Calcul Matriciel en L2 Informatique à l'ULCO
Cours et Exercices

Isar Stubbe

(version de 6-9-2023)

Table de matières.

1. Matrices, opérations

On a l'habitude de travailler avec des nombres réels pour exprimer des *quantitiés scalaires*, p.e. distance $d \in \mathbb{R}$, température $T \in \mathbb{R}$, etc. Parfois on doit combiner plusieurs nombres pour former un *vecteur*, p.e. position dans l'espace-temps $p = (x, y, z, t) \in \mathbb{R}^4$. Mais il s'avère qu'aussi des *tableaux de nombres* sont souvent utiles, p.e. un tableau qui reprend les notes de tous les étudiants à tous les cours, ou un tableau dont les éléments correspondent aux pixels d'une image sur écran. C'est ce qu'on appelle une matrice—terme introduite par le mathématicien anglais James Joseph Sylvester (1814–1897):

Définition 1.1 Une matrice réelle A à m lignes et n colonnes (ou de genre $m \times n$) est un tableau rectangulaire de mn nombres réels écrit comme

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = (a_{ij})_{ij} \in \mathbb{R}^{m \times n}.$$

Avec les nombres réels on peut calculer: on connait la somme $a + b$ et le produit ab de deux nombres, puis il y a des règles de calcul tels que $a(b + c) = ab + ac$, $a(bc) = (ab)c$, $a + 0 = a$, etc. Pour les matrices on a des opérations et règles similaires—mais les choses sont un peu plus compliquées.

Définition 1.2 La somme de deux matrices $A, B \in \mathbb{R}^{m\times n}$ est donnée par $A+B = (a_{ij}+b_{ij})_{ij} \in \mathbb{R}^{m\times n}$. La matrice nulle (de genre $m \times n$) est donnée par $O = (0)_{ij} \in \mathbb{R}^{m\times n}$.

Il y a plusieurs matrices nulles – une pour chaque genre $m \times n$ – mais nous les écrivons toutes comme O.

Proposition 1.3 Pour toutes les matrices $A, B, C \in \mathbb{R}^{m\times n}$ on a que:
1. $(A + B) + C = A + (B + C)$,
2. $A + O = A = O + A$,
3. $A + B = B + A$,
4. il existe une (unique) matrice $-A \in \mathbb{R}^{m\times n}$ telle que $A + (-A) = O$.

On voit facilement que les éléments de $-A$ sont exactement les opposés des éléments de A. Ainsi, on dit que la matrice $-A$ est la matrice opposée de A. Par ailleurs, on écrira souvent $A - B$ à la place de $A + (-B)$.

Définition 1.4 Le multiple à gauche, resp. à droite, d'une matrice $A \in \mathbb{R}^{m\times n}$ par un scalaire $\alpha \in \mathbb{R}$ est donnée par $\alpha A = (\alpha a_{ij})_{ij} \in \mathbb{R}^{m\times n}$, resp. $A\alpha = (a_{ij}\alpha)_{ij} \in \mathbb{R}^{m\times n}$.

Proposition 1.5 Pour toutes les matrices $A, B \in \mathbb{R}^{m\times n}$ et tous les scalaires $\alpha, \beta \in \mathbb{R}$ on a que:
1. $\alpha A = A\alpha$,
2. $1A = A$,
3. $(\alpha + \beta)A = \alpha A + \beta A$,
4. $\alpha(A + B) = \alpha A + \alpha B$,
5. $(\alpha\beta)A = \alpha(\beta A)$.

Il est évident que $(-1)A = -A$ (donc, "-1 fois la matrice A est la matrice opposée de A"), et de même manière $0A = O$ ("nulle fois la matrice A est la matrice nulle (de même genre que A)").

Définition 1.6 Le produit d'une matrice $A \in \mathbb{R}^{m\times n}$ avec une matrice $B \in \mathbb{R}^{n\times k}$ est donné par $AB = (\sum_{s=1}^{n} a_{is}b_{sj})_{ij} \in \mathbb{R}^{m\times k}$ (attention aux genres!). La matrice unité (de genre $n \times n$) est donnée par $I = (\delta_{ij})_{ij} \in \mathbb{R}^{n\times n}$ où $\delta_{ij} = 1$ si $i = j$ et 0 sinon.

Il y a plusieurs matrices unité (une pour chaque n); nous écrivons $I_n \in \mathbb{R}^{n\times n}$ si nous voulons indiquer explicitement son genre.

Proposition 1.7 Pour toutes les matrices $A, A' \in \mathbb{R}^{m\times n}$, $B, B' \in \mathbb{R}^{n\times k}$ et $C \in \mathbb{R}^{k\times l}$ on a que:
1. $(AB)C = A(BC)$,
2. $I_nA = A = AI_m$,
3. $(A + A')B = AB + A'B$,
4. $A(B + B') = AB + AB'$,
5. $\alpha(AB) = (\alpha A)B = A(\alpha B)$.

Le produit matriciel est la notion centrale du calcul matriciel; tout ce que nous ferons dans la suite, reposera d'une façon ou l'autre sur le produit matriciel. Par ailleurs, par analogie avec les nombres réels, on définit les *puissances* d'une matrice carrée $A \in \mathbb{R}^{n\times n}$ par récurrence comme suit: $A^0 = I_n$ et $A^k = A \cdot A^{k-1}$ pour $k > 0$.

Définition 1.8 La matrice transposée d'une matrice $A \in \mathbb{R}^{m\times n}$ est donnée par $A^t = (a_{ji})_{ij} \in \mathbb{R}^{n\times m}$ (attention à l'inversion des indices!).

Proposition 1.9 Pour toutes les matrices $A, A' \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{n\times k}$ et tout scalaire $\alpha \in \mathbb{R}$ on a que:

1. $O^t = O$,
2. $I^t = I$,
3. $(A^t)^t = A$,
4. $(A + A')^t = A^t + A'^t$,
5. $(\alpha A)^t = \alpha A^t$
6. $(AB)^t = B^t A^t$.

Les 4 opérations (somme, multiple, produit et transposée) et les 20 règles de calcul énoncées dans les Propositions ci-dessus constituent le Calcul Matriciel: tous les calculs que nous ferons dans ce cours, s'y ramènent. Attention: même si le calcul matriciel ressemble parfois aux calculs avec des nombres réels, il y a des différences importantes! Par exemple, pour des nombres réels a et b, nous avons bien sûr que $ab = ba$; mais pour des matrices il peut arriver que $AB \neq BA$. De même, pour des nombres on sait que $a \neq 0 \neq b$ implique $ab \neq 0$; mais pour des matrices on peut bien avoir $AB = O$ même si $A \neq O \neq B$. Pour $a \in \mathbb{R}$ on sait que, si $a \neq 0$ alors il existe un unique $a^{-1} \in \mathbb{R}$ tel que $aa^{-1} = 1 = a^{-1}a$. Ceci est très utile: si $ax = b$ (pour $a, b \in \mathbb{R}$ donnés) et $a \neq 0$ alors $x = a^{-1}b$. Pour les matrices, la situation est comme suit:

Définition 1.10 Une matrice carrée $A \in \mathbb{R}^{n\times n}$ est inversible s'il existe une (nécessairement unique) matrice $A^{-1} \in \mathbb{R}^{n\times n}$ telle que $AA^{-1} = I_n = A^{-1}A$.

Une définition à l'apparence plus générale est possible: une matrice $A \in \mathbb{R}^{m\times n}$ est inversible s'il existe des matrices $B, C \in \mathbb{R}^{n\times m}$ telles que $AB = I_m$ et $CA = I_n$. Mais avec la théorie du rang des matrices (voir Théorème 7.7 plus loin) on peut montrer que, dans ce cas, nécessairement $m = n$ et $B = C$. Autrement dit, on retrouve la définition donnée ci-dessus. Evidemment toute matrice unité I est inversible, et toute matrice carrée nulle O n'est pas inversible; mais il existe des matrices inversibles différentes de I, ainsi que des matrices carrées non-inversibles différentes de O.

Proposition 1.11 Pour toutes matrices inversibles $A, B \in \mathbb{R}^{n\times n}$ et tout scalaire inversible $\alpha \in \mathbb{R}$,
1. αA est inversible, et $(\alpha A)^{-1} = \alpha^{-1}A^{-1}$,
2. AB est inversible, et $(AB)^{-1} = B^{-1}A^{-1}$,
3. A^t est inversible, et $(A^t)^{-1} = (A^{-1})^t$.

Attention: une somme de matrices inversibles n'est pas nécessairement inversible!

Exercices

1. Calculer, si possible, la somme $A+B$, le produit $A\cdot B$ et le produit $B\cdot A$ des matrices suivantes:
   (a) $A = \begin{pmatrix} 1 & 0 & -1 \\ 1 & -2 & 1 \end{pmatrix}$ et $B = \begin{pmatrix} \frac{2}{3} & 0 & 1 \\ -1 & 3 & 1 \end{pmatrix}$
   (b) $A = \begin{pmatrix} 1 & 0 & -1 \\ 1 & -2 & 1 \end{pmatrix}$ et $B = \begin{pmatrix} \frac{2}{3} & -1 \\ 0 & 3 \\ 1 & 1 \end{pmatrix}$
   (c) $A = \begin{pmatrix} 1 & 0 & -1 \\ 1 & -2 & 1 \\ 0 & -3 & 1 \end{pmatrix}$ et $B = \begin{pmatrix} \frac{2}{3} & -1 & 4 \\ 0 & 3 & 5 \\ 1 & 1 & 1 \end{pmatrix}$
   (d) $A = \begin{pmatrix} 3 & 1 & 0 & -2 \end{pmatrix}$ et $B = \begin{pmatrix} 1 \\ 5 \\ 3 \\ 1 \end{pmatrix}$

2. Pour les matrices $A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ et $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, calculer $A \cdot B$, $B \cdot A$ et $B \cdot B$. Conclusion?

3. Pour A et B des matrices $n \times n$ *triangulaires supérieures*, montrer qu'aussi $A \cdot B$ et $A + B$ sont triangulaires supérieures. Déduire, par transposition, le même résultat pour les matrices *triangulaires inférieures*. Et pour les *matrices diagonales*?

4. Soit $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. Vérifier que $A^2 = A + 2I_3$, en déduire que A est inversible et calculer son inverse.

5. Soit $A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & -1 & 1 \\ 1 & -2 & 0 \end{pmatrix}$. Calculer $A^3 - A$ puis en déduire A^{-1}.

6. On dit qu'une matrice $A \in \mathbb{R}^{n \times n}$ est *nilpotente* s'il existe $k \in \mathbb{N}$ tel que $A^k = O$.
   (a) Donner un exemple d'une matrice nilpotente non-nulle (et spécifier k).
   (b) Montrer qu'aucune matrice nilpotente est inversible.

7. Pour les matrices
$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 3 & 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{et } C = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & -1 & -1 \end{pmatrix},$$
   calculer AB et AC; A peut-elle être inversible?

8. La *complexité arithmétique* d'un calcul sur ordinateur est le nombre de sommes et de produits que l'on doit faire pour exécuter ce calcul. Une somme de k termes (ou un produit de k facteurs) compte pour $k - 1$ sommes consécutives de deux termes (ou $k - 1$ produits de deux facteurs). Pour simplifier, on admettera ici qu'une différence est une somme, et qu'une division est un produit. Considérons maintenant le produit $AB = C$ de matrices 2×2.

   (a) Donner la complexité arithmétique pour la formule habituelle.

   (b) Observer qu'on peut aussi poser
$$\text{d'abord} \begin{cases} m_1 := (a_{11} + a_{22})(b_{11} + b_{22}) \\ m_2 := (a_{21} + a_{22})b_{11} \\ m_3 := a_{11}(b_{12} - b_{22}) \\ m_4 := a_{22}(b_{21} - b_{11}) \\ m_5 := (a_{11} + a_{12})b_{22} \\ m_6 := (a_{21} - a_{11})(b_{11} + b_{12}) \\ m_7 := (a_{12} - a_{22})(b_{21} + b_{22}) \end{cases} \quad \text{puis} \begin{cases} c_{11} = m_1 + m_4 - m_5 + m_7 \\ c_{12} = m_3 + m_5 \\ c_{21} = m_2 + m_4 \\ c_{22} = m_1 - m_2 + m_3 + m_6 \end{cases}$$

   C'est *l'algorithme de Strassen*. Quelle est sa complexité arithmétique?

   Puisqu'on estime que le calcul d'un produit est (beaucoup) plus "coûteux" que le calcul d'une somme, on peut (considérablement) accélérer le calcul de AB par l'algorithme de Strassen—qui par ailleurs peut être généralisé aux matrices $n \times n$. C'est un sujet de recherche d'optimiser ainsi le calcul du produit matriciel, et à ce jour on ne sait pas s'il existe un algorithme optimal pour des matrices $n \times n$.

9. Dans cet exercice, on identifie l'ensemble $\mathbb{R}^n$ des n-uples avec l'ensemble $\mathbb{R}^{n \times 1}$ des colonnes à n éléments. Ainsi, toute matrice $A \in \mathbb{R}^{m \times n}$ définit une fonction $f_A : \mathbb{R}^n \to \mathbb{R}^m : X \mapsto AX$.

(a) Montrer que, pour tout $X, Y \in \mathbb{R}^n$ et tout $\alpha \in \mathbb{R}$, on a $f_A(X+Y) = f_A(X) + f_A(Y)$ et $f_A(\alpha X) = \alpha f_A(X)$. On dit que f_A est une *transformation linéaire.*

(b) Montrer que, pour toute transformation linéaire $f \colon \mathbb{R}^n \to \mathbb{R}^m$, il existe une unique matrice A telle que $f = f_A$. Indication: puisque

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

on peut appliquer f aux deux côtés et retrouver ainsi une formule matricielle pour f. On appelle A la *matrice représentative de f par rapport aux bases canoniques.*

Ceci établit une bijection entre d'un côté les matrices réelles de genre $m \times n$, et de l'autre les applications linéaires de $\mathbb{R}^n$ à $\mathbb{R}^m$ (munis de leurs bases canoniques). En mathématiques pures, le deuxième point de vue l'emporte sur le premier (pour des raisons que nous ne pouvons développer ici); mais en mathématiques appliquées et en informatique, c'est souvent le calcul matriciel – direct et sans détours – qui est priviligié. C'est aussi notre choix dans ce cours à volume horaire réduit.

2. Résolution de systèmes linéaires

Une équation linéaire à un inconnu est donné par $ax = b$; ici a et b sont des nombres connus, et on veut calculer toutes les valeurs possibles pour l'inconnu x. Une équation linéaire à deux inconnus est du type $ax + by = c$; et s'il y a n inconnus on écrit plutôt $a_1x_1 + \cdots + a_nx_n = b$. Un système de m équations linéaires en n inconnus est donné par

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

Ici, on suppose connu tous les coefficients a_{ij} et tous les termes constants b_i, et on veut calculer toutes les valeurs possibles pour les x_j. On y reconnait facilement le produit matriciel

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

ou encore, si $A \in \mathbb{R}^{m \times n}$ désigne la matrice des coefficients, $X \in \mathbb{R}^{n \times 1}$ la colonnes des inconnus et $B \in \mathbb{R}^{m \times 1}$ la colonnes des constants, tout simplement

$$AX = B.$$

Etant donné une matrice A et une colonne B ayant le même nombre de lignes que A, nous voulons donc calculer toutes les colonnes X telles que $AX = B$.

Exemple 2.1 On peut aussi s'intéresser à un problème d'apparence plus difficile: étant donné des matrices $A \in \mathbb{R}^{m \times n}$ et $B \in \mathbb{R}^{m \times k}$, trouver toutes les matrices $X \in \mathbb{R}^{n \times k}$ telles que $AX = B$. (Autrement dit, les matrices X et B ne sont pas nécessairement des simples colonnes.)

Pourtant, par la nature même du produit matriciel, et notant les matrices $X = (X_1|\dots|X_k)$ et $B = (B_1|\dots|B_k)$ par leurs colonnes, on vérifie facilement que

$$AX = B \iff \begin{cases} AX_1 = B_1 \\ \quad\vdots \\ AX_k = B_k \end{cases}$$

Autrement dit, pour trouver les k colonnes d'une matrice X satisfaisant à $AX = B$, il faut et il suffit de résoudre les k systèmes linéaires $AX_i = B_i$.

Exemple 2.2 Par ailleurs, il est tout aussi naturel de s'intéresser au problème $XA = B$, où $A \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{k\times n}$ sont donnés et on cherche toutes les $X \in \mathbb{R}^{k\times m}$ qui satisfont à cette équation matricielle. Mais par transposition on sait que $XA = B$ si et seulement si $A^tX^t = B^t$. On se ramène ainsi à la situation de l'exemple précédent, et donc à la résolution de systèmes linéaires.

Dans la suite, nous allons donc nous intéresser spécifiquement à la résolution de systèmes linéaires.

Exemple 2.3 Si $A \in \mathbb{R}^{n\times n}$ est une matrice inversible dont on connait la matrice inverse, alors le système linéaire $AX = B$ a comme unique solution $X = A^{-1}B$.

Mais nous cherchons à résoudre des systèmes linéaires quelconques (et par ailleurs il peut être difficile de déterminer si A est inversible, et de calculer l'inverse de A). Cependant, nous observons que certains systèmes sont très faciles à résoudre.

Exemple 2.4 Si $A \in \mathbb{R}^{m\times n}$ est une matrice échelonnée – c'est à dire que toute ligne de A débute avec strictement plus de nuls que la ligne précédente, sauf si la ligne précédente est nulle, auquel cas toute ligne suivante est aussi nulle – alors on résout le système linéaire $AX = B$ facilement par substitution ("du bas vers le haut"), en introduisant des paramètres au besoin.

La stratégie pour résoudre un système $AX = B$ quelconque consiste à le transformer en un système $A'X = B'$, avec une matrice de coefficients échelonnée, ayant exactement les mêmes solutions.

Proposition 2.5 Toute matrice $A \in \mathbb{R}^{m\times n}$ peut être transformée en une matrice échelonnée en appliquant une suite finie d'opérations élémentaires:
- des permutations de lignes: $L_i \leftrightarrow L_j$,
- des réductions inférieures de lignes: $L_i := L_i + \alpha L_j$ avec $j < i$.

Démonstration. Décrivons un algorithme pour échelonner $A = (a_{ij})_{ij}$. Si la première colonne de A n'est pas nulle, alors par permutation de lignes on peut supposer que $a_{11} \neq 0$, puis par réductions inférieures $L_i = L_i + \frac{-a_{i1}}{a_{11}}L_1$ $(1 < i \leq m)$ on annule tous les autres éléments de la première colonne; si la première colonne de A est nulle, alors on ne fait rien du tout. On répète avec la sous-matrice obtenue en supprimant première ligne et première colonne de A. Après m itérations on aura échelonné la matrice. □

Cet algorithme était connu en Chine déjà en -200: dans 'Les Neuf Chapitres sur l'Art Mathématique', un livre (anonyme) détaillant la résolution de 246 problèmes mathématiques, le huitième chapitre sur 'La disposition rectangulaire' est dédié à la résolution de systèmes linéaires 5×5 avec la méthode décrite ici. En Europe c'est d'abord Sir Isaac Newton (1643-1727) puis Carl Friedrich Gauss (1777-1835) qui décrivent et developpent cette methode en toute généralité.

Le résultat suivant dit que les opérations élémentaires sont des “transformations naturelles”: elles sont compatibles d’une façon assez particulière avec le produit matriciel.

Lemme 2.6 Soient des matrices $A \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{n\times k}$. Effectuer une opération élémentaire sur A puis faire le produit avec B donne le même résultat que d’effectuer cette même opération élémentaire sur le produit AB.

Démonstration. Notons les lignes de A par L_1, ..., L_m et les colonnes de B par C_1, ..., C_k; la i-ième ligne du produit AB est donc $(L_iC_1 \cdots L_iC_k)$. Il est clair qu’une permutation de lignes $L_i \leftrightarrow L_j$ dans A aura comme effet la même permutation de lignes dans AB. Et une réduction (inférieure) de lignes $L_i := L_i + \alpha L_j$ dans A aura comme effet la même réduction (inférieure) de lignes dans AB. $\square$

Corollaire 2.7 Pour effectuer une opération élémentaire sur les lignes d’une matrice $A \in \mathbb{R}^{m\times n}$, on peut effectuer cette même opération élémentaire sur les lignes de la matrice unite I_m pour obtenir une matrice (dite élémentaire) E, puis faire le produit EA.

Toute opération élémentaire peut être “défaite”: si on permute deux lignes $L_i \leftrightarrow L_j$ dans une matrice A, alors en refaisant la même permutation on retrouve la matrice de départ; et si on fait une réduction (inférieure) $L_i := L_i + \alpha L_j$ dans A, alors en faisant ensuite $L_i := L_i - \alpha L_j$ on retrouve la matrice originale. Ainsi on voit que:

Proposition 2.8 Toute matrice élémentaire $E \in \mathbb{R}^{m\times m}$ (donc une matrice obtenue en effectuant une seule opération élémentaire sur I_m) est inversible, et

- l’inverse d’une matrice de permutation est une matrice de permutation (en fait, c’est la même matrice),
- l’inverse d’une matrice de réduction inférieure est une matrice de réduction inférieure (en fait, on l’obtient en changeant le signe de l’élément hors diagonale).

Corollaire 2.9 Pour toute matrice $A \in \mathbb{R}^{m\times n}$ il existe une suite finie de matrices élémentaires (donc inversibles) $E_1, \ldots, E_k \in \mathbb{R}^{m\times m}$ telle que $A' = E_k \ldots E_1 A$ est échelonnée.

Avec le lemme suivant on obtiendra enfin une méthode simple pour la résolution de systèmes.

Lemme 2.10 Si $M \in \mathbb{R}^{n\times n}$ est une matrice inversible, alors pour tout $U, V \in \mathbb{R}^{n\times k}$ on a $MU = MV$ si et seulement si $U = V$.

Démonstration. Pour tout M, $U = V$ implique $MU = MV$. Lorsque M est inversible, $MU = MV$ implique aussi $U = M^{-1}MU = M^{-1}MV = V$. $\square$

Proposition 2.11 Soit un système linéaire $AX = B$. Si $E_1, \ldots, E_k \in \mathbb{R}^{m\times m}$ sont des matrices élémentaires telles que $A' = E_k \ldots E_1 A$ est échelonnée, et on pose $B' = E_k \ldots E_1 B$, alors $AX = B$ si et seulement si $A'X = B'$.

Dans la pratique, pour un système linéaire $AX = B$, on “augmente” la matrice des coéfficients A par la colonne des constants B, pour obtenir la matrice $(A|B)$. Ensuite on échelonne la matrice A, tout en appliquant exactement les mêmes opérations élémentaires sur la colonne B. On obtient ainsi $(A'|B')$, d’où le système échelonné $A'X = B'$ à résoudre par substitution.

Exercices

1. Résoudre les systèmes linéaires suivants:

(a) $\begin{cases} x+y+2z=9 \\ 2x+4y-3z=1 \\ 3x+6y-5z=0 \end{cases}$ (d) $\begin{cases} x+y+2z=4 \\ x+z=-2 \\ 2x+y+3z=1 \end{cases}$

(b) $\begin{cases} x+y+2z=3 \\ x+z=-2 \\ 2x+y+3z=1 \end{cases}$ (e) $\begin{cases} x+y+2z=9 \\ 2x+4y-3z=1 \end{cases}$

(c) $\begin{cases} x_1+x_3-x_5=-2 \end{cases}$ (f) $\begin{cases} x_2+x_3-x_5=-2 \\ x_3-x_6=0 \end{cases}$

2. Pour $A = \begin{pmatrix} 2 & 1 & 2 \\ 2 & 2 & -2 \\ 3 & 1 & 1 \end{pmatrix}$ et $X \in \mathbb{R}^{3\times 1}$, résoudre l'équation $AX = 4X$.

3. Trouver toutes les matrices X telles que $\begin{pmatrix} -2 & 1 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ -2 & 3 & 7 & 3 \end{pmatrix} X = \begin{pmatrix} 2 & 2 \\ -1 & 1 \\ 0 & 1 \end{pmatrix}$.

4. Trouver toutes les matrices X telles que $X \begin{pmatrix} -2 & 1 & 3 \\ 0 & 1 & 0 \\ -2 & 3 & 7 \end{pmatrix} = \begin{pmatrix} 2 & 2 & 3 \\ -1 & 0 & -2 \end{pmatrix}$.

5. Résoudre le système $\begin{cases} x+2y=6 \\ 2x-y=1 \end{cases}$ puis tracer les droites $x+2y=6$ et $2x-y=1$ dans le plan. Constat?

3. Factorisation $PA = LU$

D'abord un constat:

Proposition 3.1 Toute matrice élémentaire de réduction inférieure est triangulaire inférieure et inversible; et il en est de même pour son inverse.

Ainsi, si une matrice $A \in \mathbb{R}^{m\times n}$ peut être échelonnée, soit $A' = E_k \dots E_1 A$, uniquement par des réductions inférieures de lignes (donc sans permutations de lignes), alors la matrice $(E_k \dots E_1)^{-1}$ est aussi triangulaire inférieure et inversible (et facile à calculer!), et la matrice échelonnée A' est triangulaire supérieure (au sens large: A n'est pas nécessairement carré). On obtient ainsi une factorisation $A = LU$ de la matrice A en un produit d'une matrice triangulaire inférieure ("lower") $L = (E_k \dots E_1)^{-1}$ avec une matrice triangulaire supérieure ("upper") $U = A'$. Mais, pour échelonner une matrice, on fait au besoin des réductions inférieures et des permutations; ainsi, en général on "mélange" reductions et permutations. Cependant, ces permutations n'agissent jamais sur les lignes que l'on a déjà reduites auparavant. Sous cette contrainte, il est possible de faire d'abord toutes les permutations, et ensuite toutes les réductions inférieures, à condition d'éventuellement modifier ces dernières.

Proposition 3.2 Soit P la matrice élémentaire de permutation $L_i \leftrightarrow L_j$, et R la matrice élémentaire de réduction inférieure $L_k := L_k + \alpha L_l$, et supposons que $i, j > l$. Il existe alors une matrice élémentaire R' de réduction inférieure $L_s := L_s + \alpha L_t$ telle que $PR = R'P$.

Démonstration. On suppose que l'on fait d'abord la réduction inférieure $L_k := L_k + \alpha L_l$ et ensuite la permutation $L_i \leftrightarrow L_j$, avec $i, j > l$. Autrement dit, la ligne L_l se trouve au-dessus des lignes L_i, L_j et L_k, et n'est pas affectée par la permutation ou la réduction inférieure. On obtient le même effet si on fait d'abord la permutation $L_i \leftrightarrow L_j$ et ensuite,

- si $L_i \neq L_k \neq L_j$: la réduction inférieure $L_k := L_k + \alpha L_l$,
- si $L_k = L_i$: la réduction inférieure $L_j := L_j + \alpha L_l$,
- si $L_k = L_j$: la réduction inférieure $L_i := L_i + \alpha L_l$.

En associant à chaque opération élémentaire sa matrice élémentaire, on obtient le résultat. $\square$

Exemple 3.3 Voici un exemple dans $\mathbb{R}^{4\times 4}$. Supposons que l'on fait d'abord la réduction inférieure $L_3 := L_3+\alpha L_1$ puis la permutation $L_2 \leftrightarrow L_3$; le schéma ci-dessous indique qu'il est équivalent de faire d'abord la permutation $L_2 \leftrightarrow L_3$ et ensuite la réduction inférieure $L_2 := L_2+\alpha L_1$:

$$\begin{tikzcd}
\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix} \arrow[r, "L_3:=L_3+\alpha L_1"] \arrow[d, "L_2\leftrightarrow L_3"] & \begin{pmatrix}1&0&0&0\\0&1&0&0\\\alpha&0&1&0\\0&0&0&1\end{pmatrix} \arrow[d, "L_2\leftrightarrow L_3"] \\
\begin{pmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{pmatrix} \arrow[r, "L_2:=L_2+\alpha L_1"] & \begin{pmatrix}1&0&0&0\\\alpha&0&1&0\\0&1&0&0\\0&0&0&1\end{pmatrix}
\end{tikzcd}$$

La raison pour la condition sur les indices dans l'énoncé ($i, j > l$) est que, par exemple, faire d'abord $L_2 := L_2 + \alpha L_1$ et ensuite $L_1 \leftrightarrow L_2$, est bien équivalent à faire d'abord $L_1 \leftrightarrow L_2$ et ensuite $L_1 := L_1 + \alpha L_2$... mais cette dernière opération n'est pas une réduction *inférieure*!

Corollaire 3.4 Pour toute matrice $A \in \mathbb{R}^{m\times n}$ il existe une matrice $P \in \mathbb{R}^{m\times m}$ de permutations de lignes de A (donc un produit de matrices élémentaires de permutation), une matrice triangulaire inférieure $L \in \mathbb{R}^{n\times n}$ et une matrice triangulaire supérieure $U \in \mathbb{R}^{m\times n}$ telles que $PA = LU$.

Esquisse de démonstration. Pour éviter les formalités très lourdes d'une preuve complète, supposons que $A \in \mathbb{R}^{m\times n}$ est échelonnée par une suite de 6 opérations élémentaires, dont 2 permutations P_1 et P_2 et 4 réductions inférieures R_1, ..., R_4, comme suit:

$$A' = R_4P_2R_3R_2P_1R_1A.$$

Par le résultat précédent, on peut "glisser" les permutations vers la droite, en modifiant au besoin les réductions:

$$\begin{aligned}
A' &= R_4P_2R_3R_2P_1R_1A \\
&= R_4P_2R_3R_2R_1'P_1A \\
&= R_4R_3'P_2R_2R_1'P_1A \\
&= R_4R_3'R_2'P_2R_1'P_1A \\
&= R_4R_3'R_2'R_1''P_2P_1A
\end{aligned}$$

Les réductions inférieures étant inversibles (et triangulaires inférieures), et leurs inverses étant toujours des réductions inférieures (et donc toujours triangulaires inférieures), il suit que

$$(P_2P_1)A = (R_4R_3'R_2'R_1'')^{-1}A'$$

exprime la matrice A, à permutations de ses lignes près, comme un produit d'une matrice triangulaire inférieure avec une matrice échelonnée (donc triangulaire supérieure au sens large). On pose donc $P = P_2P_1$, $L = (R_4R_3'R_2'R_1'')^{-1}$ et $U = A'$ pour obtenir le résultat. $\square$

Dans la pratique, étant donné une matrice $A \in \mathbb{R}^{m\times n}$ quelconque, on procède ainsi:

1. on échelonne A: $A' = E_k \dots E_1 A$,
2. parmi les matrices élémentaires $E_k, ..., E_1$ on repère les permutations P_s, ..., P_1 et on pose $P = P_s \dots P_1$,
3. pour échelonner PA on n'aura besoin que de réductions inférieures: $A'' = R_t \dots R_1 PA$,
4. on inverse ces réductions (ce qui est facile!) puis on pose $L = R_1^{-1} \dots R_t^{-1}$ et $U = A''$.

Exercices

1. Montrer qu'une matrice $A \in \mathbb{R}^{n\times n}$ est inversible si et seulement si la matrice triangulaire supérieure U, obtenue par factorisation $PA = LU$, est inversible.

2. Ecrire toutes les matrices élémentaires 2×2 et 3×3, et leurs inverses.

3. Echelonner les matrices suivantes, puis en déduire une factorisation $PA = LU$:

(a) $\begin{pmatrix} 1 & 5 \\ -3 & 2 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & -2 & 3 \\ 4 & 0 & 6 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & -1 & 0 \\ 2 & 1 & 3 \\ -1 & 0 & -1 \end{pmatrix}$

(d) $\begin{pmatrix} 2 & 0 & 1 \\ 1 & -1 & 2 \\ -1 & 1 & 0 \end{pmatrix}$

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

(f) $\begin{pmatrix} -1 & 6 & 1 & 0 \\ 0 & 2 & 3 & 0 \\ 1 & 1 & -1 & 1 \\ 3 & -2 & 1 & 2 \end{pmatrix}$

(g) $\begin{pmatrix} 1 & -1 & 0 & 2 \\ -1 & 1 & 1 & 1 \\ 3 & -1 & -2 & 7 \\ 1 & 0 & -1 & 4 \end{pmatrix}$

4. Forme réduite, matrice inverse

Jusqu'à présent nous avons échelonné une matrice $A \in \mathbb{R}^{m\times n}$ en utilisant

- des permutations de lignes: $L_i \leftrightarrow L_j$,
- des réductions inférieures de lignes: $L_i := L_i + \alpha L_j$ avec $j < i$.

Si on se permet d'utiliser aussi

- des réductions supérieures de lignes: $L_i := L_i + \alpha L_j$ avec $j > i$,
- des multiples non-nuls de lignes: $L_i := \alpha L_i$ avec $\alpha \neq 0$,

alors on peut calculer la *forme réduite* (dite de Gauss-Jordan[1]) de A: c'est une matrice échelonnée avec tous les pivots égaux à 1, et au-dessus (et en-dessous) des pivots tous les éléments sont 0. Et comme pour l'échelonnement, il est toujours vrai que, pour effectuer une telle opération élémentaire sur les lignes d'une matrice $A \in \mathbb{R}^{m\times n}$, on peut effectuer cette même opération élémentaire sur les lignes de la matrice unite I_m pour obtenir une matrice (dite élémentaire) E, puis faire le produit EA. De plus, ces matrices élémentaires sont toujours inversibles, et les inverses toujours très faciles à calculer. Bref, on peut conclure:

Proposition 4.1 Pour toute matrice $A \in \mathbb{R}^{m\times n}$ il existe une suite finie de matrices élémentaires (on permet donc les 4 types d'opérations élémentaires mentionnées ci-dessus) $E_1, \dots, E_k \in \mathbb{R}^{m\times m}$ telle que $A' = E_k \dots E_1 A$ est réduite.

Mieux encore:

Proposition 4.2 La forme réduite d'une matrice $A \in \mathbb{R}^{m\times n}$ est unique (contrairement à une forme échelonnée de A).

[1]Wilhelm Jordan (1842-1899).

Démonstration. On fait une induction sur n. Si $n = 1$ alors le résultat est trivial: la forme réduite d'une simple colonne A est soit nulle (si A est nulle), soit contient un pivot (égal à 1) en première position (si A n'est pas nulle). Pour $n > 1$, supposons que le résultat est vrai pour toute matrice à $n - 1$ colonnes. Si B et C sont deux formes d'une même matrice A (à n colonnes), alors par les mêmes opérations élémentaires on obtient des sous-matrices $B^{(1)}$ et $C^{(1)}$ en formes réduites associées à la matrice $A^{(1)}$ obtenue en supprimant la n-ième colonne; et par hypothèse $B^{(1)} = C^{(1)}$, donc B et C ne peuvent différer que dans leurs dernières colonnes, B_n et C_n. On va montrer que ceci est impossible. Toute solution $(x_1, \ldots, x_n)$ aux sytèmes équivalents

$$AX = O \iff BX = O \iff CX = O$$

est aussi une solution au système

$$(B - C)X = O.$$

Sous l'hypothèse que $B_n \neq C_n$, seule la dernière colonne de $B - C$ est non-nulle, et donc on a nécessairement que $x_n = 0$. Mais pour que toute solution à $BX = O$, resp. $CX = O$, ait $x_n = 0$, il faut un pivot dans B_n, resp. C_n (sinon x_n serait une variable libre). Les $n - 1$ premières colonnes de B et C étant identiques, ce pivot dans leurs n-ième colonnes est nécessairement en même position—et donc $B_n = C_n$. Ceci est en contradiction avec l'hypothèse $B_n \neq C_n$, qui est donc impossible. □

Exemple 4.3 La démonstration précédente étant quelque peu technique, nous illustrons le principe sur un exemple (avec un argument légèrement différent). Supposons que $A \in \mathbb{R}^{3\times 3}$ admet deux formes réduites différentes, p.e.

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

Il existe donc des matrices inversibles R et S telles que $B = RA$ et $C = SA$, ou encore, il existe une matrice (inversible) $T = RS^{-1}$ telle que $B = TC$. Exprimé colonne par colonne, cela voudrait dire que

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \overset{(a)}{=} T \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \overset{(b)}{=} T \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad , \quad \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \overset{(c)}{=} T \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Lorsqu'on regarde la troisième colonne, on en déduit facilement une contradiction:

$$\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \overset{(c)}{=} T \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = T \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right) = T \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + T \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \overset{(a,b)}{=} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Proposition 4.4 Une matrice $A \in \mathbb{R}^{n\times n}$ est inversible si et seulement si sa forme réduite est I_n.

Démonstration. Si la forme réduite de A est I_n, alors on sait que $I_n = E_k \ldots E_1 A$ pour des matrices élémentaires (et donc inversibles) E_i, d'où l'inversibilité de $A = (E_k \ldots E_1)^{-1}$ (dont l'inverse est $E_k \ldots E_1$). Si la forme réduite de A n'est pas I_n, alors c'est une matrice A' avec au plus $n - 1$ pivots (et sa dernière ligne est nulle). Le système $A'X = O$ admet donc au moins une solution non-nulle (à cause d'au moins une variable libre), qui est nécessairement aussi solution au système $AX = O$, et ainsi A ne peut pas être inversible. □

Dans la pratique, pour vérifier si une matrice $A \in \mathbb{R}^{n\times n}$ est inversible, et pour calculer la matrice inverse A^{-1} si elle existe, on effectue des opérations élémentaires sur toute la matrice augmentée $(A \mid I_n)$ dans le but de calculer la forme réduite de A. Si cette forme réduite est I_n, alors il suit que A est inversible, et on aura transformé la partie I_n en A^{-1}; sinon A n'est pas inversible.

Exercices

1. Donner une factorisation $PA = LU$ et la forme réduite des matrices suivantes:

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

2. Calculer la forme réduite et – si possible – l'inverse des matrices suivantes:

(a) $\begin{pmatrix} -1 & 3 \\ 8 & -2 \end{pmatrix}$

(b) $\begin{pmatrix} 6 & 2 \\ 9 & 3 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & -2 & 1 \\ 2 & 3 & 6 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 0 & 0 \end{pmatrix}$

(e) $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{pmatrix}$

(f) $\begin{pmatrix} 1 & 6 & 4 & -1 \\ 2 & 4 & -1 & 3 \\ -1 & 2 & 5 & 0 \end{pmatrix}$

3. Ecrire chaque système linéaire ci-dessous comme une équation matricielle, identifier les matrices inversibles parmi les matrices de coefficients, et le cas échéant calculer l'unique solution du système par un calcul de matrice inverse:

(a) $\begin{cases} x - 2y = 0 \\ 3x + 2y = 11 \end{cases}$

(b) $\begin{cases} x - y + 3z = 0 \\ x + y = 0 \end{cases}$

(c) $\begin{cases} x + y + 2z = 9 \\ 2x + 4y - 3z = 1 \\ 3x + 6y - 5z = 0 \end{cases}$

(d) $\begin{cases} x + y + 2z = 4 \\ x + z = -2 \\ 2x + y + 3z = 1 \end{cases}$

4. Soit une matrice inversible A de genre $n \times n$, et B son inverse. Si Alice dispose de la matrice A, et Bob de la matrice B, alors Alice peut envoyer des messages codés à Bob. On numérote les lettres de l'alphabet de 1 à 26, et le numéro 0 représentera un espace. Alice découpe son message en blocs de longeur n (en ajoutant eventuellement des espaces à la fin du message); puis elle remplace les lettres et les espaces par les nombres correspondants pour former des matrices $1 \times n$; ensuite elle multiplie chacune de ces lignes avec A et elle envoie les résultats à Bob. Comment fera Bob pour décoder le message? Et comment est-ce que Bob peut envoyer un message codé à Alice? Calculer l'inverse de la matrice A ci-dessous, et utiliser la bonne matrice pour coder "`hello world`" (par Alice) et décoder "$(32/3, 5, -10/3)(-4/3, 9, 23/3)(-8/3, 0, 31/3)(-8, 15, 16)$"(par Bob):

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

5. Déterminant

Pour déterminer "par un simple calcul" si une matrice $A \in \mathbb{R}^{n\times n}$ est inversible ou pas, on peut utiliser une formule dûe à Pierre-Simon Laplace (1749–1827):

Définition 5.1 Le déterminant d'une matrice $A \in \mathbb{R}^{n\times n}$ est le nombre, noté $\det(A)$, calculé récursivement comme suit:

- si $n = 1$: $\det(a) = a$,
- si $n \geq 2$:

$$\det\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \begin{cases} \sum_{i=1}^{n} (-1)^{i+j} a_{ij} A^{ij} & \text{(“développement vers la } j\text{-ième colonne”)} \\ \sum_{j=1}^{n} (-1)^{i+j} a_{ij} A^{ij} & \text{(“développement vers la } i\text{-ième ligne”)} \end{cases}$$

où $A^{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ est la matrice (dite mineure) obtenue en supprimant la i-ième ligne et la j-ième colonne dans la matrice A.

En effet, on peut montrer que chacune de ces formules récursives donne le même résultat; nous l'admettrons.

Exemple 5.2 On vérifie facilement que

$$\det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{21}a_{12} \quad \text{et}$$

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

De manière générale, $\det(A)$ est une somme de $n!$ termes contenant chacun un produit de n facteurs; ainsi si $n = 10$ on doit calculer une somme de 3628800 termes. La formule générale de Laplace a le mérite d'exister—mais elle n'est pas toujours très utile dans la pratique.

Exemple 5.3 Pour une matrice triangulaire (supérieure ou inférieure), le déterminant est le produit des éléments sur la diagonale. Ainsi on calcule facilement le déterminant des matrices élémentaires associées à des réductions (le déterminant vaut 1) et des multiplications (le déterminant vaut le facteur de la multiplication). Pour une matrice élémentaire associée à une permutation de deux lignes, on vérifie que son déterminant vaut -1. Par ailleurs, pour toute matrice élémentaire E on voit bien que $\det(E) = \det(E^t)$.

Souvent, on n'utilise pas la formule de Laplace pour calculer $\det(A)$, mais plutôt la propriété suivante que – faute de temps – nous admettrons:

Théorème 5.4 (Donnée sans preuve.) Pour toutes matrices carrées $A, B \in \mathbb{R}^{n\times n}$ on a que $\det(AB) = \det(A)\det(B)$.

Attention: on n'a pas du tout que $\det(A + B) = \det(A) + \det(B)$!

Exemple 5.5 Soit $A' = E_k \dots E_1 A$ de $A \in \mathbb{R}^{n\times n}$ une matrice échelonnée, alors $\det(A) = \det(E_1)^{-1} \dots \det(E_k)^{-1} \det(A')$. On connait les déterminants des matrices élémentaires, et puisque A' est triangulaire supérieure aussi $\det(A')$ est facile à calculer. Ceci suggère une méthode de calcul de $\det(A)$ en l'échelonnant—et c'est souvent bien plus rapide que l'application de la formule de la Proposition 5.1.

Exemple 5.6 Si on connait une factorisation $PA = LU$ de $A \in \mathbb{R}^{n\times n}$, alors $\det(P)\det(A) = \det(L)\det(U)$ et les trois facteurs $\det(P)$, $\det(L)$ et $\det(U)$ sont faciles à calculer; ainsi on trouve donc $\det(A)$. (Notons que $\det(P) = \pm 1$, selon le nombre d'échanges de lignes que l'on a dû effectuer.) Souvent les logiciels appliquent cette méthode—mais bien sûr il faut d'abord un bon algorithme pour calculer la factorisation $PA = LU$!

Nous pouvons maintenant relever le défi posé au début de cette section:

Proposition 5.7 Une matrice carrée $A \in \mathbb{R}^{n\times n}$ est inversible si et seulement si $\det(A) \neq 0$; et dans ce cas $\det(A^{-1}) = \det(A)^{-1}$.

Démonstration. Soit $A' = E_k \cdots E_1 A$ une matrice échelonnée; c'est à dire, A' est une matrice triangulaire supérieur. On sait que A est inversible si et seulement si chaque colonne de A' contient un pivot. La matrice A' étant carrée, cela veut exactement dire que tout élément sur sa diagonale est non-nul, ce qui est équivalent au fait que $\det(A') \neq 0$, puisque ce déterminant est le produit des éléments de la diagonale de A'. Mais on sait que $\det(A') = \det(E_k)\ldots\det(E_1)\det(A)$, et toute matrice élémentaire est de déterminant non-nul, donc on a $\det(A') \neq 0$ si et seulement si $\det(A) \neq 0$. Si maintenant A est inversible, on a que $1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})$ d'où $\det(A^{-1}) = \det(A)^{-1}$. □

Au 17e siècle, le mathématicien japonnais Seki Takakazu (1640–1708) utilise des déterminants pour résoudre des systèmes linéaires jusqu'à 5 variables; en Europe c'est l'allemand Gottfried Wilhelm Leibniz (1646–1716) qui développe une théorie de déterminants pour la résolution de systèmes linéaires. Le Corollaire ci-dessous (connu par Leibniz) nous sera particulièrement utile dans un chapitre ultérieur:

Corollaire 5.8 Pour $A \in \mathbb{R}^{n\times n}$, le système linéaire homogène $AX = O$ admet des solutions non-nulles si et seulement si $\det(A) = 0$.

Notons pour terminer aussi que:

Proposition 5.9 Pour toute matrice carrée $A \in \mathbb{R}^{n\times n}$ on a que $\det(A^t) = \det(A)$.

Démonstration. Si A est non-inversible, alors A^t l'est aussi, et $\det(A) = 0 = \det(A^t)$. Si A est inversible, alors $A = E_k \cdots E_1$ pour des matrices élémentaires E_i. La transposée d'une matrice élémentaire est toujours une matrice élémentaire "du même type"; il suit que $\det(E_i) = \det(E_i^t)$. On peut alors calculer que $\det(A) = \det(E_k)\cdots\det(E_1) = \det(E_1^t)\cdots\det(E_k^t) = \det(E_1^t \cdots E_k^t) = \det(A^t)$. □

Exercices

1. Calculer le déterminant des matrices

$$A = \begin{pmatrix} 2 & 3 & 5 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 4 & 5 \\ 3 & 2 & 1 \end{pmatrix} \quad C = \begin{pmatrix} 9 & 1 & 1 \\ 0 & 4 & 5 \\ 0 & 0 & 1 \end{pmatrix}$$

$$D = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 4 & 0 \\ 0 & 2 & 0 \end{pmatrix} \quad E = \begin{pmatrix} 0 & 1 & -3 & 8 \\ 4 & 2 & 1 & 2 \\ 0 & 1 & -1 & 0 \\ 0 & 3 & -2 & 0 \end{pmatrix}$$

et en déduire la valeur de $\det(\det(C^t)(2A)^2B^{-1}(-D))$.

2. Calculer les déterminants des matrices suivantes:

$$A = \begin{pmatrix} -1 & 5 & 2 \\ 2 & 0 & -2 \\ 3 & -1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 4 & 0 \\ -5 & -4 & 0 \end{pmatrix}, \quad C = A^{-1} \cdot \det(B^{-1}) \cdot (\frac{1}{2} \cdot B)^2 \cdot A^3.$$

3. Vérifier par un calcul de déterminant si les systèmes homogènes suivants admettent des solutions non-triviales; le cas échéant, donner une matrice à colonnes libres engendrant l'ensemble des solutions.

(a) $\begin{cases} 3x + 7y = 0 \\ -4x + 8y = 0 \end{cases}$

(b) $\begin{cases} 3x + 7y - 2z = 0 \\ -4x + 8y = 0 \\ -2x + 30y - 4z = 0 \end{cases}$

(c) $\begin{cases} 3x + 7y - 2z + w = 0 \\ -4x + 8y + w = 0 \\ -2x + 30y - 4z + w = 0 \\ x + y - z = 0 \end{cases}$

4. Soit $A \in \mathbb{R}^{n\times n}$. Montrer que A est inversible si et seulement si A^tA l'est.

5. Pour tout $x_0, x_1, \ldots, x_n \in \mathbb{R}$ on peut démontrer la formule du *déterminant de Vandermonde*[2]:

$$\det \begin{pmatrix} x_0^n & x_0^{n-1} & \ldots & x_0 & 1 \\ x_1^n & x_1^{n-1} & \ldots & x_1 & 1 \\ \vdots & & & & \\ x_n^n & x_n^{n-1} & \ldots & x_n & 1 \end{pmatrix} = \prod_{i>j}(x_i - x_j)$$

(a) Expliciter et montrer cette formule pour $n = 1$ et $n = 2$. (La preuve du cas général peut se faire par induction sur n, à l'aide de quelques astuces pour le calcul de déterminants.)

(b) En déduire que, pour $n + 1$ points de $\mathbb{R}^2$, disons (x_0, y_0), ..., (x_n, y_n), avec les x_i tous distincts, il existe une unique fonction polynomiale de degré n,

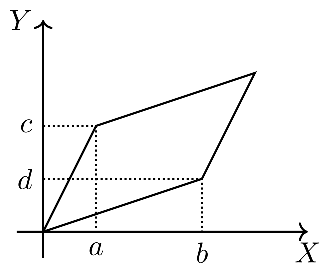
$$f : \mathbb{R} \to \mathbb{R} : x \mapsto a_n x^n + \cdots + a_1 x + a_0,$$

dont le graphe contient tous les (x_i, y_i); c'est le *polynôme d'interpolation*.

(c) Donner le polynôme d'interpolation ainsi déterminé par $(-2, 1), (-1, 2), (0, 0), (1, 2)$.

(d) Même question pour les points $(-2, 61), (-1, 3), (0, -1), (1, 4), (2, 49)$.

6. Montrer que l'aire du parallélogramme ci-dessous est égal à la valeur absolue de $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.



(Plus généralement, le volume d'un parallélépipède dans $\mathbb{R}^n$ peut être calculé comme la valeur absolue du déterminant d'une matrice $n \times n$.)

[2] Alexandre-Theophile Vandermonde (1735–1796)

7. Rappelons que la complexité (arithmétique) d'un calcul matriciel est donnée par le nombre de sommes et le nombre de produits que le calcul nécessite; pour simplifier, on admettera ici qu'une différence est une somme, et qu'une division est un produit. Pour une matrice A de genre 3×3, quelle est la complexité du calcul de $\det(A)$ lorsqu'on développe vers une colonne ou une ligne? Quelle est la complexité de la factorisation $A = LU$ par réductions inférieures (en supposant qu'aucune permutation n'est nécessaire)? Quelle est donc la complexité du calcul de $\det(A) = \det(L)\det(U)$? Mêmes questions pour une matrice 4×4.

6. Image et noyau

Le produit d'une matrice $A = (a_{ij})_{ij} \in \mathbb{R}^{m\times n}$ avec une colonne $X = (x_i)_i \in \mathbb{R}^{n\times 1}$ peut s'écrire comme

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 \cdot \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \ldots + x_n \cdot \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Autrement dit, AX est exactement la *combinaison linéaire* des colonnes de A de coefficients $x_1, \ldots, x_n$. Résoudre un système linéaire $AX = B$ revient donc à trouver toutes les manières dont on peut écrire la colonne B comme une combinaison linéaire des colonnes de A. Pour cette raison (entre autres) nous allons étudier l'ensemble de toutes les combinaisons linéaires des colonnes d'une matrice A.

Cependant, dans un exemple comme

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

il est clair que certaines colonnes de A peuvent être "redondantes" pour engendrer cet ensemble des combinaisons linéaires—ici la quatrième colonne de A est elle-même une combinaison linéaire des 3 premières, et ainsi toute combinaison linéaire des 4 colonnes peut être réécrite comme une combinaison linéaire des 3 premières colonnes. Nous devons donc aussi développer une méthode pour trouver les colonnes *strictement nécessaires* d'une matrice A pour engendrer toutes les combinaisons linéaires.

Définition 6.1 Pour une matrice $A \in \mathbb{R}^{m\times n}$, les ensembles

$$\mathrm{Im}(A) = \{AX \in \mathbb{R}^{m\times 1} \mid X \in \mathbb{R}^{n\times 1}\} \quad \text{et} \quad \mathrm{Ker}(A) = \{X \in \mathbb{R}^{n\times 1} \mid AX = O\}$$

sont l'image et le noyau (anglais: "kernel") de A.

Remarquons qu'une somme de deux éléments de $\mathrm{Im}(A)$ est encore un élément de $\mathrm{Im}(A)$, et que tout multiple d'un élément de $\mathrm{Im}(A)$ est encore un élément de $\mathrm{Im}(A)$; et de même pour $\mathrm{Ker}(A)$. Pour cette raison, on dit que $\mathrm{Im}(A) \subseteq \mathbb{R}^m$ et $\mathrm{Ker}(A) \subseteq \mathbb{R}^n$ sont des *sous-espaces*.

Proposition 6.2 Pour $A \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{m\times 1}$, le système linéaire $AX = B$ admet
- *au moins une* solution si et seulement si $B \in \mathrm{Im}(A)$,
- *au plus une* solution si et seulement si $\mathrm{Ker}(A) = \{O\}$.

Démonstration. La première assertion est une tautologie. Pour la deuxième, si $AX_1 = B = AX_2$ alors $X_1 - X_2 \in \mathrm{Ker}(A)$; donc $X_1 = X_2$ si et seulement si $\mathrm{Ker}(A) = \{O\}$. □

Proposition 6.3 Pour une matrice $A \in \mathbb{R}^{m\times n}$, aucune colonne de A n'est combinaison linéaire des autres colonnes de A si et seulement si $\mathrm{Ker}(A) = \{O\}$. Une telle matrice est dite "à colonnes libres" (ou "à colonnes indépendantes").

Démonstration. Notons les n colonnes de A par A_1, ..., A_n. Si la colonne A_k est une combinaison linéaire des $n-1$ autres colonnes, soit

$$A_k = \sum_{i\neq k} \alpha_i A_i,$$

alors $(\alpha_1, ..., \alpha_{k-1}, -1, \alpha_{k+1}, ..., \alpha_n)$ est une solution non-nulle du système $AX = O$. Réciproquement, si $X = (x_1, ..., x_n)$ est une solution non-nulle, alors au moins un x_k est non-nul, et donc

$$A_k = \sum_{i\neq k} -\frac{x_i}{x_k} A_i$$

montre que la colonne A_k dépend des autres colonnes. □

Pour vérifier si A est à colonnes libres, on résout le système homogène $AX = O$: si l'unique solution est $X = O$ alors A est à colonnes libres, si on trouve (au moins) une solution non-nulle alors A n'est pas à colonnes libres. Dans la suite, nous allons voir comment on peut calculer, pour une matrice quelconque $A \in \mathbb{R}^{m\times n}$, une matrice à colonnes libres $\overline{A} \in \mathbb{R}^{m\times k}$ ayant la même image, en supprimant les colonnes "redondantes" de A. Notons d'abord:

Lemme 6.4 Pour des matrices $A \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{m\times k}$ (donc éventuellement $n \neq k$), il existe une matrice $T \in \mathbb{R}^{k\times n}$ telle que $A = BT$ si et seulement si on a $\mathrm{Im}(A) \subseteq \mathrm{Im}(B)$.

Démonstration. Si $A = BT$ alors

$$\mathrm{Im}(A) = \{AX \mid X \in \mathbb{R}^{n\times 1}\} = \{BTX \mid X \in \mathbb{R}^{n\times 1}\} \subseteq \{BY \mid Y \in \mathbb{R}^{k\times 1}\} = \mathrm{Im}(B).$$

Réciproquement, si $\mathrm{Im}(A) \subseteq \mathrm{Im}(B)$ alors chaque colonne A_i de A est une combinaison linéaire des colonnes de B, soit $BT_i = A_i$; et donc l'assemblage des colonnes $(T_1|\ldots|T_n)$ fournit une matrice T telle que $A = BT$. □

Proposition 6.5 Si $A \in \mathbb{R}^{m\times n}$ est une matrice échelonnée, alors toute colonne sans pivot est une combinaison linéaire des colonnes à pivot qui la précèdent. Si on note $\overline{A}$ la sous-matrice de A contenant exactement les colonnes à pivot de A, on obtient ainsi que $\mathrm{Im}(A) = \mathrm{Im}(\overline{A})$ et $\mathrm{Ker}(\overline{A}) = \{O\}$.

Esquisse de démonstration. Le principe de la démonstration est facile à comprendre sur un exemple concret: supposons que

$$A = \begin{pmatrix} 1 & 2 & 1 & -1 & -4 \\ 0 & -1 & 3 & 2 & 4 \\ 0 & 0 & 0 & 3 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{et donc} \quad \overline{A} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix},$$

alors on observe que $A_3 = -3A_2 + 7A_1$ et $A_5 = 2A_4 - 2A_2 + A_1$ pour les colonnes A_i de A. Ainsi on trouve facilement des matrices S et T telles que $AS = \overline{A}$ et $\overline{A}T = A$, et donc $\mathrm{Im}(A) = \mathrm{Im}(\overline{A})$ par le Lemme 6.4. De plus, dans $\overline{A}$ aucune colonne ne peut être combinaison linéaire des autres colonnes (car il y a un pivot dans chaque colonne!): il s'agit donc d'une matrice à colonnes libres, comme voulu. □

Corollaire 6.6 Pour une matrice $A \in \mathbb{R}^{m\times n}$, soit $E_k \dots E_1 A = A'$ un échelonnement par opérations élémentaires (en particulier, A' peut être la forme réduite de A). On a alors exactement les mêmes dépendances entre les colonnes de A qu'entre les colonnes de même indices de A', et vice versa. Si on note $\overline{A}$ la sous-matrice de A contenant exactement les colonnes de même indices que les colonnes à pivot dans A', alors $\mathrm{Im}(A) = \mathrm{Im}(\overline{A})$ et $\mathrm{Ker}(\overline{A}) = \{O\}$.

Démonstration. On sait déjà que, pour tout $B \in \mathbb{R}^{m\times 1}$, on a $AX = B$ si et seulement si $A'X = B'$, où $B' = E_k \dots E_1 B$. En particulier, pour les colonnes A_i de A et A'_i de A', on a $AX = A_i$ (ce qui exprime une dépendance de la i-ième colonne de A des autres colonnes de A) si et seulement si $A'X = A'_i$ (ce qui exprime la même dépendance, mais pour les colonnes de A'). Définissant $\overline{A'}$ comme la sous-matrice de A' contenant exactement les colonnes à pivot, on a alors que $E_k \dots E_1 \overline{A} = \overline{A'}$. Ensuite on procède comme dans la proposition précédente pour construire des matrices S et T telles que $A'S = \overline{A'}$ et $\overline{A'}T = A'$; mais cela implique directement qu'aussi $AS = \overline{A}$ et $\overline{A}T = A$, d'où $\mathrm{Im}(A) = \mathrm{Im}(\overline{A})$ par le Lemme 6.4. Finalement, $\mathrm{Ker}(\overline{A}) = \mathrm{Ker}(\overline{A'}) = \{O\}$. □

Exercices

1. Exprimer chacune des colonnes B_1, B_2 et B_3 comme une combinaison linéaire des colonnes de la matrice A, puis construire une matrice M telle que $AM = B$ où B est la matrice dont les colonnes sont B_1, B_2, B_3:

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 1 & 3 \\ -1 & 0 & 1 \end{pmatrix} \quad B_1 = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} \quad B_2 = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} \quad B_3 = \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$$

2. Montrer que $A = \begin{pmatrix} 1 & -1 & 3 \\ -1 & 2 & 0 \end{pmatrix}$ et $B = \begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix}$ ont la même image.

3. Pour chacune matrices suivantes, donner une sous-matrice à colonnes libres ayant la même image:

(a) $\begin{pmatrix} 2 & 1 & 3 & -1 \\ 2 & -2 & 1 & 2 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & 3 & 5 \\ -1 & 4 & 2 \\ -3 & 4 & -2 \end{pmatrix}$

(c) $\begin{pmatrix} 5 & 4 & 7 \\ -1 & 3 & 2 \end{pmatrix}$

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

4. Démontrer: pour $A \in \mathbb{R}^{n\times n}$ on a $A^2 = O$ si et seulement si $\mathrm{Im}(A) \subseteq \mathrm{Ker}(A)$. Donner un exemple non-trivial d'une telle matrice.

7. Théorème du Rang

Deux matrices différentes peuvent avoir la même image. Observons tout de même:

Proposition 7.1 Si $A \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{m\times k}$ sont deux matrices à colonnes libres et $\mathrm{Im}(A) \subseteq \mathrm{Im}(B)$, alors $n \leq k$.

Démonstration. Puisque $\mathrm{Im}(A) \subseteq \mathrm{Im}(B)$, par le Lemme 6.4 on peut trouver $T \in \mathbb{R}^{k\times n}$ telle que $A = BT$. Il suit que $AX = O$ si et seulement si $BTX = O$, si et seulement si $TX = O$ (car

B est, par hypothèse, à colonnes libres). Si $n > k$ alors $TX = O$ admet nécessairement une solution non-nulle (car quand on l'échelonne on a nécessairement des colonnes sans pivot); il en est donc de même pour $AX = O$. Cela contredit l'hypothèse qu'aussi A est à colonnes libres. $\square$

Par conséquent, deux matrices à colonnes libres ayant la même image ont nécessairement le même nombre de colonnes. Si, pour une matrice A, on connait deux matrices à colonnes libres ayant la même image que A, alors elles ont le même nombre de colonnes. Ainsi ce nombre est un *invariant* de (l'image de) A, et mérite un nom:

Définition 7.2 Le rang d'une matrice $A \in \mathbb{R}^{m\times n}$, noté $\mathrm{rang}(A)$, est le nombre de colonnes d'une (et donc de toute) matrice à colonnes libres dont l'image est égale à $\mathrm{Im}(A)$.

Autrement dit, le rang de A est le nombre minimal de colonnes dont on a besoin pour engendrer toute l'image de A; on dit aussi que c'est la *dimension* de $\mathrm{Im}(A)$. Ce nombre exprime en quelque sorte la "complexité" de A.

Corollaire 7.3 Si $A \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{m\times k}$ sont deux matrices (pas nécessairement à colonnes libres) et $\mathrm{Im}(A) \subseteq \mathrm{Im}(B)$, alors $\mathrm{rang}(A) \leq \mathrm{rang}(B)$.

Pour calculer le rang d'une matrice A, on l'échelonne pour obtenir A', puis on compte le nombre de colonnes avec pivot dans A' (qui, par ailleurs, est égal au nombre de lignes non-nulles dans A'). Attention: on a bien $\mathrm{rang}(A) = \mathrm{rang}(A')$, mais pas nécessairement $\mathrm{Im}(A) = \mathrm{Im}(A')$! Il suit en particulier que $\mathrm{rang}(A)$ est toujours inférieur ou égal au nombre de colonnes de A. Voici une propriété surprenante du rang d'une matrice:

Proposition 7.4 Pour toute matrice $A \in \mathbb{R}^{m\times n}$ on a que $\mathrm{rang}(A) = \mathrm{rang}(A^t)$.

Démonstration. Soit une matrice $B \in \mathbb{R}^{m\times k}$ à colonnes libres telle que $\mathrm{Im}(A) = \mathrm{Im}(B)$; on a donc $\mathrm{rang}(A) = k$. Par le Lemme 6.4, il existe une matrice $T \in \mathbb{R}^{k\times n}$ telle que $A = BT$. Transposant les matrices on obtient $A^t = T^tB^t$, d'où $\mathrm{Im}(A^t) \subseteq \mathrm{Im}(T^t)$. Puisque $T^t \in \mathbb{R}^{n\times k}$, ceci implique $\mathrm{rang}(A^t) \leq \mathrm{rang}(T^t) \leq k = \mathrm{rang}(A)$. En remplaçant toute occurence de A par A^t dans ce raisonnement, on obtient aussi l'inégalité $\mathrm{rang}(A) \leq \mathrm{rang}(A^t)$. $\square$

Il suit donc que $\mathrm{rang}(A)$ est aussi inférieur ou égal au nombre de lignes de A!

Pour terminer cette section, reconsidérons le noyau d'une matrice, $\mathrm{Ker}(A)$. Clairement, toute combinaison linéaire de solutions au système $AX = O$ en est encore une solution. Ainsi on a intérêt à trouver un nombre minimal de solutions telles que toute autre solution en est une combinaison linéaire. Autrement dit, on cherche une matrice à colonnes libres $M \in \mathbb{R}^{n\times k}$ telle que $\mathrm{Im}(M) = \mathrm{Ker}(A)$. La méthode par échelonnement de A pour résoudre $AX = O$ produit "automatiquement" une telle matrice M, comme l'indique l'exemple suivant.

Exemple 7.5 Soit une matrice $A \in \mathbb{R}^{3\times 4}$ dont une matrice échelonnée est

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

On a $AX = O$ si et seulement si $A'X = O$, si et seulement si

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -3\beta + 3\alpha \\ \beta \\ -\alpha \\ \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} -3\beta \\ \beta \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 3\alpha \\ 0 \\ -\alpha \\ \alpha \\ 0 \end{pmatrix} = \beta \begin{pmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 3 \\ 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 & 3 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \beta \\ \alpha \end{pmatrix}.$$

On constate donc que l'image de la matrice

$$M = \begin{pmatrix} -3 & 3 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

est la totalité des solutions au système $AX = O$. De plus, il suffit de regarder la 4-ième et la 2-ième ligne (précisément les lignes qui correspondent aux inconnus pour lesquels on a dû introduire un paramètre) du système $MX = O$ pour se convaincre que M est à colonnes libres.

C'est encore la Proposition 7.1 qui garantit que la définition suivante est bien posée:

Définition 7.6 La nullité d'une matrice $A \in \mathbb{R}^{m\times n}$, notée $\mathrm{null}(A)$, est le nombre de colonnes d'une (et donc de toute) matrice à colonnes libres dont l'image est égale à $\mathrm{Ker}(A)$.

Répétons que $\mathrm{null}(A)$ correspond donc exactement au nombre de paramètres que l'on doit introduire lorsqu'on résout le système homogène $AX = O$ par échelonnement.

Théorème 7.7 (Théorème du rang.) Pour toute matrice $A \in \mathbb{R}^{m\times n}$ on a

$$n = \mathrm{rang}(A) + \mathrm{null}(A).$$

Démonstration. Soit $A' = E_k \dots E_1 A$ une matrice échelonnée pour A; ainsi les solutions à $AX = O$ sont identiques aux solutions à $A'X = O$. On sait que $\mathrm{rang}(A) = \mathrm{rang}(A')$, et ce dernier est égal au nombre de colonnes de A' contenant un pivot. Le nombre de colonnes de A' ne contenant pas de pivot est exactement le nombre de paramètres à choisir pour constuire la solution générale à $A'X = O$. Ainsi n, le nombre total de colonnes, est bien égal à la somme des deux. □

Corollaire 7.8 Une matrice $A \in \mathbb{R}^{n\times n}$ est inversible si et seulement si $\mathrm{rang}(A) = n$, si et seulement si $\mathrm{null}(A) = 0$.

Démonstration. Que $\mathrm{rang}(A) = n$ si et seulement si $\mathrm{null}(A) = 0$ est une conséquence directe du Théorème 7.7. Puis, soit A' la forme réduite de A; on a nécessairement $\mathrm{rang}(A) = \mathrm{rang}(A')$. Mais A est inversible si et seulement si $A' = I_n$, qui est l'unique matrice réduite de genre $n \times n$ et de rang n. □

Exercices

1. (a) Calculer une factorisation LU de la matrice $A = \begin{pmatrix} 3 & 2 & 0 & -1 & 1 \\ -6 & -3 & -1 & 4 & 1 \\ 3 & -1 & 5 & -7 & -7 \end{pmatrix}$.
   (b) Donner son rang et sa nullité.
   (c) Calculer une matrice B à colonnes libres dont l'image est le noyau de A.

2. Pour $A \in \mathbb{R}^{3\times 3}$ on sait que $\begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix} \in \mathrm{Ker}(A)$ et que les deux premières colonnes de A sont $\begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix}$ et $\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$. Calculer la troisième colonne de A, puis déterminer si $\begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} \in \mathrm{Im}(A)$.

3. Pour chacune des matrices suivantes, exprimer son noyau comme l'image d'une matrice à colonnes libres, puis en déduire sa nullité. Identifier aussi les matrices inversibles.

(a) $\begin{pmatrix} 0 & 2 & -1 & 3 \\ 4 & 2 & 1 & 0 \\ 1 & 0 & -1 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 6 & 4 & -1 \\ 2 & 4 & -1 & 3 \\ -1 & 2 & 5 & 0 \end{pmatrix}$

(b) $\begin{pmatrix} -1 & 5 & 2 & 3 \\ 3 & 0 & -1 & 1 \\ 4 & 2 & 2 & 4 \\ 0 & 2 & 4 & 3 \end{pmatrix}$ (d) $\begin{pmatrix} 4 & -2 \\ -8 & 4 \\ -1 & 1 \end{pmatrix}$

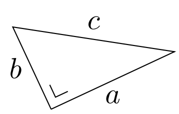
4. Soit une matrice $A \in \mathbb{R}^{m\times n}$. Montrer que, s'il existe des matrices $B, C \in \mathbb{R}^{n\times m}$ telles que $AB = I_m$ et $CA = I_n$, alors nécessairement $m = n$ et $B = C$. Autrement dit, c'est le cas si et seulement si A est inversible.

5. Soit une matrice $A \in \mathbb{R}^{n\times n}$. Montrer que, s'il existe une matrice $B \in \mathbb{R}^{n\times n}$ telle que $AB = I_n$, alors nécessairement $BA = I_n$. Autrement dit, c'est le cas si et seulement si A est inversible.

8. Orthogonalité et distance, moindres carrés

Un système $AX = B$ admet une solution si et seulement si $B \in \operatorname{Im}(A)$. Si $B \notin \operatorname{Im}(A)$, on veut souvent connaître un X_0 tel que $B_0 := AX_0$ est *presque* égal à B: ce X_0 est une *solution approchée* au système (irrésoluble) $AX = B$. Pour que cela aît un sens, il faut d'abord établir une façon d'exprimer la *proximité* de B_0 avec B. La géométrie classique nous inspirera pour accomplir cette tâche.

Le théorème de Pythagore dit qu'un triangle de côtés a, b et c est rectangulaire en $\widehat{ab}$ si et seulement si $a^2 + b^2 = c^2$:



Il permet de vérifier que, pour deux points $P = (x_1, y_1)$ et $Q = (x_2, y_2)$ dans le plan $\mathbb{R}^2$,
- la distance de O à P est $\|P\| = \sqrt{x_1^2 + y_1^2}$,
- la distance de P à Q est donc $\|P - Q\| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$,
- le segment OP est orthogonal au segment OQ (on dira plutôt que "P est orthogonal à Q", noté $P \perp Q$) si et seulement si $\|P\|^2 + \|Q\|^2 = \|P - Q\|^2$, si et seulement si $x_1x_2 + y_1y_2 = 0$.

En identifiant $\mathbb{R}^2$ avec $\mathbb{R}^{2\times 1}$ on note

$$P = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \quad \text{et} \quad Q = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

et on peut alors écrire que

$$\|P\| = \sqrt{P^tP} \quad \text{et} \quad P \perp Q \Leftrightarrow P^tQ = 0.$$

On généralise de façon évidente à n dimensions.

Définition 8.1 Soient deux colonnes $P, Q \in \mathbb{R}^{n\times 1}$. On dit que
- $P^tQ \in \mathbb{R}$ est le produit scalaire de P et Q,
- $\|P\| = \sqrt{P^tP}$ est la norme de P,
- $\|P - Q\|$ est la distance de P à Q,
- P est orthogonal à Q (noté $P \perp Q$) si $P^tQ = 0$.

Le produit scalaire étant un cas très particulier du produit matriciel, il est facile de vérifier les propriétés suivantes.

Proposition 8.2 Pour tout $P, P_1, P_2, Q, Q_1, Q_2 \in \mathbb{R}^{n\times 1}$ et tout $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$ on a que

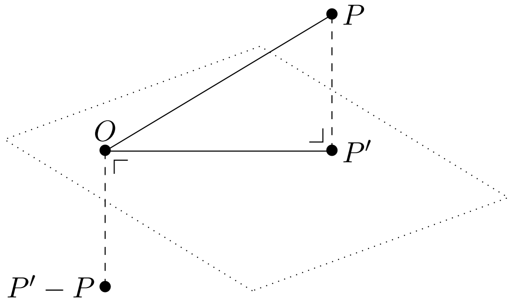
1. $(\alpha_1 P_1 + \alpha_2 P_2)^t(\beta_1 Q_1 + \beta_2 Q_2) = \sum_{i,j} \alpha_i \beta_j P_i^t Q_j$ (le produit scalaire est bilinéaire),
2. $P^tP \geq 0$ (le produit scalaire est positif),
3. $P^tP = 0 \Leftrightarrow P = O$ (le produit scalaire est défini),
4. $P^tQ = Q^tP$ (le produit scalaire est symétrique).

La positivité du produit scalaire justifie la définition de $\|P\|$. Notons aussi que $P \perp Q \Leftrightarrow Q \perp P$ et $\|P - Q\| = \|Q - P\|$, ce qui confirme l'intuition géométrique.

Corollaire 8.3 Pour un n-uple $P \in \mathbb{R}^{n\times 1}$ et une matrice $A \in \mathbb{R}^{n\times k}$, P est orthogonal à chaque élément de $\mathrm{Im}(A)$, si et seulement si P est orthogonal à chaque colonne de A, si et seulement si $A^tP = O$, si et seulement si $P \in \mathrm{Ker}(A^t)$.

L'exemple suivant indique une relation importante entre distance et orthogonalité en géométrie classique.

Exemple 8.4 La projection orthogonale d'un point P sur un plan (comprenant O) dans l'espace $\mathbb{R}^3$ donne exactement le point P' du plan le plus proche du point P:



Plus généralement on a:

Proposition 8.5 Soit une matrice $A \in \mathbb{R}^{n\times k}$ et $P \in \mathbb{R}^{n\times 1}$. Pour un point $P' \in \mathrm{Im}(A)$, les conditions suivantes sont équivalentes:

1. $P' - P$ est orthogonal à chaque élément de $\mathrm{Im}(A)$,
2. la distance $\|P' - P\|$ est minimale, c'est à dire, pour tout $Q \in \mathrm{Im}(A)$, $\|P' - P\| \leq \|Q - P\|$, l'égalité étant obtenue uniquement si $Q = P'$.

On dit alors que P' est la projection orthogonale de P sur $\mathrm{Im}(A)$, noté $P' = \mathrm{proj}_A(P)$.

Démonstration. ($1 \Rightarrow 2$) Pour tout $Q \in \mathrm{Im}(A)$ on vérifie facilement que $Q - P' \in \mathrm{Im}(A)$; ainsi, par hypothèse, $P' - P$ est orthogonal à $Q - P'$. On peut alors calculer que

$$\begin{aligned}
&\|Q - P\|^2 \\
&= \|Q - P' + P' - P\|^2 \\
&= ((Q - P') + (P' - P))^t((Q - P') + (P' - P)) \\
&= (Q - P')^t(Q - P') + (Q - P')^t(P' - P) + (P' - P)^t(Q - P') + (P' - P)^t(P' - P) \\
&= \|Q - P'\|^2 + 0 + 0 + \|P' - P\|^2.
\end{aligned}$$

On a ainsi que $\|Q - P\| = \sqrt{\|Q - P'\|^2 + \|P' - P\|^2}$, ce qui montre que $\|Q - P\| \geq \|P' - P\|$ pour tout $Q \in \mathrm{Im}(A)$, avec égalité si et seulement si $\|Q - P'\| = 0$, si et seulement si $Q = P'$.

$(2 \Rightarrow 1)$ Ecrivons $P' = AX$, et supposons donc que la distance $\|AX - P\|$ est minimale; de manière équivalente, le carré de la distance,

$$\|AX - P\|^2 = (AX - P)^t(AX - P) = X^tA^tAX + X^t(-2A^tP) + P^tP,$$

est minimal. Puisque A et P sont donnés, c'est le choix de la "variable" $X \in \mathbb{R}^{k\times 1}$ qui doit assurer la valeur minimale de cette expression. Lorsqu'on note

$$A^tA = (u_{ij})_{ij} \quad \text{et} \quad -2A^tP = (v_i)_i \quad \text{et} \quad P^tP = (w) \quad \text{et} \quad X = (x_i)_i$$

l'expression ci-dessus est le polynôme

$$f(x_1, ..., x_k) = \sum_{i,j=1}^{k} u_{ij}x_ix_j + \sum_{i=1}^{k} v_ix_i + w.$$

Par l'analyse des fonctions scalaires en plusieurs variables réelles[3], lorsque la valeur minimale de f est atteinte, on a nécessairement

$$\text{pour tout } l \in \{1, ..., k\} : \frac{\partial f}{\partial x_l} = 0.$$

En calculant ces dérivées partielles, cette condition devient

$$\text{pour tout } l \in \{1, ..., k\} : \sum_{i=1}^{k} u_{il}x_i + \sum_{j=1}^{k} u_{lj}x_j + v_l = 0.$$

Mais la matrice A^tA est symétrique, c'est à dire $u_{ij} = u_{ji}$, et on obtient

$$\text{pour tout } l \in \{1, ..., k\} : \sum_{j=1}^{k} 2u_{lj}x_j + v_l = 0.$$

En notation matricielle cela devient $2A^tAX - 2A^tP = O$; c'est équivalent à $A^t(AX - P) = O$ et dit donc (par le Corollaire 8.3) que $AX - P$ est orthogonal à chaque élément de $\mathrm{Im}(A)$. $\square$

La Proposition précédente atteste de l'*unicité* de la projection orthogonale de P sur $\mathrm{Im}(A)$, mais nous avons encore à montrer l'*existence* de ladite projection. Pour cela:

Lemme 8.6 Pour toute matrice $A \in \mathbb{R}^{n\times k}$ on a que $\mathrm{Ker}(A^tA) = \mathrm{Ker}(A)$, et donc aussi $\mathrm{null}(A^tA) = \mathrm{null}(A)$ et $\mathrm{rang}(A^tA) = \mathrm{rang}(A)$.

Démonstration. Si $AX = O$ alors $A^tAX = O$. Réciproquement, si $A^tAX = O$ alors $\|AX\| = X^tA^tAX = X^tO = 0$ et donc $AX = O$. Le reste suit par le Théorème du Rang. $\square$

Proposition 8.7 Si $A \in \mathbb{R}^{n\times k}$ est une matrice de rank k, alors pour tout $P \in \mathbb{R}^{n\times 1}$, on a $\mathrm{proj}_A(P) = A(A^tA)^{-1}A^tP$.

Démonstration. Par le Lemme 8.6, on a $\mathrm{rang}(A^tA) = \mathrm{rang}(A) = k$ ce qui confirme l'inversibilité de $A^tA \in \mathbb{R}^{k\times k}$. Le système $A^tAX = A^tP$ admet donc l'unique solution $X = (A^tA)^{-1}A^tP$ permettant de calculer $\mathrm{proj}_A(P)$ comme indiqué. $\square$

Pour manipuler les matrices quelconques, remarquons que l'on peut toujours trouver une matrice à colonnes libres ayant la même image (cf. Proposition 6.6). La Proposition 8.5 ne dépend en effet que de l'image en tant que sous-ensemble de $\mathbb{R}^{n\times 1}$, et non-pas de la matrice qui l'engendre.

[3] Voir tout bon cours sur ce sujet.

Corollaire 8.8 Soit $A \in \mathbb{R}^{n\times k}$ une matrice non-nulle et $\overline{A} \in \mathbb{R}^{n\times l}$ une matrice à colonnes libres ayant la même image que A. Pour tout $P \in \mathbb{R}^{n\times 1}$ on a

$$\mathrm{proj}_A(P) = \mathrm{proj}_{\overline{A}}(P) = \overline{A}(\overline{A}^t\overline{A})^{-1}\overline{A}^t P.$$

(Et si A est une matrice nulle, tout est trivial.)

Revenons à la question initiale de ce chapitre:

Exemple 8.9 (Méthode des moindres carrés.) Soit $AX = B$ un système linéaire quelconque. On peut *toujours* calculer la projection orthogonale de B sur $\mathrm{Im}(A)$, soit $B' = \mathrm{proj}_A(B)$; et le système $AX = B'$ admet *toujours au moins une solution.* Si $B \in \mathrm{Im}(A)$ alors $B' = B$ et on retrouve toutes les *solutions exactes* de $AX = B$. Si $B \notin \mathrm{Im}(A)$ alors $B' \neq B$ et les solutions de $AX = B'$ sont appelées *les solutions approchées au sens des moindres carrés*[4] de $AX = B$. Si, en plus, la matrice $A \in \mathbb{R}^{n\times k}$ est de rang k ("A est de rang complet") alors $B' = A(A^tA)^{-1}A^tB$ et donc $AX = B'$ admet pour *unique* solution $X = (A^tA)^{-1}A^tB$. Si A n'est pas de rang complet alors il existe plusieurs solutions approchées – qui minimisent toutes l'expression $\|AX - B\|^2$ – mais on peut démontrer qu'il existe une unique solution approchée minimisant aussi $\|X\|^2$; voir Exemple 12.4.

Un cas particulier de la projection orthogonale sera utile dans la suite:

Exemple 8.10 (Projection orthogonale sur une seule colonne) Soit un n-uple non-nulle $A \in \mathbb{R}^n$; en tant que matrice à une colonne, $A \in \mathbb{R}^{n\times 1}$ est de rang complet. Pour tout n-uple $P \in \mathbb{R}^n$ (pensé comme colonne $P \in \mathbb{R}^{n\times 1}$) on peut donc calculer la projection orthogonale de P sur $\mathrm{Im}(A)$ (qui n'est autre que la *droite passant par O et A*). La formule pour calculer cette projection se réduit alors à

$$\mathrm{proj}_A(P) = A(A^tA)^{-1}A^tP = \frac{1}{A^tA}(AA^t)P = \frac{1}{A^tA}A(A^tP) = \frac{A^tP}{A^tA}A = \frac{P^tA}{A^tA}A,$$

parce que $A^tA \in \mathbb{R}$ et $A^tP \in \mathbb{R}$. (Ceci est *faux* si A est une matrice à plusieurs colonnes!)

Exercices

1. Calculer dans $\mathbb{R}^4$:
   (a) toutes les valeurs $\alpha \in \mathbb{R}$ pour lesquelles $(1, \alpha, 3, 2\alpha)$ est orthogonal à $(-1, 3, 0, 2\alpha)$,
   (b) la distance entre $(1, 3, 2, 8)$ et $(4, 3, 2, 4)$.

2. Soient
$$A = \begin{pmatrix} 1 & 0 & 2 & -2 \\ -1 & 2 & -4 & 6 \\ 3 & -2 & 8 & -10 \\ 2 & 1 & 3 & -2 \end{pmatrix} \quad \text{et} \quad P = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$
   Sachant que le rang de A n'est pas 4, calculer la projection orthogonale de P sur $\mathrm{Im}(A)$.

3. Ecrire les systèmes suivants comme une équation matricielle $AX = B$, calculer le rang de la matrice A des coefficients, observer qu'une solution exacte n'existe pas, puis les résoudre au sens des moindres carrés:

[4]Ce nom historique s'explique par le fait que l'on minimise en effet la quantité $\|AX - B\|$, qui est (la racine carrée d')une somme de carrés.

(a) $\begin{cases} x-y &= 1 \\ x &= 0 \\ y &= 2 \end{cases}$ (b) $\begin{cases} 2x+y &= -1 \\ x+y &= 1 \\ -x+y &= 3 \\ y &= -1 \end{cases}$ (c) $\begin{cases} 4x &= 2 \\ 2y &= 0 \\ x+y &= 11 \end{cases}$

4. Faire un dessin dans le plan $\mathbb{R}^2$ des droites d'équations $y = x+1$, $y = 3x + \frac{1}{2}$ et $y = -x + \frac{5}{2}$. Calculer les trois points d'intersection de cette configuration, ainsi que "*le point dans lequel les trois droites se coupent presque*" (avec la méthode des moindres carrés).

5. Soient les points $(1,3)$, $(2,5)$, $(3,4)$, $(4,5)$ du plan $\mathbb{R}^2$. On veut construire une droite qui passe "au plus près possible" de tous ces points. Sachant qu'une droite est donnée par une équation du type $y = ax+b$ (et donc on cherche à déterminer a et b), transformer ce problème en un problème de résolution de système linéaire au sens des moindres carrés, pour trouver la réponse.

6. Généraliser l'énoncé précédent: donner la procédure à suivre pour construire une fonction polynomiale $y = a_k x^k + \cdots + a_1 x + a_0$ de degré k dont le graphe passe "au plus près possible" de n points (x_1, y_1), ..., (x_n, y_n) du plan $\mathbb{R}^2$.

   Cette technique relève de la discipline d'*ajustement de courbe* (ou *curve fitting*): on veut trouver l'équation polynomiale $y = a_k x^k + \cdots + a_1 x + a_0$ qui décrit "au mieux" la relation entre abscisses et ordonnées d'une série de données expérimentales (x_1, y_1), ..., (x_n, y_n). Si on choisit $k = 1$, et donc on cherche une droite $y = ax + b$, on parle d'un *modèle de régression linéaire*; pour le cas général on parle de *régression polynomiale.* On a vu précédemment que l'on peut toujours trouver un unique polynôme de degré $n-1$ dont le graphe passe *exactement* par les n points donnés; c'est le polynôme d'*interpolation.* L'intérêt de la *régression* est donc de déterminer un polynôme de degré $k < n-1$ dont le graphe passe *approximativement* par les n points donnés.

9. Gram-Schmidt, factorisation $A = QR$

La projection orthogonale d'un $P \in \mathbb{R}^{n\times 1}$ sur les colonnes d'une matrice $A \in \mathbb{R}^{n\times k}$ de rang k,

$$\mathrm{proj}_A(P) = A(A^tA)^{-1}A^tP,$$

serait beaucoup plus facile à calculer si $A^tA = I_k$. Si on écrit les colonnes de A comme

$$A = \left(\begin{array}{c|c|c|c} A_1 & A_2 & \cdots & A_k \end{array} \right)$$

alors $A^tA = I_k$ si et seulement si

$$\begin{cases} \|A_i\| = 1 \text{ pour tout } i, \\ A_i \perp A_j \text{ pour tout } i \neq j. \end{cases}$$

Autrement dit, chaque colonne de A est de norme 1 ("normale") et orthogonale à toute autre colonne—on dira tout simplement que A est une matrice *à colonnes orthonormales.*

Proposition 9.1 Si une matrice $A \in \mathbb{R}^{n\times k}$ est à colonnes orthonormales, alors $\mathrm{rang}(A) = k$.

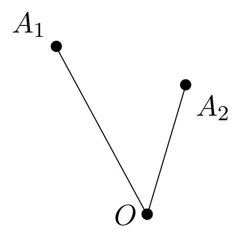
Démonstration. Si $AX = O$ alors $X = A^tAX = A^tO = O$ donc $\mathrm{Ker}(A) = \{O\}$; le résultat suit par le Théorème du Rang. □

Etant donné une matrice $A \in \mathbb{R}^{n\times k}$ de rang k, supposons qu'on puisse calculer ("efficacement") une matrice $Q \in \mathbb{R}^{n\times k}$ à colonnes orthonormales telle que $\mathrm{Im}(A) = \mathrm{Im}(Q)$. Il suit alors que, pour tout $P \in \mathbb{R}^{n\times 1}$,

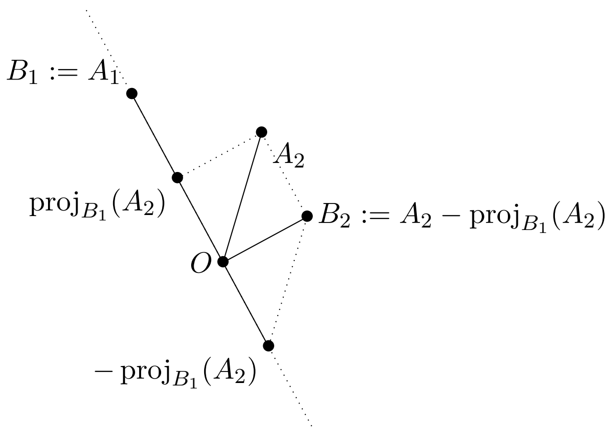
$$\mathrm{proj}_A(P) = \mathrm{proj}_Q(P) = QQ^tP.$$

Cela évite le calcul ("coûteux") de $(A^tA)^{-1}$.

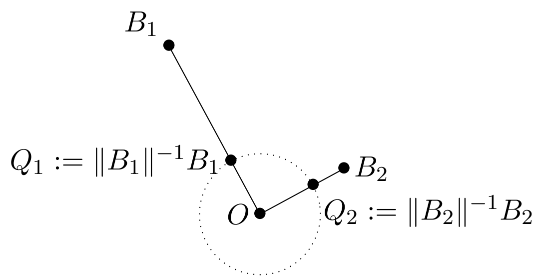
Exemple 9.2 Soit $A \in \mathbb{R}^{3\times 2}$ une matrice dont les colonnes, notées A_1 et A_2, déterminent un plan dans $\mathbb{R}^3$; on dessinera ci-dessous dans ce plan. D'abord on peut représenter les colonnes de A:



Tout en restant dans le même plan, on peut remplacer A_2 par un point orthogonal à A_1:



On peut ensuite ramener les points B_1 et B_2 à distance 1 de l'origine:



En formules, on pose donc

$$\text{d'abord} \begin{cases} B_1 = A_1 \\ B_2 = A_2 - \mathrm{proj}_{B_1}(A_2) \end{cases} \qquad \text{et ensuite} \qquad \begin{cases} Q_1 = \|B_1\|^{-1}B_1 \\ Q_2 = \|B_2\|^{-1}B_2 \end{cases}$$

Puisqu'on a

$$B_2 = A_2 - \mathrm{proj}_{B_1}(A_2) = A_2 - B_1(B_1^tB_1)^{-1}B_1^tA_2 = A_2 - \frac{A_2^tB_1}{B_1^tB_1}B_1$$

ces formules sont équivalentes à

$$\begin{cases} A_1 = 1B_1 + 0B_2 \\ A_2 = \frac{A_2^tB_1}{B_1^tB_1}B_1 + 1B_2 \end{cases} \quad \text{et} \quad \begin{cases} B_1 = \|B_1\|Q_1 + 0Q_2 \\ B_2 = 0Q_1 + \|B_2\|Q_2 \end{cases}$$

et cela s'écrit de façon matricielle comme

$$\left(\begin{array}{c|c} A_1 & A_2 \end{array}\right) = \left(\begin{array}{c|c} B_1 & B_2 \end{array}\right) \begin{pmatrix} 1 & \frac{A_2^t B_1}{B_1^t B_1} \\ 0 & 1 \end{pmatrix} \quad \text{et} \quad \left(\begin{array}{c|c} B_1 & B_2 \end{array}\right) = \left(\begin{array}{c|c} Q_1 & Q_2 \end{array}\right) \begin{pmatrix} \|B_1\| & 0 \\ 0 & \|B_2\| \end{pmatrix}.$$

Ainsi on obtient une factorisation $A = QR$ où $Q \in \mathbb{R}^{3\times 2}$ est la matrice à colonnes orthonormales Q_1 et Q_2, et $R \in \mathbb{R}^{2\times 2}$ est le produit de deux matrices triangulaires supérieures inversibles (car de déterminants non-nuls), donc elle-même triangulaire supérieure inversible. Cela implique en particulier que $\mathrm{Im}(A) = \mathrm{Im}(Q)$. Bien qu'on puisse calculer la matrice R par

$$R = \begin{pmatrix} \|B_1\| & 0 \\ 0 & \|B_2\| \end{pmatrix} \begin{pmatrix} 1 & \frac{A_2^t B_1}{B_1^t B_1} \\ 0 & 1 \end{pmatrix},$$

on peut aussi la retrouver par $R = Q^t A$ (si on connaît Q).

Cet exemple indique le résultat général suivant, dû à Jørgen Pedersen Gram (1850-1916) et Erhard Schmidt (1876-1959):

Théorème 9.3 (Procédé de Gram-Schmidt) Soit $A \in \mathbb{R}^{n\times k}$ une matrice de rang k, dont on écrit les colonnes comme A_1, ..., A_k. Par récursion on pose pour tout $i \in \{1, \ldots, k\}$:

$$B_i = A_i - \sum_{j<i} \mathrm{proj}_{B_j}(A_i) \quad \text{et} \quad Q_i = \|B_i\|^{-1} B_i.$$

Ainsi la matrice $Q \in \mathbb{R}^{n\times k}$ de colonnes Q_1, ..., Q_k est à colonnes orthonormales, $R = Q^t A$ est une matrice triangulaire supérieure inversible, et on obtient la factorisation $A = QR$. Il suit en particulier que $\mathrm{Im}(A) = \mathrm{Im}(Q)$.

Esquisse de la démonstration. Si on suppose que les n-uples B_1, ..., B_r sont orthogonaux deux-à-deux, alors il n'est pas difficile de montrer qu'aussi $B_{r+1} = A_{r+1} - \sum_{j<r+1} \mathrm{proj}_{B_j}(A_{r+1})$ est orthogonal à chacun des B_1, ..., B_r. Ainsi, par récurrence sur k on obtient l'orthogonalité des B_i's, et l'orthonormalité des Q_i's est alors immédiat. On pourra s'inspirer des exemples précédents pour détailler la factorisation $A = QR$. $\square$

Il s'agit ici de la (très fameuse) *factorisation QR* de la matrice A, qui intervient dans un tas d'applications du calcul matriciel en informatique. Cette factorisation n'est pas unique, et le Procédé de Gram-Schmidt n'est pas optimale (ni en temps de calcul, ni en stabilité, ni en mémoire utilisé). L'optimisation du calcul de $A = QR$ (et ses multiples applications) est toujours un sujet de recherche!

Exemple 9.4 Si $A \in \mathbb{R}^{n\times k}$ est de rang k et $A = QR$ est une factorisation QR, alors pour tout $P \in \mathbb{R}^{n\times 1}$ on a que $\mathrm{proj}_A(P) = QQ^t P$.

Exemple 9.5 Si $A \in \mathbb{R}^{n\times n}$ est inversible et $A = QR$ est une factorisation QR, alors $\det(A) = \det(Q)\det(R)$. On montre facilement que $\det(Q) = \pm 1$ alors que $\det(R)$ est facile à calculer puisque R est triangulaire supérieure.

Exemple 9.6 Si $A \in \mathbb{R}^{n\times k}$ est de rang k et $A = QR$ est une factorisation QR, alors le système linéaire $AX = B$ admet la solution approchée au sens des moindres carrés $X = R^{-1}Q^t B$, que l'on peut calculer en résolvant le système (triangulaire supérieure, donc facile) $RX = Q^t B$.

Exercices

1. Identifier les matrices à colonnes orthonormales:

(a) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$

(d) $\begin{pmatrix} -1 & 0 & 2 \\ 0 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}$

(e) $\begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & -1 & 0 \end{pmatrix}$

(f) $\begin{pmatrix} -1 & 2 \\ 0 & 4 \\ 1 & -1 \end{pmatrix}$

(g) $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

(h) $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$

2. Soient la matrice $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}$ et le quadruple $P = \begin{pmatrix} -3 \\ 5 \\ 2 \\ -1 \end{pmatrix}$.
   (a) Calculer le rang de A.
   (b) Par le procédé de Gram-Schmidt, donner une factorisation $A = QR$.
   (c) Montrer que $P \notin \mathrm{Im}(A)$ puis calculer la projection orthogonale de P sur $\mathrm{Im}(A)$.
   (d) Résoudre le système $AX = P$ au sens des moindres carrés.

3. Donner une factorisation $A = LU$ et une factorisation $A = QR$ de la matrice $A = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}$.

4. Donner une factorisation QR de la matrice $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ puis résoudre $AX = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}$.

5. (a) Appliquer Gram-Schmidt à la suite $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.
   (b) En déduire une factorisation $A = QR$ de la matrice $A = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$.
   (c) Calculer la projection orthogonale de $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ sur l'espace engendré par les colonnes de A.

10. Diagonalisation, factorisation $A = BDB^{-1}$

Nous avons vu la factorisation $PA = LU$ dans le contexte de la résolution de systèmes, et la factorisation $A = QR$ dans le contexte des projections orthogonales, mais ni l'une ni l'autre est très utile pour le problème suivant.

Exemple 10.1 Calculer la 100-ième puissance de la matrice

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

Sans ordinateur ce n'est pas possible—sauf si on observe d'abord que $A = BDB^{-1}$ où

$$B = \begin{pmatrix} 1 & 3 \\ -1 & 1 \end{pmatrix} \quad \text{et} \quad D = \begin{pmatrix} -1 & 0 \\ 0 & \frac{1}{3} \end{pmatrix},$$

puisque $A^n = BD^nB^{-1}$ et D^n est trivial à calculer.

Ceci est une des raisons pour introduire la notion suivante.

Définition 10.2 Une matrice $A \in \mathbb{R}^{n\times n}$ est diagonalisable s'il existe une matrice diagonale $D \in \mathbb{R}^{n\times n}$ et une matrice inversible $B \in \mathbb{R}^{n\times n}$ telles que $A = BDB^{-1}$.

Dans la suite on écrira systématiquement λ_1, ..., λ_n pour les éléments sur la diagonale de D, et $B_1, ..., B_n$ pour les colonnes de B. On observe alors que

$$\begin{aligned} A = BDB^{-1} &\iff AB = BD \\ &\iff \forall i \in \{1, \ldots, n\} : AB_i = \lambda_i B_i \\ &\iff \forall i \in \{1, \ldots, n\} : (A - \lambda_i I_n)B_i = O. \end{aligned}$$

Puisque aucune colonne B_i ne peut être nulle (sinon la matrice B ne serait pas inversible), cette dernière équation matricielle implique que les scalaires λ_i sont ceux pour lesquels le système linéaire homogène $(A - \lambda_i I_n)X = O$ admet des solutions non-nulles; et par la Proposition 5.8, ce sont exactement les scalaires λ_i pour lesquels $\det(A - \lambda_i I_n) = 0$. On observe également que les B_i sont des solutions non-nulles à des systèmes linéaires homogènes.

Définition 10.3 Soit une matrice $A \in \mathbb{R}^{n\times n}$. On appelle $p_A(\lambda) = \det(A - \lambda I_n)$ le polynôme caractéristique de A (il s'agit en effet d'une expression polynomiale de degré n en λ), et ses racines les valeurs propres de A. L'ensemble des valeurs propres de A est noté $\mathrm{Spec}(A)$ et appelé le spectre de A. Autrement dit, $\lambda \in \mathrm{Spec}(A)$ si et seulement si le système linéaire homogène $(A - \lambda I_n)X = O$ admet des solutions non-nulles. Si λ est une valeur propre de A, on appelle $E_\lambda = \mathrm{Ker}(A - \lambda I_n)$ l'espace propre pour la valeur propre λ; et on dit qu'un élément non-nul de E_λ est un vecteur propre de A pour la valeur propre λ.

Ci-dessus on a donc montré que, si $A = BDB^{-1}$, alors la i-ième colonne de B est un vecteur propre pour le i-ième élément sur la diagonale de D. Mais on peut calculer *toutes* les valeurs propres de A en résolvant l'équation $p_A(\lambda) = 0$. Puisqu'il s'agit d'un polynôme de degré n, il admet au plus n racines réelles distinctes $\lambda_1, \ldots, \lambda_k$ ($k \leq n$). Ensuite, pour chacune des valeur propres λ_i, on peut calculer *tous* les vecteurs propres qui lui sont associés, en résolvant le système linéaire $(A - \lambda_i I_n)X = O$. L'espace des solutions $E_{\lambda_i} = \mathrm{Ker}(A - \lambda_i I_n)$ étant non-vide, on peut calculer une matrice $B_{\lambda_i} \in \mathbb{R}^{n\times n_i}$ à colonnes libres engendrant cet espace: $E_{\lambda_i} = \mathrm{Im}(B_{\lambda_i})$. Le résultat suivant (dont nous ne détaillerons pas la démonstration, faute de temps) atteste qu'on peut ainsi détecter si la matrice $A \in \mathbb{R}^{n\times n}$ donnée est diagonalisable, et – le cas échéant – montre comment construire concrètement une factorisation $A = BDB^{-1}$.

Théorème 10.4 Soit une matrice $A \in \mathbb{R}^{n\times n}$ de spectre $\mathrm{Spec}(A) = \{\lambda_1, ..., \lambda_k\}$, et supposons que, pour tout $\lambda_i \in \mathrm{Spec}(A)$, $B_{\lambda_i} \in \mathbb{R}^{n\times n_i}$ est une matrice à colonnes libres telle que $E_{\lambda_i} = \mathrm{Im}(B_{\lambda_i})$. La matrice A est diagonalisable si et seulement si $n_1 + \cdots + n_k = n$. Dans ce cas, si on pose

$$D = \left(\begin{array}{ccc|c|ccc} \lambda_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda_1 & & & & \\ & & & \cdots & & & \\ & & & & \lambda_k & & \\ & & & & & \ddots & \\ & & & & & & \lambda_k \end{array}\right) \quad \text{et} \quad B = \left(\begin{array}{c|c|c} & & \\ B_{\lambda_1} & \cdots & B_{\lambda_k} \\ & & \end{array}\right)$$

(où on répète chaque λ_i exactement n_i fois) alors on a que $A = BDB^{-1}$.

La diagonalisation d'une matrice, si elle existe, n'est pas unique: on doit *choisir* des matrices B_{λ_i} engendrant les espaces propres E_{λ_i}, puis on doit *choisir* l'ordre dans lequel on place les colonnes de D et de B (bien que souvent on fait en sorte que $|\lambda_1| \geq \ldots \geq |\lambda_k|$). Mais le spectre de la matrice est unique. Cependant, le calcul de $\mathrm{Spec}(A)$ peut être difficile, voir impossible: en effet, on doit non-seulement calculer un déterminant pour obtenir le polynôme caractéristique $p_A(\lambda)$, mais – pire encore – on doit trouver les racines dudit polynôme de degré n—et on peut prouver qu'il n'existe pas d'algorithme pour trouver toutes les racines d'un polynôme quelconque de degré plus grand ou égal à 5! Heureusement, pour beaucoup d'applications, on n'a pas besoin de tout le spectre de A, mais seulement de son plus grand élément; et là on a d'autres méthodes (souvent numériques, donc approximatives) pour effectuer les calculs requis.

Exercices

1. Donner une diagonalisation de chacune des matrices suivantes, ou expliquer pourquoi ce n'est pas possible:

   (a) $\begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & 7 & -6 \\ -1 & 4 & 0 \\ 0 & 2 & -2 \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ (d) $\begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ (e) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 \end{pmatrix}$

2. Donner des factorisations $PA = LU$, $A = QR$ et $A = BDB^{-1}$ des matrices suivantes:

   (a) $A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$ (b) $A = \begin{pmatrix} 2 & 0 & 0 \\ 2 & 1 & 0 \\ 6 & -3 & -2 \end{pmatrix}$ (c) $A = \begin{pmatrix} 2/3 & 2/3 & 1/3 \\ -2/3 & 1/3 & 2/3 \\ 1/3 & -2/3 & 2/3 \end{pmatrix}$

3. Calculer la puissance A^{100} de la matrice $A = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{pmatrix}$.

4. Donner des formules explicites pour les termes des deux suites de nombres, $(x_i)_{i\in\mathbb{N}}$ et $(y_i)_{i\in\mathbb{N}}$, définies récursivement par

$$\left\{ \begin{array}{l} x_0 = 1000 \\ y_0 = 1000 \end{array} \right. \quad \text{et} \quad \left\{ \begin{array}{l} x_{i+1} = x_i + 4y_i \\ y_{i+1} = \dfrac{x_i}{2} \end{array} \right.$$

5. Si on pose $f_0 = 1$, $f_1 = 1$, et $f_{n+2} = f_{n+1} + f_n$ pour tout $n \geq 0$, donner une formule explicite (non-récursive) pour f_n.

11. Symétrie, factorisation $A = QDQ^t$

Quant à la diagonalisation d'une matrice quelconque $A \in \mathbb{R}^{n\times n}$, deux problèmes se posent (hormis le problème du calcul de son spectre):

- pas toute matrice A est diagonalisable,
- on doit calculer une matrice inverse si on veut se servir de la formule $A = BDB^{-1}$.

Pourtant, ces deux problèmes ont une solution commune. D'abord on observe qu'il y a une classe de matrices inversibles *dont l'inverse est trivial à calculer.*

Proposition 11.1 Une matrice carrée $Q \in \mathbb{R}^{n\times n}$ est à colonnes orthonormales si et seulement si Q est inversible et $Q^t = Q^{-1}$; on l'appelle alors une matrice orthogonale (même si ce terme historique peut être source de confusion).

Si on suppose maintenant qu'une matrice $A \in \mathbb{R}^{n\times n}$ est *diagonalisable par une matrice orthogonale* – qu'il existe donc une matrice diagonale $D \in \mathbb{R}^{n\times n}$ et une matrice orthogonale $Q \in \mathbb{R}^{n\times n}$ telles que $A = QDQ^t$ – alors on observe facilement que $A = QDQ^t = (QDQ^t)^t = A^t$.

Définition 11.2 Une matrice $A \in \mathbb{R}^{n\times n}$ est symétrique si $A^t = A$.

Autrement dit: si A est diagonalisable par une matrice orthogonale, alors A est symétrique. De manière suprenante, la réciproque est aussi vrai: toute matrice symétrique $A \in \mathbb{R}^{n\times n}$ est diagonalisable par une matrice orthogonale! Résumons par le théorème suivant, dont – faute de temps – nous ne donnerons pas la démonstration détaillée:

Théorème 11.3 (Théorème Spectral.) Une matrice $A \in \mathbb{R}^{n\times n}$ est symétrique si et seulement si A est diagonalisable par une matrice orthogonale. Soit, pour tout $\lambda_i \in \mathrm{Spec}(A)$, une matrice à colonnes orthonormales $Q_{\lambda_i} \in \mathbb{R}^{n\times n_i}$ telle que $E_{\lambda_i} = \mathrm{Im}(Q_{\lambda_i})$, alors pour les matrices

$$D = \left(\begin{array}{ccc|c|ccc} \lambda_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda_1 & & & & \\ & & & \cdots & & & \\ & & & & \lambda_k & & \\ & & & & & \ddots & \\ & & & & & & \lambda_k \end{array}\right) \quad \text{et} \quad Q = \left(\begin{array}{c|c|c} & & \\ Q_{\lambda_1} & \cdots & Q_{\lambda_k} \\ & & \end{array}\right)$$

(où on répète chaque λ_i exactement n_i fois) on a que $A = QDQ^t$.

Dans la pratique, si $A \in \mathbb{R}^{n\times n}$ est une matrice *symétrique*, alors on peut *toujours* calculer une diagonalisation $A = BDB^{-1}$ par son polynôme caractéristique $p_\lambda(A)$, ses valeurs propres $\mathrm{Spec}(A) = \{\lambda_1, ..., \lambda_k\}$ et ses espaces propres $E_{\lambda_i} = \mathrm{Ker}(A - \lambda_i I) = \mathrm{Im}(B_{\lambda_i})$. De plus, on peut toujours remplacer la matrice inversible B par une matrice orthogonale Q. Le calcul de cette matrice Q se fait par application du procédé de Gram-Schmidt: soit on fait Gram-Schmidt sur chaque B_{λ_i} pour produire des Q_{λ_i} dont l'assemblage est la matrice Q; soit on fait Gram-Schmidt sur la matrice B tout entière pour obtenir la matrice Q. (Ces deux méthodes donnent en effet la même matrice Q.)

Exercices

1. Démontrer que, si $A \in \mathbb{R}^{n\times n}$ est une matrice symétrique, alors A^2 est symétrique. Montrer que la réciproque n'est pas vraie.

2. Démontrer que la transposée d'une matrice orthogonale est orthogonale. Qu'en est-il pour le produit de deux matrices orthogonales? la somme de deux matrices orthogonales? un multiple scalaire d'une matrice orthogonale?

3. Montrer que le déterminant d'une matrice orthogonale est toujours ± 1. Est-ce que toute matrice de déterminant ± 1 est orthogonale?

4. Diagonaliser les matrices symétriques suivantes au moyen d'une matrice orthogonale:

(a) $\begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix}$ (b) $\begin{pmatrix} 3 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 4 \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ (d) $\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$

5. Calculer des factorisations $A = LU$, $A = QR$ et $A = BDB^t$ de la matrice symétrique

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

12. Valeurs singulières, factorisation $A = USV^t$

Dans les deux sections précédentes, on a vu que certaines matrices *carrées* $A \in \mathbb{R}^{n \times n}$ (notamment les symétriques) sont diagonalisables. Dans cette section on verra que *toute* matrice $A \in \mathbb{R}^{m \times n}$ est “diagonalisable”... d'une certaine manière! L'idée est la suivante: on cherche à construire des matrices orthogonales $U \in \mathbb{R}^{m \times m}$ et $V \in \mathbb{R}^{n \times n}$, et une matrice pseudo-diagonale $S \in \mathbb{R}^{m \times n}$, telles que

$$A = USV^t.$$

Pour fixer les notations, on pose

$$S = \left(\begin{array}{ccc|c} s_1 & & & \\ & \ddots & & O \\ & & s_k & \\ \hline & O & & O \end{array}\right) \in \mathbb{R}^{m \times n}$$

avec $s_1 \cdot ... \cdot s_k \neq 0$ (et il est alors évident que $\mathrm{rang}(A) = \mathrm{rang}(S) = k$). Les nombres $s_1, ..., s_k$ sont appelés les *valeurs singulières* de la matrice A, et la factorisation $A = USV^t$ est appelée une *factorisation par valeurs singulières* de A. Dans la suite on indiquera comment on peut calculer une telle factorisation.

Supposons d'abord qu'on dispose d'une factorisation $A = USV^t$ (avec les notations introduites ci-dessus). On observe alors tout de suite que

$$A^tA = VS^tSV^t = V \left(\begin{array}{ccc|c} s_1^2 & & & \\ & \ddots & & O \\ & & s_k^2 & \\ \hline & O & & O \end{array}\right) V^t \tag{1}$$

est *nécessairement* une diagonalisation par matrice orthogonale de la matrice symétrique A^tA. De plus, par $AV = USV^tV = US$ on voit que *nécessairement*

$$U_i = \frac{1}{s_i} AV_i \quad \text{ pour } i \in \{1, \ldots, k\}, \tag{2}$$

c'est à dire, les k premières colonnes de U sont déterminées par les k premières colonnes de V. Quant aux $m - k$ dernières colonnes de U, on observe que $A^tU = (VSU^t)U = VS$, d'où $AU_i = O$ pour $i > k$. C'est à dire, on sait que $\mathrm{Im}(U_{k+1} \mid \cdots \mid U_m) \subseteq \mathrm{Ker}(A^t)$. Mais la matrice $(U_{k+1} \mid \cdots \mid U_m)$ étant à colonnes orthonormales, et puisque $\mathrm{rang}(A^t) = \mathrm{rang}(A) = k$ donc $\mathrm{null}(A^t) = m - k$, on a *nécessairement*

$$\mathrm{Im}(U_{k+1} \mid \cdots \mid U_m) = \mathrm{Ker}(A^t). \tag{3}$$

Il se trouve que, en fait, les trois conditions *nécessaires* décrites ci-dessus sont aussi *suffisantes.* En effet, pour toute matrice $A \in \mathbb{R}^{m \times n}$, la matrice $A^tA \in \mathbb{R}^{n \times n}$ est symétrique et on peut donc la diagonaliser par une matrice orthogonale (avec les techniques des chapitres précédents), soit $A^tA = VDV^t$. En permuttant éventuellement les colonnes dans V et D, on peut toujours placer les valeurs propres non-nulles en premier, pour avoir la diagonalisation

$$A^tA = VDV^t = V \left(\begin{array}{ccc|c} d_1 & & & \\ & \ddots & & O \\ & & d_k & \\ \hline & O & & O \end{array}\right) V^t. \tag{4}$$

avec $d_1 \cdot ... \cdot d_k \neq 0$. On peut montrer que toutes les valeurs propres de A^tA sont positives (voir exercices); ainsi on peut faire en sorte que $d_1 \geq \cdots \geq d_k > 0$ dans la matrice D. La comparaison des formules dans (1) et (4) suggère que l'on retrouve ainsi la matrice V requise pour la factorisation $A = USV^t$, et pour retrouver la matrice S il suffit de poser $s_i := \sqrt{d_i}$. Finalement, s'inspirant des formules dans (2) et (3) on peut toujours former une matrice U par les formules

$$U_i := \frac{1}{\sqrt{d_i}} AV_i \text{ pour } i \in \{1, \ldots, k\} \qquad \text{et} \qquad \mathrm{Im}(U_{k+1}| \ldots |U_m) = \mathrm{Ker}(A^t).$$

Le résultat suivant (dont nous ne donnerons pas une démonstration détaillée, faute de temps) atteste qu'en effet une factorisation $A = USV^t$ est ainsi obtenue:

Théorème 12.1 (Singular Value Decomposition, SVD) Avec la méthode et les notations introduites ci-dessus, on trouve, pour toute matrice $A \in \mathbb{R}^{m\times n}$, des matrices orthogonales $U \in \mathbb{R}^{n\times n}$ et $V \in \mathbb{R}^{m\times m}$ et une matrice pseudo-diagonale $S \in \mathbb{R}^{m\times n}$ (dont les éléments non-nuls sont appelés les valeurs singulières de la matrice A) telles que $A = USV^t$.

Ce résultat a beaucoup d'applications—en voici quelques-unes:

Exemple 12.2 (Low rank approximation) Soit $A = USV^t$ avec U et V orthogonales et S pseudo-diagonale, et notons désormais

$$S = \left(\begin{array}{ccc|c} s_1 & & & \\ & \ddots & & O \\ & & s_k & \\ \hline & O & & O \end{array}\right) \in \mathbb{R}^{m\times n} \qquad \text{avec } s_1 \geq s_2 \geq \ldots \geq s_k > 0.$$

On peut alors calculer explicitement que

$$A = USV^t = \left(\begin{array}{c|c|c} U_1 & \ldots & U_m \end{array}\right) \cdot \left(\begin{array}{ccc|c} s_1 & & & \\ & \ddots & & O \\ & & s_k & \\ \hline & O & & O \end{array}\right) \cdot \left(\begin{array}{c|c|c} V_1 & \ldots & V_n \end{array}\right)^t = \sum_{i=1}^{k} s_i U_i V_i^t \tag{5}$$

et il n'est pas difficile de montrer que $\mathrm{rang}(A) = k$. Pour $k_0 \leq k$ on pose maintenant

$$A_{k_0} = \sum_{i=1}^{k_0} s_i U_i V_i^t, \tag{6}$$

autrement dit, on "coupe la queue" de la somme dans (5), en faisant bien attention (par l'ordre imposé sur les s_i) de n'éliminer que des termes précédés d'un "petit" coefficient. Ainsi cette matrice est "légèrement" différente de A. Par un même raisonnement qu'avant, on peut montrer que $\mathrm{rang}(A_{k_0}) = k_0$. Il est en fait vrai que A_{k_0} est *la meilleure approximation*[5] *de* A *par*

[5]On peut en effet définir la *norme de Frobenius* (d'après l'allemand Ferdinand Georg Frobenius, 1849–1917) d'une matrice $M \in \mathbb{R}^{m\times n}$ par $\|M\|_F = \sqrt{\sum_{i,j} m_{ij}^2}$, puis mesurer la "distance" entre deux matrices $M, N \in \mathbb{R}^{m\times n}$ par la formule $\|M - N\|_F$. L'approximation ici est faite par rapport à cette notion de "distance".

une matrice de rang $k_0 \leq k$ (mais une démonstration de ce résultat est hors portée de ce cours). L'intérêt de remplacer une matrice $A \in \mathbb{R}^{m\times n}$ de rang k donnée par une approximation $A_{k_0} \in \mathbb{R}^{m\times n}$ de rang k_0, réside dans le fait que pour calculer cette dernière matrice avec la formule (6), on n'a besoin que de $k_0(1+m+n)$ nombres, alors que la matrice A en contient $m \times n$. Cette technique est utilisée pour la réduction de la taille de fichiers d'images.

Exemple 12.3 (Pseudo-inverse et inverse) Avec les mêmes notations de l'exemple précédent, on définit

$$S^+ = \left(\begin{array}{ccc|c} s_1^{-1} & & & \\ & \ddots & & O \\ & & s_k^{-1} & \\ \hline & O & & O \end{array}\right) \in \mathbb{R}^{n\times m} \qquad \text{et} \qquad A^+ = VS^+U^t.$$

On dit que A^+ est la matrice pseudo-inverse[6] de A. Soulignons que toute matrice A – carrée ou pas – a un unique pseudo-inverse A^+. Supposons maintenant que $A \in \mathbb{R}^{n\times n}$ est carré et de rang complet—donc inversible. Puisque $\mathrm{rang}(A) = \mathrm{rang}(S)$, il suit que $S \in \mathbb{R}^{n\times n}$ a tous les éléments sur la diagonale strictement positifs. Ainsi on peut calculer que

$$A^{-1} = (USV^t)^{-1} = VS^{-1}U^t = VS^+U^t = A^+.$$

Autrement dit, si A est inversible alors A^+ est son inverse, et on peut le calculer *sans avoir à inverser la matrice* A.

Exemple 12.4 (Moindres carrés—reprise) Toujours avec les notations de l'exemple précédent, on peut calculer que

$$A^tAA^+ = A^t(USV^t)(VS^+U^t) = A^tUSS^+U^t.$$

Mais puisque

$$SS^+ = \left(\begin{array}{ccc|c} s_1s_1^{-1} & & & \\ & \ddots & & O \\ & & s_ks_k^{-1} & \\ \hline & O & & O \end{array}\right) = \left(\begin{array}{ccc|c} 1 & & & \\ & \ddots & & O \\ & & 1 & \\ \hline & O & & O \end{array}\right) \in \mathbb{R}^{m\times m}$$

on a de fait que

$$A^tAA^+ = A^t\sum_{i=1}^{k} U_iU_i^t = A^t\sum_{i=1}^{m} U_iU_i^t = A^tUU^t = A^t.$$

(On utilise que les $m-k$ dernières colonnes de U sont dans le noyau de A^t.) Soit maintenant un système $AX = B$ (avec A de rang quelconque). Pour calculer les solutions (approchées ou exactes) par la méthode des moindres carrés (voir Exemple 8.9), on doit résoudre $A^tAX = A^tB$. Mais par le calcul ci-dessus on voit que

$$A^tAA^+B = A^tB,$$

et donc $X = A^+B$ en est une solution: on peut démontrer que c'est l'unique solution de norme minimale (i.e. minimisant l'expression $\|X\|$). Par conséquent, si $B \in \mathrm{Im}(A)$ alors $X = A^+B$ est l'*unique solution exacte de norme minimale* au système $AX = B$; et si $B \notin \mathrm{Im}(A)$ alors $X = A^+B$ est l'*unique solution approchée au sens des moindres carrés et de norme minimale* au système $AX = B$. Remarquons que, dans tous les cas, $X = A^+B$ peut être calculé *sans avoir à inverser une matrice*!

[6]Attribuée à E. Hastings Moore (1862-1932) et Roger Penrose (1931-...).

Exercices

1. Soit $A \in \mathbb{R}^{m\times n}$ une matrice quelconque. Montrer que toutes les valeurs propres des matrices symétriques A^tA et AA^t sont positives, et que ces deux matrices symétriques ont les mêmes valeurs propres strictement positives.

   Solution. (1) Si X est un vecteur propre de valeur propre λ pour A^tA, alors $X \neq O$ et $A^tAX = \lambda X$; mais alors $X^tA^tAX = X^t\lambda X$ d'où $\|AX\| = \lambda\|X\|$; et toute norme étant positive, on a nécessairement $\lambda \geq 0$. (2) Si $AA^tX = \lambda X$ alors $A^tAA^tX = A^t\lambda X = \lambda A^tX$. Si $X \neq O$ et $\lambda \neq 0$ (et donc $\lambda > 0$) alors $\|A^tX\| = \sqrt{X^tAA^tX} = \sqrt{\lambda}\|X\| \neq 0$, donc $A^tX \neq O$. Ainsi un tel λ est une valeur propre de A^tA (et A^tX est un vecteur propre associé). $\square$

2. Montrer qu'une matrice A est de rang k si et seulement si elle a k valeurs singulières (pas nécessairement distinctes).

3. Montrer qu'on peut également trouver une factorisation par valeurs singulières d'une matrice $A \in \mathbb{R}^{m\times n}$ en diagonalisant la matrice symétrique AA^t (et non pas A^tA comme on a fait pour obtenir le Théorème 12.1). Quel est éventuellement l'intérêt de cette observation?

4. Calculer des factorisations LU et SVD des matrices suivantes:

   (a) $\begin{pmatrix} 1 & 2 & 2 \\ 0 & 2 & -2 \end{pmatrix}$ (b) $\begin{pmatrix} 5 & -5 & 3 \\ -5 & 5 & 3 \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 2 & 2 \end{pmatrix}$

5. Pour comprimer une matrice $m \times n$ par un facteur 4 par le procédé des valeurs singulières, quel sera le rang de la matrice approchée?

6. Soit le système linéaire

$$\begin{cases} \dfrac{3x}{2} - \dfrac{3y}{2} = 3 \\ \sqrt{2}x + \sqrt{2}y = 0 \\ -\dfrac{3x}{2} + \dfrac{3y}{2} = 0 \end{cases}$$

   (a) Ecrire le système sous forme matricielle $AX = B$, calculer le rang de A et montrer que le système n'admet pas de solution exacte.

   (b) Calculer la SVD de la matrice des coefficients du système, soit $A = USV^t$.

   (c) Calculer la matrice pseudo-inverse de A, soit $A^+ = VS^+U^t$.

   (d) Utiliser A^+ pour calculer une solution approchée au sens des moindres carrées du système donné.

7. Soit la matrice $A = \begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix}$.

   (a) Calculer sa factorisation SVD, soit $A = USV^t$.

   (b) Calculer sa pseudo-inverse $A^+ = VS^+U^t$.

   (c) Vérifier que A^+ est la matrice inverse de A.