Eléments de Calcul Matriciel et Différentiel pour l'Analyse Factorielle de Données.

Jean-François Durand

jfd@helios.ensam.inra.fr

Université Montpellier II

Contents

Chapter 1

Matrices, Définitions et Propriétés

Ce chapitre n'est pas un cours de calcul matriciel mais a pour ojectif d'une part de rappeler de façon succinte les principales définitions et propriétés étudiées au cours du premier cycle et d'autre part, d'en introduire de nouvelles utilisées en statistique et en calcul différentiel matriciel. On supposera que les notions de $I\!\!K$-espace vectoriel ($I\!\!K$ est $I\!\!R$ ou $\mathbb{C}$) et d'applications linéaires sont connues. De façon commode en dimension finie, lorsqu'une base de l'espace vectoriel est spécifiée, l'expression d'un vecteur dans la base est assimilée à une matrice colonne. La base choisie par défaut est la base canonique.

1.1 Notations et premières définitions

Une matrice $m \times n$, A, est un tableau d'éléments de $I\!\!K$, tel que

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}.$$

L'élément courant a_{ij}, $i$ème ligne, $j$ième colonne, sera parfois noté a_i^j. La somme de deux matrices de même dimensions est définie par

$$A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}].$$

Le produit d'une matrice par un scalaire $\lambda \in I\!\!K$, est défini par

$$\lambda A = A\lambda = [\lambda a_{ij}].$$

Ces deux opérations confèrent à l'ensemble $\mathcal{M}_{mn}(I\!\!K)$ des matrices de m lignes et n colonnes sur $I\!\!K$, une structure d'espace vectoriel de dimension mn qui sera aussi noté

$I\!\!K^{m\times n}$. On note $\{E_{ij}(m,n)\}_{i,j}$ la base canonique de $I\!\!K^{m\times n}$, définie par

$$E_{ij}(m,n) = E_{ij} = \begin{bmatrix} 0 & . & . & . & . & . & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & . & . & . & . & . & 0 \end{bmatrix},$$

matrice dont le seul élément non nul est l'élément ij qui vaut 1.

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{i\,j}E_{ij}(m,n).$$

Outre les opérations liées à la structure d'espace vectoriel, il existe une autre opération appelée produit matriciel. Soient $A = [a_{ij}]$, $m \times n$, et $B = [b_{ij}]$, $n \times p$, alors le produit AB est la matrice $m \times p$ définie par

$$C = AB = \left[c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}\right].$$

Soit $A = [a_{ij}]$, une matrice $m \times n$.

La matrice identité d'ordre m, $I_m = \sum_{i=1}^{m} E_{ii}(m,m)$. Alors, $A = I_m A = A I_n$.

La matrice A', transposée de A, est la matrice $n \times m$, $A' = [a'_{ij} = a_{ji}]$.

La base canonique pour $\mathcal{M}_{m1}(I\!\!K) = I\!\!K^{m\times 1}$ identifié à $I\!\!K^m$, est notée

$$\{e_i(m) = E_{i1}(m,1)\}_{i=1,\ldots,m}\ ,$$

ou $\{e_i\}_i$ lorsqu'il n'y aura pas d'ambiguté sur la dimension. On a les propriétés :

$*$ $E_{ij}(m,n) = e_i(m)e'_j(n)$.

$*$ Assimilant $I\!\!R$ et $I\!\!R^{1\times 1}$, le symbole de Kronecker, $\delta_{ij} = e'_i(n)e_j(n) = \begin{cases} 1 & si\ \ i = j \\ 0 & si\ \ i \neq j \end{cases}$

$*$ $E_{ij}e_r =$ colonne r de $E_{ij} = \delta_{jr}e_i$

$e'_r E_{ij} =$ ligne r de $E_{ij} = \delta_{ir}e'_j$.

$*$ Soient A_i la ligne i et A^j la colonne j de A.

$$a_{ij} = e'_i(m)Ae_j(n).$$

$$A^j = Ae_j(n) = \sum_{i=1}^{m} a_{ij}e_i(m) \quad \text{et} \quad A = \sum_{j=1}^{n} A^j e_j(n)'.$$

$$A_i = e'_i(m)A = \sum_{j=1}^{n} a_{ij}e'_j(n) \quad \text{et} \quad A = \sum_{i=1}^{m} e_i(m)A_i.$$

1.2 Matrice associée à une application linéaire

Rappelons la définition d'une application linéaire g du $\mathbb{K}$-espace vectoriel E dans le $\mathbb{K}$-espace vectoriel F

$$g(\lambda x + \mu y) = \lambda g(x) + \mu g(y), \qquad \forall (\lambda, \mu) \in \mathbb{K}^2, (x, y) \in E^2.$$

Soient E et F deux $\mathbb{K}$-espaces vectoriels de <u>dimensions finies</u> ($E = \mathbb{K}^n$ et $F = \mathbb{K}^m$), $\{e_1, \ldots, e_n\}$ une base de E et $\{f_1, \ldots, f_m\}$ une base de F. Soit enfin g une application linéaire de E dans F (on dira $g \in \mathcal{L}(E, F)$), telle que $x \longrightarrow y = g(x)$.

On appelle matrice de g relativement aux deux bases, la matrice $A = [a_{ij}]$ de dimensions $m \times n$ définie par colonnes

$$\begin{array}{cc} & \begin{array}{ccc} g(e_1) & \ldots & g(e_n) \end{array} \\ A = & \left[\begin{array}{ccc} A^1 & \ldots & A^n \end{array} \right] \end{array}$$

La $j$ième colonne A^j contient les coordonnées de $g(e_j)$ exprimées dans la base $\{f_1, \ldots, f_m\}$. Plus précisément, $g(e_j) = \sum_{i=1}^m a_{ij} f_i$.

Alors, la matrice colonne $Y = [y_i]$ des coordonnées de y dans la base $\{f_i\}$, s'exprime en fonction de la matrice colonne $X = [x_j]$ des coordonnées de x dans la base $\{e_j\}$, grâce au produit matriciel $Y = AX$ où $y_i = \sum_{j=1}^m a_{ij} x_j$.

En outre, soit A une matrice de $\mathcal{M}_{mn}(\mathbb{K})$. Alors A est la matrice d'une unique application linéaire g_A de $\mathbb{K}^n$ dans $\mathbb{K}^m$ relativement aux bases canoniques de ces espaces. On dit que g_A est l'application linéaire canoniquement associée à A.

1.3 Quelques matrices particulières

1.3.1 Matrice adjointe d'une matrice A

Dans ce cours, les nombres et matrices complexes sont seulement liés aux valeurs propres et vecteurs propres de matrices carrées réelles non-symétriques. Aussi, un traitement détaillé des éléments de $\mathbb{C}^{m \times n}$ est omis.

Soit $A = [a_{ij}] \in \mathbb{C}^{m \times n}$, $A^* = [a^*_{ij}]$, adjointe de A, est l'élément de $\mathbb{C}^{n \times m}$ tel que $a^*_{ij} = \overline{a}_{ji}$.

$$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} = X + iY \qquad A^* = \begin{bmatrix} \overline{a}_{11} & \ldots & \overline{a}_{m1} \\ \vdots & & \vdots \\ \overline{a}_{1n} & \ldots & \overline{a}_{mn} \end{bmatrix} = X' - iY'$$

Remarquons que $A^* = A'$ si et seulement si A est réelle. On a les propriétés suivantes

P1 : $(A^*)^* = A$,

P2 : $(A+B)^* = A^* + B^*$,

P3 : $(\lambda A)^* = \overline{\lambda} A^*$,

P4 : $(AB)^* = B^* A^*$.

1.3.2 Matrices hermitiennes

On dit que A est hermitienne (respectivement symétrique) si $A^* = A$ (respectivement $A' = A$).

Bien sûr, des matrices hermitiennes (resp. symétriques) sont carrées. Remarquons que le produit de deux matrices hermitiennes n'est pas en général une matrice hermitienne. La propriété suivante sera démontrée au chapitre suivant

P0 : Une matrice hermitienne A a toutes ses valeurs propres réelles et il existe une base orthonormée de $\mathbb{C}^n$ formée par les vecteurs propres de A.
Notons

$$\Lambda = diag(\lambda_1, \ldots, \lambda_n), \qquad \lambda_i \in \mathbb{R}$$

la matrice diagonale des valeurs propres de A hermitienne et $V = [v^1, \ldots, v^n]$ la matrice dont les colonnes sont les vecteurs propres correspondants.

$$\forall i, \quad Av^i = \lambda_i v^i \quad \Longleftrightarrow \quad AV = V\Lambda$$

$$\forall i, j \quad {v^i}^* v^j = \delta_{ij} \quad \Longleftrightarrow \quad V^*V = I_n\,,$$

où I_n est la matrice identité d'ordre n. Si A est symétrique réelle, les vecteurs propres sont réels et forment une base orthonormée de $\mathbb{R}^n$.

1.4 Image, noyau, rang d'une matrice

On définit l'image, le rang et le noyau de la matrice A comme étant respectivement l'image, le rang et le noyau de l'application linéaire g_A canoniquement associée à A. Dans le chapitre 2, nous verrons que la décomposition en valeurs singulières d'une matrice est un outil efficace permettant de déterminer le rang de la matrice et de construire des bases orthogonales pour l'image et pour le noyau lorsque celui-ci n'est pas réduit à $\{0\}$.

1.4.1 Image d'une matrice

L'image de $A \in I\!K^{m \times n}$, notée $Im\,A$, est le sous espace vectoriel de $I\!K^m$ engendré par les colonnes de A.

$$Im\,A = \mathcal{E}\{A^1, \ldots, A^n\} = \{y \in I\!K^m \,|\, \exists x \in I\!K^n, \quad Ax = y\}.$$

C'est l'ensemble des combinaisons linéaires, $y = A^1 x_1 + \ldots + A^n x_n$, des colonnes de A. La $i$ème coordonnée, x_i, de x s'appelle le poids de A^i dans l'expression de y.

Systèmes d'équations linéaires compatibles

Soit $A \in I\!R^{m \times n}$ la matrice des coefficients du système et soit $y \in I\!R^m$ le vecteur second membre. Résoudre le système linéaire défini par le couple (A, y), c'est chercher une solution x telle que $Ax = y$.
On dit que le système linéaire $Ax = y$ est compatible s'il admet au moins une solution c'est à dire si $y \in Im\,A$.

Tous les systèmes linéaires seront supposés réels, car un système sur $C\!\!\!\!I$, se décompose en deux sytèmes sur $I\!R$ l'un correspondant à la partie réelle et l'autre à la partie imaginaire des nombres complexes.
On verra au chapitre 4 comment résoudre un système compatible grâce à la notion d'inverse généralisé d'une matrice. Un système qui n'est pas compatible est dit impossible. La résolution d'un problème "aux moindres carrés" étudié au chapitre 4, permet, entre autres résultats, de décider si un système linéaire est impossible ou compatible.

Un système linéaire défini par $(A, 0)$ est dit homogène. Un système linéaire homogène est toujours compatible.
L'ensemble des solutions d'un système homogène $Ax = 0$, s'appelle le noyau de A.

1.4.2 Noyau d'une matrice

Le noyau de $A \in I\!K^{m \times n}$ est le s.e.v. de $I\!K^n$ dont l'image est le zéro de $I\!K^m$

$$Ker\,A = \{x \in I\!K^n \,|\, Ax = 0_m\}.$$

On a les propriétés suivantes,

P1 : $Ker\,A$ est non-vide car $0_n \in Ker\,A$.

P2 : g_A est injective ssi $Ker\, A = \{0_n\}$.

P3 : $\{Im\, A\}^\perp = Ker\, A^*$ et $\{Ker\, A\}^\perp = Im\, A^*$ (clef chapitre 2).

1.4.3 Rang d'une matrice

Le rang de $A \in I\!K^{m \times n}$, noté $rang(A)$, est le nombre de colonnes de A linéairement indépendantes,

$$rang(A) = dim\, Im\, A.$$

Outre que $rang(X) \leq min(m, n)$, on a les propriétés suivantes

P0 : Théorème du rang : $n = rang(A) + dim\, Ker\, A$.

P1 : g_A est surjective (injective) ssi $rang(A) = m$ $(rang(A) = n)$.

P2 : $rang(A) = rang(A^*) = rang(A^*A) = rang(AA^*)$.

P3 : $rang(AB) \leq min(rang(A), rang(B))$.

P4 : $rang(A + B) \leq rang(A) + rang(B)$.

P5 : Soient B et C deux matrices carrées inversibles. Alors, $rang(BAC) = rang(A)$.

1.5 Déterminant d'une matrice carrée

σ est une permutation sur $\{1, \ldots, n\}$ si et seulement si σ est une bijection de $\{1, \ldots, n\}$ sur lui-même.

On verra dans la section 1.11.1 les propriétés d'une matrice associée à une permutation.

Le déterminant d'une matrice $n \times n$, $A = [a_{ij}]$ est l'unique nombre noté dét(A)

$$\text{dét}(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)},$$

où S_n est l'ensemble des $n!$ permutations de $\{1, 2, \ldots, n\}$ et $\epsilon(\sigma)$ est la signature de la permutation σ. La signature vaut +1 si σ est une composition d'un nombre pair de permutations élémentaires, -1 si c'est une composition d'un nombre impair de permutations élémentaires (σ est dite élémentaire si $\sigma(i) = j$, $\sigma(j) = i$ et $\sigma(k) = k$ pour $k \neq i, j$).

Cette définition est d'un coût calculatoire très élevé. Elle exprime le déterminant comme une fonction polynomiale des éléments de la matrice. L'expression suivante permet un calcul du déterminant de proche en proche en fonction de déterminants de matrices extraites de A, ce qui dans certains cas peut simplifier le calcul.

Si $A = [a] \in \mathbb{C}^{1\times 1}$, alors $\text{dét}(A) = a$.
Si $A \in \mathbb{C}^{n\times n}$, développement du déterminant selon la colonne j

$$\text{dét}(A) = \sum_{i=1}^{n} c_{ij} a_{ij} \text{ pour } j \in \{1, \dots, n\},$$

où c_{ij} est le cofacteur de l'élément a_{ij} de A, c'est à dire $c_{ij} = (-1)^{i+j}\text{dét}(A_{ij})$. La matrice A_{ij} est la matrice extraite de A en suprimant la ligne i et la colonne j.

P0 : Le déterminant d'une matice carrée $A = [A^1 \dots A^n]$ est une forme multilinaire alternée des colonnes de A, c'est à dire, $(A^1, \dots, A^n) \longrightarrow D(A^1, \dots, A^n) = \text{dét}(A)$ avec

$$D(A^1, \dots, \lambda A^j + \mu V^j, \dots, A^n) = \lambda D(A^1, \dots, A^j, \dots, A^n) + \mu D(A^1, \dots, V^j, \dots, A^n)$$

$$D(A^1, \dots, A^i, \dots, A^j, \dots, A^n) = -D(A^1, \dots, A^j, \dots, A^i, \dots, A^n).$$

Conséquences :

* $D(A^1, \dots, \lambda A^j, \dots, A^n) = \lambda D(A^1, \dots, A^j, \dots, A^n)$.

** On ne change pas le déterminant d'une matrice carrée si à une colonne (une ligne) on ajoute une combinaison linéaire des autres colonnes (lignes).

P1 : $\text{dét}(AB) = \text{dét}(A)\text{dét}(B)$.

P2 : $\text{dét}(A^*) = \overline{\text{dét}(A)}$.

P3 : Si A est diagonale ou triangulaire alors $\text{dét}(A)$ est le produit des éléments diagonaux.

1.6 Inverse d'une matrice carrée

On dit que A carrée d'ordre n est inversible ou régulière s'il existe une matrice notée A^{-1} appelée inverse de A, telle que

$$A^{-1}A = AA^{-1} = I_n.$$

Si A^{-1} existe, cette matrice est unique.

P1 : $(AB)^{-1} = B^{-1}A^{-1}$ si A et B sont inversibles.

P2 : si A est inversible, $(A^{-1})^* = (A^*)^{-1}$.

P3 : Soit $C = [c_{ij}]$ la matrice des cofacteurs de A, alors si A est inversible, $A^{-1} = C'/\text{dét}(A)$.

P4 : A carrée est inversible ssi A est de plein rang $rang(A) = n$ c'est à dire $\text{dét}(A) \neq 0$.

P5 : Si A est inversible, $\text{dét}(A^{-1}) = 1/\text{dét}(A)$.

P6 : Systèmes d'équations linéaires de Cramer : Soient $A \in I\!R^{n\times n}$ et $y \in I\!R^n$, alors le système linéaire $Ax = y$ est dit de Cramer si de façon équivalente

(i) : $\text{dét}(A) \neq 0$,

(ii) : $rang(A) = n$,

(iii) : Il existe une unique solution $\hat{x} = A^{-1}y$.

P7 : Caractérisation du rang d'une matrice $n \times p$: Soit $A \in I\!K^{n\times p}$, Il esiste un entier naturel $r \leq \min(n, p)$, possédant les propriétés suivantes :

(i) : on peut extraire de A au moins une sous-matrice carrée régulière d'ordre r;

(ii) : toute sous-matrice carrée d'ordre supérieur à r n'est pas régulière.

Ce nombre r est appelé le rang de A. Ceci à pour conséquence la proptiété **P2** $rang(A) = rang(A^*)$ du rang d'une matrice.

P8 : Si $A = diag(a_1, \ldots, a_n)$ est inversible alors $A^{-1} = diag(1/a_1, \ldots, 1/a_n)$.

1.7 Changements de base

Soit E un espace vectoriel muni d'une base $\mathcal{B} = \{e_1, \ldots, e_n\}$ que l'on qualifie d'ancienne base. On choisit dans E une autre base $\mathcal{B}' = \{e'_1, \ldots, e'_n\}$ appelée nouvelle base. Dans tout le paragraphe 7, "prime" ne signifie pas "transposé" mais indique des objets relatifs à la nouvelle base.

On appelle matrice de passage de la base $\mathcal{B}$ à la base $\mathcal{B}'$, la matrice P carrée d'ordre n, dont les colonnes expriment les coordonnées des vecteurs de $\mathcal{B}'$ dans la base $\mathcal{B}$.
Cette matrice est régulière. La matrice de passage de $\mathcal{B}'$ à $\mathcal{B}$ est la matrice P^{-1}.

1.7.1 Effet sur les coordonnées d'un vecteur

Soit x un élément de E dont les coordonnées dans les deux bases $\mathcal{B}$ et $\mathcal{B}'$ sont stockées dans les matrices colonnes X et X'. Alors

$$X = PX' \quad \text{et} \quad X' = P^{-1}X.$$

1.7.2 Effet sur les éléments d'une matrice

Soient E et F deux $\mathbb{K}$ espaces vectoriels de dimension finie et $g \in \mathcal{L}(E, F)$. Si E et F sont munis chacun d'une base $\mathcal{B} = \{e_1, \ldots, e_n\}$ pour E et $\mathcal{C} = \{f_1, \ldots, f_m\}$ pour F, alors g s'exprime à l'aide d'une matrice $A \in \mathbb{K}^{m\times n}$ dans les deux bases précédentes. Si on change maitenant la base $\mathcal{B}$ en une base $\mathcal{B}' = \{e'_1, \ldots, e'_n\}$ et la base $\mathcal{C}$ en une base $\mathcal{C}' = \{f'_1, \ldots, f'_m\}$, la même application linéaire g va s'exprimer dans ces nouvelles bases selon une autre matrice A'. Qu'elle relation lie A et A' qui appartiennent toutes deux à $\mathbb{K}^{m\times n}$.

Soit P la matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$ et soit Q la matrice de passage de $\mathcal{C}$ à $\mathcal{C}'$, alors

$$A' = Q^{-1}AP.$$

On dit que les matrices A et A' sont équivalentes.

Dans le cas particulier d'un endomorphisme $g \in \mathcal{L}(E, E)$, l'influence du changement de la base $\mathcal{B}$ en la base $\mathcal{B}'$ se traduit par $A' = P^{-1}AP$. On dit alors que les matrices carrées A et A' sont semblables. Si A et A' sont semblables, $\text{dét}(A) = \text{dét}(A')$. S'il existe une base $\mathcal{B}'$ telle que A' soit une matrice diagonale, on dit que la matrice A est diagonalisable c'est à dire semblable à une matrice diagonale. Les conditions pour que A soit diagonalisable sont présentées au chapitre 2.

1.8 Valeurs propres, vecteurs propres

Un vecteur v non nul de $\mathbb{C}^n$ est appelé vecteur propre d'une matrice $A \in \mathbb{C}^{n\times n}$ s'il existe un scalaire $\lambda \in \mathbb{C}$ tel que

$$Av = \lambda v \,. \tag{1}$$

Le scalaire λ est appelé valeur propre de A associée à v.

L'ensemble des n valeurs propres est le spectre de la matrice.

On appelle rayon spectral de A le nombre positif ou nul $\rho(A) = \max_i |\lambda_i|$.

Remarquer que (1) s'écrit

$$v \neq 0 \qquad \text{et} \qquad v \in Ker\,(A - \lambda I_n). \tag{1}$$

Les vecteurs propres n'étant définis qu'à une homothétie près, le plus souvent, on choisit de normer à 1 un vecteur propre et la norme choisie est la norme Euclidienne (chapitre 3)

$$\|v\|_2 = 1\,.$$

Dans ce cas, un vecteur propre est défini au signe près.
Déterminer le sous-espace propre associé à λ c'est à dire le sous-espace de $\mathbb{C}^n$ engendré par l'ensemble des vecteurs propres définis par (1), revient à résoudre le système d'équations linéaires homogène

$$(A - \lambda I_n)v = 0_n.$$

Plus précisément,

$$\begin{aligned}
(a_{11} - \lambda)v_1 + a_{12}v_2 + \ldots + a_{1n}v_n &= 0 \\
a_{21}v_1 + (a_{22} - \lambda)v_2 + \ldots + a_{2n}v_n &= 0 \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots &\;.\quad . \\
a_{n1}v_1 + a_{n2}v_2 + \ldots + (a_{nn} - \lambda)v_n &= 0.
\end{aligned}$$

Pour λ fixé, ce système de n équations à n inconnues possède une solution non nulle si et seulement si il n'est pas de Cramer. Le spectre de A est donc l'ensemble des n solutions, $\{\lambda_1, \ldots, \lambda_n\}$, de l'équation caractéristique

$$\mathcal{P}_A(\lambda) = \text{dét}(A - \lambda I_n) = 0 \tag{2}$$

dont le premier membre est appelé le polynôme caractéristique de A.

Ce polynôme de degré n en λ, s'écrit en fonction des éléments de A (on n'exprime que les coefficients les plus connus)

$$\mathcal{P}_A(\lambda) = (-1)^n\lambda^n + (-1)^{n-1} trace(A)\lambda^{n-1} + \ldots + \text{dét}(A),$$

où la trace de la matrice A est la somme de ses éléments diagonaux $trace(A) = a_{11} + \ldots + a_{nn}$, voir section 1.9.

Une valeur propre de multiplicité ou d'ordre r de A est donc un zéro d'ordre r du polynôme caractéristique.
Si A est réelle, alors, à toute valeur propre complexe, λ_i, est associée sa valeur propre conjuguée, $\overline{\lambda}_i$.

P1 : Deux matrices semblables ont mêmes polynômes caractéristiques. Comme conséquence, les coefficients de ce polynôme ne dépendent pas de A mais de g. En particulier le déterminant et la trace d'une matrice ne dépendent que de g.

P2 : Si (λ, v) est un élément propre de A, alors $A^k v = \lambda^k v$, pour k entier positif.

P3 : Soit $\{\lambda_1, \ldots, \lambda_n\}$ le spectre de A, alors,

$$\text{dét}(A) = \Pi_{i=1}^{n} \lambda_i \; = \lambda_1 \ldots \lambda_n .$$

Comme conséquence, pour qu'une matrice A soit régulière, il faut et il suffit qu'elle n'admette aucune valeur propre nulle.

P4 : Si (λ, v) est un élément propre de A inversible, $A^{-1} v = \lambda^{-1} v$.

P5 : Si (λ, v) est un élément propre de A, $\mathcal{P}(A)v = \mathcal{P}(\lambda)v$ pour tout polynôme $\mathcal{P}$.

P6 : <u>Cayley-Hamilton</u> : $\mathcal{P}_A(A) = 0_{n\times n}$

P7 : <u>Perron</u> : Pour une matrice carrée à éléments positifs, la valeur propre la plus grande en module est également positive. C'est une valeur propre simple et le vecteur propre associé est à coordonnées de même signe (positives).

P8 : Les valeurs propres non-nulles des produits AB et BA sont les mêmes. Plus précisément, soient A $n \times m$ et B $m \times n$ $(m > n)$, alors

$$\text{dét}(\lambda I_m - BA) = \lambda^{m-n} \text{dét}(\lambda I_n - AB) .$$

P9 : Si A est une matrice diagonale ou triangulaire, alors $\lambda_i = a_{ii}\,, \forall i$.

1.9 Trace d'une matrice carrée

La trace de $A = [a_{ij}] \in \mathbb{C}^{n\times n}$ est définie par la somme des éléments diagonaux

$$\text{trace}(A) = \sum_{i=1}^{n} a_{ii} .$$

Propriétés

P1 : $\text{trace}(A) = \text{trace}(A')$.

P2 : Linéarité : $\text{trace}(A+B) = \text{trace}(A) + \text{trace}(B)$

$\text{trace}(\alpha A) = \alpha\, \text{trace}(A)$.

P3 : Invariance par permutation : $\text{trace}(AB) = \text{trace}(BA)$, si AB et BA sont carrées.

P4 : $\text{trace}(A) = \sum_i \lambda_i$, si $\{\lambda_1, \dots, \lambda_n\} = \text{spect}(A)$. (Clef, théorème de Schur).

P5 : Deux matrices semblables ont même trace, $\text{trace}(P^{-1}AP) = \text{trace}(A)$.

1.10 Formes linéaires, formes quadratiques

Produit scalaire défini sur un espace vectoriel E :

Soit E un espace vectoriel pas forcément de dimension finie, une application de $E \times E$ dans $\mathbb{C}$ ($\mathbb{R}$) est un produit scalaire ssi à tout couple (x, y) elle fait correspondre le nombre complexe (réel) $< x, y >$ tel que

D1 : $< x, y > = \overline{< y, x >}$.

D2 : $< \alpha_1 x_1 + \alpha_2 x_2, y > = \alpha_1 < x_1, y > + \alpha_2 < x_2, y > \qquad \forall \alpha_1, \alpha_2 \in \mathbb{C}\ (\mathbb{R})$.

D3 : $< x, x > \geq 0$ (réel d'après D1).

D4 : $< x, x > = 0 \Longleftrightarrow x = 0_E$.

Comme première propriété

$$< x, \alpha y > = \overline{\alpha} < x, y > \ .$$

La norme Euclidienne d'un élément x associée au produit scalaire est définie par

$$\|x\| = \sqrt{< x, x >}.$$

L'inégalité suivante dite de Cauchy-Schwarz est donnée par

$$| < x, y > | \leq \|x\| \|y\| \, .$$

L'égalité a lieu si les vecteurs sont colinéaires.

En dimension finie, on note par $x' = [x_1, \dots, x_n]$ le vecteur ligne transposé du vecteur colonne x et par $x^* = [\overline{x}_1 \dots, \overline{x}_n]$ le vecteur adjoint, alors le produit scalaire usuel se calcule par

$$< x, y > = y^* x = \sum_{i=1}^{n} x_i \overline{y}_i \quad \text{sur} \quad \mathbb{C}^n,$$

$$< x, y >= y'x = \sum_{i=1}^{n} x_i y_i \quad \text{sur} \quad I\!R^n.$$

On verra au chapitre 3 que la norme Euclidienne associée au produit scalaire usuel est la 2-norme de Hölder définie par

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}\,.$$

Soit $a \in \mathbb{C}^n$, l'application φ de $\mathbb{C}^n$ dans $\mathbb{C}$ est appelée une forme linéaire si

$$x \longrightarrow \varphi(x) =< x, a >= a^* x\,.$$

Soit $A \in \mathbb{C}^{n\times n}$ hermitienne, l'application ϕ de $\mathbb{C}^n$ dans $I\!R$ est appelée une forme quadratique si

$$x \longrightarrow \phi(x) =< Ax, x >= x^* Ax.$$

En effet, d'après D1, A hermitienne entraine que $< Ax, x >$ est réel

$$< Ax, x >= x^* Ax = x^* A^* x =< x, Ax >= \overline{< Ax, x >}\,.$$

Remarque : Dans le cas réel, supposons $A \in I\!R^{n\times n}$ non symétrique et $x \in I\!R$, alors

$$\varphi_A(x) = x' Ax = x'(\frac{1}{2}(A + A'))x = \varphi_{\frac{1}{2}(A+A')}(x).$$

La matrice $\frac{1}{2}(A + A')$ est symétrique. Dans le cas réel, l'hypothése de symétrie de A est donc toujours implicitement admise.

On dit que A hermitienne, est

définie positive	si $x^* Ax > 0$	$\forall x \neq 0$,
semi définie positive	si $x^* Ax \geq 0$	$\forall x$.

La forme quadratique associée est définie positive (resp. semi définie positive) .

De façon évidente, les matrices réelles BB^* et B^*B sont semi définies positives. Montrons le pour B^*B,

$$x^* B^* Bx = \|Bx\|_2^2 \geq 0\,.$$

B^*B est définie positive si et seulement si $Ker\, B = \{0\}$, c'est à dire, B est de plein rang colonne. En effet, dans ce cas : $Bx = \sum_i B^i x_i = 0 \Longleftrightarrow x = 0$.

1.11 Matrices orthogonales et unitaires

Soit une matrice U de $\mathbb{C}^{m\times n}$ $(m \geq n)$, l'espace $\mathbb{C}^m$ des colonnes de U étant muni du produit scalaire usuel. On dit que U est unitaire si les colonnes de U sont des vecteurs 2 à 2 orthogonaux et de norme unité, c'est à dire si

$$U^*U = I_n.$$

Lorsque U est réelle ont dit que U est orthogonale, dans ce cas

$$U'U = I_n.$$

Premières propriétés des matrices unitaires :

P1 : Le produit de 2 matrices unitaires est une matrice unitaire.

P2 : Une matrice carrée est unitaire si et seulement si

$$U^{-1} = U^*.$$

P3 : Le déterminant d'une matrice carrée unitaire est de module égal à 1.

1.11.1 Les matrices de permutation

Soit σ une permutation définie dans la section 1.5, par exemple sur $\{1,2,3,4\}$, $\begin{array}{c|cccc} i & 1 & 2 & 3 & 4 \\ \hline \sigma(i) & 3 & 1 & 4 & 2 \end{array}$.

La matrice de passage de la base canonique $\{e_i\}$ de $\mathbb{R}^n$ à la base $\{f_i = e_{\sigma(i)}\}$, $Q = [e_{\sigma(1)}, \ldots, e_{\sigma(n)}]$, est appelée la matrice de permutation associée à σ,

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

La permutation inverse est définie par $\begin{array}{c|cccc} i & 1 & 2 & 3 & 4 \\ \hline \sigma^{-1}(i) & 2 & 4 & 1 & 3 \end{array}$.

D'où $Q^{-1} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} = Q'$. Cette propriété est vraie pour toutes les matrices de permutation.

Soit $X = [x_1, x_2, x_3, x_4]'$ la matrice colonne des coordonnées d'un vecteur dans $\{e_i\}$ et Y la matrice colonne des coordonnées de ce vecteur dans la base permutée, alors

$Y = Q^{-1}X = [x_3, x_1, x_4, x_2]'$.
Multiplier à gauche une matrice par Q^{-1} revient à permuter ses lignes selon σ et la multiplier à droite par Q permute ses colonnes selon σ. Sur l'exemple, $X = [x_1, x_2, x_3, x_4]$ alors $XQ = [x_3, x_1, x_4, x_2]$.

1.11.2 Les matrices de rotation

La dimension 2 :

Une matrice Q, 2×2, orthogonale est une matrice de rotation si

$$Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, \qquad (\text{dét}(Q) = 1)\,.$$

Alors, $y = Qx$ est obtenu par rotation de x d'un angle θ.
La dimension n :

Une matrice $Q(i, k, \theta)$, $n \times n$, orthogonale est une matrice de rotation de Givens si elle est de la forme

$$Q(i,k,\theta) = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & \cos(\theta) & \dots & -\sin(\theta) & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & \sin(\theta) & \dots & \cos(\theta) & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix}, \ (\text{dét}(Q(i,k,\theta)) = 1)\,,$$

obtenue par modification de la matrice identité I_n à l'intersection des lignes, colonnes i et k.

Si $y = Q(i, k, \theta)x$, y est obtenu par la rotation de x d'un angle θ dans le sens direct dans le plan défini par les deux vecteurs e_i et e_k de la base canonique (y diffère de x seulement sur les coordonnées i et k).

1.11.3 Construction d'une base orthonormée par le procédé de Gram-Schmidt

Proposition : A partir d'un ensemble de vecteurs linéairement indépendants de $I\!\!K^n$ $\{f_1, \ldots, f_k\}$, on peut construire un ensemble de vecteurs $\{q_1, \ldots, q_k\}$ orthonormés qui engendrent le même espace

$$\mathcal{E}\{f_1, \ldots, f_k\} = \mathcal{E}\{q_1, \ldots, q_k\}\,. \quad \square$$

Preuve : Construisons d'abord une suite de k vecteurs $\{p_1, \ldots, p_k\}$ deux à deux orthogonaux.

$$\begin{array}{rcl} p_1 & = & f_1 \\ p_2 & = & f_2 - \dfrac{< p_1, f_2 >}{\|p_1\|_2^2} p_1 \\ . & . & \ldots \\ p_i & = & f_i - \dfrac{< p_1, f_i >}{\|p_1\|_2^2} p_1 - \dfrac{< p_2, f_i >}{\|p_2\|_2^2} p_2 \ldots - \dfrac{< p_{i-1}, f_i >}{\|p_{i-1}\|_2^2} p_{i-1} \qquad i \leq k\,. \end{array}$$

On vérifie que $< p_i, p_j >= 0$ pour $j < i$ et que $\mathcal{E}\{f_1, \ldots, f_i\} = \mathcal{E}\{p_1, \ldots, p_i\}$. La suite

$$\{q_i = \frac{p_i}{\|p_i\|_2}\}_{i=1\ldots k}$$

est bien une suite de vecteurs orthonormés. Si de plus $k = n$, c'est une base de $I\!\!K^n$. $\square$

1.12 Opérateur vec et produit de Kronecker

1.12.1 L'opérateur vec

Soit $A = [A^1, \ldots, A^p]$, une matrice $n \times p$. On définit l'opérateur *vec* d'empilement des colonnes par

$$vec(A) = \begin{bmatrix} A^1 \\ \vdots \\ A^p \end{bmatrix} \in I\!\!K^{np}$$

Si $A = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}$, l'empilement en colonne des lignes de A se fait par $vec(A') \in I\!\!K^{np}$. En général, $vec(A) \neq vec(A')$ sauf si A est une matrice colonne,

Notations : $(vec(A))' = vec'(A)$ et $(vec(A))^* = vec^*(A)$

Propriétés

P1 : Linéarité : $vec(\alpha A + \beta B) = \alpha\, vec(A) + \beta\, vec(B)$

P2 : Soient X et Y deux matrices $n \times p$

$$\boxed{\text{trace}(Y^* X) = \sum_{j=1}^{p} \sum_{i=1}^{n} \overline{y}_{ij} x_{ij} \;\; = vec^*(Y) vec(X) =< vec(X), vec(Y) >.}$$

1.12.2 Produit de Kronecker

Soient $A = [a_{ij}]$ une matrice $n \times p$ et $B = [b_{ij}]$ une matrice $m \times q$. Alors le produit de Kronecker de A par B est une matrice $nm \times pq$.

$$A \otimes B = [a_{ij} B] \;\; = \begin{bmatrix} a_{11}B & & \cdots & & a_{1p}B \\ & \ddots & & & \\ \vdots & & a_{ij}B & & \vdots \\ & & & \ddots & \\ a_{n1}B & & \cdots & & a_{np}B \end{bmatrix}.$$

Propriétés

Les trois premières propriétés justifient le nom de produit.

P1 : $(\alpha\, A) \otimes B = A \otimes (\alpha\, B) \;\; = \alpha(A \otimes B)$.

P2 : Soient A et B deux matrices de mêmes dimensions, C et D, deux autres matrices de mêmes dimensions, alors

$$(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D.$$

P3 : $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

P4 : $I_{np} = I_n \otimes I_p \;\; = I_p \otimes I_n$.

P5 : $(A \otimes B)^* = A^* \otimes B^*$.

P6 : $(AC) \otimes (BD) = (A \otimes B)(C \otimes D)$.

P7 : Soient A,B, deux matrices régulières, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

P8 : Soient A,B, deux matrices carrées, alors $\quad \text{trace}(A \otimes B) = \text{trace}(A)\text{trace}(B)$.

P9 : Si x et y sont des matrices colonnes alors

$$xy' = \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_p \\ \vdots & \ddots & \vdots \\ x_n y_1 & \cdots & x_n y_p \end{bmatrix} = x \otimes y' \;\; = y' \otimes x \;\; = [y_1 x | \cdots | y_p x].$$

$$vec(xy) = \begin{bmatrix} y_1 x \\ \vdots \\ y_p x \end{bmatrix} = y \otimes x.$$

P10 : Soient A, B, deux matrices carrées respectivement $n \times n$ et $p \times p$ avec $spect(A) = \{\lambda_1, \ldots, \lambda_n\}$ et $spect(B) = \{\mu_1, \ldots, \mu_p\}$ alors

$$spect(A \otimes B) = \{\lambda_i \mu_j \;/\; i = 1, \ldots, n; j = 1, \ldots, p\},$$

$$x_i \otimes y_j \text{ est vecteur propre de } A \otimes B,$$

si x_i et y_j sont les vecteurs propres associés respectivement à λ_i et à μ_j . $\square$

P11 : Soient A et B, deux matrices hermitiennes (semi) définies positives, alors $A \otimes B$ est une matrice hermitienne (semi) définie positive.

P12 : Si A est $n \times n$ et B est $p \times p$, $\quad \text{dét}(A \otimes B) = (\text{dét}(A))^p (\text{dét}(B))^n$.

P13 : $rang(A \otimes B) = rang(A) rang(B)$.

P14 : Le produit de Kronecker de 2 matrices unitaires est une matrice unitaire.

P15 : Le produit de Kronecker de 2 matrices diagonales (rep. triangulaires) est une matrice diagonale (resp. triangulaire).

Vectorisation d'un produit de matrices

Proposition : Soient A, B et C telles que ABC existe. Alors,

$$\boxed{vec(ABC) = (C' \otimes A) vec(B).} \quad \square$$

Preuve : Soit p le nombre de colonnes de B, $B = \sum_{j=1}^{p} B^j e_j'$. La propriété P9 donne

$$\begin{aligned} vec(ABC) &= vec\left(\sum_{j=1}^{p} AB^j e_j' C\right) = \sum_{j=1}^{p} vec\left(AB^j (C' e_j)'\right) \\ &= \sum_{j=1}^{p} (C' e_j) \otimes (AB^j) = (C' \otimes A) \sum_{j=1}^{p} e_j \otimes B^j \\ &= (C' \otimes A) \sum_{j=1}^{p} vec(B^j e_j') = (C' \otimes A) vec\left(\sum_{j=1}^{p} B^j e_j'\right). \quad \square \end{aligned}$$

Cas particulier

$vec(AB) = (I_p \otimes A) vec(B) = (B' \otimes I_m) vec(A)$ avec A matrice $m \times n$ et B matrice $n \times p$.

Corollaire : Soient A, B, C et D telles que $ABCD$ soit carrée

$$\boxed{\text{trace}(ABCD) = vec'(A')(D' \otimes B)vec(C)\ .} \qquad \square$$

1.12.3 Matrice de commutation

Lorsque l'on a besoin d'exprimer $vec(A')$ en fonction de $vec(A)$, la matrice de commutation permet ce passage.

On appelle K_{mn}, la matrice de commutation d'ordre $m \times n$ avec $\dim(K_{mn}) = mn \times mn$

$$K_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} E_{ij}(m,n) \otimes E'_{ij}(m,n).$$

Exemple : Soit $m = 3$ et $n = 2$ alors K_{32} est une matrice 6×6

$$K_{32} = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

Remarque : K_{mn} est une matrice de permutation.

Proposition : Soit A une matrice $m \times n$, alors

$$\boxed{vec(A') = K_{mn} vec(A).} \qquad \square$$

Remarque : $vec(A) = K_{nm} vec(A')$.

Propriétés

P1 : $K'_{mn} = K_{nm} = K_{mn}^{-1}$

P2 : $K_{1n} = K_{n1} = I_n$.

P3 : Soient A et B, deux matrices respectivement $n \times s$ et $m \times t$

$K_{mn}(A \otimes B) = (B \otimes A)K_{ts} \Leftrightarrow B \otimes A = K_{mn}(A \otimes B)K_{st} \Leftrightarrow A \otimes B = K_{nm}(B \otimes A)K_{ts}$.

1.13 Exercices

<u>**Exercice 1**</u> : Dans l'espace $I\!R$-espace vectoriel de dimension 4.

1) Soit $D = \{u_1, u_2, u_3, u_4\}$ une base de E et soit g une application linéaire de E dans lui-même (on dit un endomorphisme de E) ayant pour matrice N dans la base D : (a, b,

c et d désignent des réels quelconques)

$$N = \begin{bmatrix} 1 & 0 & \mathrm{a} & \mathrm{b} \\ 0 & 1 & \mathrm{c} & \mathrm{d} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

On notera $Im\, g$ l'image de g et $Ker\, g$ son noyau.

a) Calculer le rang de g ainsi que la dimension de $Ker\, g$.

b) Donner une base pour $Im\, g$.

c) Soient $v_1 = au_1 + cu_2 - u_3$, $v_2 = bu_1 + du_2 - u_4$. Démontrer que $\{v_1, v_2\}$ est une base de $Ker\, g$.

2) Soit $B = \{e_1, e_2, e_3, e_4\}$ une base de E. On pose :

$c_1 = e_2 + 2e_3 + 2e_4$, $c_2 = e_1 + e_2 + 2e_3 + 3e_4$, $c_3 = 2e_1 + 2e_2 + 2e_3 + 3e_4$ et $c_4 = 2e_1 + 3e_2 + 3e_3 + 3e_4$.

a) Démontrer que $C = \{c_1, c_2, c_3, c_4\}$ est une base de E.

b) Ecrire la matrice de passage P de B à C et calculer son inverse.

c) Soit f l'endomorphisme de E défini par

$f(e_1) = -e_1 - 3e_2 - 6e_3 - 7e_4$, $f(e_2) = 2e_1 + 5e_2 + 10e_3 + 12e_4$, $f(e_3) = -3e_1 - 5e_2 - 10e_3 - 13e_4$ et $f(e_4) = 2e_1 + 3e_2 + 6e_3 + 8e_4$. Calculer la matrice M de f dans la base B puis la matrice M' de f dans la base C.

d) Donner une base B_I pour $Im\, f$ et une base B_K pour $Ker\, f$ (les vecteurs de B_I et B_K devront être exprimés dans la base B).

e) Soit V le sous-espace vectoriel de E engendré par c_1 et c_2. Soit q la projection de E sur V. Calculer Q la matrice de q dans la base B

Exercice 2 : Soit E l'espace vectoriel des polynômes sur $I\!\!R$ de degré inférieur ou égal à 3.

Déterminer le rang du système de vecteurs $\{u_1, u_2, u_3\}$ de E défini par :

$u_1 = 4 + 3X + 2X^2 + X^3$, $u_2 = 6 + 2X + 2X^2 + 2X^3$ et $u_3 = -8 - 11X - 6X^2 + X^3$.

Exercice 3 : Soit $B = \{e_1, e_2, e_3, e_4\}$ une base de $I\!\!R^4$. On considère l'endomorphisme

f de $I\!R^4$ et sa matrice $A_{f,B}$ associée dans la base B

$$A_{f,B} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}.$$

1) Calculer les valeurs propres de f.

2) Pour chaque valeur propre λ de f, déterminer l'espace propre correspondant.

3) Déduire que f est diagonalisable.

4) Trouver en l'exprimant dans la base B, une base $C = \{u_1, u_2, u_3, u_4\}$ de $I\!R^4$ avec $A_{f,C}$ diagonale et préciser quelle est la matrice $A_{f,C}$.

<u>Exercice 4</u> : Soit la matrice

$$A = \begin{bmatrix} 1 & -3 & \text{-4} \\ -1 & 3 & 4 \\ 1 & -3 & -4 \end{bmatrix}.$$

Calculer A^2. Calculer $(I + A)(I - A)$ où I est la matrice unité d'ordre 3. En déduire que $I + A$ est inversible.

<u>Exercice 5</u> : On considère la matrice

$$A = \begin{bmatrix} 7 & 4 & 0 & 0 \\ -12 & -7 & 0 & 0 \\ 20 & 11 & -6 & -12 \\ -12 & -6 & 6 & 11 \end{bmatrix}.$$

1) Calculer les valeurs propres de A.

2) A est-elle diagonalisable? Inversible?

3) Déterminer une base de l'espace propre associé à la plus grande valeur propre de A.

<u>Exercice 6</u> : La matrice suivante est-elle inversible? Si oui, calculer son inverse

$$A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 3 & 3 & 2 \\ 2 & 4 & 3 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Exercice 7 : Montrer que la matrice suivante est inversible dans $\mathcal{M}_{33}(\mathbb{R})$ et calculer son inverse

$$A = \begin{bmatrix} -3 & 5 & 6 \\ -1 & 0 & -1 \\ -2 & 1 & 1 \end{bmatrix}.$$

Exercice 8:

Démontrer ces quelques "petits" résultats utiles, les matrices sont supposées réelles.

1.1. A $m \times n$, B $n \times p$, C $n \times p$, M $m \times m$ symétrique définie positive, x $n \times 1$.

a) $Ax = 0 \Longleftrightarrow A'Ax = 0$ et $Ker\ A = Ker\ A'MA$.

b) $AB = 0 \Longleftrightarrow A'AB = 0$.

c) $A'AB = A'AC \Longleftrightarrow AB = AC$.

1.2. A $m \times n$, B $n \times n$, C $n \times n$ symétrique.

a) $(Ax = 0, \forall x \in \mathbb{R}^n) \Longleftrightarrow A = 0$.

b) $(x'Ay = 0, \forall x \in \mathbb{R}^m, \forall y \in \mathbb{R}^n) \Longleftrightarrow A = 0$.

c) $(x'Cx = 0, \forall x \in \mathbb{R}^n) \Longleftrightarrow C = 0$.

d) $(x'Bx = 0, \forall x \in \mathbb{R}^n) \Longleftrightarrow B' = -B$.

1.3. Soient B $m \times p$ et C $p \times n$, alors $A = BC$ s'écrit

$$A = \sum_{i=1}^{p} B^i C_i\,,$$

où B^i et C_i sont respectivement la $i$ème colonne de B et la $i$ème ligne de C.

Si $D = \text{diag}(d_1, \ldots, d_p)$ alors $A = BDC$ s'écrit

$$A = \sum_{i=1}^{p} d_i B^i C_i\,.$$

Exercice 9 : Montrer que trace(AB) = trace(BA).

Exercice 10 : Soit A une matrice $n \times n$ et $spect(A) = \{\lambda_i\}_i$. Soit B une matrice $p \times p$ et $spect(B) = \{\mu_j\}_j$. Montrer que les np valeurs propres de $A \otimes B$ sont $\lambda_i \mu_j$ et que, si x est un vecteur propre de A et y un vecteur propre de B, alors, $x \otimes y$ est vecteur propre de $A \otimes B$.

Vérifier sur l'exemple $A = B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ qu'il y a des vecteurs propres de $A \otimes B$ qui ne se déduisent pas des vecteurs propres de A et B.

Exercice 11 : Soient A et B deux matrices carrées d'ordre respectifs n et p.

a) Montrer que si A et B sont symétriques (semi) définies positives, alors $A \otimes B$ est symétrique (semi) définie positive.

b) Montrer que : $\text{dét}(A \otimes B) = [\text{dét}(A)]^p [\text{dét}(B)]^n$

c) Montrer que : $rang(A \otimes B) = rang(A) rang(B)$. Indication : Montrer que $rang(A \otimes B) = rang((AA') \otimes (BB'))$ et conclure en utilisant la DVS.

Exercice 12 : Soit A une matrice $n \times n$. Démontrer que $vec(A') = K_{mn} vec(A)$ avec

$$K_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} E_{ij}(m,n) \otimes E'_{ij}(m,n).$$

Chapter 2

Décomposition de Matrices

On a vu qu'une application linéaire f de l'espace vectoriel E dans l'espace vectoriel F, tous deux de dimension finie, pouvait être représentée par une matrice A relativement aux bases choisies dans ces espaces. Une application linéaire étant ainsi représentée par différentes matrices selon les bases choisies, le problème se pose de trouver des bases dans lesquelles la matrice représentant l'application linéaire soit *la plus simple possible*. C'est le problème de la réduction d'une matrice. Le cas le plus simple est celui des matrices carrées diagonalisables. Mais toutes les matrices carrées ne sont pas diagonalisables... D'autre part, toutes les matrices ne sont pas carrées. Comment réduire des matrices rectangulaires? Ou plutôt, comment décomposer une matrice en un produit de matrices ayant de "bonnes" propriétés? Cependant, avant d'aborder les différents types de réduction ou de décomposition et pour mieux comprendre leur interprétation géométrique, il est nécessaire d'aborder la notion de projecteur qui sera développée au chapitre suivant.

2.1 Les projecteurs

La notion de projection est fondamentale tant en analyse fonctionnelle lors de l'approximation par des séries de fonctions qu'en statistique où pour prendre un exemple simple, la moyenne d'une variable s'exprime comme la projection sur la "droite des constantes". De nombreuses méthodes statistiques comme la régression linéaire, l'analyse en composantes principales, etc, sont basées sur les projecteurs. Dans tous les espaces vectoriels sur lesquels on a défini un produit scalaire, la projection est un outil pour résoudre de nombreux problèmes d'optimisation. On reviendra sur cette notion qui est abordée dans la section suivante.

2.1.1 Sous espaces supplémentaires et projecteurs

Soient F et G deux s.e.v. du $\mathbb{K}$-e.v. E.

$$F+G=\{x+y\,|\,x\in F,\,y\in G\}\qquad \text{et}\qquad F\times G=\{(x,y)\,|\,x\in F,\,y\in G\}\,.$$

On dit que F et G sont supplémentaires si $F\bigcap G=\{0_E\}$ et $F+G=E$.
De façon équivalente : tout vecteur x de E s'écrit de manière unique

$$x=u+v \text{ avec } u\in F \text{ et } v\in G.$$

E est alors somme directe de F et G noté $E=F\oplus G$

Remarquons que le supplémentaire d'un s.e.v. n'est pas unique.

Si F et G sont supplémentaires, les applications p et q de E dans E définies par

$$\forall x\in E,\qquad x=p(x)+q(x)\qquad \text{avec}\qquad p(x)\in F\qquad \text{et}\qquad q(x)\in G$$

sont linéaires (endomorphismes de E) et vérifient :

P1 $\quad p^2=p\ ;\ q^2=q$ (idempotence)

P2 $\quad p\circ q=q\circ p=0$

P3 $\quad p+q=Id_E$

P4 $\quad Im\,p=F=Ker\,q\quad$ et $\quad Im\,q=G=Ker\,p$.

On dit que p est la projection sur F parallèlement à G et que $q=Id_E-p$ est la projection sur G parallèlement à F ou encore le projecteur supplémentaire de p.
On appelle projecteur dans un $\mathbb{K}$-e.v. E tout endomorphisme idempotent de E.

Dans le cas particulier où les deux sous espaces supplémentaires sont orthogonaux (bien sûr, E est muni d'un produit scalaire)

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

alors les projecteurs p et q associés sont dits projecteurs orthogonaux.

2.1.2 Exemple fondamental

Soient u et v de $\mathbb{K}^n$ muni du produit scalaire usuel, tels que

$$<u,v>=v^*u=1\,.$$

Remarquons que, puisque $< u, v >= \|u\|_2 \|v\|_2 \cos(u, v)$, la condition précédente impose que l'angle vectoriel entre u et v est aigu. Considérons la matrice $n \times n$

$$P = uv^* .$$

Cette matrice jouit des propriétés suivantes :

$$P^2 = uv^*uv^* = uv^* = P$$

et si $x \in Im\, u$, c'est à dire si $x = \alpha u$,

$$Px = uv^*(\alpha u) = \alpha uv^*u = \alpha u = x .$$

Mais, si x est orthogonal à v, alors

$$Px = uv^*x = u(v^*x) = 0 .$$

L'image de P est donc $Im\, u$, le noyau de P est le sous espace vectoriel de dimension $n-1$ orthogonal à v.

$$I\!\!K^n = Im\, u \oplus \{Im\, v\}^\perp .$$

P est donc la matrice de l'application linéaire "projection sur u parallèlement à $\{Im\, v\}^\perp$".

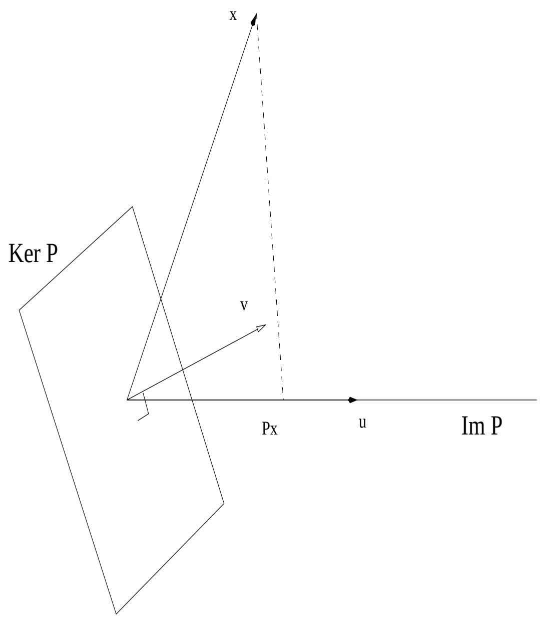


Figure 1 : Projection sur u parallèlement à $\{Im\, v\}^\perp$.

Si on choisit $v = u$ et $\|u\|_2 = 1$, dans ce cas, le <u>projecteur orthogonal</u> s'écrit $P = uu^*$.

D'une façon plus générale, soit F donné ainsi qu'une base $\{u_1, \ldots, u_r\}$ orthonormée de F. Soit $U = [u_1, \ldots, u_r]$, alors $U^*U = I_r$. On montrera au chapitre suivant que la matrice

$$P = \sum_{i=1}^{r} u_i u_i^* = UU^*$$

est le projecteur orthogonal sur $F = Im\, U$. Le projecteur ($P^2 = P$) est orthogonal car $P = P^*$.

2.1.3 D'autres matrices orthogonales : les matrices de réflexion

La dimension 2 :

Une matrice Q, 2×2, orthogonale et symétrique, est une matrice de réflexion si elle est de la forme

$$Q = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}, \qquad (\text{dét}(Q) = -1)\,.$$

Si $y = Qx = Q'x$, y est obtenu par symétrie du vecteur x par rapport à la droite vectorielle définie par

$$S = Im \begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix}.$$

La dimension n :

Soit $v \in I\!R^n$, $v \neq 0$, une matrice Q, $n \times n$, orthogonale et symétrique, est une matrice de réflexion de Householder si elle est de la forme

$$Q(v) = I_n - 2vv'/v'v\,, \qquad (\text{dét}(Q(v)) = -1)\,.$$

Par un léger abus de langage on convient que $Q(0) = I_n$ bien que $\text{dét}(Q(0)) = 1$.

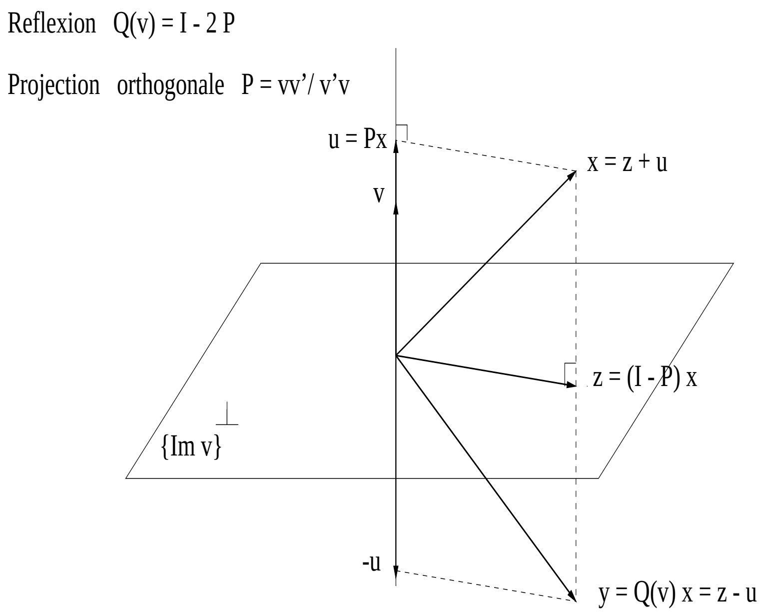


Figure 2: Symétrie par rapport à l'hyperplan vectoriel $\{Im\,v\}^\perp$.

Si $y = Q(v)x = Q'(v)x$, y est obtenu par symétrie du vecteur x par rapport à l'hyperplan vectoriel $\{Im\,v\}^\perp = \{z = (z_1, \ldots, z_n) \in I\!R^n \,|\, z'v = v_1 z_1 + \ldots v_n z_n = 0\}$. Cela résulte du fait que la matrice $P = vv'/v'v$ est la matrice de projection orthogonale sur $Im\,v$ (clef section 2 1.2).

Proposition : Toute matrice Q $n \times n$ orthogonale peut s'écrire comme le produit de n réflexions

$$\exists\, v_1, \ldots, v_n \in I\!R^n\,, \qquad Q = Q(v_1) \ldots Q(v_n)\,.$$

Remarque : Les endomorphismes de $(I\!R^n, \|.\|_2)$ canoniquement associés à des matrices carrées orthogonales Q, sont appelés des isométries de $(I\!R^n, \|.\|_2)$ car d'après P4 les normes sont conservées ($\|Qx\|_2 = \|x\|_2$) ou de façon équivalente les produits scalaires ($< Qx, Qy >= y'Q'Qx =< x, y >$). Pour les angles, on a

$$\cos(Qx, Qy) = \frac{< Qx, Qy >}{\|Qx\|_2 \|Qy\|_2} = \cos(x, y).$$

2.2 Matrices carrées diagonalisables

Une matrice carrée A d'ordre n est diagonalisable si elle est semblable à une matrice diagonale $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, c'est à dire s'il existe une matrice S inversible (matrice de passage de l'ancienne base à la base diagonalisante) telle que

$$\Lambda = S^{-1}AS \Longleftrightarrow A = S\Lambda S^{-1} \Longleftrightarrow AS = S\Lambda \Longleftrightarrow S^{-1}A = \Lambda S^{-1}\,.$$

La $i$ème colonne de S est le vecteur propre de A associé à la valeur propre λ_i.

Condition nécessaire et suffisante : Une condition nécessaire et suffisante pour que A carrée d'ordre n, soit diagonalisable est que ses n vecteurs propres soient linéairement indépendants. $\square$

Condition suffisante : Les vecteurs propres associés à des valeurs propres distinctes sont linéairement indépendants. Si toutes les valeurs propres de A sont distinctes, alors A est diagonalisable. $\square$

Décomposition spectrale de A diagonalisable : Soit A diagonalisable telle que $A = S\Lambda S^{-1}$. Associés à λ_i, notons u^i la $i$ème colonne de S et v^{i*} la $i$ème ligne de S^{-1}. La décompositon spectrale de A s'écrit

$$A = \sum_{i=1}^{n} \lambda_i u^i v^{i*}$$

Le vecteur v^i est vecteur propre de A^* associé à $\overline{\lambda}_i$ et $v^{j*}u^i = 0$, si $j \neq i$. Ceci signifie que les vecteurs propres distincts de A et de A^* sont orthogonaux. $\square$

Preuve : La diagonalisation donne $S^{-1}A = \Lambda S^{-1}$. La $i$ème ligne de cette équation matricielle s'écrit $v^{i*}A = \lambda_i v^{i*}$ et en prenant l'adjointe $A^*v^i = \overline{\lambda}_i v^i$. D'autre part, $S^{-1}S = I$ signifie que $v^{j*}u^i = \delta_{ij}$. $\square$

Remarquons que les valeurs propres λ_i ne sont pas toutes forcément distinctes. Regroupons les valeurs propres égales. Soit n_i la multiplicité de λ_i,

$$n = n_1 + \ldots + n_s\,.$$

Posons $N_0 = 0$ et pour $i = 1$ à s, $N_i = n_1 + \ldots + n_i$. A s'écrit

$$A = \sum_{i=1}^{s} \lambda_i \left(\sum_{k=N_{i-1}+1}^{N_i} u^k v^{k*} \right)\,.$$

Posons

$$P_i = \sum_{k=N_{i-1}+1}^{N_i} u^k v^{k*}$$

alors

$$A = \sum_{i=1}^{s} \lambda_i P_i$$

$$I = \sum_{i=1}^{s} P_i \, .$$

C'est l'écriture de la décomposition spectrale de A par les projecteurs sur les sous espaces propres $Ker(A - \lambda_i I)$. En effet les matrices P_i vérifient les propriétés suivantes.

$$AP_i = \lambda_i P_i \qquad P_i^2 = P_i \qquad \text{et} \qquad P_i P_j = 0 \qquad j \neq i \, .$$

2.3 Factorisation QR d'une matrice rectangulaire

Nous avons vu comment par le procédé de Gram-Schmidt, il est possible à partir d'une famille libre $\{A^1, \ldots, A^n\}$ de vecteurs de $I\!K^m$ $(m \geq n)$, de construire une famille $\{Q^1, \ldots, Q^n\}$ orthonormée qui engendre le même espace. La construction de la matrice unitaire $Q_1 = [Q^1, \ldots, Q^n]$ à partir de la matrice de plein rang colonne $A = [A^1, \ldots, A^n]$ est ainsi basée sur le système d'équations

$$\left\{ \begin{array}{lcl} A^1 & = & r_{11} Q^1 \\ A^2 & = & r_{12} Q^1 + r_{22} Q^2 \\ . & . & \cdots \\ A^n & = & r_{1n} Q^1 + r_{2n} Q^2 + \cdots + r_{nn} Q^n \, . \end{array} \right.$$

Ceci conduit à la forme "maigre" de la factorisation QR.

Version "maigre" de la factorisation QR

Soit $A \in \mathbb{C}^{m \times n}$ $(m \geq n)$ une matrice de plein rang colonne,

$$A = Q_1 R_1$$

est unique avec Q_1 matrice $m \times n$ unitaire et $R_1 \in \mathbb{C}^{n \times n}$ matrice triangulaire supérieure à éléments diagonaux réels positifs. $\square$

La construction de Q_1 et de R_1 par la procédure de Gram-Schmidt est numériquement instable à cause de la propagation des erreurs d'arrondi dues au fait que les colonnes de Q_1 sont calculées "en cascade" : Q^2 est construite à partir de de Q^1, Q^3 en fonction de Q^2 et Q^1, etc. D'autres méthodes (voir exercice) numériquement plus stables sont mises en oeuvre.

Il est difficile de connaître à priori le rang d'une matrice. La version "pleine" de la factorisation QR supprime l'hypothèse de plein rang colonne pour A.

Version "pleine" de la factorisation QR

Soit $A \in \mathbb{C}^{m\times n}$ $(m \geq n)$, il existe Q et R telles que

$$A = Q \begin{bmatrix} R \\ 0 \end{bmatrix}$$

avec Q matrice $m \times m$ unitaire et $R \in \mathbb{C}^{n\times n}$ matrice triangulaire supérieure à éléments diagonaux réels positifs ou nuls. Si A est de plein rang colonne alors les n premières colonnes de Q forment une base orthonormée de ImA et la diagonale de R est à éléments positifs. $\square$

2.4 Décomposition unitaire des matrices carrées

Le théorème suivant indique que toute matrice carrée est triangularisable.

2.4.1 Le théorème de Schur

Théorème de Schur : Si $A \in \mathbb{C}^{n\times n}$, il existe une matrice unitaire $U \in \mathbb{C}^{n\times n}$ $(U^*U = UU^* = I_n)$ et une matrice triangulaire supérieure T telles que

$$A = UTU^*\,, \qquad U^*AU = T\,, \qquad AU = UT\,, \qquad U^*A = TU^*\,,$$

où $T = \Lambda + N$ avec $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ matrice diagonale des valeurs propres de A et N matrice triangulaire supérieure stricte. $\square$

Remarque : Les colonnes de U ou vecteurs de Schur, solutions de $AU = U\Lambda + UN$, peuvent être complexes même si A est réelle. Dans le cas A réel, les valeurs propres complexes sont deux à deux conjuguées. Les vecteurs de Schur ne sont vecteurs propres de A que si $N = 0$ (matrices normales), dans ce cas, A est diagonalisable.

Premières applications

A1 $\mathrm{trace}(A) = \mathrm{trace}(UTU^*) = \mathrm{trace}(U^*UT) = \mathrm{trace}(T) = \sum_i \lambda_i$.

A2 $\mathrm{dét}(A) = \mathrm{dét}(UTU^*) = \mathrm{dét}(U)\mathrm{dét}(T)\mathrm{dét}(U^{-1}) = \mathrm{dét}(T) = \prod_i \lambda_i$.

2.4.2 Matrices normales

Une matrice carrée A est normale ssi $A^*A = AA^*$.

Exemples : Les matrices hermitiennes $A^* = A$, unitaires $A^*A = AA^* = I$, anti-hermitiennes $A^* = -A$, sont normales.

Proposition : A est normale ssi il existe une matrice carrée U unitaire telle que

$$A = U\Lambda U^*\,, \qquad U^*AU = \Lambda\,, \qquad AU = U\Lambda\,, \qquad U^*A = \Lambda U^*\,.$$

Λ est la matrice diagonale formée des valeurs propres de A. Autrement dit, une matrice normale est diagonalisable et les vecteurs propres (les colonnes de U) sont orthonormés. $\square$

Preuve : $$A = U\Lambda U^* \stackrel{?}{\Longrightarrow} A^*A = AA^*$$

$A^*A = U\Lambda^*\Lambda U^*$ et $AA^* = U\Lambda\Lambda^*U^*$.

Or $\Lambda^*\Lambda = \Lambda\Lambda^* = diag(|\lambda_i|^2)$. Donc $A^*A = AA^*$.

$$A^*A = AA^* \stackrel{?}{\Longrightarrow} A = U\Lambda U^*$$

La décomposition de Schur donne $A = UTU^*$ avec T triangulaire supérieure. Si A est normale alors T triangulaire supérieure est normale ce qui implique T diagonale (voir exercice). $\square$

Remarque : Lorsque A est normale, la décomposition spectrale de A s'écrit

$$A = \sum_{i=1}^{n} \lambda_i u^i u^{i*} = \sum_{i=1}^{s} \lambda_i \left(\sum_{k=N_{i-1}+1}^{N_i} u^k u^{k*} \right) .$$

Le projecteur

$$P_i = \sum_{k=N_{i-1}+1}^{N_i} u^k u^{k*}$$

sur le sous espace propre associé à λ_i est maintenant orthogonal car $P_i^* = P_i$.

Revenons sur les trois exemples de matrices normales.

Matrices hermitiennes

H1 Une matrice hermitienne est diagonalisable. Ses valeurs propres sont réelles et ses vecteurs propres orthogonaux. $\square$

Preuve : Seul reste à montrer que les valeurs propres sont réelles. $A^* = A$, donc $\Lambda^* = \Lambda$ ce qui implique $\lambda_i = \overline{\lambda_i}$. $\square$

H2 Une matrice symétrique et réelle est diagonalisable. Ses valeurs propres sont réelles et ses vecteurs propres sont réels et orthogonaux. $\square$

Preuve : Une matrice symétrique réelle est hermitienne. Tout vecteur propre u est réel puisque solution du système linéaire réel $(A - \lambda I)u = 0$. $\square$

H3 Une matrice hermitienne est (semi) définie positive ssi toutes ses valeurs propres sont positives (non négatives). $\square$

Preuve : Soit (λ, u) avec $\|u\|_2^2 = 1$ un élément propre de A hermitienne (semi) définie positive. Alors, $< Au, u >=< \lambda u, u >= \lambda$. Si A est (semi) définie positive λ est positif (positif ou nul).

Réciproquement, soit A hermitienne, $A = U\Lambda U^*$,

$$< Ax, x >=< U\Lambda U^{-1}x, x >=< \Lambda U^{-1}x, U^{-1}x > \ .$$

Posons $y = U^{-1}x$. Les composantes du vecteur y sont les composantes du même vecteur dans la base de $I\!\!K^n$ formée des vecteurs propres $\{u_1, \ldots, u_n\}$ de A.

$$< Ax, x >= \sum_{i=1}^{n} \lambda_i |y_i|^2 \, .$$

Le vecteur y est non nul ssi x est non nul. Supposons tous les λ_i strictement positifs, alors $\forall x \neq 0 < Ax, x >> 0$. Supposons maintenant tous les $\lambda_i \geq 0$. Notons $\mathcal{I} = \{i \in \{1, \ldots, n\} \,|\, \lambda_i > 0\}$ et supposons son complémentaire $\overline{\mathcal{I}} = \{i \in \{1, \ldots, n\} \,|\, \lambda_i = 0\}$ non vide. Alors, $ImA = \mathcal{E}\{u_i\}_{i \in \mathcal{I}}$ et $KerA = \mathcal{E}\{u_i\}_{i \in \overline{\mathcal{I}}}$ sont des sous espaces vectoriels supplémentaires orthogonaux et

$$< Ax, x >= \sum_{i \in \mathcal{I}} \lambda_i |y_i|^2 \geq 0 \, .$$

La nullité est obtenue pour x dans $KerA$. $\square$

H4 <u>Critère de Sylvester</u> : Une matrice hermitienne est (semi) définie positive ssi tous ses mineurs principaux sont positifs (non-négatifs). Preuve, voir exercice 6.

H5 Soit A une matrice hermitienne (semi) définie positive. On peut construire une matrice hermitienne (semi) définie positive notée $A^{1/2}$ telle que

$$A = (A^{1/2})^2 \, . \quad \square$$

Preuve : $A = U\Lambda U^*$ où $\Lambda = diag(\lambda_i)$ avec λ_i positf ou (positif ou nul). Définissons $\Lambda^{1/2} = diag(\sqrt{\lambda_i})$; donc, $(\Lambda^{1/2})^2 = \Lambda$. Alors, $A^{1/2} = U\Lambda^{1/2}U^*$ est hermitienne (semi) définie positive et $(A^{1/2})^2 = A$. $\square$

Matrices unitaires

U1 Une matrice carrée unitaire est diagonalisable, ses valeurs propres ont pour module 1 et ses vecteurs propres sont orthogonaux. $\square$

Preuve : Reste à montrer que les valeurs propres sont de module 1. A est carrée unitaire, $A^* = A^{-1}$. Comme $AA^* = A^*A = I$, $\Lambda\Lambda^* = \Lambda^*\Lambda = I$. L'égalité des éléments diagonaux donne $|\lambda_i|^2 = 1$. $\square$

Matrices anti-hermitiennes

AH1 Une matrice anti-hermitienne est diagonalisable, ses valeurs propres sont des imaginaires purs et les vecteurs propres sont orthogonaux. $\square$

Preuve : Reste à montrer que les valeurs propres sont des imaginaires purs. $A^* = -A$ implique que $\Lambda^* = -\Lambda$ et en écrivant les éléments diagonaux, $\overline{\lambda}_i = -\lambda_i$. $\square$

2.5 Décomposition en valeurs singulières

Pour une matrice rectangulaire la notion de valeur propre n'a pas de sens. Néanmoins, les matrices carrées A^*A et AA^* sont hermitiennes semi définies positives. De plus,

$$rang(A) = rang(A^*A) = rang(AA^*) = r\,,$$

et d'après la propriété P7 de la section 1.8, les r valeurs propres non nulles (positives) de A^*A et AA^* sont identiques.

On appelle valeurs singulières de A les racines carrées des valeurs propres non nulles de A^*A ou de A^*A

$$\mu_i = \sqrt{\lambda_i(A^*A)} = \sqrt{\lambda_i(AA^*)}.$$

2.5.1 Deux versions de la DVS

Version "maigre" (DVS1) : Soit A $m \times n$ telle que $rang(A) = r$. Alors

$$\boxed{A = U\Lambda_r^{1/2}V^* = \sum_{i=1}^{r} \mu_i u^i v^{i*}}$$

- $U = [u^1|\cdots|u^r]$ unitaire $m \times r$ est telle que u^i est vecteur propre de AA^* associé à la valeur propre non nulle λ_i.

- $V = [v^1|\cdots|v^r]$ unitaire $n\times r$ est telle que v^i est vecteur propre de A^*A associé à la valeur propre non nulle λ_i.

- $\Lambda_r = \mathrm{diag}(\lambda_1,...,\lambda_r)$ et $\Lambda_r^{1/2} = \mathrm{diag}(\mu_1,...,\mu_r)$, où $\mu_i = \sqrt{\lambda_i}$ est la $i$ème valeur singulière de A. $\square$

Version "pleine" (DVS2) : Soit A $m\times n$ de rang r. Alors

$$\boxed{A = P\Lambda Q^*\ .}$$

$P = [u^1|...|u^r|u^{r+1}|...|u^m] = [U|\tilde{U}]$ carrée $m\times m$ unitaire et $Q = [v^1|...|v^r|v^{r+1}|...|v^n] = [V|\tilde{V}]$ carrée $n\times n$ unitaire, ont leurs colonnes formées respectivement par les vecteurs propres de AA^* et de A^*A. Pour obtenir P (resp. Q) on complète les vecteurs colonnes de U (resp. V) de la DVS1 par les vecteurs colonnes de $\tilde{U}$ (resp $\tilde{V}$) qui sont les vecteurs propres de AA^* (resp A^*A) associés à la valeur propre multiple 0. On forme ainsi une base orthonormée de $I\!\!K^m$ (resp. $I\!\!K^n$) : $P^*P = PP^* = I_m$, $Q^*Q = QQ^* = I_n$. De plus

$$\Lambda = \begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{bmatrix}\ ,\ m\times n\ , \qquad \Lambda_r^{1/2} = \begin{bmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_r \end{bmatrix}\ ,\ r\times r. \quad \square$$

Remarque : Lorsque A est réelle, $A^* = A'$ est réelle ainsi que U, V, P et Q.
Preuve :

AA^* et AA^* sont hermitiennes (symétriques), semi-définies positives et ont mêmes valeurs propres non nulles. Par Schur, $\exists Q$ unitaire telle que
$$Q^*A^*AQ = \mathrm{diag}(\mu_1^2,...,\mu_n^2).$$
Posons $C = AQ$. Soit c^j la $j$ème colonne de C, alors $c^{i*}c^j = \mu_i^2\delta_{ij}$.

Comme $\mathrm{rang}(AQ) = r$, on peut supposer que $\mu_1,...,\mu_r$ sont strictement positifs et que $\mu_{r+1},...,\mu_n$ sont nuls.

Alors $c^j = 0_m$ pour $j = r+1,...,n$, car c'est un vecteur associé à une valeur propre nulle et donc $\|c^j\|_2^2 = 0$.
On construit P par colonnes :
$$p^j = \mu_j^{-1}c^j \text{ si } j = 1,...,r. \quad (*)$$
$$p^{i*}p^j = (\mu_i\mu_j)^{-1}c^{i*}c^j = \delta_{ij} \text{ pour } 1\le i,j\le r.$$
On complète la matrice P pour former une base orthonormée de $I\!\!K^m$, $P = [p^1\cdots p^m]$ et $\forall i,j\in\{1,...,m\}$ $p^{i*}p^j = \delta_{ij}$.

Vérifions DVS2 en calculant un élément de la matrice P^*AQ :

Si $1 \leq i,j \leq r$, $[P^*AQ]_{ij} = p^{i*}c^j = \mu_j \delta_{ij}$.
Si $j > r$, $[P^*AQ]_{ij} = 0$ car alors $c^j = 0_m$.
Si $j \leq r$ et si $i > r$, $p^{i*}c^j = \mu_j p^{i*}p^j = 0$.

Donc, $P^*AQ = \Lambda$.

Vérifions que les colonnes de P sont les vecteurs propres de AA^* :

$$
\begin{aligned}
AA^*P &= AI_nA^*P = AQQ^*A^*P \\
&= AQ\Lambda^* = P\Lambda\Lambda^* \\
&= P \left[\begin{array}{ccc|c} \mu_1^2 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & \mu_r^2 & 0 \\ \hline 0 & \cdots & 0 & 0 \end{array} \right] .
\end{aligned}
$$

Par construction les colonnes de Q sont vecteurs propres de A^*A. $\square$

Remarque : Dans la pratique :

- Le nombre de valeurs singulières fournit le rang de la matrice. Ces valeurs singulières sont ordonnées, $\mu_1 \geq \ldots \geq \mu_r > 0$, ce qui induit un ordre sur les colonnes de U et de V.

- Dans le calcul de U et de V, on ne calcule les vecteurs propres de AA^* ou de A^*A que pour celle de ces matrices de plus petite dimension, les vecteurs propres de l'autre se déduisent par des "formules de transition" (*) et (**). Par exemple, si $m \geq n$, on calcule les vecteurs propres $v^1, \ldots, v^r$ de A^*A, ceux de AA^* associés aux valeurs propres non nulles, sont donnés par

$$U = AV\Lambda_r^{-1/2}\,, \tag{$*$}$$

où $\Lambda_r^{-1/2} = (\Lambda_r^{1/2})^{-1} = diag(1/\mu_1, \ldots, 1/\mu_r)$. Dans l'autre cas, on utilise

$$V = A^*U\Lambda_r^{-1/2}\,. \tag{$**$}$$

Pour obtenir la DVS pleine il faut calculer les vecteurs propres associés aux valeurs propres nulles.

Corollaire 1 : La décomposition en valeurs singulières donne

$$A^*A = V\Lambda_r^2V^* = \sum_{i=1}^{r} \mu_i^2 v^i v^{i*},$$

$$AA^* = U\Lambda_r^2 U^* = \sum_{i=1}^{r} \mu_i^2 u^i u^{i*}. \quad \square$$

Cas particulier : DVS d'une matrice hermitienne semi-définie positive
$A^*A = AA^* = A^2$ et les valeurs singulières de A sont égales à ses valeurs propres. Alors, $U = V$ dans la DVS1 et la décomposition de Schur concide avec la DVS (modulo le signe des vecteurs propres).
Corollaire 2 : Soit A $m \times n$ avec rang$(A) = r$. La DVS2 de A s'écrit $A = P\Lambda Q^*$, avec $P = [U|\tilde{U}]$ et $Q = [V|\tilde{V}]$. Alors,

$$\boxed{\begin{aligned}&\mathrm{Ker}\, A = \mathrm{Ker}\, A^*A = \mathrm{Im}\, \tilde{V} = \mathcal{E}(v^{r+1}, ..., v^n),\\ &\mathrm{Im}\, A = \mathrm{Im}\, AA^* = \mathrm{Im}\, U = \mathcal{E}(u^1, ..., u^r),\\ &\mathrm{Im}\, A^* = \mathrm{Im}\, V = \mathcal{E}(v^1, ..., v^r) = \{\mathrm{Ker}\, A\}^\perp,\end{aligned}}$$

où $\mathcal{E}(x_i, \ldots, x_j)$ désigne l'espace vectoriel engendré par les vecteurs $x_i, ..., x_j$. $\square$

Preuve de la première égalité

$Ax = 0 \Leftrightarrow A^*Ax = 0$. Donc $\mathrm{Ker}\, A = \mathrm{Ker}\, A^*A$.

Comme $A = U\Lambda_r V^*$ d'après la DVS1, $x \in \mathrm{Ker}\ \mathrm{A} \Longleftrightarrow U\Lambda_r V^* x = 0$.

Comme $(AB = 0) \Longleftrightarrow (A^*AB = 0)$, on obtient :

$\Lambda_r U^* U \Lambda_r V^* x = 0 \Leftrightarrow \Lambda_r^2\ V^* x = 0 \Leftrightarrow V^* x = 0$, puisque Λ_r est inversible.

x orthogonal aux vecteurs colonnes de V, c'est-à-dire à $(v^1, ..., v^r)$, appartient à $\mathcal{E}(v^{r+1}, ..., v^n)$. $\square$

Corollaire 3 : Il y a d'importantes projections orthogonales associées à la décomposition en valeurs singulières. Soit A supposée de rang r et $A = U\Lambda_r^{1/2} V^* = P\Lambda Q^*$, la DVS de A. Rappelons les partitions colonnes de P et de Q

$$P = [U|\tilde{U}]\,, \qquad Q = [V|\tilde{V}].$$

Les résultats sur la projection orthogonale présentés dans l'exemple fondamental et qui seront démontrés au chapitre 4 doivent être associés au corollaire 2 pour donner

$$\boxed{\begin{array}{lcl} VV^* & = & \text{projection orthogonale sur } \{\text{Ker}\, A\}^{\perp} = \text{Im}\, A^* \\ \tilde{V}\tilde{V}^* & = & \text{projection orthogonale sur Ker}\, A \\ UU^* & = & \text{projection orthogonale sur Im}\, A \\ \tilde{U}\tilde{U}^* & = & \text{projection orthogonale sur } \{\text{Im}\, A\}^{\perp} = \text{Ker}\, A^* \end{array}}$$

Corollaire 4 : **Éléments propres de la matrice de projection sur Im** A
Proposition : Soit $A = U\Lambda_r^{1/2}V^* = P\Lambda Q^*$, la DVS de la matrice A de rang r. La matrice $P_A = UU^*$ est diagonalisable, ses vecteurs propres sont les colonnes de P et donc orthonormés, ses valeurs propres sont 0 ou 1. Les vecteurs propres associés à la valeur propre 1 sont les colonnes de U ceux associés à 0 sont les colonnes de $\tilde{U}$. De plus,

$$\text{trace}(P_A) = \text{rang }(P_A) = \text{rang }(A). \quad \square$$

Preuve : Bien sûr, $P_A = UU^*$ hermitienne est diagonalisable. Le fait que $P = [U|\tilde{U}]$ est carrée unitaire donne $P^* = P^{-1}$ et

$$P_A = P \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} P^{-1}.$$

En outre, $\text{trace}(P_A) = \text{trace}(I_r) = r$. $\square$

2.5.2 Décomposition polaire

Proposition : Soit A <u>carrée</u> $n \times n$, alors il existe une matrice carrée unitaire U et une matrice H hermitienne semi définie positive telles que

$$A = UH\,. \quad \square$$

Preuve : Si $A = P\Lambda Q^*$ est carrée $n \times n$, alors P, Λ et Q de la DVS2 sont carrées $n \times n$. On pose

$$A = P\Lambda Q^* = (PQ^*)\,(Q\Lambda Q^*) \doteq U\,H$$

avec $U = PQ^*$ et $H = Q\Lambda Q^*$. U est unitaire comme produit de matrices unitaires et H est par construction hermitienne semi définie positive. $\square$
Remarque : Le nom de "décomposition polaire" rappelle la représentation polaire d'un nombre complexe $z = \rho e^{i\theta}$ (ρ réel positif ou nul et $e^{i\theta}$ nombre complexe unitaire). Il y a analogie entre ces deux idées car les valeurs propres de H sont des nombres réels non négatifs, et les valeurs propres de U sont des nombres complexes unitaires. Pour une

matrice A normale, l'analogie est encore plus stricte : une valeur propre λ de A est de la forme $\lambda = \rho e^{i\theta}$ où ρ est une valeur propre de H et $e^{i\theta}$ est valeur propre de U. En effet, par Schur, pour A normale, de valeurs propres $\{\rho_j e^{i\theta_j}\}_j$, $\exists V$ unitaire telle que

$$\begin{aligned} A &= V \ \mathrm{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n}) \ \mathrm{diag}(\rho_1, \ldots, \rho_n) V^* \\ &= \left(V \mathrm{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n}) V^*\right) \left(V \ \mathrm{diag}(\rho_1, \ldots, \rho_n) \, V^*\right) = UH. \end{aligned}$$

2.6 Factorisation de Cholesky d'une matrice symétrique définie positive

Proposition : Soit A $n \times n$ réelle symétrique, définie positive. Il existe T triangulaire supérieure avec $T_{ii} > 0$ telle que

$$A = T'T\,.$$

Cette décomposition est unique. $\square$

Intérêt

A symétrique définie positive est souvent utilisée en statistique comme métrique sur $I\!R^n$ et définit ainsi le produit scalaire :

$$\forall x, y \in I\!R^n, < x, y >_A = y' A x.$$

La décomposition de Cholesky de A donne

$$< x, y >_A = < Tx, Ty >_{I_n}\ .$$

Changer de métrique revient donc à effectuer un transformation linéaire des données.

Cette transformation n'est pas unique : une autre transformation est obtenue par la factorisation

$$A = A^{1/2} A^{1/2}$$

où $A^{1/2}$ est la matrice hermitienne définie positive obtenue à partir de la décomposition en valeurs singulières de la matrice A hermitienne définie positive, ou par la décomposition de Schur, voir la propriété H4 des matrices hermitiennes. Cette décomposition donne

$$< x, y >_A = < A^{1/2} x, A^{1/2} y >_{I_n}\ .$$

2.7 Exercices

Exercice 1 : Soient θ un angle exprimé en radians et w dont l'expression dans la base canonique est

$$w = \begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix}.$$

5.1 Calculer en fonction de θ le vecteur $v \in \mathbb{R}^2$ déduit de w par une rotation de $\pi/2$.

5.2 Soit x un vecteur de $\mathbb{R}^2$ muni du produit scalaire usuel.
a) Calculer u, projection orthogonale de x sur $Im\, v$. Montrer que $u = vv'x$. On note $P = vv'$, la matrice de la projection orthogonale sur $Im\, v$. Vérifier les deux propriétés qui caractérisent un projecteur orthogonal. Calculer P en fonction de θ. Quel est le rang de P? Calculer les valeurs propres et les vecteurs propres de P.
b) Soit y le vecteur déduit de x par symétrie par rapport à la droite vectorielle $Im\, w = \{Im\, v\}^{\perp}$. Exprimer y en fonction de u et de x. On note Q la matrice telle que $y = Qx$. Exprimer Q en fonction de v et vérifier que

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

Quelles sont les propriétés de Q appelée matrice de réflexion par rapport à $\{Im\, v\}^{\perp}$?

Exercice 2 : Démonstration du théorème de Schur :
Soit $A \in \mathbb{C}^{n\times n}$, construire pour $n = 2$, une matrice U carrée unitaire telle que $T = U^*AU$ soit triangulaire supérieure dont les éléments diagonaux soient λ_1 et λ_2, valeurs propres de A. En supposant ce résultat vrai à l'ordre $n-1$ montrer qu'il est vrai à l'ordre n.

Exercice 3 : Montrer qu'une matrice triangulaire supérieure normale est diagonale.

Exercice 4 : La norme de Frobénius d'une matrice A étant définie par $\|A\|_F = (trace(A^*A))^{1/2}$, montrer que pour une matrice A carrée quelconque, on peut mesurer l'écart de A à la normalité grce à la norme de Frobénius de la partie hors diagonale N de la matrice T dans la décompsition de Schur de A,

$$\|N\|_F^2 = \|A\|_F^2 - \sum_i |\lambda_i|^2 \doteq \Delta^2(A).$$

Cette quantité ne dépend pas du choix de U, plus $\Delta^2(A)$ est proche de 0, plus A est près d'être normale. Dire pourquoi une autre mesure de l'écart de A à la normalité est $\|AA^* - A^*A\|_F^2$.

<u>Exercice 5</u> : Décomposer la matrice

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

en valeurs singulières. Dans l'espace vectoriel $I\!R^3$ muni de la base canonique, représenter $Im\, A$ et $\{Im\, A\}^\perp$. Donner l'expression des matrices de projection orthogonale sur $Im\, A$, $\{Im\, A\}^\perp$, $Im\, A'$ et $\{Im\, A'\}^\perp$.

<u>Exercice 6</u> : **Critère de Sylvester**

On se propose de montrer que pour A symétrique réelle,

$$\{A \text{ \textbf{définie positive}}\} \Longleftrightarrow \{\text{ \textbf{les mineurs principaux sont positifs} }\}.$$

Soit $A = [a_{ij}]$ une matrice réelle symétrique, $n \times n$, définie positive. On appelle $k$ième <u>matrice principale</u> de A, notée A_k, la matrice $k \times k$ obtenue en supprimant les $n - k$ dernières lignes et dernières colonnes de A. Remarquer que $A_n = A$. On appelle $k$ième <u>mineur principal</u> de A, le déterminant de A_k, noté dét(A_k).

6.1. a. Montrer que $a_{ii} > 0$ pour $i = 1, \ldots, n$.

b. Ecrire la décomposition spectrale de A. Calculer dét(A) en fonction des valeurs propres de A. Dire pourquoi dét$(A) > 0$ et A^{-1} est symétrique définie positive.

6.2. Soient B et P les matrices $(n+1) \times (n+1)$ définies par

$$B = \begin{bmatrix} A & b \\ b' & \alpha \end{bmatrix}, \qquad P = \begin{bmatrix} I_n & -A^{-1}b \\ 0' & 1 \end{bmatrix}.$$

a. Calculer $C = P'BP$ par blocs. Quel est le spectre de C? Montrer que

$$\text{dét}(C) = \text{dét}(B) = \text{dét}(A)\ (\alpha - b'A^{-1}b).$$

b. Montrer que dét$(B) \leq \alpha$ dét(A) et que légalité a lieu si et seulement si $b = 0$.

c. Montrer que B est définie positive si et seulement si C est définie positive.

d. Montrer que $\text{dét}(B) > 0$ si et seulement si C est définie positive.

e. Déduire des précédentes questions que $\text{dét}(B) > 0$ si et seulement si B est définie positive

6.3. Montrer que pour A symétrique, $\{A$ définie positive$\} \Longleftrightarrow \{\text{dét}(A_k) > 0\ ,\ k = 1, \ldots n\}$.

a. Sens $\Longrightarrow$: on note E_k la matrice $k \times n$, définie par blocs par $E_k = [I_k\ 0_{k\times(n-k)}]$, où I_k est la matrice identité d'ordre k. Calculer A_k en fonction de A (supposée définie positive) et de E_k. En déduire que toute matrice A_k est définie positive et donc que tout mineur principal est positif.

b. Sens $\Longleftarrow$: exprimer les blocs de A_{k+1} en fonction de A_k, $a_{k+1\,k+1}$ et d'un vecteur b_{k+1} à déterminer. Vérifier que A_1 est symétrique définie positive. Montrer par récurrence que toutes les matrices principales sont définies positives.

<u>Exercice 7</u> : Soit la matrice

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

et le vecteur

$$y = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

1. Calculer la décomposition en valeurs singulières (dvs) maigre de X (les valeurs singulières de X étant classées par ordre décroissant). Puis la dvs pleine.

2. Quel est le rang de X? Expliciter le noyau de X. Ecrire la matrice de la projection orthogonale sur $Ker\ X$, puis le vecteur projection de y sur $Ker\ X$.

3. Ecrire l'expression analytique du plan vectoriel $Im\ X$, l'expression de la matrice de la projection orthogonale sur $Im\ X$, puis le vecteur projection de y sur $Im\ X$.

Chapter 3

Normes de Matrices

En Analyse Numérique, en Statistique et dans d'autres disciplines, les inconnues de certains problèmes sont des vecteurs ou des matrices. De nombreuses méthodes consistent à construire un élément d'un sous espace vectoriel le plus proche d'un autre élément qui lui, est connu. Il est donc nécessaire de savoir mesurer des distances entre vecteurs ou entre matrices. Pour cela on peut définir une norme sur l'espace vectoriel auquel ils appartiennent. Les deux opérations suivantes vont permettre de définir des normes Euclidiennes sur l'espace vectoriel des matrices.

3.1 Normes de vecteurs

Soit E un $I\!K$-espace vectoriel ($I\!K$ est $I\!R$ ou $\mathbb{C}$).

Une norme sur E est une application $\|.\|$ de E dans $I\!R^+$ telle que

$$\|x\| = 0 \Longleftrightarrow x = 0_E$$

$$\|\lambda x\| = |\lambda| \, \|x\| \qquad x \in E, \lambda \in I\!K$$

$$\|x + y\| \leq \|x\| + \|y\| \qquad x, y \in E\,.$$

Une norme définit une distance entre deux éléments de E par

$$d(x, y) = \|x - y\|.$$

Ainsi, la norme de x est la distance de x au vecteur nul

$$d(x, 0_E) = \|x\|.$$

3.1.1 Normes de Hölder

Lorsque E est de dimension finie, $E = \mathbb{K}^n$, une norme de Hölder ou p-norme ($p \geq 1$) est définie par

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} .$$

Les deux premiers axiomes sont évidents, on montre le troisième par une inégalité dite de Hölder

$$\left| \sum_{i=1}^{n} x_i y_i \right| \leq \|x\|_p \|y\|_q \qquad\qquad \frac{1}{p} + \frac{1}{q} = 1\,.$$

Un cas particulier important est celui où $p = q = 2$ qui donne l'inégalité de Cauchy-Schwarz.
Les 1, 2, ∞ normes sont les plus usuelles

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$$

$$\|x\|_\infty = \max_{i=1,n} |x_i|\,.$$

La 2-norme est dite euclidienne car elle est associée à un produit scalaire $< .,. >$ sur E,

$$\|x\|_2 = \sqrt{< x, x >}\,.$$

La boule unité $B_p = \{x \in E \mid \|x\|_p \leq 1\}$ est un compact (fermé, borné) de E dont la frontière $S_p = \{x \in E \mid \|x\|_p = 1\}$ est la sphère unité. La boule unité ouverte est l'ensemble des éléments de B_p qui n'appartiennent pas à S_p ($\leq$ est remplacé par $<$ dans B_p). Lorsque $E = \mathbb{R}^2$, la figure ci dessous représente les sphères unités S_1, S_2 et S_∞.

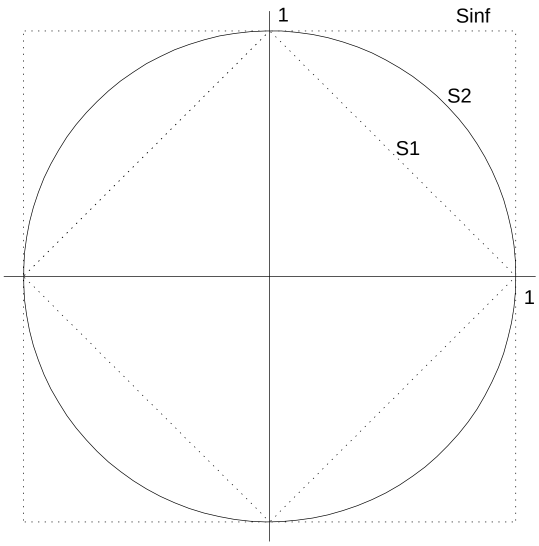


Figure 3 : sphère unité pour les p-normes usuelles, $p = 1$, $p = 2$ et $p = +\infty$.

Toutes les normes de $I\!K^n$ sont équivalentes, c'est à dire si $\| \ . \ \|_\alpha$ et $\| \ . \ \|_\beta$ sont deux normes il existe des constantes positives c_1 et c_2 telles que

$$c_1 \|x\|_\alpha \leq \|x\|_\beta \leq c_2 \|x\|_\alpha \, .$$

La relation précédente s'écrit

$$n^{(\frac{1}{p} - \frac{1}{q})} \|x\|_q \leq \|x\|_p \leq \|x\|_q \qquad p > q \geq 1 \, .$$

Supposons que $\widehat{x}$ est une approximation de x, pour une norme donnée on dit que

$$\varepsilon_{abs} = \|x - \widehat{x}\|$$

est l'erreur absolue tandis que, si $x \neq 0$

$$\varepsilon_{rel} = \frac{\|x - \widehat{x}\|}{\|x\|}$$

est l'erreur relative. En particulier, si

$$\frac{\|x - \widehat{x}\|_\infty}{\|x\|_\infty} \approx 10^{-p}$$

alors la plus grande composante a approximativement p décimales correctes. On dit qu'une suite $\{x^k\}$ de vecteurs converge vers x si

$$\lim_{k \to \infty} \|x^k - x\| = 0 \, .$$

Les normes sur $I\!K^n$ étant équivalentes, la notion de convergence ne dépend pas de la norme choisie.

3.1.2 Généralisation de la norme Euclidienne, la M-norme

Il est possible d'utiliser des produits scalaires définis à partir de formes hermitiennes définies positives, qui sont plus généraux que le produit scalaire usuel présenté dans la section 1.10. En statistique en particulier, de tels produits scalaires sont utilisés pour définir des normes, des notions d'orthogonalité et de projection plus adaptées aux problèmes étudiés.

En dimension finie $E = \mathbb{C}^n$ (ou $\mathbb{R}^n$), la donnée d'une matrice M <u>hermitienne (symétrique) définie positive</u>, appelée <u>une métrique</u>, permet de définir un produit scalaire

$$< x, y >_M = y^* M x \ = < Mx, y >_{I_n} \ .$$

La norme vectorielle euclidienne sur (E, M) est définie par

$$\|x\|_M = \sqrt{< x, x >_M} = \sqrt{x^* M x}\,.$$

Remarques :

R1 : Le produit scalaire usuel et la 2-norme vectorielle associée définis dans la section 1.10 correspondent à $M = I_n$ c'est à dire à la métrique définie par la matrice identité.

R2 : $(\mathbb{R}^n, M)$ est aussi l'espace affine d'origine o attaché à l'espace vectoriel $(\mathbb{R}^n, M)$. On définit pour deux points m_1 et m_2, $\overrightarrow{om_1} = x$, $\overrightarrow{om_2} = y$, et la M-distance entre m_1 et m_2 par $d_M(m_1, m_2)^2 = (y - x)' M (y - x)$.

La sphère unité $S_M = \{x \in E \,|\, x^* M x = 1\}$ est l'ensemble de niveau 1 de la forme quadratique définie positive $x \to \|x\|_M^2 = x^* M x$.
Par exemple, la figure suivante montre la sphère unité de $(\mathbb{R}^2, M = \frac{1}{4}\begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 7 \end{bmatrix})$ qui est une ellipse dont les axes principaux sont les vecteurs propres de M.

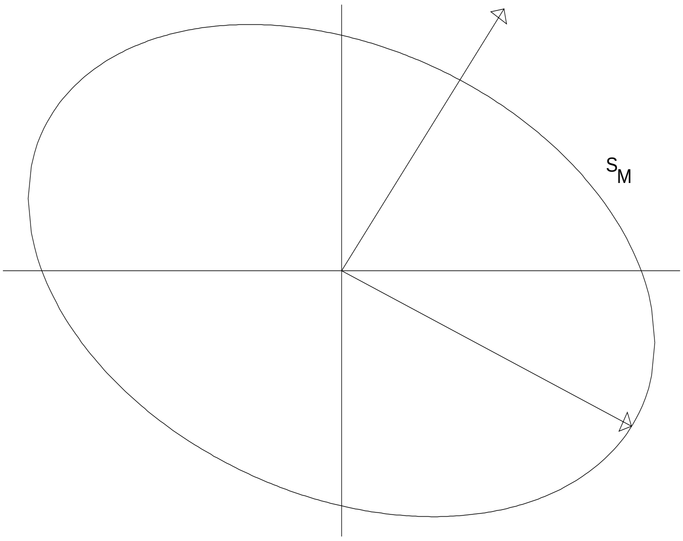


Figure 4 : sphère unité pour l'espace ($I\!R^2$, M).

Dans le cas $n = 2$ réel, la ligne de niveau k d'une forme quadratique définie positive est donc une ellipse de centre l'origine des coordonnées. Si la forme quadratique est semi définie positive (rang $M = 1$), l'ensemble de niveau k est formé de 2 droites parallèles à $Ker\, M$. Le graphe $G = \{(x, y = x'Mx) \in I\!R^3 \mid x \in I\!R^2\}$ est représenté selon le cas, par :

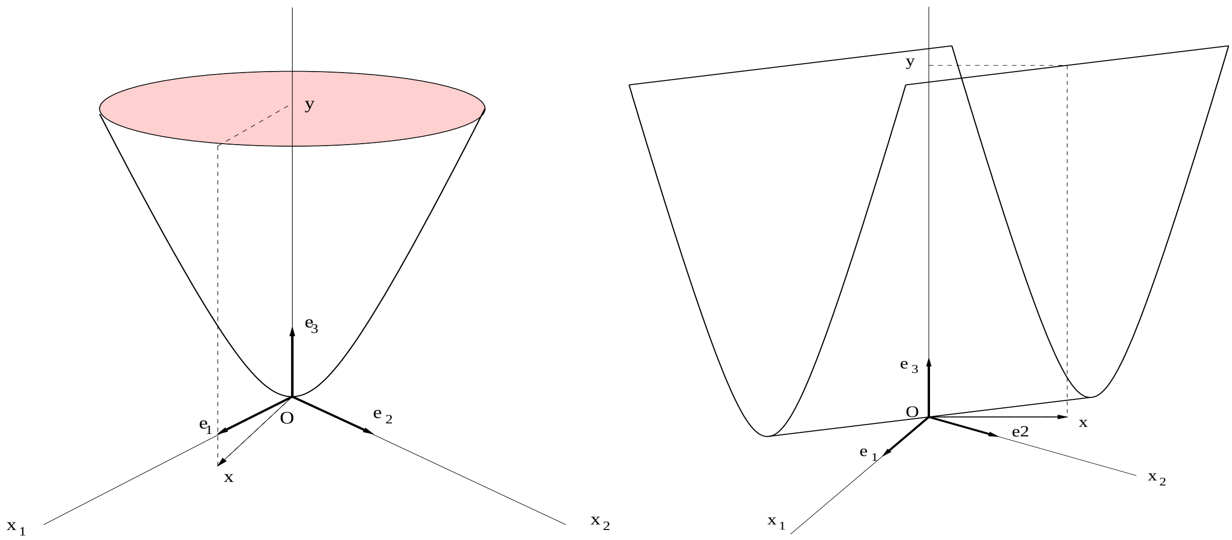


Figure 5 : graphe d'une forme quadratique définie positive puis semi définie positive de ($I\!R^2$, M).

L'orthogonalité au sens de M ou M-orthogonalité entre deux vecteurs s'écrit

$$< x, y >_M = y^* M x = 0 \, .$$

3.2 Normes de matrices

L'analyse d'algorithmes matriciels nécessite souvent l'usage de normes matricielles. Par exemple, telle méthode de résolution de systèmes linéaires pourra être peu performante si la matrice des coefficients est proche de la singularité. D'autre part en statistique exploratoire de données multivariées, il arrive que l'on désire approcher une matrice X par une matrice $\widehat{X}$ de rang plus petit. On désire donc quantifier la notion de distance sur l'espace vectoriel des matrices de dimension $n \times p$ noté $\mathbb{K}^{n\times p}$.

Il ya deux façons de construire une norme de matrice: la "**subordination**" à une norme vectorielle ou la "**vectorisation**" du tableau de nombres. La première fait appel à l'application linéaire associée à la matrice. La seconde consiste à identifier les deux espaces vectoriels $\mathbb{K}^{n\times p}$ et $\mathbb{K}^{np}$ grâce à l'opérateur *vec*. Dans cette dernière approche, nous n'étudierons que les normes Euclidiennes.

3.2.1 Normes subordonnées à des normes vectorielles

Généralités

Soient $(E, \|.\|_E)$ et $(F, \|.\|_F)$ deux $\mathbb{K}$ espaces vectoriels normés de dimension quelconque pas forcément finie. Soit f une application linéaire de E dans F

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \qquad \alpha, \beta \in \mathbb{K},\ x, y \in E\,.$$

Il est de plus grand intérêt de comparer $\|x\|_E$ et $\|f(x)\|_F$.

Supposons $x \neq 0_E$, et construisons le rapport

$$r(x) = \frac{\|f(x)\|_F}{\|x\|_E}.$$

Ce rapport est invariant par homothétie, $r(\lambda x) = r(x)$, $\lambda \in \mathbb{K} - \{0\}$.

Proposition : Pour que l'application linéaire f de E dans F soit continue, il faut et il suffit que $r(x)$ soit majoré. $\square$

Proposition : Sur $\mathcal{L}(E, \|.\|_E; F, \|.\|_F)$ espace vectoriel des applications linéaires continues de $(E, \|.\|_E)$ dans $(F, \|.\|_F)$,

$$\|f\|_{E,F} = \sup_{x \neq 0_E} \frac{\|f(x)\|_F}{\|x\|_E} = \sup_{\|x\|_E = 1} \|f(x)\|_F$$

est une norme dite subordonnée aux normes vectorielles $\|.\|_E$ et $\|.\|_F$. Comme corollaire,

$$\|f(x)\|_F \leq \|f\|_{E,F}\,\|x\|_E \qquad x \in E\,. \quad \square$$

Normes matricielles subordonnées

Soient $E = I\!\!K^n$ et $F = I\!\!K^m$ munis de la même p-norme $\|.\|_p$ pour simplifier. A chaque choix d'une base $\mathcal{E} = \{e_j\}_{j=1,n}$ pour E et d'une base $\mathcal{F} = \{f_i\}_{i=1,m}$ pour F, l'application linéaire f est caractérisée par la matrice $A = [a_{i\,j}]$ de dimensions $m \times n$ telle que

$$f(e_j) = \sum_{i=1}^{m} a_{ij}\, f_i\,.$$

Les éléments de la $j$ième colonne de A sont les composantes du vecteur $f(e_j)$ dans $\mathcal{F}$. Les composantes du vecteur $y = f(x)$ dans la base $\mathcal{F}$ sont calculées à partir des composantes de x dans la base $\mathcal{E}$ par le produit matriciel

$$y = Ax\,.$$

Toute application linéaire f de $I\!\!K^n$ dans $I\!\!K^m$ munis de n'importe quelle norme, étant continue, on définit

$$\|A\|_p = \sup_{x \neq 0_E} \frac{\|Ax\|_p}{\|x\|_p} = \sup_{\|x\|_p = 1} \|Ax\|_p = \|A\widehat{x}\|_p$$

pour un $\widehat{x}$ de la sphère unité de $I\!\!K^n$. On obtient la relation

$$\|Ax\|_p \leq \|A\|_p\,\|x\|_p \qquad x \in I\!\!K^n\,.$$

Remarquons que la matrice identité est de p-norme égale à 1 quelle que soit la valeur de p. On verra qu'il existe d'autres normes matricielles qui donnent des valeurs différentes. Le calcul des normes $p = 1$ et $p = \infty$ est très simple

$$\|A\|_1 = \max_{j=1\ldots n} \sum_{i=1}^{m} |a_{ij}|$$

$$\|A\|_\infty = \max_{i=1\ldots m} \sum_{j=1}^{n} |a_{ij}|\,.$$

On vérifie immédiatement que $\|A\|_\infty = \|A'\|_1$. Le calcul de la 2-norme n'est pas aussi explicite. Son calcul nécessite l'usage des propriétés des matrices hermitiennes.

La 2-norme subordonnée

Pour $A \in \mathbb{C}^{m\times n}$, il s'agit de maximiser la quantité

$$r(x) = \frac{\|Ax\|_2}{\|x\|_2}\,.$$

Un vecteur $x \in \mathbb{C}^n$ se décompose sur la base orthonormée $\{v^i\}$ formée par les vecteurs propres de A^*A <u>ordonnés par ordre décroissant des valeurs propres</u>, $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$.

$$x = \sum_{i=1}^{n} x_i v^i \qquad \text{et} \qquad x^* = \sum_{j=1}^{n} \overline{x}_j v^{j*}\,.$$

L'orthonormalité de la base choisie conduit à

$$\lambda_n \leq r(x)^2 = \frac{\sum_i \lambda_i |x_i|^2}{\sum_i |x_i|^2} \leq \lambda_1\,.$$

Le maximum est réalisé pour $x = v^1$, premier vecteur propre.

Pour résumer

$$\|A\|_2 = \mu_1 = \sqrt{\rho(A^*A)}\,,$$

où μ_1 est la plus grande valeur singulière de A et $\rho(A^*A)$ est le rayon spectral de A^*A.

Si de plus, <u>A est hermitienne</u>, alors

$$\|A\|_2 = \rho(A)\,.$$

3.2.2 Normes Euclidiennes par vectorisation

L'identification de $\mathbb{K}^{m\times n}$ avec $\mathbb{K}^{mn}$ grâce à l'opérateur *vec*, permet de définir par vectorisation, des normes matricielles associées aux p-normes de Hölder. Cependant, nous ne présenterons que le cas $p = 2$, Euclidien, qui conduit à la norme de Frobénius.

Norme de Frobénius

On a vu précédemment dans la propriété P2 de l'opérateur *vec*, que pour deux matrices X et Y de $\mathbb{C}^{m\times n}$,

$$\text{trace}(Y^*X) = \textstyle\sum_{j=1}^{n}\sum_{i=1}^{m} \overline{y}_{ij}x_{ij} \; = vec^*(Y)vec(X) =< vec(X), vec(Y) >,$$

ce qui définit le produit scalaire usuel sur $\mathcal{M}_{np}(\mathbb{C})$ par

$$< X, Y >= trace(Y^*X).$$

La norme de Frobénius de $A \in \mathbb{C}^{m\times n}$ est définie par

$$\|A\|_F = \sqrt{\text{trace}(A^*A)} = \left(\sum_{j=1}^{n}\sum_{i=1}^{m}|a_{ij}|^2\right)^{1/2} = \left(\sum_{j=1}^{n}\|A^j\|_2^2\right)^{1/2},$$

où A^j est la $j$ième colonne de A.

On vérifie que cette norme Euclidienne n'est pas subordonnée à une norme vectorielle car $\|I_n\|_F = \sqrt{n}$ alors que le résultat est 1 pour toute p-norme subordonnée.
L'inégalité de Cauchy-Schwarz s'écrit:

$$[\text{trace}(Y^*X)]^2 \leq \text{trace}(X^*X)\text{trace}(Y^*Y);\ \text{égalité si } Y = kX.$$

La propriété P4 de la trace permet d'écrire, si $rang(A) = r$,

$$\|A\|_F^2 = \sum_{i=1}^{n}\lambda_i(A^*A) = \sum_{i=1}^{r}\mu_i^2\,.$$

Il en résulte que

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2\,.$$

Comme conséquence, il vient

$$\|Ax\|_2 \leq \|A\|_2\|x\|_2 \leq \|A\|_F\|x\|_2\,.$$

On dit alors que la norme de Frobénius et la norme vectorielle euclidienne sont compatibles.

Norme de Hilbert-Schmidt

En analyse statistique des données on est ammené à définir des métriques sur chacun des espaces vectoriels associés aux lignes et aux colonnes d'une matrice réelle.

On définit le triplet $(\mathbb{R}^{n\times p}, M, D)$ par la donnée de
$\mathbb{R}^{n\times p}$ espace vectoriel des matrices réelles $n \times p$,
M métrique Euclidienne sur l'espace $\mathbb{R}^p$ des lignes,
D métrique Euclidienne sur l'espace $\mathbb{R}^n$ des colonnes.

D'après P11, $M \otimes D$ est une métrique sur $\mathbb{R}^{np}$ et d'autre part,

$$\text{trace}(Y'DXM) = vec'(Y)(M \otimes D)vec(X) =< vec(X), vec(Y) >_{M\otimes D}.$$

Le produit scalaire de Hilbert-Schmidt associé au triplet $(\mathbb{R}^{n\times p}, M, D)$ est obtenu par identification de ce triplet à l'e.v. Euclidien $(\mathbb{R}^{np},\ M \otimes D)$. Il est défini par

$$< X, Y >_{M\otimes D} = \text{trace}(Y'DXM).$$

La norme de Hilbert-Schmidt associée est alors

$$\|X\|_{M\otimes D} = \sqrt{\text{trace}(X'DXM)}.$$

De façon symétrique, $\|X\|_{M\otimes D} = \sqrt{\text{trace}(XMX'D)} = \|X'\|_{D\otimes M}$.

Cas particuliers

- $(\mathbb{R}^{n\times p},\ M = I_p,\ D = I_n)$. Alors, $M \otimes D = I_{np}$.
  On retrouve les produits scalaires usuels et la norme de Frobénius de X,
  $$\|X\|^2_{I_p\otimes I_n} = \text{trace}(X'X) = \|X\|^2_F.$$
- $(\mathbb{R}^{n\times p},\ M = I_p,\ D)$. La structure de $M \otimes D$ est bloc-diagonale, chaque bloc étant D.
  Dans ce cas, le carré de la norme de X s'exprime comme la somme des carrés des D-normes des colonnes X^i,
  $$\|X\|^2_{I_p\otimes D} = \text{trace}(X'DX) = \sum_{i=1}^{p} \|X^i\|^2_D.$$
- $(\mathcal{E}_S,\ D,\ D)$, où $\mathcal{E}_S$ est le s.e.v. de $\mathbb{R}^{n\times n}$ formé des matrices symétriques. Alors,
  $$\|X\|^2_{D\otimes D} = \text{trace}((XD)^2).$$

Remarque

$$\|X\|^2_{M\otimes D} = \|D^{1/2}XM^{1/2}\|^2_F$$

On retrouve la norme Euclidienne usuelle des matrices par transformation des lignes et des colonnes.

3.2.3 Normes matricielles sous multiplicatives

Les normes matricielles considérées ont été déduites des normes vectorielles soit par subordination soit par vectorisation. Ces normes sont construites sur l'espace vectoriel des matrices rectangulaires, c'est à dire que seules les opérations + et multiplication par un scalaire sont définies. Cependant une autre opération joue un rôle fondamental: la multiplication entre matrices. Que peut-on dire de la norme d'un produit AB? La définition et la proposition suivantes permettent de répondre à cette question.

Une norme matricielle est dite sous multiplicative si
$$\|AB\| \leq \|A\|\|B\| \qquad \forall A \in \mathbb{C}^{m\times n}, \forall B \in \mathbb{C}^{n\times p}.$$

Proposition : La norme de Frobénius ainsi que toute p-norme sont des normes sous multiplicatives. $\square$ (Preuve en exercice).
Remarquons que l'on peut construire des contre-exemples. Ainsi la norme non subordonnée $\|A\|_\Delta = \max_{i,j} |a_{ij}|$ (vérifier que c'est une norme obtenue par vectorisation) appliquée à
$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$
donne $\|AB\|_\Delta > \|A\|_\Delta \|B\|_\Delta$.
Proposition : A toute norme matricielle sous multiplicative on peut toujours associer une norme vectorielle qui lui soit compatible. $\square$

Preuve : Soit $\|A\|$ une norme matricielle de la matrice A. Pour une matrice B telle que le produit AB existe on a $\|AB\| \leq \|A\|\|B\|$. A un vecteur x associons la matrice $X = [x, 0, \ldots, 0]$ où les colonnes sont nulles sauf la première égale à x. Il est clair que $\|x\| = \|X\|$ est une norme vectorielle. Donc, $\|Ax\| = \|[Ax, 0, \ldots, 0]\| = \|AX\| \leq \|A\|\|X\| = \|A\|\|x\|$. $\square$
Proposition : Quelle que soit la norme matricielle $\|.\|$ sous multiplicative, pour une matrice carrée A on a
$$\rho(A) \leq \|A\|. \quad \square$$
Preuve : Pour le vecteur $Au = \lambda u$, on choisit une norme vectorielle compatible et
$$|\lambda|\,\|u\| = \|\lambda\, u\| = \|Au\| \leq \|A\|\|u\|.$$
Quelle que soit la valeur propre, $|\lambda| \leq \|A\|$, ce qui achève la preuve $\square$.
Conséquence : Si <u>A est hermitienne</u>, la 2-norme est la plus petite des normes de A sous multiplicatives.

3.2.4 Normes unitairement invariantes

Dans la section 1.11 la notion de matrice unitaire (orthogonale, dans le cas réel) a déjà été exposée dans le cadre du produit scalaire usuel. Si l'on dispose d'un M-produit scalaire, la définition de matrices M-unitaires est immédiate.

Une matrice U de $\mathbb{C}^{m\times n}$ $(m \geq n)$, $\mathbb{C}^m$ étant muni de la métrique euclidienne M, est M-unitaire si les colonnes de U sont des vecteurs 2 à 2 M-orthogonaux et de M-norme unité, c'est à dire si

$$U^* M U = I_n.$$

Lorsque U et M sont à coefficients réels ont dit que U est M-orthogonale, dans ce cas

$$U' M U = I_n.$$

Une matrice est unitaire (orthogonale) si $M = I_m$.

Les normes Euclidiennes ne changent pas lorsque la matrice concernée est multipliée par une matrice orthogonale ou unitaire.

Normes unitairement invariantes :

Voici la suite des propriétés P1, P2 et P3 concernant les matrices unitaires définies dans la section 1.11 :

P4 : La norme vectorielle Euclidienne est unitairement invariante

$$\|Ux\|_2 = \|x\|_2, \qquad \forall U \text{ unitaire.}$$

P5 : Pour les M- normes vectorielles, on obtient

$$\|Ux\|_M = \|x\|_2, \qquad \forall U\ M\text{-unitaire.}$$

P6 : La 2-norme et la norme de Frobénius sont des normes matricielles invariantes par transformation unitaire.

$$\|UAV\| = \|A\|, \qquad \forall U, V \text{ unitaires, } V \text{ carrée.}$$

3.3 Suites de matrices

Comme pour les suites de vecteurs, en considérant l'ensemble $\mathbb{C}^{m\times n}$ comme un espace vectoriel de dimension mn, la convergence d'une suite de matrices est indépendante de la norme choisie. Elle équivaut à la convergence des mn suites de scalaires formées par les éléments des matrices. Le théorème suivant donne les conditions nécessaires et suffisantes pour que la suite des puissances d'une matrice carrée donnée converge vers la matrice nulle. Le deuxième théorème concerne la convergence de la série géométrique de matrices carrées.

Proposition : Soit B une matrice carrée. Les conditions suivantes sont équivalentes.

(i) $\lim_{k \to \infty} B^k = 0$.

(ii) $\lim_{k \to \infty} B^k v = 0$ pour tout vecteur v.

(iii) $\rho(B) < 1$.

(iv) $\|B\| < 1$ pour au moins une norme matricielle subordonnée $\|.\|$. $\square$

Proposition : La série $I + B + B^2 + \ldots$ converge vers $(I - B)^{-1}$ ssi

$$\rho(B) < 1 \,. \quad \square$$

3.4 Conditionnement d'une matrice

En analyse numérique matricielle ainsi qu'en statistique, les données sont généralement entachées d'erreurs et parfois une légère perturbation sur les données peut entrainer une grande perturbation sur la solution du problème considéré.
Prenons l'exemple dû à R.S. Wilson du système linéaire

$$\begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 32 \\ 23 \\ 33 \\ 31 \end{bmatrix} \qquad \text{de solution} \qquad \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Si l'on perturbe légèrement le second membre b d'une erreur relative de l'ordre de 1/200 pour obtenir $b = [32.1, 22.9, 33.1, 30.9]'$, alors, la solution du nouveau système devient $x = [9.2, -12.6, 4.5, -1.1]'$!! Bien que la matrice A du système soit inversible ($\text{dét}(A) = 1$), c'est à dire bien que ses vecteurs colonnes soient linéairement indépendants, cependant, les colonnes de A sont deux à deux presque colinéaires. Le statisticien mesure cette notion par la proximité à 1 ou à -1 du coefficient de corrélation linéaire r. Le coefficient r entre A^1 et A^2 est de 0.985. Il est de 0.90 entre A^3 et A^4. Cette colinéarité quasi parfaite confère à la solution du système linéaire une forte instabilité numérique vis à vis de petites perturbations des données. Cette notion est mesurée numériquement par

le conditionnement d'une matrice inversible :
soit $\|.\|$ une norme matricielle, le conditionnement de A régulière associé à cette norme, est le nombre

$$\text{cond}(A) = \|A\| \|A^{-1}\|.$$

Propriétés :

P1 : $\text{cond}(A^{-1}) = \text{cond}(A)$ et $\text{cond}(\alpha A) = \text{cond}(A)$.

P2 : $\text{cond}(A) \geq 1$ si le conditionnement est calculé pour une norme sous multiplicative.

P3 : $\text{cond}_2(A) = \dfrac{\mu_1}{\mu_n}$ où μ_1 et μ_n sont respectivement la plus grande et la plus petite des valeurs singulières de A.

P4 : $\text{cond}_2(A) = 1$ ssi $A = \alpha Q$ où α est un scalaire et Q est une matrice unitaire.

Dans les propriétés 3 et 4 on note cond_p le conditionnement associé à la p-norme matricielle.

Remarque : on dira qu'une matrice est bien conditionnée si son conditionnement n'est pas beaucoup plus grand que 1. La propriété 4 montre que les matrices unitaires sont les mieux conditionnées possibles. En analyse numérique comme en statistique on utilisera le plus possible des systèmes de vecteurs orthonormés.

Proposition : Soit A une matrice inversible. Soient x et $x + \Delta x$ les solutions des systèmes linéaires

$$Ax = b \qquad \text{et} \qquad A(x + \Delta x) = b + \Delta b\,.$$

On a

$$\frac{\|\Delta x\|}{\|x\|} \leq \text{cond}(A)\frac{\|\Delta b\|}{\|b\|}\,. \quad \square$$

Proposition : Soit A une matrice inversible. Soient x et $x+\Delta x$ les solutions des systèmes linéaires

$$Ax = b \qquad \text{et} \qquad (A + \Delta A)(x + \Delta x) = b\,.$$

On a

$$\frac{\|\Delta x\|}{\|x + \Delta x\|} \leq \text{cond}(A)\frac{\|\Delta A\|}{\|A\|}\,. \quad \square$$

3.5 Exercices

Exercice 1 : Soit $E = \mathbb{R}^2$ muni de la métrique euclidienne M dans la base canonique $\{e_1, e_2\}$

$$M = \frac{1}{4}\begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 7 \end{bmatrix}.$$

Vérifier que M est une métrique euclidienne. Calculer les axes principaux de la sphère unité

$$B_M = \{x \in E \,|\, x'Mx = 1\}\,.$$

Indications : Faire un changement de repère de telle faon que dans le nouveau repère $\{u_1, u_2\}$, l'équation de B_M soit de la forme

$$\frac{y_1^2}{a^2} + \frac{y_2^2}{b^2} = 1\,.$$

On poura aussi bien poser le problème sous forme d'un problème d'optimisation du carré de la norme euclidienne usuelle avec contrainte de type égalité.

Exercice 2 : Montrer que la norme de Frobénius est sous multiplicative ainsi que toute p norme subordonnée.

Chapter 4

Inverses Généralisés, Projecteurs M-Orthogonaux

Lorsque les colonnes (lignes) d'une matrice carrée A sont linéairement indépendantes, l'inverse A^{-1} de A est définie par $A^{-1}A = AA^{-1} = I$. Si A est inversible, alors le système linéaire

$$Ax = y \qquad (*)$$

dit de Cramer, admet une solution unique $\widehat{x} = A^{-1}y$. Même lorsque A n'est pas carrée, le système (*) admet une solution pas forcément unique, si y appartient à l'espace vectoriel $Im\, A$ engendré par les colonnes de A. La notion d'inverse généralisé d'une matrice rectangulaire permet de construire de telles solutions. On supposera maitenant que toutes les matrices sont réelles car résoudre un système linéaire sur $\mathbb{C}$ revient à résoudre deux sytèmes linéaires sur $\mathbb{R}$ associés aux parties réelles et imaginaires.

4.1 Inverses Généralisés

4.1.1 Définition et propriétés

Soit A une matrice $n \times p$, A^- inverse généralisé de A, est une matrice $p \times n$ telle que
Définition 1 : $\widehat{x} = A^-y$ est solution de $Ax = y$, $\forall\, y \in ImA$.
Définition 2 : $AA^-A = A$.

Equivalence des définitions

Définition 2 $\Rightarrow$ Définition 1	Définition 1 $\Rightarrow$ Définition 2
$AA^-A = A \Rightarrow AA^-Ax = Ax$	Soit $y = A^i$ une colonne de A , $y \in ImA$
$AA^-y = y$, $\forall\, y = Ax$ (*)	$AA^-A^i = A^i\,, \qquad \forall i$
A^-y est sol. de (*) $\quad\square$	$AA^-A = A. \quad \square$

Construction et non unicité d'un inverse

Soit A une matrice $n \times p$ telle que $rang(A) = r$. D'après la DVS 2, $\exists P, Q$ orthogonales telles que $PP' = P'P = I_n$, $QQ' = Q'Q = I_p$ et

$$A = P \begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{bmatrix} Q' \ .$$

$$AXA = A \Longleftrightarrow P \begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{bmatrix} Q'XP \begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{bmatrix} Q' = P \begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{bmatrix} Q'.$$

Multipliant à gauche et à droite par P' et Q respectivement,

$$\begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{bmatrix} Q'XP \begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{bmatrix}.$$

En décomposant par blocs, $Q'XP = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$, $AXA = A \ \Leftrightarrow T_{11} = \Lambda_r^{-1/2}$ et T_{12}, T_{21}, T_{22} arbitraires. Alors,

$$\boxed{A^- = Q \begin{bmatrix} \Lambda_r^{-1/2} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} P' \ ,}$$

avec T_{12}, T_{21}, T_{22} arbitraires et on n'a pas l'unicité de l'inverse. $\quad\square$

Quelques propriétés de $\mathbf{A^-}$

P1 : $rang(A^-) \geq rang(A)$.

P2 : Si A est carrée inversible alors $A^- = A^{-1}$.

P3 : $AA^- = P \begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{bmatrix} Q'Q \begin{bmatrix} \Lambda_r^{-1/2} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} P' = P \begin{bmatrix} I_r & \Lambda_r^{1/2}T_{12} \\ 0 & 0 \end{bmatrix} P'$,

et AA^- est idempotente car $(AA^-)^2 = AA^-$.

P4 : $A^-A = Q \begin{bmatrix} I_r & 0 \\ T_{21}\Lambda_r^{1/2} & 0 \end{bmatrix} Q'$ et A^-A est idempotente.

P5 : $rang(A) = rang(AA^-) \ = rang(A^-A)$.

P6 : $A = A(A'A)^-A'A$.

En effet, d'après la définition 2, $A'A(A'A)^-A'A = A'A$.

On conclut en utilisant $A'AB = A'AC \Leftrightarrow AB = AC$.

P7 : Le projecteur orthogonal sur $Im\,A$

$P_A = A(A'A)^-A'$, matrice symétrique idempotente est la matrice de projection orthogonale sur $Im\,A$, invariante quel que soit l'inverse généralisé utilisé. Soit $A = U\Lambda_r^{1/2}V' = P\Lambda Q'$ la DVS de A, alors

$$P_A = A(A'A)^-A' = UU' = P \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} P^{-1}. \qquad \square$$

$P_A^2 = A(A'A)^-A'A(A'A)^-A' = A(A'A)^-A' = P_A$, d'après P6.

On sait que $P_AP_A' = A(A'A)^-A'A[(A'A)^-]'A' = P_A'$ d'après P6.

En transposant $P_A = P_AP_A' = P_A'$ et P_A est symétrique.

Soient $P_1 = A(A'A)_1^-A'$, $P_1A = A$,

et $P_2 = A(A'A)_2^-A'$, $P_2A = A$.

Donc $P_1A = P_2A \Leftrightarrow A'P_1 = A'P_2 \Leftrightarrow A'A(A'A)_1^-A' = A'A(A'A)_2^-A'$

et P_A est invariant quelque soit l'inverse généralisé utilisé car

$P_1 = A(A'A)_1^-A' = A(A'A)_2^-A' = P_2$.

Enfin, $P_AA = A$ d'après P6, ce qui montre que P_A est le projecteur sur $Im\,A$.

De plus, un inverse généralisé de $A'A$ s'écrit

$$(A'A)^- = Q \begin{bmatrix} \Lambda_r^{-1} & T_{12} \\ T_{12}' & T_{22} \end{bmatrix} Q'$$

ce qui implique

$$P_A = P \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} P' = UU'. \qquad \square$$

P8 : Soit A^- un inverse généralisé de A. Une solution générale du système homogène $Ax = 0$ est : $\quad x = (I - A^-A)u, \quad$ pour u arbitraire.

P9 : Soit A^- un inverse généralisé de A et $y \in Im\,A$. Une solution générale du système Ax=y est : $\quad x = A^-y + (I - A^-A)u, \quad$ pour u arbitraire.

P10 : Si A^- est un inverse généralisé de A, un inverse généralisé quelconque s'écrit

$$G = A^- + U - A^-AUAA^-,$$

$$G = A^- + V(I - AA^-) + (I - A^-A)W, \quad \text{pour } U,\ V,\ W \text{ arbitraires.}$$

4.1.2 Inverse de Moore-Penrose

$\mathbf{A}^-$ n'étant pas unique, on construit une matrice inverse particulière appelée inverse de Moore-Penrose qui possède de "bonnes" propriétés.

Soit A une matrice $n \times p$, A^+, inverse de Moore-Penrose de A, est une matrice $p \times n$ vérifiant les propriétés suivantes

$$\begin{aligned} &\mathbf{D1} : AA^+A = A. \\ &\mathbf{D2} : A^+AA^+ = A^+. \\ &\mathbf{D3} : (AA^+)' = AA^+. \\ &\mathbf{D4} : (A^+A)' = A^+A. \end{aligned}$$

Construction de l'inverse

Soit A une matrice $n \times p$ telle que $rang(A) = r$. D'après la DVS 1, $\exists U, V$ orthogonales telles que $A = U\Lambda_r^{1/2}V'$ où $U'U = V'V = I_r$ et $\Lambda_r^{1/2} = \text{diag}\{\mu_1, ..., \mu_r\}$ avec $\mu_i > 0$. D'où

$$\boxed{A^+ = V\Lambda_r^{-1/2}U',}$$

vérifie les axiomes **D1**, **D2**, **D3** et **D4**.

Unicité de l'inverse

Soit A_1^+ et A_2^+, deux matrices inverses. D'après **D2** et **D3**, $A_1^+ = A_1^+AA_1^+ = A_1^+(AA_1^+)'$,

$A_1^+ = A_1^+(A_1^+)'A' = A_1^+(A_1^+)'(AA_2^+A)'$ par **D1**

$A_1^+ = A_1^+(A_1^+)'A'(AA_2^+)' = A_1^+(A_1^+)'A'AA_2^+ = A_1^+(AA_1^+)'AA_2^+ = A_1^+AA_1^+AA_2^+$ par **D3**

$A_1^+ = A_1^+AA_2^+$ par **D1**.

De même, $A_2^+ = A_2^+AA_2^+ = (A_2^+A)'A_2^+$ par **D2** et **D4**

$A_2^+ = (A_2^+AA_1^+A)'A_2^+ = (A_1^+A)'(A_2^+A)'A_2^+$ par **D1**

$A_2^+ = A_1^+AA_2^+AA_2^+ = A_1^+AA_2^+$ par **D4** et **D2**.

L'inverse de Moore-Penrose est donc unique. $\square$

Propriétés de A^+

P0 : Si $A = kB$ alors $A^+ = k^{-1}B^+$ avec $k \in I\!R_*$.

P1 : $A^+ = A^{-1}$ si A est inversible.

P2 : $(A^+)^+ = A$.

P3 : $(A')^+ = (A^+)'$.

P4 : Si A est une matrice symétrique et idempotente alors $A^+ = A$.

P5 : AA^+ et A^+A matrices symétriques sont idempotentes. Ce sont les projecteurs orthogonaux respectivement sur $Im\,A$ et sur $Im\,A' = \{Ker\,A\}^\perp$. Soit $A = U\Lambda_r^{1/2}V'$ la DVS maigre de A, alors $AA^+ = UU'$ et $A^+A = VV'$.

P6 : A, A^+, AA^+ et A^+A sont des matrices de même rang.

P7 : $A' = A'AA^+ \; = A^+AA'$.

P8 : $A^+ = A'(A^+)'A^+ \; = A^+(A^+)'A'$.

P9 : $(A'A)^+ = A^+(A^+)'$, $\quad (AA')^+ = (A^+)'A^+$.

P10 : $A = A(A'A)^+A'A \; = AA'(AA')^+A$.

P11 : $A^+ = (A'A)^+A' \; = A'(AA')^+$.

P12 : Si A est une matrice de plein rang colonne, alors $A^+ = (A'A)^{-1}A'$.

P13 : Si A est une matrice de plein rang ligne, alors $A^+ = A'(AA')^{-1}$.

P14 : $A = 0 \Longleftrightarrow A^+ = 0$.

P15 : $AB = 0 \Longleftrightarrow B^+A^+ = 0$.

P16 : $A^+B = 0 \Longleftrightarrow A'B = 0$.

P17 : $(A \otimes B)^+ = A^+ \otimes B^+$.

P18 : Si $y \in ImA$, la solution de norme euclidienne minimum de $Ax = y$, est $\hat{x} = A^+y$.

P19 : $A'AB = A'C \Longleftrightarrow AB = AA^+C$.

P20 : Soient A $n \times p$, B $n \times r$, C $p \times r$ telles que $rang(A) = rang(B) \; = rang(C) \; = r$ et

$A = BC'$, alors $A^+ = (C^+)'B^+ = C(C'C)^{-1}(B'B)^{-1}B'$.

4.2 Projecteurs M-orthogonaux

On a vu au chapitre 2 que P est un projecteur si $P^2 = P$ (idempotence) et un projecteur orthogonal au sens du produit scalaire usuel si P est de plus, symétrique. De plus, la DVS maigre de $A = U\Lambda_r^{1/2}V'$ permet de définir $P_A = UU'$ comme la matrice de projection orthogonale sur l'espace engendré par les colonnes de A. Ce projecteur s'écrit aussi $P_A = A(A'A)^-A' = AA^+$ grce aux inverses généralisés.

En statistique, on est souvent amené à définir des produits scalaires différents du produit scalaire usuel et basés sur des métriques M, M est une matrice symétrique définie positive, différentes de l'identité

$$< x, y >_M = y'Mx \quad \text{et} \quad \|x\|_M^2 = x'Mx.$$

Soit l'espace vectoriel euclidien $\mathbb{E} = \mathbb{R}^m$ muni d'un M-produit scalaire et soit $\mathbb{E}_1$ un s.e.v. de $\mathbb{E}$ tel que $\mathbb{E} = \mathbb{E}_1 \oplus \mathbb{E}_1^\perp$ où $\mathbb{E}_1^\perp = \{y \in \mathbb{E} \mid \, < y, x >_M = 0\,,\, x \in \mathbb{E}_1\}$. Pour tout x de $\mathbb{E}$ la décomposition

$$x = x_1 + y_1, x_1 \in \mathbb{E}_1,\ y_1 \in \mathbb{E}_1^\perp$$

est unique. P est un projecteur M-orthogonal sur $\mathbb{E}_1$ ssi

$$Px = x_1 \qquad (I - P)x = y_1.$$

La notion de M-orthogonalité est liée à une notion de symétrie particulière, la M-symétrie, plus générale que la symétrie usuelle des matrices.

Un matrice $A \in \mathbb{R}^{m \times m}$ est M-symétrique si

$$MA = A'M,$$

c'est à dire si MA est symétrique (car M est symétrique).

Pour α réel non nul, une matrice $(\alpha\, I_m)$-symétrique est symétrique au sens usuel.

Propriété caractéristique d'un projecteur M-orthogonal

Proposition : Un projecteur P est un projecteur M-orthogonal ssi P est M-symétrique. $\square$

Preuve : Soit P un projecteur ($P^2 = P$) sur $\mathbb{E}_1$ tel que

$$\forall x, y \in \mathbb{E}, \qquad Px \in \mathbb{E}_1, \qquad (I - P)y \in \mathbb{E}_1^\perp \text{ au sens de } M.$$

c'est à dire, $x'P'M(I - P)y = 0 \Longleftrightarrow P'M(I - P) = 0 \Longleftrightarrow P'M = P'MP$.

Puique M est symétrique, $P'M$ est aussi symétrique, $P'M = MP$. $\square$

4.2.1 Projecteur M-orthogonal sur $Im\,A$

Proposition : Soit $A \in \mathbb{R}^{m \times n}$ et M une métrique sur $\mathbb{R}^m$, P_A projecteur M-orthogonal sur l'espace $Im\,A$ engendré par les colonnes de A, est la matrice $m \times m$

$$\boxed{P_A = A(A'MA)^- A'M\,.}$$

Ce projecteur est unique quel que soit le choix de l'inverse généralisé. $\square$

Preuve : Soit $x \in \mathbb{R}^m$ décomposé en $x = x_1 + x_2$ sur $Im\,A \oplus \{Im\,A\}^\perp$. Multipliant à gauche par $A'M$,

$$A'M\,x = A'M\,x_1 + A'M\,x_2 = A'M\,x_1 = A'MA\,\beta.$$

Le système linéaire

$$A'MA\,\beta = A'M\,x$$

admet des solutions puisque $Im\,A'MA = Im\,A'M = Im\,A'$. Pour un choix d'inverse généralisé,

$$\beta = (A'MA)^{-}A'M\,x$$

ce qui donne

$$x_1 = A\,\beta = A(A'MA)^{-}A'Mx \qquad \text{et} \qquad P_A = A(A'MA)^{-}A'M.$$

Pour montrer l'unicité, décomposons M par Cholesky, $M = T'T$, et posons $B = TA$. T triangulaire supérieure est inversible et $A = T^{-1}B$. Alors $P_A = T^{-1}B(B'B)^{-}B'T$. La décomposition de Cholesky est unique et $B(B'B)^{-}B'$ ne dépend pas de l'inverse généralisé choisi (propriété P7 des inverses généralisés). $\square$

Conséquence : Dans la pratique, l'écriture du projecteur utilise l'inverse de Moore-Penrose $P_A = A(A'MA)^{+}A'M$ et, lorsque A est de plein rang colonne $P_A = A(A'MA)^{-1}A'M$.

Cas particulier 1 : $M = \alpha\,I$, où α est un réel positif.

Pour tout α positif, le projecteur est le projecteur orthogonal usuel $A(A'A)^{-}A' = A(A'A)^{+}A' = AA^{+}$.

Cas particulier 2 : Les colonnes de A sont M-orthogonales

$$A'MA = diag(\|A^1\|^2_M, \ldots, \|A^n\|^2_M)\,.$$

Dans ce cas ImA est de dimension n supposé plus petit que m. Le projecteur sur ImA se décompose en la somme des projecteurs sur chacun des vecteurs de la base $A^1, \ldots, A^n$

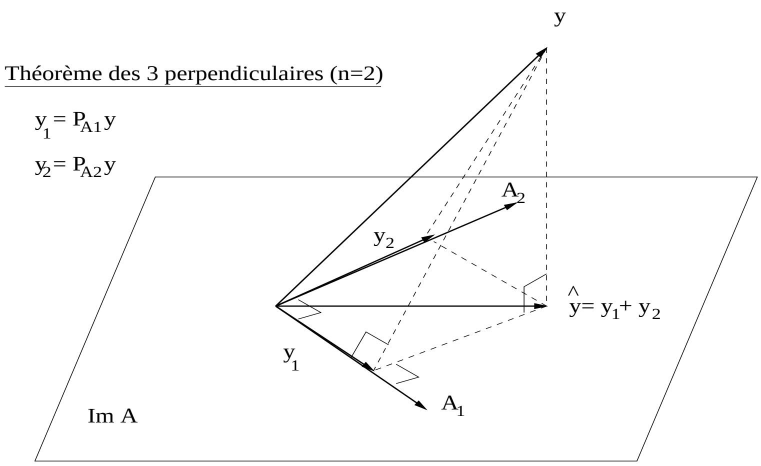


Figure 6 : Théorème des 3 perpendiculaires de la géométrie Euclidienne ($n = 2$).

$$P_A = \sum_{i=1}^{n} P_{A^i}\,,$$

où $P_{A^i} = A^i A^{i\,\prime} M / \|A^i\|^2_M$. Si de plus la base est M-orthonormée, c'est à dire si A est M-orthogonale, alors $P_{A^i} = A^i A^{i\,\prime} M$. L'orthogonalité des colonnes de A donne $\|P_A\, y\|^2_M = \sum_{i=1}^n \|P_{A^i}\, y\|^2_M$.
Lorsque $y \in Im\, A$, $\|y\|^2_M = \sum_{i=1}^n \|P_{A^i}\, y\|^2_M$, c'est le cas en particulier lorsque la matrice A est carrée (les colonnes de A forment une base de l'espace tout entier).

Eléments propres du projecteur M-orthogonal sur Im A

Proposition : La matrice $P_A = A(A'MA)^- A'M$ est diagonalisable, ses valeurs propres sont 0 ou 1 et ses vecteurs propres sont M-orthonormés. De plus,

$$\text{trace}(P_A) = \text{rang}(A) = \text{rang}(P_A). \quad \square$$

Preuve : P_A est M-symétrique ($MP_A = P_A' M$). On ne peut pas appliquer directement les propriétés des matrices symétriques. Utilisons la décomposition $M = M^{1/2} M^{1/2}$ (on peut aussi utiliser Cholesky) et posons $B = M^{1/2} A$ dont le rang est celui de A. La DVS de $B = U \Lambda_r^{1/2} V' = P \Lambda Q'$ et la propriété **P7** des inverses généralisés appliquée à B donnent

$$P_A = M^{-1/2} P \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} P^{-1} M^{1/2} = M^{-1/2} U U' M^{1/2}.$$

Les vecteurs propres de P_A sont les colonnes de $M^{-1/2} P$ qui sont M-orthonormées. $\square$

4.2.2 Un problème aux moindres carrés

Si A est une matrice réelle $m \times n$ ($m \geq n$) et $y \in \mathbb{R}^m$, le système linéaire $Ax = y$ n'a pas nécessairement de solution. Pour qu'il existe une solution il faut que $y \in Im\, A$. On peut toujours se poser le problème de chercher $\widehat{x} \in \mathbb{R}^n$ tel que

$$\|A\widehat{x} - y\|^2_M = \min_{x \in \mathbb{R}^n} \|Ax - y\|^2_M. \qquad (*)$$

Ce problème sécrit comme la recherche de l'élément $\widehat{y} = A\widehat{x}$ de ImA le plus proche de y au sens de la norme $\|.\|_M$.

$$\|\widehat{y} - y\|^2_M = \min_{v \in ImA} \|v - y\|^2_M. \qquad (*)$$

Proposition : Une solution du problème aux moindres carrés (*) est donnée par

$$\widehat{x} = (A'MA)^+ A'My.$$

La solution est unique si $rang(A) = n$. Si la solution n'est pas unique, toute solution s'écrit $\overline{x} = \widehat{x} + u$ avec $u \in Ker\,(A'MA) = Ker\, A$. L'approximation $\widehat{y} = A\overline{x} = A\widehat{x} = P_A\, y$ de y est unique, c'est la projection M-orthogonale de y sur $Im\, A$. $\square$

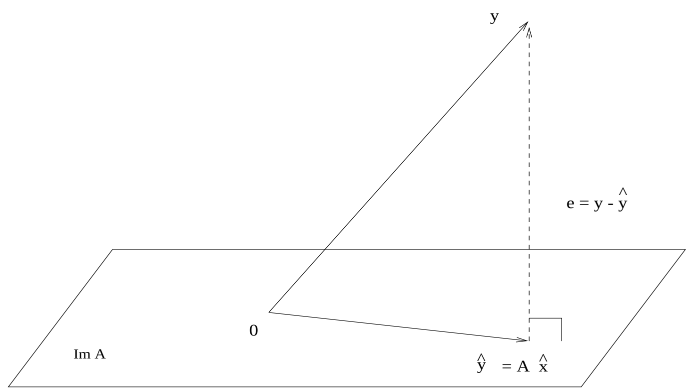


Figure 7 : solution du problème aux moindres carrés, $\widehat{y} = P_A y$.

Preuve : Notons

$$\varphi(x) = \|Ax - y\|_M^2$$

la fonction à minimiser dite fonction objectif. C'est une fonction de $I\!R^n$ dans $I\!R^+$, c'est à dire une fonction numérique des n variables $[x_1, \ldots, x_n]' = x$.

$$\varphi(x) = (Ax - y)'M(Ax - y) = x'A'MAx - 2y'MAx + y'My\,.$$

Cette fonction est une fonction quadratique différentiable sur $I\!R^n$. Les équations aux dérivées partielles dites équations normales fournissent les points stationnaires

$$\nabla\varphi(x) = 2A'MAx - 2A'My = 0_{I\!R^n}\,,$$

où $\nabla\varphi(x)$ est le vecteur gradient de φ calculé en x. Une solution x du système précédent est un minimum local si la dérivée seconde, la matrice Hessien $H\varphi(x)$, calculée en x est semi définie positive dans un voisinage de x. C'est bien le cas,

$$H\varphi(x) = 2A'MA$$

est constante indépendante de x. C'est une matrice symétrique semi définie positive. Pour obtenir un minimum, il faut donc résoudre le système linéaire

$$A'MA\,x = A'My\,.$$

Si la matrice $A'MA$ $n \times n$ est inversible, c'est à dire si $rang(A) = rang(A') = n$, alors il y a une seule solution

$$\widehat{x} = (A'MA)^{-1}A'My.$$

Dans le cas contraire, le système admet une infinité de solutions car le second membre appartient à $Im\,A'$ et puisque M est régulière $Im\,A' = Im\,A'MA$. Plutôt que d'écrire les solutions en utilisant un inverse généralisé $(A'MA)^-$ on préfère construire une solution

$$\widehat{x} = (A'MA)^+ A'My.$$

Soit $\overline{x}$ une autre solution des équations normales, alors

$$A'MA(\overline{x} - \widehat{x}) = 0_{\mathbb{R}^n} \quad \text{et} \quad \overline{x} = \widehat{x} + u, \text{ avec } u \in Ker\,(A'MA).$$

Puisque M est symétrique définie positive, en décomposant M par Cholesky ou par $M = (M^{1/2})^2$, on montre aisément que $Ker\,(A'MA) = Ker\,A$. Alors, qu'il y ait unicité ou non, l'approximation de y par $\widehat{y}$

$$\widehat{y} = A\overline{x} = A\widehat{x} = P_A y,$$

où P_A est le projecteur M-orthogonal sur $Im\,A$, est unique.

Un façon de mesurer la qualité de l'approximation est fournie par l'indice

$$R^2(y\,; ImA) = \frac{\|\widehat{y}\|_M^2}{\|y\|_M^2}$$

appelé "coefficient de détermination" entre y et ImA. On le note R^2 pour simplifier. C'est le carré du M-cosinus du M-angle formé par les vecteurs y et $\widehat{y}$.

$R^2 = 1 \quad \Longleftrightarrow \quad y = \widehat{y} \quad \Longleftrightarrow \quad y \in Im\,A.$

$R^2 = 0 \quad \Longleftrightarrow \quad \widehat{y} = 0 \quad \Longleftrightarrow \quad y \in \{Im\,A\}^\perp. \quad \square$

Propriétés du projeté $\widehat{y} = P_A\, y = A\widehat{x}$

P1 $e = y - \widehat{y} \in \{Im\,A\}^\perp$, c'est à dire $< y - \widehat{y}, u >_M = 0\,, \quad \forall u \in Im\,A$.

P2 $\|y\|_M^2 = \|y - \widehat{y}\|_M^2 + \|\widehat{y}\|_M^2$. C'est le théorème de Pythagore.

P3 $\|P_A y\|_M \leq \|y\|_M\,, \quad \forall y \in \mathbb{R}^m$, ou $0 \leq R^2 \leq 1$. On dit que l'application linéaire P_A est contractante.

4.3 Exercices

Exercice 1 : Soit l'espace vectoriel euclidien $(\mathbb{R}^p, I_p)$ et E le s.e.v. engendré par les deux vecteurs $\mathbb{1}_p = (1, \ldots, 1)'$ et $(1, 2, \ldots, p)'$. Calculer la matrice du projecteur orthogonal sur E.

Exercice 2 : Soit $E = \{(y_1, \ldots, y_{2n}) \in \mathbb{R}^{2n} \mid y_t = y_{t+2}, 1 \leq t \leq 2n-2\}$.

a. Montrer que E est un s.e.v. de $\mathbb{R}^{2n}$. Quelle est sa dimension?

b. On munit $\mathbb{R}^{2n}$ de la métrique identité. Soit $x \in \mathbb{R}^{2n}$, calculer

$$\hat{y} = arg \min_{y \in E} \|x - y\|_2^2\,.$$

Chapter 5

Dérivation Matricielle

5.1 Introduction

Dans de nombreux domaines d'application de l'analyse numérique et de la statistique comme l'économétrie, la chimiométrie, etc, on est amené à résoudre des problèmes du type

$$\min_X \varphi(X)$$

où la fonction à minimiser est une fonction numérique qui a pour argument une matrice X. Il est alors commode d'utiliser les outils de vectorisation pour tenter de résoudre le problème lorsque φ est différentiable. L'objectif de cette section n'est pas de proposer un cours de calcul différentiel ou d'optimisation mais seulement d'examiner le rôle joué par les opérateurs *vec* et $\otimes$, étudiés dans le chapitre 1, pour le calcul pratique des dérivées de fonctions matricielles lors de la résolution de certains problèmes d'optimisation.

5.2 Dérivation matricielle

5.2.1 Matrices Jacobiennes

Classification des fonctions et des variables

Les fonctions scalaires ou numériques sont notées ϕ.
Les fonctions vectorielles sont notées f.
Les fonctions matricielles sont notées F.
Les variables réelles sont notées ξ.
Les variables vectorielles sont notées x.

Les variables matricielles sont notées X.

Variables / Fonctions	numériques	vectorielles	matricielles
numériques	$\phi(\xi)$	$\phi(x)$	$\phi(X)$
vectorielles	$f(\xi)$	$f(x)$	$f(X)$
matricielles	$F(\xi)$	$F(x)$	$F(X)$

Exemples

$$\phi(\xi) = \xi^2 \qquad \phi(x) = a'x \qquad \phi(X) = a'Xa$$

$$f(\xi) = \begin{bmatrix} \xi \\ \xi^2 \end{bmatrix} \qquad f(x) = Ax \qquad f(X) = Xa$$

$$F(\xi) = \begin{bmatrix} \cos(\xi) & -\sin(\xi) \\ \sin(\xi) & \cos(\xi) \end{bmatrix} \quad F(x) = xx' \quad F(X) = X^{-1}.$$

Matrices Jacobiennes de fonctions différentiables

• Soit $\phi : x \longrightarrow \phi(x)$, une fonction numérique différentiable de $\mathbb{R}^n$ dans $\mathbb{R}$. La dérivée $D\phi(x)$ ou matrice Jacobienne s'écrit

$$D\phi(x) = [D_1\phi(x), \ldots, D_n\phi(x)] = \frac{\partial \phi(x)}{\partial x'}$$

c'est le vecteur ligne $1 \times n$ tel que

$$d\phi = D\phi(x)\; dx = \sum_{i=1}^{n} D_i\phi(x)dx_i.$$

Le gradient de ϕ en x est le vecteur colonne transposé de $D\phi(x)$

$$\nabla\phi(x) = (D\phi(x))'.$$

• Soit f une fonction vectorielle différentiable de $\mathbb{R}^n$ dans $\mathbb{R}^m$, c'est-à-dire

$$x \longmapsto f(x) = \begin{bmatrix} \phi_1(x) \\ \vdots \\ \phi_m(x) \end{bmatrix}.$$

La dérivée de f où matrice Jacobienne est la matrice d'ordre $m \times n$

$$Df(x) = \frac{\partial f(x)}{\partial x'} = \begin{bmatrix} D_1\phi_1(x) & \cdots & D_n\phi_1(x) \\ \vdots & \ddots & \vdots \\ D_1\phi_m(x) & \cdots & D_n\phi_m(x) \end{bmatrix}$$

telle que

$$df = Df(x)dx,$$

où $df = [d\phi_1, \ldots, d\phi_m]'$.

• Soit F une fonction matricielle différentiable de $I\!R^{n\times q}$ dans $I\!R^{m\times p}$,

$$X = [x_{ij}] \longrightarrow F(X) = \begin{bmatrix} \phi_{11}(X) & \cdots & \phi_{1p}(X) \\ \vdots & \ddots & \vdots \\ \phi_{m1}(X) & \cdots & \phi_{mp}(X) \end{bmatrix}$$

la matrice Jacobienne de F en X est la matrice d'ordre $mp \times nq$

$$DF(X) = \begin{bmatrix} D_{11}\phi_{11}(X) & \cdots & D_{nq}\phi_{11}(X) \\ \vdots & \ddots & \vdots \\ D_{11}\phi_{m1}(X) & \cdots & D_{nq}\phi_{m1}(X) \\ \vdots & \ddots & \vdots \\ D_{11}\phi_{1p}(X) & \cdots & D_{nq}\phi_{1p}(X) \\ \vdots & \ddots & \vdots \\ D_{11}\phi_{mp}(X) & \cdots & D_{nq}\phi_{mp}(X) \end{bmatrix}$$

où $D_{ij}\phi_{kl}(X) = \dfrac{\partial \phi_{kl}(X)}{\partial x_{ij}}$. On la note

$$DF(X) \doteq \frac{\partial vec(F(X))}{\partial vec'(X)}.$$

Le théorème d'identification pour les fonctions matricielles différentiables donne

$$d\,vec(F(X)) = A(X)\,d\,vec(X) \quad \Longleftrightarrow \quad A(X) = DF(X).$$

Formulaire

$$X = [x_{ij}]\ n \times p,\ d\,X = \begin{bmatrix} dx_{11} & \cdots & dx_{1p} \\ \vdots & \ddots & \vdots \\ dx_{n1} & \cdots & dx_{np} \end{bmatrix}.$$

$$d\,C^{ste} = 0.$$

$$d\,(X+Y) = d\,X + d\,Y.$$

$$d\,(\alpha\,X) = \alpha\,d\,X.$$

$$d\,(X') = (d\,X)'.$$

$$d\,(\mathrm{trace}(X)) = \mathrm{trace}(d\,X).$$

$$d\,(vec(X)) = vec(d\,X).$$

$$d\,(XY) = (d\,X)Y + X(d\,Y).$$

$$d\,(X \otimes Y) = ((dX) \otimes Y) + (X \otimes (d\,Y)).$$

Tableau d'identification

fonctions	différentielles	matrices Jacobiennes	ordre de D	autres notations
$\phi(\xi)$ $\phi(x)$ $\phi(X)$	$d\phi = \alpha d\xi$ $d\phi = a'dx$ $d\phi = vec'(A)dvec(X)$ $= \text{trace}(A'dX)$	$D\phi(\xi) = \alpha$ $D\phi(x) = a'$ $D\phi(X) = vec'(A)$	1×1 $1 \times n$ $1 \times nq$	 $\nabla\phi(x) = a$ $\dfrac{\partial\phi(X)}{\partial X} = A \;\; n \times q$
$f(\xi)$ $f(x)$ $f(X)$	$df = ad\xi$ $df = Adx$ $df = Advec(X)$	$Df(\xi) = a$ $Df(x) = A$ $Df(X) = A$	$m \times 1$ $m \times n$ $m \times nq$	
$F(\xi)$ $F(x)$ $F(X)$	$dvec(F) = vec(A)d\xi$ $dvec(F) = Adx$ $dvec(F) = Advec(X)$	$DF(\xi) = vec(A)$ $DF(x) = A$ $DF(X) = A$	$mp \times 1$ $mp \times n$ $mp \times nq$	$\dfrac{d\,F(\xi)}{d\,\xi} = A \;\; m \times p$

Remarque : Dans deux cas particuliers on réorganise les dérivées partielles

$*$ Soit $\phi(X)$ avec $X = [x_{ij}]$ $n \times q$, alors $d\phi = \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{q} \frac{\partial\phi\,(X)}{\partial x_{ij}} dx_{ij} = \text{trace}((\frac{\partial\phi\,(X)}{\partial X})'dX)$.

$$\frac{\partial\phi\,(X)}{\partial X} = \left[\frac{\partial\phi\,(X)}{\partial x_{ij}}\right] \; n \times q.$$

$*$ Soit $F(\xi)$ avec $F(\xi) = [F_{ij}(\xi)]$ qui est d'ordre $m \times p$

$$\frac{d\,F(\xi)}{d\,\xi} = \left[\frac{d\,F_{ij}(\xi)}{d\,\xi}\right].$$

Marche à suivre (présentée dans le cas $F(X)$) :

1) On calcule $d\,F(X)$.

2) On vectorise $d\,F(X)$ c'est-à-dire, on calcule $vec\,(d\,F(X)) = d\,vec\,(F(X))$.

3) On met sous la forme $d\,vec\,(F(X)) = A(X)d\,vec\,(X)$.

4) On identifie $A(X) = DF(X)$.

Dérivation des fonctions matricielles composées :

Soit S un ouvert de $I\!R^{n\times q}$ et supposons que $F : S \to I\!R^{m\times p}$ soit différentiable en $C \in S$. Soit T un ouvert de $I\!R^{m\times p}$ tel que $F(X) \in T$ pour tout $X \in S$ et supposons que $G : T \to I\!R^{r\times s}$ soit différentiable en $B = F(C)$. Alors, la fonction composée $H : S \to I\!R^{r\times s}$ définie par

$$H(X) = G(F(X))$$

est différentiable en C et

$$\boxed{DH(C) = [DG(B)]\,[DF(C)].}$$

La formule précédente s'appelle la règle de "dérivation en chaine" pour les Jacobiens. □

Exemple : Calcul du Jacobien de $\phi(X) = \text{trace}(AXB)$ avec A et B constantes.

<u>Première méthode</u> : $d\phi(X) = d\text{trace}(AXB) \;= \text{trace}(d(AXB)) \;= \text{trace}(A(dX)B)$

$$= \text{trace}(BA(dX)) = vec'(A'B')dvec(X)$$

$$\Longleftrightarrow D\phi(X) = vec'(A'B')$$

ou en réorganisant les dérivées partielles selon la position des éléments de X

$$\frac{\partial \text{trace}(AXB)}{\partial X} = A'B'. \quad \square$$

<u>Deuxième méthode</u> : Soit $F(X) = AXB$, alors $dF(X) = A(dX)B$

$vec(dF(X)) = dvec(F(X)) \;= (B' \otimes A)vec(dX)$.

Donc $DF(X) = B' \otimes A$.

Soit $\phi(Y) = \text{trace}(Y)$, alors $d\phi(Y) = \text{trace}(dY) \;= \text{trace}(IdY) \;= vec'(I)vec(dY)$.

Donc $D\text{trace}(Y) = vec'(I)$. Bien sûr, on peut réorganiser les dérivées partielles selon la position des éléments de Y, et

$$\frac{\partial\, \text{trace}(Y)}{\partial Y} = I\,.$$

Mais ici, il faut user du Jacobien pour utiliser la dérivation de fonctions composées

$D\phi(F(X)) = vec'(I)(B' \otimes A) \;= [(B \otimes A')vec(I)]' = vec'(A'IB') = vec'(A'B')$. □

5.2.2 Hessien de fonctions numériques

Soit $\phi : I\!R^n \longrightarrow I\!R$ une fonction deux fois différentiable, alors

$$H\phi(x) = [D^2_{ij}\phi(x)]$$

est une matrice symétrique d'ordre n appelée le Hessien de ϕ en x.

Théorème d'identification

$$d^2\phi = d\,(d\,\phi) = (dx_2)'B(dx_1) \iff H\phi(x) = \tfrac{1}{2}(B+B').$$

Tableau d'identification

Fonctions	Différentielles d'ordre 2	Hessien	Ordre
$\phi(\xi)$	$d^2\phi = \beta(d\xi_1)(d\xi_2)$	$H\phi(\xi) = \beta$	1×1
$\phi(x)$	$d^2\phi = (dx_2)'Bdx_1$	$H\phi(x) = \frac{1}{2}(B+B')$	$n \times n$
$\phi(X)$	$d^2\phi = (dvec(X_2))'B\,dvec(X_1)$	$H\phi(X) = \frac{1}{2}(B+B')$	$nq \times nq$

Proposition : Soit $\phi(X)$ une fonction numérique de X, matrice $n \times q$, alors
$d^2\phi = \text{trace}(B(dX_2)'C(dX_1)) \Leftrightarrow H\phi(X) = \frac{1}{2}(B' \otimes C + B \otimes C')$

$d^2\phi = \ \text{trace}(B(dX_2)C(dX_1)) \Leftrightarrow H\phi(X) = \frac{1}{2}K_{qn}(B' \otimes C + C' \otimes B).$ $\square$

Preuve

On sait que $\text{trace}(ABCD) = vec'(D')(C' \otimes A)vec(B) \ = vec'(D)(A \otimes C')vec(B')$.

Donc $\text{trace}(B(dX_2)'C(dX_1)) = vec'(dX_1)(B \otimes C')vec(dX_2)$

$$H\phi(X) = \tfrac{1}{2}(B \otimes C' + B' \otimes C).$$

$\text{trace}(BdX_2CdX_1) = vec'(dX_1)(B \otimes C')vec(dX_2)' = vec'(dX_1)(B \otimes C')K_{nq}vec(dX_2)$

$H\phi(X) = \frac{1}{2}((B \otimes C')K_{nq} + K_{qn}(B' \otimes C)) = \frac{1}{2}(K_{qn}(C' \otimes B) + K_{qn}(B' \otimes C))$. $\square$

5.3 Extremums de fonctions numériques

Les problèmes d'optimisation considérés dans cette section sont des problèmes de minimisation d'une fonction numérique réelle ϕ appelée fonction objectif. Pour maximiser une fonction ϕ on se ramène à la minimisation de $-\phi$.

Soient $\phi : S \subset I\!R^n \longrightarrow I\!R$ et $\overline{x}$ un point de S. On dit que ϕ présente en $\overline{x}$

un minimum local ou relatif, s'il existe une boule ouverte $B(\overline{x})$ centrée en $\overline{x}$ telle que

$$\phi(x) \geq \phi(\overline{x}) \text{ pour tout } x \in B(\overline{x}) \cap S,$$

un minimum strict local, s'il existe une boule ouverte $B(\overline{x})$ centrée en $\overline{x}$ telle que

$$\phi(x) > \phi(\overline{x}) \text{ pour tout } x \in B(\overline{x}) \cap S,\ x \neq \overline{x},$$

un minimum global ou absolu, si

$$\phi(x) \geq \phi(\overline{x}) \text{ pour tout } x \in S,$$

un minimum strict global, si

$$\phi(x) > \phi(\overline{x}) \text{ pour tout } x \in S,\ x \neq \overline{x}.$$

Par abus de langage on dira que le point $\overline{x}$ lui même est un minimum local, local strict...

Transformations strictement croissantes

Théorème 1 : Soit $S \subset I\!R^n$ et ϕ une fonction de S dans $I\!R$. Notons $T = \phi(\Omega)$. Soit φ une application strictement croissante de T dans $I\!R$. Alors ϕ présente un minimum absolu (resp. relatif) en $\overline{x}$ si et seulement si $\psi = \varphi \circ \phi$ présente un minimum absolu (resp. relatif) en $\overline{x}$. $\square$

Preuve : Soit φ une fonction croissante sur T et $\overline{x}$ un minimum local ou global pour ϕ, c'est à dire $\phi(x) \geq \phi(\overline{x})$ sur un certain ensemble $\mathcal{O}$. Alors $\psi(x) = \varphi(\phi(x)) \geq \varphi(\phi(\overline{x})) = \psi(\overline{x})$.

Dans l'autre sens, supposons que $\overline{x}$ soit un minimum local ou global pour ψ, $\psi(x) \geq \psi(\overline{x})$ pour tout x dans un certain ensemble $\mathcal{O}$. Supposons qu'il existe x_0 de $\mathcal{O}$ tel que $\phi(x_0) < \phi(\overline{x})$. La stricte croissance de φ implique $\psi(x_0) = \varphi(\phi(x_0)) < \varphi(\phi(\overline{x})) = \psi(\overline{x})$ ce qui est impossible. Donc $\phi(x) \geq \phi(\overline{x})$ pour tout $x \in \mathcal{O}$. $\square$

Applications :

A1 : **Minimisation d'une norme Euclidienne** : La fonction $\|.\|_2$ n'est pas différentiable en 0, aussi pour récupérer la différentiabilité de la fonction objectif, on minimise $\psi = \|.\|_2^2$ plutôt que $\phi = \|.\|_2$ (on choisit donc $\varphi(\xi) = \xi^2$ qui est strictement croissante sur $[0, +\infty[$).

A2 : **Maximum de vraisemblance** : Maximiser $\phi(x) = \exp(-\eta(x))$ est équivalent à maximiser $\ln(\phi(x))$ ce qui est encore équivalent à minimiser $\eta(x)$.

5.3.1 Problèmes d'extremums libres

Conditions nécessaires de minimum local

Théorème 2 : Soit Ω un ouvert de $I\!R^n$ et ϕ une fonction de Ω dans $I\!R$ différentiable sur Ω. Si ϕ admet un minimum local en un point $\overline{x}$ de Ω, alors

$$D\phi(\overline{x}) = 0_{1\times n} \quad \text{ou} \quad \nabla\phi(\overline{x}) = 0_{n\times 1}. \quad (*) \quad \square$$

Remarques:

R1 : Importance de l'hypothèse "Ω est un ouvert". Contre-exemple évident : minimiser la fonction de la variable réelle $\phi(\xi) = \xi^2$ sur l'intervalle fermé $[\,1, 2\,]$ admet pour solution $\hat{\xi} = 1$ en lequel $D\phi(1) = 2 \neq 0$.

R2 : Toute solution $\overline{x}$ du système (*) des n équations aux dérivées partielles ou équations normales ou encore équations d'Euler, est appelée point stationnaire ou point critique. Déterminer la nature d'un point critique, c'est étudier s'il s'agit d'un minimum, d'un maximum ou d'un point-selle. Le théorème 3 fournit des conditions suffisantes dans le cas où la fonction obectif est 2 fois différentiable.

R3 : Lorsque la fonction objectif $\phi(X)$ a un argument matriciel $X_{n\times p}$, la condition nécessaire d'extremum local en $\overline{X}$ de l'ouvert Ω, s'écrit

$$D\phi(\overline{X}) = 0_{1\times np} \quad \text{ou encore} \quad \frac{\partial\phi(\overline{X})}{\partial X} = 0_{n\times p}\ . \quad (*)$$

Conditions suffisantes d'optimalité (convexité) locale

Théorème 3 : Si ϕ est 2 fois différentiable sur Ω ouvert de $I\!R^n$, un point critique $\overline{x}$ de Ω est un minimum local si une des conditions suivantes est réalisée

(1) Il existe une boule $B(\overline{x})$ centrée en $\overline{x} \in \Omega$ telle que
le Hessien $H\phi(x)$ est une matrice semi définie positive pour tout $x \in B(\overline{x}) \cap \Omega$.

(2) Le Hessien $H\phi(\overline{x})$ est une matrice définie positive, c'est à dire, les mineurs principaux du déterminant de $H\phi(\overline{x})$ sont positifs. Dans ce cas le minimum est strict. $\square$

Remarque : Les conditions suffisantes précédentes s'étendent au cas $\phi(X)$ d'une fonction de matrice, par le calcul de $H\phi(X)$.

5.3.2 Problèmes d'extremums liés

Soit la fonction ojectif ϕ : Ω ouvert $\subset I\!R^n \to I\!R$ et soit la fonction vectorielle $g = [g_1, \ldots, g_m]' : \Omega \to I\!R^m$ permettant de définir l'ensemble S des contraintes

$$\boxed{S = \{x \in I\!R^n \,|\, x \in \Omega\,;\; g_1(x) = 0\,, \ldots, g_m(x) = 0\}.}$$

Le problème est maintenant le suivant :

$$\boxed{\begin{array}{c}\text{minimiser } \phi(x) \\ \text{sous la contrainte : } x \in S\end{array}}$$

La façon la plus efficace pour résoudre ce type de problème d'extremums liés par les m contraintes de type égalité, est en général d'utiliser les *multiplicateurs de Lagrange.*

Conditions nécessaires d'optimalité locale

Théorème 4 (Lagrange) :

<u>Hypothèses sur les contraintes</u>

Soient g : Ω ouvert de $I\!R^n \to I\!R^m$ $(n > m)$ et $\overline{x} \in \Omega$ tels que

H1 : $g(\overline{x}) = 0$,

H2 : g est différentiable sur une boule ouverte $B(\overline{x})$,

H3 : la matrice Jacobienne $Dg(x)$ $m \times n$ est continue en $\overline{x}$,

H4 : $Dg(\overline{x})$ est de rang m.

<u>Hypothèses sur la fonction objectif</u>

Soit $\phi : \Omega \to I\!R$ telle que

H5 : ϕ est différentiable en $\overline{x}$,

H6 : $\phi(x) \geq \phi(\overline{x})$ pour tout x de $B(\overline{x})$ vérifiant $g(x) = 0$.

<u>Conclusion</u>

Il existe un (unique dû à H4) vecteur $\overline{l} = [\overline{\lambda}_1, \ldots, \overline{\lambda}_m]'$ appelé vecteur des multiplicateurs de Lagrange, tel que

$$D\phi(\overline{x}) + \overline{l}\,'Dg(\overline{x}) = 0_{1\times n}, \qquad (**)$$

ou en transposant

$$\nabla\phi(\overline{x}) + \textstyle\sum_{i=1}^{m} \overline{\lambda}_i \nabla g_i(\overline{x}) = 0_{n\times 1}. \qquad (**) \qquad \square$$

Remarques :

R1 : **Méthode des multiplicateurs de Lagrange** : Dans la recherhe d'optimums locaux, on construit d'abord la fonction de Lagrange $\mathcal{L} : \Omega \times I\!R^m \to I\!R$

$$\boxed{\mathcal{L}(x,l) = \phi(x) + l'g(x) = \phi(x) + \sum_{i=1}^m \lambda_i g_i(x).}$$

Les conditions nécessaires d'optimalité locale pour le problème d'extremums libres

$$\boxed{\begin{array}{ll} \min & \mathcal{L}(x,l) \\ \Omega \times I\!R^m & \end{array}}$$

s'écrivent

$$\boxed{\begin{array}{c} \nabla_x \mathcal{L}(x,l) = \nabla\phi(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) = 0 \qquad (^{**}) \\ \nabla_l \mathcal{L}(x,l) = g(x) = 0. \end{array}}$$

Elles fournissent comme solutions les point critiques $(\overline{x}, \overline{l})$. Parmi eux se trouvent les optimums locaux de ϕ relatifs à S. Il faut enfin sélectionner parmi les points critiques ceux qui sont des minimums.

R2 : Cas où ϕ est fonction d'une matrice $X_{n\times p}$.
Soient $\phi : \Omega$ (ouvert de $I\!R^{n\times p}$) $\to I\!R$ et $G : \Omega \to I\!R^{m\times q}$, le problème d'optimisation devient

$$\boxed{\begin{array}{cc} \min & \phi(X) \\ G(X) = 0_{m\times q} & \end{array}}$$

Par vectorisation, ce problème est un problème d'optimisation à mq contraintes. On introduit $\Lambda = [\lambda_{ij}]$, matrice $m\times q$ des multiplicateurs de Lagrange, dans l'expression de la fonction de Lagrange $\mathcal{L}(X, \Lambda)$ definie par

$$\boxed{\mathcal{L}(X,\Lambda) = \phi(X) + vec'(\Lambda)vec(G(X)) = \phi(X) + \text{trace}(\Lambda' G(X)).}$$

Il faut déterminer les points critiques $(\overline{X}, \overline{\Lambda})$ de la fonction $\mathcal{L}(X, \Lambda)$ et enfin étudier la nature ces points.

Conditions suffisantes de minimum local sous contraintes
Théorème 5 : Supposons que $\mathcal{L}(x,l)$ soit 2 fois différentiable au point critique $(\overline{x}, \overline{l})$ et que la matrice Jacobienne $m \times n$ $Dg(\overline{x})$ soit de rang m. Si de plus

$$u' H_x \mathcal{L}(\overline{x}, \overline{l})\, u > 0 \quad \text{pour tout } u \in I\!R^n \text{ tel que } Dg(\overline{x})\, u = 0_{m\times 1},$$

où $H_x\mathcal{L}(\overline{x}, \overline{l}) = H\phi(\overline{x}) + \sum_{i=1}^m \overline{\lambda}_i H g_i(\overline{x})$, alors ϕ présente un minimum strict en $\overline{x}$ sous la contrainte $g(x) = 0$. $\square$

5.4 Exercices

Exercice 1 : Fonctions numériques d'un vecteur : Etablir le tableau d'identification suivant

$\phi(x)$	$d\,\phi(x)$	$D\,\phi(x)$	$\nabla\,\phi(x)$
$a'x$	$a'\,dx$	a'	a
$x'Ax$	$x'(A+A')dx$	$x'(A+A')$	$(A+A')x$
$\frac{\phi_1(x)}{\phi_2(x)}$	$\frac{\phi_2 d\phi_1 - \phi_1 d\phi_2}{\phi_2^2}$	$(\phi_2 D\phi_1 - \phi_1 D\phi_2)/\phi_2^2$	$\frac{\phi_2(x)\nabla\phi_1(x) - \phi_1(x)\nabla\phi_2(x)}{\phi_2^2(x)}$

Exercice 2 : Fonctions numériques d'une matrice

a) Montrer que

$$\frac{\partial \mathrm{trace}(X)}{\partial X} = I \; ; \qquad \frac{\partial \mathrm{trace}(X'X)}{\partial X} = 2X \; ; \qquad \frac{\partial \mathrm{trace}(X^2)}{\partial X} = 2X'.$$

b) Etablir le tableau d'identification suivant

$\phi(X)$	$d\,\phi(X)$	$D\,\phi(X)$	$\partial\phi(X)/\partial X$
$\mathrm{trace}(AX)$	$\mathrm{trace}(AdX)$	$vec'(A')$	A'
$\mathrm{trace}(XAX'B)$	$\mathrm{trace}[(AX'B + A'X'B')dX]$	$vec'(B'XA' + BXA)$	$B'XA' + BXA$
$\mathrm{trace}(XAXB)$	$\mathrm{trace}[(AXB + BXA)dX]$	$vec'(B'X'A' + A'X'B')$	$B'X'A' + A'X'B'$

Exercice 3 : Fonctions vectorielles

Calculer les matrices Jacobiennes de $f(x) = Ax$, $f(x) = a(x'x)$, $f(X) = Xa$.

Exercice 4 : Fonctions matricielles : Etablir le tableau d'identification suivant

$F(X)$	$d\,F(X)$	$D\,F(X)$
X	dX	I_{nq}
X'	dX'	K_{nq}
XX'	$(dX)X' + X(dX)'$	$(I_{n^2} + K_{nn})(X \otimes I_n)$
$X'X$	$(dX)'X + X'(dX)$	$(I_{q^2} + K_{qq})(I_q \otimes X')$

où X est une matrice $n \times q$.

Exercice 5 : Calculer les matrices Hessiennes des fonctions scalaires suivantes : $\phi(x) = x'Ax$, $\phi(X) = \mathrm{trace}(X'AX)$, $\phi(X) = \mathrm{trace}(AXBX')$, $\phi(X) = \mathrm{trace}(X^2)$.

Exercice 6 : Résoudre le problème d'optimum libre suivant

$$\min_X \|Y - AX\|_F^2$$

où les matrices Y $m \times n$ et A $m \times p$ sont données.

Exercice 10 : Soient A une matrice symétrique définie positive $n \times n$ et B une matrice $m \times n$. Alors,

$$\text{trace}(X'AX) \geq \text{trace}[(BA^{-1}B')^{-1}]$$

pout toute matrice X $n \times m$ satisfaisant $BX = I_m$. Le minimum étant obtenu pour

$$\widehat{X} = A^{-1}B'(BA^{-1}B')^{-1}\,.$$

Indication : exprimer le problème sous forme d'un problème d'optimisation avec contraintes de type égalité et le résoudre.

Chapter 6

Le paysage mathématique et statistique de l'Analyse Factorielle de Données : la géométrie Euclidienne

L'analyse factorielle de données a pris son essort dans le deuxième tiers du XXième siècle avec l'apparition des ordinateurs et le devoir de traiter des données dont la taille n'a fait que crotre avec les progrès de l'informatique. Traiter ou analyser les données, c'est d'abord essayer de détecter la présence d'éventuelles erreurs de saisie ou de données atypiques, ensuite tenter de visualiser certaines structures communes difficiles voire impossibles à mettre a priori en évidence à cause de la taille des données.

La diversité de nature des variables mesurées est à l'origine de la création des différentes méthodes d'analyse factorielle. Historiquement, l'Analyse en Composantes Principales (ACP), dite classique ou usuelle, est la première apparue (Hotelling, 1901, Spearman, 1904) pour traiter des mesures sur des variables *quantitatives*. Cependant d'autres méthodes d'analyse factorielle portant sur des variables *qualitatives* ou *booléennes*, par exemple les méthodes d'analyse d'une enquête, procèdent du même principe : celui de la **réduction de la dimension**, c'est à dire, la mise en évidence de **variables latentes** ou **Composantes Principales**, en petit nombre, qui résument les variables mesurées, au sens où **ces nouvelles variables synthétiques sont des combinaisons linéaires des variables originelles dont les poids sont appelés les facteurs principaux**. On emploie aussi de ce fait, la dénomination d'analyse factorielle linéaire.
L'objectif de ce chapitre est de définir vocabulaire et notations aptes à mettre en évidence les analogies entre la géométrie Euclidienne et la statistique

6.1 Le triplet (T, M, D) des données

Un jeu de données est constitué par un triplet (T, M, D) défini par les trois éléments suivants.

- $T = [T_i^j] \in I\!R^{n \times p}$ est la matrice des données brutes exprimant les n mesures de p variables, $x^1, \ldots, x^p$, par exemple quantitatives. Le tableau T pourra aussi être obtenu à partir des résultats d'une enquête; on verra plus loin dans ce cas, la nature des données.

- M, $p \times p$, est une métrique Euclidienne sur l'espace $I\!R^p$ des lignes de T.
  La $i$ème ligne de T, notée T_i, sera considérée comme l'expression dans la base canonique de l'espace $(I\!R^p, M)$, de l'échantillon du $i$ème individu.

- D, métrique sur l'espace Euclidien $I\!R^n$ des colonnes de T, est une matrice, $n \times n$, qui sera toujours diagonale $D = \text{diag}(p_1, \ldots, p_n)$.
  La $j$ème colonne de T, notée T^j, sera considérée comme l'expression dans la base canonique de $(I\!R^n, D)$, de l'échantillon de la $j$ème variable, x^j.

Les espaces Euclidiens $(I\!R^n, D)$ et $(I\!R^p, M)$ considérés sont respectivement les espaces des individus et des variables. Ces espaces sont soit des espaces vectoriels soit des espaces affines selon que l'on parlera de vecteurs ou de points, un point origine ayant alors été choisi au préalable. Par abus de langage, on confondra parfois lorsqu'il n'y aura pas d'ambiguté, un point-vecteur ligne (respectivement un point-vecteur colonne) avec la matrice ligne (colonne) de son expression dans la base canonique.
D'autre part, on notera r le rang de T, $r \leq \min(n, p)$. L'espace des variables échantillonnées, $Im\ T$, et l'espace des individus échantillonnés, $Im\ T'$, sont de même dimension r

$$dim\ Im\ T = dim\ Im\ T' = r.$$

6.2 Statistique et géométrie sur $(\mathbb{R}^n, D)$, espace des variables

6.2.1 Le simplexe des poids statistiques et la droite des constantes

Dans les applications, les poids p_i sont les poids statistiques des individus. La matrice diagonale

$$\begin{cases} D = \mathrm{diag}(p_1, \ldots, p_n) \\ \mathrm{trace}(D) = \sum_i p_i = 1 \\ p_i > 0 \quad \text{pour} \quad i = 1 \ldots n, \end{cases}$$

est appelée la matrice des poids statistiques affectés aux n individus.

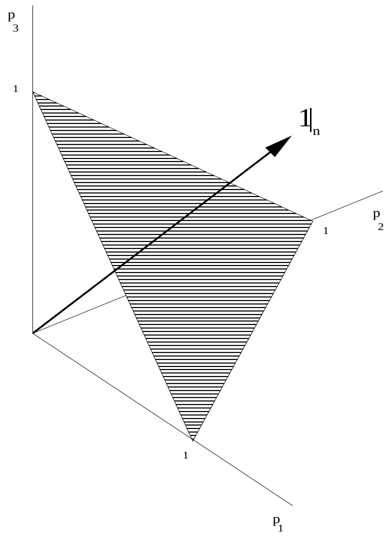


Figure 8 : simplexe des poids statistiques ($n = 3$).

On appelle simplexe des poids l'ensemble

$$\mathcal{D} = \{(p_1, \ldots, p_n) \in \mathbb{R}^n \mid p_i \geq 0,\ i = 1, \ldots, n \text{ et } \textstyle\sum_i p_i = 1\}.$$

Remarquer que si un poids est nul, la matrice D n'est que semi-définie positive et ne peut plus être considérée comme une métrique sur $\mathbb{R}^n$. Le cas le plus usité en statistique est celui des poids uniformément répartis

$$\forall i \quad p_i = \frac{1}{n}, \quad \text{et} \quad D = \frac{1}{n} I_n .$$

Dans ce cas la projection D-orthogonale concide avec la projection au sens usuel.

Soit $\mathbf{1}_n = \sum_i e_i(n) = [1, \ldots, 1]'$. Le vecteur $\mathbf{1}_n$ est I_n-orthogonal au plan vectoriel $\mathbf{1}_n' x = x_1 + \ldots + x_n = 0$ qui est associé au plan affine $x_1 + \ldots + x_n = 1$ contenant $\mathcal{D}$. De plus, ce vecteur est de D-norme égale à 1, en effet

$$\text{trace}(D) = \mathbf{1}_n' D \mathbf{1}_n = \|\mathbf{1}_n\|_D^2 = 1.$$

On appelle "droite des constantes" la droite vectorielle engendrée par $\mathbf{1}_n$.

P1 Le projecteur D-orthogonal sur $\mathbf{1}_n$ est égal à

$$P_{1_n} = \mathbf{1}_n(\mathbf{1}_n' D \mathbf{1}_n)^{-1}\mathbf{1}_n' D = \mathbf{1}_n \mathbf{1}_n' D = \begin{bmatrix} p_1 & p_2 & \dots & p_n \\ p_1 & p_2 & \dots & p_n \\ . & . & \dots & . \\ p_1 & p_2 & \dots & p_n \end{bmatrix}.$$

C'est une matrice $n \times n$ de rang égal à 1. $\square$

Dans le cas usuel de poids statistiques égaux,

$$P_{1_n} = n^{-1}\mathbf{1}_n \mathbf{1}_n{}'.$$

6.2.2 Moyenne et centrage vus comme une projection

Soit $t = [t_1, \dots, t_n]'$ un n-échantillon d'une variable statistique munie des poids statistiques définis par la diagonale de D.

La moyenne de t, notée $\bar{t}$, est définie par le D-produit scalaire entre t et le vecteur $\mathbf{1}_n$. En effet,

$$\bar{t} = \textstyle\sum_i p_i t_i = \mathbf{1}_n' D t = < \mathbf{1}_n, t >_D .$$

La projection du vecteur t sur $\mathbf{1}_n$ a pour mesure algébrique $\bar{t}$. En effet,

$$P_{1_n} t = \mathbf{1}_n \mathbf{1}_n{}' D t = \mathbf{1}_n \bar{t} = \bar{t}\mathbf{1}_n.$$

P2 Centrer une variable t au sens de D, on dira D-centrer t, c'est construire

$$x = \begin{bmatrix} t_1 - \bar{t} \\ t_2 - \bar{t} \\ \dots \\ t_n - \bar{t} \end{bmatrix} = t - \bar{t}\mathbf{1}_n = (I_n - P_{1_n})u = P_{1_n}^{\perp} t$$

c'est à dire projeter t sur le sev de $\mathbb{R}^n$ D-orthogonal à $\mathbf{1}_n$. $\square$

Preuve :

$$x = t - \bar{t}\,\mathbf{1}_n = t - \ \mathbf{1}_n \bar{t} = t - P_{1_n} t. \quad \square$$

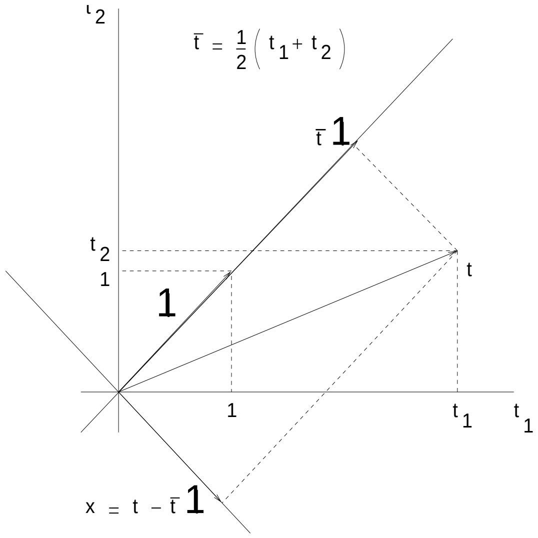


Figure 9 : La moyenne vue comme une projection sur la droite des constantes ($n = 2$).

6.2.3 Variance et écart-type

La moyenne $\bar{t}$ définit un “paramètre de position” pour la variable t à partir de la géométrie euclidienne basée sur la métrique diagonale des poids statistiques. Caractériser la position d'une variable statistique n'a de sens que si on lui associe un autre paramètre statistique caractérisant la “dispersion autour de la moyenne” appellé l'écart-type ou la déviation standard. Ce paramètre est défini à partir de la variance de t.

La variance de t, on dit aussi la D-variance ou encore la variance géométrique ou empirique, notée $var(t)$, est définie par

$$var(t) = \sum_{i=1}^{n} p_i (t_i - \bar{t})^2 = var(x).$$

L'écart-type, noté $\sigma(t)$, est défini par

$$\sigma(t) = \sqrt{var(t)} = \sigma(x).$$

Il s'exprime dans la même unité de mesure que celle des observations, $t_1, \ldots, t_n$, de la variable t.

P3 La variance de t est le carré de la D-norme de la variable centrée x,

$$var(t) = \|x\|_D^2.$$

Formule développée: $var(t) = \|t - \bar{t}\mathbf{1}_n\|_D^2 = \|t\|_D^2 - \bar{t}^2 = \sum_i p_i t_i^2 - \bar{t}^2$.
L'écart-type est la norme de x, $\sigma(t) = \sqrt{var(t)} = \|x\|_D = \sigma(x)$. □
Preuve : Par définition, $var(t) = \sum_i p_i(t_i - \bar{t})^2 = x'Dx = \|x\|_D^2$. □

Le vecteur $x/\|x\|_D$ est appelé la variable déduite de t par centrage et réduction (au sens de D), ont dit aussi variable réduite ou standardisée. Bien sûr, la moyenne d'une variable réduite est nulle, sa variance et son écart-type sont égaux à 1.

6.2.4 Proximité entre deux variables, covariance et corrélation linéaire

Soient deux n-échatillons $t = [t_1, \ldots, t_n)]'$ et $u = [u_1, \ldots, u_n)]'$ munis des mêmes poids statistiques de la diagonale de D. Notons x et y les deux variables centrées respectivement associées à t et u. On mesure la proximité entre t et u par le coefficient de corrélation linéaire $r(t, u)$ dont la définition découle de celle de la covariance entre ces deux variables.

La covariance netre t et u est définie par

$$cov(t, u) = \sum_{i=1}^{n} p_i(t_i - \bar{t})(u_i - \bar{u}) = \sum_{i=1}^{n} p_i\, x_i y_i = cov(x, y).$$

De façon évidente, $cov(t, u) = cov(u, t)$. La variance d'une variable peut être définie à partir de la covariance, $var(t) = cov(t, t)$, ce qui est la conséquence de l'interprétation géométrique suivante

P4 La covariance entre deux variables t et u est le D-produit scalaire entre les variables centrées, $cov(t, u) = x'Dy =< x, y >_D= cov(x, y)$.
Formule développée: $cov(t, u) =< t, u >_D - \bar{t}\bar{u} = \sum_i p_i t_i u_i - \bar{t}\bar{u}$. □
Preuve : Evidente. □

Le coefficient de corrélation linéaire entre t et u, noté $r(t, u)$, est défini par

$$r(t, u) == \frac{cov(t, u)}{\sigma(t)\sigma(u)} = r(x, y).$$

P5 Le coefficient de corrélation linéaire s'interprète dans $I\!R^n$ comme le cosinus du D-angle formé par les vecteurs D-centrés.

$$r(t, u) = \frac{cov(t, u)}{\sigma(t)\sigma(u)} = \frac{< x, y >_D}{\|x\|_D \|y\|_D},$$

Le théorème de Cauchy-Schwarz donne

$$-1 \leq r(t,u) \leq 1.$$

L'égalité n'a lieu que si x et y sont colinéaires. $\square$

On dit que t et u sont non corrélés si $r(t,u) = 0$ c'est à dire si les variables centrées x et y sont D-orthogonales.

P6 Droite de régression linéaire et coefficient de corrélation

La propriété **P5** de $r(t,u)$ fournit l'interprétation géométrique de la proximité entre t et u dans l'espace $(\mathbb{R}^n, D)$ qui est l'espace Euclidien des variables. Dans ce cas bivarié, on peut visualiser l'interprétation de $r(t,u)$ dans l'espace $(\mathbb{R}^2, I_2)$ des individus. Le nuage $\mathcal{N}$ des n points-individus M_i de coordonnées (t_i, u_i) dans la base canonique $\{\vec{e_1}, \vec{e_2}\}$, et de poids respectifs p_i est représenté Figure 10.

On appelle droite de régression, la droite Δ passant par G, individu moyen de coordonnées $(\overline{t}, \overline{u})$, et d'équation $y = \hat{a}x$ dans le repère (x, y). Le coefficient angulaire, $\hat{a}$, de Δ est solution du prolème de la minimisation de la fonction φ de $\mathbb{R}$ dans $\mathbb{R}$ définie par

$$\varphi(a) = \sum_{i=1}^{n} p_i (y_i - a x_i)^2 = \|y - ax\|_D^2.$$

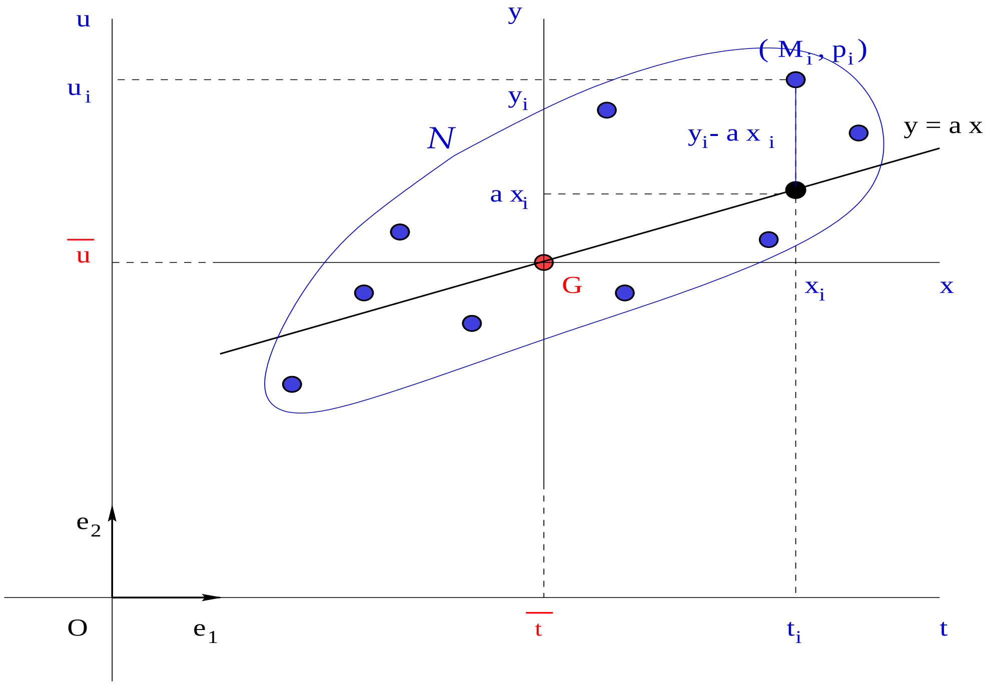


Figure 10: Ajustement linéaire d'un nuage de points bivariés.

Cette fonction mesure la somme pondérée des carrés des écarts verticaux entre les individus et leurs homologues sur une droite quelconque passant par G et de pente a. En quelque sorte, $\varphi(a)$ mesure l'écart vertical entre le nuage $\mathcal{N}$ des points pesants et une telle droite. Le minimum de φ est atteint pourla valeur $\hat{a}$ qui définit, dans le repère (t, u), la droite de régression ou d'ajustement linéaire Δ.

$$\text{Equation de } \Delta \; : \; \begin{cases} u = \hat{a}\, t + \hat{b} \\ \hat{a} = \dfrac{cov(t,u)}{var(t)} \\ \overline{u} = \hat{a}\, \overline{t} + \hat{b} \end{cases}$$

La valeur minimale de φ s'exprime en fonction de $r(t, u)$

$$\varphi(\hat{a}) = var(u)[1 - r^2(t, u)],$$

ce qui permet de mesurer grâce à $r(t, u)$ l'ajustabilité du nuage par une droite.

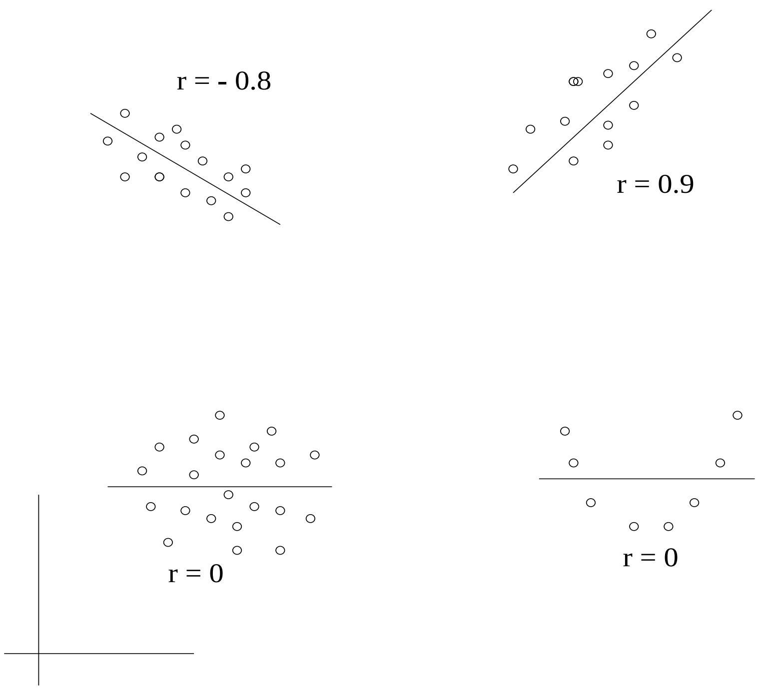


Figure 11: Ajustabilité d'un nuage bivarié suivant les valeurs de r.

Remarquer que lorsque t et u sont non corrélés, $cov(t, u) = 0$, Δ a pour équation $u = \overline{u}$.

6.2.5 Définitions et notations pour la statistique multivariée

On dispose d'un échantillon de mesures de p variables toutes mesurées sur les mêmes n individus. Ces mesures sont stockées dans la matrice T des données brutes. On suppose que les n individus sont munis des poids statistiques formés par la diagonale de D.

On appelle individu moyen le vecteur ligne, noté $\overline{T}$, formé des moyennes de p variables (colonnes de T).

$$\overline{T} = [\overline{t}_1, \ldots, \overline{t}_p] = \mathbf{1}'_n DT.$$

La matrice X obtenue par centrage des variables est définie par

$$X = (I_n - \mathbf{1}_n \mathbf{1}'_n D)T.$$

Elle s'interprète du point de vue des individus (lignes) comme la différence entre chaque individu et l'individu moyen.

On est amené à étudier les liens de proximité entre les variables prises deux à deux :

La matrice des covariances géométriques est définie par

$$\mathbb{V} = [\mathbb{V}_{ij} = cov(t^i, t^j)] = X'DX,$$

où t^i est la $i$ème colonne de T.

Cette matrice carrée d'ordre p, est symétrique semi définie positive, $rang(V) = rang(T)$. Elle est définie positive si $rang(T) = p$. Remarquer que $V_{ii} = var(t^i)$.

Standardiser les p variables, c'est à dire D-centrer réduire les variables sur l'échantillon, revient à construire

$$Z = XQ$$

où $Q = \text{diag}(\sigma_1^{-1}, \ldots, \sigma_p^{-1})$. La matrice des corrélations entre les variables est

$$R = [R_{ij} = r(t^i, t^j)] = Z'DZ = Q'VQ. \quad \square$$

Remarquer que la diagonale de R est formée de 1 car $r(t, t) = 1$.

P7 Supposons Y $n \times 1$ et $X = [X^1 | \ldots | X^p]$ $n \times p$ deux matrices D-centrées. Soit $\widehat{Y} = P_X Y$ la projection D-orthogonale de Y sur ImX. Puisque X est D-centrée, $\widehat{Y}$ est de moyenne nulle et $\|\widehat{Y}\|_D^2 = var(\widehat{Y})$. Le coefficient de détermination entre Y et ImX défini dans la section 4.2.2, s'écrit

$$R^2(Y, ImX) = \frac{var(\widehat{Y})}{var(Y)}.$$

De plus,

$$R^2(Y, ImX) = \frac{Y'D\widehat{Y}}{Y'DY} = \frac{cov(Y, \widehat{Y})}{var(Y)} = r^2(Y, \widehat{Y}).$$

Preuve : C'est la conséquence de la D-symétrie du projecteur P_X, c'est à dire $DP_X = (P_X)'D$. Alors, $\|\widehat{Y}\|_D^2 = Y'(P_X)'DP_XY = Y'DP_XP_XY = Y'D\widehat{Y}$. $\square$

On peut montrer, voir exercice, que

$$R^2(Y, ImX) = \max_{W \in ImX} r^2(Y, W)\,,$$

et que le maximum est réalisé pour $W = P_XY$.

On utilise aussi

$$R(Y, ImX) = \sqrt{R^2(Y, ImX)} = r(Y, \widehat{Y})$$

qui s'appelle le coefficient de corrélation multiple entre Y et ImX.

6.3 Exercices

<u>**Exercice 1**</u> : **Régression linéaire simple de u sur t.**
Dans le contexte de la propriété **P6** de la section 6.2.4, la régression linéaire de u sur t est présentée sur les variables centrées y et x respectivement, cet exercice la présente maintenant sur les variables initiales. On dispose de deux n-échantillons, t variable explicative et u variable réponse, dont les observations sont munies des poids statistiques $\{p_i \mid i = 1, \ldots, n\}$ stockés dans la diagonale de D.
L'objectif est de minimiser la fonction ϕ de $I\!R^2$ dans $I\!R^+$ définie par

$$\phi(a, b) = \sum_{i=1}^{n} p_i(u_i - at_i - b)^2.$$

Posons $X = [\mathbf{I}_n\ t]$, matrice $n \times 2$, et $\beta = \begin{bmatrix} b \\ a \end{bmatrix}$ vecteur colonne des inconnues a et b.
1) Montrer que $\phi(\beta) = \|u - X\beta\|_D^2$. Calculer $X'DX$, $(X'DX)^{-1}$ et $X'Du$.
2) Soit $\hat{\beta} = \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix}$ solution du problème. Calculer $\hat{\beta}$ en utilisant les résultats de la section 4.2.2. Vérifier que l'on retrouve les résultats de **P6**. Quelle est l'interprétation géométrique de $\hat{u} = X\hat{\beta}$ dans $(I\!R^n, D)$?
3) Eprimer $\phi(\hat{\beta})$ en fonction de $r(t, u)$.

<u>**Exercice 2**</u> : Dans le contexte de la propriété **P7** de la section 6.2.5, on se propose de montrer que

$$R^2(Y, ImX) = \max_{W \in ImX} r^2(Y, W)\,,$$

et que le maximum est réalisé pour $W = P_X Y$.

1) Quelle est l'interprétation géométrique de cette propriété?

2) Calculer le vecteur gradient de l'application φ de $\mathbb{R}^p$ dans $\mathbb{R}$

$$v \longrightarrow \varphi(v) = \frac{(Y'DXv)^2}{\|Y\|_D^2 (v'X'DXv)} .$$

Soit $\nabla\varphi(v)$ ce vecteur. Montrer que l'équation $\nabla\varphi(v) = 0$ s'écrit

$$P_X P_Y W = \varphi(v) W .$$

En déduire que le vecteur $W = Xv$ optimal est vecteur propre de $P_X P_Y$ associé à la plus grande valeur propre.

3) Si $\hat{Y} = P_X Y$, montrer que $P_X P_Y$ admet une seule valeur propre non nulle égale à

$$\frac{Y'D\hat{Y}}{\|Y\|_D^2} = R^2(Y, ImX) .$$

Vérifiez enfin que $W = \hat{Y}$ est le vecteur propre associé à cette valeur propre.

Chapter 7

Généralisation de la Décomposition en Valeurs Singulières. Analyse en Composantes Principales du triplet (X, M, D)

Toutes les méthodes de l'Analyse Factorielle de l'Analyse de Données peuvent être présentées dans un cadre commun : celui de l'extension du théorème de la Décomposition en Valeurs Singulières (DVS) au cadre d'espaces Euclidiens plus généraux. La présentation synthétique qui va suivre, est basée sur l'introduction de métriques sur les espaces Euclidiens envisagés. Le choix d'une métrique permettra d'adapter cette technique générale, appelée ACP du triplet (X, M, D), au problème posé par le type de données à traiter.

Historiquement, la première méthode apparue pour analyser un tableau issu de mesures sur variables quantitatives, est l'Analyse en Composantes Principales (ACP) dite usuelle. Elle correspond au triplet

- X matrice, $n \times p$, des variables centrées (éventuellement réduites),
- $M = I_p$, métrique usuelle sur l'espace des lignes,
- $D = n^{-1} I_n$, métrique sur l'espace des variables, formée par la matrice diagonale des poids égaux pour les individus.

On verra que d'autres méthodes d'analyse des données nécessitent des choix différents, en particulier lorsque les données sont issues du dépouillement d'une enquête statistique.

Toutes ces méthodes rentrent dans le cadre de la décomposition en valeurs singulières du triplet (X, M, D).

7.1 Décomposition en Valeurs Singulières du triplet

Dans la DVS usuelle examinée au chapitre 2, les matrices $X'X_{p\times p}$ et $XX'_{n\times n}$, symétriques, jouent un rôle fondamental. Dans la DVS du triplet (X, M, D), ce rôle va être attribué repectivement aux matrices $X'DXM_{p\times p}$ et $XMX'D_{n\times n}$. Ces matrices ne sont pas symétriques, sauf dans le cas où M et D sont de la forme kI comme dans la DVS usuelle et dans le cas de l'ACP usuelle. Elles sont respectivement M et D-symétriques. Il est d'autre part nécessaire de s'assurer que les valeurs propres de telles matrices sont non-négatives et que les vecteurs propres sont *orthogonaux au sens de la métrique concernée.* C'est l'objectif du Lemme suivant.

7.1.1 Lemme

La matrice $X'DXM$ (resp. $XMX'D$) est M-symétrique (resp. D-symétrique), ses r valeurs propres non-nulles sont réeelles positives et ses vecteurs propres forment une base M-orthonormée de $Im\,X'$ (resp. D-orthonormée de ImX). $\square$

Preuve :
Une matrice carrée A est M-symétrique si et seulement si $MA = A'M$, ce qui est le cas pour $X'DXM$ à cause de la symétrie de M et de D. La matrice M étant symétrique définie-positive, soit $M = M^{1/2}M^{1/2}$ sa décomposition par la DVS ($M^{1/2}$ est symétrique définie positive). On peut aussi utiliser Cholesky : $M = TT'$, où T est triangulaire inférieure à diagonale positive. Notons $\Lambda_r = diag(\lambda_1, \dots, \lambda_r)$ et $V = [V^1, \dots, V^r]$ les matrices des valeurs propres non-nulles et des vecteurs proppres associés pour $X'DXM$ de rang r

$$X'DXMV = V\Lambda_r$$

$$X'DXM^{1/2}(M^{1/2}V) = V\Lambda_r$$

$$M^{1/2}X'DXM^{1/2}(M^{1/2}V) = (M^{1/2}V)\Lambda_r$$

On pose $Y = M^{1/2}V$ ou $V = M^{-1/2}Y$. La matrice $M^{1/2}X'DXM^{1/2}$ est symétrique semi-définie positive de rang r, ses r valeurs propres non-nulles sont réelles positives et ses r vecteurs propres $\{Y^j\}$ sont orthonormés au sens usuel (métrique identité). Il en résulte

$$I_r = Y'Y = V'(M^{1/2})'M^{1/2}V = V'MV.$$

Puisque $V = X'DXMV\Lambda_r^{-1}$, $Im\, V \subset Im\, X'$ et comme $rang(V) = rang(X') = r$, il en résulte que les deux espaces vectoriels coincident, $Im\, V = Im\, X'$. $\square$

Remarque : La construction effective des vecteurs propres $\{V^j\}$ de $X'DXM$ passe d'abord par le calcul des vecteurs propres $\{Y^j\}$ de $M^{1/2}X'DXM^{1/2}$ puis par le calcul de $V^j = M^{-1/2}Y^j$.

7.1.2 La DVS du triplet (X, M, D)

Théorème : Soient $X \in I\!R^{n\times p}$ de rang r, M métrique sur $I\!R^p$ et D métrique sur $I\!R^n$. Il existe

- $U_{n\times r}$, D-orthonormée ($U'DU = I_r$), dont les colonnes sont les vecteurs propres associés aux valeurs propres $\{\lambda_1, \ldots, \lambda_r\}$ non-nulles de $XMX'D$,
- $V_{p\times r}$, M-orthonormée ($V'MV = I_r$), dont les colonnes sont les vecteurs propres associés aux mêmes valeurs propres $\{\lambda_1, \ldots, \lambda_r\}$ non-nulles de $X'DXM$,
- $\Lambda_r^{1/2} = diag(\lambda_1^{1/2}, \ldots, \lambda_r^{1/2})$ matrice diagonale des valeurs singulières $\{\lambda_i^{1/2}\}$ du triplet (X, M, D),

telles que X s'écrive

$$X = U\Lambda_r^{1/2}V' = \sum_{i=1}^{r} \sqrt{\lambda_i} U^i V^{i\prime}. \quad \square$$

Les valeurs singulières sont toujours classées par ordre décroissant

$$\lambda_1^{1/2} \geq \ldots \geq \lambda_r^{1/2} > 0$$

ce qui induit un classement sur les colonnes de $U = [U^1, \ldots, U^r]$ et de $V = [V^1, \ldots, V^r]$.

Preuve :
La décomposition spectrale de $X'DXM$ s'écrit d'après le Lemme précédent

$$\begin{cases} X'DXMV = V\Lambda_r \\ V'MV = I_r\,. \end{cases}$$

Posons

$$U = XMV\Lambda_r^{-1/2}, \qquad (*)$$

appelée *première formule de transition* qui permet d'exprimer U en fonction de V.

Propriétés de U

- Les colonnes de U sont vecteurs propres de la matrice $XMX'D$ qui à les mêmes valeurs propres non-nulles que $X'DXM$:

$$\begin{aligned} XMX'DU &= XMX'DXMV\Lambda_r^{-1/2} \\ &= XMV\Lambda_r\Lambda_r^{-1/2} \\ &= XMV\Lambda_r^{-1/2}\Lambda_r \\ &= U\Lambda_r\,. \end{aligned}$$

- Les colonnes de U forment une base D-orthonormée de $Im\,X$ (preuve immédiate).

Montrons que $X = U\Lambda_r^{1/2}V'$. Pour cela, notons $\Pi_V^M = VV'M$, le projecteur M-orthogonal sur $Im\,V = Im\,X'$.
$U\Lambda_r^{1/2}V' = XMV\Lambda_r^{-1/2}\Lambda_r^{1/2}V' = XMVV' = (\Pi_V^M X')' = (\Pi_{X'}^M X')' = (X')' = X$. □
Deuxième formule de transition : De façon duale, on peut aussi démontrer le théorème de la DVS de X en partant de la décomposition spectrale de la matrice $XMX'D$. On construit la deuxième formule de transition qui exprime V en fonction de U

$$V = X'DU\Lambda_r^{-1/2}\,. \qquad (**)$$

Corollaire : décomposition des matrices $\mathbb{V} = X'DX$ et $\mathbb{W} = XMX'$

$$\mathbb{V} = V\Lambda_r V' = \sum_{i=1}^{r} \lambda_i V^i V^{i\prime} \quad \text{et} \quad \mathbb{W} = U\Lambda_r U' = \sum_{i=1}^{r} \lambda_i U^i U^{i\prime}\,.$$

7.1.3 Relation avec la DVS usuelle

La DVS usuelle de X est la DVS du triplet

$$(X, M = I_p, D = I_n)\,.$$

La DVS usuelle déjà étudiée au chapitre 2, correspond au cadre Euclidien naturel pour les espaces des lignes et des colonnes du tableau X.

Relation entre la DVS de (X, M, D) et de $(Z = D^{1/2}XM^{1/2}, I_p, I_n)$
La DVS du triplet (X, M, D) est équivalente à la DVS usuelle de la matrice

$$Z = D^{1/2}XM^{1/2}$$

au sens suivant : toutes deux ont les mêmes valeurs singulières; si $Z = U_z\Lambda_r^{1/2}V_z'$ et $X = U_x\Lambda_r^{1/2}V_x'$ sont les deux décompositions, alors $U_x = D^{-1/2}U_z$ et $V_x = M^{-1/2}V_z$. □
Preuve : Exercice 1. □

7.1.4 Projecteurs orthogonaux associés à la DVS

Les colonnes de $V = [V^1, \ldots, V^r]$ forment une base M-orthonormée de $Im\, X'$, celles de $U = [U^1, \ldots, U^r]$ forment une base D-orthonormée de $Im\, X$. Il en résulte l'expression des projecteurs orthogonaux sur ces espaces vectoriels de dimension r.
Soit $X = U\Lambda_r^{1/2}V'$ la DVS de (X, M, D). Le projecteur M-orthogonal sur $Im\, X'$ et le projecteur D-orthogonal sur $Im\, X$ sont donnés par :

$$\Pi_{X'}^M = V(V'MV)^{-1}V'M = VV'M = \sum_{i=1}^r V^i V^{i\prime} M = \sum_{i=1}^r \Pi_{V^i}^M ,$$

$$\Pi_X^D = U(U'DU)^{-1}U'D = UU'D = \sum_{i=1}^r U^i U^{i\prime} D = \sum_{i=1}^r \Pi_{U^i}^D .$$

Notons que lorsqu'on dispose d'une base orthogonale d'un sous-espace vectoriel, le projecteur sur cet espace se décompose en la somme des projecteurs sur chacun des espaces de dimension 1 engendrés par les vecteurs de base.

7.1.5 Théorème d'approximation d'Eckart-Young

Théorème : Soient $X \in I\!R^{n\times p}$ de rang r, $X = U\Lambda_r^{1/2}V'$ la DVS de (X, M, D) et k un entier, $k \leq r$. On note $U_k = [U^1, \ldots, U^k]$ et $V_k = [V^1, \ldots, V^k]$ les matrices extraites de U et de V, et $\Lambda_k^{1/2} = diag(\lambda_1^{1/2}, \ldots, \lambda_k^{1/2})$ la matrice diagonale des k premières valeurs singulières.
On cherche un élément $\widehat{X}_k$ de $E_k = \{X_k \in I\!R^{n\times p} \,|\, rang(X_k) = k\}$, le plus proche de X au sens de la norme $\|.\|_{M\otimes D}$. Alors,

$$\min_{X_k \in E_k} \|X - X_k\|_{M\otimes D}^2 = \|X - \widehat{X}_k\|_{M\otimes D}^2 = \sum_{i=k+1}^r \lambda_i ,$$

l'optimum étant atteint par la DVS incomplète de rang k,

$$\widehat{X}_k = U_k \Lambda_k^{1/2} V_k' . \quad \square$$

Les valeurs propres (carrés des valeurs singulières) étant classées par ordre décroissant, le carré de l'erreur lors de l'approximation de X par $\widehat{X}_k$ est la somme des plus petites valeurs propres restantes.
Preuve :
Admettons le résultat dans le cadre Euclidien usuel, voir Exercice 2 : pour toute matrice

Z, $n \times p$, de rang r, élément du triplet (Z, I_p, I_n)

$$\min_{Z_k \in E_k} \|Z - Z_k\|_F^2 = \|Z - \widehat{Z}_k\|_F^2 = \sum_{i=k+1}^{r} \lambda_i\,.$$

L'optimum est atteint par la DVS incomplète de rang k de Z

$$\widehat{Z}_k = [U_z]_k \Lambda_k^{1/2} [V_z]_k{}'\,.$$

Posons $Z = D^{1/2} X M^{1/2}$ et notons que E_k est l'ensemble des matrices $X_k = D^{-1/2} Z_k M^{-1/2}$ pour toute matrice Z_k de rang k. De façon évidente

$$\|Z - Z_k\|_F^2 = trace((Z - Z_k)'(Z - Z_k)) = trace((X - X_k)' D (X - X_k) M) = \|X - X_k\|_{M \otimes D}^2.$$

L'équivalence de la DVS des triplets (X, M, D) et (Z, I_p, I_n) au sens de la Section 6.2.3, implique que la solution

$$\widehat{X}_k = D^{-1/2} \widehat{Z}_k M^{-1/2} = D^{-1/2} [U_z]_k \Lambda_k^{1/2} [V_z]_k{}' M^{-1/2} = [U_x]_k \Lambda_k^{1/2} [V_x]_k{}'\,,$$

est fournie par la DVS incomplète de rang k de X. $\square$

7.2 Analyse en Composantes Principales d'ordre k du triplet (X, M, D)

L'Analyse en Composantes Principales, en bref ACP, usuelle associée au choix $M = I_p$ et $D = n^{-1} I_n$, est une méthode d'analyse exploratoire de données multivariées dont l'un des objectifs est la vision plane à deux dimensions des points lignes et des points colonnes (photos obtenues par projection sur des plans dits factoriels).

Le fait d'envisager des métriques plus générales introduit une distorsion dans la représentation des distances, voir les remarques de la section 2.6. et la section 6.2.3. Cependant, dans la plupart des méthodes factorielles, outre D, la métrique M est diagonale. Dans ce cas, pour une vision naturelle d'un point il suffit de multiplier chaque coordonnée i par la racine carrée du $i$ème élément diagonal de la métrique. Cela est cependant inutile dans l'ACP usuelle où tous les points lignes ont le même poids ainsi que les points colonnes. Dans l'Analyse Factorielle des Correspondances (AFC) simple ou multiple, des transformations sur les données sont effectuées pour que les distances Euclidiennes associées au triplet correspondent à la distance du χ^2 entre vecteurs des fréquences conditionnelles des données d'une enquête. La vision naturelle des points n'est pas, dans ce cas, l'objectif à atteindre.

Les plans factoriels de projection ne sont autres que ceux formés par les couples de vecteurs des bases orthonormées de la DVS du triplet : (V^i, V^j) pour voir les points lignes, ou (U^i, U^j) pour voir les poins colonnes. Reste à décider quels sont les k meilleurs plans factoriels, c'est à dire ceux pour qui les photos obtenues seront porteuses d'informations interprétables : l'ACP d'ordre k est définie à partir de la DVS en éliminant la part de bruit incluse dans les données et mesurée grâce au théorème d'Eckart-Young.

7.2.1 Définitions

La matrice X est supposée D-centrée en colonnes à partir d'une matrice T des données brutes

$$X = (I_n - \mathbb{1}_n \mathbb{1}'_n D)T \,.$$

Le point origine de l'espace des points lignes de X s'interprète comme le point ligne moyen, $\mathbb{1}'_n DT$, du tableau T. La matrice des covariances entre les variables est $\mathbb{V} = X'DX$, celle des produits scalaires entre les individus $\mathbb{W} = XMX'$. A ces matrices sont associés les opérateurs aux valeurs propres-vecteurs propres de la DVS du triplet.

Opérateurs et inertie du triplet (X, M, D)

- Opérateur des covariances : $\mathbb{V}M = X'DXM$
- Opérateur des produits scalaires entre individus : $\mathbb{W}D = XMX'D$
- Inertie totale du triplet : $\|X\|^2_{M\otimes D} = trace(XMX'D) = trace(X'DXM)$

Expression tirée de la terminologie de la mécanique du point matériel, l'"inertie totale" des n points lignes M_i pesant chacun p_i

$$\sum_{i=1}^{n} p_i \|X_i\|^2_M = \sum_{i=1}^{n} p_i X_i M X'_i = trace(XMX'D) = \|X\|^2_{M\otimes D} \,,$$

est la mesure du moment d'inertie du nuage des n points par rapport à l'origine des coordonnées, ici le point ligne moyen. Cette expression mesure l'éloignement de l'origine des points M_i par les carrés de leurs distances pondérés par les poids statistiques.

Dans le cas particulier $M = I_p$, l'inertie totale trouve une interprétation duale par rapport aux colonnes de X. En effet, $trace(X'DX) = \sum_{j=1}^{p} \mathbb{V}_j^j = \sum_{j=1}^{p} \sum_{i=1}^{n} p_i (X_i^j)^2 =$

$\sum_{j=1}^{p}\sum_{i=1}^{n} p_i(T_i^j - \overline{T}^j)^2$ est aussi la variance totale c'est à dire la somme des variances des p variables. Si de plus les variables sont D-centrées réduites l'inertie totale est dans ce cas, égale à p.

Proposition :

$$\|X\|^2_{M\otimes D} = trace(\Lambda_r) = \sum_{i=1}^{r} \lambda_i \,.$$

Preuve : utiliser la DVS du triplet et l'orthogonalité des matrices U et V. $\square$

ACP d'ordre k du triplet (X, M, D)

La matrice X étant supposée de rang r et D-centrée en colonne, on appelle ACP d'ordre k, $k \leq r$, du triplet (X, M, D), la DVS incomplète de rang k

$$\widehat{X}_k = U_k \Lambda_k^{1/2} V_k' \,,$$

telle qu'elle est définie dans le théorème d'Eckart-Young.

Les deux formules de transition s'écrivent à l'ordre k

$$U_k = XMV_k\Lambda_k^{-1/2}\ (*) \qquad \text{et} \qquad V_k = X'DU_k\Lambda_k^{-1/2}\ (**)\,.$$

Proposition : L'approximation de rang k de X a pour colonnes (pour lignes) les projections des colonnes (des lignes) de X sur l'espace vectoriel $Im\,U_k$ (sur $Im\,V_k$)

$$\widehat{X}_k = \Pi^D_{U_k} X \qquad \text{et} \qquad \widehat{X}_k{}' = \Pi^M_{V_k} X'\,. \quad \square$$

Preuve : Faisons la preuve pour les colonnes. La deuxième formule de transition (**) donne

$$\begin{aligned}\Pi^D_{U_k} X &= U_k U_k' D X \\ &= U_k (X'DU_k)' \\ &= U_k \Lambda_k^{1/2} V_k' \,. \quad \square\end{aligned}$$

Au delà de la définition précédente, l'ACP est une méthode d'analyse et d'exploration des données basée sur l'examen des points lignes et des points colonnes projetés sur des espaces de dimensions 1 ou 2 obtenus à partir des bases orthonormées $\{V_1, \ldots, V_k\}$ et $\{U_1, \ldots, U_k\}$ des deux espaces respectifs $Im\,X_k{}'$ et $Im\,X_k$.

Grâce aux posibilités actuelles des logiciels informatiques concernant la vision à trois dimensions, il est parfois intéressant de visionner les projections dans les repères $\{V_1, V_2, , V_3\}$ et $\{U_1, U_2, U_3\}$.

Cas particulier : l'ACP usuelle réduite d'ordre k

L'ACP usuelle réduite, ou ACP normée, est l'ACP d'ordre k du triplet

$$(X, M = I_p, D = \frac{1}{n} I_n)$$

où X est formée par les n mesures de p variables quantitatives D-centrées réduites.

Dans ce cas la matrice des covariances $\mathbb{V} = \frac{1}{n} X'X$ est la matrice des corrélations entre les p variables. Parfois, lorsque les variables sont "homogènes", c'est à dire ont des variances du même ordre de grandeur, il n'est pas nécessaire de réduire les variables. On dit alors que l'ACP est centrée.

Remarques :

R1 Les deux opérateurs $\mathbb{V}M = \frac{1}{n} X'X$ et $\mathbb{W}D = \frac{1}{n} XX'$ jouent un rôle symétrique. On retrouvera cette symétrie des opérateurs dans le cas où, comme D pour les lignes, M est une matrice diagonale des poids statistiques des points colonnes. L'Analyse Factorielle des Correspondances est l'exemple type de ce choix.

R2 Dans l'espace des variables, la D-orthogonalité est identique à l'orthogonalité usuelle

$$< x, y >_D = \frac{1}{n} y'x = 0 \Longleftrightarrow y'x = < x, y > = 0,$$

$$\Pi_X^D = \Pi_X^{\frac{1}{n} I_n} = \frac{1}{n} X (\frac{1}{n} X'X)^+ X' = X(X'X)^+ X' = \Pi_X^{I_n}.$$

R3 Puisque $M = I_p$ et $D = n^{-1} I_n$, le carré de l'erreur d'approximation entre $\widehat{X}_k$ et X s'écrit

$$\|X - \widehat{X}_k\|^2_{M \otimes D} = trace[(X - \widehat{X}_k)' D (X - \widehat{X}_k) M] = \frac{1}{n} \|X - \widehat{X}_k\|^2_F = \sum_{i=k+1}^{r} \lambda_i \,.$$

Critère global du choix de l'ordre k

Le théorème d'Eckart-Young conduit à un critère de la qualité globale de l'ACP.

Le pourcentage de l'Inertie Reconstituée à l'ordre k est défini par

$$\% IR(k) = \frac{\|\widehat{X}_k\|^2_{M \otimes D}}{\|X\|^2_{M \otimes D}} \times 100 = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{r} \lambda_i} \times 100 \,.$$

Une cassure dans la croissance du diagramme des valeurs $\{\%IR(k) \,|\, k = 1, \ldots, r\}$ permet de déterminer l'ordre k à retenir. Ce diagramme porte souvent le nom de diagramme de l'inertie cumulée.
La règle empirique précédente fournit, lorsque cela est possible, le nombre k de vecteurs de base (vecteurs propres) à retenir. Le choix de k devra permettre de retenir des vecteurs qui pris deux à deux, fourniront de "bons" plans de projection, c'est à dire des représentations à deux dimensions des points lignes ou colonnes facilement interprétables.

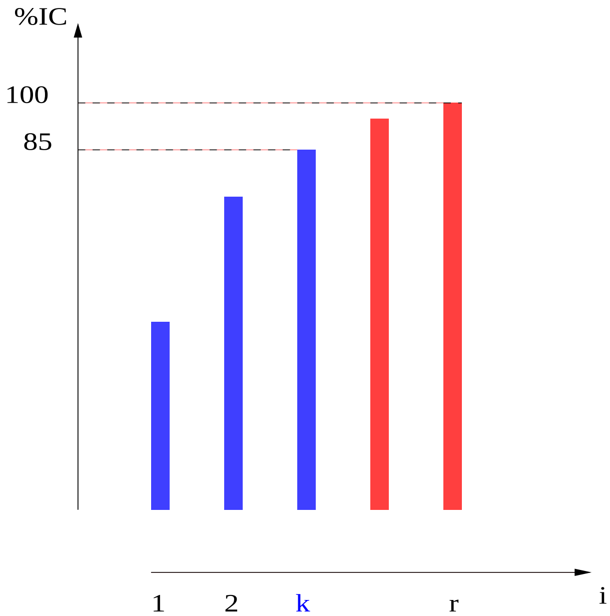


Figure 12 : diagramme des pourcentages d'inertie cumulée ou reconstituée.

Axes factoriels et principaux, composantes principales et facteurs principaux

- **Axes factoriels**

  On note $V_k = [V^1, \ldots, V^k]$ et $U_k = [U^1, \ldots, U^k]$ les matrices dont les colonnes sont les k vecteurs de base retenus : $\{V^1, \ldots, V^k\}$ vecteurs M-orthonormés dans l'espace des points lignes ($V_k' M V_k = I_k$) et $\{U^1, \ldots, U^k\}$ vecteurs D-orthonormés dans l'espace des points colonnes ($U_k' D U_k = I_k$).

  Un plan factoriel est un espace vectoriel de dimension 2 formé par l'un des couples d'axes factoriels (V^i, V^j) ou (U^i, U^j).

  On va définir des vecteurs colinéaires à ces vecteurs et de même sens mais de norme non plus égale à 1 mais égale à la valeur singulière correspondante. Ces vecteurs vont jouer un rôle capital dans le calcul des projections sur les plans factoriels des points lignes, respectivement des points colonnes.

- **Axes principaux**

Les k axes principaux $\{A^j \in (\mathbb{R}^p, M) \,|\, j = 1, \ldots, k\}$ définis par

$$A^j = \sqrt{\lambda_j} V^j$$

sont M-orthogonaux et de M-longueur $\sqrt{\lambda_j}$.

La matrice $A_k = [A^1, \ldots, A^k]$, $p \times k$, est telle que

$$A_k = V_k \Lambda_k^{1/2} \qquad \text{et} \qquad A'_k M A_k = \Lambda_k\,.$$

La deuxième formule de transition (**) devient

$$A_k = X' D U_k\,. \tag{$**$}'$$

La DVS du triplet permet d'autre part de décomposer la matrice des covariances $\mathbb{V}$ en

$$\mathbb{V} = X' D X = V \Lambda_r V' = A_r A'_r\,.$$

- **Composantes principales**

Les k composantes principales $\{C^j \in (\mathbb{R}^n, D) \,|\, j = 1, \ldots, k\}$ définies par

$$C^j = \sqrt{\lambda_j} U^j$$

sont D-orthogonales et de D-longueur $\sqrt{\lambda_j}$.

La matrice $C_k = [C^1, \ldots, C^k]$, $n \times k$, est telle que

$$C_k = U_k \Lambda_k^{1/2} \qquad \text{et} \qquad C'_k D C_k = \Lambda_k\,.$$

La première formule de transition (*) devient

$$C_k = X M V_k\,. \tag{$*$}'$$

La DVS du triplet permet aussi de décomposer la matrice des produits scalaires entre individus $\mathbb{W}$ en

$$\mathbb{W} = X M X' = U \Lambda_r U' = C_r C'_r\,.$$

- **Facteurs principaux**

On appelle matrice des facteurs principaux la matrice $p \times k$

$$F_k = [F^1, \ldots, F^k] = MV_k \,,$$

de telle sorte que (*)' s'écrire

$$C_k = XF_k \,. \qquad (*)'$$

Remarquer que dans le cas particulier $M = I_p$, $F_k = V_k$ et une composante principale s'écrit $C^j = XV^j = \sum_{i=1}^{p} X^i V_i^j$.

Interprétation des composantes principales par rapport aux variables-colonnes :

Une composante principale

$$C^j = XF^j = \sum_{i=1}^{p} X^i F_i^j$$

s'écrit comme une combinaison linéaire des p colonnes de X. Pour cette raison,

une composante principale peut être considérée comme l'expression dans la base canonique, d'une "variable latente", notée $\mathcal{C}^j$, c'est à dire une variable synthétique qui résume linéairement les variables naturelles $x^1, \ldots, x^p$.

Le scalaire F_i^j est le facteur principal qui mesure l'influence de la variable x^i sur la variable $\mathcal{C}^j$.

Propriétés statistiques des composantes principales-variables latentes

- **P1 :** $\mathcal{C}^j$ est D-centrée.
  Le fait que les variables soient D-centrées implique que (*)' a pour conséquence que les variables latentes$\{\mathcal{C}^j\}$ sont D-centrées : $\mathbf{1}_n' DC^j = \mathbf{1}_n' DXF^j = 0$.
- **P2 :** L'écart type de la variable latente $\mathcal{C}^j$ est $\sqrt{\lambda_j}$.
  Cela résulte de **P1**

$$\|C^j\|_D^2 = C^{j\prime} DC^j = \sum_{i=1}^{n} p_i (C_i^j)^2 = var(\mathcal{C}^j) = \lambda_j \,.$$

- **P3 :** Les variables latentes sont deux à deux non corrélées

$$cov(\mathcal{C}^i, \mathcal{C}^j) = C^{j\prime} DC^i = 0, \quad \text{si} \quad i \neq j.$$

7.2.2 Projections des points lignes

La matrice X étant D-centrée en colonnes, le nuage $\mathcal{N}$ des n points lignes $\{M_1, \ldots, M_n\}$ appartient à l'espace affine $Im\, X'$, sous-espace de dimension r de l'espace $(\mathbb{R}^p, M)$, et d'origine le point ligne moyen O calculé sur les données brutes. Pour visionner la position relative des points lignes par rapport à ce point de référence, on pourrait bien sûr, projeter les points du nuage $\mathcal{N}$ sur des axes ou des plans de la base canonique. Mais ces représentations ne sont pas forcément les plus simples à interpréter. On va voir pourquoi il est préférable d'utiliser des projections sur des espaces de dimension 1 ou 2 définis à partir des axes factoriels V^j.

Projection des points lignes sur l' axe factoriel V^j

Le vecteur $\overrightarrow{OM_l}$ dont l'expression dans la base canonique de $(\mathbb{R}^p, M)$ est le transposé de la ligne X_l, noté X_l', se projette orthogonalement au sens de M sur l'axe V^j selon

$$\Pi^M_{V^j} X_l' = V^j V^{j\prime} M X_l' = V^j (X_l M V^j)' = V^j C_l^j.$$

Le scalaire C_l^j est la mesure algébrique de la projection du vecteur $\overrightarrow{OM_l}$ sur le vecteur unitaire V^j.
Interprétation d'une composante principale C^j par rapport aux individus-lignes :

Le nuage $\mathcal{N}$ des points lignes se projette sur V^j selon les coordonnées du vecteur C^j, voir Figure 13. D'après la propiété **P2** d'une composante C^j, *l'inertie par rapport au point moyen O, du nuage des points projetés munis des poids p_i, est égale à λ_j.*
En effet, les n points projetés sur V^i étant donnés par les colonnes de la matrice $V^i C^{i\prime}$, l'inertie par rapport à O de ces n points s'exprime par la quantité

$$trace(V^j C^{j\prime} D C^j V^{j\prime} M) = trace(V^{j\prime} M V^j C^{j\prime} D C^j) = C^{j\prime} D C^j = \lambda_j\,.$$

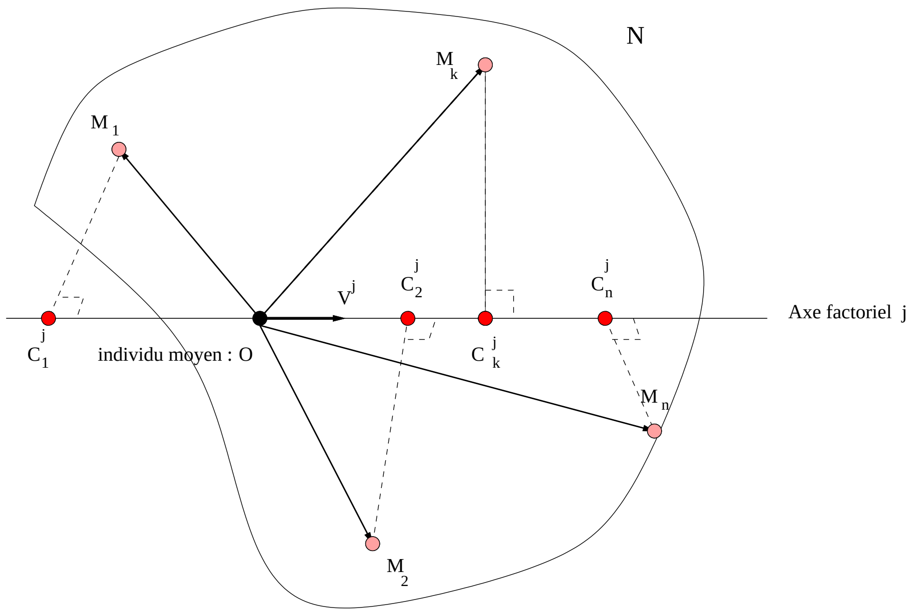


Figure 13 : Nuage $\mathcal{N}$ des points lignes de $(I\!\!R^p, M)$ et projections sur l'axe V^j.

N'oublions pas que les valeurs propres sont rangées par ordre décroissant $\lambda_1 \geq \ldots \geq \lambda_k$. On peut maintenant énoncer le principe de l'Analyse Factorielle.

<u>Principe de l'Analyse Factorielle</u> :
Si on admet que le meilleur "cliché" unidimensionnel est fourni par un axe sur lequel, en projection, le nuage des points lignes est d'inertie maximale, alors, l'axe factoriel V^1 est le meilleur axe; ensuite, V^2 est meilleur second, orthogonal au premier...

Preuve : Montrons que, parmi tous les vecteurs $V \in (I\!\!R^p, M)$ de longueur 1,

$$V^1 = arg \max_{\|V\|_M^2=1} trace((\Pi_V^M X')' M (\Pi_V^M X')D)\,.$$

La fonction objectif à maximiser qui est l'inertie des points lignes projetés sur V, s'écrit

$$\varphi(V) = trace((\Pi_V^M X')' M (\Pi_V^M X')D) = trace(XMVV'MX'D) = V'M\mathbb{V}MV\,.$$

Ecrivons les équations aux dérivées partielles pour la fonction de Lagrange

$$L(V, \lambda) = \varphi(V) + \lambda(1 - V'MV),$$

$$\nabla_V L(V, \lambda) = \nabla_V \varphi(V) - \lambda \nabla_V (V'MV) = 2M\mathbb{V}MV - 2\lambda MV = 0\,.$$

ce qui donne $\lambda = \varphi(V)$ et $\mathbb{V}MV = \lambda V$. D'où la conclusion que le maximum est donné par V^1 vecteur propre de $\mathbb{V}M$ associé à la plus grande valeur propre λ_1.
Montrons seulement que V^2 maximise $\varphi(V)$ sous les contraintes $\|V\|_M^2 = 1$ et $V^{1\prime}MV = 0$.

La restriction de la fonction objectif à l'espace vectoriel $V^{1\perp} = \{V \in I\!R^p \,|\, V^{1\prime}MV = 0\}$, peut s'écrire $\varphi_{V^{1\perp}}(V) = V'M(\mathbb{V} - \lambda_1 V^1 V^{1\prime})MV$. Sur $V^{1\perp}$, la méthode de Lagrange associée à la contrainte $\|V\|_M^2 = 1$, conduit à $\lambda = \varphi_{V^{1\perp}}(V)$ et à $(\mathbb{V} - \lambda_1 V^1 V^{1\prime})MV = \lambda V$. Le maximum est donc réalisé par le couple (V^2, λ_2), λ_2 plus grande valeur propre de $(\mathbb{V} - \lambda_1 V^1 V^{1\prime})M$, étant la deuxième valeur propre de $\mathbb{V}M$... $\square$

Remarque : Si tous les points lignes ont le même poids, le principe de l'analyse factorielle est un principe géométrique d'allongement maximum des points projetés sur chacun des axes.

Projection des points lignes sur un plan factoriel (V^i, V^j)

Notons $[V^i, V^j]$ la matrice $p\times 2$ extraite des colonnes de V_k, les propriétés des axes factoriels impliquent que $[V^i, V^j]'M[V^i, V^j] = I_2$. La matrice de la projection M-orthogonale sur le plan factoriel $Im\,[V^i, V^j]$, notée $\Pi^M_{[V^i,V^j]}$, se décompose en la somme des projecteurs sur chacun des axes

$$\Pi^M_{[V^i,V^j]} = \Pi^M_{V^i} + \Pi^M_{V^j}\,.$$

Le vecteur $\overrightarrow{OM_l}$ se projette sur le plan factoriel, appelé (i,j) pour simplifier, selon le vecteur $\overrightarrow{Om_l}$ dont les coordonnées dans le repère (V^i, V^j) sont données par (C^i_l, C^j_l)

$$\Pi^M_{[V^i,V^j]} X'_l = V^i C^i_l + V^j C^j_l\,.$$

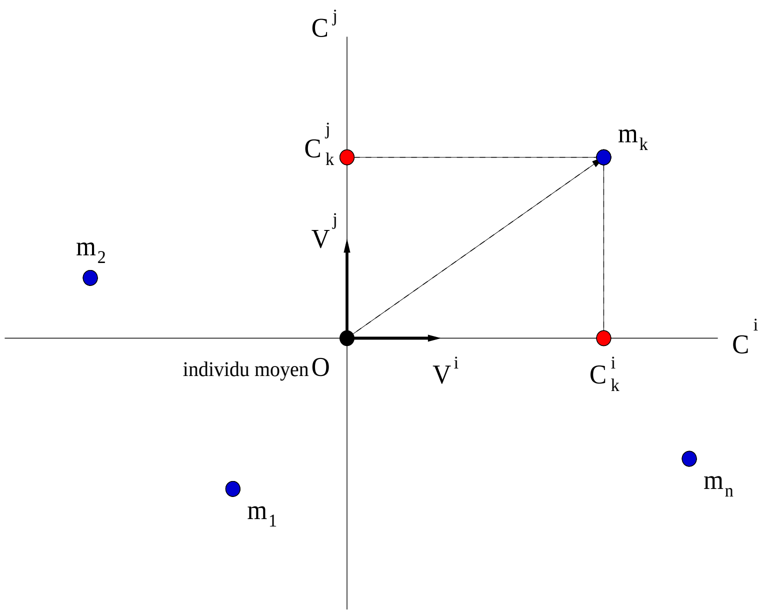


Figure 14 : Projection du nuage des individus sur le plan factoriel (V^i, V^j).

La matrice $p \times n$ des projections du nuage des n points sur le plan (i, j) est donnée par

$$\Pi^M_{[V^i,V^j]} X' = V^i C^{i\prime} + V^j C^{j\prime}.$$

La M-orthonormalité des $\{V^i\}$ permet de montrer que l'inertie par rapport à O des points lignes projetés sur le plan (i, j)

$$trace[(\Pi^M_{[V^i,V^j]} X')D(\Pi^M_{[V^i,V^j]} X')'M] = C^{i\prime}DC^i + C^{j\prime}DC^j = \lambda_i + \lambda_j \ ,$$

est la somme des inerties des points projetés sur chacun des axes.
Pour faire une bonne photo de points lignes, il faut trouver un plan sur lequel en projection, le nuage des points est d'inertie maximale : alors le plan factoriel (1,2) est le meilleur plan possible; on peut ensuite tirer les photos $(1, 3)$, $(2, 3)$...

Aides à l'interprétation des points lignes

Contributions absolues

On désire quantifier la contribution de chaque individu-ligne l à l'expression de l'inertie du nuage projeté sur l'axe factoriel V^j. On a vu que l'inertie autour du point ligne moyen est caractérisée par la variance de C^j, c'est à dire $C^{j\prime}DC^j = \lambda_j$; alors

$$1 = \sum_{l=1}^{n} \frac{p_l}{\lambda_j}(C^j_l)^2.$$

La contribution absolue de l'individu l à la dispersion du nuage projeté sur l'axe j est

$$CTA^j_l = \frac{p_l}{\lambda_j}(C^j_l)^2 \ .$$

Si tous les individus ont le même poids statistique $p_l = 1/n$ alors la dispersion ne dépend que du carré de la coordonnée de la projection de l'individu sur l'axe. Dans ce cas, les individus à forte CTA sont ceux les plus éloignés de l'individu moyen. Ce n'est plus forcément le cas si les poids statistiques ne sont pas identiques.

Contributions relatives

Tous les points d'une photo uni ou bi-dimensionnelle ne sont pas visibles avec la même précision. En conséquence, on ne pourra interpréter la position relative de deux points projetés m_{l1} et m_{l2} que si elle reflète bien la position des points M_{l1} et M_{l2} de l'espace $(I\!R^p, M)$. De façon plus pratique, on mesurera la proximité relative entre m_l et M_l par le carré du cosinus du M-angle entre les vecteurs $\overrightarrow{Om_l}$ et $\overrightarrow{OM_l}$

$$\cos^2\theta_l = \frac{\parallel \overrightarrow{Om_l} \parallel_M^2}{\parallel \overrightarrow{OM_l} \parallel_M^2} = \frac{\parallel \overrightarrow{Om_l} \parallel_M^2}{X_l M X_l'} = \frac{\parallel \overrightarrow{Om_l} \parallel_M^2}{\mathbb{W}_l^l} = \frac{\parallel \overrightarrow{Om_l} \parallel_M^2}{\sum_{j=1}^r (C_l^j)^2}.$$

On dira que l'individu l est bien représenté par m_l si cette expression, appelée aussi *contribution relative de l'axe ou du plan factoriel à la représentation de l'individu* l, est voisine de 1, mal représentée si elle est voisine de 0.
Notons m_l^i et $m_l^{i,j}$ les projections de M_l sur l'axe V^i respectivement sur le plan (V^i, V^j). L'orthonormalité de V^i et V^j implique $\overrightarrow{Om_l^{i,j}} = \overrightarrow{Om_l^i} + \overrightarrow{Om_l^j}$ dont l'expression dans la base canonique est $V^i C_l^i + V^j C_l^j$. Il vient

$$\|V^i C_l^i + V^j C_l^j\|_M^2 = \|V^i\|_M^2 (C_l^i)^2 + \|V^j\|_M^2 (C_l^j)^2 = (C_l^i)^2 + (C_l^j)^2 \,.$$

On obbtient ainsi les contributions relatives de l'axe i et du plan (i, j) à l'individu l

$$CTR_l^i = \cos^2\theta_l^i = \frac{(C_l^i)^2}{\sum_{j=1}^r (C_l^j)^2} \text{ et } CTR_l^{i,j} = \cos^2\theta_l^{i,j} = \frac{(C_l^i)^2 + (C_l^j)^2}{\sum_{j=1}^r (C_l^j)^2} = \cos^2\theta_l^i + \cos^2\theta_l^j \,.$$

De façon évidente, la somme des contributions relatives pour un même individu l est égale à 1, $\sum_{i=1}^r \cos^2\theta_l^i = 1$.

7.2.3 Projections des vecteurs colonnes

Dans l'ACP usuelle, les colonnes sont plutôt considérées comme l'expression de vecteurs-variables, à la différence des lignes considérées comme des points. La proximité de deux variables est basée sur l'examen du coeficient de corrélation qui d'un point de vue géométrique, est le cosinus du D-angle entre ces deux vecteurs. Il sera toujours possible, dans l'ACP généralisée de considérer les points extrémités de ces vecteurs et appelés points colonnes.

Projections sur les axes et plans factoriels

De façon duale à celle des lignes, les matrices $n \times n$ de projection D-orthogonale sur un axe factoriel $Im\, U^i = Im\, C^i$ et sur un plan factoriel $Im\, [U^i, U^j] = Im\, [C^i, C^j]$ sont données par

$$\Pi_{U^i}^D = U^i U^{i\prime} D \qquad \text{et} \qquad \Pi_{[U^i, U^j]}^D = \Pi_{U^i}^D + \Pi_{U^j}^D \,.$$

La mesure algébrique de la projection d'un vecteur-variable x^l sur un vecteur de base U^i est égale à A^i_l. En effet, la formule de transition (**)' donne

$$\Pi^D_{U^i} X^l = U^i A^i_l \qquad \text{et} \qquad \Pi^D_{[U^i,U^j]} X^l = U^i A^i_l + U^j A^j_l \,.$$

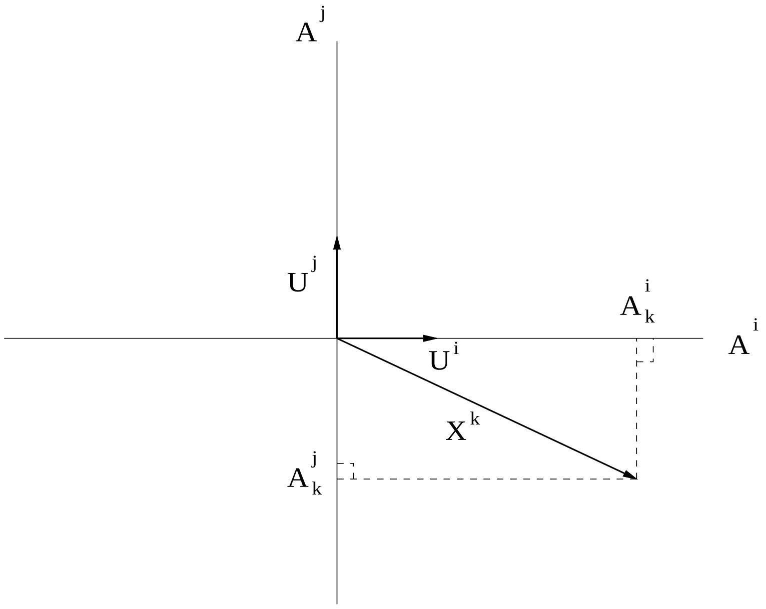


Figure 15 : Projection d'une variable x^k sur le plan factoriel (U^i, U^j).

L'axe principal A^i est le vecteur des coordonnées des projections des p points colonnes sur l'axe factoriel U^i.

Coefficient de corrélation entre la variable latente $\mathcal{C}^j$ et une variable x^i

Une composante C^j s'interprète comme le vecteur des coordonnées des projections du nuage des indvidus sur l'axe factoriel V^j mais aussi comme l'expression d'une variable latente $\mathcal{C}^j$ dans la base canonique de $(I\!R^n, D)$. Afin de pouvoir donner un nom à cette variable synthétique il est nécessaire d'examiner quelles sont les variables qui lui sont statistiquement voisines. L'examen des coefficients de corrélation entre $\mathcal{C}^j$ et les variables x^i, $i = 1, \ldots, p$,

$$r(\mathcal{C}^j, x^i) = \frac{cov(\mathcal{C}^j, x^i)}{\sigma(\mathcal{C}^j)\sigma(x^i)} = \frac{X^{i\prime} D C^j}{\sqrt{\lambda_j}\, \|X^i\|_D} = \frac{X^{i\prime} D X M V^j}{\sqrt{\lambda_j}\, \|X^i\|_D} = \frac{\lambda_j V^j_i}{\sqrt{\lambda_j}\, \|X^i\|_D} = \frac{A^j_i}{\|X^i\|_D} \,,$$

permet de sélectionner celles qui sont fortement corrélées avec la variable latente.

Cas de l'ACP usuelle réduite

Dans le cas particulier, $\sigma(x^i) = \|X^i\|_D = 1$, et

$$r(\mathcal{C}^j, x^i) = A^j_i \,.$$

Les points colonnes sont situés sur la sphère unité de $(\mathbb{R}^n, D)$. Ils se projettent sur un plan factoriel (U^i, U^j) à l'intérieur du cercle trigonométrique appelé "cercle des corrélations". Dans ce cas, la projection d'un point colonne sur un axe factoriel est la corrélation de la variable avec la variable latente correspondante.

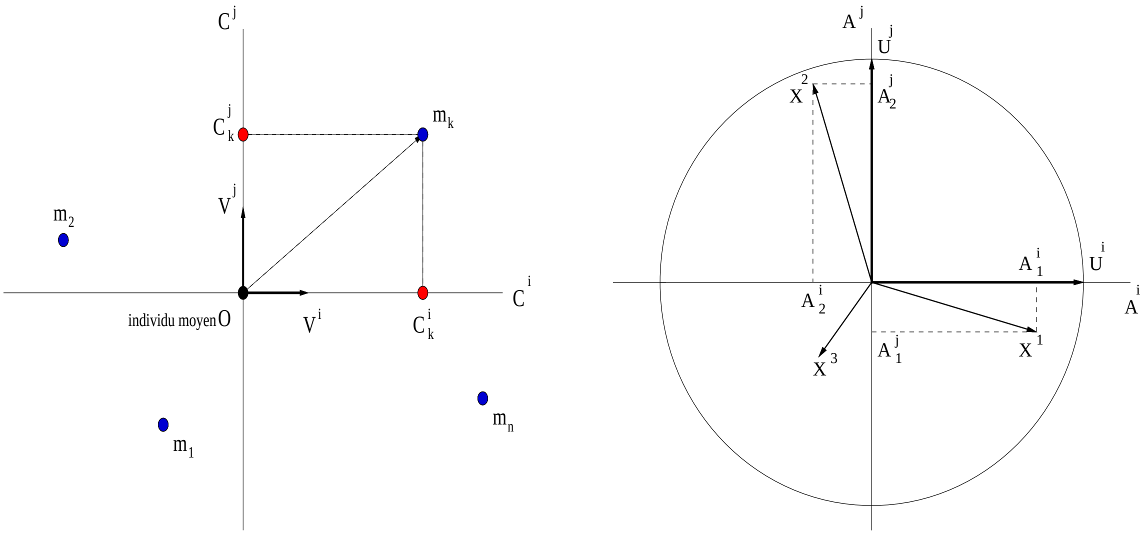


Figure 16 : Visualisation duale des plans factoriels (i, j) pour les individus et les variables dans l'ACP usuelle normée. Cercle des corrélations.

La Figure 16, présente les deux représentations duales individus-variables sur les plans factoriels (i, j). Un plan factoriel est de bonne qualité globale si $\lambda_i + \lambda_j$ est grand. Les points lignes bien représentés dans ce plan sont assez nombreux; on peut interpréter leur position relative au point moyen, en expliquant les variables latentes grâce aux variables naturelles bien représentées (voir paragraphe suivant). Dans ce cas particulier de l'ACP usuelle, une variable est bien représentée dans un plan factoriel si son point colonne associé est proche du cercle des corrélations. Sur la Figure 5, X^1 et X^2 sont bien représentées, alors que X^3 ne l'est pas. L'individu numéro n a une valeur au dessus de la moyenne pour la variable x^1, inférieure à la moyenne pour x^2...

Aides à l'interprétation des variables

Contributions absolues : cas où M est diagonale

Les ACP les plus courantes concernent le cas où $M = diag(m_1, \ldots, m_p)$: $M = I_p$ (ACP usuelle) ou bien M est la matrice des poids statistiques des colonnes (Analyses Factorielles des Correspondances simples et multiples). On a vu que la projection d'une variable x^i sur l'axe U^j à pour mesure algébrique A_i^j, voir Figure 4. Dans ce cas particulier, le nuage des points colonnes pondérés $\{(x^i, m_i)\}$ projetés sur l'axe j, a pour inertie $A^{j\prime} M A^j = \lambda_j$; ce qui donne

$$1 = \sum_{i=1}^{p} \frac{m_i}{\lambda_j} (A_i^j)^2 \,.$$

On appelle contribution absolue de la variable x^i à l'inertie sur l'axe j, l'expression

$$CTA_i^j = \frac{m_i}{\lambda_j} (A_i^j)^2 \,.$$

Comme pour les CTA des points lignes, la CTA d'un point colonne prend en compte non seulement l'éloignement de l'origine $(A_i^j)^2$ mais aussi le poids m_i d'un point colonne.

Contributions relatives

On dira qu'un point colonne x^k est bien représenté sur l'axe U^i ou sur le plan (U^i, U^j) si le carré du cosinus du D-angle entre le vecteur projeté et le vecteur colonne initial, est voisin de 1.

On appelle contribution relative de l'axe ou du plan factoriel à la variable X^k

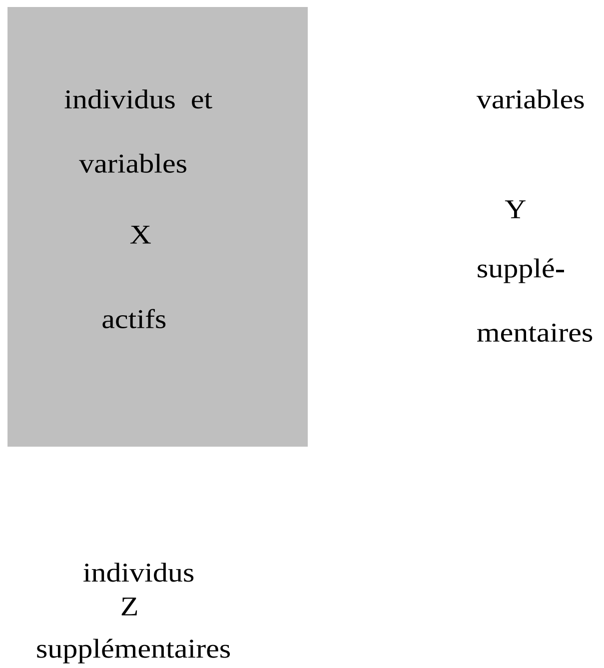
$$CTR_k^i = \cos^2 \theta_k^i = \frac{(A_k^i)^2}{\sum_{j=1}^r (A_k^j)^2} \text{ ou } CTR_k^{i,j} = \cos^2 \theta_k^{i,j} = \frac{(A_k^i)^2 + (A_k^j)^2}{\sum_{j=1}^r (A_k^j)^2} = \cos^2 \theta_k^i + \cos^2 \theta_k^j \,.$$

En effet, la contribution relative de l'axe i à la variable k est le carré du coefficient de corrélation entre x^k et $\mathcal{C}^i$. *La contribution relative du plan factoriel (i, j) à la variable k est le coefficient de détermination R^2 entre x^k et le couple de variables latentes* $(\mathcal{C}^i, \mathcal{C}^j)$. Remarquons que le dénominateur, $\sum_{j=1}^r (A_k^j)^2 = \mathbb{V}_k^k = \|X^k\|_D^2$, est égal à 1 dans le cas d'une ACP usuelle sur variables centrées réduites.

7.2.4 Eléments supplémentaires

Les individus et variables du tableau X sur lesquels a été effectué une ACP sont dits "actifs". Ils ont été visualisés et interprétés grâce aux bases orthonormées qui ont servi à les montrer en photo.

On peu représenter sur celles ci des individus ou variables n'ayant pas pris part à la détermination des axes factoriels. Ces individus ou variables dits "supplémentaires" peuvent apporter des compléments dans l'analyse.



Variables supplémentaires

Soit Y le tabeau des variables supplémentaires mesurées sur les mêmes n individus et transformées comme les variables actives l'ont été : Y est centré mais peut être éventuellement réduit... Grâce à la deuxième formule de transition (**)', la coordonnée de la variable Y^k sur l'axe factoriel actif U^j, s'écrit

$$A_k^j = \frac{Y^{k\prime} DC^j}{\sqrt{\lambda_j}} = \sum_{i=1}^{n} \frac{p_i}{\sqrt{\lambda_j}} Y_i^k C_i^j \, .$$

On peut représenter la variable suplémentaire k sur un plan factoriel et par exemple, visualiser par l'ACP usuelle, les corrélations de Y^k avec les variables actives...

Individus supplémentaires

Le taleau Z des individus supplémentaires est supposé D-centré par rapport à l'individu moyen actif $\mathbf{1}'DT$. Ce tableau Z a été réduit si les colonnes de X l'ont été, en utilisant les écarts types des variables actives. La première formule de transition (*)' permet de calculer la coordonnée sur l'axe actif V^j, de la projection M-orthogonale de l'individu l

$$C_l^j = \frac{Z_l M A^j}{\sqrt{\lambda_j}} \, .$$

On peut visualiser sur le plan factoriel (i, j) une population d'individus supplémentaires et la comparer à celle des individus actifs.

7.3 Exercices

Exercice 1

Soit X une matrice réelle $n \times p$ dont le rang est r. Comparer la décomposition en valeurs singulières des deux triplets (X, M, D) et $(Z = D^{1/2} X M^{1/2}, I_p, I_n)$.

Exercice 2 : Approximation d'une matrice par une matrice de rang donné

Théorème : Soit Z une matrice réelle $n \times p$ de rang r dont la décomposition en valeurs singulières s'écrit $Z = U_r \Lambda_r^{1/2} V_r'$ où $\Lambda_r^{1/2} = diag(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r})$ avec $\lambda_1 \geq \ldots \geq \lambda_r > 0$. On note $\mathcal{E}_k$ l'ensemble des matrices $n \times p$ de rang fixé k $(k \leq r)$ et $\{E_{ij}(r,k)\}_{i,j}$ la base canonique pour les matrices $r \times k$. Alors

$$\min_{Z_k \in \mathcal{E}_k} \|Z - Z_k\|_F^2 = \|Z - \hat{Z}_k\|_F^2 = \sum_{i=k+1}^{r} \lambda_i \ ,$$

où $\hat{Z}_k = U_k \Lambda_k^{1/2} V_k'$ avec $U_k = \sum_{i=1}^k U_r E_{ii}(r,k)$, $V_k = \sum_{i=1}^k V_r E_{ii}(r,k)$ et $\Lambda_k^{1/2} = diag(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_k})$. $\square$

1. Exprimer U_k (respectivement V_k) en fonction des colonnes de U_r (respectivement V_r). Quelle est la valeur de $\hat{Z}_k$ lorsque $k = r$?

2. Sachant que la norme de Frobénius $\|.\|_F$, $(\|X\|_F^2 = trace(X'X))$ est invariante par transformation orthogonale, montrer qu'il existe deux matrice B et C respectivement $n \times k$ et $k \times p$ de rang maximum telles que

$$\|Z - Z_k\|_F^2 = \|\Lambda - BC\|_F^2 = \phi(B, C)$$

où Λ est la matrice $n \times p$ de la décomposition en valeurs singulières complète contenant les valeurs singulières de Z.

3. L'objectif étant maintenant de minimiser $\phi(B, C)$, montrer dans un premier temps que pour B fixé,

$$\hat{C}(B) \stackrel{\text{def}}{=} arg \min_C \phi(B, C) = (B'B)^{-1} B' \Lambda$$

et que minimiser $\phi(B, \hat{C}(B))$revient à maximiser $\varphi(B) = trace(B(B'B)^{-1} B' \Lambda \Lambda')$.

4. Utiliser les propriétés de $\Pi_B = B(B'B)^{-1} B'$ pour montrer que

$$\varphi(B) = \|\Pi_B \Lambda\|_F^2 \ .$$

Ecrire Λ dans la base canonique $\{E_{ij}(n,p)\}_{i,j}$ des matrices $n \times p$ et montrer que

$$\varphi(B) = \sum_{i=1}^{r} \lambda_i a_i \ ,$$

où $a_i = \|\Pi_B e_i(n)\|_F^2$.

5. Démontrer que $0 \leq a_i \leq 1$, pour $i = 1, \ldots, n$ et que $\sum_{i=1}^{n} a_i = rang(B) = k$. En déduire que

$$\max_B \varphi(B) = \sum_{i=1}^{k} \lambda_i$$

et que la maximum est réalisé pour $\hat{B} = \sum_{i=1}^{k} E_{ii}(n,k)$. Calculer $\hat{C}(\hat{B})$.

6. Conclure en montrant que

$$\hat{Z}_k = arg \min_{Z_k \in \mathcal{E}_k} \|Z - Z_k\|_F^2 \ .$$

L'ACP usuelle normée ou non

Exercice 3

Une étude gastronomique a conduit à apprécier le service, la qualité et le prix de quatre restaurants. Pour cela, un expert a noté ces restaurants avec des notes allant de -3 à 3. Les résultats sont les suivants

Restaurant	Service	Qualité	Prix
R1	-2	3	-1
R2	-1	1	0
R3	2	-1	-1
R4	1	-3	2

La matrice des covariances est

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

et celle des corrélations (aux erreurs d'arrondi près)

$$R = \begin{bmatrix} 1 & -0.85 & 0.26 \\ -0.85 & 1 & -0.73 \\ 0.26 & -0.73 & 1 \end{bmatrix} .$$

Pour l'étude , on effectuera une ACP centrée avec des poids équirépartis.

1. Etude des valeurs propres
   a. Vérifier simplement que V admet une valeur propre $\lambda_3 = 0$.
   b. On donne $\lambda_1 = 30.5/4$. En déduire λ_2.
   c. Calculer les pourcentages d'inertie. Quelle est la dimension à retenir?
2. a. On donne, aux erreurs d'arrondi près, $v_1 = \begin{bmatrix} 0.5 \\ -0.8 \\ 0.3 \end{bmatrix}$ et $v_2 = \begin{bmatrix} 0.65 \\ 0.11 \\ -0.75 \end{bmatrix}$.
   Calculer les composantes principales.
   b. Représenter les individus dans le plan principal (1,2).
3. a. Déterminer les corrélations entre les variables et les composantes.
   b. Représenter les variables sur le cercle des corrélations dans le plan factoriel (1,2).
   c. Interpréter les résultats.

Exercice 4

Soit la matrice $X = [X^1, X^2, X^3]$ dont les variables ont pour matrice des corrélations

$$R = \begin{bmatrix} 1 & \rho & -\rho \\ \rho & 1 & \rho \\ -\rho & \rho & 1 \end{bmatrix},$$

avec$-1 \leq \rho \leq 1$. On désire effectuer une ACP centrée réduite de X.

1. Vérifier que R admet pour vecteur propre $\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.
2. Déterminer les autres valeurs propres et vecteurs propres de R.
3. Quelles sont les valeurs possibles de ρ?
4. Justifier le fait que l'ACP n'a d'intérêt que si $-1 < \rho < 0$.
5. Calculer dans ce cas les pourcentages de variance expliquée.

6. Comment s'interprète par rapport à X^1, X^2, et X^3 l'unique composante à retenir ici?

Exercice 5

Soit la matrice

$$T = \sqrt{10}\begin{bmatrix} 2 & 2 & 3 \\ 3 & 1 & 2 \\ 1 & 0 & 3 \\ 2 & 1 & 4 \\ 2 & 1 & 3 \end{bmatrix}$$

des mesures de 5 individus munis de poids statististiques égaux, sur 3 variables notées T^1, T^2 et T^3. On désire effectuer une Analyse en Composantes Principales (ACP) sur variables centrées-réduites.

1. Calculer l'individu moyen, le vecteur $(\sigma_1, \sigma_2, \sigma_3)$ des écarts types et la X des variables centrées-réduites.

2. Calculer la matrice R des corrélations.

3. Calculer les éléments propres de R.

4. Les deux premiers vecteurs propres de R associés aux valeurs propres $\lambda_1 = 1+\sqrt{2}/2$ et $\lambda_2 = 1$, sont :

$$v_1 = \frac{1}{2}\begin{bmatrix} \sqrt{2} \\ 1 \\ -1 \end{bmatrix} \qquad \text{et} \qquad v_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Déterminer les composantes principales c_1 et c_2 dont on vérifiera les propriétés statistiques.

5. Représenter les individus et les variables dans les plans factoriels (1,2). Quelle est l'interprétation des variables c_1 et c_2?

6. Représenter dans le plan (1,2) l'individu supléméntaire $(\sqrt{10}, 2\sqrt{10}, 2\sqrt{10})$.

Exercice 6 : ACP usuelle normée et ACP du triplet $(X, diag(\sigma_1^{-2}, \dots, \sigma_p^{-2}), n^{-1}I_n)$

Soit T la matrice des n mesures sur p variables quantitatives et X la matrice des variables centrées au sens des poids statistiques $\frac{1}{n}$. On note $\sigma_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i^j)^2}$, l'écart type de la variable X^j et $Y = Xdiag(\sigma_1^{-1}, \dots, \sigma_p^{-1})$ la matrice des variables centrées réduites.

1. Etude de l'ACP normée d'ordre k du triplet $(Y, M = I_p, D = n^{-1}I_n)$:

a. Ecrire l'expression des opérateurs $\mathbb{V}_Y M = Y'DYM$ et $\mathbb{W}_Y D = YMY'D$

b. Ecrire la définition des axes principaux et des composantes principales pour l'ACP normée.

2. Etude de l'ACP d'ordre k du triplet $(X, M = diag(\sigma_1^{-2}, \ldots, \sigma_p^{-2}), D = n^{-1}I_n)$:

a. Ecrire l'expression des opérateurs $\mathbb{V}_X M = X'DXM$ et $\mathbb{W}_X D = XMX'D$

b. Ecrire la définition des axes principaux et des composantes principales pour cette l'ACP.

3. Comparer les résultats des questions **1** et **2**. Conclusions quant-à la représentation des individus et des variables pour ces deux ACP?

Exercice 7 : ACP d'un ensemble de notes

Les données jointes ont été restreintes pour les besoins de ce problème à 10 étudiants. Pour chaque individu on dispose d'un ensemble de 3 notes , Mathématiques ("Maths"), Physique ("Phys") et Technologie ("Techn") ainsi que d'une information supplémentaire sur la provenance de l'étudiant fournie par la variable booléenne indicatrice de l'origine ("Orig").

Les résultats de l'Analyse en Composantes Principales des 3 premières variables actives centrées sont donnés en annexe.

a. Analyser les résultats. Dire pourquoi le premier axe factoriel a ses coordonnées positives. Dans ce cas, le premier axe est appelé "axe de taille" le second "axe de forme".

b. Pour examiner si la provenance des individus est liée aux résultats obtenus, représenter la variable "Orig" en variable supplémentaire sur le plan factoriel (1,2).

c. Situer dans le plan factoriel (1,2) un individu supplémentaire dont les notes seraient : Math = 12, Phys = 12, Techn = 12.

```
-----------------------------------------------------------------
ANNEXE :
------
 Notes et Origine
   Math Phys Techn Orig
 1   17   12    13    1
```

```
 2    9   10     8    0
 3   12   12    13    1
 4   15   12    14    1
 5    9   10    11    0
 6   13   15    12    1
 7   11    9    10    0
 8   14   15    16    1
 9    9   11    11    0
10   13   14    13    1
----------------------------------------------------------------
moyenne et variances des variables    Matrice des covariances

    Math Phys Techn                           Math Phys Techn
moy 12.2   12  12.1                     Math  6.76  2.9  3.98
var 6.76    4  4.49                     Phys  2.90  4.0  3.10
                                        Techn 3.98  3.1  4.49
------------------------------------------------------------------
Inertie totale = 15.25
------------------------------------------------------------------
   val.pro. % inert. % cumul.
1  11.9718    78.50    78.50
2   2.3033    15.10    93.60
3   0.9750     6.39   100.00
---------------------------------------------------------------
aides a l'interpretation pour les u.s. :
Composantes Principales
      c1     c2     c3
 1  3.795 -3.022  0.560
 2 -5.417 -0.263  1.278
 3  0.365  0.367 -0.763
 4  2.981 -1.409 -0.758
 5 -3.743  0.508 -1.089
 6  1.893  1.499  1.955
 7 -3.395 -1.793 -0.325
 8  4.812  1.849 -0.936
 9 -3.276  1.197 -0.534
10  1.985  1.067  0.611

Contributions absolues des 10 u.s. pour les 3 premieres composantes
   CTA1 CTA2 CTA3
 1 1203 3965  322
 2 2451   30 1675
 3   11   58  597
 4  743  863  589
 5 1170  112 1215
 6  299  975 3919
 7  963 1396  108
 8 1934 1484  899
 9  896  622  292
10  329  494  383
```

Contributions relative des 3 premieres composantes pour les 10 u.s.

	COS1	COS2	COS3
1	6039	3829	132
2	9452	22	526
3	1569	1583	6847
4	7763	1735	502
5	9066	167	767
6	3713	2328	3959
7	7764	2165	71
8	8435	1245	319
9	8620	1151	229
10	7226	2088	685

----

aides a l'interpretation pour les variables :

Axes Principaux

	a1	a2	a3
Math	2.374	-1.029	0.261
Phys	1.615	1.046	0.548
Techn	1.932	0.390	-0.779

Contributions absolues des 3 variables pour les 3 premiers axes

	CTA1	CTA2	CTA3
Math	4706	4594	700
Phys	2178	4746	3077
Techn	3117	660	6223

Contributions relative des 3 axes pour les 3 premieres variables

	COS1	COS2	COS3
Math	8334	1565	101
Phys	6517	2733	750
Techn	8310	339	1351

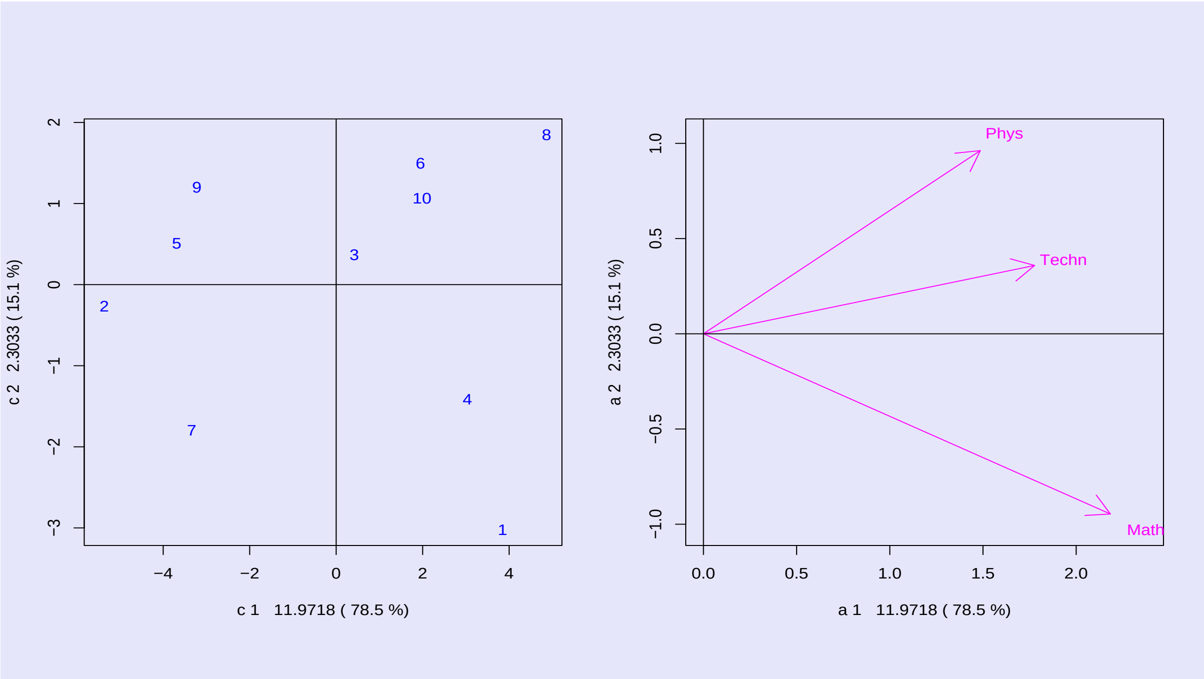


Figure 17 : plans factoriels $(1,2)$ *pour les individus et les variables de l'ACP centrée d'un ensemble de notes.*

L'Analyse Factorielle Discriminante (AFD)

Exercice 8 : L'AFD ou ACP du triplet $(G, \mathbb{V}^{-1}, \Delta)$

Soit X la matrice, supposée de rang p, de l'échantillon d'apprentissage de n mesures sur p variables quantitatives centrées au sens des poids statistiques stockés dans la matrice diagonale $D = diag(p_1, \ldots, p_n)$. On dispose en outre d'une partition des n individus en q groupes : $\{1, \ldots, n\} = I_1 \bigcup \ldots \bigcup I_q$ où $I_i = \{k \in \{1, \ldots, n\} \mid k \in \text{groupe } i\}$. Soit $\Delta = diag(P_1, \ldots, P_q)$ la matrice diagonale des poids statistiques des groupes : $P_i = \sum_{j \in I_i} p_j$ est le poids du groupe i. On note y la variable qualitative à q modalités caractérisant l'appartenance des n individus aux q groupes : $y_k = i$ signifie que l'individu k appartient au groupe i. Soit Y la matrice $n \times q$, du codage disjonctif complet de y : $Y = [Y_k^i = \delta_{y_k\, i}]$.

L'objectif de l'Analyse Factorielle Discriminante (AFD), ne pas confondre avec Analyse Factorielle des Données, est de **déterminer des axes factoriels de $I\!R^p$ sur lesquels, en projection, les individus moyens des q groupes ont une inertie maximale**. La métrique utilisée sur $I\!R^p$ est $\mathbb{V}^{-1} = (X'DX)^{-1}$.

1. Construction et propriétés de la matrice G, $q \times p$, des individus moyens:

a. Dire pourquoi $\Delta = Y'DY$ est bien une métrique sur $\mathbb{R}^q$ et $\mathbb{V}^{-1}$ est une métrique sur $\mathbb{R}^p$? L'individu moyen du groupe i, la $i$ème ligne de G, est défini par $G_i = \dfrac{1}{P_i} \sum_{j \in I_i} p_j X_j$. Vérifier que $G = \Delta^{-1} Y'DX$.

b. Montrer que G est Δ-centrée. Soit $B = G'\Delta G$ la matrice $p \times p$ des covariances entre les groupes ("Between"). Montrer que $B = X'D \prod_Y^D X$, où $\prod_Y^D$ est la matrice de projection D-orthogonale sur $Im\,Y$.

2. Etude de l'ACP d'ordre k du triplet $(G, \mathbb{V}^{-1}, \Delta)$:

a. Ecrire l'expression de l'opérateur des covariances de ce triplet en fonction de B et de $\mathbb{V}$. Soient $\{V^1, \ldots, V^k\}$ les k premiers vecteurs propres de cette matrice (les axes factoriels appelés ici axes discriminants). Quelle est l'expression du facteur principal F^i, appelé facteur discriminant, en fonction de V^i? Quelle est la matrice qui admet les $\{F^i\}$ comme vecteurs propres?

b. On représente sur l'axe factoriel V^i, les n individus de l'échantillon d'apprentissage comme individus supplémentaires. Donner l'expression du vecteur d^i de $\mathbb{R}^n$, appelé ici variable discriminante, donnant les coordonnées de ces points sur l'axe V^i.

7.4 Formulaires

7.4.1 ACP usuelle d'ordre k du triplet $(X, M = I_p, D = n^{-1}I_n)$

T ; $n \times p$ $X = (I_n - n^{-1}\mathbf{1}\mathbf{1}')T$	n individus (actifs), p variables quantitatives (actives) tableau des variables centres
$\mathbb{V} = n^{-1}X'X$	matrice des covariances (des corrlations si variables rduites)
$\mathbb{V}\ V^i = \lambda_i V^i$	
$\lambda_1 \geq \ldots \geq \lambda_k$ $\{V^1, \ldots, V^k\}$ $\{A^1, \ldots, A^k\}$	$\sum_1^p \lambda_i = trace(\mathbb{V})$ = Variance totale = Inertie Facteurs principaux, $V^i \in I\!\!R^p$, $\|V^i\|_2 = 1$, $V^{i\prime}V^j = 0$ Axes principaux, $A^i = \sqrt{\lambda_i}V^i$, $\|A^i\|_2 = \sqrt{\lambda_i}$, $A^{i\prime}A^j = 0$
$\{C^1, \ldots, C^k\}$	Composantes principales, $C^i \in I\!\!R^n$ $C^i = XV^i = X^1V_1^i + \ldots + X^pV_p^i$ $moy(C^i) = 0$; $var(C^i) = \lambda_i$; $cov(C^i, C^j) = 0$, $i \neq j$ C^i vecteur propre de $n^{-1}XX'$ associ λ_i ; $\mathbb{W} = XX'$ $r(C^j, X^i) = A_i^j/\sigma(X^i)$
$\widehat{X}_k = \sum_{i=1}^k \frac{1}{\sqrt{\lambda_i}} C^i A^{i\prime}$	Approximation de rang k de X ; $\|X - \widehat{X}_k\|_F^2 = n \sum_{i=k+1}^p \lambda_i$
Contributions absolues	de l'individu k l'axe i : $CTA_k^i = (C_k^i)^2/(n\lambda_i)$, $\sum_{k=1}^n CTA_k^i = 1$ de la variable k l'axe i : $CTA_k^i = (A_k^i)^2/\lambda_i$, $\sum_{k=1}^p CTA_k^i = 1$
Contributions relatives	de l'axe i l'individu k : $CTR_k^i = (C_k^i)^2/\mathbb{W}_k^k$, $\sum_{i=1}^p CTR_k^i = 1$ de l'axe i la variable k : $CTR_k^i = (A_k^i)^2/\mathbb{V}_k^k$, $\sum_{i=1}^p CTR_k^i = 1$
Formules de transition	$C^i = XA^i/\sqrt{\lambda_i}$; $A^i = X'C^i/(n\sqrt{\lambda_i})$

7.4.2 DVS du triplet (X, M, D)

<table>
<tr><td colspan="3">DVS du triplet $(X_{n\times p},\, M_{p\times p},\, D_{n\times n})$; rang $(X) = r$</td></tr>
<tr><td colspan="3">$X = U\Lambda_r^{1/2}V' = C\Lambda_r^{-1/2}A'$; $\Lambda_r = diag(\lambda_1, \dots, \lambda_r)$; $\lambda_1 \geq \dots \geq \lambda_r > 0$</td></tr>
<tr><td rowspan="2">Opérateurs</td><td>$\boxed{\mathbb{V}M = X'DXM}$
$X'DXMV = V\Lambda_r$
$X'DXMA = A\Lambda_r$</td><td>$\boxed{\mathbb{W}D = XMX'D}$
$XMX'DU = U\Lambda_r$
$XMX'DC = C\Lambda_r$</td></tr>
<tr><td colspan="2">Inertie totale $\|X\|^2_{M\otimes D} = \text{trace}(\mathbb{V}M) = \text{trace}(\mathbb{W}D) = \sum_i^r \lambda_i$</td></tr>
<tr><td>Espaces</td><td>Axes Principaux</td><td>Composantes Principales</td></tr>
<tr><td>Individus
Bases $M \perp$ de $Im\,(X') \subset (I\!R^p, M)$
$V = [V^1, \dots, V^r]$, $V'MV = I_r$
$A = [A^1, \dots, A^r]$, $A'MA = \Lambda_r$
$\|V^i\|_M = 1$, $\|A^i\|_M = \sqrt{\lambda_i}$</td><td>$A^i = \sqrt{\lambda_i}\,V^i$
$A = V\Lambda_r^{1/2}$
(**)' $A = X'DU = X'DC\Lambda_r^{-1/2}$
Facteurs Principaux $F = MV$
$\mathbb{V} = AA'$</td><td></td></tr>
<tr><td>Variables
Bases $D \perp$ de $Im\,(X) \subset (I\!R^n, D)$
$U = [U^1, \dots, U^r]$, $U'DU = I_r$
$C = [C^1, \dots, C^r]$, $C'DC = \Lambda_r$
$\|U^i\|_D = 1$, $\|C^i\|_D = \sqrt{\lambda_i}$</td><td></td><td>$C^i = \sqrt{\lambda_i}\,U^i$
$C = U\Lambda_r^{1/2}$
(*)' $C = XMV = XMA\Lambda_r^{-1/2} = XF$

$\mathbb{W} = CC'$</td></tr>
</table>

Bibliography

[1] X. Bry. *Analyses Factorielles Simples*, Economica, Paris, 1995.

[2] X. Bry. *Analyses Factorielles Multiples*, Economica, Paris, 1996.

[3] F. Cailliez et J. P. Pages, *Introduction à l'Analyse des Données*, SMASH, Paris, 1976.

[4] P. G. Ciarlet. *Introduction à l'Analyse Numérique Matricielle et à l'Optimisation*, Masson, Paris, 1990.

[5] European Courses in Advanced Statistics. *Methods for Multidimensional Data Analysis*, Dipartimento di Matematica e Statistica, Università di Napoli, 1987.

[6] M. J. Greenacre. *Theory and Applications of Correspondence Analysis*, Academic Press, London, 1984.

[7] P. Lascaux et R. Theodor. *Analyse Numérique Matricielle Appliquée à l'Art de l'Ingénieur*, tome 1, Masson, Paris, 1986.

[8] J. R. Magnus et H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley & Sons, Chichester, 1988

[9] C. R. Rao et S. K. Mitra. *Generalized Inverse of Matrices and Its Applications*, Wiley, New-York, 1971.

[10] G. Saporta. *Probabilités, Analyse de Données et Statistique*, Technip, Paris, 1990.