MATRICES

ARNAUD CHADOZEAU, GABRIEL PALLIER

Table des matières

1. Généralités

1.1. **Place dans l'enseignement.** *Les matrices sont étudiées sous divers points de vue : modélisation de problèmes issus des autres disciplines, systèmes linéaires, transformations géométriques. Il s'agit de mettre en valeur l'efficacité du calcul matriciel pour représenter et résoudre des problèmes.*

1.2. Programme de Terminale math expertes.

Contenus. (Extraits)

— Notion de matrice (tableau de nombres réels). Matrice carrée, matrice colonne, matrice ligne. Opérations. Inverse, puissances d'une matrice carrée.
— Exemples de représentations matricielles : matrice d'adjacence d'un graphe ; transformations géométriques du plan ; systèmes linéaires ; suites récurrentes.
— Exemples de calcul de puissances de matrices carrées d'ordre 2 ou 3.
— Suite de matrices colonnes (U_n) vérifiant une relation de récurrence du type $\mathrm{U}_{n+1} = \mathrm{AU}_n + \mathrm{C}$.
— Graphe orienté pondéré associé à une chaîne de Markov à deux ou trois états.

Date: octobre 2021.
Source : Eduscol, Éducation nationale.

— Chaîne de Markov à deux ou trois états. Distribution initiale, représentée par une matrice ligne Π_0. Matrice de transition, graphe pondéré associé.
— Pour une chaîne de Markov à deux ou trois états de matrice P, interprétation du coefficient (i, j) de P^n. Distribution après n transitions, représentée comme la matrice ligne $\Pi_0\mathrm{P}^n$.
— Distributions invariantes d'une chaîne de Markov à deux ou trois états.

Capacités attendues. (Extraits)
— Modéliser une situation par une matrice.
— Calculer l'inverse, les puissances d'une matrice carrée.
— Dans le cadre de la résolution de problèmes, utiliser le calcul matriciel, notamment l'inverse et les puissances d'une matrice carrée, pour résoudre un système linéaire, étudier une suite récurrente linéaire, calculer le nombre de chemins de longueur donnée entre deux sommets d'un graphe, étudier une chaîne de Markov à deux ou trois états (calculer des probabilités, déterminer une probabilité invariante).

— Expression du nombre de chemins de longueur n reliant deux sommets d'un graphe à l'aide de la puissance n-ième de la matrice d'adjacence.
— Pour une chaîne de Markov, expression de la probabilité de passer de l'état i à l'état j en n transitions, de la matrice ligne représentant la distribution après n transitions.

Problèmes possibles.
— Interpolation polynomiale.
— Marche aléatoire sur un graphe. Étude asymptotique.
— Modèle de diffusion d'Ehrenfest.
— Modèle « proie-prédateur » discrétisé : évolution couplée de deux suites récurrentes.
— Algorithme PageRank.

2. Définition comme tableau de nombres : représenter

2.1. **Lien avec précédents tableaux.** Les programmes suggèrent l'introduction des matrices comme une façon de représenter un tableau de nombres réels. Nous tentons de d'expliquer ce que cela signifie, et quels types de tableaux ont déjà été rencontrés par les élèves.

Caractéristiques générales d'un tableau. De façon la plus générale possible, un tableau est une disposition bidimensionnelle d'informations organisée en lignes et en colonnes, selon une grille. Il est généralement accompagné d'informations supplémentaires (légende, étiquettes, ...) permettant d'interpréter les informations par rapport à leur position dans le tableau. Il est un outil essentiel pour organiser et gérer l'information.

Les informations dans les cellules (intersections d'une ligne et d'une colonne) peut être de n'importe quel type, et éventuellement d'un type différent d'une cellule à l'autre.

Par exemple, les élèves connaissent déjà les *tableaux de signes* et *de variation*, remplis de symboles + ou − ou de 0, et de flèches ascendantes ou descendantes dans le second cas. Dans ces exemples, la représentation des colonnes n'est pas aussi rigide que dans un tableau classique, mais l'ordre des colonnes est absolument crucial.

Deux types de tableaux vont nous intéresser particulièrement : les tableaux numériques et les tableaux booléens.

Tableau à entrée simple ou à entrée double. Par *nature*, un tableau est bidimensionnel, donc à double entrée, sauf lorsque l'information se réduit à une ligne ou une colonne.
Cependant, par *usage*, on peut considérer que certains tableaux bidimensionnels sont à entrée simple.

Cela dépend essentiellement de la signification des informations qu'il contient et de l'utilisation qu'on en fait.

L'**entrée** est l'information dont doit disposer l'utilisateur pour obtenir l'information recherchée. Elle se trouve donc généralement parmi les étiquettes des lignes ou des colonnes, mais elle se trouve aussi parfois à l'intérieur même du tableau.

Le type d'entrée peut être qualitatif ou quantitatif. Un tableau ne permet quoiqu'il arrive de représenter qu'un nombre fini d'entrées possibles. On notera E l'ensemble de ces entrées possibles. Selon l'orientation du tableau, $\operatorname{card} E$ est le nombre de lignes ou de colonnes du tableau.

On distinguera <u>deux sortes de tableaux</u> numériques : ceux qui représentent une fonction de type $\mathrm{E} \to \mathbb{R}^k$ où k est un entier, et ceux qui représentent une fonction de type $\mathrm{E}_1 \times \mathrm{E}_2 \to \mathbb{R}$ où E_1 et E_2 sont deux ensembles d'entrées possibles.

Exemple 1 (**Tableau de valeurs**)**.** Le tableau de valeurs est l'archétype d'un tableau à entrée simple. L'utilisateur cherche la valeur de x qui l'intéresse et lit la valeur de la fonction $f(x)$ qui correspond.

On peut donner l'exemple concret des tables de fonctions logarithmiques ou trigonométriques, qui précédaient l'utilisation généralisée de calculatrice.
Mais un tableau de valeur peut aussi avoir d'autres emplois. Par exemple, pour construire la courbe représentative d'une fonction, on peut utiliser le tableau de valeurs, et considérer les deux nombres de chaque colonne (ou chaque ligne selon l'orientation du tableau) comme l'information recherchée : l'abscisse et l'ordonnée d'un point de la courbe.

N	$\log \mathrm{N}$
1	0
2	0,30103
3	0,47712
4	0,60206
5	0,69897
6	0,77815
7	0,84510
8	0,90309
9	0,95424
10	1

θ	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\sin(\theta)$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1

La tableau de valeurs devient le tableau de valeur de la fonction $i \mapsto (x_i, f(x_i))$, qui à chaque indice de colonne associe le vecteur de $\mathbb{R}^2$ dont les composantes sont les données de la colonne correspondante.
Ainsi, on pourra donc représenter le tableau de valeurs d'une fonction f en les valeurs $\mathrm{E} = \{x_1, \ldots, x_n\}$ par l'une des matrices

$$\begin{pmatrix} f(x_1) & f(x_2) & \cdots & f(x_n) \end{pmatrix} \quad \text{ou} \quad \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ f(x_1) & f(x_2) & \cdots & f(x_n) \end{pmatrix}$$

selon la situation.

Tableau à « sorties multiples ». L'exemple précédent montre qu'on peut avoir des tableaux à entrée simple, mais « à sorties mutiples », où l'information recherchée est sous la forme d'une série de nombres. On a dans ce cas-là une fonction sous-jacente, qui, pour une entrée choisie parmi un ensemble E, fournit l'information sous la forme d'un k-uplet de nombre, dans $\mathbb{R}^k$.
La représentation de ce tableau possèdera $\operatorname{card} \mathrm{E}$ lignes (une pour chaque entrée possible) et k colonnes (une pour chaque composante de la sortie), ou $\operatorname{card} \mathrm{E}$ colonnes et k lignes selon l'orientation choisie.

On pourrait également modéliser cette situation par une fonction $\mathrm{E} \times [\![1, k]\!] \to \mathbb{R}$, ce qui correspondrait à un tableau à double entrée (une entrée parmi E, l'autre comme indice de $[\![1, k]\!]$). Il y a cependant une différence à faire lorsque les nombres sont associés à des grandeurs.

Exemple 2 (**Tableau de proportionnalité**)**.** Dans le cadre des grandeurs, la proportionnalité est une relation entre deux (ou plus) types de grandeurs associées à une famille d'objets. Lorsque, d'un objet à un autre, une des grandeurs est multipliée par un coefficient, toute autre grandeur sera aussi multipliée par ce coefficient.

Masse de farine (en g)	200	300	500
Quantité d'œufs	2	3	5
Volume de lait (en cL)	40	60	100
Quantité de crêpes	10	15	25

Dans le cas présent, chaque colonne correspond à une recette particulière. Et chaque recette est associée à une série de données numériques. Mais ces données ne sont pas du même type, elle ne sont pas les mesures d'un même type de grandeur.

Là encore, l'entrée dépend un peu de l'utilisation : souhaite-t-on savoir quels ingrédients sont nécessaires pour faire 45 crêpes, ou combien de crêpes il est possible de faire avec 7 œufs ?

Tableau à double entrée. La situation est différente dans un tableau à double entrée, pour lequel la fonction sous-jacente est de la forme $\mathrm{E}_1 \times \mathrm{E}_2 \to \mathbb{R}$. Pour une entrée choisie parmi E_1, et une entrée choisie parmi E_2, on obtient une information numérique. Mais contrairement au cas ci-dessus, on ne s'attend qu'à des valeurs numériques d'un seul type de grandeur.

Exemple 3 (**Tableau statistique à double entrée**)**.** On représente les températures moyennes en ºC de 4 villes américaines selon les mois de l'année. Une des entrées est géographique, ordonnée alphabétiquement par convention, l'autre temporelle, ordonnée naturellement. Toutes les données sont d'un même type de grandeur.

	Jan.	Fév.	Mars	Avr.	Mai	Juin	Juil.	Août	Sep.	Oct.	Nov.	Déc.
Boston	-1	0	5	10	15	21	24	23	18	12	7	1
Las Vegas	14	17	21	26	31	37	40	39	34	27	19	14
Orlando	16	16	19	22	25	27	28	28	27	23	19	16
San Fransisco	14	16	17	18	18	20	20	21	22	21	18	15

Même dans un cadre purement numérique, on peut considérer l'exemple des tables de multiplication : on représente dans chaque cellule le produit des deux entrées. Dans ce cas, on a généralement $\mathrm{E}_1 = \mathrm{E}_2$.

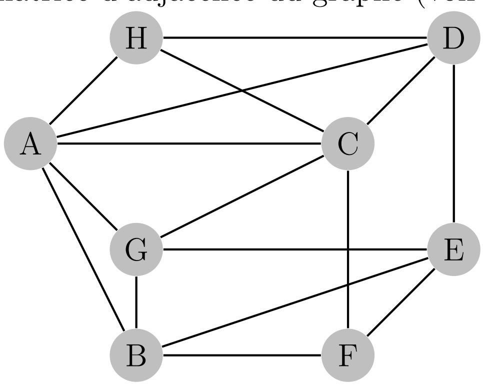
On cite finalement un dernier exemple de tableaux à double entrée : le tableau à informations booléennes, généralement représentées sous forme de croix. Il permet d'exprimer une relation entre les éléments de E_1 et ceux de E_2. Voici un exemple [1]

A, B, C, D, E, F, G et H désignent huit poissons ; dans le tableau ci-dessous, une croix signifie que les poissons ne peuvent cohabiter dans un même aquarium.

Quel nombre minimal d'aquariums faut-il ?

	A	B	C	D	E	F	G	H
A		×	×	×			×	×
B	×				×	×	×	
C	×			×		×	×	×
D	×		×		×			×
E		×		×		×	×	
F		×	×		×			
G	×	×	×		×			
H	×		×	×				

Cette situation peut se modéliser efficacement par un graphe. Le tableau peut être interprété comme tableau numérique, en remplaçant chaque croix par un 1, et chaque absence de croix par un 0. La matrice associée à ce tableau correspond alors à la matrice d'adjacence du graphe (voir exemple 7).



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

2.2. **Matrices.** Une matrice (réelle) est une famille finie d'éléments (de nombres réels) doublement indexée, ou plutôt indexée par un produit cartésien de deux ensembles finis d'indices.

À chaque tableau de nombres, on peut associer une matrice : le premier ensemble d'indices correspond à la ligne, le second à la colonne. Il s'agit d'une matrice indexée par les étiquettes des lignes et des colonnes du tableau.

Il est cependant plus courant d'indexer par les nombres de 1 à n pour les lignes, et de 1 à m pour les colonnes, lorsqu'il s'agit d'une matrice à n lignes et m colonnes. On peut alors interpréter une matrice comme une fonction $[\![1, n]\!] \times [\![1, m]\!] \to \mathbb{R}$.

Il y a cependant des exceptions, par exemple :

— il peut être pratique d'indexer de 0 à n, par exemple lorsque que l'on travaille sur des polynômes de degré inférieur ou égal à n et que l'on indexe la base canonique $(1, \mathrm{X}, \ldots, \mathrm{X}^n)$ par le degré.

— lorsqu'on associe une matrice d'adjacence à un graphe, il est naturel d'indexer lignes et colonnes par les sommets du graphe. (voir partie graphe)

Représentation graphique. Une matrice est indissociable de sa représentation sous forme de tableau, ou de grille de nombres. Pour la représenter il est nécessaire de

1. tiré de l'annexe aux documents d'accompagnement 2002 *Les graphes dans l'enseignement de spécialité*, Eduscol.

choisir un ordre pour chaque ensemble d'indice. Cet ordre est naturel lorsque ces indices sont des entiers, mais il doit être explicité dans le cas de graphe, par exemple.

Cette représentation graphique est porteuse de significations, et certaines formes plus importantes portent des noms particuliers (matrice triangulaire, diagonale, échelonnée, matrice ligne, matrice colonne, ...).

2.3. **De nombreuses représentations.** Outre les tableaux, les matrices permettent de représenter de très nombreux objets. L'une des particularités des matrices est qu'elles permettent aussi bien de représenter des objets que les transformations qui agissent sur ces objets.

Exemple 4 (**Points et vecteurs**)**.** Dans le plan repéré (ou dans l'espace), on peut identifier un point par ses deux (ou trois) coordonnées, que l'on dispose dans une matrice colonne.
Un vecteur est aussi identifié par ses coordonnées, aussi disposé dans une matrice colonne.
Une même matrice peut représenter un objet ou une action. De même, l'addition $\begin{pmatrix}2\\4\end{pmatrix} + \begin{pmatrix}-3\\8\end{pmatrix} = \begin{pmatrix}-1\\12\end{pmatrix}$ peut représenter l'effet d'une translation par le vecteur de coordonnées $\begin{pmatrix}-3\\8\end{pmatrix}$ sur le point de coordonnées $\begin{pmatrix}2\\4\end{pmatrix}$, ou la somme de deux vecteurs, ce qui correspond à la composition de deux translations.
Cette même dualité objet/action s'observe déjà avec les entiers, où un entier peut être à la fois un objet statique (un cardinal, par exemple) ou une transformation, dynamique, sur les entiers.

Dans la suite, on liste les principales représentations possibles à l'aide de matrice, en mettant l'accent sur celles rencontrées dans l'enseignement secondaire.

2.3.1. *Structure de données.* Une matrice peut être utilisée pour représenter une donnée statique. C'est d'ailleurs souvent le cas en informatique où la matrice est considérée comme une structure de données fondamentale.

Exemple 5 (**Images digitales**)**.** Les images digitales sont généralement présentées sous la forme d'une grille de pixels, et c'est sous format qu'elles sont affichées sur un écran. Si ces pixels sont généralement la donnée de 3 ou 4 nombres (pour les couleurs Rouge, Vert, Bleu par exemple), on peut se limiter à un seul nombre pour les photos monochromes (en noir et blanc notamment). On associe à une image une matrice dont les indices permettent de situer le pixel, et la valeur du coefficient est lié à la coloration du pixel.

Dans l'exemple précédent, les deux ensembles d'indices jouent des rôles similaires, et chaque cellule prend son sens par sa position dans la matrice. Mais il est des exemples où l'information contenue dans une matrice se structure par lignes ou par colonnes. Par exemple, dans la matrice associée à une famille de vecteurs, chaque colonne recueille les coordonnées d'un vecteur de la famille ; l'information contenue dans la matrice prend tout son sens lorsqu'on la considère par colonne. On donne un exemple (celui-ci au programme de Terminale - Maths expertes) pour lequel l'information de la matrice se structure en lignes.

Exemple 6 (**Systèmes linéaires**)**.** Lors de la résolution d'un système linéaire d'équations, on est amené à transformer les équations pour en faire apparaître

d'autres. Les lettres ont un rôle limité, et ne servent finalement qu'à indiquer quel coefficient interagit avec quel autre d'une autre équation lors de la combinaison de ces équations.
Ainsi, il est possible d'associer au système

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

la matrice *augmentée* des coefficients

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

où la barre verticale est juste cosmétique, et sert uniquement à rappeler que la dernière colonne possède une signification différente.
Mais c'est une autre représentation matricielle du système, sous la forme d'un produit matriciel, qui est préférée :

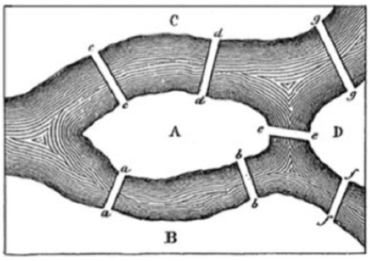
$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

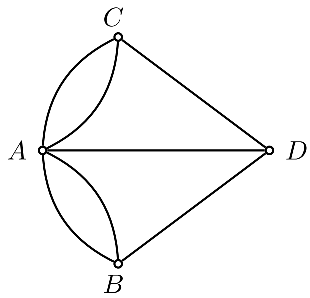
Cette représentation a l'avantage de faire apparaître la matrice des coefficients de la partie linéaire du système. On est alors proche d'une interprétation du système comme la recherche d'un antécédent par une application linéaire. C'est aussi historiquement plus proche de l'introduction de la notation matricielle.

Notre dernier exemple (détaillé dans le poly *Graphes*) concerne les graphes non-orientés.

Exemple 7. Matrice d'adjacence Soit G un graphe non-orienté d'ordre n, dont les sommets forment un ensemble S. On ordonne les sommets, par exemple en les indexant de 1 à n, de telle sorte qu'on puisse représenter la matrice d'adjacence de façon unique.
On définit la *matrice d'adjacence de G* comme la matrice carrée $A = (a_{ij})_{1\leqslant i,j\leqslant n} \in \mathcal{M}_n(\mathbb{R})$ où a_{ij} est le nombre d'arêtes entre les sommets s_i et s_j.
Dans le cas d'un graphe non-orienté, cette matrice est symétrique.

Par exemple, dans le problème des ponts de Königsberg, la situation peut être modélisée par le graphe ci-dessous, et en utilisant l'ordre alphabétique pour ordonner les sommets, on associe à ce graphe la matrice d'adjacence M, qui contient la même information.



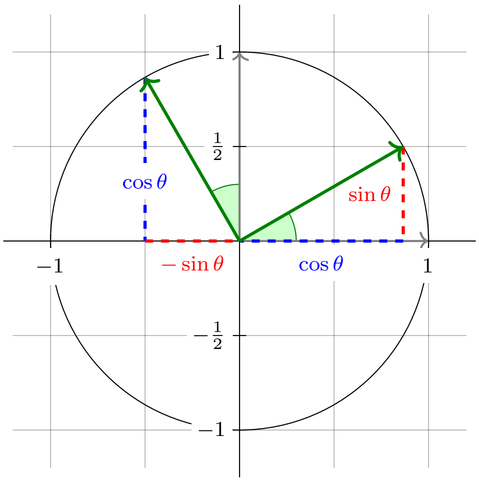


$$M = \begin{pmatrix} 0 & 2 & 2 & 1 \\ 2 & 0 & 0 & 1 \\ 2 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

2.3.2. *Transformations.* De toutes les représentations par des matrices, la plus utilisée, ou du moins celle qui vient en premier à l'esprit, est la représentation des applications linéaires, c'est-à-dire d'une transformation d'un espace vectoriel à un autre. Les programmes de mathématiques expertes en donnent quelques exemples dans des cadres particuliers.

Exemple 8 (Similitudes vectorielles du plan). On rappelle qu'une *similitude du plan euclidien* est une transformation du plan qui multiplie les distances par un coefficient constant, appelé *rapport de similitude*. Une similitude est dite *vectorielle* si elle laisse invariant l'origine du plan. Toute similitude se décompose comme la composée (en général non commutative) d'une translation et d'une similitude vectorielle. L'étude des similitudes ou de leurs représentations n'est pas au programme. Seuls les exemples suivants sont traités.
Les homothéties vectorielles. Pour $\rho \in \mathbb{R}$, l'application qui à un vecteur $\vec{v}$ associe son multiple $\rho\vec{v}$ est l'homothétie vectorielle de rapport a. Cette transformation est représentée par la matrice $H_\rho = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix}$, c'est-à-dire $\rho\mathrm{I}_2$. En effet, si le vecteur $\vec{v}$ a pour coordonnées $\begin{pmatrix} x \\ y \end{pmatrix}$, alors $\rho\vec{v}$ a pour coordonnées $H_\rho \begin{pmatrix} x \\ y \end{pmatrix}$, c'est-à-dire $\begin{pmatrix} \rho x \\ \rho y \end{pmatrix}$.
Plus généralement, pour toute matrice M à deux lignes, on a $H_\rho M = \rho M$ et pour toute matrice M à deux colonnes, on a $MH_\rho = \rho M$. En particulier, H_ρ commute avec toute matrice de $\mathcal{M}_2(\mathbb{R})$.[2]

Les rotations vectorielles. Pour $\theta \in \mathbb{R}$, on note R_θ la matrice $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.



Vue comme matrice associée à une famille de vecteurs, R_θ est associé à la base orthonormée obtenue après rotation d'angle θ dans le sens trigonométrique des vecteurs la base canonique $(\vec{\imath}, \vec{\jmath})$.
Une relation importante est que pour tout $\theta \in \mathbb{R}$, et tout $\phi \in \mathbb{R}$, on a $R_\phi R_\theta = R_{\phi+\theta}$. En effet, R_ϕ est la matrice de la base obtenue par rotation d'angle $\phi + \theta$ de la base $(\vec{\imath}, \vec{\jmath})$ exprimée dans la base obtenue par rotation d'angle θ de la base $(\vec{\imath}, \vec{\jmath})$. En exprimant toutes ces bases de nouveau dans la base $(\vec{\imath}, \vec{\jmath})$, on obtient la relation attendue.

2. Cela a aussi pour conséquence que H_ρ est la matrice de l'homothétie vectorielle de rapport ρ par rapport à n'importe quelle base de $\mathbb{R}^2$.

Le concept de transformation possède un aspect temporel, un avant et un après, qui n'est pas exploité dans le dernier exemple. Cet aspect temporel est beaucoup plus présent dans l'exemple suivant de suites récurrentes, qui sert généralement à représenter des systèmes dynamique à temps discret.

Exemple 9 (Suites récurrentes). On considère un système dont l'état au temps n est décrit par les réels (u_n, v_n)[3]. On suppose que la loi d'évolution du système se présente sous la forme de fonctions affines : pour tout $n \in \mathbb{N}$, on a

$$u_{n+1} = au_n + bv_n + e \quad \text{et} \quad v_{n+1} = cu_n + dv_n + f,$$

où a, b, c, d, e et f sont des constantes réelles.
Un tel modèle est généralement admissible lorsque les valeurs des suites u_n et v_n ne varient pas trop, et restent proches des valeurs de u_0 et v_0 ; on peut alors approcher une loi d'évolution plus générale du type $u_{n+1} = F(u_n, v_n)$ par sa partie linéaire $F(u_n, v_n) \approx F(u_0, v_0) + \frac{\partial F}{\partial u}(u_0, v_0)(u_n - u_0) + \frac{\partial F}{\partial v}(u_0, v_0)(v_n - v_0)$, et pareillement pour la relation $v_{n+1} = G(u_n, v_n)$.
On peut représenter ces relations sous forme matricielle. On définit les matrices colonnes $U_n = \begin{pmatrix} u_n \\ v_n \end{pmatrix}$ pour tout $n \in \mathbb{N}$ pour représenter l'état du système au temps n et les matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ et $B = \begin{pmatrix} e \\ f \end{pmatrix}$ pour représenter la loi d'évolution. On obtient alors une suite récurrente de matrices colonnes définie par la relation de récurrence $U_{n+1} = AU_n + B$.

Ce type de représentation s'adapte également au système dont la loi d'évolution prend en compte un nombre fini d'état dans le passé. Pour limiter la taille des matrices, on ne considère que le cas où l'état du système au temps n peut être décrit par un seul nombre u_n, et que l'état au temps $n+1$ est déterminé par l'état du système au temps n, mais aussi à celui au temps $n-1$. On considère donc une loi de la forme $u_{n+1} = au_n + bu_{n-1} + c$. On peut représenter la situation par des vecteurs colonnes $U_n = \begin{pmatrix} u_{n-1} \\ u_n \end{pmatrix}$ pour $n \geqslant 1$, et par la relation

$$U_{n+1} = \begin{pmatrix} 0 & 1 \\ b & a \end{pmatrix} U_n + \begin{pmatrix} 0 \\ c \end{pmatrix}.$$

Cela peut se comprendre par le fait que la passé immédiat du système (décrit par u_{n-1}) doit également faire partie de l'état du système au temps n, puisqu'il conditionne l'état au temps $n+1$.

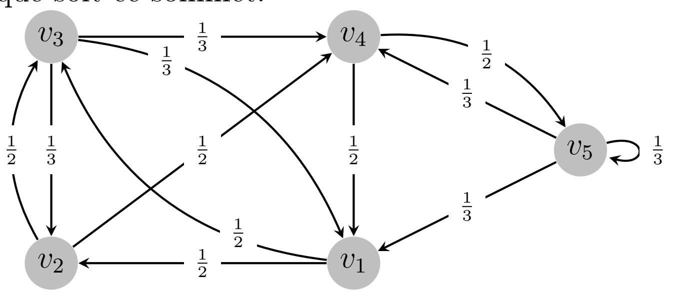
En ce qui concerne les graphes, les graphes orientés présentent cet aspect temporel, par orientation des arcs reliant les sommets. Et on peut, tout comme pour les graphes non-orientés, décrire l'information contenue dans un graphe orienté en utilisant la matrice d'adjacence de celui-ci : les coefficients d'indice (i, j) est égal au nombre d'arcs partant du i^{e} sommet et arrivant au j^{e}. La différence essentielle est ici que la matrice obtenue n'est plus nécessairement symétrique.

Cependant, la transformation associée à un graphe orienté n'est pas claire lorsque plus d'un arc partent d'un sommet. Dans ce cas, on peut définir une loi de probabilité sur les arcs partant d'un sommet.

3. On se limite à deux variables d'état dans les programmes, mais cela se généralise à un plus grand nombre de variable

Exemple 10 (Graphes probabilistes). Un graphe probabiliste est un graphe orienté sans arc multiple dans lequel :
— à chaque arc est associé un nombre dans $]0,1]$.
— pour tout sommet, la somme des nombres associés aux arcs partant de ce sommet vaut 1.

Les graphes probabilistes permettent de représenter la loi de transition d'une chaîne de Markov (voir poly *Graphes*). On peut représenter ces graphes par une matrice en suivant la même règle que pour la matrice d'adjacence, mais en prenant en compte la pondération par les probabilités : Pour tout couple d'indice (i,j), le coefficient de la matrice d'indice (i,j) vaut 0 s'il n'y a pas d'arc allant du i^{e} sommet au j^{e} sommet, et si cet arc existe, ce coefficient vaut la probabilité associée à cet arc.
Lorsque le graphe probabiliste représente la loi de transition d'une chaîne de Markov, on appelle cette matrice la **matrice de transition** de la chaîne de Markov.
On donne ci-dessous un exemple avec un graphe probabiliste associée à une marche aléatoire, pour lesquels les arcs partant d'un même sommet sont équiprobables, quelque soit ce sommet.



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

2.4. **Matrice de permutation.** Les matrices de permutation sont traitées dans la partie disciplinaire de ce cours, en lien avec les opérations élémentaires et l'algorithme de Gauss-Jordan. Il ne s'agit pas d'une notion au programme de Terminale maths experte mais il peut être utile de l'avoir à l'esprit dans les situations d'enseignement.

Par rapport à la discussion qui précède, les matrices de permutation vont prendre la place d'un certain type de tableau à informations booléennes à double entrée, avec les trois spécificités suivantes :
— Les deux entrées dans le tableau s'opèrent par la même liste, parcourue dans le même ordre.
— Chaque ligne possède une seule '×' (qui sera encodée par le 1 de l'anneau des scalaires, et dont l'absence est encodée par 0)
— Chaque colonne possède une seule '×'

Voici un exemple de tableau à information booléenne de ce type, et la matrice de permutation J qu'on peut lui associer.

	poule	renard	vipère
poule		×	
renard			×
vipère	×		

$$\mathrm{J} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Ce tableau décrit l'information "est poursuivi par..." ou "poursuit", selon le sens de lecture, qui importe réellement, comme on va le détailler plus loin.

En réalité, les tableaux du type précédent sont peu utilisés car ils décrivent simplement une bijection de l'ensemble indexant les lignes et colonnes, et si le but est simplement de fournir une telle description, il est plus efficace d'utiliser un tableau de valeurs. En voici pour les permutations « chassé » et « chasseur » de l'ensemble {poule, renard, vipère}.

x	poule	renard	vipère
chassé(x)	vipère	poule	renard

x	poule	renard	vipère
chasseur(x)	renard	vipère	poule

Le tableau de valeur est aussi efficace pour les permutations de $[\![1, \ldots, n]\!]$, que l'étudiant du supérieur décrira plus aisément par des tableaux de nombres, voire par leur décomposition en cycles. Notons toutefois que ces deux dernières représentations sont moins commodes pour déterminer l'inverse, ou bien pour s'assurer que l'application représentée est bien bijective. L'intérêt de la présentation à double entrée apparaît clairement lors de la composition et l'inversion des applications. L'aspect essentiel est que les deux copies de la liste sont ordonnées (sur notre exemple, dans l'ordre alphabétique), donc chercher un élément dans ces listes se fait à faible coût. Quand à la composition, c'est la multiplication matricielle qui va jouer son rôle. Commençons par rappeler la définition d'une matrice de permutation.

Définition 1 (Matrice de permutation)**.** Soit σ une permutation de $[\![1, \ldots, n]\!]$ et $\mathbf{K}$ un corps. La matrice de permutation P_σ est par définition l'élément de $\mathrm{GL}(n, \mathbf{K})$ égal à

$$\mathrm{P}_\sigma = \sum_{j=1}^{n} \mathrm{E}_{\sigma(j),j}. \tag{1}$$

Par définition, P_σ représente donc l'endomorphisme de $\mathbf{K}^n$ qui envoie le j-ième vecteur de base sur le $\sigma(j)$-ième vecteur de base, exprimé dans la base canonique. Notons aussi que $\mathrm{P}_\sigma \mathrm{P}_\lambda = \mathrm{P}_{\sigma\lambda}$ pour toutes $\sigma, \lambda \in \mathcal{S}_n$. Concernant les traductions en termes d'opérations élémentaires, le mantra est le suivant :

— Multiplier $\mathrm{M} \in \mathcal{M}_{n,m}(\mathbf{K})$ pour $m \geqslant 1$ par la matrice de permutation P_σ **à gauche**, c'est permuter ses lignes suivant σ.

— Multiplier une matrice $\mathrm{M} \in \mathcal{M}_{m,n}(\mathbf{K})$ pour $m \geqslant 1$ par une matrice de permutation **à droite**, c'est permuter ses colonnes suivant σ^{-1}.

Multiplication matricielle	Opérations élémentaires
Remplacer M par $\mathrm{P}_\sigma M$	$\forall j \in [\![1, \ldots, n]\!], \mathrm{L}_j \leftarrow \mathrm{L}_{\sigma(j)}$
Remplacer M par MP_σ	$\forall i \in [\![1, \ldots, n]\!], \mathrm{C}_i \leftarrow \mathrm{C}_{\sigma^{-1}(i)}$

Table 1. Opérations sur les matrices et matrices de permutation

Il est tout aussi possible d'utiliser la convention opposée (ou *représentation contragédiente*) et de définir ${}_\sigma\mathrm{P} = \sum_{j=1}^{n} \mathrm{E}_{i,\sigma(i)}$. La contrepartie que nous avons à subir quand nous nous fixons sur la Définition 1 est que, quand nous manipulons des matrices ligne (par exemple, dans le cas des chaînes de Markov que nous verrons plus loin), nous ne devons pas oublier que LP_σ, les coefficients de la matrice ligne L sont permutés par σ^{-1} et non par σ. Le choix de convention (1) est donc un arbitrage considérant que les matrices opèrent plus souvent sur les vecteurs colonne que sur les vecteurs ligne, usage lui-même hérité de l'écriture même des systèmes linéaires.

Exemple 11 (Codage d'image et transformations)**.** Reprenons l'exemple de l'image digitale, codée par une (disons M) ou trois (disons R, G et B) matrices dans $\mathcal{M}_{n,m}(\mathbf{R})$. On déplace le cadre de l'image en remplaçant M par $\mathrm{P}_\gamma^k \mathrm{M} \mathrm{P}_\rho^\ell$, où γ (resp. ρ) est la permutation cyclique de $[\![1, \ldots, n]\!]$ (resp. de $[\![1, \ldots, m]\!]$).

On notera tout de même que le codage d'une image digitale par une matrice véritablement rectangulaire (c'est-à-dire, pour laquelle $\min(m, n) \geqslant 2$) n'est pas pratique dans le cas où nous souhaiterions effectuer des opérations plus compliquées sur l'image, par exemple, la tourner d'un angle de 90 degrés ou bien obtenir une image miroir. Pour ces manipulations, il est (au moins théoriquement) plus efficace d'utiliser une seule ligne ou colonne décrivant l'image. La matrice de permutation permettant d'effectuer une rotation de l'image est alors de forme plus compliquée, mais elle est fixée une fois pour toute si le format de l'image est donné. Cela permet aussi d'englober dans un même cadre les transformations géométrique de l'image et les changements comme le floutage, qui consiste à remplacer la valeur d'un pixel par sa moyenne sur les pixels avoisinants, à l'aide d'une multiplication par une matrice qui est une petite perturbation de l'identité.

3. Opérations sur les matrices : Résoudre

Les matrices permettent donc de représenter des objets et des situations de types très différents. Cette représentation permet généralement d'effectuer des calculs qui peuvent s'interpréter dans la situation représentée, et uniquement grâce à peu d'opérations définies sur les matrices (addition, multiplication par un scalaire, multiplication de matrices).

3.1. **Espace vectoriel de matrices.** L'ensemble $\mathcal{M}_{n,m}(\mathbb{R})$ des matrices réelles de dimension $n \times m$ possède une structure naturelle d'espace vectoriel, fournie par l'addition terme à terme, et la multiplication par un scalaire distribuée sur tous les coefficients. Il s'agit donc d'un espace vectoriel de dimension nm, et qui ne diffère de $\mathbb{R}^{nm}$ que par la disposition spatiale des coefficients.

L'interprétation de ces deux opérations n'est généralement que le reflet de la structure d'espace vectoriel de l'objet représenté. C'est par exemple le cas de l'exemple 4 de représentation de vecteurs.

C'est aussi le cas de l'exemple 1 des tableaux de valeurs. Considérons deux fonctions réelles f et g et considérons leur tableau de valeur respectif pour un choix de valeurs $\mathrm{E} = \{x_1, \ldots, x_n\}$. Les matrices représentant ces tableaux sont de dimension $1 \times n$: il est important que l'on considère les valeurs x_i comme des étiquettes du tableau, et donc qu'elles n'apparaissent pas dans la matrice représentant le tableau de valeur. Il est également important que les valeurs choisies pour les x_i soient les mêmes pour les deux fonctions.
Dans ce cas, la somme de des matrices correspond alors au tableau de valeurs de la fonction somme $f + g$, et le produit par un scalaire $a \in \mathbb{R}$ de la matrice associée à f est la matrice associée au tableau de valeurs de la fonction af. Ces opérations ne font que traduire le fait que les fonctions à valeurs réelles définies sur un ensemble E forment un espace vectoriel.

Cela se généralise aux deux types de tableaux numériques : les espaces de fonctions de type $\mathrm{E} \to \mathbb{R}^k$ et de celles de type $\mathrm{E}_1 \times \mathrm{E}_2 \to \mathbb{R}$ sont munies d'une structure d'espace vectoriel.[4] Ainsi, l'espace des tableaux numériques qui partagent une même dimension et de mêmes étiquettes est muni d'une structure d'espace vectoriel, et l'on peut ajouter ces tableaux terme à terme, ou multiplier toutes leurs entrées par une même constante.

Cependant, les programmes préconisent d'introduire ces opérations à partir de tableaux, mais en s'appuyant sur le sens de ces opérations. Or, comme dans les exemples 2 et 3, les nombres remplissant le tableau ont un sens en termes de grandeurs. Cela complique la situation. On donne quelques exemples typiques.

Tableaux de grandeurs. On rappelle que lorsqu'une grandeur sur un type d'objets est mesurable, on peut estimer la grandeur d'un système formé de plusieurs objets comme la somme des grandeurs de chaque objet (voir poly *Grandeurs*). C'est par exemple le cas de la masse, du volume, du prix, de toute quantité, etc. Ce n'est pas souvent le cas de grandeurs quotients — comme la masse volumique, le prix à l'unité, etc. — ou de grandeur repérable non mesurable comme la température, la position, etc.

Pour deux tableaux similaires associés à deux objets différents, il est également possible de considérer les grandeurs associées à ces deux objets pris comme système. Puisqu'il s'agit de grandeurs mesurables, chaque grandeur est la somme des deux grandeurs associées à chacun des objets. L'opération correspond à l'addition des matrices.

Lorsqu'un tableau est rempli de grandeurs mesurables associées à un objet ou une situation, il est possible de considérer les grandeurs associées à un multiple de cet objet. Par exemple, dans le cas des crêpes de l'exemple 2, on pourra considérer le tableau multiplié par 10, adressé à un pâtissier, un restaurateur ou un industriel, peut-être. Cette opération correspond bien à la multiplication de la matrice par un scalaire. On note que ce scalaire est ici un nombre sans dimension, qui agit comme un rapport d'agrandissement.

Il est possible de considérer aussi la multiplication du tableau par un nombre qui ne soit pas sans dimension. Mais il est nécessaire que les grandeurs présentes dans le tableau soient de dimensions compatibles. C'est particulièrement possible dans des tableaux à double entrée, où toutes les valeurs du tableau proviennent de la même grandeur, mais cela nécessite une grandeur mesurable. Dans l'exemple 3, il est difficile d'imaginer un sens à la somme de matrices ou à la multiplication par un coefficient. Ce n'aurait pas été le cas si le tableau indiquait les précipitations mensuelles (en mm) : on aurait pu donner un sens au produit par une constante (pour changer d'unité par exemple) ou à la somme (en ajoutant les précipitations d'une autre année par exemple).

3.2. **Produit matrice-vecteur.** L'opération la plus importante sur les matrices est la multiplication matricielle. On définit d'abord le produit d'une matrice et d'une matrice-colonne.

4. Tout espace de fonctions à valeurs dans un espace vectoriel hérite de l'ensemble d'arrivée une structure d'espace vectoriel.

Définition 2. Soit m et n deux entiers strictement positifs. Soient $\mathrm{A} = (a_{ij})_{\substack{1\leqslant i\leqslant n\\1\leqslant j\leqslant m}} \in \mathcal{M}_{n,m}(\mathbb{R})$ et $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$. On définit le produit Av comme la matrice colonne à n lignes

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\ \vdots \\ a_{n1}x_1 + a_{12}x_2 + \cdots + a_{nm}x_m \end{pmatrix}.$$

Il est aussi possible de façon similaire définir la multiplication d'une matrice ligne avec une matrice en intervertissant le rôle des lignes et des colonnes.

Une expression avec de nombreuses interprétations. Dans cette définition, chaque coefficient possède la même forme. On considère l'expression similaire $a_1x_1 + a_2x_2 + a_3x_3$. Elle peut s'interpréter des façons suivantes :

— comme la partie linéaire d'une équation linéaire en les variables x_1, x_2 et x_3, ou bien l'expression générale d'une forme linéaire de $\mathbb{R}^3$ dans $\mathbb{R}$.

— comme le produit scalaire des vecteurs (a_1, a_2, a_3) et (x_1, x_2, x_3) de $\mathbb{R}^3$; c'est plus flagrant si l'on renomme les a_i en y_i.

— comme la combinaison linéaire des x_i par les scalaires a_i ; c'est plus flagrant si l'on renomme ces scalaires λ_i.

De plus, les rôles des a_i et x_i sont symétriques, ce qui fournit encore plus d'interprétations possibles.

Un formalisme pour les combinaisons linéaires. De toutes les interprétations possibles de la définition du produit d'une matrice par une matrice colonne, on en présente trois principales.

À chaque ligne de la matrice A, on associe la forme linéaire ϕ_i de $\mathbb{R}^m \to \mathbb{R}$ d'expression générale $a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m$. Le produit s'écrit donc

$$\mathrm{A}v = \begin{pmatrix} \phi_1(x_1, x_2, \ldots, x_m) \\ \phi_2(x_1, x_2, \ldots, x_m) \\ \vdots \\ \phi_n(x_1, x_2, \ldots, x_m) \end{pmatrix}.$$

Cette interprétation est particulièrement adaptée à la représentation d'un système linéaire sous forme de produit matriciel, ou d'une relation de récurrence linéaire pour une suite de matrices colonnes.

De façon symétrique, les x_i peuvent être considérés non plus comme les variables de la forme linéaire, mais comme ses coefficients. On considère donc la forme linéaire $\psi : (y_1, \ldots, y_m) \mapsto x_1y_1 + \cdots + x_my_m$ de $\mathbb{R}^m \to \mathbb{R}$. Le produit s'écrit alors

$$\mathrm{A}v = \begin{pmatrix} \psi(a_{11}, a_{12}, \ldots, a_{1m}) \\ \psi(a_{21}, a_{22}, \ldots, a_{2m}) \\ \vdots \\ \psi(a_{n1}, a_{n2}, \ldots, a_{nm}) \end{pmatrix}.$$

Finalement, les x_i peuvent être considérés comme les coefficients scalaires d'une combinaison linéaire. On peut alors donner une autre interprétation de la définition du produit Av : on a

$$\mathrm{A}v = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_m \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix}.$$

Cette interprétation est celle utilisée dans le calcul de l'image d'un vecteur par une application linéaire.

Exemple 12. On se place dans le cas d'une fonction linéaire f d'un espace vectoriel E de dimension 3 vers un espace vectoriel E$'$ de dimension 2. On choisit une base (e_1, e_2, e_3) de E et une base (e'_1, e'_2) de E$'$.
Soit $v \in$ E. Ce vecteur se décompose comme $x_1e_1 + x_2e_2 + x_3e_3$ dans une base (e_1, e_2, e_2) pour trois scalaires x_1, x_2 et x_3. L'application linéaire f envoie v sur $x_1f(e_1) + x_2f(e_2) + x_3f(e_3)$. Chaque vecteur $f(e_i) \in$ E$'$ se décompose dans la base (e'_1, e'_2) comme $a_{1i}e'_1 + a_{2i}e'_2$, et peut se représenter sous la forme d'une matrice colonne $\begin{pmatrix} a_{1i} \\ a_{2i} \end{pmatrix}$. Ainsi $f(v)$ se représente sous la forme

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} + x_3 \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix}, \qquad \text{c'est-à-dire} \qquad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Plus généralement, si on a deux familles $(u_j)_{1\leqslant j\leqslant m}$ et $(v_i)_{1\leqslant i\leqslant n}$ de vecteurs tels que chaque u_j est une combinaison linéaire de vecteurs de $(v_i)_{1\leqslant i\leqslant n}$, et si on se donne un élément v qui s'exprime comme combinaison linéaire de vecteurs de $(u_j)_{1\leqslant j\leqslant m}$, alors v s'exprime comme combinaison linéaire de vecteurs de $(v_i)_{1\leqslant i\leqslant n}$. Les coefficients de ces différentes combinaisons linéaires sont liés par la relation

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ a_{21} & \cdots & a_{2m} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$$

où les x_j sont les coefficients de la combinaison linéaire des $(u_j)_{1\leqslant j\leqslant m}$ représentant v, les y_i sont les coefficients de la combinaison linéaire des $(v_i)_{1\leqslant i\leqslant n}$ représentant v, et pour tout j, les a_{ij} sont les coefficients de la combinaison linéaire des $(v_i)_{1\leqslant i\leqslant n}$ représentant u_j.

Introduction par les tableaux. On donne un exemple qui permet de montrer que des tableaux dont l'information est contenue par colonne (ou par ligne) permettent d'illustrer différents types d'interprétation pour le produit.

Exemple 13. Un fleuriste propose trois types de bouquets dont les compositions sont indiquées dans le tableau suivant.

	Bouquet 1	Bouquet 2	Bouquet 3
Lys	5	0	15
Roses	8	20	10
Œillets	3	10	10

À la question « combien de fleurs de chaque type obtient-on pour une commande de 10 bouquets de type 1, 3 de type 2 et 2 de type 3 ? »correspond une multiplication par une matrice-colonne qui peut s'interpréter comme combinaison de colonnes.

À la question « sachant qu'un lys coute 4€, une rose 2€ et un œillet 1,50€, quel est le coût de chaque bouquet ? »correspond une multiplication de la matrice par une matrice ligne dont les coefficients encode une forme linéaire.

On note aussi que pour la première question, les coefficients sont des nombres sans unité. Dans la seconde, chaque coefficient possède une unité (€/lys, €/rose, €/œillet).

On remarque que si l'on veut utiliser un tableur pour résoudre l'une des questions, disons la seconde, on peut utiliser la fonction SOMMEPROD. Cependant l'orientation des matrices ne correspond plus à celle utilisée pour le produit :

	A	B	C	D	E	F
1		Bouquet 1	Bouquet 2	Bouquet 3		Prix d'une fleur
2	Lys	5	0	15		4,00 €
3	Roses	8	20	10		2,00 €
4	Œillets	3	10	10		1,50 €
5						
6	Prix			=SOMMEPROD(D2:D4;F2:F4)		

3.3. **Produit matrice-matrice.** On étend la définition de la multiplication pour toute paire de matrices, pour peu que leurs dimensions permettent de calculer le produit.

Définition 3. Une matrice est compatible à une autre si le nombre de ses colonnes est égal au nombre de lignes de l'autre.

Définition 4. Soit $\mathrm{A} = (a_{ij})_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant m}} \in \mathcal{M}_{n,m}(\mathbb{R})$ et $\mathrm{B} = (b_{ij})_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant p}} \in \mathcal{M}_{m,p}(\mathbb{R})$. On définit le produit AB comme la matrice $\mathrm{C} = (c_{ij})_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant p}} \in \mathcal{M}_{n,p}(\mathbb{R})$ telle que pour tout couple d'indices $(i, j) \in [\![1, n]\!] \times [\![1, p]\!]$

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

Là encore, on peut interpréter

— chaque colonne du produit AB comme le produit de A par la colonne correspondante de B,

— chaque ligne de AB comme le produit de la ligne correspondante de A par la matrice B,

— et chaque coefficient c_{ij} de AB comme l'expression $x_1y_1 + \cdots + x_my_m$, où les $x_1, \ldots, x_m$ sont les coefficients de la i^{e} ligne de A et les $y_1, \ldots, y_m$ sont les coefficients de la j^{e} colonne de B.

Cette dernière réécriture a le mérite de comporter moins d'indices que dans la définition, et est donc plus lisible pour un élève de terminale.

3.3.1. *Produit de matrices carrées.* On notera que la relation de compatibilité n'est en général pas symétrique.

Si on se restreint au cas de matrices carrés de même dimension, tout couple de matrice est compatible. De plus, le produit de deux telles matrices est une matrice carrée de même dimension. Cependant, ce produit n'est pas commutatif.

On se place donc dans l'espace $\mathcal{M}_n(\mathbb{R})$ des matrices carrées d'ordre n, stable par multiplication.

La matrice identité I_n est l'élément neutre pour la multiplication. Une matrice M est dite inversible si il existe une autre matrice M′ telle que $\mathrm{MM}' = \mathrm{I}_n$ et $\mathrm{M}'\mathrm{M} = \mathrm{I}_n$. Une telle matrice M′ est alors appelée inverse de M. Si la matrice M est inversible, elle ne possède qu'une seule matrice inverse, notée M^{-1}.

Le déterminant est une application non triviale de $\mathcal{M}_n(\mathbb{R})$ dans $\mathbb{R}$ qui préserve le produit (en fait qui transforme le produit matriciel en produit de nombres réels. On peut en déduire que la matrice identité est de déterminant 1, et que tout matrice inversible possède un déterminant non nul. Réciproquement, une matrice non inversible possède un déterminant nul.

Remarque 1. Toutes les propriétés évoquées ci-dessous on naturellement leur place dans un cours d'algèbre linéaire en dimension finie dans le premier cycle universitaure. Au niveau de la Terminale, l'inversion des matrices carrées (en particulier celles d'ordre au moins 3) reste conceptuellement délicate. Il en est de même de l'existence et la non-nullité de l'application déterminant, sauf pour les matrices d'ordre 2. À toutes fin utiles, nous détaillons le cas d'ordre 2 ci-dessous.

3.3.2. *Cas des matrices d'ordre 2.* Nous revenons sur ces notions dans le cas de matrices carrées d'ordre 2.

Définition 5. On définit le déterminant la matrice $\mathrm{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ par $\det \mathrm{A} = ad - bc$.

Proposition 1. *La matrice* $\mathrm{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *est inversible si et seulement si* $\det \mathrm{A} \neq 0$. *Lorsque* $\det \mathrm{A} \neq 0$, *on a* $\mathrm{A}^{-1} = \frac{1}{\det \mathrm{A}} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Preuve A (matricielle). On pose $\mathrm{B} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. On calcule le produit

$$\mathrm{BA} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = (\det \mathrm{A})\mathrm{I}_2. \tag{2}$$

On suppose que $\det \mathrm{A} \neq 0$, et on veut montrer que A est inversible, et que son inverse est $\frac{1}{\det \mathrm{A}} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Puisque $\det \mathrm{A} \neq 0$, on divise les termes de l'égalité (2) par $\det \mathrm{A}$. On obtient

$$\frac{1}{\det \mathrm{A}} \mathrm{BA} = \mathrm{I}_2.$$

De même, on a

$$\mathrm{A} \times \left(\frac{1}{\det \mathrm{A}} \mathrm{B}\right) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix} = \begin{pmatrix} a\frac{d}{ad-bc} + b\frac{-c}{ad-bc} & a\frac{-b}{ad-bc} - b\frac{a}{ad-bc} \\ c\frac{d}{ad-bc} + d\frac{-c}{ad-bc} & c\frac{-b}{ad-bc} + d\frac{a}{ad-bc} \end{pmatrix} = \mathrm{I}_2.$$

Ainsi A est inversible, d'inverse $\frac{1}{\det \mathrm{A}} \mathrm{B}$.

On suppose à présent que A est inversible et on veut montrer que $\det \mathrm{A} \neq 0$ et que $\mathrm{A}^{-1} = \frac{1}{\det \mathrm{A}} \mathrm{B}$.

Puisque A est inversible, on multiplie à droite les termes de l'égalité (2) par l'inverse de A. On obtient

$$(\mathrm{BA})\mathrm{A}^{-1} = (\det \mathrm{A})\mathrm{I}_2\mathrm{A}^{-1}.$$

Or $(\mathrm{BA})\mathrm{A}^{-1} = \mathrm{B}(\mathrm{AA}^{-1}) = \mathrm{BI}_2 = \mathrm{B}$. On a donc $\mathrm{B} = (\det \mathrm{A})\mathrm{A}^{-1}$.
On montre par l'absurde que $\det \mathrm{A} \neq 0$. En effet, supposons que $\det \mathrm{A} = 0$. Puisque $\mathrm{B} = (\det \mathrm{A})\mathrm{A}^{-1}$, B est la matrice nulle, et ces coefficients d, $-b$, $-c$ et a sont tous nuls, donc A est aussi la matrice nulle, ce qui contredit le fait qu'elle soit inversible.
On en déduit que $\det \mathrm{A} \neq 0$, et en divisant par $\det \mathrm{A}$, on obtient $\frac{1}{\det \mathrm{A}}\mathrm{B} = \mathrm{A}^{-1}$. $\square$

Cette preuve reste dans le cadre matriciel. Cependant, il y a quelques arguments qui sont parfois non explicités, ou admis :

— La nécessité d'établir l'existence d'un inverse à gauche et un inverse à droite.
— L'utilisation de l'associativité du produit matriciel (pour $(\mathrm{BA})\mathrm{A}^{-1} = \mathrm{B}(\mathrm{AA}^{-1})$ par exemple).
— La compatibilité de la multiplication externe et du produit matriciel. Ici, on utilise implicitement l'identité $(\lambda \cdot M) \times M' = \lambda \cdot (M \times M')$. On utilise aussi la compatibilité entre multiplication externe et multiplication entre nombres réels $\lambda \cdot (\lambda' \cdot M) = (\lambda\lambda') \cdot M$, mais cela est beaucoup plus facile à prouver. Si l'on ne veut pas admettre ou passer sous silence cette propriété, il est possible de le prouver ou de faire les calculs sans l'utiliser : on a choisi d'expliciter le calcul de $\mathrm{A} \times (\frac{1}{\det \mathrm{A}}\mathrm{B})$ pour ne pas utiliser l'identité $M \times (\lambda \cdot M') = \lambda \cdot (M \times M')$.

Preuve B (par système linéaire). On cherche une matrice $\mathrm{B} = \begin{pmatrix} x & y \\ z & t \end{pmatrix}$ telle que $\mathrm{AB} = \mathrm{I}_2$. L'expression du produit donne

$$\begin{pmatrix} ax+bz & ay+bt \\ cx+dz & cy+dt \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

On a égalité de matrices si et seulement si les coefficients de même indice sont égaux. Donc, l'égalité $\mathrm{AB} = \mathrm{I}_2$ est réalisée si et seulement si (x, y, z, t) est solution du couple de systèmes

$$(S_1)\begin{cases} ax+bz=1 \\ cx+dz=0 \end{cases} \qquad \text{et} \qquad (S_2)\begin{cases} ay+bt=0 \\ cy+dt=1 \end{cases}$$

On suppose que $ad - bc = 0$ et on veut montrer qu'il n'y a pas de solution.
Procédons par l'absurde et supposons qu'il existe une solution (x, z) de (S_1) et une solution (z, t) de (S_2). En éliminant z dans (S_1), on obtient $(ad - bc)x = d$. En éliminant x dans (S_1), on obtient $(ad - bc)z = -c$. On en déduit que $d = c = 0$. Ainsi l'égalité $cy + dz = 1$ fournit $0 = 1$, ce qui est absurde.
On en déduit que si $ad - bc = 0$ alors il n'y a pas de solution.

On suppose à présent que $ad - bc \neq 0$. Puisque leur déterminant est non nul, les systèmes (S_1) et (S_2) possèdent une unique solution, que l'on peut déterminer :

$$(x, z) = \left(\tfrac{d}{ad-bc}, \tfrac{-c}{ad-bc}\right) \quad \text{et} \quad (y, t) = \left(\tfrac{-b}{ad-bc}, \tfrac{a}{ad-bc}\right).$$

L'unique solution de $\mathrm{AB} = \mathrm{I}_2$ est donc $\mathrm{B} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

On a également

$$\mathrm{BA} = \begin{pmatrix} ax+cy & bx+dy \\ az+ct & bz+dt \end{pmatrix}.$$

En remplaçant (x, y, z, t) par les valeurs obtenues, on obtient $\mathrm{BA} = \mathrm{I}_2$. Puisque $\mathrm{AB} = \mathrm{BA} = \mathrm{I}_2$, B est l'inverse de A.
On en conclut que si $ad - bc \neq 0$, alors A est inversible et possède pour inverse $\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. □

Cette preuve fait venir naturellement la forme de l'inverse, et donne une nouvelle interprétation possible pour la condition $ad - bc \neq 0$. Cependant, le traitement de l'inverse à gauche est aussi peu naturel.
Pour éviter le recours au calcul, on aurait aussi pu remarquer, dans le cas où $ad - bc$ est non nul, que l'équation $\mathrm{BA} = \mathrm{I}_2$ s'écrit aussi sous forme de deux systèmes, l'un en x et y, l'autre en z et t, dont les déterminants sont tous les deux $ad - bc$. Ainsi ces systèmes admettent également une unique solution, ce qui montre qu'il existe une unique matrice B' vérifiant $\mathrm{B}'\mathrm{A} = \mathrm{I}_2$. En considérant le produit $\mathrm{B}'\mathrm{AB}$ et en utilisant l'associativité, on peut déduire que $\mathrm{B}' = \mathrm{B}$.

4. Puissances de matrices carrées

4.1. **Transformations successives.** Si une matrice carrée représente un endomorphisme, ses puissances représenteront encore des endomorphisme du même espace vectoriel. Pour les puissances de (certaines) matrices carrée d'ordre 2, il est possible sur l'interprétation au programme de certaines transformations linéaires du plan. Pour $\theta \in \mathbf{R}$ et $n \in \mathbb{N}$, on aura ainsi

$$R_\theta^n = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}^n = \begin{pmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{pmatrix} = R_{n\theta}. \tag{3}$$

Notons que c'est valable pour $n \in \mathbf{Z}$. Derrière (3) se trouve l'identité de Moivre

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta, \tag{4}$$

utile pour linéariser $\cos^n\theta$. Cette formule est au programme de Terminale, et il s'agit d'une connexion possible avec le chapitre sur les nombres complexes.

4.2. **Application aux suites récurrentes.** L'utilisation des matrices est unificatrice en ce qu'elle permet de formuler dans un même cadre les relations linéaires d'ordre supérieur avec les suites qu'on appelle en Terminale « récurrentes croisées » de premier ordre, par exemple du type

$$\begin{cases} u_{n+1} = au_n + bv_n \\ v_{n+1} = cu_n + dv_n \end{cases}$$

évoqué plus haut.

4.2.1. *Discussion générale sur les suites de la forme* $\mathrm{U}_{n+1} = \mathrm{AU}_n + \mathrm{B}$. Ici $(\mathrm{U}_n)_{n\geqslant 0}$ est une suite de matrices colonnes, $\mathrm{U}_n \in \mathcal{M}_{k,1}(\mathbf{R})$ pour un certain $k \geqslant 1$, tandis que $\mathrm{A} \in \mathcal{M}_k(\mathbf{R})$ et $\mathrm{B} \in \mathcal{M}_{k,1}(\mathbf{R})$. Quand $k = 1$ il s'agit des suites dites arithmético-géométriques. Quand $k = 2$, ces suites interviennent notamment dans les modèles proie-prédateur discret évoqués comme problème possible par les programmes.

Partant de $\mathrm{U}_{n+1} = \mathrm{AU}_n + \mathrm{B}$ il est utile de chercher C tel que la suite auxiliaire (W_n) de terme général $\mathrm{U}_n - \mathrm{C}$ vérifie la relation de récurrence plus simple. C'est notamment possible quand B est dans l'image de $(\mathrm{I}_2 - \mathrm{A})$: en choisissant alors C tel que

$$(\mathrm{I}_2 - \mathrm{A})\mathrm{C} = \mathrm{B}, \tag{5}$$

on trouve que $\mathrm{W}_{n+1} = \mathrm{AW}_n$, et on est ainsi ramené au calcul des puissances de A.

Exemple 14. Soit donnée la relation

$$\begin{cases} u_{n+1} = 2u_n - v_n + 5 \\ v_{n+1} = -u_n + v_n - 2, \end{cases}$$

c'est-à-dire $\mathrm{U}_{n+1} = \mathrm{AU}_n + \mathrm{B}$ avec $\mathrm{A} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ et $\mathrm{B} = \begin{pmatrix} 5 \\ -2 \end{pmatrix}$. Dans ce cas $(\mathrm{I}_2 - \mathrm{A})$ est inversible et $(\mathrm{I}_2 - \mathrm{A}) \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ -2 \end{pmatrix}$, donc

$$\mathrm{U}_n = \mathrm{A}^n \mathrm{U}_0 + (\mathrm{I}_2 - \mathrm{A}^n) \begin{pmatrix} 2 \\ 3 \end{pmatrix}. \tag{6}$$

Qu'il y ait un terme additif ou non dans la relation de récurrence déterminant U_n, ce qui compte pour les propriétés qualitatives de la suite U_n (convergence vers une limite ou non, périodicité...), ce sont les propriétés spectrales de A.

4.2.2. *Rappels sur les suites récurrentes d'ordre supérieur.* Soient $k \geqslant 1$ et une relation de dépendance linéaire sur l'espace des suites réelles

$$u_{n+k} = c_{k-1} u_{n+k-1} + \cdots + c_0 u_n. \tag{7}$$

On dit de l'entier k que c'est l'**ordre** de la relation (7). On peut la reformuler sous la forme matricielle suivante :

$$\begin{pmatrix} u_{n+1} \\ \vdots \\ \vdots \\ u_{n+k} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ c_0 & c_1 & \cdots & c_{k-1} \end{pmatrix} \begin{pmatrix} u_n \\ \vdots \\ \vdots \\ u_{n+k-1} \end{pmatrix}. \tag{7_{mat}}$$

Il s'agit là d'une relation de la forme $\mathrm{U}_{n+1} = \mathrm{CU}_n$, où $\mathrm{C} \in \mathcal{M}_k(\mathbf{R})$ et (U_n) est une suite de vecteurs qui encode (de manière quelque peu redondante si $k \geqslant 2$) la suite (u_n). On appelle **polynôme caractéristique** associé à la relation (7) le polynôme

$$\Pi = X^k - (c_0 + c_1 X + \cdots + c_{k-1} X^{k-1}). \tag{8}$$

Proposition 2. Π *est le polynôme caractéristique de* C.

Notons que le formalisme matriciel n'est pas strictement nécessaire pour traiter le cas d'ordre 2 ; dans un cours consacré aux suites récurrentes d'ordre 2, on peut introduire le polynôme caractéristique *ad hoc.*

Démonstration. Par récurrence sur k. Si $k = 1$, $\Pi(X) = X - k$. Si $k \geqslant 2$, en développant la première ligne de $XI_k - C$ puis en réordonnant les colonnes de la sous-matrice associée à $(1, 2)$

$$\det(XI_k - C) = X \det(XI_{k-1} - C') - c_0(-1)^{k-2} \varepsilon(\sigma),$$

n	F_n	n	F_n	n	F_n	n	F_n
0	0	5	5	10	55	15	610
1	1	6	8	11	89	16	987
2	1	7	13	12	144	17	1597
3	2	8	21	13	233	18	2584
4	3	9	34	14	377	19	4181

TABLE 2. Les premiers nombres de Fibonacci

où C' est la matrice obtenue de manière analogue à C avec $c_1, \dots, c_k$ et σ est une permutation circulaire d'ordre $k-1$, donc de signature $(-1)^{k-2}$. Par hypothèse de récurrence

$$\chi_C(X) = -c_0 + X(X^{k-1}c_{k-1} - \cdots - c_1) = X^k - (c_0 + c_1 X + \cdots + c_{k-1}X^{k-1}). \quad \square$$

Exemple 15 (Suite de Fibonacci)**.** Soit donnée la relation linéaire $u_{n+2} = u_{n+1} + u_n$. La matrice compagne et le polynôme caractéristique sont

$$C = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \qquad \Pi = X^2 - X - 1.$$

Puisque Π est scindé, nous savons que cette matrice est diagonalisable sur $\mathbf{R}$, de valeurs propres $\varphi_\pm$ et espace propres dirigés par $v_\pm$, où

$$\varphi_+ = \frac{1+\sqrt{5}}{2} \qquad \varphi_- = \frac{1-\sqrt{5}}{2} \qquad v_\pm = \begin{pmatrix} 1 \\ \varphi_\pm \end{pmatrix}.$$

Soit $P_\pm$ la matrice représentant la projection sur $\mathbf{R}v_\pm$ parallèlement à $\mathbf{R}v_\mp$. Alors $P_\pm^2 = P_\pm$, et $P_\pm P_\mp = 0$, donc si (U_n) est solution alors

$$C^n = (\varphi_+ P_+ + \varphi_- P_-)^n = \varphi_+^n P_+ + \varphi_-^n P_-.$$

Ceci permet de calculer explicitement u_n connaissant u_0 et u_1. En particulier,

$$\text{si} \quad \begin{cases} F_0 = 0 \\ F_1 = 1 \\ F_{n+2} = F_n + F_{n+1} \quad n \geqslant 0, \end{cases} \qquad \text{alors } F_n = \frac{\varphi_+^n - \varphi_-^n}{\sqrt{5}}.$$

On comparera ce cas aux deux suivants.

Exemple 16. Soit donnée la relation de récurrence linéaire

$$u_{n+2} = 2u_{n+1} - u_n.$$

Le polynôme caractéristique $\Pi(X) = X^2 - 2X + 1$ est scindé sur $\mathbf{R}$, mais possède la racine double 1 ; puisque la matrice compagne C n'est pas scalaire, elle est semblable à la matrice unipotente

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

Un calcul rapide le confirme, de sorte que

$$C^n = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1-n & n \\ -n & n+1 \end{pmatrix},$$

ce qui permet aussi de calculer rapidement le terme général.

Exemple 17. Soit donnée la relation de récurrence linéaire

$$u_{n+2} = 2u_{n+1} - 2u_n.$$

Le polynôme caractéristique est irréductaible dans $\mathbf{R}[X]$ et les deux racines sont complexes, $\rho_\pm = 1 \pm i$. Dans ce dernier cas (la notation de l') exponentielle complexe, un thème particulièrement transversal, rend un service conséquent. En effet, $\rho_+ = \sqrt{2}e^{i\pi/4}$ et $\rho_- = \sqrt{2}e^{-i\pi/4}$. (u_n) sera de la forme

$$u_n = \alpha(\sqrt{2})^n \cos\left(\theta + \frac{n\pi}{4}\right).$$

4.3. **Application aux processus de Markov.**

4.3.1. *Les chaînes de Markov et leurs représentations.*

Définition 6. Un processus de Markov discret (ou chaîne de Markov) est la donnée d'une suite de variables aléatoires $\mathrm{X}_1, \ldots, \mathrm{X}_n, \ldots$ à valeur dans un même ensemble E dit espace d'états telle que

$$\forall x \in \mathrm{E},\ \mathbf{P}(\mathrm{X}_{n+1} = x \mid \mathrm{X}_1, \ldots, X_n) = \mathbf{P}(\mathrm{X}_{n+1} = x \mid \mathrm{X}_n). \tag{9}$$

On dit de plus que le processus est **homogène** si cette quantité ne dépend pas de $\mathbf{N}$. On dit que la chaîne de Markov est **finie** si E est fini.

Dans la suite on travaillera avec des chaînes de Markov homogène finies.

Les processus de Markov sont **très généraux**. Voici des exemples.

(1) Si $(\mathrm{X}_n)_{n\geqslant 0}$ est une suite de variables aléatoires indépendantes, c'est un processus de Markov. Il est homogène si elles sont identiquement distribuées.

(2) Si X_0 est une variable aléatoire à valeurs dans E, puis $\mathrm{X}_{n+1} = f_n(\mathrm{X}_n)$ où f est une fonction, c'est une chaîne de Markov (homogène si f_n est toujours la même fonction f).

(3) Soit (Z_n) une suite de variables i.i.d à valeurs dans $\mathbf{Z}$. La suite

$$\left(\mathrm{X}_n = \sum_{k=0}^{n} \mathrm{Z}_k \text{ modulo } 2\right)_n$$

est un processus de Markov homogène et fini.

Les exemples (1) et (2) sont des caricatures. Ce qu'on va dire est utile quand on se situe quelque part entre les deux : X_n nous donne un peu d'information sur X_{n+1}, mais pas trop. En fait, on peut observer que (3) est une version « à paramètre » qui interpole entre (1) et (2). Le paramètre pertinent est $p = \mathbf{P}(X_0 \text{ est impair})$:

— Si $p = 1$, alors $\mathrm{X}_{n+1} = \mathrm{X}_n$ pour tout n.
— Si $p = 0$, alors $\mathrm{X}_{n+1} = 1 - \mathrm{X}_n$ pour tout n.
— Si $p = 1/2$, alors

$$\mathbf{P}\left(\sum_{k=0}^{n+1} Z_k \text{ est pair } \mid X_n\right) = \mathbf{P}\left(\sum_{k=0}^{n+1} Z_k \text{ est pair}\right)$$
$$\mathbf{P}\left(\sum_{k=0}^{n+1} Z_k \text{ est impair } \mid X_n\right) = \mathbf{P}\left(\sum_{k=0}^{n+1} Z_k \text{ est impair}\right).$$

Donc si $p \in \{0, 1\}$ on est dans le cas (2) et si $p = 1/2$ on est dans le cas de l'exemple (1).

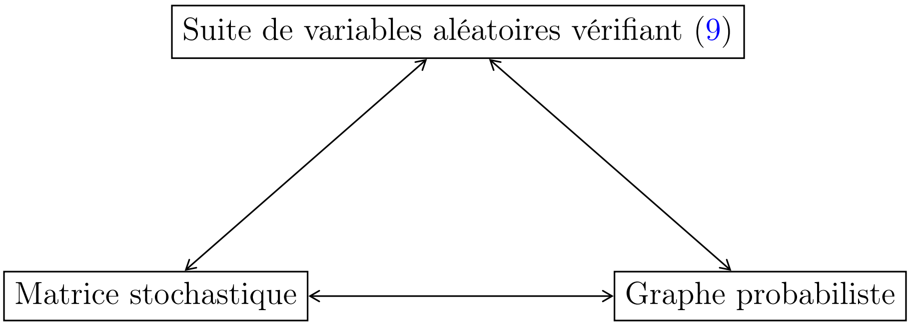


FIGURE 1. Trois modes de représentations d'un même objet.

Définition 7 (Matrice de transition)**.** Soit (X_j) une chaîne de Markov homogène finie d'espace d'états E. La matrice de transition d'une chaîne de Markov homogène finie est la matrice indexée par E dont le coefficient en position (u, v) est

$$\mathbf{P}(X_1 = v \mid X_0 = u). \tag{10}$$

La matrice de transition d'une chaîne de Markov finie a la propriété d'être **stochastique**, c'est à dire que la somme des coefficients situés sur chaque ligne est égale à 1.

Remarque 2. La convention de la Définition 7 n'est pas universelle. Dans certains ouvrages anglo-saxons[5] on trouvera la convention transposée, et les matrices stochastiques ont la somme des coefficients de chacune de leur colonnes égale à 1.

On associe aussi à une chaîne de Markov homogène et finie un **graphe probabiliste** (comme à l'Exemple 10) sur l'espace des états. Les arêtes portent les probabilités de transition, et dans les représentations graphiques de ce graphe, on omet les arêtes de probabilité nulle. Notons que la somme des probabilités portées par les arêtes sortantes est **toujours** 1.

4.3.2. *Irréductibilité.*

Définition 8. Une matrice $\mathrm{A} \in \mathcal{M}_n(\mathbf{C})$ est dite réductible s'il existe $\mathrm{J} \subset [\![1, \dots, n]\!]$ non vide ni totale telle que

$$\forall j \in \mathrm{J}, \forall i \in [\![1, \dots, n]\!] \setminus \mathrm{J}, a_{i,j} = 0. \tag{11}$$

Dans le cas contraire, A est dite **irréductible**, c'est cette dernière propriété sur laquelle nous mettrons l'accent.

En d'autres termes, une matrice est irréductible si parmi tous les $2^n - 2$ sous-espaces engendrés par la base canonique de $\mathbf{C}^n$ différents de 0 et $\mathbf{C}^n$, aucun n'est stable par l'endomorphisme associé.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

réductible irréductible irréductible

On dit qu'une chaîne de Markov est **irréductible** si sa matrice de transition est irréductible.

5. *Linear algebra and Geometry*, Cuoco et al., AMS/MAA textbolks vol. **46**, 2019.

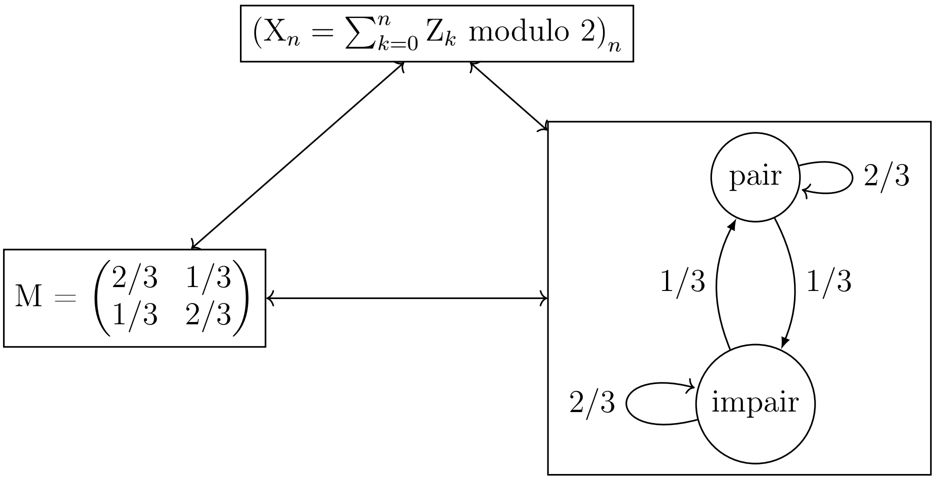


FIGURE 2. Une chaîne de Markov, sa matrice de transition et son graphe probabiliste. Ici (Z_n) est une suite de variable i.i.d., paire avec probabilité 2/3.

Étant donnée une matrice stochastique, il n'est pas toujours évident a priori qu'elle est irréductible. Cela se « voit » un peu mieux sur le graphe. Voici un critère utile pour montrer qu'une matrice est irréductible.

Proposition 3. *Soit* A *une matrice stochastique. S'il existe* $p > 0$ *tel que* A^p *a tous ses coefficients strictements positifs, alors* A *est irréductible.*

Démonstration. Supposons que A est irréductible et soit $J \subsetneq \{1, \ldots, n\}$ non vide telle que $\mathrm{vect}(e_j : j \in J)$ est A-stable. Alors $\mathrm{vect}(e_i : j \in J)$ est encore A^p-stable ; mais alors A^p possède un coefficient nul en position $(i, j) \in \overline{J} \times J$. □

Attention, la proposition ne donne pas de condition nécessaire. On pourra penser à

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

qui est irréductible, mais dont toutes les puissances sont des matrices de permutation (c'est-à-dire I_2 ou A).

Exemple 18. Soit P_σ la matrice de permutation du 3-cycle $\sigma = (1 \quad 2 \quad 3)$, et $\mathrm{A} = \mathrm{P}_\sigma + \mathrm{P}_\sigma^{-1}$. Alors

$$\mathrm{A}^2 = \mathrm{P}_\sigma^2 + \mathrm{P}_\sigma^{-2} + 2\mathrm{I}_2 = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

donc A est irréductible.

4.3.3. *Évolution, probabilité stationnaire et théorème de Perron-Frobenius.* Si (X_n) est une chaîne de Markov finie à valeurs dans E, la loi de X_n peut être encodée par une suite de matrices lignes indexées par E. Cette suite est notée (Π_n).

Proposition 4. Π_n *vérifie l'identité*

$$\Pi_{n+1} = \Pi_n A \tag{12}$$

où A *est la matrice de transition.*

Démonstration. C'est une application de la formule des probabilités totales : pour tout v dans E,

$$\mathbf{P}(\mathrm{X}_{n+1} = v) = \sum_{u \in \mathrm{E}} \mathbf{P}(\mathrm{X}_{n+1} = v \mid \mathrm{X}_n = u)\mathbf{P}(\mathrm{X}_n = u). \qquad \square$$

On appelle **probabilité invariante** un vecteur ligne π tel que $\pi A = \pi$.

Théorème 1 (Corollaire du théorème de Perron et Frobenius)**.** *Toute chaîne de Markov irréductible admet une unique probabilité invariante. De plus, cette probabilité invariante est strictement positive.*

Quelques remarques :
— En fait, l'existence ne requiert pas que A soit irréductible. L'unicité et la stricte positivité, si.
— Si A est la matrice de transition, déterminer la probabilité invariante revient à déterminer le vecteur propre de ${}^t\mathrm{A}$ associé à la valeur propre 1. C'est-à-dire à résoudre un système linéaire.
— Si les coefficients de A sont rationnels, ceux de la probabilité invariante le seront aussi.

4.3.4. *Preuve du théorème de Perron-Frobenius.* Soient $n \geqslant 1$ et $\mathrm{A} \in \mathcal{M}_n(\mathbf{C})$. On appelle **rayon spectral** de A et on note $\rho(\mathrm{A})$ le plus grand module d'une valeur propre de A.

Exemple 19. Soit R_θ la matrice de rotation d'un angle θ défini plus haut. Alors les valeurs propres de R_θ sont $e^{i\theta}$ et $e^{-i\theta}$. Donc $\rho(\mathrm{R}_\theta) = 1$.

Exemple 20. Soit σ une permutation de $[\![1, \dots, n]\!]$ et P_σ sa matrice. Les valeurs propres de P_σ sont des racines de l'unité. Donc $\rho(\mathrm{P}_\sigma) = 1$.

Lemme 1 (Formule de Gelfand)**.** *Soit* $\|\cdot\|$ *une norme d'opérateur sur* $\mathcal{M}_n(\mathbf{C})$*. Alors*

$$\rho(A) = \lim_{k \to +\infty} \|\mathrm{A}^k\|^{1/k}. \tag{13}$$

Nous admettrons ce Lemme.[6] La formule de Gelfand est aussi appelée « formule du rayon spectral » ou théorème de Househölder.

Lemme 2 (Continuité du rayon spectral)**.** *L'application* ρ *est continue sur* $\mathcal{M}_n(\mathbf{C})$.

Démonstration. Considérons l'application $\mathrm{A} \mapsto \chi_\mathrm{A}$ qui à A dans $\mathcal{M}_n(\mathbf{C})$ associe son polynôme caractéristique dans $\mathbf{C}_n[\mathrm{X}]$. Les coefficients de χ sont continus (en fait ils sont même polynomiaux) sur $\mathcal{M}_n(\mathbf{C})$. Pour $\mathrm{P} \in \mathbf{C}_n[t]$ unitaire écrivons

$$\|\mathrm{P}\| = 1 + |a_{n-1}| + \cdots + |a_0|$$

quand $\mathrm{P}(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$ et posons $\overline{\rho}(\mathrm{P})$ le plus grand module d'une racine de P, de sorte que $\rho(\mathrm{A}) = \overline{\rho}(\chi_\mathrm{A})$ pour toute $\mathrm{A} \in \mathcal{M}_n(\mathbf{C})$. Nous sommes ainsi ramenées à prouver la continuité de l'application $\overline{\rho}$. Soit $z \in \mathbf{C}$ une racine de P de module $\overline{\rho}(\mathrm{P})$ et soit $(\mathrm{P}_k)_{k \geqslant 0}$ une suite de $\mathbf{C}_n[t]$ qui converge vers P quand k tend vers l'infini. Donnons-nous $\varepsilon > 0$ et soit m la multiplicité de z ; nous allons montrer que pour k assez grand, P_k possède au moins m racines (comptées avec mutiplicité)

6. Voir par exemple : Mneimné et Testard, *Introduction à la théorie des groupes de Lie classique*, Hermann, exercice 2 page 22.

dans le disque ouvert $\mathrm{D}(z,\varepsilon)$. Supposons par l'absurde que ce n'est pas le cas, et ordonnons pour tout k la liste des racines $(z_{k,i})_{1\leqslant i\leqslant n}$ de P_k de sorte que $|z-z_{k,i}|$ soit une fonction croissante de i. Puisque (P_k) converge vers P, $\|\mathrm{P}_k\|$ est bornée à partir d'un certain rang par $\mathrm{R}=\|\mathrm{P}\|+1$. Si nous posons

$$\mathrm{P}_k(t)=t^n+a_{n-1,k}t^{n-1}+\cdots+a_{1,k}t^1+a_{0,k}$$

alors pour tout $i\in[\![0,n]\!]$, $z_{k,i}^n=-\sum_{j=0}^{n-1}a_{j,k}z_{k,i}^j$ d'où

$$|z_{k,i}|^n\leqslant\left(\sum_{j=0}^{n-1}|a_{j,k}|\right)\max(1,|z_{k,i}|)^{n-1}.$$

On en déduit que pour k assez grand, $|z_{k,i}|\leqslant\|\mathrm{P}_k\|\leqslant\mathrm{R}$. Par compacité, quitte à extraire une sous-suite, $z_{\psi(k),i}$ converge vers $z_{\infty,i}\in\mathrm{D}(0,\mathrm{R})$. Si par l'absurde $|z_{\psi(k),p}-z|\geqslant\varepsilon$ alors $|z_{\infty,p}-z|\geqslant p$. Mais par continuité et puisque $\mathrm{P}_k\to\mathrm{P}$

$$\mathrm{P}(t)=\prod_{i=1}^n(t-z_{\infty,i}),$$

ce qui amène une contradiction avec l'assertion précédente. Donc $\overline{\rho}$ est continue, puis ρ l'est également. □

On dira à partir d'ici qu'une matrice $\mathrm{A}\in\mathcal{M}_n(\mathbf{C})$ est **strictement positive** si tous ses coefficients sont réels et strictement positifs. Nous dirons aussi que $\mathrm{X}\in\mathbf{C}^n$ est positif (resp. strictement positif) si tous ses coefficients sont positifs (resp. le sont strictement). Étant donné $\mathrm{X}\in\mathbf{C}^n$, on note $|\mathrm{X}|$ le vecteur dont les coefficients sont les modules des coefficients de A.

Lemme 3. *Soit $A\in\mathcal{M}_n(\mathbf{C})$ strictement positive. Soit $\mathrm{Y}\in\mathbf{C}^n$ de coordonnées réelles et positives. Alors AY est de coordonnées réelles et positive. De plus, ou bien AY est nul, ou bien il est de coordonnées strictement positives.*

Démonstration. Les sommes et les produits de réels positifs, sont positifs. Si AY est non nul, alors Y est non nul. Posons $\eta=\min_i y_i$ et soit α le plus petit coefficient de A (strictement positif par hypothèse). Alors tous les coefficients de AY sont au moins égaux à $\alpha\eta$. □

Lemme 4. *Soit $\mathrm{A}\in\mathcal{M}_n(\mathbf{C})$ une matrice strictement positive. Soit $\lambda\in\mathbf{C}$ une valeur propre de A telle que $|\lambda|=\rho(A)$. On se donne $\mathrm{X}\in\mathbf{C}^n$ non nul tel que $\mathrm{AX}=\lambda\mathrm{X}$. Alors*

(1) $|\mathrm{X}|$ est un vecteur propre de A pour la valeur propre $\rho(\mathrm{A})$ (en particulier, le rayon spectral est valeur propre),

(2) Il existe $\theta\in\mathbf{R}$ tel que $\mathrm{X}=e^{i\theta}|\mathrm{X}|$.

Démonstration. Considérons le vecteur

$$\mathrm{Y}=\mathrm{A}|\mathrm{X}|-\rho(\mathrm{A})|\mathrm{X}|.$$

D'après l'inégalité triangulaire, Y est positif. Si par l'absurde Y est non nul alors d'après le Lemme 3 il est strictement positif, il existe donc $\rho>\rho(\mathrm{A})$ tel que $(\mathrm{A}-\rho)|\mathrm{X}|$ est positif. Mais alors, pour tout $k\in\mathbf{N}$, $A^k|X|-\rho^k|X|$ est positif, et donc

$$\rho\leqslant\liminf_k\|A^k\|^{1/k}$$

où $\|\cdot\|$ est la norme d'opérateur associée à la norme uniforme sur $\mathbf{C}^n$. Or d'après la formule de Gelfand, le terme de droite est égal à $\rho(\mathrm{A})$, contradiction. Il ya donc égalité dans l'inégalité triangulaire, ce qui n'arrive que so les composantes de X ont toutes un argument en commun. □

Lemme 5 (Théorème de Perron)**.** *Soit* $\mathrm{A} \in \mathcal{M}_n(\mathbf{C})$ *strictement positive. Alors* $\rho(A)$ *est l'unique valeur propre de module maximal, et le sous-espace propre* $\mathrm{E}_{\rho(\mathrm{A})}(\mathrm{A})$ *associé à* $\rho(\mathrm{A})$ *est une droite, dirigée par un vecteur strictement positif.*

Démonstration. [7] Soit λ une valeur propre de module maximal et X un vecteur propre associé. D'après le Lemme 4, X est de la forme $e^{i\theta}|\mathrm{X}|$ et $|\mathrm{X}|$ est un vecteur propre de A pour la valeur propre $\rho(A)$. Donc

$$\lambda\mathrm{X} = \mathrm{AX} = e^{i\theta}\mathrm{A}|\mathrm{X}| = e^{i\theta}\rho(\mathrm{A})|\mathrm{X}| = \rho(A)\mathrm{X}.$$

Puisque X est non-nul, l'égalité précédente implique que $\lambda = \rho(A)$. Soit maintenant X un vecteur propre associé à $\rho(\mathrm{A})$. D'après le Lemme 4, $\rho(\mathrm{A})\mathrm{X} = e^{i\theta}\rho(\mathrm{A})|\mathrm{X}| = e^{i\theta}\mathrm{A}|\mathrm{X}|$. Or $|\mathrm{X}|$ est positif et A strictement positive donc $\mathrm{A}|\mathrm{X}|$ est strictement positif. Donc $\rho(\mathrm{A}) > 0$, X n'a pas de coordonnée nulle. Ceci exclut que $\mathrm{E}_{\rho(\mathrm{A})}(\mathrm{A})$ puisse avoir une dimension au moins 2 ; en effet si c'était le cas, quitte à se donner deux vecteurs propres X et Y linéairement indépendants, $x_1\mathrm{Y} - y_1\mathrm{X}$ aurait au moins une coordonnée nulle, ce qui contredit ce qui précède. □

Lemme 6. *Soit* $\mathrm{A} \in \mathcal{M}_n(\mathbf{C})$ *positive. Alors* $\rho(\mathrm{A})$ *est valeur propre et le sous-espace associé contient un vecteur positif.*

Démonstration. Quand A est strictement positive, c'est le Lemme 5. On va déduire le cas général à l'aide du Lemme 2 de continuité du rayon spectral. Pour tout $k \geqslant 0$, considérons la matrice

$$\mathrm{A}_k = \mathrm{A} + \frac{1}{2^k}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}.$$

D'après le Lemme 5, il existe un vecteur propre X_k de A_k) de coordonnées strictement positives et de norme 1, associé à la valeur propre $\rho(\mathrm{A}_k) > 0$. Quitte à extraire une sous-suite $\mathrm{A}_{\psi(k)}$, on a X de norme 1 et de coordonnées positives tel que

$$\mathrm{AX} = \lim_{k\to+\infty} \mathrm{A}_{\psi(k)}\mathrm{X} = (\lim_{k\to+\infty} \rho(\mathrm{A}_{\psi(k)}))\mathrm{X} = \rho(\mathrm{A})\mathrm{X},$$

où l'on a utilisé le Lemme 2 pour la dernière égalité. □

Lemme 7. *Soit* $\mathrm{A} \in \mathcal{M}_n(\mathbf{C})$ *une matrice positive. Alors* A *est irréductible si et seulement si* $(\mathrm{I}_n + \mathrm{A})^{n-1}$ *est strictement positive.*

Démonstration. Commençons par supposer $(\mathrm{I} + \mathrm{A})^{n-1}$ strictement positive. Si par l'absurde A est réductible, alors il existe $\mathrm{J} \subseteq [\![1, n]\!]$ non vide ni total tel que $\mathrm{Vect}_{j\in\mathrm{J}}(e_j)$ est A stable, donc $(\mathrm{I} + \mathrm{A})$-stable, donc $(\mathrm{I} + \mathrm{A})^{n-1}$-stable. Mais alors la matrice $(\mathrm{I} + \mathrm{A})^{n-1}$ ne pourrait pas être strictement positive. Donc A est irréductible. Réciproquement, supposons que A est irréductible, et montrons que [8] pour tout vecteur $\mathrm{X} \in \mathbf{C}^n$ positif non strictement positif, on a l'inégalité

$$\sharp\{i : e_i^*(\mathrm{I}_n + \mathrm{A})\mathrm{X} = 0\} < \sharp\{i : e_i^*\mathrm{X} = 0\}, \tag{14}$$

7. D'après Mneimné et Mansuy, *Algèbre linéaire*, Vuibert, 2012.
8. Mneimné et Mansuy, *loc. cit*, Lemme XIII.5.8.

autrement dit, $X + AX$ a strictement moins de coefficients nuls que X. Observons déjà que $(I_n + A)X - X$ est positive, de sorte qu'on a l'inégalité large dans (14). Si par l'absurde il y a égalité, posons J_0 l'ensemble des coordonnées nulles de X. Pour tout $i \in J_0$, on a que $\sum_{i,k:i\in J_0,k\notin J_0} a_{i,k}x_k = 0$. Notons que tous les termes de cette somme nulle sont positifs, ils sont donc nuls. Posons $J = [\![1, n]\!] \setminus J_0$. Alors pour tout $k \in J$, on a que pour tout $i \notin J$, $a_{i,k} = 0$ (puisque $x_k \neq 0$). Donc A n'est pas irréductible, absurde. A présent, la quantité de coordonnées nulle calculée dans (14) diminue d'au moins 1 à chaque application de $I_n + A$. Mais alors $(I_n + A)^{n-1}X$ est strictement positif pour tout X positif non nul, en particulier A est strictement positive. □

Théorème 2 (Théorème de Perron-Frobenius)**.** *Soit* $A \in \mathcal{M}_n(\mathbf{C})$ *irréductible et positive. Alors* $\rho(A)$ *est valeur propre de module maximal, et le sous-espace propre associé est une droite, dirigée par un vecteur de coordonnées strictement positives.*

Démonstration. D'après le Lemme 6, $\rho(A)$ est valeur propre, et le sous-espace propre (appelons-le V) associé à la valeur propre $\rho(A)$ contient un vecteur de coordonnées positives (nous allons améliorer cela par la suite). D'après le Lemme 7, $(I_n + A)^{n-1}$ est strictement positive, et d'après le Lemme 5, V est de dimension 1, dirigé par un vecteur de coordonnées strictement positives. □