# Algèbre linéaire
# Prof. Boumal Nicolas — EPFL

Notes par Joachim Favre

Bachelor d'informatique — Semestre 1
Automne 2021

J'ai fait ce document pour mon usage, mais je me suis dit que des notes dactylographiées pouvaient intéresser d'autres personnes. Ainsi, je l'ai partagé (à vous, si vous lisez ces lignes !) ; puisque cela ne me coûtait rien. Je vous demande simplement de garder en tête qu'il y a des erreurs, c'est impossible de ne pas en faire. Si vous en trouvez, n'hésitez pas à me les partager (les erreurs de grammaires et de vocabulaires sont naturellement aussi bienvenues). Vous pouvez me contacter à l'adresse e-mail suivante :

`joachim.favre@epfl.ch`

Si vous n'avez pas obtenu ce document par le biais de mon repo GitHub, vous serez peut-être intéressé par le fait que j'en ai un sur lequel je mets mes notes dactylographiées. Voici le lien :

`https://github.com/JoachimFavre/EPFLNotesIN`

Notez que le contenu ne m'appartient pas. J'ai fait quelques modifications de structures, j'ai reformulé certains bouts, et j'ai ajouté quelques notes personnelles ; mais les formulations et les explications viennent principalement de la personne qui nous a donné ce cours, et du livre dont elle s'est inspirée.

Je pense qu'il est intéressant de préciser que, pour avoir ces notes dactylographiées, j'ai pris mes notes en LaTeXpendant le cours, puis j'ai fait quelques corrections. Je ne pense pas que mettre au propre des notes écrites à la main est faisable niveau quantité de travail. Pour prendre des notes en LaTeX, je me suis inspiré du lien suivant, écrit par Gilles Castel. Si vous voulez plus de détails, n'hésitez pas à me contacter à mon adresse e-mail, mentionnée ci-dessus.

`https://castel.dev/post/lecture-notes-1/`

Je tiens aussi à préciser que les mots "trivial" et "simple" n'ont, dans ce cours, pas la définition que vous trouvez dans un dictionnaire. Nous sommes à l'EPFL, rien de ce que nous faisons n'est trivial. Quelque chose de trivial, c'est quelque chose que quelqu'un pris de manière aléatoire dans la rue serait capable de faire. Dans notre contexte, comprenez plutôt ces mots comme "plus simple que le reste". Aussi, ce n'est pas grave si vous prenez du temps à comprendre quelque chose qui est dit trivial.

Puisque vous lisez ces lignes, je vais me permettre de vous donner un petit conseil. Le sommeil est un outil bien plus puissant que ce que vous pouvez imaginer, donc ne négligez jamais une bonne nuit de sommeil au profit de vos révisions (particulièrement la veille de l'examen). Je vais aussi me permettre de paraphraser mon enseignante de philosophie du gymnase, Ms. Marques, j'espère que vous vous amuserez en faisant vos examens !

Pour finir, j'aimerais remercier Paul Tissot-Daguette qui m'a fait beaucoup de corrections sur ce document. Ce dernier n'est toujours pas parfait, mais il est plus lisible que si Paul ne m'avait pas donné son aide.

*Version 2022–01–10*

# Table des matières

# Liste des cours

# Chapitre 1

# Résumé par cours



- Explication de l'organisation du cours.
- Définition du concept d'équation linéaire et de système.
- Définition des systèmes triangulaires.
- Explication des opérations élémentaires.
- Définition de la forme matricielle, ainsi que des matrices des coefficients et matrices augmentées.



- Construction des définitions de matrice échelonnée et matrice échelonnée réduite.
- Définition des vecteurs et de leurs opérations.
- Définition des combinaisons linéaires.



- Définition des équations vectorielles.
- Définition du span, ou vect en français.
- Définition du produit matrice-vecteur.
- Définition des équations matricielle.
- Observation que les systèmes d'équations, les matrices augmentées, les équations vectorielles et les équations matricielles sont des notations équivalentes.
- Définition des systèmes homogènes et de leur solution triviale.



- Démonstration que les solutions d'un système peuvent être écrites sous la forme d'une somme d'une solution particulière et de la solution générale au système homogène correspondant.
- Définition du concept de dépendance et d'indépendance linéaire.
- Démonstration qu'une famille de vecteurs est linéairement dépendantes si et seulement si au moins un de ses vecteurs est une combinaison de ceux qui le précèdent.
- Démonstration que si le vecteur nul est présent dans la famille de vecteur, ou que si la dimension de nos vecteurs est plus petite que notre nombre de vecteurs, alors ils sont nécessairement linéairement dépendants.

- Définition des applications et de la terminologie autour.
- Définition des applications linéaires et démonstration que $T$ est linéaire si et seulement s'il est sous la forme $T(\overrightarrow{x}) = A\overrightarrow{x}$, où $A$ est une matrice.
- Explication de comment calculer la transformation linéaire d'un ensemble de points.



- Définition de surjectivité et d'injectivité d'une application linéaire.
- Résumé des théorèmes liés aux pivots des matrices, ainsi que la surjectivité ou injectivité des transformation linéaires correspondantes.
- Définition des opérations matricielles, et explication de la logique derrière le produit matriciel.



- Définition des matrices transposées.
- Démonstrations de certaines propriétés des matrices transposées.
- Définition des matrices inverses.
- Démonstrations de certaines propriétés des inverses.
- Explication d'un algorithme pour trouver une matrice inverse.



1. Définition des matrices élémentaires.
2. Explication du grand théorème (12 propositions équivalentes pour dire qu'une matrice est inversible).
3. Définition du déterminant pour une matrice $n \times n$.
4. Démonstration de la formule pour les matrices $2 \times 2$ et $3 \times 3$, et les matrices triangulaires et diagonales.



- Calcul du déterminant pour les matrices élémentaires.
- Découverte que déterminant d'un produit de matrice est le produit des déterminants ($\det(AB) = \det(A)\det(B)$).
- Utilisation de ce fait pour calculer des déterminants plus simplement.
- Explication de l'interprétation géométrique des déterminants.
- Notez que le titre de ce cours essaie de faire une référence à Undertale (en *run* génocide, quand on sauvegarde, il est juste écrit "Détermination.")



- Définition des espaces vectoriels, et des propriétés qu'ils doivent suivre.
- Définition des sous-espaces vectoriels, et des propriétés qu'ils doivent suivre.

- Définition des matrices diagonalisable.



- Explication d'une méthode pour trouver les matrices permettant de diagonaliser une matrice $A$.
- Étude des propriétés d'une matrice qui font qu'elle est diagonalisable ou non.
- Définition de la multiplicité géométrique d'une valuer propre.



- Fin des matrices diagonales en expliquant leur utilité, et en revenant à Fibonacci et aux changement de base d'applications linéaires.
- Définition de distance Euclidienne.
- Définition du produit scalaire.



- Définition de la norme d'un vecteur, de la normalisation d'un vecteur et de la distance entre deux vecteurs.
- Définition de vecteurs orthogonaux et preuve du théorème de Pythagore pour les vecteurs.
- Définition de complément orthogonal, et preuve que c'est un sous-espace vectoriel de $\mathbb{R}^n$.



- Démonstrations que l'orthogonal du kernel est l'espace engendré par les lignes, et explication similaire pour les trois autres espaces d'une matrice.
- Définition de famille orthogonale, et démonstration qu'une famille orthogonale avec aucun vecteur nul est libre.
- Définition de base orthogonale, et explication de la méthode pour trouver les coefficients d'un vecteur dans une base orthogonale.
- Explication des projections orthogonales en 2D.



- Explication du calcul de projections orthogonales sur des bases orthogonales.
- Définition de base orthonormée.
- Démonstration que $U^T U = I$ pour les bases orthogonales et que $UU^T$ est la matrice de projection pour les bases orthonormales.
- Définition de matrice orthogonale.



- Explication de l'algorithme du Grand Schtroumpf (euh de Gram-Schmidt pardon).
- Explication de la méthode pour factoriser une matrice dans la forme $QR$.
- Étude du cas de la factorisation $QR$ matrice carrée, et meilleure compréhension du déterminant d'une matrice.

# Chapitre 2

# Organisation

| | |
|---|---|
| **Moodle** | Toutes les informations sont sur Moodle, y compris les slides prises en "live". Il faut aller le consulter régulièrement. |
| **Examen blanc** | En novembre, il y aura un examen blanc, pour qu'on puisse voir notre niveau. |
| **Ouvrage de référence** | L'ouvrage de référence est "Algèbre linéaire et applications" écrit par David C. Lay (référence sur Moodle). Il donne des explications détaillées et des explication supplémentaire. |

# Chapitre 3

# Équations linéaires

## 3.1 Le cas d'une unique équation

**Définition des équations linéaires**

Pour des variables (= inconnues) $x_1, \ldots, x_n$, une **équation linéaire** est une équation de la forme

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$$

pour des nombres $a_1, a_2, \ldots, a_n$ et $b$ indépendants des variables.

**Exemples**

Les équations suivantes sont des équations linéaires :

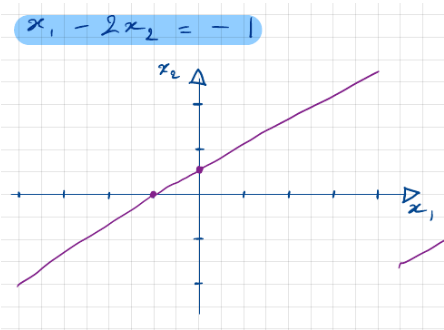
- $5x_1 + 3x_2 = 7$
- $0 = 5$
- $\sqrt{7} = x_1 + x_2$
- $e^3x_1 = 9 + x_2$
- $x_1 + x_2 + x_3 + x_4 + x_5 = 0$
- $5x_1 + x_{1000} = -\pi$

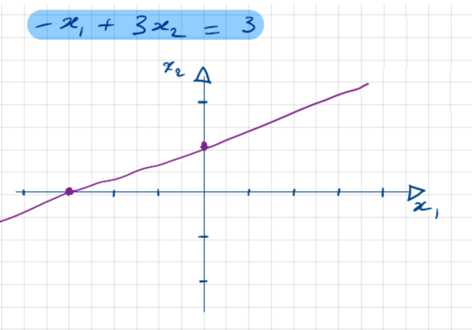
Les équations suivantes ne sont pas des équations linéaires :

- $e^{x_1} = e^{x_2}$
- $x_1x_2 + x_1 = 0$
- $\sqrt{x_1 + x_2 + x_3} = \sqrt{9}$

**Exemple en deux dimensions**

On dessine les droites $x_1 = 2x_2 = -1$ et $-x_1 + 3x_2 = 3$ sur deux repères $x_1, x_2$.




Gauche : $x_1 = 0$, donc $-2x_2 = -1 \implies x_2 = \frac{1}{2}$, donc $\left(0, \frac{1}{2}\right)$ est un point. De la même manière, en prenant $x_2 = 0$, on obtient $x_1 = -1$, donc $(-1, 0)$ est un autre point par lequel passe cette droite. On peut donc la dessiner.
Droite : Si $x_1 = 0$, alors $x_2 = 1$, donc on a le point $(0, 1)$. De la même manière, en prenant $x_2 = 0$, on a $x_1 = -3$, donc le point $(-3, 0)$ fait aussi partie de la droite, qu'on peut dessiner.
Ces droites sont donc l'ensemble des solutions à ces équations.

## 3.2 Système (d'équations) linéaires

**Définition des systèmes d'équations linéaires, des solutions et des ensembles solutions**

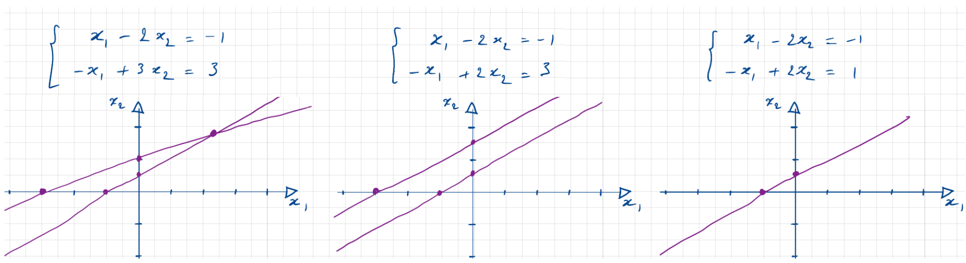
Un **système d'équation linéaires** est une collection de une ou plusieurs équations linéaire en des variables $x_1, \ldots, x_n$. S'il y a $m$ équations en $n$ variables, on écrit

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 & (3.2a) \\ \vdots & (3.2b) \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m & (3.2c) \end{cases}$$

avec $a_{ij}$, le coefficient de $x_j$ dans la ième équation.
Étant donnée une liste de nombres $(s_1, \ldots, s_n)$, on dit que cette liste est une **solution** (du système) si les $m$ équations sont vraies quand on remplace $x_1$ par $s_1$, …, et $x_n$ par $s_n$.
L'ensemble des solutions d'un système est son **ensemble solution.** Deux systèmes sont **équivalents** s'ils ont le même ensemble solutions.

**Exemples**

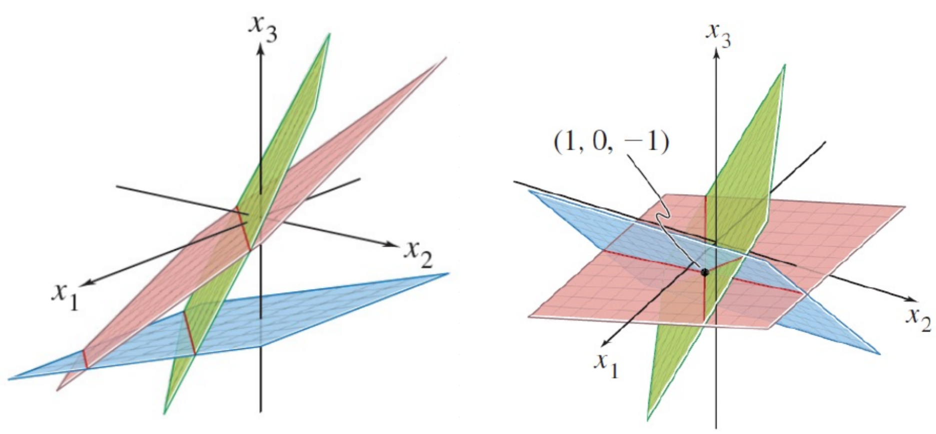
Les droites se croisent en un point. Un point d'une droite satisfait son équation, donc si un point appartient aux deux droites simultanément, il est une solution au système d'équations.




Dans le cas où les droites se croisent, il y a une solution unique. Si les deux droites sont parallèles, il n'existe aucun point qui appartient simultanément aux deux droites, donc l'ensemble de solution est vide, noté ø. Dans le cas où les droites sont les mêmes, il y a une infinité de solution.

**Trois variables**

Une équation avec $a_1x_1 + a_2x_2 + a_3x_3 = b$ définit un plan. Le croisement entre deux plan définit une droite.
Du coup, un système linéaire avec $m$ équations en $n = 3$ variables définit $m$ plans dans l'espace en 3D.




Sur le premier des deux dessins, le plan bleu est parallèle à la droite dessinée par les plans verts et rouges, donc il n'y a pas de solution.

**Théorème** Un système linéaire a zéro (aucune solution n'existe), une (la solution existe et est unique) ou une infinité (des solutions existent, mais elles ne sont pas uniques) de solutions.

*Preuve*

- Zéro solution :
- Une solution :
- Plus qu'une solution (on complètera plus tard, mais on peut la faire ; c'est une preuve par l'absurde. Si on a deux solutions, il y a une droite qui relie ces deux points dans l'espace en n-D et on peut montrer que, si on a ces deux solutions, alors tous les points de la droite qui les relient sont aussi des solutions, ce qui est une contradiction).

□

## 3.3 Système triangulaire

**Trouver toutes les solutions** On a un système linéaire de $m \times n$ (avec $m$, le nombre d'équation, et $n$ le nombre de variables) avec coefficients $a_{ij}$ (avec $i \in \{1, \ldots, m\}$ et $j \in \{1, \ldots, n\}$) :

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 & (3.4a) \\ \vdots & (3.4b) \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m & (3.4c) \end{cases}$$

Certains systèmes sont plus faciles à résoudre que d'autres. Les diagonaux sont plus simples que le triangulaires, qui sont plus simples que les “généraux”. La stratégie va être de transformer un système général en système triangulaire (ou aussi triangulaire que possible).

**Exemple triangulaire**

$$\begin{cases} 2x_1 + 4x_2 + 2x_3 = 1 & (3.6a) \\ 2x_2 + 2x_3 = 3 & (3.6b) \\ 4x_3 = 2 & (3.6c) \end{cases}$$

- $4x_3 = 2 \implies x_3 = \frac{1}{2}$
- $2x_2 + 2x_3 = 3 \implies 2x_2 = 3 - 2x_3 = 2 \implies x_2 = 1$
- $2x_1 + 4x_2 + 2x_3 = 1 \implies 2x_1 = 1 - 4x_2 - 2x_3 = 1 - 4 - 1 = -4 \implies x_1 = -2$

Le système a donc une solution unique : $\left(-2, 1, \frac{1}{2}\right)$. On peut vérifier la solution.

**Systèmes triangulaires** Plus généralement, un système est triangulaire, s'il est de la forme :

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 & (3.8a) \\ a_{22}x_2 + \ldots + a_{2n}x_n = b_2 & (3.8b) \\ \ldots & (3.8c) \end{cases}$$

On peut avoir $m \neq n$, naturellement. Si le système n'est pas triangulaire, on va modifier le système pour le rendre triangulaire ou “presque” triangulaire, sans changer l'ensemble des solution on ne veut ni “perdre” ni “créer” de solutions.

## 3.4 Opérations élémentaires

**Théorème des opérations élémentaires**

Les opérations suivantes ne changent pas l'ensemble solution :
1. Permuter deux équations.
2. Multiplier une équation par un nombre non nul.
3. Additionner à une équation un multiple d'une autre équation.

On les appelle des **opérations élémentaires**.

*Preuve* Laissée au lecteur.

*Remarque* Toutes les opérations sont reversible. Cela peut être démontré facilement.

**Exemple**

On a

$$\begin{cases} 2x_1 - 3x_2 + 5x_3 = 8 & (3.14a) \\ x_1 + x_2 - 3x_3 = -7 & (3.14b) \\ 4x_1 + 2x_2 + 2x_3 = 0 & (3.14c) \end{cases}$$

En permutant les équations $(a)$ et $(b)$, i.e. $(a) \leftrightarrow (b)$ :

$$\begin{cases} x_1 + x_2 - 3x_3 = -7 & (3.15a) \\ 2x_1 - 3x_2 + 5x_3 = 8 & (3.15b) \\ 4x_1 + 2x_2 + 2x_3 = 0 & (3.15c) \end{cases}$$

On va essayer de rentre les coefficients tout à gauche nul, la seule opération qui pourrait fonctionner est d'ajouter un multiple d'une autre équation. On va remplacer l'équation $(c)$ par l'équation $(c)$ moins quatre fois l'équation $(a)$ (les noms $(a), (b), (c)$ changent à chaque étape) et l'équation deux de manière similaire, i.e. $(c) \leftarrow (c) - 4(a)$ et $(b) \leftarrow (b) - 2(a)$ :

$$\begin{cases} x_1 + x_2 - 3x_3 = -7 & (3.16a) \\ -5x_2 + 11x_3 = 22 & (3.16b) \\ -2x_2 + 14x_3 = 28 & (3.16c) \end{cases}$$

On veut rendre le coefficient tout à gauche de la troisième équation nul. Cela semble compliqué en utilisant la première équation (le coefficient devant $x_1$ ne serait plus nul). On prend $(c) \leftarrow \frac{5}{2}(c)$ :

$$\begin{cases} x_1 + x_2 - 3x_3 = -7 & (3.17a) \\ -5x_2 + 11x_3 = 22 & (3.17b) \\ -5x_2 + 35x_3 = 70 & (3.17c) \end{cases}$$

On prend $(c) \leftarrow (c) - (b)$ :

$$\begin{cases} x_1 + x_2 - 3x_3 = -7 & (3.18a) \\ -5x_2 + 11x_3 = 22 & (3.18b) \\ 24x_3 = 48 & (3.18c) \end{cases}$$

Maintenant qu'on a rendu notre système triangulaire, il est très facile de trouver les solutions du système.

**Pour aller plus loin**

Nous avons encore plusieurs choses à faire pour aller plus loin :
1. On peut adopter une **notation** plus efficace, qui se révèlera être bien plus qu'une notation (les matrices ☺)
2. On peut **simplifier davantage** (en transformant la matrice en "diagonale" et non en "triangulaire")
3. Il n'est pas toujours possible d'obtenir un triangle parfait.

## 3.5 Matrices

**Notation matricielle** Les équations sont de la forme $a_1x_1 + \dots a_nx_n = b_n$. Vu qu'elles ont toutes la même forme, seuls les coefficients $a_1, \dots, a_n, b_n$ sont importants. Si on a un système

$$\begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = b_1 & (3.20a) \\ \vdots & (3.20b) \\ a_{m1}x_1 + \dots + a_{mn}x_n = b_m & (3.20c) \end{cases}$$

On collecte les nombres utiles dans deux tableaux appelés matrices :

$$\begin{bmatrix} a_{1,1} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & \cdot a_{mn} \end{bmatrix} \quad \begin{bmatrix} a_{1,1} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{bmatrix}$$

Sur la gauche, la **matrice des coefficients** $a_{ij}$ est de taille $m \times n$. Sur la droite, c'est la **matrice augmentée**, qui est de taille $m \times (n+1)$ (on peut mettre des traitillés juste avant la dernière colonne de cette matrice augmentée, mais on n'est pas obligé).
Chaque système linéaire correspond à une matrice augmentée, et vice versa.

**Les opérations élémentaires** Nous retrouvons ces mêmes opérations élémentaires pour une matrice augmenté :

1. Permuter deux équations. / Permuter deux lignes.
2. Multiplier une équation par un nombre non nul. / Multiplier une ligne par un nombre non nul.
3. Additionner à une équation un multiple d'une autre équation. / Additionner à une ligne un multiple d'une autre ligne.

**Exemple** On peut transformer notre système vu plus tôt en matrice augmentée :

$$\begin{bmatrix} 2 & -3 & 5 & 8 \\ 1 & 1 & -3 & -7 \\ 4 & 2 & 2 & 0 \end{bmatrix}$$

On peut permuter les deux premières lignes :

$$\begin{bmatrix} 1 & 1 & -3 & -7 \\ 2 & -3 & 5 & 8 \\ 4 & 2 & 2 & 0 \end{bmatrix}$$

On peut soustraire quatre fois la première ligne à la dernière ligne, et on peut soustraire deux fois la première à la seconde ligne :

$$\begin{bmatrix} 1 & 1 & -3 & -7 \\ 0 & -5 & 11 & 22 \\ 0 & -2 & 14 & 28 \end{bmatrix}$$

On passe les quelques étapes qu'on a vue plus haut, puis on arrive à

$$\begin{bmatrix} 1 & 1 & -3 & -7 \\ 0 & -5 & 11 & 22 \\ 0 & 0 & 24 & 48 \end{bmatrix}$$

qui est bien une matrice triangulaire, et qui est la matrice augmentée du système triangulaire qu'on avait trouvé plus tôt. On peut retourner au monde des systèmes pour résoudre nos équations.

Lundi 27 septembre 2021 — **Cours 2 : Matrices échelonnées et vecteurs**

**Suite de l'exemple** Nous avons la matrice augmentée

$$\begin{bmatrix} 1 & 1 & -3 & -7 \\ 0 & -5 & 11 & 22 \\ 0 & 0 & 24 & 48 \end{bmatrix}$$

Nous voudrions essayer d'aller plus loin, en rendant la matrice diagonale. On ne veut pas "casser" les zéros en dessous de la diagonale, cependant on voudrait faire apparaître des zéros au dessus. On peut utiliser la dernière ligne pour réduire les coefficients de la troisième colonne à zéro. Donc on peut prendre $(b) \leftarrow (b) - \frac{11}{24}(c)$ et $(a) \leftarrow (a) + \frac{3}{24}(c)$ :

$$\begin{bmatrix} 1 & 1 & 0 & -1 \\ 0 & -5 & 0 & 0 \\ 0 & 0 & 24 & 48 \end{bmatrix}$$

On peut maintenant prendre $(a) \leftarrow (a) + \frac{1}{5}(b)$ :

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -5 & 0 & 0 \\ 0 & 0 & 24 & 48 \end{bmatrix}$$

Maintenant on peut rendre les coefficients de la diagonale égaux à 1, en prenant $(a) \leftarrow (a)$, $(b) \leftarrow \frac{-1}{5}(b)$ et $(c) \leftarrow \frac{1}{24}(c)$ :

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Ce qui est équivalent au système suivant

$$\begin{cases} x_1 = -1 & (3.22a) \\ x_2 = 0 & (3.22b) \\ x_3 = 2 & (3.22c) \end{cases}$$

On ne peut plus simplifier notre matrice, puisqu'on a toutes nos solutions. En l'occurrence, nous avons une solution unique. Cependant, comme nous l'avons vu, il peut y en avoir aucune, ou en avoir une infinité.

**Exemple avec une conclusion différente** Soit le système

$$\begin{cases} x_2 - 4x_3 = 8 & (3.25a) \\ 2x_1 - 3x_2 + 2x_3 = 1 & (3.25b) \\ 4x_1 - 8x_2 + 12x_3 = 1 & (3.25c) \end{cases}$$

En le transformant en matrice augmentée et en faisant les étapes usuelles, on obtient la matrice suivante

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 15 \end{bmatrix}$$

On a donc un zéro de plus que d'habitude sur la dernière ligne (et le coefficient n'est pas nul). On s'est donc retrouvé avec le système d'équations suivant :

$$\begin{cases} 2x_1 - 3x_2 + 2x_3 = 1 & (3.26a) \\ x_2 - 4x_3 = 8 & (3.26b) \\ 0 = 15 & (3.26c) \end{cases}$$

Ce qui veut dire que ce système n'a aucune solution.
Géométriquement, on cherche l'intersection entre trois plans, mais avec un plan parallèle à l'intersection des deux autres.

**Exemple avec encore une autre conclusion**

En prenant le même système, mais en changeant le dernier coefficient :

$$\begin{cases} x_2 - 4x_3 = 8 & \text{(3.30a)} \\ 2x_1 - 3x_2 + 2x_3 = 1 & \text{(3.30b)} \\ 4x_1 - 8x_2 + 12x_3 = -14 & \text{(3.30c)} \end{cases}$$

De nouveau, on peut transformer ce système en matrice augmentée, appliquer les opérations élémentaires sur les lignes, et on obtient

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Elle ressemble fortement à la matrice qu'on a obtenue dans l'exemple juste avant, sauf que le dernier coefficient est 0 et non 15. C'est donc équivalent au système suivant :

$$\begin{cases} 2x_1 - 3x_2 + 2x_3 = 1 & \text{(3.31a)} \\ x_2 - 4x_3 = 8 & \text{(3.31b)} \\ 0 = 0 & \text{(3.31c)} \end{cases}$$

Cette dernière équation ne nous apporte rien, on peut donc l'enlever. Nous avons donc une infinité de solutions. Pour les écrire, nous allons simplement écrire toutes nos équations en fonction de $x_3$ (qu'on peut choisir comme on veut) : $x_2 = 8 + 4x_3$ et

$$2x_1 = 1 + 3x_3 - 2x_3 1 + 3(8 + 4x_3) - 2x_3 = 25 + 10x_3 \iff x_1 = \frac{25}{2} + 5x_3$$

Donc,

$$\begin{cases} x_1 = \frac{25}{2} + 5x_3 & \text{(3.32a)} \\ x_2 = 8 + 4x_3 & \text{(3.32b)} \\ x_3 \text{ est libre} & \text{(3.32c)} \end{cases}$$

On aurait pu choisir une autre variable libre, mais cela aurait été équivalent. Géométriquement, on a l'intersection entre trois plans, qui forme une droite.

**L'exemple précédent, sans bidouiller**

On avait :

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

On veut continuer à travailler sur cette matrice. On peut choisir d'enlever le deuxième coefficient de la première ligne, ou le troisième. On va choisir le premier, pour avoir quelque chose qui ressemble à une matrice diagonale sur les deux premières colonnes. On prend $(a) \leftarrow (a) + 3(b)$ :

$$\begin{bmatrix} 2 & 0 & -10 & 25 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Maintenant, on peut mettre des 1 sur la diagonale :

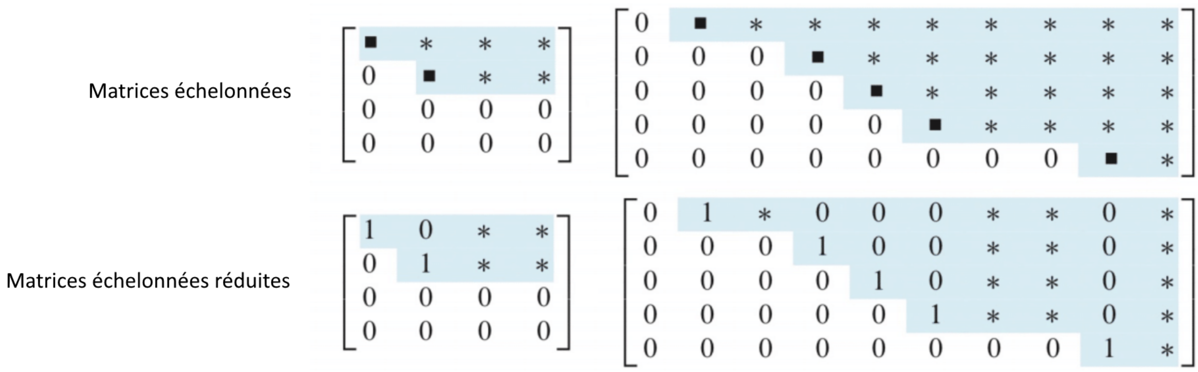
$$\begin{bmatrix} 1 & 0 & -5 & \frac{25}{2} \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

En repassant sur une notation de système :

$$\begin{cases} x_1 - 5x_3 = \frac{25}{2} & \text{(3.34a)} \\ x_2 - 4x_3 = 8 & \text{(3.34b)} \\ 0 = 0 & \text{(3.34c)} \end{cases}$$

Ce qui est équivalent au système qu'on a trouvé en bidouillant. Si on avait choisit de mettre le 0 dans l'autre entier, on aurait simplement eu une autre variable libre (mais l'ensemble solution aurait été le même, décrit différemment). Par convention, on fait comme ce qu'on a fait dans cet exemple.

## 3.6 Matrice échelonnée (réduite)




**Définition de coefficient principal** Pour une matrice de taille $m \times n$, on appelle **coefficient principal** d'une ligne non-nulle le coefficient non-nul le plus à gauche dans la ligne. Dans le cas ou la ligne est complètement nulle, alors il n'y a pas de coefficient principal.
Ce sont les carrés et les 1 dans le dessin.

**Définition de forme échelonnée** Une matrice est sous **forme échelonnée** (“aussi triangulaire que possible”) si

1. Toutes les lignes nulles (s'il y en a) sous tout en bas.
2. Le coefficient principal d'une ligne se trouve à droite du coefficient principal sur la line au dessus d'elle (les coefficients principals “descendent en escalier”, en général, toutes les “marches” ne font pas la même longueur).
3. Tous les coefficients d'une colonne sous un coefficient principal sont nuls.

**Définition de forme échelonnée réduite** Une matrice sous formée échelonnée qui satisfait en plus les deux conditions ci-dessous est sous **forme échelonnée réduite** (“aussi diagonale que possible et avec des 1 sur la diagonale”)

4. Le coefficient principal de chaque ligne non nul vaut 1.
5. Les coefficients principaux sont les seuls éléments non-nuls de leur colonne.

**Exemples** Les matrices suivantes ne sont pas échelonnées :

$$\begin{bmatrix} 0 & 4 & 3 \\ 0 & 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Les matrices suivantes sont échelonnées :

$$\begin{bmatrix} 0 & 4 & 3 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Les matrices suivantes sont échelonnées réduites :

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

**Définition d'équivalence par les lignes** Deux matrices sont **équivalentes par les lignes** si on peut obtenir l'une à partir de l'autre via une séquence d'opérations élémentaires sur les lignes.

*Remarque*

Toute matrice non-nulle est équivalent par les lignes avec au moins une matrice échelonnée, mais cette réduction n'est pas unique :

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ et } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

sont échelonnées, et équivalentes.

Cependant :

*Théorème* Toute matrice est équivalente par les lignes à exactement une matrice échelonnée réduite.

**Définition de position de pivot** Une matrice $A$ a une (unique) forme échelonnée réduite ; les emplacements de ses coefficients principaux sont les **positions de pivot** de $A$.

**Définition de colonne pivot** Une colonne de $A$ qui contient une position de pivot est une **colonne pivot**.

**Observation** Toutes les formes échelonnées de $A$ ont leur coefficient au même endroit. Donc, pour trouver les positions de pivot, il est suffisant de trouver une forme échelonnée, pas besoin de trouver la "réduite". Cela peut être compris de manière intuitive, en voyant qu'on passe par une matrice échelonnée pour obtenir une matrice échelonnée réduite.

**Exemple** Les trois matrices suivantes sont équivalentes par les lignes

$$\begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -3 & 0 & 5 \\ 0 & 1 & 2 & 0 & -3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

On peut bien voir que les coefficients principaux sont au même endroit pour la matrice échelonnée (qui n'est pas unique) que pour la matrice échelonnée réduite, qui est unique.

**Remarque** Si on met tant d'emphase sur la forme échelonnée réduite et les pivots, c'est parce que cela nous donne une description complète des solutions d'un système. Par exemple, la matrice

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

Est équivalente au système :

$$\begin{cases} x_1 - 3x_3 = 5 & (3.37a) \\ x_2 + 2x_3 = -3 & (3.37b) \\ x_4 = 0 & (3.37c) \\ 0 = 0 & (3.37d) \end{cases}$$

On peut voir que les colonnes pivot, les colonnes 1, 2 et 4, nous montrent quels variables n'apparaissent qu'une fois ($x_1$, $x_2$, $x_4$ n'apparaissent qu'une fois). On peut donc résoudre pour ces variables. Donc,

$$\begin{cases} x_1 = 5 + 3x_3 & (3.38a) \\ x_2 = -3 - 2x_3 & (3.38b) \\ x_3 \text{ est libre} & (3.38c) \\ x_4 = 0 & (3.38d) \end{cases}$$

**Solutions à partir de la forme échelonnée réduite**

Toutes les variables qui correspondent à une colonne pivot apparaissent dans exactement une équation, avec le coefficient 1. On appelle ces variables les **variables de bases** ou **variables liées**.
Les autres variables sont **les variables libres** : on peut leur donner n'importe quelle valeur, et il est alors facile de choisir les valeurs des variables liée pour obtenir une des solutions du système.

**Résumé pour résoudre un système**

La méthode qu'on utilise s'appelle le **pivot de Gauss** :
1. Écrire la matrice augmentée du système.
2. Appliquer la méthode du pivot (en utilisant les opérations élémentaires sur les lignes) pour obtenir une matrice complète équivalente sous forme échelonnée. Déterminer si le système est compatible (regarder s'il y a des équations qui disent que $1 = 0$). S'il n'y a pas de solution c'est terminé ; sinon aller à l'étape suivante.
3. Continuer la méthode du pivot pour obtenir la forme échelonnée réduite.
4. Repasser sur le système d'équations correspondant à la matrice obtenue.
5. Réécrire chaque équation non nulle, de manière à exprimer les variables liées en fonctions des variables libres.

## 3.7 Équations vectorielles

### 3.7.1 Définition et opérations

**Définition de vecteur**

$\mathbb{R}^n$ désigne l'ensemble des matrices de taille $n \times 1$ ; des matrices sous la forme :

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

On appelle un élément de $\mathbb{R}^n$, un **vecteur de $\mathbb{R}^n$**.
Les coefficients du vecteur s'appellent aussi ses composantes.

**Exemples**

On a

$$\mathbb{R}^2 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ tel que } x_1, x_2 \in \mathbb{R} \right\}$$

On peut identifier chaque vecteur avec le point de coordonnées $(x_1, x_2)$ dans le plan avec un repère cartésien. De ce point de vue, $\mathbb{R}^2$ correspond au plan, et $\mathbb{R}^3$ correspond à l'espace cartésien en 3D.

**Remarque**

On appelle aussi cela un vecteur colonne ; on aurait pu choisir de travailler avec des matrices de taille $1 \times n$m qu'on appelle vecteurs lignes. Cela va juste être plus simple pour nous avec les vecteurs colonnes.

**Opérations sur les vecteurs**

Soient $\vec{u}$ et $\vec{v}$ deux vecteurs de $\mathbb{R}^n$, on écrit $\vec{u}, \vec{v} \in \mathbb{R}^n$ et on désigne leur composantes par

$$\vec{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \text{ et } \vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

La somme de $\vec{u}$ et $\vec{v}$ est

$$\vec{u} + \vec{v} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{bmatrix}$$

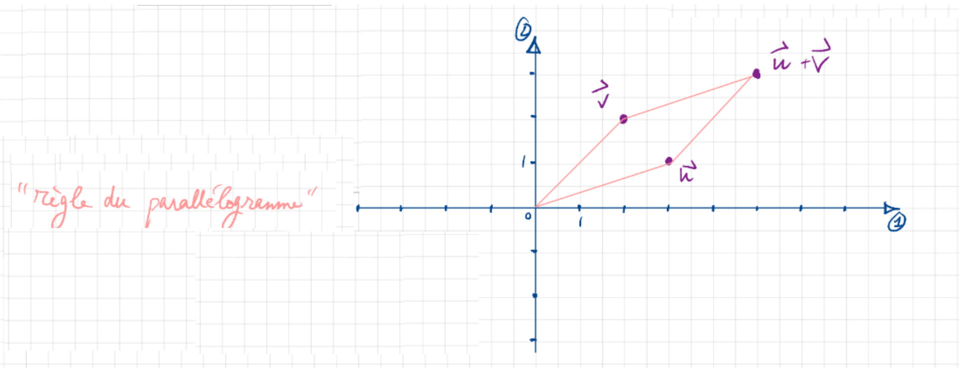
**Exemple**

Si on a

$$\vec{u} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \vec{v} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Alors,

$$\overrightarrow{u} + \overrightarrow{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3+2 \\ 1+2 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$




**Définition du produit par un scalaire**

Le produit de $\overrightarrow{u}$ par un scalaire $c \in \mathbb{R}$ est

$$c\overrightarrow{u} = c \cdot \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} cu_1 \\ \vdots \\ cu_n \end{bmatrix}$$

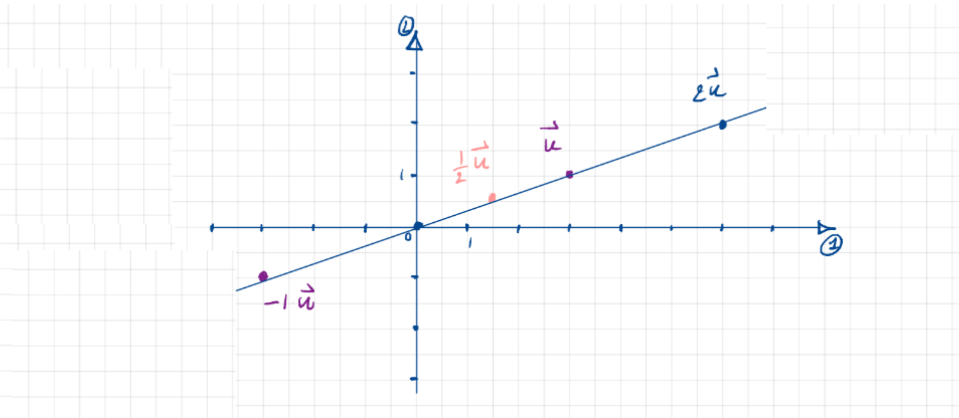
**Exemple**

Si on a

$$\overrightarrow{u} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Alors,

$$2\overrightarrow{u} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}, -1\overrightarrow{u} = \begin{bmatrix} -3 \\ -1 \end{bmatrix}, \frac{1}{2}\overrightarrow{u} = \begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \end{bmatrix}$$



**Observation**

On remarque que si $\overrightarrow{u}$ n'est pas nul, alors tous les vecteurs de la forme $c\overrightarrow{u}$ pour $c \in \mathbb{R}$ sont sur une droite qui passe par l'origine (puisque $0\overrightarrow{u} = \overrightarrow{\mathbf{0}}$).
On écrit $\overrightarrow{\mathbf{0}}$, le vecteur nul :

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

**Définition de la colinéarité**

On dit que deux vecteurs $\overrightarrow{u}$ et $\overrightarrow{v}$ sont **colinéaires** s'il existe $c \in \mathbb{R}$ tel que $\overrightarrow{u} = c\overrightarrow{v}$

**Définition de la soustraction**

La soustraction de $\overrightarrow{u}$ et $\overrightarrow{v}$ est

$$\overrightarrow{u}-\overrightarrow{v}=\begin{bmatrix}u_1-v_1\\ \vdots\\ u_n-v_n\end{bmatrix}$$

On remarque que

$$\overrightarrow{u}-\overrightarrow{v}=\overrightarrow{u}+(-1)\overrightarrow{v}$$

**Exemple**

Si on a

$$\overrightarrow{u}=\begin{bmatrix}-2\\1\end{bmatrix}\text{ et }\overrightarrow{v}=\begin{bmatrix}-4\\-1\end{bmatrix}$$

Alors,

$$\overrightarrow{u}-\overrightarrow{v}=\begin{bmatrix}-2-(-4)\\1-(-1)\end{bmatrix}=\begin{bmatrix}2\\2\end{bmatrix}$$

**Propriétés algébriques de $\mathbb{R}^n$**

Pour tout $\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\in\mathbb{R}^n$ et tout $a,b\in\mathbb{R}$ :

- $\overrightarrow{u}+\overrightarrow{v}=\overrightarrow{v}+\overrightarrow{u}$
- $(\overrightarrow{u}+\overrightarrow{v})+\overrightarrow{w}=\overrightarrow{u}+(\overrightarrow{v}+\overrightarrow{w})$
- $\overrightarrow{u}+\overrightarrow{0}=\overrightarrow{0}+\overrightarrow{u}=\overrightarrow{u}$
- $\overrightarrow{u}+(-\overrightarrow{u})=-\overrightarrow{u}+\overrightarrow{u}=\overrightarrow{0}$, où $-\overrightarrow{u}$ désigne $(-1)\overrightarrow{u}$
- $a(\overrightarrow{u}+\overrightarrow{v})=a\overrightarrow{u}+a\overrightarrow{v}$
- $(a+b)\overrightarrow{u}=a\overrightarrow{u}+b\overrightarrow{u}$
- $a(b\overrightarrow{u})=(ab)\overrightarrow{u}$
- $1\overrightarrow{u}=\overrightarrow{u}$

### 3.7.2 Combinaisons linéaires

**Définition de combinaison linéaire**

Étant donnés $p$ vecteurs $\overrightarrow{v_1},\dots,\overrightarrow{v_p}\in\mathbb{R}^n$ et $p$ scalaires $c_1,\dots,c_p\in\mathbb{R}$, on appelle $\overrightarrow{y}=c_1\overrightarrow{v_1}+\dots+c_p\overrightarrow{v_p}$ une **combinaison linéaire** de $\overrightarrow{v_1},\dots,\overrightarrow{v_p}$ avec les coefficients ou poids $c_1,\dots,c_p$.

**Exemple 1**

Si on a

$$\overrightarrow{v}_1=\begin{bmatrix}1\\0\end{bmatrix},\quad\overrightarrow{v}_2=\begin{bmatrix}0\\1\end{bmatrix}$$

On remarque que tous vecteur de $\mathbb{R}^2$ peut être écrit sous la forme d'une combinaison de ces deux vecteurs.

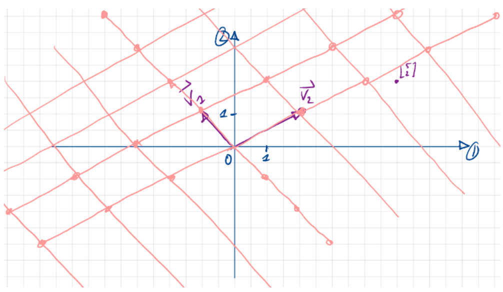
---

Jeudi 30 septembre 2021 — **Cours 3 : Équations vectorielles et matricielles**

**Exemple 2**

Dessinons les combinaisons linéaires possibles de $\overrightarrow{v}_1$ et $\overrightarrow{v}_2$ s'ils sont définis tels que

$$\overrightarrow{v}_1=\begin{bmatrix}-1\\1\end{bmatrix},\quad\overrightarrow{v}_2=\begin{bmatrix}2\\1\end{bmatrix}$$



Les combinaisons linéaires des vecteurs forment une grille, on se rend compte qu'on peut atteindre n'importe quelle vecteur.

On peut maintenant se poser la question suivante : existe-t-il $x_1, x_2 \in \mathbb{R}$ tels que $x_1 \vec{v}_1 + x_2 \vec{v}_2 = \vec{b}$ avec

$$\vec{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \text{ et } \vec{b} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

Si oui quelles valeurs de $x_1$ et $x_2$ conviennent ? On appelle ceci une **équation vectorielle**.

### 3.7.3 Équation vectorielle

**Équation vectorielle**

Nous cherchons une solution à l'équation mentionnée ci-dessus. Ainsi, on a

$$c_1 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix} \iff \begin{bmatrix} -c_1 \\ c_1 \end{bmatrix} + \begin{bmatrix} 2c_2 \\ c_2 \end{bmatrix} = \begin{bmatrix} -c_1 + 2c_2 \\ c_1 + c_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

On se retrouve donc avec le système d'équations suivantes

$$\begin{cases} -c_1 + 2c_2 = 5 & (3.40a) \\ c_1 + c_2 = 2 & (3.40b) \end{cases}$$

**Exemple en 3D avec deux vecteurs de base**

On a

$$\vec{a}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}, \vec{a}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} \text{ et } \vec{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$$

On veut savoir si $\vec{b}$ est une combinaison linéaire des deux autres vecteurs, donc on a l'équation suivante :

$$x_1 \vec{a}_1 + x_2 \vec{a}_2 = \vec{b}$$

En faisant les mêmes étapes que toute à l'heure on se retrouve avec le système

$$\begin{cases} x_1 + 2x_2 = 7 & (3.42a) \\ -2x_1 + 5x_2 = 4 & (3.42b) \\ -5x_1 + 6x_2 = -3 & (3.42c) \end{cases}$$

La matrice augmentée est donnée par

$$\begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{b} \end{bmatrix}$$

On peut donc clairement écrire la matrice directement.

**Conclusion**

Une équation vectorielle avec $\vec{a}_1, \ldots, \vec{a}_n$ et $\vec{b}$ dans $\mathbb{R}^m$

$$x_1 \vec{a}_1 + \ldots + x_n \vec{a}_n = \vec{b}$$

a le même ensemble solution que le système linéaire dont la matrice augmentée est

$$\begin{bmatrix} \vec{a}_1 & \ldots & \vec{a}_n & \vec{b} \end{bmatrix}$$

En particulier, $\vec{b}$ peut être généré comme combinaison linéaire de $\vec{a}_1, \ldots, \vec{a}_n$ si et seulement si ce système linéaire a (au moins) une solution.

**Définition** Étant donnés des vecteurs $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{R}^n$, on écrit $\text{vect}(\vec{v}_1, \ldots, \vec{v}_p)$ ou $\text{span}(\vec{v}_1, \ldots, \vec{v}_p)$ (en anglais) pour désigner **le sous-ensemble de $\mathbb{R}^n$ engendré par $\vec{v}_1, \ldots, \vec{v}_n$**, c'est-à-dire l'ensemble des vecteurs de la forme

$$c_1\vec{v}_1 + \ldots + c_n\vec{v}_n$$

où les $c_i$ peuvent être n'importe quel scalaire.

> *Remarque* Il est intéressant de noter que le vecteur nul fait toujours parti du span de n'importe quel ensemble de vecteurs.

## 3.8 Équation matricielle

**Équation matricielle** On peut rajouter une troisième manière équivalente de dire les choses.
Pour $n$ vecteurs $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^m$, on a défini la combinaison linaire. Utilisons les matrices pour simplifier la notation :

> *Définition* Soit $A$ une matrice de taille $m \times n$ (on écrit $A \in \mathbb{R}^{m\times n}$). Soient $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^m$ les colonnes de $A$, de sorte que
>
> $$A = \begin{bmatrix} \vec{a}_1 & \ldots & \vec{a}_n \end{bmatrix}$$
>
> On définit **le produit de $A$ avec tout vecteur $\vec{x}$**
>
> $$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$
>
> comme suit :
>
> $$A\vec{x} = x_1\vec{a}_1 + \ldots x_n\vec{a}_n$$
>
> Il est important de noter, que cette définition ne marche que si le nombre de colonnes de $A$ est égal au nombre de coefficients de $\vec{x}$.

**Exemple** Si on a

$$A = \begin{bmatrix} 3 & -2 & 1 \\ 2 & 9 & 8 \end{bmatrix} \text{ et } \vec{x} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$$

Alors,

$$A\vec{x} = 2\begin{bmatrix} 3 \\ 2 \end{bmatrix} - 1\begin{bmatrix} -2 \\ 9 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 8 \end{bmatrix} = \begin{bmatrix} 8 \\ -5 \end{bmatrix}$$

Si on avait un $\vec{x}$ avec des composantes quelconque (évêque) :

$$A\vec{x} = x_1\begin{bmatrix} 3 \\ 2 \end{bmatrix} + x_2\begin{bmatrix} -2 \\ 9 \end{bmatrix} + x_3\begin{bmatrix} 1 \\ 8 \end{bmatrix} = \begin{bmatrix} 3x_1 - 2x_2 + 3x_3 \\ 2x_1 + 9x_2 + 8x_3 \end{bmatrix}$$

On observe de manière plus générale que le produit $A\vec{x}$ peut être interprété en termes des lignes de $A$ également. En d'autres mots, si on a

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} \text{ et } \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Alors,

$$A\vec{x} = x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \ldots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \ldots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \end{bmatrix}$$

**Observation** Considérons le système de deux équations en trois inconnues suivant :

$$\begin{cases} 3x_1 - 2x_2 + x_3 = b_1 & (3.44a) \\ 2x_1 + 9x_2 + 8x_3 = b_2 & (3.44b) \end{cases}$$

En le transformation en équation de vecteur, on a :

$$x_1 \begin{bmatrix} 3 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 9 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 8 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Puis en équation matricielle, on a

$$\begin{bmatrix} 3 & -2 & 1 \\ 2 & 9 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = A\vec{x} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Par ailleurs, la matrice augmentée de ce système est

$$\begin{bmatrix} 3 & -2 & 1 & b_1 \\ 2 & 9 & 8 & b_2 \end{bmatrix}$$

Toutes ces notations sont équivalentes.

**En général** Soit $A \in \mathbb{R}^{m \times n}$ une matrice avec colonnes $\vec{a}_1, \ldots \vec{a}_n \in \mathbb{R}^m$, et soit $\vec{b} \in \mathbb{R}^m$, un vecteur.
L'équation matricielle

$$A\vec{x} = \vec{b}$$

est équivalent à (= a le même ensemble solution que) l'équation vectorielles

$$x_1 \vec{a}_1 + \ldots + x_n \vec{a}_n = \vec{b}$$

Les deux sont de plus équivalentes au système d'équation dont la matrice augmentée est

$$\begin{bmatrix} \vec{a}_1 & \ldots & \vec{a}_n & \vec{b} \end{bmatrix} \in \mathbb{R}^{m \times (n+1)}$$

Finalement, si on écrit

$$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}$$

ce système s'écrit encore comme :

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 & (3.46a) \\ \vdots & (3.46b) \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m & (3.46c) \end{cases}$$

**Interprétation géométrique** Un système peut s'écrire $A\vec{x} = \vec{b}$, ce qui est équivalent à l'équation

$$x_1 \vec{a}_1 + \ldots + x_n \vec{a}_n = \vec{b}$$

Donc, il existe (au moins) une solution pour $A\vec{x} = \vec{b}$ si et seulement si $\vec{b}$ est une combinaison linéaire des colonnes de $A$, c'est à dire, que $\vec{b} \in \text{vect}(\vec{a}_1, \dots, \vec{a}_n)$

**Théorème** Pour $A \in \mathbb{R}^{m \times n}$, les quatre propriétés suivantes sont équivalentes (elles sont soit toutes vraies, soit toutes fausses) :

1. Pour tout $\vec{b} \in \mathbb{R}^m$, l'équation $A\vec{x} = \vec{b}$ a (au moins) une solution.
2. Tout vecteur de $\mathbb{R}^m$ est une combinaison linéaire des colonnes de $A$.
3. Les colonnes de $A$ engendrent $\mathbb{R}^m$.
4. Il existe dans chaque ligne de $A$ une position pivot (attention, $A$ est la matrice des coefficients, pas la matrice augmentée).

Note pour (4) : on ne peut pas avoir que des zéros dans une ligne, car ça voudrait dire que la matrice échelonnée aurait un pivot dans la colonne de $\vec{x}$ (puisque, de toutes façons, il a des composantes dans chaque ligne). Donc on serait dans le cas où on aurait $0 = c$, qui est une contradiction.

**Théorème** Soit $A \in \mathbb{R}^{m \times n}$, $\vec{u}, \vec{v} \in \mathbb{R}^n$ et $c \in \mathbb{R}$. Alors,

$$A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v}$$

De manière similaire,

$$A(c\vec{u}) = cA\vec{u}$$

*Preuve* La preuve est considérée comme trivial et laissée au lecteur comme exercice.

Cependant, voici un ca particulier pour que les notation soient correctes : Supposons pour simplifier que $n = 3$ :

$$A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{bmatrix}, \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \text{ et } \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

Alors,

$$A(\vec{u} + \vec{v}) = A\left(\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\right) = A\begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{bmatrix}$$

Or, en utilisant la définition du produit matrice-vecteur, on obtient :

$$(u_1 + v_1)\vec{a}_1 + (u_2 + v_2)\vec{a}_2 + (u_3 + v_3)\vec{a}_3$$

En distribuant et en réorganisant les termes, on a

$$u_1\vec{a}_1 + u_2\vec{a}_2 + u_3\vec{a}_3 + v_1\vec{a}_1 + v_2\vec{a}_2 + v_3\vec{a}_3 = A\vec{u} + A\vec{v}$$

## 3.9 Système homogène

**Système homogène** On appelle le cas particulier suivant

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

un système **homogène**.
Cela correspond à un système de la forme

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n = 0 & (3.48a) \\ \vdots & (3.48b) \\ a_{m1}x_1 + \ldots + a_{mn}x_n = 0 & (3.48c) \end{cases}$$

Ce système a toujours une solution — appelée **solution triviale** :

$$\overrightarrow{x} = \overrightarrow{0}$$

Cela nous montre donc que les systèmes de cette forme sont toujours compatibles. Maintenant, on peut se demander quand il a d'autres solutions (donc une infinité par notre théorème).

**Théorème** L'équation homogène $A\overrightarrow{x} = \overrightarrow{0}$ admet une solution non-triviale si et seulement si elle possède au moins une variable libre (c'est-à-dire que le système a une infinité de solutions, puisqu'on sait qu'il est compatible).

**Exemple** Si on a un système

$$\begin{cases} -3x_1 + 5x_2 - 4x_3 = 0 & (3.51a) \\ -3x_1 - 2x_2 + 4x_3 = 0 & (3.51b) \\ 6x_1 + x_2 - 8x_3 & (3.51c) \end{cases}$$

On remarque que peu importe les opérations élémentaires qu'on applique sur la matrice augmentée, les zéros vont toujours rester dans lea dernière colonne. Pour notre système, on se rend compte qu'il est équivalent au système suivant :

$$\begin{cases} x_1 = \frac{4}{3}x_3 & (3.52a) \\ x_2 = 0 & (3.52b) \\ 0 = 0 & (3.52c) \end{cases}$$

On remarque donc que $x_3$ est libre. Ainsi, on peut écrire les solutions sous la forme d'un vecteur :

$$\overrightarrow{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{4}{3}x_3 \\ 0 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}$$

On voit donc clairement que l'ensemble des solutions représente une droite qui passe par l'origine (puisque $\overrightarrow{x} = \overrightarrow{0}$ est une solution).
L'ensemble solution est donc de la forme $\overrightarrow{x} = t\overrightarrow{v}$ avec $t \in \mathbb{R}$, un paramètre. Autrement dit, l'ensemble solution est la droite $\text{vect}\left(\overrightarrow{v}\right)$, qui passe par l'origine.

**Exemple 2** Si on a le système

$$\left\{ 10x_1 - 3x_2 - 2x_3 = 0 \right. \quad (3.55a)$$

En passant par une matrice augmentée, on peut la réduire (simplement en divisant par 10), puis on peut revenir en système

$$\begin{cases} x_1 = \frac{3}{10}x_2 + \frac{2}{10}x_3 & (3.56a) \\ x_2, x_3 \text{ libres} & (3.56b) \end{cases}$$

Donc, les solutions sont :

$$\overrightarrow{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{3}{10}x_2 + \frac{2}{10}x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} \frac{3}{10} \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \frac{2}{10} \\ 0 \\ 1 \end{bmatrix} = s\overrightarrow{u} + t\overrightarrow{v}$$

Où $s, t \in \mathbb{R}$ sont des paramètres. Une écriture équivalente des solution est un plan (puisque $\overrightarrow{u}$ et $\overrightarrow{v}$ ne sont pas colinéaires), le plan $\text{vect}\left(\overrightarrow{u}, \overrightarrow{v}\right)$.

Lundi 4 octobre 2021 — **Cours 4 : Dépendance linéaire**

**Dans le cas ou le système n'est pas homogène**

Il suffit de trouver **une** solution de $A\vec{x} = \vec{b}$ et **toutes** les solutions de $A\vec{x} = \vec{0}$ pour avoir **toutes** les solutions de $A\vec{x} = \vec{b}$. En effet, nous avons le théorème ci-après.

**Théorème**

Si $\vec{p}$ est solution de $A\vec{x} = \vec{b}$ (en particulier, le système est compatible), alors l'ensemble solution de $A\vec{x} = \vec{b}$ est l'ensemble des vecteurs de la forme :

$$\vec{w} = \vec{p} + \vec{h}$$

où $\vec{h}$ est une solution quelconque de $A\vec{x} = \vec{0}$ ($\vec{h}$ pour homogène).
En d'autres mots, $\vec{p}$ est une solution de $A\vec{x} = \vec{b}$ et $\vec{h}$ est une solution quelconque de $A\vec{x} = \vec{0}$ si et seulement si $\vec{w} = \vec{p} + \vec{h}$ est une solution de $A\vec{x} = \vec{b}$.

*Preuve*

**Première direction :** Si $A\vec{p} = \vec{b}$ et $A\vec{h} = \vec{0}$, alors

$$A\left(\vec{p} + \vec{h}\right) = A\vec{p} + A\vec{h} = \vec{b} + \vec{0} = \vec{b}$$

Donc, $\vec{p} + \vec{h}$ est solution de $A\vec{x} = \vec{b}$.

**Dans l'autre sens :** Si $\vec{w}$ est une solution de $A\vec{x} = \vec{b}$, alors :

$$A\vec{w} = \vec{b} \text{ et } A(\vec{w} - \vec{p}) = A\vec{w} - A\vec{p} = \vec{b} - \vec{p} = \vec{0}$$

Donc, $\vec{h} = \vec{w} - \vec{p}$ est bien une solution de $A\vec{x} = \vec{0}$. On n'a donc pas loupé de solution.

□

**Interprétation géométrique**

Géométrique, l'ensemble solution de $A\vec{x} = \vec{b}$ (s'il est non-vide) est une **translation** de celui de $A\vec{x} = \vec{0}$.

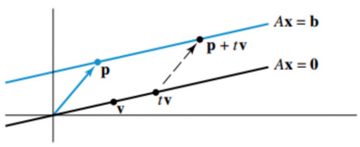



**FIGURE 5** Les ensembles de solutions de $A\mathbf{x} = \mathbf{b}$ et $A\mathbf{x} = \mathbf{0}$ sont parallèles.

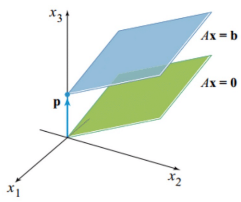



**FIGURE 6**
Les ensembles de solutions de $A\mathbf{x} = \mathbf{b}$ et $A\mathbf{x} = \mathbf{0}$ sont parallèles.

## 3.10 Indépendance linéaire

**Indépendance linéaire**

Étant donné une matrice $A$ avec colonnes $\vec{a}_1, \ldots, \vec{a}_n$, le système homogène $A\vec{x} = \vec{0}$ est équivalent à l'équation vectorielle

$$x_1\vec{a}_1 + \ldots + x_n\vec{a}_n = \vec{0}$$

Demander si $A\vec{x} = \vec{0}$ possède une solution non-triviale, est donc équivalent à demander s'il existe une façon non triviale de combiner les vecteurs $\vec{a}_1, \ldots, \vec{a}_n$ de façon à obtenir $\vec{0}$. Si c'est le cas, on dit que les vecteurs sont **linéairement dépendants**.
Plus généralement, on dit qu'une famille $\left(\vec{v}_1, \ldots, \vec{v}_p\right)$ de vecteurs de $\mathbb{R}^n$ est **libre**, ou que ses vecteurs sont **linéairement indépendants**, si l'équation vectorielle

$$x_1\vec{v}_1 + \ldots + x_p\vec{v}_p = \vec{0}$$

admet la solution trivial comme unique solution. À l'inverse, on dit que cette famille est **liée**, ou que ses vecteurs sont **linéairement dépendants** dans le cas où il existe des coefficients $x_1, \ldots, x_p$ non nuls pour certains tel que l'équation ci-dessus est vérifiée.
Les **colonnes d'une matrice** $A$ sont linéairement indépendantes si et seulement si l'équation $A\vec{x} = \vec{0}$ admet la solution triviale pour solution unique. Donc, étant donnés une famille de vecteurs, on peut vérifier s'ils sont linéairement indépendants ou non en étudiant $A\vec{x} = \vec{0}$, avec la matrice $A$ dont les colonnes sont les vecteurs donnés.

**Example** Soient

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ et } \vec{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

On prend donc la matrice

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$$

En la transformant en matrice augentée, puis en matrice échelonnée réduite, et enfin en système, on obtient

$$\begin{cases} x_1 = 0 & (3.58a) \\ x_2 = 0 & (3.58b) \\ 0 = 0 & (3.58c) \end{cases}$$

La seule solution est la solution triviale, donc $\left(\vec{v}_1, \vec{v}_2\right)$ est une famille linéairement indépendante.

**Le cas d'un unique vecteur** Une famille avec **un seul vecteur** $\vec{v}$ est linéairement indépendant si et seulement si

$$\vec{v} \neq \vec{0}$$

Cela se vérifie simplement en regardant la définition de vecteur linéairement dépendant.

**Le cas avec deux vecteurs** Une famille avec **deux vecteurs** $\left(\vec{v}_1, \vec{v}_2\right)$ est linéairement indépendante si et seulement s'ils ne sont pas colinéaires :

$$\nexists c \in \mathbb{R} \text{ tel que } c\vec{v}_1 = \vec{v}_2$$

*Démonstration* Supposons que $\vec{v}_1$ et $\vec{v}_2$ sont linéairement dépendants. Donc, il existe $c_1, c_2 \in \mathbb{R}$ pas tous les deux nuls tels que :

$$c_1\vec{v}_1 + c_2\vec{v}_2 = \vec{0}$$

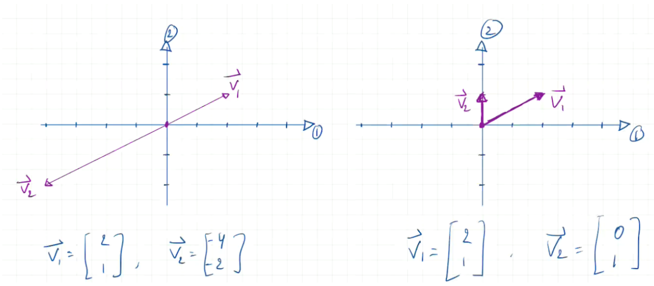
Par hypothèse, il est impossible que $c_1 = c_2 = 0$. Donc séparons notre preuve en deux cas :

- $\boldsymbol{c_1 \neq 0}$ On a simplement $\vec{v}_1 = -\frac{c_2}{c_1}\vec{v}_2$.
- $\boldsymbol{c_2 \neq 0}$ On a que $\vec{v}_2 = -\frac{c_1}{c_2}\vec{v}_1$.

Dans les deux cas, les vecteurs peuvent être écrits comme un multiple l'un de l'autre, donc ils sont colinéaires. Il n'y a pas d'autres cas possible.

□

**Illustration pour deux vecteurs**



Dans le cas où on a

$$\vec{\boldsymbol{v}}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \vec{\boldsymbol{v}}_2 = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$$

Ils sont colinéaires, donc ils sont linéairement dépendants. En effet,

$$1\vec{\boldsymbol{v}}_2 + 2\vec{\boldsymbol{v}}_1 = \vec{\mathbf{0}}$$

Cependant, si on a

$$\vec{\boldsymbol{v}}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \vec{\boldsymbol{v}}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Alors, $x_1\vec{\boldsymbol{v}}_1 + x_2\vec{\boldsymbol{v}}_2 = \vec{\mathbf{0}}$ n'a que la solution $x_1 = x_2 = 0$. Géométriquement, il est clair que si on se balade le long de $\vec{\boldsymbol{v}}_1$, puis qu'on se balade le long de $\vec{\boldsymbol{v}}_2$, il est impossible d'atteindre $\vec{\mathbf{0}}$ (sauf si on ne bouge pas du centre).

**Trois vecteurs**

Peut-on avoir trois vecteurs dans $\mathbb{R}^3$ tels que aucune paire n'est colinéaire, mais les vecteurs sont quand même linéairement dépendants ? Oui, par exemple :

$$\vec{\boldsymbol{v}}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{\boldsymbol{v}}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{\boldsymbol{v}}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Aucune paire de vecteurs n'est colinéaire, cependant

$$\vec{\boldsymbol{v}}_1 + \vec{\boldsymbol{v}}_2 - \vec{\boldsymbol{v}}_3 = \vec{\mathbf{0}}$$

Ces trois vecteurs sont tout de même un peu spéciaux, puisque $\text{vect}\left(\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \vec{\boldsymbol{v}}_3\right)$ ne donne pas $\mathbb{R}^3$, mais un plan. Cela découle directement du fait qu'ils ne soient pas linéairement indépendants.

**Théorème**

Si $\left(\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_p\right)$ sont linéairement dépendants, alors au moins un des vecteurs est une **combinaison linéaire de ceux qui le précèdent**. Cela généralise ce qu'on a dit pour deux vecteurs (qu'il faut qu'ils soient colinéaires pour être linéairement indépendants).

*Exemple* Si on reprend nos vecteurs ci-dessus :

$$\vec{\boldsymbol{v}}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{\boldsymbol{v}}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{\boldsymbol{v}}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Ils sont linéairement dépendants. Le vecteur $\vec{\boldsymbol{v}}_1$ est précédé par "rien" qui, par convention, est le vecteur nul. Or, il est impossible d'écrire $\vec{\boldsymbol{v}}_1$ sous la forme d'une combinaison linéaire du vecteur nul.

De la même manière, il est impossible d'écrire $\vec{\boldsymbol{v}}_2$ comme une combinaison linéaire du vecteur $\vec{\boldsymbol{v}}_1$.

Cependant, il est possible d'écrire le troisième vecteur comme une combinaison linéaire de $\vec{\boldsymbol{v}}_1$ et $\vec{\boldsymbol{v}}_2$ :

$$\vec{\boldsymbol{v}}_3 = \vec{\boldsymbol{v}}_1 + \vec{\boldsymbol{v}}_2$$

*Preuve* Puisque les vecteurs sont linéairement indépendants, il existe $c_1, \ldots, c_p$ non tous nuls (il en existe au moins un qui n'est pas nul) tels que

$$c_1 \overrightarrow{v}_1 + \ldots + c_p \overrightarrow{v}_p = \overrightarrow{0}$$

**Si $c_p \neq 0$**, alors on peut écrire

$$c_p \overrightarrow{v}_p = -c_1 \overrightarrow{v}_1 - \ldots - c_{p-1} \overrightarrow{v}_{p-1}$$

Puisque $c_p \neq 0$, on peut diviser par ce nombre :

$$\overrightarrow{v}_p = \frac{-c_1}{c_p} \overrightarrow{v}_1 - \ldots - \frac{c_{p-1}}{c_p} \overrightarrow{v}_{p-1}$$

Ce qui montre que $\overrightarrow{v}_p$ est combinaison linéaire de $\overrightarrow{v}_1, \ldots, \overrightarrow{v}_{p-1}$, comme souhaité.

**Si $c_p = 0$**, alors on a

$$c_1 \overrightarrow{v}_1 + \ldots + c_{p-1} \overrightarrow{v}_{p-1} = \overrightarrow{0}$$

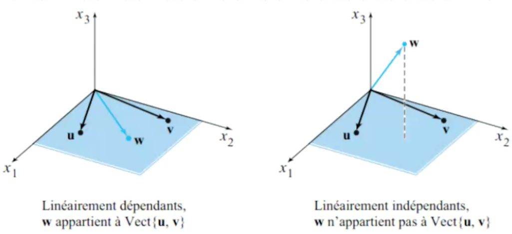
On est de retour dans notre situation de départ, donc on peut donc simplement recommencer le raisonnement avec $c_{p-1}$. Dans le cas où il est encore nul, on recommence avec $c_{p-2}$, etc. Vu qu'il y a au moins un coefficient qui est non-nul, on va forcément en avoir un à un moment.

La réciproque de ce théorème est aussi juste :

*Réciproque* Si au moins un des vecteurs $\overrightarrow{v}_1, \ldots, \overrightarrow{v}_p$ est une combinaison linéaire de ceux qui le précédent, alors $\overrightarrow{v}_1, \ldots, \overrightarrow{v}_p$ sont linéairement dépendants.

Sa démonstration est plus simple, mais elle est trouvable dans les slides du prof' si nécessaire.

*Conséquence*




Supposons que $\overrightarrow{v}_1, \ldots, \overrightarrow{v}_p$ sont linéairement indépendants. Étant donné un vecteur $\overrightarrow{w}$, les vecteurs $\overrightarrow{v}_1, \ldots, \overrightarrow{v}_p, \overrightarrow{w}$ sont linéairement dépendants **si et seulement si** $\overrightarrow{w}$ est une combinaison linéaire de $\overrightarrow{v}_1, \ldots, \overrightarrow{v}_p$.

Autrement dit, $\overrightarrow{v}_1, \ldots, \overrightarrow{v}_p, \overrightarrow{w}$ sont linéairement dépendants **si et seulement si** $\overrightarrow{w} \in \text{vect}(\overrightarrow{v}_1, \ldots, \overrightarrow{v}_p)$.

Cela se démontre facilement en utilisant le théorème vu ci-dessus. $\overrightarrow{v}_i$ ne peut pas être une combinaison linéaire des vecteurs qui le précèdent, pour n'importe quel $i$, de là on en déduit que ça doit être $\overrightarrow{w}$ qui est dépendant de ces derniers. Ce n'est qu'un sens de l'implication, mais ça donne l'idée de comment démontrer ce théorème.

**Théorème** Si un des vecteurs de $\left(\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_p\right)$ est nul, alors la famille est linéairement dépendante.

*Preuve* Si $\vec{\boldsymbol{v}}_j = \vec{\mathbf{0}}$ pour un certain $j$, alors on peut choisir $c_j = 1$ (qui n'est pas nul !), $c_1 = \ldots = c_{j-1} = 0$ et $c_{j+1} = \ldots = c_p = 0$ (tous nuls sauf $c_j$), alors

$$\underbrace{c_1 \vec{\boldsymbol{v}}_1 + \ldots}_{\vec{0}} + \underbrace{c_j \vec{\boldsymbol{v}}_j}_{1 \cdot \vec{0}} + \underbrace{\ldots + c_p \vec{\boldsymbol{v}}_p}_{\vec{0}} = \vec{\mathbf{0}}$$

□

**Théorème** Une famille $\left(\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_p\right)$ de vecteurs dans $\mathbb{R}^n$ est nécessairement linéairement dépendante si $p > n$ (si $p \leq n$, alors tout peut arriver).

*Preuve* Soit

$$A = \begin{bmatrix} \vec{\boldsymbol{v}}_1 & \ldots & \vec{\boldsymbol{v}}_p \end{bmatrix} \in \mathbb{R}^{n \times p}$$

Les colonnes sont linéairement dépendantes si (et seulement si) le système $A\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$ a une solution non-triviale. On sait que le système est compatible (car $\vec{\mathbf{0}}$ est solution). De plus, il y a seulement $n$ équations pour $p > n$ variables ; donc, au moins une des variables est libre. Puisqu'il ne peut peut pas y avoir de contradiction (le système est compatible), cela implique que $A\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$ a une infinité de solutions.

□

# Chapitre 4

# Introduction aux applications linéaires

## 4.1 Applications linéaires

**Introduction** Étant donnés un vecteur $\vec{x} \in \mathbb{R}^n$ et une matrice $A \in \mathbb{R}^{m \times n}$, on a défini le produit $A\vec{x}$ comme suit :

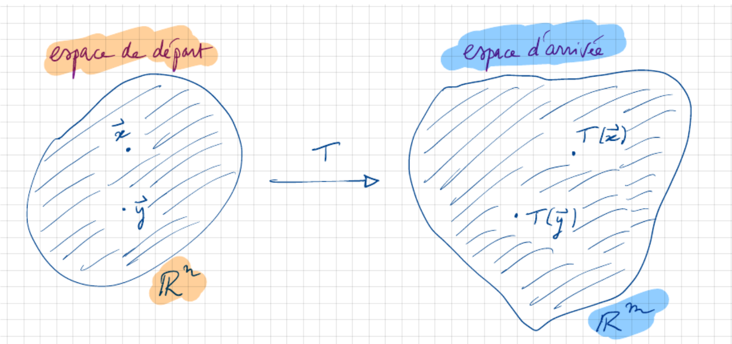
$$A\vec{x} = x_1\vec{a}_1 + \ldots + x_n\vec{a}_n$$

où $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^m$ sont les colonnes de $A$. Donc, **une matrice $A \in \mathbb{R}^{m\times n}$ permet de transformer chaque vecteur de $R^n$ en vecteur de $R^m$**

**Définition application** On appelle **application** (ou fonction, ou transformation) de $\mathbb{R}^n$ vers $\mathbb{R}^m$ une règle $T$ qui associe à tout vecteur de $\mathbb{R}^n$ un vecteur de $\mathbb{R}^m$. On écrit :

$$\begin{aligned} T : \mathbb{R}^n &\longmapsto \mathbb{R}^m \\ \vec{x} &\longmapsto T(\vec{x}) \end{aligned}$$

On appelle $\mathbb{R}^n$ l'espace de départ et $\mathbb{R}^m$ l'espace d'arrivée.




**Terminologie**

- $T(\vec{x})$ est l'image de $\vec{x}$ par $T$.
- L'image de $T$ est l'ensemble des images de tous les $\vec{x}$ par $T$.
- Étant donné une matrice $A \in \mathbb{R}^{m\times n}$, la règle $\vec{x} \mapsto A\vec{x}$ est un exemple d'application de $\mathbb{R}^n$ vers $\mathbb{R}^m$
  On appelle cela une **transformation matricielle**.

**Exemple** Si on prend

$$A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$$

Alors, on a

$$\begin{aligned} T : \mathbb{R}^2 &\longmapsto \mathbb{R}^3 \\ \vec{x} &\longmapsto T(\vec{x}) \end{aligned}$$

Où :

$$T(\vec{x}) = A\vec{x} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 3x_2 \\ 3x_1 + 5x_2 \\ -x_1 + 7x_2 \end{bmatrix}$$

Donc, par exemple,

$$\vec{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \implies T(\vec{u}) = T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 5 \\ 1 \\ -9 \end{bmatrix}$$

**Opération inverse**

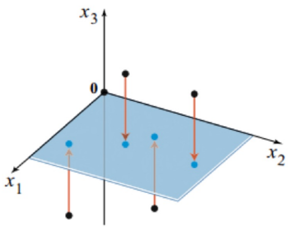
Étant donné $\vec{b} \in \mathbb{R}^m$, existe-t-il $\vec{x}$ dans l'espace de départ de $T$ tel que $T(\vec{x}) = \vec{b}$ ?
Autrement dit **le vecteur $\vec{b}$ est-il dans l'image de $T$ ?** Si un tel $\vec{x}$ existe, est-il unique ?
On peut répondre à cette question, en utilisant un système linéaire (naturellement, on est en algèbre linéaire) ! ☺

**Exemple**

Prenons

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ et } T(\vec{x}) = A\vec{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}$$

Géométriquement, $T$ **projète** les points de $\mathbb{R}^3$ sur le plan défini par $x_3 = 0$.




Donc, étant donné un vecteur $\vec{x}$, on calcule $\vec{y} = T(\vec{x})$. On sait que $T(\vec{y})$ ne va rien faire (c'est pour ça qu'on l'appelle une projection).

**Application linéaire**

On sait que les matrices ont comme propriété que

$$A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v} \text{ et } A(c\vec{u}) = cA\vec{u}$$

pour $A \in \mathbb{R}^{m\times n}$ et pour tous $\vec{u}, \vec{v} \in \mathbb{R}^n$, et pour tout $c \in \mathbb{R}$. Si on a $T(\vec{x}) = A\vec{x}$, alors cette application satisfait aussi ce propriétés.

*Définition* Une **application $T$ est linéaire** si elle suit les deux propriétés suivantes :

1. $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$
2. $T(c\vec{u}) = cT(\vec{u})$

pour tous $\vec{u}, \vec{v}$ dans l'espace de départ de $T$ et pour tout $c \in \mathbb{R}$.
On remarque que si $T(\vec{x}) = A\vec{x}$ avec $A$ étant une matrice, alors $T(\vec{x})$ est une application linéaire. On verra que, en fait, toutes les applications linéaires peuvent s'écrire sous cette forme.

**Propriétés de la transformation linéaire** La définition a ces trois conséquences directes :

1. $T\left(\vec{\mathbf{0}}\right) = \vec{\mathbf{0}}$
2. $T\left(c\vec{u} + d\vec{v}\right) = cT\left(\vec{u}\right) + dT\left(\vec{v}\right)$
3. $T\left(c_1\vec{v}_1 + \ldots + c_p\vec{v}_p\right) = c_1T\left(\vec{v}_1\right) + \ldots + c_pT\left(\vec{v}_p\right)$

En d'autres mots, la transformation linéaire d'une combinaison linéaire de vecteurs est égale à la combinaison linéaire des transformations linéaires, avec les même coefficients.

On peut les justifier comme suit :

1. $T\left(\vec{\mathbf{0}} + \vec{\mathbf{0}}\right) = T\left(\vec{\mathbf{0}}\right) + T\left(\vec{\mathbf{0}}\right) \iff T\left(\vec{\mathbf{0}}\right) = T\left(\vec{\mathbf{0}}\right) + T\left(\vec{\mathbf{0}}\right) \iff T\left(\vec{\mathbf{0}}\right) = \vec{\mathbf{0}}$
2. $T\left(c\vec{u} + d\vec{v}\right) = T\left(c\vec{u}\right) + T\left(d\vec{v}\right) = cT\left(\vec{u}\right) + dT\left(\vec{v}\right)$
3. $T\left(c_1\vec{v}_1 + \ldots + c_p\vec{v}_p\right) = T\left(c_1\vec{v}_1\right) + T\left(c_2\vec{v}_2 + \ldots + c_p\vec{v}_p\right) = \ldots = c_1T\left(\vec{v}_1\right) + \ldots + c_pT\left(\vec{v}_p\right)$

**Proposition** On sait que si $T$ est linéaire, alors $T\left(c\vec{u} + d\vec{v}\right) = cT\left(\vec{u}\right) + dT\left(\vec{v}\right)$ tient pour tous $\vec{u}, \vec{v}, c, d$. Cependant, la réciproque tient aussi (on peut utiliser cette propriété pour démontrer qu'une application est linéaire).

*Démonstration* Laissée au lecteur.

**Exemple** Soit $r \in \mathbb{R}$ et soit l'application suivante :

$$\begin{aligned} T : \mathbb{R}^2 &\longmapsto \mathbb{R}^2 \\ \vec{x} &\longmapsto r\vec{x} \end{aligned}$$

On se demande si $T$ est une application linéaire. Pour la première propriété on a :

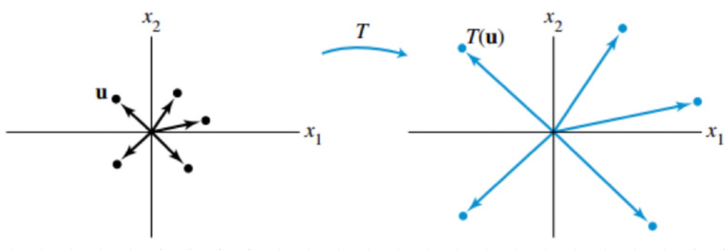
$$T\left(\vec{u} + \vec{v}\right) = r\left(\vec{u} + \vec{v}\right) = r\vec{u} + r\vec{v} = T\left(\vec{u}\right) + T\left(\vec{v}\right)$$

De la même manière, pour la deuxième propriété :

$$T\left(c\vec{u}\right) = rc\vec{u} = cr\vec{u} = cT\left(\vec{u}\right)$$

Puisque les deux propriétés tiennent, on sait que $T$ est une application linéaire.

De plus, on appelle $T$ une **homothétie du plan**. Si $r > 1$, on appelle $T$ une **dilatation**. Sinon, si $0 \leq r \leq 1$, alors on appelle $T$ est une **contraction**. $r = 0$ et $r = 1$ sont des cas spéciaux.




**Exemple** Considérons l'exemple

$$T\left(\vec{x}\right) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$$

$T$ est clairement linéaire car il est sous la forme $A\vec{x}$. Nous avons par exemple

$$T\left(\begin{bmatrix} 2 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad T\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} -2 \\ 3 \end{bmatrix} \quad \text{et} \quad T\left(\begin{bmatrix} -1/2 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} -2 \\ -1/2 \end{bmatrix}$$

Géométriquement, cette matrice applique une rotation de 90° sur les vecteurs.

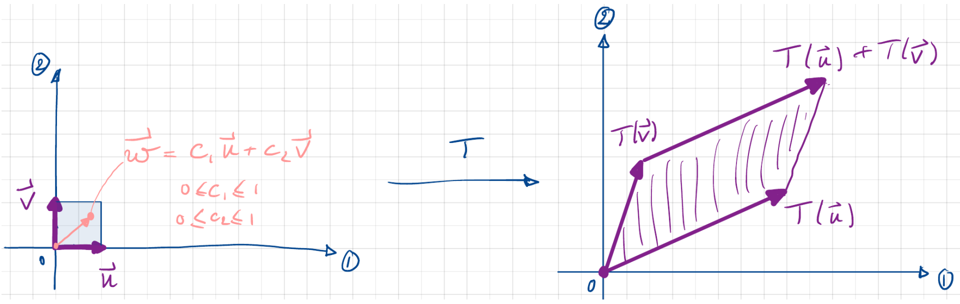
**Application sur un ensemble de points**

Disons que nous voulons calculer l'image de tous les points dans un parallélogramme à travers une application linéaire $T$.
Prenons par exemple $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ et un carré. On peut simplement calculer la transformation linéaire de deux des côtés du carré. En effet, tous les points du carré sont donnés par

$$\vec{w} = c_1 \vec{u} + c_2 \vec{v}$$

où $0 \leq c_1 \leq 1$, $0 \leq c_2 \leq 1$, et $\vec{u}$ et $\vec{v}$ sont les côtés du carré. Donc, en calculant l'image de $\vec{w}$ on a bien que :

$$T(\vec{w}) = c_1 T(\vec{u}) + c_2 T(\vec{v})$$



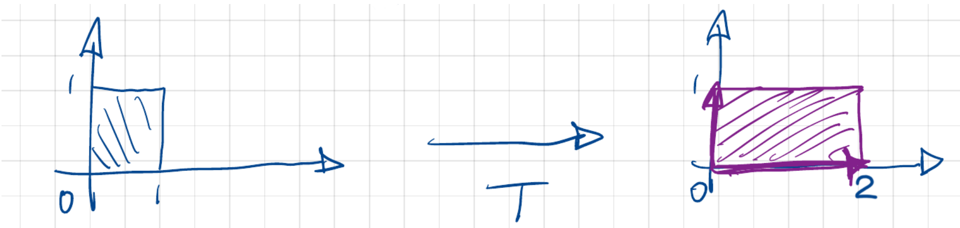

De manière générale, n'importe quelle transformation linéaire transforme un parallélogramme en parallélogramme.

**Exemple**

Disons que nous avons

$$T(\vec{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ x_2 \end{bmatrix}$$

et que nous voulons calculer l'image d'un carré de côté 1. Alors, nous aurons un rectangle deux fois plus long selon le premier axe et dont la dimension sur le premier axe n'aura pas changée.




**Matrice d"une application linéaire**

Étant donné n'importe quelle matrice $A \in \mathbb{R}^{m \times n}$. On a vu que l'application

$$\begin{aligned} T : \mathbb{R}^n &\longmapsto \mathbb{R}^m \\ \vec{x} &\longmapsto A\vec{x} \end{aligned}$$

est linéaire.
L'opposé est aussi vrai. Nous devons donc vérifier que pour toute transformation linéaire $T : R^n \mapsto R^m$, il existe une matrice $A \in \mathbb{R}^{m \times n}$ telle que $T(\vec{x}) = A\vec{x}$.

*Exemple* Définissons

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{et} \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

les colonnes de la matrice $I_2$ :

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

On considère une transformation $T : \mathbb{R}^2 \mapsto \mathbb{R}^3$ telle que

$$T\left(\vec{e}_1\right) = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} \quad \text{et} \quad T\left(\vec{e}_2\right) = \begin{bmatrix} 2 \\ 2 \\ 6 \end{bmatrix}$$

Si on sait aussi que $T$ est linéaire, alors on sait tout de $T$. En effet. pour tout $\vec{x} \in \mathbb{R}^2$, on peut écrire :

$$\vec{x} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \end{bmatrix} = x_1 \vec{e}_1 + x_2 \vec{e}_2$$

Par la linéarité de $T$, on peut en déduire que

$$T\left(\vec{x}\right) = T\left(x_1 \vec{e}_1 + x_2 \vec{e}_2\right) = x_1 T\left(\vec{e}_1\right) + x_2 T\left(\vec{e}_2\right)$$

Donc,

$$T\left(\vec{x}\right) = x_1 \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ -1 & 2 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A\vec{x}$$

Plus généralement, nous avons le théorème suivant.

**Théorème** Soit $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ une application linéaire. Il existe une unique matrice $A$ telle que

$$T\left(\vec{x}\right) = A\vec{x} \quad \forall x \in \mathbb{R}^n$$

De plus, $A$ est la matrice de taille $m \times n$ donnée par

$$A = \begin{bmatrix} \\ T\left(\vec{e}_1\right) & \dots & T\left(\vec{e}_n\right) \\ \\ \end{bmatrix}$$

où $\vec{e}_1, \ldots, \vec{e}_n$ sont les colonnes de la matrice identité $I_n$ (ils changent selon la dimension donnée par le contexte, $\vec{e}_1 = [1, 0]^T$ en deux dimensions, mais $\vec{e}_1 = [1, 0, 0]^T$ en trois dimensions). On appelle $A$ la **matrice standard** de l'application linéaire $T$.

*Preuve* Tout $\vec{x}$ peut s'écrire comme combinaison linéaire des colonnes $\vec{e}_1, \ldots, \vec{e}_n$ de la matrice identité de taille $n$ :

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} + \ldots + x_n \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} = x_1 \vec{e}_1 + \ldots + x_n \vec{e}_n$$

Puisque $T$ est linéaire, on en déduit que

$$T\left(\vec{x}\right) = T\left(x_1 \vec{e}_1 + \ldots + x_n \vec{e}_n\right) = x_1 T\left(\vec{e}_1\right) + \ldots + x_n T\left(\vec{e}_n\right)$$

On peut transformer cette somme en produit matrice-vecteur :

$$T\left(\vec{x}\right) = \begin{bmatrix} \\ T\left(\vec{e}_1\right) & \dots & T\left(\vec{e}_n\right) \\ \\ \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = A\vec{x}$$

La matrice est bien unique, puisque on a trouvé cette matrice. On a absolument aucun choix sur la manière dont on la construit, puisque $\vec{e}_1, \ldots, \vec{e}_n$ sont uniques, et les transformations linéaires transforment une valeur en une seule autre valeur (unique).

**Exemple**

L'homothétie $T(\vec{x}) = 3\vec{x}$ de $\mathbb{R}^2 \mapsto \mathbb{R}^2$ est une application linéaire. On veut trouver sa matrice standard. On a

$$T(\vec{e}_1) = T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}3\\0\end{bmatrix} \quad \text{et} \quad T(\vec{e}_2) = \begin{bmatrix}0\\3\end{bmatrix}$$

Donc, en utilisant notre théorème, on a

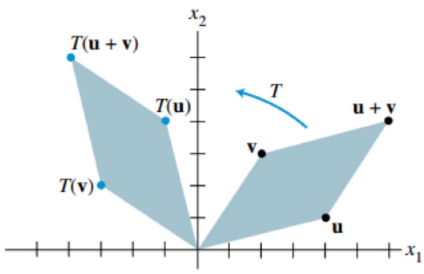
$$A = \begin{bmatrix}3 & 0\\0 & 3\end{bmatrix} = 3I_2$$

Note : on ne peut pas faire les transformations sur les lignes sur cette matrice, contrairement à ce qu'on a fait jusque là, car dans notre contexte ce n'est plus la matrice d'un système d'équation, on ne cherche plus de solution.

**Exemple 2**

Soit $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ l'application qui à tout $\vec{x} \in \mathbb{R}^2$ fait correspondre la rotation de $\vec{x}$ autour de l'origine par un angle $\theta$, dénotée $T(\vec{x})$. Cette application est linéaire. En effet,

- $T(c\vec{x}) = cT(\vec{x})$ est clair. Cela paraît logique que agrandir par $c$ un vecteur puis le faire tourner est la même chose que de faire tourner un vecteur puis de l'agrandir.
- $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$ fait du sens aussi. Voir le dessin suivant :




Pour trouver cette matrice, on a seulement besoin de trouver comment le vecteur $\vec{e}_1$ tourne et comment le vecteur $\vec{e}_2$ tourne. En faisant un peu de géométrie, et en utilisant les identités trigonométriques, on trouve : :

$$A = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$$

Lundi 11 octobre 2021 — **Cours 6 : Fin des applications, et produit matriciel**

**Exemple d'autres transformations du plan** On a vu que les rotations sont une transformation linéaire. Cependant, il y en a d'autres comme des dilatations horizontales et verticales, les réflexions selon une droite quelconque (évêque ? (ouai je la refais)), les projections, etc.

## 4.2 Injectivité et surjectivité d'applications

**Définition de surjectivité** On dit qu'une application $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ est **surjective** si tout vecteur de $\mathbb{R}^m$ est l'image *d'au moins un* vecteur de $\mathbb{R}^n$.

**Définition d'injectivité** On dit qu'une application $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ est **injective** si tout vecteur de $\mathbb{R}^m$ est l'image *d'au plus un* vecteur de $\mathbb{R}^n$. En d'autres mots, $T(\vec{x}) = \vec{b}$ a au plus une solution pour tout $\vec{b}$.

**Théorème** Soit $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ une application linéaire. $T$ est injective si et seulement si l'équation

$$T(\vec{x}) = \vec{0}$$

admet la solution triviale, $\vec{x} = \vec{0}$, pour unique solution.

*Preuve*

1. Supposons que $T$ est injective.
   On sait que de $T\left(\vec{0}\right) = \vec{0}$ puisque $T$ est linéaire. De plus, comme $T$ est injective, l'équation $T(\vec{x}) = \vec{0}$ a au plus une solution. De ce fait, elle ne peut pas avoir d'autre solution que l'équation triviale.

2. Supposons que $T(\vec{x}) = \vec{0}$ a pour seule solution $\vec{x} = \vec{0}$.
   Puisque que $T$ est une application linéaire, on sait qu'il existe une matrice $A$ telle que $T(\vec{x}) = A\vec{x}$.
   De plus, on sait que $T$ est injective si et seulement si $A\vec{x} = \vec{b}$ a au plus une solution pour tout $\vec{b}$. Comme $T(\vec{x}) = \vec{0}$ a une unique solution, $A\vec{x} = \vec{0}$ a une unique solution.
   On sait que toutes les solutions de $A\vec{x} = \vec{b}$ s'écrivent sous la forme
   $$\vec{x} = \vec{p} + \vec{h}$$
   avec $\vec{p}$, une solution de $A\vec{x} = \vec{b}$ (si elle existe) et $\vec{h}$, toutes les solutions de $A\vec{x} = \vec{0}$. Or, $\vec{h} = \vec{0}$ uniquement, donc on en déduit que $T(\vec{x}) = A\vec{x} = \vec{b}$ a au plus une solution.
   Ainsi, cela implique que $T$ est injective.

□

**Théorème** Soit $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ une application linéaire, et $A$ la matrice canoniquement associée à $T$ ($T(\vec{x}) = A\vec{x}$). Alors :

1. $T$ est surjective si et seulement si les colonnes de $A$ engendrent $\mathbb{R}^m$.
2. $T$ est injective si et seulement si les colonnes de $A$ sont linéairement indépendantes.

Pour nos preuves, prenons

$$T(\vec{x}) = A\vec{x} = x_1\vec{a}_1 + \ldots + x_n\vec{a}_n$$

où $\vec{a}_1, \ldots, \vec{a}_n$ sont les colonnes de la matrice $A$.

*Preuve de la partie (1)*

$T$ est surjective si et seulement si pour tout $\vec{\boldsymbol{b}}$ dans $\mathbb{R}^m$, il existe $\vec{\boldsymbol{x}}$ tel que $T(\vec{\boldsymbol{x}}) = \vec{\boldsymbol{b}}$. Les colonnes de $A$ engendrent $\mathbb{R}^m$ si et seulement si pour $\vec{\boldsymbol{b}}$ dans $\mathbb{R}^m$ il existe $x_1, \ldots, x_n$ tel que $x_1 \vec{\boldsymbol{a}}_1 + \ldots + x_n \vec{\boldsymbol{a}}_n = \vec{\boldsymbol{b}}$. Les deux sont donc bel et bien équivalentes.

Note personnelle : on a déjà démontré un théorème qui dit que $A\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}$ a au moins une solution si et seulement si les colonnes de $A$ engendrent $\mathbb{R}^m$, voir le théorème avec quatre propositions équivalentes.

*Preuve de la partie (2)*

Par le théorème précédent, on sait que $T$ est injective si et seulement si $T(\vec{\boldsymbol{x}}) = \vec{\mathbf{0}}$ admet la solution triviale comme unique solution. Donc, $T$ est injective si et seulement si

$$x_1 \vec{\boldsymbol{a}}_1 + \ldots + x_n \vec{\boldsymbol{a}}_n = \vec{\mathbf{0}}$$

admet $\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$ comme unique solution (ce qui est la définition d'indépendance linéaire).

□

**Résumé personnel**

Soient $A \in \mathbb{R}^{m\times n}$, une matrice, et $T$ l'application linéaire définie telle que

$$T(\vec{\boldsymbol{x}}) = A\vec{\boldsymbol{x}}$$

Les propriétés suivantes sont équivalentes :

1. $T$ est surjective.
2. Il existe dans chaque ligne de $A$ une position pivot.
3. Pour tout $\vec{\boldsymbol{b}} \in \mathbb{R}^m$, l'équation $A\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}$ a au moins une solution.
4. Tout vecteur de $\mathbb{R}^m$ est une combinaison linéaire des colonnes de $A$.
5. Les colonnes de $A$ engendrent $\mathbb{R}^m$.

De la même manière, les propriétés suivantes sont aussi équivalentes :

1. $T$ est injective.
2. Il existe dans chaque colonne de $A$ une position pivot.
3. Pour tout $\vec{\boldsymbol{b}} \in \mathbb{R}^m$, l'équation $A\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}$ a au plus une solution.
4. Les colonnes de $A$ sont linéairement indépendantes.
5. $T$ admet la solution triviale pour unique solution.
6. Aucune des colonnes de $A$ n'est une combinaison linéaire de celles qui la précèdent.

Nous avons aussi les conditions suivantes qui nous permettent de savoir instantanément si une famille de $p$ vecteurs de dimension $n$ sont linéairement dépendants :

1. Un des vecteur est le vecteur nul.
2. $p > n$.

# Chapitre 5

# Matrices

## 5.1 Opérations matricielles

**Coefficients d'une matrice**

Pour une matrice $A \in \mathbb{R}^{m \times n}$ ($m$ lignes et $n$ colonne), on note $a_{ij} \in \mathbb{R}$ le coefficient $(i, j)$ de $A$, c'est-à-dire le coefficient dans la $i$-ème ligne et $j$-ème colonne.
On note souvent les colonnes de $A$ par $\overrightarrow{\boldsymbol{a}}_1, \ldots, \overrightarrow{\boldsymbol{a}}_n \in \mathbb{R}^m$ de la manière suivante :

$$A = \begin{bmatrix} \overrightarrow{\boldsymbol{a}}_1 & \ldots & \overrightarrow{\boldsymbol{a}}_n \end{bmatrix}$$

En particulier, $a_{ij}$ est le $i$-ème coefficient du vecteur colonne $\overrightarrow{\boldsymbol{a}}_j$. Les coefficients diagonaux $a_{11}$, $a_{22}$, …forment la **diagonale principale de $A \in \mathbb{R}^{m \times n}$**

**Matrices remarquables**

Une **matrice diagonale** est une matrice carrée de taille $n \times n$ dont les entrées non diagonales sont nulles. En d'autres mots, $i \neq j \implies a_{ij} = 0$. Par exemple, la matrice identité est diagonale :

$$I_n = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

(quand un coefficient est nul, on ne le note pas forcément).
Une autre matrice utile est la matrice suivante :

$$0 = \begin{bmatrix} 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Sa taille est définie selon le contexte

**Somme de deux matrices**

Étant données deux matrices $A, B \in \mathbb{R}^{m \times n}$ de même taille (sinon la somme n'est pas définie). On définit leur somme comme la matrice de taille $m \times n$ notée $A + B$ telle que l'entrée $(i, j)$ de $A + B$ (qu'on peut aussi écrire $(A + B)_{ij}$) est égale à $a_{ij} + b_{ij}$.

*Exemple* Si on a

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 2 & 1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 1 & 7 \\ 1 & -2 & 0 \end{bmatrix}$$

Alors, leur somme est donnée par :

$$A + B = \begin{bmatrix} -1 & 0 & 0 \\ 2 & 1 & 3 \end{bmatrix} + \begin{bmatrix} 3 & 1 & 7 \\ 1 & -2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 7 \\ 3 & -1 & 3 \end{bmatrix}$$

**Multiplication d'une matrice par un scalaire** Pour une matrice $A \in \mathbb{R}^{m\times n}$ et un scalaire $r \in \mathbb{R}$, on définit leur **produit** comme la matrice de taille $m \times n$ notée $rA$ (il est habituel de mettre le scalaire à gauche de la matrice, mais ce n'est pas obligatoire) telle que l'entrée $(i, j)$ de $rA$ (aussi notée $(rA)_{ij}$) est égale à $ra_{ij}$.
On écrit aussi $-A$ pour $(-1)A$ et $A - B$ pour $A + (-1)B$.

*Exemple* Si on a

$$A = \begin{bmatrix} 3 & 2 & 0 \\ 1 & 2 & 4 \end{bmatrix} \text{ et } r = 2$$

Alors :

$$rA = 2\begin{bmatrix} 3 & 2 & 0 \\ 1 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 6 & 4 & 0 \\ 2 & 4 & 8 \end{bmatrix}$$

**Propriétés** Soit $A, B, C$ des matrices de **même taille**, et $r, s$ des scalaires. On a les propriétés suivantes :

1. $A + B = B + A$
2. $(A + B) + C = A + (B + C)$
3. $A + 0 = A$
4. $r(A + B) = rA + rB$
5. $(r + s)A = rA + sA$
6. $r(sA) = (rs)A$

**Produit de deux matrices** Le produit intuitif, celui où on deux matrices de tailles $m \times n$, et dont dont l'entrée $(i, j)$ est égale à $a_{ij}b_{ij}$, n'est pas celui qu'on va prendre. En effet, il s'appelle produit de Hadamard mais est très rarement utile. Le produit qu'on va définir ci-dessous peut sembler plus compliqué, mais il est beaucoup plus utile.
Dans le langage des transformations linéaires, une matrice est un objet qui transforme un vecteur en un autre. Donc si on transforme $\overrightarrow{\boldsymbol{x}}$ par $B$, on obtient $B\overrightarrow{\boldsymbol{x}}$. Si ensuite on utilise la matrice $A$ pour transformer ce vecteur, on va obtenir $A\big(B\overrightarrow{\boldsymbol{x}}\big)$. On peut donc se demander s'il y aurait pas une matrice qui transforme directement $\overrightarrow{\boldsymbol{x}}$ vers $A\big(B\overrightarrow{\boldsymbol{x}}\big)$. On va appeler cette matrice $AB$.
En d'autres mots, étant données deux matrices $A, B$ de tailles telles que $A\big(B\overrightarrow{\boldsymbol{x}}\big)$ est bien défini, on cherche une matrice qu'on notera $AB$ telle que $AB\overrightarrow{\boldsymbol{x}} = A\big(B\overrightarrow{\boldsymbol{x}}\big)$ pour tout $\overrightarrow{\boldsymbol{x}}$. Essayons avec $\overrightarrow{\boldsymbol{x}} \in \mathbb{R}^p$, $B \in \mathbb{R}^{n\times p}$ et $A \in \mathbb{R}^{m\times n}$. On a :

$$B\overrightarrow{\boldsymbol{x}} = x_1\overrightarrow{\boldsymbol{b}}_1 + \ldots + x_p\overrightarrow{\boldsymbol{b}}_p \in \mathbb{R}^n$$

On a pris $A \in \mathbb{R}^{m\times n}$ de manière à ce qu'on puisse encore multiplier le vecteur $B\overrightarrow{\boldsymbol{x}}$ par $A$. On a donc :

$$A\big(B\overrightarrow{\boldsymbol{x}}\big) = A\Big(x_1\overrightarrow{\boldsymbol{b}}_1 + \ldots + x_p\overrightarrow{\boldsymbol{b}}_p\Big) = x_1A\overrightarrow{\boldsymbol{b}}_1 + \ldots + x_pA\overrightarrow{\boldsymbol{b}}_p$$

par les propriétés vues ci-dessus. Donc, on en déduit $A\big(B\overrightarrow{\boldsymbol{x}}\big)$ est une combinaison linéaire des vecteurs $A\overrightarrow{\boldsymbol{b}}_1$, …, $A\overrightarrow{\boldsymbol{b}}_p$ avec $x_1, \ldots, x_p$. En notation matricielle, cela s'écrit :

$$A\big(B\overrightarrow{\boldsymbol{x}}\big) = \underbrace{\begin{bmatrix} A\overrightarrow{\boldsymbol{b}}_1 & \ldots & A\overrightarrow{\boldsymbol{b}}_p \end{bmatrix}}_{\text{On l'appelle } AB}\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} = AB\overrightarrow{\boldsymbol{x}}$$

Puisque $A$ est $m \times n$ et $B$ est $n \times p$, on a que $AB$ est $m \times p$.
Note personnelle : en écrivant les dimensions de nos matrices, il est possible de voir facilement la logique de quelle coordonnée de la dimension doit être égale, et de quelle va être la dimension du résultat :

$$\begin{matrix} A & \cdot & B & = & AB \\ m \times {\color{red}n} & & {\color{red}n} \times p & & m \times p \end{matrix}$$

**Définition du produit matriciel** Soit $A$ une matrice $m \times n$ et $B$ une matrice $n \times p$ (les $n$ doivent être les mêmes) dont on note les colonnes $\overrightarrow{\boldsymbol{b}}_1, \ldots, \overrightarrow{\boldsymbol{b}}_p$. On appelle produit de $A$ et $B$, et l'on note $AB$ la matrice $m \times p$ dont les colonnes sont $A\overrightarrow{\boldsymbol{b}}_1, \ldots, A\overrightarrow{\boldsymbol{b}}_p$, c'est-à-dire :

$$AB = A\begin{bmatrix} \overrightarrow{\boldsymbol{b}}_1 & \ldots & \overrightarrow{\boldsymbol{b}}_p \end{bmatrix} = \begin{bmatrix} A\overrightarrow{\boldsymbol{b}}_1 & \ldots & A\overrightarrow{\boldsymbol{b}}_p \end{bmatrix}$$

Cette définition est construite de sorte que le produit matriciel correspond à la composition d'applications linéaires.

**Exemple** Soient $A$, une matrice $2 \times 3$, $B$ une matrice $3 \times 2$ et $C$ une matrice $2 \times 2$. On a que $AA$ n'est pas définie, $AB$ est une matrice $2 \times 2$, $BA$ une matrice $3 \times 3$, etc.
Par exemple :

$$AB = \begin{bmatrix} -1 & 0 & 3 \\ 2 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Donc,

$$A\overrightarrow{\boldsymbol{b}}_1 = \begin{bmatrix} -1 \\ 4 \end{bmatrix} \text{ et } A\overrightarrow{\boldsymbol{b}}_2 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

Ce qui nous donne :

$$AB = \begin{bmatrix} -1 & 10 \\ 4 & 10 \end{bmatrix}$$

**Définition équivalente** Si l'on note $(AB)_{ij}$ le coefficient $(i, j)$ de $AB$ et si $A$ est une matrice $m \times n$ alors :

$$(AB)_{ij} = a_{i1}b_{1j} + \ldots + a_{1n}b_{nj}$$

**Exemple** Si on a

$$A = \begin{bmatrix} -1 & 0 & 3 \\ 2 & 1 & 4 \end{bmatrix} \text{ et } B = \begin{bmatrix} 1 & -1 \\ 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Alors :

$$(AB)_{11} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} = -1 \cdot 1 + 0 \cdot 2 + 3 \cdot 0 = -1$$

**Note personnelle** Mettre les matrices dans la position suivante, sous forme de grille, m'aide à calculer un produit matriciel (et à connaître la dimension de la matrice d'arrivée).

$$\begin{array}{ccc||cc} & & & 1 & -1 \\ & & & 2 & 0 \\ & & & 0 & 3 \\ \hline\hline -1 & 0 & 3 & -1 & 10 \\ 2 & 1 & 4 & 4 & 10 \end{array}$$

Qui représente :

$$A \cdot B = \begin{bmatrix} -1 & 0 & 3 \\ 2 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} -1 & 10 \\ 4 & 10 \end{bmatrix} = AB$$

De cette manière, les coefficients de $AB$ sont trouvés en regardant les coefficients de $A$ qui sont sur la même ligne et ceux de $B$ qui sont sur la même colonne. On a par exemple bien que :

$$(AB)_{12} = -1 \cdot (-1) + 0 \cdot 0 + 3 \cdot 3 = 1 + 9 = 10$$

**Propriétés** Soit $A$ une matrice $m \times n$ et $B$, $C$ deux matrices telles que la sommes et les produits ci-dessous aient un sens. On a :

1. $A(BC) = (AB)C$
2. $A(B+C) = AB + AC$
3. $(B+C)A = BA + CA$
4. $r(AB) = r(A)B = A(rB)$ pour tout scalaire $r$
5. $I_m A = A = A I_n$, d'où le nom "matrice identité"

*Preuve* Laissée en exercice au lecteur.

**Non-propriétés** Attention, il y a certaines propriétés qui tiennent pour les scalaires mais pas pour les matrices. Ainsi :

1. En général, $AB \neq BA$
2. On ne peut pas simplifier un produit de matrices. Autrement dit, si $AB = AC$, alors il est en général faux que $B = C$.
3. Si un produit $AB$ est égal à la matrice nulle, on ne peut pas en général en déduire que $A = 0$ ou $B = 0$.

*Exemple* On a par exemple

$$A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \neq 0, \quad B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \neq 0, \quad C = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \neq 0$$

Alors :

$$AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = AC$$

Ce qui est un contre-exemple aux propriétés 2 et 3.

**Puissance de matrice** Si $A$ est carré, alors les produits $AA$, $AAA$, ... ont du sens (ce n'est pas le cas si $A$ n'est pas carrée). Donc, on définit la notation :

$$A^k = \underbrace{AA\dots AA}_{k \text{ fois}} \qquad k \in \mathbb{N}_0$$

On appelle $A^k$, "$A$ à la puissance $k$". Par convention, on prend

$$A^0 = I_n$$

On a donc $A^k\vec{x}$ = "multiplier $\vec{x}$ $k$ fois par $A$" = "$A\dots A\vec{x}$".

## Jeudi 14 octobre 2021 — **Cours 7 : Transposées et inverses**

**Transposé d'une matrice** Soit $A \in \mathbb{R}^{m\times n}$. Sa **transposée** est $A^p \in \mathbb{R}^{n\times m}$ telle que $\left(A^T\right)_{ij} = A_{ji}$.
En d'autres mots, les lignes deviennent des colonnes et les colonnes deviennent des lignes.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad B = \begin{bmatrix} -5 & 2 \\ 1 & -3 \\ 0 & 4 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -3 & 5 & -2 & 7 \end{bmatrix}$$

$$A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \quad B^T = \begin{bmatrix} -5 & 1 & 0 \\ 2 & -3 & 4 \end{bmatrix} \quad C^T = \begin{bmatrix} 1 & -3 \\ 1 & 5 \\ 1 & -2 \\ 1 & 7 \end{bmatrix}$$

**Propriétés**

1. $\left(A^T\right)^T = A$
2. $(A+B)^T = A^T + B^T$

3. $(rA)^T = rA^T$
4. $(AB)^T = B^T A^T$

Attention, en général, $(AB)^T \neq A^T B^T$. En soit c'est logique, car on n'est même pas sûr que le produit $A^T B^T$ est défini si $AB$ l'est (par contre on est sûr que $B^T A^T$ est défini).
On peu en déduire que

$$(A_1 A_2 \ldots A_k)^T = A_k^T \ldots A_2^T A_1^T$$

*Preuve* Laissée en tant qu'exercice au lecteur.

**Vision équivalente pour le produit de matrice** Prenons la matrice suivante :

$$A = \begin{bmatrix} 3 & 0 & -1 \\ 2 & 5 & 2 \end{bmatrix}$$

Alors on que

$$\begin{bmatrix} 3 & 0 & -1 \\ 2 & 5 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & -1 \\ 2 & 5 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 0 & -1 \end{bmatrix}$$

En combinant ces résultats :

$$\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & -1 \\ 2 & 5 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = -1$$

De manière générale, si on multiplie $A$ par un vecteur colonne ayant son $j$-ème coefficient égal à 1, et tous les autres nuls, alors on va obtenir la $j$-ème colonne de $A$. De manière similaire, si on multiplie un vecteur ligne construit de la même manière par $A$, alors on va obtenir la $i$-ème ligne de $A$.
Si on utilise cette propriété pour calculer le coefficient $(i, j)$ de $AB$, on peut choisir de faire les produits dans un autre ordre (le produit est associatif). De ce fait, on peut obtenir un produit entre la $i$-ème ligne de $A$ et la $j$-ème colonne de $B$. Donc,

$$(AB)_{ij} = \begin{bmatrix} a_{i1} & \ldots & a_{in} \end{bmatrix} \begin{bmatrix} b_{ij} \\ \vdots \\ b_{nj} \end{bmatrix} = a_{1i} b_{ij} + \ldots + a_{in} b_{nj}$$

On retrouve bien la même règle.

## 5.2 Inversion de matrices

**Le cas des scalaires** L'inverse d'un scalaire non nul $a \in \mathbb{R}$ est le scalaire $c = \frac{1}{a} = a^{-1}$ tel que $ac = 1$ où 1 est l'identité pour le produit. Pour les matrices **carrées**, on généralise comme suit.

**Définition de l'inverse d'une matrice** $A \in \mathbb{R}^{n \times n}$ (carrée) est **inversible** s'il existe une matrice $C \in \mathbb{R}^{n \times n}$ telle que $AC = I_n$ et $CA = I_n$. Dans le cas où $C$ existe, on l'appelle **l'inverse** de $A$.

**Unicité** Si $A$ est inversible, la matrice $C$ telle que $CA = AC = I$ est **unique**.

*Preuve* Supposons qu'il existe une autre matrice $B \in \mathbb{R}^{n \times n}$ telle que $AB = BA = I$. Alors :

$$B = BI = B(AC) = (BA)C = IC = C$$

ce qui est une contradiction, puisqu'on avait supposé $B \neq C$.

$\square$

Dès lors, si $A$ est inversible, son inverse est donc unique et on peut la noter $A^{-1}$ (on utilise une notation qui prend avantage de cette unicité). La matrice $A^{-1}$, si elle existe, est caractérisée par

$$A^{-1}A = AA^{-1} = I$$

**Nomenclature** Une matrice non-inversible est appelée **singulière**. Une matrice inversible est appelée **non-singulière**.

**Exemple** Si on a

$$A = \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix} \quad \text{et} \quad C = \begin{bmatrix} -3 & 5 \\ 2 & -3 \end{bmatrix}$$

alors on se rend compte que $AC = CA = I$. Donc, $A$ est inversible et $C$ est son inverse. Mais aussi, $C$ est inversible et $A$ est son inverse. On a donc

$$A = C^{-1} \quad \text{et} \quad C = A^{-1}$$

On peut donc "choisir" quelle matrice est l'inverse de l'autre.

*Plus généralement* Si $A \in \mathbb{R}^{n\times n}$ est inversible, son inverse $A^{-1}$ est inversible également, et l'inverse de $A^{-1}$ est $\left(A^{-1}\right)^{-1} = A$.

**Théorème** Si $A$ est une matrice $n \times n$ inversible, alors, pour tout vecteur $\overrightarrow{\boldsymbol{b}} \in \mathbb{R}^n$, l'équation $A\overrightarrow{\boldsymbol{x}} = \overrightarrow{\boldsymbol{b}}$ admet pour unique solution le vecteur

$$\overrightarrow{\boldsymbol{x}} = A^{-1}\overrightarrow{\boldsymbol{b}}$$

*Preuve* **Est une solution** : On sait que $A^{-1}\overrightarrow{\boldsymbol{b}}$ existe, puisque $A$ est inversible par hypothèse. Or, en substituant cette égalité dans l'équation, on remarque que :

$$A\left(A^{-1}\overrightarrow{\boldsymbol{b}}\right) = \left(AA^{-1}\right)\overrightarrow{\boldsymbol{b}} = I_n\overrightarrow{\boldsymbol{b}} = \overrightarrow{\boldsymbol{b}}$$

Donc, $A\overrightarrow{\boldsymbol{x}} = \overrightarrow{\boldsymbol{b}}$ admet $A^{-1}\overrightarrow{\boldsymbol{b}}$ comme solution.

**Est unique** : Cette solution est unique. En effet, si $\overrightarrow{\boldsymbol{x}}$ est une solution quelconque, alors

$$A\overrightarrow{\boldsymbol{x}} = \overrightarrow{\boldsymbol{b}} \implies A^{-1}A\overrightarrow{\boldsymbol{x}} = A^{-1}\overrightarrow{\boldsymbol{b}} \implies \overrightarrow{\boldsymbol{x}} = A^{-1}\overrightarrow{\boldsymbol{b}}$$

qui est un vecteur bien défini et unique.

$\square$

*Corollaire* Si $A$ est inversible, alors les colonnes de $A$ sont linéairement indépendantes et elles engendrent $\mathbb{R}^n$.

**Cas particulier du 2x2** Soit la matrice suivante :

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

On peut vérifier que $A$ est inversible si et seulement si $\det(A) = ad - bc \neq 0$. Dans ce cas :

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

*Preuve*

**Si $ad - bc \neq 0$**, on a

$$\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

On peut vérifier que $A^{-1}A$ est aussi égal à la matrice identité.

**Si $ad - bc = 0$**, alors les colonnes de $A$ sont colinéaires (c'est laissé en exercice au lecteur). De ce fait, elles ne sont pas indépendantes, donc on peut utiliser la réciproque du théorème qu'on vient de voir.

**Déterminant pour les matrices 2x2** Pour une matrice $A \in \mathbb{R}^{2\times 2}$, la quantité $ad - bc$ est tellement importante qu'on lui donne un nom, on l'appel le **déterminant** de $A$, noté :

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

On a prouvé que $A$ est inversible si et seulement si $\det A \neq 0$. Plus tard, on généralisera ce résultat pour des matrice $n \times n$ avec $n$ quelconque.

**Exemple** Soit la matrice

$$A = \begin{bmatrix} 2 & 3 \\ 2 & 4 \end{bmatrix} \implies \det A = 2\cdot 4 - 3 \cdot 2 = 2 \neq 0$$

Puisque le déterminant est non nul, cette matrice est inversible, et son inverse est donnée par :

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 2 \end{bmatrix}$$

On peut vérifier notre résultat en vérifiant que $AA^{-1} = A^{-1}A = I$.

**Application aux systèmes 2x2** Si on a un système

$$\begin{cases} 2x_1 + 3x_2 = b_1 & (5.2a) \\ 2x_1 + 4x_2 = b_2 & (5.2b) \end{cases}$$

La matrice équivalente est celle de notre exemple ci-dessus. Son déterminant est non-nul et son inverse est simple à calculer. Donc,

$$\overrightarrow{x} = A^{-1}\overrightarrow{b} = \frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 4b_1 - 3b_2 \\ -2b_1 + 2b_2 \end{bmatrix}$$

Ainsi, il est maintenant très simple de résoudre un système linéaire de deux équations à deux inconnues.

**Propriétés de l'inverse**

1. Si $A$ est une matrice inversible, alors $A^{-1}$ est inversible et
$$\left(A^{-1}\right)^{-1} = A$$
2. Si $A$ et $B$ sont deux matrices $n \times n$ inversibles, alors $AB$ est inversible et
$$(AB)^{-1} = B^{-1}A^{-1}$$
3. Si $A$ est une matrice inversible, alors $A^T$ l'est aussi, et
$$\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$$

*Preuve*

1. $A^{-1}$ est inversible si et seulement s'il existe une matrice $C$ telle que
$$A^{-1}C = I \quad \text{et} \quad CA^{-1} = I$$
C'est effectivement le cas avec $C = A$, donc $\left(A^{-1}\right)^{-1} = A$.

2. $AB$ est inversible si et seulement s'il existe une matrice $C$ telle que :
$$(AB)C = I \quad \text{et} \quad C(AB) = I$$
C'est effectivement le cas avec $C = B^{-1}A^{-1}$ puisque
$$ABC = ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I$$
De plus :
$$CAB = B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I$$
3. $A^T$ est inversible si et seulement s'il existe une matrice $C$ telle que
$$A^TC = I \quad \text{et} \quad CA^T = I$$
C'est effectivement le cas avec $C = \left(A^{-1}\right)^T$ puisque :
$$CA^T = \left(A^{-1}\right)^T A^T = \left(AA^{-1}\right)^T = I^T = I$$
Le fait que $A^TC = I$ peut être démontré de la même manière.

□

**Corollaire** Le produit de matrices $n \times n$ inversibles est inversible, et l'inverse de ce produit est donné par
$$(A_1A_2 \dots A_k)^{-1} = A_k^{-1} \dots A_2^{-1}A_1^{-1}$$

*Preuve pour 3 matrices* Si on a $A, B, C$ inversibles, alors :
$$(ABC)^{-1} = (BC)^{-1}A^{-1} = C^{-1}B^{-1}A^{-1}$$

**Calcul de l'inverse** Soit $A \in \mathbb{R}^{n\times n}$. On veut calculer $A^{-1}$.
Raisonnons colonne par colonne. Soient $\vec{e}_1, \dots, \vec{e}_n$ les colonnes de $I_n$. La matrice $A^{-1}$ est de taille $n \times n$, elle a donc $n$ colonnes dans $\mathbb{R}^n$. Écrivons :
$$A^{-1} = \begin{bmatrix} \vec{y}_1 & \dots & \vec{y}_n \end{bmatrix}$$
La $k$-ème colonne est $\vec{y}_k$, qu'on peut obtenir de la manière suivante (comme on l'a vu plus haut, en début du cours) :
$$\vec{y}_k = A^{-1}\vec{e}_k \implies A\vec{y}_k = AA^{-1}\vec{e}_k = I\vec{e}_k = \vec{e}_k$$
Donc, la $k$-ème colonne de $A^{-1}$ est l'unique solution de l'équation vectorielle suivante :
$$A\vec{y}_k = \vec{e}_k$$
On peut donc résoudre cette équation pour chaque colonne de $A^{-1}$. Si $A$ n'est pas inversible, un des systèmes n'aura pas de solution unique.
Cependant, cela va nous demander de résoudre $n$ système différents, mais qui ont tous la même matrice $A$. On peut donc organiser nos calculs pour résoudre ces $n$ systèmes ensemble. En effet, pour résoudre $A\vec{x} = \vec{e}_1$, on va échelonner la matrice augmentée donnée par
$$\begin{bmatrix} A & \vec{e}_1 \end{bmatrix}$$

Sous forme échelonnée réduite, puisque le système est carré et qu'il y a une solution unique, on obtiendra

$$\begin{bmatrix} I_n & \vec{y}_1 \end{bmatrix}$$

On fera exactement le même raisonnement pour $\vec{y}_2$. Les opérations qu'on fait pour échelonner les matrices augmentées seront les mêmes, ils ne dépendent que de $A$. Donc, si on a la matrice

$$\begin{bmatrix} A & \vec{e}_1 & \ldots & \vec{e}_n \end{bmatrix} = \begin{bmatrix} A & I \end{bmatrix}$$

et qu'on la réduit, on va obtenir

$$\begin{bmatrix} I_n & A^{-1} \end{bmatrix}$$

Cet algorithme n'est pas forcément le plus efficace pour un ordinateur, mais il fonctionne !

Lundi 18 octobre 2021 — **Cours 8 : Fin des inverses et début des déterminants**

**Exemple** Si on a une matrice $A$ définie comme :

$$A = \begin{bmatrix} 2 & 3 \\ 2 & 4 \end{bmatrix}$$

On peut prendre la matrice augmentée :

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

En prenant $(b) \leftarrow (b) - (a)$ :

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

Puis on peut prendre $(a) \leftarrow (a) - 3(b)$ :

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

Enfin, on prend $(a) \leftarrow \frac{1}{2}(a)$ :

$$\begin{bmatrix} 1 & 0 & 2 & \frac{-3}{2} \\ 0 & 1 & -1 & 1 \end{bmatrix}$$

Donc on a :

$$A^{-1} = \begin{bmatrix} 2 & \frac{-3}{2} \\ -1 & 1 \end{bmatrix}$$

Ce qu'on peut vérifier simplement.

**Lien entre matrices et opération élémentaires** Pour échelonner une matrice, on effectue trois types d'opérations sur les lignes. Toutes peuvent s'écrire comme un produit matriciel avec une matrice identité à laquelle on a préalable appliqué l'opération.

*Permuter deux lignes*

Par exemple, pour permuter la ligne $(a)$ et la ligne $(b)$ dans $I_3$ :

$$E = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$E$ est la matrice identité dont la ligne $(a)$ et la ligne $(b)$ ont été permutées.
On remarque que $E$ est inversible. En effet :

$$E^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = E$$

Ce qui est logique, puisque pour faire l'inverse d'échanger deux lignes, on peut ré-échanger ces deux lignes.

*Multiplier une ligne par un coefficient non nul*

Par exemple, si on veut multiplier la ligne $(c)$ par 5 :

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

$E$ est la matrice identité, dont la ligne $(c)$ a été multipliée par 5.
On remarque que $E$ est inversible. En effet :

$$E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}$$

Pour faire l'inverse de multiplier une ligne par 5, il faut la multiplier par $\frac{1}{5}$.

*Ajouter à une ligne un multiple d'une autre ligne*

Par exemple, si on veut ajouter à la ligne $(c)$ la ligne $(a)$ multipliée par 4 :

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}$$

$E$ est la matrice identité, où quatre fois la ligne $(a)$ a été ajoutée à la ligne $(c)$ .
On remarque que $E$ est inversible. En effet :

$$E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$$

Pour annuler l'effet de $E$, il faut soustraire quatre fois la ligne $(a)$ à la ligne $(c)$.

**Exemple** Soit la matrice suivante :

$$A = \begin{bmatrix} 2 & 3 \\ 2 & 4 \end{bmatrix}$$

On veut soustraire la ligne $(a)$ de la ligne $(b)$. Donc, notre matrice $E_1$ est donnée par :

$$E_1 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \implies E_1 A = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$$

On a donc bien eu l'effet voulu. Maintenant, si on veut soustraire trois fois la ligne $(b)$ à la ligne $(a)$ :

$$E_2 = \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \implies E_2 E_1 A = E_2 \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Maintenant, on veut diviser la première ligne par 2, donc on a :

$$E_3 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \implies E_3 E_2 E_1 A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

On a donc que

$$E_3 E_2 E_1 = A^{-1}$$

*Récapitulation*

- Les matrices $E$ ci-dessus sont appelées **matrices élémentaires**.
- Elles correspondent à des opérations élémentaires sur les lignes et elles sont inversibles.
- Quand on échelonne une matrice, on la multiplie à gauche par une suite de matrices élémentaires.
- En particulière, si $A \in \mathbb{R}^{n \times n}$ est inversible, on sait qu'elle a $n$ pivots. Dès lors, sa forme échelonne réduite est $I_n$. En d'autres mots, il existes une suite d'opérations élémentaires qui réduit $A$ en $I_n$. Donc :

  $$E_k \dots E_2 E_1 A = I_n$$

  Comme $A$ est inversible, on peut multiplier par $A^{-1}$ à droite :

  $$E_k \dots E_2 E_1 A A^{-1} = I_n A^{-1} \implies A^{-1} = E_k \dots E_2 E_1$$

Donc, L'inverse de $A$ est le produit des matrices élémentaires qui réduisent $A$ sous forme échelonnée réduite. En pratique, quand on échelonne

$$\begin{bmatrix} A & I_n \end{bmatrix}$$

On multiplie à gauche par les matrices $E_1$, $E_2$, ... :

$$\begin{aligned} E_k \dots E_2 E_1 \begin{bmatrix} A & I_n \end{bmatrix} &= \begin{bmatrix} E_k \dots E_2 E_1 A & E_k \dots E_2 E_1 I_n \end{bmatrix} \\ &= \begin{bmatrix} I_n & A^{-1} \end{bmatrix} \end{aligned}$$

**Théorème** Soit $A$ une matrice carrée $n \times n$. Les propriétés suivantes sont équivalentes, c'est-à-dire qu'elles sont soit toutes vraies, soit toutes fausses.

1. $A$ est inversible.
2. $A$ est équivalente selon les lignes à la matrice unité $n \times n$, $I_n$.
3. $A$ admet $n$ positions de pivot.
4. L'équation $A\vec{x} = \vec{0}$ admet la solution triviale pour unique solution.
5. Les colonnes de $A$ sont linéairement indépendantes.
6. L'application $\vec{x} \mapsto A\vec{x}$ est injective.
7. Pour tout vecteur $\vec{b} \in \mathbb{R}^n$, l'équation $A\vec{x} = \vec{b}$ admet au moins une solution (on aurait aussi pu dire "une solution unique", ou au plus une solution).
8. Les colonnes de $A$ engendrent $\mathbb{R}^n$.
9. L'application linéaire $x \mapsto A\vec{x}$ est surjective.
10. Il existe une matrice $C$ de taille $n \times n$ telle que $CA = I_n$.
11. Il existe une matrice $D$ de taille $n \times n$ telle que $AD = I_n$.
12. $A^T$ est inversible.

Les propriétés (10) et (11) permettent de voir que notre définition de matrice inversible — que $A^{-1}$ doit être tel que $AA^{-1} = A^{-1}A = I_n$ — est trop forte. Cependant, nous en avions besoin pour démontrer ces propriétés.

## 5.3 Déterminant

**Introduction**

On a découvert qu'une matrice $2 \times 2$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

est inversible si et seulement si $ad - bc \neq 0$. On a appelé cette quantité, $ad - bc$, le déterminant de $A$, noté $\det A$.
On peut faire pareil pour une matrice $1 \times 1$, $A = \begin{bmatrix} a \end{bmatrix}$. Une telle matrice est inversible si et seulement si $a \neq 0$. On définit donc $\det A = a$.
On se demande maintenant s'il est possible de définir une quantité "$\det A$" pour $A$ de taille $n \times n$, avec $n$ quelconque, telle que $A$ est inversible si et seulement si $\det A \neq 0$.

**Définition pour une matrice $n \times n$**

Soit $A \in \mathbb{R}^{n \times n}$ :

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

Par $A_{ij}$, on désigne la matrice de taille $(n-1) \times (n-1)$ obtenu en effaçant la $i$-ème ligne et la $j$-ème colonne de $A$.
Le **déterminant de $A$** est définit récursivement comme suit :

- Si $n = 1$, $\det A = a_{11}$.
- Si $n > 1$,

$$\det A = \sum_{j=1}^{n} a_{ij} \underbrace{(-1)^{i+j} \det A_{ij}}_{\text{cofacteur } (i,j) \text{ de } A}$$

  où $i$ est une ligne qu'on a choisie. Le résultat ne dépend pas de la ligne qu'on choisit à chaque itération.

On peut se représenter le signe de la manière suivante, pour une matrice $5 \times 5$ :

$$\begin{bmatrix} + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \end{bmatrix}$$

Donc, si on a $i = 4$ et $j = 3$, alors on a que $(-1)^{4+3} = -1$.

**Exemple pour une matrice $2 \times 2$**

Si on choisit $i = 1$, on a, pour $j = 1$ :

$$A_{11} = \begin{bmatrix} a_{22} \end{bmatrix} \implies \det A_{11} = a_{22}$$

Ensuite, pour $j = 2$ :

$$A_{12} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \implies \det A_{12} = a_{21}$$

Donc, on a que :

$$\det A = a_{11}(-1)^{1+1} \det A_{11} + a_{12}(-1)^{1+2} \det A_{12} = a_{11}a_{22} - a_{12}a_{21} = ad - bc$$

Si on fait le même raisonnement, mais en choisissant $i = 2$, alors on obtiendra exactement la même chose.

**Exemple pour une matrice $3 \times 3$** Soit une matrice $3 \times 3$ quelconque (ok je ferai pas la blague cette fois, mais est-ce que je l'ai faite en disant que je n'allais pas la faire ?) :

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

Choisissons $i = 1$. On a pour $j = 1$ :

$$A_{11} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} \implies \det A_{11} = a_{22}a_{33} - a_{23}a_{32}$$

Pour $j = 2$ :

$$A_{12} = \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} \implies \det A_{12} = a_{21}a_{33} - a_{23}a_{31}$$

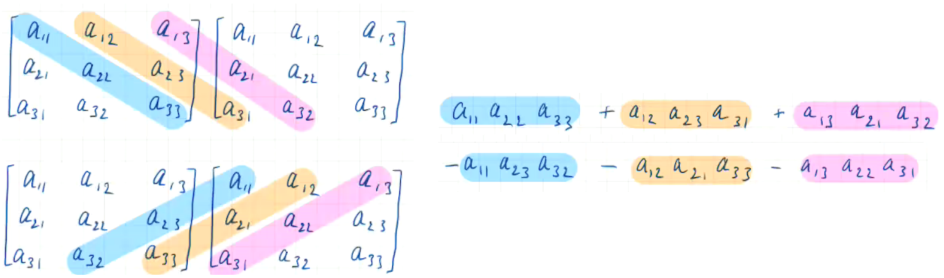
Pour $j = 3$ :

$$A_{13} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \implies \det A_{13} = a_{21}a_{32} - a_{22}a_{31}$$

Ainsi, on peut obtenir le déterminant :

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

On peut utiliser le moyen mnémotechnique suivant (valable uniquement pour les matrices $3 \times 3$) :




**Théorème** Pour $A \in \mathbb{R}^{n \times n}$ quelconque, on a

$$\det A = \det A^T$$

*Conséquence* Dans la définition du déterminant, on peut choisir de calculer "l'expansion en cofacteurs" du déterminant le long d'une ligne **ou** d'une colonne au choix.

*Exemple* Si on a la matrice

$$A = \begin{bmatrix} 0 & 1 & 3 & 0 \\ 2 & 0 & 1 & 2 \\ 0 & 1 & 4 & 2 \\ 0 & 1 & 1 & 1 \end{bmatrix} \iff A^T = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 3 & 1 & 4 & 1 \\ 0 & 2 & 2 & 1 \end{bmatrix}$$

Utiliser la première ligne de $A^T$, c'est à dire la première colonne de $A$, est beaucoup plus simple car il y a beaucoup de zéros.

**Le cas d'une matrice triangulaire** Si on a $A$, une matrice triangulaire, par exemple :

$$A = \begin{bmatrix} 5 & 2 & 3 & 2 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & 3 & 7 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Dans ce cas, si on fait l'expansion selon la première colonne, on a que

$$\det A = 5 \det \begin{bmatrix} -2 & 1 & 1 \\ 0 & 3 & 7 \\ 0 & 0 & 1 \end{bmatrix} + 0 \ldots + 0 \ldots + 0 \ldots$$

qui est à nouveau une matrice triangulaire. En répétant cette idée plusieurs fois, on se trouve avec le fait que

$$\det A = 5 \cdot (-2) \cdot 3 \cdot 1$$

soit le produit des éléments sur la diagonale de $A$.
En d'autres mots, si on a une matrice triangulaire supérieure, triangulaire inférieure ou diagonale, alors son déterminant est donné par le produit des coefficients diagonaux.

## Jeudi 21 octobre 2021 — **Cours 9 : Détermination(.) du déterminant**

**Théorème** Soit $A$ une matrice carrée.

1. Si l'on **ajoute à une ligne** de $A$ un multiple d'une autre ligne, alors la matrice $B$ obtenue vérifie
$$\det B = \det A$$
2. Si l'on **échange deux lignes** de $A$, alors la matrice $B$ obtenue vérifie
$$\det B = -\det A$$
3. Si l'on **multiplie une ligne** de $A$ par $k$, alors la matrice $B$ obtenue vérifie
$$\det B = k \det A$$

> *Remarque* Ce théorème est très pratique pour calculer un déterminant d'une matrice. On peut transformer notre matrice $A$ en matrice diagonale, puis il est très simple de retrouver le déterminant si on a pris notes des opérations qu'on a utilisée.

**Exemple pour une matrice $2 \times 2$** Si on a une matrice

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies \det A = ad - bc$$

**(1) :** Alors, si on ajoute $k$ fois la première ligne à la seconde ligne :

$$\det \begin{bmatrix} a & b \\ c + ka & d + ka \end{bmatrix} = a(d + kb) - b(c + ka) = ad + akb - bc - bka = ad - bc = \det A$$

Or, pour faire cette transformation on a utilisé la matrice $E$ :

$$E = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \implies \det E = 1$$

puisque la matrice est diagonale. On observe que, ici :

$$\det(EA) = \det(E) \det(A)$$

**(2) :** Si on permute les deux premières lignes, alors :

$$\det \begin{bmatrix} c & d \\ a & b \end{bmatrix} = cb - da = -(ad - bc) = -\det A$$

La matrice élémentaire qu'on a utilisée est donnée par $E$ :

$$E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \implies \det E = -1$$

On remarque que, à nouveau :

$$\det(EA) = \det(E)\det(A)$$

**(3) :** Si on multiplie la deuxième ligne par $k \neq 0$ :

$$\det \begin{bmatrix} a & b \\ kc & kd \end{bmatrix} = akd - bkc = k(ad - bc) = k \det A$$

La matrice élémentaire que nous avons utilisée est :

$$E = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix} \implies \det E = k$$

Or, encore une fois :

$$\det(EA) = \det(E)\det(A)$$

**Théorème (multiplication de matrice élémentaire)** Si $E$ est une matrice élémentaire, et $A$ est une matrice quelconque, alors

$$\det(EA) = \det(E)\det(A)$$

> *Généralisation* On se rendra compte que c'est vrai pour n'importe quelle matrice $A$ et $B$.

**Exemple avec des nombres** Prenons la matrice suivante

$$A = \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$$

On veut ajouter deux fois la première ligne à la deuxième ligne :

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \implies E_1 A = \begin{bmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ -1 & 7 & 0 \end{bmatrix}$$

On veut maintenant ajouter la première à la troisième ligne :

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \implies E_2 E_1 A = \begin{bmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ 0 & 3 & 2 \end{bmatrix}$$

Finalement, on peut échanger les deux dernières lignes :

$$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \implies E_3 E_2 E_1 A = \begin{bmatrix} 1 & -4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & -5 \end{bmatrix} = U$$

On a bien obtenu une matrice triangulaire, $U$. On a donc :

$$U = E_3 E_2 E_1 A \iff \det(U) = \det(E_3 E_2 E_1 A)$$

Pour calculer le déterminant de $U$, on peut multiplier les coefficients de sa ligne :

$$\det(U) = (1)(3)(-5) = -15$$

Or, pour trouver le déterminant du produit, on peut prendre le produit des déterminants :

$$\det(E_3E_2E_1A) = \det(E_3)\det(E_2E_1A) = \ldots = \det(E_3)\det(E_2)\det(E_1)\det(A)$$

Ce calcul est relativement simple à faire :

$$\det(E_3)\det(E_2)\det(E_1)\det(A) = (-1)(1)(1)\det(A) = -\det(A)$$

On peut donc en déduire que

$$\det(U) = \det(E_3E_2E_1A) \iff -15 = -\det(A) \iff \det(A) = 15$$

**Théorème** $A \in \mathbb{R}^{n\times n}$ est inversible si et seulement si $\det(A) \neq 0$.

*Preuve* Soit $M \in \mathbb{R}^{n\times n}$ la forme échelonnée réduite de $A$. Il existe une séquence de matrices élémentaires $E_1, \ldots, E_k \in \mathbb{R}^{n\times n}$ telles que

$$M = E_k \cdots E_1 A$$

On peut par exemple les trouver avec la méthode du pivot. On en déduit que :

$$\det(M) = \det(E_k \cdots E_1 A) = \det(E_k)\det(E_{k-1}\cdots E_1 A)$$

En répétant cette idée :

$$\det(M) = \underbrace{\det(E_k)}_{\neq 0} \cdots \underbrace{\det(E_1)}_{\neq 0} \det(A)$$

Par la “zero-product property”, cela veut dire que soit $\det(M) = 0 = \det(A)$, soit ils sont les deux non-nuls. En d'autres mots :

$$\det(A) \neq 0 \iff \det(M) \neq 0$$

**Si $A$ est inversible**, alors elle a $n$ pivots. Donc, $M = I_n$. Ceci implique alors que :

$$\det(M) = \det(I_n) = 1 \implies \det(A) \neq 0$$

**Si $A$ n'est pas inversible**, alors elle a strictement moins que $n$ pivots. Donc, $M$ a au moins une ligne de zéros. Ainsi :

$$\det M = 0 \implies \det A = 0$$

On a donc bien démontré que $A$ est inversible si et seulement si $\det A \neq 0$.

□

**Théorème (produit de déterminant)** Pour $A, B \in \mathbb{R}^{n\times n}$ (matrices carrées de même taille), alors on a :

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

*Preuve* **Si $A$ n'est pas inversible**, alors $\det(A) = 0$, et $AB$ n'est pas inversible (la preuve de ce deuxième point est laissée en exercice au lecteur). On en déduit donc que $\det(AB) = 0$. Or, on sait aussi que

$$\det(A) = 0 \implies \det(A)\det(B) = 0 = \det(AB)$$

Donc notre propriété tient bien.

**Si $A$ est inversible**, alors sa forme échelonnée réduite est $I_n$, et il existe une séquence de matrices élémentaires $E_1, \ldots, E_k$ telles que :

$$E_k \cdots E_1 A = I_n \implies A^{-1} = E_k \cdots E_1$$

Ceci implique que :

$$A = (E_k \cdots E_1)^{-1} = E_1^{-1} \cdots E_k^{-1}$$

Or, on sait que l'inverse d'une matrice élémentaire est une matrice élémentaire (on le sait puisqu'une opération élémentaire est toujours inversible par une autre opération élémentaire). Donc, on a :

$$\begin{aligned} \det(AB) &= \det\left(E_1^{-1} \cdots E_k^{-1} B\right) \\ &= \det\left(E_1^{-1}\right) \cdots \det\left(E_k^{-1}\right) \det(B) \\ &= \det\left(E_1^{-1} \cdots E_k^{-1}\right) \det(B) \\ &= \det(A) \det(B) \end{aligned}$$

□

**Déterminant de l'inverse** Si $A \in \mathbb{R}^{n \times n}$ est inversible, alors on a que :

$$1 = \det(I_n) = \det\left(AA^{-1}\right) = \det(A) \det\left(A^{-1}\right)$$

Donc :

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

**Déterminant d'une puissance** Si on a $A \in \mathbb{R}^{n \times n}$ et $k \in \mathbb{N}^*$, alors :

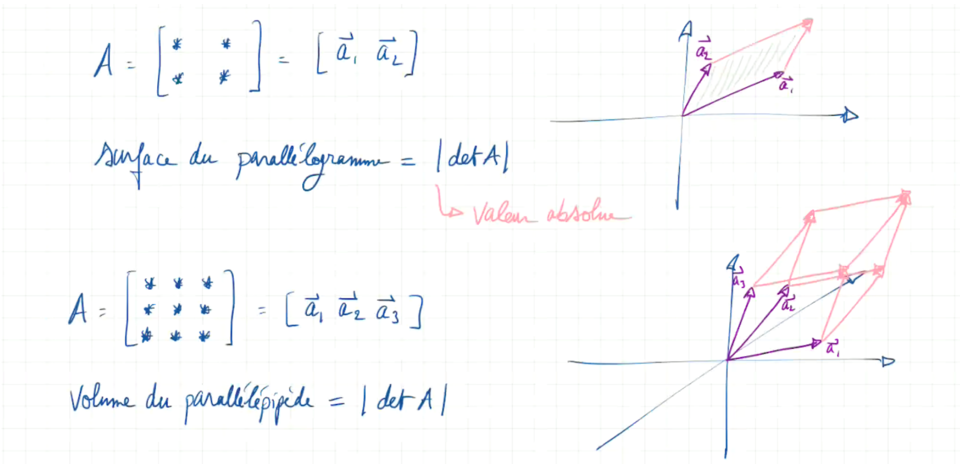
$$\det\left(A^k\right) = \det(A \cdots A) = \underbrace{\det(A) \cdots \det(A)}_{k \text{ fois}} = k \det(A)$$

*Inversibilité de la puissance* Il est donc intéressant de voir que si $A$ n'est pas inversible, alors $A^k$ ne le sera jamais. De la même manière, si $A$ est inversible, alors $A^k$ l'est forcément.

**Déterminant du produit avec un scalaire** Si on a $A \in \mathbb{R}^{n \times n}$ et $k \in \mathbb{R}$, alors :

$$\det(kA) = \det(kI_n A) = \det(kI_n) \det(A) = k^n \det(A)$$

**Théorème (interprétation géométrique)** Si $A$ est une matrice $2 \times 2$, l'aire du parallélogramme défini par les colonnes de $A$ est égale à $|\det A|$. Si $A$ est une matrice $3 \times 3$, le volume du parallélépipède défini par les colonnes de $A$ est égal à $|\det A|$.

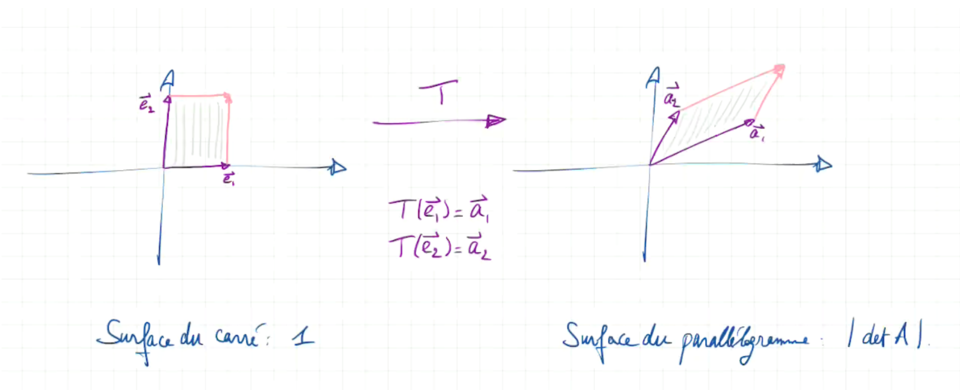



*Notation* Il arrive que $\det A$ soit noté $|A|$. Il ne faut pas confondre cette notation avec la valeur absolue d'un scalaire, comme $|\det A|$.

*Remarque* Dans le cas $2 \times 2$, si les colonnes de la matrice sont colinéaire, alors l'aire du parallélogramme qu'ils dessinent est nul. Or, s'ils sont colinéaires ont sait que la matrice n'est pas inversible, ce qui est donc cohérent.

On peut faire la même réflexion avec une matrice $3 \times 3$, dans le cas où des vecteurs sont coplanaires.

**Le cas des transformations linéaires**




On peut voir qu'une transformation linéaire $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ définie comme $T(\overrightarrow{x}) = A\overrightarrow{x}$ a pour effet de multiplier la surface par $|\det A|$, dans le cas d'un carré.
De manière générale, si $S$ est une région de $\mathbb{R}^2$ de surface finie, alors

$$\text{surface}(T(S)) = |\det A| \text{surface}(S)$$

On a le résultat analogue dans $\mathbb{R}^3$ :

$$\text{volume}(T(S)) = |\det A| \text{volume}(S)$$

# Chapitre 6

# Espaces et sous-espaces vectoriels

## 6.1 Espaces vectoriels

**Introduction**

Avec ce chapitre, nous faisons un bon en abstraction. Cela demande un effort intellectuel important. Cet effort sera récompensé parce qu'il va rendre nos outils mathématiques bien plus largement applicables.
Jusqu'ici nous avons travaillé dans

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ tel que } x_1, \ldots, x_n \in \mathbb{R} \right\}$$

On a appelé $\vec{x} \in \mathbb{R}^n$ un vecteur.
Sans y prêter attention, nous avons utilisé les propriétés suivantes de $\mathbb{R}^n$, valides pour tout $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$ et $c, d \in \mathbb{R}$ :

- $\vec{u} + \vec{v} \in \mathbb{R}^n$
- $\vec{u} + \vec{v} = \vec{v} + \vec{u}$
- $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$
- $\vec{u} + \vec{0} = \vec{u}$
- $\vec{u} + (-\vec{u}) = \vec{0}$
- $c\vec{u} \in \mathbb{R}^n$
- $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$
- $(c+d)\vec{u} = c\vec{u} + d\vec{u}$
- $c(d\vec{u}) = (cd)\vec{u}$
- $1\vec{u} = \vec{u}$

**Généralisation**

En réalité, ce sont là *toutes* les propriétés de $\mathbb{R}^n$ que nous avons utilisées. Donc, si au lieu de $\mathbb{R}^n$ on a un autre ensemble $V$ muni d'opérations "d'addition de deux éléments" et de "multiplication par un scalaire", et que ces opérations ont toutes les propriétés ci-dessus, alors on peut refaire tout notre travail des semaines précédentes dans $V$ plutôt que dans $\mathbb{R}^n$.
On aura les notions de combinaison linéaire, d'indépendance linéaire, d'espace engendré ("vect"), d'application linéaire, etc.
On appelle un ensemble $V$ muni de telles opérations un **espace vectoriel**. Un **vecteur** est un élément $\vec{v}$ dans $V$.

**Définition**

On appelle un **espace vectoriel** tout ensemble non vide $V$ constitué d'objets appelés vecteurs, sur lequel sont définies deux opérations appelées addition et multiplication par un scalaire (nombre réel). Ces opérations vérifient les dix axiomes énumérées ci-après, quels que soient les vecteurs $\vec{u}$, $\vec{v}$ et $\vec{w}$ de $V$ et les scalaires $c$ et $d$ :

1. La somme de $\vec{u}$ et $\vec{v}$, notée $\vec{u} + \vec{v}$, est dans $V$.
2. $\vec{u} + \vec{v} = \vec{v} + \vec{u}$
3. $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$
4. Il existe un vecteur de $V$ dit vecteur nul, ou zéro, noté $\vec{0}$, tel que $\vec{u} + \vec{0} = \vec{u}$
5. Pour tout vecteur de $\vec{u}$ de $V$, il existe un vecteur $-\vec{u}$ de $V$ tel que $\vec{u} + (-\vec{u}) = \vec{0}$.

6. Le produit du vecteur $\overrightarrow{u}$ par le scalaire $c$, noté $c\overrightarrow{u}$, est dans $V$.
7. $c(\overrightarrow{u}+\overrightarrow{v})=c\overrightarrow{u}+c\overrightarrow{v}$
8. $(c+d)\overrightarrow{u}=c\overrightarrow{u}+d\overrightarrow{u}$
9. $c(d\overrightarrow{u})=(cd)\overrightarrow{u}$
10. $1\overrightarrow{u}=\overrightarrow{u}$

En d'autres mots, la somme de deux vecteurs doit suivre ces 5 axiomes :

Somme de deux vecteurs :
1. Stable.
2. Commutative.
3. Associative.
4. Existence du vecteur nul.
5. Existence de l'inverse.

Produit par un scalaire :
1. Stable.
2. Distributive sur une somme de vecteur.
3. Distributive sur une somme de scalaire.
4. Associative.
5. Existence de l'unité.

**Exemple 1** Si on a $V=\mathbb{R}^n$, avec l'addition et la multiplication par un scalaire habituelles, alors c'est bien un espace vectoriel.

**Exemple 2** Prenons $V=\mathbb{R}^{m\times n}$ avec l'addition et la multiplication par un scalaire habituelles pour des matrices.
On pourrait démontrer que les axiomes tiennent pour cet ensemble et ces opérations, donc $\mathbb{R}^{m\times n}$ est un espace vectoriel, et les vecteurs sont les matrices.

**Exemple 3** Si on prend $V=\mathbb{P}_n$, l'ensemble des polynômes de degré $n$ ou moins, avec les opérations décrites ci-dessus.
Un élément $p$ de $\mathbb{P}_n$ est un polynôme de degré au plus $n$, c'est-à-dire que $p$ est une fonction de $\mathbb{R}$ vers $\mathbb{R}$ telle que

$$p(t)=a_0+a_1t+a_2t^2+\ldots+a_nt^n$$

pour tout $t\in\mathbb{R}$ avec certains coefficients fixés $a_0,\ldots,a_n\in\mathbb{R}$.
Alors on définit leur somme notée $p+q$ comme étant le polynôme tel que

$$(p+q)(t)=p(t)+q(t)=a_0+\ldots+a_nt^n+b_0+\ldots+b_nt^n=(a_0+b_0)+\ldots+(a_n+b_n)t^n$$

Pour tout $t\in\mathbb{R}$, on remarque que $p+q$ est dans $\mathbb{P}_n$.
Étant donné un scalaire $c\in\mathbb{R}$, on définit le produit de $c$ et $p$, noté $cp$, comme étant le polynôme tel que

$$(cp)(t)=cp(t)=(ca_0)+\ldots(ca_n)t^n$$

Pour tout $t\in\mathbb{R}$, on remarque que $cp$ est dans $\mathbb{P}_n$.
Il est clair que la somme des polynômes est commutative et associatives.
Soit $z$ le polynôme tel que $z(t)=0$ pour tout $t\in\mathbb{R}$. Notons que $z$ est dans $\mathbb{P}_n$, et que $(p+z)(t)=p(t)+z(t)=p(z)$, donc $p+z=p$.
On définit $-p=(-1)p$.
On peut démontrer toutes les autres propriétés, donc $\mathbb{P}_n$ est un espace vectoriel.

**Propriétés** Soit $V$, un espace vectoriel. Alors, nous avons les propriétés suivantes :
1. Le vecteur nul $\overrightarrow{0}$ est unique.
2. Pour $\overrightarrow{u}\in V$ donné, le vecteur $-\overrightarrow{u}$ est unique.
3. $0\overrightarrow{u}=0$
4. $c\overrightarrow{0}=\overrightarrow{0}$
5. $-\overrightarrow{u}=(-1)\overrightarrow{u}$

*Preuve* Laissée au lecteur.

## 6.2 Sous-espaces vectoriels

**Définition de sous-espace vectoriel**

On appelle un **sous-espace vectoriel**, ou, en abrégé, sous-espace, d'un espace vectoriel $V$ toute sous-ensemble $H$ de $V$ possédant les trois propriétés suivantes :

1. Le vecteur nul de $V$ appartient à $H$.
2. $H$ est stable par l'addition vectorielle, i.e :
$$\vec{u}, \vec{v} \in H \implies \vec{u} + \vec{v} \in H$$
3. $H$ est stable par la multiplication par un scalaire, i.e, pour tout scalaire $c$ :
$$\vec{u} \in H \implies c\vec{u} \in H$$

> *Remarque* Il est presque suffisant d'avoir les propriétés (2) et (3) pour avoir un sous-espace, puisque $0 \cdot \vec{u} = \vec{0}$. Cependant, cela pose problème si $H$ est complètement vide. Uniquement dans ce cas, les propriétés (2) et (3) tiennent, mais la propriété (1) ne tient pas.

**Exemple 1**

Soit $V = \mathbb{R}^n$ et
$$H = \left\{\vec{x} \in \mathbb{R}^n \text{ tel que } x_1 + \ldots + x_n = 0\right\}$$
On remarque que les trois propriétés tiennent :

1. $\vec{0}$ est dans $H$. En effet :
$$x_1 + \ldots + x_n = 0 + \ldots + 0 = 0$$
2. Si $\vec{u}, \vec{v} \in H$, alors $\vec{u} + \vec{v} \in H$, en effet :
$$u_1 + \ldots + u_n = 0 \text{ et } v_1 + \ldots + v_n = 0 \implies (u_1 + v_1) + \ldots + (u_n + v_n) = 0$$
3. Si $\vec{u} \in H$ et $c$ est un scalaire, alors $c\vec{u} \in H$, en effet :
$$u_1 + \ldots + u_n = 0 \implies cu_1 + \ldots + cu_n = c(u_1 + \ldots + u_n) = c \cdot 0 = 0$$

Donc, $H$ est un sous-espace vectoriel de $\mathbb{R}^n$.

**Non-exemple 1**

Soit $V = \mathbb{R}^n$ et
$$H = \left\{\vec{x} \in \mathbb{R}^n \text{ tel que } x_1 + \ldots + x_n = 1\right\}$$
Ce n'est pas un sous-espace vectoriel de $\mathbb{R}^n$, puisqu'il n'a pas le vecteur nul.

**Non-exemple 2**

Soit $V = \mathbb{R}^n$ et
$$H = \left\{\vec{x} \in \mathbb{R}^n \text{ tel que } x_1 \geq 0, \ldots, x_n \geq 0\right\}$$
Cependant, si on prend $\vec{u} \in H$, $\vec{u} \neq \vec{0}$, et qu'on le multiplie par un scalaire négatif, alors il n'appartient pas à $H$. La troisième propriété ne tient donc pas.

**Exemple 2**

Soit $V = \mathbb{R}^{n \times n}$ et
$$H = \left\{A \in \mathbb{R}^{n \times n} \text{ tel que } A = A^T\right\}$$
On appelle de telles matrices, des matrices symétriques. Alors, on a :

1. $\vec{0} \in H$ puisque la matrice nulle est égale à elle-même après avoir été transposée ($0^T = 0$).
2. Si $A, B \in \mathbb{R}^{n \times n}$ sont dans $H$, alors $C = A + B \in H$. En effet :
$$(A + B)^T = A^T + B^T = A + B$$
3. Si $A \in \mathbb{R}^{n \times n}$ est dans $H$ et $c \in \mathbb{R}$ est un scalaire, alors $B = cA$ est dans $H$. En effet :
$$B^T = (cA)^T = cA^T = cA = B$$

Donc, l'ensemble des matrices symétriques de taille $n \times n$ est un sous-espace vectoriel de $\mathbb{R}^{n \times n}$.

**Exemple 3** Soit $V = \mathbb{P}_5$, et $H = \mathbb{P}_3$. On a bien que $H$ est un sous-ensemble de $V$. De plus :

1. $z(t) = 0$ pour tout $t \in \mathbb{R}$ définit bien un polynôme nul, qui est dans $\mathbb{P}_5$ et dans $\mathbb{P}_3$.
2. Pour tous $p, q \in \mathbb{P}_3$, on a bien que $p + q \in \mathbb{P}_3$ (on pourrait le démontrer plus formellement, mais on l'a déjà fait dans le cas général de $\mathbb{P}_n$).
3. Pour tout $p \in \mathbb{P}_3$ et $c \in \mathbb{R}$, on a bien que $cp \in \mathbb{P}_3$.

Donc $\mathbb{P}_3$ est un sous-espace vectoriel de $\mathbb{P}_5$. Plus généralement, $\mathbb{P}_m$ est un sous-espace vectoriel de $\mathbb{P}_n$ si et seulement si $m \leq n$.
On aurait pu argumenter en disant que $\mathbb{P}_3$ est un sous-ensemble de $\mathbb{P}_5$, et il est un espace vectoriel quand on utilise les mêmes opérations, donc c'est un sous-espace.

**Non-exemple 4** $\mathbb{R}^2$ n'est pas un sous espace vectoriel de $\mathbb{R}^3$, puisque $\mathbb{R}^2$ n'est pas un sous-ensemble de $\mathbb{R}^3$.
Cependant, $V = \left\{x \in \mathbb{R}^3 \text{ tel que } x_3 = 0\right\}$ est un sous-espace vectoriel de $\mathbb{R}^3$.
En d'autres mots, par exemple,

$$\begin{bmatrix}1\\2\\0\end{bmatrix} \neq \begin{bmatrix}1\\2\end{bmatrix}$$

**Exemple 5** Soit $V$ est un espace vectoriel et $H = \left\{\overrightarrow{\mathbf{0}}\right\} \subset V$ où $\overrightarrow{\mathbf{0}}$ est le vecteur nul de $V$. Alors, $H$ est un sous-espace vectoriel de $V$.
On appelle $\left\{\overrightarrow{\mathbf{0}}\right\}$ le **sous-espace nul** ou **sous-espace trivial**.

**Exemple 6 (sous-espace engendré)** Soit $V$ un espace vectoriel. Soient $\overrightarrow{v}_1, \dots, \overrightarrow{v}_3 \in V$ des vecteurs de $V$. L'ensemble

$$H = \operatorname{vect}\{\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3\} = \{c_1\overrightarrow{v}_1 + c_2\overrightarrow{v}_2 + c\overrightarrow{v}_3, \quad c_1, c_2, c_3 \in \mathbb{R}\}$$

de toutes les combinaisons linéaires de ces trois vecteurs est un sous-ensemble de $V$. De plus, c'est un sous-espace vectoriel de $V$. En effet :

1. $\overrightarrow{\mathbf{0}} \in H$, puisque
$$0\overrightarrow{v}_1 + 0\overrightarrow{v}_2 + 0\overrightarrow{v}_3 = \overrightarrow{\mathbf{0}}$$
2. Si $\overrightarrow{u} \in H$, alors il existe $c_1, c_2, c_3$ tels que
$$\overrightarrow{u} = c_1\overrightarrow{v}_1 + c_2\overrightarrow{v}_2 + c_3\overrightarrow{v}_3$$
et de la même manière pour $\overrightarrow{v} \in H$, avec des scalaires $d_1, d_2, d_3 \in \mathbb{R}$. Alors, on a :
$$\overrightarrow{u} + \overrightarrow{v} = (c_1 + d_1)\overrightarrow{v}_1 + (c_2 + d_2)\overrightarrow{v}_2 + (c_3 + d_3)\overrightarrow{v}_3 \in H$$
3. La preuve est similaire, et elle est laissée en exercice au lecteur.

## Jeudi 28 octobre 2021 — **Cours 11 : Retour à l'App. (store) linéaire**

**Théorème** Si $\overrightarrow{v}_1, \dots, \overrightarrow{v}_p$ sont des vecteurs d'un espace vectoriel $V$, alors $\operatorname{vect}\{\overrightarrow{v}_1, \dots, \overrightarrow{v}_p\}$ est un sous-espace vectoriel de $V$.

**Exemples** Dans $V = \mathbb{R}^3$, les ensembles suivants sont des sous-espaces :

- L'origine : $\left\{\overrightarrow{\mathbf{0}}\right\} = \operatorname{vect} \emptyset$ ;
- Une droite passant par l'origine : $\operatorname{vect} \overrightarrow{v}$ avec $\overrightarrow{v} \neq \overrightarrow{\mathbf{0}}$ ;
- Un plan passant par l'origine : $\operatorname{vect}\{\overrightarrow{v}_1, \overrightarrow{v}_2\}$ avec $\overrightarrow{v}_1$ $\overrightarrow{v}_1, \overrightarrow{v}_2$ linéairement indépendants.
- $\mathbb{R}^3$ tout entier, obtenu comme $\operatorname{vect}\{\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3\}$ avec $\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3$ linéairement indépendants.

**Théorème** L'ensemble des solutions $H$ d'un système homogène $A\vec{x} = \vec{0}$ avec $A \in \mathbb{R}^{m \times n}$ est un sous-espace vectoriel de $\mathbb{R}^n$.

*Preuve*

0. L'ensemble solution est un sous-ensemble de $\mathbb{R}^n$ (il faut clairement que $\vec{x}$ soit en dimension $n$).
1. $\vec{0}$ appartient à $H$ puisque $A\vec{0} = \vec{0}$
2. Soient $\vec{u}, \vec{v}$ deux solutions quelconques ( :p) du système homogène. On a bien :

$$A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v} = \vec{0} + \vec{0} = \vec{0}$$

Donc $\vec{u} + \vec{v} \in H$

3. Soit $\vec{u}$ une solution de $A\vec{x} = \vec{0}$, et $c$ un réel quelconque. Alors :

$$A(c\vec{u}) = cA\vec{u} = c\vec{0} = \vec{0}$$

Donc $c\vec{u} \in H$.

## 6.3 Noyaux, images, et applications linéaires

**Définition de kernel et d'image** Soit $A \in \mathbb{R}^{m \times n}$ une matrice. On lui associe deux espaces.
**Le noyau de $A$**, "kernel" en anglais, l'ensemble solution du système homogène :

$$\ker A = \left\{ \vec{x} \in \mathbb{R}^n \text{ tel que } A\vec{x} = \vec{0} \right\}$$

C'est un sous-espace de $\mathbb{R}^n$.
**L'image de $A$**, le sous-espace engendré par les colonnes de $A$ :

$$\operatorname{Im} A = \operatorname{vect}\{\vec{a}_1, \ldots, \vec{a}_n\} = \{A\vec{x} \in \mathbb{R}^m \text{ tel que } \vec{x} \in \mathbb{R}^n\}$$

En d'autres mots :

$$\operatorname{Im} A = \left\{ \vec{b} \in \mathbb{R}^m \text{ tel que } \exists \vec{x} \in \mathbb{R}^n, A\vec{x} = \vec{b} \right\}$$

C'est un sous-espace de $\mathbb{R}^m$.

**Exemple** Pour une matrice $A \in \mathbb{R}^{m \times n}$, on a $\operatorname{Im} A = \operatorname{vect}\{\vec{a}_1, \ldots, \vec{a}_n\}$. Il est aussi possible d'écrire $\ker A$ comme $\operatorname{vect}\{\ldots\}$.
Pour cela, il nous faut trouver une forme paramétrique pour les solutions de $A\vec{x} = \vec{0}$.
On a le système suivant, et sa matrice échelonnée réduite équivalente :

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Le système suivant est équivalent :

$$\begin{cases} x_1 - 2x_2 - x_4 + 3x_5 = 0 & (6.3a) \\ x_3 + 2x_4 - 2x_5 = 0 & (6.3b) \\ 0 = 0 & (6.3c) \end{cases}$$

Or, on a donc :

$$\begin{cases} x_1 = 2x_2 + x_4 - 3x_5 & (6.4a) \\ x_2 = x_2 & (6.4b) \\ x_3 = -2x_4 + 2x_5 & (6.4c) \\ x_4 = x_4 & (6.4d) \\ x_5 = x_5 & (6.4e) \end{cases}$$

Qu'on peut aussi noter sous forme paramétrique :

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$

En d'autres mots,

$$\ker A = \left\{ \vec{x} \text{ tel que } A\vec{x} = \vec{0} \right\} = \text{vect}\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

On remarques que le nombre de vecteurs dont on a besoin pour décrire le kernel, 3 ici, est le nombre de variables libres. Ces trois vecteurs sont clairement linéairement indépendants (et le seront toujours ; cela vient du fait qu'il y a que des zéros sur la ligne des variables libres, sauf dans un des vecteurs (par exemple, la deuxième composante des trois vecteurs est nulle, sauf celle du premier)).

## 6.4 Extension aux applications linéaires

**Définition application linéaire** On appelle une **application linéaire** $T$ d'un espace vectoriel $V$ dans un espace vectoriel $W$ un procédé qui, à tout vecteur de $\vec{x}$ de $V$, associe un unique vecteur $T(\vec{x})$ de $W$ de façon que :

1. $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$ pour tout $\vec{u}, \vec{v} \in V$
2. $T(c\vec{u}) = cT(\vec{u})$ pour tout $\vec{u}$ de $V$ et tout scalaire $c$.

**Extension** On définit le noyau de $T$ :

$$\ker T = \left\{ \vec{x} \in V \text{ tel que } T(\vec{x}) = \vec{0} \right\}$$

On définit aussi l'image de $T$ :

$$\operatorname{Im} T = \left\{ T(\vec{x}) \in W \text{ tel que } \vec{x} \in V \right\}$$

On peut montrer que $\ker T$ est un sous-espace de $V$ et $\operatorname{Im} T$ est un sous-espace de $W$.

> *Note* Dans ce cas, on n'a pas $T(\vec{x}) = A\vec{x}$, puisque $V$ et $W$ peuvent ne pas être $\mathbb{R}^n$ et $\mathbb{R}^m$.

**Exemple** Soit $V \in \mathbb{P}_3$ et $W = \mathbb{P}_2$. Soit $T : V \mapsto W$ l'application qui, à un polynôme $p$, fait correspondre sa dérivée $T(p) = p'$.
Si $p(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$, alors :

$$p'(t) = a_1 + 2a_2 t + 3a_3 t^2$$

On a bien $p \in \mathbb{P}_3 \implies T(p) \in \mathbb{P}_2$. De plus, on peut remarquer que c'est bien une application linéaire :

1. Soit $p, q \in \mathbb{P}_3$. On a :

$$(T(p) + T(q))(t) = a_1 + 2a_2 t + 3a_3 t^2 + b_1 + 2b_2 t + 3b_3^2$$

   Or, c'est bien égal à :

$$(a_1 + b_1) + 2(a_2 + b_2)t + 3(a_3 + b_3)t^2 = T(p + q)(t)$$

2. On peut aussi montrer que $T(cp) = cT(p)$ pour tout $p \in \mathbb{P}_3$.

On peut maintenant regarder quel est le noyau de $T$. On a :

$$\ker T = \{p \in \mathbb{P}_3 \text{ tel que } T(p) = 0\} = \mathbb{P}_0 = \{\text{polynômes constants}\}$$

De plus, l'image de $T$ est donnée par :

$$\operatorname{Im} T = \{q \in \mathbb{P}_2 \text{ tel que } \exists p \in \mathbb{P}_3, T(p) = q\} = \mathbb{P}_2$$

## 6.5 Familles libres et bases

**Définition** Soit $V$ un espace vectoriel. On dit que les vecteurs $\vec{v}_1, \ldots, \vec{v}_p$ dans $V$ sont **linéairement indépendants** si l'équation

$$c_1 \vec{v}_1 + \ldots c_p \vec{v}_p = \vec{0}$$

admet comme seule solution la solution triviale $c_1 = \ldots = c_p = 0$.
Dans ce cas, on dit que la **famille** $\left(\vec{v}_1, \ldots, \vec{v}_p\right)$ est **libre**. Sinon, on dit que la famille est **liée.**

**Théorème** Une famille $\left(\vec{v}_1, \ldots, \vec{v}_p\right)$ d'au moins deux vecteurs, avec $\vec{v}_1 \neq \vec{0}$, est liée si et seulement s'il existe $j > 1$ tel que $\vec{v}_j$ soit une combinaison linéaire des vecteurs qui viennent avant lui, soit $\vec{v}_1, \ldots, \vec{v}_{j-1}$.

> *Remarque* On l'avait démontré pour $\mathbb{R}^n$, mais il fonctionne donc de la même manière pour n'importe quel espace vectoriel. La preuve est similaire à ce qu'on avait fait.

**Exemple** Avec $V = \mathbb{P}_2$, on considère $p_1, p_2, p_3 \in \mathbb{P}_2$ définis par :

$$p_1(t) = 1, \quad p_2(t) = t, \quad p_3(t) = 4 - t$$

On remarque qu'ils sont linéairement dépendants puisque

$$(p_3 + p_2 - 4p_1)(t) = p_3(t) + p_2(t) - 4p_1(t) = 4 - t + t - 4 = 0$$

pour tout $t$. (Le "pour tout $t$" est important, sinon on a juste une équation en $t$).

**Exemple 2** Avec $V = \mathbb{P}_2$, on considère $p_1, p_2, p_3 \in \mathbb{P}_2$ définis par :

$$p_1(t) = 1, \quad p_2(t) = t, \quad p_3(t) = t^2$$

On veut résoudre :

$$c_1 p_1 + c_2 p_2 + c_3 p_3 = 0$$

On a :

$$(c_1 p_1 + c_2 p_2 + c_3 p_3)(t) = c_1 p_1(t) + c_2 p_2(t) + c_3 p_3(t) = c_1 + c_2 t + c_3 t^2$$

On sait que ***le seul polynôme de $\mathbb{P}_n$ qui a strictement plus que $n$ racines est le polynôme nul***. Ainsi, nécessairement, chaque coefficient doit être nul. On a donc :

$$c_1 = c_2 = c_3 = 0$$

Donc, les polynômes $p_1, p_2, p_3$ sont linéairement indépendants.

**Définition des bases**

Soit $H$ un sous-espace de $V$ (on pourrait très bien avoir $H = V$).
Une famille de vecteurs $\left(\vec{\boldsymbol{b}}_1, \dots, \vec{\boldsymbol{b}}_p\right)$ de $V$ est une **base de $\boldsymbol{H}$** si :

1. La famille est libre : les vecteurs $\vec{\boldsymbol{b}}_1, \dots, \vec{\boldsymbol{b}}_p$ sont linéairement indépendants
2. La famille engendre $H$ : $H = \text{vect}\left(\vec{\boldsymbol{b}}_1, \dots, \vec{\boldsymbol{b}}_p\right)$, mais ni plus ni moins.

*Remarque* En d'autres mots, une base est le nombre minimum de vecteurs qui engendrent l'espace vectoriel. Si on en rajoute un, alors ils ne seront plus linéairement indépendants.

**Observations**

1. Si $\vec{\boldsymbol{b}}_1, \dots, \vec{\boldsymbol{b}}_p \in V$ sont linéairement indépendants, alors ils forment une base pour $\text{vect}\left(\vec{\boldsymbol{b}}_1, \dots, \vec{\boldsymbol{b}}_p\right)$.
2. Si $\vec{\boldsymbol{b}}_1, \dots, \vec{\boldsymbol{b}}_p \in V$ forment une base pour un sous-espace $H$, alors les vecteurs $\vec{\boldsymbol{b}}_1, \dots, \vec{\boldsymbol{b}}_p$ sont dans $H$ car $H = \text{vect}\left(\vec{\boldsymbol{b}}_1, \dots, \vec{\boldsymbol{b}}_p\right)$.

**Exemples et non-exemples**

1. Si on a :
$$V = \mathbb{R}^2, \quad \vec{\boldsymbol{b}}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \vec{\boldsymbol{b}}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
Ils sont linéairement indépendants, et forment une base pour $\text{vect}\left(\vec{\boldsymbol{b}}_1, \vec{\boldsymbol{b}}_n\right) = \mathbb{R}^2$.
2. Si on a :
$$V = \mathbb{R}^3, \quad \vec{\boldsymbol{b}}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{\boldsymbol{b}}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$
Ils sont bien linéairement indépendants, donc ils forment une base pour l'espace qu'ils engendrent, qui est $\text{vect}\left\{\vec{\boldsymbol{b}}_1, \vec{\boldsymbol{b}}_2\right\}$, le plan $O_{xy}$ dans $\mathbb{R}^3$.
3. Si on a :
$$V = \mathbb{R}^2, \quad \vec{\boldsymbol{b}}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \vec{\boldsymbol{b}}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \vec{\boldsymbol{b}}_3 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$
Ils sont linéairement dépendants puisque $3 > 2$. Donc, ils ne forment pas une base.
4. Si on a :
$$V = \mathbb{P}_2, \quad \vec{\boldsymbol{b}}_1(t) = 1, \quad \vec{\boldsymbol{b}}_2(t) = t, \quad \vec{\boldsymbol{b}}_3(t) = t^2$$
Ils sont linéairement indépendants, donc ils forment une base pour $\text{vect}\left\{\vec{\boldsymbol{b}}_1, \vec{\boldsymbol{b}}_2, \vec{\boldsymbol{b}}_3\right\} = \mathbb{P}_2$.
5. Si on a :
$$V = \mathbb{P}_3, \quad \vec{\boldsymbol{b}}_1(t) = 1, \quad \vec{\boldsymbol{b}}_2(t) = t, \quad \vec{\boldsymbol{b}}_3(t) = t^2$$
Ils sont linéairement indépendants, donc ils forment une base pour $\text{vect}\left\{\vec{\boldsymbol{b}}_1, \vec{\boldsymbol{b}}_2, \vec{\boldsymbol{b}}_3\right\} = \mathbb{P}_2 \subset \mathbb{P}_3$.

**Matrices inversibles**

On remarque que si on a $A \in \mathbb{R}^{n \times n}$ inversible, alors les colonnes de $A$ forment une base de $\mathbb{R}^n$, puisque ses colonnes sont donc linéairement indépendantes et engendrent $\mathbb{R}^n$.
Dans l'autre sens, si la matrice n'est pas inversible, alors les colonnes de $A$ ne sont pas linéairement indépendantes, donc elles forment pas de base.
En d'autres mots, une matrice est inversible si et seulement si ses colonnes forment une base de $\mathbb{R}^n$.

**Base canonique de $\mathbb{R}^n$**

Les colonnes de la matrice identité $I_n$ forment une base de $\mathbb{R}^n$, puisqu'elle est inversible, on l'appelle la **base canonique de $\mathbb{R}^n$**.

Par exemple, dans $\mathbb{R}^3$ :

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Lundi 1er novembre 2021 — **Cours 12 : Basique, simple. Parce que vous êtes trop cons.**

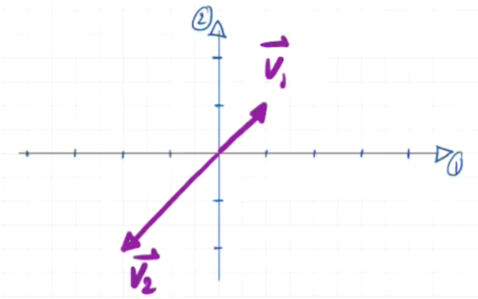
**Familles** Une famille de vecteurs $\left(\vec{v}_1, \ldots, \vec{v}_p\right)$ de $V$ engendrent toujours un sous espace

$$H = \text{vect}\left\{\vec{v}_1, \ldots, \vec{v}_p\right\}$$

Si la famille est libre, c'est une base de $H$. Si la famille est liée, on peut retirer des vecteurs pour rendre la famille libre.

**Exemple 1** Soit les deux vecteurs de $\mathbb{R}^2$ suivants :

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$




On remarque que $\text{vect}\left\{\vec{v}_1, \vec{v}_2\right\}$ est une droite qui passe par l'origine. Donc,

$$\text{vect}\left\{\vec{v}_1, \vec{v}_2\right\} = \text{vect}\left\{\vec{v}_1\right\} = \text{vect}\left\{\vec{v}_2\right\}$$

Donc chacun $\vec{v}_1$ et $\vec{v}_2$ forment une base pour la droite.

**Exemple 2** Soit les trois vecteurs de $\mathbb{R}^3$ suivant :

$$\vec{v}_1 = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 6 \\ 16 \\ -5 \end{bmatrix}$$

Pour déterminer s'ils sont linéairement dépendants, on peut calculer le déterminant de la matrice qui a ces vecteurs en tant que colonnes :

$$\det \begin{bmatrix} 0 & 2 & 6 \\ 2 & 2 & 16 \\ -1 & 0 & -5 \end{bmatrix} = 0$$

Donc cette matrice n'est pas inversible, et ses vecteurs sont donc liés. On peut voir que :

$$\vec{v}_3 = 5\vec{v}_1 + 3\vec{v}_2$$

On peut aussi exprimer $\vec{v}_1$ en fonction de $\vec{v}_2$ et $\vec{v}_3$, l'analyse qu'on fait ci-dessous serait similaire.
On veut montrer que

$$\text{vect}\left\{\vec{v}_1, \vec{v}_2, \vec{v}_3\right\} = \text{vect}\left\{\vec{v}_1, \vec{v}_2\right\}$$

Pour faire cela, on va montrer que le premier ensemble est inclus dans le deuxième, puis que le deuxième est inclus dans le premier.

Si $\vec{v}$ est dans $H = \text{vect}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, alors il existe $c_1, c_2, c_3$ tels que :

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 = c_1\vec{v}_1 + c_2\vec{v}_2 + c_3(5\vec{v}_1 + 3\vec{v}_2)$$

Ce qui nous permet de conclure que :

$$\vec{v} = (c_1 + 5c_3)\vec{v}_1 + (c_2 + 3c_3)\vec{v}_2$$

Donc, $\vec{v}$ est une combinaison linéaire de $\vec{v}_1$ et $\vec{v}_2$. En d'autres mots, cela veut dire que

$$\text{vect}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \subset \text{vect}\{\vec{v}_1, \vec{v}_2\}$$

Dans l'autre sens, n'importe quelle combinaison linéaire de $\vec{v}_1$ et $\vec{v}_2$ est aussi une combinaison linéaire de $\vec{v}_1$, $\vec{v}_2$ et $\vec{v}_3$ (il suffit de mettre le dernier coefficient à zéros). Autrement dit,

$$\text{vect}\{\vec{v}_1, \vec{v}_2\} \subset \text{vect}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$$

**Théorème de la base extraite**

Soit $F = (\vec{v}_1, \dots, \vec{v}_p)$ une famille de vecteurs de $V$ et $H = \text{vect}\{\vec{v}_1, \dots, \vec{v}_p\}$.

1. Si l'un des vecteurs de $F$ (disons $\vec{v}_k$) est une combinaison linéaire des autres vecteurs de $F$, alors la famille obtenue en supprimant dans $F$ le vecteur $\vec{v}_k$ engendre toujours $H$.
2. Si $H \neq \left\{\vec{0}\right\}$, alors il existe une sous-famille de $F$ qui est une base de $H$. Autrement dit, on peut extraire de la famille $F$ une base de $H$.

*Preuve* On peut utiliser la même idée que ce qu'on a fait dans l'exemple ci-dessus.

**Bases du kernel et de l'image**

Soit $A \in \mathbb{R}^{m\times n}$ une matrice donnée.
On a déjà trouvé une façon de déterminer une base du noyau de $A$. En effet, on peut mettre $A$ sous forme échelonnée réduite pour trouver l'ensemble solution de $A\vec{x} = \vec{0}$ en forme paramétrique. Les vecteurs obtenus engendrent le noyau et ils sont linéairement indépendants, donc ils forment une base.
On voudrait faire la même chose pour trouver une base de l'image de $A$. On sait que les colonnes de $A$ génèrent l'image de $A$, cependant ils ne sont pas nécessairement linéairement indépendants. On peut par contre utiliser le théorème de la base extraite pour extraire une base.
Par exemple, prenons la matrice suivante :

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

On appelle la $i$-ème colonne $\vec{a}_i$. On a donc :

$$\text{Im}\, A = \text{vect}\{\vec{a}_1, \vec{a}_2, \vec{a}_3, \vec{a}_4, \vec{a}_5\}$$

Par définition, les colonnes de $A$ engendrent le sous-espace $\text{Im}\, A$. Elles ne forment pas une base de $\text{Im}\, A$ puisqu'elles sont linéairement dépendantes. On peut s'en rendre compte en calculant le noyau de $A$. Dans notre exemple, le vecteur suivant fait partie du kernel :

$$A\vec{x} = \vec{0} \text{ avec } \vec{x} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \implies 2\vec{a}_1 + \vec{a}_2 = \vec{0}$$

Utilisons le théorème de la base extraite. Il est toujours possible de choisir des colonnes de $A$ qui forment une base de $\text{Im}\, A$. Mais on ne peut pas les choisir n'importe comment. Notre étude du noyau nous dit exactement procéder.

On sait que

$$\ker A = \operatorname{vect}\left\{\begin{bmatrix}2\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\-2\\1\\0\end{bmatrix}, \begin{bmatrix}-3\\0\\2\\0\\1\end{bmatrix}\right\}$$

En d'autres mots :

$$\begin{cases}\overrightarrow{a}_2 = -2\overrightarrow{a}_1 & (6.6a)\\ \overrightarrow{a}_4 = -\overrightarrow{a}_1 + 2\overrightarrow{a}_3 & (6.6b)\\ \overrightarrow{a}_5 = 3\overrightarrow{a}_1 - 2\overrightarrow{a}_3 & (6.6c)\end{cases}$$

Donc, par le théorème de la base extraite, on sait que

$$\operatorname{Im} A = \operatorname{vect}\{\overrightarrow{a}_1, \overrightarrow{a}_2, \overrightarrow{a}_3, \overrightarrow{a}_4, \overrightarrow{a}_5\} = \operatorname{vect}\{\overrightarrow{a}_1, \overrightarrow{a}_3\}$$

On remarque que, de manière générale, on ne modifie pas la dépendance linéaire des colonnes en échelonnant une matrice. En d'autres mots, si la première et la deuxième colonne sont dépendantes linéairement dans la matrice échelonnée réduite, alors elles le sont aussi dans la matrice de base. Plus que ça, si on remarque que 3 fois la première colonne donne la deuxième colonne dans la matrice échelonnée réduite, alors c'est aussi le cas dans la matrice de base.
Puisqu'on ne modifie pas l'indépendance linéaire en échelonnant une matrice, et puisque la première et la troisième colonne de la matrice échelonnée réduite sont clairement linéairement indépendants, $\overrightarrow{a}_1$ et $\overrightarrow{a}_3$ sont donc aussi linéairement indépendants.
On peut justifier cela de manière plus formelle. Étudions si $\overrightarrow{a}_1$ et $\overrightarrow{a}_3$ sont linéairement dépendants :

$$c_1\overrightarrow{a}_1 + c_3\overrightarrow{a}_3 = \overrightarrow{0} \iff A\begin{bmatrix}c_1\\0\\c_3\\0\\0\end{bmatrix} = \overrightarrow{0}$$

Or, l'ensemble solution ne change pas si on échelonne $A$, donc :

$$\begin{bmatrix}1 & -2 & 0 & -1 & 3\\ 0 & 0 & 1 & 2 & -2\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}c_1\\0\\c_3\\0\\0\end{bmatrix} = \overrightarrow{0} \implies c_1\begin{bmatrix}1\\0\\0\end{bmatrix} + c_3\begin{bmatrix}0\\1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

Ainsi, on en déduit qu'ils sont linéairement indépendants, donc ils forment une base de l'image de $A$.

**Théorème** Les colonnes pivots d'une matrice forment une base de son image.

> *Remarque* L'échelonnement permet d'identifier les colonnes pivots, mais ce sont bien les colonnes de $A$ qui engendre $\operatorname{Im} A$. Les colonnes d'une forme échelonnée de $A$ pourraient engendrer un tout autre espace.

## 6.6 Système de coordonnées

**Théorème de représentation d'un vecteur**

Soit $\mathcal{B} = \left(\vec{b}_1, \ldots, \vec{b}_n\right)$ une base d'une space vectoriel $V$. Alors, pour tout vecteur $\vec{x}$ de $V$, il existe une famille *unique* $(c_1, \ldots, c_n)$ de scalaires tels que :

$$\vec{x} = c_1 \vec{b}_1 + \ldots + c_n \vec{b}_n$$

*Preuve*

**Existence :** Comme $\mathcal{B}$ est une base de $V$, on a $V = \operatorname{vect}\left\{\vec{b}_1, \ldots, \vec{b}_n\right\}$. Donc, $\vec{x}$ est une combinaison linéaire de $\vec{b}_1, \ldots, \vec{b}_n$.

**Unicité :** Supposons par l'absurde que ce n'est pas unique, i.e. que :

$$\vec{x} = c_1 \vec{b}_1 + \ldots + c_n \vec{b}_n \quad \text{et} \quad \vec{x} = d_1 \vec{b}_1 + \ldots + d_n \vec{b}_n$$

Alors, on a :

$$\vec{0} = \vec{x} - \vec{x} = c_1 \vec{b}_1 + \ldots + c_n \vec{b}_n - d_1 \vec{b}_1 - \ldots - d_n \vec{b}_n$$

Ce qui implique que :

$$\vec{0} = (c_1 - d_1)\vec{b}_1 + \ldots + (c_n - d_n)\vec{b}_n$$

Comme $\mathcal{B}$ est une base, les vecteurs $\vec{b}_1, \ldots, \vec{b}_n$ sont linéairement indépendants. Donc,

$$c_1 - d_1 = \ldots = c_n - d_n = 0 \implies c_1 = d_1, \ldots, c_n = d_n$$

Ce qui est une contradiction.

$\square$

*Importance* Si on est d'accord sur la base, connaître le vecteur $\vec{x}$ ou connaître les coefficients $c_1, \ldots, c_n$ est équivalent.

**Notation**

Soit $\mathcal{B} = \left(\vec{b}_1, \ldots, \vec{b}_n\right)$ une base de $V$ (l'ordre des vecteurs n'est pas important, mais une fois qu'on l'a choisit on ne doit plus le changer).
Pour chaque $\vec{x}$ dans $V$, il existe un unique choix de $c_1, \ldots, c_n \in \mathbb{R}$ tels que

$$\vec{x} = c_1 \vec{b}_1 + \ldots + c_n \vec{b}_n$$

Ces $c_i$ sont les **composantes** ou **coordonnées** de $\vec{x}$ dans la base $\mathcal{B}$.
Le **vecteur de coordonnées** (visible ci-après) identifie $\vec{x}$ de façon unique (dans la base $\mathcal{B}$) : on peut trouver l'un à partir de l'autre. On écrit :

$$\left[\vec{x}\right]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n$$

De plus, $\vec{x} \mapsto \left[\vec{x}\right]_{\mathcal{B}}$ est **l'application coordonnées** définie par $B$, de $V$ vers $\mathbb{R}^n$.

**Exemple 1**

Soit $\mathcal{E} = (\vec{e}_1, \vec{e}_2)$ la base canonique de $\mathbb{R}^2$. Soit le vecteur suivant :

$$\vec{x} = \begin{bmatrix} 5 \\ 3 \end{bmatrix} \in \mathbb{R}^2$$

On se rend compte que :

$$\overrightarrow{\boldsymbol{x}} = 5\overrightarrow{\boldsymbol{e}}_1 + 3\overrightarrow{\boldsymbol{e}}_2 \implies \left[\overrightarrow{\boldsymbol{x}}\right]_{\mathcal{E}} = \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \overrightarrow{\boldsymbol{x}}$$

C'est pour cela que nous appelons notre base, la base canonique. Cette propriété est très spéciale, et elle n'est vraie que pour cette base. De manière générale, il ne faut pas confondre $\overrightarrow{\boldsymbol{x}}$ et son vecteur de coordonnées.

**Exemple 2**

Soit $\mathcal{B} = \left(\overrightarrow{\boldsymbol{b}}_1, \overrightarrow{\boldsymbol{b}}_2\right)$, avec :

$$\overrightarrow{\boldsymbol{b}}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \overrightarrow{\boldsymbol{b}}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

On peut vérifier que c'est une base de $\mathbb{R}^2$. On veut trouver le vecteur $\overrightarrow{\boldsymbol{x}}$ de coordonnées :

$$\left[\overrightarrow{\boldsymbol{x}}\right]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \implies \overrightarrow{\boldsymbol{x}} = 3\overrightarrow{\boldsymbol{b}}_1 + \overrightarrow{\boldsymbol{b}}_2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

Soit maintenant le vecteur suivant :

$$\overrightarrow{\boldsymbol{y}} = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$$

Maintenant, si on veut trouver ses coordonnées pour $\mathcal{B}$, il faut résoudre un système. On peut aussi le faire “avec les mains”, et on trouve :

$$\left[\overrightarrow{\boldsymbol{y}}\right]_{\mathcal{B}} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

Pour un $\overrightarrow{\boldsymbol{x}}$ quelconque (quoi ? j'ai rien dit. . .) dans $\mathbb{R}^2$, on a que ses coordonnées (uniques) $c_1, c_2$ dans la base $\mathcal{B} = \left(\overrightarrow{\boldsymbol{b}}_1, \overrightarrow{\boldsymbol{b}}_2\right)$ satisfont :

$$c_1\overrightarrow{\boldsymbol{b}}_1 + c_2\overrightarrow{\boldsymbol{b}}_2 = \overrightarrow{\boldsymbol{x}}$$

Qui est équivalent à :

$$\begin{bmatrix} \overrightarrow{\boldsymbol{b}}_1 & \overrightarrow{\boldsymbol{b}}_2 \end{bmatrix} \left[\overrightarrow{\boldsymbol{x}}\right]_{\mathcal{B}} = \overrightarrow{\boldsymbol{x}} \iff \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

On sait que la matrice composée des vecteurs de base est inversible (puisque ces colonnes représentent une base), on a donc :

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}^{-1}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

**Généralisation**

Plus généralement, soit $\mathcal{B} = \left(\overrightarrow{\boldsymbol{b}}_1, \ldots, \overrightarrow{\boldsymbol{b}}_n\right)$ une base de $\mathbb{R}^n$. On définit la matrice de passage, ou **matrice de changement de base**, $P$ :

$$P_{\mathcal{B}} = \begin{bmatrix} \overrightarrow{\boldsymbol{b}}_1 & \ldots & \overrightarrow{\boldsymbol{b}}_n \end{bmatrix} \in \mathbb{R}^{n\times n}$$

Alors, la relation suivante :

$$\overrightarrow{\boldsymbol{x}} = c_1\overrightarrow{\boldsymbol{b}}_1 + \ldots c_n\overrightarrow{\boldsymbol{b}}_n \text{ où } \left[\overrightarrow{\boldsymbol{x}}\right]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

s'écrit aussi :

$$\vec{x} = P_{\mathcal{B}}\left[\vec{x}\right]_{\mathcal{B}}$$

C'est donc bien une matrice de changement de base, puisqu'elle transforme un vecteur de la base $\mathcal{B}$ en un vecteur de la base canonique.
De plus, on sait que $P_{\mathcal{B}}$ est inversible (puisque ses colonnes représentent une base), on a donc :

$$\left[\vec{x}\right]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\vec{x}$$

On voit que $P_{\mathcal{B}}^{-1}$ est la matrice de **l'application coordonnées** (qu'on avait définie plus haut). Cette application est donc linéaire (car c'est le produit d'une matrice et du vecteur donné en paramètre) et bijective (puisque la matrice est inversible).

**Théorème** Soit $\mathcal{B}$ une base d'un espace vectoriel $V$. L'application coordonnées $\vec{x} \mapsto \left[\vec{x}\right]_{\mathcal{B}}$ est une application linéaire bijective de $V$ dans $\mathbb{R}^n$ ($n$ est le nombre de vecteurs de la base, plus tard on l'appellera la dimension de l'espace vectoriel).

*Preuve* Laissée en exercice au lecteur.

*Remarque* Une application linéaire bijective est un **isomorphisme** d'espaces vectoriels.
La linéarité nous dit que :

$$\left[c_1\vec{u}_1 + \ldots + c_p\vec{u}_p\right]_{\mathcal{B}} = c_1\left[\vec{u}_1\right]_{\mathcal{B}} + \ldots + c_p\left[\vec{u}_p\right]_{\mathcal{B}}$$

Ça, et la bijectivité de $\vec{x} \mapsto \left[\vec{x}\right]_{\mathcal{B}}$ donnent un sens précis à l'intuition que $V$ et $\mathbb{R}^n$ “se ressemblent fortement” (le fait que ce soit un isomorphisme, nous permet de savoir qu'il y a le même nombre d'éléments).
De plus, la linéarité nous permet de nous rendre compte que, si des vecteurs sont linéairement dépendants dans une base, alors ils le seront aussi dans n'importe quelle autre base ; et de la même manière s'ils sont linéairement indépendants. Cette propriété est très puissante.

Jeudi 4 novembre 2021 — **Cours 13 : Dimension et lignes**

**Exemple 1** La base canonique $\mathbb{P}_3$ est $\mathcal{B} = (p_0, p_1, p_2, p_3)$ avec :

$$p_0(t) = 1, \quad p_1(t) = t, \quad p_2(t) = t^2, \quad p_3 = t^3$$

Tout $p$ dans $\mathbb{P}_3$ s'écrit $p(t) = a_0 + a_1t + a_2t^2 + a_3t^3$ pour certains coefficients $a_0$, $a_1$, $a_2$ et $a_3$. Ce sont les coordonnées de $p$ dans la base $\mathcal{B}$. On peut donc écrire :

$$[p]_{\mathcal{B}} = [a_0p_0(t) + a_1p_1(t) + a_2p_2(t) + a_rp_3(t)]_{\mathcal{B}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

De cette manière, $\mathbb{P}_3$ et $\mathbb{R}^4$ sont isomorphes : on peut répondre à des questions d'algèbre linéaire dans $\mathbb{P}_3$ en les transformant en questions dans $\mathbb{R}^4$.

**Exemple 2** Soit $V = \mathbb{R}^{2\times 3}$ l'espace des matrices de taille $2 \times 3$. Sa base canonique est $\mathcal{B} = (B_1, \ldots, B_6)$ avec :

$$B_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$B_4 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad B_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad B_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

L'application coordonnées associée à $\mathcal{B}$ est laissée en exercice au lecteur. Elle établit un isomorphisme entre $\mathbb{R}^{2\times 3}$ et $\mathbb{R}^6$.

**Exemple 3** Nous voulons trouver une base pour le sous-espace des matrices symétriques $2 \times 2$. On se rend compte qu'elles sont toutes sous la forme :

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Donc on peut utiliser ces matrices comme base, ce qui montre un isomorphisme avec $\mathbb{R}^3$.

## 6.7 Dimension

**Théorème** Si un espace vectoriel $V$ admet une base de $n$ vecteurs, alors toutes les bases de $V$ comportent exactement $n$ vecteurs.

> *Isomorphisme* Si $\mathcal{B} = \left(\vec{b}_1, \ldots, \vec{b}_n\right)$ est une base de $V$, alors $V$ est isomorphe à $\mathbb{R}^n$. Le nombre $n$ ne dépend *pas* du choix de la base.

> *Remarque* On sait que tout espace vectoriel a une base, puisqu'on peut en extraire une d'une famille de vecteurs qui l'engendrent.

**Définition de la dimension** On appelle **dimension de $V$**, notée $\dim V$, le nombre d'éléments d'une base de $V$. Par convention, $\dim\left\{\vec{0}\right\} = 1$.

**Exemples** Voici les dimensions de certains espaces vectoriels :
- $\dim \mathbb{R}^n = n$
- $\dim \mathbb{R}^{m\times n} = mn$
- $\dim \mathbb{P}_n = n+1$
- L'origine dans $\mathbb{R}^3$ : 0
- Une droite passant par l'origine dans $\mathbb{R}^3$ : 1
- Un plan passant par l'origine dans $\mathbb{R}^3$ : 2
- L'espace $\mathbb{R}^3$ tout entier : 3
- L'espace des matrices symétriques de taille $2 \times 2$ : 3
- L'espace des matrices symétriques de taille $3 \times 3$ : 6
- L'espace des matrices symétriques de taille $n \times n$ : $\frac{n(n+1)}{2}$

**Complétion d'une base** Soit $H$ un sous-espace vectoriel de $V$. Toute famille libre de $H$ peut être complétée pour former une base de $H$.
Cette affirmation est celle complémentaire au théorème de la base extraite.

> *Justification* Soit $F = \left(\vec{u}_1, \ldots, \vec{u}_k\right)$ une famille libre de $H$. Si $H = \operatorname{vect} F$, alors $F$ est une base.
> Sinon, il existe un vecteur dans $H$ mais pas dans $\operatorname{vect} F$. Appelons-le $\vec{u}_{k+1}$. Alors, $\left(\vec{u}_1, \ldots, \vec{u}_k, \vec{u}_{k+1}\right)$ est toujours libre.
> On peut continuer à ajouter des vecteurs jusqu'à ce que la famille (toujours libre) engendre $H$. Alors, on obtient une base de $H$.

**Dimension d'un sous-espace** Soit $H$ un sous-espace vectoriel de $V$. Alors :

$$\dim H \leq \dim V$$

En effet, une base de $H$ est une famille libre de $V$, on peut la compléter pour obtenir une base de $V$, sans jamais devoir supprimer de vecteur.

**Nombre de vecteurs d'une base** Soit $V$ un espace vectoriel avec $\dim V = p$.
- Toute famille libre de $p$ éléments de $V$ est une base de $V$.
- Toute famille de $p$ éléments qui engendre $V$ (exactement) est une base de $V$.

*Justification*
- On peut "compléter" la famille pour obtenir une base, mais une base a $p$ éléments, donc on ne doit rien ajouter.
- On peut "extraire" une base de la famille, mais une base a $p$ éléments, donc on ne doit rien retirer.

**Dimension du kernel et de l'image** Prenons la matrice suivante, et sa forme échelonnée réduite :

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

On avait trouvé la base de son kernel et de son image :

$$\ker A = \operatorname{vect}\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\}, \quad \operatorname{Im} A = \operatorname{vect}\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} \right\}$$

De manière générale, on remarque que la **dimension de ker $A$** est égale au nombre de variables libres de l'équation $A\overrightarrow{x} = \overrightarrow{0}$, et la **dimension de Im $A$** est égale au nombre de colonnes pivots de $A$.
On se rend compte, qu'il nous aurait suffit d'un échelonnement non-réduit pour les trouver. De plus, on voit que :

$$\dim \ker A + \dim \operatorname{Im} A = n$$

où $n$ est le nombre de colonnes de $A$.

## 6.8 Rang

**Définition de rang** Le **rang** ("rank" en anglais) d'une matrice $A \in \mathbb{R}^{m\times n}$ est la dimension du sous-espace engendré par ses colonnes :

$$\operatorname{rang} A = \dim \operatorname{Im} A$$

**Observation** Comme $\operatorname{Im} A$ est un sous-espace $\mathbb{R}^m$, on a :

$$\operatorname{rang} A \leq m$$

Et, puisque $\operatorname{Im} A$ est engendré par $n$ vecteurs, on a aussi :

$$\operatorname{rang} A \leq n$$

**Théorème du rang** L'égalité suivante tiens toujours :

$$\operatorname{rang} A + \dim \ker A = n$$

*Justification*
- $\operatorname{rang} A =$ nombre de colonnes pivots
- $\dim \ker A =$ nombre de colonnes non pivots
- $n =$ nombre de colonnes

**Vrai ou faux** Est-ce que pour, $A \in \mathbb{R}^{m\times n}$, on a :
1. $\dim \ker A \leq n$ ?
2. $\dim \ker A \leq m$ ?

On a :

1. C'est vrai, on peut le déduire du théorème du rang, puisque $0 \leq \text{rang}\, A \leq n$. Une autre justification est que le kernel est un sous espace de $\mathbb{R}^n$, donc il a une dimension de au plus $n$.
2. C'est faux. On peut par exemple prendre la matrice suivante :

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

   On a l'inégalité suivante :

$$\dim \ker A = 15 > 2 = m$$

## 6.9 Espace engendré par les lignes

**Lignes des matrices** Pour une matrice $A \in \mathbb{R}^{m \times n}$, on a considéré $\text{Im}\, A$ l'espace engendré par ses colonnes. Que peut-on dire de l'espace engendré par les lignes de $A$, vues comme des vecteurs de $\mathbb{R}^n$ ?

**Exemple** Soit la matrice suivante, et sa transpose :

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix} \implies A^T = \begin{bmatrix} -3 & 1 & 2 \\ 6 & -2 & -4 \\ -1 & 2 & 5 \\ 1 & 3 & 8 \\ -7 & -1 & -4 \end{bmatrix}$$

Alors, on a :

$$\text{ligne}_1(A) = \begin{bmatrix} -3 \\ 6 \\ -1 \\ 1 \\ -7 \end{bmatrix}, \quad \text{ligne}_2(A) = \begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \\ -1 \end{bmatrix}, \quad \text{ligne}_3(A) = \begin{bmatrix} 2 \\ -4 \\ 5 \\ 8 \\ -4 \end{bmatrix}$$

**Espace des lignes** L'espace des lignes de $A \in \mathbb{R}^{m \times n}$ est le sous-espace de $\mathbb{R}^n$ engendré par les lignes, noté :

$$\text{Lgn}\, A = \text{Im}\, A^T = \text{vect}\{\text{ligne}_1(A), \ldots, \text{ligne}_m(A)\}$$

**Observation** On remarque que, puisque $\text{Lgn}\, A$ est un sous-espace de $\mathbb{R}^n$ :

$$\text{Lgn}\, A \leq n$$

De plus, puisque $\text{Lgn}\, A$ est engendré par $m$ vecteurs :

$$\text{Lgn}\, A \leq m$$

Aussi, on remarque que :

$$\dim \text{Lgn}\, A = \dim \text{Im}\, A^T = \text{rang}\, A^T$$

**Théorème** On a l'égalité suivante :

$$\text{rang}\, A = \text{rang}\, A^T$$

Autrement dit :

$$\dim \text{Im}\, A = \dim \text{Lgn}\, A$$

*Surprise* Ce résultat est quelque peu surprenant, puisque $\text{Im}\, A$ est un sous-espace de $\mathbb{R}^m$, mais $\text{Lgn}\, A$ est un sous-espace de $\mathbb{R}^n$.

*Lemme*

Soient $A$ et $B$ deux matrices dont le produit $AB$ est compatible. Si $B$ est inversible, alors

$$\operatorname{Im} A = \operatorname{Im}(AB)$$

*Preuve du lemme*

On va prouver que $\operatorname{Im}(AB) \subseteq \operatorname{Im} A$, puis que $\operatorname{Im} A \subseteq \operatorname{Im}(AB)$. On va voir qu'on n'a pas besoin que $B$ soit inversible dans le premier point, mais qu'on en aura besoin dans le deuxième.

**$\operatorname{Im}(AB) \subseteq \operatorname{Im} A$ :** Soit $\vec{y}$ un vecteur quelconque dans $\operatorname{Im}(AB)$. Par définition de l'image de $AB$, on sait qu'il existe un vecteur $\vec{z}$ tel que $\vec{y} = AB\vec{z}$.
Donc, cela implique :

$$\vec{y} = A\big(B\vec{z}\big) = A\vec{x}, \quad \text{avec } \vec{x} = B\vec{z}$$

On peut ainsi en déduire que $\vec{y}$ est une combinaison linéaire des colonnes de $A$.

**$\operatorname{Im} A \subseteq \operatorname{Im}(AB)$** Soit $\vec{y}$ un vecteur quelconque dans $\operatorname{Im} A$. Par définition, il existe $\vec{x}$ tel que $\vec{y} = A\vec{x}$.
Comme $B$ est inversible, on sait qu'il existe $\vec{z}$ tel que :

$$\vec{z} = B^{-1}\vec{x} \implies B\vec{z} = \vec{x}$$

Donc,

$$\vec{y} = A\vec{x} = AB\vec{z}$$

On en déduit que $\vec{y}$ est dans $\operatorname{Im}(AB)$.

□

*Preuve du théorème*

On veut montrer que $\operatorname{rang} A = \operatorname{rang} A^T$.
Soit $M$ une forme échelonnée de $A$ ; $A$ et $M$ sont de même taille, $m \times n$. On sait qu'il existe une matrice inversible $E \in \mathbb{R}^{m\times m}$ telle que :

$$M = EA$$

En transposant de chaque côté, on trouve :

$$M^T = (EA)^T = A^T E^T$$

On sait que, puisque $E$ est inversible, alors $E^T$ est aussi inversible. Donc, le lemme nous permet d'écrire :

$$\operatorname{Im} M^T = \operatorname{Im}\big(A^T E^T\big) = \operatorname{Im} A^T$$

On sait que $\dim \operatorname{Im} A^T = \operatorname{rang} A^T$, mais on veut encore montrer que $\dim \operatorname{Im} M^T = \operatorname{rang} A$. On peut étudier un exemple :

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

Sa matrice échelonnée réduite équivalent $M$ est :

$$M = \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad M^T = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 2 & 0 \\ 3 & -2 & 0 \end{bmatrix}$$

On déduit à raison que $\dim \operatorname{Im} M^T = \#$ pivots $= \operatorname{rang} A$.
Donc, on a que :

$$\operatorname{rang} A = \operatorname{rang} A^T$$

□

*Autre implication*

Nous avons utilisé l'égalité suivante dans la preuve du théorème :

$$\operatorname{Lgn} A = \operatorname{Im} A^T = \operatorname{Im} M^T = \operatorname{Lgn} M$$

En regardant l'exemple ci-dessus, il est clair que les colonnes non-nulles de $M^T$ sont linéairement indépendantes (puisque les coefficients sur la droite et la gauche des coefficients pivots sont toujours nuls (en effet, $M$ est une forme échelonnée réduite)).

Nous avons donc trouvé une procédure pour déterminer une base de l'espace des lignes de $A$.

Il n'est pas nécessaire d'utiliser la matrice échelonnée réduite, une matrice échelonnée quelconque (☺) est suffisante.

Dans notre exemple ci-dessus :

$$\operatorname{Lgn} A = \operatorname{vect}\left\{\begin{bmatrix} 1 \\ -2 \\ 0 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \\ -2 \end{bmatrix}\right\}$$

Lundi 8 novembre 2021 — **Cours 14 : Changements de base et applications linéaires**

**Sous-espaces intéressants**

On commence à réaliser qu'il y a **quatre sous-espaces** intéressants pour une matrice $A \in \mathbb{R}^{m\times n}$ :

- $\operatorname{Im} A^T$ et $\ker A$, sous-espaces de $\mathbb{R}^n$, et tels que :

$$\dim \operatorname{Im} A^T + \dim \ker A = n$$

- $\operatorname{Im} A$ et $\ker A^T$, sous-espaces de $\mathbb{R}^m$, et tels que :

$$\dim \operatorname{Im} A + \dim \ker A^T = m$$

Pour les équations ci-dessus, on utilise le théorème du rang, qui nous dit $\operatorname{rang} A + \dim \ker A = n$, avec $\operatorname{rang} A = \dim \operatorname{Im} A = \dim \operatorname{Im} A^T = \operatorname{rang} A^T$.
On reviendra là-dessus plus tard.

**Vrai ou faux**

On a déterminé que les opérations élémentaires sur les lignes n'affectent pas l'espace des lignes :

$$\operatorname{Lgn} A = \operatorname{Lgn} M = \operatorname{Lgn}(EA)$$

où $M$ est une forme échelonnée de $A$, $M = EA$.
Nous nous demandons donc si les affirmations suivantes sont vraies ou fausses :

1. $\operatorname{Im} A = \operatorname{Im} M$
2. $\ker A = \ker M$

*Solution au point (1)*

C'est faux. Si on a :

$$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \implies M = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Alors, clairement,

$$\operatorname{Im} A \neq \operatorname{Im} M$$

*Solution au point (2)*

C'est vrai. On peut faire la suite d'implication (qui vont dans les deux sens) suivantes :

$$\begin{aligned} & x \in \ker A \\ \iff & A\overrightarrow{\boldsymbol{x}} = \overrightarrow{\mathbf{0}} \\ \iff & EA\overrightarrow{\boldsymbol{x}} = E\overrightarrow{\mathbf{0}} \\ \iff & M\overrightarrow{\boldsymbol{x}} = \overrightarrow{\mathbf{0}} \\ \iff & \overrightarrow{\boldsymbol{x}} \in \ker M \end{aligned}$$

Il est important de noter qu'on utilise le fait que $E$ soit inversible pour affirmer que :

$$A\overrightarrow{\boldsymbol{x}} = \overrightarrow{\mathbf{0}} \iff EA\overrightarrow{\boldsymbol{x}} = E\overrightarrow{\mathbf{0}}$$

Sinon, on aurait bien une implication "$\Longrightarrow$", mais pas un biconditionnel "$\iff$".

**Résumé**

Nous pouvons résumer ce que nous avons fait à l'aide des théorèmes suivants.

*Théorème 1*

Si $A$ et $B$ sont deux matrices équivalentes selon les lignes, alors leur lignes engendrent le même espace. Si $B$ est échelonnée, alors les lignes non-nulles de $B$ forment une base de cet espace.

*Théorème 2*

Soit $A$ une matrice $m \times n$. L'espace engendré par les colonnes de $A$ et l'espace engendré par les lignes de cette matrice ont la même dimension. Cette dimension commune, le rang de $A$, est égale au nombre de positons pivots de $A$ et vérifie la relation :

$$\operatorname{rang} A + \dim \ker A = n \iff \operatorname{rang} A + \dim \ker A^T = m$$

*Théorème 3 (suite du grand théorème)*

Soit $A$ une matrice $n \times n$. Les propriétés suivantes sont équivalentes :

1. $A$ est inversible.
2. $A$ est équivalente selon les lignes à la matrice unité $n \times n$, $I_n$.
3. $A$ admet $n$ positions de pivot.
4. L'équation $A\overrightarrow{\boldsymbol{x}} = \overrightarrow{\mathbf{0}}$ admet la solution triviale pour unique solution.
5. Les colonnes de $A$ sont linéairement indépendantes.
6. L'application $\overrightarrow{\boldsymbol{x}} \mapsto A\overrightarrow{\boldsymbol{x}}$ est injective.
7. Pour tout vecteur $\overrightarrow{\boldsymbol{b}} \in \mathbb{R}^n$, l'équation $A\overrightarrow{\boldsymbol{x}} = \overrightarrow{\boldsymbol{b}}$ admet au moins une solution (on aurait aussi pu dire "une solution unique", ou au plus une solution).
8. Les colonnes de $A$ engendrent $\mathbb{R}^n$.
9. L'application linéaire $x \mapsto A\overrightarrow{\boldsymbol{x}}$ est surjective.
10. Il existe une matrice $C$ de taille $n \times n$ telle que $CA = I_n$.
11. Il existe une matrice $D$ de taille $n \times n$ telle que $AD = I_n$.
12. $A^T$ est inversible.
13. Les colonnes de $A$ forment une base de $\mathbb{R}^n$.
14. $\operatorname{Im} A = \mathbb{R}^n$
15. $\dim \operatorname{Im} A = n$
16. $\operatorname{rang} A = n$
17. $\ker A = \left\{ \overrightarrow{\mathbf{0}} \right\}$
18. $\dim \ker A = 0$

## 6.10 Changements de base

**Introduction** Plus tôt, on a considéré la matrice de passage $P_{\mathcal{B}}$ qui permet de "passer" des coordonnées de $\vec{x} \in \mathbb{R}^n$ dans une base $\mathcal{B} = \left(\vec{b}_1, \ldots, \vec{b}_n\right)$ de $\mathbb{R}^n$ à $\vec{x}$ lui même, c'est-à-dire, aux coordonnées de $\vec{x}$ dans la base canonique :

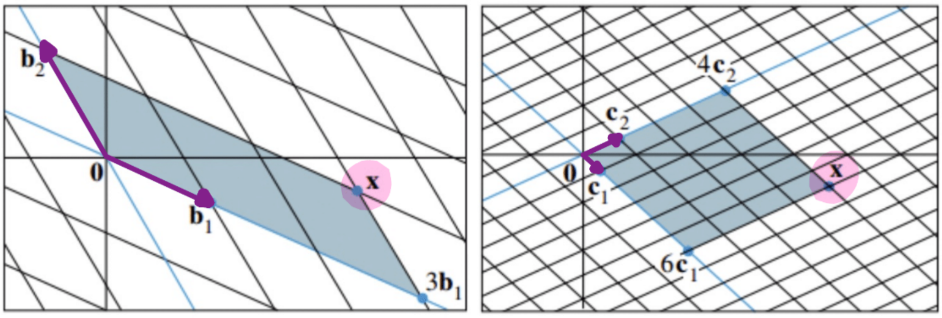
$$\vec{x} = P_{\mathcal{B}}\left[\vec{x}\right]_{\mathcal{B}} \quad \text{avec } P_{\mathcal{B}} = \begin{bmatrix} \vec{b}_1 & \ldots & \vec{b}_n \end{bmatrix}$$

Nous allons généraliser ce concept de deux façons :

1. On va passe d'une base $\mathcal{B}$ à une base $\mathcal{C}$ quelconque ().
2. On va travailler dans un espace (ou sous-espace) quelconque.

**But** Soit $V$ un espace vectoriel, et soient $\mathcal{B} = \left(\vec{b}_1, \ldots, \vec{b}_n\right)$ et $\mathcal{C} = \left(\vec{c}_1, \ldots, \vec{c}_n\right)$ deux bases de $V$ (on sait qu'elles ont le même nombre d'éléments, par le concept de dimension).
Considérons un vecteur $\vec{x} \in V$. Si on a les coordonnées de $\vec{x}$ dans la base $\mathcal{B}$, notées $\left[\vec{x}\right]_{\mathcal{B}}$, comment peut-on trouver les coordonnées de $\vec{x}$ dans la base $\mathcal{C}$, notées $\left[\vec{x}\right]_{\mathcal{C}}$ ?

**Exemple**




En regardant les deux images ci-dessus, on remarque que :

$$\vec{x} = 3\vec{b}_1 + 1\vec{b}_2 = 6\vec{c}_1 + 4\vec{c}_2$$

En d'autres mots :

$$\left[\vec{x}\right]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \quad \text{et} \quad \left[\vec{x}\right]_{\mathcal{C}} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

Le point $\vec{x}$ est le même dans les deux images, c'est la manière de le décrire qui est différente.
On veut trouver ce qui lie $\left[\vec{x}\right]_{\mathcal{B}}$ et $\left[\vec{x}\right]_{\mathcal{C}}$. On se rend compte que :

$$\left[\vec{x}\right]_{\mathcal{C}} = \left[3\vec{b}_1 + \vec{b}_2\right]_{\mathcal{C}}$$

Puisque l'application coordonnées est linéaire :

$$\left[3\vec{b}_1 + \vec{b}_2\right]_{\mathcal{C}} = 3\left[\vec{b}_1\right]_{\mathcal{C}} + \left[\vec{b}_2\right]_{\mathcal{C}} = \begin{bmatrix} \left[\vec{b}_1\right]_{\mathcal{C}} & \left[\vec{b}_2\right]_{\mathcal{C}} \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \left[\vec{b}_1\right]_{\mathcal{C}} & \left[\vec{b}_2\right]_{\mathcal{C}} \end{bmatrix} \left[\vec{x}\right]_{\mathcal{B}}$$

Si on reçoit (magiquement) l'information que $\vec{b}_1 = 4\vec{c}_1 + \vec{c}_2$ et $\vec{b}_2 = -6\vec{c}_1 + \vec{c}_2$, alors :

$$\left[\vec{b}_1\right]_{\mathcal{C}} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \quad \text{et} \quad \left[\vec{b}_2\right]_{\mathcal{C}} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$$

On peut donc affirmer pour tout $\vec{x} \in V$ que :

$$[\vec{x}]_{\mathcal{C}} = \begin{bmatrix} 4 & -6 \\ 1 & 1 \end{bmatrix} [\vec{x}]_{\mathcal{B}}$$

En particulier, avec $\vec{x}$ tel que

$$[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Alors, on obtient :

$$[\vec{x}]_{\mathcal{C}} = \begin{bmatrix} 4 & -6 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 3 \begin{bmatrix} 4 \\ 1 \end{bmatrix} + \begin{bmatrix} -6 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

Ce qui est bien ce qu'on avait trouvé graphiquement.

**Généralisation**

De manière générale, soient $\mathcal{B} = \left(\vec{b}_1, \ldots, \vec{b}_n\right)$ et $\mathcal{C} = \left(\vec{c}_1, \ldots, \vec{c}_n\right)$ deux bases d'un espace vectoriel $V$. Étant donné un vecteur $\vec{x} \in V$, soit

$$[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

son vecteur de coordonnées dans la base $\mathcal{B}$.
Autrement dit, $\vec{x} = a_1 \vec{b}_1 + \ldots + a_n \vec{b}_n$. Nous voulons maintenant trouver $[\vec{x}]_{\mathcal{C}}$.
Pour faire cela, on peut utiliser la linéarité de l'application coordonnées :

$$[\vec{x}]_{\mathcal{C}} = \left[a_1 \vec{b}_1 + \ldots + a_n \vec{b}_n\right]_{\mathcal{C}} = a_1 \left[\vec{b}_1\right]_{\mathcal{C}} = a \left[\vec{b}_1\right]_{\mathcal{C}} + \ldots + a_n \left[\vec{b}_n\right]_{\mathcal{C}}$$

Qu'on peut transformer en produit matrice-vecteur :

$$[\vec{x}]_{\mathcal{C}} = \begin{bmatrix} \left[\vec{b}_1\right]_{\mathcal{C}} & \ldots & \left[\vec{b}_n\right]_{\mathcal{C}} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \left[\vec{b}_1\right]_{\mathcal{C}} & \ldots & \left[\vec{b}_n\right]_{\mathcal{C}} \end{bmatrix} [\vec{x}]_{\mathcal{B}}$$

Cette matrice permet de passer des coordonnées $\mathcal{B}$ vers $\mathcal{C}$. Elle est de taille $n \times n$ et on la note $P_{\mathcal{C} \leftarrow \mathcal{B}}$. On l'appelle la matrice de passage, ou la matrice de changement de base de $\mathcal{B}$ à $\mathcal{C}$. Cette matrice est unique car on a construit cette matrice, et à aucun moment on n'a dû faire de choix.
On remarque que notre résultat peut être appliqué pour n'importe quel espace vectoriel, et non pas uniquement $\mathbb{R}^2$.

**Théorème**

Soit $\mathcal{B} = \left(\vec{b}_1, \ldots, \vec{b}_n\right)$ et $\mathcal{C} = \left(\vec{c}_1, \ldots, \vec{c}_n\right)$ deux bases d'un espace vectoriel $V$. Il existe une matrice $n \times n$ unique, notée $P_{\mathcal{C} \leftarrow \mathcal{B}}$, telle que pour tout vecteur $\vec{x} \in V$ :

$$[\vec{x}]_{\mathcal{C}} = \underset{\mathcal{C} \leftarrow \mathcal{B}}{P} [\vec{x}]_{\mathcal{B}}$$

Les colonnes de $P_{\mathcal{C} \leftarrow \mathcal{B}}$ sont les colonnes de composantes, dans la base $\mathcal{C}$, des vecteurs de la base $\mathcal{B}$. Autrement dit :

$$\underset{\mathcal{C} \leftarrow \mathcal{B}}{P} = \begin{bmatrix} \left[\vec{b}_1\right]_{\mathcal{C}} & \ldots & \left[\vec{b}_n\right]_{\mathcal{C}} \end{bmatrix}$$

**Inverse de la matrice de passage** La matrice suivante passe de $\mathcal{B}$ à $\mathcal{C}$ :

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} \left[\overrightarrow{\boldsymbol{b}}_1\right]_{\mathcal{C}} & \cdots & \left[\overrightarrow{\boldsymbol{b}}_n\right]_{\mathcal{C}} \end{bmatrix}$$

De plus, la matrice suivante passe de $\mathcal{C}$ à $\mathcal{B}$ :

$$\underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = \begin{bmatrix} \left[\overrightarrow{\boldsymbol{c}}_1\right]_{\mathcal{B}} & \cdots & \left[\overrightarrow{\boldsymbol{c}}_n\right]_{\mathcal{B}} \end{bmatrix}$$

On remarque que leur produit donne la matrice identité, puisque cela correspond à transformer les coordonnées de $\overrightarrow{\boldsymbol{x}}$ de la base $\mathcal{B}$ vers la base $\mathcal{C}$, puis de la matrice $\mathcal{C}$ vers la base $\mathcal{B}$ ; donc, on n'a pas bougé :

$$\underset{\mathcal{B}\leftarrow\mathcal{C}}{P}\,\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = I_n$$

On peut donc en déduire que :

$$\left(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\right)^{-1} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P} \iff \left(\underset{\mathcal{B}\leftarrow\mathcal{C}}{P}\right)^{-1} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$$

**Exemple** Plus tôt, on avait un exemple dans $V = \mathbb{R}^2$ avec :

$$\overrightarrow{\boldsymbol{b}}_1 = 4\overrightarrow{\boldsymbol{c}}_1 + \overrightarrow{\boldsymbol{c}}_2 \quad \text{et} \quad \overrightarrow{\boldsymbol{b}}_2 = -6\overrightarrow{\boldsymbol{c}}_1 + \overrightarrow{\boldsymbol{c}}_2$$

Donc :

$$\left[\overrightarrow{\boldsymbol{b}}_1\right]_{\mathcal{C}} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \quad \text{et} \quad \left[\overrightarrow{\boldsymbol{b}}_2\right]_{\mathcal{C}} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$$

Ainsi, on en déduit :

$$\left[\overrightarrow{\boldsymbol{x}}\right]_{\mathcal{C}} = \begin{bmatrix} 4 & -6 \\ 1 & 1 \end{bmatrix} \left[\overrightarrow{\boldsymbol{x}}\right]_{\mathcal{B}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P} \left[\overrightarrow{\boldsymbol{x}}\right]_{\mathcal{B}}$$

On peut alors calculer :

$$\begin{bmatrix} \left[\overrightarrow{\boldsymbol{c}}_1\right]_{\mathcal{B}} & \left[\overrightarrow{\boldsymbol{c}}_2\right]_{\mathcal{B}} \end{bmatrix} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = \left(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\right)^{-1} = \begin{bmatrix} 4 & -6 \\ 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1/10 & 3/5 \\ -1/10 & 2/5 \end{bmatrix}$$

On en déduit donc :

$$\left[\overrightarrow{\boldsymbol{c}}_1\right]_{\mathcal{B}} = \begin{bmatrix} 1/10 \\ -1/10 \end{bmatrix} \quad \text{et} \quad \left[\overrightarrow{\boldsymbol{c}}_2\right]_{\mathcal{B}} = \begin{bmatrix} 3/5 \\ 1/5 \end{bmatrix}$$

De plus :

$$\overrightarrow{\boldsymbol{c}}_1 = \frac{1}{10}\overrightarrow{\boldsymbol{b}}_1 - \frac{1}{10}\overrightarrow{\boldsymbol{b}}_2 \quad \text{et} \quad \overrightarrow{\boldsymbol{c}}_2 = \frac{3}{5}\overrightarrow{\boldsymbol{b}}_1 + \frac{2}{5}\overrightarrow{\boldsymbol{b}}_2$$

## 6.11 Applications linéaires d'espaces quelconques

**Retour aux applications**

Nous voulons exprimer une application linéaire par rapport à deux bases.
Soit $T : V \mapsto W$ une application linéaire d'un espace vectoriel $V$ vers un espace vectoriel $W$. Dans le cas particulier $T : \mathbb{R}^n \mapsto \mathbb{R}^m$, on a vu comment exprimer $T$ au moyen d'une matrice $A \in \mathbb{R}^{m\times n}$, $T(\vec{x}) = A\vec{x}$. On peut faire quelque chose de similaire dans des espaces abstraits si on a une base de $V$ et une base de $W$.
Soit $\mathcal{V} = (\vec{v}_1, \ldots, \vec{v}_n)$ une base de $V$ (donc $\dim V = n$), et soit $\mathcal{W} = (\vec{w}_1, \ldots, \vec{w}_m)$ (donc $\dim W = m$).
Tout $\vec{x} \in V$ s'écrit de manière unique dans la base $\mathcal{V}$ :

$$\vec{x} = c_1\vec{v}_1 + \ldots + c_n\vec{v}_n \quad \text{avec } [\vec{x}]_{\mathcal{V}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

On se demande maintenant que vaut $T(\vec{x})$. On utilise la linéarité de $T$ :

$$T(\vec{x}) = T(c_1\vec{v}_1 + \ldots + c_n\vec{v}_n) = c_1T(\vec{v}_1) + \ldots + c_nT(\vec{v}_n)$$

Chacun des vecteurs $T(\vec{v}_1), \ldots, T(\vec{v}_n)$ appartient à $W$. Donc, on peut les exprimer dans la base $\mathcal{W}$ :

$$[T(\vec{v}_1)]_{\mathcal{W}}, \ldots, [T(\vec{v}_n)]_{\mathcal{W}}$$

Ce sont $n$ vecteurs-colonnes dans $\mathbb{R}^m$. Par linéarité de l'application coordonnées, on trouve :

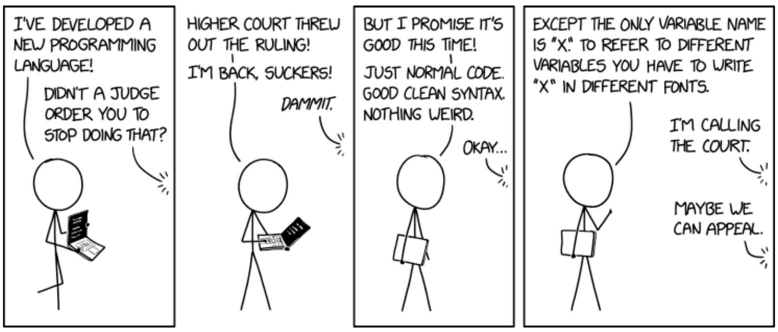
$$[T(\vec{x})]_{\mathcal{W}} = [c_1T(\vec{v}_1) + \ldots + c_nT(\vec{v}_n)]_{\mathcal{W}} = c_1[T(\vec{v}_1)]_{\mathcal{W}} + \ldots + c_n[T(\vec{v}_n)]_{\mathcal{W}}$$

On peut donc l'écrire sous la forme d'un produit matrice-vecteur :

$$\begin{bmatrix} [T(\vec{v}_1)]_{\mathcal{W}} & \ldots & [T(\vec{v}_n)]_{\mathcal{W}} \end{bmatrix} \underbrace{\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}}_{[\vec{x}]_{\mathcal{V}}}$$

*Note personnelle*

La différence entre $V$, la matrice, et $\mathcal{V}$, la base, me fait penser à ce XKCD :




`https://xkcd.com/2309/`

**Théorème**

Soit $T : V \mapsto W$ une application linéaire et soient $\mathcal{V}$ et $\mathcal{W}$ des bases de $V$ et $W$, respectivement. Il existe une matrice unique $A$ telle que, pour tout $\vec{x} \in V$ :

$$[T(\vec{x})]_{\mathcal{W}} = A[\vec{x}]_{\mathcal{V}}$$

Cette matrice est donnée par :

$$A = \begin{bmatrix} [T(\overrightarrow{\boldsymbol{v}}_1)]_{\mathcal{W}} & \dots & [T(\overrightarrow{\boldsymbol{v}}_n)]_{\mathcal{W}} \end{bmatrix}$$

où $\overrightarrow{\boldsymbol{v}}_1, \ldots, \overrightarrow{\boldsymbol{v}}_n$ sont les vecteurs de la base $\mathcal{V}$.

# Chapitre 7

# Valeurs propres et vecteurs propres

**Introduction**

Considérons la matrice suivante :

$$A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$$

L'application linéaire $\overrightarrow{x} \mapsto A\overrightarrow{x}$ de $\mathbb{R}^2$ vers $\mathbb{R}^2$ a *parfois* un effet très simple à décrire. En effet :

$$A\overrightarrow{e}_1 = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad A\overrightarrow{e}_2 = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

Jusque là, rien de spécial, mais regardons les vecteurs suivants :

$$\overrightarrow{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \implies A\overrightarrow{u} = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2\overrightarrow{u}$$

$$\overrightarrow{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \implies A\overrightarrow{u} = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \overrightarrow{v}$$

Maintenant, si nous voulons appliquer 17 fois $A$ au vecteurs $\overrightarrow{e}_1$, cela serait très compliqué, cependant si on essaie de le faire sur le vecteur $\overrightarrow{u}$ on obtient :

$$A^{17}\overrightarrow{u} = 2A^{16}\overrightarrow{u} = \ldots = 2^{17}\overrightarrow{u}$$

De la même manière, si on le fait sur $\overrightarrow{v}$ :

$$A^{17}\overrightarrow{v} = A^{16}\overrightarrow{v} = \ldots = \overrightarrow{v}$$

Cela nous motive donc à creuser les vecteurs non-nuls (sinon l'équation est triviale) tels que, avec un scalaire $\lambda$ :

$$A\overrightarrow{x} = \lambda\overrightarrow{x}$$

**Définition des vecteurs et valeurs propres**

Soit $A$ une matrice $n \times n$ (on travaille uniquement avec des matrices carrées). On appelle **vecteur propre** de $A$ tout vecteur non nul $\overrightarrow{x}$ tel que $A\overrightarrow{x} = \lambda\overrightarrow{x}$ pour un certain scalaire $\lambda$.

Un scalaire $\lambda$ est appelé **valeur propre** de $A$ si l'équation $A\overrightarrow{x} = \lambda\overrightarrow{x}$ admet au moins une solution non triviale $\overrightarrow{x}$ ; un tel vecteur $\overrightarrow{x}$ est appelé **vecteur propre associé à $\lambda$**.

**Exemple 1**

Soit la matrice et le vecteur suivants :

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}, \quad \overrightarrow{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$$

On veut savoir si $\overrightarrow{u}$ est un vecteur propre de $A$. On a :

$$A\overrightarrow{u} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\overrightarrow{u}$$

Donc, $\overrightarrow{u}$ est bien un vecteur propre de $A$, et sa valeur propre associée est $-4$. Regardons maintenant si le vecteur suivant est un vecteur propre de $A$ :

$$\overrightarrow{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \implies A\overrightarrow{v} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix}$$

On en déduit que $\overrightarrow{v}$ n'est pas un vecteur propre de $A$.

**Exemple 2** Soit la matrice suivante :

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$$

On veut savoir si 7 est une valeur propre de $A$. Si c'est le cas, on veut trouver les vecteurs propres associés. Pour répondre à ceci, on doit déterminer l'équation suivante :

$$A\overrightarrow{x} = 7\overrightarrow{x}$$

On remarque que cette équation est linéaire, et on peut mettre le $\overrightarrow{x}$ à gauche :

$$A\overrightarrow{x} - 7\overrightarrow{x} = \overrightarrow{\mathbf{0}} \iff A\overrightarrow{x} - 7I_2\overrightarrow{x} = 0 \iff (A - 7I_2)\overrightarrow{x} = \overrightarrow{\mathbf{0}}$$

En d'autres mots, on cherche le noyau de la matrice $A - 7I_2$, on veut savoir s'il existe un $\overrightarrow{x}$ non nul qui satisfait cette équation (dans la définition des vecteurs propres, $\overrightarrow{x}$ ne peut pas être nul). On a :

$$A - 7I_2 = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix}$$

Par échelonnement, on trouve que l'ensemble des solutions $\overrightarrow{x}$ est :

$$\text{vect}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

Puisque le kernel n'est pas vide, on en déduit que 7 est une valeur propre de $A$, et le vecteur suivant est un vecteur propre associé :

$$\overrightarrow{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

**Cas général** Pour une matrice $A \in \mathbb{R}^{n\times n}$ et un scalaire $\lambda$, les affirmations suivantes sont équivalentes :

- $\lambda$ est une valeur propre de $A$
- $A\overrightarrow{x} = \lambda\overrightarrow{x}$ a une solution non-nulle, par définition
- $A\overrightarrow{x} - \lambda\overrightarrow{x} = \overrightarrow{\mathbf{0}}$ a une solution non-nulle
- $A\overrightarrow{x} - \lambda I_n\overrightarrow{x} = \overrightarrow{\mathbf{0}}$ a une solution non-nulle
- $(A - \lambda I_n)\overrightarrow{x} = \overrightarrow{\mathbf{0}}$ a une solution non-nulle
- $A - \lambda I_n$ n'est pas inversible puisque son noyau n'est pas de dimension 0
- $\det(A - \lambda I_n) = 0$

**Observation 1** Les valeurs propres de $A$ sont exactement les solutions de l'équation scalaire (non linéaire) suivante :

$$\det(A - \lambda I_n) = 0$$

On appelle cette équation l'équation caractéristique, et le polynôme qui en découle (ce sera toujours un polynôme), le polynôme caractéristique.

*Exemple*

Prenons la matrice suivante :

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$$

On peut calculer $A - \lambda I_2$ :

$$A - \lambda I_2 = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 1 - \lambda & 6 \\ 5 & 2 - \lambda \end{bmatrix}$$

Calculons maintenant son déterminant :

$$\det(A - \lambda I_2) = \det \begin{bmatrix} 1 - \lambda & 6 \\ 5 & 2 - \lambda \end{bmatrix} = (1 - \lambda)(2 - \lambda) - 30$$

Ce qu'on peut simplifier :

$$\det(A - \lambda I_2) = \lambda^2 - 3\lambda - 28 = (\lambda - 7)(\lambda + 4)$$

Les racines de ce polynôme sont 7 et $-4$ ; ce sont les valeurs propres de $A$. Pour trouver les vecteurs propres associés à ces valeurs propres, on peut calculer le kernel de $A - 7I_2$ et $A + 4I_2$.

**Observation 2** Si $\lambda$ est valeur propre de $A$, alors tous les vecteurs propres $\overrightarrow{x}$ associés à $\lambda$ sont solution de $(A - \lambda I_n)\overrightarrow{x} = \overrightarrow{0}$. L'ensemble solution de cette équation est un sous-espace de $\mathbb{R}^n$ appelé **sous-espace propre** de $A$ associé à la valeur propre $\lambda$.

*Remarque* On peut justifier cette remarque en montrant qu'on cherche le kernel d'une matrice $A - \lambda I_n$, qui est un sous-espace.

**Exemple 1** Nous voulons trouver les valeurs propres de la matrice suivantes :

$$A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$$

Calculons $A - \lambda I_3$ :

$$A - \lambda I_3 = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} = \begin{bmatrix} 4 - \lambda & -1 & 6 \\ 2 & 1 - \lambda & 6 \\ 2 & -1 & 8 - \lambda \end{bmatrix}$$

Calculons son déterminant maintenant :

$$\det(A - \lambda I_3) = (4 - \lambda)(1 - \lambda)(8 - \lambda) - 12 - 12 + 6(4 - \lambda) + 2(8 - \lambda) - 12(1 - \lambda)$$

Simplifions le polynôme caractéristique de $A$ :

$$\det(A - \lambda I_3) = -\lambda^3 + 13\lambda^2 - 40\lambda + 36 = -(\lambda - 9)(\lambda - 2)^2$$

Ainsi, on en déduit que les valeurs propres de $A$ sont 9 et 2.
Cherchons maintenant l'espace propre de $A$ associé à la valeur propre $\lambda = 2$. C'est l'ensemble solution de $A\overrightarrow{x} = 2\overrightarrow{x}$, c'est-à-dire celui de $(A - 2I_2)\overrightarrow{x} = \overrightarrow{0}$ :

$$\left(\begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix} - \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}\right)\overrightarrow{x} = \overrightarrow{0} \implies \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}\overrightarrow{x} = \overrightarrow{0}$$

Puisque les colonnes sont toutes égales, le rang de cette matrice est 1, on en déduit que la dimension de son noyau est 2, et on devine la base suivante (on aurait aussi pu la trouver par échelonnement) :

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Donc, l'espace propre de $A$ associé à $\lambda = 2$ est un plan dans $\mathbb{R}^3$ est donné par :

$$\operatorname{vect}\left\{\begin{bmatrix}1\\2\\0\end{bmatrix}, \begin{bmatrix}3\\0\\-1\end{bmatrix}\right\}$$

Pour tout $\overrightarrow{x}$ dans ce sous-espace, on a $A\overrightarrow{x} = 2\overrightarrow{x}$.

**Exemple 2** Nous voulons trouver les valeurs propres de la matrice suivante :

$$A = \begin{bmatrix}1 & 4 & 6\\0 & 3 & 9\\0 & 0 & 8\end{bmatrix}$$

Ainsi, on trouve :

$$A - \lambda I_3 = \begin{bmatrix}1-\lambda & 4 & 6\\0 & 3-\lambda & 9\\0 & 0 & 8-\lambda\end{bmatrix}$$

On se rend compte que $A$ est une matrice triangulaire, et que $A-\lambda I_3$ reste triangulaire (puisqu'on soustrait juste une matrice diagonale). Donc, le déterminant est très simple à calculer :

$$\det(A - \lambda I_3) = (1-\lambda)(3-\lambda)(8-\lambda)$$

On en déduit que les valeurs propres de $A$ sont $1, 3, 8$, les éléments sur sa diagonale.

**Théorème** Si $A \in \mathbb{R}^{n\times n}$ est triangulaire (inférieure, supérieure ou diagonale), alors ses valeurs propres sont ses entités diagonales : $a_{11}, a_{22}, \ldots, a_{nn}$.

*Remarque* Attention, il est important de noter qu'échelonner une matrice ne permet pas de trouver ses valeurs propres, puisqu'un échelonnement modifie les valeurs propres.

**Exemple 1** Soit la matrice suivante :

$$A = \begin{bmatrix}-5 & 2 & 3\\3 & -4 & 1\\1 & 2 & -3\end{bmatrix} \implies \det(A - \lambda I) = -\lambda(\lambda+6)^2$$

On en déduit que les valeurs propres de cette matrice sont $-6$ et $0$. Puisque $0$ est une valeur propre nulle, cela veut dire qu'il existe un $\overrightarrow{x}$ non-nul tel que $A\overrightarrow{x} = \lambda\overrightarrow{x} = \overrightarrow{0}$. Donc, $A$ n'est pas inversible.

**Exemple 2 : Fibonacci** Nous souhaitons trouver une formule explicite à la suite de Fibonacci en utilisant des matrices.
On défini cette suite tel que $f_0 = 1$, $f_1 = 1$, et $f_{k+1} = f_k + f_{k-1}$ (notez que le professeur a utilisé la notation $n^k = f_k$, mais je préfère mettre l'index en indice plutôt qu'en exposant). Ainsi, on trouve que

$$f_2 = 2, \quad f_3 = 3, \quad f_4 = 5, \quad \ldots$$

On se rend compte que $f_{k+1} = f_k + f_{k-1}$ ressemble fortement à une relation linéaire, donc on la met sous forme matricielle :

$$\begin{bmatrix}f_1\\f_2\end{bmatrix} = \begin{bmatrix}f_1\\f_1+f_0\end{bmatrix} = \underbrace{\begin{bmatrix}0 & 1\\1 & 1\end{bmatrix}}_{A}\begin{bmatrix}f_0\\f_1\end{bmatrix} = A\begin{bmatrix}f_0\\f_1\end{bmatrix}$$

De la même manière :

$$\begin{bmatrix}f_2\\f_3\end{bmatrix} = \begin{bmatrix}f_2\\f_2+f_1\end{bmatrix} = \begin{bmatrix}0 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}f_1\\f_2\end{bmatrix} = \begin{bmatrix}0 & 1\\1 & 1\end{bmatrix}^2\begin{bmatrix}f_0\\f_1\end{bmatrix} = A^2\begin{bmatrix}f_0\\f_1\end{bmatrix}$$

Ainsi, de manière générale :

$$\begin{bmatrix} f_k \\ f_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} f_{k-1} \\ f_k \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^2 \begin{bmatrix} f_{k-2} \\ f_{k-1} \end{bmatrix} = \ldots = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^k \begin{bmatrix} f_0 \\ f_1 \end{bmatrix} = A^k \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

On a besoin d'obtenir la puissance $k$-ème de $A$. On peut y arriver en étudiant ses valeurs propres et ses vecteurs propres :

$$A - \lambda I_2 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix}$$

On peut maintenant trouver le polynôme caractéristique :

$$\det(A - \lambda I_2) = -\lambda(1-\lambda) - 1 = \lambda^2 - \lambda - 1$$

On trouve donc que les deux valeurs propres sont :

$$\lambda_1 = \frac{1+\sqrt{5}}{2} = \varphi, \quad \lambda_2 = \frac{1-\sqrt{5}}{2} = -\frac{1}{\varphi} = \overline{\varphi}$$

Les vecteurs propres associés sont (les trouver est laissé en exercice au lecteur) :

$$\overrightarrow{\boldsymbol{v}}_1 = \begin{bmatrix} 1 \\ \varphi \end{bmatrix}, \quad \overrightarrow{\boldsymbol{v}}_2 = \begin{bmatrix} -\varphi \\ 1 \end{bmatrix}$$

car $A\overrightarrow{\boldsymbol{v}}_1 = \lambda_1 \overrightarrow{\boldsymbol{v}}_1$ et $A\overrightarrow{\boldsymbol{v}}_2 = \lambda_2 \overrightarrow{\boldsymbol{v}}_2$.
On remarque que $\overrightarrow{\boldsymbol{v}}_1$ et $\overrightarrow{\boldsymbol{v}}_2$ sont linéairement indépendants, donc ils forment une base de $\mathbb{R}^2$. Ainsi, le vecteur initial peut s'écrire comme une combinaison linéaire de $\overrightarrow{\boldsymbol{v}}_1$ et $\overrightarrow{\boldsymbol{v}}_2$ :

$$c_1 \overrightarrow{\boldsymbol{v}}_1 + c_2 \overrightarrow{\boldsymbol{v}}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \implies \begin{bmatrix} 1 & -\varphi \\ \varphi & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

En résolvant ce système, on trouve :

$$c_1 = \frac{1+\varphi}{1+\varphi^2}, \quad c_2 = \frac{1-\varphi}{1+\varphi^2}$$

Ceci nous permet d'écrire la séquence de Fibonacci comme suit :

$$\begin{bmatrix} f_k \\ f_{k-1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^k \begin{bmatrix} 1 \\ 1 \end{bmatrix} = A^k \left(c_1 \overrightarrow{\boldsymbol{v}}_1 + c_2 \overrightarrow{\boldsymbol{v}}_2\right) = c_1 A^k \overrightarrow{\boldsymbol{v}}_1 + c_2 A^k \overrightarrow{\boldsymbol{v}}_2$$

Or, puisque ce sont des vecteurs propres :

$$A^k \overrightarrow{\boldsymbol{v}}_1 = \lambda_1^k \overrightarrow{\boldsymbol{v}}_1 = \varphi^k \begin{bmatrix} 1 \\ \varphi \end{bmatrix}, \quad A^k \overrightarrow{\boldsymbol{v}}_2 = \lambda_2^k \overrightarrow{\boldsymbol{v}}_2 = \overline{\varphi}^k \begin{bmatrix} -\varphi \\ 1 \end{bmatrix}$$

Donc, en mettant tout en semble :

$$\begin{bmatrix} f_k \\ f_{k+1} \end{bmatrix} = c_1 A^k \overrightarrow{\boldsymbol{v}}_1 + c_2 A^k \overrightarrow{\boldsymbol{v}}_2 = \frac{1+\varphi}{1+\varphi^2} \varphi^k \begin{bmatrix} 1 \\ \varphi \end{bmatrix} + \frac{1-\varphi}{1+\varphi^2} \left(\frac{-1}{\varphi}\right)^k \begin{bmatrix} -\varphi \\ 1 \end{bmatrix}$$

En prenant la première composante, ceci nous donne la formule directe suivante :

$$f_k = \frac{1+\varphi}{1+\varphi^2} \varphi^k + \frac{1-\varphi}{1+\varphi^2} \left(\frac{-1}{\varphi}\right)^k (-\varphi)$$

On peut se rendre compte des simplifications suivantes :

$$\frac{1+\varphi}{1+\varphi^2} \cdot \frac{1}{\varphi} = \frac{1}{\sqrt{5}}, \quad \frac{1-\varphi}{1+\varphi^2}(-\varphi) \frac{1}{-\frac{1}{\varphi}} = -\frac{1}{\sqrt{5}}$$

Ainsi, on trouve que :

$$f_k = \frac{1}{\sqrt{5}} \varphi^{k+1} - \frac{1}{\sqrt{5}} \left(-\frac{1}{\varphi}\right)^{k+1} = \frac{\varphi^{k+1} - \overline{\varphi}^{k+1}}{\sqrt{5}}$$

qui est une formule qui se trouve dans le passage sous le métro de l'arrêt EPFL (et accessoirement une formule que je trouve, personnellement, très belle).

Lundi 15 novembre 2021 — **Cours 16 : Équation caractéristique et matrices semblables**

**Observation** Dans l'exemple de Fibonacci, on a observé que les vecteurs propres de $A$ pour les deux valeurs propres distinctes étaient linéairement indépendantes, ce qui nous a permis d'exprimer le vecteur initial comme combinaison linéaire de ceux-ci. Ce n'était pas un accident.

**Théorème** Soit $A$ une matrice $n \times n$ et $\vec{v}_1, \ldots, \vec{v}_r$ des vecteurs propres de $A$ associés à des valeurs propres *distinctes* $\lambda_1$, …, $\lambda_r$. Ces vecteurs propres sont linéairement indépendants.

**Exemple** La matrice suivante est de taille $3 \times 3$ :

$$A = \begin{bmatrix} 1 & 8 & 9 \\ 0 & 2 & 7 \\ 0 & 0 & 3 \end{bmatrix}$$

Elle a 3 valeurs propres, $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$ (les valeurs sur la diagonale puisque la matrice est triangulaire), qui sont distinctes. Donc, si $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$ sont des vecteurs propres associés à $\lambda_1, \lambda_2, \lambda_3$, ils sont linéairement indépendants et forment une base de $\mathbb{R}^3$.

**Multiplication des valeurs propres** Si les vecteurs propres de $A$ sont $\lambda_1, \ldots, \lambda_p$ et qu'elles sont associées à des vecteurs propres $\vec{v}_1, \ldots, \vec{v}_p$. On se demande ce qu'on peut dire des valeurs propres et des vecteurs propres $B = 2A$ :

$$A\vec{v}_1 = \lambda_1 \vec{v}_1 \implies B\vec{v}_1 = (2A)\vec{v}_1 = 2\big(A\vec{v}_1\big) = 2\big(\lambda_1\vec{v}_1\big) = (2\lambda_1)\vec{v}_1$$

Donc, les vecteurs propres restent les mêmes, et les valeurs propres doublent.

## 7.1 Équation caractéristique

**Équation caractéristique** Soit $A$ une matrice carré de taille $n \times n$. On a défini qu'un scalaire $\lambda$ est valeur propre de $A$ si et seulement si $A\vec{x} = \vec{x}$ admet une solution $\vec{x} \neq \vec{0}$. On a déduit de cette définition que $\lambda$ est valeur propre de $A$ exactement si :

$$\det(A - \lambda I_n) = 0$$

On appelle ceci **l'équation caractéristique de $A$**. Dans tous nos exemples, le membre de gauche était un polynôme de degré $n$ en $\lambda$. C'est toujours le cas.
On appelle $p_A$ le **polynôme caractéristique de $A$** défini par :

$$p_A(\lambda) = \det(A - \lambda I_n)$$

Pour voir que c'est un polynôme de degré $n$, regardons des petites valeurs de $n$. Si $n = 1$ :

$$A = \begin{bmatrix} a_{11} \end{bmatrix} \implies A - \lambda I_1 = \begin{bmatrix} a_{11} - \lambda \end{bmatrix} \implies p_A(\lambda) = \det(A - \lambda I_1) = a_{11} - \lambda$$

Si $n = 2$ :

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \implies A - \lambda I_2 = \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix}$$

Le polynôme caractéristique de cette matrice est donc :

$$p_A(\lambda) = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21}$$

En faisant la même chose pour $n = 3$, on obtient :

$$\begin{aligned} p_A(\lambda) = {} & \det(A - \lambda I_3) \\ = {} & (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ & - (a_{11} - \lambda)a_{23}a_{32} - a_{12}a_{21}(a_{33} - \lambda) - a_{13}a_{31}(a_{22} - \lambda) \\ = {} & -\lambda^3 + \ldots\lambda^2 + \ldots\lambda + \det(A) \end{aligned}$$

Ce n'est pas très important, mais on remarque que le terme indépendant (la constante) est toujours le déterminant de la matrice. En effet, si on évalue $p_A(0) = \det(A)$ dans la définition du polynôme caractéristique, donc la constante doit bien être égale au déterminant.

**Généralisation** En regardant la définition du déterminant, on voit que $\det(A - \lambda I_n)$ est une somme de très nombreux termes ; chacun d'entre eux est un produit de $n$ coefficients de la matrice $A - \lambda I_n$.
Comme les coefficients de $A - \lambda I_n$ sont des polynômes de degré $\leq 1$, chaque terme de la somme est de degré $\leq n$. De plus, un terme de degré $n$ vient du produit des coefficients diagonaux.

**Théorème** Les valeurs propres de $A \in \mathbb{R}^{n \times n}$ sont les racines du polynôme caractéristique :

$$p_A(\lambda) = \det(A - \lambda I_n)$$

Ce polynôme est exactement de degré $n$.

> *Implication* On sait qu'un polynôme de degré $n$ a exactement $n$ racines, si on compte leur multiplicités et qu'on admet les racines complexes.
> Donc, toute matrice $A \in \mathbb{R}^{n \times n}$ a exactement $n$ valeurs propres $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$, en répétant chaque racine selon sa multiplicité (on l'appelle la **multiplicité algébrique** de la valeur propre ; on verra un autre type de multiplicité plus tard). On a donc :
>
> $$p_A(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda)$$

**Exemple 1** Soit la matrice suivante :

$$A = \begin{bmatrix} 3 & 9 & 3 & 1 \\ 0 & 6 & -2 & 0 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 6 \end{bmatrix} \implies A - \lambda I_4 = \begin{bmatrix} 3-\lambda & 9 & 3 & 1 \\ 0 & 6-\lambda & -2 & 0 \\ 0 & 0 & 0-\lambda & -2 \\ 0 & 0 & 0 & 6-\lambda \end{bmatrix}$$

Son polynôme caractéristique est donné par :

$$p_A(\lambda) = \det(A - \lambda I_4) = (3-\lambda)(6-\lambda)(0-\lambda)(6-\lambda) = (0-\lambda)(3-\lambda)(6-\lambda)^2$$

Ainsi, les valeurs propres de $A$ sont $0, 3, 6$ et leur multiplicité algébrique sont $1, 1, 2$ respectivement. $A$ est de taille $4 \times 4$ et la somme des multiplicités algébriques et $1 + 1 + 2 = 4$, comme attendu.

**Exemple 2** Soit la matrice suivante (qui n'est pas triangulaire !) :

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \implies A - \lambda I_2 = \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix}$$

On a :

$$p_A(\lambda) = \det(A - \lambda I_2) = (-\lambda)^2 + 1 = \lambda^2 + 1 = \lambda^2 - i^2 = (\lambda + i)(\lambda - i)$$

Les valeurs propres de $A$ sont donc $i$ et $-i$, et leur multiplicité algébrique sont 1 et 1.
Géométriquement, on se rend compte que $\overrightarrow{x} \mapsto A\overrightarrow{x}$ est une rotation (on peut voir son comportement en l'appliquant sur $\overrightarrow{e}_1$ et $\overrightarrow{e}_2$). Ainsi, s'il y avait une valeur propre réelle, alors cela voudrait dire qu'il existerait un vecteur $\overrightarrow{x}$ tel que quand on fait tourner $\overrightarrow{x}$ de $\frac{\pi}{2}$, alors cela donnerait un multiple de ce vecteur.

## 7.2 Matrices semblables

**Définition des matrices semblables**

Soient $A$ et $B$ deux matrices de taille $n \times n$.
On dit que $A$ est **semblable** à $B$ s'il existe $P$ inversible de taille $n \times n$ telle que :

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

*Observation* $B = P^{-1}AP$ est équivalent à :

$$A = PBP^{-1}$$

*Remarque* Ceci définit une relation d'équivalence sur les matrices de taille $n \times n$. En effet, selon cette définition on peut montrer que :

- $A$ est semblable à $A$ (réflexivité).
- Si $A$ est semblable à $B$, alors $B$ est semblable à $A$ (symétrie).
- Si $A$ est semblable à $B$ et $B$ est semblable à $C$, alors $A$ est semblable à $C$ (transitivité).

Par exemple, pour le point (2). Puisque $A$ est semblable à $B$, il existe $Q$ inversible telle que $B = Q^{-1}AQ$. Ainsi, en prenant $P = Q^{-1}$ :

$$B = Q^{-1}AQ \implies A = QBQ^{-1} = P^{-1}BP$$

Il faut noter que, puisque si $A$ est semblable à $B$ implique que $B$ est semblable à $A$, on dit donc que "$A$ et $B$ sont semblables", et non pas "$A$ est semblable à $B$".

**Théorème**

Si $A$ et $B$ sont semblables, alors $p_A = p_B$.
En particulier, $A$ et $B$ ont donc les mêmes valeurs propres, avec les mêmes multiplicités algébriques.

*Preuve* Comme $A$ et $B$ sont semblables, il existe $P$ inversible telle que :

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

De plus, on remarque que :

$$I_n = P^{-1}P = P^{-1}I_nP$$

Ainsi, on en déduit que :

$$B - \lambda I_n = P^{-1}AP - \lambda P^{-1}I_nP$$

Cependant, puisque $\lambda$ est un scalaire, on obtient que c'est égal à :

$$B - \lambda I_n = P^{-1}AP - P^{-1}(\lambda I_n)P = P^{-1}(A - \lambda I_n)P$$

À partir de là, on peut calculer le polynôme caractéristique :

$$p_B(\lambda) = \det(B - \lambda I_n) = \det\big(P^{-1}(A - \lambda I_n)P\big)$$

On sait que le déterminant est multiplicatif, donc :

$$p_B(\lambda) = \det\big(P^{-1}\big)\det(A - \lambda I_n)\det(P) = \det\big(P^{-1}\big)\det(P)\det(A - \lambda I_n)$$

De plus, on remarque que :

$$\det\big(P^{-1}\big)\det(P) = \det\big(P^{-1}P\big) = \det(I_n) = 1$$

On peut donc en déduire que :

$$p_B(\lambda) = \det(A - \lambda I_n) = p_A(\lambda)$$

*Remarque* La réciproque de ce théorème n'est pas vraie. En effet, prenons les matrices suivantes :

$$A = \begin{bmatrix} 3 & 2 \\ 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

Leur polynômes caractéristiques sont égaux, ils sont les deux $(3-\lambda)^2$. Cependant, elle ne sont pas semblables.
Supposons par l'absurde qu'elles sont semblables, donc qu'il existe $P$ inversible telle que :

$$B = P^{-1}AP \iff A = PBP^{-1}$$

Mais, $B = 3I_2$, donc :

$$A = P(3I_2)P^{-1} = 3 \cdot PI_2P^{-1} = 3I_2$$

Cependant, $A \neq 3I_2$, donc c'est une contradiction.

**Utilité des matrices semblables** On se demande d'où pourraient apparaître des matrices semblables. Une des réponses est les changements de base (on verra d'autres réponses).
Soit $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$, une application linéaire. On sait qu'il existe $A \in \mathbb{R}^{3\times 3}$ telle que :

$$T(\vec{x}) = A\vec{x}, \quad \forall \vec{x} \in \mathbb{R}^3$$

Disons maintenant que nous avons envie d'exprimer $T$ par rapport à une base $\mathcal{B} = \left(\vec{b}_1, \vec{b}_2, \vec{b}_3\right)$. Soit la matrice de passage suivante :

$$P = \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \vec{b}_3 \end{bmatrix} \in \mathbb{R}^{3\times 3} \implies \vec{v} = P[\vec{v}]_{\mathcal{B}} \text{ et } [\vec{v}]_{\mathcal{B}} = P^{-1}\vec{v} \quad \forall \vec{v} \in \mathbb{R}^3$$

Ainsi, on peut changer notre transformation linéaire de base :

$$[T(\vec{x})]_{\mathcal{B}} = P^{-1}T(\vec{x}) = P^{-1}A\vec{x} = P^{-1}AP[\vec{v}]_{\mathcal{B}}$$

On a donc trouvé que la matrice de $T$ par rapport à $\mathcal{B}$ est $P^{-1}AP$, qui est semblable à $A$.
On se demande maintenant s'il existe une base "spéciale" qui rend la matrice "simple".

## 7.3 Diagonalisation

**Définition** Soit $A \in \mathbb{R}^{n\times n}$. On dit que $A$ est **diagonalisable** si elle est semblable à une matrice diagonale. C'est-à-dire s'il existe $D$ diagonale et $P$ inversible telles que :

$$A = PDP^{-1} \iff P^{-1}AP = D \iff AP = PD$$

**Exemple** Soient les matrices suivantes :

$$A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}, \quad D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$$

On peut vérifier facilement que $A = PDP^{-1}$, donc $A$ est diagonalisable. En effet, $\det(P) = -1 \neq 0$, donc $P$ est inversible. De plus, on n'a pas besoin de calculer $P^{-1}$ :

$$A = PDP^{-1} \iff AP = PD$$

Ce qu'on peut vérifier :

$$AP = \begin{bmatrix} 5 & 3 \\ -5 & -6 \end{bmatrix}, \quad PD = \begin{bmatrix} 5 & 3 \\ -5 & -6 \end{bmatrix}$$

Jeudi 18 novembre 2021 — **Cours 17 : Diagonalisation**

**Propriété de la diagonalisation**

Si $A = PDP^{-1}$, où :

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Alors, on a :

$$A^2 = A \cdot A = PDP^{-1} \cdot PDP^{-1} = PDDP^{-1} = PD^2P^{-1}$$

Calculer $D^2$ est très facile :

$$D^2 = \begin{bmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{bmatrix}$$

De manière très générale :

$$A^k = A \cdot A^{k-1} = PDP^{-1} \cdot PD^{k-1}P^{-1} = PD^kP^{-1}$$

Ce n'est pas très important, mais on voit que le comportement de $A^k$ pour $k$ grand est déterminé par ses valeurs propres (si leur valeur absolue est strictement plus petite que 0, alors $\lambda^k$ tend vers 0).

**Exemple**

Soient les matrices suivantes :

$$A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}, \quad D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$$

On a déjà vérifié que $A = PDP^{-1}$. Donc :

$$A^k = PD^kP^{-1} = P\begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix}P^{-1}$$

**Observation**

On observe quelque chose de plus dans l'exemple ci-dessus. Les colonnes de $P$ sont des vecteurs propres de $A$, et les entrées diagonales de $D$ sont les vecteurs propres associés :

$$AP = A\begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} A\begin{bmatrix} 1 \\ -1 \end{bmatrix} & A\begin{bmatrix} 1 \\ -2 \end{bmatrix} \end{bmatrix}$$

De plus :

$$PD = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 5\begin{bmatrix} 1 \\ -1 \end{bmatrix} & 3\begin{bmatrix} 1 \\ -2 \end{bmatrix} \end{bmatrix}$$

Or, puisque $AP = PD$, cela veut dire que les deux matrices sont égales. Ainsi :

$$A\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 5\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad A\begin{bmatrix} 1 \\ -2 \end{bmatrix} = 3\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

On trouve donc que, par définition, 3 et 5 sont les valeurs propres de $A$, et les vecteurs sont des vecteurs propres associés à ces valeurs propres.

**Théorème**

$A \in \mathbb{R}^{n \times n}$ est diagonalisable si et seulement si elle a $n$ vecteurs propres $\vec{v}_1, \ldots, \vec{v}_n$ linéairement indépendants.

Dans ce cas, on peut choisir :

$$P = \begin{bmatrix} \vec{v}_1 & \ldots & \vec{v}_n \end{bmatrix}, \quad D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

où $A\vec{v}_i = \lambda_i \vec{v}_i$.
On appelle $\left(\vec{v}_1, \ldots, \vec{v}_n\right)$ une **base de vecteurs propres de $A$**.

*Idée de preuve* Si les vecteurs $\vec{v}_1, \ldots, \vec{v}_n$ — associés aux valeurs propres $\lambda_1, \ldots, \lambda_n$ respectivement — existent, alors :

$$AP = A\begin{bmatrix} \vec{v}_1 & \ldots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} A\vec{v}_1 & \ldots & A\vec{v}_n \end{bmatrix}$$

Or, puisque ce sont des vecteurs propres, c'est égal à :

$$\begin{bmatrix} \lambda_1\vec{v}_1 & \ldots & \lambda_n\vec{v}_n \end{bmatrix} = \begin{bmatrix} \vec{v}_1 & \ldots & \vec{v}_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = PD$$

De plus, nous avons besoin que $P$ soit inversible, donc ses colonnes, les vecteurs propres de $A$, doivent être linéairement indépendantes.

**Théorème** Tout matrice $n \times n$ admettant $n$ valeurs propres *distinctes* est diagonalisable.

*Preuve* Ce théorème découle directement du théorème ci-dessus, et de celui qu'on a vu plus tôt, juste après l'exemple de Fibonacci, qui nous dit que si nous avons $n$ valeurs propres distinctes, alors leurs vecteurs propres associés sont linéairement indépendants.

*Remarque* Ceci est une condition suffisante, mais non pas nécessaire.

**Exemple 1** Soit la matrice suivante :

$$A = \begin{bmatrix} -1 & 7 & 9 \\ 0 & 2 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

On remarque rapidement que la matrice est triangulaire, donc que ses valeurs propres sont sur sa diagonale, et ainsi que ses valeurs propres sont $-1$, $2$ et $0$. Puisqu'elles sont distinctes, on en déduit que cette matrice est diagonalisable.

**Exemple 2** Soit la matrice suivante :

$$A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

On se demande si elle est diagonalisable. Commençons par trouver ses valeurs propres. Avec un peu de travail, on trouve que :

$$p_A(\lambda) = -\lambda^3 - 3\lambda^2 + 4 = (1 - \lambda)(2 + \lambda)^2$$

Donc, les valeurs propres de $A$ sont $\lambda_1 = 1$ et $\lambda_2 = -2$, et leur multiplicité algébrique sont 1 et 2, respectivement. Si elles étaient distinctes on aurait gagné, mais on doit aller un peu plus loin.
Cherchons les vecteurs propres associés à $\lambda_1 = 1$.

$$\ker(A - \lambda_1 I_3) = \ker\begin{bmatrix} 0 & 3 & 3 \\ -3 & -6 & -3 \\ 3 & 3 & 0 \end{bmatrix} = \text{vect}\left\{\begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}\right\}$$

qui est le sous-espace propre de $A$ associé à $\lambda_1$. Étudions maintenant les vecteurs propres associés à $\lambda_2 = -2$ :

$$\ker(A - \lambda_2 I_3) = \ker \begin{bmatrix} 3 & 3 & 3 \\ -3 & -3 & -3 \\ 3 & 3 & 3 \end{bmatrix} = \operatorname{vect}\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\}$$

On peut vérifier notre résultat en multipliant $A$ par nos trois vecteurs.
On remarque que les vecteurs propres sont linéairement indépendants (il n'y a pas beaucoup de surprise puisque des vecteurs associés à des valeurs propres distinctes sont différents, et nous avons choisi les vecteurs du kernel de manière à ce qu'ils soient linéairement indépendants) :

$$\vec{v}_1 = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

Ainsi, ces vecteurs forment une base de vecteurs propres de $A$. On peut en déduire que $A$ est diagonalisable, et que $A = PDP^{-1}$, où :

$$P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 1 \\ -1 & 0 & -1 \end{bmatrix}, \quad D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

**Exemple 3** Soit la matrice suivante :

$$A = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

On se pose la même question, on se demande si elle est diagonalisable. Premièrement, trouvons les valeurs propres de $A$ :

$$p_A(\lambda) = \det(A - \lambda I_3) = -\lambda^3 - 3\lambda^2 + 4 = (1 - \lambda)(2 + \lambda)^2$$

On remarque c'est le même polynôme caractéristique que dans l'exemple ci-dessus. Ainsi, on en déduit que les valeurs propres de $A$ sont $\lambda_1 = 1$ et $\lambda_2 = -2$, avec, comme multiplicité algébrique, 1 et 2 respectivement.
Cherchons maintenant les vecteurs propres associés à ces valeurs propres. On trouve :

$$\ker(A - \lambda_1 I_3) = \ker \begin{bmatrix} 1 & 4 & 3 \\ -4 & -7 & -3 \\ 3 & 3 & 0 \end{bmatrix} = \operatorname{vect}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

$$\ker(A - \lambda_2 I_3) = \ker \begin{bmatrix} 4 & 4 & 3 \\ -4 & -4 & -3 \\ 3 & 3 & 3 \end{bmatrix} = \operatorname{vect}\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}$$

On a donc seulement deux vecteurs propres linéairement indépendants pour $A$, et il est impossible d'en trouver 3. Ainsi, on peut en déduire que $A$ n'est pas diagonalisable.

**Généralisation** On se demande maintenant ce qu'il faut pour qu'une matrice $A \in \mathbb{R}^{n \times n}$ soit diagonalisable. Par notre théorème, on sait qu'il nous faut $n$ vecteurs propres linéairement indépendants.
Si $\lambda$ est une valeur propre de $A$, on se demande donc combien de vecteurs propres associés à $\lambda$ qui sont linéairement indépendants on peut trouver. Tous les vecteurs propres associés à $\lambda$ sont dans l'espace associé à $\lambda$, qui est donné par $\ker(A - \lambda I_n)$.
Or, on sait qu'on peut choisir au maximum

$$\dim \ker(A - \lambda I_n)$$

vecteurs dans ce sous espace. On appelle ceci la **multiplicité géométrique** de $\lambda$ (à ne pas confondre avec la multiplicité algébrique).

Soit $\lambda_1, \ldots, \lambda_p$ les valeurs propres distinctes de $A \in \mathbb{R}^{n \times n}$. On sait que $\lambda_1$ nous donne $\dim \ker(A - \lambda_1 I_n)$ vecteurs propres linéaire indépendants. De la même manière, $\lambda_2$ nous donne $\dim \ker(A - \lambda_2 I_n)$ vecteurs propres linéairement indépendants. De plus, puisque $\lambda_1 \neq \lambda_2$, on en déduit qu'ils sont aussi linéairement indépendants une fois mis ensemble. On a donc déjà trouvé

$$\dim \ker(A - \lambda_1 I_n) + \dim \ker(A - \lambda_2 I_n)$$

vecteurs propres linéairement indépendants.
On continue avec $\lambda_3, \ldots, \lambda_p$. Au total, on a trouvé exactement la somme des multiplicité géométrique

$$\dim \ker(A - \lambda_1 I_n) + \ldots + \dim \ker(A - \lambda_p I_n)$$

vecteurs propres linéairement indépendants pour $A$.
On remarque que la multiplicité d'une valeur propre $\lambda$ est nécessairement inférieure ou égale à sa multiplicité algébrique (et plus grande ou égale à 1, par définition des valeurs propres la dimension de l'espace propre associé est toujours au moins 1). Une justification de ce fait (qui est hors de la matière du cours) peut être vue dans le paragraphe suivant.
Donc, pour une matrice $A \in \mathbb{R}^{n \times n}$ avec $p$ valeurs propres distinctes, on a :

$$[\text{somme des } p \text{ multiplicités géométriques}] \leq [\text{somme des } p \text{ multiplicités algébriques}] = n$$

Or, pour avoir $n$ vecteurs propres linéairement indépendants, il nous faut que la somme des $p$ multiplicités géométriques soit $n$, et la seule possibilité pour que ce soit possible c'est que la multiplicité géométrique soit toujours égale à la multiplicité algébrique. On peut donc conclure que $A$ est diagonalisable si et seulement si les multiplicités algébriques et géométriques sont toutes égales.

**Théorème (hors de la matière du cours)** La multiplicité géométrique d'une valeur propre est toujours plus petite ou égale à sa multiplicité algébrique.

*Preuve (toujours hors cours, ça change pas)* Supposons que $\lambda$ a une multiplicité géométrique de $p$. Donc, on a $p$ vecteurs propres $\vec{v}_1, \ldots, \vec{v}_p$ linéairement indépendants associés à cette valeur propre. On peut compléter cette famille de vecteurs, avec des vecteurs quelconques () de $\mathbb{R}^n$ (qui ne sont pas forcément des vecteurs propres) de manière à créer une base de $\mathbb{R}^n$. Appelons ces vecteurs $\vec{v}_{p+1}, \ldots, \vec{v}_n$. Ainsi, on peut prendre la matrice suivante :

$$P = \begin{bmatrix} \vec{v}_1 & \ldots & \vec{v}_p & \vec{v}_{p+1} & \ldots & \vec{v}_n \end{bmatrix}$$

On sait que $P$ est inversible puisque ses colonnes forment une base de $\mathbb{R}^n$. Prenons $B = P^{-1}AP$. Cette matrice est semblable à $A$, donc :

$$p_B(\lambda) = p_A(\lambda)$$

De plus, puisque $P^{-1}P = I_n$, on voit que :

$$P^{-1} \begin{bmatrix} \vec{v}_1 & \ldots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} \vec{e}_1 & \ldots & \vec{e}_n \end{bmatrix}$$

Ainsi :

$$\begin{aligned}
B &= P^{-1}AP \\
&= P^{-1}\begin{bmatrix} A\vec{v}_1 & \dots & A\vec{v}_p & A\vec{v}_{p+1} & \dots & A\vec{v}_n \end{bmatrix} \\
&= P^{-1}\begin{bmatrix} \lambda\vec{v}_1 & \dots & \lambda\vec{v}_p & A\vec{v}_{p+1} & \dots & A\vec{v}_n \end{bmatrix} \\
&= \begin{bmatrix} \lambda\vec{e}_1 & \dots & \lambda\vec{e}_p & A\vec{e}_{p+1} & \dots & A\vec{e}_n \end{bmatrix} \\
&= \begin{bmatrix} \lambda & \cdots & 0 & & & \\ \vdots & \ddots & \vdots & & & \\ 0 & \cdots & \lambda & A\vec{e}_{p+1} & \cdots & A\vec{e}_n \\ \vdots & & \vdots & & & \\ 0 & \cdots & 0 & & & \end{bmatrix}
\end{aligned}$$

On remarque que les $p$ premières colonnes de $B$ sont diagonales. Donc, en calculant le polynôme caractéristique de $B$ — qui est le même que celui de $A$ puisque ces matrices sont semblables — on aura :

$$p_A(x) = p_B(x) = \det(B - xI_n) = (\lambda - x)^p(\dots)$$

Ainsi, $\lambda$ a une multiplicité algébrique de, au moins, $p$.

□

## Lundi 22 novembre 2021 — **Cours 18 : Fin de la propreté et début des produits scalaires**

**Théorème** Soit $A \in \mathbb{R}^{n\times n}$ une matrice quelconque et soient $\lambda_1, \dots \lambda_p$ les valeurs propres distinctes de $A$.
On a toujours que $\dim\ker(A - \lambda_i I_n)$ (la dimension géométrique de $\lambda_i$) est inférieure ou égale à la multiplicité algébrique de $\lambda_i$ pour chaque $i$.
$A$ est diagonalisable si et seulement si $\dim\ker(A - \lambda_i I_n)$ est égale à la multiplicité algébrique de $\lambda_i$ pour tout $i \in \{1, 2, \dots, p\}$

*Remarque* Si $A$ est diagonalisable, alors pour chaque $\lambda_i$, on peut trouver une base $\mathcal{B}_i$ de $\ker(A - \lambda_i I_n)$. Mises toutes ensemble, les bases $\mathcal{B}_1, \dots, \mathcal{B}_p$ forment une base de vecteurs propres de $A$ pour $\mathbb{R}^n$.

**Exemple** Soit la matrice suivante :

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

On remarque que la matrice est triangulaire inférieure, donc que ses valeurs propres sont sur sa diagonale. Ainsi, on trouve que $\lambda_1 = 5$ et $\lambda_2 = -3$, avec une multiplicité algébrique de 2 chacune. En effet, le polynôme caractéristique de $A$ est donné par :

$$p_A(\lambda) = \det(A - \lambda I_4) = (5 - \lambda)^2(-3 - \lambda)^2$$

Calculons maintenant les espaces propres associés aux valeurs propres :

$$\ker(A - \lambda_1 I_4) = \ker\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 4 & -8 & 0 \\ -1 & -2 & 0 & -8 \end{bmatrix} = \text{vect}\left\{\begin{bmatrix} 8 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 2 \\ -1 \end{bmatrix}\right\}$$

$$\ker(A - \lambda_2 I_4) = \ker\begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ -1 & -2 & 0 & 0 \end{bmatrix} = \text{vect}\left\{\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}\right\}$$

On en déduit que l'ensemble suivant est une base de $\mathbb{R}^4$ de vecteurs propres de $A$.

$$\left\{\begin{bmatrix} 8 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}\right\}$$

Ainsi, on en déduit que $A = PDP^{-1}$, où :

$$P = \begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ -1 & -1 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

**Retour à Fibonacci**

Revenons à notre exemple de Fibonacci, voir ce qu'on peut faire plus rapidement avec nos nouveaux outils. Pour rappel, on avait trouvé :

$$\begin{bmatrix} f_k \\ f_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^k \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

On peut prendre :

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \implies p_A(\lambda) = \det\begin{bmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix} = -\lambda(1-\lambda) - 1 = \lambda^2 - \lambda - 1 = (\lambda - \varphi)(\lambda - \overline{\varphi})$$

Donc, $\lambda_1 = \varphi$ et $\lambda_2 = \overline{\varphi}$, où :

$$\varphi = \frac{1+\sqrt{5}}{2}, \quad \overline{\varphi} = \frac{1-\sqrt{5}}{2}$$

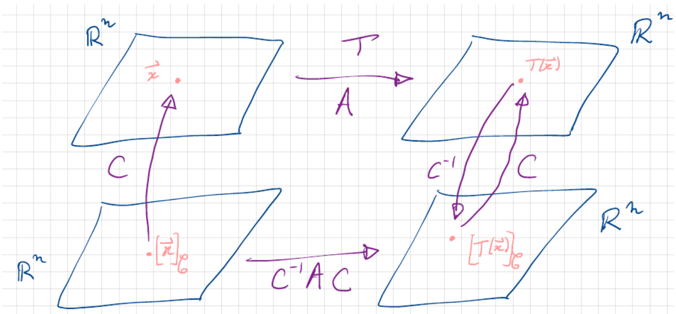
Les deux valeurs propres sont distinctes, donc $A$ est diagonalisable. Dès lors, $A = PDP^{-1}$, et on sait comment calculer $P$ et $D$. Ainsi :

$$\begin{bmatrix} f_k \\ f_{k+1} \end{bmatrix} = A^k \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \left(PDP^{-1}\right)^k \begin{bmatrix} 1 \\ 1 \end{bmatrix} = PD^kP^{-1}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = P\begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix} P^{-1}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

De manière générale, les matrices diagonalisables nous permettent d'étudier beaucoup de systèmes dynamiques. Elles nous permettent aussi de résoudre des systèmes d'équations différentielles, ou d'étudier les chaines de Markov.

**Changement de base d'application linéaire**

Soit $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ une application linéaire avec $T\left(\overrightarrow{x}\right) = A\overrightarrow{x}$, où $A \in \mathbb{R}^{n \times n}$ est la matrice de $T$ pour la base canonique.
Soit $\mathcal{C} = \left(\overrightarrow{c}_1, \ldots, \overrightarrow{c}_n\right)$ une base quelconque de $\mathbb{R}^n$. On se demande quelle est la matrice de $T$ pour la base $\mathcal{C}$. En d'autres mots, on cherche $B \in \mathbb{R}^{n \times n}$ telle que $\left[T\left(\overrightarrow{x}\right)\right]_{\mathcal{C}} = B\left[\overrightarrow{x}\right]_{\mathcal{C}}$.




On trouve donc que $B = C^{-1}AC$, et donc $B$ et $A$ sont semblables.
Note personnelle : au lieu de faire un dessin, on peut aussi utiliser les matrices de passage (dont la notation est très cohérente). Notons $\mathcal{E}$, la base canonique. On a $C = P_{\mathcal{C}\leftarrow\mathcal{E}}$, et $A$ "prend un vecteur en base $\mathcal{E}$ en entrée", et "donne un vecteur en base $\mathcal{E}$ en sortie". Donc :

$$[T(\vec{x})]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{E}}{P} A \underset{\mathcal{E}\leftarrow\mathcal{C}}{P} [\vec{x}]_{\mathcal{C}} \implies B = \underset{\mathcal{C}\leftarrow\mathcal{E}}{P} A \underset{\mathcal{E}\leftarrow\mathcal{C}}{P} = C^{-1}AC$$

**Vecteurs propres et applications linéaires**

Soit $A$ une matrice diagonalisable. On sait donc qu'elle est semblable à une matrice diagonale, et on sait que $A$ admet une base de vecteurs propres $\vec{v}_1, \dots, \vec{v}_n$ avec valeurs propres $\lambda_1, \dots, \lambda_n$ de sorte que $A = PDP^{-1}$ avec :

$$P = \begin{bmatrix} \vec{v}_1 & \dots & \vec{v}_n \end{bmatrix}, \quad D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Si $A$ est la matrice de $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ en base canonique, alors $D$ est la matrice de $T$ (de la même application) par rapport à la base de vecteurs propres $\mathcal{B} = (\vec{v}_1, \dots, \vec{v}_n)$.
Si $\vec{x} = a_1\vec{v}_1 + \ldots + a_n\vec{v}_n$, on a :

$$T(\vec{x}) = T(a_1\vec{v}_1 + \ldots + a_n\vec{v}_n) = a_1T(\vec{v}_1) + \ldots + a_nT(\vec{v}_n) = a_1\lambda_1\vec{v}_1 + \ldots + a_n\lambda_n\vec{v}_n$$

Ainsi, on en déduit que :

$$[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \implies [T(\vec{x})]_{\mathcal{B}} = \begin{bmatrix} a_1\lambda_1 \\ \vdots \\ a_n\lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} [\vec{x}]_{\mathcal{B}}$$

On en déduit donc que la matrice qui permet de décrire $T$ en base $\mathcal{B}$ est la matrice avec les valeurs propres sur la diagonale.

# Chapitre 8

# Orthogonalité et méthode des moindres carrés

**Introduction**

Considérons un système d'équations linéaires $A\overrightarrow{\boldsymbol{x}} = \overrightarrow{\boldsymbol{b}}$, où $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$et $\overrightarrow{\boldsymbol{x}} \in \mathbb{R}^n$ (l'inconnue).
Il se pourrait que ce système n'ait pas de solution, mais disons que nous ne voulons pas abandonner. En effet, il se pourrait qu'il existe $\overrightarrow{\boldsymbol{x}}$ tel que :

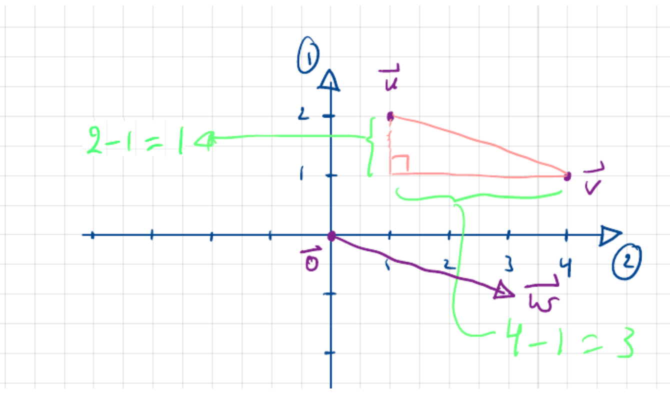
$$A\overrightarrow{\boldsymbol{x}} \approx \overrightarrow{\boldsymbol{b}}$$

Notre vecteur $\overrightarrow{\boldsymbol{b}}$ pourrait venir de mesures, qui sont rarement exactes. Cela pourrait par exemple être utile dans ce scénario.
Il nous faut définir de nouveaux concepts de “taille” et de “distance” dans $\mathbb{R}^n$ et $\mathbb{R}^m$.

**Distance**

Nous avons déjà vu une formule pour calculer la distance entre deux vecteurs dans $\mathbb{R}^2$. On peut utiliser la distance Euclidienne.
Par exemple, soient les deux vecteurs suivants :

$$\overrightarrow{\boldsymbol{u}} = \begin{bmatrix}1\\2\end{bmatrix}, \quad \overrightarrow{\boldsymbol{v}} = \begin{bmatrix}4\\1\end{bmatrix}$$




Prenons la différence entre ces deux vecteurs :

$$\overrightarrow{\boldsymbol{w}} = \overrightarrow{\boldsymbol{v}} - \overrightarrow{\boldsymbol{u}} = \begin{bmatrix}4\\1\end{bmatrix} - \begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}3\\-1\end{bmatrix}$$

Par Pythagore, la distance entre $\overrightarrow{\boldsymbol{u}}$ et $\overrightarrow{\boldsymbol{v}}$ est donc :

$$d^2 = 3^2 + 1^2 = 10 \implies d = \sqrt{10}$$

On se rend compte que la distance entre $\overrightarrow{\boldsymbol{u}}$ et $\overrightarrow{\boldsymbol{v}}$ est équivalente à la distance entre $\overrightarrow{\boldsymbol{w}}$ et l'origine, $\overrightarrow{\boldsymbol{o}}$. On appelle ceci la **norme** de $\overrightarrow{\boldsymbol{w}}$.

**Définition du produit scalaire**

Soient les deux vecteurs de $\mathbb{R}^n$ suivants :

$$\vec{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

Le **produit scalaire** de $\vec{u}$ et $\vec{v}$ est un scalaire noté $\vec{u} \bullet \vec{v}$ défini comme suit :

$$\vec{u} \bullet \vec{v} = \vec{u}^T \vec{v} = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \ldots + u_n v_n$$

**Exemple**

Soient les deux vecteurs suivants :

$$\vec{u} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} 1 \\ 4 \\ -1 \end{bmatrix}$$

Ainsi, on obtient que :

$$\vec{u} \bullet \vec{v} = 3 \cdot 1 + (-1)4 + 2(-1) = -3 = 1 \cdot 3 + 4(-1) + (-1)2 = \vec{v} \bullet \vec{u}$$

De plus :

$$\vec{u} \bullet \vec{u} = 3 \cdot 3 + (-1)(-1) + 2 \cdot 2 = 9 + 1 + 4 = 14$$

Jeudi 25 novembre 2021 — **Cours 19 : Norme et orthogonalité**

**Propriétés**

Soient $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$ et $c \in \mathbb{R}$. On a :

1. $\vec{u} \bullet \vec{v} = \vec{v} \bullet \vec{u}$
2. $(\vec{u} + \vec{v}) \bullet \vec{w} = \vec{u} \bullet \vec{w} + \vec{v} \bullet \vec{w}$
3. $c(\vec{u}) \bullet \vec{v} = c(\vec{u} \bullet \vec{v}) = \vec{u} \bullet (c\vec{v})$
4. $\vec{u} \bullet \vec{u} \geq 0$
5. $\vec{u} \bullet \vec{u} = 0 \iff \vec{u} = \vec{0}$.

**Exemple**

Soient $\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{w} \in \mathbb{R}^n$. Alors :

$$\begin{aligned}(7\vec{u}_1 - 2\vec{u}_2 + 4\vec{u}_3) \bullet \vec{w} &\overset{(2)}{=} (7\vec{u}_1 - 2\vec{u}_2) \bullet \vec{w} + (4\vec{u}_3) \bullet \vec{w} \\ &\overset{(2)}{=} (7\vec{u}_1) \bullet \vec{w} + (-2\vec{u}) \bullet \vec{w} + (4\vec{u}_3) \bullet \vec{w} \\ &\overset{(3)}{=} 7(\vec{u}_1 \bullet \vec{w}) + (-2)(\vec{u}_2 \bullet \vec{w}) + 4(\vec{u}_3 \bullet \vec{w})\end{aligned}$$

Très généralement, on trouve :

$$(c_1\vec{u}_1 + \ldots + c_p\vec{u}_p) \bullet \vec{w} = c_1(\vec{u}_1 \bullet \vec{w}) + \ldots + c_p(\vec{u}_p \bullet \vec{w})$$

**Définition de norme**

Soit $\vec{v}$ un vecteur de $\mathbb{R}^n$.
La **norme** (ou longueur) de $\vec{v}$ est le scalaire positif ou nul noté $\|\vec{v}\|$ défini par :

$$\|\vec{v}\| = \sqrt{\vec{v} \bullet \vec{v}} = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

*Observation* On sait que $\vec{v} \bullet \vec{v}$ est toujours positif pour tout $\vec{v}$, on ne devra donc jamais calculer la racine carrée d'un nombre négatif. Ainsi, la norme est toujours un nombre réel positif.

**Exemple**

Soit le vecteur suivant :

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \in \mathbb{R}^2$$

On peut calculer sa norme :

$$\left\|\overrightarrow{v}\right\|^2 = \overrightarrow{v} \bullet \overrightarrow{v} = 3^2 + 2^2 = 13 \implies \left\|\overrightarrow{v}\right\| = \sqrt{13}$$

**Observation**

Soit $\overrightarrow{v} \in \mathbb{R}^n$ et $c \in \mathbb{R}$ quelconques.
Regardons le carré de la norme de $c\overrightarrow{v}$ :

$$\left\|c\overrightarrow{v}\right\|^2 = \left(c\overrightarrow{v}\right) \bullet \left(c\overrightarrow{v}\right) = c\left(\overrightarrow{v} \bullet \left(c\overrightarrow{v}\right)\right) = c^2\left(\overrightarrow{v} \bullet \overrightarrow{v}\right) = c^2\left\|\overrightarrow{v}\right\|^2$$

On en déduit donc que :

$$\left\|c\overrightarrow{v}\right\| = |c|\left\|\overrightarrow{v}\right\|$$

**Normalisation**

Si $\left\|\overrightarrow{u}\right\| = 1$, on dit que $\overrightarrow{u}$ est un **vecteur unitaire**.
Pour $\overrightarrow{v}$ non-nul, on peut toujours prend le vecteur $\overrightarrow{u}$ suivant, de sorte que $\overrightarrow{u}$ est unitaire et pointe dans la même direction que $\overrightarrow{v}$ :

$$\overrightarrow{u} = \frac{1}{\left\|\overrightarrow{v}\right\|}\overrightarrow{v}$$

En effet, puisque $\frac{1}{\left\|\overrightarrow{v}\right\|} > 0$, alors $\overrightarrow{u}$ pointe bien dans la même direction que $\overrightarrow{v}$. De plus :

$$\left\|\overrightarrow{u}\right\| = \left\|\frac{1}{\left\|\overrightarrow{v}\right\|}\overrightarrow{v}\right\| = \left|\frac{1}{\left\|\overrightarrow{v}\right\|}\right|\left\|\overrightarrow{v}\right\| = \frac{\left\|\overrightarrow{v}\right\|}{\left\|\overrightarrow{v}\right\|} = 1$$

On appelle cela la **normalisation** de $\overrightarrow{v}$.

**Exemple**

Soit le vecteur suivant :

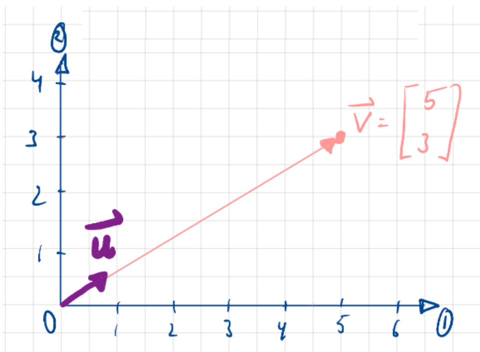
$$\overrightarrow{v} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Calculons sa norme :

$$\left\|\overrightarrow{v}\right\|^2 = \overrightarrow{v} \bullet \overrightarrow{v} = 5^2 + 3^2 = 34 \implies \left\|\overrightarrow{v}\right\| = \sqrt{34}$$

Ainsi, par normalisation, on obtient :

$$\overrightarrow{u} = \frac{1}{\left\|\overrightarrow{v}\right\|}\overrightarrow{v} = \frac{1}{\sqrt{34}}\begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 5/\sqrt{34} \\ 3/\sqrt{34} \end{bmatrix}$$




**Définition de distance**

Soient $\overrightarrow{u}$ et $\overrightarrow{v}$ deux vecteurs de $\mathbb{R}^n$.
On définit la **distance** entre $\overrightarrow{u}$ et $\overrightarrow{v}$ comme :

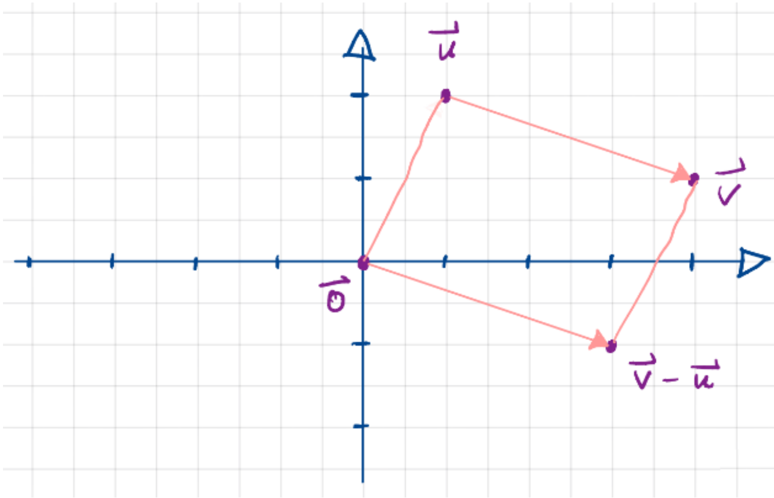
$$\operatorname{dist}\left(\overrightarrow{u}, \overrightarrow{v}\right) = \left\|\overrightarrow{v} - \overrightarrow{u}\right\|$$

*Observation*

On peut voir que :

$$\operatorname{dist}\left(\overrightarrow{u}, \overrightarrow{v}\right) = \operatorname{dist}\left(\overrightarrow{v}, \overrightarrow{u}\right) = \left\|\overrightarrow{u} - \overrightarrow{v}\right\| = \left\|\overrightarrow{v} - \overrightarrow{u}\right\|$$

*Interprétation géométrique*




**Dans $\mathbb{R}^3$**

Soient les vecteurs $\vec{u}, \vec{v} \in \mathbb{R}^3$ :

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

On peut calculer leur distance :

$$\begin{aligned} \operatorname{dist}(\vec{u}, \vec{v}) &= \|\vec{u} - \vec{v}\| \\ &= \sqrt{(\vec{v} - \vec{u}) \bullet (\vec{v} - \vec{u})} \\ &= \sqrt{(v_1 - u_1)^2 + (v_2 - u_2)^2 + (v_3 - u_3)^2} \end{aligned}$$

**Définition d'orthogonalité**

On dit que deux vecteurs $\vec{u}, \vec{v} \in \mathbb{R}^n$ sont **orthogonaux** (l'un à l'autre) si :

$$\vec{u} \bullet \vec{v} = \vec{\mathbf{0}}$$

*Observation*

On remarque que $\vec{\mathbf{0}}$ est orthogonal avec n'importe quel vecteur puisque :

$$\vec{\mathbf{0}} \bullet \vec{v} = 0 \cdot v_1 + \ldots + 0 \cdot v_n = 0$$

**Théorème de Pythagore**

$\vec{u}, \vec{v} \in \mathbb{R}^n$ sont orthogonaux si et seulement si :

$$\|\vec{u} + \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2$$

*Remarque*

On dit que l'orthogonalité capture la perpendicularité (ou les "angles droits") parce qu'elle nous donne ce théorème dans $\mathbb{R}^n$

*Preuve*

$$\begin{aligned} &\|\vec{u} + \vec{v}\|^2 - \|\vec{u}\|^2 - \|\vec{v}\|^2 \\ &= (\vec{u} + \vec{v}) \bullet (\vec{u} + \vec{v}) - \vec{u} \bullet \vec{u} - \vec{v} \bullet \vec{v} \\ &= \vec{u} \bullet \vec{u} + \vec{u} \bullet \vec{v} + \vec{v} \bullet \vec{u} + \vec{v} \bullet \vec{v} + \vec{v} \bullet \vec{v} - \vec{u} \bullet \vec{u} - \vec{v} \bullet \vec{v} \\ &= 2\vec{u} \bullet \vec{v} \end{aligned}$$

Et ceci est nul si et seulement si $\vec{u} \bullet \vec{v} = 0$, c'est-à-dire, si et seulement si $\vec{u}$ et $\vec{v}$ sont orthogonaux.

□

**Exemple** Soient les deux vecteurs suivants :

$$\vec{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

On remarque qu'ils sont orthogonaux. En effet :

$$\vec{u} \bullet \vec{v} = 2(-1) + 1(2) = 0$$

Calculons le carré de leur norme :

$$\left\|\vec{u}\right\|^2 = 2^2 + 1^2 = 5, \quad \left\|\vec{v}\right\|^2 = (-1)^2 + 2^2 = 5, \quad \left\|\vec{u} + \vec{v}\right\|^2 = 1^2 + 3^2 = 10$$

Or, puisque $5 + 5 = 10$, notre théorème est correct (et heureusement) !

## 8.1 Orthogonalité à tout un sous-espace vectoriel

**Observation** Si $\vec{u}$ est orthogonal à $\vec{v}$, alors $\vec{u}$ est orthogonal à tous les vecteurs dans $\text{vect}\{\vec{v}\}$. En effet, si $\vec{w} \in \text{vect}\{\vec{v}\}$, alors $\vec{w} = c\vec{v}$ pour un certain scalaire $c$. Ainsi :

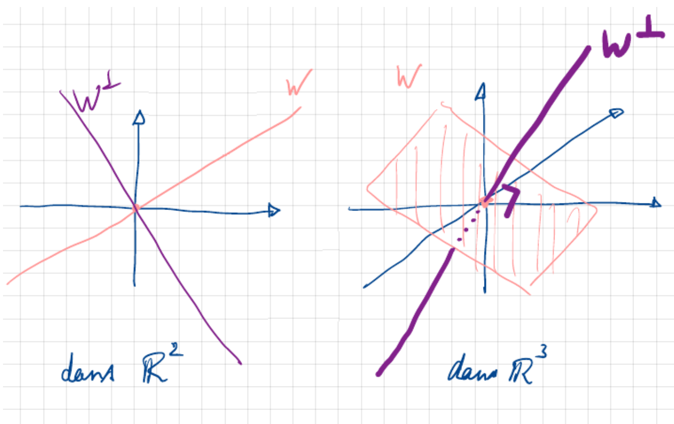
$$\vec{u} \bullet \vec{w} = \vec{u} \bullet (c\vec{v}) = c(\vec{u} \bullet \vec{v}) = c \cdot 0 = 0$$

Puisque $\vec{u}$ et $\vec{v}$ sont orthogonaux par hypothèse.

*Généralisation* On se demande alors : étant donné un sous-espace $W$ de $\mathbb{R}^n$, est-ce qu'il existe des vecteurs orthogonaux à tous les vecteurs de $W$ ?

**Définition** Soit $W$ un sous-espace vectoriel de $\mathbb{R}^n$.
On note $W^\perp$ l'ensemble des vecteurs de $\mathbb{R}^n$ qui sont orthogonaux à *tous* les vecteurs de $W$. On appelle $W^\perp$ le **complément orthogonal de $W$** (ou "l'orthogonal de $W$")

**Exemple** Par exemple, le complément orthogonal à une droite dans $\mathbb{R}^2$, c'est une droite.
De la même manière, le complètement orthogonal à un plan dans $\mathbb{R}^3$, c'est aussi une droite.
On peut voir ces exemples graphiquement :




**Théorème** $W^\perp$ est un sous-espace vectoriel de $\mathbb{R}^n$.

*Preuve* On doit vérifier les trois propriétés suivantes :
1. $\vec{0} \in W^\perp$
2. $\vec{u}, \vec{v} \in W^\perp \implies \vec{u} + \vec{v} \in W^\perp$
3. $\vec{u} \in W^\perp, c$ un scalaire $\implies c\vec{u} \in W^\perp$

Pour rappel, même si la première propriété découle de la troisième, on en a quand même besoin pour s'assurer que l'ensemble n'est pas vide.

On peut démontrer ces propriétés de la manière suivante :

1. Soit $\vec{w} \in W$ quelconque.
   On remarque que $\vec{0} \bullet \vec{w} = \vec{0}$, donc $\vec{0}$ est bien orthogonal $\vec{w}$, et donc à tous les vecteurs de $W$.
2. Soient $\vec{u}, \vec{v} \in W^\perp$, et soit $\vec{w} \in W$ quelconque. On a :
   $$\left(\vec{u} + \vec{v}\right) \bullet \vec{w} = \vec{u} \bullet \vec{w} + \vec{v} \bullet \vec{w} = 0 + 0 = 0$$
   puisque $\vec{u}$ et $\vec{v}$ appartiennent à $W^\perp$.
   Ainsi, on en déduit que $\vec{u} + \vec{v}$ est orthogonal à $\vec{w}$, donc $\vec{u} + \vec{v} \in W^\perp$
3. Soit $\vec{u} \in W^\perp$ et $c$ un scalaire quelconque. Donc :
   $$\left(c\vec{u}\right) \bullet \vec{w} = c\vec{u} \bullet \vec{w} = 0$$
   On en déduit que $c\vec{u}$ est orthogonal à $\vec{w}$ quelconque dans $W$, donc que $c\vec{u} \in W^\perp$.

□

**Théorème**

$$\left(W^\perp\right)^\perp = W$$

*Preuve* Laissée en exercice au lecteur.

Indice : pour montrer que deux ensembles $A$ et $B$ sont égaux, il est souvent plus simple de montrer que $A \subset B$ et $B \subset A$.

**Théorème** Si $W = \text{vect}\left\{\vec{w}_1, \ldots, \vec{w}_p\right\}$, alors $\vec{u} \in W^\perp$ si et seulement si :

$$\vec{u} \bullet \vec{w}_i = 0, \quad \forall i = 1, \ldots, p$$

*Preuve* Si $\vec{u} \in W^\perp$, alors en particulier $\vec{u} \bullet \vec{w}_i = 0$ pour tout $i$. Donc, ce sens de l'implication est tenu.

Supposons maintenant que $\vec{u}$ est orthogonal à $\vec{w}_1, \ldots, \vec{w}_p$. On veut montrer que $\vec{u}$ est orthogonal à tous les vecteurs de $W$. Soit $\vec{w} \in W$ quelconque . Alors :

$$\vec{w} = c_1\vec{w}_1 + \ldots + c_p\vec{w}_p$$

Calculons leur produit scalaire :

$$\begin{aligned}
\vec{w} \bullet \vec{u} &= \left(c_1\vec{w}_1 + \ldots + c_p\vec{w}_p\right) \bullet \vec{u} \\
&= c_1\vec{w}_1 \bullet \vec{u} + \ldots + c_p\vec{w}_p \bullet \vec{u} \\
&= c_1 \cdot 0 + \ldots + c_p \cdot 0 \\
&= 0
\end{aligned}$$

Donc $\vec{u} \in W^\perp$.

□

Lundi 29 novembre 2021 — **Cours 20 : Kernel, image, famille et espace orthogonal**

**Observation** Soient les trois vecteurs suivants :

$$\vec{v}_1 = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 9 \\ 10 \\ 8 \end{bmatrix}$$

Alors, on voit que :

$$\begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2^T \end{bmatrix} \vec{x} = \begin{bmatrix} 7 & 3 & 1 \\ 2 & 4 & 5 \end{bmatrix} \begin{bmatrix} 9 \\ 10 \\ 8 \end{bmatrix} = \begin{bmatrix} 7 \cdot 9 + 3 \cdot 10 + 1 \cdot 8 \\ 2 \cdot 9 + 4 \cdot 10 + 5 \cdot 8 \end{bmatrix} = \begin{bmatrix} \vec{v}_1 \bullet \vec{x} \\ \vec{v}_2 \bullet \vec{x} \end{bmatrix}$$

Ceci peut être démontré pour un cas complètement général.

**Retour aux quatre sous-espaces** Pour une matrice $A \in \mathbb{R}^{m \times n}$, on a quatre sous-espaces :

| **Sous espaces de $\mathbb{R}^n$** | **Sous espaces de $\mathbb{R}^m$** |
|---|---|
| $\ker(A)$ | $\operatorname{Im}(A)$ |
| $\operatorname{Im}(A^T) = \operatorname{Lgn}(A)$ | $\ker(A^T)$ |

On avait déjà vu que :

$$\dim(\ker A) + \dim(\operatorname{Im} A^T) = \dim \mathbb{R}^n$$

$$\dim(\operatorname{Im} A) + \dim(\ker A^T) = \dim \mathbb{R}^m$$

**Théorème** Soit une matrice $A$ quelconque. On a :

$$\left(\operatorname{Im}(A^T)\right)^\perp = \ker A$$

*Preuve de $\supseteq$* On veut montrer que $\ker A \subseteq \left(\operatorname{Im}(A^T)\right)^\perp$.

Soit $\vec{x}$ un vecteur quelconque dans $\ker A$, donc :

$$A\vec{x} = \vec{0}.$$

Soit $\vec{w}$ un vecteur quelconque dans $\operatorname{Im}(A^T)$, donc :

$$\exists \vec{c} \text{ tel que } \vec{w} = A^T \vec{c}$$

On doit vérifier que $\vec{x}$ et $\vec{w}$ sont orthogonaux :

$$\vec{w} \bullet \vec{x} = \vec{w}^T \vec{x} = \left(A^T \vec{c}\right)^T \vec{x} = \vec{c}^T A \vec{x} = \vec{c}^T \vec{0} = \vec{0}$$

Ainsi, on en déduit bien que $\vec{x}$ est orthogonal à $\vec{w}$, donc que $\vec{x} \in \left(\operatorname{Im} A^T\right)^\perp$.

*Preuve de $\subseteq$* On veut montrer que $\left(\operatorname{Im}(A^T)\right)^\perp \subseteq \ker A$.

Soit $\vec{x}$ un vecteur quelconque de $\left(\operatorname{Im}(A^T)\right)^\perp$, donc $\vec{x}$ est orthogonal à tous les vecteurs de $\operatorname{Im}(A^T)$. En particulier, $\vec{x}$ est orthogonal à chaque colonne de $A^T$, c'est-à-dire à chaque ligne de $A$.
Comme ce qu'on a vu plus haut, la $i$-ème entrée de $A\vec{x}$ est le produit scalaire de la $i$-ème ligne de $A$ avec $\vec{x}$, ce qui est à chaque fois nul puisque $\vec{x}$ est orthogonal à tous les vecteurs de $\operatorname{Im}(A^T)$.
On en déduit donc que $A\vec{x} = \vec{0}$.
Ainsi, $(\operatorname{Im} A)^\perp \subseteq \ker A$.

□

**Corollaire**

Soit $A$ une matrice. On a :

- $(\ker A)^{\perp} = \operatorname{Im}(A^T)$
- $(\operatorname{Im} A)^{\perp} = \ker(A^T)$
- $(\ker A^T)^{\perp} = \operatorname{Im}(A)$
- $(\operatorname{Im} A^T)^{\perp} = \ker(A)$

On remarque que pour chaque égalité on a échangé ker et Im, on a échangé $A$ et $A^T$, et on a échangé $V$ et $V^{\perp}$.

*Intuition*

Toutes les personnes à qui j'ai demandé et tous les liens que j'ai consulté sur internet sortaient simplement la démonstration du théorème ci-dessus comme intuition. Il va donc falloir s'en contenter.

*Preuve*

Puisqu'on sait que $\ker A = \left(\operatorname{Im}(A^T)\right)^{\perp}$, on a :

$$(\ker A)^{\perp} = \left(\operatorname{Im}(A^T)^{\perp}\right)^{\perp} = \operatorname{Im}(A^T)$$

De plus, en appliquant ces relations à $A^T$ au lieu de $A$ (puisque le théorème ci-dessus est vrai pour toutes les matrices, il l'est aussi pour $A^T$), on a :

$$\ker(A^T) = \left(\operatorname{Im}\left(\left(A^T\right)^T\right)\right)^{\perp} = (\operatorname{Im} A)^{\perp}$$

$$\left(\ker A^T\right)^{\perp} = \left((\operatorname{Im} A)^{\perp}\right)^{\perp} = \operatorname{Im} A$$

**Théorème d'Al Kashi**

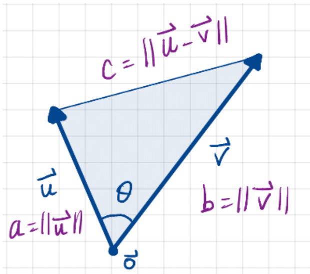
On sait que, dans n'importe quel triangle :

$$a^2 + b^2 - 2ab\cos(\gamma) = c^2$$

où $\gamma$ est l'angle opposé au côté $c$ (donc l'angle entre le côté $a$ et le côté $b$).
Ce théorème est appelé le théorème d'Al Kashi, le théorème des cosinus ou, par certains hérétiques, le théorème de Pythagore généralisé.

*Implication*

De manière générale, deux vecteur dessinent un triangle :




Ainsi, on trouve par Al Kashi que :

$$\begin{aligned}
&\left\|\overrightarrow{u}\right\|^2 + \left\|\overrightarrow{v}\right\|^2 - 2\left\|\overrightarrow{u}\right\|\left\|\overrightarrow{v}\right\|\cos\theta \\
&= \left\|\overrightarrow{u} - \overrightarrow{v}\right\|^2 \\
&= \left(\overrightarrow{u} - \overrightarrow{v}\right) \bullet \left(\overrightarrow{u} - \overrightarrow{v}\right) \\
&= \overrightarrow{u} \bullet \overrightarrow{u} - \overrightarrow{u} \bullet \overrightarrow{v} - \overrightarrow{v} \bullet \overrightarrow{u} + \overrightarrow{v} \bullet \overrightarrow{v} \\
&= \left\|\overrightarrow{u}\right\|^2 + \left\|\overrightarrow{v}\right\|^2 - 2\overrightarrow{u} \bullet \overrightarrow{v}
\end{aligned}$$

On peut donc en déduire que :

$$\overrightarrow{u} \bullet \overrightarrow{v} = \left\|\overrightarrow{u}\right\|\left\|\overrightarrow{v}\right\|\cos\theta$$

Ainsi, on remarque bien que deux vecteurs non-nuls sont orthogonaux si et seulement si le cosinus de l'angle entre les deux est nul, donc si et seulement s'ils sont perpendiculaires.

## 8.2 Familles orthogonales

**Définition** Des vecteurs $\left(\vec{u}_1, \ldots, \vec{u}_p\right)$ de $\mathbb{R}^n$ forment une **famille orthogonale** si chaque vecteur est orthogonal à tous les autres, c'est-à-dire si :

$$\vec{u}_i \bullet \vec{u}_j = 0, \quad \forall i \forall j, i \neq j$$

*Remarque* Comme $\vec{0} \bullet \vec{u} = \vec{0}$ pour tout $\vec{u}$, on peut très bien avoir des vecteurs nuls dans une famille orthogonale.

Cependant, on est souvent plus intéressé par des familles de vecteurs non-nuls à cause du théorème qui va suivre après ces exemples.

**Exemple 1** Soient les vecteurs de $\mathbb{R}^2$ suivants :

$$\vec{u}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$$

La famille $\left(\vec{u}_1, \vec{u}_2\right)$ est orthogonale car :

$$\vec{u}_1 \bullet \vec{u}_2 = 2 \cdot 2 + 1(-4) = 4 - 4 = 0$$

**Exemple 2** Soient les trois vecteurs de $\mathbb{R}^3$ suivants :

$$\vec{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

La famille $\left(\vec{u}_1, \vec{u}_2, \vec{u}_3\right)$ est orthogonale car :

$$\vec{u}_1 \bullet \vec{u}_2 = 3(-1) + 1(2) + 1(1) = 0$$

$$\vec{u}_1 \bullet \vec{u}_3 = 3\left(\frac{-1}{2}\right) + 1(-2) + 1\left(\frac{7}{2}\right) = 0$$

$$\vec{u}_2 \bullet \vec{u}_3 = -1\left(\frac{-1}{2}\right) + 2(-2) + 1\left(\frac{7}{2}\right) = 0$$

**Théorème** Si les vecteurs $\left(\vec{u}_1, \ldots, \vec{u}_p\right)$ sont *tous* non-nuls et forment une famille orthogonale, alors ils sont linéairement indépendants.

*Preuve* Considérons des coefficients $c_1, \ldots, c_p$ quelconques tels que :

$$\vec{0} = c_1 \vec{u}_1 + \ldots + c_p \vec{u}_p$$

On doit montrer qu'ils sont tous nuls. Commençons par montrer que $c_1 = 0$. On remarque que :

$$\begin{aligned} 0 &= \vec{0} \bullet \vec{u}_1 \\ &= \left(c_1 \vec{u}_1 + \ldots + c_p \vec{u}_p\right) \bullet \vec{u}_1 \\ &= c_1\left(\vec{u}_1 \bullet \vec{u}_1\right) + c_2\left(\underbrace{\vec{u}_2 \bullet \vec{u}_1}_{=0}\right) + \ldots + c_p\left(\underbrace{\vec{u}_p \bullet \vec{u}_1}_{=0}\right) \\ &= c_1\left\|\vec{u}_1\right\|^2 \end{aligned}$$

Puisque $\vec{u}_1$ est non-nul, $\|\vec{u}_1\|^2 \neq 0$. Ainsi, nécessairement, $c_1 = 0$. On peut répéter le même raisonnement pour montrer que $c_2 = 0, \ldots, c_p = 0$. Donc, la seule combinaison linéaire de $\vec{u}_1, \ldots, \vec{u}_p$ qui est égale à $\vec{0}$ est la combinaison triviale, ce qui implique que les vecteurs sont libres.

□

**Définition** Soit $(\vec{u}_1, \ldots, \vec{u}_p)$ une base d'un sous-espace $W$ de $\mathbb{R}^n$. Si cette famille est orthogonale, on dit qu'elle forme une **base orthogonale** de $W$.

*Utilité* Les bases orthogonales sont très pratiques. En effet, une des raisons est qu'il est très facile de trouver les coordonnées dans ce type de base.

**Théorème** Soit $(\vec{u}_1, \ldots, \vec{u}_p)$ une base orthogonale d'un sous-espace vectoriel $W$ de $\mathbb{R}^n$. Pour tout $\vec{y}$ de $W$, les coefficients de la combinaison linéaire $\vec{y} = c_1\vec{u}_1 + \ldots + c_p\vec{u}_p$ sont données par la relation :

$$c_j = \frac{\vec{y} \bullet \vec{u}_j}{\vec{u}_j \bullet \vec{u}_j}, \quad j = 1, \ldots, p$$

*Mnémotechnie* Notez qu'il n'est pas nécessaire d'apprendre cette formule par cœur, il est plus simple de comprendre la démonstration qui suit et de redériver cette formule à chaque fois (même le professeur la redérive à chaque fois, on voit bien qu'on est pas des étudiants en biologie vu la quantité d'appris par cœur).

*Implication* On peut donc facilement trouver $[\vec{y}]_{\mathcal{B}}$ :

$$[\vec{y}]_{\mathcal{B}} = \begin{bmatrix} \frac{\vec{y} \bullet \vec{u}_1}{\|\vec{u}_1\|^2} \\ \vdots \\ \frac{\vec{y} \bullet \vec{u}_n}{\|\vec{u}_n\|^2} \end{bmatrix}$$

*Preuve* Soit $\vec{y} \in W$ quelconque.
Comme $\mathcal{B} = (\vec{u}_1, \ldots, \vec{u}_p)$ est une base de $W$, on sait qu'il existe un choix unique de coefficients $c_1, \ldots, c_p$ tels que :

$$\vec{y} = c_1\vec{u}_1 + \ldots + c_p\vec{u}_p$$

Calculons $c_2$, par exemple :

$$\begin{aligned} \vec{y} \bullet \vec{u}_2 &= (c_1\vec{u}_1 + \ldots + c_p\vec{u}_p) \bullet \vec{u}_2 \\ &= c_1(\vec{u}_1 \bullet \vec{u}_2) + c_2(\vec{u}_2 \bullet \vec{u}_2) + \ldots + c_p(\vec{u}_p \bullet \vec{u}_2) \\ &= c_2\vec{u}_2 \bullet \vec{u}_2 \end{aligned}$$

Donc, on trouve :

$$c_2 = \frac{\vec{y} \bullet \vec{u}_2}{\vec{u}_2 \bullet \vec{u}_2}$$

Ce qui est bien défini puisque $\vec{u}_2 \neq \vec{0}$.
On peut faire le même raisonnement pour n'importe quel vecteur de la base.

□

**Exemple** Soient les trois vecteurs de $\mathbb{R}^3$ suivants :

$$\overrightarrow{\boldsymbol{u}}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \overrightarrow{\boldsymbol{u}}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \overrightarrow{\boldsymbol{u}}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

On a déjà vérifié que la famille $\left(\overrightarrow{\boldsymbol{u}}_1, \overrightarrow{\boldsymbol{u}}_2, \overrightarrow{\boldsymbol{u}}_3\right)$ est orthogonale. Comme les vecteurs sont tous non-nuls, on sait qu'ils forment une base de pour $\operatorname{vect}\left\{\overrightarrow{\boldsymbol{u}}_1, \overrightarrow{\boldsymbol{u}}_2, \overrightarrow{\boldsymbol{u}}_3\right\} = \mathbb{R}^3$. Prenons le vecteur suivant de $\mathbb{R}^3$ :

$$\overrightarrow{\boldsymbol{y}} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

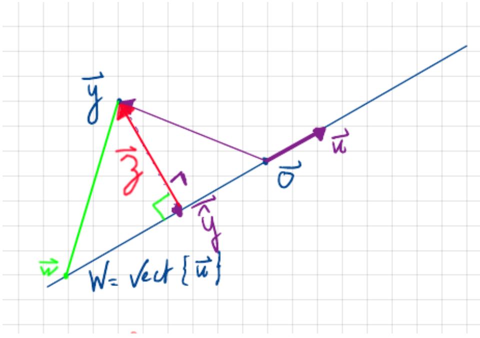
On peut trouver ses coordonnées dans la base $\mathcal{B} = \left(\overrightarrow{\boldsymbol{u}}_1, \overrightarrow{\boldsymbol{u}}_2, \overrightarrow{\boldsymbol{u}}_3\right)$ :

$$c_1 = \frac{\overrightarrow{\boldsymbol{y}} \bullet \overrightarrow{\boldsymbol{u}}_1}{\overrightarrow{\boldsymbol{u}}_1 \bullet \overrightarrow{\boldsymbol{u}}_1} = \frac{1(3) + 1(1) + 2(1)}{3(3) + 1(1) + 1(1)} = \frac{6}{11}$$

On peut faire le même raisonnement pour les autres coordonnées.

## 8.3 Projections orthogonales

**En 2D** Considérons une droite qui passe par l'origine $W$ (c'est un sous-espace vectoriel) et un point $\vec{y}$ dans $\mathbb{R}^2$. On se demande quel est le point $\vec{\hat{y}}$ sur $W$ qui est le plus proche de $\vec{y}$.




On se rend compte que le point le plus proche est celui projeté orthogonalement sur $W$. En effet, peu importe le point $\vec{w} \neq \vec{\hat{y}} \in W$ qu'on prend cela dessine un triangle rectangle entre $\vec{w}$, $\vec{y}$ et $\vec{\hat{y}}$, donc la distance entre $\vec{w}$ et $\vec{y}$ (l'hypothénuse du triangle rectangle) est forcément plus grande que celle entre $\vec{y}$ et $\vec{\hat{y}}$ (une des cathètes de ce triangle).
On prend $\vec{z}$ tel que $\vec{y} = \vec{\hat{y}} + \vec{z}$. Puisque $\vec{\hat{y}}$ est sur la droite, on voit que $\vec{\hat{y}} = c\vec{u}$. De plus, $\vec{z} \bullet \vec{u} = 0$ puisqu'ils sont orthogonaux. Ainsi, on en déduit que :

$$0 = \vec{z} \bullet \vec{u} = \left(\vec{y} - \vec{\hat{y}}\right) \bullet \vec{u} = (\vec{y} - c\vec{u}) \bullet \vec{u} = \vec{y} \bullet \vec{u} - c\vec{u} \bullet \vec{u}$$

Ainsi, on en déduit que :

$$\vec{y} \bullet \vec{u} = c\vec{u} \bullet \vec{u} \implies c = \frac{\vec{y} \bullet \vec{u}}{\vec{u} \bullet \vec{u}}$$

Ce qui nous permet de trouver la formule pour $\vec{\hat{y}}$ :

$$\vec{\hat{y}} = \frac{\vec{y} \bullet \vec{u}}{\vec{u} \bullet \vec{u}}\vec{u}$$

Jeudi 2 décembre 2021 — **Cours 21 : Projections et matrices orthogonales**

**En 3D** Nous nous posons la même question avec un plan $W = \text{vect}\{\vec{u}_1, \vec{u}_2\}$ et un point $\vec{y} \in \mathbb{R}^3$.
Puisqu'on veut $\vec{\hat{y}} \in W$, alors on peut écrire $\vec{\hat{y}} = c_1\vec{u}_1 + c_2\vec{u}_2 \in W$. Prenons aussi $\vec{z} = \vec{y} - \vec{\hat{y}}$. Par le même raisonnement qu'en 2D, $\vec{z}$ doit être orthogonal à $W$. Or, un vecteur est orthogonal à un espace vectoriel si et seulement s'il est orthogonal à tous les vecteurs d'une base qui engendre l'espace (ici $\vec{u}_1$ et $\vec{u}_2$) :

$$\vec{z} \bullet \vec{u}_1 = 0, \quad \vec{z} \bullet \vec{u}_2 = 0$$

Développons la première équation :

$$0 = \vec{z} \bullet \vec{u}_1 = \left(\vec{y} - \vec{\hat{y}}\right) \bullet \vec{u}_1 = \vec{y} \bullet \vec{u}_1 - \vec{\hat{y}} \bullet \vec{u}_1 = \vec{y} \bullet \vec{u}_1 - c_1\vec{u}_1 \bullet \vec{u}_1 - c_2\vec{u}_2 \bullet \vec{u}_1$$

On voit que c'est une équation linéaire en $c_1$ et $c_2$ :

$$c_1\vec{u}_1 \bullet \vec{u}_1 + c_2\vec{u}_2 \bullet \vec{u}_1 = \vec{y} \bullet \vec{u}_1$$

En faisant le même raisonnement avec la deuxième équation, on obtient :

$$c_1\vec{u}_1 \bullet \vec{u}_2 + c_2\vec{u}_2 \bullet \vec{u}_2 = \vec{y} \bullet \vec{u}_2$$

On a donc un système de deux équations à deux inconnues, qu'on peut mettre en forme matricielle :

$$\begin{bmatrix} \vec{u}_1 \bullet \vec{u}_1 & \vec{u}_2 \bullet \vec{u}_1 \\ \vec{u}_1 \bullet \vec{u}_2 & \vec{u}_2 \bullet \vec{u}_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \vec{y} \bullet \vec{u}_1 \\ \vec{y} \bullet \vec{u}_2 \end{bmatrix}$$

On se rend compte qu'il est plus facile de résoudre ce système si on a une base orthogonale, puisque ça nous donnerait une matrice diagonale.

**Théorème de projection orthogonale** Soit $W$, un sous-espace vectoriel de $\mathbb{R}^n$.
Tout vecteur $\vec{y}$ de $\mathbb{R}^n$ s'écrit de façon *unique* sous la forme :

$$\vec{y} = \vec{\hat{y}} + \vec{z}, \quad \vec{\hat{y}} \in W, \vec{z} \in W^{\perp}$$

De façon plus précise, si $\left(\vec{u}_1 \ldots, \vec{u}_p\right)$ est une base orthogonale quelconque de $W$, alors :

$$\vec{\hat{y}} = \frac{\vec{y} \bullet \vec{u}_1}{\vec{u}_1 \bullet \vec{u}_1}\vec{u}_1 + \ldots + \frac{\vec{y} \bullet \vec{u}_p}{\vec{u}_p \bullet \vec{u}_p}\vec{u}_p, \quad \vec{z} = \vec{y} - \vec{\hat{y}}$$

*Terminologie* On appelle $\vec{\hat{y}}$ la **projection orthogonale** de $\vec{y}$ sur $W$, notée $\text{proj}_W \vec{y}$.

*Preuve* Soit $\vec{y} \in \mathbb{R}^n$ quelconque. On cherche à écrire $\vec{y} = \vec{\hat{y}} + \vec{z}$ avec $\vec{\hat{y}} \in W$ et $\vec{z} \in W^{\perp}$, donc :

$$\vec{\hat{y}} = c_1\vec{u}_1 + \ldots + c_p\vec{u}_p, \quad \vec{u}_1 \bullet \vec{z} = \ldots = \vec{u}_p \bullet \vec{z} = 0$$

Prenons le produit scalaire entre $\vec{u_i}$, un vecteur de la base, et $\vec{z}$ :

$$0 = \vec{u}_i \bullet \vec{z} = \vec{u}_i \bullet \left(\vec{y} - \vec{\hat{y}}\right) = \vec{u}_i \bullet \vec{y} - \vec{u}_i \bullet \vec{\hat{y}}$$

On peut remplacer $\hat{\vec{y}}$ par sa forme :

$$0 = \vec{u}_i \bullet \vec{y} - \vec{u}_i \bullet \left(c_1\vec{u}_1 + \ldots + c_p\vec{u}_p\right) = \vec{u}_i \bullet \vec{y} - c_i\vec{u}_i \bullet \vec{u}_i$$

En effet, la base est orthogonale, donc le produit scalaire entre deux vecteurs de la base différents donne zéro. Ainsi, on peut résoudre cette équation :

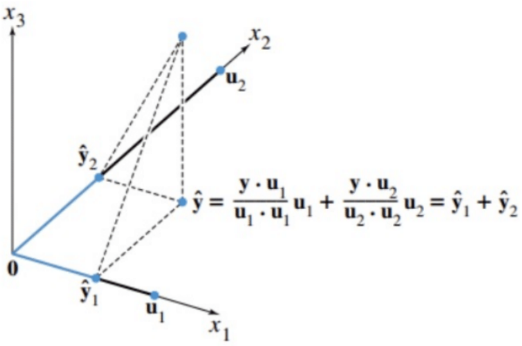
$$c_1\vec{u}_i \bullet \vec{u}_i = \vec{u}_i \bullet \vec{y} \implies c_i = \frac{\vec{u}_i \bullet \vec{y}}{\vec{u}_i \bullet \vec{u}_i}$$

Donc, non seulement il est possible de trouver de coefficients $c_i$, mais aussi ils sont uniques puisqu'on n'a jamais dû faire de choix dans notre raisonnement.
Ainsi, on peut écrire :

$$\vec{\hat{y}} = \frac{\vec{y} \bullet \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1}\vec{u}_1 + \ldots + \frac{\vec{y} \bullet \vec{u}_p}{\vec{u}_p \bullet \vec{u}_p}\vec{u}_p$$

□

*Interprétation géométrique* On fait la projection sur les lignes données par les vecteurs de la base, puis on en fait la combinaison linéaire :




**Théorème de meilleure approximation (quadratique)**

Soit $W$ un sous-espace vectoriel de $\mathbb{R}^n$, $\vec{y} \in \mathbb{R}^n$ et $\vec{\hat{y}} = \text{proj}_W \vec{y}$.
Alors, $\vec{\hat{y}}$ est le point de $W$ le plus proche de $\vec{y}$, c'est-à-dire :

$$\left\|\vec{y} - \vec{\hat{y}}\right\| < \|\vec{y} - \vec{v}\|, \quad \forall \vec{v} \in W, \vec{v} \neq \vec{\hat{y}}$$

*Preuve*

Soit $\vec{v}$ un vecteur quelconque de $W$. Regardons la distance entre $\vec{y}$ et $\vec{v}$ :

$$\|\vec{y} - \vec{v}\|^2 = \left\|\vec{y} - \vec{\hat{y}} + \vec{\hat{y}} - \vec{v}\right\|^2$$

Or, on sait que $\vec{y} - \vec{\hat{y}} = \vec{z} \in W^\perp$, et $\vec{\hat{y}} - \vec{v} \in W$ puisque $\vec{\hat{y}}$ et $\vec{v}$ sont dans $W$. Ainsi, les deux différences de vecteurs sont orthogonales, on peut donc utiliser Pythagore :

$$\|\vec{y} - \vec{v}\|^2 = \left\|\vec{y} - \vec{\hat{y}}\right\|^2 + \underbrace{\left\|\vec{\hat{y}} - \vec{v}\right\|^2}_{>0} > \left\|\vec{y} - \vec{\hat{y}}\right\|^2$$

puisque $\vec{\hat{y}} \neq \vec{v}$.

□

**Observations**

On voit que les cas suivants sont un peu spéciaux :
- Si $\vec{y} \in W$, alors $\text{proj}_W \vec{y} = \vec{y}$.
- Si $\vec{y} \in W^\perp$, alors $\text{proj}_W \vec{y} = \vec{0}$.

De plus, on peut voir que, si on connait la projection dans $W$, alors on peut facilement calculer la projection dans $W^\perp$ :

$$\text{proj}_{W^\perp} \vec{y} = \vec{y} - \text{proj}_W \vec{y}$$

Ainsi, pour projeter sur un plan dans $\mathbb{R}^3$, on peut projeter sur la droite orthogonale, puis calculer cette différence.

**Définition**

On dit qu'une base $\mathcal{B} = \left(\vec{u}_1, \ldots, \vec{u}_p\right)$ d'un sous-espace $W$ est **orthonormé** si :

$$\vec{u}_i \bullet \vec{u}_j = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

*Utilité*

Dans une base orthonormée, la formule de projection est très simple :

$$\text{proj}_W \vec{y} = (\vec{y} \bullet \vec{u}_1)\vec{u}_1 + \ldots + (\vec{y} \bullet \vec{u}_p)\vec{u}_p$$

*Exemple*

$\mathbb{R}^n$ est une base orthonormée.

**Observation** Soit $\mathcal{U} = \left\{\overrightarrow{u}_1, \overrightarrow{u}_2, \overrightarrow{u}_3\right\}$ une base d'un sous-espace de $\mathbb{R}^4$.
On peut construire la matrice suivante :

$$U = \begin{bmatrix} \overrightarrow{u}_1 & \overrightarrow{u}_2 & \overrightarrow{u}_3 \end{bmatrix} \in \mathbb{R}^{4\times 3}$$

Trouvons $U^T$ :

$$U^T = \begin{bmatrix} \overrightarrow{u}_1^T \\ \overrightarrow{u}_2^T \\ \overrightarrow{u}_3^T \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

Par definition du produit matriciel, chaque entrée de $U^T U$ est le produit entre une ligne de $U^T$ et une colonne de $U$ :

$$U^T U = \begin{bmatrix} \overrightarrow{u}_1^T \overrightarrow{u}_1 & \overrightarrow{u}_1^T \overrightarrow{u}_2 & \overrightarrow{u}_1^T \overrightarrow{u}_3 \\ \overrightarrow{u}_2^T \overrightarrow{u}_1 & \overrightarrow{u}_2^T \overrightarrow{u}_2 & \overrightarrow{u}_2^T \overrightarrow{u}_3 \\ \overrightarrow{u}_3^T \overrightarrow{u}_1 & \overrightarrow{u}_3^T \overrightarrow{u}_2 & \overrightarrow{u}_3^T \overrightarrow{u}_3 \end{bmatrix} = \begin{bmatrix} \overrightarrow{u}_1 \bullet \overrightarrow{u}_1 & \overrightarrow{u}_1 \bullet \overrightarrow{u}_2 & \overrightarrow{u}_1 \bullet \overrightarrow{u}_3 \\ \overrightarrow{u}_2 \bullet \overrightarrow{u}_1 & \overrightarrow{u}_2 \bullet \overrightarrow{u}_2 & \overrightarrow{u}_2 \bullet \overrightarrow{u}_3 \\ \overrightarrow{u}_3 \bullet \overrightarrow{u}_1 & \overrightarrow{u}_3 \bullet \overrightarrow{u}_2 & \overrightarrow{u}_3 \bullet \overrightarrow{u}_3 \end{bmatrix}$$

Dons, si $\mathcal{U}$ est orthogonale, alors $U^T U$ est diagonale. Si $\mathcal{U}$ est orthonormée, alors $U^T U = I$.

**Théorème** Les colonnes d'une matrice $U \in \mathbb{R}^{m\times n}$ sont orthonormées *si et seulement si*

$$U^T U = I$$

**Exemple** Soient les trois vecteurs de $\mathbb{R}^4$ suivants :

$$\overrightarrow{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \overrightarrow{u}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 2 \end{bmatrix}, \quad \overrightarrow{u}_3 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ -1/2 \end{bmatrix}$$

Pour un vecteur $\overrightarrow{y} \in \text{vect}\left\{\overrightarrow{u}_1, \overrightarrow{u}_2, \overrightarrow{u}_3\right\}$, on cherche les coordonnées $\left[\overrightarrow{y}\right]_{\mathcal{U}}$ telles que :

$$\overrightarrow{y} = c_1 \overrightarrow{u}_1 + c_2 \overrightarrow{u}_3 + c_3 \overrightarrow{u}_3 = \begin{bmatrix} \overrightarrow{u}_1 & \overrightarrow{u}_2 & \overrightarrow{u}_3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

Donc, on veut résoudre le système $U\left[\overrightarrow{y}\right]_{\mathcal{U}} = \overrightarrow{y}$ où :

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 2 & -1/2 \end{bmatrix}$$

On remarque que $\mathcal{U}$ est une base orthogonale :

$$U^T U = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & -1 & 2 \\ 1 & -1 & 0 & -1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 2 & -1/2 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 9/4 \end{bmatrix}$$

Donc :

$$U\left[\overrightarrow{y}\right]_{\mathcal{U}} = \overrightarrow{y} \implies U^T U \left[\overrightarrow{y}\right]_{\mathcal{U}} = U^T \overrightarrow{y}$$

Ainsi, on a que :

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 9/4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & -1 & 2 \\ 1 & -1 & 0 & -1/2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \overrightarrow{u}_1 \bullet \overrightarrow{y} \\ \overrightarrow{u}_2 \bullet \overrightarrow{y} \\ \overrightarrow{u}_3 \bullet \overrightarrow{y} \end{bmatrix}$$

Donc on peut en déduire que :

$$\begin{cases} c_1 = \dfrac{1}{3}\vec{u}_1 \bullet \vec{y} & (8.2a) \\ c_2 = \dfrac{1}{6}\vec{u}_2 \bullet \vec{y} & (8.2b) \\ c_3 = \dfrac{4}{9}\vec{u}_3 \bullet \vec{y} & (8.2c) \end{cases}$$

**Théorème** Soit $U$ une matrice $m \times n$ dont les colonnes sont orthonormées, et $\vec{x}$ et $\vec{y}$ deux vecteurs de $\mathbb{R}^n$. Alors :

1. $\|U\vec{x}\| = \|\vec{x}\|$
2. $(U\vec{x}) \bullet (U\vec{y}) = \vec{x} \bullet \vec{y}$
3. $(U\vec{x}) \bullet (U\vec{y}) = 0$ si et seulement si $\vec{x} \bullet \vec{y} = 0$

*Preuve* La preuve est laissée en exercice au lecteur.
Voici une idée de comment démontrer le premier point :

$$\|U\vec{x}\|^2 = (U\vec{x}) \bullet (U\vec{x}) = (U\vec{x})^T (U\vec{x}) = \vec{x}^T U^T U \vec{x}$$

Et donc, puisque $U^T U = I$ :

$$\|U\vec{x}\|^2 = \vec{x}^T\vec{x} = \vec{x} \bullet \vec{x} = \|\vec{x}\|^2$$

Ainsi :

$$\|U\vec{x}\| = \|\vec{x}\|$$

**Théorème** Soient $(\vec{u}_1, \ldots, \vec{u}_p)$ une base orthonormée d'un sous espace vectoriel $W$ de $\mathbb{R}^n$, et $\vec{y} \in \mathbb{R}^n$.
La projection de $\vec{y}$ sur $W$ est donnée par :

$$\text{proj}_W \vec{y} = (y \bullet \vec{u}_1)\vec{u}_1 + \ldots + (\vec{y} \bullet \vec{u}_p)\vec{u}_p$$

De plus :

$$\text{proj}_W \vec{y} = UU^T\vec{y}, \quad \text{où } U = \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_p \end{bmatrix}$$

*Preuve* On a déjà démontré que pour n'importe quelle base orthogonale :

$$\text{proj}_W \vec{y} = \frac{(y \bullet \vec{u}_1)}{\vec{u}_1 \bullet \vec{u}_1}\vec{u}_1 + \ldots + \frac{(\vec{y} \bullet \vec{u}_p)}{\vec{u}_p \bullet \vec{u}_p}\vec{u}_p$$

Ainsi, puisque notre base est en plus orthonormée, alors :

$$\vec{u}_1 \bullet \vec{u}_1 = \ldots = \vec{u}_p \bullet \vec{u}_p = 1$$

On peut donc simplifier notre résultat :

$$\text{proj}_W \vec{y} = (y \bullet \vec{u}_1)\vec{u}_1 + \ldots + (\vec{y} \bullet \vec{u}_p)\vec{u}_p$$

De plus, on se rend compte qu'on peut écrire ça sous forme matricielle :

$$\begin{aligned} UU^T\vec{y} &= \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_p \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_p^T \end{bmatrix} \vec{y} \\ &= \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_p \end{bmatrix} \begin{bmatrix} \vec{u}_1^T\vec{y} \\ \vdots \\ \vec{u}_p^T\vec{y} \end{bmatrix} \\ &= \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_p \end{bmatrix} \begin{bmatrix} \vec{u}_1 \bullet \vec{y} \\ \vdots \\ \vec{u}_p \bullet \vec{y} \end{bmatrix} \\ &= \left(\vec{u}_1 \bullet \vec{y}\right)\vec{u}_1 + \ldots + \left(\vec{u}_p \bullet \vec{y}\right)\vec{u}_p \end{aligned}$$

□

*Observation* On peut voir que, de manière générale, $U^TU = I$, mais $UU^T \neq I$.

*Remarque* Puisque $\text{proj}_W \vec{y} = UU^T\vec{y}$ est sous la forme $T(\vec{x}) = A\vec{x}$, on en déduit que $\text{proj}_W$ est une application linéaire. On peut aussi démontrer ce résultat pour n'importe quel base orthogonale.

De plus, puisque la projection ne dépend pas de la base, $UU^T$ sera toujours le même pour un sous-ensemble, peu importe la base.

**Définition** Si $U \in \mathbb{R}^{n\times n}$ est *carrée* et $U^TU = I_n$ (c'est-à-dire que ses colonnes sont orthonormées), alors on dit que la matrice $U$ est **orthogonale**.

*Remarque* Dans ce cas là, $U^{-1} = U^T$, donc :

$$U^TU = UU^T = I_n$$

*Terminologie* Il faut faire attention au fait que les colonnes sont orthonormées, mais qu'on dit que la matrice est orthogonale.

Lundi 6 décembre 2021 — **Cours 22 : La la la Schtroumpf la la**

## 8.4 Orthonormalisation

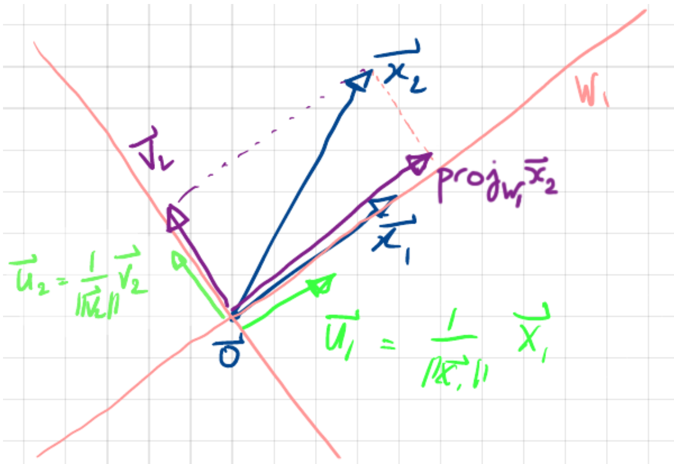
**Introduction** Soit $W = \text{vect}\left\{\vec{x}_1, \vec{x}_2\right\}$, où :

$$\vec{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

Pour calculer des projections orthogonales sur $W$, ce serait pratique d'avoir une base orthonormée. La base $\mathcal{B} = \left(\vec{x}_1, \vec{x}_2\right)$ n'est même pas orthogonale :

$$\vec{x}_1 \bullet \vec{x}_1 = 45, \quad \vec{x}_1 \bullet \vec{x}_2 = 15, \quad \vec{x}_2 \bullet \vec{x}_2 = 9$$

On va donc **orthonormaliser** cette base.




Pour commencer, normalisons $\vec{x}_1$ :

$$\|\vec{x}_1\|^2 = 3^2 + 6^2 + 0^2 = 45 \implies \vec{u}_1 = \frac{1}{\|\vec{x}_1\|}\vec{x}_1 = \frac{1}{3\sqrt{5}}\begin{bmatrix}3\\6\\0\end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix}1\\2\\0\end{bmatrix}$$

Maintenant, construisons $\vec{v}_2$. On projette $\vec{x}_2$ sur la droite donnée par $W_1$, donc la droite engendrée par $\vec{u}_1$, puis on va soustraire ce résultat à $\vec{x}_2$ :

$$\vec{v}_2 = \vec{x}_2 - \text{proj}_{W_1}(\vec{x}_2) = \vec{x}_2 - (\vec{x}_2 \bullet \vec{u}_1)\vec{u}_1 = \begin{bmatrix}1\\2\\2\end{bmatrix} - \frac{5}{\sqrt{5}} \cdot \frac{1}{\sqrt{5}}\begin{bmatrix}1\\2\\0\end{bmatrix} = \begin{bmatrix}0\\0\\2\end{bmatrix}$$

Finalement, nous pouvons normaliser $\vec{v}_2$ pour obtenir le deuxième vecteur de notre base orthonormée :

$$\vec{u}_2 = \frac{1}{\|\vec{v}\|}\vec{v} = \frac{1}{2}\begin{bmatrix}0\\0\\2\end{bmatrix} = \begin{bmatrix}0\\0\\1\end{bmatrix}$$

Ainsi, par construction, $\mathcal{U} = (\vec{u}_1, \vec{u}_2)$ est une base orthonormée, et $\text{vect}\{\vec{u}_1, \vec{u}_2\} = \text{vect}\{\vec{x}_1, \vec{x}_2\}$

*Observation 1* On remarque que :

$$\vec{x}_1 = \|\vec{x}_1\|\vec{u}_1$$

De plus, en manipulant un petit peu les équations ci-dessus :

$$\vec{x}_2 = (\vec{x}_2 \bullet \vec{u}_1)\vec{u}_1 + \vec{v}_2 = (\vec{x}_2 \bullet \vec{u}_1)\vec{u}_1 + \|\vec{v}_2\|\vec{u}_2$$

Ainsi, on peut écrire ceci sous la forme d'un produit matriciel :

$$\begin{bmatrix}\vec{x}_1 & \vec{x}_2\end{bmatrix} = \begin{bmatrix}\vec{u}_1 & \vec{u}_2\end{bmatrix}\begin{bmatrix}\|\vec{x}_1\| & \vec{x}_2 \bullet \vec{u}_1\\ 0 & \|\vec{v}_2\|\end{bmatrix} \implies A = QR$$

où $A$ est la matrice dont les colonnes forment une base quelconque de $W$, $Q$ est une matrice dont les colonnes forment une base orthonormée de $W$, et $R$ est une matrice triangulaire supérieure. On se rendra compte qu'on pourra toujours factoriser $A$ sous cette forme, si les colonnes de $A$ sont linéairement indépendants.

*Observation 2* Non seulement $(\vec{u}_1, \vec{u}_2)$ est une base orthonormée de $\text{vect}\{\vec{x}_1, \vec{x}_2\}$, mais aussi $(\vec{u}_1)$ est une base orthonormée de $\text{vect}\{\vec{x}_1\}$.

> *Remarque* Cet algorithme s'appelle formellement l'algorithme de Gram-Schmidt. Cependant, il est beaucoup plus intéressant de lui donner des noms amicaux. Un très bon ami, un *Génie en la Physique*, l'appelle "l'algorithme du Grand Schmit", ou "l'algorithme du Grand Schtroumpf".

**Avec 3 vecteurs** Soient les trois vecteurs suivants :

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

On peut se rendre compte que ces trois vecteurs sont linéairement indépendants, donc ils forment une base pour $W = \text{vect}\left\{\vec{x}_1, \vec{x}_2, \vec{x}_3\right\}$. Comme la dernière fois, on veut construire une base orthonormée $\mathcal{U} = \left(\vec{u}_1, \vec{u}_2, \vec{u}_3\right)$.
Pour commencer, calculons $\vec{u}_1$ en normalisant $\vec{x}_1$ :

$$\left\|\vec{x}_1\right\|^2 = 4 \implies \vec{u}_1 = \frac{1}{\left\|\vec{x}_1\right\|}\vec{x}_1 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

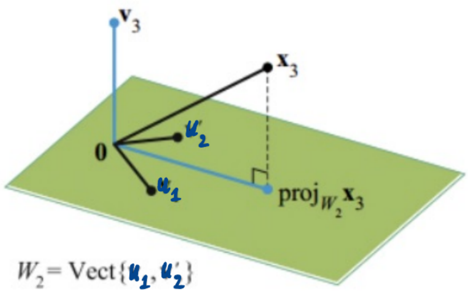
Maintenant, pour calculer $\vec{v}_2$, on retranche la projection de $\vec{x}_2$ sur $W_1 = \text{vect}\left\{\vec{u}_1\right\}$ à $\vec{x}_2$ :

$$\vec{v}_2 = \vec{x}_2 - \text{proj}_{W_1}\left(\vec{x}_2\right) = \vec{x}_2 - \left(\vec{x}_2 \bullet \vec{u}_1\right)\vec{u}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{2}\cdot\frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Ainsi, en normalisant $\vec{v_2}$, on obtient $\vec{u}_1$ :

$$\vec{u}_2 = \frac{1}{\vec{v}_2}\vec{v}_2 = \frac{2}{\sqrt{3}}\frac{1}{4}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{2\sqrt{3}}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

On remarque que, jusque là, le fait qu'on ait 3 vecteurs ne change rien : on n'a jamais utilisé le troisième vecteur.




On va faire le même raisonnement que pour les dernières fois : on va projeter $\vec{x}_3$ sur $W_2 = \text{vect}\left\{\vec{u}_1, \vec{u}_2\right\}$, puis retrancher ce résultat à $\vec{x}_3$ pour obtenir $\vec{v_3}$. La projection est calculable facilement, puisque $\left(\vec{u}_1, \vec{u}_2\right)$ est une base orthogonale.

$$\vec{v}_3 = \vec{x}_3 - \text{proj}_{W_2}\left(\vec{x}_3\right) = \vec{x}_3 - \left(\vec{x_3} \bullet \vec{u}_1\right)\vec{u}_1 - \left(\vec{x}_3 \bullet \vec{u}_2\right)\vec{u}_2$$

Ainsi, on obtient

$$\vec{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - 1\cdot\frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{\sqrt{3}}\frac{1}{2\sqrt{3}}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} - \begin{bmatrix} -1/6 \\ 1/6 \\ 1/6 \\ 1/6 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}$$

Finalement, on peut normaliser $\vec{v}_3$ pour obtenir $\vec{u}_3$ :

$$\vec{u}_3 = \frac{1}{\|\vec{v}_3\|}\vec{v}_3 = \frac{\sqrt{3}}{\sqrt{2}} \cdot \frac{1}{3}\begin{bmatrix}0\\-2\\1\\1\end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix}0\\-2\\1\\1\end{bmatrix}$$

On a donc construit la base orthonormée $U = \left(\vec{u}_1, \vec{u}_2, \vec{u}_3\right)$ pour $W = \text{vect}\left\{\vec{x}_1, \vec{x}_2, \vec{x}_3\right\}$, où :

$$\vec{u}_1 = \frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \quad \vec{u}_2 = \frac{1}{2\sqrt{3}}\begin{bmatrix}-3\\1\\1\\1\end{bmatrix}, \quad \vec{u}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix}0\\-2\\1\\1\end{bmatrix}$$

*Observation* On remarque que :

- $\left\{\vec{u}_1\right\}$ est une base orthonormée pour $W = \text{vect}\left\{\vec{x}_1\right\}$.
- $\left\{\vec{u}_1, \vec{u}_2\right\}$ est une base orthonormée pour $W = \text{vect}\left\{\vec{x}_1, \vec{x}_2\right\}$.
- $\left\{\vec{u}_1, \vec{u}_2, \vec{u}_3\right\}$ est une base orthonormée pour $W = \text{vect}\left\{\vec{x}_1, \vec{x}_2, \vec{x}_3\right\}$.

De plus, on voit aussi que :

$$\vec{x}_1 = \|\vec{x}_1\|\vec{u}_1$$

$$\vec{x}_2 = \left(\vec{x}_2 \bullet \vec{u}_1\right)\vec{u}_1 + \vec{v}_2 = \left(\vec{x}_2 \bullet \vec{u}_1\right)\vec{u}_1 + \|\vec{v}_2\|\vec{u}_2$$

$$\begin{aligned}\vec{x}_3 &= \left(\vec{x}_3 \bullet \vec{u}_1\right)\vec{u}_1 + \left(\vec{x}_3 \bullet \vec{u}_2\right)\vec{u}_2 + \vec{v}_3\\ &= \left(\vec{x}_3 \bullet \vec{u}_1\right)\vec{u}_1 + \left(\vec{x}_3 \bullet \vec{u}_3\right)\vec{u}_2 + \|\vec{v}_3\|\vec{u}_3\end{aligned}$$

On peut donc à nouveau écrire $A = QR$ :

$$\begin{bmatrix}\vec{x}_1 & \vec{x}_2 & \vec{x}_3\end{bmatrix} = \begin{bmatrix}\vec{u}_1 & \vec{u}_2 & \vec{u}_3\end{bmatrix}\begin{bmatrix}\|\vec{x}_1\| & \vec{x}_2 \bullet \vec{u}_1 & \vec{x}_3 \bullet \vec{u}_1\\ 0 & \|\vec{v}_2\| & \vec{x}_3 \bullet \vec{u}_2\\ 0 & 0 & \|\vec{v}_3\|\end{bmatrix}$$

Ainsi, puisque les colonnes de $Q$ forment une base orthonormée pour $\text{Im}\, A$, on a :

$$Q^T Q = I_3$$

$$\text{proj}_{\text{Im}\, A}\left(\vec{y}\right) = QQ^T\vec{y}$$

Le plus simple pour trouver la matrice qui projette sur l'espace image de $A$ est donc d'appliquer le Grand Schtroumpf (Gram-Schmidt), de factoriser $A$ sous la forme $QR$, puis de calculer $QQ^T$

**Dépendance linéaire** Nous pouvons nous rendre compte que, si nous appliquons le Grand Schtroumpf sur des vecteurs qui sont linéairement dépendants, alors on aurait obtenu le vecteur $\vec{0}$ à un endroit. Cependant, puisqu'en suite nous devons diviser ce vecteur par sa norme, ce qui est une division par 0, l'algorithme casse bel et bien.
En effet, la projection d'un vecteur dans un espace auquel il appartient est égal à ce vecteur. Ainsi, quand on va retrancher ce résultat à ce vecteur, alors cela va nous donner $\vec{0}$.
Ceci est cohérent puisque notre but est d'orthonormaliser une *base*, ainsi si les vecteurs n'étaient pas linéairement indépendants, alors ce n'était pas une base au départ.

**Gram-Schmidt et factorisation QR** Soit $A$ une matrice de taille $m \times n$ dont les colonnes sont linéairement indépendants. En d'autres mots, $\left(\vec{x}_1, \ldots, \vec{x}_n\right)$ est une base de $W = \text{vect}\left\{\vec{x}_1, \ldots, \vec{x}_n\right\} = \text{Im}\, A$. **L'algorithme de Gram-Schmidt** (Grand Schtroumpf) revient à calculer :

$$\vec{v}_n = \vec{x}_n - \text{proj}_{W_{n-1}}\left(\vec{x}_n\right) = \vec{x}_n - \left(\vec{x}_n \bullet \vec{u}_1\right)\vec{u}_1 - \ldots - \left(\vec{x}_n \bullet \vec{u}_{n-1}\right)\vec{u}_{n-1}$$

$$\vec{u}_n = \frac{1}{\left\|\vec{v}_n\right\|}\vec{v}_n$$

où $W_{n-1} = \text{vect}\left\{\vec{x}_1, \ldots, \vec{x}_{n-1}\right\}$.
Ces calculs nous donnent $\left(\vec{u}_1, \ldots, \vec{u}_n\right)$, une **base orthonormée** de $W = \operatorname{Im} A$.
De plus, ils révèlent la factorisation $QR$ de $A$ :

$$A = QR, \quad Q = \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_n \end{bmatrix}, \quad R = \begin{bmatrix} \left\|\vec{x}_1\right\| & \vec{x}_2 \bullet \vec{u}_1 & \ldots & \vec{x}_n \bullet \vec{u}_1 \\ 0 & \left\|\vec{v}_2\right\| & \ldots & \vec{x}_n \bullet \vec{u}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \left\|\vec{v}_n\right\| \end{bmatrix}$$

**Matrice carrée**

Étudions la factorisation $QR$ d'une matrice $A \in \mathbb{R}^{n \times n}$ carrée.
Appliquons le Grand Schtroumpf à ses colonnes. Il y a deux cas possibles. Si cet algorithme échoue (il y a une division par zéro), alors $A$ n'est pas inversible. S'il aboutit, alors on trouve $Q \in \mathbb{R}^{n\times n}$ et $R \in \mathbb{R}^{n \times n}$ telles que :

$$A = QR, \quad Q^T Q = QQ^T = I_n \implies Q^{-1} = Q^T$$

et où $R$ est triangulaire est inversible (elle n'a que des nombres strictement positifs sur sa diagonale, donc son déterminant n'est pas nul).
On peut utiliser ce résultat pour résoudre $A\vec{x} = \vec{b}$. En effet, on remarque que :

$$A\vec{x} = \vec{b} \iff QR\vec{x} = \vec{b} \iff Q^T QR\vec{x} = Q^T\vec{b} \iff R\vec{x} = Q^T\vec{b}$$

Or, $Q^T\vec{b}$ est facilement calculable, et, puisque $R$ est triangulaire, il est très facile de résoudre ce système.
De plus, on peut aussi utiliser ce résultat pour calculer (et mieux comprendre) $\det(A)$ :

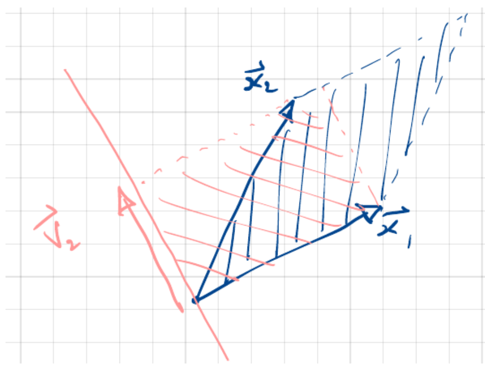
$$1 = \det(I) = \det\left(Q^T Q\right) = \det\left(Q^T\right)\det(Q) = \det(Q)^2 \implies \det(Q) = \pm 1$$

Ainsi, on peut utiliser ceci pour calculer $\det(A)$ :

$$\det(A) = \det(QR) = \det(Q)\det(R) = \pm\det(R) \implies \left|\det(A)\right| = \left|\det(R)\right|$$

Cependant, on sait que $R$ est une matrice triangulaire supérieure, ainsi il est très simple de calculer son déterminant : c'est le produit de ses entrées diagonales (mais les calculs pour y arriver rendent cette méthode plus compliquée que celles que nous avions vues jusque là). Ainsi, on en déduit que :

$$\left|\det(A)\right| = \left\|\vec{x}_1\right\|\left\|\vec{v}_2\right\|\left\|\vec{v}_3\right\|\cdots\left\|\vec{v}_n\right\|$$




On peut voir que l'aire du parallélogramme défini par $\vec{x}_1, \vec{x}_n$ est égale à l'aire du rectangle définie par $\vec{x}_1$ et $\vec{v}_2$, qui vaut $\left\|\vec{x}_1\right\| \cdot \left\|\vec{v}_2\right\| = \det(R)$.
Ceci est vrai pour tout $n$, donc $\left|\det(A)\right|$ est l'hyper-volume de l'hyper-parallélépipède de dimension $n$

Jeudi 9 décembre 2021 — **Cours 23 : Moindres carrés et début des régressions**

## 8.5 Moindres carrés

**Introduction** Maintenant que nous avons vu plus de notions, nous pouvons revenir à notre motivation de départ : on souhaite trouver $\overrightarrow{x} \in \mathbb{R}^n$ tel que $A\overrightarrow{x} = \overrightarrow{b}$. Cependant, si les colonnes de $A$ sont liées, ou si, par exemple on a $m > n$ (plus d'équations que d'inconnues), il se pourrait qu'aucune solution n'existe. Ainsi, nous voulons trouver une solution qui est la plus proche possible.

On cherche donc $\overrightarrow{x} \in \mathbb{R}^n$ tel que $A\overrightarrow{x}$ est aussi proche que possible de $\overrightarrow{b} \in \mathbb{R}^m$. En considérant tous les $\overrightarrow{x}$ dans $\mathbb{R}^n$, on peut "atteindre" n'importe quel point de $\operatorname{Im}(A)$ (mais aucun autre). Ainsi, puisque $A\overrightarrow{x} = \overrightarrow{b}$ n'a pas de solution, cela implique que $\overrightarrow{b}$ n'est pas dans l'image de $A$. Or, on sait que l'unique point de $\operatorname{Im}(A)$ le plus proche de $\overrightarrow{b}$ est la projection orthogonale de $\overrightarrow{b}$ sur $\operatorname{Im}(A)$. Ainsi, on va tenter de choisir $\overrightarrow{\hat{x}}$ tel que :

$$A\overrightarrow{\hat{x}} = \operatorname{proj}_{\operatorname{Im}(A)} \overrightarrow{b} = \overrightarrow{\hat{b}}$$

Comme $\overrightarrow{\hat{b}}$ est dans $\operatorname{Im} A$, il existe $\overrightarrow{\hat{x}} \in \mathbb{R}^n$ (qui n'est pas forcément unique) tel que $A\overrightarrow{\hat{x}} = \overrightarrow{\hat{b}}$. Si on trouve un tel $\overrightarrow{\hat{x}}$, on a gagné. En effet, pour tout $x \in \mathbb{R}^n$ :

$$\left\|A\overrightarrow{x} - \overrightarrow{b}\right\| \geq \left\|\overrightarrow{\hat{b}} - \overrightarrow{b}\right\| = \left\|A\overrightarrow{\hat{x}} - \overrightarrow{b}\right\|$$

En d'autres mots, aucun $\overrightarrow{x}$ ne fait strictement mieux que $\overrightarrow{\hat{x}}$.

**Méthode** Pour trouver $\overrightarrow{\hat{x}}$ tel que $A\overrightarrow{\hat{x}}$ est aussi proche que possible de $\overrightarrow{b}$, on peut adopter différents points de vue, tous équivalents. En voici un :

$$\begin{aligned}
&\overrightarrow{\hat{b}} = \operatorname{proj}_{\operatorname{Im} A}\left(\overrightarrow{b}\right) = A\overrightarrow{x} \\
\implies &\overrightarrow{\hat{b}} - \overrightarrow{b} \text{ est orthogonal à } \operatorname{Im} A \\
\implies &\overrightarrow{\hat{b}} - \overrightarrow{b} \in (\operatorname{Im} A)^{\perp} = \ker\left(A^T\right) \\
\implies &A^T\left(\overrightarrow{\hat{b}} - \overrightarrow{b}\right) = \overrightarrow{0} \\
\implies &A^T\left(A\overrightarrow{\hat{x}} - \overrightarrow{b}\right) = \overrightarrow{0} \\
\implies &A^T A\overrightarrow{\hat{x}} - A^T\overrightarrow{b} = \overrightarrow{0} \\
\implies &A^T A\overrightarrow{\hat{x}} = A^T\overrightarrow{b}
\end{aligned}$$

En partant de notre équation de départ, on multiplie donc simplement les deux côtés par $A^T$. Ce nouveau système a, nécessairement, au moins une solution.

**Théorème** Étant donné un système d'équation $A\overrightarrow{x} = \overrightarrow{b}$ (qui pourrait ne pas avoir de solution), on définit un autre système d'équations, appelées **équations normales** :

$$A^T A\overrightarrow{\hat{x}} = A^T\overrightarrow{b}$$

Alors, ce système a au moins une solution, et chaque solution $\overrightarrow{\hat{x}}$ a la propriété que :

$$\left\|A\overrightarrow{\hat{x}} - \overrightarrow{b}\right\| \leq \left\|A\overrightarrow{x} - \overrightarrow{b}\right\|, \quad \forall\overrightarrow{x}$$

*Observation* On peut voir que $\left\| A \hat{\vec{x}} - \vec{b} \right\|$ est l'erreur à la solution réellement attendue, donc cette erreur est minimale.

*Remarque* On appelle chaque $\hat{\vec{x}}$ une **solution au moindres carrés** de $A\vec{x} = \vec{b}$ car $\hat{\vec{x}}$ minimise (rend aussi petit que possible) :

$$\begin{aligned}\left\| A \hat{\vec{x}} \right\|^2 &= \left(A\hat{\vec{x}} - \vec{b}\right) \bullet \left(A\hat{\vec{x}} - \vec{b}\right) \\ &= \left(A\hat{\vec{x}} - \vec{b}\right)_1^2 + \ldots + \left(A\hat{\vec{x}} - \vec{b}\right)_m^2\end{aligned}$$

On minimise donc le carré des erreurs individuelles, ce qui justifie le nom.

**Théorème : unicité** Soit $A \in \mathbb{R}^{m \times n}$.
Les affirmations sont équivalentes :
1. La solution aux moindres carrés de $A\vec{x} = \vec{b}$ est unique pour tout $\vec{b}$.
2. $A^T A \in \mathbb{R}^{n \times n}$ est inversible.
3. Les colonnes de $A$ sont linéairement indépendantes.

Dans le cas où elles sont vraies (donc qu'au moins une est vraie), alors :

$$\hat{\vec{x}} = \left(A^T A\right)^{-1} A^T \vec{b}$$

*Preuve* On va démontrer que (1) $\implies$ (3) $\implies$ (2) $\implies$ (1). Dans ce cas, on aura bien démontré que si une des proposition est vraie, alors elles le seront toutes.

- (1) $\implies$ (3) est vrai car on peut considérer la contraposée (si (3) est faux, alors (1) est faux) : On sait que les colonnes de $A$ sont linéairement dépendantes, donc qu'il existe $\vec{z} \neq \vec{0}$ tel que $A\vec{z} = \vec{0}$. Ainsi, en prenant $\hat{\vec{x}}$, une solution aux moindres carrés :

$$A^T A\left(\hat{\vec{x}} + \vec{z}\right) = A^T A \hat{\vec{x}} + A^T A \vec{z} = A^T A \hat{\vec{x}} + \vec{0}$$

  Donc, $\hat{\vec{x}} + \vec{z} \neq \hat{\vec{x}}$ est une autre solution, ce qui montre bien qu'elles ne sont pas uniques.
- Démontrons aussi (3) $\implies$ (2) en passant par la contraposée. Puisque $A^T A$ n'est pas inversible, alors il existe $\vec{z} \neq \vec{0}$ tel que $A^T A \vec{z} = \vec{0}$. Ainsi :

$$A\vec{z} \in \ker\left(A^T\right) = (\operatorname{Im} A)^\perp$$

  Or, puisque $A\vec{z} \in \operatorname{Im} A$, $A\vec{z}$ appartient à la fois à un espace vectoriel et son orthogonal. On en déduit que $A\vec{z} = \vec{0}$, et donc que les colonnes de $A$ sont linéairement dépendantes.
- (2) $\implies$ (1) est clair vu que les équations normales sont sous la forme $A^T A \hat{\vec{x}} = A^T \vec{b}$. Si la matrice est inversible, alors la solution est clairement unique.

□

**Exemple** Soient la matrice et le vecteur suivants :

$$A = \begin{bmatrix} 3 & 1 \\ 6 & 2 \\ 0 & 2 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 11 \\ 17 \\ 6 \end{bmatrix}$$

On veut trouver les solutions au moindres carrés, donc on doit résoudre les équations normales :

$$A^T A \vec{\boldsymbol{x}} = A^T \vec{\boldsymbol{b}} \implies \begin{bmatrix} 45 & 15 \\ 15 & 9 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 135 \\ 57 \end{bmatrix}$$

Notez que $A^T A$ est symétrique, et sera toujours symétrique. Si $A^T A$ n'est pas symétrique dans nos calculs c'est que nous avons fait une erreur !
On peut résoudre notre système en échelonnant. On trouve :

$$\vec{\hat{\boldsymbol{x}}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

De plus, on voit que :

$$A\vec{\hat{\boldsymbol{x}}} = \begin{bmatrix} 3 & 1 \\ 6 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 18 \\ 6 \end{bmatrix}$$

qui est la projection de $\vec{\boldsymbol{b}}$ sur $\operatorname{Im} A$.
L'erreur (inévitable) est donnée par :

$$A\vec{\hat{\boldsymbol{x}}} - \vec{\boldsymbol{b}} = \begin{bmatrix} 9 \\ 18 \\ 6 \end{bmatrix} - \begin{bmatrix} 11 \\ 17 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \implies \left\| A\vec{\hat{\boldsymbol{x}}} - \vec{\boldsymbol{b}} \right\|^2 = 5$$

et tout autre $\vec{\boldsymbol{x}}$ ferait pire.

**Méthode via factorisation $QR$**

Soit $A \in \mathbb{R}^{m\times n}$. Supposons quel les colonnes de $A$ sont libres. Alors, on sait par la méthode du Grand Schtroumpf qu'il existe $Q \in \mathbb{R}^{m\times n}$ et $R \in \mathbb{R}^{n\times n}$ telles que $A = QR$, où $Q^T Q = I_n$ et $R$ est triangulaire et inversible.
De plus, par notre dernier théorème ci-dessus, on sait que $\vec{\boldsymbol{b}} \in \mathbb{R}^m$, la solution aux moindres carrés de $A\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}$, est l'unique solution de :

$$A^T A\vec{\hat{\boldsymbol{x}}} = A^T \vec{\boldsymbol{b}}$$

On peut donc écrire :

$$A^T = R^T Q^T \implies A^T A = R^T Q^T Q R = R^T I R = R^T R$$

Ainsi, notre système d'équations peut être écrit sous la forme :

$$A^T A\vec{\hat{\boldsymbol{x}}} = A^T \vec{\boldsymbol{b}} \implies R^T R\vec{\hat{\boldsymbol{x}}} = R^T Q^T \vec{\boldsymbol{b}} \implies R\vec{\hat{\boldsymbol{x}}} = Q^T \vec{\boldsymbol{b}}$$

puisque $R$ est inversible, et donc $R^T$ est aussi inversible.

**Théorème**

Si les colonnes de $A \in \mathbb{R}^{m\times n}$ sont linéairement indépendantes, alors pour tout $\vec{\boldsymbol{b}} \in \mathbb{R}^m$, la solution aux moindres carrés de $A\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}$ est la solution de :

$$R\vec{\hat{\boldsymbol{x}}} = Q^T \vec{\boldsymbol{b}}$$

où $A = QR$ est une factorisation $QR$ : $Q^T Q = I_n$ et $R$ est une matrice inversible et triangulaire.

*Observation* Ce système a bien une solution unique car $R$ est inversible. De plus, il est facile à résoudre (une fois qu'on a la factorisation $QR$) car $R$ est triangulaire.

*Remarque* À la main, cet algorithme est plus compliqué que celui qui passe par les équations normales. Cependant, pour un ordinateur, cet algorithme est bien meilleur, notamment car cela implique moins de *floating point errors* (problèmes d'arrondis dû au fait que les nombres ne peuvent pas avoir une infinité de décimales dans les ordinateurs).

**Exemple** Soient la matrice et le vecteurs suivants :

$$A = \begin{bmatrix} 3 & 1 \\ 6 & 2 \\ 0 & 2 \end{bmatrix}, \quad \overrightarrow{\boldsymbol{b}} = \begin{bmatrix} 11 \\ 17 \\ 6 \end{bmatrix}$$

On a déjà calculé la factorisation $QR$ de $A$ :

$$\begin{bmatrix} 3 & 1 \\ 6 & 2 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} & 0 \\ 2/\sqrt{5} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3\sqrt{5} & \sqrt{5} \\ 0 & 2 \end{bmatrix}$$

La solution aux moindres carrés est données par $R\overrightarrow{\hat{\boldsymbol{x}}} = Q^T \overrightarrow{\boldsymbol{b}}$ :

$$\begin{bmatrix} 3\sqrt{5} & \sqrt{5} \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 9\sqrt{5} \\ 6 \end{bmatrix} \implies \overrightarrow{\hat{\boldsymbol{x}}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Ce système est simple à résoudre puisqu'il est triangulaire.

**Observation** On voit que, si $A = QR$, alors la solution aux moindres carrés de $A\overrightarrow{\boldsymbol{x}} = \overrightarrow{\boldsymbol{b}}$ est :

$$R\overrightarrow{\hat{\boldsymbol{x}}} = Q^T \overrightarrow{\boldsymbol{b}}$$

Ceci est cohérent car :

$$A\overrightarrow{\hat{\boldsymbol{x}}} = QR\overrightarrow{\hat{\boldsymbol{x}}} = QQ^T \overrightarrow{\boldsymbol{b}} = \text{proj}_{\text{Im}\, Q} \overrightarrow{\boldsymbol{b}} = \text{proj}_{\text{Im}\, A} \overrightarrow{\boldsymbol{b}}$$

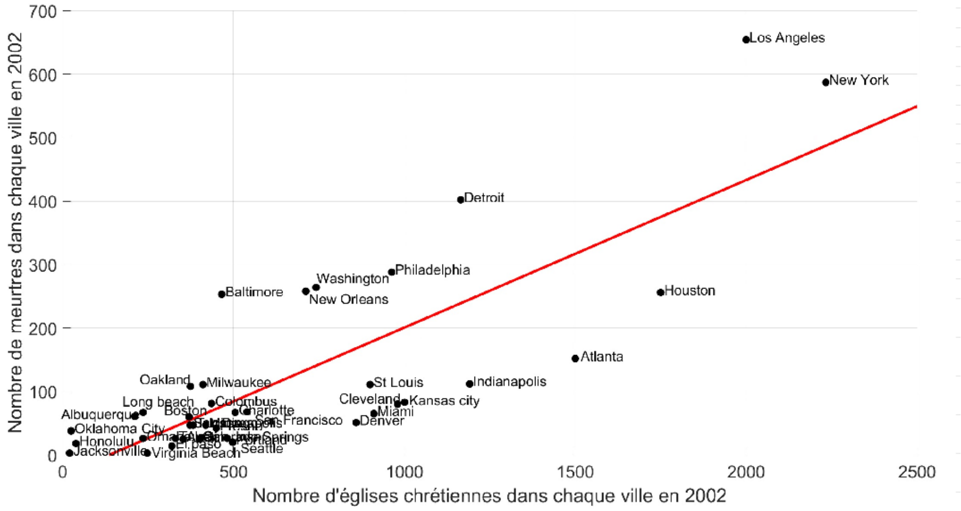
comme attendu.

## 8.6 Régression linéaire

**Introduction** Étant données des paires de scalaires $(x_i, y_i)$, on cherche la relation :

$$y_i \approx \beta_1 x_i + \beta_0$$

En d'autres mots, on cherche la meilleure droite qui passe par ces points.
Ce serait très utile pour l'analyse de données scientifique (comme par exemple pour calculer l'augmentation de la température de la planète au fil du temps).
Comme d'habitude, quand on analyse des données il faut faire attention au fait que corrélation n'implique pas forcément causalité. En effet, par exemple on peut voir la statistique suivante :




On en déduit donc que :

$$[\#\text{meurtres}] \approx 0.23 \cdot [\#\text{églises}] - 31.8$$

Cependant, cette statistique ne fait aucun sens. En effet, la variable cachée derrière est la taille de la population dans la ville ; détruire des églises ne va pas diminuer le nombre de meurtres.

**Méthode** Mathématiquement, on cherche une droite d'équation :

$$y = \beta_1 x + \beta_0$$

où $\beta_1, \beta_0 \in \mathbb{R}$ sont nos inconnues.
On a des points $(x_1, y_1), \ldots, (x_n, y_n)$, et on veut que :

$$\beta_1 x_i + \beta_0 \approx y_i, \quad \forall i = 1, \ldots, n$$

On veut minimiser l'erreur. Pour commencer, écrivons tout ceci sous forme matricielle :

$$\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \approx \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \implies A\overrightarrow{\boldsymbol{\beta}} \approx \overrightarrow{\boldsymbol{y}}$$

On cherche donc $\overrightarrow{\boldsymbol{\beta}} \in \mathbb{R}^2$ tel que $A\overrightarrow{\boldsymbol{\beta}}$ soit aussi proche que possible de $\overrightarrow{\boldsymbol{y}}$.

---

Lundi 13 décembre 2021 — **Cours 24 : Diagonalisation en base orthonormée**

**Méthode** Utilisons notre approche de moindres carrées. Les solutions $\overrightarrow{\hat{\boldsymbol{\beta}}}$ de $A\overrightarrow{\boldsymbol{\beta}} = \overrightarrow{\boldsymbol{y}}$ au sens des moindres carrés sont les solutions des équations normales :

$$A^T A \overrightarrow{\hat{\boldsymbol{\beta}}} = A^T \overrightarrow{\boldsymbol{y}}$$

On sait que la solution $\overrightarrow{\hat{\boldsymbol{\beta}}}$ existe toujours. La solution est unique, si et seulement si les colonnes de $A$ sont linéairement indépendants, donc si et seulement si les $x_i$ ne sont pas tous égaux (si et seulement si au moins deux sont différents).

**Solution explicite** Notre cas est tellement simple que nous pouvons chercher une solution explicite. Cependant, nous n'avons pas besoin de connaître le résultat par cœur.

Commençons par définir le point moyen :

$$\overline{x} = \frac{1}{n}\sum_{i} x_i, \quad \overline{y} = \frac{1}{n}\sum_{i} y_i$$

Cherchons maintenant les équations normales :

$$A^T A = \begin{bmatrix} 1 & \ldots & 1 \\ x_1 & \ldots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\overline{x} \\ n\overline{x} & \left\|\overrightarrow{\boldsymbol{x}}\right\|^2 \end{bmatrix}$$

$$A^T \overrightarrow{\boldsymbol{y}} = \begin{bmatrix} 1 & \ldots & 1 \\ x_1 & \ldots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} = \begin{bmatrix} n\overline{y} \\ \overrightarrow{\boldsymbol{x}} \bullet \overrightarrow{\boldsymbol{y}} \end{bmatrix}$$

On remarque que si on avait préalablement centré nos données, donc si on avait pris $\overline{x} = \overline{y} = 0$, cela aurait simplifié nos calculs.
Supposons que notre matrice est inversible. Alors :

$$\overrightarrow{\boldsymbol{b}} = \left(A^T A\right)^{-1} A^T \overrightarrow{\boldsymbol{y}} = \begin{bmatrix} n & n\overline{x} \\ n\overline{x} & \left\|\overrightarrow{\boldsymbol{x}}\right\|^2 \end{bmatrix}^{-1} \begin{bmatrix} n\overline{y} \\ \overrightarrow{\boldsymbol{x}} \bullet \overrightarrow{\boldsymbol{y}} \end{bmatrix}$$

On peut utiliser la formule pour inverser les matrices $2 \times 2$ :

$$\frac{1}{n\left\|\overrightarrow{\boldsymbol{x}}\right\|^2 - n^2\overline{x}^2} \begin{bmatrix} \left\|\overrightarrow{\boldsymbol{x}}\right\|^2 & -n\overline{x} \\ -n\overline{x} & n \end{bmatrix} \begin{bmatrix} n\overline{y} \\ \overrightarrow{\boldsymbol{x}} \bullet \overrightarrow{\boldsymbol{y}} \end{bmatrix} = \frac{1}{n\left\|\overrightarrow{\boldsymbol{x}}\right\|^2 - n^2\overline{x}^2} \begin{bmatrix} \left\|\overrightarrow{\boldsymbol{x}}\right\|^2 n\overline{y} - n\overline{x}\left(\overrightarrow{\boldsymbol{x}} \bullet \overrightarrow{\boldsymbol{y}}\right) \\ -n^2\overline{x}\overline{y} + n\left(\overrightarrow{\boldsymbol{x}} \bullet \overrightarrow{\boldsymbol{y}}\right) \end{bmatrix}$$

On en déduit donc que :

$$\hat{\beta}_1 = \frac{\overrightarrow{\boldsymbol{x}} \bullet \overrightarrow{\boldsymbol{y}} - n\overline{x}\overline{y}}{\left\|\overrightarrow{\boldsymbol{x}}\right\|^2 - n\overline{x}^2}$$

Notez qu'il ne faut surtout pas apprendre cette solution par cœur, que nous pouvons juste résoudre les équations normales. Ceci était juste pour nous montrer qu'on a développé assez d'outils pour obtenir un résultat, qui est de plus relativement simple.

**Justification du moindre carré**

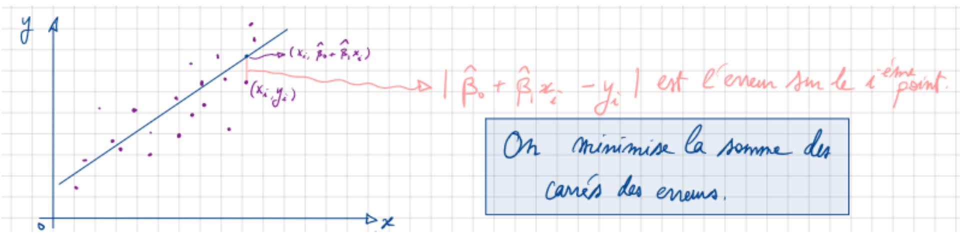
Ces formules nous permettent de calculer $\hat{b}_0, \hat{b}_1$ tels que $\left\|A\overrightarrow{\hat{\boldsymbol{\beta}}} - \overrightarrow{\boldsymbol{y}}\right\|^2$ est aussi petit que possible. Nous voulons étudier la quantité qu'on est *réellement* en train de minimiser :

$$A\overrightarrow{\hat{\boldsymbol{b}}} - \overrightarrow{\boldsymbol{y}} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 + \hat{\beta}_1 x_1 - y_1 \\ \ldots \\ \hat{\beta}_0 + \hat{\beta}_1 x_n - y_n \end{bmatrix}$$

Ainsi :

$$\left\|A\overrightarrow{\hat{\boldsymbol{\beta}}} - \overrightarrow{\boldsymbol{y}}\right\|^2 = \left(\hat{\beta}_0 + \hat{\beta}_1 x_1 - y_1\right)^2 + \ldots + \left(\hat{\beta}_0 + \hat{\beta}_1 x_n - y_n\right)^2$$

On minimise donc la somme des carrés des erreurs verticales.



# Chapitre 9

# Matrices symétriques

## 9.1 Diagonalisation des matrices symétriques

**Rappels** Une matrice $A$ est diagonalisable s'il existe une matrice $P$ inversible et une matrice $D$ diagonale telles que $A = PDP^{-1}$. Une matrice est diagonalisable si et seulement si on peut trouver $n$ vecteurs propres linéairement indépendants, et donc si et seulement si la multiplicité géométrique de chaque valeur propre est égale à sa multiplicité algébrique. Une condition suffisante est que les multiplicités algébriques soient toutes égales à 1.
Une matrice $A$ est symétrique si $A = A^T$. On remarque donc que $A$ doit nécessairement être carrée.

**But** Nous voulons démontrer deux choses dans cette section :

1. Toutes les matrices symétriques sont diagonalisables, donc on peut écrire $A = PDP^{-1}$.
2. On peut même choisir $P$ orthogonale et $D$ réelle (donc les valeurs propres de $A$ sont toutes réelles).

**Exemple** Soit la matrice suivante :

$$A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$$

On remarque que $A$ est bien symétrique puisque $A^T = A$. Calculons son polynôme caractéristique :

$$p_A(\lambda) = (6-\lambda)^2(5-\lambda) - 2 - 2 - (6-\lambda) - 4(5-\lambda) - (6-\lambda) = (8-\lambda)(6-\lambda)(3-\lambda)$$

Les valeurs propres de $A$ sont donc 8, 6, 3. Elles ont toutes une multiplicité algébrique de 1, donc la matrice est diagonalisable. En calculant $\ker(A - \lambda I_3)$, on obtient :

$$\ker(A - 8I_3) = \text{vect}\left\{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}\right\} \implies \vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

$$\ker(A - 6I_3) = \text{vect}\left\{\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}\right\} \implies \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

$$\ker(A - 3I_3) = \text{vect}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\} \implies \vec{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Ceci nous permet de diagonaliser $A$ :

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -2 & 1 \end{bmatrix}$$

On a que $A\vec{v}_i = \lambda_i \vec{v}_i$, donc $AP = PD$, ce qui nous montre bien qu'on a trouvé $A = PDP^{-1}$.
Nous voulons maintenant trouver l'inverse de $P$. Cela peut prendre beaucoup de temps, mais ici on peut faire une observation utile : les colonnes de $P$ sont orthogonales (et ce n'est pas un accident). En effet :

$$\vec{v}_1 \bullet \vec{v}_2 = 0, \quad \vec{v}_1 \bullet \vec{v}_3 = 0, \quad \vec{v}_2 \bullet \vec{v}_3 = 0$$

Ainsi, si on normalise nos vecteurs, on va obtenir des vecteurs orthonormés, et donc une matrice orthogonale (dont le calcul de l'inverse est très simple, puisque, pour de telles matrices $U^{-1} = U^T$) :

$$\left\|\vec{v}_1\right\|^2 = 1^2 + (-1)^2 + 0^2 = 2 \implies \vec{u}_1 = \frac{1}{\left\|\vec{v}_1\right\|}\vec{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

$$\left\|\vec{v}_2\right\|^2 = 1^2 + 1^2 + (-2)^2 = 6 \implies \vec{u}_2 = \frac{1}{\left\|\vec{u}_2\right\|}\vec{v}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

$$\left\|\vec{v}_3\right\|^2 = 1^2 + 1^2 + 1^2 = 3 \implies \vec{u}_3 = \frac{1}{\left\|\vec{v}_3\right\|}\vec{v}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$\vec{u}_1$, $\vec{u}_2$ et $\vec{u}_3$ sont toujours des vecteurs propres associés aux mêmes valeurs propres, puisqu'on peut les multiplier par un scalaire. Ainsi, on peut redéfinir :

$$P = \begin{bmatrix} \\ \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \\ \\ \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}$$

On a toujours que $A = PDP^{-1}$, avec le même $D$ que ci-dessus. Étudions maintenant $P^TP$ :

$$P^TP = \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vec{u}_3^T \end{bmatrix}\begin{bmatrix} \\ \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \\ \\ \end{bmatrix} = \begin{bmatrix} \vec{u}_1 \bullet \vec{u}_1 & \vec{u}_1 \bullet \vec{u}_2 & \vec{u}_1 \bullet \vec{u}_3 \\ \vec{u}_2 \bullet \vec{u}_1 & \vec{u}_2 \bullet \vec{u}_2 & \vec{u}_2 \bullet \vec{u}_3 \\ \vec{u}_3 \bullet \vec{u}_1 & \vec{u}_3 \bullet \vec{u}_2 & \vec{u}_3 \bullet \vec{u}_3 \end{bmatrix} = I_3$$

Ainsi, on trouve que $P^{-1} = P^T$. Donc :

$$A = PDP^T$$

On dit que $A$ est **diagonalisable en base orthonormée**.

*Importance de la symétrie* Nous nous demandons pourquoi il est nécessaire que $A$ soit symétrique dans cette histoire.
Si $A = PDP^T$ avec $D$ diagonale, alors :

$$A^T = \left(PDP^T\right)^T = \left(P^T\right)^T D^T P^T = PDP^T = A$$

En effet, les matrices diagonales sont symétriques. On se rend donc compte que c'est une condition nécessaire que $A$ soit symétrique pour que l'on puisse espérer diagonaliser $A$ avec une base de vecteurs propres orthonormés.
Il est plus difficile de démontrer qu'il est *suffisant* que $A$ soit symétrique : on va seulement en prouver une partie.

*Observations* Plusieurs choses se sont passées “exactement comme il faut” pour qu'on arrive à la formule $A = PDP^T$ :

1. Toutes les valeurs propres de $A$ étaient réelles.

2. Les valeurs propres de $A$ associés à des valeurs propres distinctes étaient orthogonales.
3. Les multiplicités algébriques et géométriques étaient égales.

Nous allons voir des théorèmes, qui montrent que ceci n'était pas un accident.

**Théorème** Les valeurs propres d'une matrice symétrique sont toujours toutes réelles.

*Preuve* On peut démontrer ce théorème en regardant le conjugué de $A\vec{v} = \lambda\vec{v}$, et en démontrant que $\overline{\lambda} = \lambda$, ce qui nous assure qu'elle est réelle.

**Théorème** Soit $A$ une matrice symétrique, i.e. $A^T = A$.
Si $\vec{v}_1$ et $\vec{v}_2$ sont deux vecteurs propres de $A$ associés à des valeurs propres $\lambda_1$ et $\lambda_2$ *distinctes*, alors $\vec{v}_1$ et $\vec{v}_2$ sont orthogonaux. En d'autres mots :

$$\lambda_1 \neq \lambda_2 \implies \vec{v}_1 \bullet \vec{v}_2 = 0$$

*Preuve* On a :

$$A\vec{v}_1 = \lambda_1\vec{v}_1, \quad A\vec{v}_2 = \lambda_2\vec{v}_2, \quad \lambda_1 \neq \lambda_2$$

On veut démontrer que $\vec{v}_1 \bullet \vec{v}_2 = 0$. Ainsi, considérons $\left(A\vec{v}_1\right) \bullet \vec{v}_2$ :

$$\left(A\vec{v}_1\right) \bullet \vec{v}_2 = \left(\lambda_1\vec{v}_1\right) \bullet \vec{v}_2 = \lambda_1\left(\vec{v}_1 \bullet \vec{v}_2\right)$$

De plus, on voit aussi que :

$$\begin{aligned}\left(A\vec{v}_1\right) \bullet \vec{v}_2 &= \left(A\vec{v}_1\right)^T\vec{v}_2 \\ &= \vec{v}_1^T A^T \vec{v}_2 \\ &= \vec{v}_1^T A \vec{v}_2 \\ &= \vec{v}_1^T \lambda_2 \vec{v}_2 \\ &= \lambda_2\left(\vec{v}_1^T\vec{v}_2\right) \\ &= \lambda_2\left(\vec{v}_1 \bullet \vec{v}_2\right)\end{aligned}$$

Ainsi, on en déduit que :

$$\lambda_1\left(\vec{v}_1 \bullet \vec{v}_2\right) = \lambda_2\left(\vec{v}_1 \bullet \vec{v}_2\right)$$

Or, nécessairement, pour que $\lambda_1 x = \lambda_2 x$ où $\lambda_1 \neq \lambda_2$ (ça nous permet d'éviter la division par 0), il faut nécessairement que $x = 0$. Donc, $\vec{v}_1 \bullet \vec{v}_2 = 0$ : ces vecteurs sont orthogonaux.

□

Jeudi 16 décembre 2021 — **Cours 25 : James Bond Spectre**

**Théorème spectral des matrices symétriques** Toute matrice *symétrique* $A \in \mathbb{R}^{n\times n}$ vérifie les propriétés suivantes :

1. $A$ admet $n$ valeurs propres réelles (pas forcément distinctes).
2. Pour chaque valeur propre $\lambda$, sa multiplicité géométrique est égale à sa multiplicité algébrique.
3. Les sous-espaces propres sont deux à deux orthogonaux, ce qui signifie que deux vecteurs propres associés à des valeurs propres distinctes sont orthogonaux.
4. $A$ est diagonalisable en base orthonormée.

Ce dernier fait est une condition nécessaire et suffisante : $A = A^T$ si et seulement si $A$ est diagonalisable en base orthonormée.

Notez qu'en diagonalisant une matrice symétrique, on pourrait se retrouver avec une matrice $P$ qui ne serait pas orthogonale. En effet, il faut correctement choisir les vecteurs propres associés à la même valeur propre, et il nous faudra peut-être normaliser tous les vecteurs de notre base.

*Preuves et remarques*

1. La preuve n'est pas difficile mais requiert quelques calculs avec des nombres complexes.
2. La preuve est longue, mais ce fait est important car il nous dit que $A$ est diagonalisable.
3. On l'a démontré au cours précédent.
4. C'est-à-dire qu'on peut choisir $n$ vecteurs propres $\vec{u}_1, \ldots, \vec{u}_n$ orthonormés associés à des valeurs propres $\lambda_1, \ldots, \lambda_n$ (avec répétitions) de sorte que :

$$D = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix}, \quad P = \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_n \end{bmatrix}$$

$$A = PDP^T, \quad P^{-1} = P^T$$

*Terminologie*

Le terme "spectral" fait référence aux valeurs propres.

Note personnelle : j'aime beaucoup ce terme, je pense à *James Bond : Spectre* à chaque fois.

**Exemple**

Soit la matrice suivante :

$$A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix} \implies p_A(\lambda) = (7 - \lambda)^2(-2 - \lambda)$$

Le professeur nous conseille vivement de directement regarder si notre matrice est symétrique quand on nous donne une matrice carrée et qu'on commence à nous parler de valeurs propres.
On remarque que les valeurs propres de $A$ sont $\lambda_1 = \lambda_2 = 7$ et $\lambda_3 = -2$. Par le théorème spectral, on connait déjà les multiplicités géométriques de ces valeurs propres : 2 et 1 respectivement, comme leur multiplicité algébrique.
Calculons les vecteurs propres associés à $\lambda_3 = -2$ :

$$\ker(A - (-2)I_3) = \text{vect}\left\{\begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}\right\} \implies \vec{v}_3 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}$$

Puisqu'au final on voudra trouver une base orthonormale, normalisons $\vec{v}_3$ :

$$\vec{u}_3 = \frac{1}{\|\vec{v}_3\|}\vec{v}_3 = \frac{1}{3}\begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}$$

Calculons maintenant les vecteurs propres associés à $\lambda_1 = \lambda_2 = 7$ :

$$\ker(A - 7I_3) = \text{vect}\left\{\begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\} \implies \vec{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

qu'on peut trouver en échelonnant notre matrice.
On peut vérifier que des vecteurs propres associés à des valeurs propres distinctes sont bien orthogonaux :

$$\vec{v}_3 \bullet \vec{v}_1 = 0, \quad \vec{v}_3 \bullet \vec{v}_2 = 0$$

Cependant, on peut remarquer que $\overrightarrow{v}_1$ et $\overrightarrow{v}_2$ ne sont pas orthogonaux

$$\overrightarrow{v}_1 \bullet \overrightarrow{v}_2 = 1 \neq 0$$

$\overrightarrow{v}_1$ et $\overrightarrow{v}_2$ forment une base pour l'espace propre $\ker(A - 7I_3)$, et nous aimerions avoir une base orthonormée pour le même espace propre. L'algorithme du Grand-Schtroumpf nous ne change pas l'espace engendré par les vecteurs, on peut donc l'appliquer sur $\left\{\overrightarrow{v}_1, \overrightarrow{v}_2\right\}$. Cependant, notez bien qu'on n'aurait pas pu l'appliquer sur $\left\{\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3\right\}$, puisqu'on n'aurait pas obtenu de vecteurs propres au final ; mais on peut l'appliquer sur $\left\{\overrightarrow{v}_1, \overrightarrow{v}_2\right\}$, car on va rester dans le sous-espace $\ker(A - 7I_3)$, donc notre résultat nous donnera bien des vecteurs propres.
La première étape du Grand-Schtroumpf consiste à normaliser le premier vecteur :

$$\overrightarrow{u}_1 = \frac{1}{\left\|\overrightarrow{v}_1\right\|}\overrightarrow{v}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$$

On calcule ensuite la projection de $\overrightarrow{v}_2$ sur $\text{vect}\left\{\overrightarrow{u}_1\right\}$, puis on retranche ce résultat à $\overrightarrow{v}_2$ :

$$\overrightarrow{w}_2 = \overrightarrow{v}_1 - \text{proj}_{\text{vect}\left\{\overrightarrow{u}_1\right\}}\overrightarrow{v}_1 = \overrightarrow{v}_1 - \left(\overrightarrow{u}_1 \bullet \overrightarrow{v}_2\right)\overrightarrow{u}_1 = \begin{bmatrix} 4/5 \\ 2/5 \\ 1 \end{bmatrix}$$

Finalement, on peut normaliser $\overrightarrow{w}_2$ :

$$\overrightarrow{u}_2 = \frac{1}{\left\|\overrightarrow{w}_2\right\|}\overrightarrow{w}_2 = \frac{\sqrt{5}}{3}\begin{bmatrix} 4/5 \\ 2/5 \\ 1 \end{bmatrix}$$

On a donc $A = PDP^T$, avec :

$$D = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{bmatrix}, \quad P = \begin{bmatrix} \overrightarrow{u}_1 & \overrightarrow{u}_2 & \overrightarrow{u}_3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} & 4/3\sqrt{5} & -2/3 \\ -2/\sqrt{5} & 2/3\sqrt{5} & -1/3 \\ 0 & 5/3\sqrt{5} & 2/3 \end{bmatrix}$$

**Exemple : décomposition spectrale**

On pourrait aussi écrire la factorisation $A = PDP^T$ sous une forme intéressante, qui décompose $A$ en une somme. En effet, partons de notre diagonalisation en base orthonormale :

$$A = PDP^T = \begin{bmatrix} \overrightarrow{u} & \overrightarrow{v} & \overrightarrow{w} \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}\begin{bmatrix} \overrightarrow{u}^T \\ \overrightarrow{v}^T \\ \overrightarrow{w}^T \end{bmatrix}$$

Donc :

$$\begin{aligned} A &= \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 u_1 & \lambda_2 v_1 & \lambda_3 w_1 \\ \lambda_1 u_2 & \lambda_2 v_2 & \lambda_3 w_2 \\ \lambda_1 u_3 & \lambda_2 v_3 & \lambda_3 w_3 \end{bmatrix}\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \end{aligned}$$

Qui est égal à :

$$\begin{bmatrix} \lambda_1 u_1 u_1 + \lambda_2 v_1 v_1 + \lambda_3 w_1 w_1 & \lambda_1 u_1 u_2 + \lambda_2 v_1 v_2 + \lambda_3 w_1 w_2 & \lambda_1 u_1 u_3 + \lambda_2 v_1 v_3 + \lambda_3 w_1 w_3 \\ \lambda_1 u_2 u_1 + \lambda_2 v_2 v_1 + \lambda_3 w_2 w_1 & \lambda_1 u_2 u_2 + \lambda_2 v_2 v_2 + \lambda_3 w_2 w_2 & \lambda_1 u_2 u_3 + \lambda_2 v_2 v_3 + \lambda_3 w_2 w_3 \\ \lambda_1 u_3 u_1 + \lambda_2 v_3 v_1 + \lambda_3 w_3 w_1 & \lambda_1 u_3 u_2 + \lambda_2 v_3 v_2 + \lambda_3 w_3 w_2 & \lambda_1 u_3 u_3 + \lambda_2 v_3 v_3 + \lambda_3 w_3 w_3 \end{bmatrix}$$

Ainsi, on trouve que c'est égal à :

$$\lambda_1\begin{bmatrix} u_1 u_1 & u_1 u_2 & u_1 u_3 \\ u_2 u_1 & u_2 u_2 & u_2 u_3 \\ u_3 u_1 & u_3 u_2 & u_3 u_3 \end{bmatrix} + \lambda_2\begin{bmatrix} v_1 v_1 & v_1 v_2 & v_1 v_3 \\ v_2 v_1 & v_2 v_2 & v_2 v_3 \\ v_3 v_1 & v_3 v_2 & v_3 v_3 \end{bmatrix} + \lambda_3\begin{bmatrix} w_1 w_1 & w_1 w_2 & w_1 w_3 \\ w_2 w_1 & w_2 w_2 & w_2 w_3 \\ w_3 w_1 & w_3 w_2 & w_3 w_3 \end{bmatrix}$$

Ce qui est lui même égal à :

$$\lambda_1 \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} + \lambda_2 \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} + \lambda_3 \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix}$$

Ce qui nous permet de, finalement, trouver la décomposition spectrale de $A$ :

$$A = \lambda_1 \vec{u}\,\vec{u}^T + \lambda_2 \vec{v}\,\vec{v}^T + \lambda_3 \vec{w}\,\vec{w}^T$$

**Décomposition spectrale**

Soit $A \in \mathbb{R}^{n \times n}$, une matrice symétrique. On sait qu'elle est diagonalisable en base orthonormée. Ainsi, on sait qu'on peut trouver $\lambda_1, \ldots, \lambda_n$, des valeurs propres réelles, et $\vec{u}_1, \ldots, \vec{u}_n$ des vecteurs propres orthonormés de telle manière que $A = PDP^T$ où :

$$D = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix}, \quad P = \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_n \end{bmatrix}$$

Ainsi, on trouve que :

$$\begin{aligned} A &= PDP^T \\ &= \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 \vec{u}_1 & \ldots & \lambda_n \vec{u}_n \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix} \\ &= \lambda_1 \vec{u}_1 \vec{u}_1^T + \ldots + \lambda_n \vec{u}_n \vec{u}_n^T \end{aligned}$$

**Théorème**

Si $A \in \mathbb{R}^{n \times n}$ est symétrique alors :

1. $A$ a $n$ valeurs propres réelles $\lambda_1, \ldots, \lambda_n$ (multiplicités comptées, c'est-à-dire qu'elles sont pas forcément distinctes).
2. Elle admet une base de vecteurs propres orthonormés $\left(\vec{u}_1, \ldots, \vec{u}_n\right)$.
3. On peut trouver la décomposition suivante, appelée **décomposition spectrale** :

$$A = \lambda_1 \vec{u}_1 \vec{u}_1^T + \ldots + \lambda_n \vec{u}_n \vec{u}_n^T$$

## 9.2 Décomposition en valeurs singulières

**But**

Notre but est de trouver la SVD, *Singular Value Decomposition*, qui est "le couteau Suisse de l'algèbre linéaire".

Une fois qu'on aura trouvé la décomposition en valeurs singulières, on peut répondre à toutes les questions qu'on a vues jusque là, comme la dimension du Kernel, ou la matrice de projection sur l'espace image, très simplement.

Dans notre cours, les méthodes qu'on a vues jusque là seront plus rapides. Cependant, pour un ordinateur, avec une matrice un million par un milliard (ce qui est plus courant que ce qu'on pourrait imaginer), alors cela devient très intéressant ; notamment grâce à Gene Golub (et non pas *Gollum*) qui a développé un algorithme permettant de les calculer de manière très efficace.

Le théorème de la SVD, celui que l'on veut développer, nous dit que, pour *n'importe quelle matrice* $A$, il existe deux matrices orthogonales $U \in \mathbb{R}^{m \times m}$ et $V \in \mathbb{R}^{n \times n}$, et une matrice diagonale $\Sigma \in \mathbb{R}^{m \times n}$ telles que :

$$A = U \Sigma V^T$$

**Définition : matrice diagonale** Une matrice $n \times m$ est diagonale si :

$$a_{ij} \neq 0 \implies i = j$$

Voici quelques exemples de matrices diagonales :

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 9 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

**Exemple** Soit la matrice suivante :

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

Cette matrice n'est pas carrée, donc certainement pas symétrique. Ainsi, le théorème spectral ne nous dit rien à propos de $A$ directement. Par contre, on sait que $A^T A$ est toujours carrée et symétrique ($\left(A^T A\right)^T = A^T A$) :

$$A^T A = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

Ainsi, le théorème spectral nous dit que $A^T A$ a trois valeurs propres réelles $\lambda_1, \lambda_2, \lambda_3$ et qu'on peut trouver trois vecteurs propres associés $\left(\vec{v}_1, \vec{v}_2, \vec{v}_3\right)$ orthonormés. En effet, on peut calculer les valeurs propres :

$$p_{A^T A}(\lambda) = -\lambda\left(\lambda^2 - 540\lambda + 32400\right) \implies \lambda_1 = 360, \lambda_2 = 90, \lambda_3 = 0$$

Avec un peu plus de travail, on trouve aussi trois vecteurs propres orthonormaux de $A^T A$ :

$$\vec{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

Ces trois vecteurs sont des vecteurs propres de $A^T A$, mais on se demande ce qu'ils signifient pour $A$. Essayons de regarder ce que donne $A\vec{v}_1$ :

$$A\vec{v}_1 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \end{bmatrix} \implies \left\|A\vec{v}_1\right\|^2 = 18^2 + 6^2 = 360$$

On remarque quelque chose d'intéressant, $\left\|A\vec{v}_1\right\|^2 = \lambda_1$. Faisons le même calcul pour les autres vecteurs propres :

$$A\vec{v}_2 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 3 \\ -9 \end{bmatrix} \implies \left\|A\vec{v}_2\right\|^2 = 90 = \lambda_2$$

$$A\vec{v}_3 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \left\|A\vec{v}_3\right\|^2 = 0 = \lambda_3$$

Nous voulons démontrer que $\left\|A\vec{v}_i\right\|^2 = \lambda_i$ :

$$\begin{aligned} \left\|A\vec{v}_i\right\|^2 &= \left(A\vec{v}_i\right) \bullet \left(A\vec{v}_i\right) = \left(A\vec{v}_i\right)^T \left(A\vec{v}_i\right) = \vec{v}_i^T \underbrace{A^T A\vec{v}_i}_{\lambda_i \vec{v}_i} \\ &= \lambda_i \vec{v}_i^T \vec{v}_i = \lambda_i\left(\vec{v}_i \bullet \vec{v}_i\right) = \lambda_i \left\|\vec{v}_i\right\|^2 = \lambda_i \end{aligned}$$

Cela nous permet d'en déduire que les valeurs propres de $A^T A$ sont toujours positives ou nulles.

Une autre observation qu'on peut faire, qui est très importante, c'est que les vecteurs $A\vec{v}_1, A\vec{v}_2$ et $A\vec{v}_3$ sont orthogonaux :

$$\left(A\vec{v}_1\right) \bullet \left(A\vec{v}_2\right) = 0, \quad \left(A\vec{v}_1\right) \bullet \left(A\vec{v}_3\right) = 0, \quad \left(A\vec{v}_2\right) \bullet \left(A\vec{v}_3\right) = 0$$

On remarque que ce n'est pas un accident. En effet, si $i \neq j$ :

$$\left(A\vec{v}_i\right) \bullet \left(A\vec{v}_j\right) = \left(A\vec{v}_i\right)^T \left(A\vec{v}_j\right) = \vec{v}_i^T \underbrace{A^T A \vec{v}_j}_{\lambda_j \vec{v}_j} = \lambda_j \vec{v}_i^T \vec{v}_j = \lambda_j \vec{v}_i \bullet \vec{v}_j = 0$$

puisqu'on avait choisi les vecteurs propres pour qu'ils soient orthonormaux (le théorème spectral nous garantit leur existence).

Lundi 20 décembre 2021 — **Cours 26 : Calcul de la SVD**

**Suite de l'exemple**

Inévitablement, un des $A\vec{v}_1$, $A\vec{v}_2$ et $A\vec{v}_3$ doit être être nul car ils sont orthogonaux et ils font partie de $\mathbb{R}^2$. Si les trois vecteurs étaient non-nuls, alors ils seraient linéairement indépendants (puisqu'ils sont orthogonaux) ; or, dans $\mathbb{R}^2$ on ne peut pas avoir de famille linéairement indépendantes avec plus de deux vecteurs.
Normalisons nos deux vecteurs non-nuls :

$$\vec{u}_1 = \frac{1}{\left\|A\vec{v}_1\right\|} A\vec{v}_1 = \frac{1}{6\sqrt{10}} \begin{bmatrix} 18 \\ 6 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$\vec{u}_2 = \frac{1}{\left\|A\vec{v}_2\right\|} A\vec{v}_2 = \frac{1}{3\sqrt{10}} \begin{bmatrix} 3 \\ -9 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

Définissons maintenant les **valeurs singulières** de $A$ (les valeurs propres étaient uniquement pour les matrices carrées, les valeurs singulières fonctionnent pour n'importe quelle matrice) :

$$\sigma_1 = \left\|A\vec{v}_1\right\| = \sqrt{\lambda_1} = \sqrt{360} = 6\sqrt{10}$$

$$\sigma_2 = \left\|A\vec{v}_2\right\| = \sqrt{\lambda_2} = \sqrt{90} = 3\sqrt{10}$$

Pour rappel, pour calculer les valeurs singulières d'une matrice $A$, on doit calculer les valeurs propres de $A^T A$.
Ainsi, on a avait déjà :

$$\vec{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix} : \quad \text{base orthonormée de } \mathbb{R}^3$$

$$\vec{u}_1 = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix} : \quad \text{base orthonormée de } \mathbb{R}^2$$

De plus, par construction, on peut écrire :

$$A\vec{v}_1 = \sigma_1 \vec{u}_1, \quad A\vec{v}_2 = \sigma_2 \vec{u}_2. \quad A\vec{v}_3 = \vec{0}$$

On peut écrire tout ça sous forme matricielle :

$$V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix} : \quad \text{matrice } 3 \times 3 \text{ orthogonale}$$

$$U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix} : \quad \text{matrice } 2 \times 2 \text{ orthogonale}$$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix} = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix} : \quad \text{matrice } 2 \times 3 \text{ diagonale}$$

Notez que le fait que $U$ soit symétrique est un accident. Tout ceci nous permet d'écrire :

$$AV = A\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix} = \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & A\vec{v}_3 \end{bmatrix}$$

En utilisant nos égalités, on trouve que :

$$AV = \begin{bmatrix} \sigma_1\vec{u}_1 & \sigma_2\vec{u}_2 & 0 \end{bmatrix} = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix}\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix} = U\Sigma$$

Ceci nous permet de conclure que :

$$AV = U\Sigma \implies A = U\Sigma V^{-1} = U\Sigma V^T$$

puisque $V$ est une matrice orthogonale. On a donc bien trouvé des matrices telles que :

$$A = U\Sigma V^T$$

**Cas général** Soit $A \in \mathbb{R}^{m\times n}$ une matrice symétrique.

- On sait par le théorème spectral que la matrice symétrique $A^T A$ de taille $n \times n$ a $n$ valeurs propres réelles $\lambda_1, \ldots, \lambda_n$ associées à une base de vecteurs propres orthonormées $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{R}^n$. On choisit d'ordonner les valeurs propres (on sait qu'elles sont réelles par le théorème spectral ; si elles étaient complexes on ne pourrait pas les ordonner) de sorte que :

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$$

- Les valeurs propres sont positives ou nulles car $\lambda_i = \left\|A\vec{v}_i\right\|^2$ (on l'a prouvé). On définit donc :

$$\sigma_1 = \left\|A\vec{v}_1\right\| = \sqrt{\lambda_1}, \quad \ldots, \quad \sigma_n = \left\|A\vec{v}_n\right\| = \sqrt{\lambda_n}$$

Puisqu'on a ordonné nos valeurs propres, on a aussi la relation d'ordre suivante :

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$$

On appelle $\sigma_1, \ldots, \sigma_n$ les **valeurs singulières** de $A$.

- Il est possible que certaines des valeurs singulières de $A$ soient nulles. Disons que $A$ a $r$ valeurs singulières non nulles :

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$$

En d'autres mots, si $n > r$, alors $\sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_n = 0$.

- On définit $r$ vecteurs unitaires (c'est-à-dire de norme égale à 1) :

$$\vec{u}_1 = \frac{1}{\left\|A\vec{v}_1\right\|}A\vec{v}_1 = \frac{1}{\sigma_1}A\vec{v}_1$$

$$\vdots$$

$$\vec{u}_r = \frac{1}{\left\|A\vec{v}_r\right\|}A\vec{v}_r = \frac{1}{\sigma_r}A\vec{v}_r$$

On a donc :

$$A\vec{v}_1 = \sigma_1\vec{u}_1, \quad \ldots, \quad A\vec{v}_r = \sigma_r\vec{u}_r$$

Les vecteurs $\vec{u}_1, \ldots, \vec{u}_r$ sont orthonormés dans $\mathbb{R}^m$ car ils sont orthogonaux (on a démontré que $\left(A\vec{v}_i\right) \bullet \left(A\vec{v}_j\right) = 0$ quand $i \neq j$) et car ils sont de norme égale à 1.

- Les vecteurs $\vec{u}_1, \ldots, \vec{u}_r$ sont linéairement indépendants dans $\mathbb{R}^m$. Donc, on a forcément que $r \leq m$ (par la dimension de $\mathbb{R}^m$). Il reste deux cas possibles :
  — Si $r = m$, alors $\vec{u}_1, \ldots, \vec{u}_r$ forment une base orthonormée de $\mathbb{R}^m$.

— Si $r < m$, alors $\vec{\boldsymbol{u}}_1, \ldots, \vec{\boldsymbol{u}}_r$ forment une base orthonormée d'un sous-espace de $\mathbb{R}^m$. On sait qu'il est toujours possible de compléter cette base en choisissant $m - r$ vecteurs $\vec{\boldsymbol{u}}_{r+1}, \ldots, \vec{\boldsymbol{u}}_m$ tels que :

$$\vec{\boldsymbol{u}}_1, \ldots, \vec{\boldsymbol{u}}_r, \vec{\boldsymbol{u}}_{r+1}, \ldots, \vec{\boldsymbol{u}}_m$$

forment une base orthonormée de $\mathbb{R}^m$. En effet, on peut compléter la base avec des vecteurs, puis on peut appliquer le Grand Schtroumpf : il ne fera rien aux $r$ premiers vecteurs puisqu'ils sont déjà orthonormaux, mais va rendre orthonormaux les vecteurs qu'on ajouté.

- Dans tous les cas, on a :

$$A\vec{\boldsymbol{v}}_i = \sigma_i \vec{\boldsymbol{u}}_i, \quad \forall i = 1, \ldots, \min(m, n)$$

(Cette expression ne fait du sens que quand $\vec{\boldsymbol{v}}_i$ et $\vec{\boldsymbol{u}}_i$ sont définis, c'est pourquoi on a besoin du min.)
En effet, c'est vrai pour $i \leq r$ car c'est comme ça qu'on a défini $\vec{\boldsymbol{u}}_i$ :

$$\vec{\boldsymbol{u}}_i = \frac{1}{\left\|A\vec{\boldsymbol{v}}_i\right\|} = A\vec{\boldsymbol{v}}_i, \quad \sigma_i = \left\|A\vec{\boldsymbol{v}}_i\right\|$$

De plus, c'est aussi vrai quand $i > r$ car $\sigma_i = 0$, donc :

$$\left\|A\vec{\boldsymbol{v}}_i\right\| = 0 \implies A\vec{\boldsymbol{v}}_i = \vec{\boldsymbol{0}}$$

- Prenons les matrices suivantes, qui vont nous permettre de simplifier l'écriture :

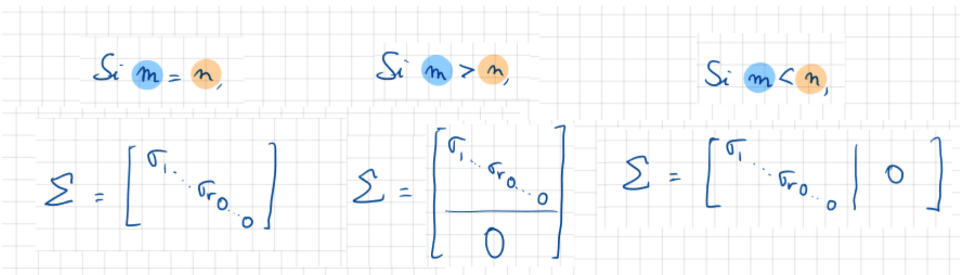
$$U = \begin{bmatrix} \vec{\boldsymbol{u}}_1 & \ldots & \vec{\boldsymbol{u}}_m \end{bmatrix} \in \mathbb{R}^{m \times m}, \quad \text{orthogonale}$$

$$V = \begin{bmatrix} \vec{\boldsymbol{v}}_1 & \ldots & \vec{\boldsymbol{v}}_n \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad \text{orthogonale}$$

$$\Sigma \in \mathbb{R}^{m \times n}, \quad \text{diagonale avec } \Sigma_{ii} = \sigma_i$$

Pour rappel, une matrice $m \times n$ diagonale est telle que :

$$i \neq j \implies \Sigma_{ij} = 0$$




- On peut écrire nos équations de manières compactes :

$$AV = U\Sigma$$

Ainsi, puisque $VV^T = I_n$ (c'est une matrice orthogonale), on sait que $V^{-1} = V^T$. Ceci nous permet d'en déduire que :

$$A = U\Sigma V^T$$

On a donc bien trouvé la décomposition SVD de $A$.

**Décomposition en somme** On sait que quand $A = A^T$, alors on peut écrire :

$$A = VDV^T, \quad V = \begin{bmatrix} \vec{v}_1 & \ldots & \vec{v}_n \end{bmatrix}, \quad D = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix}$$

On avait discuté au cours précédent qu'on peut écrire cela sous la forme de la décomposition spectrale :

$$A = \lambda_1 \vec{v}_1 \vec{v}_1^T + \ldots + \lambda_n \vec{v}_n \vec{v}_n^T$$

De manière complètement similaire, on sait que, par la SVD, on peut écrire :

$$A = U\Sigma V^T,$$

Où :

$$U = \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_m \end{bmatrix}, \quad V = \begin{bmatrix} \vec{v}_1 & \ldots & \vec{v}_n \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1 & \ldots & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \ldots & \sigma_r & \ldots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \ldots & 0 & \ldots & 0 \end{bmatrix}$$

De manière très similaire, on peut donc aussi écrire :

$$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \ldots + \sigma_r \vec{u}_r \vec{v}_r^T$$

Ceci est justifiable de la même manière que pour la décomposition spectrale.

## Jeudi 23 décembre 2021 — **Cours 27 : La puissance de la SVD**

**Rappel** Le théorème de la SVD nous assure que, si on a une matrice quelconque $A \in \mathbb{R}^{m \times n}$, alors, on peut écrire $A$ sous la forme :

$$A = U\Sigma V^T$$

Où :

$$U = \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_m \end{bmatrix} \in \mathbb{R}^{m \times m}, \quad \text{orthogonale}$$

$$V = \begin{bmatrix} \vec{v}_1 & \ldots & \vec{v}_m \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad \text{orthogonale}$$

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \\ \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad \text{diagonale}$$

De plus, on a vu qu'on peut réécrire ce produit matriciel :

$$A = U\Sigma V^T = \sigma_1 \vec{u}_1 \vec{v}_1^T + \ldots + \sigma_r \vec{u}_r \vec{v}_r^T$$

**Théorème : base de l'image** Les vecteurs $\left\{\vec{u}_1, \ldots, \vec{u}_r\right\}$ forment une base orthonormée de $\operatorname{Im} A$.

*Preuve que*

Im $\subset$ vect

On veut montrer que $\operatorname{Im} A \subset \operatorname{vect}\{\vec{u}_1, \ldots, \vec{u}_r\}$.

Soit $\vec{x} \in \mathbb{R}^n$ un vecteurs quelconque (ça aura été une blague très récurrente ; si vous vous demandez d'où ça vient, cherchez "évêque quelconque" sur YouTube (si vous voulez la solution, allez regarder *What the Cut 36*, d'Antoine Daniel)). Alors :

$$\begin{aligned} A\vec{x} &= \left(\sigma_1 \vec{u}_1 \vec{v}_1^T + \ldots + \sigma_r \vec{u}_r \vec{v}_r^T\right)\vec{x} \\ &= \underbrace{\sigma_1}_{\in\mathbb{R}} \underbrace{\vec{u}_1}_{\in\mathbb{R}^{m\times 1}} \underbrace{\vec{v}_1^T}_{\in\mathbb{R}^{1\times n}} \underbrace{\vec{x}}_{\in\mathbb{R}^{n\times 1}} + \ldots + \sigma_r \vec{u}_r \vec{v}_r^T \vec{x} \end{aligned}$$

Or, on remarque que $\vec{v}_i^T \vec{x} \in \mathbb{R}^{1\times 1}$, ce qui est donc un scalaire (on peut même reconnaitre la définition du produit scalaire). On a la commutativité pour le produit matrice-scalaire, donc :

$$A\vec{x} = \sigma_1\left(\vec{v}_1^T\vec{x}\right)\vec{u}_1 + \ldots + \sigma_r\left(\vec{v}_r^T\vec{x}\right)\vec{u}_r$$

Notez que les parenthèses ne sont pas nécessaires, elles sont uniquement là pour faciliter la lecture. On déduit de notre égalité ci-dessus que touts les vecteurs dans $\operatorname{Im} A$ sont des combinaisons linéaires des vecteurs $\vec{u}_1, \ldots, \vec{u}_r$. En d'autres mots :

$$\operatorname{Im} A \subseteq \operatorname{vect}\{\vec{u}_1, \ldots, \vec{u}_r\}$$

*Preuve que* $\operatorname{Im} \subset \operatorname{vect}$

On veut montrer que $\operatorname{vect}\{\vec{u}_1, \ldots, \vec{u}_r\} \subset \operatorname{Im} A$.

Pour cela, il nous suffit de montrer que chaque $\vec{u}_i$ fait partie de $\operatorname{Im} A$. En effet, $\operatorname{Im} A$ est un sous-espace vectoriel ; donc, toute combinaison linéaire de vecteurs qui font partie de cet ensemble, fait aussi partie de cet ensemble.

On remarque bien que $\vec{u}_i \in \operatorname{Im} A$ (pour $i = 1, \ldots, r$) car on peut écrire notre formule comme :

$$A\vec{x} = \sigma_1\left(\vec{v}_1 \bullet \vec{x}\right)\vec{u}_1 + \ldots + \sigma_r\left(\vec{v}_r \bullet \vec{x}\right)\vec{u}_r$$

Ainsi, en prenant $\vec{x} = \frac{1}{\sigma_i}\vec{v}_i$, on a :

$$\begin{aligned} A\vec{x} &= \sigma_1\left(\vec{v}_1^T\vec{x}\right)\vec{u}_1 + \ldots + \sigma_r\left(\vec{v}_r^T\vec{x}\right)\vec{u}_r \\ &= \vec{0} + \ldots + \sigma_i\left(\vec{v}_i^T \frac{1}{\sigma_i}\vec{v}_i\right)\vec{u}_i + \ldots + \vec{0} \\ &= \vec{u}_i \end{aligned}$$

En effet, les vecteurs $\{\vec{v}_1, \ldots, \vec{v}_r\}$ sont orthonormaux, donc $\vec{v}_i \bullet \vec{v}_j = 0$, sauf si $i = j$, auquel cas $\vec{v}_i \bullet \vec{v}_i = 1$.

On a trouvé un $\vec{x}$ tel que $A\vec{x} = \vec{u}_i$, ce qui nous permet d'en déduire que les vecteurs $\vec{u}_i$ font bien partie de l'image de $A$.

Pour démontrer ce fait, on aurait aussi pu partir de la manière dont on a construit la SVD. On avait construit nos vecteurs de telle manière que $A\vec{v}_i = \sigma_i\vec{u}_i$. Ainsi, on avait directement que :

$$A\left(\frac{1}{\sigma_i}\vec{v}_i\right) = \vec{u}_i$$

Ceci nous permet donc de déduire que $\operatorname{vect}\{\vec{u}_1, \ldots, \vec{u}_r\} \subseteq \operatorname{Im} A$, par l'argument donné au début de cette démonstration.

*Preuve* Par les deux points précédents, on sait que :

$$\operatorname{vect}\{\vec{u}_1, \ldots, \vec{u}_r\} = \operatorname{Im} A$$

De plus, on sait que les vecteurs $\{\vec{u}_1, \ldots, \vec{u}_r\}$ sont orthonormés et tous non-nuls, ils sont donc linéairement indépendants.
Ceci nous permet de conclure que les vecteurs $\{\vec{u}_1, \ldots, \vec{u}_r\}$ forment une base orthonormée pour l'image de $A$.

□

**Corollaire 1** Puisqu'on a $r$ vecteurs qui nous font une base, on trouve :

$$\operatorname{rang} A = \dim \operatorname{Im} A = r$$

Ce qui est le nombre de valeurs singulières strictement positives de $A$.

**Corollaire 2** Les vecteurs $\{\vec{u}_{r+1}, \ldots, \vec{u}_m\}$ forment une base de $\ker\left(A^T\right)$.
En particulier, cela nous dit que :

$$\dim\left(\ker A^T\right) = m - r$$

*Preuve* Nos vecteurs $\{\vec{u}_{r+1}, \ldots, \vec{u}_m\}$ forment aussi une base orthonormée, celle du complément orthogonal générées par $\{\vec{u}_1, \ldots, \vec{u}_r\}$. En effet, nous avons bien $m - r$ vecteurs et, si un vecteur est orthogonal aux vecteurs d'une base, alors il est aussi orthogonal à tous les vecteurs de l'espace. Ceci nous permet d'en déduire que $\{\vec{u}_{r+1}, \ldots, \vec{u}_m\}$ forment une base orthonormée pour :

$$(\operatorname{Im} A)^{\perp} = \ker\left(A^T\right)$$

□

**SVD de la transposée** On remarque que, si on connaît la SVD d'une matrice, alors il est très facile de trouver la SVD de sa transposée :

$$A = U\Sigma V^T \implies A^T = \left(U\Sigma V^T\right)^T = \left(V^T\right)^T \Sigma^T U^T = V\Sigma^T U^T$$

En effet, la transposée d'une matrice diagonale $m \times n$ est une matrice diagonale $n \times m$. En particulier, on voit que les valeurs singulières de $A$ et de $A^T$ sont les mêmes.

**Théorème** Les vecteurs $\{\vec{v}_1, \ldots, \vec{v}_r\}$ forment une base orthonormée pour :

$$\operatorname{Im}\left(A^T\right)$$

*Preuve* Nous pouvons utiliser un raisonnement similaire à celui ci-dessus.

**Corollaire** Les vecteurs $\{\vec{v}_{r+1}, \ldots, \vec{v}_n\}$ forment une base orthonormée pour :

$$\left(\operatorname{Im}\left(A^T\right)\right)^{\perp} = \ker(A)$$

**Résumé** Nous avons trouvé des base orthonormées pour les 4 sous-espaces intéressants de $A$ :

| **Sous-espace** | **Base orthonormée** |
| --- | --- |
| $\operatorname{Im} A$ | $\{\vec{u}_1, \ldots, \vec{u}_r\}$ |
| $\ker A^T$ | $\{\vec{u}_{r+1}, \ldots, \vec{u}_m\}$ |
| $\operatorname{Im} A^T$ | $\{\vec{v}_1, \ldots, \vec{v}_r\}$ |
| $\ker A$ | $\{\vec{v}_{r+1}, \ldots, \vec{v}_n\}$ |

**Implications** Non-seulement nous avons trouvé des bases pour ces sous-espaces, mais en plus elles sont orthonormées, ce qui est extrêmement pratique. Par exemple, nous pourrions facilement calculer les matrices de projection sur ces sous-espaces.

**Cas d'une matrice carrée** On sait que le SVD des matrices carrées $A \in \mathbb{R}^{n\times n}$ existent aussi. Ainsi, il existe $U, V$ orthogonales de taille $n \times n$ et $\Sigma$ une matrice diagonale carrée telles que :

$$A = U\Sigma V^T$$

*Déterminant* On peut utiliser notre théorème pour séparer le produit de matrices :

$$\det(A) = \det\left(U\Sigma V^T\right) = \det(U)\det(\Sigma)\det\left(V^T\right)$$

On peut retrouver que le déterminant d'une matrice orthogonale est $\pm 1$ :

$$1 = \det(I_n) = \det\left(U^T U\right) = \underbrace{\det\left(U^T\right)}_{\det(U)}\det(U) = \det(U)^2$$

Ainsi, on trouve que :

$$\det(A) = \pm\sigma_1 \cdots \sigma_n$$

*Inversibilité* En utilisant notre formule ci-dessus, on trouve que $A$ est inversible si et seulement is $\sigma_1, \ldots, \sigma_n$ sont toutes non-nulles. Cela peut aussi se voir avec le fait que le nombre de valeurs singulières non-nulles est égal à la dimension de son espace image, qui doit être égal à $n$ pour les matrices inversibles.

*Inverse* On peut facilement calculer l'inverse :

$$A^{-1} = \left(U\Sigma V^T\right)^{-1} = \left(V^T\right)^{-1}\Sigma^{-1}U^{-1} = V\Sigma^{-1}U^T$$

De plus, l'inverse d'une matrice diagonale peut être facilement calculée :

$$\Sigma^{-1} = \begin{bmatrix} \sigma_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_n \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \frac{1}{\sigma_n} \end{bmatrix}$$

*Valeurs singulières de l'inverse* On a trouvé la SVD de l'inverse ci-dessus, donc les valeurs singulières de l'inverse sont :

$$\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_n}$$

*Inverse de la transpose* On peut faire une preuve différente que celle qu'on avait faite pour $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$ en utilisant la SVD. Cette preuve est laissée en exercice au lecteur (comment terminer *parfaitement* un semestre !).