A Fixed Effect Curvilinear Model

Philippe C. Besse and Frédéric Ferraty

Laboratoire de Statistique et Probabilités
U.A. CNRS 745, Université Paul Sabatier
31062 Toulouse cedex France
email: besse@cict.fr

Summary

This paper presents a fixed effect curvilinear model, a generalization of the fixed effect model or functional model, which attempts to build smooth optimal transformations of data for the purpose of dimension reduction. More precisely, we look for a set of curves, one per variable, such that they maximize the percentage of variance explained by the first q principal components computed on the transformed data. We show how the fixed effect curvilinear model synthesizes several methods such as non metric principal component analysis by taking appropriate shape constraints for the transformations. Finally, this model allows a non parametric estimation of these non linear functions by using a spline smoother; an optimal smoothing parameter is obtained by building a bootstrap procedure adapted to the model.

Keywords: fixed effect curvilinear model, non metric principal component analysis, spline smoother, bootstrap.

1 Introduction

Different kinds of generalized additive models were studied by Hastie and Tibshirani (1990) in the framework of regression. In the context of principal component analysis (PCA), several generalizations, which introduce non linear transformations for each variable, were developped as "non metric PCA", (De Leeuw and al., 1981, De Leeuw, 1982, Rijckvorsel, 1988, Gifi, 1990) or "dimension reduction" (Winsbeg and Ramsay, 1983, Ramsay, 1988). In these two approaches, the solutions are expressed in terms of splines basis coefficients: B-splines (De Boor, 1978, Schumaker 1981) or I-splines (Winsbeg and Ramsay, 1983).

We introduce an extension of a fixed effect model for PCA (Caussinus, 1986), called a fixed effect curvilinear model, which builds smooth optimal transformations of the data and then derive graphical displays. This model provides a unified presentation to former approaches, which enables to reformulate them in terms of a least squares estimation problem . This leads to a new computational approach replacing the B-splines (or I-splines) coding by a spline smoother (Breiman and Peters, 1992) in which the transformation flexibility is controlled by smoothing parameters instead of knots (number and position). The estimation is performed by a smoothing alternative least squares algorithm (SALSA) easily implemented in the software S+ (Becker et al., 1988).

Regarding the optimality of the smoothing parameter, we consider a risk function (Besse, 1992) based on the expectation of the usual matrix norm and measuring distances between projectors, i.e. between the associated subspaces. Estimates of this risk function are computed via resampling methods such as bootstrap. More precisely, the structure of the fixed effect curvilinear model suggests to bootstrap residuals in the same fashion as Efron and Tibshirani (1986) but in a different context. In our presentation, we will assess the efficiency of our approach on simulated data by comparing the true risk function values and their estimates.

On the other hand, the SAS (1989) procedure namely `prinqual` computes a solution of the least squares estimation problem associated to the context of non metric PCA. In fact, `prinqual` proposes optimal transformations spanned by B-splines or I-splines whereas SALSA uses a spline smoother. Thus, it is interesting to compare `prinqual` and SALSA with the previous risk function once more on simulated data.

2 Smooth transformation

Let Y be a real random variable, taking its values in the range T. n realizations of this variable are collected in the vector $\mathbf{y}$ of $\mathbb{R}^n$. Several approaches, in different frameworks have been proposed to define transformations of a statistical variable. They deal with power functions, polynomial piecewise functions based on B-splines, with monotony constraint or not.... In this paper, we look for functions f belonging to a Sobolev space $W^m(T)$ to define data transformations

$$\tilde{\mathbf{y}} = f(\mathbf{y}) = [f(y_1), \ldots, f(y_n)]'.$$

These functions are thus assumed to have $(m-1)$ absolutely continuous derivatives and a square integrable mth derivative, $(m \geq 1)$. They are built by using a spline smoother which allows to control the smoothness by imposing the following constraint

$$||f||_m^2 = \int_T (f^{(m)})^2(t)dt < c, (c > 0), \tag{1}$$

where $||.||_m$ defines a semi-norm on $W^m(T)$. The most popular case of the cubic splines is obtained with ($m = 2$) and the positive constant c controls the transformation smoothness. Cubic splines are degree 3 piecewise polynomial functions but one could consider any shape of —linear, higher degree polynomial, exponential...— piecewise functions (Wahba, 1990) by considering any other semi-norm of $W^m(T)$. Nevertheless, as in any functional estimation problem, the semi-norm (or the kernel) shape appears to be less important than the smoothing parameter value in the fit quality and thus, illustrative examples will be computed using cubic splines.

With matrix notation, the constraint (1) simply amounts to

$$||f||_m^2 = \tilde{\mathbf{y}}'\mathbf{N}\tilde{\mathbf{y}} = ||\tilde{\mathbf{y}}||_{\mathbf{N}}^2$$

where $\mathbf{N}$ depends on $\mathbf{y}$ and m; it is the symmetric semi-definite and positive matrix which denotes the representation, in the canonical basis, of the induced semi-norm of $||.||_m$ on a n-dimensional subspace (see section 4 for more details).

We denote by $(\mathcal{C})$ the following set of constraints:

$$(\mathcal{C}) \Longleftrightarrow \left\{ \begin{array}{c} f \in W^m(T), \\ ||f||_m^2 \leq c, \\ \mathrm{var}(f(Y)) = 1. \end{array} \right. \tag{2}$$

3 The fixed effect curvilinear model

3.1 Notations

Consider now $\{\mathbf{y}_i;\ i = 1, \ldots, n\}$ n independent random variables in $\mathbb{R}^p$. Their realization is collected in a $(n \times p)$ matrix $\mathbf{Y}$ which is assumed, without loss of generality, to be column centered when each row is weighted by w_i

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

Let p smooth transformation functions f_j operating on the p random variables Y^j taking respectivly their values in a range T_j. The columns $\mathbf{y}^j$ of the data matrix become after transformation

$$\tilde{\mathbf{Y}} = f(\mathbf{Y}) = [f_1(\mathbf{y}^1), \ldots, f_p(\mathbf{y}^p)].$$

Each function f_j much satisfy $\mathcal{C}_j$. It belongs to a subspace of $W^m(T_j)$ whose semi-norm matrix is $\mathbf{N}_j$; the smoothing parameter is c_j; $\mathbf{I}$ denotes the identity matrix in $\mathbb{R}^p$.

3.2 Model

The rows of the data matrix are assumed to be the observations of n random vectors $\{\mathbf{y}_i;\ i = 1, \ldots, n\}$. Moreover, we assume that the transformed variables $\{f(\mathbf{y}_i);\ i = 1, \ldots, n\}$ have the same covariance matrix $\sigma^2\mathbf{I}$ but different expectations $\mathbf{z}_i$ belonging to a common q-dimensional ($q < p$) subspace of $\mathbb{R}^p$.

To summarize, the curvilinear fixed effect model is defined by

$$\begin{array}{l} \{\mathbf{y}_i; i = 1, \ldots, n\}\ n \text{ independent random variables in } \mathbb{R}^p, \\ \exists f; \forall j,\ f_j \text{ obeying } (\mathcal{C}_j),\ \exists E_q,\ dim(E_q) = q \text{ such that:} \\ f(\mathbf{y}_i) = \mathbf{z}_i + \varepsilon_i\ ; i = 1, \ldots, n \text{ and } \forall i, \mathbf{z}_i \in Eq \\ \ \text{with } E(\varepsilon_i) = 0,\ \mathrm{var}(\varepsilon_i) = \sigma^2\mathbf{I}. \end{array} \tag{3}$$

3.3 Estimation

$\mathbf{Z}$ denotes the $(n \times p)$ matrix whose rows are the $\mathbf{z}_i$ and columns are denoted by $\mathbf{z}^j$. A least squares estimation of the transformations together with the fixed effects leads to the solution of the following problem

$$\min_{\mathbf{z}^j, f^j} \left\{ \sum_{j=1}^{p} \left\| f^j(\mathbf{y}^j) - \mathbf{z}^j \right\|_{\mathbf{D}}^2\ ; f^j \text{ satisfying } C_j \text{ and } \mathrm{rank}(\mathbf{Z}) = q \right\}. \tag{4}$$

Problem (4) is equivalent to the following partial Lagrangian problem when denoting the transformed data columns by $\tilde{\mathbf{y}}^j$

$$\min_{\mathbf{z}^j, \tilde{\mathbf{y}}^j} \left\{ \sum_{j=1}^{p} \|\tilde{\mathbf{y}}^j - \mathbf{z}^j\|_{\mathbf{D}}^2 + \rho_j \|\tilde{\mathbf{y}}^j\|_{\mathbf{N_j}}^2\ ; \|\tilde{\mathbf{y}}_j\|_{\mathbf{D}} = 1 \text{ and } \mathrm{rank}(\mathbf{Z}) = q \right\}, \tag{5}$$

where the smoothing parameters ρ_j, depending on c_j, are positive constants. With matrix notations, this problem becomes

$$\min_{\mathbf{Z}, \tilde{\mathbf{Y}}} \left\{ \left\| \tilde{\mathbf{Y}} - \mathbf{Z} \right\|_{\mathbf{I},\mathbf{D}}^2 + \sum_{j=1}^{p} \rho_j \|\tilde{\mathbf{y}}^j\|_{\mathbf{N_j}}^2\ ; \|\tilde{\mathbf{y}}_j\|_{\mathbf{D}} = 1 \text{ and } \mathrm{rank}(\mathbf{Z}) = q \right\}, \tag{6}$$

where $\|\mathbf{Z}\|_{\mathbf{I},\mathbf{D}}^2 = \mathrm{tr}\mathbf{Z}'\mathbf{DZI}$ is the usual matrix norm.

Proposition 1 *Problem (6) is equivalent to the following one*

$$\min_{\mathbf{Q}, \tilde{\mathbf{Y}}} \left\{ \left\| \tilde{\mathbf{Y}} - \mathbf{Q}\tilde{\mathbf{Y}} \right\|_{\mathbf{I},\mathbf{D}}^2 + \sum_{j=1}^{p} \rho_j \|\tilde{\mathbf{y}}^j\|_{\mathbf{N_j}}^2\ ; \|\tilde{\mathbf{y}}_j\|_{\mathbf{D}} = 1 \right\} \tag{7}$$

where $\mathbf{Q}$ *is a rank* q $\mathbf{D}$*-orthogonal projection matrix.*

Proof:

On one hand we consider a solution $(\widetilde{\mathbf{Y}}^*, \mathbf{Z}^*)$ of problem (6). Let $\mathbf{Q}^0$ a rank q $\mathbf{D}$-orthogonal projection matrix such that

$$\left\|\widetilde{\mathbf{Y}}^* - \mathbf{Q}^0\widetilde{\mathbf{Y}}^*\right\|^2_{\mathbf{I},\mathbf{D}} + \sum_{j=1}^{p} \rho_j \left\|\widetilde{\mathbf{y}}^{j*}\right\|^2_{\mathbf{N_j}} \leq \left\|\widetilde{\mathbf{Y}}^* - \mathbf{Z}^*\right\|^2_{\mathbf{I},\mathbf{D}} + \sum_{j=1}^{p} \rho_j \left\|\widetilde{\mathbf{y}}^{j*}\right\|^2_{\mathbf{N_j}}.$$

Since $\mathbf{Z}^*$ is solution of (6), then $\mathbf{Z}^0 = \mathbf{Q}^0\widetilde{\mathbf{Y}}^* = \mathbf{Z}^*$. Thus, $(\widetilde{\mathbf{Y}}^*, \mathbf{Q}^*)$ is the solution of problem (7) where $\mathbf{Q}^*$ is the rank q $\mathbf{D}$-orthogonal projection matrix onto the subspace spanned by the $\mathbf{Z}^*$ columns.

On the other hand, we consider a solution $(\widetilde{\mathbf{Y}}^*, \mathbf{Q}^*)$ of problem (7). Let $\mathbf{Z}^0$ be a $(n \times p)$ matrix of rank q such that

$$\left\|\widetilde{\mathbf{Y}}^* - \mathbf{Z}^0\right\|^2_{\mathbf{I},\mathbf{D}} + \sum_{j=1}^{p} \rho_j \left\|\widetilde{\mathbf{y}}^{j*}\right\|^2_{\mathbf{N_j}} \leq \left\|\widetilde{\mathbf{Y}}^* - \mathbf{Q}^*\widetilde{\mathbf{Y}}^*\right\|^2_{\mathbf{I},\mathbf{D}} + \sum_{j=1}^{p} \rho_j \left\|\widetilde{\mathbf{y}}^{j*}\right\|^2_{\mathbf{N_j}}. \tag{8}$$

Let be $\mathbf{Q}^0$ be the rank q $\mathbf{D}$-orthogonal projection matrix onto the subspace spanned by the $\mathbf{Z}^0$ columns:

$$\left\|\widetilde{\mathbf{Y}}^* - \mathbf{Z}^0\right\|^2_{\mathbf{I},\mathbf{D}} = \left\|\widetilde{\mathbf{Y}}^* - \mathbf{Q}^0\widetilde{\mathbf{Y}}^*\right\|^2_{\mathbf{I},\mathbf{D}} + \left\|\mathbf{Z}^0 - \mathbf{Q}^0\widetilde{\mathbf{Y}}^*\right\|^2_{\mathbf{I},\mathbf{D}},$$

and by (8) we have:

$$\left\|\widetilde{\mathbf{Y}}^* - \mathbf{Q}^0\widetilde{\mathbf{Y}}^*\right\|^2_{\mathbf{I},\mathbf{D}} + \left\|\mathbf{Z}^0 - \mathbf{Q}^0\widetilde{\mathbf{Y}}^*\right\|^2_{\mathbf{I},\mathbf{D}} \leq \left\|\widetilde{\mathbf{Y}}^* - \mathbf{Q}^*\widetilde{\mathbf{Y}}^*\right\|^2_{\mathbf{I},\mathbf{D}}.$$

It follows that $\mathbf{Q}^0 = \mathbf{Q}^*$ and thus $(\widetilde{\mathbf{Y}}^*, \mathbf{Z}^* = \mathbf{Q}^*\widetilde{\mathbf{Y}}^*)$ is solution of problem (6).

∎

Furthermore, it is easy to prove that other sets of constraints than $(\mathcal{C}_j)$ lead to equivalent formulations of some methods in this modeling framework. They amount to write one of the least squares problem (6) or (7). Let $S_j = \{t_k; k = 1, \ldots, r_j\}$ be a sequence of "knots" belonging to T_j.

- If f_j is now constrained to be a step function based on intervals $[t_k, t_{k+1}]$, the transformation amounts to a "crisp coding" of the data which transforms a quantitative variable into a categorical one. The model estimation becomes the non metric PCA as defined by De Leeuw (1982).
- If f_j is assumed to belong to a B-spline subspace of $W^m(T_j)$ (De Boor, 1978) which is also based on a prior sequence S_j of knots, the transformation now amounts to a "fuzzy coding" (Rijckevorsel, 1988) of the data. The least squares estimation of the model amounts to the approach by De Leeuw et al. (1981).

4 Algorithm

In that class of problem, the solution is generally achieved through an alternate least squares algorithm (Gifi, 1990). An implementation of the "classical" non metric PCA can be found in a SAS procedure (`prinqual`) in which the user must set several prior parameter values;

- number and positions of knots which define the B-spline basis for each variable transformation,
- shape and properties of transformations (degree of B or I-splines),
- the dimension q (note that solutions are not nested).

The introduction of a spline smoother rather than B or I-splines leads us to propose a new algorithm in order to estimate the curvilinear fixed effect model. In our case, only the dimension q and the p smoothing parameters must be choosen.

Proposition 2 *Smoothing alternative least Square algorithm (SALSA)*

i. Let $\widetilde{\mathbf{Y}}_{(0)} = \left[\mathbf{y}^1|\cdots|\mathbf{y}^p\right]$ *be the original data set. A matrix* $\mathbf{N}_j$ *is associated to each vector* $\widetilde{\mathbf{y}}^j$.

ii. k^{th} *iteration:*

- *Given* $\widetilde{\mathbf{Y}}_{(k)}$, *let* $\mathbf{Q}$ *be a rank* q $\mathbf{D}$*-orthogonal projection matrix. Then solve*

$$\min_{\mathbf{Q}} \left\{ \left\| \widetilde{\mathbf{Y}}_{(k)} - \mathbf{Q}\widetilde{\mathbf{Y}}_{(k)} \right\|^2_{\mathbf{D},\mathbf{I_p}} \right\} \tag{9}$$

  by the first q *terms of the singular value decomposition of* $\widetilde{\mathbf{Y}}_{(k)}$ *;*

$$\widehat{\mathbf{Z}}_{(k)} = \mathbf{U}_q \mathbf{L}_q^{1/2} \mathbf{V}'_q \text{ and } \mathbf{Q}_{(k)} = \mathbf{U}_q \mathbf{U}'_q \mathbf{D}.$$

- *For this value of* $\mathbf{Q}_{(k)}$, *solve the following* p *optimization problems*

$$\min \left\{ \left\| \mathbf{Q}_{(k)}\widetilde{\mathbf{y}}^j - \widetilde{\mathbf{y}}^j \right\|^2_{\mathbf{D}} + \rho_j \left\| \widetilde{\mathbf{y}}^j \right\|^2_{\mathbf{N_j}} ; \|\widetilde{\mathbf{y}}_j\|_{\mathbf{D}} = 1 \right\}. \tag{10}$$

  The solution $\widetilde{\mathbf{y}}^j_{(k+1)}$ *is the eigenvector of*

$$\mathbf{D}\left(\mathbf{I} - \mathbf{Q}_{(k)}\right) + \rho_j \mathbf{N}_j$$

  corresponding to the smallest eigenvalue.

This algorithm is locally convergent.

Proof. The equivalent expression (6) of problem (7) is the classical approximation of a matrix by another one with the same dimensions but a reduced rank q ($q < p$). Problem (10) amounts to minimize the following quadratic form

$$\left\|\widetilde{\mathbf{y}}^j - \mathbf{Q}\widetilde{\mathbf{y}}^j\right\|_{\mathbf{D}}^2 + \rho_j \left\|\widetilde{\mathbf{y}}^j\right\|_{\mathbf{N_j}}^2 = \widetilde{\mathbf{y}}^{j'} \left\{\mathbf{D}\left(\mathbf{I} - \mathbf{Q}\right) + \rho_j \mathbf{N}_j\right\} \widetilde{\mathbf{y}}^j = \widetilde{\mathbf{y}}^{j'} \mathbf{A}_{\rho_j} \widetilde{\mathbf{y}}^j,$$

where $\mathbf{A}_{\rho_j}$ is a symmetric positive matrix. The minimum is reached when $\widetilde{\mathbf{y}}^j$ is the eigenvector corresponding to the smallest eigenvalue.

Since the least squares criterion is non negative and decreasing, it reaches a minimum but this could be a local one.

∎

When a minimum is reached, the last SVD provides a rank q estimate of the fixed effect $\widehat{\mathbf{Z}}$ and also an estimate of the graphical representation subspace $\widehat{E_q}$. It is spanned by the first q eigenvectors $\mathbf{v}^1, \ldots, \mathbf{v}^q$ of the correlation matrix of the transformed data $\mathbf{R} = \widetilde{\mathbf{Y}}'\mathbf{D}\widetilde{\mathbf{Y}}$.

Practically, there are different ways of building matrices $\mathbf{N}_j$ depending on the basis used to span the n-dimensional spline subspace of $W^m(T_j)$. In the case of a small data set ($n < 50$), it is easy and possible to use a reproducing kernel basis (Wahba, 1990).

To this end, we consider the same notations as in section 2, and we introduce the operator φ which maps any h in $W^m(T)$ into $h^{(m)}$ belonging to $L_2(T)$. Thus, we have

$$W^m(T) = \ker\varphi \oplus \ker\varphi^\perp = H_1 \oplus H_2.$$

Furthermore, let k (resp. k_2) a reproducing kernel of $W^m(T)$ (resp. H_2) and $\mathbf{K}$ (resp. $\mathbf{K}_2$) the following matrix:

$$[\mathbf{K}]_{ij} = k(y_i, y_j) \text{ (resp. } [\mathbf{K}_2]_{ij} = k_2(y_i, y_j)).$$

Then, we have:

$$\mathbf{N} = \mathbf{K}^{-1}\mathbf{K}_2\mathbf{K}^{-1}.$$

For example, we can take $m = 2$, $T = [\alpha, \beta]$ and the following scalar product :

$$\forall (f, g) \in H \times H \ : < f, g >_H = f(\alpha)g(\alpha) + f'(\alpha)g'(\alpha) + \int_\alpha^\beta f''(u)g''(u)du.$$

Then, if k_1 is a reproducing kernel of H_1, we have for $i = 1, \ldots, n$ and $j = 1, \ldots, n$:

$$\begin{aligned} [\mathbf{K}]_{i,j} &= [\mathbf{K}_1]_{i,j} + [\mathbf{K}_2]_{i,j} \\ &= k_1(y_i, y_j) + k_2(y_i, y_j), \end{aligned}$$

with :

$$\begin{array}{rcl} k_1(y_i, y_j) & = & 1 + (y_i - \alpha)(y_j - \alpha), \\ k_2(y_i, y_j) & = & \left\{ \begin{array}{cll} y_i y_j (y_i - \alpha) - (y_i + y_j)\frac{y_i^2 - \alpha^2}{2} + \frac{y_i^3 - \alpha^3}{3} & \text{if} & y_i \leq y_j \\ k_2(y_j, y_i) & \text{if} & y_i > y_j. \end{array} \right. \end{array}$$

In other cases, with a larger data set, it is computationally faster and more accurate to build tridiagonal matrices as described, for example, by Buja et al. (1989). For this, assume that the y_i's are in the increasing order and let $h_i = y_{i+1} - y_i$ for $i = 1, \ldots, n-1$. It is easy to compute the two matrices $\mathbf{\Delta}$ and $\mathbf{\Gamma}$ as follows:

- $[\mathbf{\Delta}]_{i,i} = \frac{1}{h_i}$, $[\mathbf{\Delta}]_{i,i+1} = -(\frac{1}{h_i} + \frac{1}{h_{i+1}})$, $[\mathbf{\Delta}]_{i,i+2} = \frac{1}{h_{i+1}}$,
- $[\mathbf{\Gamma}]_{i,i} = \frac{(h_i + h_{i+1})}{3}$, $[\mathbf{\Gamma}]_{i-1,i} = [\mathbf{\Gamma}]_{i,i-1} = \frac{h_i}{6}$,

where $\mathbf{\Delta}$ (resp. $\mathbf{\Gamma}$) is a tridiagonal $(n-2) \times n$ (resp. $(n-2) \times (n-2)$) matrix. Then:

$$\mathbf{N} = \mathbf{\Delta}'\mathbf{\Gamma}\mathbf{\Delta}.$$

A more convenient algorithm, dealing with B-spline approximation of spline smoothing, is beeing studied to deal with very large data sets.

5 Quality criterion and optimality

The major interest of the adopted modeling approach lies in the possibility to define quality criteria as the measure of goodness of fit. This gives us a rationale to be able to discuss the optimality of the parameter values concerning dimensionality or smoothing. Different kinds of criteria could be defined but we only consider quadratic risks, i.e. the expectation of a quadratic loss function. We consider the following loss functions which measure or the fixed effect estimation quality

$$\mathcal{L}_1(q, \rho) = \left\| \mathbf{Z} - \widehat{\mathbf{Z}}_q \right\|^2_{\mathbf{I},\mathbf{D}}, \tag{11}$$

either the subspace estimation quality

$$\mathcal{L}_2(q, \rho) = \frac{1}{2} \left\| \mathbf{P}_q - \widehat{\mathbf{P}_q} \right\|^2_{\mathbf{I},\mathbf{I}} = q - \mathrm{tr}\mathbf{P}_q \widehat{\mathbf{P}_q} \tag{12}$$

where $\widehat{\mathbf{P}_q}$ (resp. $\mathbf{P}_q$) is the orthogonal projection onto $\widehat{E_q}$ (resp. E_q). This last function is based on the usual norm of matrices when it is applied to measure distances between projectors, and thus to measure distances between the associated subspaces.

Risk functions are then defined by taking the expectation

$$R_2 = E(\mathcal{L}_2). \tag{13}$$

It seems interesting to use $\mathcal{L}_1$ in cases of simulation studies when the "true" fixed effect is known. The second risk $\mathcal{L}_2$ had been successfully used in the case of linear PCA (Besse, 1992). Its estimation, in the case of real life data, seems more convenient and since the number of parameters to estimate does not depend on the sample size, we may hope better asymptotic behaviors. They both can be easily computed in cases of simulated data when the true fixed effect and the space E_q are known. In practical cases, even if they are computationally expensive, resampling methods such as the bootstrap are natural candidates to compute estimates for the risk function R_2 since it is symmetrically defined; its value is invariant under any permutation of the observations y_i.

Model (3) suggests to boostrap residuals, which are assumed to be i.i.d., as was proposed by Efron and Tibshirani (1986). This raises a problem since our resampling procedure requires the inversion of fonctions (f_j) which are not necessarily constrained to monotonicity[1]. Nevertheless, this problem is overridden in the following way: let $\widehat{\widehat{\mathbf{Y}}}$ (resp. $\widehat{\mathbf{Z}}$) the transformed data (resp. fixed effect) estimation matrix obtained by SALSA. The bootstrap estimate of the risk R_2 is then otained with the following algorithm:

i. computation of residuals:
$\widehat{\mathbf{E}} = \widehat{\widehat{\mathbf{Y}}} - \widehat{\mathbf{Z}}$; as the rank of the matrix $\widehat{\mathbf{E}}$ is equal to $p-q$, residuals are linearly dependent, we replace its null singular values by the smallest and obtain $\widetilde{\mathbf{E}}$, a full rank matrix.

ii. bootstrap estimate:
pour $b = 1, \ldots, B$:

- resampling of the residuals matrix $\widetilde{\mathbf{E}}$; we obtain $\widetilde{\mathbf{E}}^{(b)}$,
- $\widehat{\widehat{\mathbf{Y}}}^{(b)} = \widehat{\mathbf{Z}} + \widetilde{\mathbf{E}}^{(b)} \Leftrightarrow \{\widehat{\widehat{y}}_{ij}^{(b)} = \widehat{f}_j(y_{ij}^{(b)}) = \widehat{z}_{ij} + \widetilde{\epsilon}_{ij}^{(b)} \; ; \; i = 1, \ldots, n \text{ et } j = 1, \ldots, p\}$,
- if $\widehat{\widehat{y}}_{ij}^{(b)}$ has more than one preimage by $\widehat{f}_j$, we take the closest to y_{ij} denoted by $y_{ij}^{(b)}$, thus constructing the matrix $\mathbf{Y}^{(b)}$,
- SALSA is applied to get $\widehat{\mathbf{Z}}^{(b)}$ and $\widehat{\mathbf{P}}_q^{(b)}$; the estimate of $\mathcal{R}$ computed with B samples is:

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

In the application, we take $B = 30$.

[1]This possibility will be proposed in a forthcomming paper.

6 Example

We first discuss a method of choice for the smoothing parameter value and show its optimality on simulated data. Then, we compare results obtained by `prinqual` (SAS) and SALSA.

The simulated data set consists of a $(50, 6)$ matrix; $p = 6$ variables are observed on $n = 50$ units. They are obtained by transforming each column of the sum of a matrix of fixed effects with known rank ($q = 3$) and a simulated white noise of controlled standard deviation:

i. The fixed effects are constructed as follows:

   - we first generate a $(50, 6)$ matrix of pseudo-random numbers uniformly and independently distributed in the range $[0, 20]$,
   - then the $(50, 6)$ matrix of non centered fixed effects is derived from its SVD by constructing its best least squares approximation of rank $q = 3$.

ii. The noise is simulated as follows:

   - we generate 300 observations of independently and identically normally distributed pseudo-random variables with mean 0 and variance σ^2,

iii. The following transformations $f(z + \varepsilon)$ of each column give the matrix $\mathbf{Y}$:

$$
\begin{aligned}
f_1(t) &= \exp(t) \\
f_2(t) &= \log((t + c_1)^4 + 1) \\
f_3(t) &= 4\exp(t) + (t + c_2)^2 \\
f_4(t) &= 2\log(t + c_3) + t^3 \\
f_5(t) &= 3\log(t + c_4) \\
f_6(t) &= (t + c_5)\exp(t + c_6)
\end{aligned}
$$

   where $c_1, \ldots, c_6$ are real constants, we check that rank($\mathbf{Y}$)> 3.

Since the data are simulated, the true quadratic risk can be computed and reported. If we look at the figure 1, it seems that the true risk function R_2 and its bootstrap estimate $\widehat{R}_2$ reach their minimum for the same value. So, we deduce a value for the optimal smoothing parameter ρ^* ; in this example, ρ^* is close to 0.001.

Now let us compare different kinds of estimations proposed in the context of this model. In particular, `prinqual`, a procedure implemented in SAS, offers transformations spanned by B-splines or I-splines basis, the last one allowing monotonicity constraint. We choose the optimal smoothing parameter previously estimated for SALSA. As regards `prinqual`, different options are offered: degree of transformations, number and location of knots. The degree is choosen to be the same as the one used to build the spline smoother, namely 3. Moreover, it is usual to place the knots at the quantiles of each variable. Finally, we choose the number of knots such that the loss function R_2, computed on the transformed data, reaches its minimum. In this way, we may now compare `prinqual` and SALSA. Figure 2 displays the behavior of the true risk function R_2 computed on results obtained by `prinqual` and SALSA. It is comforting to observe that the quality criterion recognizes the true dimension (q=3). Moreover, the SALSA estimate of the eigensubspace appears to be more stable than the `prinqual` one.

In fact, as regards the transformed data obtained by SALSA, the first 3 principal axes provide about 80 percent of the variance, which corresponds with the part of variance explained by the first 4 principal components deduced of the classical PCA. It means that, in addition nonlinear relationship between variables, some singular behavior, invisible in a 3-dimensional eigensubspace computed by a classical PCA, can appear in the SALSA estimation.

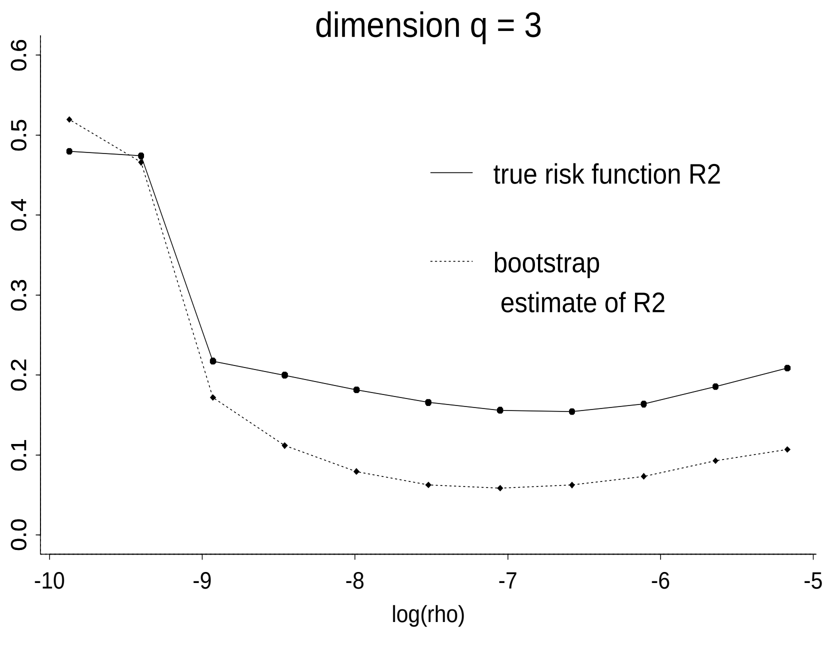

Figure 1: *comparison between the true risk function R_2 and its bootstrap estimate*

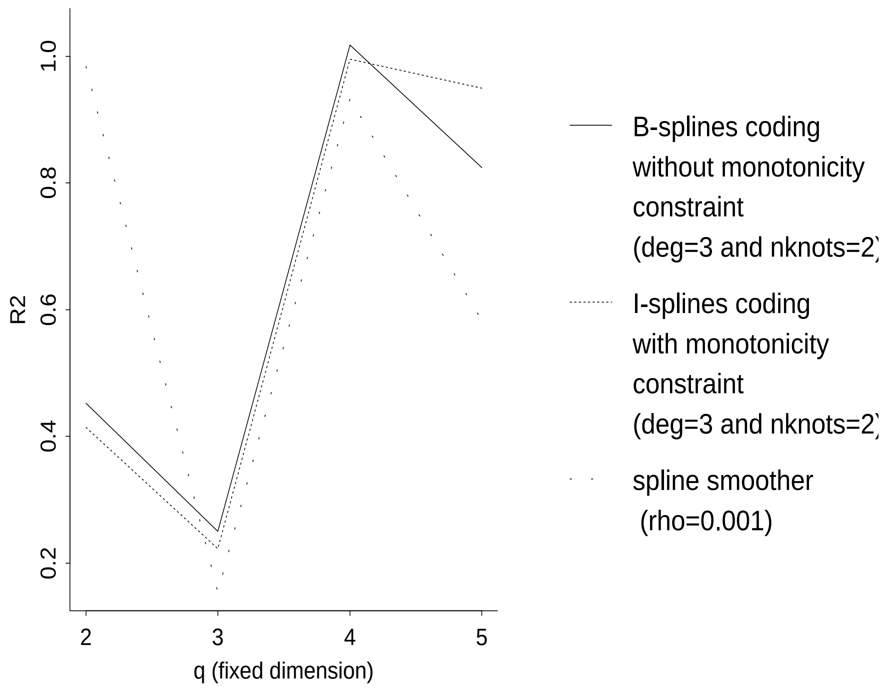

Figure 2: *SAS (*`prinqual`*) versus SALSA*