

Éléments de Calcul Matriciel
et
d'Analyse Factorielle de Données

Jean-François Durand

jfd@helios.ensam.inra.fr

Université Montpellier II

Licence MASS

Maîtrise MASS

Maîtrise d'Ingénierie Mathématique

DEA de Biostatistique

Novembre 2002

Table des matières

1	Matrices, Définitions et Propriétés	9
1.1	Notations et premières définitions	9
1.2	Matrice associée à une application linéaire	11
1.3	Quelques matrices particulières	11
1.3.1	Matrice adjointe	11
1.3.2	Matrices hermitiennes	12
1.4	Image, noyau, rang d'une matrice	12
1.4.1	Image d'une matrice	13
1.4.2	Noyau d'une matrice	13
1.4.3	Rang d'une matrice	14
1.5	Déterminant d'une matrice carrée	14
1.6	Inverse d'une matrice carrée	15
1.7	Changements de base	16
1.7.1	Effet sur les coordonnées d'un vecteur	16
1.7.2	Effet sur les éléments d'une matrice	17
1.8	Valeurs propres, vecteurs propres	18
1.9	Trace d'une matrice carrée	20
1.10	Formes linéaires, formes quadratiques	20
1.11	Matrices orthogonales et unitaires	22
1.11.1	Les matrices de permutation	22
1.11.2	Les matrices de rotation	23
1.11.3	Construction d'une base orthonormée par le procédé de Gram-Schmidt	24
1.12	Opérateur vec et produit de Kronecker	24
1.12.1	L'opérateur vec	24
1.12.2	Produit de Kronecker	25
1.12.3	Matrice de commutation	27

1.13	Exercices	28
2	Décomposition de Matrices	33
2.1	Les projecteurs	33
2.1.1	Sous espaces supplémentaires et projecteurs	34
2.1.2	Exemple fondamental	34
2.1.3	D'autres matrices orthogonales : les matrices de réflexion	36
2.2	Matrices carrées diagonalisables	37
2.3	Factorisation QR d'une matrice rectangulaire	38
2.4	Décomposition unitaire des matrices carrées	39
2.4.1	Le théorème de Schur	39
2.4.2	Matrices normales	40
2.5	Décomposition en valeurs singulières	42
2.5.1	Deux versions de la DVS	43
2.5.2	Décomposition polaire	46
2.6	Factorisation de Cholesky d'une matrice symétrique définie positive	47
2.7	Exercices	48
3	Normes de Matrices	51
3.1	Normes de vecteurs	51
3.1.1	Normes de Hölder	52
3.1.2	Généralisation de la norme Euclidienne, la M -norme	54
3.2	Normes de matrices	56
3.2.1	Normes subordonnées à des normes vectorielles	56
3.2.2	Normes Euclidiennes par vectorisation	58
3.2.3	Normes matricielles sous multiplicatives	60
3.2.4	Normes unitairement invariantes	61
3.3	Suites de matrices	62
3.4	Conditionnement d'une matrice	63
3.5	Exercices	64
4	Inverses Généralisés, Projecteurs M-Orthogonaux	67
4.1	Inverses Généralisés	67
4.1.1	Définition et propriétés	67
4.1.2	Inverse de Moore-Penrose	70
4.2	Projecteurs M -orthogonaux	71

4.2.1	Projecteur M -orthogonal sur $Im A$	72
4.2.2	Un problème aux moindres carrés	74
4.3	Exercices	76
5	Dérivation Matricielle	77
5.1	Introduction	77
5.2	Dérivation matricielle	77
5.2.1	Matrices Jacobiennes	77
5.2.2	Hessien de fonctions numériques	81
5.3	Extremums de fonctions numériques	82
5.3.1	Problèmes d'extremums libres	84
5.3.2	Problèmes d'extremums liés	84
5.4	Exercices	86
6	Le paysage mathématique et statistique de l'Analyse Factorielle de Données : la géométrie Euclidienne	89
6.1	Le triplet (T, M, D) des données	90
6.2	Statistique et géométrie sur (\mathbb{R}^n, D) , espace des variables	91
6.2.1	Le simplexe des poids statistiques et la droite des constantes	91
6.2.2	Moyenne et centrage vus comme une projection	92
6.2.3	Variance et écart-type	93
6.2.4	Proximité entre deux variables, covariance et corrélation linéaire	94
6.2.5	Définitions et notations pour la statistique multivariée	96
6.3	Exercices	98
7	Généralisation de la Décomposition en Valeurs Singulières. Analyse en Composantes Principales du triplet (X, M, D)	101
7.1	Décomposition en Valeurs Singulières du triplet	102
7.1.1	Lemme	102
7.1.2	La DVS du triplet (X, M, D)	103
7.1.3	Relation avec la DVS usuelle	104
7.1.4	Projecteurs orthogonaux associés à la DVS	105
7.1.5	Théorème d'approximation d'Eckart-Young	105
7.2	Analyse en Composantes Principales d'ordre k du triplet (X, M, D)	106
7.2.1	Définitions	107
7.2.2	Principe fondamental de l'Analyse Factorielle	108

7.2.3	L'ACP usuelle d'ordre k	110
7.3	Représentations factorielles	111
7.3.1	Définitions	111
7.3.2	Projections des points lignes	114
7.3.3	Projections des vecteurs colonnes	118
7.3.4	Éléments supplémentaires	121
7.4	Exercices	123
7.5	Formulaires	132
7.5.1	ACP usuelle d'ordre k du triplet $(X, M = I_p, D = n^{-1}I_n)$	132
7.5.2	DVS du triplet (X, M, D)	132

**8 Traitement d'enquêtes, Analyse Factorielle des Correspondances
Simples et Multiples 135**

8.1	Variables d'une enquête, codage	135
8.1.1	Variables qualitatives	135
8.1.2	Indicatrice des modalités	136
8.2	Table de contingence, liaison entre deux variables qualitatives	137
8.2.1	Définitions et notations	138
8.2.2	Indépendance de deux variables qualitatives \mathcal{L} et \mathcal{C}	140
8.2.3	Profils lignes et colonnes, distributions conditionnelles	141
8.3	Analyse Factorielle des Correspondances	143
8.3.1	Définition et propriétés	143
8.3.2	Éléments propres des opérateurs en dualité	145
8.3.3	Pratique de l'AFC	146
8.3.4	Cas d'une variable ordinale, rapport de corrélation et "optimal scoring"	149
8.4	Analyse Factorielle des Correspondances Multiples	152
8.4.1	Définitions et propriétés	152
8.4.2	Pratique de l'AFCM	155
8.5	Exercices	156
8.5.1	Analyse Factorielle des Correspondances	156
8.5.2	Analyse Factorielle des Correspondances Multiples	161

9 La régression Partial Least-Squares linéaire 165

9.1	Motivations pour les régressions factorielles	166
-----	---------------------------------------------------------	-----

9.2	La régression sur composantes principales	167
9.3	Le contexte et le modèle PLS	168
9.4	L'algorithme PLS	169
9.4.1	Le centrage des variables	170
9.4.2	Construction de la composante t^i	171
9.4.3	Les régressions partielles	173
9.5	Première écriture des modèles PLS en les composantes	175
9.6	Orthogonalité des composantes principales et des poids	176
9.7	Écriture définitive des modèles PLS en les composantes	178
9.8	Les composantes PLS, compromis linéaires des variables explicatives initiales	180
9.8.1	Expression du vecteur des poids	181
9.8.2	Orthogonalité des vecteurs des poids au sens de \mathbb{V}	182
9.8.3	Propriétés des vecteurs des poids	183
9.9	Le modèle de régression PLS en les variables explicatives	184
9.9.1	Le modèle sur variables centrées, éventuellement réduites	185
9.9.2	Le modèle en les variables initiales	185
9.10	Régression PLS et Analyse en Composantes Principales usuelle	186
9.11	Représentations factorielles	187
9.11.1	Représentation des prédicteurs et des réponses	188
9.11.2	Représentation des individus	190
9.12	Métriques pour les individus et optiques photographiques associées	193
9.12.1	Métriques générales pour les individus	193
9.12.2	Régression PLS discriminante	194
9.13	Choix du nombre de composantes	195
9.13.1	Critères basés sur l'ajustement	195
9.13.2	Critères basés sur la prédiction	197
9.14	Pratique de la régression PLS	199
9.14.1	PLS univarié, les données de Cornell	199
9.14.2	"Calibration" PLS en spectroscopie proche infrarouge	207
9.15	Exercices	211

Chapitre 1

Matrices, Définitions et Propriétés

Ce chapitre n'est pas un cours de calcul matriciel mais a pour objectif d'une part de rappeler de façon succincte 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 \mathbb{K} -espace vectoriel (\mathbb{K} est \mathbb{R} ou \mathbb{C}) et d'applications linéaires sont connues. De façon commode, pour éviter d'alourdir les notations, lorsqu'une base de l'espace vectoriel \mathbb{K}^n est spécifiée, **on notera de façon identique le vecteur et la matrice colonne donnant les coordonnées de ce vecteur dans la base**. 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 \mathbb{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 ème 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 $\lambda \in \mathbb{K}$, est défini par

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

Ces deux opérations confèrent à l'ensemble $\mathcal{M}_{mn}(\mathbb{K})$ des matrices de m lignes et n colonnes sur \mathbb{K} , une structure d'espace vectoriel de dimension mn qui sera aussi noté $\mathbb{K}^{m \times n}$. On note $\{E_{ij}(m, n)\}_{i,j}$ la base canonique de $\mathbb{K}^{m \times n}$, définie par

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

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

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

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 $\mathcal{M}_{m1}(\mathbb{K}) = \mathbb{K}^{m \times 1}$ identifié à \mathbb{K}^m , est notée

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

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

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

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

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

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

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

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

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

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

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

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

Soient E et F deux \mathbb{K} -espaces vectoriels de dimensions finies ($E = \mathbb{K}^n$ et $F = \mathbb{K}^m$), $\{e_1, \dots, e_n\}$ une base de E et $\{f_1, \dots, f_m\}$ 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

$$A = \begin{bmatrix} g(e_1) & \dots & g(e_n) \\ A^1 & \dots & A^n \end{bmatrix}$$

La j ème colonne A^j contient les coordonnées de $g(e_j)$ exprimées dans la base $\{f_1, \dots, f_m\}$. 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 $y_i = \sum_{j=1}^n a_{ij} x_j$.

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

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 $\mathcal{C}^{m \times n}$ est omis.

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

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

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

$$\mathbf{P1} : (A^*)^* = A,$$

$$\mathbf{P2} : (A + B)^* = A^* + B^*,$$

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

$$\mathbf{P4} : (AB)^* = B^* A^*.$$

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 \mathcal{C}^n formée par les vecteurs propres de A .

Notons

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n), \quad \lambda_i \in \mathbb{R}$$

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

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

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

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

$$Im A = \mathcal{E}\{A^1, \dots, A^n\} = \{y \in \mathbb{K}^m \mid \exists x \in \mathbb{K}^n, Ax = y\}.$$

C'est l'ensemble des combinaisons linéaires, $y = A^1x_1 + \dots + A^nx_n$, 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 $A \in \mathbb{R}^{m \times n}$ la matrice des coefficients du système et soit $y \in \mathbb{R}^m$ 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 \mathcal{C} , se décompose en deux systèmes sur \mathbb{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 $A \in \mathbb{K}^{m \times n}$ est le s.e.v. de \mathbb{K}^n dont l'image est le zéro de \mathbb{K}^m

$$Ker A = \{x \in \mathbb{K}^n \mid Ax = 0_m\}.$$

On a les propriétés suivantes,

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

P2 : g_A est injective si et seulement si $\text{Ker } A = \{0_n\}$.

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

1.4.3 Rang d'une matrice

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

$$\text{rang}(A) = \dim \text{Im } A.$$

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

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

P1 : g_A est surjective (injective) si et seulement si $\text{rang}(A) = m$ ($\text{rang}(A) = n$).

P2 : $\text{rang}(A) = \text{rang}(A^*) = \text{rang}(A^*A) = \text{rang}(AA^*)$.

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

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

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

1.5 Déterminant d'une matrice carrée

σ est une permutation sur $\{1, \dots, n\}$ si et seulement si σ est une bijection de $\{1, \dots, n\}$ 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 $\det(A)$

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)},$$

où S_n est l'ensemble des $n!$ permutations de $\{1, 2, \dots, n\}$ 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 $A = [a] \in \mathcal{C}^{1 \times 1}$, alors $\det(A) = a$.

Si $A \in \mathcal{C}^{n \times n}$, développement du déterminant selon la colonne j

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

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

P0 : Le déterminant d'une matrice carrée $A = [A^1 \dots A^n]$ est une forme multilinéaire alternée des colonnes de A , c'est à dire, $(A^1, \dots, A^n) \longrightarrow D(A^1, \dots, A^n) = \det(A)$ 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 : $\det(AB) = \det(A)\det(B)$.

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

P3 : Si A est diagonale ou triangulaire alors $\det(A)$ 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, $A^{-1} = C' / \det(A)$.

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

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

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

(i) : $\det(A) \neq 0$,

(ii) : $\text{rang}(A) = n$,

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

P7 : Caractérisation du rang d'une matrice $n \times p$: Soit $A \in \mathbb{K}^{n \times p}$, Il existe 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 propriété **P2** $\text{rang}(A) = \text{rang}(A^*)$ du rang d'une matrice.

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

1.7 Changements de base

Soit E un espace vectoriel muni d'une base $\mathcal{E} = \{e_1, \dots, e_n\}$ que l'on qualifie d'ancienne base. On choisit dans E une autre base $\mathcal{F} = \{f_1, \dots, f_n\}$ appelée nouvelle base.

On appelle matrice de passage de la base \mathcal{E} à la base \mathcal{F} , la matrice P carrée d'ordre n , dont les colonnes expriment les coordonnées des vecteurs de \mathcal{F} dans la base \mathcal{E} . Cette matrice est régulière. La matrice de passage de \mathcal{F} à \mathcal{E} 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{E} et \mathcal{F} sont stockées dans les matrices colonnes X et Y . Alors

$$X = PY \quad \text{et} \quad Y = P^{-1}X.$$

Exemple : Une base utile en statistique

En statistique, une série de n mesures est pratiquée sur une variable x . On note

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

les coordonnées de x dans la base canonique $\mathcal{E} = \{e_i\}_i$. Le statisticien préfère exprimer les mesures dans la base $\mathcal{F} = \{f_i = \sqrt{n}e_i\}_i$. La matrice de passage de la base \mathcal{E} à la base \mathcal{F} est $P = \sqrt{n}I_n$ de telle sorte que, dans la nouvelle base, l'échantillon de x devient $Y = P^{-1}X = 1/\sqrt{n}X$.

Soient maintenant, deux variables u et v dont les n -échantillons s'expriment dans \mathcal{E} par les deux matrices colonnes U et V . Ces deux échantillons s'écrivent $(1/\sqrt{n})U$ et $(1/\sqrt{n})V$ dans la nouvelle base \mathcal{F} . C'est cette base que retiendra le statisticien car le produit scalaire $\langle u, v \rangle$ s'exprime dans \mathcal{F} par

$$\sum_{i=1}^n (1/\sqrt{n}u_i)(1/\sqrt{n}v_i) = \frac{1}{n} \sum_{i=1}^n u_i v_i$$

et prend un sens statistique pour calculer des moyennes, des covariances et des variances.

Il est parfois utile d'individualiser le poids statistique de chaque mesure, p_i au lieu de $1/n$. La nouvelle base est alors $\mathcal{F} = \{f_i = (1/\sqrt{p_i})e_i\}_i$, avec $0 < p_i < 1$ et $\sum_i p_i = 1$. La matrice de passage s'écrit $P = \text{diag}(1/\sqrt{p_1}, \dots, 1/\sqrt{p_n})$ et l'expression de Y devient $Y = P^{-1}X = \text{diag}(\sqrt{p_1}, \dots, \sqrt{p_n})X$. Le produit scalaire entre u et v s'écrit dans la nouvelle base

$$\sum_{i=1}^n p_i u_i v_i.$$

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 $\mathcal{E} = \{e_1, \dots, e_n\}$ pour E et $\mathcal{F} = \{f_1, \dots, f_m\}$ 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 maintenant la base \mathcal{E} en une base $\mathcal{U} = \{u_1, \dots, u_n\}$ et la base \mathcal{F} en une base $\mathcal{V} = \{v_1, \dots, v_m\}$, la même application linéaire g va s'exprimer dans ces nouvelles bases selon une autre matrice B . Qu'elle relation lie A et B qui appartiennent toutes deux à $\mathbb{K}^{m \times n}$?

Soit P la matrice de passage de \mathcal{E} à \mathcal{U} et soit Q la matrice de passage de \mathcal{F} à \mathcal{V} , alors

$$B = Q^{-1}AP.$$

On dit que les matrices A et B sont équivalentes.

Dans le cas particulier d'un endomorphisme $g \in \mathcal{L}(E, E)$, l'influence du changement de la base \mathcal{E} en la base \mathcal{F} se traduit par $B = P^{-1}AP$. On dit alors que les matrices carrées

A et B sont semblables. Si A et B sont semblables, $\det(A) = \det(B)$. S'il existe une base \mathcal{F} telle que B 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 \mathcal{C}^m est appelé vecteur propre d'une matrice $A \in \mathcal{C}^{m \times n}$ s'il existe un scalaire $\lambda \in \mathcal{C}$ 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

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

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 \mathcal{C}^m 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,

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

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, $\{\lambda_1, \dots, \lambda_n\}$, de l'équation caractéristique

$$\mathcal{P}_A(\lambda) = \det(A - \lambda I_n) = 0 \quad (2)$$

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)

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

où la trace de la matrice A est la somme de ses éléments diagonaux $\text{trace}(A) = a_{11} + \dots + a_{nn}$, 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, $\bar{\lambda}_i$.

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 $\{\lambda_1, \dots, \lambda_n\}$ le spectre de A , alors,

$$\det(A) = \prod_{i=1}^n \lambda_i = \lambda_1 \dots \lambda_n.$$

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 : Cayley-Hamilton : $\mathcal{P}_A(A) = 0_{n \times n}$

P7 : Perron : 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

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

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 $A = [a_{ij}] \in \mathcal{C}^{n \times n}$ 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 \mathcal{C} (\mathbb{R}) est un produit scalaire si et seulement si à tout couple (x, y) elle fait correspondre le nombre complexe (réel) $\langle x, y \rangle$ tel que

$$\text{D1} : \langle x, y \rangle = \overline{\langle y, x \rangle}.$$

$$\text{D2} : \langle \alpha_1 x_1 + \alpha_2 x_2, y \rangle = \alpha_1 \langle x_1, y \rangle + \alpha_2 \langle x_2, y \rangle \quad \forall \alpha_1, \alpha_2 \in \mathcal{C} \text{ (\mathbb{R})}.$$

$$\text{D3} : \langle x, x \rangle \geq 0 \quad (\text{réel d'après D1}).$$

$$\text{D4} : \langle x, x \rangle = 0 \iff x = 0_E.$$

Comme première propriété

$$\langle x, \alpha y \rangle = \bar{\alpha} \langle x, y \rangle.$$

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

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

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

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

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 $x^* = [\bar{x}_1 \dots, \bar{x}_n]$ le vecteur adjoint, alors le produit scalaire usuel se calcule par

$$\begin{aligned} \langle x, y \rangle &= y^* x = \sum_{i=1}^n x_i \bar{y}_i \quad \text{sur } \mathcal{C}^n, \\ \langle x, y \rangle &= y' x = \sum_{i=1}^n x_i y_i \quad \text{sur } \mathbb{R}^n. \end{aligned}$$

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

$$x \longrightarrow \varphi(x) = \langle x, a \rangle = a^* x.$$

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

$$x \longrightarrow \phi(x) = \langle Ax, x \rangle = x^* Ax.$$

En effet, d'après D1, A hermitienne entraîne que $\langle Ax, x \rangle$ est réel

$$\langle Ax, x \rangle = x^* Ax = x^* A^* x = \langle x, Ax \rangle = \overline{\langle Ax, x \rangle}.$$

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

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

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

On dit que A hermitienne, est

définie positive si $x^* Ax > 0 \quad \forall x \neq 0,$

semi définie positive si $x^* Ax \geq 0 \quad \forall x.$

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 ,

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

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

Application :

$$A = (P^{-1})'P^{-1}$$

où P est la matrice de passage de la base de x dans une autre base (voir paragraphe 1.7.1), $P^{-1}x$ est le vecteur des coordonnées dans la nouvelle base. A est bien symétrique définie positive.

1.11 Matrices orthogonales et unitaires

Soit une matrice U de $\mathcal{O}^{m \times n}$ ($m \geq n$), l'espace \mathcal{O}^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 on 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 $\{1, 2, 3, 4\}$,

$$\frac{i}{\sigma(i)} \left| \begin{array}{cccc} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{array} \right.$$

La matrice de passage de la base canonique $\{e_i\}$ de \mathbb{R}^n à la base $\{f_i = e_{\sigma(i)}\}$, $Q = [e_{\sigma(1)}, \dots, e_{\sigma(n)}]$, 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 $\frac{i}{\sigma^{-1}(i)} \left| \begin{array}{cccc} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{array} \right.$.

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

$$Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, \quad (\det(Q) = 1).$$

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

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

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 \mathbb{K}^n $\{f_1, \dots, f_k\}$, on peut construire un ensemble de vecteurs $\{q_1, \dots, q_k\}$ orthonormés qui engendrent le même espace

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

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

$$\begin{aligned} p_1 &= f_1 \\ p_2 &= f_2 - \frac{\langle p_1, f_2 \rangle}{\|p_1\|_2^2} p_1 \\ &\dots \\ p_i &= f_i - \frac{\langle p_1, f_i \rangle}{\|p_1\|_2^2} p_1 - \frac{\langle p_2, f_i \rangle}{\|p_2\|_2^2} p_2 \dots - \frac{\langle p_{i-1}, f_i \rangle}{\|p_{i-1}\|_2^2} p_{i-1} \quad i \leq k. \end{aligned}$$

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

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

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

1.12 Opérateur vec et produit de Kronecker

1.12.1 L'opérateur vec

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

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

Si $A = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}$, l'empilement en colonne des lignes de A se fait par $vec(A) \in \mathbb{K}^{np}$. 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 $n \times p$

$$\text{trace}(Y^*X) = \sum_{j=1}^p \sum_{i=1}^n \bar{y}_{ij} x_{ij} = vec^*(Y)vec(X) = \langle vec(X), vec(Y) \rangle.$$

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 $nm \times pq$.

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

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 $\text{trace}(A \otimes B) = \text{trace}(A)\text{trace}(B)$.

P9 : Si x et y sont des matrices colonnes alors

$$xy' = \begin{bmatrix} x_1y_1 & \cdots & x_1y_p \\ \vdots & \ddots & \vdots \\ x_ny_1 & \cdots & x_ny_p \end{bmatrix} = x \otimes y' = y' \otimes x = [y_1x | \cdots | y_px].$$

$$\text{vec}(xy') = \begin{bmatrix} y_1x \\ \vdots \\ y_px \end{bmatrix} = y \otimes x.$$

P10 : Soient A, B , deux matrices carrées respectivement $n \times n$ et $p \times p$ avec

$$\text{spect}(A) = \{\lambda_1, \dots, \lambda_n\} \text{ et } \text{spect}(B) = \{\mu_1, \dots, \mu_p\} \text{ alors}$$

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

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

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 $p \times p$, $\det(A \otimes B) = (\det(A))^p(\det(B))^n$.

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

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,

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

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

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

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

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

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

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

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

Exercice 1 : Dans l'espace \mathbb{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)

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

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 :

$$c_1 = e_2 + 2e_3 + 2e_4, \quad c_2 = e_1 + e_2 + 2e_3 + 3e_4, \quad c_3 = 2e_1 + 2e_2 + 2e_3 + 3e_4 \text{ et } c_4 = 2e_1 + 3e_2 + 3e_3 + 3e_4.$$

- 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 \mathbb{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 :

$$u_1 = 4 + 3X + 2X^2 + X^3, \quad u_2 = 6 + 2X + 2X^2 + 2X^3 \text{ et } u_3 = -8 - 11X - 6X^2 + X^3.$$

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

f de \mathbb{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) Dédire 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 \mathbb{R}^4 avec $A_{f,C}$ diagonale et préciser quelle est la matrice $A_{f,C}$.

Exercice 4 : Soit la matrice

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

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

Exercice 5 : 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 .

Exercice 6 : 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) $Ax = 0 \iff A'Ax = 0$ et $\text{Ker } A = \text{Ker } A'MA$.
- b) $AB = 0 \iff A'AB = 0$.
- c) $A'AB = A'AC \iff AB = AC$.

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

- a) $(Ax = 0, \forall x \in \mathbb{R}^n) \iff A = 0$.
- b) $(x'Ay = 0, \forall x \in \mathbb{R}^m, \forall y \in \mathbb{R}^n) \iff A = 0$.
- c) $(x'Cx = 0, \forall x \in \mathbb{R}^n) \iff C = 0$.
- d) $(x'Bx = 0, \forall x \in \mathbb{R}^n) \iff B' = -B$.

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

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

Exercice 9 : Montrer que $\text{trace}(AB) = \text{trace}(BA)$.

Exercice 10 : Soit A une matrice $n \times n$ et $\text{spect}(A) = \{\lambda_i\}_i$. Soit B une matrice $p \times p$ et $\text{spect}(B) = \{\mu_j\}_j$. 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 : $\det(A \otimes B) = [\det(A)]^p [\det(B)]^n$

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

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

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

Chapitre 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 quatre.

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 dans la section quatre.

2.1.1 Sous espaces supplémentaires et projecteurs

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

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

On dit que F et G sont supplémentaires si $F \cap G = \{0_E\}$ 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

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

sont linéaires (endomorphismes de E) et vérifient :

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

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

P3 $p + q = Id_E$

P4 $Im p = F = Ker q \quad \text{et} \quad Im q = G = Ker p.$

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

$$\langle u, v \rangle = v^* u = 1.$$

Remarquons que, puisque $\langle u, v \rangle = \|u\|_2 \|v\|_2 \cos(u, v)$, 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 $x \in \text{Im } u$, 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 $\text{Im } u$, le noyau de P est le sous espace vectoriel de dimension $n - 1$ orthogonal à v .

$$\mathbb{K}^n = \text{Im } u \oplus \{\text{Im } v\}^\perp.$$

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

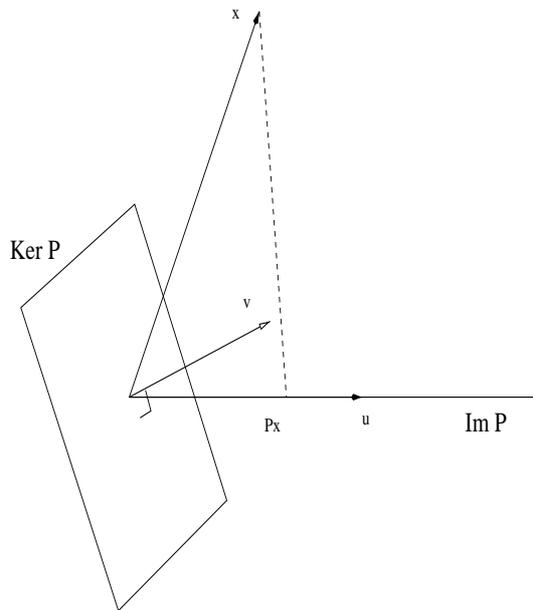


Figure 1 : Projection sur u parallèlement à $\{\text{Im } v\}^\perp$.

Si on choisit $v = u$ et $\|u\|_2 = 1$, dans ce cas, le projecteur orthogonal s'écrit

$$P = uu^*.$$

D'une façon plus générale, soit F donné ainsi qu'une base $\{u_1, \dots, u_r\}$ orthonormée de F . Soit $U = [u_1, \dots, u_r]$, 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 $F = \text{Im } U$. 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

$$Q = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}, \quad (\det(Q) = -1).$$

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

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

La dimension n :

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

$$Q(v) = I_n - 2vv'/v'v, \quad (\det(Q(v)) = -1).$$

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

Reflexion $Q(v) = I - 2P$

Projection orthogonale $P = vv'/v'v$

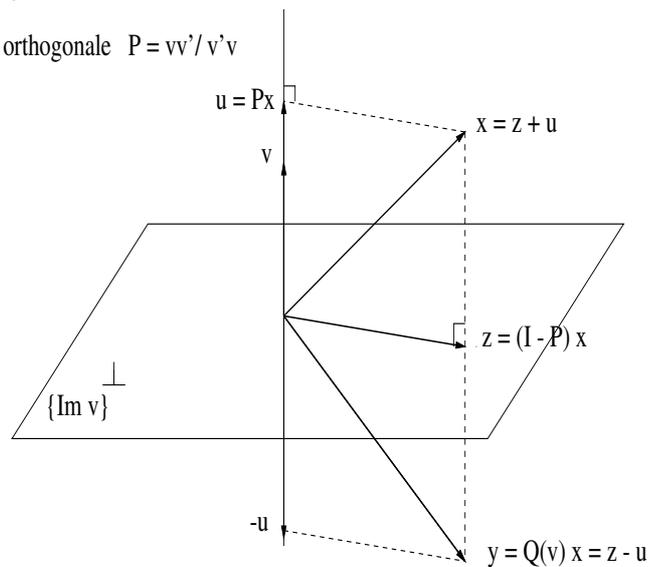


Figure 2 : Symétrie par rapport à l'hyperplan vectoriel $\{\text{Im } v\}^\perp$.

Si $y = Q(v)x = Q'(v)x$, y est obtenu par symétrie du vecteur x par rapport à l'hyperplan vectoriel $\{Im v\}^\perp = \{z = (z_1, \dots, z_n) \in \mathbb{R}^n \mid z'v = v_1z_1 + \dots + v_nz_n = 0\}$. 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

$$\exists v_1, \dots, v_n \in \mathbb{R}^n, \quad Q = Q(v_1) \dots Q(v_n).$$

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

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

2.2 Matrices carrées diagonalisables

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

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

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 décomposition 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 $\bar{\lambda}_i$ 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 $A^*v^i = \bar{\lambda}_i v^i$. 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 ,

$$n = n_1 + \dots + n_s.$$

Posons $N_0 = 0$ et pour $i = 1$ à s , $N_i = n_1 + \dots + n_i$. 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 $\text{Ker}(A - \lambda_i I)$. En effet les matrices P_i vérifient les propriétés suivantes.

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

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 $\{A^1, \dots, A^n\}$ de vecteurs de \mathbb{K}^m ($m \geq n$), de construire une famille $\{Q^1, \dots, Q^n\}$ orthonormée qui engendre le même espace. La construction de la matrice unitaire $Q_1 = [Q^1, \dots, Q^n]$ à partir de la matrice de plein rang colonne $A = [A^1, \dots, A^n]$ est ainsi basée sur le système d'équations

$$\begin{cases} A^1 &= r_{11}Q^1 \\ A^2 &= r_{12}Q^1 + r_{22}Q^2 \\ \cdot &\cdot \quad \dots \\ A^n &= r_{1n}Q^1 + r_{2n}Q^2 + \dots + r_{nn}Q^n. \end{cases}$$

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

Version "maigre" de la factorisation QR

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

$$A = Q_1 R_1$$

est unique avec Q_1 matrice $m \times n$ unitaire et $R_1 \in \mathcal{C}^{n \times n}$ 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 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 $A \in \mathcal{C}^{m \times n}$ ($m \geq n$), il existe Q et R telles que

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

avec Q matrice $m \times m$ unitaire et $R \in \mathcal{C}^{n \times n}$ 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 $A \in \mathcal{C}^{n \times n}$, il existe une matrice unitaire $U \in \mathcal{C}^{n \times n}$ ($U^*U = UU^* = I_n$) et une matrice triangulaire supérieure T telles que

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

où $T = \Lambda + N$ avec $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ 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

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

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

2.4.2 Matrices normales

Une matrice carrée A est normale si et seulement si $A^*A = AA^*$.

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

Proposition : A est normale si et seulement si il existe une matrice carrée U unitaire telle que

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

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

$$A^*A = U\Lambda^*\Lambda U^* \text{ et } AA^* = U\Lambda\Lambda^*U^*.$$

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

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

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 $\lambda_i = \bar{\lambda}_i$. \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 si et seulement si 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, $\langle Au, u \rangle = \langle \lambda u, u \rangle = \lambda$. Si A est (semi) définie positive λ est positif (positif ou nul).

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

$$\langle Ax, x \rangle = \langle U\Lambda U^{-1}x, x \rangle = \langle \Lambda U^{-1}x, U^{-1}x \rangle .$$

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

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

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

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

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

H4 Critère de Sylvester : Une matrice hermitienne est (semi) définie positive si et seulement si 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 $\Lambda = \text{diag}(\lambda_i)$ avec λ_i positif ou (positif ou nul). Définissons $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_i})$; 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, $\bar{\lambda}_i = -\lambda_i$. \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,

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

et d'après la propriété P8 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 AA^*

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

2.5.1 Deux versions de la DVS

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

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

- $U = [u^1 | \dots | u^r]$ unitaire $m \times r$ est telle que u^i est vecteur propre de AA^* associé à la valeur propre non nulle λ_i .
- $V = [v^1 | \dots | v^r]$ unitaire $n \times r$ est telle que v^i est vecteur propre de A^*A associé à la valeur propre non nulle λ_i .
- $\Lambda_r = \text{diag}(\lambda_1, \dots, \lambda_r)$ et $\Lambda_r^{1/2} = \text{diag}(\mu_1, \dots, \mu_r)$, 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

$$A = P\Lambda Q^*$$

$P = [u^1 | \dots | u^r | u^{r+1} | \dots | u^m] = [U | \tilde{U}]$ carrée $m \times m$ unitaire et $Q = [v^1 | \dots | v^r | v^{r+1} | \dots | v^n] = [V | \tilde{V}]$ 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 \mathbb{K}^m (resp. \mathbb{K}^n) : $P^*P = PP^* = I_m$, $Q^*Q = QQ^* = I_n$. De plus

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

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

Preuve :

AA^* et A^*A sont hermitiennes (symétriques), semi-définies positives et ont mêmes valeurs propres non nulles. Par Schur, $\exists Q$ unitaire telle que

$$Q^*A^*AQ = \text{diag}(\mu_1^2, \dots, \mu_n^2).$$

Posons $C = AQ$. Soit c^j la j ème colonne de C , alors $c^{i*}c^j = \mu_i^2 \delta_{ij}$.

Comme $\text{rang}(AQ) = r$, on peut supposer que μ_1, \dots, μ_r sont strictement positifs et que μ_{r+1}, \dots, μ_n sont nuls.

Alors $c^j = 0_m$ pour $j = r + 1, \dots, n$, car c'est un vecteur associé à une valeur propre

nulle et donc $\|c^j\|_2^2 = 0$.

On construit P par colonnes :

$$p^j = \mu_j^{-1} c^j \text{ si } j = 1, \dots, r. \quad (*)$$

$$p^{i*} p^j = (\mu_i \mu_j)^{-1} c^{i*} c^j = \delta_{ij} \text{ pour } 1 \leq i, j \leq r.$$

On complète la matrice P pour former une base orthonormée de \mathbb{K}^m , $P = [p^1 \dots p^m]$ et $\forall i, j \in \{1, \dots, m\}$ $p^{i*} p^j = \delta_{ij}$.

Vérifions DVS2 en calculant un élément de la matrice $P^* A Q$:

$$\text{Si } 1 \leq i, j \leq r, [P^* A Q]_{ij} = p^{i*} c^j = \mu_j \delta_{ij}.$$

$$\text{Si } j > r, [P^* A Q]_{ij} = 0 \text{ car alors } c^j = 0_m.$$

$$\text{Si } j \leq r \text{ et si } i > r, p^{i*} c^j = \mu_j p^{i*} p^j = 0.$$

Donc, $P^* A Q = \Lambda$.

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

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

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, $\mu_1 \geq \dots \geq \mu_r > 0$, 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, \dots, v^r de $A^* A$, ceux de AA^* associés aux valeurs propres non nulles, sont donnés par

$$U = A V \Lambda_r^{-1/2}, \quad (*)$$

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

$$V = A^* U \Lambda_r^{-1/2}. \quad (**)$$

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

Corollaire 1 : La décomposition en valeurs singulières donne

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

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

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 coïncide avec la DVS (modulo le signe des vecteurs propres).

Corollaire 2 : Soit A $m \times n$ avec $\text{rang}(A) = r$. La DVS2 de A s'écrit $A = P\Lambda Q^*$, avec $P = [U|\tilde{U}]$ et $Q = [V|\tilde{V}]$. Alors,

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

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

Preuve de la première égalité

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

$$\text{Comme } A = U\Lambda_r^{1/2}V^* \text{ d'après la DVS1, } x \in \text{Ker } A \Leftrightarrow U\Lambda_r^{1/2}V^*x = 0.$$

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

$$\Lambda_r^{1/2}U^*U\Lambda_r^{1/2}V^*x = 0 \Leftrightarrow \Lambda_r V^*x = 0 \Leftrightarrow V^*x = 0, \text{ puisque } \Lambda_r \text{ est inversible.}$$

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

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

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

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

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 carrée $n \times n$, 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

$$A = P\Lambda Q^* = (PQ^*)(Q\Lambda Q^*) \doteq UH$$

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

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

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 \mathbb{R}^n et définit ainsi le produit scalaire :

$$\forall x, y \in \mathbb{R}^n, \langle x, y \rangle_A = y'Ax.$$

La décomposition de Cholesky de A donne

$$\langle x, y \rangle_A = \langle Tx, Ty \rangle_{I_n}.$$

Changer de métrique revient donc à effectuer une 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

$$\langle x, y \rangle_A = \langle A^{1/2}x, A^{1/2}y \rangle_{I_n}.$$

2.7 Exercices

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

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

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

1.2 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 $A \in \mathcal{C}^{n \times n}$, 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 $\|A\|_F = (\text{trace}(A^*A))^{1/2}$, montrer que pour une matrice A carrée quelconque, on peut mesurer l'écart de A à la normalité grâce à la norme de Frobénius de la partie hors diagonale N de la matrice T dans la décomposition 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 $\|AA^* - A^*A\|_F^2$.

Exercice 5 : Décomposer la matrice

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

en valeurs singulières. Dans l'espace vectoriel \mathbb{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 $\{Im A'\}^\perp$.

Exercice 6 : Critère de Sylvester

On se propose de montrer que pour A symétrique réelle,

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

Soit $A = [a_{ij}]$ une matrice réelle symétrique, $n \times n$, définie positive. On appelle k ième matrice principale 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 mineur principal de A , le déterminant de A_k , noté $\det(A_k)$.

- 6.1. a. Montrer que $a_{ii} > 0$ pour $i = 1, \dots, n$.
 b. Ecrire la décomposition spectrale de A . Calculer $\det(A)$ en fonction des valeurs propres de A . Dire pourquoi $\det(A) > 0$ et A^{-1} est symétrique définie positive.
- 6.2. Soient B et P les matrices $(n + 1) \times (n + 1)$ définies par

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

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

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

- b. Montrer que $\det(B) \leq \alpha \det(A)$ et que l'égalité 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 $\det(B) > 0$ si et seulement si C est définie positive.
 e. Dédurre des précédentes questions que $\det(B) > 0$ si et seulement si B est définie positive

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

a. Sens \implies : 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 \impliedby : 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.

Exercice 7 : 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 $\text{Ker } X$, puis le vecteur projection de y sur $\text{Ker } X$.
 3. Ecrire l'expression analytique du plan vectoriel $\text{Im } X$, l'expression de la matrice de la projection orthogonale sur $\text{Im } X$, puis le vecteur projection de y sur $\text{Im } X$.
-

Chapitre 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.

3.1 Normes de vecteurs

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

Une norme sur E est une application $\|\cdot\|$ de E dans \mathbb{R}^+ telle que

$$\|x\| = 0 \iff x = 0_E$$

$$\|\lambda x\| = |\lambda| \|x\| \quad x \in E, \lambda \in \mathbb{K}$$

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

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

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

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 $\langle \cdot, \cdot \rangle$ sur E ,

$$\|x\|_2 = \sqrt{\langle x, x \rangle}.$$

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_∞ .

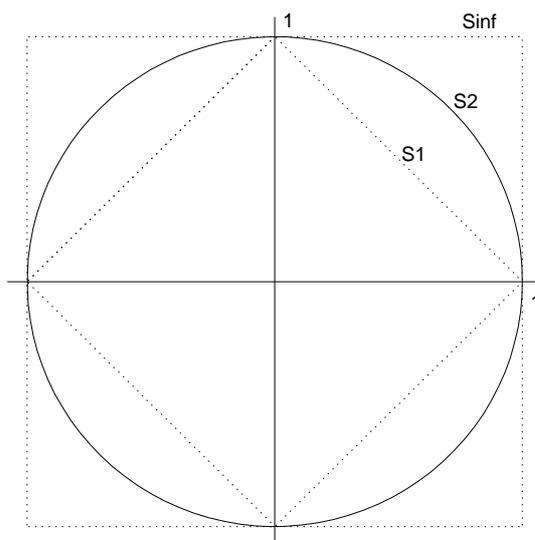


Figure 3 : sphère unité pour les p -normes usuelles, $p = 1$, $p = 2$ et $p = +\infty$.

Toutes les normes de \mathbb{K}^n sont équivalentes, c'est à dire si $\| \cdot \|_\alpha$ et $\| \cdot \|_\beta$ 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

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

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

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

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

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

est l'erreur relative. En particulier, si

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

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

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

Les normes sur \mathbb{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, $E = \mathcal{C}^n$ (ou \mathbb{R}^n), la donnée d'une matrice M hermitienne (symétrique) définie positive, appelée une métrique, permet de définir un produit scalaire

$$\langle x, y \rangle_M = y^* M x = \langle M x, y \rangle_{I_n} .$$

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

$$\|x\|_M = \sqrt{\langle x, x \rangle_M} = \sqrt{x^* M x} .$$

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 $S_M = \{x \in E \mid x^* M x = 1\}$ est l'ensemble de niveau 1 de la forme quadratique définie positive $x \rightarrow \|x\|_M^2 = x^* M x$.

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 .

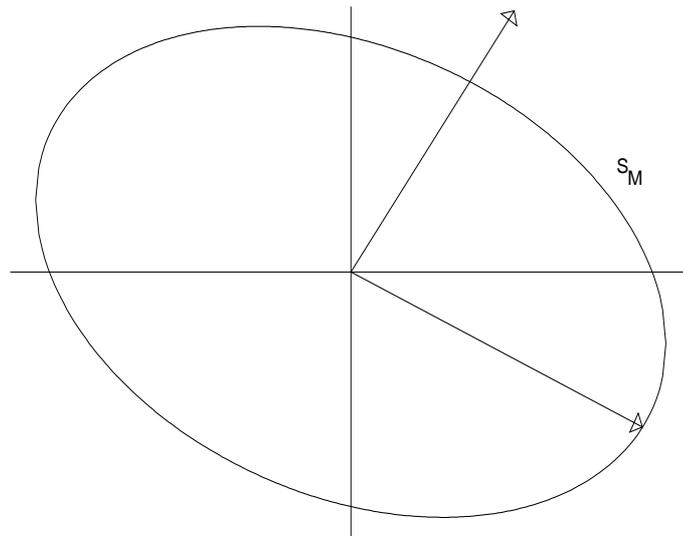


Figure 4 : sphère unité pour l'espace (\mathbb{R}^2, M) .

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 $\text{Ker } M$. Le graphe $G = \{(x, y = x'Mx) \in \mathbb{R}^3 \mid x \in \mathbb{R}^2\}$ est représenté selon le cas, par :

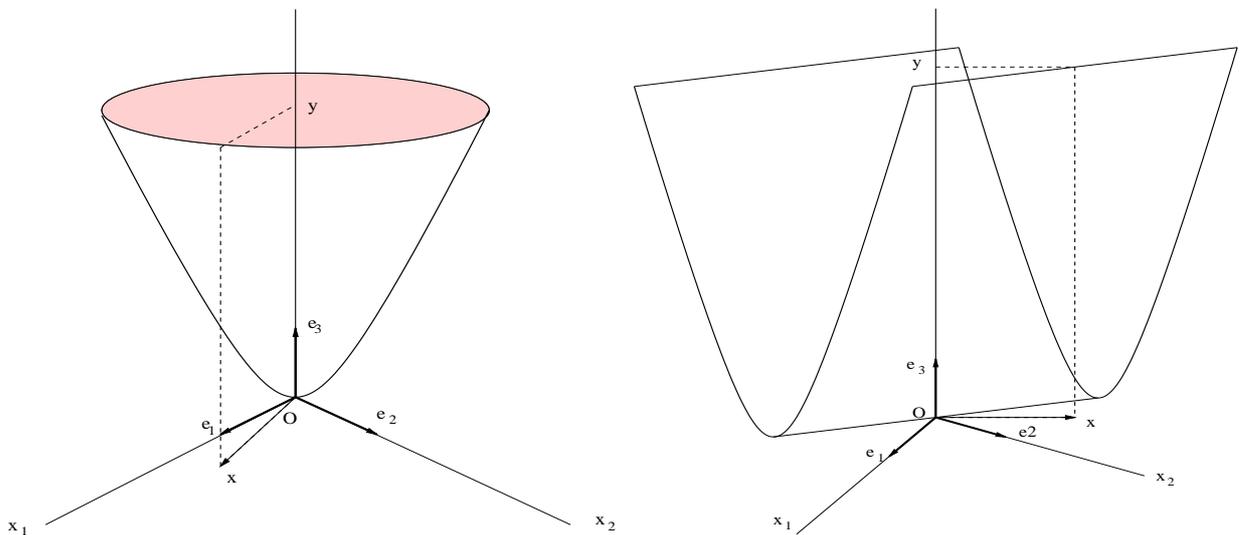


Figure 5 : graphe d'une forme quadratique définie positive puis semi définie positive de (\mathbb{R}^2, M) .

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

$$\langle x, y \rangle_M = y^* M x = 0.$$

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

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

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

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

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

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

Normes matricielles subordonnées

Soient $E = \mathbb{K}^n$ et $F = \mathbb{K}^m$ munis de la même p -norme $\|\cdot\|_p$ 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 $A = [a_{ij}]$ de dimensions $m \times n$ telle que

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

Les éléments de la j ème 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 \mathbb{K}^n dans \mathbb{K}^m munis de n'importe quelle norme, étant continue, on définit

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

pour un \hat{x} de la sphère unité de \mathbb{K}^n . On obtient la relation

$$\|Ax\|_p \leq \|A\|_p \|x\|_p \quad x \in \mathbb{K}^n.$$

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

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

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

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

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

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

$$x = \sum_{i=1}^n x_i v^i \quad \text{et} \quad x^* = \sum_{j=1}^n \bar{x}_j v^{j*}.$$

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, A est hermitienne, 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 $\mathcal{C}^{m \times n}$,

$$\text{trace}(Y^*X) = \sum_{j=1}^n \sum_{i=1}^m \bar{y}_{ij} x_{ij} = \text{vec}^*(Y) \text{vec}(X) = \langle \text{vec}(X), \text{vec}(Y) \rangle,$$

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

$$\langle X, Y \rangle = \text{trace}(Y^*X).$$

La norme de Frobenius de $A \in \mathcal{O}^{m \times n}$ 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 $\text{rang}(A) = r$,

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

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 Frobenius et la norme vectorielle Euclidienne sont compatibles.

Norme de Hilbert-Schmidt

En analyse statistique des données on est amené 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,

$$\text{trace}(Y'DXM) = \text{vec}'(Y)(M \otimes D)\text{vec}(X) = \langle \text{vec}(X), \text{vec}(Y) \rangle_{M \otimes D}.$$

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

$$\langle X, Y \rangle_{M \otimes D} = \text{trace}(Y'DXM).$$

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 ,

$$\|X\|_{I_p \otimes I_n}^2 = \text{trace}(X'X) = \|X\|_F^2.$$

– $(\mathbb{R}^{n \times p}, M = I_p, D)$. 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 ,

$$\|X\|_{I_p \otimes D}^2 = \text{trace}(X'DX) = \sum_{i=1}^p \|X^i\|_D^2.$$

– (\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,

$$\|X\|_{D \otimes D}^2 = \text{trace}((XD)^2).$$

Remarque

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

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

$$\|AB\| \leq \|A\|\|B\| \quad \forall A \in \mathcal{C}^{m \times n}, \forall B \in \mathcal{C}^{n \times p}.$$

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 $X = [x, 0, \dots, 0]$ où les colonnes sont nulles sauf la première égale à x . Il est clair que $\|x\| = \|X\|$ est une norme vectorielle. Donc, $\|Ax\| = \|[Ax, 0, \dots, 0]\| = \|AX\| \leq \|A\|\|X\| = \|A\|\|x\|$. \square

Proposition : Quelle que soit la norme matricielle $\|\cdot\|$ 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 A est hermitienne, 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 $\mathcal{C}^{m \times n}$ ($m \geq n$), \mathcal{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

$$U^*MU = I_n.$$

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

$$U'MU = I_n.$$

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

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

P5 : Pour les M -normes vectorielles, on obtient

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

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

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

3.3 Suites de matrices

Comme pour les suites de vecteurs, en considérant l'ensemble $\mathcal{C}^{m \times n}$ 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) $\lim_{k \rightarrow \infty} B^k = 0.$
- (ii) $\lim_{k \rightarrow \infty} B^k v = 0$ pour tout vecteur $v.$
- (iii) $\rho(B) < 1.$
- (iv) $\|B\| < 1$ pour au moins une norme matricielle subordonnée $\|\cdot\|.$ \square

Proposition : La série $I + B + B^2 + \dots$ converge vers $(I - B)^{-1}$ si et seulement si

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

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 entraîner une grande perturbation sur la solution du problème considéré.

Prenons l'exemple dû à R.S. Wilson du système linéaire

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

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 ($\det(A) = 1$), 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 $\|\cdot\|$ 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 : $\text{cond}_2(A) = \frac{\mu_1}{\mu_n}$ où μ_1 et μ_n sont respectivement la plus grande et la plus petite des valeurs singulières de A .

P4 : $\text{cond}_2(A) = 1$ si et seulement si $A = \alpha Q$ 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

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

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

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

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

$$B_M = \{x \in E \mid x' M x = 1\}.$$

Indications : Faire un changement de repère de telle façon 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 pourra 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 Frobenius est sous multiplicative ainsi que toute p norme subordonnée.

Chapitre 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

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

dit de Cramer, admet une solution unique $\hat{x} = A^{-1}y$. 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 maintenant que toutes les matrices sont réelles car résoudre un système linéaire sur \mathcal{C} revient à résoudre deux systèmes 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 : $\hat{x} = A^-y$ est solution de $Ax = y$, $\forall y \in ImA$.

Définition 2 : $AA^-A = A$.

Équivalence des définitions

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

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

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

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

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

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

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

Quelques propriétés de 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

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

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

On sait que $P_AP'_A = A(A'A)^-A'[(A'A)^-]'A' = P'_A$ 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

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

P8 : Soit A^- un inverse généralisé de A . Une solution générale du système homogène

$$Ax = 0 \text{ est : } \quad x = (I - A^-A)u, \quad \text{pour } u \text{ arbitraire.}$$

P9 : Soit A^- un inverse généralisé de A et $y \in Im A$. Une solution générale du système

$$Ax=y \text{ est : } \quad x = A^-y + (I - A^-A)u, \quad \text{pour } u \text{ arbitraire.}$$

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

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 $\text{rang}(A) = r$. 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 $\Lambda_r^{1/2} = \text{diag}\{\mu_1, \dots, \mu_r\}$ avec $\mu_i > 0$. D'où

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

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**, $A_1^+ = A_1^+AA_1^+ = A_1^+(AA_1^+)'$,

$$A_1^+ = A_1^+(A_1^+)'A' = A_1^+(A_1^+)'(AA_2^+)' \text{ par } \mathbf{D1}$$

$$A_1^+ = A_1^+(A_1^+)'A'(AA_2^+)' = A_1^+(A_1^+)'A'AA_2^+ = A_1^+(AA_1^+)'AA_2^+ = A_1^+AA_1^+AA_2^+ \text{ par } \mathbf{D3}$$

$$A_1^+ = A_1^+AA_2^+ \text{ par } \mathbf{D1}.$$

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

$$A_2^+ = (A_2^+AA_1^+)'A_2^+ = (A_1^+A)'(A_2^+A)'A_2^+ \text{ par } \mathbf{D1}$$

$$A_2^+ = A_1^+AA_2^+AA_2^+ = A_1^+AA_2^+ \text{ par } \mathbf{D4} \text{ et } \mathbf{D2}.$$

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

Propriétés de A^+

P0 : Si $A = kB$ alors $A^+ = k^{-1}B^+$ avec $k \in \mathbb{R}_*$.

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

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 : $A = 0 \iff A^+ = 0$.

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

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

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

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

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

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

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

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 $P_A = A(A'A)^-A' = AA^+$ grâce 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

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

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 $\mathbb{E}_1^\perp = \{y \in \mathbb{E} \mid \langle y, x \rangle_M = 0, x \in \mathbb{E}_1\}$. 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 si et seulement si

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

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

Une 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 (αI_m) -symétrique est symétrique au sens usuel.

Propriété caractéristique d'un projecteur M -orthogonal

Proposition : Un projecteur P est un projecteur M -orthogonal si et seulement si P est M -symétrique. \square

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

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

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

Puisque 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 , Π_A^M projecteur M -orthogonal sur l'espace $Im A$ engendré par les colonnes de A , est la matrice $m \times m$

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

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 $A'M$,

$$A'Mx = A'Mx_1 + A'Mx_2 = A'Mx_1 = A'MA\beta.$$

Le système linéaire

$$A'MA\beta = A'Mx$$

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

$$\beta = (A'MA)^{-} A'Mx$$

ce qui donne

$$x_1 = A\beta = A(A'MA)^{-} A'Mx \quad \text{et} \quad \Pi_A^M = A(A'MA)^{-} A'M.$$

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 $\Pi_A^M = T^{-1}B(B'B)^{-}B'T$. 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 $\Pi_A^M = A(A'MA)^+ A'M$ et, lorsque A est de plein rang colonne $\Pi_A^M = A(A'MA)^{-1} A'M$.

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

$$A'MA = \text{diag}(\|A^1\|_M^2, \dots, \|A^n\|_M^2).$$

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

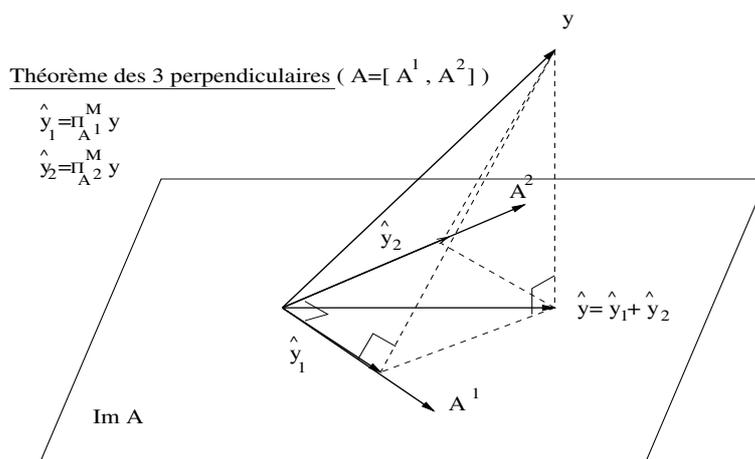


Figure 6 : Théorème des 3 perpendiculaires : $\Pi_{[A^1, A^2]}^M y = \Pi_{A^1}^M y + \Pi_{A^2}^M y$.

$$\Pi_A^M = \sum_{i=1}^n \Pi_{A^i}^M,$$

où $\Pi_{A^i}^M = A^i A^{i'}/\|A^i\|_M^2$. Si de plus la base est M -orthonormée, c'est à dire si A est M -orthogonale, alors $\Pi_{A^i}^M = A^i A^{i'}/M$. L'orthogonalité des colonnes de A donne $\|\Pi_A^M y\|_M^2 = \sum_{i=1}^n \|\Pi_{A^i}^M y\|_M^2$.

Lorsque $y \in \text{Im } A$, $\|y\|_M^2 = \sum_{i=1}^n \|\Pi_{A^i}^M y\|_M^2$, c'est le cas en particulier lorsque la matrice A est carrée (les colonnes de A forment une base de l'espace tout entier).

Éléments propres du projecteur M -orthogonal sur $\text{Im } A$

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

$$\text{trace}(\Pi_A^M) = \text{rang}(A) = \text{rang}(\Pi_A^M). \quad \square$$

Preuve : Π_A^M est M -symétrique ($M\Pi_A^M = \Pi_A^M M$). 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

$$\Pi_A^M = M^{-1/2}P \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} P^{-1}M^{1/2} = M^{-1/2}UU'M^{1/2}.$$

Les vecteurs propres de Π_A^M sont les colonnes de $M^{-1/2}P$ qui sont M -orthonormés. \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 $y \in \text{Im } A$. On peut toujours se poser le problème de chercher $\hat{x} \in \mathbb{R}^n$ tel que

$$\|A\hat{x} - y\|_M^2 = \min_{x \in \mathbb{R}^n} \|Ax - y\|_M^2. \quad (*)$$

Ce problème s'écrit comme la recherche de l'élément $\hat{y} = A\hat{x}$ de $\text{Im } A$ le plus proche de y au sens de la norme $\|\cdot\|_M$.

$$\|\hat{y} - y\|_M^2 = \min_{v \in \text{Im } A} \|v - y\|_M^2. \quad (*)$$

Proposition : Une solution du problème aux moindres carrés (*) est donnée par

$$\hat{x} = (A'MA)^+ A'My.$$

La solution est unique si $\text{rang}(A) = n$. Si la solution n'est pas unique, toute solution s'écrit $\bar{x} = \hat{x} + u$ avec $u \in \text{Ker}(A'MA) = \text{Ker } A$. L'approximation $\hat{y} = A\bar{x} = A\hat{x} = \Pi_A^M y$ de y est unique, c'est la projection M -orthogonale de y sur $\text{Im } A$. \square

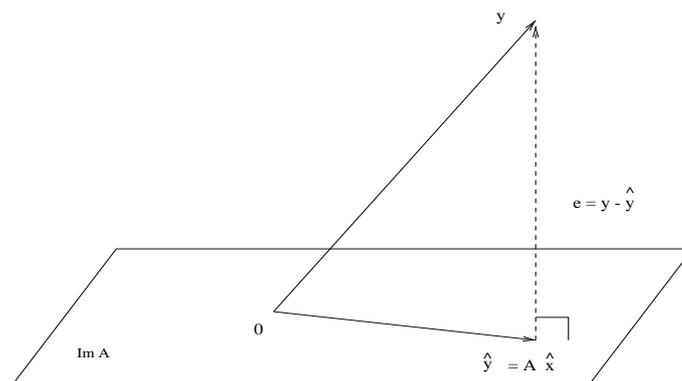


Figure 7 : solution du problème aux moindres carrés, $\hat{y} = \Pi_A^M y$.

Preuve : Notons

$$\varphi(x) = \|Ax - y\|_M^2$$

la fonction à minimiser dite fonction objectif. C'est une fonction de \mathbb{R}^n dans \mathbb{R}^+ , c'est à dire une fonction numérique des n variables $[x_1, \dots, x_n]' = x$.

$$\varphi(x) = (Ax - y)'M(Ax - y) = x'A'MAx - 2y'MAx + y'My.$$

Cette fonction est une fonction quadratique différentiable sur \mathbb{R}^n . Les équations aux dérivées partielles dites équations normales fournissent les points stationnaires

$$\nabla\varphi(x) = 2A'MAx - 2A'My = 0_{\mathbb{R}^n},$$

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 Hessian $H\varphi(x)$, 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

$$A'MAx = A'My.$$

Si la matrice $A'MA$ $n \times n$ est inversible, c'est à dire si $\text{rang}(A) = \text{rang}(A') = n$, alors il y a une seule solution

$$\hat{x} = (A'MA)^{-1}A'My.$$

Dans le cas contraire, le système admet une infinité de solutions car le second membre appartient à $\text{Im } A'$ et puisque M est régulière $\text{Im } A' = \text{Im } A'MA$. Plutôt que d'écrire les solutions en utilisant un inverse généralisé $(A'MA)^-$ on préfère construire une solution

$$\hat{x} = (A'MA)^+ A'My.$$

Soit \bar{x} une autre solution des équations normales, alors

$$A'MA(\bar{x} - \hat{x}) = 0_{\mathbb{R}^n} \quad \text{et} \quad \bar{x} = \hat{x} + u, \text{ avec } u \in \text{Ker}(A'MA).$$

Puisque M est symétrique définie positive, en décomposant M par Cholesky ou par $M = (M^{1/2})^2$, on montre aisément que $\text{Ker}(A'MA) = \text{Ker} A$. Alors, qu'il y ait unicité ou non, l'approximation de y par \hat{y}

$$\hat{y} = A\bar{x} = A\hat{x} = \Pi_A^M y,$$

où Π_A^M est le projecteur M -orthogonal sur $\text{Im} A$, est unique.

Une façon de mesurer la qualité de l'approximation est fournie par l'indice

$$R^2(y; \text{Im} A) = \frac{\|\hat{y}\|_M^2}{\|y\|_M^2}$$

appelé "coefficient de détermination" entre y et $\text{Im} A$. On le note R^2 pour simplifier. C'est le carré du M -cosinus du M -angle formé par les vecteurs y et \hat{y} .

$$R^2 = 1 \iff y = \hat{y} \iff y \in \text{Im} A.$$

$$R^2 = 0 \iff \hat{y} = 0 \iff y \in \{\text{Im} A\}^\perp. \quad \square$$

Propriétés du projeté $\hat{y} = \Pi_A^M y = A\hat{x}$

P1 $e = y - \hat{y} \in \{\text{Im} A\}^\perp$, c'est à dire $\langle y - \hat{y}, u \rangle_M = 0, \quad \forall u \in \text{Im} A$.

P2 $\|y\|_M^2 = \|y - \hat{y}\|_M^2 + \|\hat{y}\|_M^2$. C'est le théorème de Pythagore.

P3 $\|\Pi_A^M y\|_M \leq \|y\|_M, \quad \forall y \in \mathbb{R}^m$, ou $0 \leq R^2 \leq 1$. On dit que l'application linéaire Π_A^M 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 $\mathbf{1}_p = (1, \dots, 1)'$ et $(1, 2, \dots, p)'$. Calculer la matrice du projecteur orthogonal sur E .

Exercice 2 : Soit $E = \{(y_1, \dots, y_{2n}) \in \mathbb{R}^{2n} \mid y_t = y_{t+2}, 1 \leq t \leq 2n - 2\}$.

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

$$\hat{y} = \arg \min_{y \in E} \|x - y\|_2^2.$$

Chapitre 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 .

Fonctions	Variables	numériques	vectérielles	matricielles
numériques		$\phi(\xi)$	$\phi(x)$	$\phi(X)$
vectérielles		$f(\xi)$	$f(x)$	$f(X)$
matricielles		$F(\xi)$	$F(x)$	$F(X)$

Exemples

$$\begin{aligned} \phi(\xi) &= \xi^2 & \phi(x) &= a'x & \phi(X) &= a'Xa \\ f(\xi) &= \begin{bmatrix} \xi \\ \xi^2 \end{bmatrix} & f(x) &= Ax & f(X) &= Xa \\ F(\xi) &= \begin{bmatrix} \cos(\xi) & -\sin(\xi) \\ \sin(\xi) & \cos(\xi) \end{bmatrix} & F(x) &= xx' & F(X) &= X^{-1}. \end{aligned}$$

Matrices Jacobiennes de fonctions différentiables

- Soit $\phi : x \rightarrow \phi(x)$, une fonction numérique différentiable de \mathbb{R}^n dans \mathbb{R} . La dérivée $D\phi(x)$ ou matrice Jacobienne s'écrit

$$D\phi(x) = [D_1\phi(x), \dots, D_n\phi(x)] = \frac{\partial\phi(x)}{\partial x'}$$

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

$$x \mapsto f(x) = \begin{bmatrix} \phi_1(x) \\ \vdots \\ \phi_m(x) \end{bmatrix}.$$

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 $df = [d\phi_1, \dots, d\phi_m]'$.

- Soit F une fonction matricielle différentiable de $\mathbb{R}^{n \times q}$ dans $\mathbb{R}^{m \times p}$,

$$X = [x_{ij}] \longrightarrow F(X) = \begin{bmatrix} \phi_{11}(X) & \cdots & \phi_{1p}(X) \\ \vdots & \ddots & \vdots \\ \phi_{m1}(X) & \cdots & \phi_{mp}(X) \end{bmatrix}$$

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 $D_{ij}\phi_{kl}(X) = \frac{\partial \phi_{kl}(X)}{\partial x_{ij}}$. On la note

$$DF(X) \doteq \frac{\partial \text{vec}(F(X))}{\partial \text{vec}'(X)}.$$

Le théorème d'identification pour les fonctions matricielles différentiables donne

$$d \text{vec}(F(X)) = A(X) d \text{vec}(X) \iff A(X) = DF(X).$$

Formulaire

$$X = [x_{ij}] \ n \times p, \ dX = \begin{bmatrix} dx_{11} & \cdots & dx_{1p} \\ \vdots & \ddots & \vdots \\ dx_{n1} & \cdots & dx_{np} \end{bmatrix}.$$

$$dC^{ste} = 0.$$

$$d(X + Y) = dX + dY.$$

$$d(\alpha X) = \alpha dX.$$

$$d(X') = (dX)'$$

$$d(\text{trace}(X)) = \text{trace}(dX).$$

$$d(\text{vec}(X)) = \text{vec}(dX).$$

$$d(XY) = (dX)Y + X(dY).$$

$$d(X \otimes Y) = ((dX) \otimes Y) + (X \otimes (dY)).$$

Tableau d'identification

fonctions	différentielles	matrices Jacobiennes	ordre de D	autres notations
$\phi(\xi)$	$d\phi = \alpha d\xi$	$D\phi(\xi) = \alpha$	1×1	
$\phi(x)$	$d\phi = a'dx$	$D\phi(x) = a'$	$1 \times n$	$\nabla\phi(x) = a$
$\phi(X)$	$d\phi = \text{vec}'(A)d\text{vec}(X)$ $= \text{trace}(A'dX)$	$D\phi(X) = \text{vec}'(A)$	$1 \times nq$	$\frac{\partial\phi(X)}{\partial X} = A \quad n \times q$
$f(\xi)$	$df = ad\xi$	$Df(\xi) = a$	$m \times 1$	
$f(x)$	$df = Adx$	$Df(x) = A$	$m \times n$	
$f(X)$	$df = A\text{dvec}(X)$	$Df(X) = A$	$m \times nq$	
$F(\xi)$	$d\text{vec}(F) = \text{vec}(A)d\xi$	$DF(\xi) = \text{vec}(A)$	$mp \times 1$	$\frac{dF(\xi)}{d\xi} = A \quad m \times p$
$F(x)$	$d\text{vec}(F) = Adx$	$DF(x) = A$	$mp \times n$	
$F(X)$	$d\text{vec}(F) = A\text{dvec}(X)$	$DF(X) = A$	$mp \times nq$	

Remarque : Dans deux cas particuliers on réorganise les dérivées partielles

* Soit $\phi(X)$ avec $X = [x_{ij}] \quad n \times q$, alors $d\phi = \sum_{i=1}^n \sum_{j=1}^q \frac{\partial\phi(X)}{\partial x_{ij}} dx_{ij} = \text{trace}((\frac{\partial\phi(X)}{\partial X})'dX)$.

$$\frac{\partial\phi(X)}{\partial X} = \left[\frac{\partial\phi(X)}{\partial x_{ij}} \right] \quad n \times q.$$

* Soit $F(\xi)$ avec $F(\xi) = [F_{ij}(\xi)]$ qui est d'ordre $m \times p$

$$\frac{dF(\xi)}{d\xi} = \left[\frac{dF_{ij}(\xi)}{d\xi} \right].$$

Marche à suivre (présentée dans le cas $F(X)$) :

- 1) On calcule $dF(X)$.
- 2) On vectorise $dF(X)$ c'est-à-dire, on calcule $\text{vec}(dF(X)) = d\text{vec}(F(X))$.
- 3) On met sous la forme $d\text{vec}(F(X)) = A(X)d\text{vec}(X)$.
- 4) On identifie $A(X) = DF(X)$.

Dérivation des fonctions matricielles composées :

Soit S un ouvert de $\mathbb{R}^{m \times q}$ et supposons que $F : S \rightarrow \mathbb{R}^{m \times p}$ soit différentiable en $C \in S$. Soit T un ouvert de $\mathbb{R}^{m \times p}$ tel que $F(X) \in T$ pour tout $X \in S$ et supposons que $G : T \rightarrow \mathbb{R}^{r \times s}$ soit différentiable en $B = F(C)$. Alors, la fonction composée $H : S \rightarrow \mathbb{R}^{r \times s}$ définie par

$$H(X) = G(F(X))$$

est différentiable en C et

$$DH(C) = [DG(B)][DF(C)].$$

La formule précédente s'appelle la règle de "dérivation en chaîne" pour les Jacobiens. \square

Exemple : Calcul du Jacobien de $\phi(X) = \text{trace}(AXB)$ avec A et B constantes.

Première méthode : $d\phi(X) = d\text{trace}(AXB) = \text{trace}(d(AXB)) = \text{trace}(A(dX)B)$
 $= \text{trace}(BA(dX)) = \text{vec}'(A'B')d\text{vec}(X)$
 $\iff D\phi(X) = \text{vec}'(A'B')$

ou en réorganisant les dérivées partielles selon la position des éléments de X

$$\frac{\partial \text{trace}(AXB)}{\partial X} = A'B'. \quad \square$$

Deuxième méthode : Soit $F(X) = AXB$, alors $dF(X) = A(dX)B$

$$\text{vec}(dF(X)) = d\text{vec}(F(X)) = (B' \otimes A)\text{vec}(dX).$$

Donc $DF(X) = B' \otimes A$.

Soit $\phi(Y) = \text{trace}(Y)$, alors $d\phi(Y) = \text{trace}(dY) = \text{trace}(IdY) = \text{vec}'(I)\text{vec}(dY)$.

Donc $D\text{trace}(Y) = \text{vec}'(I)$. 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

$$D\phi(F(X)) = \text{vec}'(I)(B' \otimes A) = [(B' \otimes A)\text{vec}(I)]' = \text{vec}'(A'IB') = \text{vec}'(A'B'). \quad \square$$

5.2.2 Hessien de fonctions numériques

Soit $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction deux fois différentiable, alors

$$H\phi(x) = [D_{ij}^2\phi(x)]$$

est une matrice symétrique d'ordre n appelée le Hessien de ϕ en x .

Théorème d'identification

$$d^2\phi = d(d\phi) = (dx_2)'B(dx_1) \iff H\phi(x) = \frac{1}{2}(B + B').$$

Tableau d'identification

<i>Fonctions</i>	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

$$d^2\phi = \text{trace}(B(dX_2)'C(dX_1)) \iff H\phi(X) = \frac{1}{2}(B' \otimes C + B \otimes C')$$

$$d^2\phi = \text{trace}(B(dX_2)C(dX_1)) \iff H\phi(X) = \frac{1}{2}K_{qn}(B' \otimes C + C' \otimes B). \quad \square$$

Preuve

On sait que $\text{trace}(ABCD) = \text{vec}'(D')(C' \otimes A)\text{vec}(B) = \text{vec}'(D)(A \otimes C')\text{vec}(B')$.

Donc $\text{trace}(B(dX_2)'C(dX_1)) = \text{vec}'(dX_1)(B \otimes C')\text{vec}(dX_2)$

$$H\phi(X) = \frac{1}{2}(B \otimes C' + B' \otimes C).$$

$\text{trace}(BdX_2CdX_1) = \text{vec}'(dX_1)(B \otimes C')\text{vec}(dX_2)' = \text{vec}'(dX_1)(B \otimes C')K_{nq}\text{vec}(dX_2)$

$$H\phi(X) = \frac{1}{2}((B \otimes C')K_{nq} + K_{qn}(B' \otimes C)) = \frac{1}{2}(K_{qn}(C' \otimes B) + K_{qn}(B' \otimes$$

$C))$. \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 $\phi : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ et \bar{x} un point de S . On dit que ϕ présente en \bar{x} un minimum local ou relatif, s'il existe une boule ouverte $B(\bar{x})$ centrée en \bar{x} telle que

$$\phi(x) \geq \phi(\bar{x}) \text{ pour tout } x \in B(\bar{x}) \cap S,$$

un minimum strict local, s'il existe une boule ouverte $B(\bar{x})$ centrée en \bar{x} telle que

$$\phi(x) > \phi(\bar{x}) \text{ pour tout } x \in B(\bar{x}) \cap S, x \neq \bar{x},$$

un minimum global ou absolu, si

$$\phi(x) \geq \phi(\bar{x}) \text{ pour tout } x \in S,$$

un minimum strict global, si

$$\phi(x) > \phi(\bar{x}) \text{ pour tout } x \in S, x \neq \bar{x}.$$

Par abus de langage on dira que le point \bar{x} lui même est un minimum local, local strict...

Transformations strictement croissantes

Théorème 1 : Soit $S \subset \mathbb{R}^n$ et ϕ une fonction de S dans \mathbb{R} . Notons $T = \phi(S)$. Soit φ une application strictement croissante de T dans \mathbb{R} . Alors ϕ présente un minimum absolu (resp. relatif) en \bar{x} si et seulement si $\psi = \varphi \circ \phi$ présente un minimum absolu (resp. relatif) en \bar{x} . \square

Preuve : Soit φ une fonction croissante sur T et \bar{x} un minimum local ou global pour ϕ , c'est à dire $\phi(x) \geq \phi(\bar{x})$ sur un certain ensemble \mathcal{O} . Alors $\psi(x) = \varphi(\phi(x)) \geq \varphi(\phi(\bar{x})) = \psi(\bar{x})$.

Dans l'autre sens, supposons que \bar{x} soit un minimum local ou global pour ψ , $\psi(x) \geq \psi(\bar{x})$ pour tout x dans un certain ensemble \mathcal{O} . Supposons qu'il existe x_0 de \mathcal{O} tel que $\phi(x_0) < \phi(\bar{x})$. La stricte croissance de φ implique $\psi(x_0) = \varphi(\phi(x_0)) < \varphi(\phi(\bar{x})) = \psi(\bar{x})$ ce qui est impossible. Donc $\phi(x) \geq \phi(\bar{x})$ pour tout $x \in \mathcal{O}$. \square

Applications :

A1 : Minimisation d'une norme Euclidienne : La fonction $\|\cdot\|_2$ n'est pas différentiable en 0, aussi pour récupérer la différentiabilité de la fonction objectif, on minimise $\psi = \|\cdot\|_2^2$ plutôt que $\phi = \|\cdot\|_2$ (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 \mathbb{R}^n et ϕ une fonction de Ω dans \mathbb{R} différentiable sur Ω . Si ϕ admet un minimum local en un point \bar{x} de Ω , alors

$$D\phi(\bar{x}) = 0_{1 \times n} \quad \text{ou} \quad \nabla\phi(\bar{x}) = 0_{n \times 1}. \quad (*) \quad \square$$

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 \bar{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 objectif 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 \bar{X} de l'ouvert Ω , s'écrit

$$D\phi(\bar{X}) = 0_{1 \times np} \quad \text{ou encore} \quad \frac{\partial\phi(\bar{X})}{\partial X} = 0_{n \times p}. \quad (*)$$

Conditions suffisantes d'optimalité (convexité) locale

Théorème 3 : Si ϕ est 2 fois différentiable sur Ω ouvert de \mathbb{R}^n , un point critique \bar{x} de Ω est un minimum local si une des conditions suivantes est réalisée

- (1) Il existe une boule $B(\bar{x})$ centrée en $\bar{x} \in \Omega$ telle que le Hessien $H\phi(x)$ est une matrice semi définie positive pour tout $x \in B(\bar{x}) \cap \Omega$.
- (2) Le Hessien $H\phi(\bar{x})$ est une matrice définie positive, c'est à dire, les mineurs principaux du déterminant de $H\phi(\bar{x})$ sont positifs. Dans ce cas le minimum est strict.

□

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 objectif $\phi : \Omega$ ouvert $\subset \mathbb{R}^n \rightarrow \mathbb{R}$ et soit la fonction vectorielle $g = [g_1, \dots, g_m]' : \Omega \rightarrow \mathbb{R}^m$ permettant de définir l'ensemble S des contraintes

$$S = \{x \in \mathbb{R}^n \mid x \in \Omega; g_1(x) = 0, \dots, g_m(x) = 0\}.$$

Le problème est maintenant le suivant :

$$\begin{array}{c} \text{minimiser } \phi(x) \\ \text{sous la contrainte : } x \in S \end{array}$$

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) :

Hypothèses sur les contraintes

Soient $g : \Omega \rightarrow \mathbb{R}^m$ ($n > m$) et $\bar{x} \in \Omega$ tels que

H1 : $g(\bar{x}) = 0$,

H2 : g est différentiable sur une boule ouverte $B(\bar{x})$,

H3 : la matrice Jacobienne $Dg(x)$ $m \times n$ est continue en \bar{x} ,

H4 : $Dg(\bar{x})$ est de rang m .

Hypothèses sur la fonction objectif

Soit $\phi : \Omega \rightarrow \mathbb{R}$ telle que

H5 : ϕ est différentiable en \bar{x} ,

H6 : $\phi(x) \geq \phi(\bar{x})$ pour tout x de $B(\bar{x})$ vérifiant $g(x) = 0$.

Conclusion

Il existe un (unique dû à H4) vecteur $\bar{l} = [\bar{\lambda}_1, \dots, \bar{\lambda}_m]'$ appelé vecteur des multiplicateurs de Lagrange, tel que

$$D\phi(\bar{x}) + \bar{l}' Dg(\bar{x}) = 0_{1 \times n}, \quad (**)$$

ou en transposant

$$\nabla\phi(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i \nabla g_i(\bar{x}) = 0_{n \times 1}. \quad (**)$$

Remarques :

R1 : Méthode des multiplicateurs de Lagrange : Dans la recherche d'optimums locaux, on construit d'abord la fonction de Lagrange $\mathcal{L} : \Omega \times \mathbb{R}^m \rightarrow \mathbb{R}$

$$\mathcal{L}(x, l) = \phi(x) + l'g(x) = \phi(x) + \sum_{i=1}^m \lambda_i g_i(x).$$

Les conditions nécessaires d'optimalité locale pour le problème d'extremums libres

$$\begin{array}{c} \min \quad \mathcal{L}(x, l) \\ \Omega \times \mathbb{R}^m \end{array}$$

s'écrivent

$$\begin{array}{c} \nabla_x \mathcal{L}(x, l) = \nabla\phi(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) = 0 \quad (**) \\ \nabla_l \mathcal{L}(x, l) = g(x) = 0. \end{array}$$

Elles fournissent comme solutions les point critiques (\bar{x}, \bar{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 $\mathbb{R}^{n \times p}$) $\rightarrow \mathbb{R}$ et $G : \Omega \rightarrow \mathbb{R}^{m \times q}$, le problème d'optimisation devient

$$\begin{array}{c} \min \quad \phi(X) \\ G(X) = 0_{m \times q} \end{array}$$

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 $\mathcal{L}(X, \Lambda)$ définie par

$$\mathcal{L}(X, \Lambda) = \phi(X) + \text{vec}'(\Lambda) \text{vec}(G(X)) = \phi(X) + \text{trace}(\Lambda' G(X)).$$

Il faut déterminer les points critiques $(\bar{X}, \bar{\Lambda})$ 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 (\bar{x}, \bar{l}) et que la matrice Jacobienne $m \times n$ $Dg(\bar{x})$ soit de rang m . Si de plus

$$u' H_x \mathcal{L}(\bar{x}, \bar{l}) u > 0 \quad \text{pour tout } u \in \mathbb{R}^n \text{ tel que } Dg(\bar{x}) u = 0_{m \times 1},$$

où $H_x \mathcal{L}(\bar{x}, \bar{l}) = H\phi(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i H g_i(\bar{x})$, alors ϕ présente un minimum strict en \bar{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

$$\frac{\partial \text{trace}(X)}{\partial X} = I ; \quad \frac{\partial \text{trace}(X'X)}{\partial X} = 2X ; \quad \frac{\partial \text{trace}(X^2)}{\partial X} = 2X'$$

b) Etablir le tableau d'identification suivant

$\phi(X)$	$d\phi(X)$	$D\phi(X)$	$\partial\phi(X)/\partial X$
$\text{trace}(AX)$	$\text{trace}(AdX)$	$\text{vec}'(A')$	A'
$\text{trace}(XAX'B)$	$\text{trace}[(AX'B + A'X'B')dX]$	$\text{vec}'(B'XA' + BXA)$	$B'XA' + BXA$
$\text{trace}(XAXB)$	$\text{trace}[(AXB + BXA)dX]$	$\text{vec}'(B'X'A' + A'X'B')$	$B'X'A' + A'X'B'$

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

$F(X)$	$dF(X)$	$DF(X)$
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 :

$\phi(x) = x'Ax$, $\phi(X) = \text{trace}(X'AX)$, $\phi(X) = \text{trace}(AXBX')$, $\phi(X) = \text{trace}(X^2)$.

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

$$\hat{X} = A^{-1}B'(BA^{-1}B')^{-1}.$$

Indication : exprimer le problème sous forme d'un problème d'optimisation avec contraintes de type égalité et le résoudre.

Chapitre 6

Le paysage mathématique et statistique de l'Analyse Factorielle de Données : la géométrie Euclidienne

L'analyse factorielle de données a pris son essor dans le deuxième tiers du *XX*^{ème} siècle avec l'apparition des ordinateurs et le devoir de traiter des données dont la taille n'a fait que croître 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.

- $T = [T_i^j] \in \mathbb{R}^{n \times p}$ est la matrice des données brutes exprimant les n mesures de p variables, x^1, \dots, 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 \mathbb{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 (\mathbb{R}^p, M) , de l'échantillon du i ème individu.
- D , métrique sur l'espace Euclidien \mathbb{R}^n des colonnes de T , est une matrice, $n \times n$, qui sera toujours diagonale $D = \text{diag}(p_1, \dots, p_n)$.
La j ème colonne de T , notée T^j , sera considérée comme l'expression dans la base canonique de (\mathbb{R}^n, D) , de l'échantillon de la j ème variable, x^j .

Les espaces Euclidiens (\mathbb{R}^n, D) et (\mathbb{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'ambiguïté, 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, $\text{Im } T$, et l'espace des individus échantillonnés, $\text{Im } T'$, sont de même dimension r

$$\dim \text{Im } T = \dim \text{Im } T' = r.$$

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

$$\begin{cases} D = \text{diag}(p_1, \dots, p_n) \\ \text{trace}(D) = \sum_i p_i = 1 \\ p_i > 0 \quad \text{pour } i = 1 \dots n, \end{cases}$$

est appelée la matrice des poids statistiques affectés aux n individus.

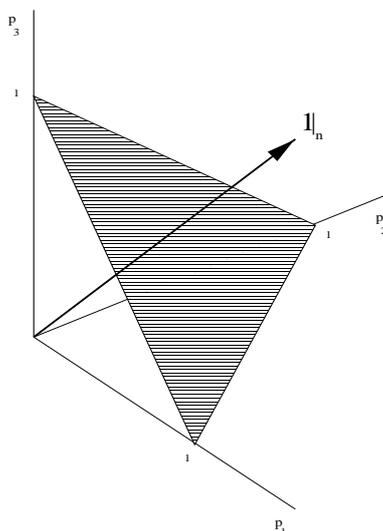


Figure 8 : simplexe des poids statistiques ($n = 3$).

On appelle simplexe des poids l'ensemble

$$\mathcal{D} = \{(p_1, \dots, p_n) \in \mathbb{R}^n \mid p_i \geq 0, i = 1, \dots, n \text{ et } \sum_i p_i = 1\}.$$

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 coïncide avec la projection au sens usuel.

Soit $\mathbf{1}_n = \sum_i e_i(n) = [1, \dots, 1]'$. Le vecteur $\mathbf{1}_n$ est I_n -orthogonal au plan vectoriel $\mathbf{1}'_n x = x_1 + \dots + x_n = 0$ qui est associé au plan affine $x_1 + \dots + x_n = 1$ 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

$$P_{1_n} = \mathbf{1}_n (\mathbf{1}'_n D \mathbf{1}_n)^{-1} \mathbf{1}'_n D = \mathbf{1}_n \mathbf{1}'_n D = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \\ p_1 & p_2 & \cdots & p_n \\ \cdot & \cdot & \cdots & \cdot \\ p_1 & p_2 & \cdots & p_n \end{bmatrix}.$$

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,

$$\bar{t} = \sum_i p_i t_i = \mathbf{1}'_n D t = \langle \mathbf{1}_n, t \rangle_D.$$

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

$$x = \begin{bmatrix} t_1 - \bar{t} \\ t_2 - \bar{t} \\ \cdots \\ t_n - \bar{t} \end{bmatrix} = t - \bar{t} \mathbf{1}_n = (I_n - P_{1_n}) t = P_{1_n}^\perp t$$

c'est à dire projeter t sur le sev de \mathbb{R}^n D -orthogonal à $\mathbf{1}_n$. \square

Preuve :

$$x = t - \bar{t} \mathbf{1}_n = t - \mathbf{1}_n \bar{t} = t - P_{1_n} t. \quad \square$$

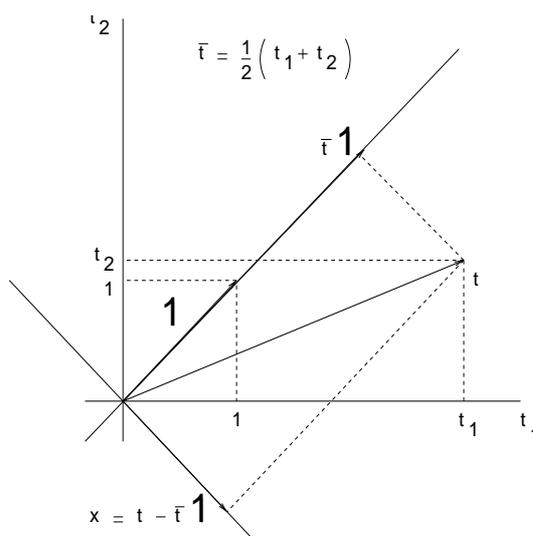


Figure 9 : La moyenne vue comme une projection sur la droite des constantes ($n = 2$).

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” appelé 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, \dots, 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 , $\sigma(t) = \sqrt{\text{var}(t)} = \|x\|_D = \sigma(x)$. \square

Preuve : Par définition, $\text{var}(t) = \sum_i p_i (t_i - \bar{t})^2 = x'Dx = \|x\|_D^2$. \square

Le vecteur $x/\|x\|_D$ est appelé la variable déduite de t par centrage et réduction (au sens de D), on 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 -échantillons $t = [t_1, \dots, t_n]'$ et $u = [u_1, \dots, u_n]'$ 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 entre t et u est définie par

$$\text{cov}(t, u) = \sum_{i=1}^n p_i (t_i - \bar{t})(u_i - \bar{u}) = \sum_{i=1}^n p_i x_i y_i = \text{cov}(x, y).$$

De façon évidente, $\text{cov}(t, u) = \text{cov}(u, t)$. La variance d'une variable peut être définie à partir de la covariance, $\text{var}(t) = \text{cov}(t, t)$, 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, $\text{cov}(t, u) = x'Dy = \langle x, y \rangle_D = \text{cov}(x, y)$.

Formule développée : $\text{cov}(t, u) = \langle t, u \rangle_D - \bar{t}\bar{u} = \sum_i p_i t_i u_i - \bar{t}\bar{u}$. \square

Preuve : Évidente. \square

Le coefficient de corrélation linéaire entre t et u , noté $r(t, u)$, est défini par

$$r(t, u) = \frac{\text{cov}(t, u)}{\sigma(t)\sigma(u)} = r(x, y).$$

P5 Le coefficient de corrélation linéaire s'interprète dans \mathbb{R}^n comme le cosinus du D -angle formé par les vecteurs D -centrés.

$$r(t, u) = \frac{\text{cov}(t, u)}{\sigma(t)\sigma(u)} = \frac{\langle x, y \rangle_D}{\|x\|_D \|y\|_D},$$

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 (\bar{t}, \bar{u}) , et d'équation $y = \hat{a}x$ dans le repère (x, y) . Le coefficient angulaire, \hat{a} , de Δ est solution du problème de la minimisation de la fonction φ de \mathbb{R} dans \mathbb{R} définie par

$$\varphi(a) = \sum_{i=1}^n p_i (y_i - ax_i)^2 = \|y - ax\|_D^2.$$

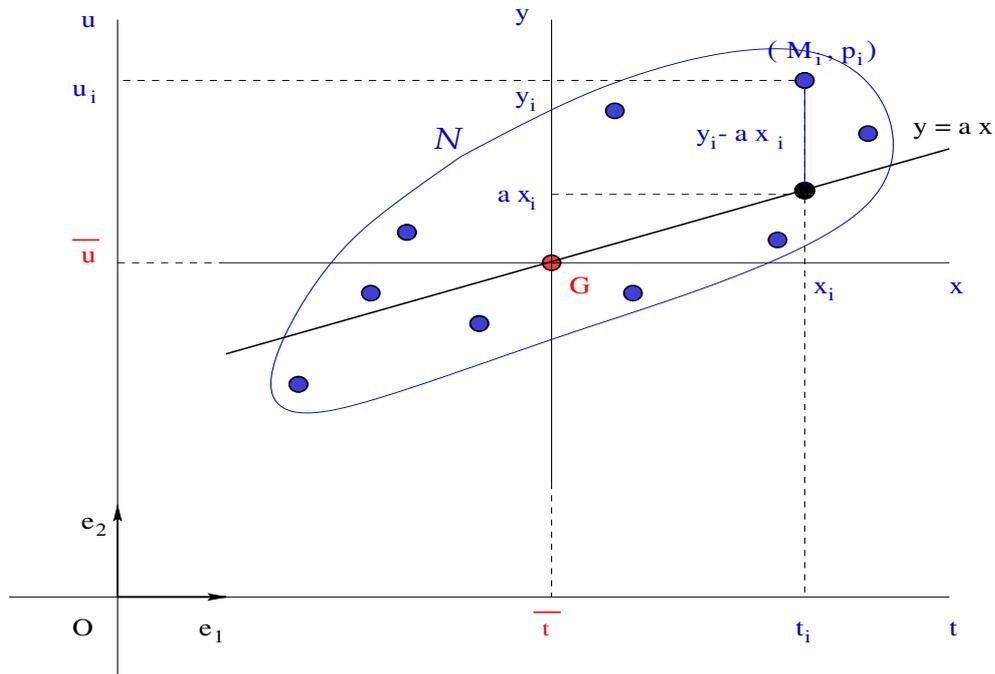


Figure 10 : Ajustement linéaire d'un nuage de points bivariés.

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 pour la valeur \hat{a} qui définit, dans

le repère (t, u) , la droite de régression ou d'ajustement linéaire Δ .

$$\text{Équation de } \Delta : \begin{cases} u = \hat{a} t + \hat{b} \\ \hat{a} = \frac{\text{cov}(t, u)}{\text{var}(t)} \\ \bar{u} = \hat{a} \bar{t} + \hat{b} \end{cases}$$

La valeur minimale de φ s'exprime en fonction de $r(t, u)$

$$\varphi(\hat{a}) = \text{var}(u)[1 - r^2(t, u)],$$

ce qui permet de mesurer grâce à $r(t, u)$ l'ajustabilité du nuage par une droite.

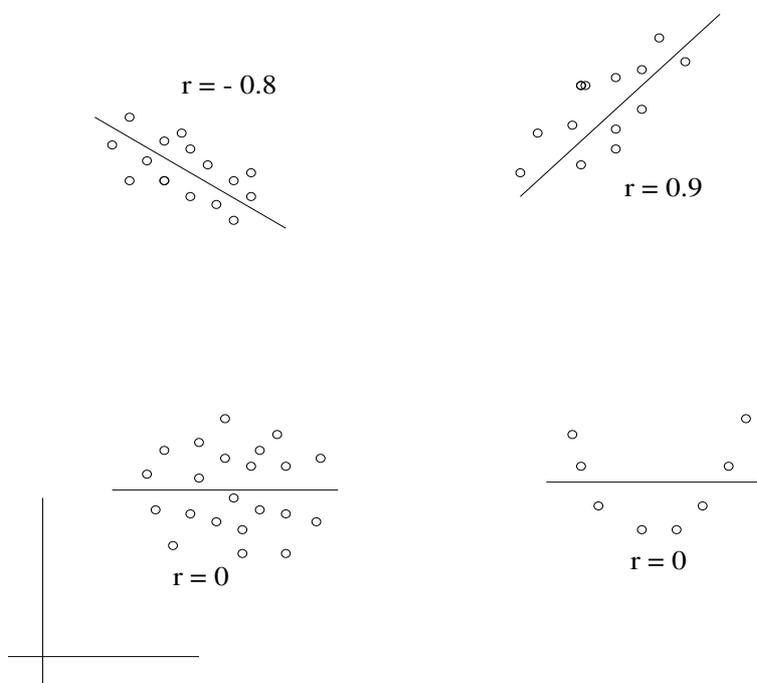


Figure 11 : Ajustabilité d'un nuage bivarié suivant les valeurs de r .

Remarquer que lorsque t et u sont non corrélés, $\text{cov}(t, u) = 0$, Δ a pour équation $u = \bar{u}$.

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 \bar{T} , formé des moyennes de p variables (colonnes de T).

$$\bar{T} = [\bar{t}_1, \dots, \bar{t}_p] = \mathbf{1}'_n DT.$$

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, $rang(\mathbb{V}) = rang(X)$. Elle est définie positive si $rang(X) = p$. Remarquer que $\mathbb{V}_{ii} = var(t^i)$. On appelle variance totale la somme des variances des p variables

$$\text{Variance totale} = trace(\mathbb{V}).$$

Standardiser les p variables, c'est à dire D -centrer réduire les variables sur l'échantillon, revient à construire

$$Z = XQ$$

où $Q = \text{diag}(\sigma_1^{-1}, \dots, \sigma_p^{-1})$. 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 $X = [X^1 | \dots | X^p]$ $n \times p$ deux matrices D -centrées. Soit $\hat{Y} = P_X Y$ la projection D -orthogonale de Y sur ImX . Puisque X est D -centrée, \hat{Y} est de moyenne nulle et $\|\hat{Y}\|_D^2 = var(\hat{Y})$. Le coefficient de détermination entre Y et ImX défini dans la section 4.2.2, s'écrit

$$R^2(Y, ImX) = \frac{var(\hat{Y})}{var(Y)}.$$

De plus,

$$R^2(Y, ImX) = \frac{Y'D\hat{Y}}{Y'DY} = \frac{cov(Y, \hat{Y})}{var(Y)} = r^2(Y, \hat{Y}).$$

Preuve : C'est la conséquence de la D -symétrie du projecteur P_X , c'est à dire $DP_X = (P_X)'D$. Alors, $\|\hat{Y}\|_D^2 = Y'(P_X)'DP_X Y = Y'DP_X P_X Y = Y'D\hat{Y}$. Il en résulte

$$R^2(Y, ImX) = \frac{var(\hat{Y})}{var(Y)} = \frac{cov(Y, \hat{Y})}{var(Y)} = \frac{cov^2(Y, \hat{Y})}{var(Y)var(\hat{Y})} = r^2(Y, \hat{Y}). \quad \square$$

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 $W = P_X Y$.

On utilise aussi

$$R(Y, ImX) = \sqrt{R^2(Y, ImX)} = r(Y, \hat{Y})$$

qui s'appelle le coefficient de corrélation multiple entre Y et ImX .

6.3 Exercices

Exercice 1 : 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 $\{p_i \mid i = 1, \dots, n\}$ stockés dans la diagonale de D .

L'objectif est de minimiser la fonction ϕ de \mathbb{R}^2 dans \mathbb{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 (\mathbb{R}^n, D) ?

3) Exprimer $\phi(\hat{\beta})$ en fonction de $r(t, u)$.

Exercice 2 : Dans le contexte de la propriété **P7** de la section 6.2.5, on se propose de montrer que

$$R^2(Y, \text{Im}X) = \max_{W \in \text{Im}X} r^2(Y, W),$$

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

$$\frac{Y'D\hat{Y}}{\|Y\|_D^2} = R^2(Y, \text{Im}X).$$

Vérifiez enfin que $W = \hat{Y}$ est le vecteur propre associé à cette valeur propre.

Chapitre 7

Généralisation de la Décomposition en Valeurs Singulières. Analyse en Composantes Principales du triplet (X, M, D)

La plupart des méthodes de l'Analyse Factorielle des 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 respectivement aux matrices $X'DXM_{p \times p}$ et $XX'D_{n \times n}$. 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. $XX'D$) est M -symétrique (resp. D -symétrique), ses r valeurs propres non-nulles sont réelles positives et ses vecteurs propres forment une base M -orthonormée de $Im X'$ (resp. D -orthonormée de $Im X$). \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 $\Lambda_r = \text{diag}(\lambda_1, \dots, \lambda_r)$ et $V = [V^1, \dots, V^r]$ les matrices des valeurs propres non-nulles et des vecteurs propres 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 $\text{rang}(V) = \text{rang}(X') = r$, il en résulte que les deux espaces vectoriels coïncident, $Im V = Im X'$. \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 $X \in \mathbb{R}^{n \times p}$ de rang r , M métrique sur \mathbb{R}^p et D métrique sur \mathbb{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 $\{\lambda_1, \dots, \lambda_r\}$ non-nulles de $XM'X'D$,
- $V_{p \times r}$, M -orthonormée ($V'MV = I_r$), dont les colonnes sont les vecteurs propres associés aux mêmes valeurs propres $\{\lambda_1, \dots, \lambda_r\}$ non-nulles de $X'DXM$,
- $\Lambda_r^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_r^{1/2})$ matrice diagonale des valeurs singulières $\{\lambda_i^{1/2}\}$ du triplet (X, M, D) ,

telles que X s'écrit

$$X = U\Lambda_r^{1/2}V' = \sum_{i=1}^r \sqrt{\lambda_i} U^i V^{i'}. \quad \square$$

Les valeurs singulières sont toujours classées par ordre décroissant

$$\lambda_1^{1/2} \geq \dots \geq \lambda_r^{1/2} > 0$$

ce qui induit un classement sur les colonnes de $U = [U^1, \dots, U^r]$ et de $V = [V^1, \dots, V^r]$.

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

$$U = XMV\Lambda_r^{-1/2}, \quad (*)$$

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 $XM'X'D$ 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 $Im V = Im X'$.

$$U\Lambda_r^{1/2}V' = XMV\Lambda_r^{-1/2}\Lambda_r^{1/2}V' = XMVV' = (\Pi_V^M X')' = (\Pi_{X'}^M X')' = (X')' = X. \quad \square$$

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

$$V = X'DU\Lambda_r^{-1/2}. \quad (**)$$

Corollaire : décomposition des matrices $\mathbb{V} = X'DX$ et $\mathbb{W} = XMX'$

$$\mathbb{V} = V\Lambda_r V' = \sum_{i=1}^r \lambda_i V^i V^{i'} \quad \text{et} \quad \mathbb{W} = U\Lambda_r U' = \sum_{i=1}^r \lambda_i U^i U^{i'}.$$

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 $V_x = M^{-1/2}V_z$. \square

Preuve : Exercice 1. \square

7.1.4 Projecteurs orthogonaux associés à la DVS

Les colonnes de $V = [V^1, \dots, V^r]$ forment une base M -orthonormée de $Im X'$, celles de $U = [U^1, \dots, U^r]$ 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 :

$$\Pi_{X'}^M = V(V'MV)^{-1}V'M = VV'M = \sum_{i=1}^r V^i V^{i'} M = \sum_{i=1}^r \Pi_{V^i}^M,$$

$$\Pi_X^D = U(U'DU)^{-1}U'D = UU'D = \sum_{i=1}^r U^i U^{i'} D = \sum_{i=1}^r \Pi_{U^i}^D.$$

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 $X \in \mathbb{R}^{n \times p}$ 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 $U_k = [U^1, \dots, U^k]$ et $V_k = [V^1, \dots, V^k]$ les matrices extraites de U et de V , et $\Lambda_k^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_k^{1/2})$ la matrice diagonale des k premières valeurs singulières.

On cherche un élément \widehat{X}_k de $E_k = \{X_k \in \mathbb{R}^{n \times p} \mid \text{rang}(X_k) = k\}$, le plus proche de X au sens de la norme $\|\cdot\|_{M \otimes D}$. 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

$$\|Z - Z_k\|_F^2 = \text{trace}((Z - Z_k)'(Z - Z_k)) = \text{trace}((X - X_k)' D (X - X_k) M) = \|X - X_k\|_{M \otimes D}^2.$$

L'équivalence de la DVS des triplets (X, M, D) et (Z, I_p, I_n) au sens de la Section 7.1.3, 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 7.1.3. 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 points 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

$$X = (I_n - \mathbf{1}_n \mathbf{1}'_n D)T.$$

Le point origine de l'espace des points lignes de X s'interprète comme le point ligne moyen, $\mathbf{1}'_n D T$, du tableau T . La matrice des covariances entre les variables est $\mathbb{V} = X' D X$, celle des produits scalaires entre les individus $\mathbb{W} = X M X'$. À ces matrices sont associés les opérateurs aux valeurs propres-vecteurs propres de la DVS du triplet.

Opérateurs en dualité et inertie du triplet (X, M, D)

- Opérateur des covariances : $\mathbb{V} M = X' D X M$
- Opérateur des produits scalaires entre individus : $\mathbb{W} D = X M X' D$
- Inertie totale du triplet : $\|X\|_{M \otimes D}^2 = \text{trace}(X M X' D) = \text{trace}(X' D X M)$

Expression tirée de la terminologie de la mécanique du point matériel, l'"inertie totale" des n points lignes M_i pesant chacun p_i

$$\sum_{i=1}^n p_i \|X_i\|_M^2 = \sum_{i=1}^n p_i X_i M X'_i = \text{trace}(X M X' D) = \|X'\|_{D \otimes M}^2 = \|X\|_{M \otimes D}^2,$$

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, $\text{trace}(X' D X) = \sum_{j=1}^p \mathbb{V}_j^j = \sum_{j=1}^p \sum_{i=1}^n p_i (X_i^j)^2 = \sum_{j=1}^p \sum_{i=1}^n p_i (T_i^j - \bar{T}^j)^2$ 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 :

$$\|X\|_{M \otimes D}^2 = \text{trace}(\Lambda_r) = \sum_{i=1}^r \lambda_i.$$

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

$$U_k = X M V_k \Lambda_k^{-1/2} (*) \quad \text{et} \quad V_k = X' D U_k \Lambda_k^{-1/2} (**).$$

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 $Im V_k$)

$$\widehat{X}_k = \Pi_{U_k}^D X \quad \text{et} \quad \widehat{X}_k' = \Pi_{V_k}^M X'. \quad \square$$

Preuve : Faisons la preuve pour les colonnes. La deuxième formule de transition (**)
donne

$$\begin{aligned} \Pi_{U_k}^D X &= U_k U_k' D X \\ &= U_k (X' D U_k)' \\ &= U_k \Lambda_k^{1/2} V_k'. \quad \square \end{aligned}$$

7.2.2 Principe fondamental de l'Analyse Factorielle

Ce principe est la justification de la projection du nuage des individus sur les axes factoriels $\{V^1, \dots, V^k\}$, classés par ordre décroissant des valeurs propres, $\lambda_1 \geq \dots \geq \lambda_k$.

Principe de l'Analyse Factorielle :

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 $V \in (\mathbb{R}^p, M)$ de longueur 1,

$$V^1 = \arg \max_{\|V\|_M^2=1} \text{trace}((\Pi_V^M X')' M (\Pi_V^M X') D).$$

La fonction objectif à maximiser qui est l'inertie des points lignes projetés sur V , s'écrit

$$\varphi(V) = \text{trace}((\Pi_V^M X')' M (\Pi_V^M X') D) = \text{trace}(X M V V' M X' D) = V' M \mathbb{V} M V.$$

Écrivons les équations aux dérivées partielles pour la fonction de Lagrange

$$L(V, \lambda) = \varphi(V) + \lambda(1 - V' M V),$$

$$\nabla_V L(V, \lambda) = \nabla_V \varphi(V) - \lambda \nabla_V (V' M V) = 2M \mathbb{V} M V - 2\lambda M V = 0.$$

ce qui donne $\lambda = \varphi(V)$ et $\mathbb{V} M V = \lambda V$. 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 $V^1' M V = 0$. La restriction de la fonction objectif à l'espace vectoriel $V^{1\perp} = \{V \in \mathbb{R}^p \mid V^1' M V = 0\}$, peut s'écrire $\varphi_{V^{1\perp}}(V) = V' M (\mathbb{V} - \lambda_1 V^1 V^1') M V$. 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 $(\mathbb{V} - \lambda_1 V^1 V^1') M V = \lambda V$. Le maximum est donc réalisé par le couple (V^2, λ_2) , λ_2 plus grande valeur propre de $(\mathbb{V} - \lambda_1 V^1 V^1') M$, 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.

ACP du triplet et "déflations" successives de X : Notons $X_{(0)} = X$, on appelle déflation de X à l'ordre j , $j = 1, \dots, k$, la matrice, $n \times p$, définie par récurrence,

$$X_{(j)} = X_{(j-1)} - \Pi_{U^j}^D X_{(j-1)} = (I_n - \Pi_{U^j}^D) X_{(j-1)}.$$

La matrice des covariances déflatée est notée $\mathbb{V}_{(j)} = X_{(j)}' D X_{(j)}$.

Proposition :

L'orthonormalité des axes factoriels $\{U^1, \dots, U^k\}$ a pour conséquences :

pour $j = 1, \dots, k$,

a)

$$\Pi_{U^j}^D X_{(j)} = \Pi_{U^j}^D X = \sqrt{\lambda_j} U^j V^{j'}$$

b)

$$X_{(j)} = X_{(j-1)} - \sqrt{\lambda_{j-1}} U^{j-1} V^{j-1'} = X - \sqrt{\lambda_1} U^1 V^{1'} - \dots - \sqrt{\lambda_k} U^k V^{k'} = X - \hat{X}_j.$$

c)

$$\mathbb{V}_{(j)} = \mathbb{V}_{(j-1)} - \lambda_j V^j V^{j'} = \mathbb{V} - \sum_{i=1}^j \lambda_i V^i V^{i'}$$

□

Preuve :

a) L'orthogonalité donne $\Pi_{U^j}^D X_{(j)} = \Pi_{U^j}^D (I_n - \Pi_{U^{j-1}}^D) X_{(j-1)} = \Pi_{U^j}^D X_{(j-1)} = \dots = \Pi_{U^j}^D X = \sqrt{\lambda_j} U^j V^{j'}$ grâce à la formule de transition (**).

b) De façon évidente, en ajoutant les matrices déflatées, $X_{(j)} = X - \sum_{i=1}^j \sqrt{\lambda_i} U^i V^{i'}$.

c) Évident. □

L'ACP d'ordre k du triplet (X, M, D) peut donc être considérée d'un double point de vue. D'abord de façon directe, par la DVS du triplet et le théorème d'Eckart-Young, comme de l'approximation de rang k de X . Ensuite de façon itérative, comme une suite de k "régressions partielles", $\Pi_{U^j}^D X_{(j-1)}$ étant la régression numéro j de la matrice déflatée $X_{(j-1)}$ sur la variable U^j . La matrice $X_{(j)}$, de rang $r - j$, est la matrice des résidus de la régression partielle numéro j . La dernière matrice des résidus, $X_{(k)}$, donne l'approximation de rang k de X par la relation

$$\hat{X}_k = U_k \Lambda_k^{1/2} V_k' = X - X_{(k)}.$$

7.2.3 L'ACP usuelle d'ordre k

L'ACP usuelle, dite réduite ou 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

$$\langle x, y \rangle_D = \frac{1}{n} y' x = 0 \iff y' x = \langle x, y \rangle = 0,$$

$$\Pi_X^D = \Pi_X^{\frac{1}{n} I_n} = \frac{1}{n} X (\frac{1}{n} X' X)^+ X' = X (X' X)^+ X' = \Pi_X^{I_n}.$$

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

$$\|X - \widehat{X}_k\|_{M \otimes D}^2 = \text{trace}[(X - \widehat{X}_k)' D (X - \widehat{X}_k) M] = \frac{1}{n} \|X - \widehat{X}_k\|_F^2 = \sum_{i=k+1}^r \lambda_i.$$

7.3 Représentations factorielles

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 $\{V_1, \dots, V_k\}$ et $\{U_1, \dots, U_k\}$ des deux espaces respectifs $\text{Im } X_k'$ et $\text{Im } X_k$.

Grâce aux possibilités 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 $\{U_1, U_2, U_3\}$.

7.3.1 Définitions

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

$$\%IR(k) = \frac{\|\widehat{X}_k\|_{M \otimes D}^2}{\|X\|_{M \otimes D}^2} \times 100 = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^r \lambda_i} \times 100.$$

Une cassure dans la croissance du diagramme des valeurs $\{\%IR(k) \mid k = 1, \dots, r\}$ 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.

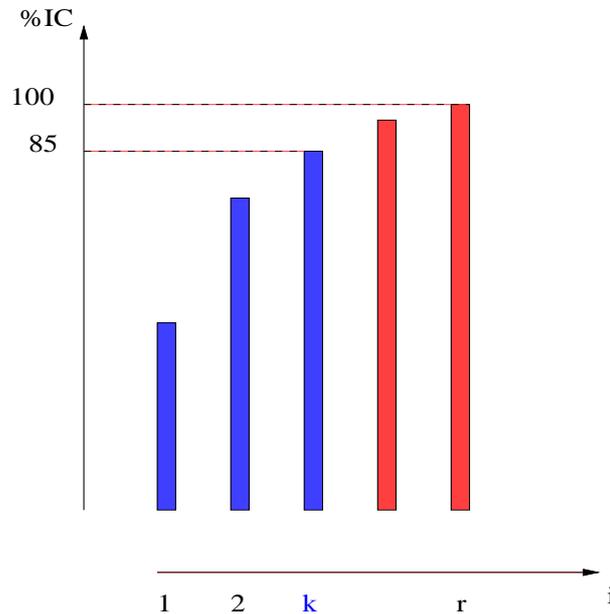


Figure 12 : diagramme des pourcentages d’inertie cumulée ou reconstituée.

Axes factoriels et principaux, composantes principales et facteurs principaux

– **Axes factoriels**

On note $V_k = [V^1, \dots, V^k]$ et $U_k = [U^1, \dots, U^k]$ les matrices dont les colonnes sont les k vecteurs de base retenus : $\{V^1, \dots, V^k\}$ vecteurs M -orthonormés dans l’espace des points lignes ($V_k' M V_k = I_k$) et $\{U^1, \dots, U^k\}$ 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 $\{A^j \in (\mathbb{R}^p, M) \mid j = 1, \dots, k\}$ définis par

$$A^j = \sqrt{\lambda_j} V^j$$

sont M -orthogonaux et de M -longueur $\sqrt{\lambda_j}$.

La matrice $A_k = [A^1, \dots, A^k]$, $p \times k$, est telle que

$$A_k = V_k \Lambda_k^{1/2} \quad \text{et} \quad A'_k M A_k = \Lambda_k.$$

La deuxième formule de transition (***) devient

$$A_k = X' D U_k. \quad (**)'$$

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 $\{C^j \in (\mathbb{R}^n, D) \mid j = 1, \dots, k\}$ définies par

$$C^j = \sqrt{\lambda_j} U^j$$

sont D -orthogonales et de D -longueur $\sqrt{\lambda_j}$.

La matrice $C_k = [C^1, \dots, C^k]$, $n \times k$, est telle que

$$C_k = U_k \Lambda_k^{1/2} \quad \text{et} \quad C'_k D C_k = \Lambda_k.$$

La première formule de transition (*) devient

$$C_k = X M V_k. \quad (*)'$$

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 $p \times k$

$$F_k = [F^1, \dots, F^k] = M V_k,$$

de telle sorte que (*)' s'écrive

$$C_k = X F_k. \quad (*)'$$

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, \dots, 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 .

Une composante principale voisine du 0 de \mathbb{R}^n fournit une relation de dépendance linéaire approchée entre les variables : $C^j = XF^j = \sum_{i=1}^p X^i F_i^j \approx 0$.

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 : $\mathbf{1}'_n D C^j = \mathbf{1}'_n D X F^j = 0$.

– **P2** : L'écart type de la variable latente \mathcal{C}^j est $\sqrt{\lambda_j}$.

Cela résulte de **P1**

$$\|C^j\|_D^2 = C^{j'} D C^j = \sum_{i=1}^n p_i (C_i^j)^2 = \text{var}(\mathcal{C}^j) = \lambda_j.$$

– **P3** : Les variables latentes sont deux à deux non corrélées

$$\text{cov}(\mathcal{C}^i, \mathcal{C}^j) = C^{j'} D C^i = 0, \quad \text{si } i \neq j.$$

7.3.2 Projections des points lignes

La matrice X étant D -centrée en colonnes, le nuage \mathcal{N} des n points lignes $\{M_1, \dots, M_n\}$ appartient à l'espace affine $\text{Im } X'$, 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

$$\Pi_{V^j}^M X'_l = V^j V^{j'} M X'_l = V^j (X_l M V^j)' = V^j C_l^j.$$

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 propriété **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^j étant donnés par les colonnes de la matrice $V^i C^{i'}$, l'inertie par rapport à O de ces n points s'exprime par la quantité

$$\text{trace}(V^j C^{j'} D C^j V^{j'} M) = \text{trace}(V^{j'} M V^j C^{j'} D C^j) = C^{j'} D C^j = \lambda_j.$$

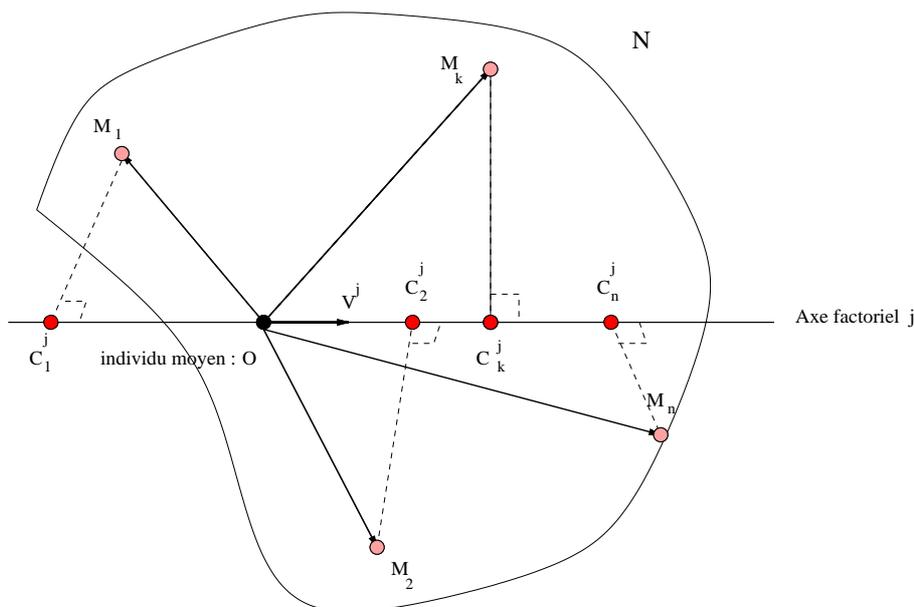


Figure 13 : Nuage \mathcal{N} des points lignes de (\mathbb{R}^p, M) et projections sur l'axe V^j .

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 $\Pi_{[V^i, V^j]}^M$, se décompose en la somme des projecteurs sur chacun des axes

$$\Pi_{[V^i, V^j]}^M = \Pi_{V^i}^M + \Pi_{V^j}^M.$$

Le vecteur \vec{OM}_l se projette sur le plan factoriel, appelé (i, j) pour simplifier, selon le vecteur \vec{Om}_l dont les coordonnées dans le repère (V^i, V^j) sont données par (C_l^i, C_l^j)

$$\Pi_{[V^i, V^j]}^M X'_l = V^i C_l^i + V^j C_l^j.$$

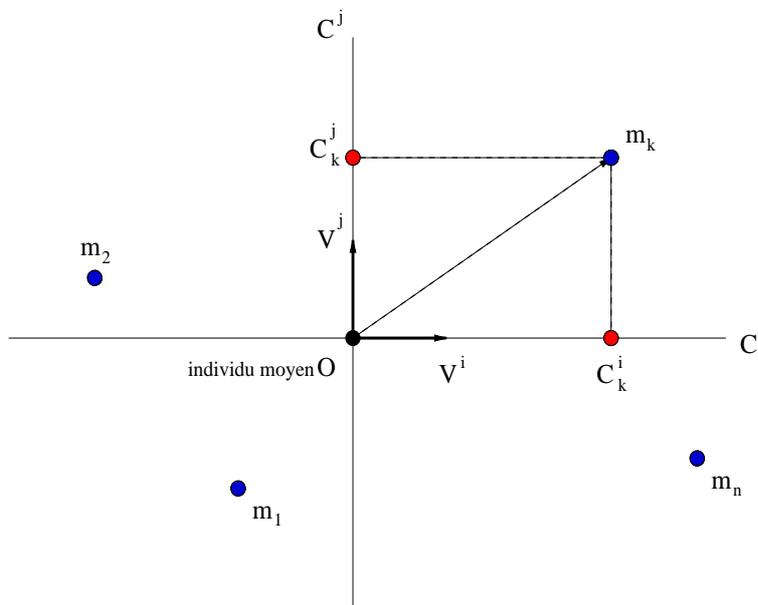


Figure 14 : Projection du nuage des individus sur le plan factoriel (V^i, V^j) .

La matrice $p \times n$ des projections du nuage des n points sur le plan (i, j) est donnée par

$$\Pi_{[V^i, V^j]}^M X' = V^i C'^i + V^j C'^j.$$

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)

$$trace[(\Pi_{[V^i, V^j]}^M X')'D(\Pi_{[V^i, V^j]}^M X')M] = C'^i D C^i + C'^j D C^j = \lambda_i + \lambda_j,$$

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 $C^{j'}DC^j = \lambda_j$; alors

$$1 = \sum_{l=1}^n \frac{p_l}{\lambda_j} (C_l^j)^2.$$

La contribution absolue de l'individu l à la dispersion du nuage projeté sur l'axe j est

$$CTA_l^j = \frac{p_l}{\lambda_j} (C_l^j)^2.$$

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 (\mathbb{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 \vec{Om}_l et \vec{OM}_l

$$\cos^2 \theta_l = \frac{\|\vec{Om}_l\|_M^2}{\|\vec{OM}_l\|_M^2} = \frac{\|\vec{Om}_l\|_M^2}{X_l M X_l'} = \frac{\|\vec{Om}_l\|_M^2}{\mathbb{W}_l^l} = \frac{\|\vec{Om}_l\|_M^2}{\sum_{j=1}^r (C_l^j)^2}.$$

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 $\vec{Om}_l^{i,j} = \vec{Om}_l^i + \vec{Om}_l^j$ 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 obtient 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, $\sum_{i=1}^r \cos^2 \theta_l^i = 1$.

7.3.3 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 coefficient 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

$$\Pi_{U^i}^D = U^i U^{i'} D \quad \text{et} \quad \Pi_{[U^i, U^j]}^D = \Pi_{U^i}^D + \Pi_{U^j}^D.$$

La mesure algébrique de la projection d'un vecteur-variable x^l sur un vecteur de base U^i est égale à A_l^i . En effet, la formule de transition (**)' donne

$$\Pi_{U^i}^D X^l = U^i A_l^i \quad \text{et} \quad \Pi_{[U^i, U^j]}^D X^l = U^i A_l^i + U^j A_l^j.$$

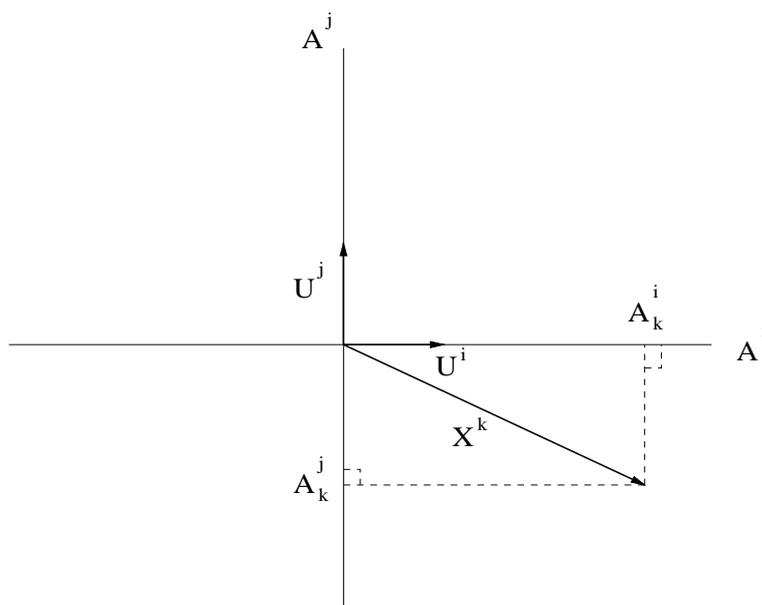


Figure 15 : Projection d'une variable x^k sur le plan factoriel (U^i, U^j) .

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 C^j et une variable x^i

Une composante C^j s'interprète comme le vecteur des coordonnées des projections du nuage des individus sur l'axe factoriel V^j mais aussi comme l'expression d'une variable latente C^j dans la base canonique de (\mathbb{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 C^j et les variables $x^i, i = 1, \dots, p$,

$$r(C^j, x^i) = \frac{\text{cov}(C^j, x^i)}{\sigma(C^j)\sigma(x^i)} = \frac{X^{i'}DC^j}{\sqrt{\lambda_j}\|X^i\|_D} = \frac{X^{i'}DXMV^j}{\sqrt{\lambda_j}\|X^i\|_D} = \frac{\lambda_j V_i^j}{\sqrt{\lambda_j}\|X^i\|_D} = \frac{A_i^j}{\|X^i\|_D},$$

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

$$r(C^j, x^i) = A_i^j.$$

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.

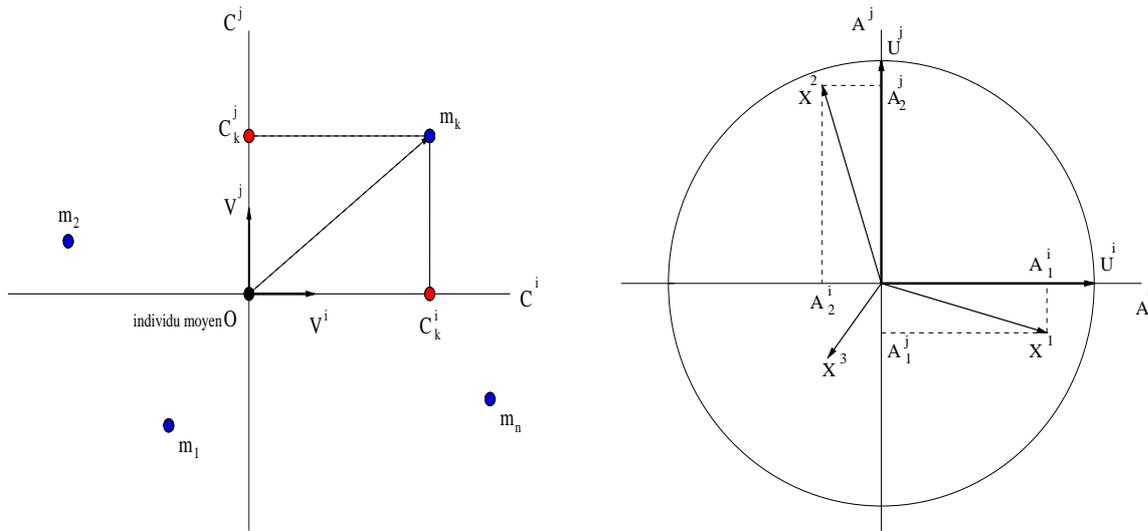


Figure 16 : Visualisation duale des plans factoriels (i, j) pour les individus et les variables dans l'ACP usuelle normée. Cercle des corrélations.

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 $M = \text{diag}(m_1, \dots, m_p)$: $M = I_p$ (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 $A^{j' } M A^j = \lambda_j$; 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

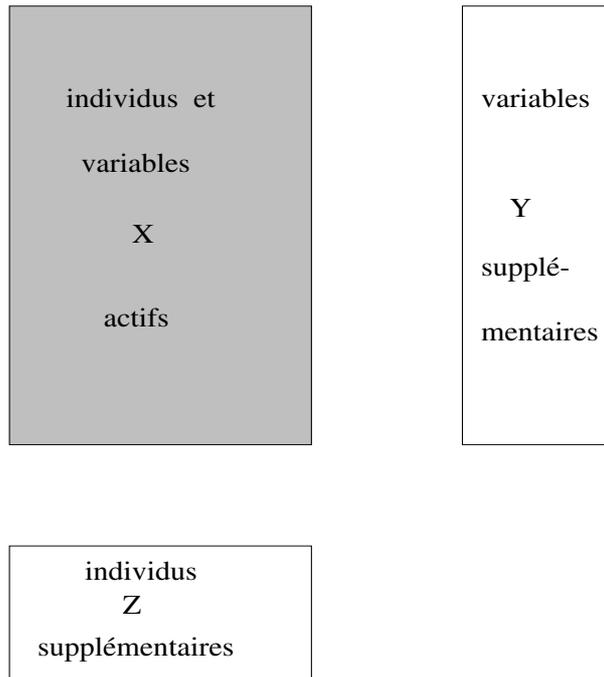
$$CTR_k^i = \cos^2 \theta_k^i = \frac{(A_k^i)^2}{\sum_{j=1}^r (A_k^j)^2} \text{ ou } CTR_k^{i,j} = \cos^2 \theta_k^{i,j} = \frac{(A_k^i)^2 + (A_k^j)^2}{\sum_{j=1}^r (A_k^j)^2} = \cos^2 \theta_k^i + \cos^2 \theta_k^j.$$

En effet, la contribution relative de l'axe i à la variable k est le carré du coefficient de corrélation entre x^k et 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 (C^i, 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.

7.3.4 Éléments 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 peut 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 tableau 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

$$A_k^j = \frac{Y^{k'}DC^j}{\sqrt{\lambda_j}} = \sum_{i=1}^n \frac{p_i}{\sqrt{\lambda_j}} Y_i^k C_i^j.$$

On peut représenter la variable supplémentaire 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 tableau 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.

7.4 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 $(Z = D^{1/2}XM^{1/2}, I_p, I_n)$.

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 $\Lambda_r^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})$ avec $\lambda_1 \geq \dots \geq \lambda_r > 0$. 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 $\Lambda_k^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_k})$. \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 Frobenius $\|\cdot\|_F$, ($\|X\|_F^2 = \text{trace}(X'X)$) est invariante par transformation orthogonale, montrer que'il existe deux matrices 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 $\varphi(B) = \text{trace}(B(B'B)^{-1} B' \Lambda)$.

4. Utiliser les propriétés de $\Pi_B = B(B'B)^{-1} B'$ pour montrer que

$$\varphi(B) = \|\Pi_B \Lambda\|_F^2.$$

Écrire 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 $i = 1, \dots, n$ et que $\sum_{i=1}^n a_i = \text{rang}(B) = k$. 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

$$\hat{Z}_k = \arg \min_{Z_k \in \hat{\mathcal{E}}_k} \|Z - Z_k\|_F^2.$$

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. Étude 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 statistiques 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 :

$$v_1 = \frac{1}{2} \begin{bmatrix} \sqrt{2} \\ 1 \\ -1 \end{bmatrix} \quad \text{et} \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

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 supplémentaire $(\sqrt{10}, 2\sqrt{10}, 2\sqrt{10})$.
-

Exercice 6 : ACP usuelle normée et ACP du triplet $(X, \text{diag}(\sigma_1^{-2}, \dots, \sigma_p^{-2}), n^{-1}I_n)$

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 $Y = X \text{diag}(\sigma_1^{-1}, \dots, \sigma_p^{-1})$ 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 $\mathbb{V}_Y M = Y' D Y M$ et $\mathbb{W}_Y D = Y M Y' D$
 - 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 $(X, M = \text{diag}(\sigma_1^{-2}, \dots, \sigma_p^{-2}), D = n^{-1}I_n)$:
 - a. Ecrire l'expression des opérateurs $\mathbb{V}_X M = X' D X M$ et $\mathbb{W}_X D = X M X' D$
 - 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

	Math	Phys	Techn
moy	12.2	12	12.1
var	6.76	4	4.49

Matrice des covariances

	Math	Phys	Techn
Math	6.76	2.9	3.98
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 à l'interprétation 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

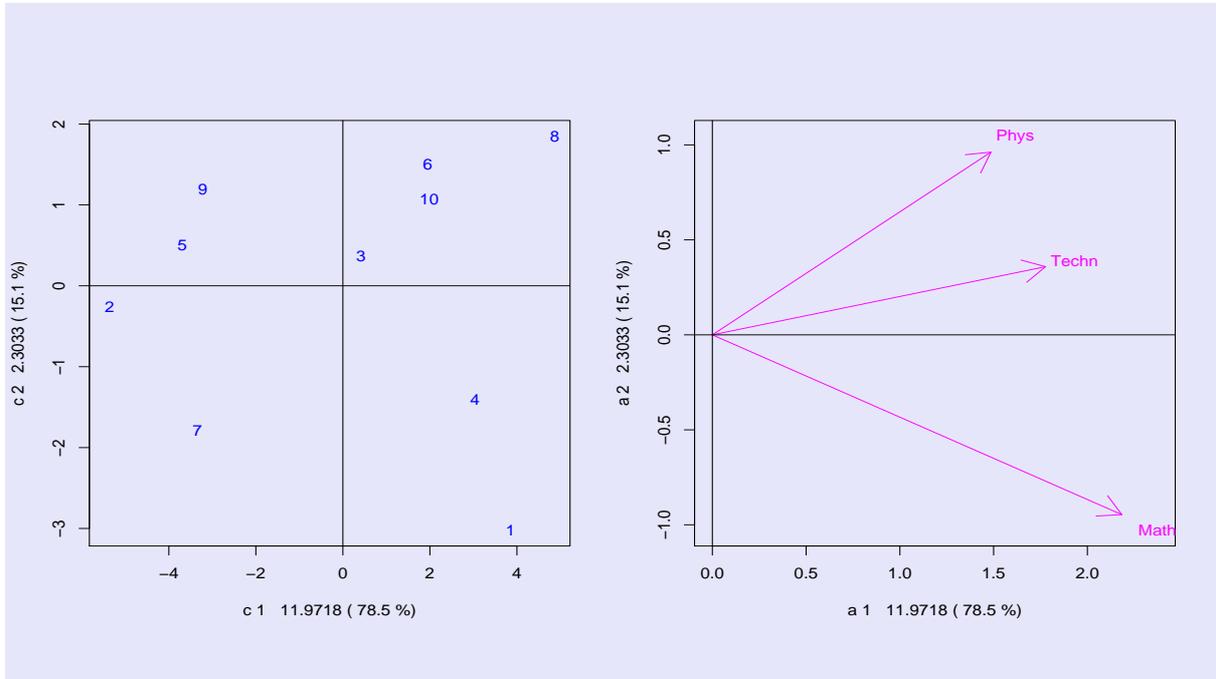


Figure 17 : plans factoriels (1, 2) pour les individus et les variables de l'ACP centrée d'un ensemble de notes.

Exercice 8 : L'Analyse Factorielle Discriminante (AFD)

Notation : Pour une matrice A , l'élément courant sera noté A_i^j , une colonne A^j , une ligne A_i , la matrice transposée A' .

Soit $X \in \mathbb{R}^{n \times p}$ matrice, de rang p , d'un échantillon de n mesures sur p variables quantitatives centrées au sens des poids statistiques stockés dans la matrice diagonale $D = \text{diag}(p_1, \dots, p_n)$. La matrice des covariances est notée \mathbb{V} .

On dispose en outre d'une partition des n individus en q groupes : $\{1, \dots, n\} = I_1 \cup \dots \cup I_q$ où $I_i = \{k \in \{1, \dots, n\} \mid k \in \text{groupe } i\}$. Soit y le vecteur à valeurs entières dont la coordonnée $y_k = i$ signifie que $k \in I_i$ et soit $\Delta = \text{diag}(P_1, \dots, P_q)$ 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 matrice $n \times q$, du codage booléen caractérisant l'appartenance des n individus aux q groupes : $Y = [Y_i^j]$, $Y_i^j = 1$ si i appartient à I_j , $Y_i^j = 0$ si non.

L'objectif de l'Analyse Factorielle Discriminante est de **déterminer des axes factoriels de \mathbb{R}^p sur lesquels, en projection, les individus moyens des groupes munis chacun du poids du groupe correspondant, ont une inertie maximale.** La métrique Euclidienne utilisée sur \mathbb{R}^p est provisoirement notée M .

1. Propriétés de la matrice G , $q \times p$, des individus moyens (centres de gravité) des groupes :

a. Sur le petit exemple suivant

$$X = \begin{pmatrix} 1 & 2 \\ -1 & -2 \\ 0 & 1 \\ 2 & 0 \\ -2 & -1 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \\ 2 \end{pmatrix}, \quad D = \text{diag}(1/5, 1/5, 1/5, 1/5, 1/5),$$

calculer Y et vérifier que $Y'DY = \Delta$. Dire pourquoi Δ est une métrique sur \mathbb{R}^q et \mathbb{V}^{-1} une métrique sur \mathbb{R}^p ? L'individu moyen du groupe i , la i ème ligne de G , est défini par $G_i = \frac{1}{P_i} \sum_{j \in I_i} p_j X_j$. Vérifier sur l'exemple que $G = \Delta^{-1} Y' D X$.

b. Montrer que G est Δ -centrée. On note $\hat{X} = \prod_Y^D X$, où \prod_Y^D est la matrice de projection D -orthogonale sur $Im Y$. Calculer \hat{X} en fonction de Y et de G et en déduire l'interprétation de la ligne courante \hat{X}_i . Soit $\mathbb{B} = G' \Delta G$ la matrice $p \times p$ des covariances entre les groupes ("Between"). Dans quel cas n'est-elle sûrement pas inversible? Montrer que $\mathbb{B} = X' D \hat{X} = \hat{X}' D \hat{X}$.

c. On appelle matrice des covariances à l'intérieur des groupes ("Within") $\mathbb{W} = (X - \hat{X})' D (X - \hat{X})$. Démontrer le théorème dit de la "décomposition de la variance totale" :

$$\mathbb{V} = \mathbb{B} + \mathbb{W}.$$

2. Recherche du meilleur axe factoriel discriminant v^1 :

a. Soit $v \in (\mathbb{R}^p, M)$ unitaire, $v' M v = 1$, et \mathcal{D}_v la droite vectorielle définie par v . On sait que l'inertie par rapport à l'origine (l'individu moyen) du nuage des n individus de X projetés M -orthogonalement sur \mathcal{D}_v , est donnée par $I_V(v) = v' M \mathbb{V} M v$.

a1. On appelle variable discriminante l'élément de (\mathbb{R}^n, D) défini par $d = X M v$ et $f = M v$ le facteur discriminant de l'axe v . Donner deux interprétations de d , par rapport aux variables et respectivement aux individus de X . Quelle est l'interprétation de $I_V(v)$ en fonction de d ?

a2. Utiliser la décomposition de la variance totale pour donner l'expression de $I_B(v)$ et $I_W(v)$ de telle sorte que $I_V(v)$ se décompose en fonction de $I_B(v)$, inertie "entre", inertie des individus moyens projetés munis des poids de leurs groupes, et de $I_W(v)$, inertie "dans" les groupes projetés, somme des inerties de chaque classe autour de son individu moyen projeté (théorème d'Huygens).

b. On considère les problèmes d'optimisation suivants :

$$(1) \quad \max_{v' M v = 1} I_B(v), \quad (2) \quad \max_v \frac{I_B(v)}{I_W(v)}, \quad (3) \quad \max_v \frac{I_B(v)}{I_V(v)}.$$

La valeur de la fonction objectif du problème (i), notée $\varphi_i(v)$, mesure le "pouvoir discriminant" de l'axe v , c'est à dire, quantifie sa capacité à séparer les groupes projetés sur v . L'optimum de (i) est réalisé par un vecteur noté $\hat{v}^{(i)}$.

b1. Equivalence entre (2) et (3) : En utilisant l'identité $\frac{a}{b} = \frac{1}{1-a/(a+b)} - 1$, montrer l'équivalence entre (2) et (3). On exprimera la relation qui lie les valeurs des deux fonctions objectifs $\varphi_2(v)$ et $\varphi_3(v)$. Montrer que ces fonctions sont homogènes de degré 0. Donner les bornes de leurs intervalles de variation. Quels types de configurations géométriques du nuage de points et de l'axe v correspondent à ces valeurs extrêmes? On choisira de résoudre le problème (3) en exprimant le système d'équations aux dérivées partielles en fonction du facteur discriminant f sous la forme d'un problème aux valeurs propres.

b2. Résolution de (1) : Résoudre (1) par la méthode des multiplicateurs de Lagrange et montrer que les trois problèmes sont équivalents en prenant $M = \mathbb{V}^{-1}$.

3. On choisit maintenant $M = \mathbb{V}^{-1}$ et on note v^1 le meilleur axe discriminant solution des trois problèmes équivalents de la question précédente. L'objectif est de chercher un meilleur second, v^2 , \mathbb{V}^{-1} -orthogonal au premier, puis un troisième orthogonal aux précédents, etc ... On va vérifier comment les ACPs d'ordre k de deux triplets conduisent de deux manières différentes, au même problème aux valeurs propres que celui de la question 2. pour ainsi réaliser l'objectif fixé.

a. ACP des individus moyens : le triplet $(G, \mathbb{V}^{-1}, \Delta)$.

b. ACP des "variables instrumentales" X : le triplet $(X, \mathbb{V}^{-1} \mathbb{B} \mathbb{V}^{-1}, D)$.

Pour chacun des triplets écrire l'expression de "l'opérateur des covariances" en fonction de \mathbb{B} et de \mathbb{V} . Soient $\{V^1, \dots, V^k\}$ les k premiers vecteurs propres (axes factoriels) de cette matrice et $\{C^1, \dots, C^k\}$ les composantes principales. Comment peut-on faire le lien entre V^1 , le premier axe discriminant v^1 et le facteur discriminant f^1 ? Comment retrouver la variable discriminante d^1 .

Pour le triplet b. on écrira en outre "l'opérateur des produits scalaires entre individus" et l'on vérifiera qu'il est identique à celui du triplet $(\hat{Y}, \Delta^{-1}, D)$ où $\hat{Y} = \prod_X^D Y$ est la projection D -orthogonale de Y sur l'espace vectoriel des colonnes de X . Qu'en déduire quant-aux représentations factorielles des individus pour ces deux derniers triplets.

7.5 Formulaires

7.5.1 ACP usuelle d'ordre k du triplet $(X, M = I_p, D = n^{-1}I_n)$

$T; n \times p$	n individus (actifs), p variables <u>quantitatives</u> (actives)
$X = (I_n - n^{-1}\mathbf{1}\mathbf{1}')T$	tableau des variables centrées
$\mathbb{V} = n^{-1}X'X$	matrice des covariances (des corrélations si variables réduites)
$\mathbb{V} V^i = \lambda_i V^i$	
$\lambda_1 \geq \dots \geq \lambda_k$	$\sum_1^p \lambda_i = \text{trace}(\mathbb{V}) = \text{Variance totale} = \text{Inertie}$
$\{V^1, \dots, V^k\}$	Facteurs principaux, $V^i \in \mathbb{R}^p$, $\ V^i\ _2 = 1$, $V^{i'}V^j = 0$
$\{A^1, \dots, A^k\}$	Axes principaux, $A^i = \sqrt{\lambda_i}V^i$, $\ A^i\ _2 = \sqrt{\lambda_i}$, $A^{i'}A^j = 0$
$\{C^1, \dots, C^k\}$	Composantes principales, $C^i \in \mathbb{R}^n$ $C^i = XV^i = X^1V_1^i + \dots + X^pV_p^i$ $\text{moy}(C^i) = 0$; $\text{var}(C^i) = \lambda_i$; $\text{cov}(C^i, C^j) = 0$, $i \neq j$ C^i vecteur propre de $n^{-1}XX'$ associé à λ_i ; $\mathbb{W} = XX'$ $r(C^j, X^i) = A_i^j / \sigma(X^i)$
$\hat{X}_k = \sum_{i=1}^k \frac{1}{\sqrt{\lambda_i}} C^i A^{i'}$	Approximation de rang k de X ; $\ X - \hat{X}_k\ _F^2 = n \sum_{i=k+1}^p \lambda_i$
Contributions absolues	de l'individu k à 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 à 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 à l'individu k : $CTR_k^i = (C_k^i)^2 / \mathbb{W}_k^i$, $\sum_{i=1}^p CTR_k^i = 1$ de l'axe i à la variable k : $CTR_k^i = (A_k^i)^2 / \mathbb{V}_k^i$, $\sum_{i=1}^p CTR_k^i = 1$
Formules de transition	$C^i = XA^i / \sqrt{\lambda_i}$; $A^i = X'C^i / (n\sqrt{\lambda_i})$

7.5.2 DVS du triplet (X, M, D)

DVS du triplet $(X_{n \times p}, M_{p \times p}, D_{n \times n})$; $\text{rang}(X) = r$		
$X = U\Lambda_r^{1/2}V' = C\Lambda_r^{-1/2}A'$; $\Lambda_r = \text{diag}(\lambda_1, \dots, \lambda_r)$; $\lambda_1 \geq \dots \geq \lambda_r > 0$		
Opérateurs	<div style="border: 1px solid black; display: inline-block; padding: 2px; margin-bottom: 5px;">$\mathbb{V}M = X'DXM$</div> $X'DXMV = V\Lambda_r$ $X'DXMA = A\Lambda_r$ Inertie totale $\ X\ _{M \otimes D}^2 = \text{trace}(\mathbb{V}M) = \text{trace}(\mathbb{W}D) = \sum_i^r \lambda_i$	<div style="border: 1px solid black; display: inline-block; padding: 2px; margin-bottom: 5px;">$\mathbb{W}D = XMX'D$</div> $XMX'DU = U\Lambda_r$ $XMX'DC = C\Lambda_r$
Espaces	Axes Principaux	Composantes Principales
<div style="border: 1px solid black; display: inline-block; padding: 2px; margin-bottom: 5px;">Individus</div> Bases $M \perp$ de $\text{Im}(X') \subset (\mathbb{R}^p, M)$ $V = [V^1, \dots, V^r]$, $V'MV = I_r$ $A = [A^1, \dots, A^r]$, $A'MA = \Lambda_r$ $\ V^i\ _M = 1$, $\ A^i\ _M = \sqrt{\lambda_i}$	$A^i = \sqrt{\lambda_i} V^i$ $A = V\Lambda_r^{1/2}$ (**)' $A = X'DU = X'DC\Lambda_r^{-1/2}$ Facteurs Principaux $F = MV$ $\mathbb{V} = AA'$	$C^i = \sqrt{\lambda_i} U^i$ $C = U\Lambda_r^{1/2}$ (*)' $C = XMV = XMA\Lambda_r^{-1/2} = XF$ $\mathbb{W} = CC'$
<div style="border: 1px solid black; display: inline-block; padding: 2px; margin-bottom: 5px;">Variables</div> Bases $D \perp$ de $\text{Im}(X) \subset (\mathbb{R}^n, D)$ $U = [U^1, \dots, U^r]$, $U'DU = I_r$ $C = [C^1, \dots, C^r]$, $C'DC = \Lambda_r$ $\ U^i\ _D = 1$, $\ C^i\ _D = \sqrt{\lambda_i}$		

Chapitre 8

Traitement d'enquêtes, Analyse Factorielle des Correspondances Simples et Multiples

L'ACP généralisée d'un triplet (X, M, D) étudiée au chapitre précédent, n'a jusqu'à présent puisé ses applications que dans le cas de variables quantitatives traitées par une ACP usuelle. Ce chapitre présente deux nouvelles méthodes, l'Analyse Factorielle des Correspondances (AFC) dites simples par opposition à la suivante qui est l'Analyse Factorielle des Correspondances Multiples (AFCM). Toutes deux sont utilisées dans le traitement de variables qualitatives issues du dépouillement des questions d'une enquête statistique.

L'AFC permet d'analyser le comportement de la population confrontée à deux questions seulement alors que l'AFCM est dédiée au traitement de deux questions et plus.

8.1 Variables d'une enquête, codage

Une enquête portant sur une population de N individus consiste pour un individu i , à choisir une réponse à chaque question. Le fait que les réponses possibles soient imposées aux personnes interrogées, est caractéristique d'une enquête à questions "fermées".

8.1.1 Variables qualitatives

En langage statistique, on parle de **variable qualitative** (question) à m **modalités** (m réponses possibles).

Remarque : un individu ne peut choisir qu'une seule réponse à chaque question

La contrainte précédente rend possible le codage booléen de l'appartenance ou non d'un individu à telle modalité de telle variable qualitative. Prenons l'exemple du dépouillement des réponses des six premiers individus pour la variable qualitative "couleur des yeux" à quatre modalités, "noir", "noisette", "vert" ou "bleu".

Individus	"couleur des yeux"	Codage
1	"noisette"	0 1 0 0
2	"noir"	1 0 0 0
3	"bleu"	0 0 0 1
4	"vert"	0 0 1 0
5	"noisette"	0 1 0 0
6	"bleu"	0 0 0 1
.....		

Il est clair que l'on a imposé un choix à chaque personne interrogée qui doit se déterminer parmi les 4 réponses possibles.

Remarque : variable quantitative \longrightarrow variable qualitative

On est souvent en présence d'un jeu de données comportant des variables qualitatives et des variables quantitatives. Une possibilité consiste à transformer le quantitatif en qualitatif. On effectue une partition de l'étendue d'une variable quantitative en m intervalles grâce au choix de $m - 1$ "valeurs seuils", construisant ainsi une variable qualitative à m modalités caractérisant chacune l'intervalle considéré. Souvent, on est amené à choisir 3 modalités "faible", "moyen" et "fort".

8.1.2 Indicatrice des modalités

La matrice de **codage disjonctif complet** d'une variable qualitative à N observations et m modalités est la matrice $N \times m$

$$U = [U_{ij}],$$

où $U_{ij} = 1$ si le i ème individu est dans la modalité j , $U_{ij} = 0$ sinon. Les N individus de la population sont munis de poids statistiques égaux $(1/N)$. On note $D = (1/N)I_N$ la métrique de l'espace \mathbb{R}^N des colonnes de U .

Propriétés de U

- **Marge ligne** : la marge ligne est le vecteur colonne somme des colonnes de U . Elle vaut

$$U\mathbf{1}_m = \mathbf{1}_N.$$

L'information contenue dans U est de fait contenue dans $m - 1$ colonnes. Il suffit en effet de connaître $m - 1$ colonnes pour calculer la colonne restante, d'où le qualificatif de "complet" pour le codage à m colonnes.

- **Marge colonne de U** : la marge colonne de U est le vecteur ligne

$$\mathbf{1}_N'U = [U_{.1}, \dots, U_{.m}]$$

des effectifs des modalités de la variable.

- Soit U^c la matrice des colonnes D -centrées, U^c est de rang plus petit ou égal à $m - 1$ puisque les colonnes de U^c somment à 0,

$$U^c\mathbf{1}_m = 0_N. \quad \square$$

Preuve : Le fait que $U\mathbf{1}_m = \mathbf{1}_N$ donne $U^c\mathbf{1}_m = (U - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N'U)\mathbf{1}_m = 0_N. \square$

- $U'U = \text{diag}(U_{.1}, \dots, U_{.m})$, car les colonnes de U sont deux à deux orthogonales.

8.2 Table de contingence, liaison entre deux variables qualitatives

On s'intéresse maintenant au tableau croisé ventilant une population selon les modalités de deux caractères qualitatifs. L'individu va disparaître en se fondant dans l'effectif correspondant au croisement d'une modalité de la première variable et d'une modalité de la seconde.

8.2.1 Définitions et notations

Soient deux variables qualitatives notées \mathcal{L} (pour Ligne) et \mathcal{C} (pour Colonne) ayant respectivement I et J modalités et des matrices de codage $N \times I$ et $N \times J$ notées $U_{\mathcal{L}}$ et $U_{\mathcal{C}}$. La **table de contingence** T croisant les effectifs des modalités de \mathcal{L} et \mathcal{C} est la matrice

$$T = [n_{ij}] = U_{\mathcal{L}}'U_{\mathcal{C}}.$$

L'élément n_{ij} est l'effectif des individus dans la modalité i de \mathcal{L} et dans la modalité j de \mathcal{C} .

$$N = \sum_{i=1}^I \sum_{j=1}^J n_{ij}.$$

On appelle **tableau des fréquences** associé à T , la matrice $I \times J$

$$P = \frac{1}{N}T,$$

d'élément courant $p_{ij} = n_{ij}/N$.

$$1 = \sum_{i=1}^I \sum_{j=1}^J p_{ij}.$$

marges ligne et marges colonnes :

On appelle marge ligne de P (resp. de T) le vecteur colonne $P\mathbf{1}_J$ (resp. $T\mathbf{1}_J$), dont l'élément courant i

$$p_{i.} = \sum_{j=1}^J p_{ij}$$

(resp. $n_{i.}$), est la fréquence (resp. l'effectif) des individus qui sont dans la modalité i de \mathcal{L} . On a $p_{i.} = n_{i.}/N$.

On appelle marge colonne de P (resp. de T) le vecteur ligne $\mathbf{1}_I'P$ (resp. $\mathbf{1}_I'T$), dont l'élément courant j

$$p_{.j} = \sum_{i=1}^I p_{ij}$$

(resp. $n_{.j}$), est la fréquence (resp. l'effectif) des individus qui sont dans la modalité j de \mathcal{C} . On a $p_{.j} = n_{.j}/N$.

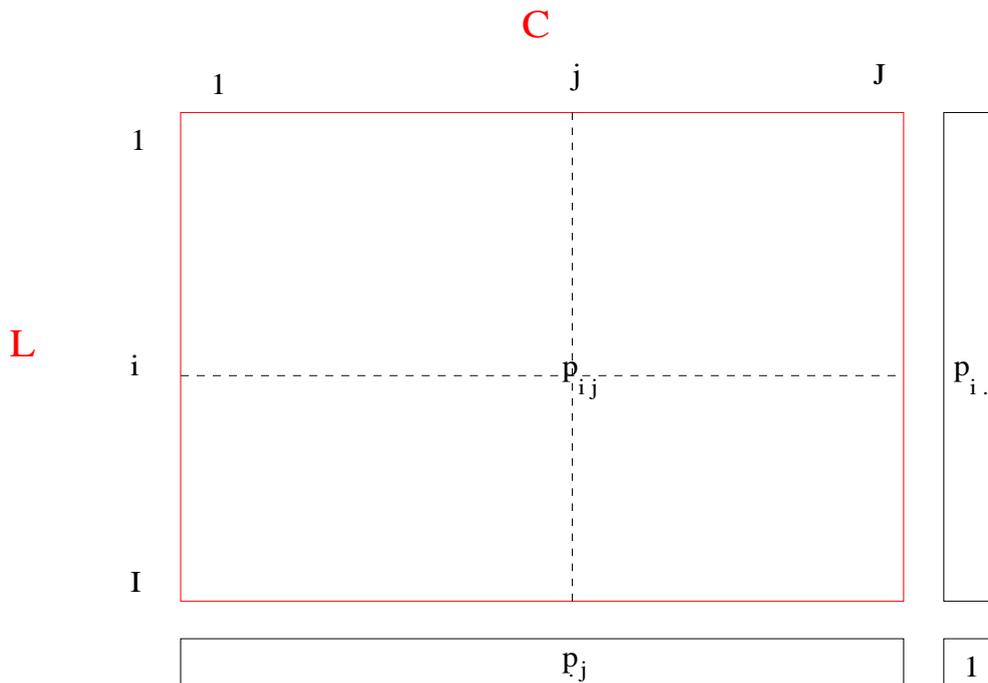


Figure 18 : Tableau P des fréquences, marge ligne et marge colonne de P .

Métriques associées à P :

L'espace \mathbb{R}^I des colonnes de P est muni de la métrique diagonale des poids statistiques de la marge ligne,

$$D_I = \text{diag}(p_{1.}, \dots, p_{I.}).$$

L'espace \mathbb{R}^J des lignes de P est muni de la métrique diagonale des poids statistiques de la marge colonne,

$$D_J = \text{diag}(p_{.1}, \dots, p_{.J}).$$

Propriétés des métriques

$$\mathbf{M1} : 1 = \mathbf{1}_I' D_I \mathbf{1}_I = \mathbf{1}_J' D_J \mathbf{1}_J = \mathbf{1}_I' P \mathbf{1}_J \text{ ou encore } 1 = \sum_{i=1}^I p_{i.} = \sum_{j=1}^J p_{.j} = \sum_{i=1}^I \sum_{j=1}^J p_{ij}.$$

$$\mathbf{M2} : \text{La marge ligne s'écrit } P \mathbf{1}_J = D_I \mathbf{1}_I \text{ et la marge colonne } \mathbf{1}_I' P = \mathbf{1}_J' D_J.$$

Les propriétés précédentes vont jouer un rôle clé dans l'Analyse Factorielle des Correspondances du tableau de contingence T . En fait, les métriques diagonales formées par les marges de P fournissent les poids statistiques au nuage des points lignes et au nuage des points colonnes représentés dans les plans factoriels de l'AFC.

8.2.2 Indépendance de deux variables qualitatives \mathcal{L} et \mathcal{C}

Soient \mathcal{L} et \mathcal{C} deux variables qualitatives et P leur tableau $I \times J$ des fréquences. \mathcal{L} et \mathcal{C} sont empiriquement indépendantes si et seulement si

$$p_{ij} = p_{i.}p_{.j} \quad \forall i = 1, \dots, I, \quad \forall j = 1, \dots, J.$$

Soit Π le tableau de l'indépendance construit à partir des marges de P

$$\Pi = [\pi_{ij} = p_{i.}p_{.j}] = D_I \mathbf{1}_I \mathbf{1}_J' D_J.$$

L'indépendance empirique signifie $P = \Pi$.

On construit une mesure de l'écart à l'indépendance par

$$d^2 = N \sum_{i=1}^I \sum_{j=1}^J \frac{1}{p_{i.}p_{.j}} (p_{ij} - p_{i.}p_{.j})^2.$$

Ainsi, $d^2 = 0$ si et seulement si $P = \Pi$. L'indépendance empirique est exceptionnellement réalisée en pratique, la question qui se pose est de savoir si la valeur calculée du d^2 de l'indépendance est proche ou non de 0!

Test de l'indépendance entre \mathcal{L} et \mathcal{C}

Pearson a montré que sous l'hypothèse " $H_0 : \mathcal{L}$ et \mathcal{C} sont indépendantes", d^2 est une réalisation d'une variable aléatoire D^2 qui suit une loi du χ^2 à $(I - 1)(J - 1)$ degrés de liberté.

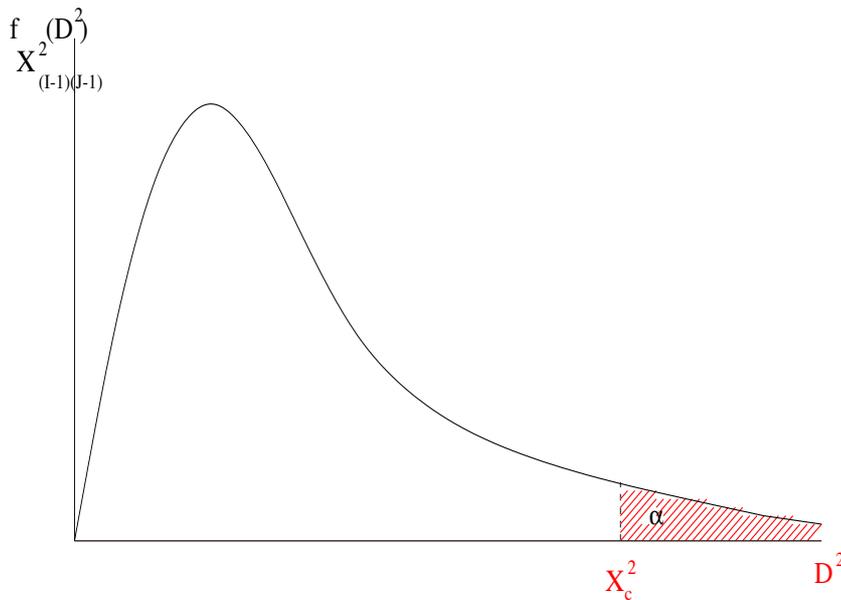


Figure 19 : Densité de probabilité de la loi du χ^2 , région (hachurée) peu probable des valeurs élevées.

Pour un risque d'erreur α , le plus souvent 0.05, on rejette l'hypothèse H_0 si d^2 est supérieur à la valeur critique χ_c^2 qu'une variable $\chi_{(I-1)(J-1)}^2$ a une probabilité α de dépasser, voir Figure 19.

Si $d^2 < \chi_c^2$ on ne peut rejeter l'hypothèse d'indépendance. Dans ce cas, l'information contenue dans P est bien résumée par ses marges ligne et colonne que l'on visualise par des histogrammes classiques.

8.2.3 Profils lignes et colonnes, distributions conditionnelles

On appelle **profil ligne** i le vecteur ligne des fréquences conditionnelles à i fixé

$$L_i = [p_{i1}/p_{i.}, \dots, p_{ij}/p_{i.}, \dots, p_{iJ}/p_{i.}].$$

La matrice $I \times J$ des I profils lignes est notée $L = D_I^{-1}P$ (L pour "Ligne").

L'espace Euclidien des profils lignes (points lignes de L) est R^J . Dans cet espace, chaque point ligne i est affecté de son poids $p_{i.}$, pour former le nuage \mathcal{N}_L .

On appelle **profil colonne** j le vecteur colonne des fréquences conditionnelles à j fixé

$$C^j = [p_{1j}/p_{.j}, \dots, p_{ij}/p_{.j}, \dots, p_{Ij}/p_{.j}]'$$

La matrice $I \times J$ des J profils colonnes est notée $C = PD_J^{-1}$ (C pour "Colonne").

L'espace Euclidien des profils colonnes (points colonnes de C) est R^I . Dans cet espace, chaque point colonne j est affecté de son poids $p_{.j}$, pour former le nuage \mathcal{N}_C .

Il ne doit pas y avoir d'ambiguïté sur les notations \mathcal{L} et \mathcal{C} , nom des variables qualitatives et matrices des profils.

Propriétés de L et C :

- $L\mathbf{1}_J = \mathbf{1}_I$ et $\mathbf{1}_I' C = \mathbf{1}_J'$, (Preuve évidente, propriétés des métriques).
- Le point ligne moyen de L est la marge colonne de P . De façon duale, le point colonne moyen de C est la marge ligne de P . \square

Preuve : Montrons le pour les points lignes. Soit \bar{L} le point ligne moyen, $\bar{L} = \mathbf{1}_I' D_I L = \mathbf{1}_I' D_I D_I^{-1} P = \mathbf{1}_I' P$, qui est bien la marge colonne de P . \square

- **Dans le cas de l'indépendance empirique**, les lignes de L sont égales à la marge colonne de P , les colonnes de C sont identiques à la marge ligne de P . \square

Preuve : Montrons le pour les points lignes. Dans le cas de l'indépendance empirique, $P = \Pi$, alors $L = D_I^{-1} \Pi = D_I^{-1} D_I \mathbf{1}_I \mathbf{1}_J' D_J = \mathbf{1}_I \mathbf{1}_J' D_J$. \square

Le nuage des points lignes de L est concentré au point ligne moyen (la marge colonne de P). Idem pour le nuage des points colonnes de C concentré au point moyen qui est la marge ligne de P .

distance du χ^2 entre profils

On est en présence de deux nuages de points pesants. Le nuage des I profils lignes, chaque profil ligne i est muni du poids $p_{i.}$. Le nuage des J profils colonnes, chaque profil colonne j est muni de son poids $p_{.j}$. Se pose maintenant la question de mesurer des distances Euclidiennes entre profils. Le type de distance utilisé est la distance du χ^2 qui est une distance Euclidienne.

Espace des profils lignes, (\mathbb{R}^J, D_J^{-1}) : Soient L_{i_1} et L_{i_2} deux lignes de L ,

$$d_{\chi^2}^2(L_{i_1}, L_{i_2}) = \sum_{j=1}^J \frac{1}{p_{.j}} \left(\frac{p_{i_1 j}}{p_{i_1.}} - \frac{p_{i_2 j}}{p_{i_2.}} \right)^2 .$$

Espace des profils colonnes, (\mathbb{R}^I, D_I^{-1}) : Soient C^{j_1} et C^{j_2} deux colonnes de C ,

$$d_{\chi^2}^2(C^{j_1}, C^{j_2}) = \sum_{i=1}^I \frac{1}{p_{i.}} \left(\frac{p_{i j_1}}{p_{.j_1}} - \frac{p_{i j_2}}{p_{.j_2}} \right)^2 .$$

La raison essentielle pour choisir la distance du χ^2 est qu'elle vérifie la propriété suivante équivalence distributionnelle :

Lorsque l'on effectue dans le tableau P , la somme de deux colonnes proportionnelles (ou de deux lignes), les distances du χ^2 entre profils lignes (ou colonnes) restent inchangées.

Lien avec le χ^2 de contingence :

$$\frac{d^2}{N} = \sum_{i=1}^I p_{i.} d_{\chi^2}^2(L_i, \bar{L}) = \sum_{j=1}^J p_{.j} d_{\chi^2}^2(C^j, \bar{C})$$

Le coefficient d^2/N est égal à l'inertie du nuage des profils lignes (des profils colonnes). \square

Preuve :

$$\frac{d^2}{N} = \sum_{i=1}^I \sum_{j=1}^J \frac{1}{p_{i.} p_{.j}} (p_{ij} - p_{i.} p_{.j})^2 = \sum_{i=1}^I p_{i.} \sum_{j=1}^J \frac{1}{p_{.j}} \left(\frac{p_{ij}}{p_{i.}} - p_{.j} \right)^2 .$$

On termine en rappelant que \bar{L} est la marge colonne de P . \square

Résumé	Points lignes	Points colonnes
matrice	$L = D_I^{-1}P$	$C = PD_J^{-1}$
nuages	$\mathcal{N}_L = \{L_1, \dots, L_I\}$	$\mathcal{N}_C = \{C^1, \dots, C^J\}$
poids	D_I	D_J
espace	(\mathbb{R}^J, D_J^{-1})	(\mathbb{R}^I, D_I^{-1})
point moyen	$\bar{L} = \mathbf{1}_I'P$	$\bar{C} = P\mathbf{1}_J$
inertie	d^2/N	d^2/N

8.3 Analyse Factorielle des Correspondances

L'Analyse Factorielle des Correspondances (AFC) est une méthode d'analyse de tables de contingences en termes de profils. Lorsque le test d'indépendance entre les deux variables qualitatives a donné comme résultat que l'hypothèse d'indépendance ne pouvait être rejetée, l'information contenue dans le tableau des fréquences P est bien résumée par les marges ligne et colonne que l'on visualise par des diagrammes en bâtons classiques. En effet, les nuages des points profils lignes et profils colonnes sont dans le cas de l'indépendance empirique, confondus avec leurs points moyens correspondants qui sont les marges de P .

L'objectif de l'AFC est de visualiser dans des plans factoriels, les nuages \mathcal{N}_L et \mathcal{N}_C des points profils et de situer ces nuages par rapport à leurs points moyens respectifs. La notion de proximité est basée sur la distance du χ^2 . On peut donc effectuer deux ACP généralisées, celle du triplet $(X = L, M = D_J^{-1}, D = D_I)$ puis celle du triplet $(X = C', M = D_I^{-1}, D = D_J)$ dont les composantes principales fourniront respectivement les représentations en projection de \mathcal{N}_L et de \mathcal{N}_C . La définition choisie dans ce cours présente l'avantage de n'effectuer qu'une ACP généralisée dont les représentations duales des lignes et des colonnes fourniront les projections des deux nuages.

8.3.1 Définition et propriétés

L'AFC d'ordre k du tableau T est l'ACP généralisée d'ordre k du triplet

$$(X = D_I^{-1}PD_J^{-1}, M = D_J, D = D_I).$$

La matrice $X = \left[\frac{p_{ij}}{p_{i.}p_{.j}} \right]$ est de dimensions $I \times J$.

Propriétés du triplet

- L'espace des points lignes de X est (\mathbb{R}^J, D_J) celui des points colonnes (\mathbb{R}^I, D_I) .
- **Où l'on retrouve les distances du χ^2 entre profils.**

Soient X_{i_1} et X_{i_2} deux vecteurs lignes de X ,

$$\begin{aligned} d_{D_J}^2(X_{i_1}, X_{i_2}) &= (X_{i_1} - X_{i_2}) D_J (X_{i_1} - X_{i_2})' \\ &= \sum_{j=1}^J p_{.j} \left(\frac{p_{i_1 j}}{p_{i_1.} p_{.j}} - \frac{p_{i_2 j}}{p_{i_2.} p_{.j}} \right)^2 \\ &= \sum_{j=1}^J \frac{1}{p_{.j}} \left(\frac{p_{i_1 j}}{p_{i_1.}} - \frac{p_{i_2 j}}{p_{i_2.}} \right)^2 \\ &= d_{\chi^2}^2(L_{i_1}, L_{i_2}). \end{aligned}$$

On vérifie de même que pour deux colonnes, $d_{D_I}^2(X^{j_1}, X^{j_2}) = d_{\chi^2}^2(C^{j_1}, C^{j_2})$. \square

On notera donc \mathcal{N}_L et \mathcal{N}_C les deux nuages de points lignes et colonnes puisque la position relative de leurs points est mesurée de façon identique à celle des lignes de L et celle des colonnes de C .

- **Centrage des colonnes de X au sens de D_I** : (utiliser **M2**, paragraphe 8.2.1)
 $\mathbf{1}_J'$ est le point ligne moyen de X car $\mathbf{1}_J' D_I X = \mathbf{1}_J'$. La matrice X^c obtenue par D_I -centrage des colonnes de X est égale à

$$X^c = X - \mathbf{1}_I \mathbf{1}_J' = \left[\frac{p_{ij}}{p_{i.} p_{.j}} - 1 \right].$$

Remarquer que $X D_J \mathbf{1}_J = \mathbf{1}_I$ ce qui signifie $\mathbf{1}_I$ est le point colonne moyen de X et que X^c est D_J -centrée en lignes.

Interprétation du point ligne (colonne) moyen :

$$d_{D_J}^2(X_i, \mathbf{1}_J') = d_{\chi^2}^2(L_i, \bar{L}) \quad \text{et} \quad d_{D_I}^2(X^j, \mathbf{1}_I) = d_{\chi^2}^2(C^j, \bar{C}).$$

Le point ligne (colonne) moyen s'interprète donc comme \bar{L} (comme \bar{C}), c'est à dire la marge colonne (ligne) de P .

- **Opérateurs du triplet** ($X^c = X - \mathbf{1}_I \mathbf{1}_J'$, $M = D_J$, $D = D_I$) :

De fait, c'est la DVS du triplet sur matrice centrée qui est mise en oeuvre :

$$\mathbb{V}M = X^c' D_I X^c D_J = X' D_I X D_J - \mathbf{1}_J \mathbf{1}_J' D_J = D_J^{-1} P' D_I^{-1} P - \mathbf{1}_J \mathbf{1}_J' D_J$$

$$\mathbb{W}D = X^c D_J X^c' D_I = X D_J X' D_I - \mathbf{1}_I \mathbf{1}_I' D_I = D_I^{-1} P D_J^{-1} P' - \mathbf{1}_I \mathbf{1}_I' D_I$$

Par dualité, on passe de l'un à l'autre en permutant I avec J , P avec P' .

- **Inertie des nuages \mathcal{N}_L et \mathcal{N}_C :**

D'après le paragraphe 7.2.1, l'inertie des nuages \mathcal{N}_L et \mathcal{N}_C de poids D_I et D_J , par rapport à leurs points moyens respectifs, est égale à

$$\text{trace}(\mathbb{W}D) = \text{trace}(\mathbb{V}M) = \frac{d^2}{N}.$$

Cette relation fournit le lien entre le χ^2 de l'indépendance et la somme des valeurs propres des opérateurs

$$\sum_{\alpha} \lambda_{\alpha} = \frac{d^2}{N}.$$

8.3.2 Éléments propres des opérateurs en dualité

Les deux opérateurs en dualité dans la DVS du triplet, calculés au paragraphe précédent, fournissent comme vecteurs propres les axes factoriels de l'AFC.

$$\mathbb{V}M V^{\alpha} = \lambda_{\alpha} V^{\alpha} \quad \text{et} \quad \mathbb{W}D U^{\alpha} = \lambda_{\alpha} U^{\alpha}.$$

Le formulaire du paragraphe 7.4.2 résume les propriétés de ces vecteurs. Présentons maintenant leurs propriétés spécifiques dans le cadre de l'AFC.

- $0 \leq \lambda_{\alpha} \leq 1$, voir paragraphe 8.5.1, Exercice 1.

- **Valeur propre triviale et vecteurs propres triviaux :**

0 est valeur propre dite triviale associée aux vecteurs propres $\mathbf{1}_J$ pour $\mathbb{V}M$ et $\mathbf{1}_I$ pour $\mathbb{W}D$. \square

Preuve : Montrons le seulement pour $\mathbf{1}_J$. L'opérateur $\mathbb{V}M$ est la matrice $J \times J$ telle que $\mathbb{V}M = D_J^{-1} P' D_I^{-1} P - \mathbf{1}_J \mathbf{1}_J' D_J$. Alors, les propriétés **M1** et **M2** des métriques D_I et D_J énoncées au paragraphe 8.2.1, donnent $\mathbb{V}M \mathbf{1}_J = 0 \mathbf{1}_J$. \square

L'opérateur $\mathbb{V}M$, calculé sur la matrice X^c centrée en colonnes, s'exprime simplement en fonction de la matrice X non centrée. On a donc

$$\mathbb{V}M \mathbf{1}_J = X' D_I X D_J \mathbf{1}_J - \mathbf{1}_J = 0 \mathbf{1}_J,$$

Ce qui signifie que 1 est valeur propre triviale associée au vecteur trivial $\mathbf{1}_J$, pour l'opérateur $\mathbb{V}M$ calculé sur la matrice X non centrée.

Ces propriétés se reconduisent par dualité pour l'opérateur $\mathbb{W}D$.

- Soient λ_{α} valeur propre non triviale, $V^{\alpha} \neq \mathbf{1}_J$ et $U^{\alpha} \neq \mathbf{1}_I$ les vecteurs propres de $\mathbb{V}M$ et de $\mathbb{W}D$ associés à λ_{α} . Alors, λ_{α} est aussi valeur propre des opérateurs calculés sur la matrice X non centrée, avec les mêmes vecteurs associés V^{α} et U^{α} .

□

Preuve : Montrons le seulement sur $\mathbb{V}M$. V^α est vecteur propre non trivial de $\mathbb{V}M$

$$\mathbb{V}MV^\alpha = X^c D_I X^c D_J V^\alpha = X' D_I X D_J V^\alpha - \mathbf{1}_J \mathbf{1}_J' D_J V^\alpha = \lambda_\alpha V^\alpha.$$

V^α et $\mathbf{1}_J$ sont deux vecteurs propres de $\mathbb{V}M$ distincts. Les vecteurs propres forment une base D_J -orthonormée de \mathbb{R}^J . Il en résulte que $\mathbf{1}_J' D_J V^\alpha = 0$. □

En résumé, que l'on centre ou non la matrice $X = D_I^{-1} P D_J^{-1}$, les axes factoriels sont les mêmes. La seule différence vient de la valeur propre triviale qui est 0 ou 1 suivant que X est D_I -centrée en colonnes ou non. Dans la pratique, la matrice X est toujours centrée.

8.3.3 Pratique de l'AFC

L'AFC permet de visualiser les écarts des deux nuages des profils par rapport à leurs moyennes respectives (les marges de P). Il est donc primordial d'analyser avant tout les deux distributions marginales de P .

Une fois analysé les marges, on peut se poser le problème du choix du nombre k d'axes factoriels à retenir, on sait en outre que $k \leq \min(I - 1, J - 1)$. Le critère global de qualité basé sur le pourcentage d'inertie reconstruite avec k axes factoriels, est identique à celui de toute ACP généralisée, voir paragraphe 7.2.1.

Examinons maintenant les particularismes de l'AFC qui en font une méthode qui se démarque de l'ACP usuelle.

– **Les axes principaux sont centrés :**

C'est une conséquence du fait que $\mathbf{1}_J$ est vecteur propre de $\mathbb{V}M$ et que les vecteurs propres sont D_J -orthogonaux

$$\mathbf{1}_J' D_J A^\alpha = \sum_{j=1}^J p_{.j} A_j^\alpha = 0.$$

Ce qui signifie que les axes principaux sont D_J -centrés. En particulier, pour $\alpha = 1$, on ne peut avoir $A_j^1 > 0 \quad \forall j = 1 \dots, J$. **Le premier axe principal de l'AFC ne peut donc être un axe de taille** comme c'est le cas, parfois, en ACP usuelle, voir paragraphe 7.3, Exercice 7.

– **Représentations simultanées des points lignes et colonnes**

Les profils lignes barycentres des profils colonnes :

Considérons, dans l'espace \mathbb{R}^J , l'axe factoriel V^α sur lequel le point ligne L_i se projette au point de coordonnée C_i^α , où $C^\alpha = \sqrt{\lambda_\alpha} U^\alpha$ est la composante principale numéro α . Pour ne pas confondre avec les profils colonnes C^j , ceux ci sont

indiqués supérieurement par une lettre latine, les composantes principales sont indiquées supérieurement par une lettre grecque. La formule de transition (*), voir Formulaire 7.4.2, appliquée au cas de l'AFC où $X^c = D_I^{-1} P D_J^{-1} - \mathbf{1}_I \mathbf{1}_J'$ et $M = D_J$, donne

$$C^\alpha = X^c M \frac{A^\alpha}{\sqrt{\lambda_\alpha}} = D_I^{-1} P \frac{A^\alpha}{\sqrt{\lambda_\alpha}} = L \frac{A^\alpha}{\sqrt{\lambda_\alpha}}.$$

Il résulte

$$C_i^\alpha = L_i \frac{A^\alpha}{\sqrt{\lambda_\alpha}} = \sum_{j=1}^J \frac{p_{ij}}{p_i} \frac{A_j^\alpha}{\sqrt{\lambda_\alpha}}.$$

Ainsi, grâce à un changement d'échelle $\frac{1}{\sqrt{\lambda_\alpha}}$ (dilatation car $0 \leq \lambda_\alpha \leq 1$), C_i^α est au barycentre des points $\{A_j^\alpha, j = 1, \dots, J\}$ munis des poids du profil ligne L_i . Mais, dans l'espace \mathbb{R}^I , A_j^α est la projection du point colonne C^j sur l'axe factoriel U^α . Il en résulte qu'il est légitime de représenter simultanément dans un même plan factoriel (α, β) , Figure 20, les profils lignes au barycentre des profils colonnes. Figure 20, un seul profil ligne L_i est représenté, pour ne pas alourdir la figure.

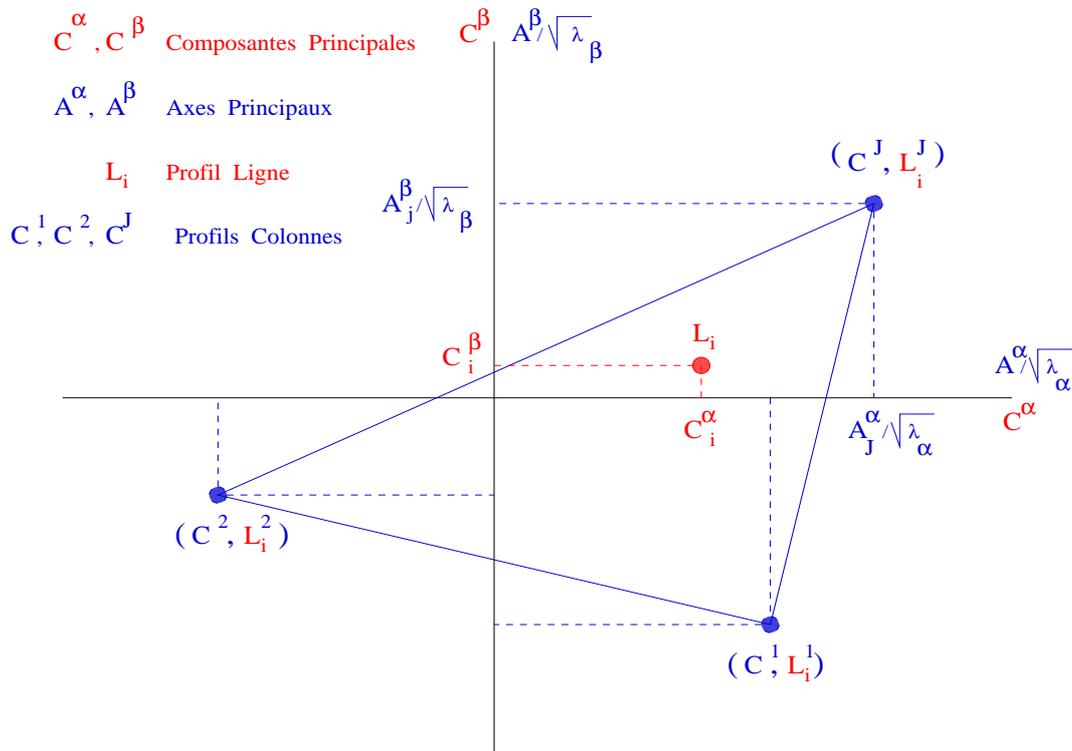


Figure 20 : représentation barycentrique dans le plan factoriel (α, β) du profil ligne L_i au barycentre des profils colonnes munis des poids puisés dans les valeurs du profil L_i .

Les profils colonnes barycentres des profils lignes :

Par dualité, grâce à la formule de transition (**)', on obtient

$$A_j^\alpha = \sum_{i=1}^I \frac{p_{ij}}{p_{.j}} \frac{C_i^\alpha}{\sqrt{\lambda_\alpha}}.$$

Ce qui légitime la représentation simultanée sur un plan factoriel des profils colonnes au barycentre des points lignes (toujours à un facteur de dilatation).

Le facteur de dilatation associé à un axe factoriel est négligeable si la valeur propre est voisine de 1 ; il est de plus en plus important au fur et à mesure que l'axe factoriel est associé à une valeur propre de plus en plus proche de 0.

Dans la pratique, on admet comme légitime la représentation simultanée des profils lignes et des profils colonnes sur un même plan factoriel, sans facteur de dilatation, bien que pour les premiers axes, les valeurs propres sont souvent plus proches de 0 que de 1.

Au vu de leur proximité sur un plan factoriel, on ne peut légitimer la proximité entre un profil ligne et un profil colonne que si ceux ci sont **bien représentés** sur ce plan.

– **Éléments supplémentaires**

Les formules de transition, dans le cas particulier de l'AFC, s'écrivent

$$C^\alpha = L \frac{A^\alpha}{\sqrt{\lambda_\alpha}} \quad (*)', \quad A^\alpha = C' \frac{C^\alpha}{\sqrt{\lambda_\alpha}} \quad (**)'.$$

Si les éléments supplémentaires sont des fréquences conditionnelles (profils), il est parfaitement valable d'utiliser ces formules dans lesquelles on remplace L , matrice des profils lignes actifs, par une matrice de profils lignes supplémentaires et C , matrice des profils colonnes actifs, par une matrice de profils colonnes supplémentaires. Un point ligne supplémentaire se trouvera au barycentre des points colonnes actifs munis des poids de cette ligne. Par dualité, un point colonne supplémentaire sera situé au barycentre des points lignes actifs munis des poids de cette colonne.

8.3.4 Cas d'une variable ordinale, rapport de corrélation et "optimal scoring"

Rapport de corrélation : Soient x et y deux variables quantitatives à I et J valeurs dont les fréquences des réalisations (x_i, y_j) sont stockées dans P défini en 8.2.1. Dû à Guttman, le rapport de corrélation de y par rapport à x est l'indice statistique dont le carré est

$$\eta_{y/x}^2 = \frac{\text{var}(E(y/x))}{\text{var}(y)} = \frac{\sum_{i=1}^I p_i (E(y/x = x_i) - \bar{y})^2}{\text{var}(y)}.$$

Remarquons que x peut être une variable qualitative, dans ce cas x_i est la i ème modalité de x .

La variance de y se décompose en

$$\text{var}(y) = \sum_{j=1}^J p_{.j} (y_j - \bar{y})^2 = \sum_{i=1}^I p_i \text{var}(y/x = x_i) + \sum_{i=1}^I p_i (E(y/x = x_i) - \bar{y})^2 = V_w + V_b,$$

où $V_b = \text{var}(E(y/x))$ est la variance **entre** (between) les modalités représentées par les moyennes conditionnelles $E(y/x = x_i)$ et $V_w = E(\text{var}(y/x))$ est la variance **à l'intérieur** (within) des modalités $x = x_i$ encore appelée moyenne des variances conditionnelles. Ceci a pour conséquence

$$0 \leq \eta_{y/x}^2 \leq 1.$$

- Si $\eta_{y/x}^2 = 1$, cela signifie que $V_w = 0$ c'est à dire que, à i fixé,

$$\text{var}(y/x = x_i) = \sum_{j=1}^J \frac{p_{ij}}{p_i} (y_j - E(y/x = x_i))^2 = 0.$$

Cela est réalisé si sur chaque ligne i de P , il n'y a qu'une seule valeur p_{ij} non nulle. Dans ce cas, la connaissance de x détermine celle de y par une relation fonctionnelle.

- Si $\eta_{y/x}^2 = 0$, alors, pour tout i , $E(y/x = x_i) = \bar{y}$. Ce cas est réalisé lorsque les variables y et x sont indépendantes empiriquement. En effet, sous cette hypothèse,

$$E(y/x = x_i) = \sum_{j=1}^J \frac{p_{ij}}{p_i} y_j = \sum_{j=1}^J p_{.j} y_j = \bar{y}.$$

Lorsque x est une variable quantitative, on peut montrer la relation suivante entre le rapport de corrélation et le coefficient de corrélation linéaire

$$0 \leq r^2(x, y) \leq \eta_{y/x}^2 \leq 1 \quad \text{et} \quad 0 \leq r^2(x, y) \leq \eta_{x/y}^2 \leq 1.$$

“Optimal scoring” en AFC

Il arrive souvent dans l’AFC du tableau P qu’une des deux variables qualitatives, par exemple \mathcal{C} , soit une variable ordinale, par exemple le temps ou toute autre variable dont les modalités présentent un ordre naturel, voir Exercice 3 paragraphe 8.5.1. On peut lui associer une variable quantitative $y = [y_1, \dots, y_J]'$ à valeurs ordonnées

$$y_1 < \dots < y_j < \dots < y_J ,$$

que l’on supposera D_J -centrée réduite

$$\bar{y} = \mathbf{1}_J' D_J y = 0, \quad \text{var}(y) = y' D_J y = 1.$$

On peut alors associer à la variable qualitative \mathcal{L} une variable quantitative t formée par les “scores” lignes

$$t = D_I^{-1} P y = L y$$

c’est à dire, le score t_i est la moyenne des y sachant la modalité \mathcal{L}_i

$$t_i = L_i y = \sum_{j=1}^J \frac{p_{ij}}{p_{i.}} y_j = E(y / \mathcal{L} = \mathcal{L}_i).$$

Remarquer que t est D_I -centrée car, d’après M2, $\bar{t} = \mathbf{1}_I' D_I t = \mathbf{1}_I' P y = \mathbf{1}_J' D_J y = 0$.

La variance de t , notée $\text{var}(t)$, ou variance des moyennes conditionnelles de y à \mathcal{L} fixé, est définie par

$$\text{var}(t) = \text{var}(E(y / \mathcal{L})) = t' D_I t = y' P' D_I^{-1} P y.$$

Le problème appelé ”optimal scoring” est celui-ci : Comment choisir y pour que les scores lignes $\{t_1, \dots, t_I\}$ soient optimaux au sens “le plus séparés possibles” ? Le critère à maximiser est donc celui de la variance de t

$$\max_{y' D_J y = 1} y' P' D_I^{-1} P y.$$

Par la méthode des multiplicateurs de Lagrange, la solution \hat{y} est donnée par le premier vecteur propre non trivial, le premier axe factoriel, de l’opérateur $\mathbb{V}M$ dans l’AFC de P

$$D_J^{-1} P' D_I^{-1} P \hat{y} = \lambda_1 \hat{y}.$$

Il faut noter que la relation d’ordre n’est pas forcément conservée sur les valeurs de \hat{y} puisque le domaine des contraintes du problème d’optimisation ne la prend pas en compte. Elle est cependant conservée dans de très nombreux cas, et il convient d’analyser

précisément les anomalies éventuelles.

Les scores lignes de dispersion maximale sont donnés par la formule de transition $\hat{t} = D_I^{-1} P \hat{y} = L \hat{y}$, ce qui signifie que \hat{t} est la première composante principale de l'AFC de P .

On note que la première valeur propre non triviale est la valeur optimale du carré du rapport de corrélation de y par rapport à \mathcal{L} , $\lambda_1 = \hat{y}' P' D_I^{-1} P \hat{y} = \text{var}(E(\hat{y}/\mathcal{L}))/\text{var}(\hat{y}) = \eta_{\hat{y}/\mathcal{L}}^2$.

L'effet Guttman

La représentation conjointe du plan factoriel (1, 2), Figure 21, est typique de l'effet Guttman encore appelé effet "fer à cheval".

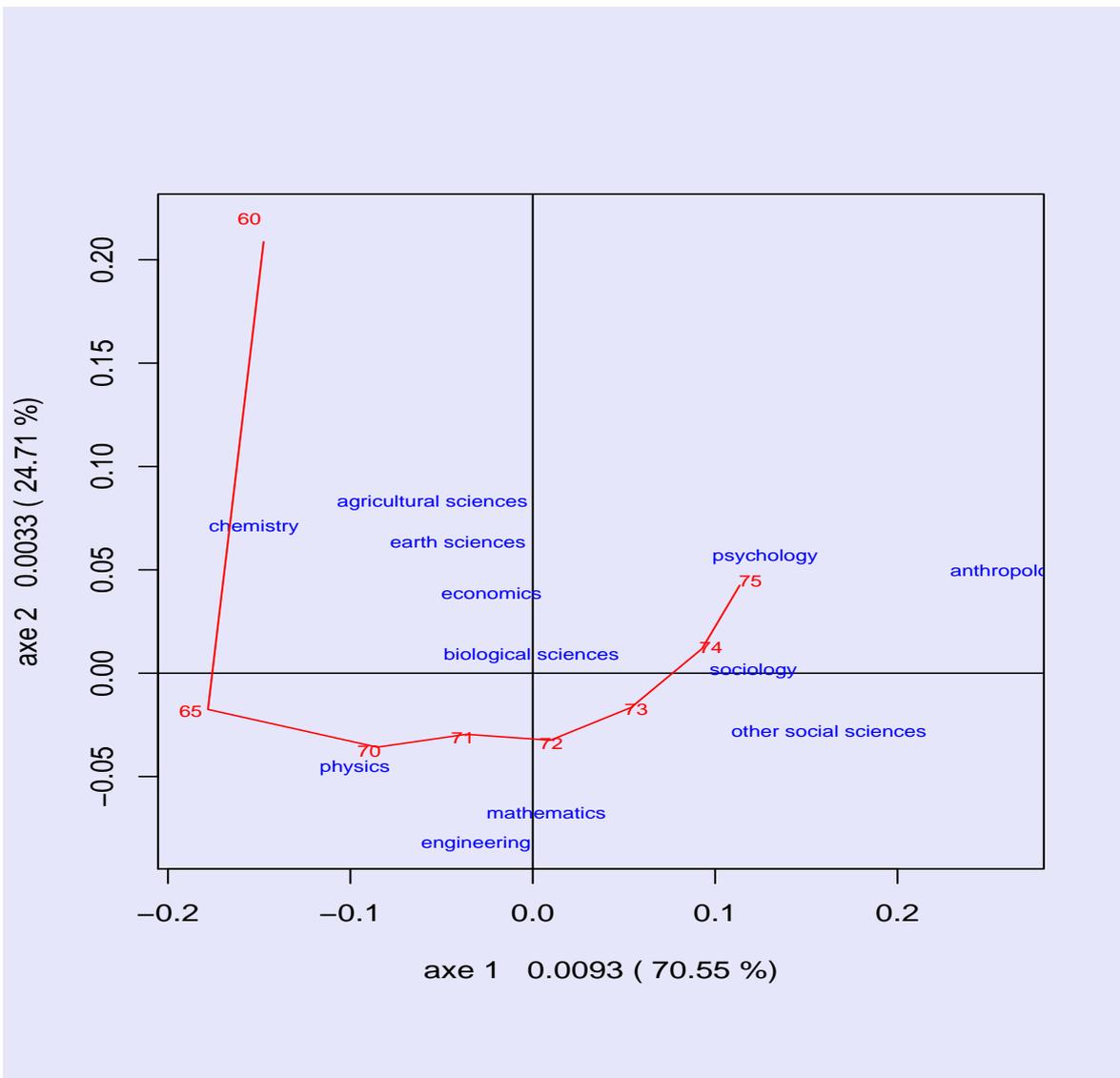


Figure 21 : Représentation conjointe (1,2) des modalités des variables "année" et "doctorats es-sciences" aux USA.

Il s'agit de l'AFC d'un jeu de données dû à Gabriel et Zamir, voir aussi [6, Greenacre], croisant les effectifs des types de doctorats es-sciences accordés aux USA et les modalités 1960,1965-1975 de la variable année. Une fourche parabolique représente souvent les modalités de la variable ordinale lorsque l'on peut réordonner les modalités de l'autre variable (selon l'ordre des valeurs sur l'axe 1) pour faire apparaître une diagonale "forte" dans le tableau des profils.

L'axe 1 oppose les valeurs faibles aux valeurs fortes de la variable ordinale, alors que l'axe 2 oppose ses valeurs extrêmes aux valeurs moyennes. L'axe 1 s'interprète comme une évolution au cours du temps des profils types de doctorats, la tendance de la courbe régulière traduit un changement allant des sciences "dures" traditionnelles vers les sciences "humaines". Les doctorats "agricultural sciences", "earth sciences", et "economics", situés à l'intérieur du fer à cheval du côté des années anciennes, sont des profils situés au dessus du profil ligne moyen pour les années extrêmes.

8.4 Analyse Factorielle des Correspondances Multiples

Cette méthode, en abrégé AFCM, est une AFC appliquée non pas à une table de contingence, mais au tableau des fréquences issu de la super-matrice de codage d'un nombre de variables qualitatives supérieur ou égal à deux. Dans le cas d'un questionnaire, cela revient à juxtaposer en colonnes les matrices de codage disjonctif complet de toutes les questions. Au contraire de l'AFC, les individus interrogés sont directement pris en compte dans l'AFCM en tant que lignes du tableau analysé.

8.4.1 Définitions et propriétés

Q : nombre de variables qualitatives observées sur les mêmes N individus ($Q \geq 2$).

La variable j , $1 \leq j \leq Q$, à m_j modalités, est codée par la matrice U^j , $N \times m_j$.

$M = \sum_{j=1}^Q m_j$ est le nombre total de modalités.

$U = [U^1 \dots U^j \dots U^Q]$ est la super-matrice $N \times M$ de codage des Q variables.

Clairement, $\mathbf{1}_N' U \mathbf{1}_M = NQ$ est la somme de tous les éléments de U . On est amené à définir la matrice des fréquences et ses marges lignes et colonnes.

La matrice des fréquences associée à U est

$$F = \frac{1}{NQ}U.$$

La marge ligne de F est le vecteur colonne $F\mathbf{1}_M = \frac{1}{N}\mathbf{1}_N$.

La marge colonne de F est le vecteur ligne

$$\frac{1}{NQ} \left[U_{\cdot 1}^1 \dots U_{\cdot m_1}^1 \dots U_{\cdot 1}^Q \dots U_{\cdot m_Q}^Q \right],$$

où $[U_{\cdot 1}^j \dots U_{\cdot m_j}^j]$ est la marge colonne de U^j .

Métriques associées aux marges de F

L'espace \mathbb{R}^N des colonnes de F est muni de la métrique diagonale D_I des poids statistiques des individus,

$$D_I = \frac{1}{N}I_N.$$

L'espace \mathbb{R}^M des lignes de F est muni de la métrique diagonale D_J des poids statistiques puisés dans la marge colonne de F ,

$$D_J = \frac{1}{NQ} \text{diag} \left(U_{\cdot 1}^1 \dots U_{\cdot m_1}^1 \mid \dots \mid U_{\cdot 1}^Q \dots U_{\cdot m_Q}^Q \right).$$

La marge colonne de F s'écrit, $\mathbf{1}_N'F = \mathbf{1}_M'D_J$.

Proposition : Soit F^c la matrice des colonnes de F , D_I -centrées. Alors

$$\text{rang}(F^c) \leq \min(N - 1, M - Q). \quad \square$$

Preuve : $F^c = [U^{1c} \dots U^{Qc}]$, où U^{jc} est la matrice centrée en colonnes du bloc U^j . Les propriétés du codage, voir paragraphe 8.1.2, impliquent que $U^{jc}\mathbf{1}_{m_j} = 0_N$, pour $j = 1, \dots, Q$, ce qui implique que $\text{rang}(F^c) \leq M - Q$. Puisque F^c est centrée, $N^{-1}\mathbf{1}_N'F^c = 0_M$, ce qui termine la preuve. \square

L'AFCM d'ordre k du tableau U est l'AFC d'ordre k du tableau F des fréquences, c'est à dire l'ACP généralisée d'ordre k du triplet

$$(X = D_I^{-1}FD_J^{-1}, D_J, D_I).$$

Les métriques utilisées étant les matrices diagonales des marges de F .

Comme dans l'AFC, c'est le triplet $(X^c = X - \mathbf{1}_N\mathbf{1}_M', D^J, D^I)$ qui est effective-

ment mis en oeuvre. Comme

$$X^c = NF^c D_J^{-1} = D_I^{-1} F^c D_J^{-1},$$

on en déduit que le rang de X^c est égal à celui de F^c .

L'expression des opérateurs en dualité s'obtient à partir de celle de l'AFC, en remplaçant P , \mathbf{I}_I et \mathbf{I}_J respectivement par F , \mathbf{I}_N et \mathbf{I}_M .

$$\mathbb{V}M = X^c D_I X^c D_J = X' D_I X D_J - \mathbf{I}_M \mathbf{I}_M' D_J = D_J^{-1} F' D_I^{-1} F - \mathbf{I}_M \mathbf{I}_M' D_J$$

$$\mathbb{W}D = X^c D_J X^c D_I = X D_J X' D_I - \mathbf{I}_N \mathbf{I}_N' D_I = F D_J^{-1} F' D_I^{-1} - \mathbf{I}_N \mathbf{I}_N' D_I.$$

Valeurs propres et inertie des nuages

Comme en AFC, les valeurs propres sont comprises entre 0 et 1 ; le nombre de valeurs propres non nulles, égal au rang de X^c , vérifie

$$\text{nombre de valeurs propres non nulles} \leq \min(N - 1, M - Q).$$

D'autre part, l'inertie du nuage \mathcal{N}_C des modalités colonnes et celle du nuage \mathcal{N}_L des individus ont pour valeur,

$$\mathcal{I} = \text{trace}(\mathbb{W}D) = \text{trace}(\mathbb{V}M) = \frac{M}{Q} - 1 = \sum_{\alpha} \lambda_{\alpha}.$$

Si l'on s'intéresse au nuage \mathcal{N}_C on peut décomposer l'inertie \mathcal{I} en une somme d'inerties partielles représentant la part de chaque variable

$$\mathcal{I} = \sum_{q=1}^Q \mathcal{I}(q) \quad \text{où} \quad \mathcal{I}(q) = \frac{m_q - 1}{Q}.$$

L'inertie $\mathcal{I}(q)$ imputée à la variable q est d'autant plus importante que la variable possède de modalités. On peut tenir compte de cette remarque dans la construction d'un questionnaire d'enquête.

L'inertie $\mathcal{I}(q)$ peut elle aussi se décomposer pour faire apparaître l'influence des modalités

$$\mathcal{I}(q) = \sum_{j=1}^{m_q} \mathcal{I}(q; j) \quad \text{où} \quad \mathcal{I}(q; j) = \frac{1 - U_{.j}^q / N}{Q}.$$

Une modalité à effectif faible aura relativement plus d'influence qu'une modalité à fort effectif.

8.4.2 Pratique de l'AFCM

Représentation des individus

Il arrive parfois que le nombre des individus est si important que leur représentation dans des plans factoriels est difficile voire impossible à analyser. Cependant, la représentation barycentrique des individus prend un sens particulier en AFCM.

Les projections du nuage \mathcal{N} des individus sur l'axe factoriel V^α sont données par la composante principale C^α .

$$C^\alpha = X D_J \frac{A^\alpha}{\sqrt{\lambda_\alpha}} = \frac{1}{Q} U \frac{A^\alpha}{\sqrt{\lambda_\alpha}}.$$

L'individu i se projette donc selon

$$C_i^\alpha = \sum_{j \in \mathcal{M}_i} \frac{1}{Q} \frac{A_j^\alpha}{\sqrt{\lambda_\alpha}},$$

où $\mathcal{M}_i = \{j \in \{1, \dots, M\} \mid U_{ij} = 1\}$ est l'ensemble, de cardinal égal à Q , des modalités prises par l'individu i . Au facteur $1/\sqrt{\lambda_\alpha}$ près, l'individu i est donc à l'iso-barycentre de ses choix de réponses.

Représentation des modalités

Centrage "global" des modalités :

Comme en AFC, les modalités projetées sur l'axe factoriel U^α sont D_J -centrées. En effet, $\mathbf{1}_J$, vecteur propre trivial, est D_J -orthogonal à l'axe principal A^α dont les coordonnées donnent les projections du nuage \mathcal{N}_C sur l'axe factoriel. De là résulte le centrage de A^α .

$$\mathbf{1}_J' D_J A^\alpha = 0.$$

Centrage des modalités "à l'intérieur" d'une variable :

Notons D_q la matrice diagonale des fréquences associée à la variable qualitative q ,

$$D_q = \text{diag}\left(\frac{U_{.1}^q}{N}, \dots, \frac{U_{.m_q}^q}{N}\right).$$

L'axe principal A^α est découpé en Q blocs

$$A^\alpha = \begin{bmatrix} A^\alpha(1) \\ \vdots \\ A^\alpha(q) \\ \vdots \\ A^\alpha(Q) \end{bmatrix}.$$

Le vecteur colonne $A^\alpha(q)$, de dimension m_q , fournit les coordonnées des projections des modalités de la variable q , sur l'axe factoriel U^α . Alors,

$$\mathbf{1}_{m_q}' D_q A^\alpha(q) = 0,$$

ce qui signifie que les modalités d'une variable sont centrées en projection par rapport à leurs fréquences, voir Exercice 1 du paragraphe 8.5.2.

8.5 Exercices

8.5.1 Analyse Factorielle des Correspondances

Exercice 1 : Soient L et C deux variables qualitatives à I et J modalités. Une enquête portant sur une population de N individus a donné comme dépouillement les deux matrices de codage disjonctif complet U_L et U_C .

1. Expliciter $U_L'U_L$, $U_C'U_C$, $U_L'U_C$, $\mathbf{1}_N'U_L$ et $\mathbf{1}_N'U_C$.
2. On note A^α et C^α l'axe principal et la composante principale associés à la valeur propre λ_α . Rappeler la définition de ces deux vecteurs. Soient $\xi_\alpha = U_L C^\alpha$ et $\eta_\alpha = U_C A^\alpha$, vecteurs de $(\mathbb{R}^N, N^{-1}I_N)$. Montrer que ces deux vecteurs sont centrés, ont pour variance λ_α et que $\sqrt{\lambda_\alpha}$ est le coefficient de corrélation linéaire entre ξ_α et η_α . En déduire que en AFC, $0 \leq \lambda_\alpha \leq 1$.

Exercice 2 :

Analyse des correspondances d'une matrice "condensée", [6, Greenacre].

Soit P , $I \times J$, la matrice des fréquences associée à un tableau de contingence T et P_0 , $H \times J$, $H < I$, la matrice déduite de P en ajoutant ensemble des groupes disjoints de lignes de P , ce qui peut s'écrire : $P_0 = U_0'P$, où U_0 , $I \times H$, est la matrice du codage disjonctif complet des H groupes (une matrice de 0 et de 1 avec un seul 1 dans chaque ligne, $U_0 \mathbf{1}_H = \mathbf{1}_I$).

1. Soit D_I la matrice diagonale telle que $D_I \mathbf{1}_I = P \mathbf{1}_J$. Comment sont formées les lignes de la matrice $U_0' D_I$? Quelle est la marge ligne de cette matrice? Écrire l'expression matricielle des profils lignes de P . Montrer que chaque profil ligne de P_0 est le centre de gravité du groupe des profils lignes de P qui ont été regroupés pour former la ligne correspondante.
2. Soient C , A et $\Lambda_r = \text{diag}(\lambda_1, \dots, \lambda_r)$ respectivement les matrices des composantes principales, des axes principaux et des inerties non nulles non triviales associées à l'AFC de P . Écrire les formules de transition de l'AFC de P .

3. Montrer comment les lignes de P_0 peuvent être représentées comme points lignes supplémentaires à l'AFC de P et, réciproquement, comment les lignes de P peuvent être représentées comme points supplémentaires dans l'AFC de P_0 .

Exercice 3 :

AFC d'une enquête sur le tabagisme dans une entreprise, [6, Greenacre].

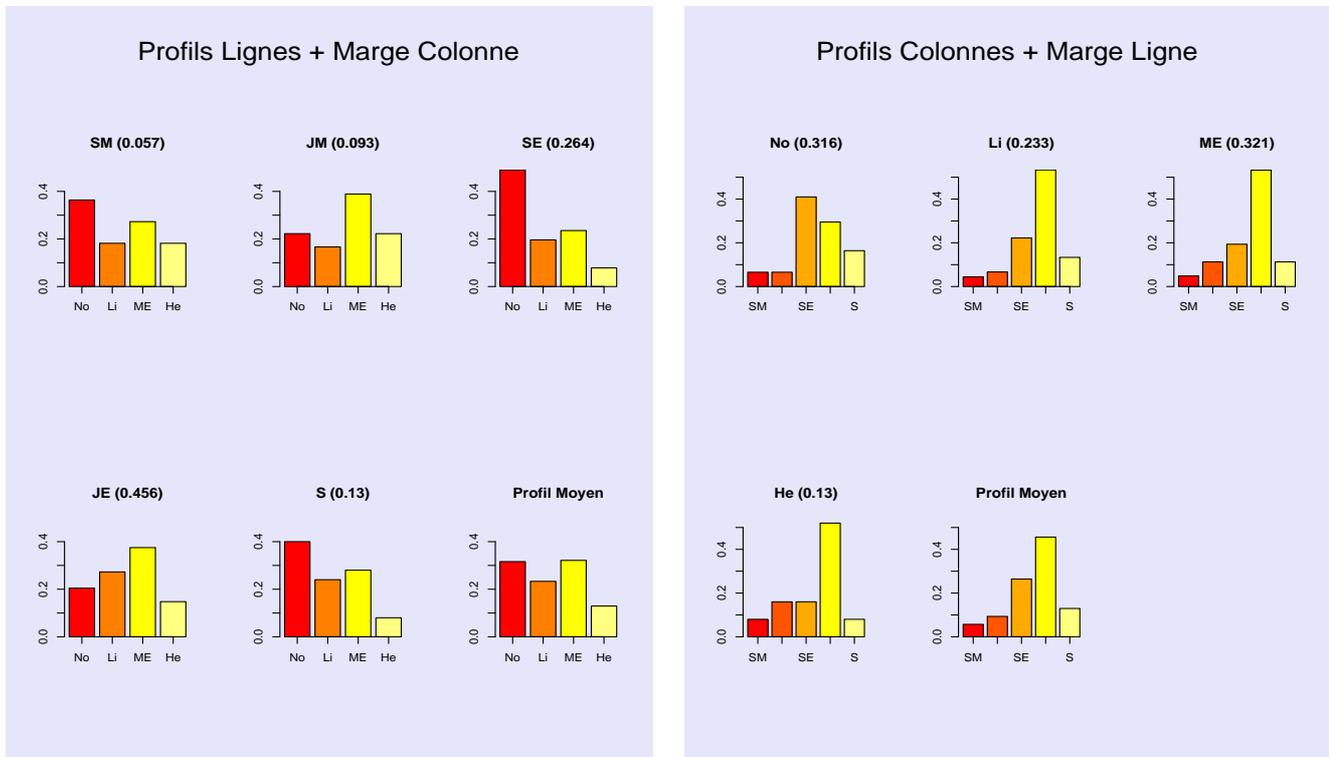
Après la publication des résultats d'une enquête nationale sur le tabagisme, le directeur du personnel d'une grande entreprise américaine a décidé de mener une enquête à l'intérieur de son établissement. Ayant consulté une société de consulting en statistique, il décida de partager les membres du personnel en 5 catégories : (1) senior management, (2) junior management, (3) senior employees, (4) junior employees, (5) secretarial staff. Un échantillon aléatoire de 10% est tiré au sort à l'intérieur de chaque groupe et chacune des personnes est interrogée pour savoir si il ou elle (a) ne fume pas, (b) fume 1-10 cigarettes par jour, (c) fume 11-20 cigarettes par jour, (d) fume plus de 20 cigarettes par jour. Ces seuils ont été choisis pour séparer respectivement non-fumeurs, fumeurs légers, fumeurs moyens, gros fumeurs. L'enquête portant sur 193 individus est résumée dans le tableau de contingence.

Staff group	Smoking				Consum. Alcool	
	None	Light	Medium	Heavy	NO	YES
Senior managers	4	2	3	2	0	11
Junior managers	4	3	7	4	1	17
Senior employees	25	10	12	4	5	46
Junior employees	18	24	33	13	10	78
Secretaries	10	6	7	2	7	18
% national	42	29	20	9	Not relevant	

D'autre part, les pourcentages de non fumeurs, fumeurs légers, fumeurs moyens et gros fumeurs fournis par l'enquête nationale sont rapportés dans la ligne supplémentaire du tableau.

Enfin, l'échantillon d'enquête sur l'entreprise a été classé en deux groupes suivant qu'une personne consomme des boissons alcoolisées ou non. Les résultats sont présentés dans les deux dernières colonnes du tableau.

Voici les diagrammes des profils des données actives. Commenter.



Donner l'interprétation des résultats numériques et graphiques pour l'AFC de cette enquête.

- AFC de Fumeurs -

Inertie totale = 0.0851899

nb de lignes = 5

nb de colonnes = 4

effectif total = 193

D2 d'independance = 16.44165 , d.d.l. = 12 , Chi2 critique = 21.02607 (0.05)

	val.pro.	% inert.	% cumul.
1	0.07476	87.76	87.76
2	0.01002	11.76	99.51
3	0.00041	0.49	100.00
4	0.00000	0.00	100.00

Axes principaux

	a1	a2	a3
No	-0.39330844	0.030492161	0.0008904801
Li	0.09945589	-0.141064311	-0.0219980424
ME	0.19632095	-0.007359155	0.0256590885
He	0.29377603	0.197765591	-0.0262108345

Composantes principales

	c1	c2	c3
SM	-0.06576826	0.19373722	-0.070981088
JM	0.25895861	0.24330479	0.033705222
SE	-0.38059494	0.01065995	0.005155762
JE	0.23295191	-0.05774398	-0.003305374
S	-0.20108920	-0.07891130	0.008081082

Contributions absolues (x 10000) des 5 modalités pour les 3 prem. compo.

	CTA1	CTA2	CTA3
SM	33	2136	6943
JM	837	5512	2562
SE	5120	30	170
JE	3310	1518	120
S	701	805	205

Contributions relative (x 10000) des 5 modalités pour les 3 prem. compo.

	COS1	COS2	COS3
SM	922	8003	1074
JM	5264	4647	89
SE	9990	8	2
JE	9419	579	2
S	8653	1333	14

Contributions absolues (x 10000) des 4 modalités pour les 3 premiers axes

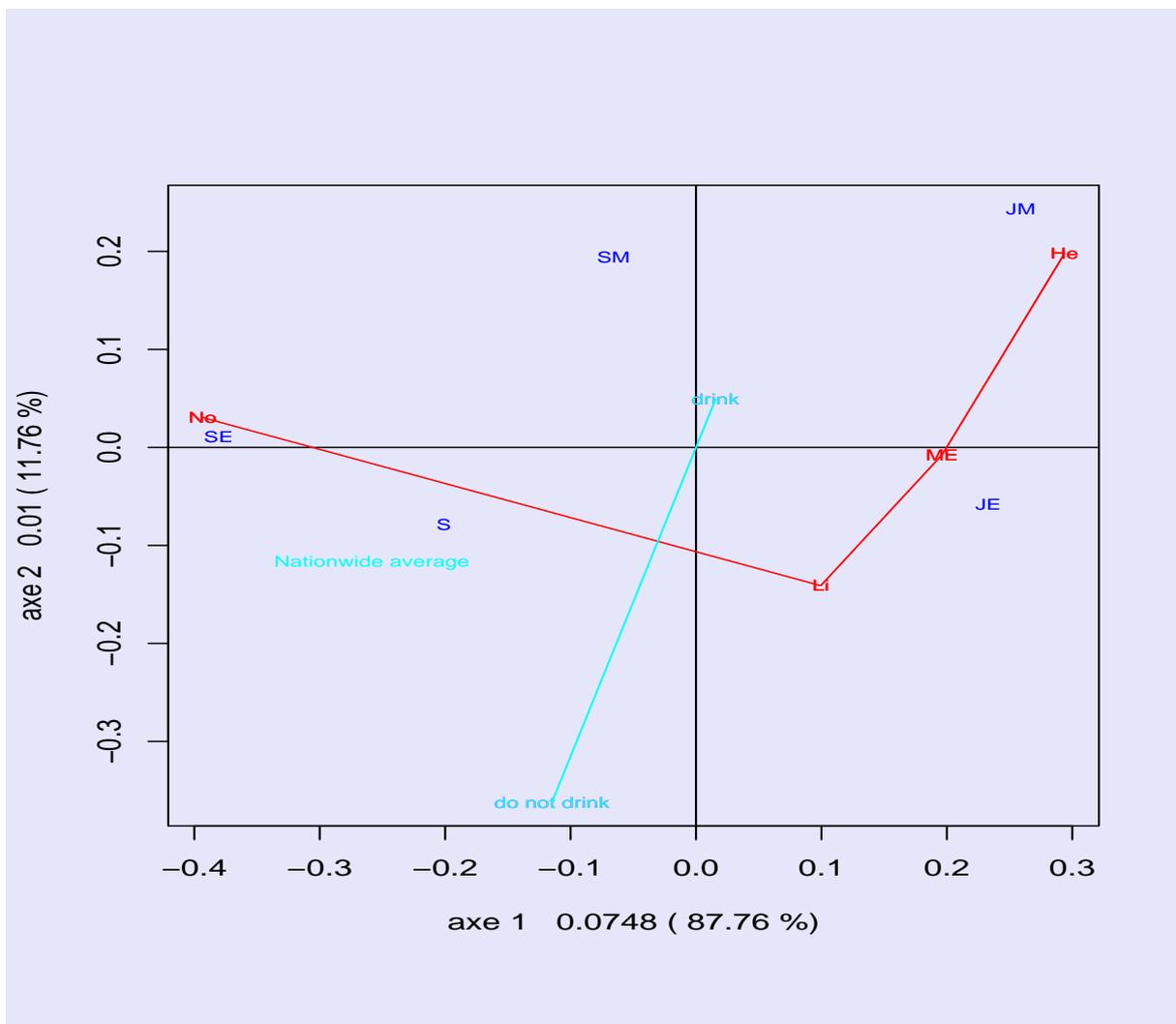
	CTA1	CTA2	CTA3
No	6540	293	6
Li	308	4632	2728

ME 1656 17 5114

He 1495 5058 2152

Contributions relative (x 10000) des 4 modalités pour les 3 premiers axes

	COS1	COS2	COS3
No 9940	60	0	
Li 3267	6573	160	
ME 9818	14	168	
He 6844	3102	54	



A l'aide des formules de transition, calculer la position du profil ligne et des profils colonnes supplémentaires sur ce plan factoriel. Enrichir l'analyse par l'apport de ces éléments supplémentaires.

8.5.2 Analyse Factorielle des Correspondances Multiples

Exercice 1 : Le centrage des modalités “à l’intérieur” d’une variable q

On appelle tableau de Burt associé à la super-matrice de codage U , la matrice $B = U'U$.

1. Quelles sont les propriétés des blocs de cette matrice indicés par les numéros des variables ?
2. Calculer en fonction de B , l’opérateur $\mathbb{V}M$ basé sur la matrice $X = D_I^{-1}FD_J^{-1}$ non centrée. Soit A^α un vecteur propre non trivial de $\mathbb{V}M$ associé à λ_α . Calculer $A^\alpha(q)$, le vecteur bloc colonne q de A^α , de dimension m_q , en fonction du bloc ligne B_q de B . En déduire que

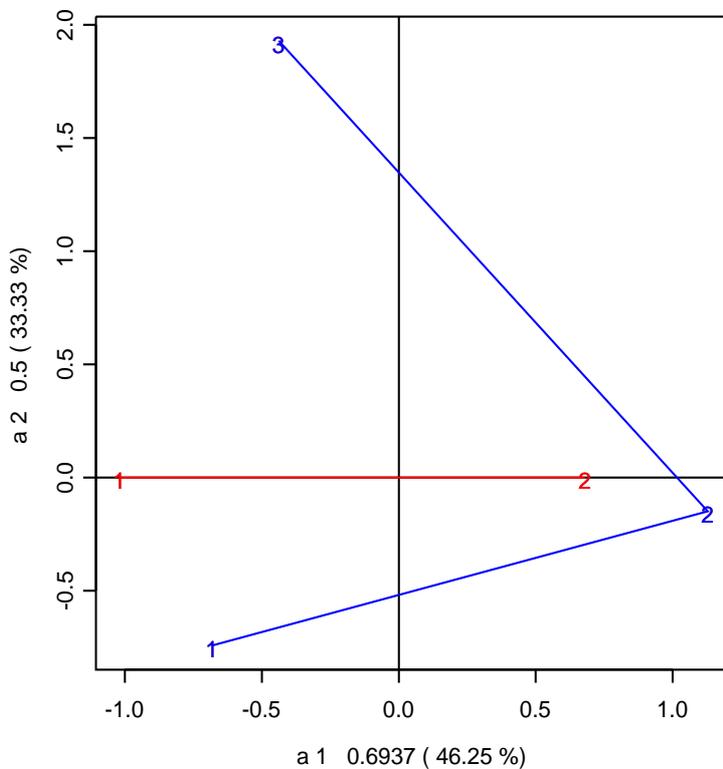
$$\mathbf{1}_{m_q}' D_q A^\alpha(q) = 0,$$

où $D_q = \text{diag}(\frac{U_{.1}^q}{N}, \dots, \frac{U_{.m_q}^q}{N})$ est la matrice diagonale des fréquences de q .

Exercice 2 : Deux variables qualitatives T^1 et T^2 ont été observées sur 20 individus. La première a trois modalités, la seconde deux. Le résultat des observations est donné dans la table suivante

Individus	T^1	T^2	Individus	T^1	T^2
1	1	2	11	1	1
2	2	2	12	2	2
3	1	2	13	3	2
4	1	1	14	3	1
5	1	1	15	1	2
6	2	2	16	1	2
7	3	1	17	1	1
8	1	1	18	2	2
9	2	2	19	2	1
10	3	2	20	2	2

1. Construire la matrice de codage disjonctif complet sur laquelle a été mise en oeuvre une AFCM dont les résultats numériques partiels associés aux variables, sont présentés en Annexe. On constate que seulement trois valeurs propres sont non nulles. Pouvait-on prévoir ce résultat ?
2. Comparer le total des contributions absolues des deux variables. Commenter.
3. La représentation des modalités de la seconde variable est-elle bonne dans le plan (1, 2) ? Donner une brève interprétation de la représentation des variables dans ce plan factoriel. Déduire la représentation des individus dans ce plan.



ANNEXE :

=====

	val.prop.	%I	%Icum.
1	0.69373	46.25	46.25
2	0.50000	33.33	79.58
3	0.30627	20.42	100.00
4	0.00000	0.00	100.00
5	0.00000	0.00	100.00

Contributions absolues (x 10000) des 5 modalités pour les 3 premiers axes

	CTA1	CTA2	CTA3
1	1511	2478	1511
2	3211	77	3211
3	278	7445	278
1	3000	0	3000
2	2000	0	2000

Contributions relative (x 10000) des 5 modalités pour les 3 premiers axes

	COS1	COS2	COS3
1	3812	4505	1683
2	6855	119	3026
3	481	9306	212
1	6937	0	3063
2	6937	0	3063

Chapitre 9

La régression Partial Least-Squares linéaire

Due à Herman Wold et à Svante Wold, la régression Partial Least-Squares, en abrégé PLS, est apparue dans les années 70. Dans ce chapitre, la régression PLS est présentée d'une façon différente de celle usuellement pratiquée dans la littérature statistique où l'approche algorithmique est le plus souvent développée et expliquée, voir par exemple [12, Tenenhaus], avec pour objectif le traitement des données manquantes et aussi pour "coller" le plus possible aux logiciels développés dans le commerce. Notre objectif est de situer PLS dans le contexte et les notations des chapitres précédents dédiés à l'analyse du triplet (X, M, D) avec ses propriétés de dualité. Cette optique permet la mise en évidence d'une représentation factorielle des individus, nouvelle à ce jour pour PLS, qui a pour avantage d'être une représentation "exacte", c'est à dire basée sur une projection du nuage des individus, et interprétable car en relation, certes non duale mais explicitée et mesurée, avec l'autre représentation factorielle, celle des variables explicatives et des réponses. Il en résulte, de façon pratique, un nouveau critère pour le choix du nombre de composantes qui complète les critères existants, ainsi que des aides à l'interprétation pour mesurer la qualité de représentation des individus. La version 9.9 de la fonction *pls()* programmée par l'auteur dans le langage Splus, contient la mise en oeuvre numérique et graphique des résultats présentés dans ce chapitre.

Nous avons vu jusqu'à présent des méthodes pour l'analyse factorielle "exploratoire" d'un seul tableau de données. Ce chapitre procède d'un contexte de régression, c'est à dire que le statisticien est en présence de deux jeux de variables, les premières (x^1, \dots, x^p) appelées "explicatives", les secondes (y^1, \dots, y^q) "à expliquer", toutes mesurées sur les mêmes n individus dont les poids statistiques seront $1/n$ sauf avis contraire, mais usuel-

lement stockés, comme dans les chapitres précédents, dans la diagonale de la matrice D . On note X , $n \times p$, et Y , $n \times q$, les matrices des observations sur les variables centrées réduites. On utilisera les néologismes “prédicteurs” pour désigner les premières variables et “réponses” pour les secondes et l’objectif est d’expérimenter un **modèle linéaire de prédiction des réponses par les prédicteurs basé sur la construction de composantes principales qui résument les prédicteurs**. L’ajustement de la réponse y^j est alors donné par

$$\hat{y}^j(k) = \hat{\beta}_0^j(k) + \hat{\beta}_1^j(k)x^1 + \dots + \hat{\beta}_p^j(k)x^p, \quad (9.1)$$

où les coefficients $\{\hat{\beta}_i^j(k)\}_i$ et donc la réponse estimée $\hat{y}^j(k)$, dépendent du nombre k de composantes principales utilisées.

On verra que cette approche factorielle de la régression offre une alternative intéressante à la régression linéaire multiple usuelle dans le cas, par exemple, où l’on désire conserver dans le modèle tous les prédicteurs bien que certains d’entre eux soient fortement corrélés, ou encore lorsque l’on dispose de peu d’observations par rapport au nombre, parfois très grand, de variables explicatives (plusieurs centaines dans le cas des problèmes de calibration en spectroscopie proche de l’infrarouge). Deux méthodes factorielles de régression sont en compétition : la Régression sur Composantes Principales ou RCP, et la régression PLS, encore appelée en anglais, “Projections onto Latent Structures”. La méthode RCP sera brièvement présentée ainsi que les motivations pour les régressions factorielles.

9.1 Motivations pour les régressions factorielles

Considérons pour simplifier le cas $q = 1$ d’une seule réponse et le modèle linéaire

$$Y = X\beta + \varepsilon.$$

Sous les hypothèses $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ et X de plein rang colonne, l’estimateur aux moindres carrés de β , obtenu par la régression linéaire multiple, est

$$\hat{\beta} = (X'X)^{-1}X'Y$$

qui est sans biais, $E(\hat{\beta}) = \beta$. La matrice X étant centrée réduite, il est possible d’interpréter les coefficients $\beta = [\beta_1, \dots, \beta_p]'$ sous réserve d’une faible variance

$$\text{var}(\hat{\beta}_i) = \sigma^2[(X'X)^{-1}]_{ii} = \frac{\sigma^2}{1 - R_{(-i)}^2}, \quad \forall i = 1, \dots, p,$$

où $R_{(-i)}^2$ est le coefficient de détermination de la régression de la variable explicative x^i sur les autres variables explicatives, voir Exercice 1.

L'erreur quadratique moyenne ou Mean Squared Error,

$$MSE = E(\|\beta - \hat{\beta}\|_2^2) = \sigma^2 \operatorname{trace}[(X'X)^{-1}] = \sigma^2 \sum_i \frac{1}{1 - R_{(-i)}^2}$$

est donc grande lorsque certaines variables explicatives sont fortement corrélées. Cela signifie que dans ce cas, l'estimateur aux moindres carrés, noté $\hat{\beta}_{OLS}$, OLS pour "Ordinary Least-Squares", est très imprécis.

Le domaine d'application de la régression linéaire multiple exclue les cas

- la matrice des covariances n'est pas inversible (X n'est pas de plein rang colonne),
- certains prédicteurs sont très fortement corrélés.

Pour contourner ces difficultés, appelées problème de la "multicolinéarité", une solution possible est d'utiliser une régression linéaire pas à pas qui éliminera du modèle un certain nombre de variables explicatives. Notre objectif est ici de **conserver toutes les variables** et la façon de procéder est d'effectuer la régression des réponses sur des variables latentes non corrélées construites sur les variables explicatives. C'est ce que l'on appelle des méthodes factorielles de régression. Un des avantages de ces méthodes, outre de construire **des modèles robustes face au problème de la multicolinéarité** des prédicteurs, est de proposer aussi un outil exploratoire des données grâce aux plans factoriels construits sur les composantes principales.

9.2 La régression sur composantes principales

Comme son nom l'indique la RCP consiste dans un premier temps à effectuer une ACP usuelle sur la matrice X des variables explicatives centrées réduites pour régresser ensuite une réponse, on supposera $q = 1$, sur les composantes principales retenues. Soit $C(k) = [C^1 \dots C^k]$ la matrice dont les colonnes sont formées de k composantes principales et $V(k) = [V^1 \dots V^k]$ celle des vecteurs propres de la matrice \mathbb{V} associés à ces composantes principales. Le modèle de la RCP s'écrit

$$Y = C(k)\alpha(k) + \varepsilon.$$

L'estimateur RCP des coefficients du modèle devient

$$\hat{\alpha}(k) = (C(k)'C(k))^{-1}C(k)'Y \quad \text{et} \quad \hat{Y}(k) = C(k)\hat{\alpha}(k).$$

L'orthogonalité des composantes donne pour $\hat{\alpha}(k) = [\hat{\alpha}(k)_1, \dots, \hat{\alpha}(k)_k]'$

$$\hat{\alpha}(k)_i = \frac{C^{i'}Y}{n\lambda_i} = \frac{\text{cov}(Y, C^i)}{\lambda_i}, \quad i = 1, \dots, k.$$

Lorsque les composantes principales ont une interprétation naturelle, il est peut être souhaitable de conserver ce modèle. Cependant, La formule de transition $C(k) = XV(k)$ permet de retrouver le modèle en les variables explicatives. L'estimateur du modèle en les prédicteurs s'écrit

$$\hat{\beta}(k) = V(k)\hat{\alpha}(k) \quad \text{et} \quad \hat{Y}(k) = X\hat{\beta}(k).$$

On exprime simplement le carré du coefficient de corrélation entre Y et C^i ,

$$r^2(Y, C^i) = \frac{\lambda_i(\hat{\alpha}(k)_i)^2}{\text{var}(Y)}.$$

L'orthogonalité des composantes principales permet la décomposition de la variance de $\hat{Y}(k)$

$$\text{var}(\hat{Y}(k)) = \sum_{i=1}^k (\hat{\alpha}(k)_i)^2 \lambda_i$$

ce qui permet d'exprimer la proportion de la variance de Y expliquée par le modèle à k composantes

$$R^2(Y, \text{Im } C(k)) = \frac{\text{var}(\hat{Y}(k))}{\text{var}(Y)} = \sum_{i=1}^k r^2(Y, C^i).$$

L'inconvénient de l'approche RCP vient de ce que les k plus grandes composantes, celles de plus grande variance, ne sont pas forcément celles qui expliquent le mieux la réponse Y . On est amené à sélectionner les composantes selon un autre critère, celui de la plus forte corrélation d'une composante avec la réponse. Cette démarche n'est pas bien pratique en particulier lorsque l'on cherche un même groupe de composantes qui puisse prédire conjointement plusieurs réponses (régression multiréponses, $q > 1$). C'est pour cette raison que la régression PLS est préférée à la RCP, car elle sélectionne automatiquement les composantes principales les plus explicatives des réponses.

9.3 Le contexte et le modèle PLS

La régression PLS a pour objectif de pallier les inconvénients de la RCP. Les composantes principales permettant de reconstruire l'espace des prédicteurs, sont construites dans le même temps que des régressions "partielles" sont effectuées ce qui conduit à une meilleure approximation des réponses.

On suppose que les n individus fournissant les observations sur les p variables explicatives et des q réponses sont munis des poids statistiques stockés dans la diagonale de la matrice $D = \text{diag}(d_1, \dots, d_n)$. On suppose que la matrice $X = [X^1 \dots X^p]$, $n \times p$ des observations sur les variables explicatives ainsi que celle des observations sur les réponses $Y = [Y^1 \dots Y^q]$, $n \times q$, sont D -centrées en colonnes

$$\mathbf{1}_n'DX = 0_p \quad \text{et} \quad \mathbf{1}_n'DY = 0_q.$$

Si les variables sont D -centrées réduites, la matrice, $\mathbb{V} = X'DX$, des covariances entre les prédicteurs est la matrice des corrélations et la variance totale vaut dans ce cas

$$\text{trace}(\mathbb{V}) = \sum_{i=1}^p \text{var}(X^i) = p.$$

La version matricielle sur variables D -centrées, du modèle PLS (9.1), à k composantes, est, pour la j ième réponse,

$$\hat{Y}^j(k) = X\hat{\beta}^j(k) = \hat{\beta}_1^j(k)X^1 + \dots + \hat{\beta}_p^j(k)X^p. \quad (9.2)$$

Nous verrons par la suite qu'une propriété attractive de PLS, comme pour la RCP d'ailleurs, est que lorsque $k = \text{rang}(X)$, on retrouve le modèle classique de la régression linéaire multiple

$$\text{si } k = \text{rang}(X), \quad \hat{\beta}^j(k) = \hat{\beta}_{OLS}^j.$$

Une autre propriété intéressante est que PLS fait le lien entre le modèle linéaire de régression et l'analyse exploratoire, au sens où l'ACP usuelle peut être considérée comme l'auto-régression PLS de X sur lui même, ce qui se résume par la formule

$$PLS(X, Y = X) \equiv ACP(X).$$

9.4 L'algorithme PLS

On note $X_{(0)} = X$ et $Y_{(0)} = Y$ les matrices de l'échantillon des variables explicatives et des réponses D -centrées. Le plus souvent ces variables sont standardisées. On supposera qu'il n'y a pas de données manquantes pour une exposition plus claire de la méthode.

L'algorithme PLS calcule les composantes t^1, \dots, t^k étape par étape, l'étape i ($i = 1, \dots, k$) permet de construire t^i et d'effectuer deux "régressions partielles" sur cette nouvelle variable.

Initialisation	$X_{(0)} = X, \quad Y_{(0)} = Y$	
Etape i	1) Construction de t^i à $X_{(i-1)}$ et $Y_{(i-1)}$ fixés	$t = X_{(i-1)}w, \quad u = Y_{(i-1)}v \quad (9.3)$ $(w^i, v^i) = \arg \max_{w'w=v'v=1} cov(t, u) \quad (9.4)$ $t^i = X_{(i-1)}w^i, \quad u^i = Y_{(i-1)}v^i \quad (9.5)$
$i = 1, \dots, k$	2) Régressions partielles, actualisation de $X_{(i)}$ et $Y_{(i)}$	$X_{(i)} = X_{(i-1)} - \Pi_{t^i}^D X_{(i-1)} \quad (9.6)$ $Y_{(i)} = Y_{(i-1)} - \Pi_{t^i}^D Y_{(i-1)}. \quad (9.7)$

Notons que bien que u^i soit, comme t^i , un compromis linéaire de variables, on réserve le nom de composante principale uniquement à t^i .

La partie 2) nécessite un commentaire immédiat. Il s'agit de régressions sur la variable t^i , car $\Pi_{t^i}^D$ est la matrice (de rang 1) de la projection D -orthogonale sur cette variable

$$\Pi_{t^i}^D = \frac{1}{\|t^i\|_D^2} t^i t^{i'}. \quad (9.8)$$

D'après (9.6) et (9.7), la matrice $X_{(i)}$ (respect. $Y_{(i)}$), matrice dont les colonnes sont les prédicteurs (réponses) actualisés, est construite de la façon suivante : **chaque variable actualisée (i) est le résidu de la régression de son homologue ($i - 1$) sur la variable t^i .**

Deux matrices vont jouer un rôle clé dans la méthode PLS, celle, $n \times k$, qui stocke en colonnes les k composantes successivement construites,

$$T(k) = [t^1 \dots t^k],$$

et celle, $p \times k$, dont les colonnes sont vecteurs des poids correspondants

$$W(k) = [w^1 \dots w^k].$$

9.4.1 Le centrage des variables

P1 : Pour $i = 1, \dots, k$, les variables actualisées $X_{(i)}$ et $Y_{(i)}$, les composantes t^i et les variables u^i sont D -centrées. \square

Preuve : Montrons le par récurrence, seulement “du côté des X ”. Supposons les colonnes de $X_{(i-1)}$ D -centrées, ce qui est vrai pour $X_{(0)} = X$, alors $t^i = X_{(i-1)}w$ est D -centrée car $\mathbf{1}_n' D t^i = \mathbf{1}_n' D X_{(i-1)}w = 0$. Ce qui implique que les colonnes de $X_{(i)}$ sont aussi centrées car, d’après (9.6) et (9.8), $\mathbf{1}_n' D X_{(i)} = \mathbf{1}_n' D X_{(i-1)} - \mathbf{1}_n' D t^i t^{i'} X_{(i)} / \|t^i\|_D^2 = 0_p$. \square

Comme conséquence, toutes ces variables appartenant à (\mathbb{R}^n, D) ont pour covariances et variances respectivement les produits scalaires et les normes Euclidiennes associées.

Ainsi, pour $i = 1, \dots, k$,

$$\begin{aligned} - \text{var}(t^i) &= \|t^i\|_D^2 \text{ et } \text{var}(u^i) = \|u^i\|_D^2, \\ - \mathbb{V}_{(i)}^{xx} &= X_{(i)}' D X_{(i)}, \quad \mathbb{V}_{(i)}^{yy} = Y_{(i)}' D Y_{(i)} \text{ et } \mathbb{V}_{(i)}^{xy} = X_{(i)}' D Y_{(i)} = (\mathbb{V}_{(i)}^{yx})' \end{aligned} \quad (9.9)$$

sont les matrices des covariances entre variables actualisées, respectivement à l’intérieur des prédicteurs, à l’intérieur des réponses et entre les prédicteurs et les réponses.

Bien sûr, est aussi D -centrée toute combinaison linéaire des prédicteurs actualisés ainsi que des réponses actualisées. D’après (9.3), $\bar{t} = \bar{u} = 0$ et si l’on suppose que les matrices des variables actualisées sont de plein rang colonne,

$$\text{var}(t) = \|t\|_D^2 = \|w\|_{\mathbb{V}_{(i-1)}^{xx}}^2, \quad \text{var}(u) = \|u\|_D^2 = \|v\|_{\mathbb{V}_{(i-1)}^{yy}}^2, \quad \text{cov}(t, u) = t' D u = w' \mathbb{V}_{(i-1)}^{xy} v.$$

Il faut noter que dans de nombreuses applications, les matrices des variables explicatives ne sont pas de plein rang colonne quand, par exemple, il y a plus de variables que d’individus ce qui rend structurellement $\mathbb{V}_{(i)}^{xx}$ semi définie positive pour tout i .

9.4.2 Construction de la composante t^i

Le critère du problème d’optimisation (9.4), est la covariance entre $t = X_{(i-1)}w$ et $u = Y_{(i-1)}v$, compromis linéaires des variables $(i-1)$ (à l’étape 1, ce sont des compromis des variables naturelles ou initiales)

$$\text{cov}(t, u) = \sigma(t)\sigma(u)r(t, u) = \|t\|_D \|u\|_D \cos((t, u)).$$

C’est la fonction de $\mathbb{R}^p \times \mathbb{R}^q$ dans \mathbb{R}

$$(w, v) \longrightarrow \varphi(w, v) = w' \mathbb{V}_{(i-1)}^{xy} v = v' \mathbb{V}_{(i-1)}^{yx} w,$$

et l’ensemble compact des contraintes est le produit cartésien des deux sphères unités

$$\{(w, v) \in (\mathbb{R}^p, I_p) \times (\mathbb{R}^q, I_q) \mid w'w = 1 \text{ et } v'v = 1\}$$

des espaces Euclidiens (\mathbb{R}^p, I_p) et (\mathbb{R}^q, I_q) .

Construisons la fonction de Lagrange de $\mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R} \times \mathbb{R}$ dans \mathbb{R} , notée L ,

$$(w, v, \lambda, \mu) \longrightarrow L(w, v, \lambda, \mu) = \varphi(w, v) + \frac{\lambda}{2}(1 - w'w) + \frac{\mu}{2}(1 - v'v),$$

où λ et μ sont les multiplicateurs de Lagrange associés aux deux contraintes. Le problème (9.4) est équivalent à la maximisation de $L(w, v, \lambda, \mu)$ sans contrainte dont les équations aux dérivées partielles, ou équations normales, s'écrivent

$$\left\{ \begin{array}{l} \nabla_w L = \mathbb{V}_{(i-1)}^{xy} v - \lambda w = 0_p \quad (e.1) \\ \nabla_v L = \mathbb{V}_{(i-1)}^{yx} w - \mu v = 0_q \quad (e.2) \end{array} \right.$$

$$\left\{ \begin{array}{l} 2 \frac{dL}{d\lambda} = 1 - w'w = 0 \quad (e.3) \end{array} \right.$$

$$\left\{ \begin{array}{l} 2 \frac{dL}{d\mu} = 1 - v'v = 0 \quad (e.4) \end{array} \right.$$

Les équations (e.1) et (e.2) donnent les formules de transition, à l'étape i , entre les vecteurs w et v . Les solutions de ce système fournissent les points critiques ou stationnaires du problème (9.4). Il faudra sélectionner ceux qui donnent un maximum.

Calcul et interprétation des multiplicateurs :

Multiplions (e.1) à gauche par le vecteur ligne w' et utilisons (e.3), de même, multiplions (e.2) à gauche par le vecteur ligne v' et utilisons (e.4). Il vient

$$\lambda = \mu = w' \mathbb{V}_{(i-1)}^{xy} v = v' \mathbb{V}_{(i-1)}^{yx} w = \text{cov}(t, u).$$

Les multiplicateurs de Lagrange sont égaux à la valeur de la fonction objectif à maximiser dans le problème d'optimisation (9.4). On suppose maintenant que $\lambda = \mu \neq 0$.

Calcul de w et de v :

Multiplions (e.1) par λ , et remplaçons dans cette équation λv par son expression puisée dans (e.2). Faisons de même en dualité sur (e.2). Il vient

$$\mathbb{V}_{(i-1)}^{xy} \mathbb{V}_{(i-1)}^{yx} w = \lambda^2 w \quad (9.10)$$

$$\mathbb{V}_{(i-1)}^{yx} \mathbb{V}_{(i-1)}^{xy} v = \lambda^2 v. \quad (9.11)$$

Si l'on note $\mathbb{W}_{(i-1)}^x = X_{(i-1)} X_{(i-1)}'$ et $\mathbb{W}_{(i-1)}^y = Y_{(i-1)} Y_{(i-1)}'$, t et u définis par (9.3), vérifient

$$\mathbb{W}_{(i-1)}^x D \mathbb{W}_{(i-1)}^y D t = \lambda^2 t \quad (9.12)$$

$$\mathbb{W}_{(i-1)}^y D \mathbb{W}_{(i-1)}^x D u = \lambda^2 u. \quad (9.13)$$

Il suffit pour cela de multiplier respectivement (9.10) et (9.11) par $X_{(i-1)}$ et par $Y_{(i-1)}$. On peut maintenant, énoncer la proposition donnant la solution de la partie 1) de la i ème étape de l'algorithme PLS.

Proposition 9.1 : Le triplet $(\lambda_i = cov(t^i, u^i), w^i, v^i)$ solution du problème (9.4) est donné par celui associé à **la plus grande valeur singulière** dans la décomposition en valeurs singulières de la matrice $\mathbb{V}_{(i-1)}^{xy} = X_{(i-1)}'DY_{(i-1)}$. \square

Remarque : Dans la pratique on ne recherche la plus grande valeur propre que pour une seule des équations (9.10), (9.11), celle de plus faible dimension, et on calcule l'autre vecteur solution par la formule de transition (e.1) ou (e.2) adéquate.

Dans le cas d'une seule réponse ($q = 1$), historiquement appelé PLS1 par opposition à PLS2 qui correspond au cas multi-réponses, (9.11) devient triviale, $v = 1$ avec une seule valeur propre

$$\lambda_i = \sqrt{\sum_{j=1}^p cov^2(Y_{(i-1)}, X_{(i-1)}^j)}.$$

À l'étape 1, on a aussi dans ce cas, $u^1 = Y_{(0)} = Y$.

9.4.3 Les régressions partielles

Visitons maintenant la partie 2) de l'algorithme PLS, celle des régressions partielles sur la composante $t^i = X_{(i-1)}w^i$ précédemment calculée dans la partie 1). Notons $X_{(i)} = [X_{(i)}^1 \dots X_{(i)}^p]$ et $Y_{(i)} = [Y_{(i)}^1 \dots Y_{(i)}^q]$ les matrices dont les colonnes sont les variables actualisées à l'étape i .

On appelle régression partielle de l'étape i , la régression simple d'une variable actualisée de l'étape $i - 1$ sur la composante t^i .

On note $\hat{X}_{(i)} = [\hat{X}_{(i)}^1 \dots \hat{X}_{(i)}^p]$ et $\hat{Y}_{(i)} = [\hat{Y}_{(i)}^1 \dots \hat{Y}_{(i)}^q]$ les matrices des modèles partiels de l'étape i

$$\hat{X}_{(i)} = \Pi_{t^i}^D X_{(i-1)}, \quad \hat{Y}_{(i)} = \Pi_{t^i}^D Y_{(i-1)}. \quad (9.14)$$

Notons :

$$p^i = (X_{(i-1)})' D t^i / \|t^i\|_D^2 = [\text{cov}(X_{(i-1)}^1, t^i) / \text{var}(t^i), \dots, \text{cov}(X_{(i-1)}^p, t^i) / \text{var}(t^i)]' \quad (9.15)$$

le vecteur des coefficients des régressions partielles des p variables explicatives et $P(k) = [p^1 \dots p^k]$ la matrice des vecteurs obtenus après k étapes ;

$$c^i = (Y_{(i-1)})' D t^i / \|t^i\|_D^2 = [\text{cov}(Y_{(i-1)}^1, t^i) / \text{var}(t^i), \dots, \text{cov}(Y_{(i-1)}^q, t^i) / \text{var}(t^i)]' \quad (9.16)$$

le vecteur des coefficients des régressions partielles des q variables réponses, et $C(k) = [c^1 \dots c^k]$ la matrice des vecteurs obtenus après k étapes.

Les matrices des variables estimées dans les régressions partielles de l'étape i , s'écrivent

$$\hat{X}_{(i)} = t^i p^{i'}, \quad \hat{Y}_{(i)} = t^i c^{i'}. \quad (9.17)$$

L'actualisation des variables à l'étape i , consiste à enlever l'information apportée par la composante t^i en prenant les résidus des régressions partielles

$$X_{(i)} = X_{(i-1)} - \hat{X}_{(i)}, \quad Y_{(i)} = Y_{(i-1)} - \hat{Y}_{(i)}. \quad (9.18)$$

Les expressions (9.17) se déduisent directement de (9.14) grâce à l'expression (9.8) du projecteur sur t^i .

La Figure 22 représente une régression partielle (simple) de l'étape i pour une variable seulement, par exemple la variable explicative $X_{(i-1)}^j$, du double point de vue de l'espace (\mathbb{R}^n, D) des variables (partie gauche) et de celui de l'espace \mathbb{R}^2 des individus (partie droite).

Projection d'une variable explicative actualisée

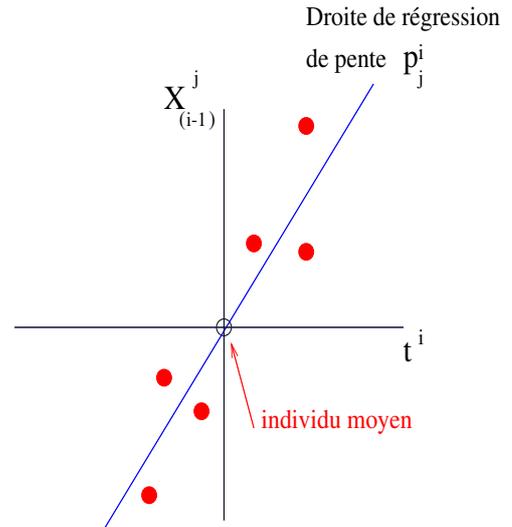
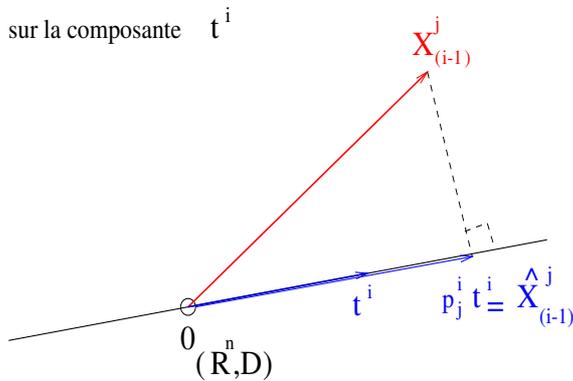


Figure 22 : Régression partielle, à l'étape i , de la variable $X_{(i-1)}^j$ sur la composante t^i .

Bien sûr, on peut présenter une figure analogue pour les régressions partielles “du côté des Y ”.

9.5 Première écriture des modèles PLS en les composantes

La régression PLS construit pas à pas, k composantes, t^1, \dots, t^k , où k , appelé la dimension du modèle, est le super-paramètre de la méthode qu'il faudra déterminer avec soin. Supposons connue, pour le moment, la dimension du modèle.

Examinons, tout d'abord, la propriété qui exprime que les variables actualisées s'expriment en fonction des variables initiales et des composantes.

P2 : Pour tout i variant de 1 à k ,

$$X_{(i)} = \Pi_{t^i}^{D\perp} \Pi_{t^{i-1}}^{D\perp} \dots \Pi_{t^1}^{D\perp} X, \quad Y_{(i)} = \Pi_{t^i}^{D\perp} \Pi_{t^{i-1}}^{D\perp} \dots \Pi_{t^1}^{D\perp} Y, \quad (9.19)$$

où $\Pi_{t^i}^{D\perp} = I_n - \Pi_{t^i}^D$ est le projecteur D -orthogonal sur le sous espace vectoriel de \mathbb{R}^n orthogonal à la droite vectorielle $Im t^i$ \square

Preuve : Il suffit d'écrire les matrices des résidus des régressions partielles (9.6) et (9.7) de proche en proche. \square

La régression PLS construit deux modèles à partir des composantes principales, l'un pour reconstruire X , l'autre pour expliquer et prédire Y .

D'abord, avec (9.6) et (9.5), pour $i = j - 1$,

$$\begin{aligned}
 t^{j-1'}DX_{(j)} &= t^{j-1'}D(X_{(j-1)} - \Pi_{t^j}^D X_{(j-1)}) \\
 &= -t^{j-1'}D\Pi_{t^j}^D X_{(j-1)} \\
 &= -t^{j-1'}Dt^j t^{j'}DX_{(j-1)}/var(t^j) \\
 &= -t^{j-1'}DX_{(j-1)}w^j t^{j'}DX_{(j-1)}/var(t^j) = 0_{1 \times p}.
 \end{aligned}$$

Ainsi de suite jusqu'à $t^1DX_{(j)} = 0_{1 \times p}$. \square

On ne montrera pas la preuve de la propriété suivante qui est un résultat accessoire pour la suite.

P4 : Pour $i \leq j$, $t^{i'}DY_{(j)} = 0_{1 \times q}$, \square

La propriété P3 a pour conséquence :

Proposition 9.2 Les composantes PLS sont deux à deux orthogonales,

$$t^{i'}Dt^j = 0, \quad \text{si } i < j. \quad (9.23)$$

Et , résultat accessoire,

$$t^{i'}Du^j = 0, \quad \text{si } i < j. \quad (9.24)$$

Preuve : Montrons seulement (9.23). Il résulte de la propriété P3 et de (9.5) que, pour $i \leq j$,

$$t^{i'}Dt^{j+1} = t^{i'}DX_{(j)}w^{j+1} = 0.$$

La variable latente t^{j+1} est donc non corrélée avec t^1, \dots, t^j . \square

La D -orthogonalité des composantes a des conséquences sur l'écriture des modèles. Rappelons d'abord, sans démonstration, quelques propriétés des projecteurs lorsque l'on dispose d'une base orthogonale.

P5 : Notons $\Pi_{T(k)}^D = T(k)[T(k)'DT(k)]^{-1}T(k)'D$ la matrice de la projection D -orthogonale sur le sous-espace vectoriel de \mathbb{R}^n , $Im\{t^1, \dots, t^k\}$, de dimension k , engendré par les composantes.

$$\Pi_{t^i}^D \Pi_{t^j}^D = 0_{n \times n}, \quad \text{si } i \neq j \quad (9.25)$$

$$\sum_{i=1}^k \Pi_{t^i}^D = \Pi_{T(k)}^D, \quad (9.26)$$

$$\Pi_{t^k}^{D\perp} \Pi_{t^{k-1}}^{D\perp} \dots \Pi_{t^1}^{D\perp} = I_n - \Pi_{T(k)}^D = \Pi_{T(k)}^{D\perp}. \quad (9.27)$$

\square

Comme conséquence de (9.27) et (9.26) la propriété P2 se réécrit.

P2 : Pour tout i variant de 1 à k ,

$$X_{(i)} = \Pi_{t^i}^{D\perp} \Pi_{t^{i-1}}^{D\perp} \dots \Pi_{t^1}^{D\perp} X = \Pi_{T^{(k)}}^D \perp X = X - \Pi_{T^{(k)}}^D X = X - \sum_{j=1}^i \Pi_{t^j}^D X \quad (9.28)$$

$$Y_{(i)} = \Pi_{t^i}^{D\perp} \Pi_{t^{i-1}}^{D\perp} \dots \Pi_{t^1}^{D\perp} Y = \Pi_{T^{(k)}}^D \perp Y = Y - \Pi_{T^{(k)}}^D Y = Y - \sum_{j=1}^i \Pi_{t^j}^D Y. \quad (9.29)$$

□

Les vecteurs des poids $\{w^i\}$ sont des éléments de l'espace Euclidien (\mathbb{R}^p, I_p) . Ils sont orthogonaux, d'après la proposition suivante, au sens habituel selon la métrique I_p .

Proposition 9.3 : Pour $j \leq i$, $w^{j'} X'_{(i)} = 0_{1 \times n}$, et, pour $j < i$, $w^{j'} w^i = 0$. □

Preuve : Faire l'exercice. □

La proposition suivante permet de préciser la nature de la matrice $P(k)'W(k)$ qui joue un rôle important le calcul des modèles PLS.

Proposition 9.4 : La matrice $P(k)'W(k)$, carrée d'ordre k , est triangulaire supérieure, à diagonale unité. □

Preuve : L'élément (i, j) de cette matrice est, d'après (9.15),

$$p^{i'} w^j = \frac{t^{i'} D X_{(i-1)} w^j}{\|t^i\|_D^2}.$$

Les éléments diagonaux sont donc égaux à 1 et la proposition 9.3 indique que si $i > j$, $X_{(i-1)} w^j = 0_{n \times 1}$ ce qui termine la démonstration. □

9.7 Écriture définitive des modèles PLS en les composantes

Revisitons, les régressions partielles et les modèles PLS en les composantes principales au vu de leur orthogonalité.

Les régressions partielles des variables actualisées donnent les mêmes résultats que les régressions des variables initiales sur les composantes.

En effet, à l'étape i , les régressions partielles (9.14) deviennent avec (9.25), (9.28) et (9.29)

$$\hat{X}_{(i)} = \Pi_{t^i}^D X_{(i-1)} = \Pi_{t^i}^D (I_n - \sum_{j=1}^{i-1} \Pi_{t^j}^D) X = \Pi_{t^i}^D X = t^i p^{i'}, \quad (9.30)$$

$$\hat{Y}_{(i)} = \Pi_{t^i}^D Y_{(i-1)} = \Pi_{t^i}^D (I_n - \sum_{j=1}^{i-1} \Pi_{t^j}^D) Y = \Pi_{t^i}^D Y = t^i c^{i'}. \quad (9.31)$$

Les vecteurs des coefficients des régressions partielles sont les mêmes que ceux des régressions des variables initiales sur la composante t^i

$$p^i = X' D t^i / \text{var}(t^i), \quad (9.32)$$

$$c^i = Y' D t^i / \text{var}(t^i). \quad (9.33)$$

La Figure 23 reprend le schéma de la Figure 22 et illustre, du double point de vue de l'espace (\mathbb{R}^n, D) des variables et de l'espace (\mathbb{R}^2, I_2) des individus, le fait que les deux régressions simples sur la composante t^i , de la variable naturelle X^j et de la variable actualisée $X_{(i-1)}^j$, donnent la même estimation p_j^i .

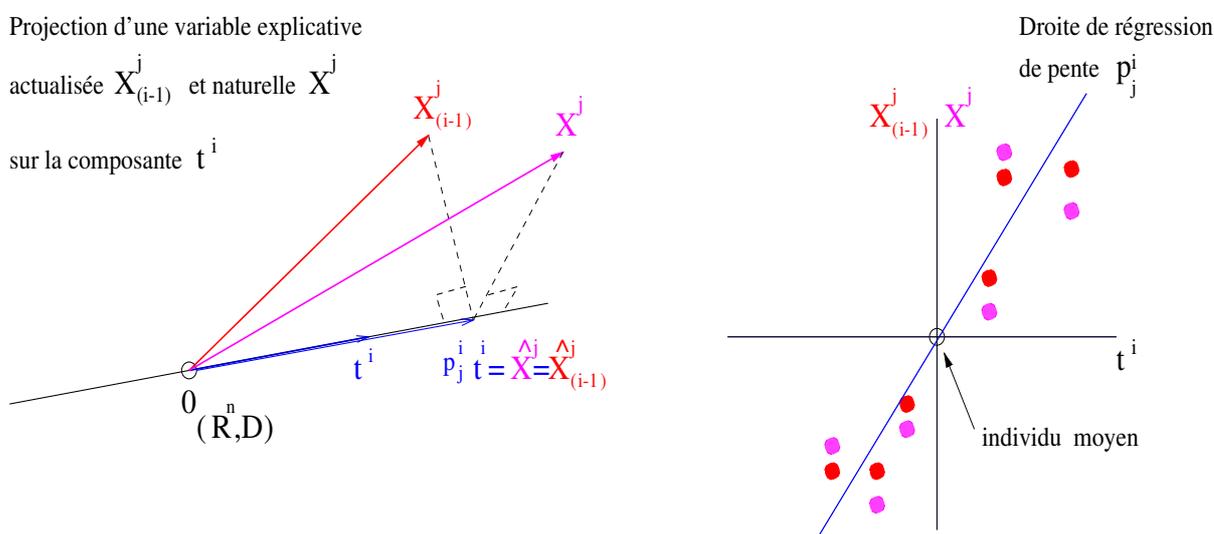


Figure 23 : La régression partielle, à l'étape i , de la variable $X_{(i-1)}^j$ et la régression de la variable naturelle X^j sur la composante t^i , donnent le même coefficient p_j^i .

Le même schéma est valable “du côté de Y ”, où c_j^i est le même coefficient dans les deux régressions simples de Y^j d'une part et de $Y_{(i-1)}^j$ d'autre part sur la composante t^i .

Grâce à (9.15) et (9.16), les matrices des covariances des variables actualisées définies en (9.9) s'écrivent

$$\mathbb{V}_{(i)}^{xx} = \mathbb{V}_{(i-1)}^{xx} - \|t^i\|_D^2 p^i p^{i'} \quad \text{et} \quad \mathbb{V}_{(i)}^{yy} = \mathbb{V}_{(i-1)}^{yy} - \|t^i\|_D^2 c^i c^{i'} \quad (9.34)$$

$$\mathbb{V}_{(i)}^{xy} = \mathbb{V}_{(i-1)}^{xy} - \|t^i\|_D^2 p^i c^{i'}. \quad (9.35)$$

Les modèles PLS, (9.20), (9.21), (9.22), en les k composantes s'expriment comme la régression linéaire multiple des variables naturelles sur ces composantes

$$X = \hat{X}(k) + X_{(k)}, \quad Y = \hat{Y}(k) + Y_{(k)}, \quad (9.36)$$

$$\hat{X}(k) = \sum_{i=1}^k \hat{X}_{(i)} = \Pi_{T(k)}^D X = t^1 p^{1'} + \dots + t^k p^{k'} = T(k)P(k)', \quad (9.37)$$

$$\hat{Y}(k) = \sum_{i=1}^k \hat{Y}_{(i)} = \Pi_{T(k)}^D Y = t^1 c^{1'} + \dots + t^k c^{k'} = T(k)C(k)', \quad (9.38)$$

où les matrices $P(k) = [p^1 \dots p^k]$ et $C(k) = [c^1 \dots c^k]$, définies en (9.15) et 9.16) se réécrivent

$$P(k) = X' D T(k) \text{diag}\left(\frac{1}{\text{var}(t^1)}, \dots, \frac{1}{\text{var}(t^k)}\right), \quad (9.39)$$

$$C(k) = Y' D T(k) \text{diag}\left(\frac{1}{\text{var}(t^1)}, \dots, \frac{1}{\text{var}(t^k)}\right). \quad (9.40)$$

Grâce à (9.34) et (9.35), les matrices des covariances sont modélisées en fonction des matrices des covariances actualisées qui jouent le rôle de résidus

$$X' D X = \sum_{i=1}^k \|t^i\|_D^2 p^i p^{i'} + \mathbb{V}_{(k)}^{xx} \quad (9.41.a)$$

$$Y' D Y = \sum_{i=1}^k \|t^i\|_D^2 c^i c^{i'} + \mathbb{V}_{(k)}^{yy} \quad (9.41.b)$$

$$X' D Y = \sum_{i=1}^k \|t^i\|_D^2 p^i c^{i'} + \mathbb{V}_{(k)}^{xy} \quad (9.41.c)$$

Pour que le modèle (9.38) soit réellement un modèle de régression sur les prédicteurs X , il reste à montrer que les composantes t^1, \dots, t^k appartiennent à l'espace $\text{Im } X$ engendré par les variables explicatives naturelles.

9.8 Les composantes PLS, compromis linéaires des variables explicatives initiales

Une des raisons qui a fait que la régression PLS a été considérée pendant longtemps comme scientifiquement mal établie, est qu'elle a été souvent présentée seulement du point de vue algorithmique. Ainsi il n'apparaissait pas de façon très claire comment les composantes, qui étaient des compromis linéaires des résidus des régressions partielles, étaient aussi des combinaisons linéaires des variables explicatives naturelles.

9.8.1 Expression du vecteur des poids

Par (9.5), la composante construite à l'étape i , $t^i = X_{(i-1)}w^i$, est une combinaison linéaire des variables actualisées à l'étape $i - 1$. Lorsque $i = 1$, t^1 est bien un compromis linéaire de $X_{(0)} = X$. Qu'en est il pour les étapes suivantes? la proposition suivante répond à cette question en construisant les vecteurs des poids par récurrence.

Proposition 9.5 : $Im T(k)$ est un sous-espace vectoriel de $Im X$, de dimension k . Plus précisément,

$$t^i = Xw^{*i}, \quad (9.42)$$

avec

$$w^{*1} = w^1, \quad \text{et} \quad w^{*i} = \left[I_p - \left(\sum_{j=1}^{i-1} \frac{w^{*j} w^{*j'}}{var(t^j)} \right) \mathbb{V} \right] w^i, \quad i > 1, \quad (9.43)$$

où $\mathbb{V} = X'DX$ est la matrice des covariances des variables explicatives initiales. \square

Preuve : De façon évidente, les composantes, deux à deux orthogonales, sont linéairement indépendantes et $Im T(k)$ est de dimension k . Montrons 9.43 par récurrence. Supposons que, pour $j = 1, \dots, i - 1$, $t^j = Xw^{*j}$. Alors (9.5) et (9.28) donnent

$$\begin{aligned} t^i &= X_{(i-1)}w^i = \left(X - \sum_{j=1}^{i-1} \frac{t^j t^{j'}}{var(t^j)} DX \right) w^i \\ &= X \left[I_p - \left(\sum_{j=1}^{i-1} \frac{w^{*j} w^{*j'}}{var(t^j)} \right) X'DX \right] w^i. \end{aligned}$$

L'hypothèse de récurrence est vérifiée pour $j = 1$, ce qui termine la preuve. \square

Corollaire : Si $k = rang(X)$, les composantes PLS permettent de reconstruire X car $Im T(k) = Im X$. En effet, $X_{(k)} = 0_{n \times p}$ dans (9.36) car, dans (9.37),

$$\hat{X}_{(k)} = \Pi_{T(k)}^D X = \Pi_X^D X = X.$$

Alors,

$$\hat{Y}(k) = \Pi_{T(k)}^D Y = \Pi_X^D Y = \hat{Y}_{OLS}.$$

Le modèle PLS de dimension $k = rang(X)$ coïncide avec le modèle de la régression linéaire multiple, sous réserve que cette dernière soit applicable.

$$PLS(X, Y) = OLS(X, Y), \quad \text{si} \quad k = rang(X). \quad (9.44)$$

\square

Ce corollaire ouvre la question du choix du nombre k de composantes. Il apporte une borne supérieure à l'étendue des choix possibles : la dimension k du modèle sera inférieure ou égale au rang de X .

9.8.2 Orthogonalité des vecteurs des poids au sens de \mathbb{V}

Grâce à (9.42), on va pouvoir représenter les individus de X dans des plans factoriels (w^{*i}, w^{*j}) si l'on peut montrer que la famille des $\{w^{*i}\}$ est orthogonale. C'est l'objectif de la proposition suivante.

Proposition 9.6 : Si $\text{rang}(X) = p$, deux vecteurs $\{w^{*i}\}$ distincts sont \mathbb{V} -orthogonaux et (9.43) s'écrit

$$w^{*1} = w^1, \quad \text{et} \quad w^{*i} = \Pi_{W^{*(i-1)}}^{\mathbb{V}\perp} w^i = [I_p - \Pi_{W^{*(i-1)}}^{\mathbb{V}}] w^i, \quad i > 1, \quad (9.45)$$

où $W^{*(i)} = [w^{*1} \dots w^{*i}]$ est la matrice des vecteurs de poids et $\Pi_{W^{*(i)}}^{\mathbb{V}}$ est le projecteur \mathbb{V} -orthogonal sur $\text{Im } W^{*(i)}$. \square

Preuve : La matrice \mathbb{V} , de rang p , symétrique définie positive, fournit une métrique sur l'espace \mathbb{R}^p . Les w^{*i} sont \mathbb{V} -orthogonaux car, d'après (9.42),

$$\langle w^{*i}, w^{*j} \rangle_{\mathbb{V}} = w^{*j\prime} \mathbb{V} w^{*i} = t^{j\prime} D t^i = \langle t^i, t^j \rangle_D = 0 \quad \text{si} \quad i \neq j.$$

Le projecteur \mathbb{V} -orthogonal sur la droite vectorielle $\text{Im } w^{*i}$ est la matrice

$$\Pi_{w^{*i}}^{\mathbb{V}} = \frac{w^{*i} w^{*i\prime}}{\text{var}(t^i)} \mathbb{V}.$$

D'après (9.43), si $i > 1$, w^{*i} s'écrit

$$w^{*i} = \left[I_p - \sum_{j=1}^{i-1} \Pi_{w^{*j}}^{\mathbb{V}} \right] w^i.$$

Enfin, soit $W^{*(i)} = [w^{*1} \dots w^{*i}]$ la matrice $p \times i$ des vecteurs des poids, alors, puisque les $\{w^{*j}\}$ sont \mathbb{V} -orthogonaux,

$$\Pi_{W^{*(k)}}^{\mathbb{V}} = \sum_{j=1}^k \Pi_{w^{*j}}^{\mathbb{V}}$$

et

$$w^{*i} = \Pi_{W^{*(i-1)}}^{\mathbb{V}\perp} w^i.$$

\square

Remarque : Très souvent dans les applications, la régression PLS est utilisée hors du contexte de la proposition 9.6, c'est à dire lorsque X n'est pas de plein rang colonne. Cela arrive, en particulier, lorsque $n < p$, ce qui est le cas, par exemple, pour des jeux de données issus de la calibration de spectres proche de l'infrarouge. La matrice des covariances est, dans ce cas, seulement semi définie positive, ce qui signifie que $\|\cdot\|_{\mathbb{V}}$ est une semi-norme

sur \mathbb{R}^p . Cependant, tant que les composantes t^1, \dots, t^k sont non nulles, c'est à dire tant que $k \leq r = \text{rang}(X) = \text{rang}(X')$, les vecteurs w^{*1}, \dots, w^{*k} produits par l'algorithme PLS sont \mathbb{V} -orthogonaux et dans $\text{Im } X'$ comme le montre la proposition 9.7. Le \mathbb{V} -produit scalaire sur l'espace $\text{Im } X'$ est défini par le produit scalaire de ses vecteurs de base

$$\text{pour } i, j \in \{1, \dots, r\}, \quad \langle w^{*i}, w^{*j} \rangle_{\mathbb{V}} = \langle t^i, t^j \rangle_D.$$

9.8.3 Propriétés des vecteurs des poids

Une composante s'exprime de deux façons possibles, $t^i = X_{(i-1)} w^i = X w^{*i}$. Les vecteurs w^{*i} des poids sont calculés de proche en proche, au fur et à mesure du déroulement de l'algorithme. La proposition suivante précise quelques propriétés de ces vecteurs en liaison avec les poids w^i et donne une formule explicite pour leur calcul.

Proposition 9.7 : Les matrices $W(k) = [w^1 \dots w^k]$, $W^*(k) = [w^{*1} \dots w^{*k}]$ et $P(k) = [p^1 \dots p^k]$, de dimensions $p \times k$, issues de la régression PLS vérifient les propriétés :

- a) $\Pi_{W^*(k)}^{\mathbb{V}} = W^*(k)P(k)'$
- b) $\text{Im } W(k) = \text{Im } W^*(k)$ est un sous-espace vectoriel de dimension k de $\text{Im } X'$
- c) $W^*(k) = W(k) [(P(k)'W(k))]^{-1}$.

□

Preuve :

a)

La matrice $P(k)$ s'exprime de deux façons en (9.15) et (9.39). C'est la seconde qui est utile ici,

$$W^*(k)P(k)' = \sum_{j=1}^k w^{*j} p^{j'} = \sum_{j=1}^k w^{*j} w^{*j'} \mathbb{V} / \|t^j\|_D^2 = \Pi_{W^*(k)}^{\mathbb{V}}.$$

b)

De façon évidente, la propriété est vraie pour $k = 1$ puisque $w^{*1} = w^1$.

Supposons la vraie pour $k - 1$, $\text{Im } W(k - 1) = \text{Im } W^*(k - 1)$. Alors, on a l'égalité des projecteurs

$$\Pi_{W^*(k-1)}^{\mathbb{V}} = \Pi_{W(k-1)}^{\mathbb{V}}$$

et (9.45) donne

$$w^{*k} = w^k - \Pi_{W(k-1)}^{\mathbb{V}} w^k,$$

ce qui implique que $w^{*k} \in \text{Im } W(k)$ et que les k vecteurs $\{w^{*1}, \dots, w^{*k}\}$ sont dans $\text{Im } W(k)$. Le fait que ces vecteurs soient linéairement indépendants, à cause de l'orthogonalité, implique $\text{Im } W(k) = \text{Im } W^*(k)$. La propriété, vraie pour $k = 1$, est donc vraie

pour k .

De façon évidente, tant que, pour $i = 1, \dots, k$, $t^i = Xw^*{}^i \neq 0$, alors $w^*{}^i$ n'est pas dans $\text{Ker } X = \{\text{Im } X'\}^\perp$. Il est donc dans $\text{Im } X'$. Les vecteurs w^i qui engendrent le même espace que celui engendré par les $w^*{}^i$, sont aussi dans $\text{Im } X'$. En conclusion, les familles orthogonales $\{w^1, \dots, w^k\}$ et $\{w^*{}^1, \dots, w^*{}^k\}$ engendrent le même espace de dimension k , sous-espace de $\text{Im } X'$.

c)

D'après (b), $W(k) = \Pi_{W^*(k)}^\vee W(k)$. La propriété (a) donne

$$W(k) = W^*(k)P(k)'W(k).$$

La proposition 9.4 a montré que $P(k)'W(k)$ est une matrice triangulaire supérieure à diagonale unité. Elle est donc inversible, ce qui termine la preuve. \square

On peut montrer, [12, Tenenhaus], que $P(k)'W(k)$ est en fait bidiagonale à droite, seuls sont non nuls les termes diagonaux égaux à 1 et les termes immédiatement à droite de ceux ci.

9.9 Le modèle de régression PLS en les variables explicatives

Notons X_b et Y_b les matrices des données brutes sur les variables initiales, $\bar{X}_b = \mathbf{1}_n' D X_b$ et $\bar{Y}_b = \mathbf{1}_n' D Y_b$ les individus moyens pour les variables explicatives et pour les variables réponses.

Soient $X_c = X_b - \mathbf{1}_n \bar{X}_b$ et $Y_c = Y_b - \mathbf{1}_n \bar{Y}_b$ les matrices des variables D -centrées. Les matrices X et Y utilisées dans PLS sont nécessairement centrées, éventuellement standardisées si de trop grandes disparités dans la mesure des variances des variables risquent de perturber l'interprétation des coefficients des modèles. On écrit,

$$X = X_c Q_x^{-1} \quad \text{et} \quad Y = Y_c Q_y^{-1},$$

où $Q_x = I_p$ et $Q_y = I_q$ si les variables sont seulement centrées,

$$Q_x = \text{diag}(\|X_c^1\|_D, \dots, \|X_c^p\|_D) \quad \text{et} \quad Q_y = \text{diag}(\|Y_c^1\|_D, \dots, \|Y_c^q\|_D),$$

matrices diagonales des écart-types, si les variables sont standardisées.

9.9.1 Le modèle sur variables centrées, éventuellement réduites

Le modèle de la régression PLS, exprimé en (9.36) et (9.38) en fonction des composantes,

$$Y = \hat{Y}(k) + Y_{(k)}, \quad (9.36)$$

$$\hat{Y}(k) = \sum_{i=1}^k \hat{Y}_{(i)} = \Pi_{T(k)}^D Y = t^1 c^{1'} + \dots + t^k c^{k'} = T(k)C(k)', \quad (9.38)$$

s'exprime simplement en fonction des variables explicatives puisque $T(k) = XW^*(k)$,

$$\hat{Y}(k) = X\hat{\beta}(k) \quad (9.46)$$

avec pour expression de la matrice $\hat{\beta}(k) = [\hat{\beta}^1(k) \dots \hat{\beta}^q(k)]$, $p \times q$, des coefficients

$$\hat{\beta}(k) = W^*(k)C(k)' = W(k)[P(k)'W(k)]^{-1}C(k)'. \quad (9.47)$$

La propriété (9.44) s'écrit pour les coefficients,

$$\hat{\beta}(k) = \hat{\beta}_{OLS} \quad \text{si} \quad k = \text{rang}(X).$$

Dans ce cas, la régression PLS multi-réponses est équivalente à q régressions linéaires multiples.

Notons que, généralement, le modèle obtenu pour une réponse n'est pas le même suivant que cette réponse est prédite séparément (PLS uni-réponse) ou simultanément avec d'autres réponses (PLS multi-réponses). Dans la pratique, plus de composantes sont nécessaires pour obtenir une même qualité d'ajustement dans le cas $q > 1$ et on effectue une régression PLS multi-réponses lorsque les q réponses à prédire simultanément sont fortement corrélées deux à deux.

9.9.2 Le modèle en les variables initiales

Le modèle défini par (9.36), (9.46) et (9.47), se transforme en les variables initiales pour s'écrire

$$Y_b = \hat{Y}_b(k) + \varepsilon \quad (9.48)$$

avec

$$\hat{Y}_b(k) = \mathbf{I}_n \hat{\alpha}(k) + X_b \hat{\beta}_b(k), \quad \text{et} \quad \varepsilon = Y_{(k)} Q_y, \quad (9.49)$$

et

$$\hat{\beta}_b(k) = Q_x^{-1} \hat{\beta}(k) Q_y, \quad \text{et} \quad \hat{\alpha}(k) = \bar{Y}_b - \bar{X}_b \hat{\beta}_b(k). \quad (9.50)$$

Le vecteur $\hat{\alpha}(k)$, $1 \times q$, donne les "ordonnées à l'origine" et la matrice $\hat{\beta}_b(k)$, $p \times q$, est la matrice des coefficients des variables naturelles pour les q modèles.

9.10 Régression PLS et Analyse en Composantes Principales usuelle

Nous allons étudier dans cette section une régression PLS multi-réponses très particulière dans laquelle la matrice des réponses est prise identique à la matrice des variables explicatives. Cette “auto-régression PLS” des prédicteurs sur eux mêmes est l’ACP usuelle

$$ACP(X) \equiv PLS(X, Y = X).$$

Rappelons que l’ACP usuelle est l’ACP du triplet $(X, M = I_p, D = n^{-1}I_n)$ dont les propriétés sont résumées dans le paragraphe 7.4.1. Nous supposons donc, que les individus sont munis des mêmes poids statistiques,

$$D = \frac{1}{n}I_n,$$

ce qui est le plus souvent le cas en régression PLS.

Proposition 9.8 : La régression $PLS(X, Y=X)$ est l’ACP usuelle de X puisque, pour $i = 1, \dots, k$, l’algorithme PLS de la section 9.4, conduit à :

$w^i = v^i = p^i = c^i = w^*{}^i$ est le vecteur propre de \mathbb{V} , associé à la i ème valeur propre, $\lambda_i = var(t^i)$, dans l’ordre décroissant et les matrices actualisées s’écrivent,
 $X_{(i)} = Y_{(i)} = X - t^1 w^{1'} - \dots - t^i w^{i'}$.

Les matrices des covariances actualisées sont données par

$$\mathbb{V}_{(i)}^{xx} = \mathbb{V}_{(i)}^{yy} = \mathbb{V} - \lambda_1 w^1 w^{1'} \dots - \lambda_i w^i w^{i'}.$$

□

Preuve : Examinons tout d’abord, l’étape $i = 1$ de l’algorithme. Puisque $Y = X$, la maximisation du critère de covariance conduit à $w^1 = v^1$ qui maximise la variance de $t = Xw$. Le vecteur w^1 est donc le premier axe factoriel et $t^1 = u^1$, la première composante principale de l’ACP. Alors, (9.32) et (9.33) s’écrivent

$$p^1 = c^1 = X'Dt^1/\lambda_1 = X'DXw^1/\lambda_1 = w^1.$$

Le (c) de la proposition 9.7 donne $w^*{}^1 = w^1/(p^{1'}w^1) = w^1/(w^{1'}w^1) = w^1$. L’actualisation des variables par (9.6) et (9.7), donne, grâce à (9.30) et (9.31), $X_{(1)} = Y_{(1)} = X - t^1 w^{1'}$. Les covariances actualisées (9.34) et (9.35) deviennent

$$\mathbb{V}_{(1)}^{xx} = \mathbb{V}_{(1)}^{yy} = \mathbb{V} - \lambda_1 w^1 w^{1'}.$$

Supposons la proposition vraie pour $j = 1, \dots, i-1$. L’algorithme (9.4) a pour solution, par (9.10) et (9.11), puisque $X_{(i-1)} = Y_{(i-1)}$, $w^i = v^i$ vecteur propre associé à la plus

grande valeur propre de la matrice des covariances déflatée $\mathbb{V}_{(i-1)}^{xx}$. Cette valeur propre, $\lambda_i = \text{var}(t_i)$, est la i ème valeur propre par ordre décroissant de \mathbb{V} , voir le paragraphe 7.2.2 qui présente aussi l'ACP du triplet comme une succession de régressions partielles. La formule (9.45) montre que $w^{*i} = w^i$ puisque d'après la proposition (9.3), w^i est orthogonal au sens usuel à tous ses précédents. Il est donc aussi \mathbb{V} -orthogonal aux vecteurs $\{w^{*1} = w^1, \dots, w^{*i-1} = w^{i-1}\}$ puisque ces derniers sont vecteurs propres de \mathbb{V} . Par construction (9.39), $P(i) = X'DT(i)[T(i)'DT(i)]^{-1} = X'DXW(i)[T(i)'DT(i)]^{-1} = W(i)$. Les déflations de X et de \mathbb{V} à l'ordre i sont évidentes d'après (9.37) et (9.41.a) puisque $p^j = w^j$ pour $j = 1, \dots, i$.

La proposition est donc vraie à tous les ordres. \square

9.11 Représentations factorielles

Les représentations factorielles sont les projections orthogonales des points-variables et des points-individus sur des espaces vectoriels à deux dimensions, appelés plans factoriels engendrés par des couples d'axes factoriels orthogonaux. Il faut donc préciser, d'une part, quels sont les vecteurs qui vont jouer le rôle d'axes factoriels et, d'autre part, quelles sont les métriques qui vont définir les projections orthogonales dans les espaces concernés.

On va voir que la dualité parfaite de la DVS du triplet (X, M, D) qui donnait une double interprétation symétrique aux vecteurs propres des opérateurs “ $\mathbb{V}M$ ” et “ $\mathbb{W}D$ ” associés à la même valeur singulière, à savoir, axe factoriel d'une part et coordonnées des projections sur l'autre axe factoriel d'autre part, ne tient plus dans PLS. En effet, les composantes PLS, t^1, \dots, t^k , permettent une représentation factorielle des variables mais ne fournissent pas directement les coordonnées des individus projetés sur la famille \mathbb{V} -orthogonale, $\{w^{*1}, \dots, w^{*k}\}$ d'axes factoriels. Cependant, le vecteur t^{*i} des coordonnées des individus projetés sur w^{*i} s'exprime simplement en fonction de t^i .

Pour visualiser les variables explicatives et les individus, on dispose donc dans PLS de deux représentations factorielles associées à deux triplets. Le premier triplet, (X, I_p, D) , a pour rôle fondamental la construction des composantes et la visualisation des prédicteurs, le second, (X, \mathbb{V}, D) , sert d'optique photographique auxiliaire pour voir les individus. Ces deux triplets ne fournissent des représentations duales que dans des cas extrêmes, mais il est possible de quantifier un “écart à la dualité” pour chaque axe factoriel i , par le coefficient de corrélation linéaire entre t^i et t^{*i} .

9.11.1 Représentation des prédicteurs et des réponses

L'espace Euclidien des variables centrées, éventuellement réduites, est (\mathbb{R}^n, D) . Dans cet espace, le D -produit scalaire entre deux vecteurs donne la covariance empirique entre les deux variables correspondantes. Les modèles (9.36), (9.37), (9.38) associés au fait que les composantes t^1, \dots, t^k sont D -orthogonales, vont permettre de projeter les colonnes de X et de Y sur les plans factoriels (t^i, t^j) . Notons

$$\tilde{t}^i = t^i / \|t^i\|_D \quad \text{et} \quad \tilde{T}(k) = [\tilde{t}^1 \dots \tilde{t}^k] = T(k)(T(k)'DT(k))^{-1/2}$$

les vecteur normés, c'est à dire, les vecteurs des mesures sur les variables latentes standardisées. Effectuons la projection des colonnes de X et de Y sur l'axe factoriel i , c'est à dire, sur la droite vectorielle engendrée par \tilde{t}^i

$$\Pi_{\tilde{t}^i}^D X = \tilde{t}^i \tilde{t}^{i'} DX, \quad \text{et} \quad \Pi_{\tilde{t}^i}^D Y = \tilde{t}^i \tilde{t}^{i'} DY.$$

Les coordonnées des projections sont données par les vecteurs lignes $\tilde{t}^{i'} DX$ et $\tilde{t}^{i'} DY$

$$[r(t^i, X^1) \sigma(X^1), \dots, r(t^i, X^p) \sigma(X^p)], \quad \text{et} \quad [r(t^i, Y^1) \sigma(Y^1), \dots, r(t^i, Y^q) \sigma(Y^q)],$$

qui lorsque les variables sont standardisées, sont les coefficients de corrélations linéaires entre les variables et la variable latente t^i . D'autre part,

$$r(t^i, X^l) \sigma(X^l) = \|t^i\|_D p_l^i, \quad \text{et} \quad r(t^i, Y^h) \sigma(Y^h) = \|t^i\|_D c_h^i. \quad (9.51)$$

Différentes "cartes" des variables sont proposées par les logiciels pour représenter prédicteurs et réponses sur l'axe i .

Carte des variables $X^l, l = 1 \dots p$, et $Y^h, h = 1 \dots q$, sur l'axe i	
projections	$(r(t^i, X^l) \sigma(X^l), r(t^i, Y^h) \sigma(Y^h))$
coefficients de régression	(p_l^i, c_h^i)
(poids des prédicteurs, coeff. de rég. des réponses)	(w_l^i, c_h^i)

Seules les deux premières méritent le nom de représentations factorielles car, différant d'un facteur d'échelle $\|t^i\|_D$, elles donnent bien les projections des variables sur l'axe factoriel numéro i .

Nous retiendrons la première, qui lorsque les variables sont standardisées, conduit à représenter les variables sur le plan factoriel (i, j) à l'intérieur du cercle des corrélations.

En effet, l'orthogonalité des composantes permet de décomposer la projection sur le plan factoriel (i, j) en la somme des projections sur chacun des axes. Pour les prédicteurs, par exemple, la projection des variables sur le plan (i, j) est

$$\Pi_{(\tilde{i}, \tilde{j})}^D X = \Pi_{\tilde{i}}^D X + \Pi_{\tilde{j}}^D X.$$

Mesure de la qualité de la représentation d'une variable

Il s'agit d'exprimer la proximité d'une variable avec un axe i ou un plan (i, j) . C'est la contribution relative de l'axe ou du plan à la représentation de la variable qui permet de quantifier cette proximité. Elle s'exprime en termes de cosinus carrés. Prenons l'exemple d'une réponse h ,

$$R^2(Y^h; \{t^i\}) = r^2(Y^h, t^i), \quad \text{et} \quad R^2(Y^h; \{t^i, t^j\}) = r^2(Y^h, t^i) + r^2(Y^h, t^j).$$

L'examen de ces valeurs est une aide à l'interprétation des composantes PLS : une composante t^i s'interprète par rapport aux variables dont elle est proche.

Décomposition de l'inertie du nuage des points-variables

Supposons maintenant que les variables sont affectées de poids égaux à 1. On peut considérer les triplets (X, I_p, D) et (Y, I_q, D) pour calculer et décomposer l'inertie des nuages de points-variables, c'est à dire la variance totale des variables.

L'inertie des nuages s'écrit grâce à (9.41.a) et (9.41.b) avec $k = r = \text{rang}(X)$

$$\mathcal{I}^x = \text{trace}(\mathbb{V}) = \sum_{l=1}^p \text{var}(X^l) = \sum_{i=1}^r \|t^i\|_D^2 \|p^i\|^2 \quad (9.52.a)$$

$$\mathcal{I}^y = \text{trace}(Y' D Y) = \sum_{h=1}^q \text{var}(Y^h) = \sum_{i=1}^r \|t^i\|_D^2 \|c^i\|^2 + \text{trace}(\mathbb{V}_{(r)}^{yy}) \quad (9.52.b)$$

où l'inertie des variables projetées sur l'axe i est

$$\mathcal{I}_i^x = \|t^i\|_D^2 \|p^i\|^2 = \sum_{l=1}^p r^2(X^l, t^i) \text{var}(X^l), \quad (9.52.c)$$

$$\mathcal{I}_i^y = \|t^i\|_D^2 \|c^i\|^2 = \sum_{h=1}^q r^2(Y^h, t^i) \text{var}(Y^h). \quad (9.52.d)$$

On peut mettre en évidence la contribution de $\text{Im } T(k)$ dans l'approximation de la variance totale par

$$\mathcal{I}^x(k) = \sum_{l=1}^p \text{var}(X^l) R^2(X^l; \text{Im } T(k)) \quad \text{et} \quad \mathcal{I}^x(r) = \mathcal{I}^x, \quad (9.53.a)$$

$$\mathcal{I}^y(k) = \sum_{h=1}^q \text{var}(Y^h) R^2(Y^h; \text{Im } T(k)) \quad \text{et} \quad \mathcal{I}^y(r) \leq \mathcal{I}^y. \quad (9.53.b)$$

On va voir maintenant comment les composantes PLS permettent de représenter les individus du tableau des variables explicatives.

9.11.2 Représentation des individus

Par analogie à l'Analyse en Composantes Principales usuelle, la plupart des logiciels présentent une carte des individus fournie par la représentation graphique (t^i, t^j) . On va voir pourquoi ce n'est pas une représentation factorielle basée sur une projection, on la qualifie de "pseudo factorielle". Elle est cependant, souvent globalement proche de la carte factorielle exacte (t^{*i}, t^{*j}) définie dans ce paragraphe et à laquelle est associée une mesure de la qualité de la représentation de chaque individu.

Pour $k \in \{1, \dots, r\}$, on dispose par la proposition 9.7 (b), de deux familles orthogonales, $\{w^1, \dots, w^k\}$ et $\{w^{*1}, \dots, w^{*k}\}$, dans l'espace, $\text{Im } X'$, des individus du tableau des prédicteurs. La première est orthogonale au sens usuel, la seconde \mathbb{V} -orthogonale. D'autre part, une composante s'écrit

$$t^i = X_{(i-1)} w^i = X w^{*i}.$$

Il est clair que si l'on projette X' sur w^i , les coordonnées du nuage des points projetés ne s'exprime pas simplement en fonction de t^i et il est impossible d'interpréter les individus projetés par rapport aux variables les plus proches de t^i .

Nous utiliserons donc, l'autre famille, $\{w^{*1}, \dots, w^{*k}\}$, comme axes factoriels et pour cela, munirons l'espace $\text{Im } X'$ de la métrique \mathbb{V} . La représentation des individus est ainsi associée au triplet (X, \mathbb{V}, D) qui sert d'optique photographique.

Proposition 9.9 : Les points du nuage des individus de X projetés sur l'axe factoriel défini par le vecteur unitaire $\tilde{w}^{*i} = w^{*i} / \|w^{*i}\|_{\mathbb{V}}$, ont pour mesures algébriques les coordonnées du vecteur

$$t^{*i} = \mathbb{W} D \tilde{t}^i = X X' D t^i / \|t^i\|_D = X p^i \|t^i\|_D. \quad (9.54)$$

□

Preuve :

$$\Pi_{\tilde{w}^{*i}}^{\mathbb{V}} X' = \tilde{w}^{*i} \tilde{w}^{*i'} \mathbb{V} X' = \tilde{w}^{*i} (X X' D X \tilde{w}^{*i})' = \tilde{w}^{*i} (X X' D t^i / \|w^{*i}\|_{\mathbb{V}})'$$

L'expression de p^i donnée par (9.32) termine la preuve. □

Hélas! Le vecteur t^{*i} n'est pas, en général, colinéaire à t^i sauf dans deux cas limites pour PLS, celui de l'ACP de X et celui où les variables explicatives standardisées sont non corrélées deux à deux ($\mathbb{V} = I_p$). Ainsi, représenter les individus par la carte (t^i, t^j) n'est légitimement fondé que dans ces deux cas extrêmes où l'on retrouve la dualité d'interprétation d'une composante, à la fois axe factoriel du côté des variables et vecteur des coordonnées des individus projetés.

L'écart à la dualité pour une composante t^i , appelé "saut de dualité" et noté SD_i , est l'expression comprise entre 0 et 1,

$$SD_i = 1 - r_i,$$

où $r_i = r(t^i, t^{*i})$ est le coefficient de corrélation linéaire entre t^{*i} et t^i .

D'après (9.32), $cov(t^i, t^{*i}) = \|p^i\|^2 var(t^i) > 0$ si $i \leq rang(X)$. Comme conséquence, r_i est positif, l'angle entre t^i et t^{*i} est aigu et SD_i est compris entre 0 et 1.

<u>Cas $SD_i = 0$: $\exists \alpha_i > 0$ tel que $XX'Dt^i = \alpha_i t^i$</u>	
$Y = X$	$PLS(X, Y) \equiv ACP(X)$, alors $\forall i, \alpha_i = \lambda_i$
$\mathbb{V} = I_p$	non corrélation, alors $\forall i, \alpha_i = 1$

Plus SD_i sera voisin de 0 (r_i voisin de 1) et plus il sera justifié d'interpréter la représentation des individus donnée par t^{*i} , grâce aux variables explicatives et aux variables réponses projetées sur t^i .

La \mathbb{V} -orthogonalité de deux axes factoriels (w^{*i}, w^{*j}) permet de décomposer la projection des individus sur le plan (i, j) comme la somme des projections sur chacun des axes

$$\Pi_{(\tilde{w}^{*i}, \tilde{w}^{*j})}^{\mathbb{V}} X' = \Pi_{\tilde{w}^{*i}}^{\mathbb{V}} X' + \Pi_{\tilde{w}^{*j}}^{\mathbb{V}} X'.$$

Mesure de la qualité de la représentation d'un individu

De façon habituelle, une mesure de la qualité de la représentation de l'individu l sur l'axe i ou sur le plan factoriel (i, j) , est donnée par le carré du cosinus du \mathbb{V} -angle formé par les deux vecteurs d'origine l'origine des coordonnées, et dont les extrémités sont le point-individu l d'une part et sa projection d'autre part. Soit $r = rang(X)$, les contributions relatives de l'axe i et du plan factoriel (i, j) à la représentation de l'individu l sont

$$\cos^2 \theta_l^i = \frac{(t_l^{*i})^2}{\sum_{j=1}^r (t_l^{*j})^2} \quad \text{et} \quad \cos^2 \theta_l^{i,j} = \cos^2 \theta_l^i + \cos^2 \theta_l^j. \quad (9.55)$$

Remarquons que dans (9.55), le dénominateur est égal à $X_l \mathbb{V} X_l'$ où X_l est la l ème ligne de X . En effet, si l'on note $\tilde{W}^*(k)$ la matrice des vecteurs unitaires, et $T^*(k)$ la matrice

des vecteurs donnant les coordonnées des projections,

$$\widetilde{W}^*(k) = [\tilde{w}^{*1} \dots \tilde{w}^{*k}] = W^*(k)[T(k)'DT(k)]^{-1/2}, \quad (9.56.a)$$

$$T^*(k) = [t^{*1} \dots t^{*k}] = XX'D\widetilde{T}(k), \quad (9.56.b)$$

alors, puisque $Im \widetilde{W}^*(r) = Im X'$, $X' = \Pi_{\widetilde{W}^*(r)}^{\mathbb{V}} X' = \widetilde{W}^*(r)T^*(r)'$ et

$$T^*(r)T^*(r)' = T^*(r)I_r T^*(r)' = T^*(r)[\widetilde{W}^*(r)'\mathbb{V}\widetilde{W}^*(r)]T^*(r)' = X\mathbb{V}X'. \quad (9.57)$$

L'élément diagonal (l, l) de ces matrices donne le résultat.

Décomposition de l'inertie du nuage des individus

Le nuage des individus a pour inertie, notée \mathcal{I}^{ind} , le carré de la norme de Frobénius de \mathbb{V}

$$\mathcal{I}^{ind} = trace(X\mathbb{V}X'D) = trace(\mathbb{V}^2) = \|\mathbb{V}\|_F^2. \quad (9.58)$$

En outre, $\mathcal{I}^{ind} \geq trace(\mathbb{V}) = \mathcal{I}^x$ si pour tout i , $\sigma(X^i) \geq 1$, ce qui est le cas dans PLS sur variables standardisées. L'inertie des individus \mathcal{I}^{ind} , égale à \mathcal{I}^x dans l'ACP usuelle, incorpore dans PLS, non seulement les variances mais aussi les covariances des variables explicatives.

$$\mathcal{I}^{ind} = \sum_{i=1}^p var^2(X^i) + 2 \sum_{i \neq j} cov^2(X^i, X^j) \geq \sum_{i=1}^p var^2(X^i).$$

Il est intéressant de regarder si \mathcal{I}^{ind} est assez proche de $\sum_{i=1}^p var^2(X^i)$ (de p dans le cas standardisé) c'est à dire si \mathbb{V} est proche d'être diagonale (de la matrice identité dans le cas standardisé). Dans ce cas, toutes les composantes PLS sont certes proches de la dualité au sens défini plus haut mais PLS perd de son intérêt.

D'autre part, l'inertie se décompose en la somme des inerties des points projetés sur chacun des r axes possibles,

$$\mathcal{I}^{ind} = trace(T^*(r)T^*(r)'D) = \sum_{i=1}^r trace(t^{*i}t^{*i}'D) = \sum_{i=1}^r \mathcal{I}_i^{ind}.$$

L'inertie des individus projetés sur l'axe i est, puisque t^{*i} est D -centré,

$$\mathcal{I}_i^{ind} = trace(t^{*i}t^{*i}'D) = t^{*i}'Dt^{*i} = var(t^{*i}),$$

On définit, en pourcentage, la qualité globale du plan factoriel (i, j) par

$$100 \frac{\mathcal{I}_{i,j}^{ind}}{\mathcal{I}^{ind}} = 100 \frac{\mathcal{I}_i^{ind}}{\mathcal{I}^{ind}} + 100 \frac{\mathcal{I}_j^{ind}}{\mathcal{I}^{ind}}.$$

Proposition 9.10 : Expression des inerties associées aux deux triplets

L'inertie des variables explicatives et l'inertie des individus correspondants peuvent s'exprimer de deux façons différentes en fonction de t^i ou de t^{*i}

$$\mathcal{I}^x = \text{trace}(\mathbb{V}) = \sum_{i=1}^r \|t^i\|_D^2 \|p^i\|^2 = \sum_{i=1}^r \|t^{*i}\|_D r_i \quad (9.59)$$

$$\mathcal{I}^{ind} = \|\mathbb{V}\|_F^2 = \sum_{i=1}^r \|t^{*i}\|_D^2 = \sum_{i=1}^r \|t^i\|_D^2 \|p^i\|_V^2. \quad (9.60)$$

□

Preuve : Seule la dernière égalité dans chacune des formules est à démontrer, les premières ayant déjà été obtenues.

Si l'on prend $k = r = \text{rang}(X)$, X' est invariant par projection sur $\text{Im } \widetilde{W}^*(r)$ de même, X est invariant par projection sur $\text{Im } \widetilde{T}(r)$. Il vient, $X' = \widetilde{W}^*(r)T^*(r)'$ et $X = \widetilde{T}(r)\widetilde{W}^*(r)'\mathbb{V}$.

Alors,

$$\begin{aligned} \text{trace}(X'DX) &= \text{trace}(\widetilde{W}^*(r)T^*(r)'\widetilde{D}\widetilde{T}(r)\widetilde{W}^*(r)'\mathbb{V}) = \text{trace}(T^*(r)'\widetilde{D}\widetilde{T}(r)) \\ &= \sum_{i=1}^r \frac{\text{cov}(t^{*i}, t^i)}{\|t^i\|_D} = \sum_{i=1}^r \|t^{*i}\|_D r_i. \end{aligned}$$

Pour (9.60), avec (9.54), $\sum_{i=1}^r \|t^{*i}\|_D^2 = \sum_{i=1}^r t^{*i'}Dt^{*i} = \sum_{i=1}^r \|t^i\|_D^2 p^{i'}\mathbb{V}p^i$.

□

9.12 Métriques pour les individus et optiques photographiques associées

9.12.1 Métriques générales pour les individus

Nous avons vu dans les paragraphes précédents, que l'espace des variables est (\mathbb{R}^n, D) et que celui des individus est (\mathbb{R}^p, I_p) pour les mesures sur les prédicteurs et (\mathbb{R}^q, I_q) pour celles des réponses. Sur ces espaces sont construites les suites $\{w^i\}_i$ et $\{v^i\}_i$, des vecteurs des poids et, par là, les composantes $\{t^i\}_i$ et les $\{u^i\}_i$ de (\mathbb{R}^n, D) définis par (9.5)

$$t^i = X_{(i-1)}w^i \quad \text{et} \quad u^i = Y_{(i-1)}v^i.$$

La métrique $\mathbb{V} = X'DX$ sert d'optique photographique adaptée à la visualisation exacte des individus de X : elle permet de projeter les individus sur la famille $\{w^{*i}\}_i$ qui est \mathbb{V} -orthogonale. Une composante s'écrit avec (9.42)

$$t^i = Xw^{*i}$$

et les coordonnées des projections du nuage des individus sur $Im w^{*i}$ sont données par (9.54)

$$t^{*i} = \mathbb{W}Dt^i / \|t^i\|_D = XX'Dt^i / \|t^i\|_D.$$

Dans ce paragraphe, nous adoptons la démarche classique en Analyse Factorielle de Données, qui consiste à choisir des métriques plus générales notées M^x et M^y en remplacement de I_p et I_q , pour mesurer des distances sur les individus de X et de Y . Dans ce cas, les formules (9.5) deviennent

$$t^i = X_{(i-1)}M^xw^i \quad \text{et} \quad u^i = Y_{(i-1)}M^yv^i.$$

Cela induit l'optique photographique adaptée à la vision exacte des individus de X

$$M^x\mathbb{V}M^x,$$

en remplacement de \mathbb{V} , car (9.42) devient

$$t^i = XM^xw^{*i}$$

et la D -norme d'une composante s'écrit

$$\|t^i\|_D = \|w^{*i}\|_{M^x\mathbb{V}M^x}.$$

Du coté des prédicteurs, le triplet (X, M^x, D) ne peut définir une inertie des variables que si M^x est diagonale. L'inertie totale des individus pour le triplet $(X, M^x\mathbb{V}M^x, D)$ devient

$$\mathcal{I}^{ind} = \text{trace}((\mathbb{V}M^x)^2) = \|X\|_{(M^x\mathbb{V}M^x) \otimes D}^2$$

et les projections du nuage sur $Im w^{*i}$, données par (9.54), s'écrivent

$$t^{*i} = \mathbb{W}Dt^i / \|t^i\|_D = XM^xX'Dt^i / \|t^i\|_D.$$

9.12.2 Régression PLS discriminante

Une application de la démarche précédente est donnée par $PLS(X, Y)$ dans le contexte de l'Analyse Factorielle Discriminante, voir chapitre 7.4 Exercice 8. La matrice Y , $n \times q$, est l'indicatrice des q groupes d'individus mesurés sur p variables explicatives, la matrice X des observations est supposée de plein rang colonne. Du coté de Y , $M^y = (Y'DY)^{-1}$, alors que du coté de X , $M^x = \mathbb{V}^{-1}$, appelée métrique de Mahalanobis, est aussi dans ce cas, l'optique photographique nécessaire pour une vision exacte des individus

$$M^x\mathbb{V}M^x = \mathbb{V}^{-1} = M^x.$$

Les axes PLS discriminants $\{w^{*i}\}_i$ sont \mathbb{V}^{-1} -orthogonaux. Le saut de dualité entre une variable discriminante $t^i = X\mathbb{V}^{-1}w^{*i}$ et les projections des individus est toujours nul car

$$t^{*i} = X\mathbb{V}^{-1}X'Dt^i/\|t^i\|_D = \Pi_X^D t^i/\|t^i\|_D = t^i/\|t^i\|_D.$$

En régression PLS discriminante, les représentations pseudo-factorielles des individus sont donc des représentations factorielles. L'inertie des individus est égale à p car $\mathcal{I}^{ind} = \text{trace}((I_p)^2) = p$. Lorsque $\text{rang}(X) < p$, $M^x = \mathbb{V}^+$ et les propriétés ci dessus tiennent toujours, sauf pour l'inertie, $\mathcal{I}^{ind} = \text{trace}((\mathbb{V}\mathbb{V}^+)^2) = \text{rang}(X)$.

9.13 Choix du nombre de composantes

Tout d'abord, on connaît une borne supérieure de k ,

$$k \leq r = \text{rang}(X).$$

Lorsque l'on choisit $k = \text{rang}(X)$, le modèle de régression PLS est identique à celui de la régression aux moindres carrés usuelle lorsque cette dernière est applicable.

Trois types de critères permettent de déterminer le nombre k de composantes, aussi appelé la dimension du modèle. Le premier type est basé sur le "fit" c'est à dire, l'ajustement de l'échantillon d'apprentissage (X, Y) par $(\hat{X}(k), \hat{Y}(k))$ construit par PLS à k composantes. Les deux autres sont basés sur la prédiction. L'un, basé sur la prédiction interne aux données d'apprentissage, est appelé critère de "validation croisée". L'autre, basé sur la prédiction externe, nécessite un jeu de données supplémentaire, appelé échantillon test ou échantillon de validation et noté (X_t, Y_t) .

Finalement, la détermination de k est une décision qui fait la synthèse des informations recueillies par l'examen des critères disponibles.

9.13.1 Critères basés sur l'ajustement

Trois critères permettent de mesurer l'apport des composantes dans l'ajustement aux données. Les deux premiers concernent la reconstruction de X et un troisième est associé à l'ajustement de Y .

Critères sur X

- \mathbb{V} -inertie des individus reconstruite par k composantes

$$\mathcal{I}^{ind}(k) = \sum_{i=1}^k \text{var}(t^{*i}) \quad \text{et} \quad \mathcal{I}^{ind}(r) = \|\mathbb{V}\|_F^2 = \|X\|_{\mathbb{V} \otimes D}^2 = \sum_{h,l} \text{cov}^2(X^l, X^h).$$

Le pourcentage d'inertie totale reconstruite par k axes

$$\%I^{ind}(k) = \sum_{i=1}^k 100 \frac{\mathcal{I}_i^{ind}}{\mathcal{I}^{ind}} \quad (9.61)$$

donne le critère mesurant l'apport des k composantes dans la représentation des individus.

– D -variance reconstruite par k composantes

$$\mathcal{I}^x(k) = \sum_{i=1}^k \left(\sum_{l=1}^p r^2(X^l, t^i) \text{var}(X^l) \right) \quad \text{et} \quad \mathcal{I}^x(r) = \text{trace}(\mathbb{V}) = \|X\|_D^2 = \sum_{l=1}^p \text{var}(X^l).$$

Le pourcentage de variance totale reconstruite par k axes est donné par

$$\%I^x(k) = 100 \frac{\sum_{j=1}^k \mathcal{I}_j^x}{\mathcal{I}^x}. \quad (9.62)$$

Ainsi PLS possède comme propriété secondaire, mais précieuse, le fait de calculer le rang de X comme étant le nombre de composantes reconstruisant 100% des deux critères basés sur l'ajustement de X .

Critère sur Y

– D -variance reconstruite par k composantes

$$\mathcal{I}^y(k) = \sum_{i=1}^k \left(\sum_{h=1}^q r^2(Y^h, t^i) \text{var}(Y^h) \right) \quad \text{et} \quad \mathcal{I}^y(r) \leq \text{trace}(Y' D Y) = \sum_{h=1}^q \text{var}(Y^h).$$

Le pourcentage de variance totale reconstruite par k axes est donné par

$$\%I^y(k) = 100 \frac{\sum_{j=1}^k \mathcal{I}_j^y}{\mathcal{I}^y}. \quad (9.63)$$

Grâce à (9.61), (9.62) et (9.63), un premier critère du choix du nombre de composantes peut être énoncé comme suit :

“on choisira k de telle façon que l'inertie des individus et la variance de X soient suffisamment reconstruites pour un gain faible dans l'approximation de la variance de Y ”.

La Figure 24 présente le diagramme de l'évolution de ces critères en fonction du nombre de composantes. Basé sur un exemple réel, il donne une première indication pour le choix de k .

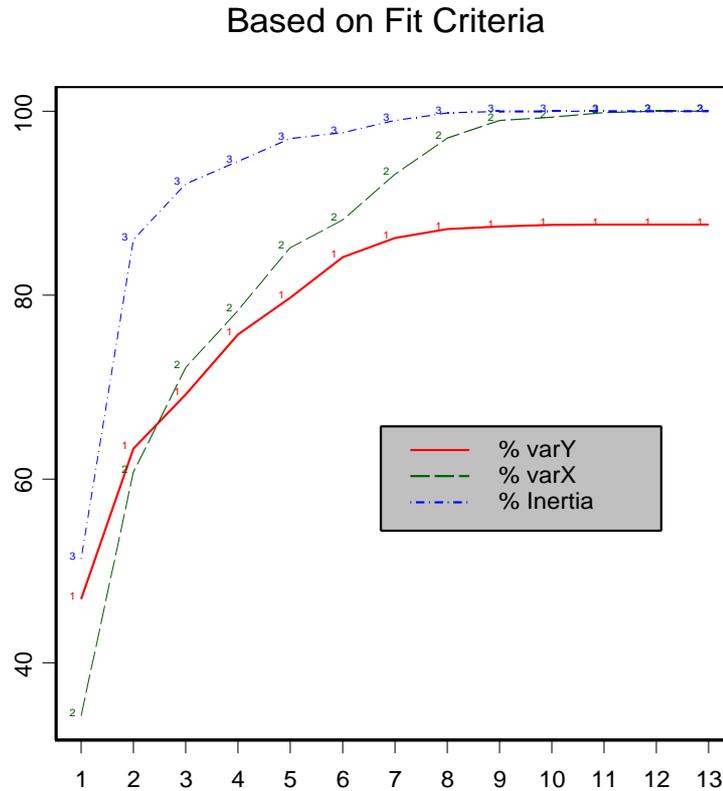


Figure 24 : Évolution des trois critères basés sur l'ajustement aux données en fonction du nombre de composantes. Sur l'exemple, le rang de X est égal à 13.

Il semble raisonnable, pour éviter un sur-ajustement aux données, de choisir k dans l'intervalle $[5, 7]$ pour 80% de la variance de Y reconstituée et pour plus de 80% de X reconstruit par les deux critères, inertie et variance.

Notons que pour $k = 1, \dots, r$, ces trois critères fournissent, par construction, des suites croissantes de valeurs.

9.13.2 Critères basés sur la prédiction

Le modèle de régression PLS (9.46)

$$\hat{Y}(k) = X\hat{\beta}(k),$$

bâti sur l'échantillon d'apprentissage (X, Y) permet de faire de la prédiction. Il est indispensable de valider le choix de k par une mesure basée sur la prédiction.

Validation externe

On dispose parfois d'un deuxième jeu de données, appelé échantillon test et noté (X_t, Y_t) ,

mesuré sur les mêmes variables mais sur N individus supplémentaires supposés de poids identique. Les matrices X_t et Y_t sont supposées centrées (éventuellement réduites) par rapport aux individus moyens (aux écart-types) des données brutes de l'échantillon d'apprentissage tels qu'ils ont été définis au paragraphe 1.9. On dispose en outre, d'une famille de modèles $\{\hat{\beta}(k) = [\hat{\beta}_i^j(k)]\}$, $k = 1, \dots, r$, pour lesquels on évalue l'erreur quadratique moyenne de prédiction sur l'échantillon (X_t, Y_t)

$$MSE(k) = \frac{1}{q} \sum_{j=1}^q MSE^j(k) \quad \text{où} \quad MSE^j(k) = \frac{1}{N} \sum_{i=1}^N (Y_{t|i}^j - X_{t|i} \hat{\beta}^j(k))^2. \quad (9.64)$$

Le minimum de $\{MSE(k), k = 1, \dots, r\}$ est généralement obtenu pour une valeur unique k_{opt} qui dépend, bien sûr, de l'échantillon test sélectionné. Cependant, cette valeur optimale (au sens du critère) n'est pas forcément celle à retenir lors du choix du nombre de composantes, tout dépend de la façon dont elle a été obtenue et de la valeur $MSE(k_{opt})$.

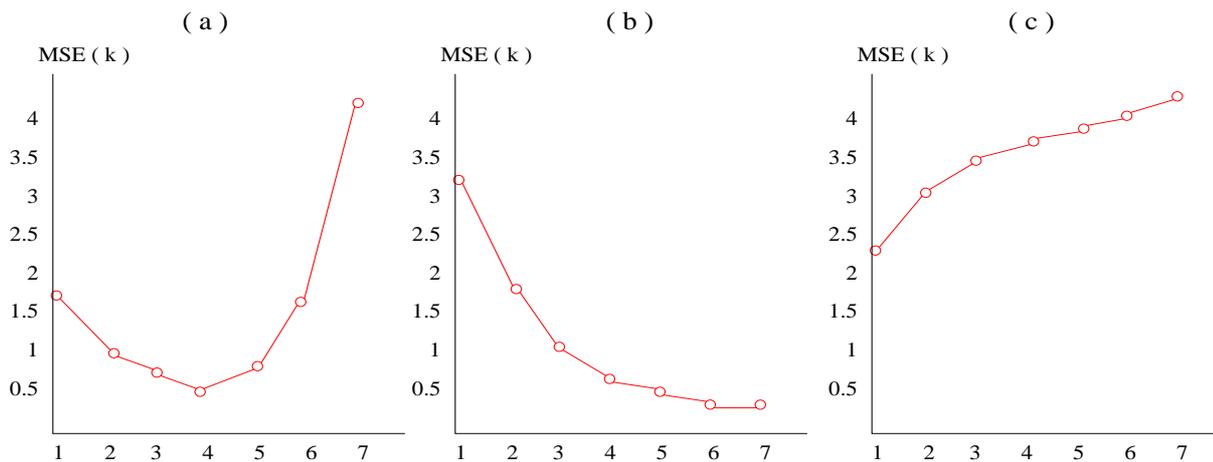


Figure 25 : Exemples d'évolution des critères de prédiction externe avec k ; (a) et (b), validation externe acceptable ; (c), remise en cause du modèle et/ou de l'échantillon test.

Figure 25, trois exemples de validation externe sont présentés qui tous trois montrent une évolution régulière du critère avec le nombre de composantes. Ce n'est pas toujours le cas, il arrive parfois que l'arrivée d'une composante dans le modèle capture l'influence d'une donnée atypique et provoque un brusque saut dans l'évolution du critère.

Une valeur limite pour refuser le modèle PLS, sous l'hypothèse d'un échantillon test fiable, est $MSE(k_{opt}) = 1$, dans le cas standardisé. En effet, une telle valeur signifie que, en moyenne, l'approximation d'une réponse donne sa valeur moyenne. Ainsi, Figure 25 (c), le modèle PLS est rejeté car $k_{opt} = 1$ et $MSE(k_{opt}) > 1$. Il faut dans le cas multi-réponses,

examiner les graphiques des $MSE^j(k)$ pour savoir quelle réponse est mal prédite ou bien prédite.

Les cas (a) et (b), Figure 25, sont ceux pour lesquels le modèle PLS est bien validé par l'échantillon test. Si le choix du nombre de composantes est clair pour (a), $k = k_{opt} = 4$, il est moins évident pour (b) et dépend dans ce cas des autres critères et aussi du "principe d'économie" : choisir la plus petite dimension dans le cas de valeurs du critère très voisines.

Validation interne ou "validation croisée"

Le principe est le même que pour la validation externe mais basé cette fois sur l'échantillon d'apprentissage. On partage l'échantillon en deux groupes d'individus, l'un pour bâtir le modèle, l'autre pour le valider et on mesure l'erreur de prédiction. On recommence le procédé sur d'autres bi-partitions de telle sorte qu'un individu ne soit prédit qu'une seule fois... À la fin, on calcule la somme (ou la moyenne) des erreurs quadratiques moyennes obtenues pour les q réponses, avec $k = 1, \dots, r$. Ce procédé appelé validation croisée, est le plus souvent utilisé en enlevant, pour être prédit, un seul individu à la fois. Dans ce cas, la méthode est appelée "leave-one-out". En général, on tolère 10% d'individus sélectionnés à chaque étape.

Écrivons le critère dans le cas "leave-one-out". On suppose que tous les individus ont le même poids et on note $\hat{\beta}_{(-i)}(k) = [\hat{\beta}_{(-i)}(k)|_i^j]$ la matrice $p \times q$ des coefficients du modèle obtenu pour l'individu i "out". Alors, le PRESS (Predictive Error Sum of Squares), s'écrit

$$PRESS(k) = \sum_{j=1}^q PRESS^j(k) \quad \text{et} \quad PRESS^j(k) = \frac{1}{n} \sum_{i=1}^n (Y_i^j - X_i \hat{\beta}_{(-i)}(k))^2. \quad (9.65)$$

La discussion sur le choix de k est identique à celle de la validation externe. Cependant, la façon dont les groupes d'individus sont enlevés-prédits à une influence sur le choix de k . Il est souvent conseillé de recommencer un certain nombre de fois la procédure avec permutation des individus. On obtient ainsi une statistique sur le PRESS ce qui robustifie la décision finale en ce qui concerne le choix des partitions.

9.14 Pratique de la régression PLS

9.14.1 PLS univarié, les données de Cornell

Les données de Cornell

On trouve analysé dans [12, Tenenhaus] cet exemple traité par PLS. L'indice d'octane moteur y de douze différents mélanges a été enregistré pour déterminer l'influence de sept composants. Les sept variables représentent des proportions et somment à 1.

n ^o	x1	x2	x3	x4	x5	x6	x7	y
1	0.00	0.23	0.00	0.00	0.00	0.74	0.03	98.7
2	0.00	0.10	0.00	0.00	0.12	0.74	0.04	97.8
3	0.00	0.00	0.00	0.10	0.12	0.74	0.04	96.6
4	0.00	0.49	0.00	0.00	0.12	0.37	0.02	92.0
5	0.00	0.00	0.00	0.62	0.12	0.18	0.08	86.6
6	0.00	0.62	0.00	0.00	0.00	0.37	0.01	91.2
7	0.17	0.27	0.10	0.38	0.00	0.00	0.08	81.9
8	0.17	0.19	0.10	0.38	0.02	0.06	0.08	83.1
9	0.17	0.21	0.10	0.38	0.00	0.06	0.08	82.4
10	0.17	0.15	0.10	0.38	0.02	0.10	0.08	83.2
11	0.21	0.36	0.12	0.25	0.00	0.00	0.06	81.4
12	0.00	0.00	0.00	0.55	0.00	0.37	0.08	88.1
moy.	0.07	0.22	0.04	0.25	0.04	0.31	0.06	88.58
stdev	0.09	0.19	0.05	0.22	0.05	0.28	0.03	6.24

La matrice des corrélations montre que la variable x_1 est très fortement corrélée avec x_3 ainsi que x_4 avec x_7 . La réponse y l'est avec x_6 et à un degré moindre avec x_1 et x_3 . Le prédicteur x_5 est peu corrélé avec les autres variables.

	x2	x3	x4	x5	x6	x7	y
x1	0.10	0.999	0.37	-0.55	-0.80	0.60	-0.84
x2		0.10	-0.54	-0.29	-0.19	-0.59	-0.07
x3			0.37	-0.55	-0.81	0.61	-0.84
x4				-0.21	-0.65	0.92	-0.71
x5					0.46	-0.27	0.49
x6						-0.66	0.99
x7							-0.74

Le tableau ci-dessus présentant les corrélations donne, pour ce qui concerne les variables explicatives, les éléments hors-diagonaux de la matrice $\mathbb{V} = n^{-1}X'X$. Les deux expressions déduites de \mathbb{V} qui jouent un rôle clé dans PLS valent

$$\mathcal{I}^x = \text{trace}(\mathbb{V}) = 7, \quad \text{et} \quad \mathcal{I}^{ind} = \|\mathbb{V}\|_F^2 = 20.411.$$

Remarquons d'abord que le rang de X est six, puisque les variables somment à un, ce qui implique que la régression linéaire multiple de y sur les sept variables explicatives ne peut être effectuée. La régression linéaire pas à pas descendante retient les variables x_1, x_2, x_4, x_5 ce qui n'est pas satisfaisant car x_6 , fortement corrélée à y , n'est pas retenue. De plus, le chimiste désire un modèle qui l'aide à réaliser un mélange qui doit intégrer les sept composants.

Choix de la dimension du modèle

Pour déterminer le nombre de composantes PLS, on ne dispose pas, ici, d'un jeu de données test. Seuls les critères basés sur l'ajustement aux données et sur la validation croisée sont utilisables.

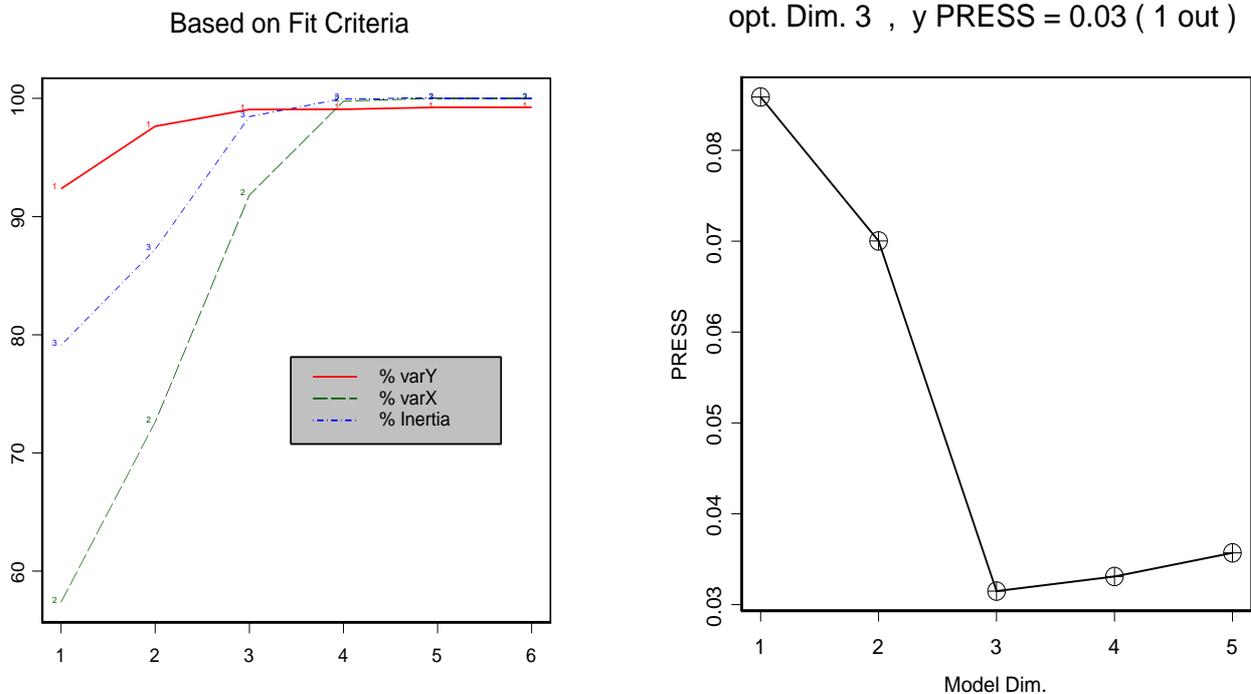


Figure 26 : Critères d'aide à la décision pour le choix du nombre de composantes pour les données de Cornell.

Avec trois composantes, PLS reconstitue 99% de variance de y ainsi que 98.4% de l'inertie des individus et 91.2% de la variance de X . La valeur optimale du $PRESS$ est obtenue pour trois composantes avec $PRESS(3) = 0.03$. L'évolution des différents critères, Figure 26, indique sans ambiguïté que trois est le meilleur choix pour la dimension du modèle.

La validation croisée permet aussi d'examiner quels sont les individus qui sont correctement prédits et quels sont ceux qui le sont moins. Pour cela, la partie gauche de la Figure 26 présente le graphique des valeurs observés sur y contre les valeurs prédites par validation croisée à trois composantes. L'écart-type de l'erreur de prédiction est égal à 0.177. Ce sont les individus 1, 6 et 12 qui sont le moins bien prédits par le modèle à trois dimensions.

Une autre indication pour le choix du modèle est fournie par l'évolution des coefficients $\hat{\beta}^j(i)$ au cours de sa construction, c'est à dire pour pour $k = 1, \dots, rang(X)$. Sur l'exemple de Cornell, une seule réponse ($j = 1$) est à prédire et la partie droite de la Figure 27 présente l'évolution des coefficients suivant le nombre de composantes qui

entrent dans le modèles. Notons à l'étape 6 une explosion des coefficients des variables X^1 et X^3 , qui présentaient auparavant des signes identiques négatifs. Dès l'étape 5, la variable X^7 voit son coefficient changer de signe. On note donc une zone d'instabilité pour les variables explicatives les plus corrélées qui commence sur cet exemple à $k = 5$. Le type de graphique précédent incite donc à choisir k dans la zone de stabilité des coefficients, c'est à dire, ici, $k \leq 4$.

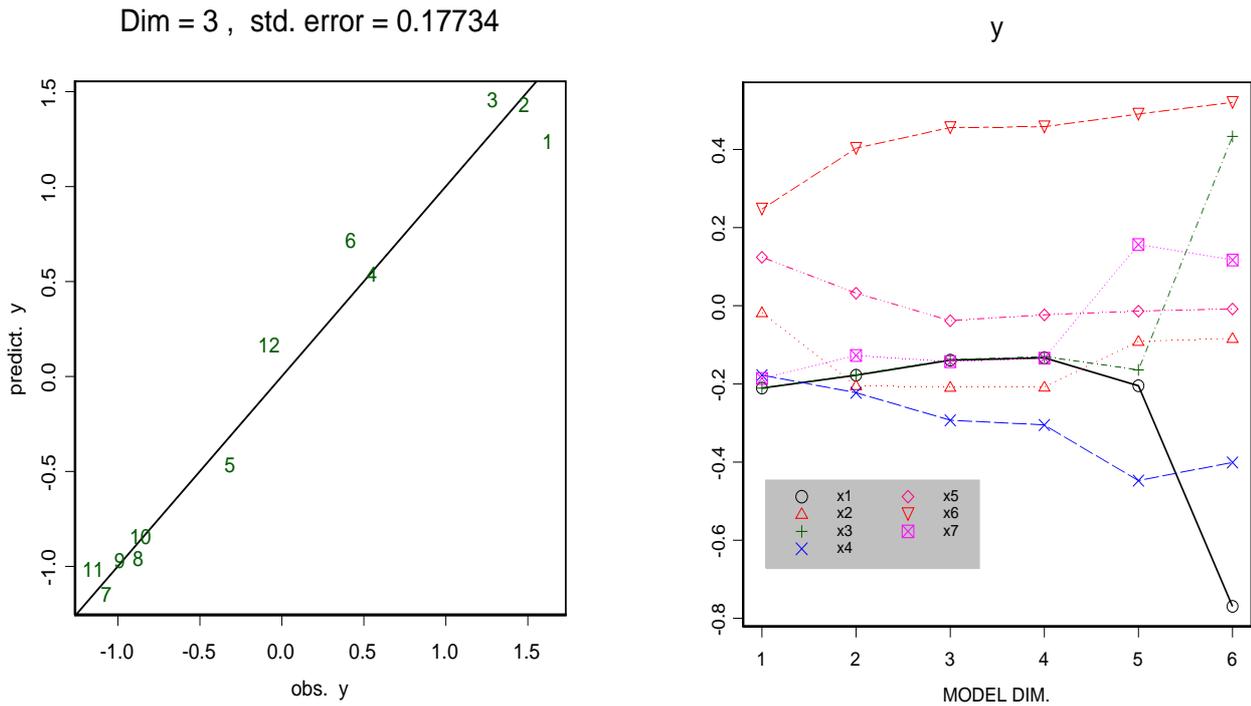


Figure 27 : À gauche, individus observés contre individus prédits par validation croisée à trois composantes. À droite, évolution des coefficients $\hat{\beta}(k)$ suivant le nombre de composantes rentrant dans le modèle.

Le modèle PLS

Finalement, le modèle retenu, à trois composantes, s'écrit sur les variables standardisées,

$$\hat{Y}(3) = -0.1391 X^1 - 0.2087 X^2 - 0.1376 X^3 - 0.2932 X^4 - 0.0384 X^5 + 0.4564 X^6 - 0.1434 X^7$$

Il est tout à fait cohérent avec les coefficients de corrélation et X^6 est le prédicteur qui contribue le plus à la construction de Y alors que l'influence de X^5 est négligeable. Figure 28, les graphiques des "fonctions coordonnées" ($x^i, \hat{\beta}_i(3)x^i$) montrent l'influence additive des observations sur la réponse. Les variables explicatives sont classées, de gauche à droite et de haut en bas, par ordre décroissant d'influence selon la valeur absolue des coefficients.

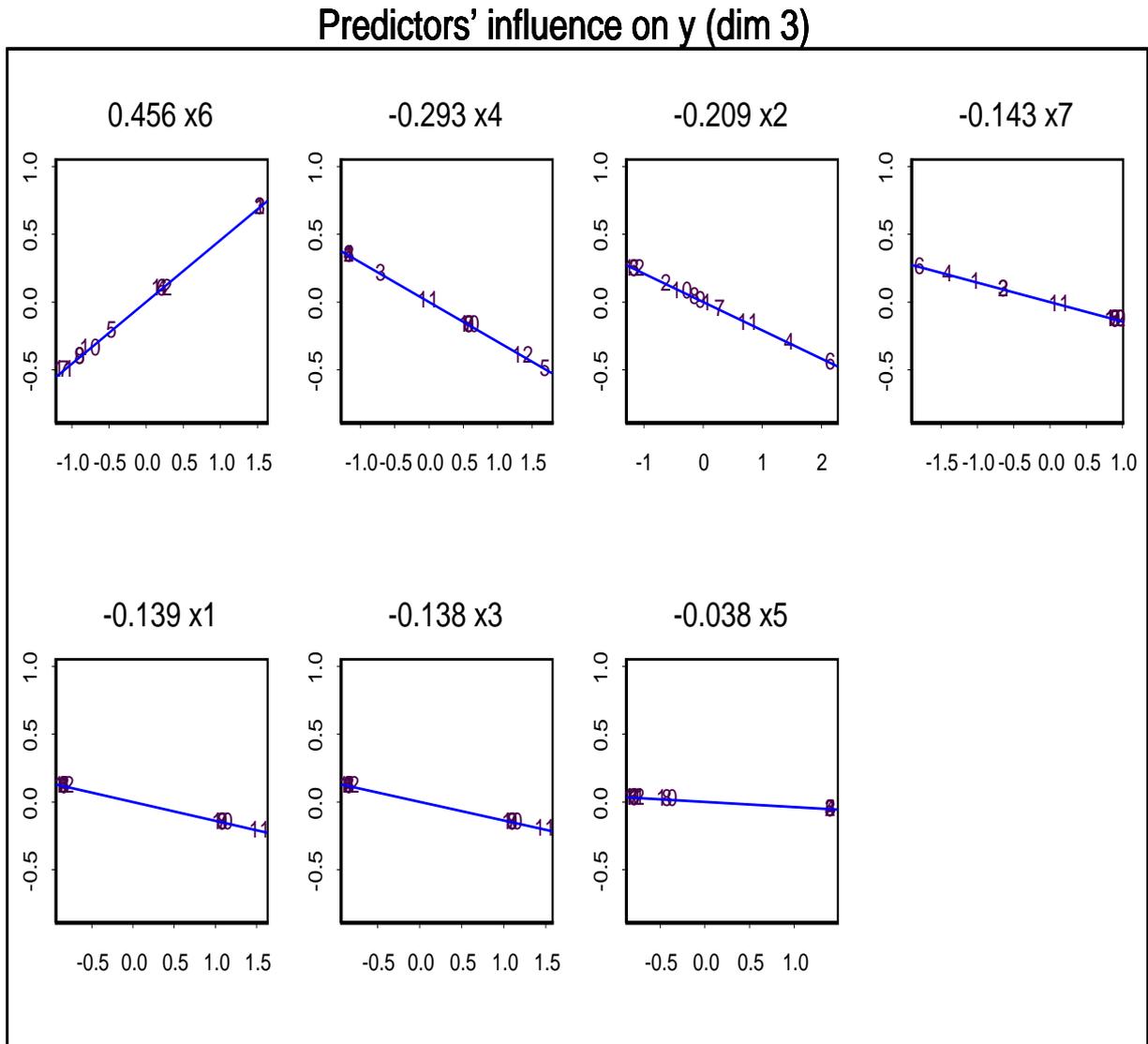


Figure 28 : Graphiques des observations sur les “fonctions coordonnées” $(x^i, \hat{\beta}_i(3)x^i)$ classées par ordre décroissant d’influence sur la réponse y standardisée.

Sur les données d’origine, le modèle devient,

$$\hat{Y}(3) = 92.676 - 9.828 X^1 - 6.96 X^2 - 16.666 X^3 - 8.422 X^4 - 4.389 X^5 + 10.161 X^6 - 34.529 X^7$$

Représentation des variables

La régression PLS ayant été effectuée sur variables centrées réduites, une variable est bien représentée sur un plan factoriel (i, j) si sa projection est proche du cercle des corrélations. La Figure 29, présente les deux plans factoriels (1,2) pour les individus, à

gauche, et pour les variables, à droite.

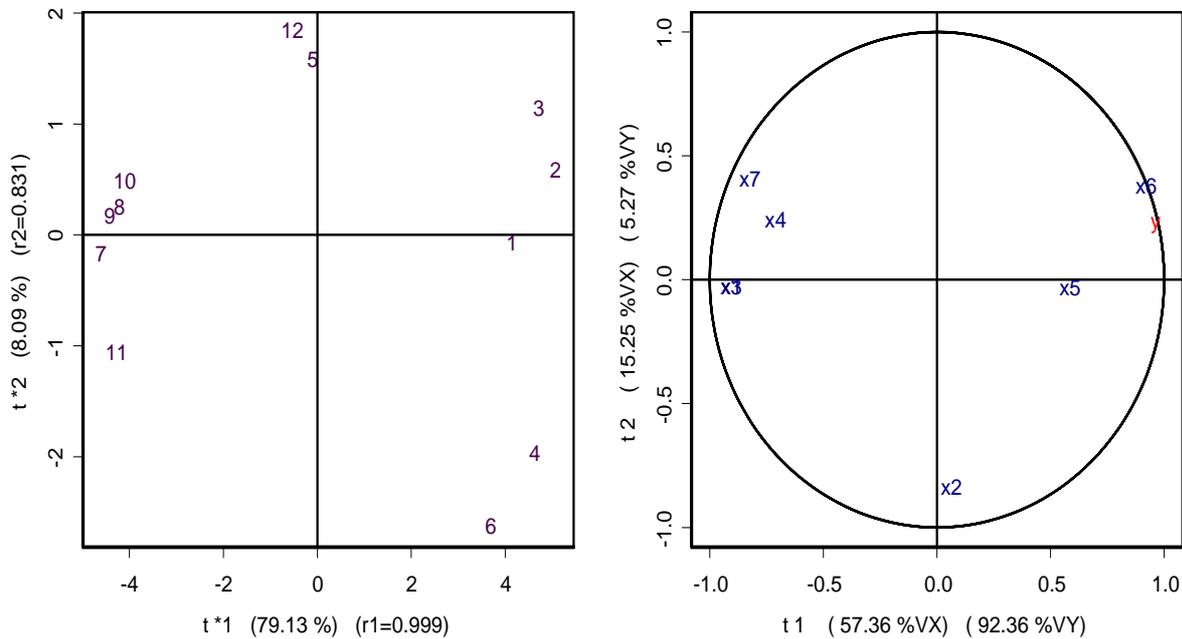


Figure 29 : Plans factoriels (1,2) des individus et des variables, les “sauts de dualité” sont très faibles à faibles, ils correspondent à $r_1 = 0.999$ et $r_2 = 0.831$.

Toutes les variables sont bien représentées sur le plan (1,2) sauf X^5 qui est très peu corrélée avec les autres variables. Pour chaque axe de projection des variables, est indiqué, grâce à (9.52.c) et (9.52.d), le pourcentage de la variance reconstruite. Par exemple, t^1 apporte 57.36% de la variance de X et 92.36% de celle de Y .

Représentation des individus

Rappelons que le plan factoriel (i, j) est la projection \mathbb{V} -orthogonale des individus sur le plan (w^{*i}, w^{*j}) . Les coordonnées du nuage des points projetés sont données par le couple (t^{*i}, t^{*j}) . Les deux plans factoriels (i, j) individus-variables ne sont pas en dualité. On ne peut interpréter un axe-coordonnée t^* , par rapport aux variables projetées sur l'axe-projection t , de même nom, que si le “saut de dualité” entre t^* et t est faible, c’est à dire si le coefficient de corrélation entre ces deux variables est voisin de un.

Pour les données de Cornell, les sauts de dualité mesurés par les coefficients $SD_i = 1 - r_i$ donnés dans le tableau ci-dessous, sont faibles sauf le cinquième

SD_1	SD_2	SD_3	SD_4	SD_5	SD_6
0.001	0.169	0.112	0.008	0.651	0.016
0.999	0.831	0.888	0.992	0.349	0.984
r_1	r_2	r_3	r_4	r_5	r_6

Seul l'axe 5 n'est pas interprétable mais ne présente ici aucun intérêt. Les trois axes utiles sont interprétables et la Figure 30 montre les différences pour les individus entre la représentation pseudo-factorielle (t^1, t^2) et la représentation factorielle (t^{*1}, t^{*2}) .

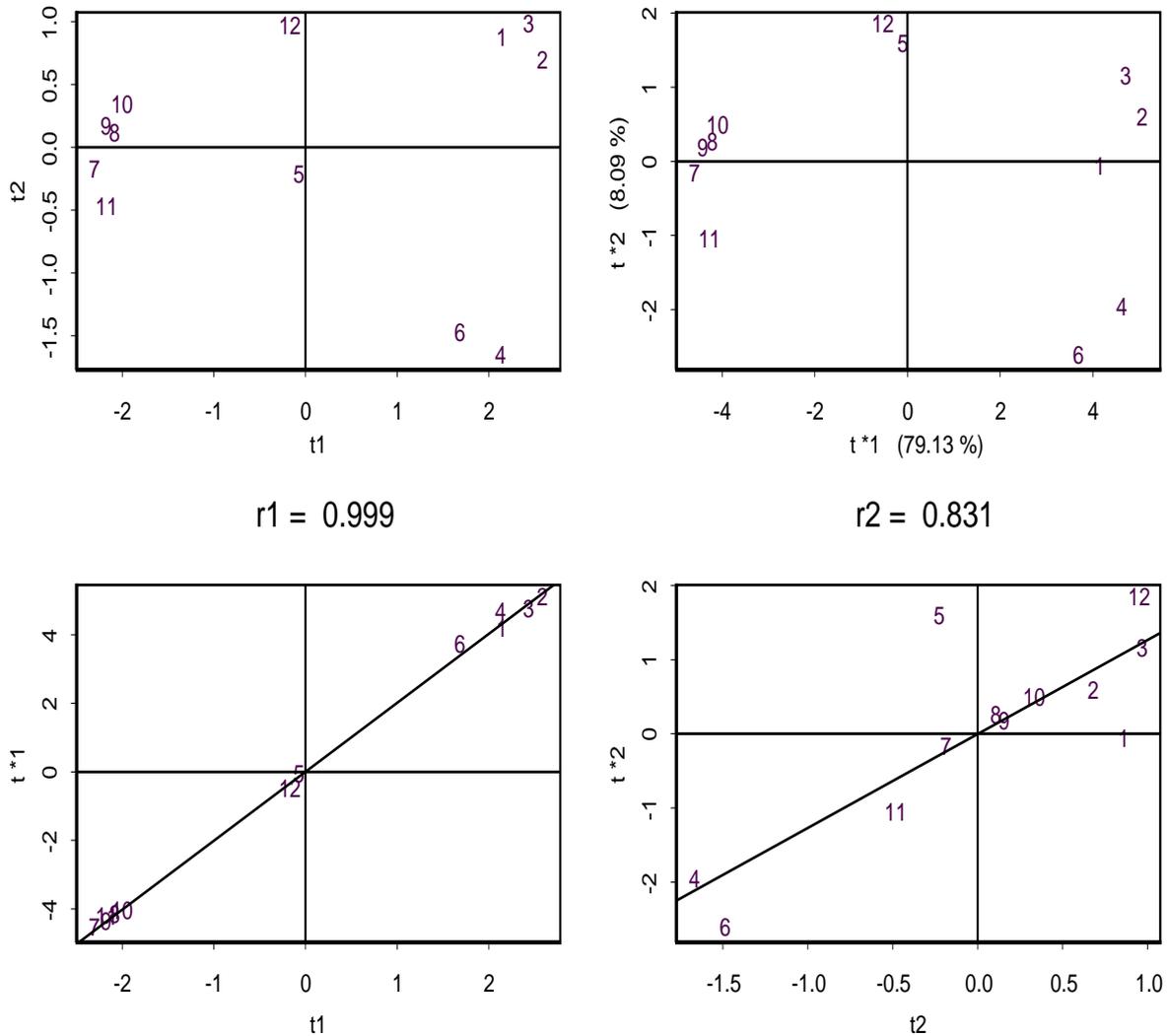


Figure 30 : Comparaison entre les représentations pseudo-factorielles et factorielles pour le plan (1,2) des individus. Visualisation des individus dans les régressions de t^* sur t pour les axes 1 et 2.

Ces deux représentations, globalement très proches à cause des faibles sauts de dualité, diffèrent, cependant, sur quelques individus : l'individu 5 et à un degré moindre, les individus 1,2 et 3. Le graphique en bas à droite, Figure 30, qui montre les observations de la régression simple de t^{*2} sur t^2 , met en évidence les décalages verticaux pour ces individus. Figure 29, le plan (1,2) qui représente 87.2% de l'inertie totale permet de voir les individus 1, 2, 3 qui sont bien représentés comme le montrent les \cos^2 ci-dessous.

	\cos_1^2	\cos_2^2	\cos_3^2	\cos_4^2	\cos_5^2	\cos_6^2
1	0.8951	0.0004	0.0837	0.0207	0.0001	0
2	0.9601	0.0123	0.0094	0.0182	0.0001	0
3	0.8868	0.0511	0.0453	0.0165	0.0002	0
4	0.8211	0.1511	0.0200	0.0078	0.0000	0
5	0.0007	0.1768	0.8224	0.0001	0.0000	0
6	0.4946	0.2524	0.2250	0.0279	0.0001	0
7	0.9893	0.0017	0.0089	0.0000	0.0000	0
8	0.9937	0.0031	0.0001	0.0032	0.0000	0
9	0.9942	0.0013	0.0045	0.0000	0.0000	0
10	0.9828	0.0129	0.0002	0.0040	0.0000	0
11	0.8245	0.0521	0.1190	0.0043	0.0001	0
12	0.0341	0.4020	0.4106	0.1520	0.0013	0

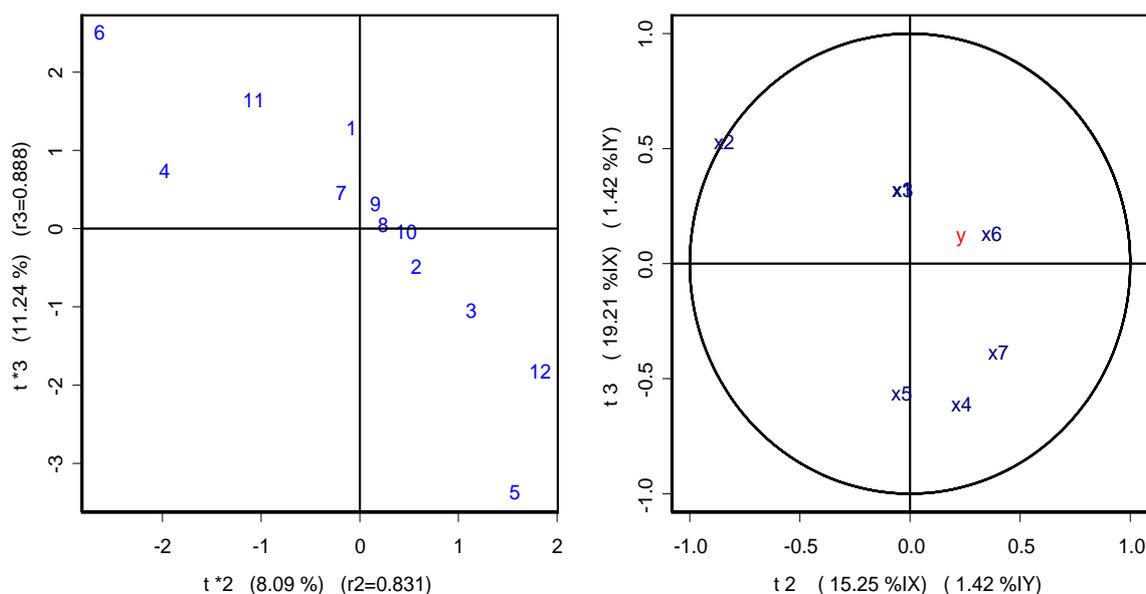


Figure 31 : Plans factoriels (2,3) pour les individus et les variables des données de Cornell. Ils permettent de “voir” les individus 5 et 12 et la variable x^2 .

Sur la Figure 29, ces trois individus se caractérisent par de faibles valeurs pour les variables x^1, x^3, x^4, x^7 , par opposition à de fortes valeurs pour x^6 et y . La Figure 30 est l’homologue de la Figure 28, mais pour les axes 2 et 3. Les individus 12 et 5, bien représentés se caractérisent par de faibles valeurs sur x^2 . Il est à noter que cette analyse de l’individu 5 ne peut se faire par l’examen de la représentation pseudo-factorielle (t^1, t^2). En outre, ce type de carte des individus ne permet pas de quantifier la qualité de la représentation comme il est habituel de procéder dans l’ACP du triplet (X, M, D) .

9.14.2 “Calibration” PLS en spectroscopie proche infrarouge

L'étalonnage multidimensionnel (traduction de l'expression “multivariate calibration”), permet de modéliser sur un échantillon d'apprentissage ou de calibration, (X, Y) , une ou plusieurs réponses, Y , à partir de mesures spectroscopiques ou chromatographiques.

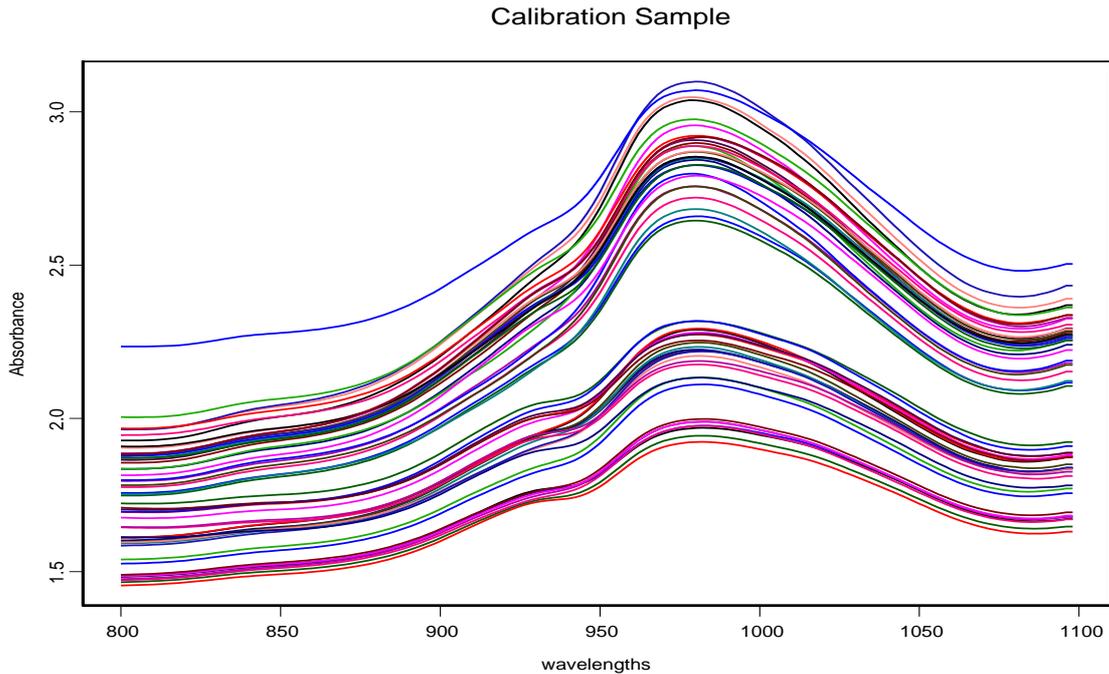


Figure 32 : Spectres d'absorbances proches de l'infrarouge pour les données de Poligny.

La matrices X contient, par exemple, les valeurs discrétisées de n spectres d'absorbance proches de l'infrarouge. En calibration PLS, les p variables explicatives sont d'un type particulier, puisque chacune correspond à une valeur numérique qui est soit une longueur d'onde (“wavelength”) discrétisée, soit un temps discrétisé. La Figure 32, présente l'échantillon d'apprentissage des prédicteurs formé de $n = 56$ spectres qui fournissent les valeurs des prédicteurs $x^1 = 800, x^2 = 802, x^3 = 804, \dots, x^{150} = 1098$.

L'exemple traité provient d'une analyse effectuée au centre de recherche sur le lait à l'INRA de Poligny [9, Mazerolles et al.], des taux d'humidité, Hum , et de matière grasse, Gr , de 56 fromages provenant pour une part du commerce et pour une autre part de fabrication expérimentale, la fabrication de ces derniers ayant eu pour objectif d'augmenter la variabilité de la variable Gr à Hum fixé. L'échantillon d'apprentissage des réponses est donc une matrice Y de dimensions 56×2 et celui des prédicteurs est une matrice X , 56×150 . D'autre part, on dispose d'un échantillon test ou de validation,

formé de 35 fromages d'origine mixte comme pour l'échantillon de calibration.

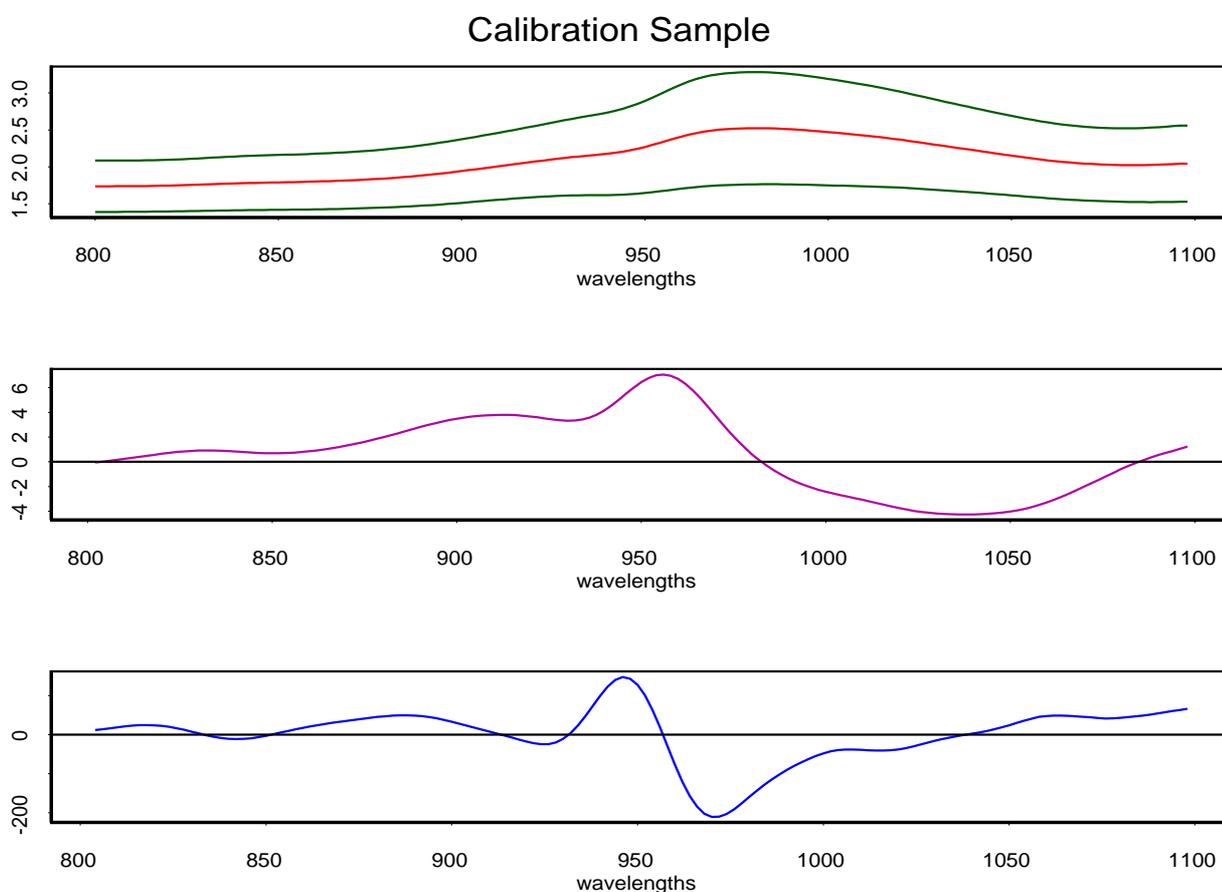


Figure 33 : Du haut vers le bas, le spectre moyen ($\pm 2\sigma$), sa dérivée première et sa dérivée seconde.

Nous remercions G. Mazerolles pour nous avoir aimablement communiqué les données sur les fromages illustrant l'application de PLS en spectroscopie proche de l'infrarouge ("NIR spectroscopy").

Pour une raison d'économie de place, seule est traitée ici la réponse *Hum*, taux d'humidité des fromages. Sur des données issues de la calibration, les prédicteurs étant homogènes, il est habituel de ne pas les standardiser. L'utilisation de la dérivée seconde du spectre moyen, Figure 33, apporte des précisions sur la nature des plages de longueurs d'onde constituant le "massif" centré sur 970 et attribué, clé [9, Mazerolles et al.], au 2^{ème} harmonique de la vibration de l'élongation de la liaison $O - H$.

Cinq composantes sont retenues par la validation croisée et aussi par la prédiction sur l'échantillon de validation.

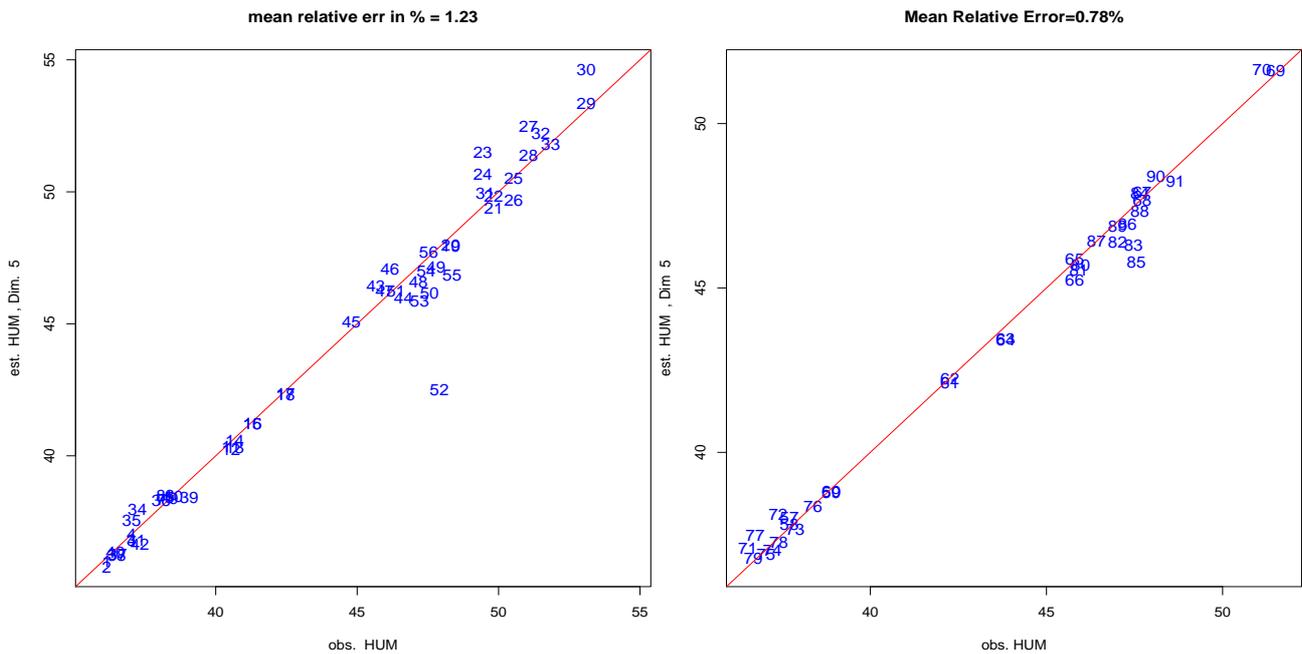


Figure 34 : Hum observé contre Hum prédit par la validation croisée, à gauche, et sur l'échantillon test, à droite.

Figure 34 est présentée une comparaison de la qualité de prédiction à 5 dimensions, par la validation croisée, figure de gauche, et sur l'échantillon test, figure de droite. L'écart type de l'erreur résiduelle est donné, on note que l'individu 52 de l'échantillon d'apprentissage est le moins bien prédit par la validation croisée : c'est le spectre atypique situé au dessus des autres, Figure 32.

Outre la prédiction, un des objectifs de PLS en calibration est de sélectionner les variables explicatives (longueurs d'onde) influentes. Le moyen utilisé consiste à examiner, Figure 35, les coefficients de ces variables dans le modèle obtenu pour détecter les plus grands en valeur absolue. On en déduit que les bandes d'absorption aux environ de 925, 970 et 1020 sont utilisées par le modèle destiné à la prédiction du taux d'humidité. Le signe des coefficients trouve son interprétation, [9, Mazerolles et al.], dans le fait que la bande caractéristique de l'eau, signe positif autour de 970, ne peut être dissociée des bandes caractéristiques de matière sèche, signe négatif autour de 925 pour la matière grasse et et de 1020 pour les protéines.

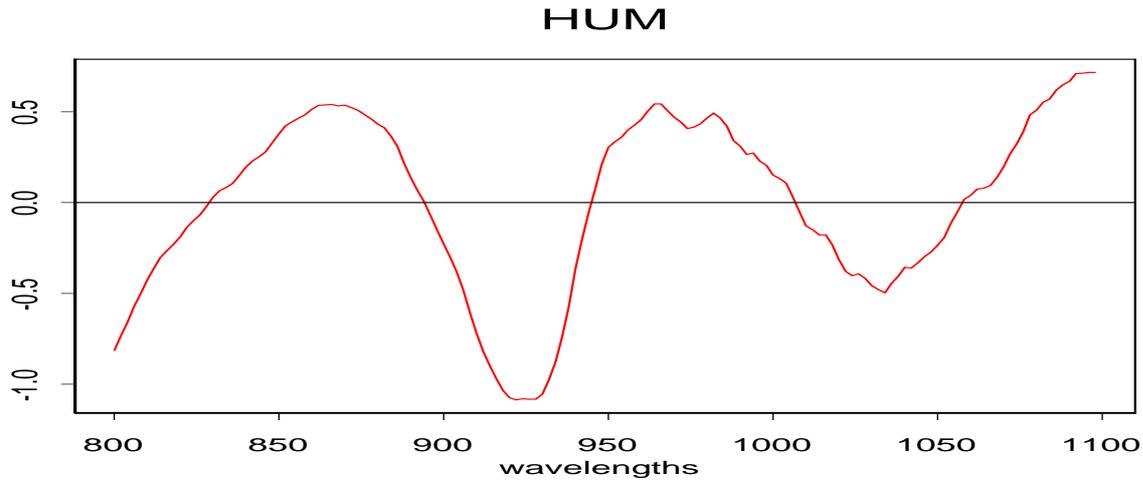


Figure 35 : Évolution des coefficients $\hat{\beta}(5)$ selon les différentes longueurs d'onde.

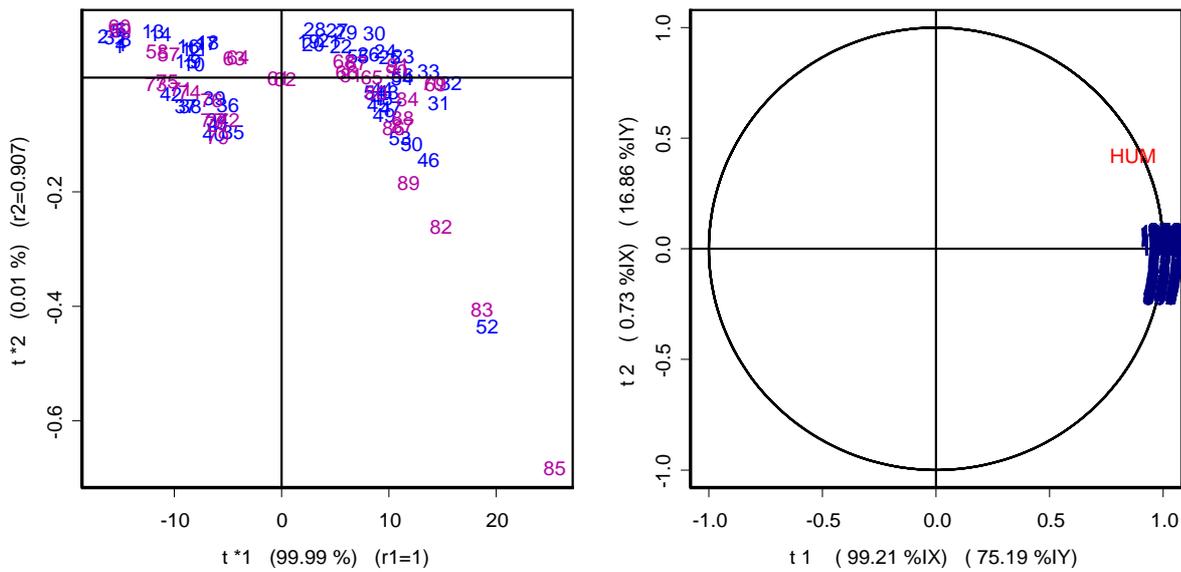


Figure 36 : Représentations factorielles (1,2) des individus et des variables.

Les individus actifs et supplémentaires sont représentés Figure 36, ainsi que le cercle des corrélations. On note que l'axe 1 fournit 99.99% de l'inertie des individus et que les sauts de dualité sont très faibles. Du côté des variables, les axes 1 et 2 apportent 92.05% de la variance de *Hum* ce qui signifie que les trois composantes supplémentaires utilisés par le modèle permettent de capturer une part petite, mais significative, de la variance de la réponse.

9.15 Exercices

Exercice 1 : Images et anti-images

Soit $T \in \mathbb{R}^{n \times p}$ la matrice des n mesures sur p variables. On supposera que chaque colonne T^j de $T = [T^1, \dots, T^p]$ est de moyenne nulle (le poids statistique de chaque observation est $1/n$). La matrice $X = [X^1, \dots, X^p]$ est déduite de T par $X^j = T^j / \sqrt{\sum_i (T_i^j)^2}$. On supposera que le rang de X est égal à p . Les espaces \mathbb{R}^n et \mathbb{R}^p sont munis du produit scalaire usuel $\langle \cdot, \cdot \rangle$ et de la norme Euclidienne $\|\cdot\|$ associée. On note P_X le projecteur orthogonal sur $Im(X)$, le sous espace vectoriel de \mathbb{R}^n engendré par les colonnes de X , et $P_{(-j)}$ le projecteur orthogonal sur $Im(X^{(-j)}) = Im\{X^1, \dots, X^{j-1}, X^{j+1}, \dots, X^p\}$ qui est un s.e.v. de $Im(X)$.

On appelle image, respectivement anti-image, de X^j , le projeté sur $Im(X^{(-j)})$

$$X_I^j = P_{(-j)} X^j,$$

respectivement, le projeté sur l'orthogonal de $Im(X^{(-j)})$

$$X_A^j = (I_n - P_{(-j)}) X^j = P_{(-j)}^\perp X^j.$$

1. Montrer que $\mathbb{V} = X'X$ est la matrice des corrélations empiriques. Montrer que

$$P_X P_{(-j)} = P_{(-j)}, \quad \langle X_I^j, X_A^j \rangle = 0 \quad \text{et} \quad X^j = X_I^j + X_A^j.$$

2. Dans toute la suite on notera $X_I = [X_I^1, \dots, X_I^p]$ et $X_A = [X_A^1, \dots, X_A^p]$.

a) Montrer que le coefficient de détermination $R_{(-j)}^2 = \|X_I^j\|^2 / \|X^j\|^2$ entre X^j et $Im(X^{(-j)})$ vérifie

$$R_{(-j)}^2 = \langle X_I^j, X^j \rangle = X^{j'} X_I^j.$$

b) On note $d_j = 1 - R_{(-j)}^2$ et $D_I = \text{diag}(d_1, \dots, d_p)$. Montrer que

$$X_I = P_X X_I, \quad X_A = P_X X_A \quad \text{et que} \quad X' X_A = D_I.$$

En déduire que

$$X_A = X \mathbb{V}^{-1} D_I, \quad X_A' X_A = D_I \mathbb{V}^{-1} D_I \quad \text{et que} \quad X_I' X_I = \mathbb{V} - 2D_I + D_I \mathbb{V}^{-1} D_I.$$

c) Montrer que les éléments diagonaux de \mathbb{V}^{-1} s'écrivent

$$[\mathbb{V}^{-1}]_{jj} = d_j^{-1}.$$

Discuter les deux cas suivants :

– X^j est non corrélé aux autres variables ;

– X^j est fortement corrélé avec d'autres variables.

Exercice 2 : Un exemple d'école

a) Mettre en oeuvre la régression PLS sur les données suivantes qui constituent un exemple d'école permettant de trouver la solution "à la main"

$$X = \frac{1}{\sqrt{2}} \begin{bmatrix} -2 & 1 \\ 1 & -2 \\ 1 & 1 \end{bmatrix} \quad Y = \sqrt{\frac{3}{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} .$$

b) Effectuer la régression linéaire multiple sur ces données et comparer le modèle obtenu avec ceux de PLS.

Exercice 3 : Équivalence entre la régression PLS1 et la régression non orthogonale de Martens

Nous sommes dans le contexte de la régression multiple, une seule réponse est à prédire à partir de p prédicteurs. Soit y ($n \times 1$) et X ($n \times p$) les matrices de l'échantillon des observations. Toutes les variables étant supposées centrées réduites au sens de la matrice diagonale D des poids statistiques.

1. La régression PLS1

Soit $X_0 = X$ et $y_0 = y$, les composantes $\{t^1, \dots, t^A\}$ sont construites en séquence, l'étape k étant définie par

<p>(1.1) $w^k = \arg \max_{w'w=1} y_{k-1}' D X_{k-1} w$</p> <p>(1.2) $t^k = X_{k-1} w^k$</p> <p>(1.3) $X_k = X_{k-1} - P_{t^k} X_{k-1}$</p> <p>(1.4) $y_k = y_{k-1} - P_{t^k} y_{k-1}$</p>

ou P_{t^k} est la matrice de la projection D -orthogonale sur t^k .

a. Écrire le problème de Lagrange associé à (1) et les équations aux dérivées partielles correspondantes.

b. On note P_k la matrice de la projection D -orthogonale sur le sous espace vectoriel de (\mathbb{R}^n, D) engendré par $\{t^1, \dots, t^k\}$. Dire pourquoi (1.3) et (1.4) s'écrivent respectivement $X_k = (I_n - P_k)X$ et $y_k = (I_n - P_k)y$.

2. La régression PLS non orthogonale de Martens

Soit $\tilde{X}_0 = X$ et $\tilde{y}_0 = y$, les composantes $\{\tilde{t}^1, \dots, \tilde{t}^A\}$ sont construites en séquence, l'étape k étant définie par

$$\begin{array}{l}
 (2.1) \quad \tilde{w}^k = \tilde{X}_{k-1}' D \tilde{y}_{k-1} \\
 (2.1) \quad \tilde{t}^k = \tilde{X}_{k-1} \tilde{w}^k / \tilde{w}^{k'} \tilde{w}^k \\
 (2.3) \quad \tilde{X}_k = \tilde{X}_{k-1} - \tilde{t}^k \tilde{w}^{k'} \\
 (2.4) \quad \tilde{y}_k = \tilde{y}_{k-1} - \tilde{P}_k \tilde{y}_{k-1}
 \end{array}$$

où \tilde{P}_k est la matrice de la projection D -orthogonale sur le sous espace vectoriel de (\mathbb{R}^n, D) engendré par $\{\tilde{t}^1, \dots, \tilde{t}^k\}$.

a. Dans l'espace \mathbb{R}^p des individus munis du produit scalaire usuel, on note $\Pi_{\tilde{w}^k}$ la matrice de projection orthogonale sur la droite vectorielle engendrée par \tilde{w}^k . Montrer que $\Pi_{\tilde{w}^k} \tilde{X}_{k-1}' = \tilde{w}^k \tilde{t}^{k'}$. Quelle est l'interprétation géométrique de la i ème coordonnée \tilde{t}_i^k de \tilde{t}^k ? En déduire que $\tilde{X}_k' = (I_p - \Pi_{\tilde{w}^k}) \tilde{X}_{k-1}'$.

b. Montrer que

$$\tilde{X}_k' = X' - \tilde{w}^1 \tilde{t}^{1'} - \dots - \tilde{w}^k \tilde{t}^{k'}. \quad (2.5)$$

Montrer que \tilde{w}^2 est orthogonal à \tilde{w}^1 et par récurrence que \tilde{w}^k est orthogonal à tous les \tilde{w}^l précédents. En déduire que $\tilde{w}^k = X' D \tilde{y}_{k-1}$. Montrer que $\tilde{w}^{k'} \tilde{X}_k' = 0$ et par (2.5) déduire

$$\tilde{t}^k = X \tilde{w}^k / \tilde{w}^{k'} \tilde{w}^k. \quad (2.6)$$

Dire pourquoi, au contraire de (1.3), (2.3) ne permet pas d'assurer l'orthogonalité des vecteurs \tilde{t}^k , d'où le nom de l'algorithme proposé par Martens.

3. Équivalence des deux régressions

Montrer par récurrence sur l'indice k , que l'algorithme de la régression PLS1 et l'algorithme de Martens sont équivalents au sens suivant

a. $\tilde{w}^k = X_{k-1}' D y$ et $w^k = \tilde{w}^k / \|\tilde{w}^k\|$.

b. Les vecteurs $\{t^1, \dots, t^k\}$ et $\{\tilde{t}^1, \dots, \tilde{t}^k\}$ engendrent le même espace, ce qui implique $P_k = \tilde{P}_k$.

c. $y_k = \tilde{y}_k$.

Bibliographie

- [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] G. Mazerolles, G. Duboz et S. Hugot. *Détermination des taux d'humidité et de matière grasse de fromages type pâte pressée par spectroscopie proche de l'infrarouge en mode transmission*, *Lait* 80, 371-379, 2000.
- [10] C. R. Rao et S. K. Mitra. *Generalized Inverse of Matrices and Its Applications*, Wiley, New-York, 1971.
- [11] G. Saporta. *Probabilités, Analyse de Données et Statistique*, Technip, Paris, 1990.
- [12] M. Tenenhaus. *La régression PLS, Théorie et Pratique*, Technip, Paris, 1998.