Vademecum* du Calcul Matriciel en Informatique

Isar Stubbe

(version de 7-11-2018)

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.

Définition 1.1 Une matrice réelle A à m lignes et n colonnes (ou de genre $m \times n$) est un tableau

* *vademecum (vade-mecum, vade mecum)*, *subst. masc.*: Recueil contenant des renseignements sur les règles d'un art ou d'une technique à observer ou sur une conduite à suivre et qu'on garde sur soi ou à portée de main pour le consulter.

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 $A + (-A) = O$.

La matrice $-A$ est la matrice opposée de A; 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. $I_n A = A = AI_m$,
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.

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 constituent le calcul matriciel: tous les calculs que nous ferons dans ce cours, s'y ramènent.

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 5.13 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 les 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!

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 A et B, nous voulons donc calculer tous les X tels que $AX = B$.

Exemple 2.1 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 2.2 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").

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 2.3 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. $\square$

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 2.4 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 2.5 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 2.6 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,
- l'inverse d'une matrice de réduction inférieure est une matrice de réduction inférieure.

Corollaire 2.7 Pour toute matrice $A \in \mathbb{R}^{m\times n}$ il existe une suite finie de matrices élémentaires (donc inversibles) $E_1, \dots, E_k \in \mathbb{R}^{m\times m}$ telle que $A' = E_k \dots E_1 A$ est échelonnée.

Corollaire 2.8 Soit un système linéaire $AX = B$. Si $E_1, \dots, E_k \in \mathbb{R}^{m\times m}$ sont des matrices élémentaires telles que $A' = E_k \dots E_1 A$ est échelonnée, et on pose $B' = E_k \dots E_1 B$, 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 $(A \mid B)$. Ensuite on échelonne la matrice A, tout en appliquant exactement les mêmes opérations élémentaires sur la colonne B. On obtient ainsi $(A' \mid B')$, d'où le système échelonnée $A'X = B'$ à résoudre par substitution.

3. Factorisation $PA = LU$

Proposition 3.1 Toute matrice élémentaire de réduction inférieure est triangulaire inférieure et inversible; et 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). 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. □

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: $A' = E_k \dots E_1A$,
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: $A'' = R_t \dots R_1PA$,
4. on inverse ces réductions (ce qui est facile!) puis on pose $L = R_1^{-1} \dots R_t^{-1}$ et $U = A''$.

4. Forme réduite et calcul de 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 $A' = E_k \dots E_1A$ 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).

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 en première position (si A n'est pas nulle). En effet, si A' est la colonne réduite, alors l'équivalence des systèmes $Ax = O \iff A'x = O$ produit immédiatement ce résultat. Pour $n > 1$, supposons que le résultat est vrai pour toute matrice à $n-1$ colonnes. Si B et C sont deux matrices en

[1] Wilhelm Jordan (1842-1899).

formes réduites associées à 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 $(x_1, \dots, x_n)$ 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. $\square$

Exemple 4.3 La démonstration précédente étant quelque peu technique, nous illustrons le principe sur un exemple (avec un argument légèrement différent). Supposons que $A \in \mathbb{R}^{3\times3}$ 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

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

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

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

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 $I_n = E_k \dots E_1 A$ pour des matrices élémentaires (et donc inversibles) E_i, d'où l'inversibilité de $A = (E_k \dots E_1)^{-1}$ (dont l'inverse est $E_k \dots E_1$). 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, qui est nécessairement aussi solution au système $AX = O$, et ainsi A ne peut pas être inversible. $\square$

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 transformer la partie A en I_n. Si on y parvient, il suit que A est inversible, et on aura transformé la partie I_n en A^{-1}; si on n'y parvient pas, alors A n'est pas inversible.

5. Image et noyau, rang et nullité

Le produit d'une matrice $A = (a_{ij})_{ij} \in \mathbb{R}^{m\times n}$ avec une colonne $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ 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

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

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 5.1 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.

Proposition 5.2 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 5.3 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 5.4 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 $(T_1|\dots|T_n)$ fournit une matrice T telle que $A = BT$. □

Proposition 5.5 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 5.4. De plus, le système $\overline{A}X = O$ ne peut avoir que la solution triviale, car – par construction – il y a autant de pivots que de colonnes. □

Corollaire 5.6 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 5.4. Finalement, $\mathrm{Ker}(\overline{A}) = \mathrm{Ker}(\overline{A'}) = \{O\}$. □

Ce résultat indique comment on peut remplacer une matrice quelconque $A \in \mathbb{R}^{m\times n}$ par une matrice $\overline{A} \in \mathbb{R}^{m\times k}$ à colonnes libres ayant la même image. Mais il peut y avoir beaucoup de matrices différentes ayant la même image que A. Pourtant:

Proposition 5.7 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 $n \le k$.

Démonstration. Puisque $\mathrm{Im}(A) \subseteq \mathrm{Im}(B)$, par le Lemme 5.4 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 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 $AX = O$, et cela contredit l'hypothèse qu'aussi A est à colonnes libres. □

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. Autrement dit, ce nombre est un *invariant* de (l'image de) A, et mérite un nom:

Définition 5.8 Le rang d'une matrice $A \in \mathbb{R}^{m\times n}$, noté $\text{rang}(A)$, est le nombre de colonnes d'une (et donc de toute) matrice à colonnes libres dont l'image est égale à $\text{Im}(A)$.

Corollaire 5.9 Si $A \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{m\times k}$ sont deux matrices et $\text{Im}(A) \subseteq \text{Im}(B)$, alors $\text{rang}(A) \leq \text{rang}(B)$.

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 $\text{rang}(A) = \text{rang}(A')$, mais pas nécessairement $\text{Im}(A) = \text{Im}(A')$!

Il suit en particulier que $\text{rang}(A)$ est toujours inférieur ou égal au nombre de colonnes de A. Voici une propriété surprenante du rang d'une matrice:

Proposition 5.10 Pour toute matrice $A \in \mathbb{R}^{m\times n}$ on a que $\text{rang}(A) = \text{rang}(A^t)$.

Démonstration. Soit une matrice $B \in \mathbb{R}^{m\times k}$ à colonnes libres telle que $\text{Im}(A) = \text{Im}(B)$; on a donc $\text{rang}(A) = k$. Par le Lemme 5.4, 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 $\text{Im}(A^t) \subseteq \text{Im}(T^t)$. Puisque $T^t \in \mathbb{R}^{n\times k}$, ceci implique $\text{rang}(A^t) \leq \text{rang}(T^t) \leq k = \text{rang}(A)$. En remplaçant toute occurence de A par A^t dans ce raisonnement, on obtient aussi l'inégalité $\text{rang}(A) \leq \text{rang}(A^t)$. $\square$

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

Pour terminer cette section, reconsidérons le noyau d'une matrice, $\text{Ker}(A)$. 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 $\text{Im}(M) = \text{Ker}(A)$. La méthode par échelonnement de A pour résoudre $AX = O$ produit "automatiquement" une telle matrice M, comme l'indique l'exemple suivant.

Exemple 5.11 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, M est échelonnée et chaque colonne de M a un pivot, donc M est à colonnes libres.

C'est encore la Proposition 5.7 qui garantit que la définition suivante est bien posée:

Définition 5.12 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 $\mathrm{null}(A)$ est exactement le nombre de paramètres que l'on doit choisir lorsqu'on résout le système homogène $AX = O$ par échelonnement.

Théorème 5.13 (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. $\square$

Ce résultat dit que: "si n inconnus sont liés par r équations indépendantes, alors il reste $n - r$ paramètres intrinsèques".

Corollaire 5.14 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 5.13. 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. $\square$

6. Déterminant

On veut déterminer par le calcul d'un nombre si une matrice $A \in \mathbb{R}^{n\times n}$ est inversible ou pas.

Définition 6.1 Le déterminant d'une matrice $A \in \mathbb{R}^{n\times n}$, noté $\det(A)$, est défini récursivement comme suit:

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

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

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.

On peut en effet démontrer que les $2n$ formules récursives données ci-dessus pour le calcul de $\det(A)$ donnent le même résultat; nous l'admettrons.

Exemple 6.2 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.

Exemple 6.3 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 peut vérifier sur un exemple que son déterminant vaut -1; une matrice de (plusieurs) permutations P a donc un déterminant de ± 1, le signe étant déterminé par la parité du nombre de permutations effectuées.

Théorème 6.4 (Donné sans preuve.) Pour des matrices $A, B \in \mathbb{R}^{n\times n}$ et un scalaire $\alpha \in \mathbb{R}$ on a que

1. $\det(AB) = \det(A)\det(B)$,
2. $\det(\alpha A) = \alpha^n \det(A)$,
3. $\det(A^t) = \det(A)$.

Corollaire 6.5 Une matrice $A \in \mathbb{R}^{n\times n}$ est inversible si et seulement si $\det(A) \neq 0$; dans ce cas, $\det(A^{-1}) = \det(A)^{-1}$.

Démonstration. Soit $A' = E_k \dots E_1 A$ la forme réduite de A, obtenue par opérations élémentaires successives; on sait que $A' = I_n$ si A est inversible, et sinon A' est un matrice triangulaire supérieure contenant un 0 sur sa diagonale. Ainsi $\det(A') \neq 0$ si et seulement si $A' = I_n$, si et seulement si A est inversible. Mais on sait que $\det(A') = \det(E_k) \dots \det(E_1) \det(A)$, 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 $\det(A^{-1}) = \det(A)^{-1}$. □

Exemple 6.6 Soit un échelonnement $A' = E_k \dots E_1 A$ de $A \in \mathbb{R}^{n\times n}$, alors on sait que $\det(A) = \det(E_1)^{-1} \dots \det(E_k)^{-1} \det(A')$. Pour les matrices élémentaires on a calculé les déterminants auparavant; 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 souvent c'est bien plus rapide que l'application de la formule donnée dans la définition même du déterminant.

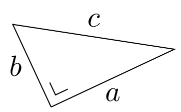
Exemple 6.7 Quand on connait une factorisation $PA = LU$ de $A \in \mathbb{R}^{n\times n}$, alors $\det(A) = \det(P)^{-1} \det(L) \det(U)$ et ces trois facteurs sont faciles à calculer. Souvent les logiciels appliquent cette méthode—mais bien sûr il faut d'abord un bon algorithme pour calculer la factorisation $PA = LU$!

Corollaire 6.8 Pour $A \in \mathbb{R}^{n\times n}$, le système linéaire homogène $AX = O$ admet des solutions non-nulles si et seulement si $\det(A) = 0$.

7. Orthogonalité et distance, moindres carrés

Un système $AX = B$ admet une solution si et seulement si $B \in \mathrm{Im}(A)$. Si $B \notin \mathrm{Im}(A)$, 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, puis développer une technique pour calculer X_0. La géométrie classique nous inspirera.

Le théorème de Pythagore dit qu'un triangle de cotés 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

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

On généralise de façon évidente à n dimensions.

Définition 7.1 Soient deux colonnes $P, Q \in \mathbb{R}^{n\times 1}$. On dira que
- $P^t Q \in \mathbb{R}$ est le produit scalaire de P et Q,
- $\|P\| = \sqrt{P^t P}$ est la norme de P,
- $\|P - Q\|$ est la distance de P à Q,
- P est orthogonal à Q (noté $P \perp Q$) si $P^t Q = 0$.

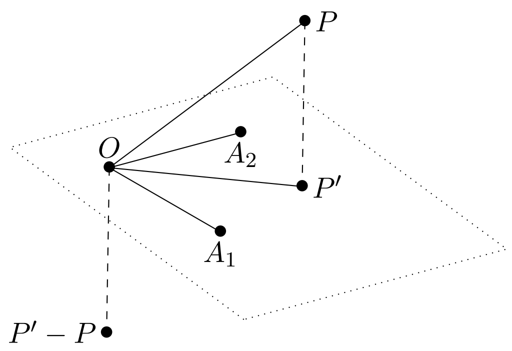
Le produit scalaire étant un cas très particulier du produit matriciel, il est facile de vérifier les propriétés suitvantes.

Proposition 7.2 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. $P^t P \geq 0$ et $P^t P = 0 \Leftrightarrow P = O$ (le produit scalaire est défini positif),
3. $P^t Q = Q^t P$ (le produit scalaire est symétrique).

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

Corollaire 7.3 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 $A^t P = O$.

Exemple 7.4 Dans l'espace $\mathbb{R}^3$ (identifié avec $\mathbb{R}^{3\times 1}$), la projection orthogonale d'un point P sur un plan déterminé par deux directions A_1 et A_2 est l'unique point P' tel que $P' = Ax_2 + A_2 x_2$ et $P' - P$ est orthogonal à A_1 et à A_2:



Autrement dit, si on pose $A = (A_1|A_2)$ la matrice ayant A_1 et A_2 comme colonnes, et $X = \binom{x_1}{x_2}$, alors P' est tel que $P' = AX$ et $A^t(AX - P) = O$. On voit donc que ce X est l'unique solution au système linéaire $A^t AX = A^t P$.

Cette intuition géométrique justifie la généralisation suivante.

Définition 7.5 Soit une matrice non-nulle $A \in \mathbb{R}^{n\times k}$. La projection orthogonale de $P \in \mathbb{R}^{n\times 1}$ sur $\mathrm{Im}(A)$ est un (unique) élément $\mathrm{proj}_A(P)$ de $\mathrm{Im}(A)$ tel que $\mathrm{proj}_A(P) - P$ soit orthogonal à tout élément de $\mathrm{Im}(A)$.

Partant de cette définition, par le Corollaire 7.3 on voit facilement que $\mathrm{proj}_A(P) = AX$ pour un X qui est nécessairement solution au système linéaire $A^tAX = A^tP$. On peut supposer que A soit de colonnes libres (sinon on remplace A par une matrice à colonnes libres $\overline{A}$ telle que $\mathrm{Im}(\overline{A}) = \mathrm{Im}(A)$), et il suit alors que $\mathrm{proj}_A(P)$ est en effet déterminé de façon unique:

Lemme 7.6 Pour toute matrice $A \in \mathbb{R}^{n\times k}$ on a que $\mathrm{Ker}(A^tA) = \mathrm{Ker}(A)$ (et, par dualité, aussi $\mathrm{Ker}(AA^t) = \mathrm{Ker}(A^t)$).

Démonstration. Si $AX = O$ alors $A^tAX = O$. Réciproquement, si $A^tAX = O$ alors $\|AX\| = X^tA^tAX = X^tO = 0$ et donc $AX = O$. □

Proposition 7.7 Une matrice $A \in \mathbb{R}^{n\times k}$ est de rang k si et seulement si $A^tA \in \mathbb{R}^{k\times k}$ est inversible. Dans ce cas, 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 précédent, $\mathrm{null}(A) = \mathrm{null}(A^tA)$, donc par le Théorème du Rang aussi $\mathrm{rang}(A) = \mathrm{rang}(A^tA)$. Ainsi $\mathrm{rang}(A) = k$ si et seulement si $A^tA \in \mathbb{R}^{k\times k}$ est inversible. □

De manière équivalente, on peut caractériser $\mathrm{proj}_A(P)$ comme étant l'unique point P' de $\mathrm{Im}(A)$ pour lequel la distance entre P et P' est minimale, comme le dit la proposition suivante.

Proposition 7.8 Soit une matrice $A \in \mathbb{R}^{n\times k}$, et $P' \in \mathbb{R}^{n\times 1}$ la projection orthogonale de $P \in \mathbb{R}^{n\times 1}$ sur $\mathrm{Im}(A)$. Pour tout $Q \in \mathrm{Im}(A)$ on a que $\|P' - P\| \leq \|Q - P\|$, l'égalité étant obtenue uniquement si $Q = P'$.

Démonstration. Pour tout $Q \in \mathrm{Im}(A)$ on peut calculer que

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

puisque $P' - P$ est orthogonal à tous les éléments de $\mathrm{Im}(A)$, dont $Q - P'$. On a donc également 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 $Q = P'$. □

Enfin on peut répondre à la question initiale de ce chapitre:

Exemple 7.9 (Méthode des moindres carrés.) Soit $AX = B$ un système linéaire quelconque. Si une solution exacte existe, alors elle exprime comment B est une combinaison linéaire des colonnes de A; si aucune solution exacte n'existe, alors B ne peut pas être écrit comme une combinaison linéaire des colonnes de A. Mais 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* une solution. Si $A \in \mathbb{R}^{n\times k}$ est de rang $k \leq n$, cela revient à résoudre le système linéaire $A^tAX = A^tB$, et ce système admet toujours une *unique* solution $X = (A^tA)^{-1}A^tB$. Puisque cette solution minimise l'expression $\|AX - B\|^2$ (qui est un polynôme de degré 2 en n variables), on l'appelle la solution approchée du système $AX = B$ au sens des moindres carrés. Et si $AX = B$ admet une (et une seule, car $\mathrm{null}(A) = 0$ par hypothèse) solution exacte, la solution approchée lui est identique.

8. 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

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

En mots, chaque colonne de A est de norme 1 ("normale"), et orthogonale à toute autre colonne.

Définition 8.1 Une matrice $A \in \mathbb{R}^{n\times k}$ est à colonnes orthonormales si $A^tA = I_k$.

Proposition 8.2 Si une matrice $A \in \mathbb{R}^{n\times k}$ est à colonnes orthonormales, alors $\mathrm{rang}(A) = k$.

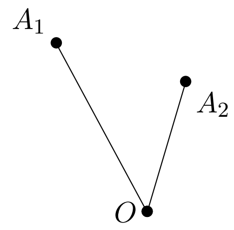
Démonstration. Si $AX = O$ alors $X = A^tAX = A^tO = O$ donc $\mathrm{null}(A) = 0$; le résultat suit par le Théorème du Rang. □

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 $P \in \mathbb{R}^{n\times 1}$,

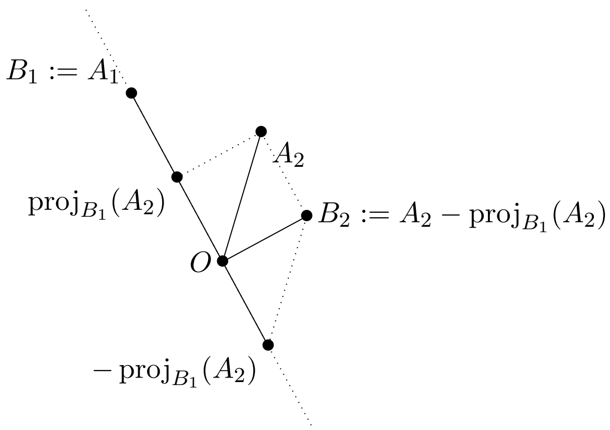
$$\mathrm{proj}_A(P) = \mathrm{proj}_Q(P) = QQ^tP.$$

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

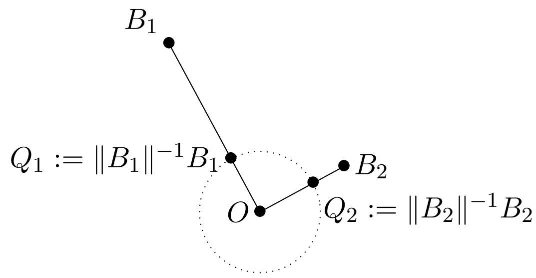
Exemple 8.3 Soit $A \in \mathbb{R}^{3\times 2}$ une matrice dont les colonnes, notées A_1 et A_2, déterminent un plan dans $\mathbb{R}^3$; on dessinera ci-dessous dans ce plan. D'abord on peut représenter les colonnes de A:



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

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

Puisqu'on a

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

ces formules sont équivalentes à

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

et cela s'écrit de façon matricielle comme

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

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 prouve 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 8.4 (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 8.5 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 8.6 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 8.7 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 $X = R^{-1}Q^tB$, que l'on peut calculer en résolvant le système (triangulaire supérieure, donc facile) $RX = Q^tB$.

9. 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 9.1 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 9.2 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

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

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; c'est à dire, ceux 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 9.3 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 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 9.4 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 $A = BDB^{-1}$.

Démonstration. On peut toujours calculer les matrices à colonnes libres $B_{\lambda_i} \in \mathbb{R}^{n\times n_i}$, puis on peut toujours assembler la matrice $B = (B_{\lambda_1}|\cdots|B_{\lambda_k})$: c'est une matrice à $n_1 + \cdots + n_k$ colonnes, et le lemme ci-dessous montre que B est, elle aussi, à colonnes libres (et nécessairement $n_1 + \cdots + n_k \leq n$). Il est alors clair que $AB = DB$ avec les notations du théorème, mais attention: B et D ne sont pas nécessairement des matrices carrées. Il suit cependant que B et D sont de genre $n \times n$ si et seulement si $n_1 + \cdots + n_k = n$; et dans ce cas D est diagonale et B est inversible (car à colonnes libres), d'où la formule $A = BDB^{-1}$. □

Lemme 9.5 Soit une matrice $A \in \mathbb{R}^{n\times n}$, $\mathrm{Spec}(A) = \{\lambda_1, \ldots, \lambda_k\}$ et $\mathrm{Im}(B_{\lambda_i}) = E_{\lambda_i}$ pour des matrices à colonnes libres $B_{\lambda_i} \in \mathbb{R}^{n\times n_i}$. La matrice $B = (B_{\lambda_1}|\cdots|B_{\lambda_k})$ est alors aussi à colonnes libres.

Démonstration. (Preuve par récurrence sur k, le nombre de valeurs propres distinctes.) Si $k = 1$ il n'y a rien à montrer. Sinon, en supposant que le lemme est vrai pour $k - 1$ valeurs propres distinctes, considérons la situation où $BX = O$ pour un certain $X \in \mathbb{R}^{n\times 1}$ avec $n = \sum_{i=1}^{k} n_i$. Il suit alors que $ABX = O$, mais $AB = (\lambda_1 B_{\lambda_1}|\cdots|\lambda_{k-1}B_{\lambda_{k-1}}|\lambda_k B_{\lambda_k})$ puisque les colonnes de B_{λ_i} sont des vecteurs propres pour la valeur propre λ_i. Il suit également que $\lambda_k BX = O$, et $\lambda_k B = (\lambda_k B_{\lambda_1}|\cdots|\lambda_k B_{\lambda_{k-1}}|\lambda_k B_{\lambda_k})$. Ainsi, par soustraction on obtient $((\lambda_k - \lambda_1)B_{\lambda_1}|\cdots|(\lambda_k - \lambda_{k-1})B_{\lambda_{k-1}}|O)X = O$, et par hypothèse d'induction les $n - n_k$ premiers éléments de X doivent être nuls. Mais alors aussi les n_k derniers éléments doivent être nuls, car B_{λ_k} est à colonnes libres. $\square$

Attention: il n'est pas vrai qu'un assemblage $(U_1|\cdots|U_k)$ de matrices à colonnes libres $U_1,\ldots,U_k$ est toujours à colonnes libres; dans la démonstration ci-dessus il est crucial que les matrices $B_{\lambda_1},\ldots,B_{\lambda_k}$ que l'on assemble sont (à colonnes libres et) ont pour image les différents espaces propres d'une matrice A!

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 (et peut être calculé même si A n'est pas diagonalisable), et contient beaucoup d'information utile concernant la matrice.

Exemple 9.6 Si une matrice $A \in \mathbb{R}^{n\times n}$ a n valeurs propres distinctes, alors elle est diagonalisable.

Exemple 9.7 Si $A \in \mathbb{R}^{n\times n}$ est diagonalisable, alors (avec les notations du théorème) $\det(A) = \det(D) = \lambda_1^{n_1}\ldots\lambda_k^{n_k}$. On a également que $\mathrm{tr}(A) := a_{11} + \cdots + a_{nn} = n_1\lambda_1 + \cdots + n_k\lambda_k$; c'est la trace de A. Soulignons que l'on ne doit pas calculer B^{-1} pour cette application!

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. Ou bien on travaille avec des matrices dont la forme fait que son polynôme caractéristique est trivial à résoudre.

Exemple 9.8 Soit $A \in \mathbb{R}^{n\times n}$ une matrice triangulaire (supérieure ou inféreure), alors $\mathrm{Spec}(A)$ contient exactement tous les éléments distincts de la diagonale de A.

10. 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): (1) pas toute matrice A est diagonalisable, et (2) 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 10.1 Une matrice $Q \in \mathbb{R}^{n\times n}$ 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 10.2 Une matrice $A \in \mathbb{R}^{n\times n}$ est symétrique si $A^t = A$.

De manière suprenante, il est aussi vrai que toute matrice symétrique est diagonalisable, et ce par une matrice orthogonale.

Théorème 10.3 (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 $A = QDQ^t$.

Esquisse de démonstration. Par la théorie des polynômes on peut montrer, pour A symétrique, que le polynôme $p_A(\lambda)$ se décompose en facteurs linéaires (sur $\mathbb{R}$); cela implique que A admet au moins une valeur propre réelle, et donc aussi un vecteur propre. Par la théorie des orthocompléments (que nous n'avons pas le temps d'aborder dans ce cours) on peut ensuite prouver, par récurrence sur n, qu'une matrice symétrique $A \in \mathbb{R}^{n\times n}$ est nécessairement diagonalisable; nous l'admettons. Soit donc $A = BDB^{-1}$ une diagonalisation; rappelons que $B = (B_{\lambda_1}|\cdots|B_{\lambda_k})$ avec $\mathrm{Im}(B_{\lambda_i}) = E_{\lambda_i}$, et chaque matrice $B_{\lambda_i} \in \mathbb{R}^{n\times n_i}$ est à colonnes libres. Par le procédé de Gram-Schmidt on peut calculer des matrices à colonnes orthonormales (donc nécessairement à colonnes libres) $Q_{\lambda_i} \in \mathbb{R}^{n\times n_i}$ telles que $\mathrm{Im}(Q_{\lambda_i}) = \mathrm{Im}(B_{\lambda_i}) = E_{\lambda_i}$; l'assemblage $Q = (Q_{\lambda_i}|\cdots|Q_{\lambda_k})$ satisfait donc automatiquement à $A = QDQ^{-1}$ aussi. Enfin, le lemme ci-dessous montre que la matrice Q est aussi à colonnes orthonormales, et donc orthogonale (puisque $n \times n$)—d'où la formule $A = QDQ^t$. $\square$

Lemme 10.4 Soit $A \in \mathbb{R}^{n\times n}$ une matrice symétrique, $\mathrm{Spec}(A) = \{\lambda_1, \ldots, \lambda_k\}$, et $\mathrm{Im}(Q_{\lambda_i}) = E_{\lambda_i}$ pour des matrices à colonnes orthonormales $Q_{\lambda_i} \in \mathbb{R}^{n_i\times n}$. La matrice $Q = (Q_{\lambda_i}|\cdots|Q_{\lambda_k})$ est alors aussi à colonnes orthonormales.

Démonstration. Si U et V sont deux colonnes de Q, on distingue trois cas. (1) Si $U = V$, alors $U^tV = 1$ puisque toute colonne de Q est une colonne d'un des Q_{λ_i} dont les colonnes sont de norme 1. (2) Si $U \neq V$ sont des colonnes d'une même composante Q_{λ_i}, alors $U^tV = 0$ car chaque Q_{λ_i} est à colonnes orthonormales. (3) Si U est une colonne de Q_{λ_i}, et V est une colonne de Q_{λ_j}, et $\lambda_i \neq \lambda_j$, alors

$$\lambda_i(U^tV) = (\lambda_iU)^tV = (AU)^tV = U^t(A^tV) \overset{(*)}{=} U^t(AV) = U^t(\lambda_jV) = \lambda_j(U^tV)$$

implique que $U^tV = 0$ (et notons que $A^t = A$ est crucial pour $(*)$). $\square$

Attention: pour toute matrice diagonalisable, disons $A = BDB^{-1}$ avec $B = (B_{\lambda_1}|\cdots|B_{\lambda_k})$, on peut appliquer le Procédé Gram-Schmidt sur chaque B_{λ_i} pour obtenir des matrices à colonnes orthonormales Q_{λ_i}; bien que leur assemblage $Q = (Q_{\lambda_1}|\cdots|Q_{\lambda_k})$ satisfera à $A = QDQ^{-1}$, il n'est pas vrai que Q est orthogonale, sauf si A est symétrique. Et par ailleurs, si on applique le Procédé de Gram-Schmidt sur l'assemblage $B = (B_{\lambda_1}|\cdots|B_{\lambda_k})$, on obtiendra certainement une matrice orthogonale Q, mais elle ne diagonalisera pas la matrice A, sauf si cette dernière est symétrique. Par contre, pour A symétrique, au lieu d'appliquer le Procédé de Gram-Schmidt sur chaque B_{λ_i} pour obtenir les Q_{λ_i} qui forment finalement Q, on peut aussi appliquer le Procédé de Gram-Schmidt sur la matrice B assemblée des B_{λ_i}: on obtiendra la même matrice Q.

11. 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 $s_1 \cdot \ldots \cdot s_k \neq 0$. Si on a une telle factorisation, alors on observe tout de suite que

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

est nécessairement une diagonalisation par matrice orthogonale de la matrice symétrique A^tA. De la même manière on vérifie facilement qu'aussi

$$AA^t = USS^tU^t = U \left(\begin{array}{ccc|c} s_1^2 & & & \\ & \ddots & & O \\ & & s_k^2 & \\ \hline & O & & O \end{array}\right) U^t$$

est une diagonalisation par matrice orthogonale. De plus, les matrices orthogonales $U \in \mathbb{R}^{m\times m}$ et $V \in \mathbb{R}^{n\times n}$ sont intimement liées, malgré la différence de genre. En effet, de $AV = USV^tV = US$ on obtient que

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

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 (par le Lemme 7.6) que

$$\mathrm{Im}(U_{k+1}|\dots|U_m) = \text{espace propre pour valeur propre 0 de } AA^t = \mathrm{Ker}(AA^t) = \mathrm{Ker}(A^t);$$

autrement dit, on les retrouve par résolution du système $A^tX = O$.

Dans la suite on fixe une matrice $A \in \mathbb{R}^{m\times n}$ quelconque, et on va montrer que, réciproquement, la connaissance d'une diagonalisation par matrice orthogonale de A^tA et du noyau de A^t suffisent pour trouver une factorisation $A = USV^t$ avec U et V orthogonale et S pseudo-diagonale.

D'abord un petit resultat innocent mais crucial.

Lemme 11.1 La matrice symétrique $A^tA \in \mathbb{R}^{n\times n}$ a toutes ses valeurs propres positives. (Par dualité la même chose est vraie pour $AA^t \in \mathbb{R}^{m\times m}$.)

Démonstration. 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 $\lambda \geq 0$. $\square$

Ainsi, on peut (avec les techniques des chapitres précédents) calculer une diagonalisation par matrice orthogonale

$$A^tA = VDV^t,$$

et pour fixer les notations on pose

$$D = \left(\begin{array}{ccc|c} d_1 & & & \\ & \ddots & & O \\ & & d_k & \\ \hline & O & & O \end{array}\right)$$

avec $d_1 \geq \dots \geq d_k > 0$. Notons ici que les $n-k$ dernières colonnes de V sont donc des vecteurs propres de A^tA pour la valeur propre nulle.

Ensuite, pour $i \in \{1, \dots, k\}$ on pose

$$U_i := \frac{1}{\sqrt{d_i}}AV_i$$

et on vérifie facilement que

$$(AA^t)U_i = d_i U_i \quad \text{et} \quad U_i^t U_j = \begin{cases} 1 \text{ si } i = j, \\ 0 \text{ sinon.} \end{cases}$$

C'est à dire, ces colonnes orthonormales deux-à-deux $U_1, \ldots, U_k$ sont des vecteurs propres de la matrice symétrique $AA^t \in \mathbb{R}^{m\times m}$ associés aux valeurs propres non-nulles $d_1, \ldots, d_k$. En fait, AA^t ne peut avoir d'autres valeurs propres non-nulles:

Lemme 11.2 Toute valeur propre (nécessairement positive) non-nulle de $AA^t \in \mathbb{R}^{m\times m}$ est aussi une valeur propre (positive non-nulle) de $A^tA \in \mathbb{R}^{n\times n}$. (Par dualité la réciproque est aussi vraie.)

Démonstration. Si $AA^tX = \lambda X$ alors $A^tAA^tX = A^t\lambda X = \lambda A^tX$. Si $X \neq O$ et $\lambda \neq 0$ (donc $\lambda > 0$ par un lemme ci-dessus) 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). □

Pour former une matrice orthogonale $U \in \mathbb{R}^{m\times m}$ diagonalisant la matrice symétrique $AA^t \in \mathbb{R}^{m\times m}$, il faut donc ajouter aux k premières colonnes orthonormales $U_1, \ldots, U_k$ éventuellement des vecteurs propres pour la valeur propre nulle. C'est à dire, il faut calculer les $m-k$ colonnes orthonormales $U_{k+1}, \ldots, U_m$ telles que $\mathrm{Im}(U_{k+1}|\ldots|U_m) = \mathrm{Ker}(AA^t)$. Par le Lemme 7.6 il suffit de calculer des colonnes orthonormales $U_{k+1}, \ldots, U_m$ telles que

$$\mathrm{Ker}(A^t) = \mathrm{Im}(U_{k+1}|\ldots|U_m)$$

pour compléter la matrice U.

Théorème 11.3 (Singular Value Decomposition, SVD) Avec les notations introduites ci-dessus, on définit la matrice pseudo-diagonale

$$S := \left(\begin{array}{ccc|c} \sqrt{d_1} & & & \\ & \ddots & & O \\ & & \sqrt{d_k} & \\ \hline & O & & O \end{array}\right) \in \mathbb{R}^{m\times n}$$

dont les éléments non-nuls sont appelés les valeurs singulières de la matrice $A \in \mathbb{R}^{m\times n}$. Il suit alors que $A = USV^t$, avec U et V des matrices orthogonales.

Démonstration. Nous avons déjà montré que U et V sont des matrices orthogonales. Il suffit donc de calculer que

$$
\begin{aligned}
USV^t &= \left(\begin{array}{c|c|c} U_1 & \dots & U_m \end{array}\right) \cdot \left(\begin{array}{ccc|c} \sqrt{d_1} & & & \\ & \ddots & & O \\ & & \sqrt{d_k} & \\ \hline & O & & O \end{array}\right) \cdot \left(\begin{array}{c|c|c} V_1 & \dots & V_n \end{array}\right)^t \\
&= \left(\begin{array}{c|c|c} U_1 & \dots & U_m \end{array}\right) \cdot \left(\begin{array}{ccc|c} \sqrt{d_1} & & & \\ & \ddots & & O \\ & & \sqrt{d_k} & \\ \hline & O & & O \end{array}\right) \cdot \left(\begin{array}{c} V_1^t \\ \hline \vdots \\ \hline V_n^t \end{array}\right) \\
&= \left(\begin{array}{c|c|c} U_1 & \dots & U_m \end{array}\right) \cdot \left(\begin{array}{c} \sqrt{d_1}V_1^t \\ \hline \vdots \\ \sqrt{d_k}V_k^t \\ \hline O \end{array}\right) \\
&= U_1\sqrt{d_1}V_1^t + \dots + U_k\sqrt{d_k}V_k^t \\
&= \sum_{i=1}^k \sqrt{d_i}U_iV_i^t \\
&= \sum_{i=1}^k \sqrt{d_i}\frac{1}{\sqrt{d_i}}AV_iV_i^t \\
&= \sum_{i=1}^k AV_iV_i^t \\
&\overset{*}{=} \sum_{i=1}^n AV_iV_i^t \\
&= A\Big(\sum_{i=1}^n V_iV_i^t\Big) \\
&= A\left(\begin{array}{c|c|c} V_1 & \dots & V_n \end{array}\right) \cdot \left(\begin{array}{c} V_1^t \\ \hline \vdots \\ \hline V_n^t \end{array}\right) \\
&= AVV^t \\
&= AI_n \\
&= A.
\end{aligned}
$$

L'égalité $(*)$ suit parce que $AV_i = O$ pour $i > k$, les colonnes $V_{k+1}, \dots, V_n$ appartenant à $\mathrm{Ker}(A^tA) = \mathrm{Ker}(A)$. $\square$

Dans ce théorème, il est important de construire la matrice orthogonale U à partir de la diagonalisation $A^tA = VDV^t$ par la formule $U_i = \frac{1}{\sqrt{d_i}}AV_i$ (pour les k premières colonnes). En effet, il peut y avoir d'autres diagonalisations de AA^t, disons par une matrice orthogonale U'—mais il se pourrait alors que $U'_i \neq \frac{1}{\sqrt{d_i}}AV_i$, et donc aussi $A \neq U'SV^t$!

Par contre, au lieu de construire la décomposition $A = USV^t$ à partir d'une diagonalisation $A^tA = VDV^t$, on peut aussi partir d'une diagonalisation $AA^t = UDU^t$, en déduire une décomposition $A^t = VSU^t$, et donc par transposition $A = USV^t$. Pour réduire les calculs, on a intérêt de choisir parmi A^tA et AA^t la matrice la plus petite!

Exemple 11.4 Pour toute matrice $A \in \mathbb{R}^{m\times n}$ on sait donc qu'il existe $U \in \mathbb{R}^{m\times m}$ et $V \in \mathbb{R}^{n\times n}$ orthogonales et $S \in \mathbb{R}^{m\times n}$ pseudo-diagonale, notée désormais

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

telles que

$$A = \sum_{i=1}^{k} s_i U_i V_i^t.$$

On montre que $\operatorname{rang}(A) = k$: puisque $\operatorname{Im}(A) = \operatorname{Im}(USV^t) = \operatorname{Im}(US)$ (car V^t est inversible) on a $\operatorname{rang}(A) = \operatorname{rang}(US)$; mais $\operatorname{Im}(S^tU^t) = \operatorname{Im}(S^t)$ (puisque U^t est inversible) donc $\operatorname{rang}(US) = \operatorname{rang}((US)^t) = \operatorname{rang}(S^t) = \operatorname{rang}(S)$; et le résultat suit puisque S est échelonnée et le nombre de lignes non-nulles est égal à k. Pour $k_0 \leq k$ on pose maintenant

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

autrement dit, on "coupe la queue" de la première somme, 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 $\operatorname{rang}(A_{k_0}) = k_0$. Il est en fait vrai que A_{k_0} est *la meilleure approximation de A par 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 $(*)$, 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 11.5 Avec les mêmes hypothèses et 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 de A, et on peut calculer que

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

Mais puisque

$$SS^+ = \left(\begin{array}{ccc|c} s_1 s_1^{-1} & & & \\ & \ddots & & O \\ & & s_k s_k^{-1} & \\ \hline & O & & O \end{array}\right) = \left(\begin{array}{ccc|c} 1 & & & \\ & \ddots & & O \\ & & 1 & \\ \hline & O & & O \end{array}\right) \in \mathbb{R}^{m\times m}$$

on peut calculer (comme dans la démonstration du théorème ci-dessus) que

$$A^t A A^+ = A^t \sum_{i=1}^{k} U_i U_i^t = A^t \sum_{i=1}^{m} U_i U_i^t = A^t U U^t = A^t.$$

Pour une matrice A de rang complet on sait (par la Proposition 7.7) que $A^t A$ est inversible, et donc $A^t A A^+ = A^t$ implique $A^+ = (A^t A)^{-1} A^t$. Par conséquent, la solution approchée au sens des moindres carrés d'un système $AX = B$ (toujours avec A de rang complet) est exactement donnée par $X = A^+ B$ *que l'on peut donc calculer sans avoir à inverser une matrice.*

Exemple 11.6 Avec les notations des exemples précédents, supposons 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

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

et *on trouve l'inverse de A sans avoir à inverser une matrice*!