F. FERRATY

# Estimations de transformations optimales en ACP curvilinéaire

*Revue de statistique appliquée*, tome 45, nº 1 (1997), p. 5-39
<http://www.numdam.org/item?id=RSA_1997__45_1_5_0>

*Rev. Statistique Appliquée*, 1997, **XLV** (1), 5-39

# ESTIMATIONS DE TRANSFORMATIONS OPTIMALES EN ACP CURVILINÉAIRE

F. Ferraty

*Laboratoire de Statistique et Probabilités, UA CNRS D0745, Université Paul Sabatier, 31062 Toulouse cedex France, email : ferraty@cict.fr*

## RÉSUMÉ

Cet article concerne la recherche de transformations optimales, une par variable, possédant certaines propriétés de régularité et telles qu'elles maximisent le pourcentage de variance expliquée par les $q$ premières composantes principales calculées sur les données transformées. Ces fonctions non linéaires sont estimées dans le cadre d'un modèle curvilinéaire au moyen de différents lisseurs splines : splines de lissage, splines hybrides, splines de lissage monotones et splines hybrides monotones. Une procédure bootstrap adaptée au modèle conduit à un choix optimal du paramètre de lissage.

De plus, en particularisant notre classe de fonctions, le problème d'estimation par les moindres carrés associé à ce modèle permet de retrouver en cas particulier les méthodes liées à la notion de codage flou et poursuivant le même objectif.

Un jeu de données simulées permet de comparer ces algorithmes alors que deux exemples concrets illustrent l'emploi de splines hybrides avec ou sans contrainte de monotonie; nous montrons alors comment une telle pratique rend l'ACP plus performante et précise son interprétation en présence de relations non-linéaires ou de points suspects.

***Mots-clés :*** *Modèle curvilinéaire à effet fixe, réduction de dimension, transformations optimales, splines hybrides, splines monotones, codage flou.*

## ABSTRACT

This article concerns smooth optimal transformations, one per variable, which maximizes the percentage of explained variance by the first $q$ principal components computed on the transformed data. These nonlinear functions are estimated in the context of a curvilinear model by means splines smoothers : smoothing splines, hybrid splines, monotone smoothing splines and monotone hybrid splines. A bootstrap procedure adapted to the model leads to an optimal choice of the smoothing parameter.

Moreover, by particularizing our class of functions, the problem of the least squares estimation allows to introduce several methods into relations with fuzzy coding.

A simulated data set allows to compare these algorithms when two practical examples illustrate the use of hybrid splines and monotone hybrid splines. We display how a such practice improves the interpretation of the PCA in the presence of nonlinear relationships and outliers.

***Keywords :*** *Fixed effect curvilinear model, dimension reduction, optimal transformations, hybrid splines, monotone splines, fuzzy coding.*

## 1. Introduction

Cet article s'intéresse aux méthodes ayant pour but de rechercher des transformations non linéaires des observations afin d'optimiser l'Analyse en Composantes Principales (ACP) des données transformées. Ainsi, l'ACP «non métrique» connue sous le nom de *nonmetric PCA* ou encore appelée parfois ACP non linéaire (De Leeuw *et al.*, 1981, De Leeuw, 1982, Rijckvorsel, 1982, Rijckvorsel, 1988, Gifi, 1990) s'efforçe de construire des codages optimaux de chaque variable. Les solutions peuvent alors être engendrées par une base de fonctions indicatrices (codage disjonctif complet), B-splines (De Boor, 1978, Schumaker, 1981) ou encore I-splines (Winsberg et Ramsay, 1983, Ramsay, 1988). En plus des nombreux paramètres nécessaires à la construction de ces fonctions de codage, l'utilisateur doit donner *a priori* le nombre $q$ ($q < p$ où $p$ désigne le nombre de variables) de composantes principales retenues, le critère à maximiser étant la proportion de variance expliquée par ces $q$ premières composantes principales.

Dans cet article, on introduit ces méthodes comme solution de l'estimation des paramètres du modèle curvilinéaire à effet fixe. Ayant observé $n$ vecteurs aléatoires $\{\mathbf{y}_i; i = 1, \ldots, n\}$ indépendamment distribués et prenant leurs valeurs dans $\mathcal{T} = T_1 \times \cdots \times T_p \subset \mathbb{R}^{\otimes p}$, ce modèle suppose qu'il existe $p$ transformations, une par variable, telles que, après transformation des variables et donc des individus, chaque ligne de la nouvelle matrice ainsi obtenue se décompose en la somme d'un effet fixe et d'une erreur. Les $n$ effets fixes sont supposés appartenir à un sous-espace $E_q$ de $\mathbb{R}^p$ de dimension $q$ ($q < p$), les erreurs aléatoires sont supposées indépendantes et identiquement distribuées. Pour résumer, soit $\mathbf{I}$ la matrice identité de $\mathbb{R}^p$ et soit $f = (f_1, \ldots, f_p)$ une fonction qui au vecteur $\mathbf{x} = [x_1, \ldots, x_p]'$ appartenant à $\mathcal{T}$ associe le vecteur $f(\mathbf{x}) = [f_1(x_1), \ldots, f_p(x_p)]'$ appartenant à $\mathbb{R}^p$. Le modèle curvilinéaire à effet fixe s'écrit alors :

$$\begin{array}{l}
\{\mathbf{y}_i; i = 1, \ldots, n\}\ n \text{ vecteurs aléatoires indépendants dans } \mathbb{R}^p, \\
\text{il existe } f; \forall j \in \{1, \ldots, p\},\ f_j \text{ satisfaisant } (\mathcal{C}_j), \\
\text{il existe un sous-espace } E_q \text{ de dimension } q \text{ tels que :} \\
f(\mathbf{y}_i) = \mathbf{z}_i + \varepsilon_i\ ; i = 1, \ldots, n \text{ et } \forall i \in \{1, \ldots, n\}, \mathbf{z}_i \in E_q \\
\ \text{avec } E(\varepsilon_i) = 0, \text{var}(\varepsilon_i) = \sigma^2 \mathbf{I}.
\end{array} \tag{1}$$

où $(\mathcal{C}_j)$ est un ensemble de contraintes agissant sur le caractère lisse de la fonction $f_j$.

L'estimation des paramètres $\mathbf{z}_i$, $E_q$ du modèle, ainsi que celle des transformations $f_j$, par les moindres carrés, conduit à l'expression de différents problèmes d'optimisation caractérisés par le jeu de contraintes $(\mathcal{C}_j)$ considéré. En conséquence, la recherche d'estimations non paramétriques des transformations optimales mène naturellement à l'emploi de différents types de lisseurs splines.

Besse et Ferraty (1995) proposent un algorithme de résolution appelé SALSA pour *Smoothing Alternative Least Squares Algorithm* dans le cas des fonctions splines de lissage (Wahba, 1990) et conduisant à une solution lorsque le nombre d'individus est restreint. En effet, le calcul de splines de lissage nécessite l'inversion de matrices de grandes tailles souvent mal conditionnées, ce qui peut entraîner une faible précision numérique associée à un temps de calcul élevé.

De plus, il est important de pouvoir, dans certaines situations, imposer une contrainte de monotonie sur les transformations afin de conserver l'ordre des valeurs observées (Winsberg et Ramsay, 1983, Ramsay, 1988).

Pour répondre à ces questions : précision numérique, temps de calcul, contrainte de monotonie, cet article propose d'utiliser les splines hybrides (Kelly et Rice, 1990) qui sont un compromis efficace entre les splines de lissage et les B-splines, ainsi que deux approches pour prendre en compte la contrainte de monotonie.

La section 2 présente donc les différents types de contraintes conduisant aux différents types de lisseurs, la section 3 résout pour chacun d'eux le problème d'optimisation qui lui est associé tandis que la section 4 décrit la version de SALSA qui conduit à la solution. Quant à la section 5, elle étudie les rapports existant entre la démarche proposée et plusieurs méthodes liées à la notion de codage flou en montrant l'équivalence des problèmes d'optimisation associés à chacune d'elle.

L'application des quatres algorithmes proposés aux données du cylindre de Thurstone (1947, p. 117), tous écrits en S+ (Becker *et al.*, 1988), permettent de comparer leurs performances. On montre ainsi que l'algorithme le plus efficace et numériquement le plus fiable utilisant les splines hybrides ne pénalise en rien la qualité des estimations obtenues.

Enfin, deux exemples, avec et sans contrainte de monotonie, montrent comment cette approche non linéaire de l'ACP permet en deux étapes :

1) d'identifier et de mieux comprendre les aspects curvilinéaires ou semi linéaires présents dans les données ainsi que de révéler des individus suspects (outliers),

2) de réaliser et interpréter, sur les données transformées, les graphiques de l'ACP dans le cadre strictement linéaire qui leur est adapté.

## 2. Transformations lisses

### *2.1. Splines de lissage*

Soit $Y^j$ la variable statistique prenant ses valeurs dans l'intervalle $T_j$ inclus dans IR. On note $\mathbf{y}^j = [y_1^j, \ldots, y_n^j]'$ les valeurs prises par cette variable sur un échantillon de taille $n$. Nous considérons l'espace de Sobolev $W_{T_j}^m$ défini par l'ensemble des fonctions de $T_j$ dans IR, absolument continues jusqu'à la dérivée $(m-1)^{\text{ème}}$ et telle que sa dérivée $m^{\text{ème}}$ soit de carré intégrable sur $T_j$ c'est-à-dire appartient à $L_2(T_j)$. Afin de simplifier les notations, nous remplacerons désormais $W_{T_j}^m$ par $W_j$. Pour toute fonction appartenant à $W_j$, nous mesurons sa régularité au moyen d'une semi-norme définie sur cet espace fonctionnel de la manière suivante :

$$\|f_j\|_m^2 = \int_{T_j} \left(f_j^{(m)}(x)\right)^2 dx.$$

La spline de lissage est définie de manière classique (Wahba, 1990) comme la solution du problème suivant :

$$\min_{f_j \in W_j} \left\{ \sum_{i=1}^{n} (z_i^j - f_j(y_i^j))^2 ; \|f_j\|_m^2 \leqslant c_j \right\}, \tag{2}$$

où $\mathbf{z}^j = [z_1^j, \ldots, z_n^j]$ est un vecteur de $\mathbb{R}^n$.

En introduisant le multiplicateur de Lagrange $\rho_{c_j}$, le problème (2) devient :

$$\min_{f_j \in W_j} \left\{ \sum_{i=1}^{n} (z_i^j - f_j(y_i^j))^2 + \rho_{c_j} \|f_j\|_m^2 \right\}.$$

De plus, soit $\widetilde{\mathbf{y}}^j$ le vecteur dont chaque coordonnée est transformée par $f_j$ :

$$\widetilde{\mathbf{y}}^j = f_j(\mathbf{y}^j) = \left[ f_j(y_1^j), \ldots, f_j(y_n^j) \right]' ;$$

avec des notations matricielles, il vient que :

$$\|f_j\|_m^2 = \|\widetilde{\mathbf{y}}^j\|_{\mathbf{N}_j}^2,$$

où $\mathbf{N}_j$ est la matrice $(n \times n)$ provenant de la semi- norme de l'espace de Sobolev. Sa construction dépend de $\mathbf{y}^j$ et fait appel à la notion de noyau reproduisant dans $W_j$ (Hastie et Tibshirani, 1990, Wahba, 1990). De façon plus précise, soit $\varphi$ l'opérateur qui à toute fonction $f$ appartenant à $W_j$ associe $f^{(m)}$ appartenant à $L_2(T_j)$. On considère alors $\theta_j$ (resp. $k_j$) un noyau reproduisant de $\ker\varphi^{\perp}$ (resp. $W_j$) et soit $\mathbf{\Theta}_j$ (resp. $\mathbf{K}_j$) la matrice telle que l'élément situé dans la i$^{\text{ème}}$ ligne et la l$^{\text{ème}}$ colonne soit égale à $\theta_j(y_i, y_l)$ (resp. $k_j(y_i, y_l)$). Ainsi, nous avons :

$$\mathbf{N}_j = \mathbf{K}_j^{-1} \mathbf{\Theta}_j \mathbf{K}_j^{-1}.$$

Notons que la taille des matrices $\mathbf{K}_j$ pour $n$ grand pose à la fois des problèmes de temps de calcul mais surtout de précision lorsqu'il s'agit d'inverser de telles matrices souvent mal conditionnées.

Un premier ensemble $(c = 1)$ de contraintes :

$$(C_j^{(1)}) \Longleftrightarrow \begin{cases} f_j \in W_j, \\ \|f_j\|_m \leqslant c_j, \\ \operatorname{var}(f_j(Y^j)) = 1 \end{cases}$$

est celui considéré par Besse et Ferraty (1995). La régularité de chaque fonction est contrôlée par le paramètre $c_j$ qui est remplacé par $\rho_{c_j}$ dans le problème d'optimisation.

### 2.2. Splines hybrides

Les splines hybrides proposées par Kelly et Rice (1990) puis Champely (1994) dans le contexte de régression non-paramétrique sont une approximation des splines de lissage dans une base de B-splines. Plus précisément, considérons le sous-espace $\mathcal{S}_{k,d}(T_j)$ engendré par la base de B-splines normalisées $\mathcal{B}_{k,d}(T_j) = \{B^j_{d,l}; l = 1, \ldots, r\}$ où $k$ désigne le nombre de nœuds intérieurs de l'intervalle $T_j$, $d$ le degré des splines et $r = k + d + 1$. Alors, avec les notations du paragraphe précédent, nous appelons spline hybride la solution du problème suivant :

$$\min_{f_j \in \mathcal{S}_{k,d}(T_j)} \left\{ \sum_{i=1}^{n} (y_i - f_j(x_i))^2 \; ; \; \|f_j\|_m^2 \leqslant c_j \right\}. \qquad (3)$$

La solution de ce problème apparaît comme une approximation par des B-splines dans $\mathcal{S}_{k,d}(T_j) \subset W_j$ du problème de lissage (2).

Le problème (3) est alors équivalent à :

$$\min_{f_j \in \mathcal{S}_{k,d}(T_j)} \left\{ \sum_{i=1}^{n} (y_i - f_j(x_i))^2 + \rho_{c_j} \|f_j\|_m^2 \right\}.$$

La semi-norme s'écrit dans ce cas :

$$\|f_j\|_m^2 = \|\mathbf{s}_j\|_{\mathbf{G}_j}^2,$$

où $\mathbf{s}_j$ est un vecteur de $\mathbb{R}^r$ et $\mathbf{G}_j$ la matrice $(r \times r)$ induite par le semi-produit scalaire entre les éléments de la base de B-splines évalués aux points $x_i$.

Le principe de ce lissage est de choisir un nombre de nœuds relativement important de sorte que ce soit le paramètre $\rho_{c_j}$ qui contrôle la régularité de la solution comme dans le cas de la spline de lissage classique. En revanche, ce nombre de nœuds et donc $r$ sont choisis de taille raisonnable $(r < n)$ afin de conserver les bonnes propriétés numériques des B-splines usuelles et des temps de calcul plus faibles liés à la taille $(r \times r)$ des matrices.

Nous considérons donc un autre ensemble de contraintes $(c = 2)$ :

$$(\mathcal{C}_j^{(2)}) \Longleftrightarrow \begin{cases} f_j \in \mathcal{S}_{k,d}(T_j), \\ \|f_j\|_m \leqslant c_j, \\ \operatorname{var}(f_j(Y^j)) = 1. \end{cases}$$

### 2.3. Splines de lissage monotones

L'objectif de ce paragraphe est de construire une version monotone d'une spline de lissage en la projetant sur le cône des fonctions monotones. Malheureusement, nous n'avons pas d'expression explicite de cette projection. La démarche consiste alors à calculer une approximation de cette projection en discrétisant le cône considéré. Delecroix *et al.* (1996) décrivent cette procédure dans le contexte de la régression non-paramétrique et donnent des résultats asymptotiques. De plus, grâce à cette technique, il est possible d'obtenir d'autres types de contraintes comme la convexité. Plus précisément, soit $k_j(.,.)$ un noyau reproduisant de $W_j$ et soit $\mathcal{M}_j$ le cône des fonctions décroissantes de $W_j$ c'est-à-dire l'ensemble des fonctions $f_j$ appartenant à $W_j$ telle que, pour tout $t$ appartenant à $T_j$, la dérivée de $f_j$ au point $t$ soit négative. Notons $\{t_i^j; i = 1, \dots, n_j\}$ une discrétisation formée par $n_j$ points de $T_j$ et $\mathcal{M}_{j,n_j}$ la version discrétisée de $\mathcal{M}_j$ :

$$\mathcal{M}_{j,n_j} = \{f_j \in W_j / \forall i, i = 1, \dots, n_j, f_j'(t_i^j) \leqslant 0\}.$$

La proposition suivante permet de calculer la projection de $f_j$ sur le cône discrétisé $\mathcal{M}_{j,n_j}$.

**Proposition 1**

Soit $f_{j,\mathcal{M}_{j,n_j}}$ la projection de $f_j$ sur $\mathcal{M}_{j,n_j}$ ; alors

$$f_{j,\mathcal{M}_{j,n_j}} = f_j - \sum_{i=1}^{n_j} \widehat{a}_i^j \left. \frac{\partial k_j(t,.)}{\partial t} \right|_{t=t_i^j}, \tag{4}$$

où $\widehat{\mathbf{a}}^j = [\widehat{a}_1^j, \dots \widehat{a}_{n_j}^j]$ est le vecteur de $\mathbb{R}^{n_j}$ solution du problème :

$$\min_{a_1^j \geqslant 0, \dots, a_{n_j}^j \geqslant 0} \left\{ \mathbf{a}^{j\prime} \mathbf{M}^j \mathbf{a}^j + \mathbf{a}^{j\prime} \mathbf{b}^j \right\}$$

avec $\mathbf{b}^j$ vecteur de $\mathbb{R}^{n_j}$ dont la $k^{\text{ème}}$coordonnées est égale à $-2f_j'(t_k^j)$ et $\mathbf{M}^j$ la matrice $(n_j \times n_j)$ de terme général

$$[\mathbf{M}^j]_i^k = \left. \frac{\partial^2 k_j(x,y)}{\partial x \partial y} \right|_{(x,y)=(t_i^j, t_k^j)}.$$

Remarquons qu'on s'intéresse ici uniquement au cône des fonctions décroissantes. Pour obtenir la projection de $f_j$ sur le cône discrétisé des fonctions croissantes, il suffit de remplacer dans (4) le signe «$-$» par le signe «$+$» et $\mathbf{b}^j$ par $-\mathbf{b}^j$. Pour plus de précision ainsi que pour la preuve de cette proposition, nous renvoyons le lecteur à Delecroix *et al.* (1996).

Considérons le cas particulier où $f_j$ est une spline de lissage évaluée aux points $y_1^j, \dots, y_n^j$ et soit $H_{j,n}$ le sous-espace de $W_j$ engendré par $\{k_j(., y_i^j); i = 1, \dots, n\}$. Alors, en reprenant les notations du paragraphe 2.1 et en posant

$$\mathbf{e}^j = \mathbf{K}_j^{-1} \widetilde{\mathbf{y}}^j,$$

pour tout $t$ appartenant à $T_j$ :

$$f_j(t) = \sum_{i=1}^{n} e_i^j k_j(t, y_i^j),$$

où $e_i^j$ désigne la $i^{\text{ème}}$coordonnée du vecteur $\mathbf{e}^j$ de $\mathbb{R}^n$. Ainsi, nous pouvons écrire :

$$f_j'(t_k^j) = \sum_{i=1}^{n} e_i^j \left( \left. \frac{\partial k_j(t, y_i^j)}{\partial t} \right|_{t=t_k^j} \right)$$

et en déduire aisément les valeurs de $\mathbf{b}^j$ puis celles des $\widehat{\mathbf{a}}^j$ en utilisant un algorithme de type gradient projeté.

Ce type de splines est associé au jeu de contraintes ($c = 3$) suivant :

$$(\mathcal{C}_j^{(3)}) \iff \begin{cases} f_j \in W_j, \\ f_j \text{ monotone } (f_j' \geqslant 0 \text{ ou } f_j' \leqslant 0), \\ \|f_j\|_m \leqslant c_j, \\ \text{var}(f_j(Y^j)) = 1. \end{cases}$$

### *2.4. Splines hybrides monotones*

Cette seconde méthode s'appuie sur la construction de splines hybrides calculées cette fois-ci à partir d'une base de I-splines (Winsberg et Ramsay, 1983) afin d'inclure la contrainte de monotonie. Avec les mêmes notations qu'au paragraphe 2.2, les I-splines se définissent analytiquement de la façon suivante :

$$I_{d,l}^j(t) = \int_{a_j}^{t} B_{d-1,l}^j(u)du,$$

quelque soit $l$ variant de 1 à $r-1$ et quelque soit $t$ dans $T_j = [a_j, b_j]$; $\{B_{d-1,l}^j; l = 1, \ldots, r-1\}$ est la base de B-splines normalisées (Schumaker, 1981) qui engendre $\mathcal{S}_{k,d-1}(T_j)$. Puisque les B-splines prennent toujours des valeurs positives, les I-splines sont des fonctions croissantes. Soit alors $\mathcal{I}_{k,d}(T_j) = \{I_{d,l}^j; l = 1, \ldots, r-1\}$ une base de I-splines et $\mathcal{J}_{k,d}(T_j)$ le sous-espace fonctionnel de $W_j$ engendré par cette base. Par analogie avec les splines hybrides, nous proposons de résoudre le problème d'optimisation suivant :

$$\min_{f_j \in \mathcal{J}_{k,d}(T_j)} \left\{ \sum_{i=1}^{n} (y_i - f_j(x_i))^2 + \rho_{c_j} \|f_j\|_m^2 \right\},$$

en imposant à $f_j'$ d'avoir un signe constant. La solution de ce problème est une approximation par I-splines de la spline de lissage obtenue sous contrainte de monotonie. Naturellement, cette fonction monotone hérite des même propriétés que celles des splines hybrides : un seul paramètre contrôle la régularité et son calul est rapide. D'autre part, la semi-norme peut s'écrire :

$$\|f_j\|_m^2 = \|\mathbf{u}\|_{\mathbf{H}_j}^2,$$

où $\mathbf{u}$ est un vecteur de $\mathbb{R}^{(r-1)}$ et $\mathbf{H}_j$ la matrice $((r-1)\times(r-1))$ induite par le semi-produit scalaire entre les éléments de la base de I-splines évalués aux points $x_i$.

Ce dernier type de splines est associé aux contraintes $(c = 4)$ :

$$(\mathcal{C}_j^{(4)}) \iff \begin{cases} f_j \in \mathcal{J}_{k,d}(T_j), \\ f_j \text{ monotone } (f_j' \geqslant 0 \text{ ou } f_j' \leqslant 0), \\ \|f_j\|_m \leqslant c_j, \\ \operatorname{var}(f_j(Y^j)) = 1. \end{cases}$$

## 3. Estimation

Dans ce paragraphe, nous nous intéressons à l'estimation des paramètres du modèle curvilinéaire à effet fixe associé aux différents jeux de contraintes. Dans ce but, considérons $\mathbf{Y}$ (resp. $\mathbf{Z}$) la matrice dont les lignes sont notées $\mathbf{y}_i$ (resp. $\mathbf{z}_i$) et $\mathbf{D}$ la matrice diagonale des poids des individus :

$$\{w_i \in [0,1]; i = 1,\ldots,n\}, \sum_{i=1}^{n} w_i = 1, \mathbf{D} = \operatorname{diag}(w_1,\ldots,w_n).$$

Une estimation par les moindres carrés pondérés nous conduit à résoudre le problème suivant :

$$\min_{\mathbf{z}^j, f_j} \left\{ \sum_{j=1}^{p} \|f_j(\mathbf{y}^j) - \mathbf{z}^j\mathbf{D}^2 \; ; f_j \text{ vérifiant } (\mathcal{C}_j^{(c)}) \text{ et } \operatorname{rang}(\mathbf{Z}) = q \right\} \tag{5}$$

avec $c = 1$ si on emploie des splines de lissage, $c = 2$ si on prend des splines hybrides, $c = 3$ si on impose aux splines de lissage d'être monotones et finalement, $c = 4$ si on veut obtenir des splines hybrides monotones.

Besse et Ferraty (1995) ont montré que le problème (5) est équivalent au problème :

$$\min_{\mathbf{Q}, f_j} \left\{ \sum_{j=1}^{p} \|f_j(\mathbf{y}^j) - \mathbf{Q} f_j(\mathbf{y}^j)\|_{\mathbf{D}}^2 \right\}. \tag{6}$$

avec les mêmes contraintes agissant sur $f_j$ et où $\mathbf{Q}$ est une matrice de projection $\mathbf{D}$-orthogonale de rang $q$. En remarquant que

$$\sum_{j=1}^{p} \|\mathbf{Q}f_j(\mathbf{y}^j)\|_{\mathbf{D}}^2 + \sum_{j=1}^{p} \|f_j(\mathbf{y}^j) - \mathbf{Q}f_j(\mathbf{y}^j)\|_{\mathbf{D}}^2 = \sum_{j=1}^{p} \|f_j(\mathbf{y}^j)\|_{\mathbf{D}}^2 = p, \qquad (7)$$

il est aussi équivalent au problème :

$$\max_{\mathbf{Q},f_j} \left\{ \sum_{j=1}^{p} \|\mathbf{Q}f_j(\mathbf{y}^j)\|_{\mathbf{D}}^2 \right\}, \qquad (8)$$

avec $\mathbf{Q}$ et $f_j$ obéissant aux mêmes contraintes que précédemment; cette dernière formulation nécessite moins de calculs.

Compte tenu de (7), le problème de maximisation (8) revient à trouver les transformations $f_j$ telles que le pourcentage de variance expliquée par les $q$ premiers axes fournis par l'ACP des variables transformées $f_1(\mathbf{y}^1), \ldots, f_p(\mathbf{y}^p)$ soit maximum.

En ce qui concerne les splines de lissage sans contrainte de monotonie ($c = 1$), la procédure d'estimation ainsi que l'algorithme de moindres carrés alternés correspondant appelé $\text{SALSA}_{(1)}$ sont décrits dans Besse et Ferraty (1995). Il suffit alors d'ajouter l'étape de projection sur un cône décrite précédemment et conduisant à $\text{SALSA}_{(3)}$ pour obtenir des estimations monotones des transformations par splines de lissage.

Les deux paragraphes suivants détaillent la procédure d'estimation non-paramétrique lorsqu'on considère les splines hybrides avec ou sans contrainte de monotonie.

### *3.1. Cas des splines hybrides*

Le problème à résoudre intégrant les contraintes $\mathcal{C}_j^{(2)}$ s'écrit :

$$\max_{\mathbf{Q},f_j} \left\{ \sum_{j=1}^{p} \|\mathbf{Q}f_j(\mathbf{y}^j)\|_{\mathbf{D}}^2 ; f_j \in \mathcal{S}_{k,d}(T_j),\ \|f_j\|_{\mathbf{m}}^2 \leqslant c_j \right\} \qquad (9)$$

sous les contraintes que les variances des variables transformées $f_j(Y^j)$ soient égales à 1 et que la matrice $\mathbf{Q}$ soit celle d'une projection $\mathbf{D}$-orthogonale de rang $q$.

Soit $\mathbf{B}^j$ la matrice $(n \times r)$ de terme général

$$[\mathbf{B}^j]_i^l = B_{d,l}^j(y_i^j)$$

et $\mathbf{G}_j$ la matrice $(r \times r)$ telle que

$$[\mathbf{G}_j]_h^l = \int_{T_j} B_{d,h}^{j(m)}(t) B_{d,l}^{j(m)}(t)dt, \forall h = 1, \ldots, r \text{ et } \forall l = 1, \ldots, r.$$

**Proposition 2**

En utilisant les notations précédentes, le problème (9) est équivalent à :

$$\max_{\mathbf{Q},\mathbf{s}_j \in \mathbb{R}^r} \left\{ \sum_{j=1}^{p} \left\{ \|\mathbf{Q}\mathbf{B}^j \mathbf{s}_j\|_{\mathbf{D}}^2 - \rho_j \|\mathbf{s}_j\|_{\mathbf{G}_j}^2 \right\} \right\} \tag{10}$$

*Preuve*

Comme les fonctions $f_j$ sont des éléments de $\mathcal{S}_{k,d}(T_j)$, elles s'écrivent comme des combinaisons linéaires des éléments de la base de B-splines :

$$f_j = \sum_{l=1}^{r} s_j^l B_{d,l}^j(.), \tag{11}$$

dont la formulation matricielle s'écrit :

$$f_j(\mathbf{y}^j) = \widetilde{\mathbf{y}}^j = \mathbf{B}^j \mathbf{s}_j.$$

En utilisant l'expression (11), la semi-norme de $f_j$ peut se réécrire de la façon suivante :

$$\|f_j\|_{\mathbf{m}}^2 = \int_{T_j} (\sum_{h=1}^{r} s_j^h B_{d,h}^j(t))^{(m)} (\sum_{l=1}^{r} s_j^l B_{d,l}^j(t))^{(m)} dt$$

ce qui entraîne :

$$\|f_j\|_{\mathbf{m}}^2 = \sum_{h=1}^{r} \sum_{l=1}^{r} s_j^h s_j^l [\mathbf{G}_j]_h^l.$$

On déduit aisément de cette dernière écriture que :

$$\|f_j\|_{\mathbf{m}}^2 = \|\mathbf{s}_j\|_{\mathbf{G}_j}^2.$$

Il suffit d'introduire les multiplicateurs de Lagrange $\rho_j$ pour obtenir l'équivalence entre (9) et (10).

∎

### *3.2. Cas des splines hybrides monotones*

Dans ce but, nous devons résoudre le problème suivant :

$$\max_{\mathbf{Q},f_j} \left\{ \sum_{j=1}^{p} \|\mathbf{Q} f_j(\mathbf{y}^j)\|_{\mathbf{D}}^2 ; f_j \in \mathcal{J}_{k,d}(T_j), \ \|f_j\|_{\mathbf{m}}^2 \leqslant c_j, \right\} \tag{12}$$

sous contrainte de monotonie en ce qui concerne les fonctions $f_j$ en plus de celles énoncées au problème (9) à savoir : les variances des variables transformées $f_j(Y^j)$ égales à 1 et $\mathbf{Q}$ une matrice de projection $\mathbf{D}$-orthogonale de rang $q$.

Soit $\mathbf{I}^j$ la matrice $(n \times (r-1))$ de terme général

$$[\mathbf{I}^j]_i^l = I_{d,l}^j(y_i^j)$$

et $\mathbf{H}_j$ la matrice $((r-1) \times (r-1))$ telle que

$$[\mathbf{H}_j]_h^l = \int_{T_j} I_{d,h}^{j(m)}(t) I_{d,l}^{j(m)}(t)dt, \forall\, h = 1, \dots, r-1 \text{ et } \forall\, l = 1, \dots, r-1.$$

**Proposition 3**

Le problème (12) est équivalent à :

$$\max_{\mathbf{Q}, \mathbf{u}_j \in \mathrm{I\!R}^{(r-1)}} \left\{ \sum_{j=1}^{p} \left\{ \|\mathbf{Q}\mathbf{I}^j \mathbf{u}_j\|_{\mathbf{D}}^2 - \rho_j \|\mathbf{u}_j\|_{\mathbf{H}_j}^2 \right\} \right\}, \tag{13}$$

en imposant d'une part, que les vecteurs $\mathbf{I}^j\mathbf{u}_j$ soient $\mathbf{D}$-normés et d'autre part, que les coordonnées des $\mathbf{u}_j$ soient toutes de même signe.

*Preuve*

Puisque $f_j$ appartient à $\mathcal{J}_{k,d}(T_j)$, sous-espace fonctionnel de $W_j$ engendré par $\mathcal{I}_{k,d}(T_j) = \{I_{d,l}^j; l = 1, \dots, r-1\}$, il existe $\mathbf{u}_j = [u_j^1, \dots, u_j^{(r-1)}]'$ dans $\mathrm{I\!R}^{(r-1)}$ tel que, pour tout $t$ dans $T_j$ :

$$f_j(t) = \sum_{l=1}^{(r-1)} u_j^l I_{d,l}^j(t), \tag{14}$$

ce qui s'écrit matriciellement :

$$f_j(\mathbf{y}^j) = \widetilde{\mathbf{y}}^j = \mathbf{I}^j \mathbf{u}_j.$$

De plus, (14) permet de reformuler la semi-norme de $f_j$ :

$$\|f_j\|_{\mathbf{m}}^2 = \int_{T_j} \Big(\sum_{h=1}^{(r-1)} u_j^h I_{d,h}^j(t)\Big)^{(m)} \Big(\sum_{l=1}^{(r-1)} u_j^l I_{d,l}^j(t)\Big)^{(m)} dt,$$

ce qui implique que

$$\|f_j\|_{\mathbf{m}}^2 = \sum_{h=1}^{(r-1)} \sum_{l=1}^{(r-1)} u_j^h u_j^l [\mathbf{H}_j]_h^l$$

et par conséquent :

$$\|f_j\|^2_{\mathbf{m}} = \|\mathbf{u}_j\|^2_{\mathbf{H}_j}.$$

Puisque les $I^j_{d,l}$ sont des fonctions croissantes, $f_j$ est monotone dès que le signe des coefficients $u^l_j$ est constant. De ce fait, si on impose à toutes les coordonnées du vecteur $\mathbf{u}_j$ d'être positives (resp. négatives), alors $f_j$ est une transformation croissante (resp. décroissante).

Enfin, l'équivalence entre (12) et (13) est obtenue en prenant le lagrangien partiel associé au problème (12).

■

Cependant, afin d'éviter des calculs inutiles, il est possible de déduire l'expression de $\mathbf{H}_j$ en fonction de celle de $\mathbf{G}_j$ dont la construction a été vue au paragraphe précédent. Pour ce faire, remarquons qu'on peut reformuler les I-splines appartenant à $\mathcal{J}_{k,d}(T_j)$ à l'aide de la base de B-splines $\mathcal{B}_{k,d}(T_j)$. En effet, nous avons :

$$I^j_{d,h}(.) = \sum_{b=h+1}^{r} B^j_{d,b}(.).$$

Alors, le semi-produit scalaire entre deux I-splines appartenant à la base $\mathcal{J}_{k,d}(T_j)$ s'écrit :

$$\int_{T_j} I^{j(m)}_{d,h}(t) I^{j(m)}_{d,l}(t) dt = \int_{T_j} \big( \sum_{b=h+1}^{r} B^j_{d,b}(t)\big)^{(m)} \big( \sum_{c=l+1}^{r} B^j_{d,c}(t)\big)^{(m)} dt$$

ce qui implique :

$$[\mathbf{H}_j]^l_h = \sum_{b=h+1}^{r} \sum_{c=l+1}^{r} [\mathbf{G_j}]^{\mathbf{c}}_{\mathbf{b}}$$

pour tout $h = 1, \ldots, r-1$ et $l = 1, \ldots, r-1$.

## 4. Algorithmes

Dans un premier temps, nous donnons l'allure générale de l'algorithme dans lequel s'insèrent les différentes sortes d'estimation décrites précédemment. Il s'agit d'un algorithme de type moindres carrés alternés. Après initialisation, l'algorithme enchaîne itérativement deux étapes. La première consiste à réaliser une décomposition en valeurs singulières de la matrice des données transformées pour estimer le sous-espace optimal de dimension réduite. Cette projection étant fixée, la seconde étape cherche les estimations non-paramétriques des transformations optimales. C'est cette dernière procédure qui diffère suivant le type de contraintes considérées.

L'algorithme enchaîne donc les étapes suivantes :

i) Initialisation : soit $\tilde{\mathbf{Y}}_{(0)} = \left[\tilde{\mathbf{y}}^1_{(0)}|\cdots|\tilde{\mathbf{y}}^p_{(0)}\right]$ la matrice des données initiales; à partir de $\tilde{\mathbf{y}}^j_{(0)}$, on calcule un certain nombre d'éléments nécessaire à la construction des splines considérées.

ii) $k^{\text{ème}}$itération :

2.(a) $\tilde{\mathbf{Y}}_{(k)} = [\tilde{\mathbf{y}}^1_{(k)}|\cdots|\tilde{\mathbf{y}}^p_{(k)}]$ étant fixée, nous résolvons :

$$\max_{\mathbf{Q}} \left\{ \sum_{j=1}^{p} \|\mathbf{Q}\tilde{\mathbf{y}}^{\mathbf{j}}_{(\mathbf{k})}\|^2_{\mathbf{D}} \right\}, \tag{15}$$

où $\mathbf{Q}$ est une matrice de projection $\mathbf{D}$-orthogonale de rang $q$. La solution $\mathbf{Q}_{(k)}$ est donnée par la décomposition en valeurs singulières (DVS) de $\tilde{\mathbf{Y}}_{(k)}$;

$$\hat{\tilde{\mathbf{Y}}}^q_{(k)} = \mathbf{U}_q\mathbf{\Lambda}_q\mathbf{V}'_q \text{ et } \mathbf{Q}_{(k)} = \mathbf{U}_q\mathbf{U}'_q\mathbf{D}.$$

2.(b) étant donné $\mathbf{Q}_{(k)}$, on résout pour $j = 1, \ldots, p$ :

$$\max_{f_j} \{\|\mathbf{Q}_{(k)}f_j(\mathbf{y}^j)\|^2_{\mathbf{D}} - \rho_j\|f_j\|^2_m\} \tag{16}$$

avec $f_j$ satisfaisant $\mathcal{C}^{(c)}_j$.

Notons que la phase 2.(a) correspond à l'approximation d'une matrice par une autre de même dimension mais de rang inférieur fixé à $q$.

### 4.1. SALSA$_{(1)}$

Nous rappelons brièvement la structure de SALSA dans le cas où on s'intéresse à l'estimation des transformations par des splines de lissage. Pour cela, dans la phase d'initialisation, on rajoute pour $j = 1, \ldots, p$, le calcul de la matrice $\mathbf{N}_j$ intervenant dans l'évaluation de la semi-norme $\|f_j\|_m$. Ensuite, 2.(a) restant inchangé, il suffit de remplacer (16) dans la deuxième étape par :

$$\max_{\tilde{\mathbf{y}}^j \in \mathbb{R}^n} \left\{\|\mathbf{Q}_{(k)}\tilde{\mathbf{y}}^j\|^2_{\mathbf{D}} - \rho_j\|\tilde{\mathbf{y}}^j\|^2_{\mathbf{N}_j}; \|\tilde{\mathbf{y}}^j\|^2_{\mathbf{D}} = 1\right\}.$$

Dans ce cas, la solution $\tilde{\mathbf{y}}^j_{(k+1)}$ est donnée par le vecteur propre de la matrice $\mathbf{DQ}_{(k)} - \rho_j\mathbf{N}_j$ de taille $(n \times n)$ correspondant à la plus grande valeur propre (Besse et Ferraty, 1995).

La construction des matrices $\mathbf{N}_j$ nécessaires au calcul des splines de lissage est fournie en partie par l'inversion de matrices $(n \times n)$. Or, celles-ci sont souvent mal conditionnées, ce qui rend les calculs des estimations longs et peu fiables lorsque $n$ est grand $(n > 50)$.

### 4.2. *SALSA*$_{(2)}$

On doit donner *a priori* $(p+2)$ constantes : la dimension $q$ du sous-espace $E_q$ contenant les effets fixes que l'on cherche à estimer, les $p$ paramètres de lissage $\rho_j$ intervenant dans l'estimation non-paramétrique des transformations optimales ainsi que le degré $m$ de l'espace de Sobolev. Cependant, afin d'obtenir des transformations possédant des régularités homogènes, on se limite à un même paramètre de lissage pour l'ensemble de celles- ci. Finalement, on a besoin d'un paramètre $\rho$ pour contrôler la régularité et du degré $m$. Comme la forme du noyau dans le contexte de régression non-paramétrique, le choix du degré $m$ est secondaire. Quant à la structure de SALSA$_{(2)}$, il suffit de la calquer sur celle de la présentation générale de SALSA en ajoutant d'une part, le calcul des matrices $\mathbf{B}^j$, $\mathbf{G}_j$ et $\mathbf{F}_j$ dans l'étape d'initialisation et d'autre part, en remplaçant $(16)$ par :

$$\max_{\mathbf{s}_j \in \mathbb{R}^r} \left\{ \|\mathbf{Q}_{(k)}\mathbf{B}^j\mathbf{s}_j\|^2_{\mathbf{D}} - \rho_j\|\mathbf{s}_j\|^2_{\mathbf{G}_j} \right\} \tag{17}$$

sous la contrainte $\|\mathbf{B}^j\mathbf{s}_j\|^2_{\mathbf{D}} = 1$.

**Proposition 4**

La solution de $(17)$ est donnée par $\mathbf{s}_j^{(k+1)} = \mathbf{F}_j^{-\frac{1}{2}}\mathbf{v}_j^{(k+1)}$ où $\mathbf{v}_j^{(k+1)}$ est le vecteur propre de la matrice

$$\mathbf{F}_j^{-\frac{1}{2}} \left\{ \mathbf{B}^{j'}\mathbf{D}\mathbf{Q}_{(k)}\mathbf{B}^j - \rho_j\mathbf{G}_j \right\} \mathbf{F}_j^{-\frac{1}{2}} \tag{18}$$

de taille $(r \times r)$ associé à la plus grande valeur propre avec $\mathbf{F}_j = \mathbf{B}^{j'}\mathbf{D}\mathbf{B}^j$.

*Preuve*

La solution de $(17)$ maximise la forme quadratique suivante :

$$\|\mathbf{Q}_{(k)}\mathbf{B}^j\mathbf{s}_j\|^2_{\mathbf{D}} - \rho_j\|\mathbf{s}_j\|^2_{\mathbf{G}_j} = \mathbf{s}_j'\mathbf{S}_{\rho_j}\mathbf{s}_j,$$

sous la contrainte $\|\mathbf{B}^j\mathbf{s}_j\mathbf{D}^2 = 1$ et où $\mathbf{S}_{\rho_j}$ est la matrice symétrique

$$\mathbf{S}_{\rho_j} = \mathbf{B}^{j'}\mathbf{D}\mathbf{Q}_{(k)}\mathbf{B}^j - \rho_j\mathbf{G}_j.$$

Pour trouver la solution de ce problème, nous utilisons la méthode des multiplicateurs de Lagrange; soit

$$L = \mathbf{s}_{j'}\mathbf{S}_{\rho_j}\mathbf{s}_j - \lambda\mathbf{s}_{j'}\mathbf{F}_j\mathbf{s}_j$$

le lagrangien associé à notre problème avec $\lambda$ une constante positive; alors :

$$\begin{aligned} \partial L/\partial \mathbf{s}_j = 0 \quad & \Leftrightarrow \mathbf{S}_{\rho_j}\mathbf{s}_j = \lambda\mathbf{F}_j\mathbf{s}_j, \\ & \Leftrightarrow \mathbf{F}_j^{-1}\mathbf{S}_{\rho_j}\mathbf{s}_j = \lambda\mathbf{s}_j, \\ & \Leftrightarrow \mathbf{F}_j^{-\frac{1}{2}}\mathbf{S}_{\rho_j}\mathbf{F}_j^{-\frac{1}{2}}\mathbf{v}_j = \lambda\mathbf{v}_j, \text{ où } \mathbf{v}_j = \mathbf{F}_j^{\frac{1}{2}}\mathbf{s}_j. \end{aligned}$$

De plus,

$$\mathbf{s}_{j'}\mathbf{S}_{\rho_j}\mathbf{s}_j = \lambda$$

et

$$\mathbf{v}_{j'}\mathbf{v}_j = \mathbf{s}_{j'}\mathbf{F}_j\mathbf{s}_j = 1 \Rightarrow \mathbf{v}_{j'}\mathbf{F}_j^{-\frac{1}{2}}\mathbf{S}_{\rho_j}\mathbf{F}_j^{-\frac{1}{2}}\mathbf{v}_j = \lambda.$$

D'où, la solution $\mathbf{s}_j^{(k+1)}$ est donnée par $\mathbf{F}_j^{-\frac{1}{2}}\mathbf{v}_j^{(k+1)}$, où $\mathbf{v}_\mathbf{j}^{(\mathbf{k+1})}$ est le vecteur propre I-orthonormal associé à la plus grande valeur propre de $\mathbf{F}_j^{-\frac{1}{2}}\mathbf{S}_{\rho_j}\mathbf{F}_j^{-\frac{1}{2}}$.

■

Le principal intérêt lié à l'emploi des splines hybrides apparaît dans cette dernière opération. En effet, on calcule ici les éléments propres d'une matrice $(r \times r)$ alors que l'usage des splines de lissage (SALSA$_{(1)}$) nous oblige d'extraire les éléments propres de matrices $(n \times n)$ et d'inverser la matrice $\mathbf{F}_j^{\frac{1}{2}}$. L'algorithme proposé rend ainsi réellement opérationnelles ces techniques non-linéaires.

### *4.3. SALSA*$_{(3)}$

Ce paragraphe décrit l'étape d'estimation non-paramétrique fournissant une version monotone des splines de lissage. Choisissons pour chaque variable $Y^j$ $(j = 1, \dots, p)$ une partition $\{t_1^j, \dots, t_{n_j}^j\}$ de $T_j$ constituée de $n_j$ points distincts et reprenons les notations du paragraphe 2.3; soit $\tilde{\mathbf{y}}^{j+}$ le vecteur de $\mathbb{R}^n$ dont la $i^{\text{ème}}$coordonnée s'écrit $f_{j,\mathcal{M}_{j,n_j}}(y_i^j)$, $\tilde{\mathbf{y}}^j$ le vecteur de $\mathbb{R}^n$ dont la $i^{\text{ème}}$coordonnée est $f_j(y_i^j)$ et enfin soit $\mathbf{D}^j$ la matrice $(n \times n_j)$ telle que :

$$[\mathbf{D}^j]_{ik} = \left.\frac{\partial k_j(t, y_i^j)}{\partial t}\right|_{t=t_k^j}.$$

Alors, (4) s'écrit matriciellement :

$$\tilde{\mathbf{y}}^{j+} = \tilde{\mathbf{y}}^j - \mathbf{D}^j\widehat{\mathbf{a}}^j,$$

et on obtient :

$$\mathbf{b}^j = -2\mathbf{D}^{j\prime}\mathbf{K}_j^{-1}\tilde{\mathbf{y}}^j.$$

Finalement, pour obtenir SALSA$_{(3)}$, on calcule les matrices $\mathbf{N}_j$, $\mathbf{K}_j$, $\mathbf{D}^j$ et $\mathbf{M}^j$ dans la phase initiale, puis, on ajoute à SALSA$_{(1)}$ l'étape 2.(c) qui contient la séquence suivante d'opérations :

$$\begin{aligned}
\mathbf{b}^j_{(k+1)} &= -2\mathbf{D}^{j\prime}\mathbf{K}_j^{-1}\tilde{\mathbf{y}}^j_{(k+1)}, \\
\mathbf{a}^j_{(k+1)} &= \operatorname{Arg}\min_{a_1^j \geqslant 0, \dots, a_l^j \geqslant 0} \left\{\mathbf{a}^{j\prime}\mathbf{M}^j\mathbf{a}^j + \mathbf{a}^{j\prime}\mathbf{b}^j_{(k+1)}\right\}, \\
\tilde{\mathbf{y}}^{j+}_{(k+1)} &= \tilde{\mathbf{y}}^j_{(k+1)} - \mathbf{D}^j\mathbf{a}^j_{(k+1)},
\end{aligned}$$

ce qui fournit une nouvelle matrice $\tilde{\mathbf{Y}}^+_{(k+1)}$ à partir de laquelle on réitère l'étape 2.

### 4.4. SALSA$_{(4)}$

La construction de SALSA$_{(4)}$ est la version monotone de SALSA$_{(2)}$. Cet algorithme est obtenu en remplaçant, pour tout $j = 1, \ldots, p$, $\mathbf{B}^j$ (resp. $\mathbf{G}_j$) par $\mathbf{I}^j$ (resp. $\mathbf{H}_j$); la phase 2.(b) devient :

$$\max_{\mathbf{u}_j \geqslant 0 \text{ ou } \mathbf{u}_j \leqslant 0} \left\{ \|\mathbf{Q}_{(k)}\mathbf{I}^j\mathbf{u}_j\|_{\mathbf{D}}^2 - \rho_j\|\mathbf{u}_j\|_{\mathbf{H}_j}^2 \right\}, \tag{19}$$

sous la contrainte $\|\mathbf{I}^j\mathbf{u}_j\|_{\mathbf{D}}^2 = 1$. Ceci revient à remplacer la base de B-splines par une base de I-splines dans la construction des splines hybrides. De plus, "$\mathbf{u}_j \geqslant 0$ ou $\mathbf{u}_j \leqslant 0$" signifie que l'on cherche une solution dans le cas décroissant ainsi qu'une solution dans le cas croissant; on retient celle qui minimise (19). Une méthode de type gradient conjugué comme pour SALSA$_{(3)}$ permet de résoudre ce problème. Remarquons enfin que, comme les splines hybrides vis-à-vis des splines de lissage, l'usage de splines hybrides monotones met en jeu des matrices de taille moins importante ($(r-1) \times (r-1)$) que celles liées à l'emploi des splines de lissage monotones ($n \times n$).

### 4.5. Convergences

Les algorithmes décrits ci-dessus fonctionnent tous sur le même principe. Ils alternent deux étapes de maximisation où, par construction, le critère

$$\|\mathbf{Q}f_j(\mathbf{y}^j)\|_{\mathbf{D}}^2 - \rho_j\|f_j\|_m^2$$

croît à chaque étape. De plus, ce critère est majoré par le nombre de variables $p$, montrant ainsi la convergence de ces différentes versions de SALSA vers une solution qui peut être locale et donc dépendre du point d'initialisation.

## 5. Modèle curvilinéaire et codage flou

Lorsqu'on s'intéresse à des transformations engendrées par une base de B-splines ou I-splines, on fait intervenir implicitement la notion de codage, appelé aussi codage flou (Lafaye de Michaux, 1978, De Leeuw *et al.*, 1981, De Leeuw, 1982, De Leeuw et Rijckvorsel, 1988, Cazes, 1990). Voici donc quelques brefs rappels sur la notion de codage et sur les transformations définies à partir d'un codage.

Considérons maintenant une subdivision de $T_j = [a_j,\ b_j]$ :

$$a_j = \gamma_j^1 < \cdots < \gamma_j^{m_j+1} = b_j,$$

laquelle définit une partition

$$\left\{ T_j^k = [\gamma_j^k, \gamma_j^{k+1}[\ ;\ k = 1, \ldots, m_j - 1 \text{ et } T_j^{m_j} = [\gamma_j^{m_j}, \gamma_j^{m_j+1}] \right\}$$

de $T_j$.

**Définition 1**

Un codage de $Y^j$ est une application multivoque $\phi^j$ telle que :
$\mathbb{R} \longmapsto \mathbb{R}^{m_j}$
$Y^j \longrightarrow \phi^j(Y) = [\phi_1^j(Y), \ldots, \phi_{m_j}^j(Y)]'$.
Pour $k = 1, \ldots, m_j$, $\phi_k^j$ est appelée fonction de codage relative à $T_j^k$.

L'exemple le plus usité consiste à choisir pour fonction de codage $\phi_k^j$ la fonction indicatrice $\mathbf{1}_{T_j^k}$ de l'intervalle $T_j^k$; on obtient alors le codage disjonctif complet de $Y^j$ utilisé en Analyse Factorielle des Correspondances Multiples (AFCM) pour transformer une variable quantitative en une variable qualitative.

La matrice obtenue par codage notée $\mathbf{\Phi}^j$ a pour terme général

$$[\mathbf{\Phi}^j]_i^k = \phi_k^j(Y(i)).$$

**Définition 2**

Une transformation $h$ de $Y^j$ est une combinaison linéaire des fonctions codages $\phi_k^j$ :

$$h = \sum_{k=1}^{m_j} a_k^j \phi_k^j ; h \in \mathcal{B}^j = \text{vect}\{\phi_1^j, \ldots, \phi_{m_j}^j\}.$$

où vect$\{\phi_1^j, \ldots, \phi_{m_j}^j\}$ désigne le sous-espace vectoriel de $L_2(T_j)$ (classe de fonctions de carré intégrable à valeurs dans $T_j$) engendré par les fonctions de codage $\phi_1^j, \ldots, \phi_{m_j}^j$.

Les propriétés de la transformation ainsi construite sont induites par celles des fonctions de base; ainsi, des transformations étagées conduisent au codage disjonctif complet.

On note $\tilde{\mathbf{y}}^j$ le vecteur $\mathbf{y}^j$ transformé :

$$\tilde{\mathbf{y}}^j = \mathbf{\Phi}^j \mathbf{a}^j,$$

où $\mathbf{a}^j = [a_1^j, \ldots, a_{m_j}^j]'$ est un vecteur de $\mathbb{R}^{m_j}$.

Soit $\|.\|_{\mathbf{D},\mathbf{I}}$ la norme usuelle dans l'espace $\mathcal{M}_{n,p}$ des matrices $n \times p$ :

$$\|\mathbf{X}\|_{\mathbf{D},\mathbf{I}}^2 = \text{tr}\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{I} = \sum_{j=1}^{p} \|\mathbf{x}^j\|_{\mathbf{D}}^2,$$

$\mathbf{x}^j$ étant la $j^{\text{ème}}$colonne de $\mathbf{X}$ avec $j$ variant de 1 à $p$.

Dans la pratique, les fonctions de codage les plus usuelles (procédure *prinqual* de SAS, 1989) peuvent être un ensemble de fonctions indicatrices (codage disjonctif complet), une base de B-splines de degré 1 ou 3, une base de I-splines (splines croissantes) permettant d'engendrer des fonctions monotones. Tous ces exemples de bases dépendent d'une partition, fixée par l'utilisateur, de l'intervalle dans lequel $Y^j$ prend ses valeurs.

Reprenons les notations précédentes et posons :

$$(\mathcal{C}_j) \Leftrightarrow \begin{cases} f_j \in \mathcal{B}^j, \\ \mathrm{var}\left(f_j(Y^j)\right) = 1. \end{cases}$$

Alors, le problème (5) s'écrit :

$$\min_{\mathbf{Z},\mathbf{a}^j} \left\{ \sum_{j=1}^{p} \|\mathbf{\Phi}^j \mathbf{a}^j - \mathbf{z}^j\|^2_{\mathbf{D}} \; ; \; \mathrm{rang}(\mathbf{Z}) = q \text{ et } \|\mathbf{\Phi}^j \mathbf{a}^j\|^2_{\mathbf{D}} = 1 \right\}, \tag{20}$$

et, comme la matrice $\tilde{\mathbf{Y}} = \left[\mathbf{\Phi}^1\mathbf{a}^1 | \cdots | \mathbf{\Phi}^p \mathbf{a}^p\right]$ des données transformées est centrée par construction, $\mathbf{Z}$ l'est également. La contrainte sur la norme des $\mathbf{\Phi}^j\mathbf{a}^j$ (ou une autre du même type) est nécessaire afin d'échapper à une solution triviale.

**Proposition 5**

Les problèmes suivants sont équivalents au problème (20) :

$$\min_{\mathbf{U}_q,\mathbf{T}_q,\mathbf{a}^j} \{ \|\mathbf{\Phi}^j\mathbf{a}^j - \mathbf{U}_q\mathbf{t}_j\|^2_{\mathbf{D}} ; \mathbf{U}'_q\mathbf{D}\mathbf{U}_q = \mathbf{I}_q$$
$$\text{et } \mathbf{T}'_q\mathbf{T}_q \text{ diagonale avec } \|\mathbf{\Phi}^j\mathbf{a}^j\|^2_{\mathbf{D}} = 1 \}, \tag{21}$$

où $\{\mathbf{t}_j ; j = 1, \ldots, p\}$ représentent les $p$ vecteurs lignes de la matrice $\mathbf{T}_q$;

$$\max_{\mathbf{a}^j,\mathbf{Q}_q} \left\{ \mathrm{tr}\mathbf{W}\mathbf{Q}_q ; \mathbf{W} = \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}'\mathbf{D} \text{ avec } \|\mathbf{\Phi}^j\mathbf{a}^j\|^2_{\mathbf{D}} = 1 \right\}, \tag{22}$$

où $\tilde{\mathbf{Y}} = \left[\mathbf{\Phi}^1\mathbf{a}^1 | \cdots | \mathbf{\Phi}^p\mathbf{a}^p\right]$ et $\mathbf{Q}_q$ projection $\mathbf{D}$-orthogonale de $\mathbb{R}^n$ de rang $q$;

$$\max_{\mathbf{a}^j,\mathbf{P}_q} \left\{ \mathrm{tr}\mathbf{R}\mathbf{P}_q ; \mathbf{R} = \tilde{\mathbf{Y}}'\mathbf{D}\tilde{\mathbf{Y}} \text{ avec } \|\mathbf{\Phi}^j\mathbf{a}^j\|^2_{\mathbf{D}} = 1 \right\}, \tag{23}$$

avec $\mathbf{P}_q$ projection orthogonale de $\mathbb{R}^p$ de rang $q$;

$$\max_{\mathbf{a}^j,\mathbf{u}^k} \left\{ \sum_{j=1}^{p} \sum_{k=1}^{q} \mathrm{cor}(\mathbf{u}^k, \mathbf{\Phi}^j\mathbf{a}^j)^2 ; \mathrm{cor}(\mathbf{u}^k, \mathbf{u}^l) = \delta_{k,l} \text{ avec } \|\mathbf{\Phi}^j\mathbf{a}^j\|^2_{\mathbf{D}} = 1 \right\}, \tag{24}$$

avec $\mathbf{U}_q = [\mathbf{u}^1 | \cdots | \mathbf{u}^q]$ et où cor désigne le cœfficient de corrélation.

Ces problèmes équivalents, qui correspondent aux différents formulations adoptées dans la littérature, s'interprètent comme la recherche des transformations des variables initiales qui rendent l'ACP des données transformées optimale au sens de la somme des $q$ premières valeurs propres qui est aussi la somme des coefficients de corrélation canonique entre variables initiales transformées et variables principales (problème (24)).

*Preuve*

i) Montrons (20) $\Longleftrightarrow$ (21)

Par décomposition en valeurs singulières de $\mathbf{Z}$ on a :

$$\mathbf{Z} = \mathbf{U}_q \mathbf{\Lambda}_q \mathbf{V}_q' = \mathbf{U}_q \mathbf{T}_q', \tag{25}$$

avec $\mathbf{U}_q'\mathbf{D}\mathbf{U}_q = \mathbf{V}_q'\mathbf{V}_q = \mathbf{I}_q$ puisque, par hypothèse, rang$(\mathbf{Z}) = q$ et en posant $\mathbf{T}_q = \mathbf{V}_q\mathbf{\Lambda}_q$; on a :

$$\mathbf{T}_q'\mathbf{T}_q = \mathbf{\Lambda}_q \mathbf{V}_q' \mathbf{V}_q \mathbf{\Lambda}_q = \mathbf{\Lambda}_q^2.$$

D'où

$$\|\tilde{\mathbf{Y}} - \mathbf{Z}\|^2_{\mathbf{D},\mathbf{I}} = \mathrm{tr}(\tilde{\mathbf{Y}} - \mathbf{Z})'\mathbf{D}(\tilde{\mathbf{Y}} - \mathbf{Z}) = \sum_{j=1}^{p} (\mathbf{\Phi}^j \mathbf{a}^j - \mathbf{z}^j)'\mathbf{D}(\mathbf{\Phi}^j \mathbf{a}^j - \mathbf{z}^j)$$

et, compte tenu de (25)$\mathbf{z}^j = \mathbf{U}_q \mathbf{t}_j$, il vient que :

$$\|\tilde{\mathbf{Y}} - \mathbf{Z}\|^2_{\mathbf{D},\mathbf{I}} = \sum_{j=1}^{p} \|\mathbf{\Phi}^j \mathbf{a}^j - \mathbf{U}_q \mathbf{t}_j\|^2_{\mathbf{D}}.$$

ii) Montrons (20) $\Longrightarrow$ (22).

Soit $\mathbf{Q}_q$ la projection $\mathbf{D}$-orthogonale de $\mathbb{R}^n$ sur Im$(\mathbf{Z})$ alors :

$$\|\tilde{\mathbf{Y}} - \mathbf{Z}\|^2_{\mathbf{D},\mathbf{I}} = \|\mathbf{Q}_q \tilde{\mathbf{Y}} - \mathbf{Z}\|^2_{\mathbf{D},\mathbf{I}} + \|\mathbf{Q}_q^{\perp} \tilde{\mathbf{Y}}\|^2_{\mathbf{D},\mathbf{I}},$$

et comme

$$\|\mathbf{Q}_q^{\perp} \tilde{\mathbf{Y}}\|^2_{\mathbf{D},\mathbf{I}} = \mathrm{tr}\mathbf{Q}_q^{\perp} \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}' \mathbf{Q}_q^{\perp'} \mathbf{D} = \mathrm{tr}\mathbf{Q}_q^{\perp}\mathbf{W},$$

alors

$$\|\tilde{\mathbf{Y}} - \mathbf{Z}\|^2_{\mathbf{D},\mathbf{I}} = \|\mathbf{Q}_q \tilde{\mathbf{Y}} - \mathbf{Z}\|^2_{\mathbf{D},\mathbf{I}} + \mathrm{tr}\mathbf{Q}_q^{\perp}\mathbf{W}. \tag{26}$$

Soit $\tilde{\mathbf{Y}}^*$ et $\mathbf{Z}^*$ solutions de (20), $\mathbf{W}^*$ la matrice $\tilde{\mathbf{Y}}^*\tilde{\mathbf{Y}}^{*'}\mathbf{D}$ et $\mathbf{Q}_q^*$ la projection sur Im$(\mathbf{Z}^*)$. Alors, si $\mathbf{Q}_q^0$ et $\tilde{\mathbf{Y}}^0$ sont solutions de (22) avec $\mathbf{W}^0$ la matrice $\tilde{\mathbf{Y}}^0\tilde{\mathbf{Y}}^{0'}\mathbf{D}$,

$\mathrm{tr}\mathbf{W}^0\mathbf{Q}_q^0$ atteint le maximum et donc $\mathrm{tr}\mathbf{W}^0\mathbf{Q}_q^{0\perp}$ le minimum car $\mathrm{tr}\mathbf{W}^0 = p$, du fait de la condition de normalisation agissant sur les données transformées :

$$\mathrm{tr}\mathbf{W}^0\mathbf{Q}_q^{0\perp} \leqslant \mathrm{tr}\mathbf{W}^*\mathbf{Q}_q^{*\perp}. \tag{27}$$

On pose $\mathbf{Z}^0 = \mathbf{Q}_q^0\tilde{\mathbf{Y}}^0$, l'équation (26) montre alors :

$$\|\tilde{\mathbf{Y}}^0 - \mathbf{Z}^0\|^2_{\mathbf{D},\mathbf{I}} = \mathrm{tr}\mathbf{Q}_q^{0\perp}\mathbf{W}^0 \leqslant \mathrm{tr}\mathbf{Q}_q^{*\perp}\mathbf{W}^*$$

mais aussi

$$\|\tilde{\mathbf{Y}}^* - \mathbf{Z}^*\|^2_{\mathbf{D},\mathbf{I}} = \|\mathbf{Q}_q^*\tilde{\mathbf{Y}}^* - \mathbf{Z}^*\|^2_{\mathbf{D},\mathbf{I}} + \mathrm{tr}\mathbf{Q}_q^{*\perp}\mathbf{W}^*$$

et donc

$$\|\tilde{\mathbf{Y}}^0 - \mathbf{Z}^0\|^2_{\mathbf{D},\mathbf{I}} \leqslant \|\tilde{\mathbf{Y}}^* - \mathbf{Z}^*\|^2_{\mathbf{D},\mathbf{I}}.$$

Ceci contredit l'hypothèse initiale $\tilde{\mathbf{Y}}^*$ et $\mathbf{Z}^*$ solutions de (20), $\mathrm{tr}\mathbf{W}^*\mathbf{Q}_q^{*\perp}$ est donc le minimum atteint par (22).

De plus, $\mathbf{Z}^* = \mathbf{Q}_q^*\tilde{\mathbf{Y}}^*$ car $\|\mathbf{Q}_q^*\tilde{\mathbf{Y}}^* - \mathbf{Z}^*\|^2_{\mathbf{D},\mathbf{I}} = 0$.

iii) la preuve (22) $\Longrightarrow$ (20) est analogue :

Soit $\tilde{\mathbf{Y}}^*$ et $\mathbf{Q}_q^*$ solutions de (22) qui rendent $\mathrm{tr}\mathbf{W}^*\mathbf{Q}_q^*$ maximum et donc $\mathrm{tr}\mathbf{W}^*\mathbf{Q}_q^{*\perp}$ minimum. On pose $\mathbf{Z}^* = \mathbf{Q}_q^*\tilde{\mathbf{Y}}^*$ et (26) donne :

$$\|\tilde{\mathbf{Y}}^* - \mathbf{Z}^*\|^2_{\mathbf{D},\mathbf{I}} = \mathrm{tr}\mathbf{Q}_q^{*\perp}\mathbf{W}^*.$$

Soit $\tilde{\mathbf{Y}}^0$ et $\mathbf{Z}^0$ solutions de (20) alors :

$$\|\tilde{\mathbf{Y}}^0 - \mathbf{Z}^0\|^2_{\mathbf{D},\mathbf{I}} \leqslant \|\tilde{\mathbf{Y}}^* - \mathbf{Z}^*\|^2_{\mathbf{D},\mathbf{I}}$$

d'où $\|\tilde{\mathbf{Y}}^0 - \mathbf{Z}^0\|^2_{\mathbf{D},\mathbf{I}} \leqslant \mathrm{tr}\mathbf{Q}_q^{*\perp}\mathbf{W}^*$.

Mais, si $\mathbf{Q}_q^0$ est la projection sur $\mathrm{Im}(\mathbf{Z}^0)$, on a $\tilde{\mathbf{Y}}^0 = \mathbf{Q}_q^0\mathbf{Z}^0$, ce qui implique

$$\|\tilde{\mathbf{Y}}^0 - \mathbf{Z}^0\|^2_{\mathbf{D},\mathbf{I}} = \mathrm{tr}\mathbf{Q}_q^{0\perp}\mathbf{W}^0$$

et donc

$$\mathrm{tr}\mathbf{Q}_q^{0\perp}\mathbf{W}^0 \leqslant \mathrm{tr}\mathbf{Q}_q^{*\perp}\mathbf{W}^*.$$

Ceci est impossible car $\mathbf{Q}_q^*$ et $\tilde{\mathbf{Y}}^*$ sont solutions de (22), et donc $\mathbf{Q}_q^*\tilde{\mathbf{Y}}^* = \mathbf{Z}^*$ et $\tilde{\mathbf{Y}}^*$ sont solutions de (20).

iv) D'une manière analogue, on montre que (20) $\Longleftrightarrow$ (23).

v) Montrons (22) $\Longleftrightarrow$ (24) :

$$\mathrm{tr}\mathbf{W}\mathbf{Q}_q = \mathrm{tr}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}'\mathbf{D}\mathbf{Q}_q = \mathrm{tr}\tilde{\mathbf{Y}}'\mathbf{Q}_q'\mathbf{D}\mathbf{Q}_q\tilde{\mathbf{Y}} = \sum_{j=1}^{p} \|\mathbf{Q}_q\tilde{\mathbf{y}}^j\|_{\mathbf{D}}^2$$

avec $\|\mathbf{Q}_q\tilde{\mathbf{y}}^j\|_{\mathbf{D}}^2 = (\mathbf{Q}_q\tilde{\mathbf{y}}^j)'\mathbf{D}\tilde{\mathbf{y}}^j$ et où $\tilde{\mathbf{y}}^j = \Phi^j\mathbf{a}^j$ est la $j^{\text{ème}}$colonne de $\tilde{\mathbf{Y}}$. Soit $\{\mathbf{u}^1, \ldots, \mathbf{u}^q\}$ une base $\mathbf{D}$-orthonormée d'un sous-espace vectoriel de $\mathbb{R}^n$ de dimension $q$; on peut écrire alors :

$$\|\mathbf{Q}_q\tilde{\mathbf{y}}^j\|_{\mathbf{D}}^2 = \sum_{k=1}^{q}(\mathbf{u}^k\mathbf{u}^{k'}\mathbf{D}\tilde{\mathbf{y}}^j)'\mathbf{D}\tilde{\mathbf{y}}^j = \sum_{k=1}^{q}(\mathbf{u}^{k'}\mathbf{D}\tilde{\mathbf{y}}^j)^2 \text{ où } \mathbf{Q}_q = \sum_{k=1}^{q}\mathbf{u}^k\mathbf{u}^{k'}\mathbf{D}.$$

Finalement :

$$\mathrm{tr}\mathbf{W}\mathbf{Q}_q = \sum_{j=1}^{p}\sum_{k=1}^{q} < \mathbf{u}^k, \tilde{\mathbf{y}}^j >_{\mathbf{D}}^2 .$$

■

On a vu précédemment qu'il suffisait de prendre les espaces $\mathcal{B}^j$ engendrés par une base de fonctions indicatrices (adéquates) pour obtenir un codage disjonctif complet de chaque variable. Ainsi, si l'on s'intéresse au problème (24) avec $q = 1$ (on regarde uniquement la plus grande valeur propre), on retrouve la recherche du premier axe de l'AFCM. En revanche, le problème (21) de la proposition (5) minimise la fonction perte de la définition usuelle de la «nonmetric PCA» (De Leeuw et Rijckvorsel, 1988). Il suffit par la suite de prendre les espaces $\mathcal{B}^j$ engendrés par une base de B-splines pour obtenir les résultats de De Leeuw *et al.* (1981).

Notons que les méthodes de codage présentées ici introduisent, en plus de la dimension $q$, de nombreux paramètres pour chaque variable :

– les nombres de nœuds en chaque point, ce qui conditionne la régularité des transformations,

– la position de ces nœuds,

– le degré des splines.

Bien qu'il existe des algorithmes qui optimisent le choix de ces nœuds (procédure MACCA, Rijckvorsel et Tessitore, 1993), le coût en temps de calcul est trop élevé pour pouvoir les appliquer dans notre contexte. En revanche, les estimations non-paramétriques des transformations optimales proposées dans cet article limitent ce nombre important de paramètres que l'on doit fixer *a priori* à un seul : le paramètre de lissage.

## 6. Exemples

### *6.1. Le cylindre de Thurstone*

Dans cette section, nous comparons les résultats obtenus par ces quatre algorithmes sur des données simulées dont la structure est calquée sur celle du cylindre de Thurstone, lesquelles ont déjà fait l'objet de plusieurs études (Kruskal et Shepard, 1974, Winsberg et Ramsay, 1983). On définit une famille paramétrique de cylindres à partir de sa hauteur notée $a$ ainsi que de la superficie de sa base que l'on désigne par $b$. Pour chaque cylindre, on considère les caractéristiques suivantes :

| | |
|---|---|
| 1. hauteur | $a$ |
| 2. superficie de la base | $b$ |
| 3. circonférence | $(2\sqrt{\pi})b^{\frac{1}{2}}$ |
| 4. aire du cylindre sans les bases | $(2\sqrt{\pi})ab^{\frac{1}{2}}$ |
| 5. volume | $ab$ |
| 6. «taux d'étroitesse» | $(1/\sqrt{2\pi})ab^{-\frac{1}{2}}$ |
| 7. angle diagonale-base | $\arctan[(\sqrt{\pi}/2)ab^{-\frac{1}{2}}]$ |
| 8. résistance électrique | $ab^{-1}$ |
| 9. conductance | $a^{-1}b$ |

Pour chacune de ces variables mesurées sur les cylindres, à l'exception de la $7^{\text{ème}}$, une transformation logarithmique permet de les écrire comme des fonctions linéaires des paramètres $\log(a)$ et $\log(b)$.

Finalement, les données simulées concernant $n = 50$ cylindres sont générées comme suit :

i) soit $a_i$ (resp. $b_i$) $n$ nombres pseudo-aléatoires uniformément et indépendemment distribués,

ii) soit $x^j = (x_1^j, \ldots, x_n^j)', 1 = 1, \ldots, 9$ le vecteur contenant $n$ observations prises par les variables précédentes :

$$
\begin{aligned}
x_i^1 &= a_i, \\
x_i^2 &= b_i, \\
x_i^3 &= (2\sqrt{\pi})b_i^{\frac{1}{2}}, \\
x_i^4 &= (2\sqrt{\pi})a_i b_i^{\frac{1}{2}}, \\
x_i^5 &= a_i b_i, \\
x_i^6 &= (1/\sqrt{2\pi})a_i b_i^{-\frac{1}{2}}, \\
x_i^7 &= \arctan[(\sqrt{\pi}/2)a_i b_i^{-\frac{1}{2}}], \\
x_i^8 &= a_i b_i^{-1}, \\
x_i^9 &= a_i^{-1} b_i.
\end{aligned}
$$

iii) soit $\mathbf{y}^j = (y_1^j, \ldots y_n^j)', j = 1, \ldots, 9$ les colonnes de la matrice de données perturbée $\mathbf{Y}$ que nous obtenons après la procédure suivante :

$$
\begin{cases}
y_i^j = \exp(\log(x_i^j) + \varepsilon_i^j), \forall i, i = 1, \ldots, n \text{ et } \forall j, j \neq 7, \\
y_i^7 = \arctan \circ \exp(\log \circ \tan(x_i^7) + \varepsilon_i^7),
\end{cases}
$$

où $\varepsilon_i^j, i = 1, \ldots, n$, sont $n$ observations indépendemment et identiquement distribuées suivant une normale de moyenne nulle et d'écart type $\sigma_j$ correspondant à 25% de celui de la variable originale $\mathbf{x}^j$.

Nous savons ici que les vraies transformations sont la fonction logarithme pour toutes les variables exceptée la septième pour laquelle la transformation optimale est $\log \circ \tan$. De plus, les vrais effets fixes du modèle curvilinéaire sont ici connus; ils engendrent un sous-espace de dimension 2. La figure 1 représente, simultanément pour chaque variable, la vraie transformation et ses estimations obtenues respectivement par les quatre versions de SALSA. Pour chaque algorithme, la valeur optimale du paramètre de lissage ainsi que la meilleure dimension de représentation sont choisies en minimisant la distance euclidienne usuelle entre les vraies transformations calculées sur les données et leurs estimations. Plus précisément, soit $\tilde{\mathbf{Y}}$ la matrice constituée des données transformées par les vraies fonctions et soit $\widehat{\tilde{\mathbf{Y}}}$ son estimation; nous définissons alors le critère de qualité $d$ comme suit :

$$d = (\tilde{\mathbf{Y}} - \widehat{\tilde{\mathbf{Y}}})'\mathbf{D}(\tilde{\mathbf{Y}} - \widehat{\tilde{\mathbf{Y}}}).$$

Nous avons constaté que le critère $d$ indique la dimension 2 comme étant la meilleure, ce qui est cohérent avec les données étudiées. De plus, les algorithmes fournissent des solutions très voisines entre elles, lesquelles sont très proches des vraies transformations. Ainsi, à qualité égale, l'usage de splines hybrides réduit considérablement les temps de calcul par rapport à ceux induits par l'emploi des splines de lissage; au lieu de considérer des matrices de taille $50 \times 50$, on utilise des matrices de taille $29 \times 29$ sachant que $k = 25$ et $d = 3$. Ainsi, dans cet exemple, $\text{SALSA}_{(4)}$ divise par 3 les temps de calcul effectué par $\text{SALSA}_{(2)}$. La figure 2 focalise notre attention sur les variables 3 et 6; même dans ces cas les plus défavorables, les différents lisseurs splines conduisent encore à des estimations très similaires; les splines hybrides fournissent de bonnes approximations. Notons qu'en ce qui concerne la variable 3, les transformations optimales fournies par $\text{SALSA}_{(1)}$ et $\text{SALSA}_{(2)}$ sont décroissantes dans l'intervalle [1,2]. La figure 3 représente le pourcentage cumulatif de variance expliquée de l'ACP classique calculée d'une part sur les données initiales et d'autre part sur les données transformées issues de $\text{SALSA}_{(4)}$. Nous remarquons ainsi que les deux premières composantes principales issues de l'ACP des données transformées expliquent 94% de la variance totale et que nous devons retenir les quatre premiers axes principaux pour obtenir le même pourcentage dans le cas de l'ACP classique. Ainsi, diverses informations, ignorées par l'ACP usuelle, peuvent être révélées par SALSA pour la même dimension fixée.

Dans la pratique, le choix optimal du paramètre de lissage est déterminé par une procédure bootstrap détaillée dans Besse and Ferraty (1995). On considère le risque quadratique

$$\mathcal{L}_2(q,p) = \frac{1}{2}\|\mathbf{P}_q - \widehat{\mathbf{P}}_q\|^2 = q - \mathrm{tr}\mathbf{P}_q\widehat{\mathbf{P}}_q \tag{28}$$

où $\mathbf{P}_q$ (resp. $\widehat{\mathbf{P}}_q$) est la projection orthogonale sur le vrai sous-espace $E_q$ (resp. sur le sous-espace estimé $\widehat{E}_q$). Le critère $\mathcal{L}_2$ mesure la distance entre les projecteurs et

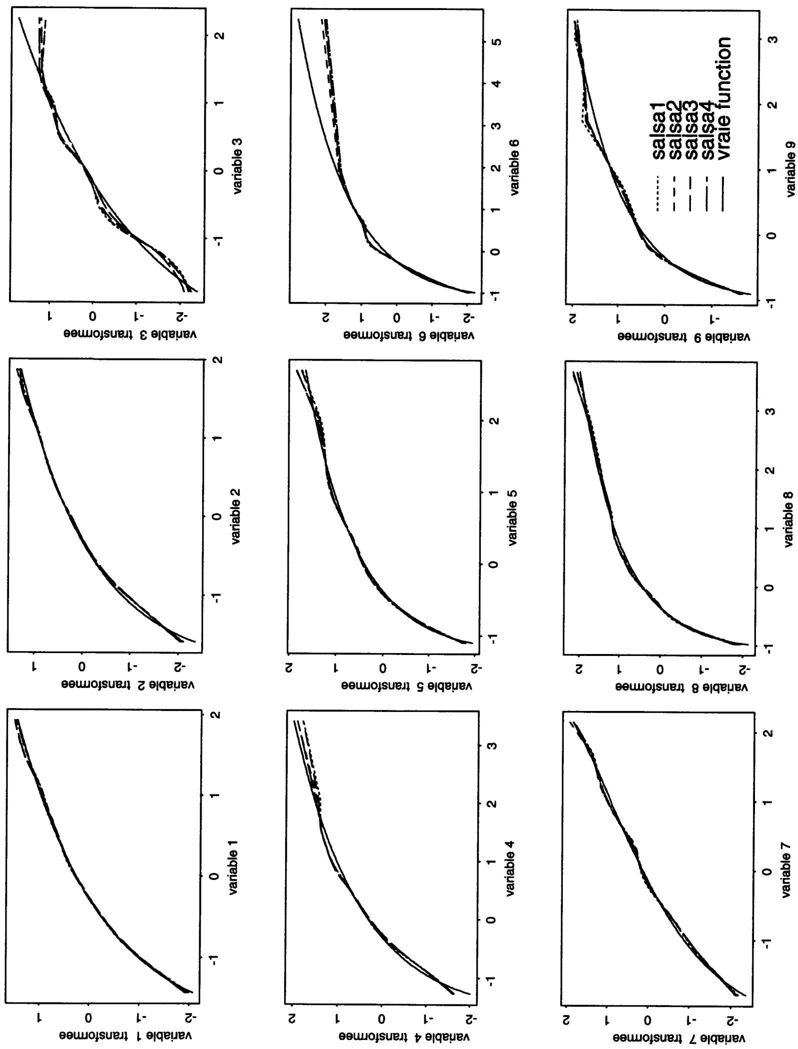


FIGURE 1

*Comparaison pour chaque variable entre la vraie transformation et ses estimations calculées par les quatre versions de SALSA.*

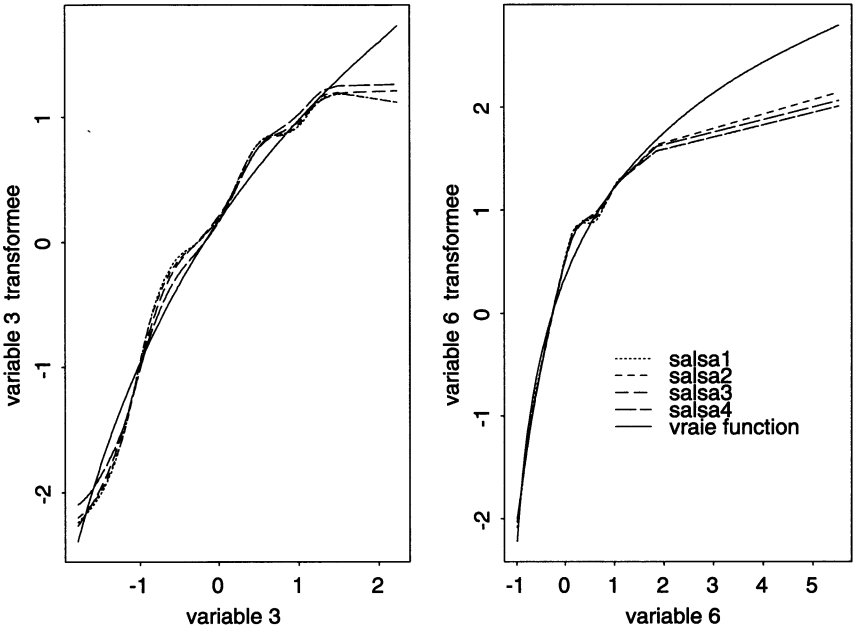


FIGURE 2
*Représentation des transformations agissant sur les variables 3 et 6.*

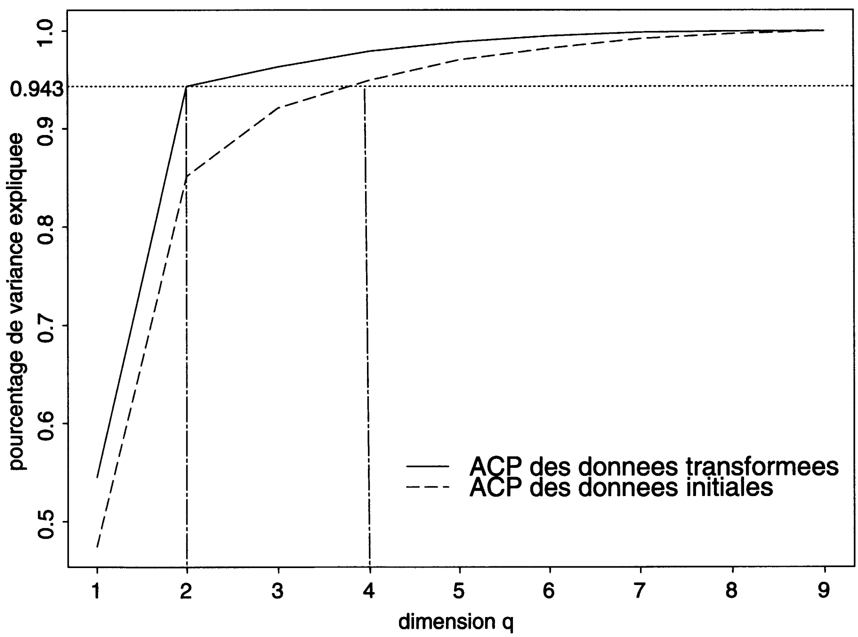


FIGURE 3
*Pourcentage cumulatif de variance expliquée*
*par les q premières composantes principales.*

ainsi, la distance entre les sous-espaces associés. Nous définissons alors le critère de qualité $R_2$ en prenant l'espérance de $\mathcal{L}_2$ :

$$R_2 = E(\mathcal{L}_2).$$

Ce critère est estimé par une procédure bootstrap analogue à celle de Efron et Tibshirani (1986) dans le cas de la régression non-paramétrique. En accord avec le modèle curvilinéaire, on calcule une estimation de $R_2$ en rééchantillonnant sur les résidus $\varepsilon_i$, pour tout $i$ variant de 1 à $n$ :

$$\widehat{R}(q,p) = q - \frac{1}{B}\sum_{b=1}^{B} \mathrm{tr}\widehat{\mathbf{P}}_q^{(b)}\widehat{\mathbf{P}}_q,$$

où $\widehat{\mathbf{P}}_q^{(b)}$ est le projecteur estimé en considérant le $b^{\text{ème}}$ échantillon de résidus. Nous disposons ainsi d'un outil permettant de déterminer une valeur optimale du paramètre de lissage. La figure 4 compare, en dimension $q = 2$, l'estimation issue de $\mathrm{SALSA}_{(4)}$ de ce critère pour différentes valeurs de $\rho$ avec la part de variance expliquée; le minimum est atteint pour une valeur du paramètre de lissage $\rho^{**}$ proche de $7.10^{-4}$. De plus, contrairement au cas non monotone ($\mathrm{SALSA}_{(1)}$ et $\mathrm{SALSA}_{(2)}$) pour lequel le pourcentage de variance expliquée croît au fur et à mesure que $\rho$ décroît, il existe systématiquement une valeur notée ici $\rho^{***}$ qui atteint le maximum; il est intéressant de remarquer :

$$\rho^* \simeq \rho^{**} \simeq \rho^{***},$$

où $\rho^*$ désigne la valeur optimale du paramètre de lissage déduite du critère $d$.

### *6.2. Données socio-économiques*

Cette section illustre l'usage des splines hybrides non monotones pour l'estimation des transformations. Les données concernent les villes de province (de plus de 100 000 habitants) et ont été publiées dans le *Nouvel Observateur* du 23-29 septembre 1993 pour illustrer un article : «Ces villes où il fait bon vivre». Elles se composent des 12 variables suivantes observées sur les 56 individus :

1. Nombre de créations d'entreprise pour 1 000 hab.,
2. Taux de réussite au bac,
3. Taux de chômage du département,
4. Salaire annuel moyen par hab. en francs,
5. Nombre d'entreprises industrielles,
6. Nombre d'entreprises de service,
7. Montant moyen des impôts locaux en francs,
8. Nombre de crimes et délits pour 1 000 hab., par département,
9. Proportion de moins de 20 ans,

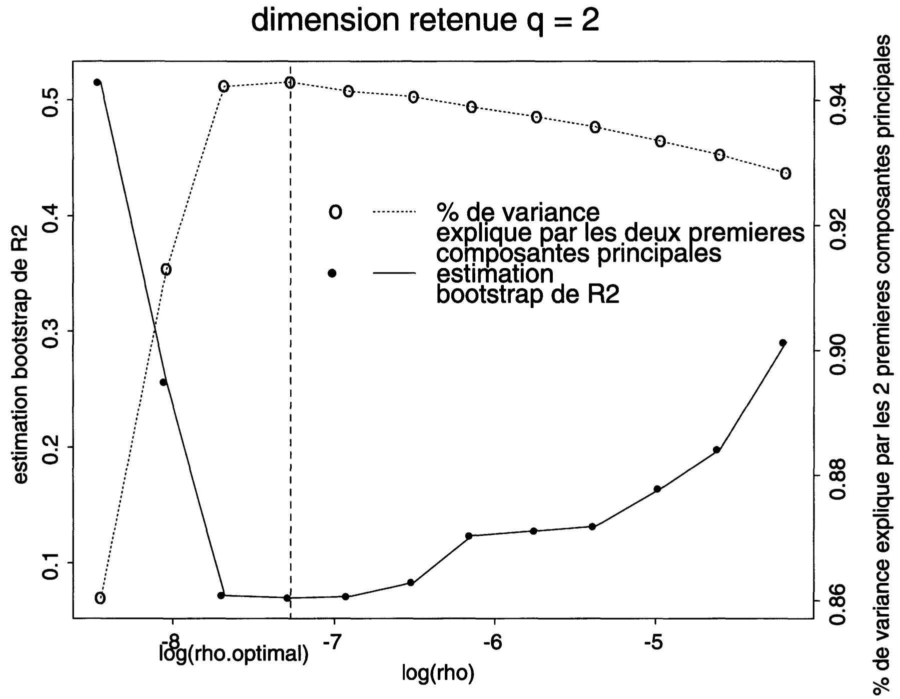


FIGURE 4

*Comparaison entre le comportement du pourcentage de variance expliquée par les 2 premières composantes principales et l'estimation bootstrap de $R_2$.*

10. Recettes moyennes de cinéma par hab. en francs,
11. Nombre de licenciés (sportifs) pour 100 hab.,
12. Quantité de polluants émis par les entreprises.

Au regard des éboulis des valeurs propres (figure 5) issues de l'ACP du tableau initial et de celles après transformation des variables, on constate que la décroissance du spectre de la matrice des covariances des données transformées est plus rapide. Ainsi, on réalise une séparation plus nette entre les $q$ premières valeurs propres retenues et les $p - q$ dernières.

Les résultats concernant l'étude de ces données ont été obtenus avec $q = 2$, la valeur du paramètre de lissage optimal étant fournie par la procédure bootstrap estimant le risque quadratique moyen $R_2$.

La plupart des transformations obtenues sont très linéaires et modifient donc peu les données initiales. En revanche, deux variables sont fortement modifiées.

Focalisons notre attention sur la transformation opérant sur la variable «Nombre de licenciés sportifs» (figure 6); elle met en exergue deux villes : *Annecy* et *Chambéry*, ayant un comportement suspect («outliers») qui n'apparaît que pour cette variable.

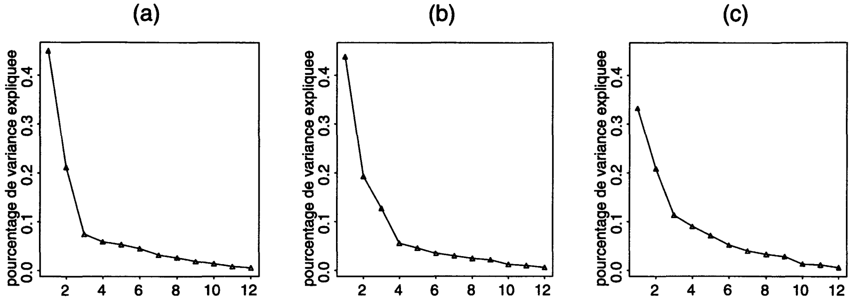


FIGURE 5
*Eboulis des valeurs propres issues de l'ACP du tableau (a) transformé avec $q = 2$, (b) transformé avec $q = 3$, (c) initial.*

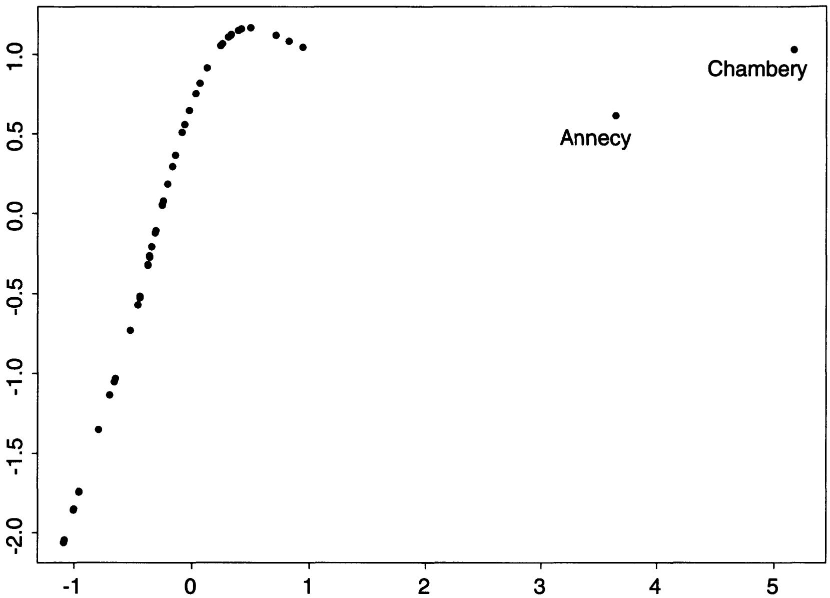


FIGURE 6
*Représentation de la variable transformée «Nombre de licenciés sportifs».*

La transformation a alors pour effet de pénaliser ces points singuliers et le modèle curvilinéaire apparaît, d'une certaine façon, comme une version robuste de l'ACP linéaire.

La transformation opérant sur la variable «Taux de chômage» est représentée figure 7. Elle modifie la hiérarchie initiale des villes relative à cette variable en distinguant deux catégories parmi les villes à fort taux de chômage : l'une concerne les villes du nord (*Bethune, Calais, Douai,...*), l'autre les villes du sud (*Marseille, Montpellier, Nîmes,...*).

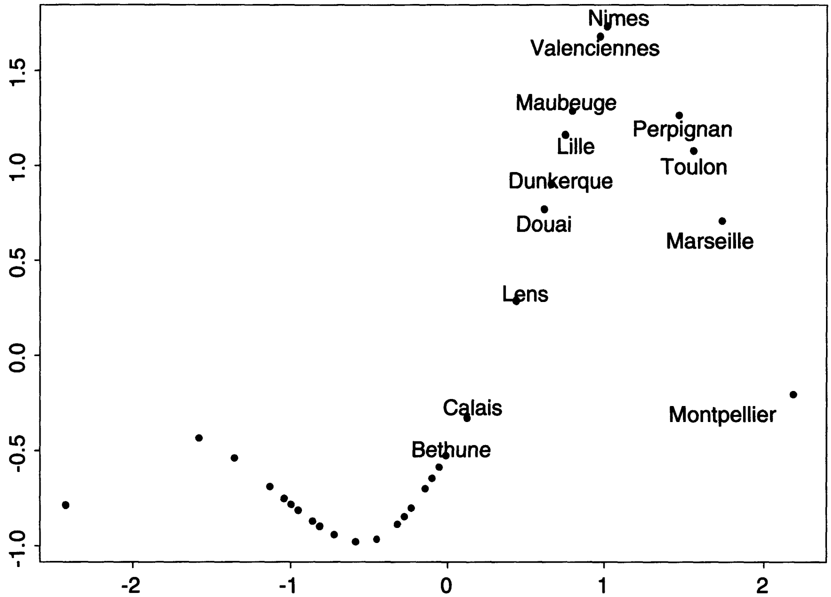


FIGURE 7
*Représentation de la transformation agissant sur la variable «Taux de chômage».*

Cette distinction nord-sud, provient essentiellement des différences structurelles du chômage : le sud participe plus activement que le nord à la corrélation entre le chômage et le nombre d'entreprises (industrielles et de services) et contribue de façon positive à celle entre la proportion de moins de 20 ans et le chômage, alors que le nord agit négativement sur cette quantité.

La représentation des individus (figures 8 et 9) devient alors plus lisible une fois que les comportements atypiques de certaines villes pour certaines variables aient été gommés. C'est le cas, par exemple, d'*Annecy* et *Chambéry*.

### *6.3. Données de pollution*

Nous proposons d'illustrer la démarche dans le cas de la recherche de transformations monotones par l'étude d'un jeu de données (McDonald and Schwing, 1973) concernant 60 agglomérations aux Etats-Unis. On s'intéresse à des variables socio-économiques, climatologiques et à des mesures de la pollution de l'air. Ces variables, au nombre de 15 sont les suivantes :

1. Moyenne annuelle des précipitations en inches,
2. Moyenne des températures de Janvier en degrés Fahrenheit,
3. Moyenne des températures de Juillet en degrés Fahrenheit,
4. Proportion de la population ayant 65 ans et plus,
5. Population par ménage,
6. Nombre médian d'années d'étude effectuées par les plus de 25 ans,

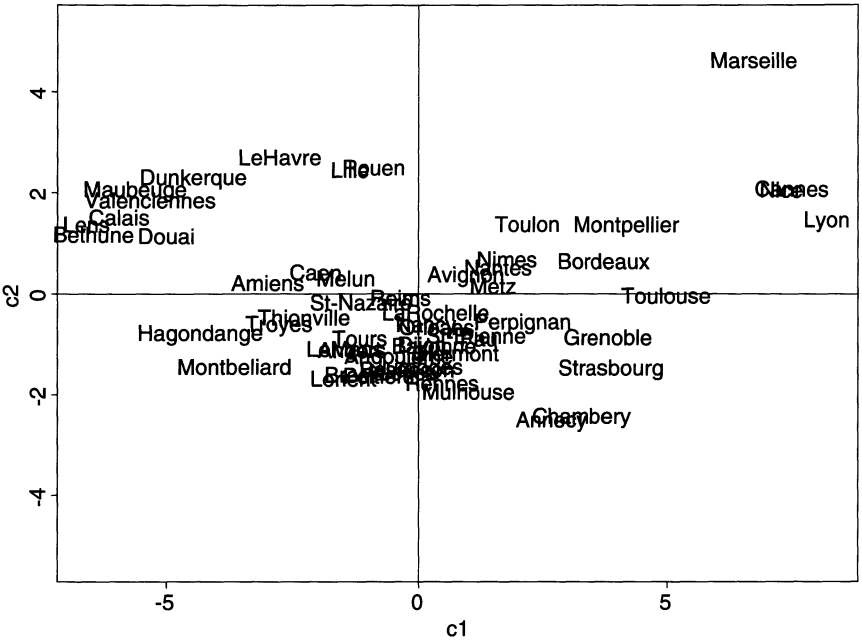


FIGURE 8
*Représentation des individus sur le sous-espace obtenu par ACP classique*

7. Pourcentage de foyers possédant toutes les commodités,

8. Habitants par miles$^2$,

9. Pourcentage d'habitants n'étant pas de race blanche,

10. Pourcentage de la population occupant un emploi de type «colblanc»,

11. Pourcentage de familles ayant un revenu inférieur à 3 000$,

12. Potential de pollution[1] relatif d'oxyde de nitrogène, appelé NOX,

13. Potentiel de pollution relatif de dioxyde de sulfure,

14. Pourcentage relatif d'humidité, moyenne annuelle prise à 13h,

15. Taux de mortalité, toute cause confondue, ajusté à l'âge et exprimé en nombre de décès pour 100 000 habitants.

Les transformations que nous obtenons figure 10 proviennent de la mise en œuvre de $SALSA_{(4)}$. L'estimation non-paramétrique des transformations a donc été réalisée à l'aide de splines hybrides sous contrainte de monotonie. Quant au choix

[1] Le potentiel de pollution est déterminé comme étant le produit entre le nombre de tonnes émises par jour et par km$^2$ de chaque polluant et un facteur de dispersion qui prend en compte l'altitude, la vitesse du vent, le nombre de jour d'atteinte d'un certain seuil et la dimension de chaque agglomération. Chaque agglomération possède le même facteur de dispersion pour chaque polluant.

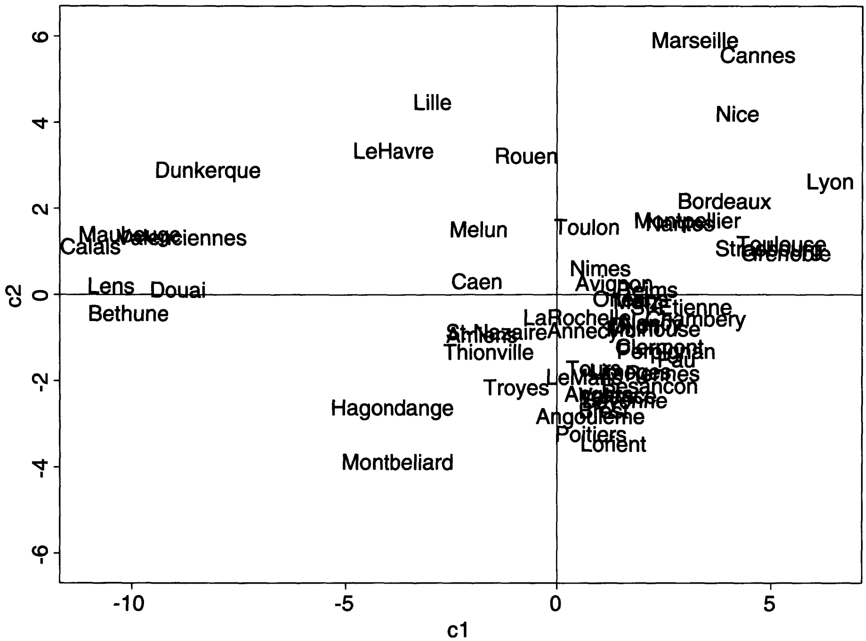


FIGURE 9
*Représentation des individus sur le sous-espace obtenu par ACP réalisée sur les données transformées (avec q fixé à 2).*

optimal du paramètre de lissage, il est fourni par la procédure bootstrap estimant le risque quadratique moyen $R_2$ pour la dimension $q = 3$. La part de variance expliquée par les trois premiers axes principaux issus de l'ACP des données transformées est de 70% contre 60% en ce qui concerne l'ACP classique. La figure 11 compare les nuages des individus projetés sur les deuxième et troisième axes principaux produits par l'ACP des données initiales et de celle des données transformées.

La première étape de l'interprétation considère la figure 10 qui fait apparaître des transformations de différents types. Certaines sont presque linéaires (variables 5,6, 9, ...) ce qui signifie qu'il n'est pas nécessaire, dans ce cas, de modifier les variables pour améliorer la réduction de dimension; l'ACP linéaire est adaptée à la structure de corrélation de ces variables.

D'autres transformations (variable 12, figure 12) présentent une courbure importante, elles révèlent l'existence de point influents (Los Angeles, San Francisco) qui perturbent fortement la structure de corrélation. Une fois identifiés, ces points sont ramenés à des valeurs compatibles avec une étude linéaire. On retrouve dans ce cas une transformation logarithmique habituellement conseillée pour analyser des variables de type concentration (de polluants). Le modèle curvilinéaire est encore sur cet exemple une adaptation robuste de l'ACP linéaire.

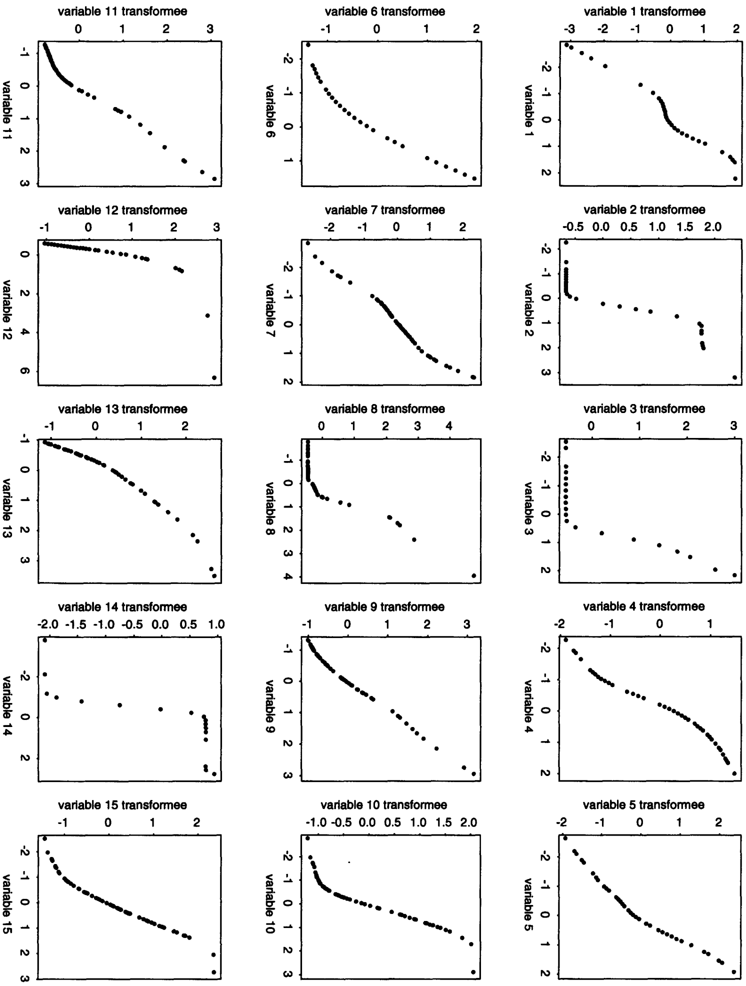


FIGURE 10

*Transformations optimales des données de pollution estimées par $SALSA_{(4)}$*

Enfin, certaines transformations présentent des paliers sans doutes imposés par la contrainte de monotonie. Elles mettent ainsi en évidence le rôle spécifique de ces variables climatologiques (température, humidité) sur la pollution. Sans entrer dans des considérations physico-chimiques, on peut penser que les variables climatologiques présentent des valeurs critiques autour desquelles il y a une forte interaction avec les autres variables.

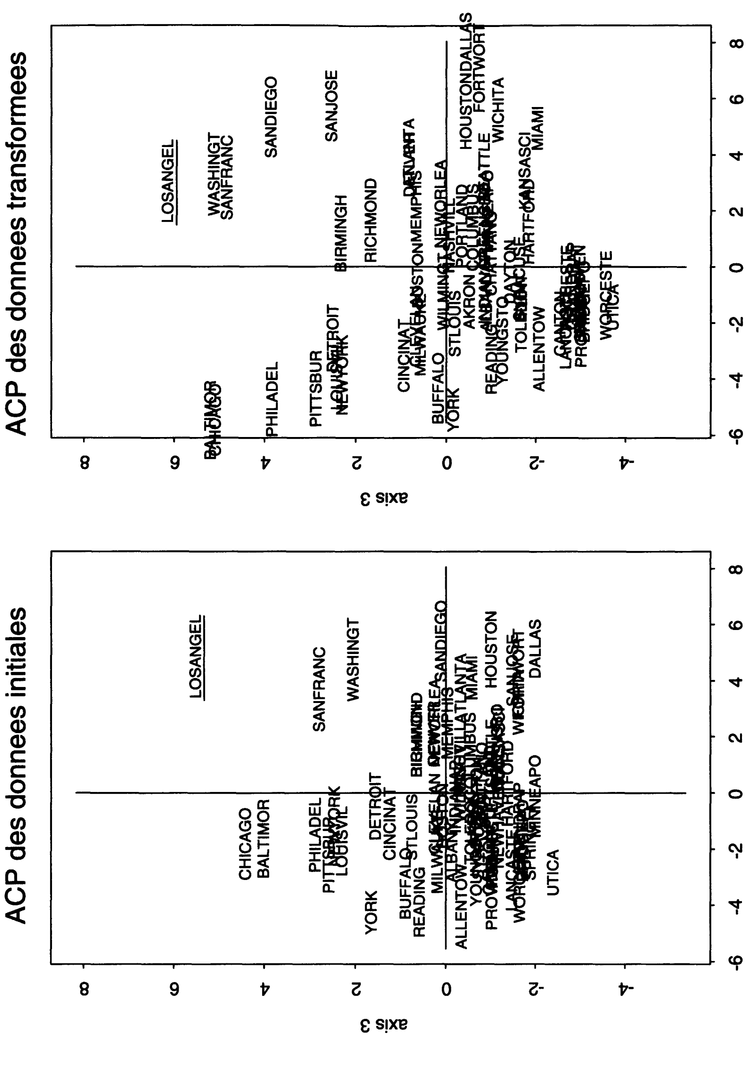


FIGURE 11

*Données de pollution; deuxième plan factoriel des ACP calculées sur les données initiales et les données transformées.*

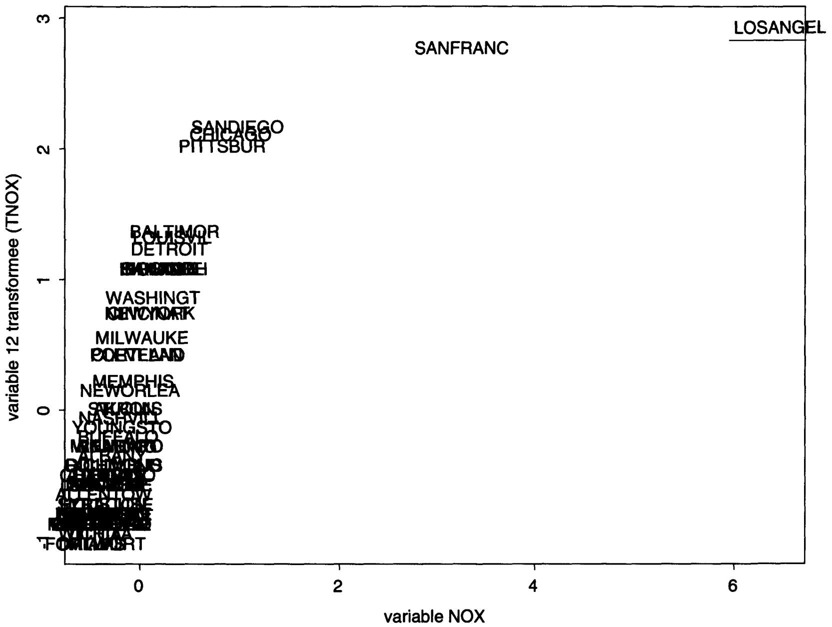


FIGURE 12
*Transformation optimale de NOX estimée par SALSA*$_{(4)}$

## 7. Conclusion

Une fois que tous les phénomènes non-linéaires ont été identifiés sur les graphes des transformations optimales, on s'intéresse dans une deuxième étape à l'analyse linéaire classique des données transformées. Ceci revient à adapter les échelles des appareils de mesure aux possibilités de représentation offertes par l'ACP. On analyse ainsi de façon plus fiable la structure globale des données car les points faibles de l'ACP classique : relations non-linéaires, individus suspects, sont traités à part.

Ce travail montre qu'il est important de considérer ces aspects non-linéaires souvent présents dans des données mais qui sont gommés par une ACP classique ou sur-représentés dans le cas d'individus suspects. Par ailleurs, les outils proposés rendent cette démarche numériquement fiable, facile à utiliser et donc réellement opérationnelle.

## Références bibliographiques

Becker R., Chambers J. et Wilks A. (1988), *The New S Language, a Programming Environment for Data Analysis and Graphics*, Wadsworth and Brooks/Cole.

Besse P. et Ferraty F. (1995), A Fixed Effect Curvilinear Model, *Computational Statistics*, 10, 339-351.

Cazes P. (1990), Codage d'une variable continue en vue d'une analyse des correspondances, *Revue de Statistique Appliquée*, 38(3), 35-51.

Champely S. (1994), *Analyse de données fonctionnelles; approximation par les splines de régression*, PhD Thesis, Université Claude Bernard – Lyon 1.

De Boor C. (1978), *A Practical Guide to Splines*, Springer-Verlag.

De Leeuw J. (1982), Non-linear Principal Component Analysis, In Caussinus, H *et al.*, editors, *Compstat 82, Proceedings in Comp. Stat.*, p. 77-85, Physica-Verlag.

De Leeuw J. et Rijckvorsel J. v. (1988), *Component and Correspondence Analysis*, New York, Wiley.

De Leeuw J., Rijckvorsel J. v. et Wonder H. (1981), Non-linear Principal Components Analysis with b-spline, *Methods of Operations Research*, 33, 379-393.

Delecroix M., Simioni M. et Thomas-Agnan C. (1996), Functional Estimation under Shape Constraint, *To Appear in Journal of Nonparametric Statistics*.

Efron B. et Tibshirani R. (1986), Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy, *Statistical Science*, 1(1), 54-77.

Gifi A. (1990), *Non-linear Multivariate Analysis*, Wiley.

Hastie T.J. et Tibshirani R. (1990), *Generalized Additive Models*, Chapman and Hall.

Kelly C. et Rice J. (1990), Monotone Smoothing with Application to Dose-Response curves and the Assessment of Synergian, *Biometrics*, 18, 75-86.

Kruskal J. et Shepard R. (1974), A Nonmetric Variety of Linear Factor Analysis, *Psychometrika*, 38, 123-157.

Lafaye de Michaux D. (1978), *Approximation d'Analyse Canonique non Linéaire de Variables Aléatoires et Analyses Factorielles Privilégiantes*, PhD thesis, Université de Nice.

McDonald G.C. et Schwing R.C. (1973), Instabilities of Regression Estimates Relating Air Pollution to Mortality, *Technometrics*, 15, 463-481.

Ramsay J. (1988), Monotone Regression Splines in Action, *Statistical Sciences*, 3, 425-461.

Rijckvorsel J.L.A. v. (1982), Canonical Analysis with b-spline, In Caussinus, H. e. a., editor, *Compstat'82*, p. 393-398, Physica-Verlag.

Rijckvorsel J.L.A. v. (1988), Fuzzy Coding and b-spline, In Rijckvorsel, J.L.A. v. and de Leeuw J., editors, *Component and Correspondence Analysis*, p. 33-54, Wiley.

Rijckvorsel J.L.A. v. et Tessitore G. (1993), A Technical Description of the Algorithm for Macca, Technical Report, Dept. of Statistics and Computer Science, TNO Institute of Preventive Health Care, Leiden.

SAS (1989), *SAS.STAT User's Guide*, fourth edition, version 6.

Schumaker L. (1981), *Spline Functions : Basic Theory*, Wiley.

Thurstone L. (1947), *Multiple Factor Analysis*, Univ. Chicago Press.

Wahba G. (1990), *Spline Models for Observational Data*, SIAM.

Winsberg S. et Ramsay J. (1983), Monotone Spline Transformations for Dimension Reduction, *Psychometrika*, 48, 575-599.