.15. Matrice identité :

$$\det(I_n) = \det \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} = 1^n = 1 \,.$$

◇

Les propriétés énoncées jusqu'ici fournissent déjà de quoi calculer un déterminant en évitant de le développer systématiquement à l'aide des relations de récurrence. En effet, on a vu que les déterminants les plus simples à calculer sont ceux des matrices triangulaires, et aussi que des opérations sur les colonnes et les lignes correspondent à certaines modifications simples du déterminant. On pourra donc appliquer des opérations sur les lignes et les colonnes, dans le but de rendre la matrice triangulaire supérieure, ou au moins avec autant de zéros que possible, ce qui ensuite d'utiliser les relations de récurrence pour une matrice simplifiée.

Exemple 9.16. Utilisons les propriétés pour calculer le déterminant de la matrice

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

On fait déjà apparaître quelques zéros en soustrayant la troisième ligne de la deuxième, ce qui ne change pas le déterminant :

$$\det \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 2 & 6 & 4 & 8 \\ 3 & 1 & 1 & 2 \end{pmatrix} = \det \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 0 & 3 & 0 \\ 2 & 6 & 4 & 8 \\ 3 & 1 & 1 & 2 \end{pmatrix}$$

Ensuite, en soustrayant la première colonne à la troisième,

$$\det \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 0 & 3 & 0 \\ 2 & 6 & 4 & 8 \\ 3 & 1 & 1 & 2 \end{pmatrix} = \det \begin{pmatrix} 1 & 2 & 2 & 4 \\ 3 & 0 & 0 & 0 \\ 2 & 6 & 2 & 8 \\ 3 & 1 & -2 & 2 \end{pmatrix}$$

Maintenant, on peut développer selon la deuxième ligne, puisqu'elle contient beaucoup de zéros :

$$= \det \begin{pmatrix} 1 & 2 & 2 & 4 \\ 3 & 0 & 0 & 0 \\ 2 & 6 & 2 & 8 \\ 3 & 1 & -2 & 2 \end{pmatrix} = -3 \det \begin{pmatrix} 2 & 2 & 4 \\ 6 & 2 & 8 \\ 1 & -2 & 2 \end{pmatrix}$$

En mettant en évidence un 2 dans les deux premières lignes, puis dans la dernière colonne,

$$\det \begin{pmatrix} 2 & 2 & 4 \\ 6 & 2 & 8 \\ 1 & -2 & 2 \end{pmatrix} = 2^2 \det \begin{pmatrix} 1 & 1 & 2 \\ 3 & 1 & 4 \\ 1 & -2 & 2 \end{pmatrix} = 2^3 \det \begin{pmatrix} 1 & 1 & 1 \\ 3 & 1 & 2 \\ 1 & -2 & 1 \end{pmatrix}$$

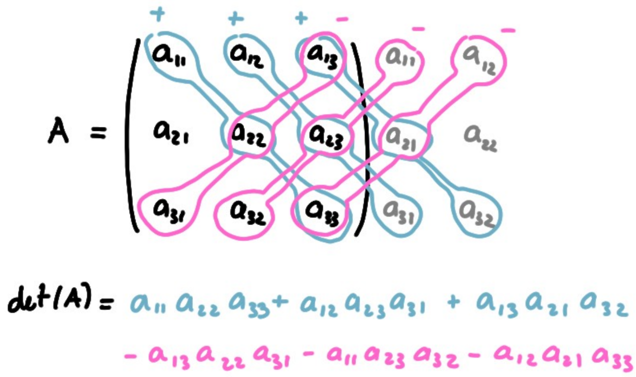
En soustrayant la dernière ligne à la première, et en développant selon la première ligne,

$$\begin{aligned} \det \begin{pmatrix} 1 & 1 & 1 \\ 3 & 1 & 2 \\ 1 & -2 & 1 \end{pmatrix} &= \det \begin{pmatrix} 0 & 3 & 0 \\ 3 & 1 & 2 \\ 1 & -2 & 1 \end{pmatrix} \\ &= -3 \det \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix} = -3 \cdot 1 = -3\,. \end{aligned}$$

donc $\det(A) = (-3) \cdot 2^3 \cdot (-3) = 72$ ◇

Une curiosité dans le cas $n = 3$

Dans le cas d'une matrice 3×3, le développement du déterminant peut se faire à l'aide de la **règle de Sarrus** (lien web). Celle-ci ne contient rien de profond, mais permet de calculer un déterminant 3×3 de façon systèmatique, facile à mémoriser. On écrit la matrice A, à la suite de laquelle on rajoute la première et la deuxième colonne. On parcours ensuite ce tableau 3×5 selon certaines diagonales :



Remarque 9.17. Il n'existe pas d'équivalent de la règle de Sarrus pour des déterminants de matrices de tailles supérieures. ◇

9.4 La formule $\det(AB) = \det(A)\det(B)$

Dans cette section, nous allons démontrer la propriété qui rend le déterminant réellement utile en algèbre linéaire :

Théorème 9.18. *Pour toute paire de matrices $n \times n$,*

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

Cas simple : A est élémentaire

Nous commencerons par démontrer le résultat dans un cas particulier :

Proposition 10. *Si E est élémentaire, alors $\det(E) \neq 0$, et pour toute matrice B,*

$$\det(EB) = \det(E)\det(B)\,.$$

Preuve:

1) Lorsque E est de Type I, $E = T_{i\leftrightarrow j}$, la matrice $T_{i\leftrightarrow j}B$ étant B avec les lignes i et j échangées, on a

$$\det(T_{i\leftrightarrow j}B) = -\det(B)$$

En utilisant cette relation avec $B = I_n$, on obtient en particulier

$$\det(T_{i\leftrightarrow j}) = -1 \neq 0\,.$$

On peut donc écrire

$$\det(T_{i\leftrightarrow j}B) = -\det(B) = \det(T_{i\leftrightarrow j})\det(B)\,.$$

2) Lorsque E est de Type II, $E = D_i(\lambda)$ (avec $\lambda \neq 0$), $D_i(\lambda)B$ est la matrice B, dans laquelle la i-ème ligne a été multipliée par λ, et donc

$$\det(D_i(\lambda)B) = \lambda\det(B)\,.$$

En utilisant cette relation avec $B = I_n$, on obtient en particulier

$$\det(D_i(\lambda)) = \lambda \neq 0$$

On peut donc écrire

$$\det(D_i(\lambda)B) = \lambda\det(B) = \det(D_i(\lambda))\det(B)\,.$$

3) Finalement, lorsque E est de Type III, $E = L_{ij}(\lambda)$, $L_{ij}(\lambda)B$ est la matrice B, dans laquelle on a rajouté λ fois la j-ème ligne à la i-ème. On sait que cette opération ne modifie pas le déterminant, et donc

$$\det(L_{ij}(\lambda)B) = \det(B)\,.$$

En prenant $B = I_n$, ceci donne en particulier

$$\det(L_{ij}(\lambda)) = 1 \neq 0\,,$$

et donc aussi

$$\det(L_{ij}(\lambda)B) = \det(L_{ij}(\lambda))\det(B)\,.$$

La proposition est démontrée. □

Preuve du théorème

Pour montrer que $\det(AB) = \det(A)\det(B)$ est vraie généralement, on va distinguer deux cas :

⋆ *Si A est inversible :* Dans cas, A peut s'écrire comme un produit de matrices élémentaires : $A = E^{(k)}\cdots E^{(1)}$. Ceci permet d'écrire, en utilisant $k-1$ fois la proposition,

$$\begin{aligned}\det(A) &= \det(E^{(k)}\cdots E^{(1)})\\ &= \det(E^{(k)})\det(E^{(k-1)}\cdots E^{(1)})\\ &= \det(E^{(k)})\det(E^{(k-1)})\det(E^{(k-2)}\cdots E^{(1)})\\ &\ \vdots\\ &= \det(E^{(k)})\cdots\det(E^{(1)})\,.\end{aligned}$$

En particulier, puisque les déterminants des matrices élémentaires de ce produit sont tous non-nuls, on a montré que **si** A **est inversible, alors** $\det(A) \neq 0$.

Mais en reproduisant essentiellement le même calcul,

$$\begin{aligned}\det(AB) &= \det(E^{(k)}\cdots E^{(1)}B)\\ &= \underbrace{\det(E^{(k)})\det(E^{(k-1)})\cdots\det(E^{(1)})}_{=\det(A)}\det(B)\\ &= \det(A)\det(B)\,.\end{aligned}$$

⋆ *Si A est singulière* : Dans ce cas, A ne peut pas être réduite à l'identité. Notons $\widetilde{E}^{(1)},\dots,\widetilde{E}^{(l)}$ les transformations qui réduisent A. Puisque

$$\widetilde{A} = \widetilde{E}^{(l)}\cdots\widetilde{E}^{(1)}A$$

n'est pas la matrice identité, c'est une matrice triangulaire supérieure possédant au moins un zéro sur sa diagonale. Ceci implique que son déterminant est nul, $\det(\widetilde{A}) = 0$. On peut donc écrire que

$$0 = \det(\widetilde{A}) = \det(\widetilde{E}^{(l)})\cdots\det(\widetilde{E}^{(1)})\det(A)\,.$$

Puisque $\det(\widetilde{E}^{(i)}) \neq 0$ pour tout i, on en déduit que $\det(A) = 0$. On a donc démontré que **si** A **est singulière, alors** $\det(A) = 0$.

Mais si A n'est pas inversible, alors AB n'est pas inversible non plus (exercice), et donc $\det(AB) = 0$. La relation suivante est donc vérifiée :

$$\underbrace{\det(AB)}_{=0} = \underbrace{\det(A)}_{=0}\det(B)\,,$$

ce qui complète la preuve du théorème.

Déterminant et inversibilité

La preuve ci-dessus (voir les passages en gras) a comme conséquence la généralisation que nous espérions, à savoir celle du critère que nous avions établi pour les matrices 2×2 :

Théorème 9.19. *A est inversible si et seulement si* $\det(A) \neq 0$.

Exemple 9.20. La matrice

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

est inversible, puisqu'en soustrayant à la première colonne la somme de toutes les autres,

$$\begin{aligned} \det(A) &= \det \begin{pmatrix} -5 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\ &= (-5)\det \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\ &= -5 \neq 0 \end{aligned}$$

◇

L'utilisation du déterminant permet maintenant d'étudier l'inversibilité de matrices contenant un *paramètre*, en évitant de devoir étudier un système.

Exemple 9.21. Pour quelles valeurs du paramètre t la matrice

$$A = \begin{pmatrix} 0 & t & -1 \\ t & 10 & 0 \\ t-1 & 1 & t \end{pmatrix}$$

est-elle inversible ?

En développant selon la première colonne,

$$\begin{aligned} \det(A) &= (-1)t\det \begin{pmatrix} t & -1 \\ 1 & t \end{pmatrix} + (t-1)\det \begin{pmatrix} t & -1 \\ 10 & 0 \end{pmatrix} \\ &= (-t)(t^2+1) + 10(t-1) \\ &= -t^3 + 9t - 10 \,. \end{aligned}$$

On sait sait donc que A est inversible si et seulement si t n'est pas racine du polynôme $P(t) = -t^3 + 9t - 10$. On remarque que $t = 2$ est racine de P, ce qui permet de factoriser (par division Euclidienne par exemple) :

$$P(t) = (t-2)(-t^2 - 2t + 5)\,.$$

Comme les racines de $-t^2 - 2t + 5$ sont $-1 \pm \sqrt{6}$, on en déduit que A est inversible si et seulement si $t \notin \{2, -1-\sqrt{6}, -1+\sqrt{6}\}$. ◇

Le déterminant de l'inverse

Lorsque A est inversible, $AA^{-1} = I_n$, et la formule démontrée plus haut permet d'écrire

$$1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})\,,$$

qui donne :

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

Le déterminant comme invariant de similitude

Définition 9.22. Deux matrices $n \times n$, A et B sont **semblables** si il existe une matrice $n \times n$ inversible M telle que

$$A = M^{-1}BM\,.$$

Lorsque A et B sont semblables, on note $A \sim B$.

Exemple 9.23. Les matrices

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

sont semblables. En effet, en prenant $M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, qui est inversible, on obtient

$$M^{-1}BM = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = A$$

◇

Proposition 11. *Si $A \sim B$, alors $\det(A) = \det(B)$.*

(On dit que le déterminant est un **invariant de similitude**.)
Preuve:

$$\begin{aligned}
\det(A) &= \det(M^{-1}BM) \\
&= \det(M^{-1})\det(B)\det(M) \\
&= \det(B)\det(M^{-1})\det(M) \\
&= \det(B)\det(M^{-1}M) \\
&= \det(B)\det(I_n) \\
&= \det(B)\,.
\end{aligned}$$

□

Le déterminant peut donc être utilisé pour démontrer à moindre frais que deux matrices ne sont *pas* semblables :

Exemple 9.24. Les matrices

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 3 & 1 \end{pmatrix} \qquad \text{et } B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 13 \\ 0 & 0 & -1 \end{pmatrix}$$

ne sont pas semblables, car $\det(A) = 6$ (en développant selon la première ligne), alors que $\det(B) = -2$. ◇

Exemple 9.25. Un exemple important de matrices semblables sont les matrices associées à une application linéaire $T : \mathbb{R}^n \to \mathbb{R}^n$, relativement à des bases différentes $\mathcal{B}$ et $\mathcal{C}$. En effet, on sait que par la formule du changement de base,

$$[T]_{\mathcal{B}} = P_{\mathcal{CB}}{}^{-1}[T]_{\mathcal{C}} P_{\mathcal{CB}}\,,$$

et donc $[T]_{\mathcal{B}} \sim [T]_{\mathcal{C}}$. Par le lemme,

$$\det([T]_{\mathcal{B}}) = \det([T]_{\mathcal{C}})\,.$$

◇

9.5 Formule de Cramer et conséquences

Dans cette section, on présente une application intéressante de la théorie du déterminant à la résolution des systèmes linéaires.

Si A est une matrice *inversible* $n \times n$, on sait que pour tout $\mathbf{b} \in \mathbb{R}^n$, le système

$$A\mathbf{x} = \mathbf{b}$$

possède exactement une solution $\mathbf{x}$, donnée par

$$\mathbf{x} = A^{-1}\mathbf{b}\,.$$

Nous allons voir comment il est possible de calculer chacune des composantes x_j de cette solution, sans passer par la connaissance de A^{-1}.

Définition 9.26. Si M est une matrice $n \times n$, et $\mathbf{z} \in \mathbb{R}^n$, $M_j(\mathbf{z})$ est la matrice $n \times n$ obtenue à partir de M en remplaçant la j-ème colonne par $\mathbf{z}$ (sans toucher aux autres colonnes).

Exemple 9.27. Si $M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$, $\mathbf{z} = \begin{pmatrix} \sqrt{2} \\ \pi \\ e \end{pmatrix}$, alors

$$M_2(\mathbf{z}) = \begin{pmatrix} 1 & \sqrt{2} & 3 \\ 4 & \pi & 6 \\ 7 & e & 9 \end{pmatrix}\,.$$

◇

Proposition 12. *(Formule de Cramer) Soit A une matrice $n \times n$ inversible, $\mathbf{b} \in \mathbb{R}^n$, et soit $\mathbf{x} \in \mathbb{R}^n$ l'unique solution du système $A\mathbf{x} = \mathbf{b}$. Alors pour tout $j \in \{1, 2, \dots, n\}$, la j-ème composante de $\mathbf{x}$ est égale à*

$$x_j = \frac{\det(A_j(\mathbf{b}))}{\det(A)}\,.$$

Preuve: Notons $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$. Calculons le produit de A par $(I_n)_j(\mathbf{x})$:

$$\begin{aligned} A((I_n)_j(\mathbf{x})) &= A\big[\mathbf{e}_1 \cdots \mathbf{e}_{j-1}\, \mathbf{x}\, \mathbf{e}_{j+1} \cdots \mathbf{e}_n\big] \\ &= \big[A\mathbf{e}_1 \cdots A\mathbf{e}_{j-1}\, A\mathbf{x}\, A\mathbf{e}_{j+1} \cdots A\mathbf{e}_n\big] \\ &= \big[\mathbf{a}_1 \cdots \mathbf{a}_{j-1}\, \mathbf{b}\, \mathbf{a}_{j+1} \cdots \mathbf{a}_n\big] \\ &= A_j(\mathbf{b})\,. \end{aligned}$$

On a donc

$$\det(A)\det\big((I_n)_j(\mathbf{x})\big) = \det\big(A((I_n)_j(\mathbf{x}))\big) = \det(A_j(\mathbf{b}))\,.$$

Or en développant selon la j-ème colonne,

$$\begin{aligned} \det\big((I_n)_j(\mathbf{x})\big) &= \sum_{k=1}^{n} (-1)^{k+j} x_k \det\big(((I_n)_j(\mathbf{x}))_{kj}\big) \\ &= \sum_{k=1}^{n} (-1)^{k+j} x_k \det\big((I_n)_{kj}\big)\,. \end{aligned}$$

Dans la deuxième ligne, on a utilisé le fait que l'on trace la colonne contenant $\mathbf{x}$, et donc cela revient au même de travailler avec I_n qu'avec $(I_n)_j(\mathbf{x})$. Ensuite, remarquons que si $k \neq j$, alors $(I_n)_{kj}$ contient une colonne et une ligne de zéros, et donc en développant celon cette ligne, on voit que $\det((I_n)_{kj}) = 0$.

Il ne subsiste donc, dans la somme ci-dessus, que le terme $k = j$:

$$\begin{aligned} \det\big((I_n)_j(\mathbf{x})\big) &= (-1)^{j+j} x_j \det\big((I_n)_{jj}\big) \\ &= x_j \det(I_{n-1}) \\ &= x_j \end{aligned}$$

Ceci démontre la formule. □

Exemple 9.28. Considérons le système linéaire $A\mathbf{x} = \mathbf{b}$ donné par

$$(*) \qquad \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \\ 7 \\ 8 \end{pmatrix}$$

Comme $\det(A) = 4! = 24 \neq 0$, la matrice est inversible et la solution du système est unique. Si on s'intéresse par exemple à la quatrième composante x_4,

$$\begin{aligned} x_4 = \frac{\det(A_4(\mathbf{b}))}{\det(A)} &= \frac{1}{24} \det \begin{pmatrix} 1 & 2 & 3 & 5 \\ 1 & 2 & 3 & 6 \\ 1 & 2 & 0 & 7 \\ 1 & 0 & 0 & 8 \end{pmatrix} \\ &= \frac{1}{4} \det \begin{pmatrix} 1 & 1 & 1 & 5 \\ 1 & 1 & 1 & 6 \\ 1 & 1 & 0 & 7 \\ 1 & 0 & 0 & 8 \end{pmatrix} \\ &= \frac{1}{4} \det \begin{pmatrix} 0 & 0 & 1 & 5 \\ 0 & 0 & 1 & 6 \\ 0 & 1 & 0 & 7 \\ 1 & 0 & 0 & 8 \end{pmatrix}\,. \end{aligned}$$

On a d'abord extrait un 2 de la deuxième colonne et un 3 de la troisième, puis on a soustrait la deuxième colonne de la première, et la troisième de la deuxième. Maintenant, en développant selon la première colonne,

$$
\begin{aligned}
x_4 &= \frac{1}{4}\det\begin{pmatrix}0&0&1&5\\0&0&1&6\\0&1&0&7\\1&0&0&8\end{pmatrix}\\
&= \frac{1}{4}(-1)\det\begin{pmatrix}0&1&5\\0&1&6\\1&0&7\end{pmatrix}\\
&= \frac{1}{4}(-1)(1)\det\begin{pmatrix}1&5\\1&6\end{pmatrix}\\
&= \frac{1}{4}(-1)(1)(6-5)\\
&= -\frac{1}{4}\,.
\end{aligned}
$$

Bien-sûr, on trouve la même chose qu'en résolvant complètement le système $(*)$, qui serait par exemple de faire $L_1 \leftarrow L_1 - L_2$, $L_2 \leftarrow L_2 - L_3$, $L_3 \leftarrow L_3 - L_4$, qui donne

$$
\begin{pmatrix}0&0&0&4\\0&0&3&0\\0&2&0&0\\1&0&0&0\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix} = \begin{pmatrix}-1\\-1\\-1\\8\end{pmatrix}
$$

◇

Une application intéressante : formule pour A^{-1}

Si le système considéré est grand, la formule de Cramer pour x_j représente un intérêt limité du point de vue calculatoire. En effet elle implique le calcul de deux déterminants, qui comme on sait représente un nombre d'opérations croissant factoriellement avec la taille du système.

Par contre, d'un point de vue théorique elle permet de dériver une formule explicite pour l'inverse d'une matrice :

Théorème 9.29. *Soit A une matrice $n \times n$ inversible, et soit C la* ***matrice*** *$n \times n$* ***des cofacteurs****, dont les coefficients sont définis par*

$$
c_{ij} := (-1)^{i+j}\det(A_{ij})\,.
$$

Alors l'inverse de A est donné par

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

Preuve: Considérons A inversible, et notons son inverse plutôt $G = (g_{ij})$, en nommant ses colonnes :

$$
G = [\mathbf{g}_1 \cdots \mathbf{g}_n]\,,
$$

où

$$
\mathbf{g}_k := \begin{pmatrix}g_{1k}\\g_{2k}\\\vdots\\g_{nk}\end{pmatrix}\,.
$$

Comme on sait, la condition $AG = I_n$ devient

$$
[A\mathbf{g}_1 \cdots A\mathbf{g}_n] = [\mathbf{e}_1 \cdots \mathbf{e}_n]\,,
$$

ce qui revient à résoudre les n systèmes

$$A\mathbf{g}_k = \mathbf{e}_k\,, \quad k = 1, \dots, n\,.$$

Par la formule de Cramer, la j-ème composante de $\mathbf{g}_k$ est donnée par

$$g_{jk} = \frac{\det(A_j(\mathbf{e}_k))}{\det(A)}\,.$$

Or, en développant selon la j-ème colonne,

$$\det(A_j(\mathbf{e}_k)) = (-1)^{j+k} \cdot 1 \cdot \det(A_{kj})\,,$$

ce qui implique que

$$g_{jk} = \frac{(-1)^{j+k}\det(A_{kj})}{\det(A)} = \frac{c_{kj}}{\det(A)} = \left(\frac{1}{\det(A)}C^T\right)_{jk}\,,$$

et démontre la formule. □

Exemple 9.30. La matrice

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

est inversible puisque $\det(A) = -14 \neq 0$. Utilisons la formule pour calculer son inverse. Indiquons en rouge le signe de chaque coefficient, venant du $(-1)^{i+j} = \pm 1$ dans la matrice des cofacteurs.

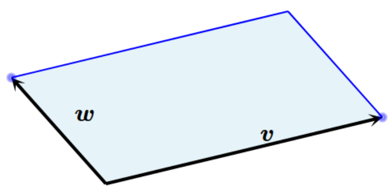
$$\begin{aligned}
A^{-1} &= \frac{1}{\det(A)}C^T \\
&= \frac{1}{-14}\begin{pmatrix} +\det\begin{pmatrix} -1 & 1 \\ 4 & -2 \end{pmatrix} & -\det\begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} & +\det\begin{pmatrix} 1 & -1 \\ 1 & 4 \end{pmatrix} \\ -\det\begin{pmatrix} 1 & 3 \\ 4 & -2 \end{pmatrix} & +\det\begin{pmatrix} 2 & 3 \\ 1 & -2 \end{pmatrix} & -\det\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \\ +\det\begin{pmatrix} 1 & 3 \\ -1 & 1 \end{pmatrix} & -\det\begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} & +\det\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \end{pmatrix}^T \\
&= \frac{1}{-14}\begin{pmatrix} -2 & 3 & 5 \\ 14 & -7 & -7 \\ 4 & 1 & -3 \end{pmatrix}^T \\
&= \frac{1}{-14}\begin{pmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{pmatrix}\,.
\end{aligned}$$

◇

9.6 Interprétation géométrique du déterminant

Le cas 2×2

Dans le plan, considérons deux vecteurs $\mathbf{v}$, $\mathbf{w}$, et considérons le parallélogramme qu'ils définissent : (animation en construction)



L'aire de ce parallélogramme, notée Aire$(\mathbf{v}, \mathbf{w})$, est reliée au déterminant de la matrice 2×2 dont les colonnes sont $\mathbf{v}$ et $\mathbf{w}$:

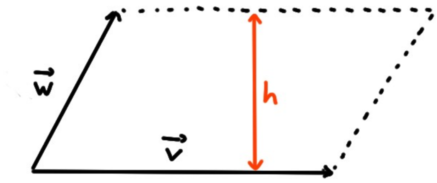
Théorème 9.31. *L'aire du parallélogramme est donnée par*

$$Aire(\mathbf{v}, \mathbf{w}) = |\det([\mathbf{v}\,\mathbf{w}])| = |v_1 w_2 - v_2 w_1| ,$$

où les coordonnées sont relativement à la base canonique.

Preuve: Considérons d'abord le cas simple, où un des vecteurs, disons $\mathbf{v}$, est parallèle $\mathbf{e}_1$. Dans ce cas,

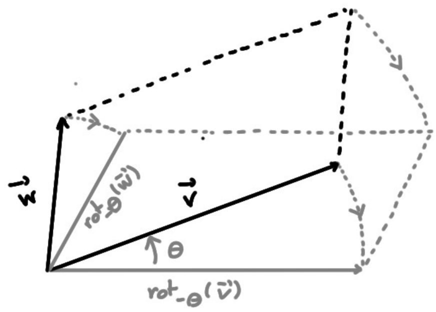
$$\mathbf{v} = \begin{pmatrix} v_1 \\ 0 \end{pmatrix} , \qquad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} .$$



L'aire géométrique vaut donc

$$\begin{aligned} \text{Aire}(\mathbf{v}, \mathbf{w}) = \|\mathbf{v}\| \cdot h = |v_1| \cdot |w_2| &= |v_1 w_2| \\ &= |v_1 w_2 - 0 w_1| \\ &= |\det[\mathbf{v}\,\mathbf{w}]| \ . \end{aligned}$$

Dans le cas général, considérons la rotation $\text{rot}_{-\theta}$, où θ est l'angle orienté entre $\mathbf{e}_1$ et $\mathbf{v}$. On a alors que $\text{rot}_{-\theta}(\mathbf{v})$ est parallèle à $\mathbf{e}_1$:



De plus, puisque l'aire du parallélogramme ne change pas si ses deux côtés subissent la même rotation :

$$\text{Aire}(\mathbf{v}, \mathbf{w}) = \text{Aire}(\text{rot}_{-\theta}(\mathbf{v}), \text{rot}_{-\theta}(\mathbf{w})) \, .$$

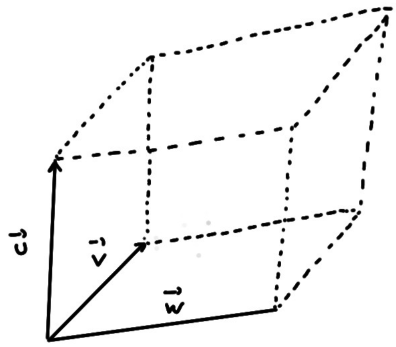
Maintenant, le parallélogramme tourné de $-\theta$ a un de ses côtés parallèle à $\mathbf{e}_1$, et donc on peut calculer son aire avec la formule vue ci-dessus :

$$\begin{aligned} \text{Aire}(\text{rot}_{-\theta}(\mathbf{v}), \text{rot}_{\theta}(\mathbf{w})) &= |\det[\text{rot}_{-\theta}(\mathbf{v})\, \text{rot}_{\theta}(\mathbf{w})]| \\ &= |\det([\text{rot}_{-\theta}]_{\mathcal{B}_{\text{can}}} [\mathbf{v}\,\mathbf{w}])| \\ &= \Big|\underbrace{\det([\text{rot}_{-\theta}]_{\mathcal{B}_{\text{can}}})}_{1} \det([\mathbf{v}\,\mathbf{w}])\Big| \\ &= |\det([\mathbf{v}\,\mathbf{w}])| \ . \end{aligned}$$

□

Le cas 3×3

Dans l'espace, considérons trois vecteurs $\mathbf{u}, \mathbf{v}, \mathbf{w}$, et le parallélépipède qu'ils définissent :



Le volume de ce parallélogramme, noté $\mathrm{Vol}(\mathbf{u}, \mathbf{v}, \mathbf{w})$, est reliée au déterminant de la matrice 3×3 dont les colonnes sont $\mathbf{u}$,$\mathbf{v}$ et $\mathbf{w}$:

Théorème 9.32. *Le volume du parallépipède est donnée par*

$$Vol(\mathbf{u}, \mathbf{v}, \mathbf{w}) = |\det[\mathbf{u}\,\mathbf{v}\,\mathbf{w}]|$$

où les coordonnées sont relativement à la base canonique.

Chapitre 10

Vecteurs et valeurs propres

10.1 Motivation

Les notions introduites jusqu'ici permettent de dire des choses très globales sur une application linéaire

$$T : V \to V'$$

Si V et V' sont de dimensions finies, une telle application peut être représentée par une matrice, et nous savons l'utiliser pour étudier l'injectivité, la surjectivité ; nous avons plusieurs critères permettant de déterminer quand l'application est bijective (via l'inversibilité de sa matrice et le déterminant).

Mais ce que nous n'avons pas encore c'est un outil, un peu comme la dérivée en analyse, qui nous permette de dire des choses plus fines sur cette application.

Nous nous concentrerons sur les applications linéaires

$$T : \mathbb{R}^n \to \mathbb{R}^m .$$

Pour motiver les nouvelles notions que nous allons introduire, voyons un exemple simple dans le plan :

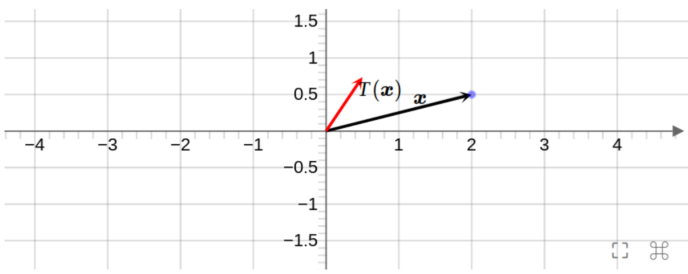
Exemple 10.1. Considérons l'application $T : \mathbb{R}^2 \to \mathbb{R}^2$ définie par

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x} \mapsto T(\mathbf{x}) := \begin{pmatrix} x_2 \\ \frac{1}{2}x_1 - \frac{1}{2}x_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}}_{=A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Les colonnes de A étant indépendantes, cette application est bijective.

Mais ne peut-on rien dire de plus ? Par exemple, peut-on dire plus précisément comment $A\mathbf{x}$ est relié *géométriquement* à $\mathbf{x}$?

Pour essayer de mieux comprendre cette application, faisons varier $\mathbf{x}$ sur l'animation ci-dessous, et observons comment l'image $A\mathbf{x}$ se comporte :



On se rend compte que certaines directions semblent jouer un rôle particulier. Sous l'action de T, c'est-à-dire lorsqu'on multiplie par A,

- ⋆ tout vecteur $\mathbf{x}$ sur la droite dirigée par $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ subit uniquement une modification de longueur, par un facteur $\frac{1}{2}$,
$$A\mathbf{x} = \tfrac{1}{2}\mathbf{x}\,.$$
- ⋆ tout vecteur $\mathbf{x}$ sur la droite dirigée par $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ subit uniquement une modification de sens :
$$A\mathbf{x} = -\mathbf{x}\,.$$

En d'autres termes, les deux directions spécifiées par $\mathbf{v}_1$ et $\mathbf{v}_2$ sont particulières puisqu'elles définissent des vecteurs *dont la direction ne change pas sous l'action de T*. Leur longueur et leur sens, par contre, peuvent être altérés.

Ces vecteurs particuliers $\mathbf{v}_1$ et $\mathbf{v}_2$, que nous appellerons *vecteurs propres*, fournissent un point de départ pour comprendre la géométrie de l'application T. Au chapitre suivant, sur la *diagonalisation*, nous utiliserons ces vecteurs propres pour construire une nouvelle base dans laquelle nous exprimerons T. ◇

10.2 Définition, espace propre

En général, lorsqu'on multiplie un vecteur $\mathbf{x} \in \mathbb{R}^n$ par une matrice A ($n \times n$), on *change* la direction de $\mathbf{x}$.
Or on a vu dans l'exemple de la section précédente qu'il peut exister des vecteurs $\mathbf{v}$ particuliers dont la direction n'est pas modifiée lorsqu'ils sont multipliés par A. En d'autres termes, pour ces vecteurs, $A\mathbf{v}$ est colinéaire à $\mathbf{v}$.

Définition 10.2. Soit A une matrice $n \times n$. Un vecteur $\mathbf{v} \in \mathbb{R}^n$ est appelé **vecteur propre de** A si il existe $\lambda \in \mathbb{R}$ tel que
$$A\mathbf{v} = \lambda\mathbf{v}\,.$$
Le scalaire λ est appelé **valeur propre de** A, et $\mathbf{v}$ est **un vecteur propre associé** à λ.

- ⋆ Le vecteur nul est toujours vecteur propre, puisque $A\mathbf{0} = \lambda\mathbf{0}$ quel que soit λ. On sera donc intéressé, par la suite, par l'existence de vecteurs propres *non-nuls* : $\mathbf{v} \neq \mathbf{0}$.
- ⋆ Puisqu'à toute matrice A ($n \times n$) correspond une application linéaire $T : \mathbb{R}^n \to \mathbb{R}^n$, définie par $T(\mathbf{x}) := A\mathbf{x}$, on définit les **vecteurs propres** (resp. **valeurs propres**) de T comme étant ceux (resp. celles) de A. Donc un vecteur propre $\mathbf{v}$ de T, associé à une valeur propre λ, est tel que
$$T(\mathbf{v}) = \lambda\mathbf{v}\,.$$

Exemple 10.3. Soit $A = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix}$.

- ⋆ Si $\mathbf{v} = \begin{pmatrix} 6 \\ -5 \end{pmatrix}$, alors
$$A\mathbf{v} = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix}\begin{pmatrix} 6 \\ -5 \end{pmatrix} = \begin{pmatrix} -24 \\ 20 \end{pmatrix} = -4\begin{pmatrix} 6 \\ -5 \end{pmatrix} = -4\mathbf{v}\,,$$
et donc $\mathbf{v}$ est vecteur propre, avec valeur propre $\lambda = -4$.

⋆ Si $\mathbf{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, alors

$$A\mathbf{v} = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \end{pmatrix},$$

qui n'est pas colinéaire à $\mathbf{v}$, donc $\mathbf{v}$ n'est pas vecteur propre.

◇

Exemple 10.4. Pour une matrice $n \times n$ diagonale,

$$A = \mathrm{diag}(d_1, \dots, d_n) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{pmatrix},$$

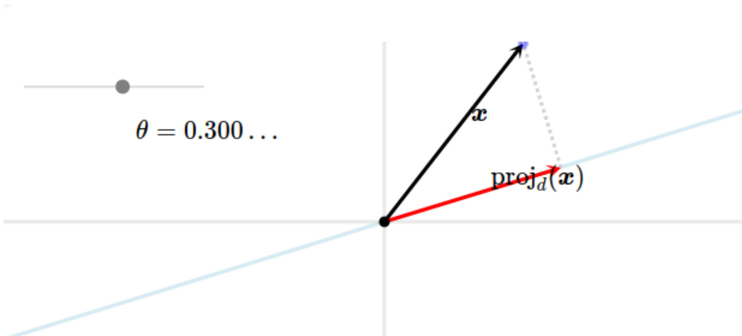
on a

$$A\mathbf{e}_k = d_k \mathbf{e}_k, \qquad \forall k = 1, \dots, n,$$

et donc chaque vecteur de la base canonique $\mathbf{e}_k$ est vecteur propre, avec valeur propre d_k. ◇

Nous verrons bientôt comment calculer les vecteurs et valeurs propres d'une matrice. Mais parfois, lorsque l'application associée a un sens géométrique direct, on peut les connaître sans faire de calculs, par simple observation.

Exemple 10.5. Considérons la projection sur une droite passant par l'origine :



⋆ Nous avons déjà remarqué que les vecteurs sur d ne sont pas modifiés par la projection :

$$\mathrm{proj}_d(\mathbf{v}) = \mathbf{v} = 1\mathbf{v} \qquad \forall \mathbf{v} \in d,$$

Ainsi, tous les vecteurs de d sont vecteurs propres de proj_d, avec valeur propre $\lambda = 1$.

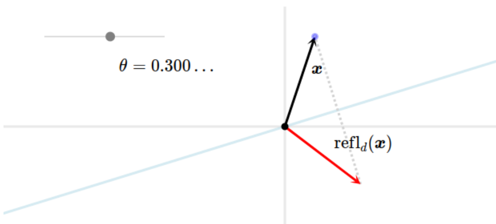
⋆ Les vecteurs qui sont *perpendiculaires* à d ont tous comme projection le vecteur nul :

$$\mathrm{proj}_d(\mathbf{v}) = \mathbf{0} = 0\mathbf{v}, \qquad \forall \mathbf{v} \perp d,$$

Ainsi, tous les vecteurs perpendiculaires à d sont vecteurs propres de proj_d, avec valeur propre $\lambda = 0$.

Il n'y a, apparemment en tout cas, pas d'autres vecteurs propres. ◇

Exemple 10.6. On peut faire de même avec la réflexion par rapport à une droite :



- ⋆ Tout vecteur sur d est vecteur propre, avec valeur propre $\lambda = 1$:

$$\mathrm{refl}_d(\mathbf{v}) = \mathbf{v} = 1\mathbf{v} \qquad \forall \mathbf{v} \in d\,,$$

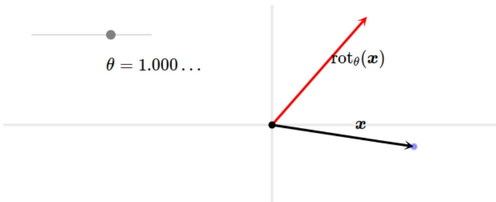
- ⋆ Tout vecteur perpendiculaire à d est vecteur propre, avec valeur propre $\lambda = -1$:

$$\mathrm{refl}_d(\mathbf{v}) = -\mathbf{v} = (-1)\mathbf{v} \qquad \forall \mathbf{v} \perp d\,.$$

◇

Une matrice ne possède pas toujours des vecteurs et valeurs propres. En effet, l'existence de vecteurs $\mathbf{v}$ qui soient colinéaires à leur image $A\mathbf{v}$ est une propriété géométrique particulière que beaucoup de transformations, même naturelles, ne satisfont pas.

Exemple 10.7. Considérons la rotation d'angle θ, $\mathbf{x} \mapsto T(\mathbf{x}) = \mathrm{rot}_\theta(\mathbf{x})$:



Pour les valeurs de $\theta \in [-\pi, \pi]$ qui sont différentes de 0 et $\pm\pi$, $\mathrm{rot}_\theta(\mathbf{x})$ pointe toujours dans une direction différente de $\mathbf{x}$. Donc pour ces valeurs de θ, sa matrice n'a pas de vecteurs propres. Par contre,

- ⋆ Si $\theta = 0$, alors évidemment la rotation ne fait rien,

$$\mathrm{rot}_0(\mathbf{x}) = \mathbf{x}\,, \qquad \forall \mathbf{x} \in \mathbb{R}^2\,,$$

  et donc n'importe quel vecteur du plan est vecteur propre, avec valeur propre $\lambda = 1$.
- ⋆ Si $\theta = \pm\pi$, alors l'effet de la rotation est de renverser $\mathbf{x}$,

$$\mathrm{rot}_{\pm\pi}(\mathbf{x}) = -\mathbf{x}\,, \qquad \forall \mathbf{x} \in \mathbb{R}^2\,,$$

  et donc n'importe quel vecteur du plan est vecteur propre, avec valeur propre $\lambda = -1$.

Nous reviendrons plus tard sur ces cas particuliers. ◇

La question se pose maintenant de savoir comment calculer les vecteurs propres et valeurs propres de façon systématique, pour une matrice donnée.

Espace propre

Par linéarité, si $\mathbf{v}$ est vecteur propre avec valeur propre λ, alors tout vecteur colinéaire à $\mathbf{v}$ est aussi vecteur propre avec valeur propre λ .

Donc dès qu'une matrice possède une valeur propre, il y a une infinité de vecteurs propres qui lui sont associés. Ceci mène à considérer, pour une valeur propre λ donnée, l'ensemble de tous les vecteurs propres associés à λ :

Définition 10.8. Soit A une matrice $n \times n$, et λ une valeur propre de A. L'ensemble

$$E_\lambda := \left\{\mathbf{v} \in \mathbb{R}^n \;:\; A\mathbf{v} = \lambda\mathbf{v}\right\} \subset \mathbb{R}^n$$

est appelé **espace propre** associé à λ.

Remarque 10.9. Comme dit précédemment, E_λ contient toujours $\mathbf{0}$. ◇

Exemple 10.10. Nous avons vu plus haut que $\lambda = -4$ était valeur propre de la matrice

$$A = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix} .$$

Calculons son espace propre associé. Pour ce faire, on cherche tous les $\mathbf{v}$ solutions de

$$A\mathbf{v} = -4\mathbf{v} .$$

Comme on sait, ce système *doit* posséder une infinité de solutions ! En nommant les composantes de $\mathbf{v}$, on peut l'écrire

$$\begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = -4 \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} .$$

En passant le second membre du côté gauche,

$$\begin{pmatrix} 5 & 6 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

L'espace propre associé à $\lambda = -4$ est donc une droite :

$$E_{-4} = \left\{ \mathbf{v} = t \begin{pmatrix} 6 \\ -5 \end{pmatrix} \middle| t \in \mathbb{R} \right\} = \mathrm{Vect} \left\{ \begin{pmatrix} 6 \\ -5 \end{pmatrix} \right\}$$

◇

Exemple 10.11. Considérons l'application associée à

$$A = \begin{pmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{pmatrix} .$$

Supposons que l'on ait déjà montré que $\lambda = 2$ est valeur propre. Calculons son espace propre associé, E_2 : on cherche tous les $\mathbf{v} \in \mathbb{R}^3$ solutions de

$$A\mathbf{v} = 2\mathbf{v} ,$$

c'est-à-dire

$$\begin{pmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 2 \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} .$$

En passant le second membre du côté gauche,

$$\begin{pmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} ,$$

qui est équivalent à

$$\begin{pmatrix} 2 & -1 & 6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} .$$

On peut donc prendre v_2 et v_3 comme variables libres, et prendre $v_1 = \frac{1}{2}(v_2 - 6v_3)$ comme variable de base. Ainsi, tout vecteur de la forme

$$\mathbf{v} = \begin{pmatrix} \frac{1}{2}v_2 - 3v_3 \\ v_2 \\ v_3 \end{pmatrix} = v_2 \begin{pmatrix} 1/2 \\ 1 \\ 0 \end{pmatrix} + v_3 \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix}$$

est vecteur propre de A, avec valeur propre 2. Ceci montre que l'espace propre E_2 est un plan :

$$\begin{aligned} E_2 &= \left\{ \mathbf{v} = s \begin{pmatrix} 1/2 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \middle| s, t \in \mathbb{R} \right\} \\ &= \operatorname{Vect} \left\{ \begin{pmatrix} 1/2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \right\}. \end{aligned}$$

$\diamond$

Remarque 10.12. On l'a observé sur ces deux premiers exemples : une fois la valeur propre connue, la recherche des vecteurs propres qui lui sont associés mène *toujours* à un système possédant une infinité de solutions. $\diamond$

Donc une fois une valeur propre connue, un calcul explicite d'espace propre n'est que la résolution d'un système du type $A\mathbf{v} = \lambda\mathbf{v}$. La question naturelle, à laquelle nous répondrons dans la section suivante, est de savoir *comment trouver les valeurs propres*.

Mais avant ça, remarquons que dans les deux exemples ci-dessus, l'espace propre trouvé était engendré par certains vecteurs, et avait donc une structure de sous-espace vectoriel. C'est l'origine du terme "espace" propre :

Lemme 15. *L'espace propre associé à λ peut s'écrire*

$$E_\lambda = \ker(A - \lambda I_n).$$

Par conséquent, c'est un sous-espace vectoriel de $\mathbb{R}^n$.

Preuve:

$$\begin{aligned} \mathbf{v} \in E_\lambda &\Leftrightarrow A\mathbf{v} = \lambda\mathbf{v} \\ &\Leftrightarrow A\mathbf{v} - \lambda\mathbf{v} = \mathbf{0} \\ &\Leftrightarrow (A - \lambda I_n)\mathbf{v} = \mathbf{0} \\ &\Leftrightarrow \mathbf{v} \in \ker(A - \lambda I_n). \end{aligned}$$

Comme $A - \lambda I_n$ est une application linéaire, nous savons depuis les chapitres précédents que son noyau est un sous-espace vectoriel de $\mathbb{R}^n$ $\square$

Matrices inversibles et la valeur propre nulle

Théorème 10.13. *Une matrice A est inversible si et seulement si $\lambda = 0$ n'est pas valeur propre.*

Preuve: On sait que A est inversible si et seulement si son noyau ne contient que le vecteur nul. Or le noyau pouvant être défini comme l'ensemble des vecteurs $\mathbf{v}$ tels que $A\mathbf{v} = 0\mathbf{v}$, ceci donne l'équivalence. $\square$

On trouvera **ici (3Blue1Brown)** (lien web) une discussion qui pourra aider (ou pas) à comprendre certaines des choses faites ici, et qui motive aussi l'usage que nous ferons plus tard des vecteurs et valeurs propres.

10.3 Le polynôme caractéristique

Voyons le résultat qui rendra la recherche de valeurs propres un problème purement algébrique :

Théorème 10.14. *Soit A une matrice $n \times n$. Alors $\lambda \in \mathbb{R}$ est valeur propre de A si et seulement si*

$$\det(A - \lambda I_n) = 0\,.$$

Preuve: En effet, λ est valeur propre de A si et seulement si il existe un vecteur non-nul $\mathbf{v}$ tel que $A\mathbf{v} = \lambda\mathbf{v}$. On a vu plus haut que ceci est équivalent à dire que $\mathbf{v} \in \ker(A - \lambda I_n)$. Mais l'existence de vecteurs non-nuls dans le noyau d'une matrice (ici : la matrice $A - \lambda I_n$) implique que celle-ci n'est pas inversible, ce qui est équivalent à dire que son déterminant est nul. □

Exemple 10.15. Considérons encore une fois la matrice

$$A = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix}\,.$$

Par le théorème, toutes les valeurs propres se trouvent en étudiant l'équation

$$\det(A - \lambda I_2) = 0\,.$$

Comme

$$\begin{aligned} \det(A - \lambda I_2) &= \det \begin{pmatrix} 1-\lambda & 6 \\ 5 & 2-\lambda \end{pmatrix} \\ &= (1-\lambda)(2-\lambda) - 30 \\ &= \lambda^2 - 3\lambda - 28 \\ &= (\lambda + 4)(\lambda - 7)\,, \end{aligned}$$

A possède exactement deux valeurs propres : $\lambda_1 = -4$ (comme nous avions déjà observé) et $\lambda_2 = 7$. ◇

Comme dans ce dernier exemple, la fonction $\lambda \mapsto \det(A - \lambda I_n)$ sera toujours un polynôme en λ.

Définition 10.16. Soit A une matrice $n \times n$. Le polynôme

$$P_A(\lambda) := \det(A - \lambda I_n)$$

est appelé le **polynôme caractéristique de** A.

Les valeurs propres d'une matrice se trouvent donc en cherchant les racines de son polynôme caractéristique.

Exemple 10.17. Soit $A = \begin{pmatrix} 1 & -5 \\ 1 & 1 \end{pmatrix}$. On a

$$P_A(\lambda) = \det \begin{pmatrix} 1-\lambda & -5 \\ 1 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 + 5\,.$$

Comme $P_A(\lambda) \geqslant 5$ pour tout λ, P_A n'a pas de racines. Donc A ne possède aucune valeur propre, et aucun vecteur propre. ◇

Exemple 10.18. Pour une matrice $n \times n$ diagonale, $A = \operatorname{diag}(d_1, \dots, d_n)$, on a

$$\begin{aligned} P_A(\lambda) &= \det(A - \lambda I_n) \\ &= \det(\operatorname{diag}(d_1 - \lambda, \dots, d_n - \lambda)) \\ &= (d_1 - \lambda) \cdots (d_n - \lambda)\,, \end{aligned}$$

donc les valeurs propres de A sont ses éléments diagonaux $d_1, \dots, d_n$. ◇

Recherche de vecteurs et valeurs propres

Pour trouver les vecteurs propres et valeurs propres (s'il y en a) d'une matrice A, on pourra donc procéder comme suit :

1) Calculer le **spectre** de A, noté $\mathrm{spectre}(A)$, et défini comme l'ensemble de toutes ses valeurs propres, racines du polynôme caractéristique, $P_A(\lambda) = 0$.
2) Si $\mathrm{spectre}(A) \neq \emptyset$, calculer pour chaque valeur propre $\lambda \in \mathrm{spectre}(A)$ l'espace propre associé E_λ, en trouvant toutes les solutions de $A\mathbf{v} = \lambda\mathbf{v}$.

Informel 10.19. A priori, si A est $n \times n$, $P_A(\lambda)$ est un polynôme de degré n. Donc puisqu'on cherche les racines de $P_A(\lambda)$, on a avantage à le calculer avec précaution, de façon à tout de suite l'obtenir sous une forme factorisée (ou aussi factorisée que possible). Dans l'exemple suivant, un choix judicieux d'opérations sur la matrice $A - \lambda I_3$ évite de devoir étudier un polynôme de degré 3.

Exemple 10.20. Cherchons les vecteurs et valeurs propres de

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

Commençons par la recherche des valeurs propres, en calculant

$$\begin{aligned} P_A(\lambda) &= \det \begin{pmatrix} 1-\lambda & -1 & -1 \\ -1 & 1-\lambda & -1 \\ -1 & -1 & 1-\lambda \end{pmatrix} \\ &= \det \begin{pmatrix} -1-\lambda & -1 & -1 \\ -1-\lambda & 1-\lambda & -1 \\ -1-\lambda & -1 & 1-\lambda \end{pmatrix} \\ &= -(1+\lambda) \det \begin{pmatrix} 1 & -1 & -1 \\ 1 & 1-\lambda & -1 \\ 1 & -1 & 1-\lambda \end{pmatrix} \\ &= -(1+\lambda) \det \begin{pmatrix} 1 & -1 & -1 \\ 0 & 2-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{pmatrix} \\ &= -(1+\lambda)(2-\lambda)^2 . \end{aligned}$$

(Dans la deuxième ligne nous avons rajouté les colonnes 2 et 3 à la première, dans la troisième nous avons extrait un $(1+\lambda)$ de la première colonne, et dans la quatrième nous avons soustrait la première de la deuxième et troisième ligne. Dans la dernière ligne, nous avons profité du fait que la matrice était triangulaire.)

Nous avons donc deux valeurs propres, $\lambda_1 = -1$ et $\lambda_2 = 2$. On calcule facilement leurs espaces propres associés :

$$E_{-1} = \ker(A + I_n) = \mathrm{Vect}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

$$E_2 = \ker(A - 2I_n) = \mathrm{Vect}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\} .$$

◇

Le polynôme caractéristique est un invariant de similitude

Rappelons que deux matrices carrées sont **semblables**, $A \sim B$, si il existe une matrice inversible M telle que $A = MBM^{-1}$.

Théorème 10.21. *Si deux matrices sont semblables, $A \sim B$, alors elles ont le même polynôme caractéristique :*

$$P_A(\lambda) = P_B(\lambda) \qquad \forall \lambda \in \mathbb{R}\,.$$

Par conséquent, elles ont le même spectre : $\text{spectre}(A) = \text{spectre}(B)$.

Preuve: Si $A = MBM^{-1}$, alors en récrivant $I_n = MM^{-1} = MI_nM^{-1}$,

$$\begin{aligned}
P_A(\lambda) &= \det(A - \lambda I_n) \\
&= \det(MBM^{-1} - \lambda MM^{-1}) \\
&= \det(M(B - \lambda I_n)M^{-1}) \\
&= \det(M)\det(B - \lambda I_n)\det(M^{-1}) \\
&= \det(B - \lambda I_n)\det(M)\det(M^{-1}) \\
&= \det(B - \lambda I_n)\det(MM^{-1}) \\
&= \det(B - \lambda I_n) \\
&= P_B(\lambda)\,.
\end{aligned}$$

□

Considérons une application linéaire

$$T : \mathbb{R}^n \to \mathbb{R}^n\,.$$

Si on possède deux bases dans $\mathbb{R}^n$, notées $\mathcal{B}$ et $\mathcal{C}$, T peut se représenter comme une matrice,

- ⋆ $[T]_{\mathcal{B}}$ relativement à $\mathcal{B}$, ou
- ⋆ $[T]_{\mathcal{C}}$ relativement à $\mathcal{C}$,

La formule du changement de base nous dit que

$$\begin{aligned}
[T]_{\mathcal{B}} &= P_{\mathcal{BC}}[T]_{\mathcal{C}}P_{\mathcal{CB}} \\
&= P_{\mathcal{CB}}{}^{-1}[T]_{\mathcal{C}}P_{\mathcal{CB}}\,.
\end{aligned}$$

Ceci implique que $[T]_{\mathcal{B}} \sim [T]_{\mathcal{C}}$, et donc, par le théorème ci-dessus, que ces deux matrices ont le même spectre.

Ceci montre que le spectre est bel et bien associé à l'*application*, pas à la matrice qui est utilisée pour la représenter relativement à une base ou une autre.

10.4 Vecteurs propres associés à des valeurs propres distinctes

Théorème 10.22. *Soient $\lambda_1, \dots, \lambda_k$ des valeurs propres distinctes ($\lambda_i \neq \lambda_j$ si $i \neq j$) d'une matrice A, et soient $\mathbf{v}_1, \dots, \mathbf{v}_k$ des vecteurs non-nuls tels que*

- ⋆ *$\mathbf{v}_1$ est vecteur propre associé à λ_1,*
- ⋆ $\vdots$
- ⋆ *$\mathbf{v}_k$ est vecteur propre associé à λ_k.*

Alors la famille $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ est libre.

Preuve: On démontre le résultat par récurrence sur k, c'est-à-dire sur le nombre de vecteurs propres dans la famille.

Cas $k = 2$: Soient $\lambda_1 \neq \lambda_2$ deux valeurs propres de A, et soient $\mathbf{v}_1, \mathbf{v}_2$ deux vecteurs non nuls tels que

$$A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1 , \qquad A\mathbf{v}_2 = \lambda_2 \mathbf{v}_2 .$$

Considérons la relation linéaire

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 = \mathbf{0} .$$

Agissons de deux manières sur cette relation :

- ⋆ En multipliant par A des deux côtés,

$$\alpha_1 \underbrace{A\mathbf{v}_1}_{=\lambda_1\mathbf{v}_1} + \alpha_2 \underbrace{A\mathbf{v}_2}_{=\lambda_2\mathbf{v}_2} = A\mathbf{0} = \mathbf{0} ,$$

  on a

$$\alpha_1\lambda_1\mathbf{v}_1 + \alpha_2\lambda_2\mathbf{v}_2 = \mathbf{0} .$$

- ⋆ En multipliant par λ_2 des deux côtés,

$$\alpha_1\lambda_2\mathbf{v}_1 + \alpha_2\lambda_2\mathbf{v}_2 = \mathbf{0} .$$

En faisant la différence de ces deux expressions, on obtient

$$\alpha_1 \underbrace{(\lambda_1 - \lambda_2)}_{\neq 0} \underbrace{\mathbf{v}_1}_{\neq \mathbf{0}} = \mathbf{0} ,$$

d'où on tire que $\alpha_1 = 0$. En refaisant la même chose mais en multipliant par λ_1 au lieu de λ_2, on montre que $\alpha_2 = 0$. Donc $\{\mathbf{v}_1, \mathbf{v}_2\}$ est libre.

Cas $k > 2$: Supposons que le résultat est vrai pour des famille de k valeurs propres et k vecteurs propres, et considérons une famille qui en contient $k + 1$:

- ⋆ $\mathbf{v}_1$ est vecteur propre associé à λ_1,
- ⋆ $\vdots$
- ⋆ $\mathbf{v}_{k+1}$ est vecteur propre associé à λ_{k+1},

où tous les λ_j sont distincts, et tous les $\mathbf{v}_j$ sont non-nuls.

Considérons la relation

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_{k+1}\mathbf{v}_{k+1} = \mathbf{0} .$$

Comme avant, agissons de deux manière sur cette relation :

- ⋆ En multipliant par A des deux côtés,

$$\alpha_1\lambda_1\mathbf{v}_1 + \cdots + \alpha_{k+1}\lambda_{k+1}\mathbf{v}_{k+1} = \mathbf{0} .$$

- ⋆ En multipliant par λ_{k+1} des deux côtés,

$$\alpha_1\lambda_{k+1}\mathbf{v}_1 + \cdots + \alpha_{k+1}\lambda_{k+1}\mathbf{v}_{k+1} = \mathbf{0} .$$

En faisant la différence de ces deux expressions, le terme $\alpha_{k+1}\lambda_{k+1}\mathbf{v}_{k+1}$ disparaît, et il reste

$$\alpha_1(\lambda_1 - \lambda_{k+1})\mathbf{v}_1 + \cdots + \alpha_k(\lambda_k - \lambda_{k+1})\mathbf{v}_k = \mathbf{0} ,$$

et comme l'hypothèse d'induction garantit que $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ est libre, tous les coefficients de cette combinaison linéaire sont nuls :

$$\alpha_1(\lambda_1 - \lambda_{k+1}) = 0 , \qquad \cdots , \qquad \alpha_k(\lambda_k - \lambda_{k+1}) = 0 .$$

Mais λ_{k+1} est distinct de toutes les autres valeurs propres : $\lambda_j - \lambda_{k+1} \neq 0$. De là, on tire que $\alpha_1 = 0, \dots, \alpha_k = 0$. En réinjectant ces zéros dans la relation de départ, elle devient

$$\alpha_{k+1}\mathbf{v}_{k+1} = \mathbf{0} .$$

Comme $\mathbf{v}_{k+1} \neq \mathbf{0}$, on conclut que α_{k+1} est nul comme les autres, et donc que la famille $\{\mathbf{v}_1, \cdots, \mathbf{v}_{k+1}\}$ est libre. □

On peut effectivement remarquer que dans les quelques exemples vus précédemment, des familles de vecteurs propres associés à des valeurs propres distinctes étaient toujours libres.

Exemple 10.23. On a vu que la matrice

$$A = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix} .$$

possède exactement deux valeurs propres, $\lambda_1 = -4$ et $\lambda_2 = 7$. Les espaces propres associés sont

$$E_{-4} = \mathrm{Vect}\left\{ \begin{pmatrix} 6 \\ -5 \end{pmatrix} \right\} , \qquad E_7 = \mathrm{Vect}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} .$$

Or si $\mathbf{v}_1 \in E_{-4}$ et $\mathbf{v}_2 \in E_7$ sont tous deux non-nuls, alors $\{\mathbf{v}_1, \mathbf{v}_2\}$ est toujours libre. ◇

10.5 Multiplicités algébriques et géométriques

L'utilisation des valeurs et vecteurs propres, dans l'étude d'une application linéaire, sera

Définition 10.24. Soit λ_k une valeur propre d'une matrice A. La **multiplicité algébrique de** λ_k est le plus grand entier n tel que $(\lambda - \lambda_k)^n$ divise $P_A(\lambda)$; on note cet entier $\mathrm{mult}_a(\lambda_k)$.

En d'autres termes, si la factorisation complète du polynôme caractéristique contient

$$P_A(\lambda) = \cdots (\lambda - \lambda_k)^n \cdots ,$$

alors $\mathrm{mult}_a(\lambda_k) = n$.

Remarque 10.25. On sait par le théorème fondamental de l'algèbre que $P_A(\lambda)$ possède au plus n racines réelles. Ceci signifie que si A possède les valeurs propres $\lambda_1, \dots, \lambda_k$, alors

$$\sum_{j=1}^{k} \mathrm{mult}_a(\lambda_j) \leqslant n .$$

◇

Exemple 10.26. Pour notre matrice

$$A = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix} ,$$

nous avions trouvé

$$P_A(\lambda) = (\lambda + 4)^1 (\lambda - 7)^1 ,$$

qui donne $\mathrm{mult}_a(-4) = 1$, $\mathrm{mult}_a(7) = 1$. ◇

Exemple 10.27. Le polynôme caractéristique de la matrice identité I_n étant

$$P_{I_n}(\lambda) = (1 - \lambda)^n ,$$

l'unique valeur propre $\lambda_1 = 1$ est de multiplicité algébrique $\mathrm{mult}_a(1) = n$. ◇

Définition 10.28. Soit λ_k une valeur propre d'une matrice A. La **multiplicité géométrique de** λ_k est la dimension de son espace propre :

$$\mathrm{mult}_g(\lambda_k) := \dim(E_\lambda) = \dim(\ker(A - \lambda_k I_n)) .$$

Remarque 10.29. Par définition, une multiplicité géométrique est toujours $\geqslant 1$. ◇

Théorème 10.30. *Soit λ_k une valeur propre de A. Alors*

$$\mathrm{mult}_g(\lambda_k) \leqslant \mathrm{mult}_a(\lambda_k) .$$

Preuve: Considérons une valeur propre de A, qu'on notera λ_0 pour simplifier, et son espace propre son espace propre associé, E_{λ_0}. Posons

$$k := \operatorname{mult}_g(\lambda_0) = \dim(E_{\lambda_0}),$$

et considérons des vecteurs propres $\mathbf{v}_1, \dots, \mathbf{v}_k$ formant une base de E_{λ_0}. Complétons cette famille en une base de $\mathbb{R}^n$:

$$\mathcal{B} = (\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_{k+1}, \dots, \mathbf{w}_n)$$

Soit A' la matrice de l'application linéaire $T(\mathbf{x}) = A\mathbf{x}$ relative à la base $\mathcal{B}$:

$$A' = \left[[T(\mathbf{v}_1)]_{\mathcal{B}} \cdots [T(\mathbf{v}_k)]_{\mathcal{B}} [T(\mathbf{w}_{k+1})]_{\mathcal{B}} \cdots [T(\mathbf{w}_n)]_{\mathcal{B}}\right]$$

Puisque chaque $\mathbf{v}_j 1$ est vecteur propre de T, $T(\mathbf{v}_j) = \lambda_0 \mathbf{v}_j$, A' a la structure suivante :

$$A' = \left(\begin{array}{ccc|c} \lambda_0 & & & \\ & \ddots & & B \\ & & \lambda_0 & \\ \hline & 0 & & C \end{array}\right)$$

Maintenant, rappelons que A et A' sont semblables, et possèdent donc le même polynôme caractéristique :

$$P_A(\lambda) = P_{A'}(\lambda).$$

Mais par la structure de A' donnée ci-dessus,

$$\begin{aligned} P_{A'}(\lambda) &= \det(A' - \lambda I_n) \\ &= \det \left(\begin{array}{ccc|c} \lambda_0 - \lambda & & & \\ & \ddots & & B \\ & & \lambda_0 - \lambda & \\ \hline & 0 & & C - \lambda I_{n-k} \end{array}\right) \\ &= (\lambda_0 - \lambda)^k \det(C - \lambda I_{n-k}). \end{aligned}$$

□

Exemple 10.31. Reprenons la matrice vue plus haut :

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

Nous avions calculé

$$P_A(\lambda) = -(1+\lambda)^1 (2-\lambda)^2,$$

Nous avons donc deux valeurs propres,

- ⋆ $\lambda_1 = -1$, de multiplicité algébrique $\operatorname{mult}_a(\lambda_1) = 1$,
- ⋆ $\lambda_2 = 2$, de multiplicité algébrique $\operatorname{mult}_a(\lambda_2) = 2$,

En ce qui concerne les espaces propres,

$$E_{-1} = \ker(A + I_n) = \operatorname{Vect}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\},$$

qui implique $\mathrm{mult}_g(\lambda_1) = 1$, et

$$E_2 = \ker(A - 2I_n) = \mathrm{Vect}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\},$$

qui implique $\mathrm{mult}_g(\lambda_2) = 2$. Donc dans cet exemple,

$$\mathrm{mult}_a(\lambda_1) = \mathrm{mult}_g(\lambda_1),$$
$$\mathrm{mult}_a(\lambda_2) = \mathrm{mult}_g(\lambda_2).$$

◇

Exemple 10.32. Considérons la matrice

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

D'une part, son polynôme caractéristique est donné par

$$P_B(\lambda) = (3 - \lambda)^2,$$

et donc B ne possède qu'une valeur propre $\lambda_1 = 3$, de multiplicité algébrique $\mathrm{mult}_a(\lambda_1) = 2$. Mais on a d'autre part que

$$E_1 = \ker(A - 3I_2) = \mathrm{Vect}\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\},$$

qui implique $\mathrm{mult}_g(\lambda_1) = 1$. Donc dans ce cas,

$$\mathrm{mult}_g(\lambda_1) < \mathrm{mult}_a(\lambda_1).$$

◇

Chapitre 11

Diagonalisation

11.1 Motivation, définition

Nous l'avions dit au début du chapitre sur les vecteurs et valeurs propres, notre but était de un outil permettant d'étudier une application linéaire de façon plus géométrique.

Un idéal : les matrices diagonales

Commençons par décrire les applications qui, même en grande dimensions, sont très simples à comprendre : les applications dont la matrice (relativement à la base canonique) est diagonale. En effet, considérons une application $T : \mathbb{R}^n \to \mathbb{R}^n$ définie par

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto T(\mathbf{x}) := \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & d_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} d_1x_1 \\ d_2x_2 \\ \vdots \\ d_nx_n \end{pmatrix}$$

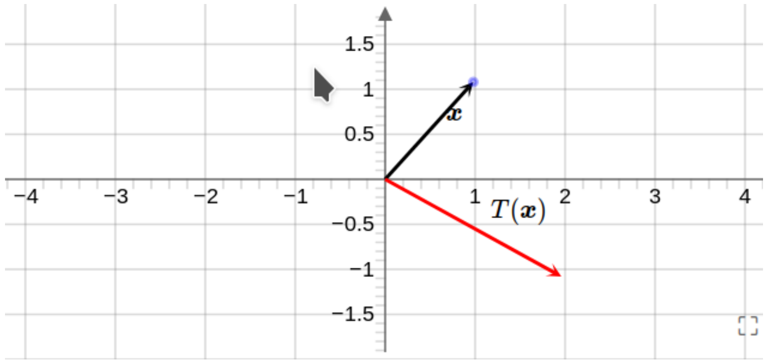
Une telle application se comprend simplement dans le sens suivant : chaque variable x_k n'est que multipliée par d_k, et n'interfère pas avec les autres variables.

Informel 11.1. Donc une application linéaire dont la matrice dans une base est diagonale correspond dans cette base à faire, indépendamment pour chaque k, une simple "dilatation" ("stretching") par un facteur d_k selon la composante k.

Exemple 11.2. Dans le plan, considérons

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto T(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ -x_2 \end{pmatrix} .$$

Dans ce cas, l'effet de T simple à décrire : elle multiplie x_1 par 2, et change le signe de x_2. Ceci permet de construire l'image d'un $\mathbf{x}$ quelconque à la règle et au compas :



◇

Objectif

On sait que la représentation d'une application linéaire relativement à la base canonique n'est qu'une représentation parmi d'autres. Au vu de la discussion ci-dessus, on peut donc se poser la question de savoir si, *pour une application linéaire donnée, il existe une base dans laquelle sa matrice est diagonale*. Si c'est le cas (parce que ça ne sera pas toujours possible), alors on a tout avantage à choisir cette base pour travailler, puisque dans cette base l'application ne devient qu'une modification multiplicative de chacune des composantes, indépendamment des autres. Le but de la **diagonalisation**, que nous présentons dans ce chapitre, est de déterminer si une application donnée peut (ou ne peut pas) être rendue diagonale dans une base bien choisie.

Puisque la diagonalisation a pour but de réduire une application à une base dans laquelle elle "multiplie simplement les composantes par des nombres", c'est sans surprise que les notions de vecteur propre et valeur propre joueront un rôle central dans son développement.

Avant de passer à l'étude générale de la diagonalisation, voyons comment elle s'implémente dans un cas simple.

Diagonaliser une application dans le plan

Exemple 11.3. Reprenons l'application utilisée comme motivation de la notion de vecteur propre, dans le chapitre précédent :

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto T(\mathbf{x}) := \begin{pmatrix} x_2 \\ \frac{1}{2}x_1 - \frac{1}{2}x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Relativement à la base canonique, on a donc

$$[T]_{\mathcal{B}_{\text{can}}} = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} .$$

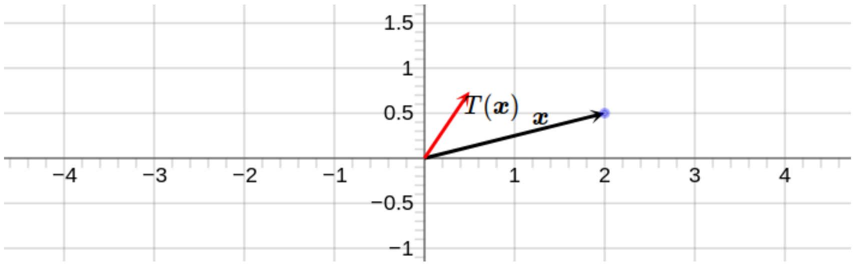
Nous avions remarqué que certains vecteurs subissaient, sous l'action de T, une simple multiplication par un scalaire. Maintenant que l'on sait que ces vecteurs sont les vecteurs propres, on peut les calculer explicitement. Puisque

$$P_{[T]_{\mathcal{B}_{\text{can}}}}(\lambda) = 2\lambda^2 + \lambda - 1 ,$$

on a deux valeurs propres :

- ⋆ $\lambda_1 = \frac{1}{2}$, avec espace propre associé $E_{\frac{1}{2}} = \text{Vect}\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\}$
- ⋆ $\lambda_2 = -1$, avec espace propre associé $E_{-1} = \text{Vect}\left\{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\}$

On peut représenter ces espaces propres, et vérifier comment ils sont modifiés sous l'action de T :



Ensuite, choisissons deux vecteurs propres,

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \in E_{\frac{1}{2}}\,, \qquad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \in E_{-1}\,.$$

Ces vecteurs étant indépendants (évident, mais surtout vrai parce qu'ils sont associés à des valeurs propres distinctes !), ils forment une base de $\mathbb{R}^2$: $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2)$.

Exprimons T dans cette base $\mathcal{B}$ formée de vecteurs propres :

$$[T]_{\mathcal{B}} = {P_{\mathcal{B}_{\mathrm{can}}\mathcal{B}}}^{-1}[T]_{\mathcal{B}_{\mathrm{can}}} P_{\mathcal{B}_{\mathrm{can}}\mathcal{B}}$$

Puisque

$$P_{\mathcal{B}_{\mathrm{can}}\mathcal{B}} = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}\,, \qquad {P_{\mathcal{B}_{\mathrm{can}}\mathcal{B}}}^{-1} = \frac{1}{3}\begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}\,,$$

on a

$$\begin{aligned} [T]_{\mathcal{B}} &= \frac{1}{3}\begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix} \\ &= \frac{1}{3}\begin{pmatrix} \frac{3}{2} & 0 \\ 0 & -3 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -1 \end{pmatrix} \end{aligned}$$

On a ainsi *diagonalisé* T ; sur la diagonale de $[T]_{\mathcal{B}}$ apparaissent précisément les valeurs propres.

Maintenant, lorsqu'on est dans la base $\mathcal{B}$, l'effet de T sur un vecteur devient transparent puisque sa matrice est diagonale. En effet, si

$$[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\,,$$

alors

$$[T(\mathbf{x})]_{\mathcal{B}} = [T]_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}x_1 \\ -x_2 \end{pmatrix}$$

Avec cette information, on peut maintenant retourner sur l'animation du dessus, et observer comment effectivement, sous l'action de T, relativement à $\mathcal{B}$, la première composante de $\mathbf{x}$, est multipliée par $\frac{1}{2}$, et la deuxième est multipliée par -1. $\diamond$

Définition générale de la diagonalisabilité

Définition 11.4. Une matrice A est **diagonalisable** si elle est semblable à une matrice diagonale, c'est-à-dire si il existe une matrice diagonale D, et une matrice inversible M telles que

$$A = MDM^{-1}.$$

Remarque 11.5.
- ⋆ La condition peut aussi s'exprimer par "$A = M^{-1}DM$", mais on verra que celle-ci est plus naturelle.
- ⋆ Toute matrice diagonale est diagonalisable.

◇

Maintenant se pose la question : *comment savoir si une matrice est diagonalisable ?*

On s'en doute, cette question est reliée à l'existence de valeurs et de vecteurs propres. Mais ça n'est pas suffisant, comme on verra dans la section suivante.

11.2 Critère de base

Le résultat suivant est une caractérisation de la diagonalisabilité d'une matrice, qui utilise les vecteurs propres de cette matrice :

Théorème 11.6. *Soit A une matrice $n \times n$. Alors A est diagonalisable si et seulement si A possède n vecteurs propres linéairement indépendants.*

De plus, dans le cas où A est diagonalisable, $A = MDM^{-1}$, alors

- ⋆ *D a sur sa diagonale des valeurs propres de A,*
- ⋆ *les colonnes de M sont les n vecteurs propres indépendants de A.*

Preuve: Supposons que A est diagonalisable : il existe donc $D = \mathrm{diag}(d_1, \dots, d_n)$ et $M = [\mathbf{m}_1 \cdots \mathbf{m}_n]$, inversible, telle que $A = MDM^{-1}$. Remarquons alors que puisque M est inversible, ses colonnes sont indépendantes. Ensuite, si on multiplie à droite par M, on obtient

$$AM = MD.$$

Si on exprime D comme

$$D = [d_1\mathbf{e}_1 \cdots d_n\mathbf{e}_n],$$

alors

$$\begin{aligned} MD &= M\big[d_1\mathbf{e}_1 \cdots d_n\mathbf{e}_n\big] \\ &= \big[d_1M\mathbf{e}_1 \cdots d_nM\mathbf{e}_n\big] \\ &= \big[d_1\mathbf{m}_1 \cdots d_n\mathbf{m}_n\big]. \end{aligned}$$

On peut donc exprimer $AM = MD$ comme suit :

$$\big[A\mathbf{m}_1 \cdots A\mathbf{m}_n\big] = \big[d_1\mathbf{m}_1 \cdots d_n\mathbf{m}_n\big],$$

qui implique bien que $A\mathbf{m}_j = d_j\mathbf{m}_j$ pour tout $j = 1, \dots, n$, et donc que A possède n vecteurs propres linéairement indépendants.

Inversément, supposons que A possède n vecteurs propres linéairement indépendants, que l'on peut noter $\mathbf{v}_1, \dots, \mathbf{v}_n$. Nommons leurs valeurs propres respectives $\lambda_1, \dots, \lambda_n$:

$$A\mathbf{v}_j = \lambda_j\mathbf{v}_j, \qquad j = 1, \dots, n.$$

Posons

$$M := [\mathbf{v}_1 \cdots \mathbf{v}_n]\,, \qquad D := \operatorname{diag}(\lambda_1, \dots, \lambda_n)\,.$$

Puisque les $\mathbf{v}_j$ sont indépendants, M est inversible. Calculons :

$$\begin{aligned} AM &= A[\mathbf{v}_1 \cdots \mathbf{v}_n] \\ &= [A\mathbf{v}_1 \cdots A\mathbf{v}_n] \\ &= [\lambda_1 \mathbf{v}_1 \cdots \lambda_n \mathbf{v}_n] \\ &= MD\,. \end{aligned}$$

En multipliant à droite par M^{-1}, on obtient $A = MDM^{-1}$, qui signifie bien que A est diagonalisable. □

Informel 11.7. Donc une matrice est diagonalisable si et seulement si il est possible de construire une base de $\mathbb{R}^n$ composée uniquement de vecteurs propres de cette matrice.

Remarque 11.8. Ce qui n'est pas précisé, dans l'énoncé du théorème ci-dessus, mais que nous avons observé dans la preuve, c'est que l'*ordre* dans lequel les valeurs propres sont rangées sur la diagonale de D doit respecter l'ordre dans lequel les vecteurs propres sont rangés pour former M. On le fera explicitement dans des cas particuliers, plus bas. ◇

Avant de voir quelques exemples, donnons une conséquence directe du théorème :

Corollaire 6. *Si A est $n \times n$ et possède n valeurs propres distinctes, alors elle est diagonalisable.*

Preuve: Si A possède n valeurs propres distinctes, alors elle possède aussi n vecteurs propres. Puisque ces vecteurs propres sont associés à des valeurs propres distinctes, ils sont linéairement indépendants. Par le théorème ci-dessus, ceci implique que A est diagonalisable. □

Exemple 11.9. Soit $A = \begin{pmatrix} 5 & -3 \\ 1 & 5 \end{pmatrix}$. Comme $P_A(\lambda) = (5-\lambda)^2 + 3 \geqslant 3$, A n'a aucune valeur propre, donc aucun vecteur propre. Par conséquent, A n'est pas diagonalisable. ◇

Exemple 11.10. Nous avons aussi vu que $B = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$ possède une seule valeur propre, $\lambda_1 = 3$, mais que

$$E_3 = \ker(A - 3I_2) = \operatorname{Vect}\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}.$$

Ceci implique que B ne possède pas deux vecteurs propres linéairement indépendants, donc B n'est pas diagonalisable. ◇

Informel 11.11. Dans ce dernier exemple, on a une matrice qui possède une infinité de vecteurs propres, mais qui n'est pas diagonalisable parce que ses vecteurs propres ne “remplissent” pas assez $\mathbb{R}^2$ (ils ne permettent pas de former une base).

Exemple 11.12. Étudions la diagonalisabilité de

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

On calcule :

$$\begin{aligned} P_B(\lambda) &= \det \begin{pmatrix} 1-\lambda & 2 & 0 \\ 0 & 3-\lambda & 0 \\ 2 & -4 & 2-\lambda \end{pmatrix} \\ &= (3-\lambda) \det \begin{pmatrix} 1-\lambda & 0 \\ 2 & 2-\lambda \end{pmatrix} \\ &= (1-\lambda)(2-\lambda)(3-\lambda)\,. \end{aligned}$$

Puisque B est 3×3 et possède 3 valeurs propres distinctes, le corollaire ci-dessus implique que B est diagonalisable. Écrivons la diagonalisation explicitement.

D'abord, calculons les espaces propres :

$$E_1 = \mathrm{Vect}\left\{\begin{pmatrix}1\\0\\-2\end{pmatrix}\right\},$$
$$E_2 = \mathrm{Vect}\left\{\begin{pmatrix}0\\0\\1\end{pmatrix}\right\},$$
$$E_3 = \mathrm{Vect}\left\{\begin{pmatrix}1\\1\\-2\end{pmatrix}\right\}.$$

Pour ce faire, il nous faut un vecteur propre pour chaque valeur propre. Choisissons, pour chaque valeur propre, un vecteur propre associé :

- ⋆ Pour $\lambda_1 = 1$, on peut prendre $\mathbf{v}_1 = \begin{pmatrix}1\\0\\-2\end{pmatrix}$.
- ⋆ Pour $\lambda_2 = 2$, on peut prendre $\mathbf{v}_2 = \begin{pmatrix}0\\0\\1\end{pmatrix}$.
- ⋆ Pour $\lambda_3 = 3$, on peut prendre $\mathbf{v}_3 = \begin{pmatrix}1\\1\\-2\end{pmatrix}$.

(Les vecteurs $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ sont automatiquement indépendants, puisqu'ils sont associés à des valeurs propres distinctes.)

Maintenant, pour réaliser la diagonalisation, on place ces valeurs propres sur une diagonale, et les vecteurs propres associés, *dans le même ordre*, dans une matrice de changement de base :

$$D := \begin{pmatrix}1&0&0\\0&2&0\\0&0&3\end{pmatrix}, \qquad M := [\mathbf{v}_1\,\mathbf{v}_2\,\mathbf{v}_3] = \begin{pmatrix}1&0&1\\0&0&1\\-2&1&-2\end{pmatrix},$$

qui donne la diagonalisation $B = MDM^{-1}$.

Mais on pourrait aussi organiser les valeurs propres dans un ordre différent ; la seule condition à respecter est que *le placement des vecteurs propres dans la matrice de changement de base respecte l'ordre choisi pour les valeur propres*. Par exemple,

$$\widetilde{D} := \begin{pmatrix}2&0&0\\0&3&0\\0&0&1\end{pmatrix}, \qquad \widetilde{M} := [\mathbf{v}_2\,\mathbf{v}_3\,\mathbf{v}_1] = \begin{pmatrix}0&1&1\\0&1&0\\1&-2&-2\end{pmatrix},$$

qui donne la diagonalisation $B = \widetilde{M}\widetilde{D}\widetilde{M}^{-1}$. ⋄

Exemple 11.13. Étudions la diagonalisabilité de

$$C = \begin{pmatrix}1&3&3\\-3&-5&-3\\3&3&1\end{pmatrix}.$$

Cette fois,

$$P_C(\lambda) = (1-\lambda)(\lambda+2)^2\,.$$

On n'a que deux valeurs propres (et pas 3), donc les hypothèses du corollaire ne sont pas satisfaites. Pour voir si l'hypothèse du théorème est satisfaites, on doit voir s'il est possible de former une base de $\mathbb{R}^3$ avec des vecteurs propres.

Or l'étude des espaces propres révèle que

- ⋆ Pour $\lambda_1 = 1$, E_1 est engendré par $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$.
- ⋆ Pour $\lambda_2 = -2$, E_{-2} est engendré par $\mathbf{w}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ et $\mathbf{w}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$.

Puisque $\{\mathbf{v}_1, \mathbf{w}_1, \mathbf{w}_2\}$ est libre, donc forme une base de $\mathbb{R}^3$; ainsi, le théorème implique que C est diagonalisable. La diagonalisation peut se faire par exemple avec

$$D := \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \qquad M := [\mathbf{v}_1\ \mathbf{w}_1\ \mathbf{w}_2] = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

qui donne $C = MDM^{-1}$. Bien-sûr, d'autres choix sont possibles. ⋄

11.3 Deuxième critère

Le deuxième critère est essentiellement une conséquence du premier, mais prend une forme dans laquelle on peut déterminer la diagonalisabilité uniquement à partir de la connaissance des multiplicités géométriques des valeurs propres :

Théorème 11.14. *Soit A une matrice $n \times n$. Alors A est diagonalisable si et seulement si son spectre est non-vide, et si*

$$\sum_{\lambda \in \mathrm{spectre}(A)} \mathrm{mult}_g(\lambda) = n\,.$$

De plus, cette dernière égalité est vérifiée si et seulement si

$$\mathrm{mult}_g(\lambda) = \mathrm{mult}_a(\lambda) \qquad \forall \lambda \in \mathrm{spectre}(A).$$

Preuve: Supposons que $\mathrm{spectre}(A) = \{\lambda_1, \dots, \lambda_k\}$. $\Rightarrow$: Supposons que A est diagonalisable. Par le théorème de la section précédente, il existe donc une base de $\mathbb{R}^n$, formée de vecteurs propres de A :

$$\mathcal{B} = (\mathbf{v}_1, \dots, \mathbf{v}_n)\,.$$

Puisque chaque $\mathbf{v}_j$ est vecteur propre, il doit être associé à une des valeurs propres de $\mathrm{spectre}(A)$. Pour $i = 1, \dots, k$, définissons m_i comme étant le nombre de vecteurs de la famille $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ qui sont associés à la valeur propre λ_i. On a donc

$$\sum_{i=1}^k m_i = n\,.$$

Puisque $\mathcal{B}$ est une base, les vecteurs de $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ qui sont associés à une même valeur propre forment une famille libre, donc

$$m_i \leqslant \mathrm{mult}_g(\lambda_i)\,.$$

Mais comme on sait aussi que $\text{mult}_g(\lambda_i) \leqslant \text{mult}_a(\lambda_i)$, on peut écrire

$$n = \sum_{i=1}^{k} m_i \leqslant \sum_{i=1}^{k} \text{mult}_g(\lambda_i) \leqslant \sum_{i=1}^{k} \text{mult}_a(\lambda_i) \leqslant n\,,$$

qui implique

$$\sum_{i=1}^{k} \text{mult}_g(\lambda_i) = \sum_{i=1}^{k} \text{mult}_a(\lambda_i) = n\,.$$

Remarquons aussi que cette dernière implique que

$$\text{mult}_g(\lambda_i) = \text{mult}_a(\lambda_i) \qquad \forall i = 1, \dots, k\,.$$

En effet, si il existe un i tel que

$$\text{mult}_g(\lambda_i) < \text{mult}_a(\lambda_i)\,,$$

alors

$$\sum_{i=1}^{k} \text{mult}_g(\lambda_i) < \sum_{i=1}^{k} \text{mult}_a(\lambda_i)\,.$$

$\Leftarrow$: (Le paragraphe qui suit est un peu lourd en notations, même si l'idée est simple.) Supposons maintenant que cette dernière égalité est vraie. Pour chaque $i = 1, \dots, k$, définissons $g_i := \text{mult}_g(\lambda_i) = \dim(E_{\lambda_i})$, et considérons une base de E_{λ_i}, notée

$$\mathcal{B}_i = (\mathbf{v}_1^{(i)}, \mathbf{v}_2^{(i)}, \dots, \mathbf{v}_{g_i}^{(i)})\,.$$

Montrons que l'union de toutes ces bases,

$$\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$$

qui contient par définition n vecteurs, est libre.
On considère donc la relation linéaire

$$(*) \qquad \lambda_1^{(1)}\mathbf{v}_1^{(1)} + \cdots + \lambda_{g_k}^{(k)}\mathbf{v}_{g_k}^{(k)} = \mathbf{0}\,.$$

Plus précisément,

$$(*) \qquad \sum_{i=1}^{k}\sum_{j=1}^{g_i} \lambda_j^{(i)}\mathbf{v}_j^{(i)} = \mathbf{0}\,.$$

Si on introduit les vecteurs

$$\mathbf{w}_i := \sum_{j=1}^{g_i} \lambda_j^{(i)}\mathbf{v}_j^{(i)}\,,$$

alors $(*)$ devient

$$(*) \qquad \mathbf{w}_1 + \cdots + \mathbf{w}_k = \mathbf{0}$$

Mais chaque $\mathbf{w}_i \in E_{\lambda_i}$, et donc les $\mathbf{w}_1, \dots, \mathbf{w}_k$ sont des vecteurs propres associés à des valeurs propres distinctes, ils forment donc une famille libre. Ceci signifie que si leur somme est nulle, alors ils sont tous nuls :

$$\mathbf{w}_i = \mathbf{0} \qquad \forall i = 1, \dots, k.$$

Mais comme $\mathcal{B}_i = (\mathbf{v}_1^{(i)}, \mathbf{v}_2^{(i)}, \dots, \mathbf{v}_{g_i}^{(i)})$ est une base, ses vecteurs sont indépendants, et donc

$$\mathbf{w}_i = \lambda_1^{(i)}\mathbf{v}_1^{(i)} + \cdots + \lambda_{g_i}^{(i)}\mathbf{v}_{g_i}^{(i)} = \mathbf{0}$$

implique que $\lambda_1^{(i)} = \cdots = \lambda_{g_i}^{(i)} = 0$. Ceci montre que $\mathcal{B}$ est libre ; puisqu'elle contient n vecteurs, c'est une base de $\mathbb{R}^n$. Par conséquent, A est diagonalisable. $\square$

Exemple 11.15. Dans la section précédente, on avait considéré

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

On a vu que cette matrice possède trois valeurs propres, chacune de multiplicité géométrique égale à 1, ce qui implique

$$\sum_{\lambda \in \mathrm{spectre}(A)} \mathrm{mult}_g(\lambda) = 1 + 1 + 1 = 3 .$$

Par le théorème ci-dessus, on en déduit que B est diagonalisable. ◇

Exemple 11.16. Étudions la diagonalisabilité de

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

Puisque

$$P_A(\lambda) = (1-\lambda)(\underbrace{\lambda^2 - 2\lambda + 2}_{\Delta<0!}) ,$$

A ne possède qu'une valeur propre : $\lambda_1 = 1$, avec $\mathrm{mult}_a(1) = 1$. Or comme

$$E_1 = \ker(A - I_n) = \mathrm{Vect}\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} ,$$

on a $\mathrm{mult}_g(1) = 1$. Puisqu'ici $n = 3$, on a

$$\underbrace{\sum_{\lambda \in \mathrm{spectre}(A)} \mathrm{mult}_g(\lambda)}_{=1} < 3 ,$$

le théorème implique que A n'est pas diagonalisable. ◇

NumChap: chap-diagonalisation, Dernière compilation: 2024-08-12 07:19:48+02:00. (Version Web: botafogo.saitis.net)

Chapitre 12

Produit scalaire et orthogonalité

12.1 Norme et distance

Définition 12.1. Si $\mathbf{x} \in \mathbb{R}^n$, et si $x_1, \ldots, x_n$ sont ses composantes relativement à la base canonique, alors sa **norme** est définie par le réel

$$\|\mathbf{x}\| := \sqrt{x_1^2 + \cdots + x_n^2}\,.$$

Proposition 13. *(Propriétés de la norme)*

- ⋆ $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$ *pour tous* $\lambda \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n$
- ⋆ $\|\mathbf{x}\| \geqslant 0$ *pour tout* $\mathbf{x}$*, et* $\|\mathbf{x}\| = 0$ *si et seulement si* $\mathbf{x} = 0$*.*
- ⋆ *(Inégalité triangulaire)* $\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$ *pour tous* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*.*

Preuve: Pour la première propriété,

$$\begin{aligned}
\|\lambda\mathbf{x}\| &= \sqrt{(\lambda x_1)^2 + \cdots + (\lambda x_n)^2} \\
&= \sqrt{\lambda^2 x_1^2 + \cdots + \lambda^2 x_n^2} \\
&= \sqrt{\lambda^2(x_1^2 + \cdots + x_n^2)} \\
&= |\lambda|\sqrt{x_1^2 + \cdots + x_n^2} \\
&= |\lambda|\|\mathbf{x}\|\,.
\end{aligned}$$

Ensuite, $\|\mathbf{x}\| \geqslant 0$ est évidente, et remarquons que $\|\mathbf{x}\| = 0$ si et seulement si $\|\mathbf{x}\|^2 = 0$, qui est équivalente à

$$x_1^2 + \cdots + x_n^2 = 0\,.$$

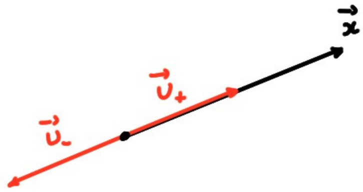
Or une somme de nombres non-négatifs est nulle si et seulement chacun de ces nombres est nul, $x_k^2 = 0$, et donc $x_k = 0$ pour chaque $k = 1, \ldots, n$.

On démontrera l'inégalité triangulaire dans la section suivante. □

Définition 12.2. $\mathbf{x} \in \mathbb{R}^n$ est **unitaire** (ou **normalisé**) si $\|\mathbf{x}\| = 1$.

Remarque 12.3. Pour tout vecteur non-nul $\mathbf{x}$, il existe exactement deux vecteurs unitaires qui sont colinéaires à $\mathbf{x}$, donnés par

$$\mathbf{u}_\pm := \pm\frac{\mathbf{x}}{\|\mathbf{x}\|}\,.$$

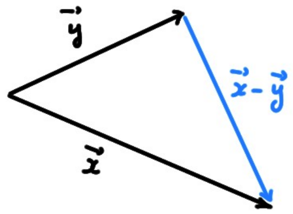


◇

La notion de norme permet de définir encore deux notions géométriques classiques :

Définition 12.4. La **distance** entre $\mathbf{x}$ et $\mathbf{y}$ est définie par

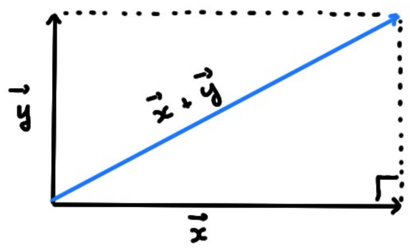
$$\mathrm{dist}(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| .$$



Définition 12.5. Deux vecteurs $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ sont **orthogonaux** (ou **perpendiculaires**) si

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 .$$

Si $\mathbf{x}$ et $\mathbf{y}$ sont orthogonaux, on écrit $\mathbf{x} \perp \mathbf{y}$.



12.2 Définition du produit scalaire

Définition 12.6. Soient $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Le **produit scalaire** de $\mathbf{x}$ et $\mathbf{y}$ est défini par

$$\mathbf{x} \cdot \mathbf{y} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n .$$

Remarque 12.7. Il sera souvent utile de récrire le produit scalaire en le réinterprétant comme un produit matriciel un peu particulier :

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \cdots + x_n y_n = \underbrace{[x_1 \cdots x_n]}_{1\times n} \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{n\times 1} = \mathbf{x}^T \mathbf{y} .$$

◇

Proposition 14. *(Propriétés du produit scalaire)*

1) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
2) $(\lambda \mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (\lambda \mathbf{y}) = \lambda(\mathbf{x} \cdot \mathbf{y})$
3) $\mathbf{x} \cdot (\mathbf{y}_1 + \mathbf{y}_2) = \mathbf{x} \cdot \mathbf{y}_1 + \mathbf{x} \cdot \mathbf{y}_2$
4) $(\mathbf{x}_1 + \mathbf{x}_2) \cdot \mathbf{y} = \mathbf{x}_1 \cdot \mathbf{y} + \mathbf{x}_2 \cdot \mathbf{y}$
5) *(Lien avec la norme)* $\mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2$
6) *(Inégalité de Cauchy-Schwartz)*
$$|\mathbf{x} \cdot \mathbf{y}| \leqslant \|\mathbf{x}\| \|\mathbf{y}\| .$$

Preuve: Les cinq premières propriétés suivent directement de la définition du produit scalaire. Démontrons l'inégalité de Cauchy-Schwartz.

Pour commencer, remarquons que l'inégalité est triviale dès que $\mathbf{y}$ (ou $\mathbf{x}$) est nul. On peut donc supposer que $\mathbf{y} \neq \mathbf{0}$. Ensuite, remarquons que pour tout $t \in \mathbb{R}$,

$$\begin{aligned} 0 \leqslant \|\mathbf{x} - t\mathbf{y}\|^2 &= (\mathbf{x} - t\mathbf{y}) \cdot (\mathbf{x} - t\mathbf{y}) \\ &= \mathbf{x} \cdot \mathbf{x} - 2t(\mathbf{x} \cdot \mathbf{y}) + t^2(\mathbf{y} \cdot \mathbf{y}) \\ &= \|\mathbf{x}\|^2 - 2t(\mathbf{x} \cdot \mathbf{y}) + t^2\|\mathbf{y}\|^2 \end{aligned}$$

Comme cette inégalité est vraie pour tout t, on peut l'utiliser pour la valeur t_* qui minimise le polynôme du deuxième degré défini par
$$t \mapsto t^2\|\mathbf{y}\|^2 - 2t(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{x}\|^2 .$$
Ce dernier est minimal lorsque
$$t_* = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{y}\|^2} .$$
Si on récrit l'inégalité du haut avec t_*, on obtient
$$0 \leqslant \|\mathbf{x}\|^2 - 2\frac{(\mathbf{x} \cdot \mathbf{y})^2}{\|\mathbf{y}\|^2} + \frac{(\mathbf{x} \cdot \mathbf{y})^2}{\|\mathbf{y}\|^2} = \|\mathbf{x}\|^2 - \frac{(\mathbf{x} \cdot \mathbf{y})^2}{\|\mathbf{y}\|^2} ,$$
qui entraîne
$$(\mathbf{x} \cdot \mathbf{y})^2 \leqslant \|\mathbf{x}\|^2\|\mathbf{y}\|^2 .$$
On obtient l'inégalité de Cauchy-Schwartz en prenant la racine carrée des deux côtés. □

Remarque 12.8. $\mathbb{R}^n$, muni du produit scalaire, est un cas particulier de ce que nous appellerons plus tard un **espace Euclidien**. ◇

Informel 12.9. En petites dimensions ($n = 2$ ou 3), le produit scalaire est relié à l'angle θ fait par $\mathbf{x}$ et $\mathbf{y}$, par la relation fondamentale suivante :

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\theta) .$$

Nous ne ferons pas usage de cette relation, puisque nous utiliserons le produit scalaire surtout en grandes dimensions, et qu'on ne sait pas vraiment comment définir l'angle entre deux vecteurs.

On peut utiliser l'inégalité de Cauchy-Schwartz pour démontrer l'inégalité triangulaire de la section précédente :

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) \\ &= \|\mathbf{x}\|^2 + 2(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{y}\|^2 \\ &\leqslant \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 . \end{aligned}$$

Orthogonalité

Le produit scalaire est surtout utilisé, en algèbre linéaire, *pour résoudre des problèmes dans $\mathbb{R}^n$ à l'aide d'arguments géométriques* empruntés à la géométrie du plan et de l'espace. Et la première notion qui joue un rôle en géométrie est celle *d'orthogonalité*.

Commençons par écrire l'égalité du dessus sous une forme un peu différente :

$$(\mathbf{x} \cdot \mathbf{y}) = \frac{1}{2}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\right)\right) .$$

Ceci implique en particulier que $\mathbf{x} \perp \mathbf{y}$ si et seulement si $\mathbf{x}\cdot\mathbf{y} = 0$. On peut donc donner une définition alternative de l'orthogonalité, basée uniquement sur le produit scalaire :

Définition 12.10. Deux vecteurs $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ sont **orthogonaux** (ou **perpendiculaires**) si $\mathbf{x} \cdot \mathbf{y} = 0$.

Exemple 12.11. Dans $\mathbb{R}^5$, les vecteurs

$$\mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ -2 \\ 0 \\ 2 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} 2 \\ 2 \\ 5 \\ 3 \\ 1 \end{pmatrix}$$

sont orthogonaux, puisque $\mathbf{x} \cdot \mathbf{y} = 0$. ◇

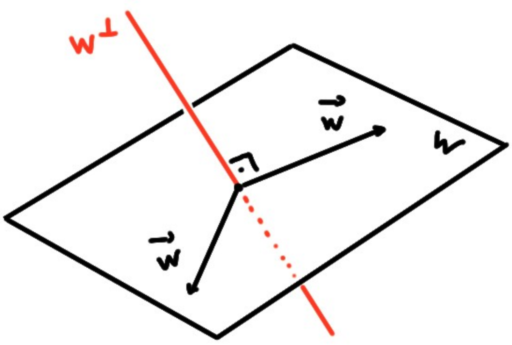
En géométrie, on considère souvent un objet géométrique, généralement une droite ou un plan, défini comme étant perpendiculaire à un autre. En algèbre linéaire, on définit un ensemble de vecteurs qui sont tous orthogonaux aux vecteurs d'un autre ensemble :

Définition 12.12. Soit W un sous-espace vectoriel de $\mathbb{R}^n$. Le **complément orthogonal de** W est l'ensemble

$$W^\perp := \left\{ \mathbf{v} \in \mathbb{R}^n \,\middle|\, \mathbf{v} \perp \mathbf{w} \ \forall \mathbf{w} \in W \right\} .$$

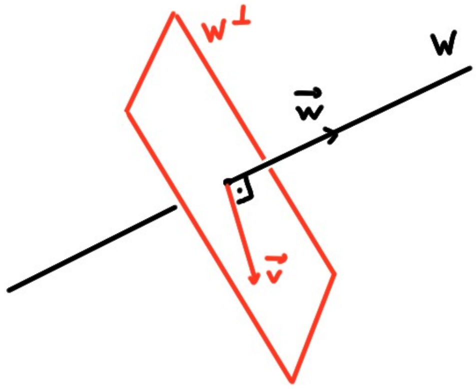
Commençons par comprendre intuitivement le sens de $W^\perp$, en petites dimensions :

Exemple 12.13. Si W est un plan (passant par l'origine) de $\mathbb{R}^3$, alors $W^\perp$ est la droite perpendiculaire à W, passant par l'origine :



(En effet, un vecteur $\mathbf{v}$ quelconque sur la droite est perpendiculaire à tous les vecteurs $\mathbf{w}$ du plan.) ◇

Exemple 12.14. Si W est une droite (passant par l'origine) de $\mathbb{R}^3$, alors $W^\perp$ est le plan perpendiculaire à W, passant par l'origine :



(En effet, un vecteur $\mathbf{v}$ quelconque sur le plan est perpendiculaire à tous les vecteurs $\mathbf{w}$ de la droite.)

◇

Ces deux derniers exemples illustrent bien les propriétés générales ci-dessous :

Proposition 15. *(Propriétés du complément orthogonal)*

1) $W^\perp$ *est un sous-espace vectoriel de* $\mathbb{R}^n$
2) $(W^\perp)^\perp = W$
3) $\dim(W) + \dim(W^\perp) = n$

Preuve: On vérifie que $W^\perp$ est un sous-espace vectoriel de $\mathbb{R}^n$:

⋆ Clairement, le vecteur nul appartient à W puisque $\mathbf{0} \cdot \mathbf{w} = 0$ pour tout $\mathbf{w} \in W$.

⋆ Si $\mathbf{v} \in W^\perp$, alors pour tout scalaire $\lambda \in \mathbb{R}$,

$$(\lambda v) \cdot \mathbf{w} = \lambda(\mathbf{v} \cdot \mathbf{w}) = 0\,,$$

donc $\lambda \mathbf{v} \in W^\perp$.

⋆ Si $\mathbf{v}_1, \mathbf{v}_2 \in W^\perp$, alors pour tout $\mathbf{w} \in W$,

$$(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w} = 0 + 0 = 0\,,$$

donc $\mathbf{v}_1 + \mathbf{v}_2 \in W^\perp$.

Les autes propriétés seront démontrées en exercice. □

Dans la définition, $W^\perp$ est défini comme l'ensemble des vecteurs qui sont orthogonaux à *tous* les vecteurs de W. Ceci implique que d'un point de vue calculatoire, on devrait a priori vérifier une infinité de conditions pour savoir si un vecteur appartient à $W^\perp$. Mais lorsqu'on possède une base les choses sont plus simples :

Lemme 16. *Soit W un sous-espace vectoriel de $\mathbb{R}^n$, et soit $\mathcal{B} = (\mathbf{w}_1, \dots, \mathbf{w}_k)$ une base de W. Alors $\mathbf{v} \in W^\perp$ si et seulement si $\mathbf{v} \perp \mathbf{w}_j$ pour tout $j = 1, \dots, k$.*

Preuve: Supposons que $W = \mathrm{Vect}\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$.

Si $\mathbf{v}$ est orthogonal à tous les vecteurs de W, il est en particulier orthogonal à chacun des éléments de la base.

Inversément, supposons que $\mathbf{v}$ est orthogonal à chacun des éléments de la base. Comme un élément quelconque $\mathbf{w} \in W$ peut se décomposer dans la base, $\mathbf{w} = \alpha_1 \mathbf{w}_1 + \cdots + \alpha_k \mathbf{w}_k$, la linéarité du produit scalaire implique que

$$\begin{aligned} \mathbf{v} \cdot \mathbf{w} &= \mathbf{v} \cdot (\alpha_1 \mathbf{w}_1 + \cdots + \alpha_k \mathbf{w}_k) \\ &= \alpha_1 \underbrace{(\mathbf{v} \cdot \mathbf{w}_1)}_{=0} + \cdots + \alpha_k \underbrace{(\mathbf{v} \cdot \mathbf{w}_k)}_{=0} \\ &= 0\,, \end{aligned}$$

et donc $\mathbf{v} \in W^\perp$. □

Exemple 12.15. Dans $\mathbb{R}^3$, considérons les vecteurs

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix},$$

et considérons le plan

$$W = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}.$$

Le lemme précédent dit que

$$W^\perp = \{\mathbf{v} \in \mathbb{R}^3 \,:\, \mathbf{v} \perp \mathbf{w}_1,\, \mathbf{v} \perp \mathbf{w}_2\}$$

Donc on cherche les vecteurs $\mathbf{v} \in \mathbb{R}^3$ tels que les deux conditions suivantes soient satisfaites simultanément :

$$\begin{cases} \mathbf{v} \cdot \mathbf{w}_1 = 0\,, \\ \mathbf{v} \cdot \mathbf{w}_2 = 0\,. \end{cases}$$

Si $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$, ceci est équivalent à

$$\left\{ \begin{array}{rcrcrcl} v_1 & + & 2v_2 & & & = & 0 \\ -v_1 & + & 3v_2 & + & v_3 & = & 0 \end{array} \right.$$

On peut prendre v_1 comme variable libre, et donc on voit que $W^\perp$ est une droite :

$$W^\perp = \left\{ \mathbf{v} = \begin{pmatrix} x_1 \\ -x_1/2 \\ 5x_1/2 \end{pmatrix} \,\middle|\, x_1 \in \mathbb{R} \right\} = \mathrm{Vect}\left\{ \begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix} \right\}.$$

On vérifie bien dans ce cas que

$$\dim(W) + \dim(W^\perp) = 2 + 1 = 3\,.$$

$\diamond$

Informel 12.16. Dans ce dernier exemple, l'intuition géométrique aurait peut-être suggéré de trouver un vecteur directeur de la droite $W^\perp$ en calculant le **produit vectoriel** de $\mathbf{w}_1$ et $\mathbf{w}_2$. Mais ce produit (que nous ne traiterons pas dans ce cours) n'existe que dans $\mathbb{R}^3$, alors que la méthode que nous avons utilisée fonctionne en toute dimension.

Exemple 12.17. Dans $\mathbb{R}^4$, considérons les vecteurs

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \\ 0 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 2 \end{pmatrix},$$

et considérons le plan

$$W = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}.$$

Puisque $\dim(W) = 2$ et que

$$\dim(W) + \dim(W^\perp) = 4\,,$$

on sait que $\dim(W^\perp) = 2$, et donc $W^\perp$ doit aussi être un plan. Et effectivement, un calcul semblable à celui de l'exemple précédent (voir exercices) montre que

$$W^\perp = \mathrm{Vect}\{\mathbf{p}_1, \mathbf{p}_2\}.$$

où

$$\mathbf{p}_1 = \begin{pmatrix}1\\0\\1\\0\end{pmatrix}, \qquad \mathbf{p}_2 = \begin{pmatrix}2\\-1\\0\\1\end{pmatrix}.$$

◇

12.3 À propos de $\mathrm{Col}(A)$ et $\mathrm{Lgn}(A)$

Rappelons que si A est une matrice $m \times n$,

- ⋆ $\mathrm{Col}(A) \subset \mathbb{R}^m$ est le sous-espace engendré par ses colonnes, et
- ⋆ $\mathrm{Lgn}(A) \subset \mathbb{R}^n$ est le sous-espace engendré par ses lignes.

Théorème 12.18. *Si A est $m \times n$, alors*

1) $\mathrm{Lgn}(A)^\perp = \ker(A)$,

2) $\mathrm{Col}(A)^\perp = \ker(A^T)$.

Preuve: Nous avons déjà vu (**ici** (lien web)) que l'on peut toujours exprimer une matrice $m \times n$ à l'aide de ses lignes :

$$A = \begin{pmatrix}\ell_1^T\\ \vdots \\ \ell_m^T\end{pmatrix},$$

où $\ell_1, \dots, \ell_m \in \mathbb{R}^n$.

1) On a

$$\begin{aligned}\mathbf{v} \in \mathrm{Lgn}(A)^\perp &\iff \mathbf{v}\cdot\ell_j = 0 \quad \forall j = 1,\dots,m\\ &\iff \ell_j^T\mathbf{v} = 0 \quad \forall j = 1,\dots,m\,.\end{aligned}$$

On peut exprimer ces m conditions simultanément en écrivant

$$\begin{pmatrix}\ell_1^T\\ \vdots \\ \ell_m^T\end{pmatrix}\mathbf{v} = \begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix},$$

qui n'est autre que $A\mathbf{v} = \mathbf{0}$.

2) Puisque $\mathrm{Col}(A) = \mathrm{Lgn}(A^T)$, l'affirmation suit de la première partie :

$$\mathrm{Col}(A)^\perp = (\mathrm{Lgn}(A^T))^\perp = \ker(A^T)\,.$$

□

12.4 Familles orthogonales

Définition 12.19. Une famille de vecteurs $\{\mathbf{w}_1, \dots, \mathbf{w}_k\} \subset \mathbb{R}^n$ est dite

- ⋆ **orthogonale** si ses vecteurs sont orthogonaux deux à deux ($\mathbf{w}_i \perp \mathbf{w}_j$ pour tout $i \neq j$),
- ⋆ **orthonormale** (ou **orthonormée**) si elle est orthogonale et si, de plus, tous ses vecteurs sont unitaires ($\|\mathbf{w}_j\| = 1$ pour tout j).

Exemple 12.20. La base canonique de $\mathbb{R}^n$, $\mathcal{B}_{\text{can}} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$, est une famille orthonormale, puisque

$$\mathbf{e}_i \cdot \mathbf{e}_j = \begin{cases} 1 & \text{si } i = j\,, \\ 0 & \text{si } i \neq j\,. \end{cases}$$

◇

Remarque 12.21. Si $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ est orthogonale, et si aucun de ses vecteurs n'est le vecteur nul, alors on la rend orthonormale en divisant chacun de ses vecteurs par sa norme :

$$\left\{ \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}, \dots, \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \right\}$$

◇

Exemple 12.22. Dans $\mathbb{R}^3$, la famille

$$\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix} \right\}$$

est orthogonale, mais pas orthonormale. Comme aucun de ses vecteurs n'est nul, on peut le diviser par sa norme,

$$\left\{ \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \frac{1}{\sqrt{12}} \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix} \right\}$$

pour obtenir une famille orthonormale. ◇

Une propriété importante des familles orthogonales :

Lemme 17. *Si $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ est orthogonale, et si aucun de ses vecteurs n'est nul, alors elle est libre.*

Preuve: Considérons la relation

$$\alpha_1 \mathbf{w}_1 + \dots + \alpha_k \mathbf{w}_k = \mathbf{0}\,.$$

Si on effectue le produit scalaire de cette relation avec $\mathbf{w}_j$,

$$\alpha_1 \underbrace{(\mathbf{w}_j \cdot \mathbf{w}_1)}_{=0} + \dots + \alpha_j \underbrace{(\mathbf{w}_j \cdot \mathbf{w}_j)}_{\neq 0} + \dots + \alpha_k \underbrace{(\mathbf{w}_j \cdot \mathbf{w}_k)}_{=0} = \mathbf{0}\,,$$

qui donne $\alpha_j \|\mathbf{w}_j\|^2 = 0$. Puisque par hypothèse $\mathbf{w}_j \neq \mathbf{0}$, ceci implique $\alpha_j = 0$. Comme ceci vaut pour tout $j = 1, \dots, k$, on a bien montré que la famille est libre. □

Le grand avantage de travailler avec des bases orthogonales :

Théorème 12.23. *Soit W un sous-espace vectoriel de $\mathbb{R}^n$, et soit $\mathcal{B} = (\mathbf{w}_1, \dots, \mathbf{w}_k)$ une base de W, orthogonale. Alors la décomposition d'un $\mathbf{w} \in W$ relativement à $\mathcal{B}$,*

$$\mathbf{w} = \gamma_1 \mathbf{w}_1 + \dots + \gamma_k \mathbf{w}_k\,,$$

a ses coefficients γ_j donnés par

$$\gamma_j = \frac{\mathbf{w} \cdot \mathbf{w}_j}{\mathbf{w}_j \cdot \mathbf{w}_j} = \frac{\mathbf{w} \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2}\,.$$

En particulier, si $\mathcal{B}$ est orthonormale, alors $\gamma_j = \mathbf{w} \cdot \mathbf{w}_j$.

Preuve: Considérons la décomposition

$$\mathbf{w} = \gamma_1 \mathbf{w}_1 + \cdots + \gamma_k \mathbf{w}_k \,.$$

En prenant le produit scalaire de cette expression avec $\mathbf{w}_j$, l'orthogonalité de la base fait qu'il ne survit qu'un seul terme dans le membre de droite :

$$\mathbf{w}_j \cdot \mathbf{w} = \gamma_j (\mathbf{w}_j \cdot \mathbf{w}_j) \,.$$

Ceci démontre l'affirmation. □

Informel 12.24. Rappelons qu'en principe, trouver les composantes d'un vecteur relativement à une base se fait en résolvant un système. Ici, on voit le grand avantage de travailler avec des bases orthogonales : pour avoir une composante, il suffit de calculer un produit scalaire.

Exemple 12.25. Considérons

$$\mathcal{B} = \left(\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix} \right) .$$

On a vu plus haut que cette famille est orthogonale, et donc libre puisqu'aucun de ses vecteurs n'est nul, ce qui en fait une base de $\mathbb{R}^3$. Si on prend un vecteur quelconque de $\mathbb{R}^n$, par exemple

$$\mathbf{v} = \begin{pmatrix} 7 \\ -5 \\ 3 \end{pmatrix} ,$$

on calcule ses composantes relativement à $\mathcal{B}$:

$$\begin{aligned} \gamma_1 &= \frac{\mathbf{v} \cdot \mathbf{b}_1}{\|\mathbf{b}_1\|^2} = \frac{0}{6} = 0 \,, \\ \gamma_2 &= \frac{\mathbf{v} \cdot \mathbf{b}_2}{\|\mathbf{b}_2\|^2} = \frac{4}{2} = 2 \,, \\ \gamma_3 &= \frac{\mathbf{v} \cdot \mathbf{b}_3}{\|\mathbf{b}_3\|^2} = \frac{-30}{12} = \frac{-5}{2} \,, \end{aligned}$$

ce qui donne

$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 2 \\ -5/2 \end{pmatrix} ,$$

c'est-à-dire

$$\mathbf{v} = 0 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \frac{5}{2} \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix} .$$

Bien-sûr, on trouverait la même chose en cherchant les composantes comme on le faisait avant, en étudiant le système

$$\gamma_1 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + \gamma_2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + \gamma_3 \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix} = \begin{pmatrix} 7 \\ -5 \\ 3 \end{pmatrix} .$$

◇

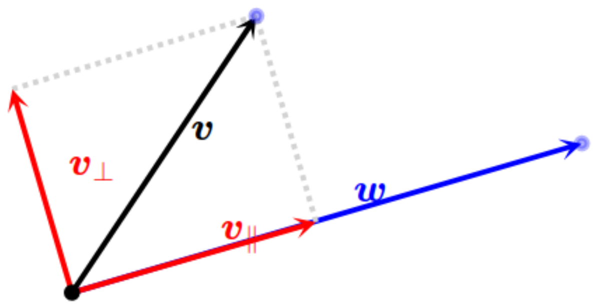
12.5 Projection sur un vecteur

La notion d'orthogonalité permet d'introduire en algèbre linéaire plusieurs notions géométriques très utiles. La première est celle de **projection**.

Comme motivation, fixons un vecteur $\mathbf{w} \in \mathbb{R}^n$, et posons la question suivante : Pour un deuxième $\mathbf{v} \in \mathbb{R}^n$ donné, comment définir la *projection orthogonale de* $\mathbf{v}$ *sur* $\mathbf{w}$?

Informel 12.26. On a déjà considéré, dans le plan, la projection d'un vecteur sur une droite. Mais ici, on est en dimension quelconque n ! Et nous allons commencer par projeter sur un vecteur, mais plus loin nous projetterons sur un sous-espace vectoriel quelconque de $\mathbb{R}^n$.

Un schéma peut aider à comprendre comment nous allons procéder (attention : cette image est représentée dans le plan, mais l'argument qui suit fonctionne en toute dimension !) :



La projection orthogonale de $\mathbf{v}$ sur $\mathbf{w}$, que nous noterons $\mathbf{v}_\parallel$ pour commencer, doit permettre de décomposer $\mathbf{v}$ en deux composantes vectorielles,

$$\mathbf{v} = \mathbf{v}_\parallel + \mathbf{v}_\perp \,,$$

où

1) $\mathbf{v}_\parallel$ est colinéaire (parallèle) à $\mathbf{w}$,
2) $\mathbf{v}_\perp$ est orthogonal à $\mathbf{w}$.

Il se trouve que ces deux conditions caractérisent entièrement $\mathbf{v}_\parallel$ et $\mathbf{v}_\perp$.
En effet, pour que $\mathbf{v}_\parallel$ soit colinéaire à $\mathbf{w}$, il doit exister un scalaire α tel que

$$\mathbf{v}_\parallel = \alpha \mathbf{w} \,.$$

Puis, pour que $\mathbf{v}_\perp$ soit orthogonal à $\mathbf{w}$, il faut que

$$0 = \mathbf{v}_\perp \cdot \mathbf{w} = (\mathbf{v} - \mathbf{v}_\parallel) \cdot \mathbf{w} = (\mathbf{v} - \alpha \mathbf{w}) \cdot \mathbf{w} \,.$$

De cette dernière relation, on tire que

$$\alpha = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{w}\|^2} \,.$$

En utilisant ce scalaire particulier dans $\mathbf{v}_\parallel = \alpha \mathbf{w}$, ceci motive la définition suivante :

Définition 12.27. Soit $\mathbf{w} \in \mathbb{R}^n$, $\mathbf{w} \neq \mathbf{0}$. La **projection orthogonale de $\mathbf{v} \in \mathbb{R}^n$ sur $\mathbf{w}$** est définie par

$$\mathrm{proj}_{\mathbf{w}}(\mathbf{v}) := \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w} = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{w}\|^2}\mathbf{w}\,.$$

Exemple 12.28. Dans $\mathbb{R}^5$, la projection orthogonale de

$$\mathbf{v} = \begin{pmatrix} 0 \\ 2 \\ -3 \\ 1 \\ -1 \end{pmatrix} \qquad \text{sur } \mathbf{w} = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

est donnée par

$$\mathrm{proj}_{\mathbf{w}}(\mathbf{v}) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{w}\|^2}\mathbf{w} = \frac{2}{3}\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2/3 \\ 0 \\ -2/3 \\ 0 \\ 2/3 \end{pmatrix}$$

◇

Remarque 12.29. La définition de $\mathrm{proj}_{\mathbf{w}}(\mathbf{v})$ dépend uniquement de la *direction* de $\mathbf{w}$, pas de son sens ni de sa norme. En effet, la projection sur un vecteur colinéaire à $\mathbf{w}$, $\mathbf{w}' = \lambda\mathbf{w}$, donne le même résultat, puisque

$$\begin{aligned} \mathrm{proj}_{\mathbf{w}'}(\mathbf{v}) &= \frac{\mathbf{v} \cdot \mathbf{w}'}{\|\mathbf{w}'\|^2}\mathbf{w}' \\ &= \frac{\mathbf{v} \cdot (\lambda\mathbf{w})}{\|\lambda\mathbf{w}\|^2}(\lambda\mathbf{w}) \\ &= \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{w}\|^2}\mathbf{w} \\ &= \mathrm{proj}_{\mathbf{w}}(\mathbf{v})\,. \end{aligned}$$

Donc il est plus juste de penser à la projection sur un vecteur comme à la projection sur *la droite engendrée par* ce vecteur. ◇

La projection de $\mathbf{v}$ sur $\mathbf{w}$ est aussi optimale, dans le sens où c'est elle qui réalise la distance minimale entre $\mathbf{v}$ et un vecteur quelconque de la droite dirigée par $\mathbf{w}$:

Théorème 12.30. *Soit $\mathbf{w} \in \mathbb{R}^n$ non-nul, et soit $W = \mathrm{Vect}\{\mathbf{w}\}$. Alors*

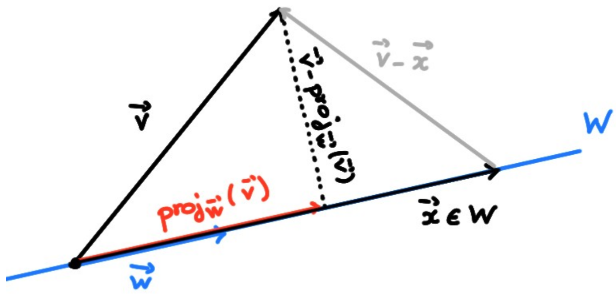
$$\|\mathbf{v} - \mathrm{proj}_{\mathbf{w}}(\mathbf{v})\| \leqslant \|\mathbf{v} - \mathbf{x}\| \qquad \forall \mathbf{x} \in W\,.$$

Puisque $\mathrm{proj}_{\mathbf{w}}(\mathbf{v}) \in W$, ceci implique

$$\|\mathbf{v} - \mathrm{proj}_{\mathbf{w}}(\mathbf{v})\| = \min_{\mathbf{x} \in W} \|\mathbf{v} - \mathbf{x}\|\,.$$

De plus, $\mathrm{proj}_W(\mathbf{v})$ est l'unique vecteur de W qui réalise ce minimum.

En d'autres termes, $\mathrm{proj}_{\mathbf{w}}(\mathbf{v})$ est *le* vecteur de W dont la distance à $\mathbf{v}$ est minimale :



Preuve: Pour tout $\mathbf{x} \in W$, on peut écrire

$$\mathbf{v} - \mathbf{x} = \underbrace{(\mathbf{v} - \text{proj}_{\mathbf{w}}(\mathbf{v}))}_{\in W^\perp} + \underbrace{(\text{proj}_{\mathbf{w}}(\mathbf{v}) - \mathbf{x})}_{\in W}$$

On a donc

$$\begin{aligned} \|\mathbf{v} - \mathbf{x}\|^2 &= \|\mathbf{v} - \text{proj}_{\mathbf{w}}(\mathbf{v})\|^2 + \underbrace{\|\text{proj}_{\mathbf{w}}(\mathbf{v}) - \mathbf{x}\|^2}_{\geqslant 0} \\ &\geqslant \|\mathbf{v} - \text{proj}_{\mathbf{w}}(\mathbf{v})\|^2 \end{aligned}$$

Supposons maintenant qu'il existe, en plus de $\mathbf{v}_\parallel = \text{proj}_{\mathbf{w}}(\mathbf{v})$, un autre vecteur de W satisfaisant la même propriété ; notons-le $\mathbf{v}'_\parallel$. Alors par définition,

$$\|\mathbf{v} - \mathbf{v}_\parallel\| = \min_{\mathbf{x} \in W} \|\mathbf{v} - \mathbf{x}\| = \|\mathbf{v} - \mathbf{v}'_\parallel\| \,.$$

Aussi,

$$\begin{aligned} \|\mathbf{v} - \mathbf{v}'_\parallel\|^2 &= \|\underbrace{(\mathbf{v} - \mathbf{v}_\parallel)}_{\in W^\perp} + \underbrace{(\mathbf{v}_\parallel - \mathbf{v}'_\parallel)}_{\in W}\|^2 \\ &= \|\mathbf{v} - \mathbf{v}_\parallel\|^2 + \|\mathbf{v}_\parallel - \mathbf{v}'_\parallel\|^2 \end{aligned}$$

On a donc

$$\|\mathbf{v}_\parallel - \mathbf{v}'_\parallel\|^2 = 0 \,,$$

qui implique $\mathbf{v}_\parallel = \mathbf{v}'_\parallel$. □

Le fait que $\text{proj}_{\mathbf{w}}(\mathbf{v})$ réalise un minimum indique que certains *problèmes d'optimisation* pourront trouver une solution par l'utilisation de projections. (Voir plus loin, *Méthode des moindres carrés*.)

12.6 Projection sur un sous-espace vectoriel

Motivation : projection sur un plan de $\mathbb{R}^3$

Pour motiver la définition générale de projection sur un sous-espace vectoriel W, nous commencerons par un cas légèrement plus compliqué que la projection sur une droite (section précédente), en considérant une projection sur un plan.

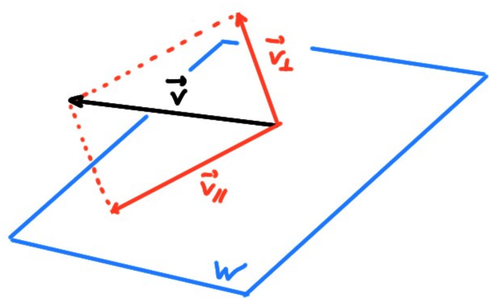
Exemple 12.31. Considérons les deux vecteurs non-colinéaires

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} , \qquad \mathbf{w}_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} ,$$

ainsi que le plan contenant l'origine, engendré par ces deux vecteurs :

$$W = \text{Vect}\{\mathbf{w}_1, \mathbf{w}_2\} \,.$$

Si $\mathbf{v} = \begin{pmatrix} 5 \\ 2 \\ -3 \end{pmatrix}$, comment calculer sa **projection orthogonale sur** W ?



Comme dans la section précédente, nous commencerons par représenter la projection de $\mathbf{v}$ sur W à l'aide du symbole $\mathbf{v}_{\parallel}$. Cette projection doit permettre de décomposer $\mathbf{v}$ en deux composantes vectorielles,

$$\mathbf{v} = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp}\,,$$

où

1) $\mathbf{v}_{\parallel} \in W$
2) $\mathbf{v}_{\perp} \in W^{\perp}$

La première condition impose que $\mathbf{v}_{\parallel}$ soit une combinaison linéaire de $\mathbf{w}_1$ et $\mathbf{w}_2$:

$$\mathbf{v}_{\parallel} = \alpha_1 \mathbf{w}_1 + \alpha_2 \mathbf{w}_2\,,$$

et la deuxième impose que

$$\begin{cases} 0 = \mathbf{v}_{\perp} \cdot \mathbf{w}_1 = (\mathbf{v} - \alpha_1\mathbf{w}_1 - \alpha_2\mathbf{w}_2) \cdot \mathbf{w}_1\,, \\ 0 = \mathbf{v}_{\perp} \cdot \mathbf{w}_2 = (\mathbf{v} - \alpha_1\mathbf{w}_1 - \alpha_2\mathbf{w}_2) \cdot \mathbf{w}_2\,. \end{cases}$$

Comme $\|\mathbf{w}_1\|^2 = 3$, $\|\mathbf{w}_2\|^2 = 6$, $\mathbf{w}_1 \cdot \mathbf{w}_2 = -2$, $\mathbf{v} \cdot \mathbf{w}_1 = 10$, $\mathbf{v} \cdot \mathbf{w}_2 = -11$, on en déduit que les coefficients α_1, α_2 sont solutions du système

$$(*)\left\{ \begin{array}{rcrcr} 3\alpha_1 & - & 2\alpha_2 & = & 10 \\ -2\alpha_1 & + & 6\alpha_2 & = & -11\,. \end{array} \right.$$

On a donc $\alpha_1 = \frac{19}{7}$, $\alpha_2 = -\frac{13}{14}$. Ainsi, la projection de $\mathbf{v}$ sur le plan W est donnée par

$$\mathbf{v}_{\parallel} = \frac{19}{7}\mathbf{w}_1 - \frac{13}{14}\mathbf{w}_2 = \begin{pmatrix} 32/7 \\ 25/14 \\ -51/14 \end{pmatrix}\,.$$

◇

Cas général

Dans le cas général, énonçons d'abord le résultat général qui garantit que la projection orthogonale sur un sous-espace vectoriel existe toujours :

Théorème 12.32. *Soit W un sous-espace vectoriel de $\mathbb{R}^n$, et soit $\mathbf{v} \in \mathbb{R}^n$. Alors il existe une unique paire de vecteurs, $\mathbf{v}_\parallel$ et $\mathbf{v}_\perp$, telle que*

$$\mathbf{v} = \mathbf{v}_\parallel + \mathbf{v}_\perp \,,$$

et telle que

1) $\mathbf{v}_\parallel \in W$
2) $\mathbf{v}_\perp \in W^\perp$

*Le vecteur $\mathbf{v}_\parallel$ est appelé **projection orthogonale de $\mathbf{v}$ sur W**, et sera noté*

$$\mathbf{v}_\parallel \equiv \mathrm{proj}_W(\mathbf{v}) \,.$$

De plus, $\mathrm{proj}_W(\mathbf{v})$ est l'unique vecteur de W qui minimise la distance à $\mathbf{v}$:

$$\|\mathbf{v} - \mathrm{proj}_W(\mathbf{v})\| = \min_{\mathbf{x} \in W} \|\mathbf{v} - \mathbf{x}\| \,.$$

Dans le cas où on connaît une famille génératrice pour W,

$$W = \mathrm{Vect}\{\mathbf{w}_1, \dots, \mathbf{w}_k\} \,,$$

on peut calculer $\mathrm{proj}_W(\mathbf{v})$ comme on l'a fait dans la section précédente, en commençant par l'écrire comme une combinaison linéaire

$$\mathbf{v}_\parallel = \alpha_1 \mathbf{w}_1 + \cdots \alpha_k \mathbf{w}_k \,,$$

où les coefficients doivent satisfaire simultanément aux k conditions suivantes :

$$\begin{cases} 0 & = (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_1 \,, \\ 0 & = (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_2 \,, \\ & \vdots \\ 0 & = (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_k \,. \end{cases}$$

Sans présenter de difficulté particulière, cette approche requiert malgré tout la résolution d'un système $n \times k$.

Cas où W est décrit par une base orthogonale

Lorsque W est décrit à l'aide d'une base orthogonale, la projection sur W prend une forme plus explicite :

Théorème 12.33. *Soit W un sous-espace vectoriel de $\mathbb{R}^n$, et soit $\mathcal{B} = (\mathbf{w}_1, \dots, \mathbf{w}_k)$ une base **orthogonale** de W. Alors la projection orthogonale d'un vecteur $\mathbf{v} \in \mathbb{R}^n$ sur W est donnée par*

$$\mathrm{proj}_W(\mathbf{v}) = \sum_{j=1}^{k} \frac{\mathbf{v} \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j \,.$$

En particulier, l'application $\mathbf{v} \mapsto \mathrm{proj}_W(\mathbf{v})$ est linéaire.

Preuve: Comme décrit plus haut, la projection est de la forme

$$\mathrm{proj}_W(\mathbf{v}) = \alpha_1 \mathbf{w}_1 + \cdots \alpha_k \mathbf{w}_k \,,$$

où les α_j doivent satisfaire

$$\begin{cases} 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_1 \,, \\ 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_2 \,, \\ &\vdots \\ 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_k \,. \end{cases}$$

Mais puisque la base $\mathcal{B}$ est orthogonale, $\mathbf{w}_i \cdot \mathbf{w}_j = 0$ si $i \neq j$. Il reste donc

$$\begin{cases} 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1) \cdot \mathbf{w}_1 \,, \\ &\vdots \\ 0 &= (\mathbf{v} - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_k \,, \end{cases}$$

qui donne bien $\alpha_j = \frac{\mathbf{v} \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2}$ pour tout $j = 1, \ldots, k$.

Vérifions la linéarité :

$$\begin{aligned} \mathrm{proj}_W(\beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2) &= \sum_{j=1}^k \frac{(\beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2) \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j \\ &= \sum_{j=1}^k \left(\beta_1 \frac{\mathbf{v}_1 \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j + \beta_2 \frac{\mathbf{v}_2 \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j \right) \\ &= \beta_1 \sum_{j=1}^k \frac{\mathbf{v}_1 \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j + \beta_2 \sum_{j=1}^k \frac{\mathbf{v}_2 \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j \\ &= \beta_1 \mathrm{proj}_W(\mathbf{v}_1) + \beta_2 \mathrm{proj}_W(\mathbf{v}_2) \,. \end{aligned}$$

□

La linéarité de la projection fait qu'on peut chercher sa matrice relativement à une base.

Exemple 12.34. Considérons les deux vecteurs non-colinéaires

$$\mathbf{w}_1 = \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix} , \qquad \mathbf{w}_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} ,$$

ainsi que le plan contenant l'origine, engendré par ces deux vecteurs :

$$W = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\} \,.$$

Commençons par prendre un vecteur, par exemple

$$\mathbf{v} = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} ,$$

et calculons sa projection sur W. On pourrait procéder comme on l'a fait plus haut, mais on remarque tout de suite que cette fois, $\{\mathbf{w}_1, \mathbf{w}_2\}$ est orthogonale car $\mathbf{w}_1 \cdot \mathbf{w}_2 = 0$. On peut donc écrire la projection directement à l'aide de la formule du théorème :

$$\begin{aligned} \mathrm{proj}_W(\mathbf{v}) &= \frac{\mathbf{v} \cdot \mathbf{w}_1}{\|\mathbf{w}_1\|^2} \mathbf{w}_1 + \frac{\mathbf{v} \cdot \mathbf{w}_2}{\|\mathbf{w}_2\|^2} \mathbf{w}_1 \\ &= \frac{-11}{30} \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix} + \frac{-1}{6} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2/5 \\ -2 \\ 1/5 \end{pmatrix} . \end{aligned}$$

Considérons ensuite la matrice de la projection, relativement à la base canonique :

$$\begin{aligned}[\mathrm{proj}_W]_{\mathcal{B}_{\mathrm{can}}} &= [[\mathrm{proj}_W(\mathbf{e}_1)]\,[\mathrm{proj}_W(\mathbf{e}_2)]\,[\mathrm{proj}_W(\mathbf{e}_3)]]] \\ &= \begin{pmatrix} 4/5 & 0 & -2/5 \\ 0 & 1 & 0 \\ -2/5 & 0 & 1/5 \end{pmatrix}\end{aligned}$$

◇

Cas où W est décrit par une base orthonomale

Si on exige en plus que la famille qui engendre W soit formée de vecteurs unitaires, alors la projection est encore plus simple à décrire :

Théorème 12.35. *Soit W un sous-espace vectoriel de $\mathbb{R}^n$, et soit $\mathcal{B} = (\mathbf{u}_1, \dots, \mathbf{u}_k)$ une base **orthonormée** de W. Alors la projection orthogonale d'un vecteur $\mathbf{v} \in \mathbb{R}^n$ sur W est donnée par*

$$\mathrm{proj}_W(\mathbf{v}) = \sum_{j=1}^{k} (\mathbf{v} \cdot \mathbf{u}_j)\mathbf{u}_j \, .$$

De plus, la matrice de proj_W, relativement à la base canonique, est donnée par

$$[\mathrm{proj}_W]_{\mathcal{B}_{\mathrm{can}}} = UU^T \, ,$$

où U est la matrice $n \times k$ dont les colonnes sont les vecteurs de la base $\mathcal{B}$ exprimés dans la base canonique :

$$U = \big[[\mathbf{u}_1]_{\mathcal{B}_{\mathrm{can}}} \cdots [\mathbf{u}_k]_{\mathcal{B}_{\mathrm{can}}}\big] \, .$$

Preuve: Par le théorème précédent, et puisque $\|\mathbf{u}_j\| = 1$ pour tout j, la projection est bien donnée par

$$\mathrm{proj}_W(\mathbf{v}) = (\mathbf{v} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{v} \cdot \mathbf{u}_k)\mathbf{u}_k \, .$$

Pensons maintenant en composantes relativement à la base canonique, et profitons de la structure de cette expression pour la récrire sous forme d'un produit matriciel :

$$\begin{aligned}[\mathrm{proj}_W(\mathbf{v})]_{\mathcal{B}_{\mathrm{can}}} &= (\mathbf{v} \cdot \mathbf{u}_1)[\mathbf{u}_1]_{\mathcal{B}_{\mathrm{can}}} + \cdots + (\mathbf{v} \cdot \mathbf{u}_k)[\mathbf{u}_k]_{\mathcal{B}_{\mathrm{can}}} \\ &= \big[[\mathbf{u}_1]_{\mathcal{B}_{\mathrm{can}}} \cdots [\mathbf{u}_k]_{\mathcal{B}_{\mathrm{can}}}\big] \begin{pmatrix} \mathbf{v} \cdot \mathbf{u}_1 \\ \vdots \\ \mathbf{v} \cdot \mathbf{u}_k \end{pmatrix} \\ &= \big[[\mathbf{u}_1]_{\mathcal{B}_{\mathrm{can}}} \cdots [\mathbf{u}_k]_{\mathcal{B}_{\mathrm{can}}}\big] \begin{pmatrix} [\mathbf{u}_1]^T_{\mathcal{B}_{\mathrm{can}}} [\mathbf{v}]_{\mathcal{B}_{\mathrm{can}}} \\ \vdots \\ [\mathbf{u}_k]^T_{\mathcal{B}_{\mathrm{can}}} [\mathbf{v}]_{\mathcal{B}_{\mathrm{can}}} \end{pmatrix} \\ &= \underbrace{\big[[\mathbf{u}_1]_{\mathcal{B}_{\mathrm{can}}} \cdots [\mathbf{u}_k]_{\mathcal{B}_{\mathrm{can}}}\big]}_{=:U} \underbrace{\begin{bmatrix} [\mathbf{u}_1]^T_{\mathcal{B}_{\mathrm{can}}} \\ \vdots \\ [\mathbf{u}_k]^T_{\mathcal{B}_{\mathrm{can}}} \end{bmatrix}}_{=U^T} [\mathbf{v}]_{\mathcal{B}_{\mathrm{can}}}\end{aligned}$$

□

Remarque 12.36. La projection est une application de $\mathbb{R}^n$ dans $\mathbb{R}^n$, donc la matrice qui la représente est $n \times n$. C'est bien le cas ici puisque

$$[\mathrm{proj}_W]_{\mathcal{B}_{\mathrm{can}}} = \underbrace{\underbrace{U}_{n\times k}\underbrace{U^T}_{k\times n}}_{n\times n}$$

◇

Exemple 12.37. Considérons la même projection orthogonale que celle vue plus haut, sur le plan W engendré par

$$\mathbf{w}_1 = \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix},$$

On sait que la base

$$\mathcal{B} = (\mathbf{w}_1, \mathbf{w}_2)$$

est orthogonale, et on peut la rendre orthonormée en divisant chaque vecteur par sa norme :

$$\mathcal{B}' = \left(\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}, \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|}\right).$$

On peut maintenant utiliser le théorème pour obtenir la matrice de la projection sur W relativement à la base canonique :

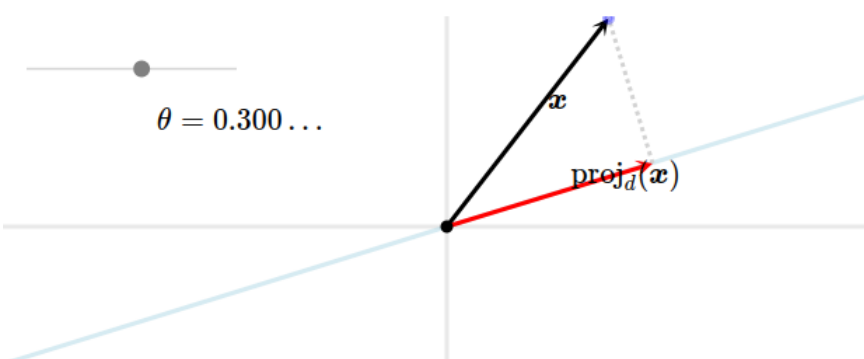
$$\begin{aligned}
[\mathrm{proj}_W]_{\mathcal{B}_{\mathrm{can}}} &= \underbrace{U}_{3\times 2}\underbrace{U^T}_{2\times 3} \\
&= \begin{bmatrix} \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} & \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} \end{bmatrix} \begin{bmatrix} (\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|})^T \\ (\frac{\mathbf{w}_2}{\|\mathbf{w}_2\|})^T \end{bmatrix} \\
&= \begin{pmatrix} 2/\sqrt{30} & -2/\sqrt{6} \\ 5/\sqrt{30} & 1/\sqrt{6} \\ -1/\sqrt{30} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 2/\sqrt{30} & 5/\sqrt{30} & -1/\sqrt{30} \\ -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} \\
&= \begin{pmatrix} 4/5 & 0 & -2/5 \\ 0 & 1 & 0 \\ -2/5 & 0 & 1/5 \end{pmatrix},
\end{aligned}$$

qui est bien la même que nous avions trouvé plus haut. On peut maintenant utiliser cette matrice pour projeter n'importe quel vecteur sur W. Par exemple,

$$\mathrm{proj}_W\left(\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}\right) = \begin{pmatrix} 4/5 & 0 & -2/5 \\ 0 & 1 & 0 \\ -2/5 & 0 & 1/5 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} -2/5 \\ -2 \\ 1/5 \end{pmatrix}$$

◇

Exemple 12.38. Considérons la projection proj_d sur une droite d passant par l'origine et faisant un angle de θ avec e_1 :



Cette droite d est un sous-espace de $\mathbb{R}^2$, engendrée par le vecteur unitaire

$$\mathbf{u}_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} .$$

Par la formule du théorème ci-dessus, sa matrice relativement à la base canonique est donc donnée par

$$\begin{aligned} [\mathrm{proj}_d]_{\mathcal{B}_{can}} &= UU^T \\ &= \mathbf{u}_1\mathbf{u}_1^T \\ &= \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix} . \end{aligned}$$

◇

Remarque 12.39. Il est important d'apprécier l'ordre des matrices apparaissant dans la formule

$$[\mathrm{proj}_W]_{\mathcal{B}_{\mathrm{can}}} = UU^T$$

Le fait que les matrices soient dans cet ordre ("U fois U^T") font de leur produit une application linéaire non-triviale, qui projette sur l'espace engendré par les colonnes de U. Car si on multiplie ces matrices dans l'ordre inverse, on obtient une matrice $k \times k$ contenant les produits scalaires

$$\mathbf{u}_i^T\mathbf{u}_j = \mathbf{u}_i \cdot \mathbf{u}_j = \begin{cases} \|\mathbf{u}_i\| = 1 & \text{si } i = j\,, \\ 0 & \text{sinon,} \end{cases}$$

et donc

$$\begin{aligned} U^TU &= \begin{pmatrix} \mathbf{u}_1\cdot\mathbf{u}_1 & \mathbf{u}_1\cdot\mathbf{u}_2 & \cdots & \mathbf{u}_1\cdot\mathbf{u}_k \\ \mathbf{u}_2\cdot\mathbf{u}_1 & \cdots & & \mathbf{u}_2\cdot\mathbf{u}_k \\ \vdots & & & \vdots \\ \mathbf{u}_k\cdot\mathbf{u}_1 & \cdots & & \mathbf{u}_k\cdot\mathbf{u}_k \end{pmatrix} \\ &= \begin{pmatrix} \|\mathbf{u}_1\|^2 & 0 & \cdots & 0 \\ 0 & \cdots & & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & \|\mathbf{u}_k\|^2 \end{pmatrix} \\ &= I_k\,. \end{aligned}$$

◇

12.7 Le procédé d'orthogonalisation de Gram-Schmidt

Les sections précédentes ont montré tout l'avantage de travailler avec une base orthogonale (ou orthonormale) pour un sous-espace W, puisque cela permet par exemple d'accéder directement aux composantes d'un vecteur relativement à cette base, ou de calculer plus facilement des projections orthogonales sur W.

Mais il se peut que le sous-espace W soit défini dès le départ par une base $\mathcal{B}$ qui n'est *pas* orthogonale. Pour profiter des avantages décrits ci-dessus, il est donc naturel de chercher une autre base de W, $\mathcal{B}'$, qui soit elle orthogonale.

Nous allons voir qu'une telle base existe toujours, et nous verrons comment la construire en *modifiant* la base de départ, par un algorithme appelé le *procédé d'orthogonalisation de Gram-Schmidt*.

Informel 12.40. L'idée est de "tordre" un à un les vecteurs de $\mathcal{B}$, de façon à les rendre progressivement orthogonaux deux-à-deux, et en garantissant *qu'ils engendrent toujours* W.

Voyons comment faire sur un exemple très simple d'une base ne contenant que deux vecteurs.

L'idée, sur un exemple où $\dim(W) = 2$

Considérons le plan de $\mathbb{R}^3$, $W = \text{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}$, où

$$\mathbf{w}_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.$$

La paire $\mathcal{B} = (\mathbf{w}_1, \mathbf{w}_2)$ est une base de W, mais elle n'est pas orthogonale car

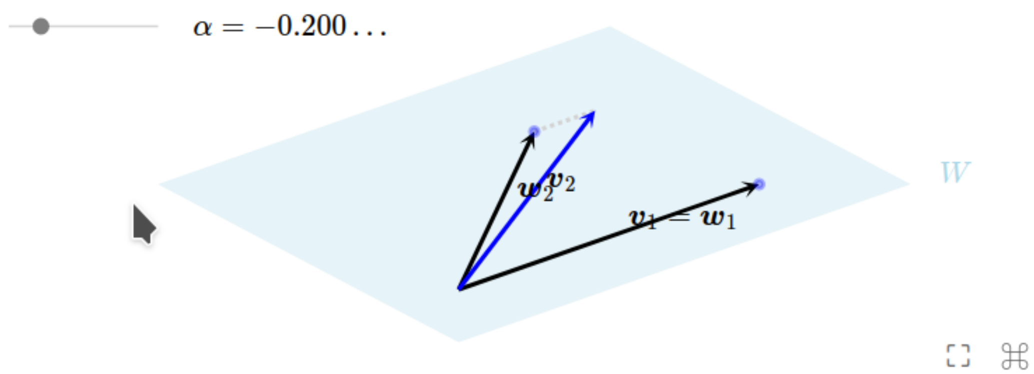
$$\mathbf{w}_1 \cdot \mathbf{w}_2 = 3 \neq 0\,.$$

Voyons comment modifier $\mathcal{B}$ de façon à la transformer en une autre base pour W, orthogonale cette fois.

La nouvelle base sera $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2)$, avec

$$\begin{aligned} \mathbf{v}_1 &:= \mathbf{w}_1\,, \\ \mathbf{v}_2 &:= \mathbf{w}_2 - \alpha \mathbf{w}_1\,. \end{aligned}$$

Voyons ce qui se passe lorsque α varie :



Informel 12.41. Remarquons que l'on "tord" $\mathbf{w}_2$ en lui rajoutant un multiple de $\mathbf{w}_1$, ce qui fait que $\mathbf{v}_2$ *reste dans le plan* W !

C'est évident sur l'animation ci-dessus, mais écrivons-le explicitement :

Lemme 18. *Peu importe la valeur du scalaire α, $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2)$ est toujours une base de W.*

Preuve: (en exercice) □

Il s'agit ensuite de choisir α de façon à ce que $\mathcal{B}'$ soit orthogonale. Or la seule condition à satisfaire est que

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = 0\,,$$

qui se traduit par

$$\mathbf{w}_1 \cdot (\mathbf{w}_2 - \alpha \mathbf{w}_1) = 0\,,$$

et qui implique

$$\alpha = \frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{\|\mathbf{w}_1\|^2}$$

Ainsi, $\alpha \mathbf{w}_1 = \frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{\|\mathbf{w}_1\|^2} \mathbf{w}_1$, qui n'est autre que la projection de $\mathbf{w}_2$ sur $\mathbf{w}_1$ (c'est-à-dire sur $\mathbf{v}_1$). En résumé, on a pris $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2)$, où

$$\begin{aligned} \mathbf{v}_1 &:= \mathbf{w}_1\,, \\ \mathbf{v}_2 &:= \mathbf{w}_2 - \mathrm{proj}_{\mathbf{v}_1}(\mathbf{w}_2)\,, \end{aligned}$$

qui donne

$$\mathbf{v}_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}, \qquad \mathbf{v}_2 = \frac{1}{14} \begin{pmatrix} 5 \\ 31 \\ 8 \end{pmatrix}.$$

Maintenant, $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2)$ est orthogonale puisque $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

La construction décrite dans l'exemple ci-dessus n'a rien de particulier à $\mathbb{R}^3$, et peut s'utiliser pour orthogonaliser la base de n'importe quel sous-espace de dimension 2 :

Exemple 12.42. Considérons le plan de $\mathbb{R}^5$ engendré par

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} 0 \\ -2 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

La base $\mathcal{B} = (\mathbf{w}_1, \mathbf{w}_2)$ de ce plan n'est pas orthogonale, mais en prenant $\mathbf{v}_1 := \mathbf{w}_1$, et

$$\mathbf{v}_2 := \mathbf{w}_2 - \mathrm{proj}_{\mathbf{w}_1}(\mathbf{w}_2) = \begin{pmatrix} 0 \\ -2 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \frac{-1}{3} \begin{pmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/3 \\ -2 \\ 1/3 \\ 2/3 \\ 1 \end{pmatrix},$$

on obtient une base $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2)$ orthogonale. ◇

Cas général

Dans le cas général, considérons un sous-espace W de $\mathbb{R}^3$, de dimension $k \leqslant n$, muni d'une base (a priori pas orthogonale)

$$\mathcal{B} = (\mathbf{w}_1, \dots, \mathbf{w}_k)\,,$$

et voyons comment l'utiliser pour construire une nouvelle base de W,

$$\mathcal{B}' = (\mathbf{v}_1, \dots, \mathbf{v}_k)\,,$$

qui soit orthogonale. Cette construction est le **procédé d'orthogonalisation de Gram-Schmidt**.

L'idée est de procéder de manière **inductive**, le j-ème vecteur $\mathbf{v}_j$ de $\mathcal{B}'$ étant construit à partir des j premiers vecteurs de $\mathcal{B}$, de façon à ce que pour tout $j = 1, \dots, k$, deux conditions soient satisfaites :

- ⋆ $\{\mathbf{v}_1, \dots, \mathbf{v}_j\}$ est orthogonale (et donc libre),

⋆ $\mathrm{Vect}\{\mathbf{v}_1, \dots, \mathbf{v}_j\} = \mathrm{Vect}\{\mathbf{w}_1, \dots, \mathbf{w}_j\}$.

La vérification de ces conditions implique qu'à la fin, lorsque $j = k$, on a bien construit une famille $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ qui est orthogonale (et donc libre), et qui engendre W.

L'exemple précédent a suggéré de commencer par modifier $\mathbf{w}_2$ en lui soustrayant sa projection sur $\mathbf{w}_1$. Pour les suivants, on peut continuer à soustraire à chaque vecteur sa projection sur l'espace engendré par les précédents :

$$\begin{aligned}
\mathbf{v}_1 &:= \mathbf{w}_1 \,, \\
\mathbf{v}_2 &:= \mathbf{w}_2 - \mathrm{proj}_{\mathrm{Vect}\{\mathbf{w}_1\}}(\mathbf{w}_2) \,, \\
\mathbf{v}_3 &:= \mathbf{w}_3 - \mathrm{proj}_{\mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}}(\mathbf{w}_3) \,, \\
&\ \ \vdots \\
\mathbf{v}_j &:= \mathbf{w}_j - \mathrm{proj}_{\mathrm{Vect}\{\mathbf{w}_1, \dots, \mathbf{w}_{j-1}\}}(\mathbf{w}_j) \\
&\ \ \vdots \\
\mathbf{v}_k &:= \mathbf{w}_k - \mathrm{proj}_{\mathrm{Vect}\{\mathbf{w}_1, \dots\dots, \mathbf{w}_{k-1}\}}(\mathbf{w}_k) \,.
\end{aligned}$$

Remarquons que le procédé nécessite, à l'étape j, de calculer la projection sur le sous-espace $\mathrm{Vect}\{\mathbf{w}_1, \dots, \mathbf{w}_{j-1}\}$. Or, puisque

$$\mathrm{Vect}\{\mathbf{w}_1, \dots, \mathbf{w}_{j-1}\} = \mathrm{Vect}\{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}\} \,,$$

on a, pour tout j,

$$\mathrm{proj}_{\mathrm{Vect}\{\mathbf{w}_1, \dots, \mathbf{w}_{j-1}\}}(\mathbf{w}_j) = \mathrm{proj}_{\mathrm{Vect}\{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}\}}(\mathbf{w}_j) \,.$$

Maintenant, comme $\{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}\}$ est orthogonale, la formule de la section précédente permet d'écrire cette dernière projection comme

$$\mathrm{proj}_{\mathrm{Vect}\{\mathbf{v}_1, \dots, \mathbf{v}_{j-1}\}}(\mathbf{w}_j) = \sum_{i=1}^{j-1} \frac{\mathbf{w}_j \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2} \mathbf{v}_i \,.$$

Donc on peut écrire le procédé comme suit :

$$\begin{aligned}
\mathbf{v}_1 &:= \mathbf{w}_1 \,, \\
\mathbf{v}_2 &:= \mathbf{w}_2 - \frac{\mathbf{w}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 \\
\mathbf{v}_3 &:= \mathbf{w}_3 - \sum_{i=1}^{2} \frac{\mathbf{w}_3 \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2} \mathbf{v}_i \,, \\
&\ \ \vdots \\
\mathbf{v}_k &:= \mathbf{w}_k - \sum_{i=1}^{k-1} \frac{\mathbf{w}_k \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2} \mathbf{v}_i \,,
\end{aligned}$$

Remarque 12.43. ⋆ Une fois le procédé terminé, on peut toujours normaliser les vecteurs de $\mathcal{B}'$ pour en faire une base *orthonormée* de W :

$$\mathcal{B}'' = \left(\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \dots, \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|} \right) \,.$$

* $\star$ La convention est que l'algorithme du procédé de Gram-Schmidt se fait en respectant l'ordre qui est est fixé dans la base de départ.

$\diamond$

Exemple 12.44. Considérons, dans $\mathbb{R}^4$, le sous-espace W défini par

$$W = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}\,,$$

où

$$\mathbf{w}_1 := \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}\,, \quad \mathbf{w}_2 := \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}\,, \quad \mathbf{w}_3 := \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}\,.$$

Appliquons le procédé de Gram-Schmidt. D'abord, $\mathbf{v}_1 := \mathbf{w}_1$, puis

$$\begin{aligned}
\mathbf{v}_2 &:= \mathbf{w}_2 - \frac{\mathbf{w}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\
&= \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} - \frac{(-1)}{2}\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 0 \\ 1/2 \\ 1 \end{pmatrix}\,,
\end{aligned}$$

et pour finir

$$\begin{aligned}
\mathbf{v}_3 &:= \mathbf{w}_3 - \frac{\mathbf{w}_3 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\mathbf{w}_3 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 \\
&= \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} - \frac{-1}{2}\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} - \frac{-1/2}{3/2}\begin{pmatrix} -1/2 \\ 0 \\ 1/2 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} 1/3 \\ 1 \\ -1/3 \\ 1/3 \end{pmatrix}\,.
\end{aligned}$$

Remarquons que $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ est bien orthogonale puisque, par construction,

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot \mathbf{v}_3 = \mathbf{v}_2 \cdot \mathbf{v}_3 = 0\,.$$

$\diamond$

Dans ce dernier exemple, on aurait pu remarquer dès le début que $\mathbf{w}_2 \perp \mathbf{w}_3$, et donc obtenir une base orthogonale $(\mathbf{v}'_1, \mathbf{v}'_2, \mathbf{v}'_3)$, en gardant deux vecteurs inchangés, et en ne modifiant que $\mathbf{w}_1$:

$$\begin{aligned}
\mathbf{v}'_2 &:= \mathbf{w}_2\,, \\
\mathbf{v}'_3 &:= \mathbf{w}_3\,, \\
\mathbf{v}'_1 &:= \mathbf{w}_1 - \frac{\mathbf{w}_1 \cdot \mathbf{v}'_2}{\|\mathbf{v}'_2\|^2}\mathbf{v}'_2 - \frac{\mathbf{w}_1 \cdot \mathbf{v}'_3}{\|\mathbf{v}'_3\|^2}\mathbf{v}'_3\,.
\end{aligned}$$

Donc en général, il y a plusieurs façons d'orthogonaliser une base, mais en général, lorsqu'on implémente le procédé de Gram-Schmidt, la convention est *de modifier les vecteurs dans l'ordre donné par la base de départ*.

12.8 La décomposition QR

La décomposition QR est une méthode qui permet de *factoriser* une matrice, c'est-à-dire de l'écrire comme un produit de deux autres matrices particulières (un peu comme une matrice carrée inversible peut être factorisée en un produit de matrices élémentaires).

On le verra, pouvoir écrire une matrice comme un produit de matrices plus simples possèdera de nombreux avantages.

Lorsque les colonnes de A sont indépendantes

Soit A une matrice $m \times n$, que l'on écrit à l'aide de ses colonnes :

$$A = [\mathbf{a}_1 \cdots \mathbf{a}_n] \,,$$

où chaque $\mathbf{a}_j \in \mathbb{R}^m$.

Si les colonnes de A sont linéairement indépendantes, elles forment une base de $\mathrm{Im}(A)$. Cette base n'est a priori pas orthogonale, on peut donc lui appliquer le procédé de Gram-Schmidt :

$$\begin{aligned}
\mathbf{v}_1 &:= \mathbf{a}_1 \,, \\
\mathbf{v}_2 &:= \mathbf{a}_2 - \frac{\mathbf{a}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\
\mathbf{v}_3 &:= \mathbf{a}_3 - \sum_{i=1}^{2} \frac{\mathbf{a}_3 \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2}\mathbf{v}_i \,, \\
&\vdots \\
\mathbf{v}_n &:= \mathbf{a}_n - \sum_{i=1}^{n-1} \frac{\mathbf{a}_n \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2}\mathbf{v}_i \,,
\end{aligned}$$

Normalisons chacun des $\mathbf{v}_j$, en définissant

$$\mathbf{u}_1 := \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} \,, \quad \dots \quad , \mathbf{u}_k := \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|} \,.$$

Les n relations ci-dessus permettent maintenant d'écrire

$$\begin{aligned}
\mathbf{a}_1 &= \|\mathbf{v}_1\|\mathbf{u}_1 \,, \\
\mathbf{a}_2 &= (\mathbf{a}_2 \cdot \mathbf{u}_1)\mathbf{u}_1 + \|\mathbf{v}_2\|\mathbf{u}_2 \\
\mathbf{a}_3 &= \left(\sum_{i=1}^{2} (\mathbf{a}_3 \cdot \mathbf{u}_i)\mathbf{u}_i\right) + \|\mathbf{v}_3\|\mathbf{u}_3 \,, \\
&\vdots \\
\mathbf{a}_n &= \left(\sum_{i=1}^{n-1} (\mathbf{a}_n \cdot \mathbf{u}_i)\mathbf{u}_i\right) + \|\mathbf{v}_n\|\mathbf{u}_n \,,
\end{aligned}$$

Cette décomposition des colonnes de A en combinaisons linéaires des vecteurs $\mathbf{u}_1, \dots, \mathbf{u}_n$ peut s'exprimer à l'aide de deux matrices.

- ⋆ D'abord, considérons la matrice Q, $m \times n$, définie par
$$Q := [\mathbf{u}_1 \cdots \mathbf{u}_n]\,.$$
Les colonnes de cette matrice étant unitaires et orthogonales deux à deux, on a (lire la remarque de la fin de la section précédente) :
$$Q^T Q = I_n\,.$$
- ⋆ Ensuite, considérons la matrice R, triangulaire supérieure $n \times n$, formée à partir des coefficients apparaissant dans les combinaisons linéaires ci-dessus :
$$R := \begin{pmatrix} \|\mathbf{v}_1\| & \mathbf{a}_2 \cdot \mathbf{u}_1 & \mathbf{a}_3 \cdot \mathbf{u}_1 & \cdots & \mathbf{a}_n \cdot \mathbf{u}_1 \\ 0 & \|\mathbf{v}_2\| & \mathbf{a}_3 \cdot \mathbf{u}_2 & \cdots & \mathbf{a}_n \cdot \mathbf{u}_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & \|\mathbf{v}_{n-1}\| & \mathbf{a}_n \cdot \mathbf{u}_{n-1} \\ 0 & \cdots & & 0 & \|\mathbf{v}_n\| \end{pmatrix}.$$
En d'autres termes, les coefficients apparaissant dans la k-ème colonne de R sont les coefficients de la combinaison linéaire donnant $\mathbf{a}_k$.

Par définition, $QR = A$. Pour le voir explicitement, on peut écrire
$$R = [\mathbf{r}_1 \cdots \mathbf{r}_n]\,,$$
où
$$\begin{aligned} \mathbf{r}_1 &= \|\mathbf{v}_1\|\mathbf{e}_1 \\ \mathbf{r}_2 &= (\mathbf{a}_1 \cdot \mathbf{u}_1)\mathbf{e}_1 + \|\mathbf{v}_2\|\mathbf{e}_2 \\ &\vdots \\ \mathbf{r}_n &= \left(\sum_{i=1}^{n-1} (\mathbf{a}_i \cdot \mathbf{u}_i)\mathbf{e}_i\right) + \|\mathbf{v}_n\|\mathbf{e}_n\,. \end{aligned}$$
Ainsi, la k-ème colonne de $QR = [Q\mathbf{r}_1 \cdots Q\mathbf{r}_n]$ est donnée par
$$\begin{aligned} Q\mathbf{r}_k &= Q\left(\sum_{i=1}^{k-1} (\mathbf{a}_i \cdot \mathbf{u}_i)\mathbf{e}_i + \|\mathbf{v}_k\|\mathbf{e}_k\right) \\ &= \sum_{i=1}^{k-1} (\mathbf{a}_i \cdot \mathbf{u}_i)(Q\mathbf{e}_i) + \|\mathbf{v}_k\|Q\mathbf{e}_k \\ &= \sum_{i=1}^{k-1} (\mathbf{a}_i \cdot \mathbf{u}_i)\mathbf{u}_i + \|\mathbf{v}_k\|\mathbf{u}_k \\ &= \mathbf{a}_k\,. \end{aligned}$$
On a donc démontré :

Théorème 12.45. *Soit A une matrice $m \times n$ de rang $\operatorname{rang}(A) = n$. Alors il existe*

1) *une matrice Q ($m \times n$) telle que $Q^T Q = I_n$ (c'est-à-dire dont les colonnes forment une famille orthonormale), et*
2) *une matrice R ($n \times n$) triangulaire supérieure,*

telles que
$$A = QR\,.$$
Cette factorisation est une ***factorisation** QR **de** A.*

- ⋆ Une décomposition QR n'est pas unique, puisque l'on peut toujours par exemple changer les signes des colonnes de Q, ce qui change les coefficients de R aussi mais qui ne change pas la validité de l'identité $A = QR$.
- ⋆ A priori, R s'obtient à l'aide de l'expression ci-dessus, pleine de produits scalaires. Mais on peut l'obtenir plus simplement. En effet, remarquons que si on multiplie (à gauche) les deux côtés de "$A = QR$" par Q^T,
$$Q^T A = Q^T(QR) = (Q^T Q)R = I_n R = R\,.$$

Donc la factorisation QR d'une matrice A dont les colonnes sont linéairement indépendantes peut s'obtenir comme suit :

1) Appliquer le procédé de Gram-Schmidt aux colonnes de $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$, et normaliser les vecteurs obtenus, pour obtenir $Q := [\mathbf{u}_1 \cdots \mathbf{u}_n]$.
2) Calculer $R = Q^T A$.

Exemple 12.46. Calculons une factorisation QR de
$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 2 \end{pmatrix}.$$

Ici, $A = [\mathbf{a}_1\, \mathbf{a}_2]$, où $\mathbf{a}_1$ et $\mathbf{a}_2$ sont indépendants, donc le théorème s'applique. Le procédé de Gram-Schmidt donne
$$\mathbf{v}_1 := \mathbf{a}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{v}_2 := \mathbf{a}_2 - \frac{\mathbf{a}_2 \cdot \mathbf{a}_1}{\|\mathbf{a}_1\|^2}\mathbf{a}_1 = \begin{pmatrix} -1/2 \\ 1/2 \\ 2 \end{pmatrix}.$$

On a donc, après normalisation,
$$Q = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/6 \\ \sqrt{2}/2 & \sqrt{2}/6 \\ 0 & 2\sqrt{2}/3 \end{pmatrix}$$

Ensuite,
$$\begin{aligned} R = Q^T A &= \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/6 & \sqrt{2}/6 & 2\sqrt{2}/3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 2 \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{2} & 3\sqrt{2}/2 \\ 0 & 3\sqrt{2}/2 \end{pmatrix} \end{aligned}$$

Remarquons que cette dernière est bien
$$R = \begin{pmatrix} \|\mathbf{v}_1\| & \mathbf{a}_2 \cdot \mathbf{u}_1 \\ 0 & \|\mathbf{v}_2\| \end{pmatrix},$$

où $\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$. ◇

Cas général

La décomposition QR existe même si les colonnes de la matrice ne sont pas linéairement indépendantes :

Théorème 12.47. *Soit A une matrice $m \times n$ de rang $\operatorname{rang}(A) = r$. Alors il existe*

1) *une matrice Q ($m \times r$) telle que $Q^T Q = I_r$ (c'est-à-dire dont les colonnes forment une famille orthonormale),*
2) *une matrice R ($r \times n$) triangulaire supérieure,*

telles que

$$A = QR\,.$$

La différence avec le cas précédent est que lorsqu'on calcule Q en appliquant le procédé de Gram-Schmidt, on tombe sur des colonnes nulles, que l'on peut éliminer pour ne garder qu'une base orthonormale de $\operatorname{Col}(A)$. Après, R s'obtient aussi par $R = Q^T A$.

Chapitre 13

La méthode des moindres carrés

13.1 Introduction

La méthode des *moindres carrés* (également appelée *régression linéaire*, **linear regression** (lien web), ou *least squares* en anglais) est une technique qui permet de modéliser des données expérimentales à l'aide d'un modèle linéaire optimal (dans un sens que nous préciserons). Elle est utilisée dans beaucoup de domaines, et constitue en particulier un des piliers des méthodes de base rencontrées en *machine learning*.

De notre point de vue, la méthode des moindres carrés sera une application de l'algèbre linéaire à des problèmes d'optimisation.

Avant de la décrire en toute généralité, nous allons la motiver sur un exemple simple, de petite dimension, qui nous permettra de comprendre l'idée de base, qui sera ensuite généralisée.

Celcius vs Fahrenheit ?

Supposons que l'on souhaite étudier la relation permettant de convertir les unités de mesure d'une température, de **Celcius** (notée T_C) en **Fahrenheit** (notée T_F).

On se souvient que cette relation est du type suivant :

$$(t): \qquad T_F = \alpha T_C + \beta\,,$$

mais on ne se souvient plus des valeurs de α et β.

Si on dispose de deux thermomètre, un qui mesure en Celcius, l'autre en Fahrenheit, on peut prendre des mesures et les utiliser pour essayer de retrouver les valeurs des coefficients α et β. Si ces thermomètres permettaient de faire des mesures "parfaites", il suffirait de faire deux mesures de températures assez différentes, $(T_C^{(1)}, T_F^{(1)})$ (au milieu du laboratoire par exemple) et $(T_C^{(2)}, T_F^{(2)})$ (dans le frigo par exemple), de les injecter dans (t),

$$\begin{aligned} \alpha T_C^{(1)} + \beta &= T_F^{(1)} \\ \alpha T_C^{(2)} + \beta &= T_F^{(2)}\,, \end{aligned}$$

et de résoudre ce système pour trouver α et β.

Mais on sait que des mesures empiriques ne sont par définition pas parfaites : un processus de mesure de ce genre peut contenir de multiples sources d'erreur : mauvaise calibration des appareils, minivariations de températures entre les points où la température est mesurée, imprécisions lors de la lecture de la température sur les thermomètres, etc.

Supposons pour simplifier que l'on fasse trois mesures. On les reporte dans un tableau :

T_C	T_F
2	30
12	52
65	147

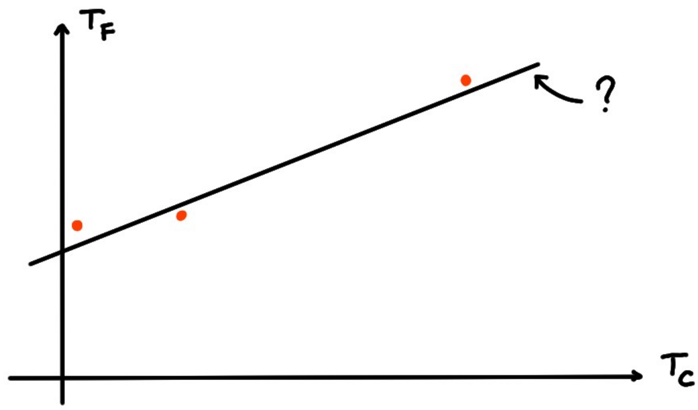
Encore une fois, comme nos mesures ne sont pas exactes, il est très peu probable que les trois points satisfassent simultanément la relation $T_F = \alpha T_C + \beta$, pour des coefficients α, β bien définis. En d'autres termes, le système

$$\begin{cases} 2\alpha & + & \beta & = & 30 \\ 12\alpha & + & \beta & = & 52 \\ 65\alpha & + & \beta & = & 147 \end{cases}$$

est incompatible.

Mais on ne doit pas pour autant abandonner la recherche de la vraie relation qui lie ces températures ! Car si des mesures expérimentales ne permettent pas de retrouver exactement une relation théorique, elles permettent néanmoins de s'en *approcher*.

D'un point de vue graphique, le problème rencontré ci-dessus peut s'exprimer comme suit : les trois paires (T_C, T_F) mesurées en laboratoire peuvent être représentées comme des points dans le plan :



Si ces points ne sont pas sur une même droite, ils doivent quand-même être *proches* de la droite théorique "$T_F = \alpha T_C + \beta$". Et on peut donc se poser la question de savoir si il est possible, à partir de nos trois mesures, de calculer une paire (α_*, β_*) qui donne une droite $T_F = \alpha_* T_C + \beta_*$ qui *approxime au mieux* ce nuage de trois points. Comment définir cette droite ?

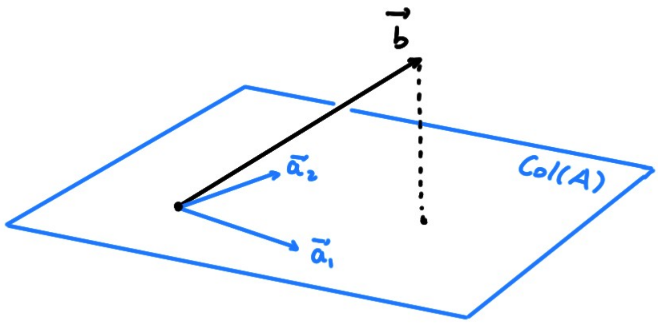
Pour répondre à cette question, utilisons le langage de l'algèbre linéaire pour formuler précisément le problème. On l'a dit, avec nos trois mesures, on est mené au système 3×2

$$\underbrace{\begin{pmatrix} 2 & 1 \\ 12 & 1 \\ 65 & 1 \end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}_{=\mathbf{x}} = \underbrace{\begin{pmatrix} 30 \\ 52 \\ 147 \end{pmatrix}}_{=\mathbf{b}},$$

qui est incompatible, et qui le sera en général dès que ces trois mesures sont faites en laboratoire. Il est utile de formuler géométriquement l'absence de solution au problème $A\mathbf{x} = \mathbf{b}$ ci-dessus, en reprenant la définition de base du produit matriciel :

$$\alpha \underbrace{\begin{pmatrix} 2 \\ 12 \\ 65 \end{pmatrix}}_{=:\mathbf{a}_1} + \beta \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}_{=:\mathbf{a}_2} = \underbrace{\begin{pmatrix} 30 \\ 52 \\ 147 \end{pmatrix}}_{=:\mathbf{b}}$$

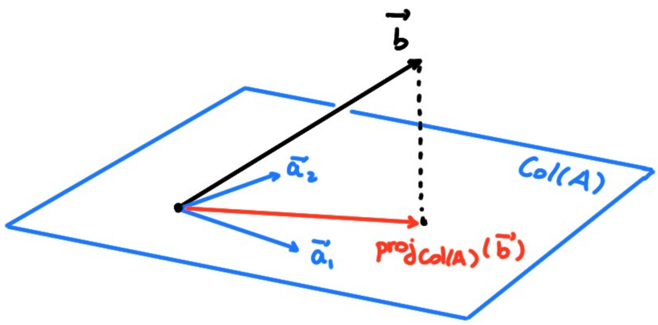
Ce système posséderait une solution (α, β) si $\mathbf{b}$ appartenait à $\mathrm{Col}(A)$, c'est-à-dire au plan engendré par $\mathbf{a}_1$ et $\mathbf{a}_2$. Mais le plus probable est que $\mathbf{b}$ ne soit *pas* dans ce plan :



Cette image suggère que malgré tout, si on ne peut pas trouver de paire telle que la combinaison linéaire $\alpha\mathbf{a}_1 + \beta\mathbf{a}_2$ soit exactement égale à $\mathbf{b}$, on pourrait chercher la paire telle que la combinaison linéaire $\alpha\mathbf{a}_1 + \beta\mathbf{a}_2$ soit *aussi proche que possible de* $\mathbf{b}$, c'est à dire la paire (α, β) qui minimise la distance

$$\|(\alpha\mathbf{a}_1 + \beta\mathbf{a}_2) - \mathbf{b}\| .$$

On sait, par les résultats démontrés dans le chapitre précédent, que la combinaison linéaire qui réalise ce minimum est précisément celle qui est égale à la projection de $\mathbf{b}$ sur l'espace engendré par $\mathbf{a}_1$ et $\mathbf{a}_2$, à savoir $\mathrm{Col}(A)$:



Pour résumer, au lieu de résoudre le système incompatible

$$(*) : \qquad A\mathbf{x} = \mathbf{b} ,$$

on cherche le $\mathbf{x}$ qui minimise la distance

$$\|A\mathbf{x} - \mathbf{b}\| ,$$

et on sait que ce $\mathbf{x}$ correspond à la solution de

$$(*)_{MC} : \qquad A\mathbf{x} = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b}) .$$

Ce deuxième système possède *toujours* une solution $\mathbf{x}$, puisque par définition, la projection $\mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b}) \in \mathrm{Col}(A)$.

Calculons donc la projection de $\mathbf{b}$ sur $W = \mathrm{Col}(A) = \mathrm{Vect}\{\mathbf{a}_1, \mathbf{a}_2\}$.

Informel 13.1. Attention, les calculs qui suivent sont simples, mais mènent à des fractions que l'on ne peut pas forcément simplifier. Pas grave, c'est la vie ! La plupart du temps, dès qu'on s'attaque à un problème venu d'une situation pratique, il apparaît toujours des nombres moins jolis que ceux qu'on est habitués à trouver dans les séries d'exercices. (Et le plus probable est que l'on implémente l'algorithme sur un ordinateur, donc on ne fera pas à la main ces calculs de fractions.)

Comme les colonnes de A ne sont pas orthogonales, on peut d'abord faire (Gram-Schmidt) :

$$\mathbf{a}_2' := \mathbf{a}_2 - \mathrm{proj}_{\mathbf{a}_1}(\mathbf{a}_2) = \begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{79}{4373}\begin{pmatrix}2\\12\\65\end{pmatrix} = \begin{pmatrix}4215/4373\\3425/4373\\-762/4373\end{pmatrix}.$$

La projection peut maintenant se calculer :

$$\begin{aligned}\mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b}) &= \frac{\mathbf{b}\cdot\mathbf{a}_1}{\|\mathbf{a}_1\|^2}\mathbf{a}_1 + \frac{\mathbf{b}\cdot\mathbf{a}_2'}{\|\mathbf{a}_2'\|^2}\mathbf{a}_2' \\ &= \frac{10239}{4373}\begin{pmatrix}2\\12\\65\end{pmatrix} + \frac{192536}{30077494}\begin{pmatrix}4215\\3425\\-762\end{pmatrix} \\ &= \begin{pmatrix}31.6644\ldots\\50.0215\ldots\\147.3140\ldots\end{pmatrix}\end{aligned}$$

Maintenant, on peut résoudre le système $A\mathbf{x} = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$:

$$\begin{pmatrix}2&1\\12&1\\65&1\end{pmatrix}\begin{pmatrix}\alpha\\\beta\end{pmatrix} = \begin{pmatrix}31.6644\ldots\\50.0215\ldots\\147.3140\ldots\end{pmatrix}$$

On trouve :

$$\begin{aligned}\alpha &= 1.8357\ldots\\ \beta &= 27.9930\ldots\end{aligned}$$

Donc nos trois mesures, et le raisonnement géométrique menant à projeter sur les colonnes de la matrice A, nous ont mené à la version suivante de la relation entre degrés Celsius et Fahrenheit :

$$T_F = 1.8357\ldots T_C + 27.9930\ldots,$$

Cette droite est celle qui approxime le mieux nos mesures, **au sens des moindre carrés** (voir la section suivante pour l'explication de cette terminologie).

Pour information, la vraie relation, que l'on trouve par exemple **ici** (lien web), est la suivante :

$$T_F = \frac{9}{5}T_C + 32 = 1.8T_C + 32\,.$$

Avec seulement trois points, notre méthode fournit donc des coefficients dont l'erreur avec la relation théorique est d'environ 2% pour α, et 13% pour β.

Informel 13.2. Bien-sûr, on obtiendrait un bien meilleur résultat en faisant beaucoup plus que trois mesures ! Si on faisait 100 mesures par exemple, l'erreur sur α et β serait bien plus petite. Pourtant, on traiterait le problème exactement de la même façon : avec 100 mesures, on devrait considérer un système incompatible

$$\underbrace{A}_{100\times 2}\underbrace{\mathbf{x}}_{\in\mathbb{R}^2} = \underbrace{\mathbf{b}}_{\in\mathbb{R}^{100}}.$$

On projetterait alors $\mathbf{b}\in\mathbb{R}^{100}$ sur le plan $\mathrm{Col}(A)\subset\mathbb{R}^{100}$, pour finalement obtenir une droite qui approxime notre nuage, formé par les 100 points des mesures faites en laboratoire.

13.2 Méthode générale

Considérons un système $m \times n$,

$$(*) : \quad A\mathbf{x} = \mathbf{b}\,,$$

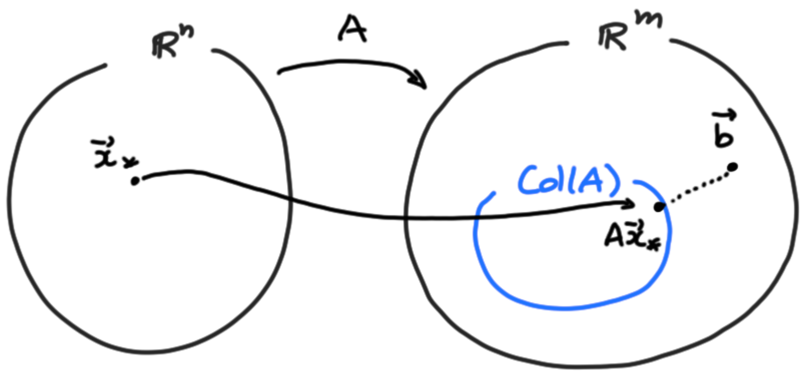
que l'on supposera incompatible, ce qui signifie

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\| > 0\,.$$

Définition 13.3. On dit que $\mathbf{x}_* \in \mathbb{R}^n$ est une **pseudo-solution de** $(*)$, ou **solution de** $(*)$ **au sens des moindres carrés** si

$$\|A\mathbf{x}_* - \mathbf{b}\| = \min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|\,.$$

Shématiquement :



Remarque 13.4. À propos de la terminologie "moindres carrés" : Trouver le $\mathbf{x}$ qui minimise une certaine norme revient au même que de trouver le $\mathbf{x}$ qui minimise le *carré* de cette norme , donc la recherche d'une pseudo-solution revient à minimiser la fonction

$$\mathbf{x} \mapsto \|A\mathbf{x} - \mathbf{b}\|^2 = \sum_{k=1}^{m} \big((A\mathbf{x})_k - b_k\big)^2\,,$$

qui est une somme de *carrés*. ◇

Exemple 13.5. (Généralisation du problème de l'exemple de motivation.) Supposons que l'on ait un nuage de points dans le plan, obtenu en prenant des mesures

$$\mathcal{P} = \{(x_1, y_1), \dots, (x_N, y_N)\}\,,$$

sensées obéir à une relation affine théorique de la forme

$$y = \alpha x + \beta\,.$$

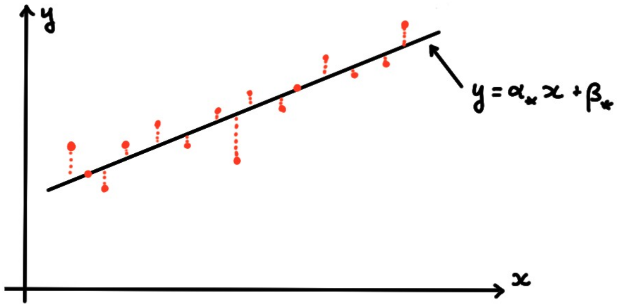
Dans ce cas, le système $N \times 2$

$$\begin{cases} \alpha x_1 & + & \beta & = & y_1 \\ \alpha x_2 & + & \beta & = & y_2 \\ & & & \vdots & \\ \alpha x_N & + & \beta & = & y_N \end{cases}$$

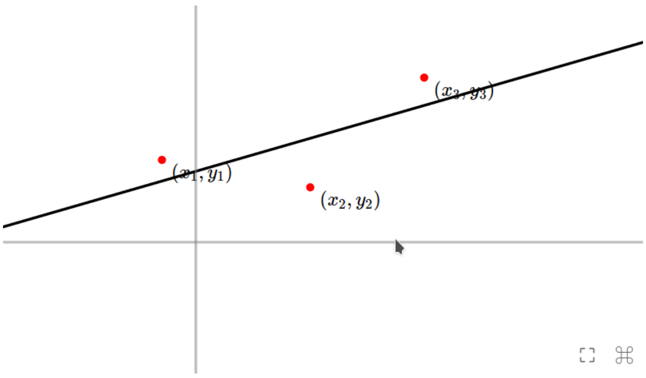
est en général incompatible, et la solution au sens des moindres carrés correspond à minimiser

$$\mathbf{x} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mapsto \sum_{k=1}^{N} \big((\alpha x_k + \beta) - y_k \big)^2 .$$

La paire (α_*, β_*) qui minimise cette fonction fournit donc une droite qui approxime le nuage $\mathcal{P}$, *au sens des moindres carrés* :



Par exemple, avec seulement $N = 3$ (comme dans l'exemple de motivation) :



Pour une animation semblable, mais fonctionnant avec un nombre arbitraire de points, cliquer **ici (Stats applets)** (lien web). ◇

L'équation normale

Considérons une pseudo-solution $\mathbf{x}_*$:

$$\|A\mathbf{x}_* - \mathbf{b}\| = \min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\| .$$

Comme on sait, considérer tous les produits $\mathbf{q} := A\mathbf{x}$ possibles, lorsque $\mathbf{x}$ varie, revient à considérer toutes les combinaisons linéaires possibles des colonnes de A, et donc à parcourir tout le sous-espace $\mathrm{Col}(A)$. Donc on peut tout aussi bien écrire

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\| = \min_{\mathbf{q} \in \mathrm{Col}(A)} \|\mathbf{q} - \mathbf{b}\| .$$

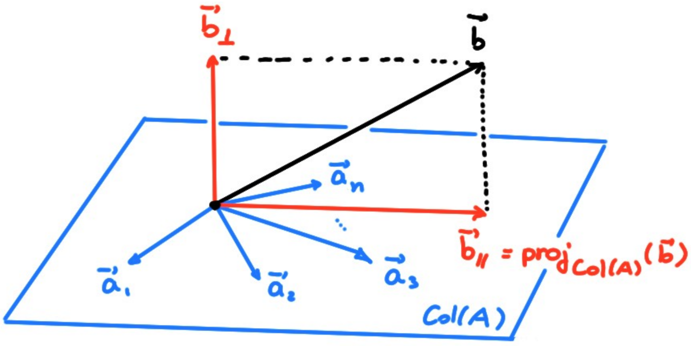
Or on a vu que le minimum de cette distance est réalisé lorsque $\mathbf{q}$ est la projection de $\mathbf{b}$ sur $\mathrm{Col}(A)$:

$$\mathbf{q}_* = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$$

On peut toujours calculer cette projection, typiquement en extrayant une base de $\mathrm{Col}(A)$, et en l'orthogonalisant avec le procédé de Gram-Schmidt. (C'est ce que nous avons fait dans l'exemple de l'introduction.)

Mais nous allons voir qu'il est possible de passer outre le calcul explicite de cette projection.

En effet, commençons par rappeler que $\mathbf{b}_{\|} = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$ est caractérisé par le fait que $\mathbf{b}_{\perp} := \mathbf{b} - \mathbf{b}_{\|}$ est orthogonal à $\mathrm{Col}(A)$, i.e. $\mathbf{b}^{\perp} \in \mathrm{Col}(A)^{\perp}$:



Or on a **montré** (lien web) précédemment que

$$\mathrm{Col}(A)^{\perp} = \ker(A^T)\,.$$

Ainsi, $\mathbf{b}^{\perp}$ doit satisfaire $A^T\mathbf{b}^{\perp} = \mathbf{0}$, qui donne

$$A^T(\mathbf{b} - \mathbf{b}_{\|}) = \mathbf{0}\,,$$

c'est-à-dire

$$A^T\mathbf{b}_{\|} = A^T\mathbf{b}\,.$$

Comme $\mathbf{b}_{\|} = A\mathbf{x}_*$, on a démontré :

Théorème 13.6. *Un vecteur $\mathbf{x} \in \mathbb{R}^n$ est solution de $A\mathbf{x} = \mathbf{b}$ au sens des moindres carrés si et seulement si $\mathbf{x}$ est solution de l'**équation normale**, donnée par*

$$A^TA\mathbf{x} = A^T\mathbf{b}\,.$$

Exemple 13.7. Considérons l'exemple de l'introduction, où le système incompatible $A\mathbf{x} = \mathbf{b}$ de départ était

$$\begin{pmatrix} 2 & 1 \\ 12 & 1 \\ 65 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 30 \\ 52 \\ 147 \end{pmatrix}\,.$$

Donnons la solution de cette équation sans passer par la projection, en utilisant le théorème ci-dessus. On obtient l'équation normale en multipliant des deux côtés par A^T, qui donne

$$\begin{pmatrix} 2 & 12 & 65 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 12 & 1 \\ 65 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 2 & 12 & 65 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 30 \\ 52 \\ 147 \end{pmatrix}\,,$$

c'est-à-dire

$$\begin{pmatrix} 4373 & 79 \\ 79 & 3 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 10239 \\ 229 \end{pmatrix}$$

La solution de ce dernier est donnée par

$$\alpha = \frac{12626}{6878} = 1.8357\dots$$
$$\beta = \frac{192536}{6878} = 27.9930\dots\,,$$

comme nous avions trouvé en utilisant la projection. ◇

Informel 13.8. Si A est $m \times n$, $A^T A$ est $n \times n$, et donc l'équation normale représente un système carré $n \times n$ qui possède *toujours* une solution.

Dans l'exemple précédent, la solution de l'équation normale était unique. Mais il peut arriver que l'équation normale possède plus d'une solution, ce que l'on aimerait éviter dans les problèmes pratiques. Voyons comment garantir l'unicité de la pseudo-solution :

Théorème 13.9. *Soit A une matrice $m \times n$. Sont équivalents :*

1) *Pour tout $\mathbf{b} \in \mathbb{R}^m$, la pseudo-solution de $A\mathbf{x} = \mathbf{b}$ est unique.*
2) *Pour tout $\mathbf{b} \in \mathbb{R}^m$, la solution de l'équation normale $A^T A\mathbf{x} = A^T\mathbf{b}$ est unique.*
3) *$A^T A$ (qui est $n \times n$) est inversible.*
4) *Les colonnes de A sont linéairement indépendantes.*

Ainsi, lorsque la pseudo-solution $\mathbf{x}_*$ du système $A\mathbf{x} = \mathbf{b}$ est unique, elle s'exprime explicitement par

$$\mathbf{x}_* = (A^T A)^{-1} A^T \mathbf{b}\,.$$

Pour la preuve des équivalences énoncées dans le théorème, nous aurons besoin du résultat préliminaire suivant :

Lemme 19. *Pour toute matrice A $(m \times n)$, $A\mathbf{x} = \mathbf{0}$ si et seulement si $A^T A\mathbf{x} = \mathbf{0}$.*

Preuve: Il est évident que si $A\mathbf{x} = \mathbf{0}$, alors $A^T A\mathbf{x} = \mathbf{0}$.

Inversément, si $A^T A\mathbf{x} = \mathbf{0}$, alors

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T(A^T A\mathbf{x}) = \mathbf{0}\,,$$

ce qui implique $\|A\mathbf{x}\| = 0$, c'est-à-dire $A\mathbf{x} = \mathbf{0}$. □

Passons maintenant à la preuve du théorème :
Preuve: $1. \Leftrightarrow 2.$: clair par ce que nous avons montré ci-dessus (un $\mathbf{x}$ est pseudo-solution si et seulement si c'est une solution de l'équation normale).

$2. \Leftrightarrow 3.$: On sait qu'un système carré $M\mathbf{x} = \mathbf{y}$ possède une unique solution pour tout $\mathbf{y}$ si et seulement si M est inversible.

$3. \Leftrightarrow 4.$: En effet, les colonnes de A sont indépendantes si et seulement si l'équation $A\mathbf{x} = \mathbf{0}$ ne possède que la solution triviale, et par le lemme ci-dessus, ceci est équivalent à dire que $A^T A\mathbf{x} = \mathbf{0}$ ne possède que la solution triviale, qui encore une fois est équivalent à dire que $A^T A$ est inversible, qui est 2. □

Exemple 13.10. Le système incompatible

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & -3 & -2 \\ 0 & 5 & 5 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

possède une infinité de pseudo-solutions. En effet, la troisième colonne de A est égale à la somme des deux premières ; par le théorème, ceci implique que la solution n'est pas unique. On conclut que le nombre de solutions est infini, par le Théorème "$0, 1, \infty$" appliqué à $A^T A\mathbf{x} = A^T\mathbf{b}$. ⋄

Plus tard, nous appliquerons la méthode des moindres carrés pour résoudre d'autres problèmes d'optimisation, inspirés de l'analyse.

13.3 Utilisation de la décomposition QR

La décomposition QR intervient dans la recherche des solutions d'un système au sens des moindres carrés.

En effet, considérons un système incompatible

$$A\mathbf{x} = \mathbf{b}\,.$$

Théorème 13.11. *Soit A une matrice $m \times n$ quelconque, et soit $A = QR$ une décomposition QR de A. Alors un vecteur $\mathbf{x} \in \mathbb{R}$ est la solution de $A\mathbf{x} = \mathbf{b}$ au sens des moindres carrés si et seulement si il est solution du système*

$$R\mathbf{x} = Q^T\mathbf{b}\,.$$

Remarque 13.12. L'avantage du système $R\mathbf{x} = Q^T\mathbf{b}$ est qu'il est *triangulaire*. ⋄

Preuve: Supposons d'abord que $\mathbf{x}$ est pseudo-solution de $A\mathbf{x} = \mathbf{b}$. On sait que cela signifie que

$$\mathbf{b} - A\mathbf{x} \in \mathrm{Col}(A)^\perp = \mathrm{Col}(Q)^\perp = \ker(Q^T)\,.$$

Dans la première égalité, on a utilisé le fait que les colonnes de Q, par définition, engendrent le même sous-espace que celles de A.

On a donc

$$Q^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}\,,$$

et comme $Q^TA = R$, cette dernière implique que $\mathbf{x}$ est solution de

$$R\mathbf{x} = Q^T\mathbf{b}\,.$$

Inversément, supposons que $\mathbf{x}$ est solution de ce dernier système, que l'on écrit plutôt

$$Q^TA\mathbf{x} = Q^T\mathbf{b}\,.$$

En multipliant des deux côtés par R^T et en utilisant $R^TQ^T = (QR)^T = A^T$, on obtient

$$A^TA\mathbf{x} = A^T\mathbf{b}\,,$$

donc $\mathbf{x}$ est solution de l'équation normale. □

Exemple 13.13. Considérons le système $A\mathbf{x} = \mathbf{b}$ incompatible suivant :

$$\begin{pmatrix} 2 & 1 \\ 2 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -5 \\ 2 \end{pmatrix}\,.$$

Puisque les colonnes de A sont indépendantes, la solution au sens des moindres carrés est unique, et on va la calculer en utilisant le théorème ci-dessus.

Le procédé de Gram-Schmidt appliqué aux colonnes de A, suivi d'une normalisation, donne

$$Q = \begin{pmatrix} 2/3 & 1/3 \\ 2/3 & -2/3 \\ 1/3 & 2/3 \end{pmatrix}\,.$$

On a donc

$$R = Q^TA = \begin{pmatrix} 3 & 1 \\ 0 & 1 \end{pmatrix}\,, \qquad \text{et} \qquad Q^T\mathbf{b} = \begin{pmatrix} -2 \\ 5 \end{pmatrix}\,.$$

Ainsi, $R\mathbf{x} = Q^T\mathbf{b}$ est le système triangulaire donné par

$$\begin{pmatrix} 3 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2 \\ 5 \end{pmatrix}\,.$$

La solution de ce dernier est $x_1 = -7/3$, $x_2 = 5$. On peut bien-sûr vérifier que cette solution est la même que celle de l'équation normale associée au système incompatible initial. ⋄

Chapitre 14

Matrices symétriques et orthogonales

14.1 Définitions

Dans ce chapitre, on ne traitera que des matrices carrées.

Définition 14.1. Une matrice A $(n \times n)$ est **symétrique** si $A^T = A$, c'est-à-dire si

$$a_{ji} = a_{ij} \qquad \forall i, j = 1, 2, \ldots, n\,.$$

Donc une matrice symétrique a ses coefficients symétriques par rapport à la diagonale.

Exemple 14.2. $\star$ La matrice identité I_n est symétrique.

$\star$ $B = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & -5 \\ 3 & -5 & 7 \end{pmatrix}$ est symétrique.

$\star$ $C = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 0 & 1 \\ 1 & 2 & 2 & 0 & 1 \\ 1 & 0 & 0 & 0 & 2 \\ 1 & 2 & 1 & 2 & 1 \end{pmatrix}$ n'est pas symétrique.

$\diamond$

Avant de commencer l'étude des propriétés remarquables des matrices symétriques, introduisons une autre classe de matrices, intimement liées (comme on le verra) aux matrices symétriques :

Définition 14.3. Une matrice A $n \times n$ est **orthogonale** si

$$A^T A = AA^T = I_n\,.$$

De par sa définition, une matrice A orthogonale est inversible, et son inverse est égale à sa transposée :

$$A^{-1} = A^T\,.$$

Aussi, on sait qu'il suffit qu'un des conditions ($A^T A = I_n$ ou $AA^T = I_n$) soit satisfaite pour garantir l'orthogonalité.

De plus, on a déjà vu que si on écrit une matrice carrée à l'aide de ses colonnes, $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$, alors on peut interpréter chaque coefficient du produit $A^T A$ comme un produit scalaire :

$$A^T A = \begin{pmatrix} \mathbf{a}_1 \cdot \mathbf{a}_1 & \mathbf{a}_1 \cdot \mathbf{a}_2 & \cdots & \mathbf{a}_1 \cdot \mathbf{a}_n \\ \mathbf{a}_2 \cdot \mathbf{a}_1 & \mathbf{a}_2 \cdot \mathbf{a}_2 & \cdots & \mathbf{a}_2 \cdot \mathbf{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_n \cdot \mathbf{a}_1 & \cdots & \cdots & \mathbf{a}_n \cdot \mathbf{a}_n \end{pmatrix}$$

Ainsi, A est orthogonale ($A^T A = I_n$) si et seulement si

$$\mathbf{a}_i \cdot \mathbf{a}_j = \begin{cases} 1 & \text{si } i = j\,, \\ 0 & \text{si } i \neq j\,, \end{cases}$$

Autrement dit :

Lemme 20. *$A = [\mathbf{a}_1 \dots \mathbf{a}_n]$ est orthogonale si et seulement si ses colonnes sont des vecteurs unitaires ($\|\mathbf{a}_k\| = 1$ pour tout $k = 1, \dots, n$) et orthogonaux deux à deux ($\mathbf{a}_i \cdot \mathbf{a}_j = 0$ si $i \neq j$).*

Il y a donc autant de matrices $n \times n$ orthogonales qu'il y a de familles orthonormales de n vecteurs dans $\mathbb{R}^n$.

Exemple 14.4. $A = \begin{pmatrix} 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ -1/\sqrt{3} & 0 & 2/\sqrt{6} \end{pmatrix}$ est orthogonale, puisque ses colonnes sont unitaires et orthogonales deux-à-deux. Par conséquent, son inverse est donné par

$$A^{-1} = A^T = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \end{pmatrix}.$$

◇

En général, dans le cas $n \times n$, on parle des matrices orthogonales comme des **rotations**, puisqu'elles représentent des transformations *rigides, qui préservent l'orthogonalité*. Nous reviendrons là-dessus.

14.2 Sur les espaces propres d'une matrice symétrique

Commençons par une propriété élémentaire du produit scalaire :

Lemme 21. *Soit B une matrice $n \times n$ quelconque. Alors pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,*

$$(B\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (B^T \mathbf{y})\,.$$

En particulier, si B est symétrique, alors

$$(B\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (B\mathbf{y})\,.$$

Preuve: Par l'interprétation matricielle du produit scalaire,

$$(B\mathbf{x}) \cdot \mathbf{y} = (B\mathbf{x})^T \mathbf{y} = \mathbf{x}^T B^T \mathbf{y} = \mathbf{x}^T (B^T \mathbf{y}) = \mathbf{x} \cdot (B^T \mathbf{y})\,.$$

□

Une conséquence immédiate :

Corollaire 7. *Soit G une matrice $n \times n$ orthogonale. Alors pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,*

- ⋆ $(G\mathbf{x}) \cdot (G\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$,
- ⋆ $\|G\mathbf{x}\| = \|\mathbf{x}\|$.

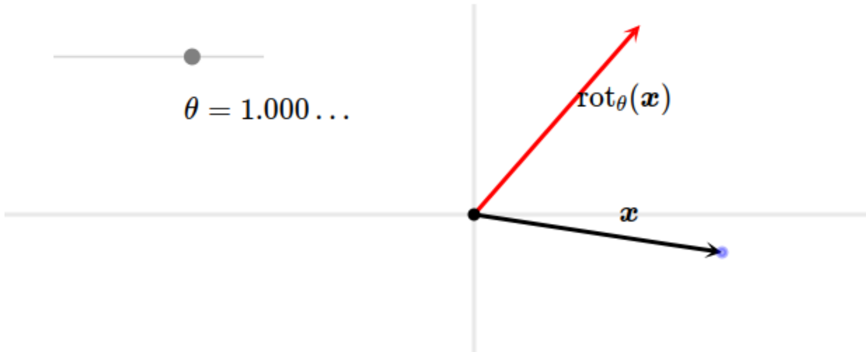
Preuve: Supposons que G est orthogonale. Par le lemme précédent,

$$(G\mathbf{x}) \cdot (G\mathbf{y}) = \mathbf{x} \cdot (G^T G\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}\,.$$

La deuxième identité s'obtient en prenant $\mathbf{y} = \mathbf{x}$. □

Remarque 14.5. La deuxième propriété montre qu'une application linéaire définie par une matrice orthogonale est une **isométrie**, c'est-à-dire qu'elle ne change pas la longueur d'un vecteur (seulement sa direction). ◇

Exemple 14.6. Un exemple typique d'isométrie est la rotation d'angle θ dans le plan :



Rappelons que relativement à la base canonique, la matrice de cette rotation est donnée par

$$[\mathrm{rot}_\theta]_{\mathcal{B}_{\mathrm{can}}} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} .$$

Cette matrice est orthogonale puisque ses colonnes sont unitaires et perpendiculaires entre elles :

$$\begin{aligned} {[\mathrm{rot}_\theta]_{\mathcal{B}_{\mathrm{can}}}}^T [\mathrm{rot}_\theta]_{\mathcal{B}_{\mathrm{can}}} &= \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \\ &= I_2 . \end{aligned}$$

◇

On sait que pour une matrice quelconque, des vecteurs propres associés à des valeurs propres distinctes sont indépendants. Pour une matrice symétrique, cette propriété est vérifiée dans un sens plus fort :

Corollaire 8. *Soit A une matrice $n \times n$ symétrique. Si $\mathbf{v}_1$ et $\mathbf{v}_2$ sont deux vecteurs propres de A associés à des valeurs propres distinctes, alors $\mathbf{v}_1 \perp \mathbf{v}_2$.*

Preuve: Si $A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1$, $A\mathbf{v}_2 = \lambda_2 \mathbf{v}_2$, alors

$$\begin{aligned} \lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) = (\lambda_1 \mathbf{v}_1) \cdot \mathbf{v}_2 = (A\mathbf{v}_1) \cdot \mathbf{v}_2 \\ = \mathbf{v}_1 \cdot (A\mathbf{v}_2) = \mathbf{v}_1 \cdot (\lambda_2 \mathbf{v}_2) = \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2) , \end{aligned}$$

qui implique $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$. Donc si $\lambda_1 \neq \lambda_2$, on a forcément que $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. □

Dans l'exemple suivant, nous vérifierons ce résultat sur un exemple concret, et nous observerons encore une propriété qui sera énoncée comme un résultat général dans la prochaine section.

Exemple 14.7. Étudions les espaces propres de la matrice symétrique

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

On calcule son polynôme caractéristique,

$$\begin{aligned} P_A(\lambda) &= \det \begin{pmatrix} 3-\lambda & -2 & 4 \\ -2 & 6-\lambda & 2 \\ 4 & 2 & 3-\lambda \end{pmatrix} \\ &= \det \begin{pmatrix} 7-\lambda & -2 & 4 \\ 0 & 6-\lambda & 2 \\ 7-\lambda & 2 & 3-\lambda \end{pmatrix} \\ &= \det \begin{pmatrix} 7-\lambda & -2 & 4 \\ 0 & 6-\lambda & 2 \\ 0 & 4 & -1-\lambda \end{pmatrix} \\ &= -(\lambda+2)(\lambda-7)^2 \,. \end{aligned}$$

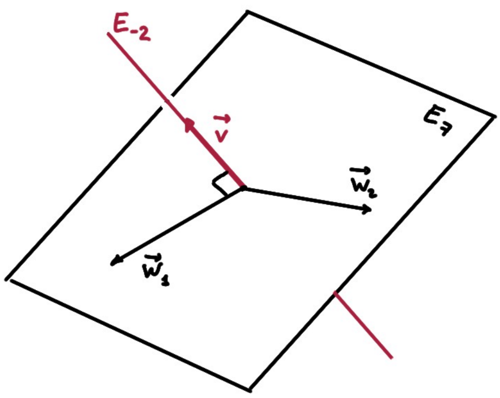
Donc A possède deux valeurs propres, $\lambda_1 = -2$ et $\lambda_2 = 7$. Les espaces propres associés se calculent facilement :

- ⋆ $E_{-2} = \mathrm{Vect}\{\mathbf{v}\}$, où $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}$,
- ⋆ $E_7 = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}$, où $\mathbf{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $\mathbf{w}_2 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}$.

On remarque qu'effectivement, $\mathbf{v} \perp \mathbf{w}_1$, et $\mathbf{v} \perp \mathbf{w}_2$, et donc n'importe quel vecteur de E_{-2} est orthogonal à n'importe quel autre vecteur de E_7. En d'autres termes :

$$E_{-2}{}^{\perp} = E_7 \,, \qquad E_7{}^{\perp} = E_{-2} \,.$$

(Pourtant, $\mathbf{w}_1$ et $\mathbf{w}_2$ ne sont pas orthogonaux entre eux.)



Remarquons aussi que

$$\sum_{k=1}^{2} \mathrm{mult}_g(\lambda_k) = 1 + 2 = 3 \,,$$

ce qui implique que A est diagonalisable. En prenant par exemple

$$D = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{pmatrix} \,, \quad P = [\mathbf{v}\, \mathbf{w}_1\, \mathbf{w}_2] = \begin{pmatrix} 1 & 1 & -1/2 \\ 1/2 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \,,$$

on obtient la diagonalisation $A = PDP^{-1}$.

Mais rappelons que l'on peut former la matrice de changement de base en choisissant les vecteurs propres que l'on veut, tant qu'ils forment une base des espaces propres concernés, et que l'on respecte l'ordre des valeurs propres dans la matrice diagonale D.

Donc on peut très bien, si on veut, commencer par orthogonaliser la base de E_7 avant de mettre en place P :

$$\begin{aligned} \mathbf{w}_1' &:= \mathbf{w}_1\,, \\ \mathbf{w}_2' &:= \mathbf{w}_2 - \mathrm{proj}_{\mathbf{w}_1}(\mathbf{w}_2) = \begin{pmatrix} -1/2 \\ 2 \\ 1/2 \end{pmatrix} \end{aligned}$$

Ainsi, une autre diagonalisation de A serait $A = QDQ^{-1}$, avec la même matrice D qu'avant, et

$$Q = [\mathbf{v}\,\mathbf{w}_1'\,\mathbf{w}_2'] = \begin{pmatrix} 1 & 1 & -1/2 \\ 1/2 & 0 & 2 \\ -1 & 1 & 1/2 \end{pmatrix}.$$

Cette fois, les colonnes de Q *sont orthogonales deux à deux*. Or rien ne nous empêche de les normaliser avant de définir Q :

$$R = \left[\frac{\mathbf{v}}{\|\mathbf{v}\|}\,\frac{\mathbf{w}_1'}{\|\mathbf{w}_1'\|}\,\frac{\mathbf{w}_2'}{\|\mathbf{w}_2'\|}\right],$$

qui donne une troisième diagonalisation de A : $A = RDR^{-1}$ (avec D la même matrice qu'avant). Mais ici, R étant orthogonale, son inverse est $R^{-1} = R^T$, et donc le changement de base devient

$$A = RDR^T.$$

On a donc pu *diagonaliser A dans une base orthonormée de $\mathbb{R}^3$*. ◇

Nous verrons, dans la section suivante, que ce que nous avons fait sur ce dernier exemple peut se faire avec *n'importe quelle matrice symétrique*.

14.3 Théorème et décomposition spectrale

Le Théorème Spectral

Un des résultats importants de l'algèbre linéaire :

Théorème 14.8. *(Théorème spectral) Soit A une matrice $n \times n$. Alors A symétrique si et seulement si elle peut se diagonaliser à l'aide d'une matrice de changement de base orthogonale.*

On dit que les matrices symétriques sont **orthogonalement diagonalisables**.
Preuve: $\Leftarrow$: Supposons que A peut se diagonaliser à l'aide d'une matrice de changement de base G orthogonale : $A = GDG^T$. Alors

$$A^T = (GDG^T)^T = (G^T)^T D^T G^T = GDG^T = A\,,$$

donc A est symétrique.

$\Rightarrow$: Donnons la preuve dans le cas $n = 2$. Soit A une matrice 2×2 symétrique, que l'on écrit comme suit :

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Montrons que A est toujours diagonalisable, quelles que soient les valeurs de $a, b, c \in \mathbb{R}$. Commençons donc par calculer les valeurs propres, à l'aide du polynôme caractéristique :

$$\begin{aligned} P_A(\lambda) &= \det \begin{pmatrix} a-\lambda & b \\ b & c-\lambda \end{pmatrix} \\ &= (a-\lambda)(c-\lambda) - b^2 \\ &= \lambda^2 - (a+c)\lambda + (ac - b^2)\,. \end{aligned}$$

Calculons le discriminant

$$\Delta = (a+c)^2 - 4(ac-b^2) = (a-c)^2 + 4b^2\,.$$

Cette dernière ligne montre que l'on a toujours $\Delta \geqslant 0$, et donc toujours au moins une valeur propre. Distinguons les cas.

1) $\Delta = (a-c)^2 + 4b^2 = 00$. Ceci signifie que $a = c$ et $b = 0$, et donc que A est en fait la matrice

$$A = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}\,,$$

qui est déjà diagonale ! On peut évidemment l'écrire comme $A = I_2 A I_2^T$.

2) $\Delta > 0$. Dans ce cas, P_A possède deux racines distinctes

$$\lambda_\pm = \frac{a + c \pm \sqrt{\Delta}}{2}\,.$$

Si on considère un vecteur propre quelconque $\mathbf{v}_+$ associé à λ_+, et un vecteur propre quelconque $\mathbf{v}_-$ associé à λ_-, on sait par le corollaire de la section précédente que $\mathbf{v}_+ \perp \mathbf{v}_-$. Ainsi, la matrice

$$G = \begin{bmatrix} \frac{\mathbf{v}_+}{\|\mathbf{v}_+\|} & \frac{\mathbf{v}_-}{\|\mathbf{v}_-\|} \end{bmatrix}$$

est orthogonale, et permet de diagonaliser A : $A = GDG^T$, où $D = \mathrm{diag}(\lambda_+, \lambda_-)$.

La preuve du cas général $n \geqslant 3$ est plus compliquée, nous ne la donnons pas ici.

□

Exemple 14.9. La matrice

$$\begin{pmatrix} 1 & \sqrt{2} & \sqrt{3} & \sqrt{5} & \sqrt{7} & \sqrt{8} & \sqrt{11} \\ \sqrt{2} & \sqrt{\pi} & \pi & \pi^2 & \pi^3 & \pi^2 & \pi \\ \sqrt{3} & \pi & -1 & e & -e & e & -e \\ \sqrt{5} & \pi^2 & e & 0 & 0 & 0 & 0 \\ \sqrt{7} & \pi^3 & -e & 0 & 1 & 1 & 1 \\ \sqrt{8} & \pi^2 & e & 0 & 1 & 2 & 2 \\ \sqrt{11} & \pi & -e & 0 & 1 & 2 & 3 \end{pmatrix}$$

étant symétrique, le Théorème Spectral s'applique : elle est diagonalisable. Il existe une matrice diagonale D et une matrice orthogonale G telles que $A = GDG^T$. ⋄

Décomposition spectrale

Voyons comment le Théorème Spectral permet de *représenter* l'application linéaire $T : \mathbb{R}^n \to \mathbb{R}^n$ associée à une matrice $n \times n$ symétrique A,

$$\mathbf{x} \mapsto T(\mathbf{x}) := A\mathbf{x}\,.$$

En effet, le Théorème Spectral garantit que A peut être diagonalisée à l'aide d'une matrice de changement de base orthogonale :

$$A = GDG^{-1} = GDG^T$$

Ici, $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ est formée de valeurs propres de A (pas forcément distinctes), et G est formée de vecteurs propres associés, formant une base orthonormale $(\mathbf{u}_1 \cdots \mathbf{u}_n)$ de $\mathbb{R}^n$:

$$G = \left[\mathbf{u}_1 \cdots \mathbf{u}_n\right],$$

On peut donc écrire, pour un $\mathbf{x} \in \mathbb{R}^n$ quelconque,

$$\begin{aligned} T(\mathbf{x}) = A\mathbf{x} = GDG^T\mathbf{x} &= \left[\mathbf{u}_1 \cdots \mathbf{u}_n\right] D \begin{pmatrix} \mathbf{u}_1^T\mathbf{x} \\ \vdots \\ \mathbf{u}_n^T\mathbf{x} \end{pmatrix} \\ &= \left[\lambda_1\mathbf{u}_1 \cdots \lambda_n\mathbf{u}_n\right] \begin{pmatrix} \mathbf{u}_1^T\mathbf{x} \\ \vdots \\ \mathbf{u}_n^T\mathbf{x} \end{pmatrix} \\ &= \sum_{k=1}^n \lambda_k \mathbf{u}_k \mathbf{u}_k^T \mathbf{x} \\ &= \left(\sum_{k=1}^n \lambda_k \mathbf{u}_k \mathbf{u}_k^T \right) \mathbf{x}\,. \end{aligned}$$

On peut donc écrire A comme une combinaison linéaire de matrices :

$$\boxed{A = \sum_{k=1}^n \lambda_k \mathbf{u}_k \mathbf{u}_k^T\,.}$$

On sait que chaque $\mathbf{u}_k\mathbf{u}_k^T$ est une matrice $n \times n$, et représente le projecteur sur $\mathbf{u}_k$. En effet, puisque chaque $\mathbf{u}_k$ est unitaire,

$$\mathbf{u}_k\mathbf{u}_k^T\mathbf{x} = (\mathbf{x}\cdot\mathbf{u}_k)\mathbf{u}_k = \frac{\mathbf{x}\cdot\mathbf{u}_k}{\|\mathbf{u}_k\|^2}\mathbf{u}_k = \mathrm{proj}_{\mathbf{u}_k}(\mathbf{x})\,.$$

On a donc pu récrire l'applications linéaire T comme la combinaison linéaire de projecteurs :

$$\boxed{T = \sum_{k=1}^n \lambda_k \mathrm{proj}_{\mathbf{u}_k}\,.}$$

Définition 14.10. Les représentations de A et T à l'aide de projecteurs sur les espaces propres de A sont appelées **décompositions spectrales**.

Remarque 14.11. La représentation spectrale dépend bien-sûr du choix des vecteurs propres pour la matrice ; elle n'est donc pas unique. ◇

Une décomposition spectrale fournit une interprétation très *géométrique* de comment A agit sur un vecteur $\mathbf{x}$.

En effet, l'expression

$$A\mathbf{x} = \sum_{k=1}^n \lambda_k \mathrm{proj}_{\mathbf{u}_k}(\mathbf{x})$$

montre que $A\mathbf{x}$ est une somme vectorielle, dans laquelle chaque terme, $\lambda_k\mathrm{proj}_{\mathbf{u}_k}(\mathbf{x})$, a une interprétation très claire :

1) Projeter $\mathbf{x}$ sur $\mathbf{u}_k$.

2) Amplifier cette projection par la valeur propre λ_k.

L'intérêt est que l'on peut travailler indépendamment pour chaque $k = 1, 2, \ldots, n$, puis les sommer.

Voyons comment réaliser concrètement cette décomposition, dans des cas particuliers.

Exemple 14.12. Considérons la matrice symétrique

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

On vérifie facilement que les valeurs propres de A sont $\lambda_1 = -1$, $\lambda_2 = 3$, et que leurs espaces propres associés sont

- ⋆ $E_{-1} = \mathrm{Vect}\{\mathbf{v}\}$, où $\mathbf{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.
- ⋆ $E_3 = \mathrm{Vect}\{\mathbf{w}\}$, où $\mathbf{w} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

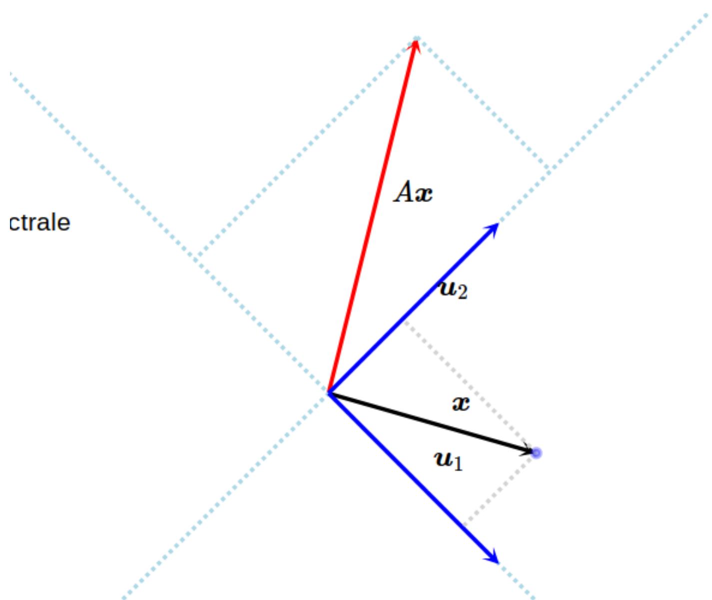
(On observe à nouveau, comme on sait, que ces espaces sont orthogonaux.) Pour faire la décomposition spectrale, on a besoin de vecteurs propres unitaires. On peut par exemple prendre

$$\mathbf{u}_1 := \frac{\mathbf{v}}{\|\mathbf{v}\|} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} , \qquad \mathbf{u}_2 := \frac{\mathbf{w}}{\|\mathbf{w}\|} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} .$$

Maintenant, la décomposition spectrale de A est donnée par

$$\begin{aligned} A\mathbf{x} &= \sum_{k=1}^{2} \lambda_k \mathbf{u}_k \mathbf{u}_k^T \mathbf{x} \\ &= (-1)\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T \\ &= (-1)\mathrm{proj}_{\mathbf{u}_1}(\mathbf{x}) + 3\mathrm{proj}_{\mathbf{u}_2}(\mathbf{x}) . \end{aligned}$$

L'interprétation géométrique de la transformation $\mathbf{x} \mapsto A\mathbf{x}$ devient limpide :



Vérifions encore, pourquoi pas :

$$\begin{aligned}
&(-1)\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T = \\
&\quad = (-1)\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} + 3\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \\
&\quad = (-1)\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} + 3\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \\
&\quad = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} = A\,.
\end{aligned}$$

◇

Exemple 14.13. Considérons la matrice symétrique déjà étudiée plus haut :

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

A possède deux valeurs propres, $\lambda_1 = -2$ et $\lambda_2 = 7$, et nous avions appliqué le procédé de Gram-Schmidt pour trouver

- ⋆ $E_{-2} = \mathrm{Vect}\{\mathbf{v}\}$, où $\mathbf{v} = \begin{pmatrix} 1 \\ 1/2 \\ -1 \end{pmatrix}$,
- ⋆ $E_7 = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}$, où $\mathbf{w}_1' = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $\mathbf{w}_2' = \begin{pmatrix} -1/2 \\ 2 \\ 1/2 \end{pmatrix}$.

En normalisant ces vecteurs,

$$\mathbf{u}_1 := \frac{\mathbf{v}}{\|\mathbf{v}\|}\,, \qquad \mathbf{u}_2 := \frac{\mathbf{w}_1'}{\|\mathbf{w}_1'\|} \qquad \mathbf{u}_3 := \frac{\mathbf{w}_2'}{\|\mathbf{w}_2'\|}$$

on a obtenu une matrice de changement de base, orthogonale,

$$R = [\mathbf{u}_1\,\mathbf{u}_2\,\mathbf{u}_3]\ ,$$

qui donne $A = RDR^T$.

Donc la décomposition spectrale obtenue est

$$A = (-2)\mathbf{u}_1\mathbf{u}_1^T + 7\mathbf{u}_2\mathbf{u}_2^T + 7\mathbf{u}_3\mathbf{u}_3^T\,.$$

◇

Chapitre 15

La décomposition en valeurs singulières

15.1 Introduction

“Today, singular value decomposition has spread through many branches of science, in particular psychology and sociology, climate and atmospheric science, and astronomy. It is also extremely useful in machine learning and in both descriptive and predictive statistics. ”

Peter Mills (lien web)

“Eigenvalues and eigenvectors are restricted to square matrices. But data comes in rectangular matrices. ”

Gilbert Strang (lien web)

Si la diagonalisation a permis de comprendre la nature géométrique de certaines applications linéaires, elle exige malheureusement que l'application considérée se prête à cette analyse (qu'elle soit *diagonalisable* justement), et surtout : elle ne s'applique qu'à des matrices carrées.

Le résultat

La décomposition en valeurs singulières (SVD=Singular Value Decomposition) est une méthode très générale de factorisation qui donne une nouvelle interprétation géométrique de n'importe quelle application linéaire $T : \mathbb{R}^n \to \mathbb{R}^m$. Elle consiste à factoriser une matrice quelconque A $(m \times n)$ en un produit,

$$A = U\Sigma V^T\,,$$

où

1) U est $m \times m$, orthogonale : $U^TU = UU^T = I_m$,
2) Σ est $m \times n$, **diagonale** dans le sens où $\Sigma_{ij} = 0$ si $i \neq j$. De plus, $\Sigma_{ij} \geqslant 0$.
3) V est $n \times n$, orthogonale : $V^TV = VV^T = I_n$.

On sait d'une part que les matrices orthogonales représentent des *isométries*, c'est-à-dire des transformations rigides de l'espace, et doivent être comprises essentiellement comme des *rotations*. D'autre part, une matrice diagonale $m \times n$ a pour effet de stretcher certaines directions (avec un changement de dimension, voir plus bas). Donc la décomposition en valeurs singulières permet de décomposer l'application $T : \mathbb{R}^n \to \mathbb{R}^m$ définie par $T(\mathbf{x}) = A\mathbf{x}$ en trois parties :

$$\mathbb{R}^n \xrightarrow[]{\substack{V^T\\ \text{(rotation)}}} \mathbb{R}^n \xrightarrow[]{\substack{\Sigma\\ \text{(stretch)}}} \mathbb{R}^m \xrightarrow[]{\substack{U\\ \text{(rotation)}}} \mathbb{R}^m$$

Il est important d'insister sur le fait que la décomposition en valeurs singulières ne suppose rien sur A ; elle est toujours possible. En particulier, elle s'applique à des matrices qui ne sont pas forcément carrées.

Structure

Dans la section suivante, nous établirons rigoureusement la décomposition en valeurs singulières. Pour l'instant, supposons qu'une décomposition

$$A = U\Sigma V^T$$

soit donnée, et voyons ce que cela dit déjà sur les matrices U, Σ et V.

Nommons les colonnes de V, U et Σ :

$$\begin{aligned} U &= [\mathbf{u}_1 \cdots \mathbf{u}_m] \qquad \mathbf{u}_k \in \mathbb{R}^m \\ \Sigma &= [\sigma_1 \cdots \sigma_n] \qquad \sigma_i \in \mathbb{R}^m \\ V &= [\mathbf{v}_1 \cdots \mathbf{v}_n] \qquad \mathbf{v}_j \in \mathbb{R}^n \,. \end{aligned}$$

Puisque U et V sont orthogonales, les familles $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subset \mathbb{R}^m$ et $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbb{R}^n$ sont orthogonales.

La matrice Σ représente une application $\mathbb{R}^n \to \mathbb{R}^m$ dont la simplicité rappelle celle des matrices diagonales *carrées*. Nous noterons σ_i les éléments diagonaux de Σ. Notons que si $m > n$ (resp. $m < n$), alors certaines lignes (resp. colonnes) de Σ sont nulles.

Exemple 15.1. Si $m = 7$ et $n = 4$, alors les 3 dernières lignes de Σ sont nulles :

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & 0 & \sigma_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Notons $\mathcal{B}_{\text{can}}$ et $\mathcal{B}'_{\text{can}}$ les bases canoniques de $\mathbb{R}^4$ et $\mathbb{R}^7$:

$$\begin{aligned} \mathcal{B}_{\text{can}} &= (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4)\,, \\ \mathcal{B}'_{\text{can}} &= (\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3, \mathbf{e}'_4, \mathbf{e}'_5, \mathbf{e}'_6, \mathbf{e}'_7)\,. \end{aligned}$$

L'application $\mathbf{x} \mapsto \Sigma\mathbf{x}$ représente des "stretches" pour les 4 vecteurs de $\mathcal{B}_{\text{can}}$,

$$\Sigma\mathbf{e}_1 = \sigma_1\mathbf{e}'_1\,, \quad \Sigma\mathbf{e}_2 = \sigma_2\mathbf{e}'_2\,, \quad \Sigma\mathbf{e}_3 = \sigma_3\mathbf{e}'_3\,, \quad \Sigma\mathbf{e}_4 = \sigma_4\mathbf{e}'_4\,.$$

◇

Exemple 15.2. Si $m = 3$, $n = 5$, alors les 2 dernières colonnes de Σ sont nulles :

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$$

Notons $\mathcal{B}_{\text{can}}$ et $\mathcal{B}'_{\text{can}}$ les bases canoniques de $\mathbb{R}^5$ et $\mathbb{R}^3$:

$$\begin{aligned} \mathcal{B}_{\text{can}} &= (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5)\,, \\ \mathcal{B}'_{\text{can}} &= (\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3)\,. \end{aligned}$$

On a des “stretches” pour les 3 premiers vecteurs de $\mathcal{B}_{\text{can}}$,

$$\Sigma\mathbf{e}_1 = \sigma_1\mathbf{e}_1'\,, \quad \Sigma\mathbf{e}_2 = \sigma_2\mathbf{e}_2'\,, \quad \Sigma\mathbf{e}_3 = \sigma_3\mathbf{e}_3'\,,$$

mais les deux derniers sont tous envoyés sur le vecteur nul :

$$\Sigma\mathbf{e}_4 = \Sigma\mathbf{e}_5 = \mathbf{0}\,.$$

◇

Remarque 15.3. Les relations ci-dessus, “$\Sigma\mathbf{e}_j = \sigma_j\mathbf{e}_j'$”, rappellent celles du type “$A\mathbf{v} = \lambda\mathbf{v}$”. La grande différence ici est que $\mathbf{e}_j$ et $\mathbf{e}_j'$ vivent dans des espaces différents ! ◇

Pour comprendre les relations entre les $\mathbf{u}_k$, les $\mathbf{v}_j$ et la matrice Σ (toujours en supposant que la décomposition $A = U\Sigma V^T$ est déjà connue), on multiplie A par sa transposée pour obtenir une matrice $n \times n$ donnée par

$$\begin{aligned} A^T A &= (U\Sigma V^T)^T(U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ &= V(\Sigma^T\Sigma)V^T \end{aligned}$$

On a, dans le terme de droite, trois matrices $n \times n$. Puisque $\Sigma^T\Sigma$ est diagonale, et puisque V^T est l'inverse de V (car cette dernière est orthogonale), on voit que ce produit de trois matrices carrées représente une *diagonalisation de la matrice symétrique* A^TA. En particulier, **les colonnes de V sont des vecteurs propres orthonormés de A^TA**, associés à des valeurs propres qui sont les éléments diagonaux de $\Sigma^T\Sigma$, à savoir σ_i^2 :

$$(A^TA)\mathbf{v}_j = \sigma_j^2\mathbf{v}_j \qquad \forall j = 1, \dots, n\,.$$

De même pour AA^T : c'est une matrice $m \times m$, et

$$AA^T = U(\Sigma\Sigma^T)U^T\,,$$

qui implique que **les colonnes de U sont des vecteurs propres orthonormés de AA^T**, associés à des valeurs propres qui sont les éléments diagonaux de $\Sigma\Sigma^T$:

$$AA^T\mathbf{u}_k = \sigma_k^2\mathbf{u}_k \qquad \forall k = 1, \dots, m\,.$$

Remarquons encore que dans les deux cas, les matrices diagonales $\Sigma^T\Sigma$ et $\Sigma\Sigma^T$ ont des coefficients diagonaux donnés par les carrés σ_i^2.

Cette discussion montre que si une décomposition en valeurs singulières existe, alors les matrices U et V se calculent en diagonalisant AA^T et A^TA. (On verra comment simplifier un peu ce procédé par la suite.)

Ce qui n'est pas du tout démontré par l'argument ci-dessus, c'est si la décomposition existe effectivement ; nous le démontrerons dans la section suivante.

Matrices définies par blocs

Dans ce chapitre, nous définirons et manipulerons des matrices définies **par blocs**, ce qui signifie définies comme composées de sous-matrices. On utilisera l'indice “□” pour indiquer qu'une matrice est définie par blocs.

Exemple 15.4. Avec

$$A = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

on peut définir

$$\left[\begin{array}{cc} A & B \end{array}\right]_{\square} = \begin{pmatrix} a & b & c & 1 & 2 \\ d & e & f & 3 & 4 \end{pmatrix}.$$

◇

Les blocs qui composent une matrice par blocs doivent avoir des dimensions compatibles.

Plus généralement, si on possède quatre matrices,

$$A\,(m \times k), \quad B\,(m \times l), \quad C\,(h \times k), \quad D\,(h \times l),$$

on peut définir

- ⋆ la matrice $m \times (k+l)$:
$$\left[\begin{array}{cc} A & B \end{array}\right]_{\square},$$
- ⋆ la matrice $(m+h) \times k$:
$$\left[\begin{array}{c} A \\ C \end{array}\right]_{\square}.$$
- ⋆ la matrice $(m+h) \times (k+l)$:
$$\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]_{\square}.$$

15.2 Existence

Dans cette section, on montre que toute matrice possède une décomposition en valeurs singullières :

Théorème 15.5. *(Existence d'une SVD) Toute matrice A $(m \times n)$ peut s'écrire comme un produit,*

$$A = U\Sigma V^T,$$

où

1) *U est $m \times m$, orthogonale : $U^TU = UU^T = I_m$; ses colonnes sont appelées **left singular vectors** de A.*
2) *Σ est $m \times n$, diagonale, les coefficients situés sur sa diagonale sont $\geqslant 0$, et sont appelés les **valeurs singulières** de A.*
3) *V est $n \times n$, orthogonale : $V^TV = VV^T = I_n$, ses colonnes sont appelées **right singular vectors** de A.*

Les matrices A^TA et AA^T

Notre point de départ :

Lemme 22. *Pour une matrice A $(m \times n)$ quelconque,*

- ⋆ *A^TA $(n \times n)$ est symétrique,*
- ⋆ *AA^T $(m \times m)$ est symétrique.*

Preuve: Par les propriétés de la transposée,

$$(A^TA)^T = A^T(A^T)^T = A^TA\,,$$
$$(AA^T)^T = (A^T)^TA^T = AA^T\,.$$

□

Exemple 15.6. Si $A = \begin{pmatrix} -1 & 2 \\ 2 & 0 \\ 3 & 5 \end{pmatrix}$, alors

$$A^TA = \begin{pmatrix} -1 & 2 & 3 \\ 2 & 0 & 5 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 2 & 0 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 14 & 13 \\ 13 & 29 \end{pmatrix}\,,$$

et

$$AA^T = \begin{pmatrix} -1 & 2 \\ 2 & 0 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} -1 & 2 & 3 \\ 2 & 0 & 5 \end{pmatrix} = \begin{pmatrix} 5 & -2 & 7 \\ -2 & 4 & 6 \\ 7 & 6 & 34 \end{pmatrix}$$

◇

Étant symétriques, le **Théorème Spectral** (lien web) garantit que A^TA et AA^T sont diagonalisables :

⋆ Il existe une matrice orthogonale V $(n \times n)$ et une matrice diagonale D $(n \times n)$ telle que
$$A^TA = VDV^T\,.$$

⋆ Il existe une matrice orthogonale U $(m \times m)$ et une matrice diagonale D' $(m \times m)$ telle que
$$AA^T = UD'U^T\,.$$

On sait que les éléments diagonaux de D (resp. D') sont les valeurs propres de A^TA (resp. AA^T), avec éventuellement des répétitions selon les dimensions des espaces propres associés. Or ces valeurs propres ont des propriétés particulières :

Lemme 23. *Pour toute matrice A,*

1) *Un scalaire $\lambda \neq 0$ est valeur propre de A^TA si et seulement si il est également valeur propre de AA^T.*
2) *Si λ est valeur propre de A^TA ou de AA^T, alors $\lambda \geqslant 0$.*

Preuve: 1. Supposons que $\lambda \neq 0$ est valeur propre de A^TA. Alors il existe $\mathbf{v} \in \mathbb{R}^n$, non-nul, tel que

$$(A^TA)\mathbf{v} = \lambda\mathbf{v}\,.$$

Remarquons que $A\mathbf{v} \neq \mathbf{0}$ puisque $A^TA\mathbf{v} \neq \mathbf{0}$.
Ensuite, en multipliant les deux côtés de l'identité de dessus par A, on obtient

$$AA^T(A\mathbf{v}) = \lambda(A\mathbf{v})\,,$$

qui signifie que λ est aussi valeur propre de AA^T, associée au vecteur propre $A\mathbf{v} \in \mathbb{R}^m$ (qui est non-nul comme on a dit). Le même argument montre que toute valeur propre non-nulle de AA^T est également valeur propre de A^TA.

2. Maintenant avec une valeur propre λ de A^TA, et un vecteur propre $\mathbf{v} \neq \mathbf{0}$, $A^TA\mathbf{v} = \lambda\mathbf{v}$, on peut écrire

$$\begin{aligned} \lambda\|\mathbf{v}\|^2 = \lambda(\mathbf{v}\cdot\mathbf{v}) &= \mathbf{v}\cdot(\lambda\mathbf{v}) \\ &= \mathbf{v}\cdot(A^TA\mathbf{v}) \\ &= (A\mathbf{v})\cdot(A\mathbf{v}) \\ &= \|A\mathbf{v}\|^2 \geqslant 0\,. \end{aligned}$$

Puisque $\|\mathbf{v}\| > 0$, on en déduit que $\lambda \geqslant 0$. □

Les deux lemmes ci-dessus impliquent :

Corollaire 9. *Pour toute matrice A ($m \times n$), il existe au plus $\min\{m, n\}$ valeurs propres non-nulles communes de A^TA et AA^T.*

Preuve: On sait que A^TA est $n \times n$ et possède donc au maximum n valeurs propres non-nulles, et AA^T est $m \times m$ et possède donc au maximum m valeurs propres non-nulles. Comme ces matrices ont les mêmes valeurs propres non-nulles, le nombre de ces valeurs propres non-nulles est plus petit que n et que m. □

Preuve du théorème :

(La preuve ci-dessous est tirée de **wikipedia** (lien web).) Considérons la diagonalisation de A^TA :

$$A^TA = VDV^T .$$

En multipliant à gauche par V^T puis à droite par V,

$$V^T(A^TA)V = D .$$

Par le lemme ci-dessus, toutes les valeurs propres de A^TA, sur la diagonale de D, sont $\geqslant 0$.

Sans perte de généralité, on peut supposer que la valeur propre nulle (éventuellement répétée) apparaît en bas de la diagonale :

$$D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_l, 0, \dots, 0) ,$$

avec $\lambda_1 > 0, \dots, \lambda_l > 0$. Distinguons ensuite la sous-matrice diagonale de D qui contient les valeurs propres strictement positives, en écrivant :

$$D = \begin{bmatrix} D_* & 0 \\ 0 & 0 \end{bmatrix}_{\square} ,$$

où $D_* = \operatorname{diag}(\lambda_1, \dots, \lambda_l)$ est $l \times l$, et les "0" sont des matrices nulles.

À l'ordre fixé par les valeurs propres dans D correspond un ordre des colonnes dans la matrice de changement de base V :

$$V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}_{\square} ,$$

où

- ⋆ V_1 est une matrice $n \times l$ dont les colonnes forment une famille libre de vecteurs propres associés aux valeurs propres non-nulles $\lambda_1, \dots, \lambda_l$.
- ⋆ V_2 est une matrice $n \times (n - l)$ dont les colonnes forment une famille libre de vecteurs propres associés à la valeur propre $\lambda = 0$.

L'orthonormalité des colonnes de V implique que

$$V_1^TV_1 = I_l , \qquad V_2^TV_2 = I_{n-l} ,$$

mais la relation $VV^T = I_n$ implique aussi que

$$I_n = VV^T = \begin{bmatrix} V_1 & V_2 \end{bmatrix}_{\square} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}_{\square} = V_1V_1^T + V_2V_2^T .$$

Utilisons ces matrices V_1 et V_2 pour récrire la diagonalisation $A^TA = VDV^T$, qui est équivalente à $V^T(A^TA)V = D$. Comme

$$\begin{aligned}
V^T(A^TA)V &= \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}_{\square} A^TA \begin{bmatrix} V_1 & V_2 \end{bmatrix}_{\square} \\
&= \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}_{\square} \begin{bmatrix} A^TAV_1 & A^TAV_2 \end{bmatrix}_{\square} \\
&= \begin{bmatrix} V_1^TA^TAV_1 & V_1^TA^TAV_2 \\ V_2^TA^TAV_1 & V_2^TA^TAV_2 \end{bmatrix}_{\square},
\end{aligned}$$

et comme cette matrice est égale à

$$D = \begin{bmatrix} D_* & 0 \\ 0 & 0 \end{bmatrix}_{\square},$$

ceci implique que

$$V_1^TA^TAV_1 = D_* \quad (l \times l)$$

et

$$V_2^TA^TAV_2 = 0 \quad ((n-l) \times (n-l))$$

De cette dernière, on tire que $(AV_2)^T(AV_2) = 0$, qui implique que

$$AV_2 = 0\,.$$

(En effet, on sait que pour toute matrice M, M^TM contient tous les produits scalaires possibles entre les colonnes de M, en particulier, sur sa diagonale, les carrés des normes des colonnes. Si $M^TM = 0$, cela implique que la norme de chaque colonne de M est nulle, et donc que M est la matrice nulle.)

Définissons maintenant la matrice $m \times l$:

$$U_1 := AV_1D_*^{-1/2}\,,$$

où

$$D_*^{-1/2} := \mathrm{diag}(1/\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_l})$$

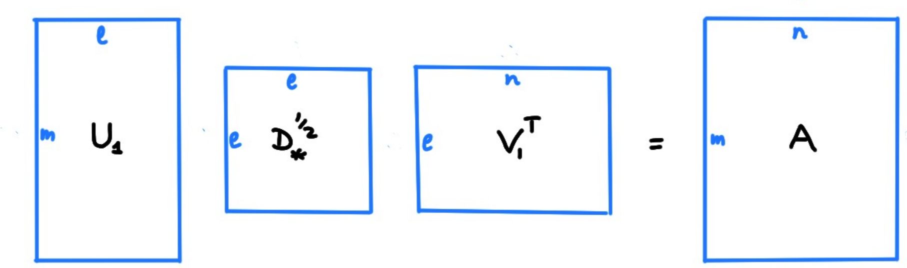
est bien définie puisque $\lambda_k > 0$ pour tout $k = 1, \dots, l$, et n'est rien d'autre que l'inverse de

$$D_*^{1/2} := \mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_l}).$$

On remarque maintenant que les colonnes de U_1 forment une famille orthonormale, puisque

$$\begin{aligned}
U_1^TU_1 &= (AV_1D_*^{-1/2})^TAV_1D_*^{-1/2} \\
&= D_*^{-1/2}(V_1^TA^TAV_1)D_*^{-1/2} \\
&= D_*^{-1/2}D_*D_*^{-1/2} \\
&= I_l\,.
\end{aligned}$$

Montrons que U_1, D_* et V_1 fournissent déjà une première factorisation de A :



En effet,

$$
\begin{aligned}
U_1 D_*^{1/2} V_1^T &= A V_1 \underbrace{D_*^{-1/2} D_*^{1/2}}_{=I_l} V_1^T \\
&= A V_1 V_1^T \\
&= A(I_n - V_2 V_2^T) \\
&= A - \underbrace{(A V_2)}_{=0} V_2^T \\
&= A
\end{aligned}
$$

Cette première factorisation constitue la base de l'argument ; il reste maintenant à modifier le produit $U_1 D_*^{1/2} V_1^T$, en augmentant les tailles des matrices, de façon à ce qu'il devienne $U\Sigma V^T$.

On rajoute d'abord à V_1^T le bloc V_2^T, ce qui donne

$$
V^T = \left[\begin{array}{c} V_1^T \\ V_2^T \end{array} \right]_\square
$$

Passons à U. Si $l = m$, alors on peut prendre $U = U_1$. Mais, si $l < m$, U_1 n'est pas carrées : ses colonnes forment une base orthonormée de $\mathrm{Col}(U_1) \subset \mathbb{R}^m$, mais pas de $\mathbb{R}^m$, on peut donc compléter cette base en une base de $\mathbb{R}^m$, et même, via un procédé de Gram-Schmidt si nécessaire, la compléter en une base orthonormée de $\mathbb{R}^n$. Les $m-l$ vecteurs rajoutés peuvent être rangés dans une matrice U_2, $m \times (m-l)$, qui permet de définir la matrice $m \times m$ orthogonale

$$
U := \left[\begin{array}{cc} U_1 & U_2 \end{array} \right]_\square .
$$

Finalement, Σ ($m \times n$) est construite à partir de $D_*^{1/2}$ ($l \times l$), en rajoutant des blocs nuls, si nécessaire (rappelons que $l \leqslant \min\{m, n\}$) :

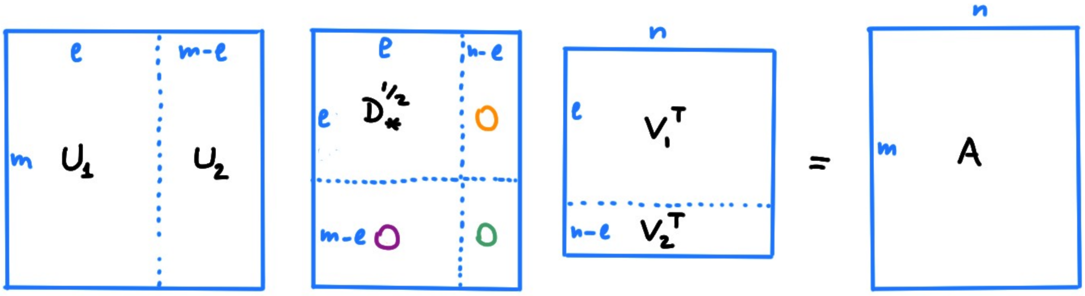
$$
\Sigma := \left[\begin{array}{cc} D_*^{1/2} & 0 \\ 0 & 0 \end{array} \right]_\square .
$$

où

- ⋆ 0 est $l \times (n-l)$,
- ⋆ 0 est $(m-l) \times l$,
- ⋆ 0 est $(m-l) \times (n-l)$.

Remarquons que ceci peut a priori faire apparaître des valeurs singulières nulles sur la diagonale de Σ.

Montrons que l'on a ce qu'on voulait :



En effet,

$$
\begin{aligned}
U\Sigma V^T &= \begin{bmatrix} U_1 & U_2 \end{bmatrix}_{\square} \begin{bmatrix} D_*^{1/2} & 0 \\ 0 & 0 \end{bmatrix}_{\square} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}_{\square} \\
&= \begin{bmatrix} U_1 & U_2 \end{bmatrix}_{\square} \begin{bmatrix} D_*^{1/2} V_1^T \\ 0 \end{bmatrix}_{\square} \\
&= U_1 D_*^{1/2} V_1^T \\
&= A\,.
\end{aligned}
$$

Ceci termine la preuve de l'existence d'une SVD.

Quelques remarques au vu de la preuve que l'on vient de donner :

- ⋆ Les valeurs singulières non-nulles de A sont **les racines carrées des valeurs propres de** A^TA (et AA^T) :
$$\sigma_j = \sqrt{\lambda_j}\,.$$
Les valeurs singulières nulles sont possibles (elles apparaissent au moment où on complète $D_*^{1/2}$ avec des blocs de zéros), et sont liées au fait que A^TA ou AA^T peuvent posséder $\lambda = 0$ comme valeur propre.
- ⋆ On l'a dit, les colonnes de V forment une base orthonormée de $\mathbb{R}^n$, formée de vecteurs propres de A^TA, et les colonnes de U forment une base orthonormale de $\mathbb{R}^m$, formée de vecteurs propres de AA^T.
- ⋆ La définition du bloc $U_1 = [\mathbf{u}_1 \cdots \mathbf{u}_l]$ peut aussi s'exprimer comme suit :
$$
\begin{aligned}
U_1 &= AV_1 D_*^{-1/2} \\
&= [A\mathbf{v}_1 \cdots A\mathbf{v}_l] D_*^{-1/2} \\
&= [\tfrac{1}{\sqrt{\lambda_1}} A\mathbf{v}_1 \cdots \tfrac{1}{\sqrt{\lambda_l}} A\mathbf{v}_l]\,.
\end{aligned}
$$
On a donc toujours un moyen direct de calculer les l premières colonnes de U_1 :
$$\mathbf{u}_j := \frac{1}{\sqrt{\lambda_j}} A\mathbf{v}_j\,, \qquad j = 1, 2, \dots, l\,,$$
où $\mathbf{v}_j$ est la j-ème colonne de V_1, correspondant au vecteur propre orthonormé de A^TA, associé à la j-ème valeur propre $\lambda_j > 0$.
- ⋆ Comme la décomposition QR, la décomposition en valeurs singulières existe toujours, mais n'est pas unique. En effet, le choix des vecteurs propres, dans la construction de V, peut toujours se faire de multiples façons, menant à autant de SVD différentes.

15.3 Exemples

La preuve de la section précédente a montré clairement quelles sont les étapes menant à une décomposition singulière d'une matrice A :

1) Diagonaliser A^TA : distinguer ses valeurs propres positives, $\lambda_1 > 0, \dots, \lambda_l > 0$, qui donnent les valeurs singulières de A
$$\sigma_j = \sqrt{\lambda_j}\,, \qquad j = 1, 2, \dots, l\,.$$
Les vecteurs propres de A^TA donnent $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$, les $n - l$ dernières colonnes étant associées à la valeur propre nulle (si besoin est).

2) En principe, les colonnes de U sont les vecteurs propres de AA^T, mais pour faire plus simple, on peut d'abord calculer les l premières colonnes de U par

$$\mathbf{u}_j := \tfrac{1}{\sqrt{\lambda_j}} A\mathbf{v}_j\,, \qquad j = 1, 2, \ldots, l\,,$$

et ensuite, si nécessaire, déterminer les $m - l$ dernières en complétant $(\mathbf{u}_1, \ldots, \mathbf{u}_l)$ en une base orthonormale $(\mathbf{u}_1, \ldots, \mathbf{u}_l, \mathbf{v}_{l+1}, \ldots, \mathbf{v}_n)$ (via Gram-Schmidt + normalisation).

3) La matrice Σ s'obtient à partir des valeurs singulières non-nulles $\sigma_1, \ldots, \sigma_l$, en rajoutant les blocs de zéros nécessaires.

Remarque 15.7. Le calcul des premiers $\mathbf{u}_j$ peut également se faire comme suit :

$$\mathbf{u}_j := \tfrac{1}{\|A\mathbf{v}_j\|} A\mathbf{v}_j\,, \qquad j = 1, 2, \ldots, l\,.$$

En effet, par un calcul que l'on a déjà fait,

$$\begin{aligned}
\|A\mathbf{v}_j\|^2 &= (A\mathbf{v}_j) \cdot (A\mathbf{v}_j) \\
&= \mathbf{v}_j \cdot (A^T A\mathbf{v}_j) \\
&= \mathbf{v}_j \cdot (\lambda_j \mathbf{v}_j) \\
&= \lambda_j \|\mathbf{v}_j\|^2 \\
&= \lambda_j\,,
\end{aligned}$$

et donc $\|A\mathbf{v}_j\| = \sqrt{\lambda_j}$. ◇

Informel 15.8. Remarquons que le travail nécessaire pour diagonaliser $A^T A$ et AA^T peut être très différent, étant donné que ces matrices sont a priori de tailles différentes !

Exemple 15.9. Calculons la décomposition en valeurs singulières de

$$A = \begin{pmatrix} \frac{3}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{9}{10\sqrt{2}} & \frac{13}{10\sqrt{2}} \end{pmatrix}$$

Comme A est 2×2, sa décomposition $A = U\Sigma V^T$ sera un produit de trois matrices 2×2.

On commence par calculer V, qui on le rappelle est formée de vecteurs propres de $A^T A$. Or

$$A^T A = \frac{1}{100} \begin{pmatrix} 153 & 96 \\ 96 & 97 \end{pmatrix}\,,$$

et on sait (voir exercices) que cette dernière possède deux valeurs propres, $\lambda_1 = 9/4$, $\lambda_2 = 1/4$, et que les espaces propres associés sont

$$E_{9/4} = \text{Vect}\left\{ \begin{pmatrix} 4 \\ 3 \end{pmatrix} \right\} \qquad E_{1/4} = \text{Vect}\left\{ \begin{pmatrix} -3 \\ 4 \end{pmatrix} \right\}$$

qui donne, après normalisation,

$$E_{9/4} = \text{Vect}\left\{ \begin{pmatrix} 4/5 \\ 3/5 \end{pmatrix} \right\} \qquad E_{1/4} = \text{Vect}\left\{ \begin{pmatrix} -3/5 \\ 4/5 \end{pmatrix} \right\}$$

On peut donc prendre

$$V = \begin{pmatrix} 4/5 & -3/5 \\ 3/5 & 4/5 \end{pmatrix}\,,$$

qui correspond à une rotation d'angle $\theta = \arccos(4/5)$. Ainsi, $V^T = V^{-1}$ correspond à une rotation de $-\theta$.

Étant connues les valeurs propres de $A^T A$, les valeurs singulières de A sont données par

$$\sigma_1 = \sqrt{\lambda_1} = \frac{3}{2}\,, \qquad \sigma_2 = \sqrt{\lambda_2} = \frac{1}{2}\,,$$

ce qui donne

$$\Sigma = \begin{pmatrix} 3/2 & 0 \\ 0 & 1/2 \end{pmatrix}\,.$$

Ensuite, U a pour colonnes des vecteurs propres de AA^T, or

$$AA^T = \begin{pmatrix} 5/4 & 1 \\ 1 & 5/4 \end{pmatrix}\,,$$

qui possède comme valeurs propres $\lambda_1 = 9/4$ et $\lambda_2 = 1/4$ (comme on sait, les mêmes que $A^T A$!). Ses espaces propres correspondants sont donnés par

$$E_{9/4} = \text{Vect}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\} \qquad E_{1/4} = \text{Vect}\left\{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\}\,,$$

ou encore, après normalisation :

$$E_{9/4} = \text{Vect}\left\{\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\right\} \qquad E_{1/4} = \text{Vect}\left\{\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\right\}\,,$$

On a donc

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\,,$$

qui n'est autre qu'une rotation de $\phi = \frac{\pi}{4}$.

Remarquons qu'on aurait aussi pu trouver les colonnes de U en faisant

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{3/2}\begin{pmatrix} \frac{3}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{9}{10\sqrt{2}} & \frac{13}{10\sqrt{2}} \end{pmatrix}\begin{pmatrix} 4/5 \\ 3/5 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\,,$$

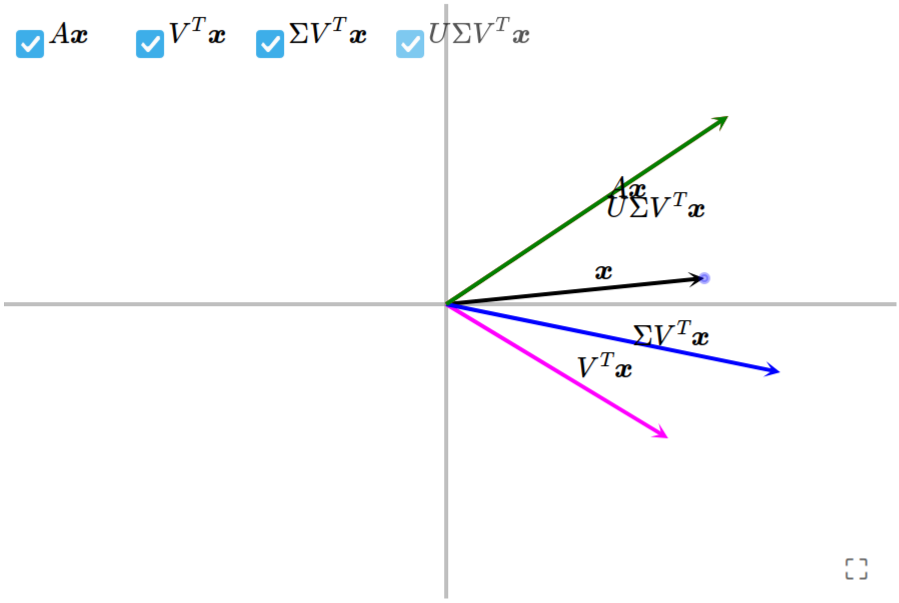
pareil pour $\mathbf{u}_2$.

On a donc la décomposition en valeurs singulières de A, qui permet de voir la transformation

$$\mathbf{x} \mapsto A\mathbf{x} = \underbrace{\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \underbrace{\begin{pmatrix} 3/2 & 0 \\ 0 & 1/2 \end{pmatrix} \underbrace{\begin{pmatrix} 4/5 & 3/5 \\ -3/5 & 4/5 \end{pmatrix}\mathbf{x}}_{V^T\mathbf{x}}}_{\Sigma V^T\mathbf{x}}}_{U\Sigma V^T\mathbf{x}}\,.$$

comme une composition

1) d'une rotation d'angle $-\theta \simeq -36.9^o$, suivie
2) d'un stretch le long des axes de coordonnées ($3/2$ selon $\mathbf{e}_1$, $1/2$ selon $\mathbf{e}_2$), suivi
3) d'une rotation d'angle $\phi = +45^o$.

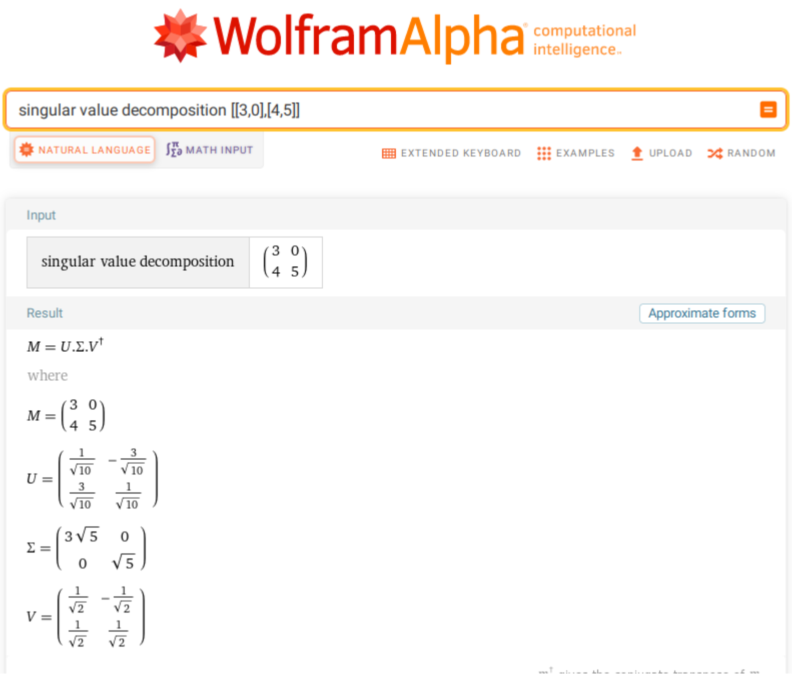


◇

Remarque 15.10. Wolfram Alpha (lien web) peut donner une décomposition en valeurs singulières de n'importe quelle matrice. Par exemple, pour obtenir la décomposition de

$$M = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix},$$

il suffit d'entrer CODE$>singularvaluedecomposition[[3,0],[4,5]]<$CODE



◇

Exemple 15.11. Calculons la décomposition en valeurs singulières de

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

On commence par

$$A^T A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} ,$$

qui possède deux valeurs propres, $\lambda_1 = 4$ et $\lambda_2 = 0$. Ainsi, A possède une seule valeur singulière non-nulle : $\sigma_1 = \sqrt{\lambda_1} = 2$. On trouve un vecteur propre unitaire pour chaque valeur propre, par exemple :

$$\mathbf{v}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} , \qquad \mathbf{v}_2 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} ,$$

qui donne déjà

$$V = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} ,$$

Pour calculer U, on peut soit passer par l'étude de AA^T, ou alors commencer par obtenir une de ses colonnes en prenant

$$\mathbf{u}_1 = \frac{1}{\sqrt{\lambda_1}} A\mathbf{v}_1 = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$$

On doit maintenant trouver deux colonnes $\mathbf{u}_2$ et $\mathbf{u}_3$, de façon à ce que $U = [\mathbf{u}_1\, \mathbf{u}_2\, \mathbf{u}_3]$ soit orthogonale. On peut par exemple prendre

$$\mathbf{u}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} , \qquad \mathbf{u}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} .$$

Il reste à produire Σ. Puisque A n'a qu'une seule valeur singulière non-nulle, et que Σ doit être 3×2, on rajoute des blocs appropriés :

$$\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} .$$

On a donc une SVD pour A :

$$\begin{aligned} A &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}^T , \end{aligned}$$

◇

Exemple 15.12. Étudions la SVD de

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

pour laquelle on aura U (2×2), Σ (2×3), V (3×3). Commençons par

$$A^T A = \begin{pmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{pmatrix} ,$$

qui a pour polynôme caractéristique

$$P_{A^T A}(\lambda) = -\lambda(\lambda - 10)(\lambda - 12) .$$

On a donc les valeurs propres, en ordre décroissant, $\lambda_1 = 12$, $\lambda_2 = 10$, $\lambda_3 = 0$. On a donc deux valeurs singulières strictement positives, $\sigma_1 = \sqrt{12} = 2\sqrt{3}$, $\sigma_2 = \sqrt{10}$.

Les espaces propres sont :

$$E_{12} = \ker(A^T A - 12 I_3) = \text{Vect}\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right\} ,$$

$$E_{10} = \ker(A^T A - 10 I_3) = \text{Vect}\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} \right\} ,$$

$$E_{0} = \ker(A^T A) = \ker(A) = \text{Vect}\left\{ \begin{pmatrix} -1 \\ -2 \\ 5 \end{pmatrix} \right\} .$$

On peut donc normaliser et obtenir

$$V = [\mathbf{v}_1\, \mathbf{v}_2\, \mathbf{v}_3] = \begin{pmatrix} 1/\sqrt{6} & -2/\sqrt{5} & -1/\sqrt{30} \\ 2/\sqrt{6} & 1/\sqrt{5} & -2/\sqrt{30} \\ 1/\sqrt{6} & 0 & 5/\sqrt{30} \end{pmatrix}$$

La matrice $U = [\mathbf{u}_1\, \mathbf{u}_2]$ s'obtient par

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A \mathbf{v}_1 = \frac{1}{2\sqrt{3}} \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} ,$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2} A \mathbf{v}_2 = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{pmatrix} = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} .$$

Finalement, les deux valeurs singulières positives permettent d'écrire

$$\Sigma = \begin{pmatrix} 2\sqrt{3} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} .$$

On a donc une SVD pour A :

$$\begin{aligned} A &= \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 2\sqrt{3} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{6} & -2/\sqrt{5} & -1/\sqrt{30} \\ 2/\sqrt{6} & 1/\sqrt{5} & -2/\sqrt{30} \\ 1/\sqrt{6} & 0 & 5/\sqrt{30} \end{pmatrix}^T \end{aligned}$$

◇

15.4 Rang et représentation

Soit A une matrice $m \times n$ dont une décomposition en valeurs singulières est donnée :

$$A = U \Sigma V^T .$$

NumChap: chap-decomp-val-singulieres, Dernière compilation: 2024-08-12 07:19:48+02:00. (Version Web: botafogo.saitis.net)

Comme nous avons fait avec la décomposition spectrale, nous allons profiter de la SVD pour écrire A comme une combinaison linéaire de matrices plus simples.

Pour commencer, remarquons que l'on peut toujours imposer, dans une SVD, que les valeurs singulières apparaissent sur la diagonale de Σ en ordre décroissant :

$$\sigma_1 \geqslant \sigma_1 \geqslant \dots$$

En nommant les colonnes de U et de V :

$$U = \begin{bmatrix}\mathbf{u}_1 \cdots \mathbf{u}_m\end{bmatrix}, \qquad V = \begin{bmatrix}\mathbf{v}_1 \cdots \mathbf{v}_n\end{bmatrix}.$$

Définissons alors l'indice de la plus petite valeur singulière strictement positive,

$$r := \max\{1 \leqslant k \leqslant l : \sigma_k > 0\}.$$

et procédons comme on l'a fait pour la décomposition spectrale en écrivant, pour tout $\mathbf{x} \in \mathbb{R}^n$,

$$\begin{aligned}
A\mathbf{x} = U\Sigma V^T\mathbf{x} &= \begin{bmatrix}\mathbf{u}_1 \cdots \mathbf{u}_m\end{bmatrix}\Sigma \begin{pmatrix}\mathbf{v}_1^T\mathbf{x} \\ \vdots \\ \mathbf{v}_n^T\mathbf{x}\end{pmatrix} \\
&= \begin{bmatrix}\sigma_1\mathbf{u}_1 \cdots \sigma_r\mathbf{u}_r\, \mathbf{0} \cdots \mathbf{0}\end{bmatrix} \begin{pmatrix}\mathbf{v}_1^T\mathbf{x} \\ \vdots \\ \mathbf{v}_n^T\mathbf{x}\end{pmatrix} \\
&= \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^T \mathbf{x} \\
&= \left(\sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^T\right)\mathbf{x}.
\end{aligned}$$

On a donc pu écrire A comme une somme :

$$\boxed{A = \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^T\,,}$$

On retrouve bien, pour tout j,

$$\begin{aligned}
A\mathbf{v}_j &= \sum_{k=1}^r \sigma_k \mathbf{u}_k \underbrace{\mathbf{v}_k^T\mathbf{v}_j}_{=\delta_{kj}} \\
&= \sigma_j \mathbf{u}_j\,.
\end{aligned}$$

Remarquons qu'à l'inverse de la décomposition spectrale, une matrice $\mathbf{u}_k\mathbf{v}_k^T$ ne représente *pas* une projection puisqu'elle transforme un vecteur de $\mathbb{R}^n$ en un vecteur de $\mathbb{R}^m$. Son seul point commun avec une projection est que

$$\operatorname{rang}(\mathbf{u}_k\mathbf{v}_k^T) = 1\,.$$

SVD fournit donc une représentation d'une matrice comme une somme de matrices de rang égal à 1. Aussi, comme on n'a gardé que les valeurs singulières $\sigma_k > 0$, et comme les $\mathbf{u}_k$ sont indépendants, on a montré :

Théorème 15.13. *Le rang de A est égal au nombre de valeurs singulières non-nulles.*

Approximation optimale par une matrice de rang fixé

Définissons, pour tout $k \leqslant r$,

$$A(k) := \sum_{j=1}^{k} \sigma_j \mathbf{u}_j \mathbf{v}_j^T .$$

Alors $A(k)$ est la matrice de rang k qui approxime le mieux A, dans le sens suivant (c'est le Théorème d'*Eckart-Young*) :

Théorème 15.14. *Soit A $(m \times n)$ de rang $l \leqslant n$. Pour tout $1 \leqslant k \leqslant l$, $A(k)$ est la matrice de rang k qui approxime le mieux A :*

$$\min_B \|A - B\| = \|A - A(k)\| ,$$

où le minimum est pris sur toutes les matrices $m \times n$ de rang au plus égal à k.

Preuve: □

15.5 Élongations et ellipsoïdes

Dans cette section nous utiliserons SVD pour répondre à deux questions géométrique naturelles à propos d'une application linéaire $T : \mathbb{R}^n \to \mathbb{R}^m$ définie par une matrice A, $T(\mathbf{x}) = A\mathbf{x}$:

1) Comment se transforme **la sphère unité**, définie par

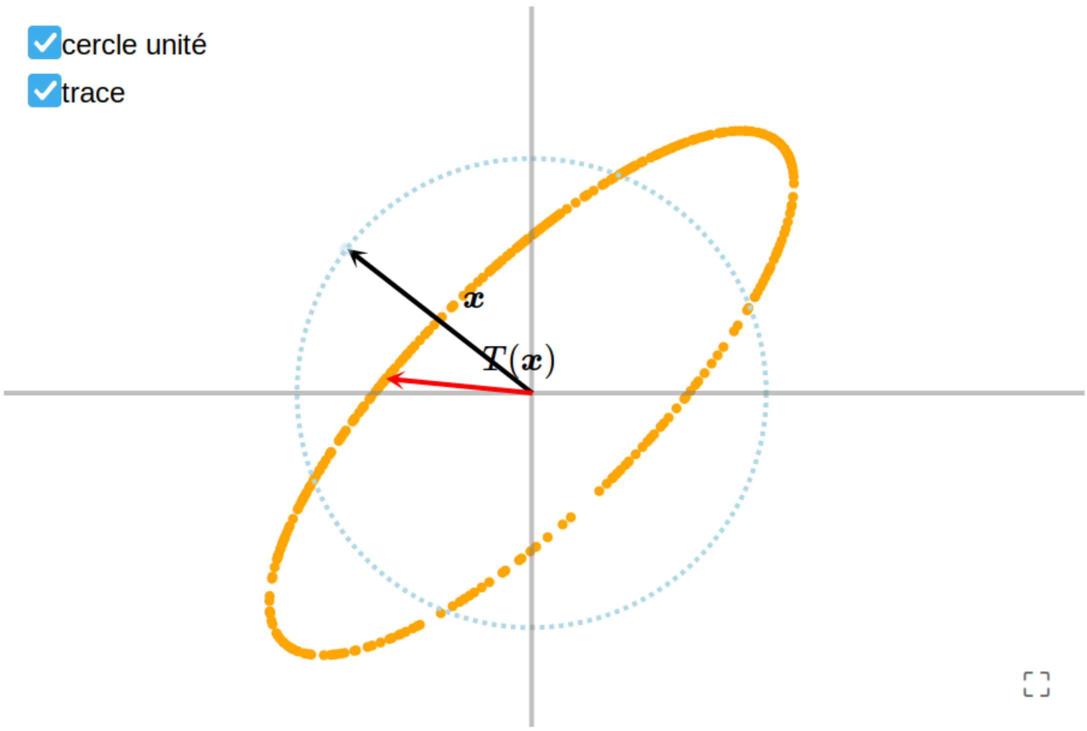
$$\begin{aligned} \mathcal{S} &:= \{\mathbf{x} \in \mathbb{R}^n \,:\, \|\mathbf{x}\| = 1\} \\ &= \{\mathbf{x} \in \mathbb{R}^n \,:\, x_1^2 + \cdots + x_n^2 = 1\} , \end{aligned}$$

   sous l'action de T ? (En $d = 2$, $\mathcal{S}$ est le cercle de rayon 1 centré à l'origine.)

2) Parmi les vecteurs $\mathbf{x}$ situés sur cette sphère, quels sont ceux qui subissent une élongation maximale/minimale, à savoir ceux pour lesquels $\|A\mathbf{x}\|$ est maximal/minimal ?

Ces deux questions pourront être étudiées simultanément.

Exemple 15.15. Sur l'animation suivante, on observe que l'application linéaire $T : \mathbb{R}^2 \to \mathbb{R}^2$ donnée transforme le cercle $\mathcal{S}$ en *ellipse*. Les axes de cette ellipse doivent donner les directions d'élongation maximale (grand axe) et minimale (petit axe) :



◇

Soit $A = U\Sigma V^T$ une décomposition en valeurs singulières de A, dans laquelle on suppose, comme précédemment, que les valeurs singulières sur la diagonale de Σ sont arrangées en ordre décroissant :

$$\sigma_1 \geqslant \sigma_2 \geqslant \dots$$

On rappelle qu'avec cet ordre, $\sigma_j = \sqrt{\lambda_j}$, où $\lambda_j > 0$ est la j-ème plus grande valeur propre de $A^T A$.

Proposition 16. *L'élongation maximale d'un vecteur sur la sphère unité est donnée par*

$$\max_{\mathbf{x}\in\mathcal{S}} \|A\mathbf{x}\| = \max_k \sigma_k = \sigma_1 .$$

L'élongation minimale d'un vecteur sur la sphère unité est donnée par 0 *si* $\ker(A) \neq \{\mathbf{0}\}$, *et sinon par*

$$\min_{\mathbf{x}\in\mathcal{S}} \|A\mathbf{x}\| = \min_k \sigma_k .$$

De plus,

- ⋆ *le maximum est réalisé lorsque* $\mathbf{x}$ *est un vecteur propre unitaire de* $A^T A$ *associé à la plus grande valeur propre de* $A^T A$ *(en l'occurence* λ_1*).*
- ⋆ *le minimum est réalisé lorsque* $\mathbf{x}$ *est un vecteur propre unitaire de* $A^T A$ *associé à la plus petite valeur propre de* $A^T A$*.*

Nous utiliserons l'entier $r = \mathrm{rang}(A)$, qui implique comme on sait que
$\sigma_r > 0$, $\sigma_{r+1} = 0$.
Preuve: Par l'orthogonalité de U (qui implique $\|U\mathbf{z}\| = \|\mathbf{z}\|$ pour tout $\mathbf{z} \in \mathbb{R}^m$), on peut écrire

$$\|A\mathbf{x}\| = \|U\Sigma V^T\mathbf{x}\| = \|\Sigma V^T\mathbf{x}\| \qquad \forall \mathbf{x} \in \mathbb{R}^n .$$

On a donc

$$\begin{aligned}\max_{\mathbf{x}\in\mathcal{S}}\|A\mathbf{x}\| &= \max_{\mathbf{x}\in\mathcal{S}}\|\Sigma V^T\mathbf{x}\|\\ &= \max_{\mathbf{y}\in\mathcal{S}}\|\Sigma\mathbf{y}\|\\ &= \max_{\mathbf{y}\in\mathcal{S}}\sqrt{\sigma_1^2y_1^2+\cdots+\sigma_r^2y_r^2}\,.\end{aligned}$$

Dans la deuxième égalité, on a effectué le changement de variable $\mathbf{y} := V^T\mathbf{x}$ (l'orthogonalité de V^T implique que cette transformation est bijective, et que la condition $\|\mathbf{x}\| = 1$ est préservée puisque $\|V^T\mathbf{x}\| = \|\mathbf{x}\|$).
Ensuite, remarquons que si $\mathbf{y} \in \mathcal{S}$, alors

$$\begin{aligned}\sigma_1^2y_1^2+\cdots+\sigma_r^2y_r^2 &\leqslant \sigma_1^2y_1^2+\cdots+\sigma_1^2y_r^2\\ &= \sigma_1^2(y_1^2+\cdots+y_r^2)\\ &\leqslant \sigma_1^2\|\mathbf{y}\|^2\\ &= \sigma_1^2\,.\end{aligned}$$

Ensuite, soit $\mathbf{z} \in \mathbb{R}^n$ le vecteur qui a toutes ses composantes nulles sauf la première, qui vaut 1. Alors $\mathbf{z} \in \mathcal{S}$, et donc

$$\begin{aligned}\max_{\mathbf{y}\in\mathcal{S}}(\sigma_1^2y_1^2+\cdots+\sigma_r^2y_r^2) &\geqslant (\sigma_1^2z_1^2+\cdots+\sigma_r^2z_r^2)\\ &= \sigma_1^2\|\mathbf{z}\|^2\\ &= \sigma_1^2\,.\end{aligned}$$

Ceci montre que $\max_{\mathbf{x}\in\mathcal{S}}\|A\mathbf{x}\| = \sigma_1$. Ensuite, on a déjà fait plusieurs fois ce calcul : si $\mathbf{v}_1$ est le vecteur propre unitaire de A^TA associé à λ_1, alors

$$\begin{aligned}\|A\mathbf{v}_1\|^2 &= (A\mathbf{v}_1)\cdot(A\mathbf{v}_1)\\ &= \mathbf{v}_1\cdot(A^TA\mathbf{v}_1)\\ &= \mathbf{v}_1\cdot(\lambda_1\mathbf{v}_1)\\ &= \lambda_1\|\mathbf{v}_1\|^2\\ &= \lambda_1\,,\end{aligned}$$

ce qui montre que

$$\max_{\mathbf{x}\in\mathcal{S}}\|A\mathbf{x}\| = \sigma_1 = \sqrt{\lambda_1} = \|A\mathbf{v}_1\|\,.$$

Pour l'élongation minimale, le cas où $\ker(A) \neq \{\mathbf{0}\}$ est immédiat puisque dans ce cas il existe $\mathbf{x}_* \in \mathcal{S}$ tel que $A\mathbf{x}_* = \mathbf{0}$Dans le cas contraire, on commence de la même façon, en utilisant SVD pour écrire

$$\min_{\mathbf{x}\in\mathcal{S}}\|A\mathbf{x}\| = \min_{\mathbf{x}\in\mathcal{S}}\sqrt{\sigma_1^2y_1^2+\cdots+\sigma_r^2y_r^2} = \min_k\sigma_k\,.$$

□

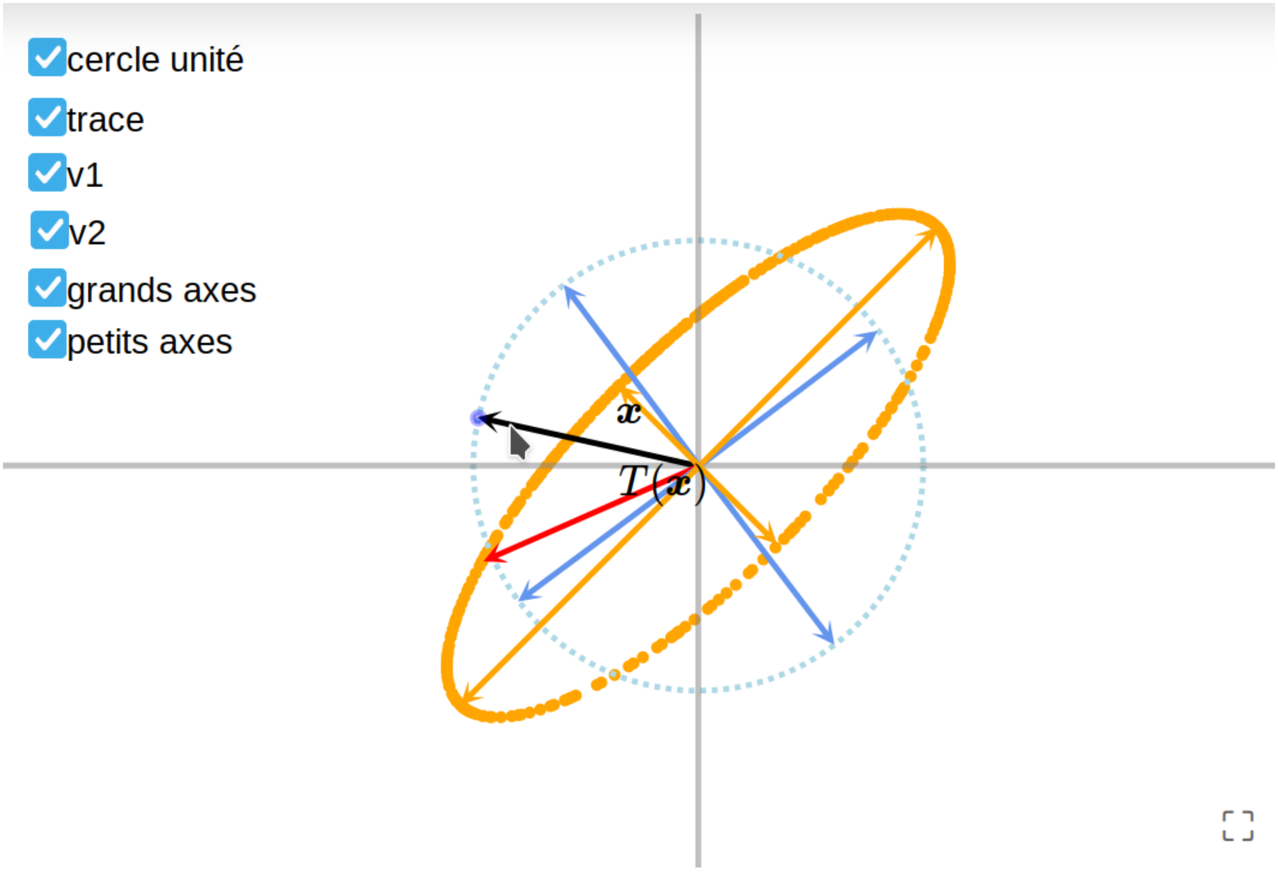
Exemple 15.16. On a déjà rencontré la matrice

$$A = \begin{pmatrix}\frac{3}{2\sqrt{2}} & \frac{1}{2\sqrt{2}}\\ \frac{9}{10\sqrt{2}} & \frac{13}{10\sqrt{2}}\end{pmatrix},$$

qui possède comme valeurs singulières $\sigma_1 = \frac{3}{2}$ et $\sigma_2 = \frac{1}{2}$. Par le théorème ci-dessus, les vecteur du cercle unité qui subissent l'élongation maximale (d'amplitude $\frac{3}{2}$) sous l'action de A sont

$$\pm\mathbf{v}_1 = \pm\begin{pmatrix}4/5\\3/5\end{pmatrix},$$

dont l'image est
$$\pm A\mathbf{v}_1 = \pm\sigma_1\mathbf{u}_1 = \pm\frac{3}{2}\begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix},$$
et les vecteur du cercle unité qui subissent l'élongation minimale (d'amplitude $\frac{1}{2}$) sous l'action de A sont
$$\pm\mathbf{v}_2 = \pm\begin{pmatrix}-3/5\\4/5\end{pmatrix},$$
dont l'image est
$$\pm A\mathbf{v}_2 = \pm\sigma_2\mathbf{u}_2 = \pm\frac{1}{2}\begin{pmatrix}-1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix},$$



◇

Chapitre 16

Espaces Euclidiens

16.1 Définition et exemples

Les chapitres précédents ont montré à quel point l'introduction d'un *produit scalaire* sur $\mathbb{R}^n$ s'est avérée utile, tant du point de vue du traitement des notions de base de l'algèbre linéaire (bases orthogonales, orthonormées, etc) que des applications (calcul de projection, approximation, moindres carrés, SVD, etc.).

Dans ce chapitre, nous allons introduire la notion de produit scalaire sur un espace vectoriel quelconque. Ceci permettra de définir la notion de perpendicularité dans un cadre très général, et d'utiliser une approche semblable à celle des derniers chapitres pour la résolution de nombreux problèmes d'approximation.

Définition 16.1. Soit V un espace vectoriel. On appelle **produit scalaire** une application qui à toute paire de vecteurs $u, v \in V$ associe un réel noté $(u|v) \in \mathbb{R}$, satisfaisant aux propriétés suivantes :

1) $(u|v) = (v|u)$ pour tous $u, v \in V$.
2) Pour tout $v \in V$, l'application $u \mapsto (u|v)$ est linéaire.
3) Pour tout $u \in V$, l'application $v \mapsto (u|v)$ est linéaire.
4) $(u|u) \geqslant 0$ pour tout $u \in V$, et $(u|u) = 0$ si et seulement si $u = 0$.

Un espace vectoriel muni d'un produit scalaire est un **espace pré-Hilbertien**. Un espace pré-Hilbertien de dimension finie est un **espace Euclidien**.

Exemple 16.2. L'espace $\mathbb{R}^n$, muni du produit scalaire

$$(\mathbf{u}|\mathbf{v}) := \mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n \,,$$

est notre premier exemple d'espace Euclidien. ◇

Exemple 16.3. Sur l'espace vectoriel des matrices $\mathbb{M}_{n\times n}(\mathbb{R})$, on peut vérifier que

$$(A|B) := \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$$

définit un produit scalaire. En effet, la symétrie et la bilinéarité sont clairement satisfaites, et

$$(A|A) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \geqslant 0 \,,$$

et cette somme de carrés est nulle si et seulement chacun des carrés $a_{ij}^2 = 0$, c'est-à-dire $a_{ij} = 0$, et donc $A = 0$ (la matrice nulle). ◇

Si V est muni d'un produit scalaire $(\cdot|\cdot)$, on peut maintenant définir

- ⋆ une **norme**,
$$\|v\| := \sqrt{(v|v)}\,,$$
ce qui permet ensuite de parler de la **distance** entre deux vecteurs $u, v \in V$, définie par $\|u - v\|$.
- ⋆ la notion d'orthogonalité : deux vecteurs $u, v \in V$ sont **orthogonaux**, noté $u \perp v$, si
$$(u|v) = 0\,.$$
- ⋆ pour un sous-espace vectoriel $W \subset V$, le **complément orthogonal** dans V est
$$W^\perp := \{v \in V : v \perp w \forall w \in W\}\,.$$

Structure Euclidienne sur les espaces de fonctions

Considérons l'espace vectoriel de toutes les fonctions continues sur un intervalle fermé et borné, $f : [a, b] \to \mathbb{R}$, noté $C([a, b])$.

Notons (voir le cours d'Analyse 1) que les fonctions continues sont intégrables. Donc si $f, g \in C([a, b])$, leur produit étant aussi une fonction continue, on peut définir le nombre
$$(f|g) := \int_a^b f(t)g(t)\,dt\,.$$

Lemme 24. *Cette expression définit un produit scalaire sur $C([a, b])$.*

Preuve: D'abord,
$$(f|g) = \int_a^b f(t)g(t)\,dt = \int_a^b g(t)f(t)\,dt = (g|f)\,.$$
Ensuite, si on fixe g, alors pour tous $f_1, f_2 \in C([a, b])$, $\lambda_1, \lambda_2 \in \mathbb{R}$, les propriétés de linéarité de l'intégrale impliquent
$$\begin{aligned}(\lambda_1 f_1 + \lambda_2 f_2|g) &= \int_a^b (\lambda_1 f_1(t) + \lambda_2 f_2(t))g(t)\,dt\\ &= \lambda_1 \int_a^b f_1(t)g(t)\,dt + \lambda_2 \int_a^b f_2(t)g(t)\,dt\\ &= \lambda_1(f_1|g) + \lambda_2(f_2|g)\,.\end{aligned}$$
En utilisant la symétrie (première propriété), et la propriété précédente,
$$\begin{aligned}(f \mid \lambda_1 g_1 + \lambda_2 g_2) &= (\lambda_1 g_1 + \lambda_2 g_2 \mid f)\\ &= \lambda_1(g_1 \mid f) + \lambda_2(g_2 \mid f)\\ &= \lambda_1(f \mid g_1) + \lambda_2(f \mid g_2)\,.\end{aligned}$$
Puisque l'intégrale d'une fonction non-négative est non-négative,
$$(f|f) = \int_a^b \underbrace{f(t)^2}_{\geqslant 0}\,dt \geqslant 0\,.$$
De plus, l'intégrale de $f(t)^2$ est nulle si et seulement si $f(t) = 0$ pour tout $t \in [a, b]$ (voir cours d'analyse), ce qui implique que f est la fonction identiquement nulle : $f = 0$. □

Ainsi, muni de ce produit scalaire, $C([a, b])$ est un espace pré-Hilbertien (mais pas Euclidien puisque $C([a, b])$ est de dimension infinie). En particulier :

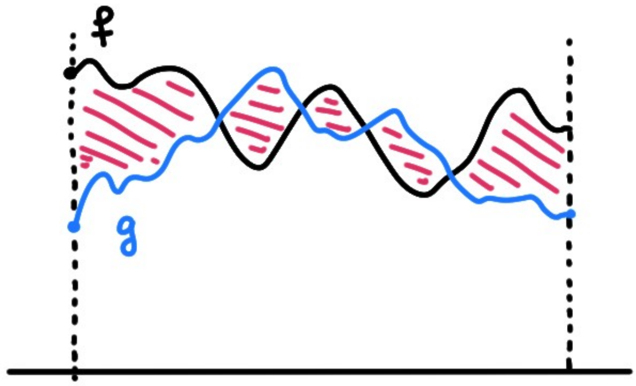
- ⋆ La **norme** de $f \in C([a,b])$ se calcule avec

$$\|f\| = \sqrt{\int_a^b f(t)^2 \, dt} \, .$$

- ⋆ La **distance** entre $f, g \in C([a,b])$ est donnée par

$$\|f - g\| = \sqrt{\int_a^b |f(t) - g(t)|^2 \, dt} \, .$$

Intuitivement, deux fonctions f, g continues sur $[a,b]$ sont *proches*, au sens de $\|\cdot\|$, si l'aire géométrique qui sépare leurs graphes est petite :



(Malgré tout, ce n'est pas exactement cette aire qui apparait puisqu'on intègre le *carré* $|f(t) - g(t)|^2$.)

L'interprétation du produit scalaire, par contre, est moins évidente.

Exemple 16.4. Sur $C([0,\pi])$, considérons les fonctions $f(t) := t$ et $g(t) := \sin(t)$, et calculons leur produit scalaire en utilisant une intégration par parties :

$$\begin{aligned}(f|g) &= \int_0^\pi t \sin(t) \, dt \\ &= t(-\cos(t))\Big|_0^\pi + \int_0^\pi \cos(t) \, dt = \pi \, .\end{aligned}$$

Ensuite, si $h(t) = \cos(t)$, alors

$$\begin{aligned}(g|h) &= \int_0^\pi \sin(t)\cos(t) \, dt \\ &= \frac{1}{2}\int_0^\pi \sin(2t) \, dt \\ &= \frac{1}{4}(-\cos(2t))\Big|_0^\pi = 0 \, ,\end{aligned}$$

donc $g \perp h$. ◇

16.2 Projections et applications

Passons à l'utilisation du produit scalaire pour introduire des *projections*.

Pour commencer, la **projection orthogonale** d'un $\mathbf{v} \in V$ sur un $w \in V$ est définie par

$$\mathrm{proj}_w(v) := \frac{(v|w)}{(w|w)} w \,.$$

Exemple 16.5. Si $f, g \in C([0, \pi])$ sont définies comme dans l'exemple précédent, $f(t) = t$, $g(t) = \sin(t)$, on peut calculer

$$\mathrm{proj}_g(f) = \frac{(f|g)}{(g|g)} g = \frac{\pi}{\frac{\pi}{2}} g = 2g \,,$$

donc $\mathrm{proj}_g(f)$ est la fonction $\mathrm{proj}_g(f)(t) = 2\sin(t)$.

Ou encore,

$$\mathrm{proj}_f(g) = \frac{(g|f)}{(f|f)} f = \frac{\pi}{\frac{\pi^3}{3}} g = \frac{3}{\pi^2} f \,,$$

donc $\mathrm{proj}_g(f)(t) = \frac{3}{\pi^2} t$. ◇

Nous pouvons ensuite définir la **projection orthogonale** d'un vecteur $v \in V$ sur un sous-espace $W \subset V$ (qu'on supposera de dimension *finie*) comme étant l'unique élément $v_{\|} \in W$ qui réalise le minimum

$$\|v - v_{\|}\| = \min_{w \in W} \|v - w\| \,.$$

On note cette projection

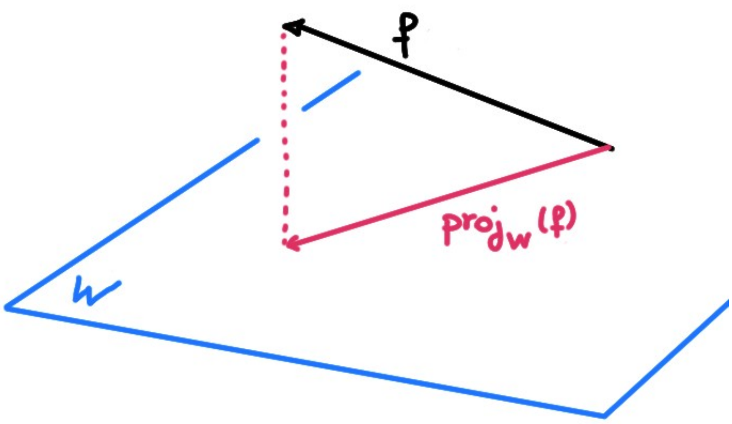
$$v_{\|} = \mathrm{proj}_W(v) \,.$$

Si $\mathcal{B} = (w_1, \dots, w_k)$ est une base orthogonale de W, la projection s'obtient comme dans le cas de $\mathbb{R}^n$:

$$\mathrm{proj}_W(v) = \sum_{j=1}^{k} \frac{(v|w_j)}{(w_j|w_j)} w_j \,.$$

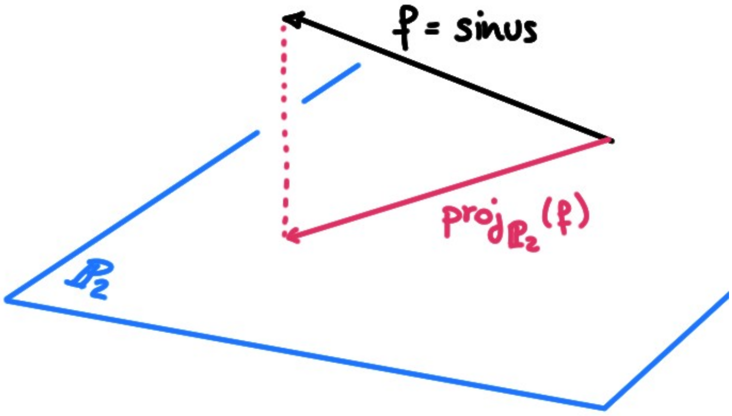
Si la base de W n'est pas orthogonale, on pourra l'orthogonaliser à l'aide du procédé de Gram-Schmidt.

L'intérêt d'une projection est qu'elle fournit, comme on sait, la meilleurs approximation d'un vecteur $v \notin W$ par un élément de W, **au sens des moindres carrés**.



Exemple 16.6. Considérons, dans $C([0,\pi])$, le sous-espace $\mathbb{P}_2$ des polynômes de degré au plus égal à 2. Choisissons $f \in C([0,\pi])$, par exemple $f(t) = \sin(t)$, et cherchons l'élément de $p \in \mathbb{P}_2$ qui approxime le mieux f, au sens des moindres carrés. Pour ce faire, nous allons projeter f sur $\mathbb{P}_2$, pour trouver

$$p(t) = \mathrm{proj}_{\mathbb{P}_2}(f)(t) = \lambda_0 + \lambda_1 t + \lambda_2 t^2 .$$



Remarquons pour commencer que dans $\mathbb{P}_2$, la base canonique $\mathcal{B}_{\text{can}} = (e_0, e_1, e_2)$ n'est pas orthogonale, puisque

$$\begin{aligned}
(e_0|e_1) &= \int_0^\pi 1 \cdot t\, dt = \frac{\pi^2}{2} \neq 0 \\
(e_0|e_2) &= \int_0^\pi 1 \cdot t^2\, dt = \frac{\pi^3}{3} \neq 0 \\
(e_1|e_2) &= \int_0^\pi t \cdot t^2\, dt = \frac{\pi^4}{4} \neq 0 .
\end{aligned}$$

Appliquons le procédé de Gram-Schmidt pour produire une base orthogonale $\mathcal{B} = (e_0', e_1', e_2')$. On prend $e_0' = e_0$, $e_1' = e_1 - \mathrm{proj}_{e_0}(e_1)$, qui donne

$$e_1'(t) = t - \frac{(e_1|e_0)}{(e_0|e_0)} 1 = t - \frac{\pi^2/2}{\pi} = t - \frac{\pi}{2} ,$$

puis

$$\begin{aligned}
e_2' &= e_2 - \mathrm{proj}_{\mathrm{Vect}\{e_0,e_1\}}(e_2) \\
&= e_2 - \mathrm{proj}_{\mathrm{Vect}\{e_0',e_1'\}}(e_2) \\
&= e_2 - \frac{(e_2|e_0')}{(e_0'|e_0')} e_0' - \frac{(e_2|e_1')}{(e_1'|e_1')} e_1' ,
\end{aligned}$$

qui donne

$$\begin{aligned}
e_2'(t) &= t^2 - \frac{(e_2|e_0')}{(e_0'|e_0')} 1 - \frac{(e_2|e_1')}{(e_1'|e_1')} (t - \tfrac{\pi}{2}) \\
&= t^2 - \frac{\pi^3/3}{\pi} 1 - \frac{\pi^4/12}{\pi^3/12} (t - \tfrac{\pi}{2}) \\
&= t^2 - \pi t + \frac{\pi^2}{6}
\end{aligned}$$

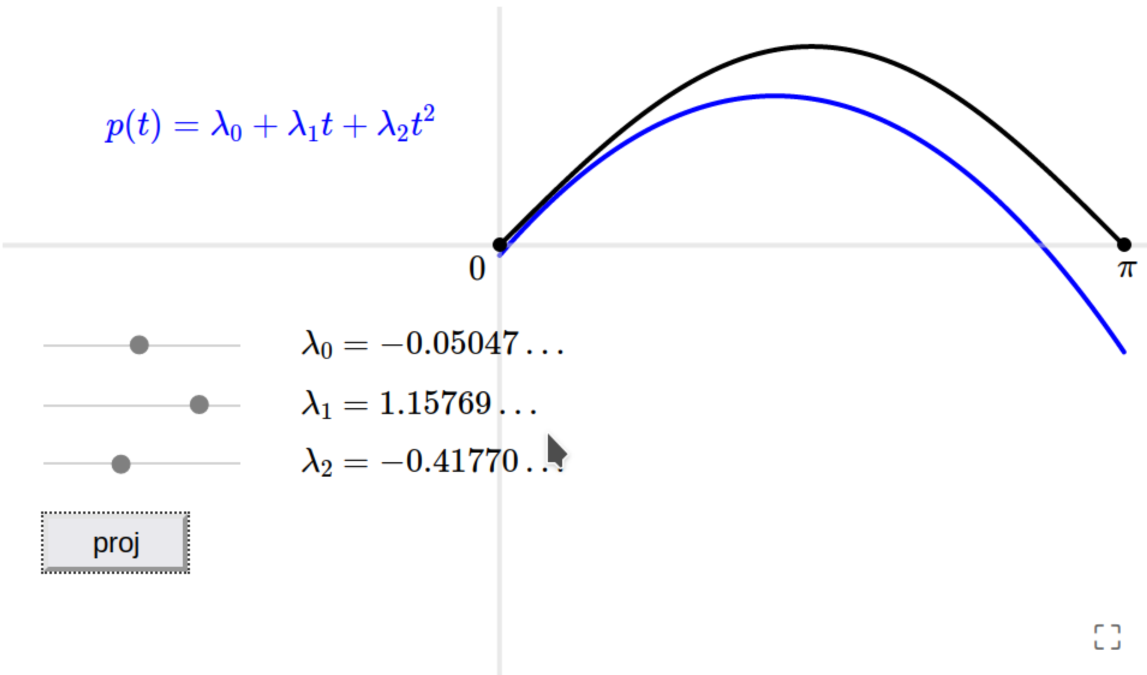
On sait maintenant que

$$p = \mathrm{proj}_{\mathbb{P}_2}(f) = \frac{(f|e_0')}{(e_0'|e_0')}e_0' + \frac{(f|e_1')}{(e_1'|e_1')}e_1' + \frac{(f|e_2')}{(e_2'|e_2')}e_2',$$

qui donne la projection :

$$\begin{aligned} p(t) &= \frac{2}{\pi}1 + \frac{0}{\pi^3/12}(t - \frac{\pi}{2}) + \frac{-4+\pi^2/3}{\pi^5/180}(t^2 - \pi t + \pi^2/6) \\ &= \frac{-4+\pi^2/3}{\pi^5/180}(t^2 - \pi t) + \frac{2}{\pi} + \frac{-4+\pi^2/3}{\pi^3/30} \\ &= -0.0504 + 1.3122\ldots t - 0.4176\ldots t^2 \end{aligned}$$

◇



Remarque 16.7. Dans cet exemple, on a approximé la fonction $\sin(t)$, sur $[0,\pi]$, à l'aide d'un polynôme de degré 2, au sens des moindres carrés : le polynôme $p(t)$ trouvé est celui qui minimise l'intégrale

$$\int_0^\pi |\sin(t) - p(t)|^2\,dt$$

Cette méthode est globale et *très* différente de la méthode du **développement limité** (lien web) de l'analyse, qui est plus locale puisqu'elle a pour but de fournir une approximation précise d'une fonction au voisinage d'un *point*. ◇