Université Mohammed V de Rabat
Faculté des Sciences

Département de Mathématiques
Laboratoire de Mathématiques, Statistique et Applications
Master: Statistique et Econométrie

POLYCOPIÉ: ALGÈBRE POUR LA STATISTIQUE

Nabil Machrafi

Années universitaires: 2018-2019 et 2019-2020

Remerciements

Je souhaite exprimer mes remerciements à **Pr. Zine El Abidine Abdelali** pour les discussions et améliorations en relation avec le contenu de ce polycopié.

Table des matières

Chapitre Préliminaire

Dans le contexte de la statistique, les données observées sont souvent sous forme de mesures quantitatives représentées dans des matrices. Par suite, et sauf mention contraire, les matrices considérées sont des matrices réelles (i.e. à coefficients dans $\mathbb{R}$) non nulles. On adopte la notation ${}^tA$ pour la transposée d'une matrice A. La taille de cette dernière est notée (m, n) indiquant que A a m lignes et n colonnes. Le cas où $m = n$, A est une matrice carrée et est dite simplement de taille n. On note $\mathcal{M}_{m,n}(\mathbb{R})$ l'espace vectoriel des matrices réelles de taille (m, n). On identifiera $\mathbb{R}^n$ à $\mathcal{M}_{n,1}(\mathbb{R})$; ainsi, lors du calcul matriciel les vecteurs sont des vecteurs colonnes.

Les résultats non démontrés dans ce document sont standards et peuvent être trouvés dans les ouvrages de l'algèbre linéaire et de l'analyse matricielle; voir par exemple [1, 3].

0.1 Produit matriciel

Soient $A = (a_{ij})$ et $B = (b_{ij})$ deux matrices de taille (m, n) et (n, p) respectivement. Le produit des matrices A et B est la matrice de taille (m, p) définie par:

$$AB = (c_{ij}) \text{ avec } c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} \ , \ 1 \leq i \leq m, \ 1 \leq j \leq p.$$

Les matrices A et B sont dites dans ce cas compatibles pour la multiplication.

0.1.1 Cas particuliers:

1. Produit tensoriel de deux vecteurs colonnes $x = {}^t(x_1, ..., x_m)$ et $y = {}^t(y_1, ..., y_n)$:

$$\begin{aligned} x \otimes y \quad : &= x\ {}^t y \\ &= \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{bmatrix} = (x_i y_j)_{\substack{1 \le i \le m \\ 1 \le j \le n}} . \end{aligned}$$

   Le résultat étant, pour deux vecteurs non nuls, une matrice de taille (m, n) et de rang égal à 1.

2. Produit scalaire (canonique) sur $\mathbb{R}^n$:
   Pour $x = {}^t(x_1, ..., x_n)$ et $y = {}^t(y_1, ..., y_n)$, le produit scalaire de x et y est le scalaire:

$$\langle x, y \rangle := {}^t x . y = \sum_{i=1}^{n} x_i y_i .$$

Remarque 0.1 *1. Le produit tensoriel est défini pour n'importe quels deux vecteurs colonnes, alors que le produit scalaire n'est défini que pour deux vecteurs colonnes de même taille.*

2. Contrairement au produit tensoriel, le produit scalaire sur $\mathbb{R}^n$ est symétrique, i.e.,

$$\langle x, y \rangle = \langle y, x \rangle \ \ , \ x, y \in \mathbb{R}^n .$$

3. Pour $x, y \in \mathbb{R}^n$, on a la relation

$$tr\,(x \otimes y) = tr(x\ {}^t y) = \langle x, y \rangle .$$

La notation $tr A$ désigne la trace de la matrice A.

Plus de détails sur le produit scalaire seront présentés dans le chapitre qui suit.

0.1.2 Différentes expressions du produit matriciel:

Dans toute la suite, on adoptera pour une matrice A de taille (m, p) l'écriture suivante

$$A = \begin{bmatrix} a_{*1} & \cdots & a_{*p} \end{bmatrix} = \begin{bmatrix} {}^t a_{1*} \\ \vdots \\ {}^t a_{m*} \end{bmatrix}$$

où l'on a exprimé A par concaténation de vecteurs colonnes $a_{*i} \in \mathbb{R}^m$ et de vecteurs lignes ${}^t a_{i*},\ a_{i*} \in \mathbb{R}^p$. Si $B = \begin{bmatrix} b_{*1} & \cdots & b_{*n} \end{bmatrix} = \begin{bmatrix} {}^t b_{1*} \\ \vdots \\ {}^t b_{p*} \end{bmatrix}$ est de taille (p, n), Alors le produit matriciel des deux matrices A et B peut être exprimé

1. via le produit scalaire:

$$AB = \begin{bmatrix} {}^t a_{1*} \\ \vdots \\ {}^t a_{m*} \end{bmatrix} \begin{bmatrix} b_{*1} & \cdots & b_{*n} \end{bmatrix} = \begin{bmatrix} {}^t a_{1*} b_{*1} & \cdots & {}^t a_{1*} b_{*n} \\ \vdots & & \vdots \\ {}^t a_{m*} b_{*1} & \cdots & {}^t a_{m*} b_{*n} \end{bmatrix}.$$

2. via le produit tensoriel:

$$AB = \begin{bmatrix} a_{*1} & \cdots & a_{*p} \end{bmatrix} \begin{bmatrix} {}^t b_{1*} \\ \vdots \\ {}^t b_{p*} \end{bmatrix} = \sum_{i=1}^{p} a_{*i} \otimes b_{i*}\ ,$$

   chaque terme $a_{*i} \otimes b_{i*}$ étant une matrice de taille (m, n).

3. Par concaténation de vecteurs colonnes (resp. lignes):

   (a)

$$AB = A \begin{bmatrix} b_{*1} & \cdots & b_{*n} \end{bmatrix} = \begin{bmatrix} Ab_{*1} & \cdots & Ab_{*n} \end{bmatrix}\ ,$$

   chaque Ab_{*i} étant un vecteur colonne de taille m.

(b)

$$AB = \begin{bmatrix} {}^t a_{1*} \\ \vdots \\ {}^t a_{m*} \end{bmatrix} B = \begin{bmatrix} {}^t a_{1*} B \\ \vdots \\ {}^t a_{m*} B \end{bmatrix},$$

chaque ${}^t a_{i*} B$ étant un vecteur ligne de taille n.

0.2 Matrices par blocs

Soit $A = \begin{bmatrix} a_{*1} & \cdots & a_{*n} \end{bmatrix}$ une matrice de taille (m, n). Remarquons que A peut être vue comme une matrice par blocs a_{*i} qui sont des matrices de taille $(m, 1)$. Similairement, A peut être vue comme une matrice par blocs de matrices de taille $(1, n)$ (concaténation de vecteurs lignes). On peut aussi faire un partitionnement de taille $(2, 2)$ (deux partitions des lignes et deux partitions des colonnes) de A en la matrice par blocs:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{0.1}$$

avec A_{11} est une matrice de taille (m_1, n_1), A_{12} est de taille (m_1, n_2), A_{21} est de taille (m_2, n_1) et A_{22} est de taille (m_2, n_2). On a donc $m = m_1 + m_2$ et $n = n_1 + n_2$.

Proposition 0.2 *Soit A une matrice de taille (m, n) partitionnée en une matrice par blocs de la forme* (0.1)*. Alors la transposée de A est donné par*

$${}^t A = \begin{bmatrix} {}^t A_{11} & {}^t A_{21} \\ {}^t A_{12} & {}^t A_{22} \end{bmatrix}$$

(en faisant la transposée de chacun des quatre blocs et en permutant les positions des matrices sur l'anti-diagonale).

On dit que deux matrices par blocs A et B sont conformément partitionnées pour l'addition si leurs blocs correspondants sont des matrices de même taille.

Ainsi, si $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ et $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$ sont conformément partitionnées pour l'addition, alors pour $\alpha, \beta \in \mathbb{R}$ on a

$$\alpha A + \beta B = \begin{bmatrix} \alpha A_{11} + \beta B_{11} & \alpha A_{12} + \beta B_{12} \\ \alpha A_{21} + \beta B_{21} & \alpha A_{22} + \beta B_{22} \end{bmatrix}.$$

Aussi, les deux matrices A et B peuvent être conformément partitionnées pour la multiplication. Dans ce cas, si A prend la forme (0.1) alors B est de taille (n, p) telle que B_{11} est de taille (n_1, p_1), B_{12} est de taille (n_1, p_2), B_{21} est de taille (n_2, p_1) et B_{22} est de taille (n_2, p_2) et $p_1 + p_2 = p$. Ainsi, les produits $A_{ik}B_{kj}$, $i, j, k = 1, 2$, sont bien définis et le produit AB se prête à la définition du produit matriciel ordinaire:

$$\begin{aligned} AB &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} . \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}. \end{aligned}$$

Plus généralement, si une matrice par blocs a la forme

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1c} \\ \vdots & & \vdots \\ A_{r1} & \cdots & A_{rc} \end{bmatrix}$$

où A est de taille (m, n) , chaque matrice A_{ij} est de taille (m_i, n_j) , $1 \leq i \leq r$, $1 \leq j \leq c$, avec $\sum\limits_{i=1}^{r} m_i = m$ et $\sum\limits_{i=1}^{c} n_j = n$, alors la transposée de A est donné par

$${}^tA = \begin{bmatrix} {}^tA_{11} & {}^tA_{21} & \cdots & {}^tA_{r1} \\ {}^tA_{12} & {}^tA_{22} & \cdots & {}^tA_{r2} \\ \vdots & \vdots & \vdots & \vdots \\ {}^tA_{1c} & {}^tA_{2c} & \cdots & {}^tA_{rc} \end{bmatrix}.$$

L'addition et le produit de deux matrices par blocs, lorsque celles-ci sont conformément partitionnées pour ces deux opérations, sont obtenues par une manière analogue à celle du cas du partitionnement de taille $(2, 2)$ en une matrice par blocs.

0.2.1 Déterminant d'une matrice par blocs

Rappelons que, pour un ensemble non vide fini E, une permutation σ de E est une bijection de E vers lui même. L'ensemble $S(E)$ des permutations de E muni de la loi de composition $\circ$ des applications de E vers lui même est un groupe. Dans le cas particulier $E = \{1, 2, ..., n\}$, $S(E)$ est noté simplement S_n. Une transposition $\tau_{i,j}$, $\{i, j\} \subset \{1, 2, ..., n\}$, est une permutation de S_n laissant invariant tout élément de $\{1, 2, ..., n\} \setminus \{i, j\}$ telle que $\tau_{i,j}(i) = j$ et $\tau_{i,j}(j) = i$. Toute permutation σ de S_n peut être décomposée en un produit (pour la loi de composition $\circ$) de transpositions. Si p est le nombre de transpositions dans ce dernier produit, alors le nombre $\varepsilon(\sigma) = (-1)^p$ ne dépend pas d'une décomposition particulière de σ. C'est donc par définition la signature de la permutation σ.

Définition 0.3 *Le déterminant d'une matrice carrée* $A = (a_{ij})_{1 \le i,j \le n}$, *noté* $|A|$ *(ou parfois* $\det A$*), est le scalaire*

$$|A| = \sum_{\sigma \in S_n} \varepsilon(\sigma)\, a_{1\sigma(1)} a_{2\sigma(2)} ... a_{n\sigma(n)}.$$

Rappelons que le déterminant d'une matrice carrée A satisfait les propriétés suivantes:

1. laissant fixe toute les autres colonnes (resp. lignes) de A, $|A|$ est linéaire par rapport à chaque colonne (resp. ligne) de A.
2. si au moins une des colonnes (resp. lignes) de A est une combinaison linéaire des autres, alors $|A| = 0$ (en particulier, c'est le cas lorsqu'au moins une des colonnes (resp. lignes) de A est nulle ou au moins deux colonnes (resp. lignes) de A sont identiques).

3. $|A| = |\, {}^tA|$.
4. $|AB| = |A|\,|B|$, B étant carrée de taille n.
5. A est inversible si et seulement si $|A| \neq 0$, et si A est inversible alors $|A^{-1}| = \frac{1}{|A|}$.

Théorème 0.4 *Soient A une matrice carrée de taille m, B une matrice de taille (m,n) et D une matrice carrée de taille n. Alors*

$$\begin{vmatrix} A & B \\ 0 & D \end{vmatrix} = |A|\,|D|\,.$$

Preuve. Soient $A = (a_{ij})_{1\leq i,j\leq m}$, $D = (d_{ij})_{1\leq i,j\leq n}$ et $P = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = (p_{ij})_{1\leq i,j\leq m+n}$.
On a

$$|P| = \sum_{\sigma\in S_{m+n}} \varepsilon(\sigma)\, p_{1\sigma(1)} p_{2\sigma(2)} ... p_{m\sigma(m)} p_{m+1\sigma(m+1)} ... p_{m+n\sigma(m+n)}\ . \tag{0.2}$$

Assertion 1 *Les termes non nuls de* (0.2) *s'écrivent*

$$\varepsilon(\theta)\,\varepsilon(\gamma)\, a_{1\theta(1)} a_{2\theta(2)} ... a_{m\theta(m)} d_{1\gamma(1)} d_{2\gamma(2)} ... d_{n\gamma(n)}$$

où $\theta \in S_m$, $\gamma \in S_n$ sont définies par

$$\theta(i) = \sigma(i)\,,\ \ i \in \{1,2,...,m\}\ \ \textit{et}\ \gamma(i) = \sigma(m+i) - m,\ \ i \in \{1,2,...,n\}\,.$$

Il s'ensuit que

$$\begin{aligned} |P| &= \sum_{\substack{\theta\in S_m \\ \gamma\in S_n}} \varepsilon(\theta)\,\varepsilon(\gamma)\, a_{1\theta(1)} a_{2\theta(2)} ... a_{m\theta(m)} d_{1\gamma(1)} d_{2\gamma(2)} ... d_{n\gamma(n)} \\ &= \sum_{\theta\in S_m} \varepsilon(\theta)\, a_{1\theta(1)} a_{2\theta(2)} ... a_{m\theta(m)} \sum_{\gamma\in S_n} \varepsilon(\gamma)\, d_{1\gamma(1)} d_{2\gamma(2)} ... d_{n\gamma(n)} \\ &= |A|\,|D|\,. \end{aligned}$$

$\square$

Exercice 1 *Montrer l'assertion 1.*

Exercice 2 *Soient A une matrice carrée de taille m, C une matrice de taille (n,m) et D une matrice carrée de taille n. Montrer que*

$$\begin{vmatrix} A & 0 \\ C & D \end{vmatrix} = |A|\,|D|\,.$$

Théorème 0.5 (Déterminant d'une matrice par blocs) *Soient A une matrice carrée de taille m, B une matrice de taille (m,n), C une matrice de taille (n,m) et D une matrice carrée de taille n. Alors*

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{cases} |A|\,|D - CA^{-1}B| & \textit{si } A \textit{ est inversible;} \\ |D|\,|A - BD^{-1}C| & \textit{si } D \textit{ est inversible.} \end{cases}$$

Preuve. Traitons le cas où A est inversible. On a

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ CA^{-1} & I_n \end{bmatrix} \begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix}.$$

Il s'ensuit que

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{vmatrix} I_m & 0 \\ CA^{-1} & I_n \end{vmatrix} \begin{vmatrix} A & B \\ 0 & D - CA^{-1}B \end{vmatrix} = |A|\left|D - CA^{-1}B\right|.$$

L'autre cas est traité d'une façon similaire. □

Une situation bien particulière est celle du déterminant des matrices dites bordées que l'on retrouve dans le contexte statistique avec l'actualisation d'un tableau de données statistiques. Une matrice bordée est obtenue en augmentant la taille d'une matrice carrée A par une ligne et une colonne, obtenant ainsi une matrice sous la forme $\begin{bmatrix} A & u \\ {}^t v & \alpha \end{bmatrix}$, où u et v sont deux vecteurs colonnes et α est un scalaire. Lorsque A est inversible, le déterminant d'une telle matrice est donné par

$$\begin{vmatrix} A & u \\ {}^t v & \alpha \end{vmatrix} = |A| \, (\alpha - \ {}^t v A^{-1} u).$$

Ce résultat trouve son intérêt dans les modèles de régression linéaire où l'on a besoin d'une actualisation de l'inverse $({}^t X X)^{-1}$ dans le but d'actualiser l'équation normale (Chapitre 1), et ce lorsque la matrice de données statistiques X est actualisée par l'ajout d'une ligne (i.e. l'observation sur un nouveau individu). Dans ce cas la matrice ${}^t X X$ est actualisée par l'ajout d'une matrice de rang 1 (on parle d'actualisations de rang 1). En effet, X est actualisée en la matrice $X_u = \begin{bmatrix} X \\ {}^t v \end{bmatrix}$ et on a donc

$${}^t X_u X_u = \ {}^t X X + v \ {}^t v.$$

Théorème 0.6 (Actualisations de rang 1 et formule de Sherman–Morrison) *Soient A une matrice carrée inversible de taille n et u, v deux vecteurs colonne de taille $(n, 1)$. Alors*

$$\left| A + u \ {}^t v \right| = |A| \, (1 + \ {}^t v A^{-1} u),$$

et donc $A + u \ {}^t v$ est inversible si et seulement si $1 + \ {}^t v A^{-1} u \neq 0$.

Si $1 + \ {}^t v A^{-1} u \neq 0$, l'inverse de $A + u \ {}^t v$ est obtenu par la correction de celui de A comme suit:

$$\left(A + u^t v \right)^{-1} = A^{-1} - \frac{1}{1 +^t v A^{-1} u} A^{-1} \left(u^t v \right) A^{-1}.$$

Preuve. Le résultat sur le déterminant est une conséquence immédiate du Théorème 0.5 appliqué à la matrice $\begin{bmatrix} A & u \\ -\,{}^t v & 1 \end{bmatrix}$. La formule de l'inverse est obtenue facilement en vérifiant que

$$(A+u^tv)(A^{-1}-\frac{1}{1+\ {}^tvA^{-1}u}A^{-1}u^tvA^{-1}) = (A^{-1}-\frac{1}{1+\ {}^tvA^{-1}u}A^{-1}u^tvA^{-1})(A+u^tv) = I.$$

□

Exercice 3 *Soit la matrice*

$$A = \begin{bmatrix} 1+\lambda_1 & 1 & \dots & 1 \\ 1 & 1+\lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 1+\lambda_n \end{bmatrix},$$

les réels λ_i *étant non nuls. Calculer* $\det A$ *(ind. exprimer* A *comme actualisation de rang 1 d'une matrice).*

Exercice 4 (Contrôle continu, automne 2018) *Calculer les inverses des matrices par blocs* $\begin{bmatrix} I_{n_1} & 0 \\ X & I_{n_2} \end{bmatrix}$; $\begin{bmatrix} I_{n_1} & Y \\ 0 & I_{n_2} \end{bmatrix}$.

1. *Soit* R *une matrice carrée partitionnée selon une matrice par blocs* $R = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ *telles que* A *et* D *sont des matrices carrées de tailles* n_1 *est* n_2 *resp. Définissons le complément de Schur de* A*, lorsque celle-ci est inversible, comme étant* $F = D - CA^{-1}B$*. Supposons que* A *et son complément de Schur sont inversibles. Donner la décomposition* LDU *par blocs de* R *:*

$$R = \begin{bmatrix} I_{n_1} & 0 \\ X & I_{n_2} \end{bmatrix} \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \begin{bmatrix} I_{n_1} & Y \\ 0 & I_{n_2} \end{bmatrix}.$$

2. *En déduire que l'inverse de R est donné par:*

$$R^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BF^{-1}CA^{-1} & -A^{-1}BF^{-1} \\ -F^{-1}CA^{-1} & F^{-1} \end{bmatrix}.$$

3. *Supposons que D est inversible et définissons de même le complément de Schur de D comme étant $G = A - BD^{-1}C$.*
   (a) *Écrire le déterminant de R de deux façons différentes et en déduire que G est inversible.*
   (b) *Montrer que*

$$R^{-1} = \begin{bmatrix} G^{-1} & -G^{-1}BD^{-1} \\ -D^{-1}CG^{-1} & D^{-1} + D^{-1}CG^{-1}BD^{-1} \end{bmatrix}.$$

4. *En déduire la formule Sherman-Woodbury-Morrison:*

$$\left(A - BD^{-1}C\right)^{-1} = A^{-1} + A^{-1}B\left(D - CA^{-1}B\right)^{-1}CA^{-1}.$$

0.3 Systèmes linéaires

L'un des fameux problèmes dans les sciences mathématiques consiste à résoudre simultanément m équations algébriques à n inconnues:

$$a_{i1}x_1 + ... + a_{in}x_n = b_i,\ i = 1, 2, ..., m, \tag{0.3}$$

les inconnues étant $x_1, ..., x_n$ et les a_{ij} , b_i étant des constantes réelles données. Ce problème est reformulé matriciellement en l'équation $Ax = b$, où $A = (a_{ij})_{\substack{1\leq i\leq m \\ 1\leq j\leq n}}$ est la matrice des coefficients du système (0.3) , $x = {}^t(x_1, ..., x_n)$ est le vecteur des inconnues et $b = {}^t(b_1, ..., b_m)$ est le vecteur du second membre. La solution du système (0.3) ne sort de l'un des trois cas suivant:

1. solution unique;
2. un nombre infini de solutions (l'ensemble de solutions est dans ce cas convexe);

3. pas de solution.

Dans les deux premiers cas, le système (0.3) est dit consistent. Si non, il est dit inconsistent. Dans ce dernier cas, le système peut être résolu dans un autre sens via la méthode des moindres carrés (Chapitre 1, Paragraphe 1.8).

0.3.1 Méthode d'élimination de Gauss

Résoudre le système (0.3) à l'aide de la méthode d'élimination de Gauss consiste à transformer le système à un autre système linéaire équivalent (i.e. ayant même ensemble de solutions) qui prend une forme échelonnée, et ce en appliquant aux lignes du système (0.3) les opérations élémentaires:

1. Type I: intérchanger deux lignes du système (0.3) .
2. Type II: multiplier une ligne par un scalaire non nul.
3. Type III: remplacer une ligne par la somme de cette dernière et une autre ligne multipliée par un scalaire.

La forme échelonnée ciblée pour le système équivalent est définie par les deux conditions suivantes:

1. si une ligne est nulle alors toutes les lignes en dessous le sont;
2. si dans une ligne non nulle L_i, le premier coefficient non nul figure dans la $j^{\text{ième}}$ position, alors dans les colonnes $C_1, ..., C_j$ tous les coefficients en dessous de la ligne L_i sont nuls.

Un exemple illustratif d'une forme échelonnée est le suivant:

$$\begin{bmatrix} * & * & * & * & * & * & * & * \\ 0 & * & * & * & * & * & * & * \\ 0 & 0 & 0 & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

La solution du système (0.3) sous l'une de ses formes échelonnées devient simple en procédant par une substitution à l'envers. Dans une forme échelonnée d'un système linéaire:

1. les premiers coefficients non nuls des lignes non nulles (les coefficients entourés dans l'exemple précédent) sont dits les pivots de ce système.
2. les colonnes contenant ces pivots sont dites colonnes basiques.

Le nombre de pivots d'un système linéaire ne dépend pas de la forme échelonnée obtenue par le processus d'élimination de Gauss. Il est donc par définition le rang de ce système.

Si dans le système (0.3) , $b = 0$ alors ce système est dit homogène. Les systèmes linéaires homogènes sont toujours consistents du fait qu'ils ont $x = 0$ comme solution triviale. L'ensemble de solutions d'un système linéaire homogène à n inconnues est un sous espace vectoriel de $\mathbb{R}^n$ engendré par $n-r$ solutions particulières linéairement indépendantes, le nombre r étant le rang de ce système linéaire.

Il est à noter que le rang d'un système linéaire est égal au rang de la matrice de ses coefficients. La section suivante porte sur la notion du rang d'une matrice.

Exercice 5 *En résolvant un système linéaire, trouver les coefficients dans l'équation de la parabole* $y = \alpha + \beta x + \gamma x^2$ *qui passe par les points* $(1, 1)$, $(2, 2)$ *et* $(3, 0)$.

Exercice 6 *Soit* A *une matrice de taille* (m, n) .

1. *Supposons que* $n > m$. *Montrer que le système homogène* $Ax = 0$ *a une solution non triviale.*
2. *Supposons que* $n = m$. *Montrer que le système* $Ax = 0$ *n'a pas de solutions non triviales si et seulement si* A *est inversible.*

Exercice 7 *Soit la matrice de Vandermonde*

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & & & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix},$$

les x_i*,* $i = 1, ..., n$ *(*$n > 1$*), étant distincts deux à deux.*

1. *Montrer que le système homogène* $Ax = 0$ *n'a pas de solutions non triviales.*
2. *Soit* $\mathcal{N}$ *le nuage de points* (x_i, y_i) *résultat de* n *observations* x_i *et* y_i*,* $i = 1, 2, ..., n$*, de deux variables statistiques quantitatives* X *et* Y *respectivement, telles que les réels* x_i *sont distincts deux à deux. Montrer qu'il existe un seul polynôme* P *de degré* $n-1$ *intérpolant le nuage de points* $\mathcal{N}$*, i.e.,* $P(x_i) = y_i$ *pour tout* $i = 1, ..., n$*.*

0.4 Rang d'une matrice

Pour deux espaces vectoriels E et F, la notation $E \simeq F$ signifie que E et F sont isomorphe. Notons que $\mathcal{M}_{m,n}(\mathbb{R}) \simeq \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ l'espace vectoriel des applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}^m$, via l'application qui à $A \in \mathcal{M}_{m,n}(\mathbb{R})$ associe l'application linéaire $x \longmapsto Ax$ de $\mathbb{R}^n$ dans $\mathbb{R}^m$. De ce fait, on a la définition suivante.

Définition 0.7 *Le rang d'une matrice* A *de taille* (m, n)*, noté* rg(A)*, est la dimension du sous espace vectoriel image de* A *:*

$$\mathrm{rg}(A) := \dim ImA = \dim \{Ax : x \in \mathbb{R}^n\}.$$

Notons que l'on a toujours $\mathrm{rg}(A) \leq \min(m, n)$.

Théorème 0.8 (Théorème du rang) *Soit A une matrice de taille (m,n). Alors*

$$\operatorname{rg}(A) + \dim \operatorname{Ker} A = n,$$

où $\operatorname{Ker} A$ est le sous espace vectoriel noyau de A :

$$\operatorname{Ker} A := \{x \in \mathbb{R}^n : Ax = 0\}.$$

Corollaire 0.9 *Une matrice carrée A de taille n est inversible si et seulement si* $\operatorname{rg}(A) = n$.

Exercice 8 *Soient A une matrice de taille (m,n) et G un supplémentaire de $\operatorname{Ker} A$ dans $\mathbb{R}^n$.*

1. *Montrer que $G \simeq \operatorname{Im} A$.*
2. *En déduire le théorème du rang.*

Exercice 9 *Soit A une matrice de taille (m,n).*

1. *Montrer que $\operatorname{Ker}({}^tAA) = \operatorname{Ker} A$.*
2. *En déduire que*

$$\operatorname{rg}(A) = \operatorname{rg}({}^tA) = \operatorname{rg}({}^tAA) = \operatorname{rg}(A\,{}^tA).$$

Exercice 10 *Deux matrice A et B de même taille (m,n) sont dites équivalentes s'il existe deux matrices inversibles P et Q de tailles m et n respectivement, telles que $A = P^{-1}BQ$.*

1. *Montrer que pour une matrice A de taille (m,n), $\operatorname{rg} A = r$ si et seulement si A est équivalente à $J_r = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$.*
2. *En déduire que deux matrices de même taille sont équivalentes si et seulement si elles ont même rang.*

Définition 0.10 *Une matrice A de taille (m, n) est dite de rang plein si* $\mathrm{rg}(A) = n$.

D'après le théorème du rang, une matrice A est de rang plein si et seulement si $\mathrm{Ker}A = \{0\}$, i.e.

$$Ax = 0 \Rightarrow x = 0.$$

Proposition 0.11 *Soit A une matrice de taille (m, n). Alors*

$$A \text{ est de rang plein } \Leftrightarrow {}^tAA \text{ est inversible.}$$

Preuve. Comme $\mathrm{rg}(A) = \mathrm{rg}({}^tAA)$, alors A est de rang plein si et seulement si $\mathrm{rg}({}^tAA) = n$, i.e., ${}^tAA$ est inversible. □

Les matrices sous la forme ${}^tAA$ jouent un rôle primordial dans certaines applications statistiques. En particulier, de telles matrices sont d'une utilité particulière dans l'étude des modèles de régression linéaire. Ces matrices jouissent d'un certain nombre de propriétés intéressantes, qui sont entre autres: les propriétés bien connues des matrices réelles symétriques, positivité semi-définie, les valeurs singulières de A sont (par définition) les racines carrées des valeurs propres strictement positives de ${}^tAA$ (Chapitre 2).

Exercice 11 *Soient A et B deux matrices de taille (m, n) et (n, p) respectivement.*

1. *Montrer que* $\mathrm{Im}(AB) \subset \mathrm{Im}A$ *et qu'il y a égalité si B est une matrice carrée inversible.*
2. *Montrer que* $\mathrm{Ker}B \subset \mathrm{Ker}(AB)$ *et qu'il y a égalité si A est une matrice carrée inversible.*

Exercice 12 *Soient A et B deux matrices de taille (n, p) et de rang plein telles que* $\mathrm{Im}A = \mathrm{Im}B$. *Montrer qu'il existe une matrice carrée C de taille p inversible telle que $AC = B$.*

0.5 Trace d'une matrice

Définition 0.12 *Soit $A = (a_{ij})$ une matrice carrée de taille n. La trace de A, notée trA, est la somme de ses coefficients diagonaux:*

$$trA = \sum_{i=1}^{n} a_{ii}.$$

Il s'agit d'une forme linéaire sur l'espace vectoriel $\mathcal{M}_n(\mathbb{R})$ des matrices carrées de taille n invariante au passage à la transposée. Elle s'applique énormément en statistique, notamment en analyse multivariée.

Proposition 0.13 (Propriété cyclique de la trace) *Soient A et B deux matrices de tailles (m, n) et (n, m) respectivement. Alors*

$$tr(AB) = tr(BA).$$

Plus généralement, la trace est invariante par permutations cycliques. Par exemple,

$$tr(ABC) = tr(BCA) = tr(CAB),$$

où les produits sous la trace donnent bien des matrices carrées.

Exercice 13 *1. Montrer que pour $x, y \in \mathbb{R}^n$, $tr(x\ {}^t y) = {}^t x\ y$.*

2. Soient A et B deux matrices de tailles (m, n) et (n, m) respectivement.

(a) Montrer que $tr(AB) = \sum_{i=1}^{m} {}^t a_{i*}\ b_{*i}$, où l'on a écrit $A = \begin{bmatrix} {}^t a_{1*} \\ \vdots \\ {}^t a_{m*} \end{bmatrix}$ et $B = \begin{bmatrix} b_{*1} & \cdots & b_{*m} \end{bmatrix}$.

(b) En déduire que $tr(AB) = tr(BA)$.

Exercice 14 *Montrer que l'application tr est la seule forme linéaire sur $\mathcal{M}_n(\mathbb{R})$*

vérifiant:

$$\begin{cases} tr\left(AB\right) = tr\left(BA\right), \ \ A, B \in \mathcal{M}_n\left(\mathbb{R}\right); \\ tr\left(I_n\right) = n. \end{cases}$$

Exercice 15 *Montrer que l'application* $(A, B) \longrightarrow \langle A, B\rangle := tr(\ {}^tAB)$ *définit un produit scalaire sur* $\mathcal{M}_{m,n}\left(\mathbb{R}\right)$*.*

Une expression équivalente de la trace d'une matrice carrée A de taille n est

$$trA = \sum_{i=1}^{n} \lambda_i$$

où les λ_i sont les n valeurs propres de A (répétition permise). Cela peut être vu via la trigonalisation de A (toujours possible dans $\mathbb{C}$) par une matrice complexe inversible U de taille n:

$$A = UTU^{-1},$$

T étant une matrice complexe triangulaire supérieure de taille n dont les coefficients diagonaux sont $\lambda_1, ..., \lambda_n$. Il s'ensuit que

$$trA = tr(UTU^{-1}) = tr(TU^{-1}U) = trT = \sum_{i=1}^{n} \lambda_i.$$

Exercice 16 *1. Montrer que pour tout* $f \in \mathcal{M}_n\left(\mathbb{R}\right)^*$ *(le dual algébrique de* $\mathcal{M}_n\left(\mathbb{R}\right)$*) il existe* $A \in \mathcal{M}_n\left(\mathbb{R}\right)$ *telle que*

$$f\left(X\right) = AX, \quad X \in \mathcal{M}_n\left(\mathbb{R}\right).$$

2. En déduire que $\mathcal{M}_n\left(\mathbb{R}\right)^* \simeq \mathcal{M}_n\left(\mathbb{R}\right)$*.*

0.6 Projecteurs

Soient F et G deux sous espaces vectoriels supplémentaires d'un espace vectoriel E de dimension finie (on écrit $E = F \oplus G$) de telle façon que tout élément $x \in E$

a une représentation unique sous la forme:

$$x = y + z, \ y \in F \text{ et } z \in G.$$

- Le vecteur y est appelé la projection de x sur F parallèlement à G.
- Le vecteur z est appelé la projection de x sur G parallèlement à F.

Définition 0.14 *Soit* $\mathbb{R}^n = E \oplus F$. *Une matrice carrée* P *de taille* n *est dite matrice de projection sur* E *parallèlement à* F, *si pour tout* $x \in \mathbb{R}^n$, Px *est la projection de* x *sur* E *parallèlement à* F. *Une matrice carrée* P *de taille* n *est dite un projecteur s'elle est une matrice de projection sur un sous espace vectoriel de* $\mathbb{R}^n$ *parallèlement à un supplémentaire.*

Les projecteurs jouissent des propriétés suivantes:

Théorème 0.15 *Pour une matrice carrée* P *de taille* n, *les assertions suivantes sont équivalentes:*

1. P *est un projecteur;*
2. P *est une matrice idempotente, i.e.* $P^2 = P$*;*
3. $\mathbb{R}^n = \mathrm{Im}P \oplus \mathrm{Ker}P$ *et* P *est la matrice de projection sur* $\mathrm{Im}P$ *parallèlement à* $\mathrm{Ker}P$.

Exercice 17 *Montrer qu'une matrice carrée* P *de taille* n *est un projecteur si et seulement si* $I_n - P$ *l'est aussi. Le cas où* P *est un projecteur,* $I_n - P$ *est un projecteur sur quel sous espace vectoriel de* $\mathbb{R}^n$ *parallèlement à quel sous espace vectoriel de* $\mathbb{R}^n$*?*

Exercice 18 (Somme de projecteurs) *Soient* P *et* Q *deux projecteurs de taille* n.

1. *Montrer que*

$$\begin{aligned} P+Q \textit{ est un projecteur } &\Leftrightarrow \quad \mathrm{Im}P \subset \mathrm{Ker}Q \textit{ et } \mathrm{Im}Q \subset \mathrm{Ker}P \\ &\Leftrightarrow \; PQ = QP = 0. \end{aligned}$$

2. *Supposons que $P+Q$ est un projecteur. Montrer que* $\mathrm{rg}(P+Q) = \mathrm{rg}P + \mathrm{rg}Q$ *et en déduire que $P+Q$ est le projecteur sur* $\mathrm{Im}P \oplus \mathrm{Im}Q$ *parallèlement à* $\mathrm{Ker}P \cap \mathrm{Ker}Q$.
3. *Que seront-t-ils les résultats si l'on remplace la somme de P et Q par leur différence?*

Chapitre 1

Espaces euclidiens et orthogonalité

1.1 Produit scalaire d'un espace euclidien

Dans toute la suite E désigne un espace vectoriel réel de dimension finie $n \geq 1$. La notation $Vect\,(X)$ désignera le sous espace vectoriel engendré par un sous ensemble X de E.

Définition 1.1 *Un produit scalaire sur E est une application $\langle .,. \rangle : E \times E \longrightarrow \mathbb{R}$ satisfaisant les trois propriétés suivantes:*

1. $\langle x, y \rangle = \langle y, x \rangle$ *(symétrie).*
2. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$ *(linéarité).*
3. $\langle x, x \rangle \geq 0$ *et* $\langle x, x \rangle = 0$ *si et seulement si* $x = 0$.

En d'autres termes, un produit scalaire sur E est une forme bilinéaire symétrique définie positive sur E. L'espace vectoriel E muni du produit scalaire $\langle .,. \rangle$ est dit un espace euclidien. L'application $\|.\| : E \longrightarrow \mathbb{R}^+$, $\|x\| = \langle x, x \rangle^{\frac{1}{2}}$ définie une norme sur E dite la norme issue du produit scalaire de E. L'espace vectoriel E peut être muni de n'importe quel autre norme qui est en général donner par la définition suivante:

Définition 1.2 *Une norme sur E est une application $\|.\| : E \longrightarrow \mathbb{R}^+$ satisfaisant les trois propriétés suivantes:*

1. $\|x\| = 0$ *si et seulement si* $x = 0$.
2. $\|\lambda x\| = |\lambda| \|x\|$, $\lambda \in \mathbb{R}$.
3. $\|x + y\| \leq \|x\| + \|y\|$ *(inégalité triangulaire).*

Exemple 1.3 *Sur* $\mathbb{R}^n$: *les* p*-normes* $\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$, $p \geq 1$, *la norme* $\|x\|_\infty = \max_{i=1}^{n} |x_i|$, *avec* $x = (x_1, ..., x_n)$.

Du point de vue topologique, comme E est de dimension finie, toutes les normes de E sont équivalentes dans le sens où elles définissent la même topologie sur E. L'espace euclidien E est un espace métrique; la distance d (ou métrique) sur E étant celle induite par la norme issue du produit scalaire de E :

$$d(x, y) = \|x - y\| = \langle x - y, x - y \rangle^{\frac{1}{2}} .$$

Un exemple essentiel d'un produit scalaire est celui du produit scalaire canonique (ou euclidien) sur $\mathbb{R}^n$:

Définition 1.4 (Produit scalaire canonique) *Le produit scalaire canonique sur* $\mathbb{R}^n$ *est défini pour* $x = {}^t(x_1, ..., x_n)$, $y = {}^t(y_1, ..., y_n)$ *par*

$$\langle x, y \rangle = {}^t x.y = \sum_{i=1}^{n} x_i y_i.$$

La norme issue de ce produit scalaire est la norme euclidienne donnée par:

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}}.$$

La distance sur $\mathbb{R}^n$ *induite par la norme euclidienne est aussi dite distance euclidienne et est donnée par:*

$$d(x, y) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}.$$

On a besoin dans certains cas, pour des raisons statistiques, de munir $\mathbb{R}^n$ d'une variante du produit scalaire canonique:

Définition 1.5 (M-produit scalaire) *Le M-produit scalaire sur $\mathbb{R}^n$ est défini par*

$$\langle x, y \rangle_M = {}^t x M y,$$

où M est une matrice carrée de taille n symétrique définie positive (Chapitre 2) générant ce produit scalaire. La norme issue de ce produit scalaire est dite la M-norme et est donnée par:

$$\|x\|_M = \left({}^t x M x \right)^{\frac{1}{2}}.$$

Comme la matrice M peut s'écrire $M = {}^t N N$, N étant une matrice inversible de taille n (Théorème 2.17), alors la M-norme est donnée par:

$$\|x\|_M = \left({}^t x \, {}^t N N x \right)^{\frac{1}{2}} = \|Nx\|.$$

Notons que dans les problèmes utilisant la fameuse méthode des moindres carrés, la norme qui reste toujours utile est celle euclidienne ou, dans certains cas, sa variante la M-norme. Pour cela, dans tout ce qui suit, et sauf mention contraire, $\mathbb{R}^n$ sera muni de son produit scalaire canonique et sa norme euclidienne notée $\|.\|$. Cette même notation est utilisée pour désigner la norme issue du produit scalaire d'un espace euclidien abstrait E qui sera, et sauf mention contraire, normé de cette manière standard.

Théorème 1.6 (Inégalité de Cauchy Schwarz) *Soit E un espace euclidien. Pour $x, y \in E$ on a*

$$|\langle x, y \rangle| \leq \|x\| . \|y\| .$$

Il y a égalité si et seulement si x et y sont linéairement dépendants.

Une généralisation du produit scalaire canonique sur $\mathbb{R}^n$ est le produit scalaire canonique sur l'espace vectoriel $\mathcal{M}_{m,n}(\mathbb{R})$ des matrices réelles de taille (m,n) que l'on dérive naturellement du premier produit scalaire via la vectorisation des matrices. En effet, une matrice $A=(a_{ij})\in\mathcal{M}_{m,n}(\mathbb{R})$ s'écrit

$$A=[a_{*1},...,a_{*n}]=\begin{bmatrix} {}^t a_{1*} \\ \vdots \\ {}^t a_{m*} \end{bmatrix},$$

où les $a_{*j}\in\mathbb{R}^m$ et les $a_{i*}\in\mathbb{R}^n$. Il s'ensuit que $\mathcal{M}_{m,n}(\mathbb{R})$ s'identifie à $\mathbb{R}^{mn}$ via l'isomorphisme qui à A associe le vecteur $vec(A)$ obtenu par la concaténation verticale des vecteurs colonnes $a_{*1},...,a_{*n}$ (resp. $a_{1*},...,a_{m*}$):

$$vec(A) := {}^t[\,{}^t a_{*1},...,\ {}^t a_{*n}] \text{ (resp. } {}^t[\,{}^t a_{1*},...,\ {}^t a_{m*}]).$$

On a donc la défintion naturelle suivante:

Définition 1.7 (Produit scalaire matriciel canonique) *Le produit scalaire canonique sur $\mathcal{M}_{m,n}(\mathbb{R})$ est défini par:*

$$\langle A,B\rangle := \langle vec(A), vec(B)\rangle = \sum_{i,j} a_{ij}b_{ij} = \sum_j {}^t a_{*j}b_{*j} = \sum_i {}^t a_{i*}b_{i*} = tr({}^t AB),$$

où l'on a écrit $B=(b_{ij})$ de la même manière par concaténation de ses vecteurs colonnes (resp. lignes).

1.2 Normes matricielles

Définition 1.8 (Norme de Frobenius) *La norme de Frobenius (ou norme matricielle standard) est la norme issue du produit scalaire canonique sur $\mathcal{M}_{m,n}(\mathbb{R})$, i.e., la norme donnée par*

$$\|A\|^2 = \sum_{i,j} a_{ij}^2 = \sum_j \|a_{*j}\|^2 = \sum_i \|a_{i*}\|^2 = tr({}^tAA).$$

Comme la trace d'une matrice carrée est la somme de ses valeurs propres, alors la norme de Frobenius de A se réduit à

$$\|A\| = (\sigma_1^2 + ... + \sigma_n^2)^{\frac{1}{2}},$$

où les σ_i^2 ($\sigma_i \geq 0$) sont les n valeurs propres positives de la matrice semi-définie positive ${}^tAA$ (Chapitre 2). La norme de Frobenius d'une matrice non nulle A se réduit encore à

$$\|A\| = (\sigma_1^2 + ... + \sigma_r^2)^{\frac{1}{2}},$$

les σ_i étant les valeurs singulières de A et r =rg(A) (voir Exercice 41). Si la matrice A est de plus symétrique, alors

$$\|A\| = (\lambda_1^2 + ... + \lambda_n^2)^{\frac{1}{2}},$$

les λ_i étant les valeurs propres de A.

La norme matricielle précédente, en tant qu'application sur l'ensemble de toutes les matrices (de n'importe quelle taille), satisfait une quatrième propriété pértinente qui est l'équivalent de l'inégalité triangulaire pour le produit matriciel; la propriété de sous-multiplicativité:

$$\|AB\| \leq \|A\| \, \|B\| \quad \text{si } A \text{ et } B \text{ sont compatibles pour la multiplication.}$$

Proposition 1.9 *1. la norme de Frobenius est compatible avec la norme (vec-*

torielle) euclidienne dans le sens où l'on a l'inégalité

$$\|Ax\| \leq \|A\| \|x\|, \ A \in \mathcal{M}_{m,n}(\mathbb{R}), \ x \in \mathbb{R}^n.$$

2. *la norme de Frobenius est sous-multiplicative.*

Preuve.

1. Comme $Ax = {}^t[{}^ta_{1*}x, ..., {}^ta_{m*}x]$, alors grâce à l'inégalité de Cauchy Schwarz on a
$$\|Ax\|^2 = \sum_i \left|{}^ta_{i*}x\right|^2 \leq \sum_i \|a_{i*}\|^2 \|x\|^2 = \|A\|^2 \|x\|^2.$$
2. En écrivant A et B par concaténation de vecteurs colonnes et en remarquant que $(AB)_{*j} = Ab_{*j}$, on a
$$\begin{aligned}\|AB\|^2 &= \sum_j \left\|(AB)_{*j}\right\|^2 = \sum_j \|Ab_{*j}\|^2 \leq \sum_j \|A\|^2 \|b_{*j}\|^2 \\ &= \|A\|^2 \sum_j \|b_{*j}\|^2 = \|A\|^2 \|B\|^2.\end{aligned}$$

□

Certains auteurs ajoutent aux trois axiomes de la définition d'une norme la propriété de sous multiplicativité pour adopter une définition non standard d'une norme matricielle en tant qu'application sur l'ensemble de toutes les matrices. L'addition et la multiplication des matrices sont donc restreintes aux matrices compatibles pour ces opérations.

Notons que l'espace vectoriel $\mathcal{M}_{m,n}(\mathbb{R})$ peut être normé par la norme d'opérateur en identifiant $\mathcal{M}_{m,n}(\mathbb{R})$ à $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, et ce en se fixant une norme quelconque du même type $\|.\|$ (on note abusivement différente norme par ce même symbôle universel) sur $\mathbb{R}^n$ et $\mathbb{R}^m$ munis de leurs bases canoniques. Ainsi, on a la définition suivante:

Définition 1.10 *La norme d'opérateur d'une matrice $A \in \mathcal{M}_{m,n}(\mathbb{R})$ est donnée par*

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\| \leq 1} \|Ax\| = \max_{\|x\|=1} \|Ax\|.$$

Dans ce cas, la norme d'opérateur sur $\mathcal{M}_{m,n}(\mathbb{R})$ est dite induite par la norme vectorielle.

On utilisera par la suite la notation $\|.\|_p$, $1 \leq p \leq \infty$, pour dénoter la norme d'opérateur induite par la norme vectorielle $\|.\|_p$ sur $\mathbb{R}^n$. Une norme d'opérateur est donc toujours compatible avec la norme vectorielle qu'elle l'induit. La norme matricielle $\|.\|_2$ est reconnue souvent pour la norme spectrale et est donnée par (voir Exercice 39):

$$\|A\|_2 = \sigma_1,$$

où σ_1 est la plus grande valeur singulière de A. Le cas où A est symétrique, sa norme spectrale n'est autre que son rayon spectral:

$$\|A\|_2 = \rho(A) := \max\{|\lambda_i| : \lambda_i \text{ est une valeur propre de } A\}.$$

L'une des préocupations en analyse matricielle consiste à l'étude de l'invariance d'une norme matricielle par passage à certaines opérations matricielles.

Proposition 1.11 *Soit A une matrice de taille (m,n). Alors les deux normes de Frobenius et spectrale sont*

1. *invariantes par passage à la transposée, i.e. $\|A\| = \| {}^tA \|$.*
2. *orthogonalement invariantes, i.e., $\|A\| = \| UAV \|$, U, V étant deux matrices de tailles m et n respectivement telles que $U\,{}^tU = I_m$ et ${}^tVV = I_n$.*

La norme d'opérateur ne satisfait pas en général les deux propriétés d'invariances précédentes; e.g. $\|.\|_p$, $p = 1, \infty$ pour l'invariance par passage à la transposée (Exercice 20).

Exercice 19 *Montrer Proposition 1.11.*

Exercice 20 *Soit* $A = (a_{ij})$ *une matrice de taille* (m, n). *Montrer que*

1. $\|A\|_1 = \max_j \sum\limits_i |a_{ij}|$ *(la plus grande* 1*-norme de ses vecteurs colonnes).*
2. $\|A\|_\infty = \max_i \sum\limits_j |a_{ij}|$ *(la plus grande* 1*-norme de ses vecteurs lignes).*
3. *En déduire que* $\| \ {}^tA \ \|_\infty = \|A\|_1$.

1.3 Orthogonalité dans un espace euclidien

Notons que, même si l'on présentera certaines propriétés géométriques et algébriques dans le cadre d'un espace euclidien abstrait $(E, \langle ., . \rangle)$, on ne s'intéressera dans le contexte de la statistique qu'aux espaces euclidien $\mathbb{R}^n$ et $\mathcal{M}_{m,n}(\mathbb{R})$ munis dans la plupart des cas de leurs produits scalaires et normes canoniques. Notons aussi que, puisque tout espace euclidien $(E, \langle ., . \rangle)$ de dimension $n \geq 1$ admet une base orthonormale $\mathcal{B}$ (Paragraphe 1.5), alors E est isomorphe à $\mathbb{R}^n$ muni de son produit scalaire canonique, et ce via l'isomorphisme qui à $u \in E$ associe $x \in \mathbb{R}^n$ le vecteur des coordonnées de u dans la base $\mathcal{B}$.

Définition 1.12 *Soit* $(E, \langle ., . \rangle)$ *un espace euclidien.*

1. *L'angle entre deux vecteurs non nuls* $x, y \in E$ *est défini par*
$$(x \overset{\wedge}{,} y) = \arccos \frac{\langle x, y \rangle}{\|x\| \, \|y\|}.$$
2. *Deux vecteurs* $x, y \in E$ *sont dits orthogonaux, et on note* $x \perp y$, *si* $\langle x, y \rangle = 0$ *(i.e.* $(x \overset{\wedge}{,} y) = \frac{\pi}{2}$ *si* x *et* y *sont non nuls).*
3. *Une famille de vecteurs* $X = (x_i)_{i \in I}$ *est dite orthogonale si ses vecteurs* x_i *sont deux à deux orthogonaux. Si de plus, les vecteurs* x_i *sont unitaires, i.e.* $\|x_i\| = 1$ *pour tout* i, *alors* X *est dite orthonormale. On a donc dans ce cas* $\langle x_i, x_j \rangle = \delta_{ij}$, *où* $\delta_{ij} = 0$ *si* $i \neq j$ *et* $\delta_{ij} = 1$ *si* $i = j$ *(le delta de Kronecker).*

Exemple 1.13 *La base canonique* $(e_1, ..., e_n)$ *de* $\mathbb{R}^n$ *est une famille orthonormale.*

Le vecteur e_i étant le vecteur ayant sa $i^{ième}$ coordonnée égale à 1 et des zéros ailleurs.

Théorème 1.14 (Identité de Pythagore) *Soit E un espace euclidien. Pour deux vecteurs orthogonaux $x, y \in E$, on a*

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 .$$

Plus généralement, si $(x_1, x_2, ..., x_p)$ est une famille orthogonale de E, alors

$$\left\| \sum_{i=1}^{p} x_i \right\|^2 = \sum_{i=1}^{p} \|x_i\|^2 .$$

Proposition 1.15 *Toute famille orthogonale $(x_1, x_2, ..., x_p)$ de vecteurs non nuls d'un espace euclidien E est libre. Il en est de même, en particulier, pour toute famille orthonormale de E.*

Preuve. Soient α_i, $i = 1, ..., p$ des scalaires tels que $\sum\limits_{i=1}^{p} \alpha_i x_i = 0$. Fixons un $j \in \{1, ..., p\}$. Alors

$$< x_j, \sum_{i=1}^{p} \alpha_i x_i >= \sum_{i=1}^{p} \alpha_i < x_j, x_i >= \alpha_j < x_j, x_j >= 0.$$

D'où, comme $x_j \neq 0$, $\alpha_j = 0$. L'indice j étant arbitraire, la famille $(x_1, ..., x_p)$ est alors libre. □

1.4 Projection orthogonale sur une droite vectorielle

Soit $(x_1, x_2, ..., x_n)$ une famille orthogonale de vecteurs non nuls de E tel que $\dim E = n$. Pour $x \in E$, calculons $p(x)$ la projection de x sur $Vect(x_1)$ parallèlement à $Vect(x_2, x_3, ..., x_n)$. Soit $\alpha_1, ..., \alpha_n \in \mathbb{R}$ tels que $x = \alpha_1 x_1 + ... + \alpha_n x_n$. D'où

$\langle x, x_1 \rangle = \alpha_1 \langle x_1, x_1 \rangle$ et donc $\alpha = \frac{\langle x,x_1 \rangle}{\langle x_1,x_1 \rangle}$. Similairement, on a $\alpha_i = \frac{\langle x,x_i \rangle}{\langle x_i,x_i \rangle}$ pour tout i. Par suite on trouve que

$$x = \frac{\langle x, x_1 \rangle}{\langle x_1, x_1 \rangle} x_1 + ... + \frac{\langle x, x_n \rangle}{\langle x_n, x_n \rangle} x_n.$$

Il s'ensuit que $p(x) = \frac{\langle x,x_1 \rangle}{\langle x_1,x_1 \rangle} x_1$. Le vecteur $p(x)$ est dit la projection orthogonale de x sur x_1 et on le note dorénavant $proj_{x_1}(x)$. Dans le cas $E = \mathbb{R}^n$, la matrice de projection sur $Vect(x_1)$ parallèlement à $Vect(x_2, x_3, ..., x_n)$ est dite matrice de projection orthogonale sur x_1.

Proposition 1.16 *Soit $a \in \mathbb{R}^n$ un vecteur non nul. Alors*

1. *La matrice de projection orthogonle sur a est donnée par*

$$P_a = \frac{a\ {}^t a}{{}^t a.a}. \tag{1.1}$$

2. *P_a est idempotente, symétrique et de rang égal à 1.*

Preuve.

1. Pour $x \in \mathbb{R}^n$, on a

$$proj_a(x) = \frac{\langle x, a \rangle}{\langle a, a \rangle} a = \frac{{}^t a.x}{{}^t a.a} a = \frac{1}{{}^t a.a} \left({}^t a.x\right) a = \frac{1}{{}^t a.a}\ a \left({}^t a.x\right) = \frac{a\ {}^t a}{{}^t a.a} x = P_a x.$$

2. Idempotence: $P_a^2 = \frac{1}{\left({}^t a.a\right)^2}\ a\ {}^t aa\ {}^t a = \frac{{}^t aa}{\left({}^t a.a\right)^2}\ a\ {}^t a = P_a$.
   Symétrie: ${}^t P_a = {}^t\left(\frac{a\ {}^t a}{{}^t a.a}\right) = \frac{1}{{}^t a.a}\ {}^t\left({}^t a\right)\ {}^t a = \frac{a\ {}^t a}{{}^t a.a} = P_a$.
   Le rang: $\text{rg}(P_a) = \text{rg}(a\ {}^t a) = \text{rg}(a) = 1$.

$\square$

Exemple 1.17 *Déterminons la projection orthogonale du vecteur $b = {}^t(1, 1, 1)$ sur la droite vectorielle engendrée par $a = {}^t(1, 2, 2)$:*

La matrice de projection orthogonale sur a est:

$$P_a = \frac{a\ {}^t a}{{}^t a.a} = \frac{1}{9}\ {}^t(1,2,2)\,(1,2,2) = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}.$$

D'où $proj_a(b) = P_a b = \ {}^t\left(\frac{5}{9}, \frac{10}{9}, \frac{10}{9}\right).$

La projection orthogonale sur n'importe quel sous espace vectoriel de $\mathbb{R}^n$ sera traitée dans Paragraphe 1.8.

1.5 Orthogonalisation de Gram-Schmidt

Partant d'une famille libre de vecteurs d'un espace euclidien E, on peut obtenir une famille orthonormale de E via l'algorithme de Gram-Schmidt, et ce sans modifier le sous espace vectoriel engendré par la famille initiale:

Théorème 1.18 (Orthogonalisation de Gram-Schmidt) *Soit* $(x_1, x_2, ..., x_p)$ *une famille libre d'un espace euclidien* E*. Alors les deux familles* $(y_1, y_2, ..., y_p)$ *et* $(e_1, e_2, ..., e_p)$ *construites comme suit:*

$$\begin{aligned} y_1 &= x_1; \quad e_1 = \frac{y_1}{\|y_1\|}, \\ y_2 &= x_2 - proj_{y_1}(x_2); \quad e_2 = \frac{y_2}{\|y_2\|}, \\ &\vdots \\ y_p &= x_p - \sum_{i=1}^{p-1} proj_{y_i}(x_p); \quad e_p = \frac{y_p}{\|y_p\|}, \end{aligned}$$

sont respectivement une famille orthogonale et une famille orthonormale sat-

isfaisant

$$Vect\,(x_1, x_2, ..., x_p) = Vect\,(y_1, y_2, ..., y_p) = Vect\,(e_1, e_2, ..., e_p)\,.$$

Exemple 1.19 $x_1 = {}^t(1,2,3)\,;\ x_2 = {}^t(3,7,10)\,;\ x_3 = {}^t(0,1,0)$:

$$\begin{aligned}
y_1 &= x_1; \quad e_1 = \frac{y_1}{\|y_1\|} = \frac{1}{\sqrt{14}}\ {}^t(1,2,3)\,;\\
y_2 &= x_2 - \frac{{}^ty_1.x_2}{{}^ty_1.y_1}.y_1 = \frac{1}{14}\ {}^t(-5,4,-1)\,; \quad e_2 = \frac{y_2}{\|y_2\|} = \frac{1}{\sqrt{42}}\ {}^t(-5,4,-1)\,;\\
y_3 &= x_3 - \frac{{}^ty_1.x_3}{{}^ty_1.y_1}.y_1 - \frac{{}^ty_2.x_3}{{}^ty_2.y_2}.y_2 = \frac{1}{3}\ {}^t(1,1,-1)\,; \quad e_3 = \frac{y_3}{\|y_3\|} = \frac{1}{\sqrt{3}}\ {}^t(1,1,-1)\,.
\end{aligned}$$

La famille $\left({}^t(1,2,3)\,,\ {}^t(-5,4,-1)\,,\ {}^t(1,1,-1)\right)$ *est une orthogonalisation de la famille* (x_1, x_2, x_3) *et* (e_1, e_2, e_3) *en est une orthonormalisation.*

Si l'on se donne un espace euclidien E, en considérant une base $(x_1, x_2, ..., x_n)$ de E, celui-ci admet une base orthonormale $(e_1, e_2, ..., e_n)$ en appliquant le procédé de Gram-Schmidt à la famille $(x_1, x_2, ..., x_n)$. Aussi, partant d'une famille orthonormale $(e_1, e_2, ..., e_p)$ de E, on peut la compléter en une base $(e_1, e_2, ..., e_p, y_{p+1}, ...y_n)$ de E, et en appliquant le procédé de Gram-Schmidt à celle-ci, on obtient une base orthonormale $(e_1, e_2, ..., e_p, e_{p+1}, ...e_n)$ de E. Les vecteurs $e_1, e_2, ..., e_p$ n'étant pas affectés par le procédé de Gram-Schmidt car la famille $(e_1, e_2, ..., e_p)$ est déjà orthonormale. Notons que les coordonnées d'un vecteur $x \in E$ dans une base orthonormale $\mathcal{B} = (e_1, e_2, ..., e_n)$ sont simplement les produits scalaires de x et les éléments de la base $\mathcal{B}$, i.e.,

$$x = \sum_{i=1}^{n} \langle x, e_i \rangle\, e_i.$$

Cette dernière écriture est dite développement de Fourier de x et les coordonnées de x dans la base $\mathcal{B}$ sont dites ses coefficients de Fourier. Géométriquement, chaque composante $\langle x, e_i \rangle\, e_i$ du développement de Fourier de x représente sa projection orthogonale sur l'axe e_i.

En développant $\langle x, x \rangle$, $x \in E$ étant remplacé par son développement de Fourier, on obtient l'identité suivante reconnue pour l'identité de Parseval:

$$\|x\|^2 = \sum_{i=1}^{n} \langle x, e_i \rangle^2 .$$

L'identité de Parseval montre que la norme euclidienne du vecteur des coefficients de Fourier de $x \in E$ est invariante au choix de la base orthonormale utilisée.

Théorème 1.20 *Soit $(E, \langle ., . \rangle)$ un espace euclidien (de dimension $n \geq 1$). Alors*

1. *E admet une base orthonormale $(e_1, e_2, ..., e_n)$.*
2. *Tout vecteur $x \in E$ s'ecrit*

$$x = \sum_{i=1}^{n} \langle x, e_i \rangle e_i.$$

Cette dernière écriture étant le développement de Fourier de x et les scalaires $\langle x, e_i \rangle$ sont ses coefficients de Fourier.

3. *La norme de $x \in E$ est donnée par l'identité de Parseval:*

$$\|x\|^2 = \sum_{i=1}^{n} |\langle x, e_i \rangle|^2 .$$

4. *Toute famille orthonormale de E peut être complétée en une base orthonormale de E.*

1.6 Projection orthogonale sur un sous espace vectoriel

Définition 1.21 *Soit X un sous ensemble non vide d'un espace euclidien E. L'ensemble orthogonal de X, noté $X^\perp$, est défini par*

$$X^\perp := \{y \in E : \langle y, x \rangle = 0 \ \forall x \in X\} .$$

Proposition 1.22 *Soit X et Y deux sous ensembles non vides d'un espace euclidien E. Alors*

1. $X \subset Y \Rightarrow Y^{\perp} \subset X^{\perp}$.
2. $X^{\perp} = Vect\,(X)^{\perp}$.
3. $X^{\perp}$ *est un sous espace vectoriel de* E.
4. $Vect\,(X) = X^{\perp\perp} := \left(X^{\perp}\right)^{\perp}$. *En particulier,* $X^{\perp\perp} = X$ *si et seulement si* X *est un sous espace vectoriel de* E.

Preuve.

1. Facile.
2. D'après 1. il suffit de voir $X^{\perp} \subset Vect\,(X)^{\perp}$. Soient $y \in X^{\perp}$, $x_1, x_2, ..., x_p \in X$ et $\alpha_1, \alpha_2, ..., \alpha_p \in \mathbb{R}$. Alors

$$< y, \sum_{i=1}^{p} \alpha_i x_i > \quad = \sum_{i=1}^{p} \alpha_i \langle y, x_i \rangle = 0.$$

   D'où $y \in Vect\,(X)^{\perp}$.
3. Clairement $0 \in X^{\perp}$. Soient $x, y \in X^{\perp}$, $u \in X$ et $\alpha \in \mathbb{R}$. Alors

$$\langle x + \alpha y, u \rangle = \langle x, u \rangle + \alpha \langle y, u \rangle = 0.$$

   D'où $x + \alpha y \in X^{\perp}$.
4. Il suffit de voir $Vect\,(X) \subset X^{\perp\perp}$. Soit $x \in Vect\,(X)$. Pour tout $y \in X^{\perp}$ on a $\langle x, y \rangle = 0$ et donc $x \in \left(X^{\perp}\right)^{\perp}$.

$\square$

Théorème 1.23 *Pour tout sous espace vectoriel F d'un espace euclidien $(E, \langle .,. \rangle)$, l'orthogonal $F^{\perp}$ en est un supplémentaire, i.e.,*

$$E = F \oplus F^{\perp}.$$

Preuve. Remarquons que $F \cap F^{\perp} = \{0\}$. En effet, si $x \in F \cap F^{\perp}$ alors $\langle x, x \rangle = 0$ i.e. $x = 0$. Montrons que $E = F + F^{\perp}$. Considérons une base orthonormale $(e_1, e_2, ..., e_p)$ de F et complétons la en une base orthonormale $(e_1, e_2, ..., e_n)$ de E. Soit maintenant $x \in E$. Alors x s'écrit

$$x = \sum_{i=1}^{p} \langle x, e_i \rangle e_i + \sum_{i=p+1}^{n} \langle x, e_i \rangle e_i.$$

Comme $\sum\limits_{i=1}^{p} \langle x, e_i \rangle e_i \in F$, il suffit de montrer que $\sum\limits_{i=p+1}^{n} \langle x, e_i \rangle e_i \in F^{\perp}$. Remarquons que pour tout $i \in \{p+1, ..., n\}$ $e_i \in F^{\perp}$. En effet, si $u = \sum\limits_{i=1}^{p} \alpha_i e_i \in F$ alors pour $k \in \{p+1, ..., n\}$, $\langle e_k, u \rangle = \sum\limits_{i=1}^{p} \alpha_i \langle e_k, e_i \rangle = 0$ (car la famille $(e_1, e_2, ..., e_n)$ est orthonormale). Maintenant, comme $F^{\perp}$ est un sous espace vectoriel de E alors $\sum\limits_{i=p+1}^{n} \langle x, e_i \rangle e_i \in F^{\perp}$. $\square$

Définition 1.24 (Projection orthogonale) *Soit F un sous espace vectoriel de E. La projection orthogonale d'un vecteur $x \in \mathbb{R}^n$ sur F est la projection de x sur F parallèlement à $F^{\perp}$. Si $E = \mathbb{R}^n$, la matrice de cette projection est dite matrice de projection orthogonale sur F.*

Géométriquement, la projection orthogonale $p(x) = proj_F(x)$ de $x \in E$ sur F est le point de F le plus proche de x pour la distance d induite par la norme issue du produit scalaire de E, i.e.

$$d(x, p(x)) = \min_{u \in F} d(x, u).$$

Théorème 1.25 *Soient F un sous espace vectoriel de $\mathbb{R}^n$ et $x \in \mathbb{R}^n$. La* *projection orthogonale de x sur F est l'unique vecteur $p(x)$ de F qui réalise*

$$\|x - p(x)\| = \min_{u \in F} \|x - u\| .$$

Preuve. Soient $x \in E$ et $u \in F$. Comme $x - p(x) \in F^{\perp}$, $x - p(x) \perp p(x) - u$. D'où d'après l'identité de Pythagore, on a

$$\begin{aligned} \|x-u\|^2 &= \|x - p(x) + p(x) - u\|^2 \\ &= \|x - p(x)\|^2 + \|p(x) - u\|^2 \\ &\geq \|x - p(x)\|^2 . \end{aligned}$$

Clairement, ce dernier minimum n'est atteint que pour $u = p(x)$. D'où le résultat. □

Exercice 21 *Soit $(x_1, ..., x_n)$ une famille de vecteurs d'un espace euclidien E. On appelle matrice de Gram associée à $(x_1, ..., x_n)$ la matrice $G(x_1, ..., x_n) = (\langle x_i, x_j \rangle)_{1 \leq i,j \leq n}$.*

1. *Soit $\mathcal{B}$ une base orthonormale de l'espace vectoriel F engendré par la famille $(x_1, ..., x_n)$ et M la matrice de cette famille dans $\mathcal{B}$. Montrer que $G(x_1, ..., x_n) = {}^tMM$ et en déduire que $\det G(x_1, ..., x_n)$ est toujours positif.*
2. *En déduire que $G(x_1, ..., x_n)$ et $(x_1, ..., x_n)$ ont le même rang et que $G(x_1, ..., x_n)$ est inversible si et seulement si $(x_1, ..., x_n)$ est libre.*
3. *Supposons que $(x_1, ..., x_n)$ est libre. Montrer que pour $x \in E$,*

$$d(x, F) = \sqrt{\frac{\det G(x, x_1, ..., x_n)}{\det G(x_1, ..., x_n)}}.$$

1.7 Les quatre sous espaces vectoriels fondamentaux d'une matrice

Soit A une matrice de taille (m,n) et de rang r. Les quatre sous espaces vectoriels fondamentaux associés à la matrice A sont les images et noyaux de A et de sa transposée:

- ImA un sous espace vectoriel de $\mathbb{R}^m$ de dimension r.
- Ker$A$ un sous espace vectoriel de $\mathbb{R}^n$ de dimension $n-r$.
- Im ${}^tA$ un sous espace vectoriel de $\mathbb{R}^n$ de dimension r.
- Ker ${}^tA$ un sous espace vectoriel de $\mathbb{R}^m$ de dimension $m-r$.

1.7.1 Détermination des bases des quatre sous espaces vectoriels fondamentaux

Dans cette sous-section, pour une matrice A de taille (m,n), G désigne la matrice produit des matrices élémentaires représentant les opérations élémentaires rendant A sous sa forme échelonnée R. On a donc $R = GA$ et G est une matrice inversible.

Proposition 1.26 *Soit A une matrice de taille (m,n) et de rang r. Alors*

1. Im ${}^tA$ =Im ${}^tR$ *et donc les r vecteurs lignes pivots de R forment une base de* Im ${}^tA$.
2. Ker$A$ =Ker$R$ *et donc $n-r$ solutions particulières et linéairement indépendantes du système homogène $Rx = 0$ forment une base de* Ker$A$.

Preuve. Appliquer Exercice 11. □

Une base de ImA est determinée à l'aide de la proposition suivante:

Proposition 1.27 *Soient A et B deux matrices ayant le même nombre de colonnes. Alors*

1. *si* Ker$A$ =Ker$B$, *alors une famille de vecteurs colonnes de* A *est libre si et seulement si la famille de vecteurs colonnes de* B *dans les positions correspondantes est libre;*
2. *une famille de vecteurs colonnes de* A *est libre si et seulement si la famille de vecteurs colonnes de* R *dans les positions correspondantes est libre. Donc, les vecteurs colonnes pivots de* A, *i.e., les vecteurs colonnes occupant les positions correspondantes aux vecteurs colonnes de* R *contenant les pivots, forment une base de* ImA.

Preuve.

1. Soient $A = [a_1, ..., a_n]$ et $B = [b_1, ..., b_n]$. Il suffit de montrer que si $\left(a_{i_1}, ..., a_{i_p}\right)$ est une famille libre de vecteurs colonnes de A alors la famille $\left(b_{i_1}, ..., b_{i_p}\right)$ est libre. Pour ce faire, supposons que $\left(a_{i_1}, ..., a_{i_p}\right)$ est libre et soient $\alpha_{i_1}, ..., \alpha_{i_p} \in \mathbb{R}$ tels que $\alpha_{i_1} b_{i_1} + ... + \alpha_{i_p} b_{i_p} = 0$. Si $x = {}^t(x_1, ..., x_n)$ avec $x_k = \alpha_{i_r}$ si $k = i_r$ et $x_k = 0$ si $k \notin \{i_1, ..., i_p\}$, alors $Bx = 0$. Par hypothèse $Ax = 0$, i.e. $\alpha_{i_1} a_{i_1} + ... + \alpha_{i_p} a_{i_p} = 0$. D'où $\alpha_{i_1} = \alpha_{i_2} = ... = \alpha_{i_p} = 0$ et la famille $\left(b_{i_1}, ..., b_{i_p}\right)$ est libre.
2. Comme $R = GA$ et G est inversible, alors $Ax = 0 \Leftrightarrow Rx = 0$, i.e., Ker$A$ =Ker$R$. Le résultat s'ensuit donc de 1.

□

Il est clair qu'à l'instar du cas du noyau de A, on peut résoudre le système homogène ${}^tAx = 0$ pour aboutir à une base de Ker ${}^tA$. Cependant, les vecteurs formant la base de Ker ${}^tA$ apparaissent dans la matrice G. En effet, on a la proposition suivante:

Proposition 1.28 *Ecrivons* $G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$ *avec* G_1 *est de taille* (r, m) *(et donc* G_2 *est de taille* $(m - r, m)$*). Alors,* Ker ${}^tA$ =Im ${}^tG_2$ *et donc les* $m - r$ *vecteurs lignes de* G_2 *forment une base de* Ker ${}^tA$.

Preuve. Comme $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ avec R_1 est de taille (r, n), alors de $R = GA$ on a $G_2A = 0$. Soit $y \in$Im ${}^tG_2$, i.e., $y = {}^tG_2x$. D'où ${}^tAy = {}^tA\,{}^tG_2x = {}^t(G_2A)x = 0$, i.e., $y \in$Ker ${}^tA$. Pour l'autre inclusion, écrivons $G^{-1} = \begin{bmatrix} F_1, & F_2 \end{bmatrix}$ avec F_1 est de taille (m, r). De $GG^{-1} = G^{-1}G = I_m$ on tire aisement

$$G_1F_1 = I_r \;\text{ et }\; F_1G_1 = I_m - F_2G_2.$$

Soit maintenant $y \in$Ker ${}^tA$, i.e.,

$${}^tAy = {}^t(G^{-1}R)y = {}^t(F_1R_1)y = {}^tR_1\,{}^tF_1y = 0.$$

Comme ${}^tR_1$ est de rang plein (égal à r) alors de ${}^tR_1\,{}^tF_1y = 0$ on tire ${}^tF_1y = 0$. D'où ${}^t(F_1G_1)y = {}^tG_1\,{}^tF_1y = 0$. On en déduit que ${}^t(I_m - F_2G_2)y = y - {}^tG_2\,{}^tF_2y = 0$, i.e., $y \in$Im ${}^tG_2$. En conclusion, Ker ${}^tA =$Im ${}^tG_2$ comme souhaité. $\square$

Exemple 1.29 *Trouvons les bases des 4 sous espaces vectoriels fondamentaux de la matrice:*

$$A = \begin{bmatrix} 2 & -2 & -5 & -3 & -1 & 2 \\ 2 & -1 & -3 & 2 & 3 & 2 \\ 4 & -1 & -4 & 10 & 11 & 4 \\ 0 & 1 & 2 & 5 & 4 & 0 \end{bmatrix}.$$

L'élimination de Gauss donne la matrice échelonnée de A :

$$R = \begin{bmatrix} 2 & -2 & -5 & -3 & -1 & 2 \\ 0 & 1 & 2 & 5 & 4 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Une base de Im ${}^tA$ *est constituée des trois* (rg$A = 3$) *premières lignes pivots de*

R, i.e.,

$$\left({}^{t}(2,-2,-5,-3,-1,2)\,,\ {}^{t}(0,1,2,5,4,0)\,,\ {}^{t}(0,0,0,1,1,0)\right).$$

Les colonnes pivots de R étant la première, la deuxième et la quatrième colonne, une base de $\mathrm{Im}A$ *est donc*

$$\left({}^{t}(2,2,4,0)\,,\ {}^{t}(-2,-1,-1,1)\,,\ {}^{t}(-3,2,10,5)\right).$$

La solution dy système homogène $Rx = 0$ donne les trois ($6-\mathrm{rg}A$) solutions particulières:

$${}^{t}(\frac{1}{2},-2,1,0,0,0),\ {}^{t}(0,1,0,-1,1,0)\,,\ {}^{t}(-1,0,0,0,0,1)\,,$$

qui sont bien linéairement indépendantes et forment donc une base de $\mathrm{Ker}A$.

La determination d'une base de $\mathrm{Ker}\ {}^{t}A$ *nécessite la determination de la matrice G. Pour ce faire, on part de l'identité*

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -2 & -5 & -3 & -1 & 2 \\ 2 & -1 & -3 & 2 & 3 & 2 \\ 4 & -1 & -4 & 10 & 11 & 4 \\ 0 & 1 & 2 & 5 & 4 & 0 \end{bmatrix}.$$

L'étape 1 du processus d'élimination de Gauss consiste à annuler les coefficients sous le pivot de la première colonne de A, et ce en effectuant les opérations élémentaires $E_{21}(-1)$ (i.e. remplacer la ligne 2 de A par " ligne $2+(-1)\times$ligne 1") et $E_{31}(-2)$. Cela se traduit par la pré-multiplication de A par la matrice

$$G_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

en remplaçant dans la matrice identité $I_4 = (\delta_{ij})$, δ_{21} *par* -1 *(relativement à* $E_{21}(-1)$*) et* δ_{31} *par* -2 *(relativement à* $E_{31}(-2)$*). Le résultat est donc*

$$R_1 = G_1 A = \begin{bmatrix} 2 & -2 & -5 & -3 & -1 & 2 \\ 0 & 1 & 2 & 5 & 4 & 0 \\ 0 & 3 & 6 & 16 & 13 & 0 \\ 0 & 1 & 2 & 5 & 4 & 0 \end{bmatrix}.$$

L'étape 2 *de l'élimination de Gauss est d'annuler les coefficients sous le pivot de la deuxième colonne de* R_1, *et ce en effectuant les opérations élémentaires* $E_{32}(-3)$ *et* $E_{42}(-1)$. *Cela se traduit par la pré-multiplication de* R_1 *par la matrice*

$$G_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -3 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}.$$

Le résultat donne donc la matrice échelonnée de A :

$$G_2 G_1 A = R.$$

Il s'ensuit que la matrice de pré-multiplication d'élimination de Gauss est donnée par

$$G = G_2 G_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & -3 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}.$$

Par suite, la dernière ligne $(4 - \mathrm{rg} A)$ *de* G, *i.e.* ${}^t(1, -1, 0, 1)$, *constitue une base de* $\mathrm{Ker}\ {}^tA$.

1.7.2 Le théorème fondamental de l'algèbre linéaire

Théorème 1.30 *Soit A une matrice de taille (m, n) et de rang r. Alors*

1. $(\mathrm{Im}A)^{\perp} = \mathrm{Ker}\ {}^{t}A$ *et on a* $\mathbb{R}^m = \mathrm{Im}A \oplus \mathrm{Ker}\ {}^{t}A$;
2. $(\mathrm{Ker})^{\perp} =$*Im* ${}^{t}A$ *et on a* $\mathbb{R}^n =$Ker$A \oplus \mathrm{Im}\ {}^{t}A$.

Preuve. Notons que la seconde partie de chaque question est Théorème 1.23.

1. Montrons que $(\mathrm{Im}A)^{\perp}$ =Ker ${}^{t}A$. On a

$$\begin{aligned} y \in (\mathrm{Im}A)^{\perp} &\Leftrightarrow \quad \langle y, Ax\rangle = 0 \ \ \forall x \in \mathbb{R}^n \\ &\Leftrightarrow \quad {}^{t}x\ {}^{t}Ay = 0 \ \ \forall x \in \mathbb{R}^n \\ &\Leftrightarrow \ {}^{t}Ay = 0 \\ &\Leftrightarrow \ y \in \mathrm{Ker}\ {}^{t}A. \end{aligned}$$

   D'où $(\mathrm{Im}A)^{\perp}$ =Ker ${}^{t}A$.

2. On a d'après 1.

$$(\mathrm{Ker}A)^{\perp} = (\mathrm{Im}\ {}^{t}A)^{\perp\perp} = \mathrm{Im}\ {}^{t}A.$$

□

Figure 1.1 illustre l'apport du théorème fondamental de l'algèbre linéaire quant à l'action de la matrice A sur un vecteur $x \in \mathbb{R}^n$. En effet, celui-ci se décompose en une somme de x_r sa projection orthogonale sur Im ${}^{t}A$ et x_n sa projection orthogonale sur Ker$A$. Il s'ensuit que tout vecteur Ax de ImA est obtenu par l'image par A d'un vecteur x_r de Im ${}^{t}A$. Plus que cela, ce dernier vecteur x_r est unique!; s'il y en a un autre x'_r ∈Im ${}^{t}A$ vérifiant $Ax_r = Ax'_r$ alors $x_r - x'_r$ ∈Ker$A\cap$ Im ${}^{t}A = \{0\}$, i.e. $x_r = x'_r$. Ce fait montre que A cache une matrice carrée inversible de taille r (égale à dimIm ${}^{t}A$ = dimImA) représentant l'isomorphisme de Im ${}^{t}A$ vers ImA. L'inverse de cette matrice donne naissance à une matrice de taille (n, m) qui applique isomorphiquement ImA à Im ${}^{t}A$, notée A^{+} et reconnue pour le pseudo-inverse de Moore-Penrose de la matrice A. Cet inverse donne une

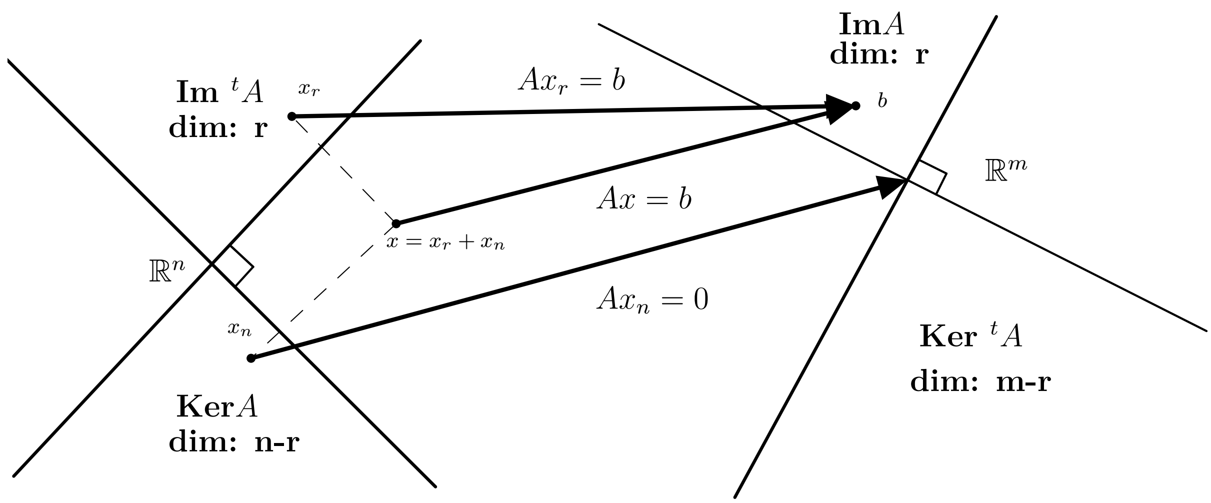


Figure 1.1: L'action de A sur un vecteur $x \in \mathbb{R}^n$

solution au sens des moindres carrés du système linéaire $Ax = b$ lorsque A n'est pas nécessairement de rang plein (on se limitera au cas A est de rang plein qui sera traité dans le paragraphe qui suit).

Exercice 22 (Alternative de Fredholm) *Soient A une matrice de taille (m, n) et $b \in \mathbb{R}^m$. Montrer qu'exactement un des deux problèmes suivants a une solution:*

1. $Ax = b$, $x \in \mathbb{R}^n$.
2. ${}^ty.A = 0$, $y \in \mathbb{R}^m$ *avec* ${}^ty.b \neq 0$.

1.8 Projection orthogonale et l'equation normale

Il est fréquemment commode dans le contexte de la statistique de faire la projection orthogonale sur l'espace image ImA d'une matrice A de rang plein. En effet, si l'on projète orthogonalement sur un sous espace vectoriel F de $\mathbb{R}^n$, quite à considérer une base $(a_1, a_2, ..., a_r)$ de F, la matrice A est donnée par $A = [a_1, a_2, ..., a_r]$. Dorénavant, la matrice d'une telle projection orthogonale sera simplement notée P_A. Soient maintenant $b \in \mathbb{R}^n$ et $p = P_Ab$ sa projection orthogonale sur $F =$ImA.

Le vecteur p est donné par les conditions:

$$p = A\hat{x}\ ,\ \hat{x} \in \mathbb{R}^r \quad \text{et} \quad b - p \perp \mathrm{Im}A = Vect\,(a_1, a_2, ..., a_r)\,.$$

Cela signifie que $\hat{x}$ est solution du système linéaire à r équations suivant:

$$\begin{cases} {}^t a_1 \left(b - A\hat{x}\right) = 0 \\ {}^t a_2 \left(b - A\hat{x}\right) = 0 \\ \qquad \vdots \\ {}^t a_r \left(b - A\hat{x}\right) = 0. \end{cases}$$

Le dernier système s'écrit matriciellement:

$${}^t A \left(b - A\hat{x}\right) = 0$$

soit

$${}^t AA\hat{x} \ = \ {}^t Ab.$$

La dernière équation est reconnue pour l'équation normale associée au système linéaire $Ax = b$, qui a une solution unique même si le système linéaire $Ax = b$ est inconsistent. En effet, comme A est de rang plein, ${}^t AA$ est inversible et l'équation normale a pour solution

$$\hat{x} \ = (\ {}^t AA)^{-1} \ {}^t Ab.$$

La solution $\hat{x}$ réalise bien $\min_{x \in \mathbb{R}^r} \|Ax - b\| = ||A\hat{x} - b||$. Elle est donc dite la solution au sens des moindres carrés du système linéaire $Ax = b$. Ainsi, la projection orthogonale de b sur $\mathrm{Im}A$ est donnée par

$$p = A\hat{x} \ = A(\ {}^t AA)^{-1} \ {}^t Ab.$$

On en déduit que la matrice de projection orthogonale sur $\mathrm{Im}A$ est donnée par

$$P_A = A(\ {}^t AA)^{-1} \ {}^t A. \tag{1.2}$$

Les propriétés de la matrice P_A, à l'instar de celles de P_a de Proposition 1.16, sont laissées à titre d'exercice:

Exercice 23 *Soit A une matrice de taille (n, r) et de rang plein. Vérifier que*

1. *P_A est une matrice idempotente, symétrique et de rang égal à r.*
2. *$P_A b = b$ si $b \in \mathrm{Im} A$ et $P_A b = 0$ si $b \in (\mathrm{Im} A)^{\perp}$.*

En résumé, on a les deux théorèmes suivants:

Théorème et définition 1 *Soit A une matrice de taille (n, r) et de rang plein. Alors*

1. *La projection orthogonale de $b \in \mathbb{R}^n$ sur $\mathrm{Im} A$ est donnée par $A\hat{x}$ où $\hat{x} \in \mathbb{R}^r$ est l'unique solution de l'équation dite équation normale:*

$$ {}^tAA\hat{x} = {}^tAb. $$

2. *La solution $\hat{x}$ réalise*

$$ \min_{x \in \mathbb{R}^r} \|Ax - b\| = ||A\hat{x} - b|| $$

et est dite solution au sens des moindres carrés du système linéaire (inconsistent en général) $Ax = b$.

Théorème 1.31 *Soit A une matrice de taille (n, r) et de rang plein. Alors*

1. *La matrice de projection orthogonale sur $\mathrm{Im} A$ est donnée par*

$$ P_A = A(\,{}^tAA)^{-1}\,{}^tA. $$

2. *P_A est une matrice idempotente, symétrique et de rang égal à r.*

Remarque 1.32 *Si l'on projète sur une droite vectoriel, la matrice A est dans ce cas réduite à un vecteur colonne a, et on retrouve bien l'expression* (1.1) *de P_a.*

Exemple 1.33 *Déterminons la projection orthogonale $proj_F\,(b)$:*

$$b = {}^t(6,0,0) \quad , \quad F = Vect({}^t(1,1,1)\,,\ {}^t(0,1,2)).$$

On projète sur ImA *avec* $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ *est une matrice de rang plein. On a* ${}^tAA = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$ *et* ${}^tAb = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$. *L'équation normale s'écrit:*

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix},$$

et sa solution est $\hat{x} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$. *D'où* $proj_F\,(b) = A\hat{x} = {}^t(5,2,-1)\,.$

La matrice de projection orthogonale ne dépend pas de la base choisie pour le sous espace vectoriel sur lequel on projète. En effet, la formule matricielle (1.2) est invariante au choix de la matrice A, ce que l'on démontre dans la proposition suivante:

Proposition 1.34 *Soient A et B deux matrices de taille (n, r) et de rang plein telles que* ImA =ImB. *Alors $P_A = P_B$.*

Preuve. Comme ImA =ImB, d'après Exercice 12 il existe une matrice carrée inversible C de taille r telle que $AC = B$. D'où l'on a

$$\begin{aligned} P_B &= B(\,{}^tBB)^{-1}\,{}^tB \\ &= AC(\,{}^tC\,{}^tAAC)^{-1}\,{}^tC\,{}^tA \\ &= ACC^{-1}(\,{}^tAA)^{-1}(\,{}^tC)^{-1}\,{}^tC\,{}^tA \quad (\,{}^tC \text{ est aussi inversible}) \\ &= A(\,{}^tAA)^{-1}\,{}^tA = P_A. \end{aligned}$$

□

Exercice 24 (Contrôle final, automne 2017) *1. Soient A et B deux matrices de tailles (m,p) et (m,q) respectivement. Montrer que* $\mathrm{Im}A \perp \mathrm{Im}B$ *si et seulement si $({}^tA)\,B = 0$ et $({}^tB)\,A = 0$.*

2. Supposons que la matrice $C = [A, B]$ est de rang plein. Montrer que $P_C = P_A + P_B$.

Exercice 25 (Formule matricielle d'un projecteur) *Soient $\mathbb{R}^n = F \oplus G$, $(a_1, ..., a_r)$ une base de F $(r < n)$ et $(b_1, ..., b_{n-r})$ une base de G. Soient les matrices $A = [a_1, ..., a_r]$, $B = [b_1, ..., b_{n-r}]$, $S = [A, B]$ et P est la matrice de projection sur F parallèlement à G.*

1. *Calculer PS et en déduire que $P = AC$ où C est une matrice à déterminer.*
2. *Soient $(v_1, ..., v_r)$ une base de $G^{\perp}$ et $V = [v_1, ..., v_r]$.*
   (a) *Montrer que ${}^tV = {}^tV\ AC$.*
   (b) *En déduire que* $\mathrm{Im}(\,{}^tVA) = \mathrm{Im}(\,{}^tV)$ *et que ${}^tVA$ est inversible.*
3. *En déduire que la matrice P est donnée par la formule:*

$$P = A(\,{}^tVA)^{-1}\,{}^tV.$$

4. *Commenter ce résultat.*

Exercice 26 (Contrôle final, automne 2018) *Soient A une matrice réelle de taille (m,n) $(m > n)$ et de rang plein, $Ax = b$ un système linéaire surdéterminé et $\hat{x}$ sa solution au sens des moindres carrés. On actualise le système $Ax = b$ par* *l'ajout de k équations représentées par le système linéaire $Vx = c$ où V est une matrice de taille (k, n) et $c \in \mathbb{R}^k$. Montrer que la solution actualisée au sens des moindres carrés est donnée par*

$$\hat{x}_a = \hat{x} + K(c - V\hat{x})$$

où $K = B^{-1}\, {}^tV(I_k + VB^{-1}\, {}^tV)^{-1}$ (la matrice $I_k + VB^{-1}\, {}^tV$ étant inversible) et $B = {}^tAA$ (ind. utiliserla formule de Sherman-Woodbury-Morrison; Exercice...).

1.8.1 Application: ajustement affine par la méthode des moindres carrés

Soit $\mathcal{N}$ un nuage de points (x_i, y_i) résultat de n observations x_i et y_i, $i = 1, 2, ..., n$, de deux variables statistiques quantitatives X et Y respectivement, telles que les réels x_i ne sont pas tous égaux. On suppose qu'on peut faire un ajustement affine (à des fins de prédiction par exemple) de la variable Y en fonction de la variable X. On cherche donc à trouver la droite D ajustant le mieux possible le nuage de points $\mathcal{N}$. La méthode des moindres carrés consiste à trouver la droite D, dite droite des moindres carrés, parmi les droites d'équation $Y = a + bX$ où a, b sont des constantes réelles, telle que D minimise la somme des carrés des résidus (ou erreurs) $e = \sum\limits_{i=1}^{n} (y_i - (a + bx_i))^2 = \|\varepsilon\|^2$, où $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)$, $\varepsilon_i = y_i - (a + bx_i)$, est le vecteur des résidus.

Il s'agit donc de résoudre au sens des moindres carrés le système linéaire d'inconnues a et b :

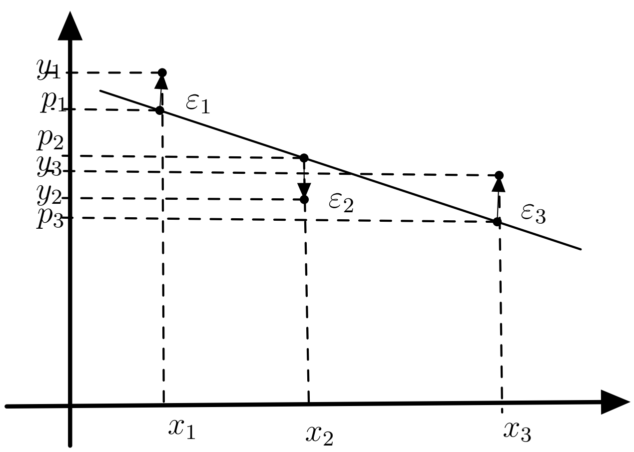
$$\begin{cases} a + bx_1 = y_1 \\ a + bx_2 = y_2 \\ \quad\vdots \\ a + bx_n = y_n \end{cases}$$

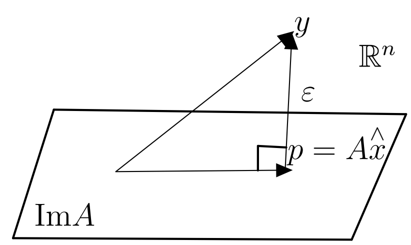
qui s'écrit matriciellement

$$Ax = y$$

où $x = \begin{bmatrix} a \\ b \end{bmatrix}$, $y = {}^t(y_1, y_2, ..., y_n)$ et $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ est une matrice de rang plein.

D'après Théorème 1.25, la somme des carrés des résidus $e = \|y - Ax\|^2$ est minimum lorsque Ax est la projection orthogonale de y sur ImA, i.e., lorsque $x = \hat{x}$ où $\hat{x}$ est la solution unique de l'équation normale associée au système linéaire $Ax = y$.





Projection orthogonale de l'équation $Ax = y$ sur ImA

Figure 1.2: Droite des moindres carrées et projection orthogonale

On a donc ${}^tAA = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$ et ${}^tAb = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$. D'où l'équation normale s'écrit:

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix},$$

et sa solution déterminant la droite des moindres carrés D est donnée par

$$\begin{aligned} a &= \frac{\left(\sum\limits_{i=1}^{n} y_i\right)\left(\sum\limits_{i=1}^{n} x_i^2\right) - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} x_i y_i\right)}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \\ b &= \frac{n\sum\limits_{i=1}^{n} x_i y_i - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} y_i\right)}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2}. \end{aligned}$$

Exercice 27 (Contrôle continu, automne 2017) *Soit le nuage de points* (x_i, y_i) *résultat de* n *observations* x_i *et* y_i, $i = 1, 2, ..., n$, *de deux variables statistiques quantitatives* X *et* Y *respectivement, telles que les réels* x_i *ne sont pas tous égaux. Posons* $x = {}^t(x_1, x_2, ..., x_n)$ *et* $y = {}^t(y_1, y_2, ..., y_n)$.

1. (a) Donner les écritures matricielles, en fonction de x et d'un vecteur colonne e à déterminer, de la moyenne arithmétique $\overline{X}$ et de la variance $V(X)$ des x_i, $i = 1, 2, ..., n$.
   (b) Donner l'écriture matricielle, en fonction de x et y, de la covariance $Cov(X, Y)$ des deux variables X et Y.
2. On suppose qu'on peut faire un ajustement affine de la variable Y en fonction de la variable X. Soit D la droite des moindres carrés d'équation $Y = a + bX$, où a, b sont des constantes réelles. Écrire le système linéaire $Au = y$, d'inconnue u, à résoudre au sens des moindres carrés, où A et u sont respectivement une matrice et un vecteur colonne à déterminer.
3. Soit $\varepsilon(u) = y - Au$ le vecteur des résidus et posons $\varepsilon(u) = {}^t(\varepsilon_1(u), \varepsilon_2(u), ..., \varepsilon_n(u))$. Montrer que $\sum\limits_{i=1}^{n} \varepsilon_i\left(\hat{u}\right) = 0$, où $\hat{u}$ est l'unique solution de l'équation normale (E) associée au système linéaire $Au = y$.
4. Écrire la matrice A à l'aide des deux vecteurs colonnes e et x et montrer que

l'équation normale (E) s'écrit:

$$\begin{cases} na + \left({}^t e.x\right) b = {}^t e.y \\ \left({}^t e.x\right) a + \left({}^t x.x\right) b = {}^t x.y \end{cases}$$

5. En déduire que la droite D a pour équation:

$$Y = \frac{Cov\,(X,Y)}{V\,(X)} \left(X - \overline{X}\right) + \overline{Y}.$$

Exercice 28 (Méthode des moindres carrés - approche différentielle) *Soient $A = (a_{ij})$ une matrice carrée de taille n et $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, définie par*

$$f\,(x) = \langle Ax, x\rangle = {}^t xAx.$$

1. *Donner $f\,(x)$ en fonction des x_i et a_{ij}, où $x = {}^t\,(x_1, ..., x_n)$. En déduire l'expression de la dérivée partielle $\frac{\partial f}{\partial x_p}$, $p = 1, 2, ..., n$.*
2. *Soit ∇f le gradient de f défini par $\nabla f\,(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$, $x \in \mathbb{R}^n$. Montrer que*

$$\nabla f\,(x) = \left(A + {}^t A\right) x, \quad x \in \mathbb{R}^n.$$

3. *Soient B une matrice de taille (m, n), $x \in \mathbb{R}^n$ et $b \in \mathbb{R}^m$. Montrer que*

$$\|Bx - b\|^2 = {}^t x\, {}^t BBx \;\; - 2\, {}^t x\, {}^t Bb + {}^t bb.$$

4. *En déduire que si $\|Bx - b\|^2$ est minimum alors x est solution de l'équation normale associée au système linéaire $Bx = b$.*

1.9 Matrices orthogonales et décomposition QR

Soit $(q_i)_{i=1}^n$ une famille orthonormale de $\mathbb{R}^m$. Considérons la matrice de taille (m,n) $Q = [q_1, ..., q_n]$. Alors, on a

$$ {}^tQQ = \begin{bmatrix} {}^tq_1 \\ \vdots \\ {}^tq_n \end{bmatrix} [q_1, ..., q_n] = \left({}^tq_i.q_j \right)_{1 \leq i,j \leq n} = (\delta_{ij})_{1 \leq i,j \leq n} = I_n. $$

Si de plus Q est une matrice carrée alors elle est inversible et $Q^{-1} = {}^tQ$. On a donc la proposition suivante:

Proposition 1.35 *Pour une matrice carrée Q de taille n, les assertions suivantes sont équivalentes:*

1. *les vecteurs colonne de Q forment une famille orthonormale de $\mathbb{R}^n$.*
2. ${}^tQQ = Q\,{}^tQ = I_n$ *, i.e., Q est inversible et $Q^{-1} = {}^tQ$.*

Définition 1.36 *Une matrice carrée A de taille n est dite matrice orthogonale s'elle vérifie l'une des deux assertions de la proposition précédente.*

Exemple 1.37 1. *Matrices de rotation dans $\mathbb{R}^2$:*

$$ R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. $$

On a $R(\theta)^{-1} = {}^tR(\theta) = R(-\theta)$.

2. *Matrice de permutation:*

$$ Q = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}. $$

$$On\ a\ Q^{-1} = {}^tQ = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.\ Notons\ que\ Q\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix}.$$

Les matrices orthogonales de taille n sont des transformations vectorielles de $\mathbb{R}^n$ conservant le produit scalaire et donc la norme, généralisant ainsi à une dimension quelconque les isométries vectorielles de $\mathbb{R}^2$ et $\mathbb{R}^3$ (e.g. rotations, réflexions...):

Proposition 1.38 *Pour une matrice orthogonale Q de taille n, on a*

1. $\langle Qx, Qy\rangle = \langle x, y\rangle$, $x, y \in \mathbb{R}^n$;
2. *en particulier,* $\|Qx\| = \|x\|$, $x \in \mathbb{R}^n$.

1.9.1 Décomposition QR d'une matrice de rang plein

Soit $A = [a_1, ..., a_p]$ une matrice de taille (n, p) et de rang plein. Appliquons l'algorithme de Gram-Schmidt aux colonnes de A :

$$\begin{aligned} b_1 &= a_1; \quad q_1 = \frac{b_1}{\|b_1\|}; \\ b_2 &= a_2 - proj_{b_1}(a_2); \quad q_2 = \frac{b_2}{\|b_2\|}; \\ &\vdots \\ b_p &= a_p - \sum_{i=1}^{p-1} proj_{b_i}(a_p); \quad q_p = \frac{b_p}{\|b_p\|}. \end{aligned}$$

Pour $k = 1, 2, ..., p$, on a $\langle q_k, b_k\rangle = \langle q_k, a_k\rangle = \|b_k\|$ et par suite $b_k = \langle q_k, a_k\rangle .q_k$. D'où l'on obtient

$$\begin{aligned} a_1 &= \langle q_1, a_1\rangle .q_1 \\ a_2 &= \langle q_1, a_2\rangle .q_1 + \langle q_2, a_2\rangle .q_2 \\ &\vdots \\ a_p &= \langle q_1, a_p\rangle .q_1 + \langle q_2, a_p\rangle .q_2 + ... \langle q_p, a_p\rangle .q_p. \end{aligned}$$

Cela s'écrit matriciellement $A = QR$, où $Q = [q_1, ..., q_p]$ (la matrice obtenue par l'orthonormalisation des colonnes de A) et

$$R = \begin{bmatrix} \langle q_1, a_1\rangle & \langle q_1, a_2\rangle & ... & \langle q_1, a_p\rangle \\ 0 & \langle q_2, a_2\rangle & ... & \langle q_2, a_p\rangle \\ \vdots & 0 & \ddots & \vdots \\ 0 & ... & 0 & \langle q_p, a_p\rangle \end{bmatrix}$$

est une matrice triangulaire suppérieure dont les éléments diagonaux sont tous strictement positifs ($\langle q_k, a_k\rangle = \|b_k\| > 0$ pour tout k).

Théorème 1.39 (Décomposition QR) *Soit A une matrice de taille (n,p) et de rang plein. Alors A admet une factorisation $A = QR$ avec Q est une matrice de taille (n,p) dont les vecteurs colonne forment une base orthonormale de* ImA *et R est une matrice triangulaire suppérieure de taille p dont les éléments diagonaux sont tous strictement positifs.*

Exemple 1.40 *Soit la matrice* $A = \begin{bmatrix} -1 & 1 \\ 2 & 1 \\ 2 & 4 \end{bmatrix} = [a_1, a_2]$. *L'algorithme d'orthogonalisation de Gram-Schmidt donne:*

$$\begin{aligned} b_1 &= a_1; \quad q_1 = \frac{b_1}{\|b_1\|} = \frac{1}{3}\ {}^t(-1, 2, 2)\,; \\ b_2 &= a_2 - proj_{y_1}(a_2) = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}; \quad q_2 = \frac{b_2}{\|b_2\|} = \frac{1}{3}\ {}^t(2, -1, 2)\,. \end{aligned}$$

D'où $Q = \frac{1}{3}\begin{bmatrix} -1 & 2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix}$. *De $A = QR$ on tire $R = {}^tQA = \begin{bmatrix} 3 & 3 \\ 0 & 3 \end{bmatrix}$. La décompo-*

sition QR de A s'écrit donc

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

La dite décomposition QR est particulièrement utile dans la résolution au sens des moindres carrés d'un système linéaire $Ax = y$. En effet, A étant une matrice de taille (n, p) et de rang plein, elle admet une décomposition $A = QR$ comme est décrit dans le théorème précédent. L'équation normale s'écrit donc

$${}^tR\,{}^tQQR.x = {}^tR\,{}^tQ.y \quad \text{i.e.} \quad {}^tR\,R.x = {}^tR\,{}^tQ.y.$$

Comme ${}^tR$ est inversible, alors la dernière équation devient $R.x = {}^tQ.y$. Cette équation est un système linéaire triangulaire qui se résout facilement par une substitution à l'envers.

Exercice 29 (Unicité de la décomposition QR) 1. *Montrer que pour toute matrice R triangulaire suppérieure dont les éléments diagonaux sont tous strictement positifs, si ${}^tRR = I$ alors $R = I$.*

2. *En déduire que la décomposition QR du Théorème 1.39 est unique.*

Exercice 30 (Décomposition QR- autre version) *Soit A une matrice de taille (n, p) et de rang plein. Soit $A = Q_1R_1$ la décomposition du Théorème 1.39.*

1. *MATLAB est programmé pour donner une décomposition $A = QR$ telle que $Q = [Q_1, Q_2]$ est une matrice orthogonale de taille n et $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ est une matrice de taille (n, p) , R_1 étant une matrice triangulaire suppérieure inversible de taille p. Montrer qu'une telle décomposition existe.*
2. *Quel sous espace véctoriel parmi les quatre sous espaces véctoriels fondamentaux de A les vecteurs colonnes de Q_2 en forment une base?*
3. *La décomposition $A = QR$ est-elle unique?*

Exercice 31 (Rattrapage, automne 2018) *Soient A une matrice de taille (n, m) et de rang plein et $y \in \mathbb{R}^n$. Soit la décomposition $A = QR$ de l'exercice précédent. Posons* ${}^tQy = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$, $c_1 \in \mathbb{R}^m$ *et* $c_2 \in \mathbb{R}^{n-m}$.

1. Montrer que pour tout $x \in \mathbb{R}^m$, on a

$$\|y - Ax\|^2 = \|c_1 - R_1 x\|^2 + \|c_2\|^2 .$$

2. En déduire que la solution au sens des moindres carrées du système linéaire $Ax = y$ est donnée par $\hat{x} = R_1^{-1} c_1$ et la somme des carrés des résidus $\|y - A\hat{x}\|^2$ est égale à $\|c_2\|^2$.

Exercice 32 (Inégalité de Hadamard) *Montrer que pour une matrice carrée* $A = [a_1, ..., a_n]$, *on a*

$$|\det A| \leq \prod_{i=1}^{n} \|a_i\| .$$

Etudier l'égalité.

Chapitre 2

Décompositions spectrale et en valeurs singulières

2.1 Préliminaire

Définition 2.1 *Soit A une matrice (réelle) carrée de taille n. Un scalaire $\lambda \in \mathbb{C}$ est dit une valeur propre de A s'il existe $x \in \mathbb{C}^n \backslash \{0\}$ solution de l'équation $Ax = \lambda x$. Le vecteur x est dit un vecteur propre de A associé à la valeur propre λ.*

Notons que si une valeur propre λ de A est réelle, alors il existe toujours un vecteur propre réel de A associé à λ, qui est une solution du système linéaire $(A - \lambda I)\, x = 0$ d'inconnue x et à coefficients réels. Deux types essentiels de matrices (réelles) ayant toutes ses valeurs propres réelles sont les matrices triangulaires et les matrices symétriques. Le sous espace vectoriel $\mathrm{Ker}(A - \lambda I_n)$ de $\mathbb{C}^n$ est dit le sous espace propre de A associé à λ. La dimension de $\mathrm{Ker}(A - \lambda I_n)$ est dite la multiplicité géométrique de la valeur propre λ et est notée $mg_A(\lambda)$. Le polynôme

$$\chi_A(z) = |zI_n - A| = (z - \lambda_1)(z - \lambda_2) \ldots (z - \lambda_n), \quad z \in \mathbb{C}$$

est dit le polynôme caractéristique de A. Les λ_i étant les n racines (éventuellement complexes) de χ_A qui sont aussi les n valeurs propres de A (répétition permise). Comme les coefficients de χ_A sont réels, alors les valeurs propres complexes non réelles de A, si elles existent, doivent être des paires λ, $\overline{\lambda}$ avec $\lambda \in \mathbb{C} \backslash \mathbb{R}$.

Exercice 33 *Soient A et B deux matrices de tailles (m,n) et (n,m) respectivement.*

1. *Montrer que $z^m\chi_{BA}(z) = z^n\chi_{AB}(z)$, $z \in \mathbb{C}$ (ind. calculer $\begin{vmatrix} zI_m & zA \\ B & zI_n \end{vmatrix}$, $z \neq 0$).*
2. *En déduire que AB et BA ont les mêmes valeurs propres non nulles.*

Le caractère spécifique du corps $\mathbb{C}$ en tant que corps algébriquement clos (i.e. tout polynôme non constant à coefficient dans $\mathbb{C}$ admet au moins une racine) justifie la considération des valeurs complexes dans la définition d'une valeur propre. De ce fait, on aura besoin dans cette section de considérer l'espace hermitien canonique $\mathbb{C}^n$, i.e. munir le $\mathbb{C}$-espace vectoriel $\mathbb{C}^n$ de son produit scalaire canonique:

Définition 2.2 *Le produit scalaire (ou produit hermitien) canonique sur le $\mathbb{C}$-espace vectoriel $\mathbb{C}^n$ est défini par*

$$\langle x, y\rangle := y^*x = \sum_{i=1}^{n} \overline{y_i}x_i$$

où pour $x = {}^t(x_1, ..., x_n) \in \mathbb{C}^n$,

$$x^* := {}^t\overline{x} = \overline{{}^tx} = (\overline{x_1}, ..., \overline{x_n})$$

dénote le vecteur adjoint de x. La norme issue de ce produit scalaire est donnée par

$$\|x\| = (x^*x)^{\frac{1}{2}} = (\sum_{i=1}^{n} |x_i|^2)^{\frac{1}{2}}$$

où $|x_i|$ dénote le module du complexe x_i.

Notons ici que $x^*x = \sum\limits_{i=1}^{n} |x_i|^2$ est toujours un réel (alors que ${}^tx.x$ ne l'est pas en général). Le produit scalaire canonique de $\mathbb{C}^n$ diffère de celui de $\mathbb{R}^n$ du fait qu'il s'agit d'une forme hermitienne définie positive

Définition 2.3 *Une application* $\langle .,. \rangle : E \times E \longrightarrow \mathbb{C}$*,* E *étant un espace vectoriel complexe, est dite une forme hermitienne définie positive si elle satisfait les trois propriétés suivantes:*

1. $\langle x, y \rangle = \overline{\langle y, x \rangle}$ *(symétrie hermitienne).*
2. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$ *(linéarité à gauche).*
3. $\langle x, x \rangle \geq 0$ *et* $\langle x, x \rangle = 0$ *si et seulement si* $x = 0$*.*

Il s'ensuit des deux premières propriétés d'un produit hermitien que celui-ci est linéaire par conjugaison à droite, i.e.

$$\langle x, \alpha y + \beta z \rangle = \overline{\alpha} \langle x, y \rangle + \overline{\beta} \langle x, z \rangle .$$

Le calcul matriciel nécessite de définir à l'instar de l'adjoint d'un vecteur l'adjoint d'une matrice A à coefficents dans $\mathbb{C}$ qui est la matrice

$$A^* := \overline{{}^t A}$$

où l'on dénote par $\overline{B}$ la matrice dont les coefficients sont les conjugués de ceux de B. Comme dans le contexte statistique les matrices sont réelles, alors leurs matrices adjointes ne sont autres que leurs transposées.

Le nombre de fois que se répète une valeur propre λ en tant que racine du polynôme caractéristique de A est dit la multiplicité algébrique de λ et est notée $ma_A(\lambda)$. On a toujours $1 \leq mg_A(\lambda) \leq ma_A(\lambda)$. Si cette dernière inégalité est une égalité, la valeur propre λ est dite régulière.

Définition 2.4 *Une matrice carrée* A *de taille* n *est dite diagonalisable si elle est semblable à une matrice diagonale* D *de taille* n *à coefficients dans* $\mathbb{C}$ *via une matrice inversible* P *de taille* n *à coefficients dans* $\mathbb{C}$*, i.e.,* $A = PDP^{-1}$*.*

Exercice 34 *Montrer qu'une matrice* A *de taille* n *non nulle et nilpotente n'est pas diagonalisable.*

Théorème 2.5 *Soit A une matrice carrée de taille n. Alors, les assertions suivantes sont équivalentes:*

1. *A est diagonalisable;*
2. *toute valeur propre de A est régulière;*
3. *si $\sigma(A) = \{\lambda_1, \lambda_2, ..., \lambda_k\}$ est l'ensemble des valeurs propres de A distinctes deux à deux, alors*

$$\mathbb{C}^n = \mathrm{Ker}(A - \lambda_1 I_n) \oplus \mathrm{Ker}(A - \lambda_2 I_n) \oplus ... \oplus \mathrm{Ker}(A - \lambda_k I_n);$$

4. *il existe une base $(x_1, ..., x_n)$ de $\mathbb{C}^n$ constituée de vecteurs propres de A.*

2.1.1 Décomposition spectrale des matrices réelles symétriques

Les matrices réelles symétriques sont d'une importance particulière dans la statistique. Les matrices de covariance en sont un exemple fréquent. Ce type de matrices admet un certain nombre de propriétés que l'on ne retrouve pour d'autres.

Proposition 2.6 *Soit A une matrice réelle symétrique de taille n. Alors*

1. *les valeurs propres de A sont réelles;*
2. *deux vecteurs propres réels de A associés à deux valeurs propres distinctes sont orthogonaux.*

Preuve.

1. Soit x un vecteur propre de A associé à une valeur propre λ. Alors comme $x^*Ax = \lambda x^*x$, passant à l'adjoint dans cette égalité, et comme $A^* = A$, on a

$$\bar{\lambda} x^*x = (x^*Ax)^* = x^*Ax = \lambda x^*x.$$

Comme $x \neq 0$ alors $x^*x > 0$. D'où $\bar{\lambda} = \lambda$, i.e. λ est réelle.

2. Soient x_1, x_2 deux vecteurs propres réels de A associés respectivement à deux valeurs propres distinctes λ_1, λ_2. On a donc $Ax_1 = \lambda_1 x_1$ et $Ax_2 = \lambda_2 x_2$. D'où ${}^t x_2 A x_1 = \lambda_1 \, {}^t x_2 x_1$ et ${}^t x_1 A x_2 = \lambda_2 \, {}^t x_1 x_2$. Comme A est symétrique, alors ${}^t x_2 A x_1 = {}^t x_1 A x_2$. Il s'ensuit que $(\lambda_1 - \lambda_2) \, {}^t x_1 x_2 = 0$, et donc ${}^t x_1 x_2 = 0$, i.e. $x_1 \perp x_2$.

□

La propriété fondamentale des matrices réelles symétriques consiste à leur diagonalisabilité via une matrice orthogonale de changement de base.

Théorème 2.7 (Le théorème spectral) *Soit A une matrice réelle symétrique de taille n. Alors, il existe une base orthonormale $(p_1, ..., p_n)$ de $\mathbb{R}^n$ formée de vecteurs propres de A associés respectivement à ses valeurs propres $\lambda_1, ..., \lambda_n$. Les équations*

$$Ap_i = \lambda_i p_i \tag{2.1}$$

s'écrivent matriciellement

$$AP = PD \textit{ soit } A = PD \, {}^t P \tag{2.2}$$

où $P = [p_1, ..., p_n]$ est la matrice orthogonale de taille n dite matrice de changement de bases, et $D = diag(\lambda_1, ..., \lambda_n)$.

Géométriquement, le théorème spectral fournit une nouvelle base orthonormale $\mathcal{B}$ de $\mathbb{R}^n$ (ou nouveau système orthonormale de coordonnées) constituée de vecteurs propres de A figurant comme vecteurs colonnes de la matrice de changement de bases P. La matrice de la transformation vectorielle $A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $x \mapsto Ax$ relativement à la base $\mathcal{B}$ est diagonale. Cela veut dire que si $X, Y \in \mathbb{R}^n$ sont respectivement les vecteurs des coordonnées de x et Ax dans $\mathcal{B}$, alors x et Ax s'écrivent $x = PX$ et $Ax = PY$. D'où $APX = PY$, soit

$$Y = {}^t PAPX = DX = diag(\lambda_1, ..., \lambda_n) X.$$

La transformation vectorielle $A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ se comprend mieux dans le nouveau système de coordonnées comme étant un simple opérateur de multiplication des coordonnées de X par des scalaires qui ne sont autres que les valeurs propres de A. Les vecteurs propres de A constituant la base $\mathcal{B}$ servent de directions des axes du nouveau système de coordonnées. Une rotation dans $\mathbb{R}^2$ ou $\mathbb{R}^3$ étant un exemple de matrice orthogonale, un exemple illustratif de changement de bases dans ces deux derniers espaces consiste à appliquer une rotation au système initial de coordonnées (la base canonique) pour obtenir le nouveau système de coordonnées.

Exercice 35 *Soit dans* $\mathbb{R}^2$ *l'ellipse* (E) *d'équation cartésienne* $5x^2+8xy+5y^2 = 1$.

1. *Donner une matrice symétrique* A *qui permet l'écriture matricielle* ${}^t uAu = 1$, $u = \begin{bmatrix} x \\ y \end{bmatrix}$, *de l'équation de* (E).
2. *Donner une décomposition spéctrale de* A.
3. *En déduire deux vecteurs directeurs des deux axes de* (E) *et sa nouvelle équation relativement au système de coordonnées constitué de ces deux axes.*

2.2 Le quotient de Rayleigh

Pour une matrice carrée A de taille n et $x \in \mathbb{R}^n \setminus \{0\}$, on considère le problème de minimisation:

$$\arg\min_{\alpha \in \mathbb{R}} \ \|Ax - \alpha x\| . \tag{2.3}$$

La solution α_0 de ce problème nous donne une "estimation d'une valeur propre de A" dans le sens où si x est proche (pour la norme) d'un vecteur propre de A associé à une valeur propre λ, alors α_0 donne une "estimation" de λ.

Proposition 2.8 *La solution du problème* (2.3) *est donné par le quotient*

$$R_A(x) = \frac{{}^t xAx}{{}^t xx}.$$

Le quotient $R_A(x)$, $x \in \mathbb{R}^n \backslash \{0\}$, est appelé le quotient de Rayleigh.

Preuve. Résoudre le problème (2.3) revient à résoudre au sens des moindres carrés le système linéaire d'inconnue α :

$$x\alpha = Ax.$$

La matrice coéfficient de ce système étant le vecteur x. L'équation normale s'écrit donc

$$^t xx\alpha = {}^t xAx \text{ soit } \alpha = \frac{^t xAx}{^t xx}.$$

□

Notons que si x est un vecteur propre de A alors $R_A(x)$ est exactement sa valeur propre associée. Le théorème suivant est fréquement utilisé dans le contexte de la statistique, notamment dans l'analyse en composantes principales.

Théorème 2.9 (Extremums du quotient de Rayleigh) *Soit A une matrice réelle symétrique de taille n. Alors le quotient de Rayleigh satisfait:*

1. $\max_{x \neq 0} \frac{^t xAx}{^t xx} = \lambda_1$, *$\lambda_1$ étant la plus grande valeur propre de A et le maximum est atteint en tout x_1 vecteur propre de A associé à λ_1.*
2. $\min_{x \neq 0} \frac{^t xAx}{^t xx} = \lambda_n$, *$\lambda_n$ étant la plus petite valeur propre de A et le minimum est atteint en tout x_n vecteur propre de A associé à λ_n.*

Preuve. Soit $A = PD\,{}^tP$ la décomposition spectrale de A telle que $D = diag(\lambda_1, ..., \lambda_n)$ et $\lambda_1 \geq ... \geq \lambda_n$ sont les valeurs propres de A. On a donc pour $x \in \mathbb{R}^n \backslash \{0\}$,

$$\begin{aligned} ^t xAx &= {}^t xPD\,{}^tPx = {}^t yDy \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 ... + \lambda_n y_n^2 \\ &\leq \lambda_1 (y_1^2 + y_2^2 ... + y_n^2) = \lambda_1\,{}^t yy \end{aligned}$$

où l'on a effectué le changement de variable $y = {}^tPx = {}^t(y_1, y_2..., y_n)$. Comme P est orthogonale, ${}^tyy = {}^txx$ et donc $\frac{{}^txAx}{{}^txx} \leq \lambda_1$. Clairement, cette dernière inégalité devient une égalité si x est un vecteur propre associé à λ_1. On démontre 2. par la même démarche. □

Exercice 36 (Extremums du quotient de Rayleigh - approche analytique) *Soient A une matrice carrée de taille n et $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ définie par $f(x) = {}^txAx$.*

1. *Montrer que f admet un maximum et un minimum sur la sphère $S = \{x \in \mathbb{R}^n : \|x\| = 1$*
2. *En déduire que le quotient de Rayleigh $R_A(x)$, $x \in \mathbb{R}^n \setminus \{0\}$, admet un maximum λ_1 atteint en x_1 et un minimum λ_n atteint en x_n.*
3. *Supposons que A est symétrique. Calculer $\nabla R_A(x)$, pour $x \in \mathbb{R}^n \setminus \{0\}$, et en déduire que λ_1 et λ_n sont respectivement la plus grande et la plus petite valeur propre de A et que x_1 et x_n en sont respectivement des vecteurs propres associés.*

2.3 Matrices (semi) définies positives

Définition 2.10 *Soit A une matrice symétrique de taille n.*

1. *A est dite semi-définie positive si ${}^txAx \geq 0$ pour tout $x \in \mathbb{R}^n$.*
2. *A est dite définie positive si ${}^txAx > 0$ pour tout $x \in \mathbb{R}^n \setminus \{0\}$.*

Exemple 2.11 1. *$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$: pour $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, on a*

$$ {}^txAx = [x_1, x_2] \begin{bmatrix} x_1 + x_2 \\ x_1 + x_2 \end{bmatrix} = (x_1 + x_2)^2 \geq 0. $$

Donc A est semi-définie positive sans être définie positive.

2. *$A = I_n$: pour tout $x \in \mathbb{R}^n \setminus \{0\}$ on a*

$$ {}^txI_nx = {}^txx = \|x\| > 0. $$

D'où I_n est définie positive.

Sur $\mathbb{S}_n(\mathbb{R})$ l'espace vectoriel des matrices réelles symétrique de taille n, l'ensemble des matrices semi-définies positives, noté $\mathbb{S}_n^+(\mathbb{R})$, définit un cône (convexe), i.e.

1. si A est semi-définie positive et $\lambda \geq 0$ alors λA l'est;
2. si A et B sont semi-définies positives alors $A + B$ l'est;
3. si A et $-A$ sont semi-définies positives alors $A = 0$.

Définition 2.12 (Ordre de Löwner) *L'ordre de Löwner sur l'espace vectoriel $\mathbb{S}_n(\mathbb{R})$ est la relation d'ordre partiel (notée abusivement par $\geq$) induite par le cône $\mathbb{S}_n^+(\mathbb{R})$, i.e.,*

$$A \geq B \quad \text{si} \quad A - B \text{ est semi-définie positive.}$$

Les éléments positifs pour cet ordre sont donc exactement les matrices semi-définies positives. Cet ordre rend $\mathbb{S}_n(\mathbb{R})$ un espace vectoriel ordonné, i.e., l'ordre $\geq$ est compatible avec sa structure vectorielle:

1. si $A \geq B$ alors $A + C \geq B + C$;
2. si $A \geq 0$ et $\lambda \geq 0$ alors $\lambda A \geq 0$.

Cet ordre est reconnue dans la litérature pour l'ordre de Löwner (le mathématicien qui l'introduisait en 1934). Il est à noter qu'en général l'espace vectoriel $\mathcal{M}_{m,n}(\mathbb{R})$ peut être ordonné autrement via le cône des matrices à coefficients positifs (l'order standard sur l'espace vectoriel des opérateurs linéaires entre espaces vectoriels ordonnés). Cependant, l'ordre de Löwner demeure le plus naturel et pertinent en analyse matricielle du fait qu'il permet l'extension de l'étude des fonctions réelles aux fonctions de matrices (puissance d'une matrice, exponentielle d'une matrice,...).

Pour les matrices diagonales, on a aisément la proposition suivante:

Proposition 2.13 *Soit la matrice diagonale $D = diag(d_1, ..., d_n)$. Alors*

1. D est semi-définie positive si et seulement si $d_i \geq 0$ pour tout i.

2. *D est définie positive si et seulement si $d_i > 0$ pour tout i.*

Théorème 2.14 *Soit A une matrice symétrique de taille n.*

1. *A est semi-définie positive si et seulement si toutes les valeurs propres λ_i de A sont positives.*
2. *A est définie positive si et seulement si toutes les valeurs propres λ_i de A sont strictement positives.*

Preuve.

1. Soit $A = PD\,{}^tP$, P étant une matrice orthogonale de taille n et $D = diag\,(\lambda_1, ..., \lambda_n)$. Si $\lambda_i \geq 0$ pour tout i, alors pour $x \in \mathbb{R}^n$ on a
$$ {}^txAx = {}^txPD\,{}^tPx = {}^tyDy = \sum_{i=1}^{n} \lambda_i y_i^2 \geq 0, $$
où $y = {}^tPx = {}^t(y_1, ..., y_n)$. D'où A est semi-définie positive. Inversement, si A est semi-définie positive alors pour une valeur propre λ de A, soit u un vecteur propre de A associé. On a donc $\lambda = \frac{{}^tuAu}{\|u\|^2} \geq 0$.
2. Par des arguments similaires.

$\square$

Corollaire 2.15 *Soit A une matrice symétrique de taille n. Alors*

1. *si A est définie positive alors $|A| > 0$ (et donc A est inversible) et A^{-1} est aussi définie positive.*
2. *si A est semi-définie positive et inversible alors A est définie positive.*

Preuve.

1. $|A| = \prod_{i=1}^{n} \lambda_i > 0$. Etant les inverses des valeurs propres de A, les valeurs propres de A^{-1} sont de même strictement positives. D'où A^{-1} est définie positive.

2. De $|A| = \prod_{i=1}^{n} \lambda_i$ il résulte que toutes les valeurs propres λ_i de A sont non nulles, donc strictement positives (elles sont déjà positives). Par suite, A est définie positive.

$\square$

L'exemple typique des matrices semi-définies positives est celui des matrices sous la forme ${}^tAA$ (A étant de taille (m, n)) intervenant dans les modèles de régression linéaire.

Proposition 2.16 *Pour toute matrice A de taille (m, n) on a*

1. *${}^tAA$ est semi-définie positive;*
2. *si A est de rang plein alors ${}^tAA$ est définie positive.*

Preuve.

1. Pour $x \in \mathbb{R}^n$, on a

$$ {}^tx\, {}^tAAx = {}^t(Ax)\, Ax = \|Ax\|^2 \geq 0. $$

2. Comme A est de rang plein alors ${}^tAA$ est inversible et donc définie positive.

$\square$

Théorème 2.17 *Soit A une matrice symétrique de taille n. Alors*

1. *A est semi-définie positive si et seulement s'il existe une matrice carrée B de taille n telle que $A = {}^tBB$;*
2. *A est définie positive si et seulement s'il existe une matrice carrée inversible B de taille n telle que $A = {}^tBB$.*

Preuve.

1. Soit $A = PD\ {}^tP$, P étant une matrice orthogonale de taille n et $D = diag\left(\lambda_1, ..., \lambda_n\right)$, les λ_i étant les valeurs propres positives de A. Considérons la matrice $D^{\frac{1}{2}} = diag\left(\lambda_1^{\frac{1}{2}}, ..., \lambda_n^{\frac{1}{2}}\right)$. On a donc
$$A = PD^{\frac{1}{2}}D^{\frac{1}{2}}\ {}^tP = {}^tBB,$$
où $B = D^{\frac{1}{2}}\ {}^tP$ est la matrice voulue.
2. Si A est définie positive, $\lambda_i > 0$ pour tout i et donc $D^{\frac{1}{2}}$ est inversible. La matrice $B = D^{\frac{1}{2}}\ {}^tP$ est inversible étant le produit de deux matrices qui le sont. Les sens inverse étant déjà traités.

□

Exercice 37 *Soient A une matrice symétrique de taille n et P une matrice de taille (n, m). Montrer que*

1. *si A est semi-définie positive alors ${}^tPAP$ l'est aussi.*
2. *Si A est définie positive et P est de rang plein, alors ${}^tPAP$ est définie positive.*

Exercice 38 *Soit A une matrice symétrique semi-définie positive de taille n. Montrer que pour toute matrice X de taille (n, m) on a*

$$AX = 0 \textit{ si et seulement si } {}^tXAX = 0.$$

Exercice 39 *Soit A une matrice de taille (m, n). Considérons le problème*

$$\begin{cases} \textit{maximiser } f(x) = \|Ax\|^2, \ x \in \mathbb{R}^n \\ \textit{sous la contrainte } g(x) = {}^txx = 1. \end{cases} \tag{2.4}$$

1. *Soit la fonction $h(x, \lambda) = f(x) - \lambda g(x)$, où λ est le multiplicateur de Lagrange. Calculer $\frac{\partial}{\partial x} h(x, \lambda)$ le gradient de la fonction $x \mapsto h(x, \lambda)$.*
2. *En déduire que la solution du problème (2.4) est la plus grande valeur propre*

σ_1^2 ($\sigma_1 \geq 0$) de ${}^tAA$, et que

$$\|A\|_2 = \sigma_1.$$

3. *Montrer que si A est symétrique, alors*

$$\|A\|_2 = r(A) := \max\{|\lambda_i| : \lambda_i \in \sigma(A)\}.$$

4. *Montrer que si A est une matrice carrée inversible, alors*

$$\left\|A^{-1}\right\|_2 = \frac{1}{\sigma_n},$$

où σ_n^2 ($\sigma_n > 0$) est la plus petite valeur propre de ${}^tAA$.

2.4 Décomposition en valeurs singulières d'une matrice

L'objectif de cette scetion est de démontrer que toute matrice A de taille (m, n) et de rang $r \geq 1$ admet une décomposition

$$A = U\Sigma\, {}^tV, \tag{2.5}$$

U, V étant deux matrices orthogonales de tailles m et n respectivement, et Σ est une matrice " presque diagonale " de taille (m, n) donnée par

$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}, \quad \Sigma_r = \mathrm{diag}\,(\sigma_1, ..., \sigma_r)$$

où $\sigma_1 \geq ... \geq \sigma_r > 0$ sont dits les valeurs singulières de A.

Pour ce faire, on considère la matrice semi-définie positive ${}^tAA$. Sa décomposition spectrale donne une matrice orthogonale $V = [v_1, ..., v_n]$ telle que ${}^tAA = VD\,{}^tV$ avec $D =$diag$(\lambda_1, ..., \lambda_n)$, les λ_i étant les valeurs propres positives de ${}^tAA$, et pour tout i, v_i est un vecteur propre de ${}^tAA$ associé à λ_i tels que $(v_1, ..., v_n)$ est une base orthonormale de $\mathbb{R}^n$. Comme rg(A) =rg$({}^tAA)$ =rg$(D) = r$, ${}^tAA$ a exactement r valeurs propres strictement positives. Sans perdre de généralité, on peut supposer que $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r > \lambda_{r+1} = ... = \lambda_n = 0$.

On aura besoin du lemme suivant:

Lemme 1 *1. La famille* $(v_1, ..., v_r)$ *est une base orthonormale de* $\mathrm{Im}\ {}^tA$;

2. la famille $(v_{r+1}, ..., v_n)$ *est une base orthonormale de* $\mathrm{Ker}A$.

Preuve.

1. Comme ${}^tAAv_i = \lambda_i v_i$ et $\lambda_i \neq 0$ pour tout $i = 1, ..., r$ alors $v_1, ..., v_r \in \mathrm{Im}\ {}^tAA \subset \mathrm{Im}\ {}^tA$. Etant orthonormale, la famille $(v_1, ..., v_r)$ est libre dans $\mathrm{Im}\ {}^tA$. Comme $\dim \mathrm{Im}\ {}^tA = \mathrm{rg}\ {}^tA$ r, alors $(v_1, ..., v_r)$ est une base orthonormale de $\mathrm{Im}\ {}^tA$.
2. Comme la famille $(v_1, ..., v_n)$ est orthonormale, alors

$$Vect\,(v_{r+1}, ..., v_n) \subset (Vect\,(v_1, ..., v_r))^{\perp} = (\mathrm{Im}\ {}^tA)^{\perp} = \mathrm{Ker}A.$$

Comme $\dim \mathrm{Ker}A = n - r$ et la famille $(v_{r+1}, ..., v_n)$ est libre, alors $(v_{r+1}, ..., v_n)$ est une base orthonormale de $\mathrm{Ker}A$.

□

Maintenant, pour tout i, j on a

$${}^t(Av_j)\,Av_i = {}^tv_j\ {}^tAAv_i = \lambda_i\ {}^tv_j v_i = \lambda_i \delta_{ij}.$$

En particulier, pour tout $i = 1, 2, ..., r$, $\|Av_i\| = \sqrt{\lambda_i} > 0$. Soient donc, pour $i = 1, 2, ..., r$, $u_i = \frac{Av_i}{\|Av_i\|}$ et $\sigma_i = \sqrt{\lambda_i}$. On a donc les équations:

$$Av_i = \sigma_i u_i \quad (\text{ou son alternative } {}^tAu_i = \sigma_i v_i), \quad i = 1, 2, ..., r \tag{2.6}$$

qui s'écrivent matriciellement

$$AV_r = U_r \Sigma_r \tag{2.7}$$

où $V_r = [v_1, ..., v_r]$ est la trancature de la matrice V jusqu'à l'order r, $U_r = [u_1, ..., u_r]$ et $\Sigma_r = \mathrm{diag}(\sigma_1, ..., \sigma_r)$. A ce stade, on a donc abouti pour la matrice réctangulaire A à une situation que l'on peut comparer au théorème spectral pour les matrices réelles symétriques, où les équations (2.1) et (2.2) sont remplacées par (2.6) et (2.7) respectivement. L'insuffisance ici est que les vecteurs colonnes des deux matrices U_r et V_r forment respectivement des bases orthonormales de $\mathrm{Im}A$ et $\mathrm{Im}\ {}^tA$ seulement. Pour compléter la tâche, on doit fournir des bases orthonormales des espaces entiers $\mathbb{R}^n$ et $\mathbb{R}^m$. Il ne reste donc qu'à compléter la famille orthonormale $(u_i)_{i=1}^r$ en une base orthonormale $(u_i)_{i=1}^m$ de $\mathbb{R}^m$. D'où, si l'on considère la matrice $U = [u_1, ..., u_m]$, on a

$$\begin{aligned} AV &= [Av_1, ..., Av_r, Av_{r+1}, ..., Av_n] \\ &= [\sigma_1 u_1, ..., \sigma_r u_r, 0, ..., 0] \\ &= U \begin{bmatrix} \sigma_1 & & 0 & \\ & \ddots & & 0_{r,n-r} \\ 0 & & \sigma_r & \\ & 0_{m-r,r} & & 0_{m-r,n-r} \end{bmatrix} \\ &= U\Sigma \end{aligned}$$

où $\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}$. Il s'ensuit qu'on a factorisé la matrice A selon la décomposition (2.5).

Notons que, par les mêmes arguments de la preuve du Lemme 1, on a

1. la famille $(u_1, ..., u_r)$ est une base orthonormale de $\mathrm{Im}A$;
2. la famille $(u_{r+1}, ..., u_m)$ est une base orthonormale de $\mathrm{Ker}\ {}^tA$.

On a donc établi le résultat suivant:

Théorème 2.18 (Décomposition en valeurs singulières) *Toute matrice A de taille (m, n) et de rang $r \geq 1$ admet une décomposition $A = U\Sigma\ {}^tV$, dite décomposition en valeurs singulières, où $U = [u_1, ..., u_m]$, $V = [v_1, ..., v_n]$ sont deux matrices orthogonales de tailles m et n respectivement, $\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}$ est* *une matrice de taille* (m,n) *, et* $\Sigma_r = diag(\sigma_1, ..., \sigma_r)$*,* $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$ *sont les racines carrées des* r *valeurs propres strictement positives de* ${}^tAA$ *et sont dites les valeurs singulières de* A*.*

Les vecteurs $v_1, ..., v_n$ *sont dits des vecteurs singuliers à droite de* A *et les vecteurs* $u_1, ..., u_m$ *sont dits des vecteurs singuliers à gauche de* A*, et sont reliés par les équations*

$$Av_i = \sigma_i u_i \quad , \quad {}^tAu_i = \sigma_i v_i, \quad i = 1, 2, ..., r.$$

De plus, on a

- *les* r *premiers vecteurs* $v_1, ..., v_r$ *forment une base orthonormale de* $\mathrm{Im}\,{}^tA$.
- *les* $n-r$ *derniers vecteurs* $v_{r+1}, ..., v_n$ *forment une base orthonormale de* $\mathrm{Ker}A$.
- *les* r *premiers vecteurs* $u_1, ..., u_r$ *forment une base orthonormale de* $\mathrm{Im}A$.
- *les* $m-r$ *derniers vecteurs* $u_{r+1}, ..., u_m$ *forment une base orthonormale de* $\mathrm{Ker}\,{}^tA$.

Géométriquement, la DVS réalise pour une matrice réctangulaire ce que la décomposition spectral le fait pour une matrice réelle symétrique. Plus précisement, si l'on considère la transformation vectorielle $A : \mathbb{R}^n \to \mathbb{R}^m$, $x \mapsto Ax$, alors la DVS fournit des bases orthonormales $\mathcal{B}_1 = (v_1, ..., v_n)$ et $\mathcal{B}_2 = (u_1, ..., u_m)$ de $\mathbb{R}^n$ et $\mathbb{R}^m$ respectivement, telles que la matrice de $A : \mathbb{R}^n \to \mathbb{R}^m$ relativement à ces deux bases soit la matrice " presque diagonale " Σ. Ainsi, si l'on exprime les vecteurs dans $\mathcal{B}_1$ et $\mathcal{B}_2$, i.e., si $x = VX$ et $Ax = UY$, X et Y étant les deux vecteurs des coordonnées, alors

$$Y = {}^tUAVX = \Sigma X = {}^t(\sigma_1 X_1, ..., \sigma_r X_r, 0, ..., 0)$$

où $X = {}^t(X_1, ..., X_n)$. Le vecteur des coordonnées de l'image Ax est obtenu donc par une simple multiplication des r premiers coordonnées de x par des scalaires qui sont les valeurs singulières de A. Les coordonnées restantes de Ax dans la base

$\mathcal{B}_2$ étant toutes nulles car $(u_{r+1}, ..., u_m)$ est une base de Ker ${}^tA = (ImA)^{\perp}$.

Exemple 2.19 $A = \begin{bmatrix} 1 & 1 \\ \sqrt{3} & 0 \\ 0 & \sqrt{3} \end{bmatrix}$:

On a ${}^tAA = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$ *et* $A\,{}^tA = \begin{bmatrix} 2 & \sqrt{3} & \sqrt{3} \\ \sqrt{3} & 3 & 0 \\ \sqrt{3} & 0 & 3 \end{bmatrix}$. *De plus,*

$$\chi_{\,{}^tAA}(X) = (X-4)^2 - 1 = (X-5)(X-3).$$

D'où les valeurs propres de ${}^tAA$ *sont* $5, 3$. *La solution des deux systèmes linéaires*

$$\begin{bmatrix} 4-\lambda & 1 \\ 1 & 4-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \lambda \in \{5, 3\}$$

donne deux vecteurs propres unitaires de ${}^tAA$ *qui sont deux vecteurs singuliers à droite de* A :

$$\begin{aligned} v_1 &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \lambda = 5, \\ v_2 &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \lambda = 3. \end{aligned}$$

Deux vecteurs singuliers à gauche de A *sont donc donnés par:*

$$\begin{aligned} u_1 &= \frac{Av_1}{\|Av_1\|} = \frac{1}{\sqrt{10}} \begin{bmatrix} 2 \\ \sqrt{3} \\ \sqrt{3} \end{bmatrix}, \\ u_2 &= \frac{Av_2}{\|Av_2\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}. \end{aligned}$$

Un troisième vecteur singuliers à gauche de A est un vecteur propre unitaire de la matrice $A\,{}^tA$ associé à sa troisième valeur propre qui est 0 (notons que ses deux autres valeurs propres sont les mêmes que celles de ${}^tAA$; Exercice 33). De tel vecteur u_3 est donné par la résolution du système linéaire:

$$\begin{bmatrix} 2 & \sqrt{3} & \sqrt{3} \\ \sqrt{3} & 3 & 0 \\ \sqrt{3} & 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

soit $u_3 = \frac{1}{\sqrt{5}} \begin{bmatrix} -\sqrt{3} \\ 1 \\ 1 \end{bmatrix}$. *Une DVS de A est donc*

$$A = \begin{bmatrix} \frac{2}{\sqrt{10}} & 0 & -\frac{\sqrt{3}}{\sqrt{5}} \\ \frac{\sqrt{3}}{\sqrt{10}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{5}} \\ \frac{\sqrt{3}}{\sqrt{10}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

2.4.1 Remarques sur la DVS d'une matrice

1. Si pour les matrices U, V de la DVS (2.5), on pose $U = [U_1, U_2]$, $V = [V_1, V_2]$, avec $U_1 = [u_1, ..., u_r]$, $U_2 = [u_{r+1}, ..., u_m]$, $V_1 = [v_1, ..., v_r]$, et $V_2 = [v_{r+1}, ..., v_n]$, alors la DVS (2.5) s'écrit

$$\begin{aligned} A &= U\Sigma\,{}^tV = [U_1, U_2] \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} {}^tV_1 \\ {}^tV_2 \end{bmatrix} \\ &= U_1 \Sigma_r\,{}^tV_1 = [u_1, ..., u_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} {}^tv_1 \\ \vdots \\ {}^tv_r \end{bmatrix} \\ &= \sum_{i=1}^{r} \sigma_i u_i\,{}^tv_i. \end{aligned}$$

Cela montre qu'une DVS de la matrice A est une somme de r matrices de rang 1 obtenues par le produit tensoriel de chacun des r premiers vecteurs singuliers à gauche de A et du vecteur singulier à droite de A du même ordre.

2. D'après ce qui précéde, dans une DVS de A, les vecteurs singuliers à droite de A sont des vecteurs propres de ${}^tAA$ formant une base orthonormale de $\mathbb{R}^n$. Montrons de même que les vecteurs singuliers à gauche de A sont des vecteurs propres de $A\,{}^tA$ formant une base orthonormale de $\mathbb{R}^m$: on a

$$A\,{}^tA = U\Sigma\,{}^tVV\,{}^t\Sigma\,{}^tU = U\Sigma\,{}^t\Sigma\,{}^tU = U\Delta\,{}^tU, \tag{2.8}$$

où l'on a la matrice diagonale de taille m :

$$\Delta = \Sigma\,{}^t\Sigma = \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{bmatrix}.$$

Il s'ensuit que (2.8) est une décomposition spectrale de $A\,{}^tA$ et les vecteurs singuliers à gauche de A, $u_1, ..., u_m$, sont bien des vecteurs propres de $A\,{}^tA$ associés aux valeurs propres $\lambda_1, ..., \lambda_r, 0, ..., 0$.

3. La décomposition en valeurs singulières d'une matrice fournit des bases orthonormales de ses quatre sous espaces vectoriels fondamentaux. Il s'agit donc d'un raffinement du théorème fondamental de l'algèbre linéaire.

4. Le cas particulier où A est une matrice carrée: alors les matrices U, V dans la DVS (2.5) sont de même taille n et la matrice $\Sigma = diag\,(\sigma_1, ..., \sigma_r, 0, ..., 0)$ est diagonale de taille n. La DVS est donc une alternative de la décomposition spectrale pour les matrices carrées non nécesserement symétriques.

   Le cas où A est symétrique est traité dans l'exercice suivant:

Exercice 40 *Soient A une matrice symétrique de taille n et de rang $r \geq 1$, $A = PD\,{}^tP$ une décomposition spectrale avec P est une matrice orthogonale de taille n et $D = diag\,(\lambda_1, ..., \lambda_n)$, les λ_i étant les valeurs propres de A arrangées de telle façon que $|\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_r| > \lambda_{r+1} = ... = \lambda_n = 0$.*

1. Donner une décomposition en valeurs singulières $A = U\Sigma\,{}^tV$, où l'on déterminera les matrices U, V en fonction de P et d'une matrice Δ à déterminer, et Σ en fonction des λ_i. (ind. comme les valeurs singulières de A sont strictement positives, on multipliera D par une matrice convenable Δ pour avoir Σ).
2. En déduire les valeurs singulières de A en fonction des λ_i.
3. Sous quelle condition supplémentaire sur A, ses deux décompositions spéctrale et en valeurs singulières ci-dessus sont les mêmes?

Exercice 41 *Montrer que la norme d'une matrice A (non nulle) de taille (m, n) est donnée par*

$$\|A\| = (\sigma_1^2 + ... + \sigma_r^2)^{\frac{1}{2}}$$

où les σ_i sont les valeurs singulières de A et $r = \mathrm{rg}A$.

Exercice 42 (Rattrapage, automne 2017) *Soit $A = U\Sigma\,{}^tV$ une décomposition en valeurs singulières d'une matrice A de taille (m, n) telles que $U = [u_1, u_2, ..., u_m]$ et $V = [v_1, v_2, ..., v_n]$. Montrer que la famille $\mathcal{F} = (u_i{}^tv_j)_{1\leq i\leq m,\ 1\leq j\leq n}$ est une base orthonormale de $\mathcal{M}_{m,n}(\mathbb{R})$.*

Exercice 43 (Rattrapage, automne 2017) *Soit A une matrice de taille (n, m) $(m \leq n)$. Montrer que le quotient $\frac{{}^txAy}{\|x\|\|y\|}$, $x \in \mathbb{R}^n \backslash \{0\}$ et $y \in \mathbb{R}^m \backslash \{0\}$, admet un maximum et déterminer deux vecteurs x et y pour lesquels ce maximum est atteint.*

Exercice 44 (Rattrapage, automne 2018) *On se propose dans cet exercice de montrer le résultat de minimisation suivant, utile en analyse factorielle: pour deux* *matrices réelles* A *et* B *de taille* (n,p), *on a*

$$\min_{X\in\mathcal{O}_p} \|A-BX\|^2 = \|A-BX_*\|^2$$

où $\mathcal{O}_p$ *est l'ensemble des matrices orthogonales de taille* p *et* $X_* = U\,{}^tV$, U *et* V *étant obtenues de la décomposition en valeurs singulières* ${}^tBA = U\Sigma\,{}^tV$.

1. Montrer que pour tout $X\in\mathcal{O}_p$, on a

$$\|A-BX\|^2 = tr(A\,{}^tA) + tr(B\,{}^tB) - 2tr(X\,{}^t({}^tBA)).$$

2. Montrer que pour tout $X\in\mathcal{O}_p$, $tr(X\,{}^t({}^tBA)) \le tr(\Sigma)$ et conclure (ind. les coefficients d'une matrice orthogonale sont majorés par 1).

Bibliographie

[1] S. Banerjee et A. Roy, Linear Algebra and Matrix Analysis for Statistics, CRC Press, (2014).

[2] D. A. Harville, Matrix Algebra From a Statistician's Perspective, Springer-Science, BusinessMedia, LLC, (2008).

[3] C.D. Meyer, Matrix Analysis and Applied Linear Algebra, Siam, (2000).

[4] J. M. Monier, Mathématiques Méthodes et Exercices MP, Dunod, (2009).

[5] S. Puntanen, G. P. H. Styan J. Isotalo, Matrix Tricks for Linear Statistical Models, Springer-Verlag Berlin Heidelberg, (2011).

[6] G. Strang, Introduction to Linear Algebra, Wellesley- Cambridge Press, (2009).