Functional linear regression with derivatives

André Mas, Besnik Pumo

Functional linear regression with derivatives

André MAS*, Besnik PUMO†

(*) Institut de Modélisation Mathématique de Montpellier,
cc 051, Place Eugène Bataillon,
34095, Montpellier Cedex 5, France,
mas@math.univ-montp2.fr

(†) Unité de Statistiques,
UMR A 462 SAGAH,
INH, Angers, France

Abstract

We introduce a new model of linear regression for random functional inputs taking into account the first order derivative of the data. We propose an estimation method which comes down to solving a special linear inverse problem. Our procedure tackles the problem through a double and synchronized penalization. An asymptotic expansion of the mean square prevision error is given. The model and the method are applied to a benchmark dataset of spectrometric curves and compared with other functional models.

Keywords : Functional data, Linear regression model, Differential operator, Penalization, Spectrometric curves.

1 Introduction

Functional Data Analysis is a well-known area of modern statistics. Advances in computer sciences make it now possible to collect data from an underlying continuous-time processe, say $(\xi_t)_{t\geq 0}$, at high frequencies. The traditional point of view consisting in discretizing (ξ_t) at $t_1, ..., t_p$ and studying it by classical multidimensional tools is outperformed by interpolation methods (such as splines or wavelets). These techniques provide the statistician with

a reconstructed curve on which inference may be carried out through what we may call "functional models" i.e. versions of the classical multidimensional models designed and suited for data that are curves. Thus, functional PCA, ANOVA or Canonical Analysis -even density estimation for curves or processes have been investigated. We refer to Ramsay, Silverman (1997, 2002), Bosq (2000), Ferraty Vieu (2006) for monographs on functional data analysis. Recently many authors focused on various versions of the regression model introduced by Ramsay and Dalzell (1991) :

$$y_i = \int_0^T X_i(t)\,\rho(t)\,dt + \varepsilon_i \tag{1}$$

where we assume that the sample $((y_1,X_1),...,(y_n,X_n))$ is made of independent copies from (y,X). Each $X_i = (X_i(t))_{t\in[0,T]}$ is a curve defined on the set $[0,T]$, $T>0$, y_i is a real number, ε_i is a white noise and ρ is an unknown function to be estimated. In other words the X_i's are random elements defined on an abstract probability space and taking values in a function space, say $\mathcal{F}$. The vector space $\mathcal{F}$ endowed with norm $\|\cdot\|_{\mathcal{F}}$ will be described soon.We refer for instance to Cardot, Mas, Sarda (2006) or Cai, Hall (2006) for recent results.

In this article we study a new (linear) regression model defined below derived from (1) and echoing the recent paper of Mas and Pumo (2006). The key idea relies on the fact that most statisticians dealing with functional data do not fully enjoy their functional properties. For instance in several models integrals such as

$$\int X_i(s)\,X_j(s)\,ds$$

are computed. The integral above is nothing but a scalar product. Nevertheless derivatives were not given the same interest. Explicit calculations of derivatives sometimes appear indirectly in kernel methods (when estimating the derivatives of the density or the regression function) or through seminorms or norms on $\mathcal{F}$. But surprisingly X_i' (or $X_i^{(m)}$) never appear in the models themselves whereas people dealing with functional data often say that "derivatives contain much information, sometimes more than the initial curves themselves". Our starting idea is the following. Since in a functional data framework, the curve-data are explicitely known and not just discretized, their derivatives may also be explicitely computed. As a consequence these derivatives may be "injected" in the model, which may enhance its prediction power. The reader is referred to the forthcoming display (5) for an immediate illustration and to Mas, Pumo (2006) for a first article dealing with a functional autoregressive model including derivatives.

The paper is rather theoretic even if it is illustrated by a real case study. It is organized as follows. The next section provides the mathematical material, dealing with Hilbert spaces and linear operators, then the model is introduced. The next section is devoted to presenting the estimation method and its stumbling stones. The main results are given before we focus on a real case application to food industry. The last section contains the derivation of the theorems.

2 About Hilbert spaces and linear operators

Silverman (1996) provided a theoretical framework for a smoothed PCA. Jim Ramsay (2000) enlightened the very wide scope of differential equations in statistical modelling. Our work is in a way based on this mathematically involved article. We are aiming at proving that derivatives may be handled in statistical models quite easily when the space $\mathcal{F}$ is well-chosen.

The choice of the space $\mathcal{F}$ is crucial. We have to think that if $X \in \mathcal{F}$, X' does not necessarily belong to $\mathcal{F}$ but to another space $\mathcal{F}'$ that may be tremendously different (larger) than $\mathcal{F}$. We decide to take $\mathcal{F} = W^{2,1}$, the Sobolev space of order $(2,1)$ defined by

$$W^{2,1} = \left\{ u \in L^2[0,1], u' \in L^2[0,1] \right\}$$

for at least three reasons :

- If $X \in \mathcal{F}$, $X' \in L^2[0,1]$ which is a well known space.
- Both spaces are Hilbert spaces as well as

  $$W^{2,p} = \left\{ u \in L^2[0,1], u^{(p)} \in L^2[0,1] \right\}.$$

  This is of great interest for mathematical reasons : bases are denumerable, projections operators are easy to handle, covariance operators admit spectral representations, etc.
- The classical interpolation methods mentioned above (splines and wavelets) provide estimates belonging to Sobolev spaces. So from a practical point of view $W^{2,1}$ -and in general $W^{m,p}$, $(m,p) \in \mathbb{N}^2$, (see Adams and Fournier (2003) for definitions)- is a natural space in which our curves should be imbedded.

In the sequel $W^{2,1}$ will be denoted W and $W^{2,0} = L^2$ will be denoted L for the sake of simplicity. We keep in mind that W (resp. L) could be

replaced by a space of higher smoothness index : $W^{2,p}$ where $p > 1$ (resp. $W^{2,p-1}$). The spaces W and L are separable Hilbert spaces endowed with scalar product :

$$\langle u, v \rangle_W = \int_0^1 u(t)\, v(t)\, dt + \int_0^1 u'(t)\, v'(t)\, dt.$$
$$\langle u, v \rangle_L = \int_0^1 u(t)\, v(t)\, dt$$

and with associated norms $\|\cdot\|_W$ and $\|\cdot\|_L$. We refer to Ziemer (1989) or to Adams and Fournier (2003) for monographs dedicated to Sobolev spaces. Obviously if we set $Du = u'$ then D maps W onto L (D is the ordinary differential operator). Furthermore Sobolev's imbedding theorem ensures that (see Adams and Fournier (2003) Theorem 4.12 p.85) that

$$\|Du\|_L \leq C \|u\|_W$$

(where C is some constant which does not depend on u) i.e. D is a bounded operator from W to L. This is a crucial point to keep in mind and the fourth reason why the functional space was chosen to be $W^{2,1}$: the differential operator D may be viewed as a continuous linear mapping from W to L.

Within all the paper and especially all along the proofs we will need basic notions about operator theory. We recall a few important facts. A linear mapping T from a Hilbert space H to another Hilbert space H' is continuous whenever

$$\|T\|_\infty = \sup_{x \in H} \frac{\|Tx\|_{H'}}{\|x\|_H} < +\infty. \tag{2}$$

The adjoint of operator T will be classically denoted T^*. Some finite rank operators are defined by means of the tensor product : if u and v belong to H and H' respectively $u \otimes_H v$ is the operator defined on H by, for all $h \in H$:

$$(u \otimes_H v)(h) = \langle u, h \rangle_H v.$$

Compact operators : Amongst linear operators the class of compact operators is one of the best known. Compact operators generalize matrix to the infinite-dimensional setting and feature nice properties. The general definition of compact operators may be found in Dunford Schwartz (1988) or Gohberg, Goldberg and Kaashoek (1991) for instance. By $\mathcal{C}_H$ (resp. $\mathcal{C}_{HH'}$) we denote the space of compact operators on the Hilbert space H (resp. mapping the Hilbert space H onto H'). If T is a compact operator from a Hilbert space H_1 to another Hilbert space H_2, T admits the Schmidt decomposition :

$$T = \sum_{k \in \mathbb{N}} \mu_k (u_k \otimes v_k) \tag{3}$$

where u_k (resp. v_k) is a complete orthonormal system in H_1 (resp. in H_2) and μ_k are the characteristic numbers of T (i.e. the square root of the eigenvalues of T^*T) and

$$\lim_{k\to+\infty} \mu_k = 0.$$

From (2) we obtain

$$\|T\|_\infty = \sup_k \{\mu_k\}.$$

When T is symmetric μ_k is the k^{th} eigenvalue of T (then $u_k = v_k$). In this situation and from (3) one may define the square root of T whenever T maps H ont H and is positive : $T^{1/2}$ is still a linear operator defined by :

$$T^{1/2} = \sum_{k\in\mathbb{N}} \sqrt{\mu_k}\,(u_k \otimes u_k). \tag{4}$$

Note that finite rank operators are always compact.

Hilbert-Schmidt operators : We also mention the celebrated space of Hilbert-Schmidt operators $\mathcal{HS}(H_1, H_2)$ - a subspace of $\mathcal{C}_{H_1 H_2}$. Let $(u_i)_{i\geq 0}$ be a basis of H_1 then $T \in \mathcal{HS}(H_1, H_2)$ whenever

$$\sum_{i=1}^{+\infty} \|T(u_i)\|_{H_2}^2 < +\infty.$$

The space $\mathcal{HS}$ is itself a separable Hilbert space endowed with scalar product

$$\langle T, S\rangle_{\mathcal{HS}} = \sum_{i=1}^{+\infty} \langle T(u_i), S(u_i)\rangle_{H_2}$$

and $\langle T, S\rangle_{\mathcal{HS}}$ does not depend on the choice of the basis $(u_i)_{i\geq 0}$. Finally the following bound is valid for all $T \in \mathcal{HS}$:

$$\|T\|_\infty \leq \|T\|_{\mathcal{HS}}.$$

Unbounded operators : If T is a one to one (injective) selfadjoint compact operator mapping a Hilbert space H onto H, T admits an inverse T^{-1}. The operator T^{-1} is defined on a dense (and distinct) subspace of H :

$$\mathcal{D}\left(T^{-1}\right) = \left\{ x = \sum_{p\in\mathbb{N}} x_p u_p : \sum_{p\in\mathbb{N}} \frac{x_p^2}{\mu_p^2} < +\infty \right\}.$$

It is unbounded which also means that T^{-1} is continuous at no point for which it is defined and $\|T^{-1}\|_\infty = +\infty$.

3 The model

We are now in position to introduce this (random input - linear) regression model :

$$y_i = \langle \phi, X_i \rangle_W + \langle \psi, X_i' \rangle_L + \varepsilon_i \tag{5}$$

where all random variables are assumed to be centered. The main result of the paper (see next section) gives an asymptotic expansion for the mean square prediction error in (5).

The unknown functions ϕ and ψ belong to W and L respectively.

Obviously we are going to face two issues :

- Studying the identifiability of ϕ and ψ in the model above.
- Providing a consistent estimation procedure for ϕ and ψ.

From now on we suppose that :

$$\mathbf{A1} : \|X\|_W < M \quad a.s.$$

This assumption could be relaxed for milder moment assumptions. We claim that our main result holds whenever

$$\mathbf{A'1} : \mathbb{E}\,\|X\|_W^8 < M.$$

is true. But considering $\mathbf{A'1}$ would lead us to longer and more intricate methods of proof.

4 Estimation procedure

4.1 The moment method

Inference is based on moment formulas. From (5) we derive the two following normal equation -multiply with $\langle X_i, \cdot \rangle$ and $\langle X_i', \cdot \rangle$ successively then take expectation :

$$\begin{cases} \delta = \Gamma\phi + \Gamma'\psi, \\ \delta' = \Gamma'^*\phi + \Gamma''\psi. \end{cases} \tag{6}$$

where Γ, Γ', Γ'^*, Γ'' are the covariance and cross-covariance of the couple $(X_i, X_i')_{1\le i\le n}$ defined by :

$$\Gamma = \mathbb{E}\,(X \otimes_W X)\,,\ \Gamma'^* = \mathbb{E}\,(X \otimes_W X')\,,$$
$$\Gamma' = \mathbb{E}\,(X' \otimes_L X)\,,\ \Gamma'' = \mathbb{E}\,(X' \otimes_L X')\,,$$

and

$$\delta = \mathbb{E}(yX) \in W, \quad \delta' = \mathbb{E}(yX') \in L.$$

Under assumption **A1** or **A'1** the covariance operators belong to $\mathcal{HS}(W)$, $\mathcal{HS}(W, L)$, $\mathcal{HS}(L, W)$ or to $\mathcal{HS}(L)$. Besides the covariance and cross-covariance mentioned above are linked through the relation

$$\Gamma'^* = D\Gamma, \quad \Gamma'' = D\Gamma'.$$

Resolving the system (6) is apparently easy but we should be aware of two facts :

- Operators (here, $\Gamma, \Gamma'...$) do not commute !
- The inverse operators of Γ, and Γ'' do not necessarily exist and when they do, they are unbounded, i.e. not continuous (recall that Γ, and Γ'' are compact operators and that compact operators have no bounded inverses).

Before trying to solve (6) we will first study identifiability of the unknown infinite dimensional parameter $(\phi, \psi) \in W \times L$ in the next subsection. We complete our definitions and notations first.

We start from a sample $(X_i, X_i')_{1 \leq i \leq n}$. By $\Gamma_n, \Gamma'_n, \Gamma'^*_n, \Gamma''_n, \delta_n$ and δ'_n we denote the empirical counterparts of the operators and vectors introduced above and based on the sample $(y_i, X_i, X_i')_{1 \leq i \leq n}$. For example :

$$\begin{aligned} \Gamma_n &= \frac{1}{n} \sum_{k=1}^{n} X_k \otimes_W X_k, \\ \Gamma'_n &= \frac{1}{n} \sum_{k=1}^{n} X'_k \otimes_L X_k, \\ \delta_n &= \frac{1}{n-1} \sum_{k=1}^{n-1} y_k X_k. \end{aligned} \tag{7}$$

4.2 Identifiability

Both equations in (6) are the starting point of the estimation procedure. We should make sure that solutions to these equations are well and uniquely defined. Suppose for instance that $\mathrm{Ker}\Gamma \neq \{0\}$ and take h in it. Now set $\widetilde{\phi} = \phi + h$. Then

$$\Gamma\widetilde{\phi} = \Gamma\phi + \Gamma h = \Gamma\phi.$$

So $\Gamma\widetilde{\phi} = \Gamma\phi$ and since $\Gamma'^* = D\Gamma$ it is plain that $\Gamma'^*\widetilde{\phi} = \Gamma'^*\phi$. Consequently $\widetilde{\phi}$ is another solution to (6). There are indeed even infinitely many solutions in the space ϕ+KerΓ. For similar reasons about ψ we should impose Ker$T = \{0\}$ for $T = \{\Gamma, \Gamma', \Gamma'^*, \Gamma''\}$. It turns out that the only necessary assumption is

$$\mathbf{A2} : \mathrm{Ker}\Gamma = \mathrm{Ker}\Gamma'' = \{0\}.$$

It is easily seen that **A2** implies Ker$\Gamma' =$ Ker$\Gamma'^* = \{0\}$. With other words we suppose that both operators Γ and Γ'' above are one to one.

We are now ready to solve the identification problem.

Proposition 1 *The couple $(\phi, \psi) \in W \times L$ is identifiable for the moment method proposed in (6) if and only if* **A2** *holds and $(\phi, \psi) \notin \mathcal{N}$ where $\mathcal{N}$ is the vector subspace of $W \times L$ defined by :*

$$\mathcal{N} = \{(\phi, \psi) : \phi + D^*\psi = 0\}. \tag{8}$$

The above Proposition is slightly abstract but (8) may be simply rewritten: $(\phi, \psi) \in \mathcal{N}$ whenever for all function f in W,

$$\int (f\phi + f'\phi' + f'\phi) = 0$$

Note that $\mathcal{N}$ is a closed set in $W \times L$. From now on we will assume that :

$$\mathbf{A3} : (\phi, \psi) \notin \mathcal{N}.$$

5 Definition of the estimates

The estimates stem from (6) which is a non invertible system. Under assumption **A2** the solution exists and is unique :

$$\begin{cases} \phi = (\Gamma - \Gamma'\Gamma''^{-1}\Gamma'^*)^{-1}\left[\delta - \Gamma'\Gamma''^{-1}\delta'\right], \\ \psi = (\Gamma'' - \Gamma'^*\Gamma^{-1}\Gamma')^{-1}\left[\delta' - \Gamma'^*\Gamma^{-1}\delta\right]. \end{cases} \tag{9}$$

Let us denote

$$S_\phi = \Gamma - \Gamma'\Gamma''^{-1}\Gamma'^*,$$
$$S_\psi = \Gamma'' - \Gamma'^*\Gamma^{-1}\Gamma'.$$

The reader should note two crucial facts. On the one hand Γ^{-1} and Γ''^{-1} are unbouded operators but closed graphs argument ensure that $\Gamma'\Gamma''^{-1}\delta'$ and $\Gamma'^*\Gamma^{-1}\delta$ exist in W and L respectively. On the other hand $\delta - \Gamma'\Gamma''^{-1}\delta'$ (resp.

$\delta' - \Gamma'^{*}\Gamma^{-1}\delta$) belong to the domain of the unbounded operator S_{ϕ}^{-1} (resp. S_{ψ}^{-1}) which also ensures the finiteness of both solutions given in the display above.

Finding approximations to the solutions of (9) is known in the mathematical literature as "solving a linear inverse problem". The book by Tikhonov and Arsenin (1977) -as many other references therein- is devoted to this theory well-known in image reconstruction. The unboundedness of S_{ϕ}^{-1} may cause large variation of $S_{\phi}^{-1}x$ even for small variations of x. This lack of stability turns out to damage, as well as the traditional "curse of dimensionality", the rates of convergence of our estimates.

Unfortunately we cannot simply replace "theoretical" operators and vectors by their empirical estimates because Γ_n and Γ''_n are not invertible. Indeed they are finite-rank operators (for example the image of Γ is $\text{span}(X_1, ..., X_n)$) hence not even injective. We are classically going to add a small perturbation to regularize Γ_n and Γ''_n (see Tikhonov and Arsenin (1977)) and another one for S_{ϕ}^{-1} and make them invertible. At last Γ^{-1} is approximated by $\Gamma_n^{\dagger} = (\Gamma_n + \alpha_n I)^{-1}$, Γ''^{-1} by $\Gamma_n''^{\dagger} = (\Gamma''_n + \alpha_n I)^{-1}$ and S_{ϕ}^{-1} by $(S_{n,\phi} + \beta_n I)^{-1}$ where

$$S_{n,\phi} = \Gamma_n - \Gamma'_n \left(\Gamma_n''^{\dagger}\right) \Gamma'^{*}_n.$$

and $\alpha_n > 0$, $\beta_n > 0$. We also set :

$$S_{n,\psi} = \Gamma''_n - \Gamma'^{*}_n \left(\Gamma_n^{\dagger}\right) \Gamma'_n, \tag{10}$$

$$u_{n,\phi} = \delta_n - \Gamma'_n \left(\Gamma_n''^{\dagger}\right) \delta'_n, \tag{11}$$

$$u_{n,\psi} = \delta'_n - \Gamma'^{*}_n \left(\Gamma_n^{\dagger}\right) \delta_n. \tag{12}$$

In the sequel we will assume that both strictly positive sequences α_n and β_n decay to zero in order to get the asymptotic convergence of the estimates.

Definition 2 *The estimate of the couple* (ϕ, ψ) *is* $\left(\widehat{\phi}_n, \widehat{\psi}_n\right)$ *based on (9) and defined by :*

$$\begin{cases} \widehat{\phi}_n = (S_{n,\phi} + \beta_n I)^{-1} u_{n,\phi}, \\ \widehat{\psi}_n = (S_{n,\psi} + \beta_n I)^{-1} u_{n,\psi}. \end{cases} \tag{13}$$

The predictor is defined as

$$\widehat{y}_{n+1} = \left\langle \widehat{\phi}_n, X_{n+1} \right\rangle_W + \left\langle \widehat{\psi}_n, X'_{n+1} \right\rangle_L.$$

6 Main results and comments

In Mas, Pumo (2006) the authors obtained convergence in probability for their estimates in a quite different model. We are now in position to assess deeper results. Mean square prediction error is indeed given an asymptotic development depending on both smoothing sequences α_n and β_n.

Before stating the main result of this article, we give and comment the next and last assumption :

$$\mathbf{A4} : \begin{cases} \left\|\Gamma^{-1/2}\phi\right\|_W < +\infty \\ \left\|(\Gamma'')^{-1/2}\psi\right\|_L < +\infty \end{cases} \tag{14}$$

For the definition of $\Gamma^{-1/2}$ and $\Gamma''^{-1/2}$ we refer to (4). Let us explain briefly what both conditions in (14) mean. To that aim we rewrite the first by developing $\Gamma^{-1/2}\phi$ in a basis of eigenvectors of Γ, say u_p

$$\Gamma^{-1/2}\phi = \sum_{p=1}^{+\infty} \frac{\langle \phi, u_p \rangle}{\sqrt{\lambda_p}} u_p$$

hence

$$\left\|\Gamma^{-1/2}\phi\right\|_W^2 = \sum_{p=1}^{+\infty} \frac{\langle \phi, u_p \rangle^2}{\lambda_p}$$

The first part of assumption **A4** tells us that "$\langle \phi, u_p \rangle$ should tend to zero quickly enough with respect to λ_p". In other words ϕ should belong to an ellipsoïd of W which may be more or less "flat" depending on the rate of decay of the λ_p's to zero. Assumption **A4** is in fact **a regularity condition** on functions ϕ and ψ : function ϕ (resp. ψ) should be smoother than X (resp. X').

We could try and state convergence results for $\widehat{\phi}_n$ and $\widehat{\psi}_n$ separatedly but it turns out that :

- The real statistical interest of the model relies on its predictive power. The statistician is mainly interested in $\widehat{y}_{n+1}$, not in $\widehat{\phi}_n$ and $\widehat{\psi}_n$ in a first attempt. The issue of goodness of fit tests (involving ϕ and ψ alone) is beyond the scope of this article.

- Considering the mean square norm of $\left\langle \widehat{\phi}_n, X_{n+1} \right\rangle_W$ (instead of $\widehat{\phi}_n$ or even of $\left\langle \widehat{\phi}_n, x \right\rangle_W$ for a nonrandom x) has a smoothing effect on our

estimates and partially counterbalance the side effects of the underlying inverse problem as will be seen within the proofs (especially along Lemma 14).

Turning to $\widehat{y}_{n+1}$, the next question is : what should we compare $\widehat{y}_{n+1}$ with ? The right answer is not y_{n+1}. Obviously we could, but it is also plain that, due to the random ε_{n+1} the best possible prediction for y_{n+1} knowing X_{n+1} (or even the "past" i.e. $X_1, ..., X_n$) is the conditional expectation :

$$y^*_{n+1} = \mathbb{E}\left(y_{n+1}|X_1, ..., X_{n+1}\right) = \langle \phi, X_{n+1} \rangle_W + \left\langle \psi, X'_{n+1} \right\rangle .$$

We are now ready to state the main theoretical result of this article.

Theorem 3 *When assumptions* $\mathbf{A1} - \mathbf{A4}$ *hold the following expansion is valid for the prediction mean square error :*

$$\mathbb{E}\left(\widehat{y}_{n+1} - y^*_{n+1}\right)^2 = O\left(\frac{\beta^2}{\alpha^2}\right) + O\left(\frac{1}{\alpha^2\beta^2 n}\right)$$

Remark 4 *Replacing* y^*_{n+1} *with* y_{n+1} *is still possible. We may easily prove that :*

$$\mathbb{E}\left(\widehat{y}_{n+1} - y_{n+1}\right)^2 = \mathbb{E}\left(\widehat{y}_{n+1} - y^*_{n+1}\right)^2 + \sigma^2_\varepsilon.$$

Corollary 5 *From Theorem 3 above an optimal choice for* β *is* $\beta^* \asymp n^{-1/4}$, *then the convergence rate is :*

$$\mathbb{E}\left(\widehat{y}_i - y^*_i\right)^2 = O\left(\frac{1}{\alpha^2 n^{1/2}}\right)$$

and may be quite close from $1/n^{1/2}$.

The proof of the Corollary will be omitted. Studying the optimality of this rate of convergence over the classes of functions defined by **A4** is beyond the scope of this article but could deserve more attention.

Remark 6 *Originally the linear model (5) is subject to serious multicolinearity troubles since* $X'_n = DX_n$. *Even if the curve* X'_n *usually looks quite different from* X_n, *there is a total stochastic dependence between them. The method used in this article to tackle this problem (as well as the intrinsic "inverse problem" aspects related to the inversion of the covariance operators* Γ *and* Γ''*) is new up to the authors' knowledge. As it can be seen through above at display (13) or in the proofs below, it relies on a double penalization technique first by the index* α_n *then by* β_n *linking both indexes in order to suppress the bias terms asymptotically .*

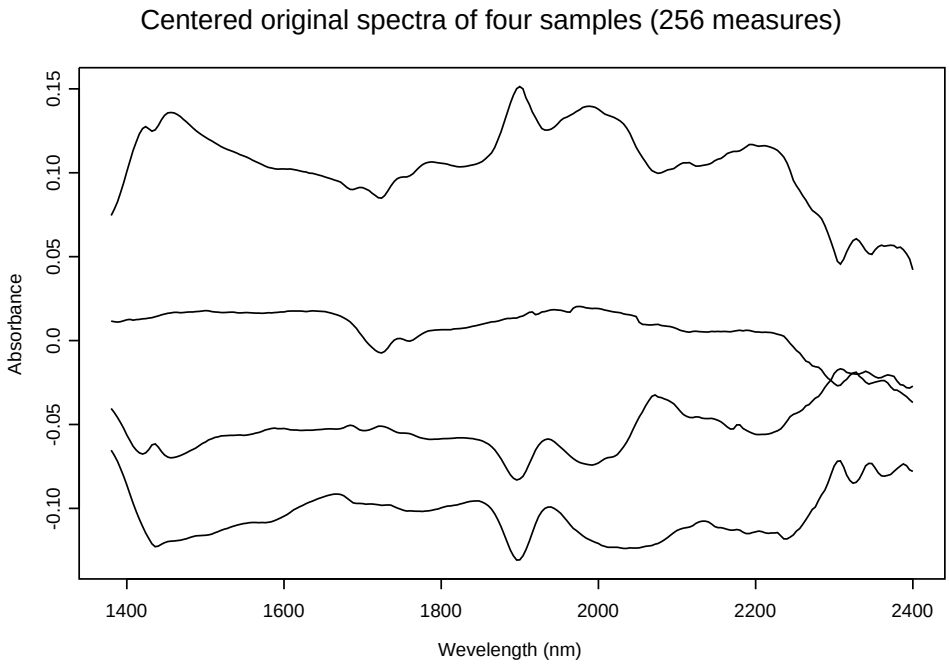

Figure 1: Centered original spectra of four samples (256 measures)

7 An application to spectrometric data

In this section we will present an application of the Functional Linear Regression with Derivatives (FLRD) introduced in this paper to a spectroscopic calibration problem. Quantitative NIR (near-infrared) spectroscopy is used to analyze food and agricultural materials. The NIR spectrum of a sample is a continuous curve giving the absorption, that is $\log_{10} 1/R$ where R is the reflection of the sample, against wavelength measured in nanometers (nm).

In the `cookie` example considered here the aim is to predict the percentage of each ingredient y given the NIR spectrum x of the sample (see Osborne et al. (1984) for a full description of the experiment). The constituents under investigation are: fat, sucrose, dry flour, and water. There were 39 samples in the calibration set, sample number 23 having been excluded from the original 40 as an outlier, and a further validation set with 31 samples, again after the exclusion of one outlier.

An NIR reflectance spectrum is available for each dough. The original spectral data consists of 700 points measured from 1100 to 1498 nm in steps of 2 nm. Following Brown et al. (2001) we reduced the number of spectral points to 256 by considering only the spectral range 1380-2400 nm in step of 4 nm. Samples of centered spectra are plotted in Figure 1.

A classical tool employed in the chemiometric literature for the prediction of y knowing the associated NIR spectra $(x_j, j = 1, \ldots, 256)$ is the linear model:

$$y = \sum_{j=1,256} \theta_j x_j + \epsilon \tag{15}$$

The problem then is to use the calibration data to estimate the unknown parameters θ_j. Clearly in this application since $39 \ll 256$ the ordinary least squares fails and many authors proposed to use alternative methods to tackle the problem: principal component regression (PCR) or partial least squares regression (PLS). We invite the reader to look at the paper of Frank and Friedman (1993) for a statistical view of some chemiometrics regression tools.

Following an idea of Hastie and Mallows, in their discussion of Frank and Friedman's paper, we consider a spectrum as a functional observation. The functional Linear Regression (FLR) corresponding to the model 15 defined above is:

$$y = \int_\delta x(t)\theta(t)dt + \varepsilon$$

where y is a scalar random variable, x a real function defined on $\delta = [1100, 2400]$ and $\theta(t)$ the unknown parameter function. Brown et al. (2001), Ferraty and Vieu (2003), Marx and Eilers (2002) or Amato et al. (2006) used such a model for a prediction problem with spectrometric data.

The model FLRD introduced in this paper can be written as:

$$y = \int_\delta x(t)\phi(t)dt + \int_\delta x'(t)\psi(t)dt + \varepsilon$$

where $\phi(t)$ and $\psi(t)$ are unknown functions (see display (5) for an equivalent definition). In this paragraph we compare the performance of PCR, PLS, FLR, FLRD, Spline Smoothing model proposed by Cardot, Ferraty and Sarda (2006) and Bayes wavelet predictions proposed by Brown et al. (2001).

We used the calibration data set for the estimation of parameter functions $\phi(t)$ and $\psi(t)$ and validation data for calculation of the MSEP (Mean Squared Error of Predictions):

$$MSEP = \frac{1}{31}\sum_{j=1}^{31}(y_j - \hat{y}_j)^2$$

where $\hat{y}_j$ is the prediction of y_j obtained by the model with estimated parameters. The choice of the parameters α and β is crucial for the prediction

model. We used a cross-validation approach based on the evaluation of the standard error of prediction $CVMSEP$:

$$CVMSEP(\alpha,\beta) = \frac{1}{39}\sum_{i=1}^{39}[\frac{1}{38}\sum\nolimits_{j=1}^{38}(y_i^c - \hat{y}_j^c(i;\alpha,\beta))^2],$$

where $\hat{y}_j^c(i;\alpha,\beta)$ denotes the prediction of y_j^c in the calibration set without sample i. Results for different methods of prediction of four ingredients are displayed in Table 1. We used B-spline basis ($k = 100$) for obtaining predictions with Spline Smoothing, Spline Ridge RLF and Spline RLFD methods. For each of those methods we give the values of the smoothing or penalty parameters based on an analogous cross-validation approach.

	MSE Validation			
Method and parameters	Fat	Sugar	Flour	Water
PLS	0.151	0.583	0.375	0.105
PCR	0.160	0.614	0.388	0.106
Spline Smoothing ($k_n = 8$)	0.546	0.471	2.226	0.183
Spline Ridge FLR ($\beta = 0.00002$)	0.044	0.494	0.318	0.087
Spline FLRD ($\alpha = 0.07, \beta = 0.15$)	0.092	0.450	0.332	0.069
Bayes Wavelet	0.063	0.449	0.348	0.050

Table 1: $MSEP$ criterion for all models (see Brown et al. for results of PLS, PCR and Bayes wavelet methods).

We note that functional approaches work better then PLS or PCR methods for the four predicted variables with respect to $MSEP$ criterion. Our simulation, as noted also by Marx and Eilers (2002), show that functional methods lead to more stable prediction. The Spline FLRD method produces in general equivalent results in terms of predictions with the best methods presented in table 1.

8 Proofs

In the sequel M and M' will stand for constants.

Let S and T be two selfadjoint linear operators on a Hilbert space H, we denote $T \ll S$ whenever for all x in H, $\langle Tx, x\rangle \leq \langle Sx, x\rangle$ then $\|T\|_\infty \leq \|S\|_\infty$.

The norm in the space $L^2(B)$ where $(B, \|\cdot\|_B)$ is a Banach space is defined the following way : let X be a random element in the Banach space B, then

$$\|X\|_{L^2(B)} = \left(E\|X\|_B^2\right)^{1/2}$$

When the notation is not ambiguous we systematically drop the index B i.e : $\|X\|_{L^2} = \left(E\|X\|_B^2\right)^{1/2}$.

8.1 Preliminary facts :

In order to gain some clarity in the proofs and to alleviate them we first list a few results stemming from operator or probabillity theory.

Fact 1: If T is a positive operator (either random or not), $T + \gamma I$ is invertible for all $\gamma > 0$ with bounded inverse and $\left\|(T+\gamma I)^{-1}\right\|_\infty \leq \gamma^{-1}$. Hence

$$\left\|\Gamma_n^\dagger\right\|_\infty = \left\|\Gamma^\dagger\right\|_\infty = \left\|\Gamma_n''^\dagger\right\|_\infty = \left\|\Gamma''^\dagger\right\|_\infty = \alpha^{-1} \tag{16}$$

Fact 2: As a consquence of assumption **A1** and of the strong law of large numbers for Hilbert valued random elements (see Ledoux, Talagrand (1991) Chapter 7),

$$T_n \underset{n\to+\infty}{\to} T \quad a.s.$$

whenever $T_n = \Gamma_n, \Gamma'_n, \Gamma'^*_n, \Gamma''_n$ (resp. $T = \Gamma, \Gamma', \Gamma'^*, \Gamma''$) since all theses random operators may be rewritten as sums of i.i.d. random variables. These sequences of random operators are almost surely bounded

$$\sup_n \|T_n\|_\infty \leq M \quad a.s. \tag{17}$$

which also means that

$$\max\left(\sup_n \|\delta_n\|_W, \sup_n \|\delta'_n\|_L\right) \leq M' \tag{18}$$

since (for instance) $\delta_n = \Gamma_n\phi + \Gamma'_n\psi + e_n$ where e_n is again a sum of i.i.d random elements :

$$e_n = \frac{1}{n}\sum_{k=1}^n X_k\varepsilon_k$$

We also set

$$e'_n = \frac{1}{n}\sum_{k=1}^n X'_k\varepsilon_k$$

(see below for details).

Fact 3: The Central Limit Thorem in Hilbert spaces (or standards results on rates of convergence for Hilbert valued random elements in square norm) provide a rate in the L^2 convergence of several random variables of interest in the proofs. See for instance Ledoux, Talagrand (1991) or Bosq

(2000) . Whenever $T_n = \Gamma_n, \Gamma'_n, \Gamma'^*_n, \Gamma''_n$ (resp. $T = \Gamma, \Gamma', \Gamma'^*, \Gamma''$) we have $\mathbb{E}\left\|T_n - T\right\|^2_{\mathcal{HS}} = O\left(\frac{1}{n}\right)$ hence

$$\left\|T_n - T\right\|_{L^2(\mathcal{HS})} = O\left(\frac{1}{\sqrt{n}}\right) \tag{19}$$

since all theses random operators may be rewritten as sums of i.i.d. random variables.

We begin with proving Proposition 1.

Proof of Proposition 1 :

The method of the proof may be adapted from the model studied in Mas, Pumo (2006). The couple (ϕ, ψ) will be identified whenever, for any other couple (ϕ_a, ψ_a), if

$$\left\{ \begin{array}{l} \delta = \Gamma\phi + \Gamma'\psi = \Gamma\phi_a + \Gamma'\psi_a, \\ \delta' = \Gamma'^*\phi + \Gamma''\psi = \Gamma'^*\phi_a + \Gamma''\psi_a. \end{array} \right.$$

$(\phi_a, \psi_a) = (\phi, \psi)$. This will be true if

$$\left\{ \begin{array}{l} \Gamma\left(\phi - \phi_a\right) + \Gamma'\left(\psi - \psi_a\right) = 0, \\ \Gamma'^*\left(\phi - \phi_a\right) + \Gamma''\left(\psi - \psi_a\right) = 0. \end{array} \right.$$

This means that the couple $(\phi - \phi_a, \psi - \psi_a)$ belongs to the kernel of the linear operator defined blockwise on $W \times L$ by :

$$\left(\begin{array}{cc} \Gamma & \Gamma' \\ \Gamma'^* & \Gamma'' \end{array} \right).$$

As $\Gamma'^* = D\Gamma$ and $\Gamma'' = D\Gamma'$, the Proposition will be proved if the blockwise operator defined on $W \times L$ and with values in W :

$$\left(\begin{array}{cc} \Gamma & \Gamma' \end{array} \right) = \left(\begin{array}{cc} \Gamma & \Gamma D^* \end{array} \right)$$

is one to one. It is plain that the kernel of this operator is precisely the space $\mathcal{N}$ that appears at display (8).

This finishes the proof of the Proposition.

The next two general Propositions are proved for further purpose.

Proposition 7

$$\begin{aligned} \sup_n \left\| \left(\Gamma''^{\dagger}_n\right)^{1/2} \Gamma'^*_n \right\|_\infty &< M \quad a.s., \\ \sup_n \left\| \Gamma'_n \left(\Gamma''^{\dagger}_n\right)^{1/2} \right\|_\infty &< M \quad a.s., \\ \sup_n \left\| \left(\Gamma''^{\dagger}\right)^{1/2} \Gamma'^* \right\|_\infty &< M, \\ \sup_n \left\| \Gamma' \left(\Gamma''^{\dagger}\right)^{1/2} \right\|_\infty &< M. \end{aligned}$$

Proof. We prove only the first bound since the method may be copied for the other ones. Set $R_n = D\Gamma_n^{1/2}$ then :

$$\Gamma_n'' = R_n R_n^*,$$
$$\left(\Gamma_n''^{\dagger}\right)^{1/2} \Gamma_n'^{*} = \left(R_n R_n^* + \alpha I\right)^{-1/2} R_n \Gamma_n^{1/2}.$$

At last,

$$\left\| \left(\Gamma_n''^{\dagger}\right)^{1/2} \Gamma_n'^{*} \right\|_\infty \leq \left\| \left(R_n R_n^* + \alpha I\right)^{-1/2} R_n \right\|_\infty \left\| \Gamma_n^{1/2} \right\|_\infty .$$

It is plain that

$$\sup_n \left\| \Gamma_n^{1/2} \right\|_\infty \leq M \quad a.s.$$

If the Schmidt decomposition of R_n is :

$$R_n = \sum_{k \in \mathbb{N}} \mu_{k,n} \left(u_{k,n} \otimes v_{k,n} \right),$$

($u_{k,n} \in W,\ v_{k,n} \in L$) it is simple algebra to get :

$$\left(R_n R_n^* + \alpha I\right)^{-1/2} R_n = \sum_{k \in \mathbb{N}} \frac{\mu_{k,n}}{\sqrt{\mu_{k,n}^2 + \alpha}} \left(u_{k,n} \otimes v_{k,n} \right) \tag{20}$$

which yields $\left\| \left(R_n R_n^* + \alpha I\right)^{-1/2} R_n \right\|_\infty = \sup_k \left\{ \frac{\mu_{k,n}}{\sqrt{\mu_{k,n}^2 + \alpha}} \right\} \leq 1.$ ■

Proposition 8

$$\left\| \left(S_n + \beta I\right)^{-1} \right\|_\infty \leq \frac{1}{\beta}.$$

Proof. The proof of this Lemma is similar to Lemma 7.4 in Mas, Pumo (2006). It was then proved for S instead of S_n and all operators should be changed to their empirical counterparts (e.g : Γ_n insted of Γ). We give a sketch of it. The proof relies on the Schmidt decomposition of S_n. One would get

$$S_n = \Gamma_n^{1/2} \Lambda_n \left(\alpha\right) \Gamma_n^{1/2}$$

where $\Lambda_n \left(\alpha\right)$ and $\Gamma_n^{1/2}$ are symmetric positive operators, which implies that S_n itself is positive. It suffices then to apply **Fact 2** (see the "Preliminary facts" subsection) to get the desired result. ■

8.2 Outline of the proof of Theorem 3 :

The following bound is valid :

$$\begin{aligned}
&\left[\widehat{y}_{n+1} - \left(\langle \phi, X_{n+1}\rangle_W + \langle \psi, X'_{n+1}\rangle_L\right)\right]^2 \\
&= \left(\left\langle \phi - \widehat{\phi}, X_{n+1}\right\rangle_W + \left\langle \psi - \widehat{\psi}, X'_{n+1}\right\rangle_L\right)^2 \\
&\leq 2\left[\left\langle \phi - \widehat{\phi}, X_{n+1}\right\rangle_W^2 + \left\langle \psi - \widehat{\psi}, X'_{n+1}\right\rangle_L^2\right].
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{E}\left\langle \phi - \widehat{\phi}, X_{n+1}\right\rangle_W^2 &= \mathbb{E}\left[\mathbb{E}\left\langle \phi - \widehat{\phi}, X_{n+1}\right\rangle_W^2 |X_1, ..., X_n\right] \\
&= \mathbb{E}\left[\mathbb{E}\left\langle \phi - \widehat{\phi}, X_{n+1}\right\rangle_W^2 |\widehat{\phi}\right] \\
&= \mathbb{E}\left[\left\|\Gamma^{1/2}\left(\phi - \widehat{\phi}\right)\right\|_W^2\right]
\end{aligned}$$

Similarly,

$$\mathbb{E}\left\langle \psi - \widehat{\psi}, X'_{n+1}\right\rangle_L^2 = \mathbb{E}\left[\left\|\Gamma''^{1/2}\left(\psi - \widehat{\psi}\right)\right\|_L^2\right]$$

Both preceding equations feature similar expressions. We focus on the term involving ϕ ; we will prove that :

$$\mathbb{E}\left[\left\|\Gamma^{1/2}\left(\phi - \widehat{\phi}\right)\right\|_W^2\right] = O\left(\frac{\beta^2}{\alpha^2}\right) + O\left(\frac{1}{\alpha^2\beta^2 n}\right).$$

Within the proof the reader will easily be convinced that the method would lead to an analogous result for the term with ψ. From now in order to alleviate notations we drop the index ϕ in $S_{n,\phi}$ and $u_{n,\phi}$. The sequences $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ will be denoted α and β respectively and for short.

We start from

$$\begin{aligned}
\widehat{\phi}_n &= \left(\Gamma_n - \Gamma'_n\Gamma''^{\dagger}_n\Gamma'^{*}_n + \beta I\right)^{-1}\left(\delta_n - \Gamma'_n\Gamma''^{\dagger}_n\delta'_n\right) \\
&= \left(S_n + \beta I\right)^{-1} u_n \\
\phi &= \left(\Gamma - \Gamma'\Gamma''^{\dagger}\Gamma'^{*}\right)^{-1}\left(\delta - \Gamma'\Gamma''^{\dagger}\delta'\right) \\
&= S^{-1}u
\end{aligned}$$

where we recall that :

$$
\begin{aligned}
u &= \delta - \Gamma' \Gamma''^{\dagger} \delta', \\
u_n &= \delta_n - \Gamma'_n \Gamma''^{\dagger}_n \delta'_n, \\
S &= \Gamma - \Gamma' \Gamma''^{\dagger} \Gamma'^{*}, \\
S_n &= \Gamma_n - \Gamma'_n \Gamma''^{\dagger}_n \Gamma'^{*}_n.
\end{aligned}
$$

The proof relies on the following decomposition :

$$
\begin{aligned}
\widehat{\phi}_n - \phi &= (S_n + \beta I)^{-1} (u_n - u) + \left((S_n + \beta I)^{-1} - S^{-1} \right) u \\
&= (S_n + \beta I)^{-1} (u_n - u) + (S_n + \beta I)^{-1} (S - S_n - \beta I) S^{-1} u \\
&= A_n + B_n + C_n \qquad (21)
\end{aligned}
$$

where

$$
\begin{aligned}
A_n &= (S_n + \beta I)^{-1} (u_n - u) \qquad (22) \\
B_n &= (S_n + \beta I)^{-1} (S - S_n) \phi \qquad (23) \\
C_n &= \beta (S_n + \beta I)^{-1} \phi \qquad (24)
\end{aligned}
$$

Along the forthcoming Lemmas we determine rates of convergence for these three terms. We will prove that the rate of decrease to zero in L^2 norm is $(\alpha\beta\sqrt{n})^{-1}$ for A_n and B_n. The rest of the proof of the main Theorem is postponed to the end of the next and last subsection.

8.3 Proof of the main Theorem

The first Lemma gives a reta of convergence for $S_n - S$.

Lemma 9 *The following holds :*

$$
S_n - S = \Gamma_n - \Gamma'_n \Gamma''^{\dagger}_n \Gamma'^{*}_n - \Gamma + \Gamma' \Gamma''^{\dagger} \Gamma'^{*} = O_{L^2}\left(\frac{1}{\alpha\sqrt{n}} \right)
$$

Proof. First of all by (19) :

$$
\| \Gamma_n - \Gamma \|_{L^2(\mathcal{HS})} = O\left(\frac{1}{\sqrt{n}} \right)
$$

We focus on

$$
\begin{aligned}
&\Gamma'_n \Gamma''^{\dagger}_n \Gamma'^{*}_n - \Gamma' \Gamma''^{\dagger} \Gamma'^{*} \\
&= \Gamma'_n \Gamma''^{\dagger}_n \Gamma'^{*}_n - \Gamma' \Gamma''^{\dagger}_n \Gamma'^{*}_n + \Gamma' \Gamma''^{\dagger}_n \Gamma'^{*}_n - \Gamma' \Gamma''^{\dagger}_n \Gamma'^{*} \\
&+ \Gamma' \Gamma''^{\dagger}_n \Gamma'^{*} - \Gamma' \Gamma''^{\dagger} \Gamma'^{*}.
\end{aligned}
$$

Then dealing with each of these three terms separatedly we get

$$\begin{aligned}\left\|\Gamma'_n\Gamma''^{\dagger}_n\Gamma'^{*}_n - \Gamma'\Gamma''^{\dagger}_n\Gamma'^{*}_n\right\|_\infty &\leq \|\Gamma'_n - \Gamma'\|_\infty \left\|\Gamma''^{\dagger}_n\Gamma'^{*}_n\right\|_\infty \\ &\leq \|\Gamma'_n - \Gamma'\|_\infty \left\|\left(\Gamma''^{\dagger}_n\right)\right\|_\infty \|\Gamma'^{*}_n\|_\infty \\ &\leq C\frac{\|\Gamma'_n - \Gamma'\|_\infty}{\alpha} \quad a.s.\end{aligned}$$

The last bound was derived from (17) and (16).

$$\begin{aligned}&\left\|\Gamma'\Gamma''^{\dagger}_n\left(\Gamma'^{*}_n - \Gamma'^{*}\right)\right\|_\infty \\ &\leq C\frac{\|\Gamma'^{*}_n - \Gamma'^{*}\|_\infty}{\alpha} \quad a.s.\end{aligned}$$

At last,

$$\begin{aligned}\Gamma'\left(\Gamma''^{\dagger}_n - \Gamma''^{\dagger}\right)\Gamma'^{*} &= \Gamma'\Gamma''^{\dagger}_n\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}\Gamma'^{*} \\ &= \Gamma'\Gamma''^{\dagger}_n\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}\Gamma'^{*} \\ &= \left(\Gamma' - \Gamma'_n\right)\Gamma''^{\dagger}_n\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}\Gamma'^{*} \\ &+ \Gamma'_n\Gamma''^{\dagger}_n\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}\Gamma'^{*}\end{aligned}$$

Then,

$$\begin{aligned}&\left\|\Gamma'\left(\Gamma''^{\dagger}_n - \Gamma''^{\dagger}\right)\Gamma'^{*}\right\|_\infty \\ &\leq \left\|\left(\Gamma' - \Gamma'_n\right)\Gamma''^{\dagger}_n\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}\Gamma'^{*}\right\|_\infty \\ &+ \left\|\Gamma'_n\left(\Gamma''^{\dagger}_n\right)^{1/2}\right\|_\infty \left\|\left(\Gamma''^{\dagger}_n\right)^{1/2}\left(\Gamma'' - \Gamma''_n\right)\left(\Gamma''^{\dagger}\right)^{1/2}\right\|_\infty \left\|\left(\Gamma''^{\dagger}\right)^{1/2}\Gamma'^{*}\right\|_\infty\end{aligned}$$

By Proposition 7 the second term may be bounded by

$$C\left\|\left(\Gamma''^{\dagger}_n\right)^{1/2}\left(\Gamma'' - \Gamma''_n\right)\left(\Gamma''^{\dagger}\right)^{1/2}\right\|_\infty = O_{L^2}\left(\frac{1}{\alpha\sqrt{n}}\right)$$

since

$$\left\|\Gamma''^{\dagger 1/2}_n\right\|_\infty = \left\|\Gamma''^{\dagger 1/2}\right\|_\infty = \alpha^{-1/2}.$$

Cauchy-Schwartz inequality yields for the first :

$$\begin{aligned}&\mathbb{E}\left\|\left(\Gamma' - \Gamma'_n\right)\Gamma''^{\dagger}_n\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}\Gamma'^{*}\right\|^2_\infty \\ &\leq M\left(\mathbb{E}\left\|\left(\Gamma' - \Gamma'_n\right)\Gamma''^{\dagger}_n\right\|^4_\infty \mathbb{E}\left\|\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}\right\|^4_\infty\right)^{1/2} \\ &\leq M\frac{1}{n^2\alpha^4},\end{aligned}$$

hence

$$\left\|\left(\Gamma' - \Gamma'_n\right)\Gamma''^{\dagger}_n\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}\Gamma'^{*}\right\| = O_{L^2}\left(\frac{1}{\alpha^2 n}\right).$$

The proof of Lemma 9 is finished. ■

Lemma 10 *We have :*

$$u_n - u = O_{L^2}\left(\frac{1}{\alpha\sqrt{n}}\right).$$

Proof. We start with :

$$u_n - u = \delta_n - \delta + \Gamma'\Gamma''^{\dagger}\delta' - \Gamma'_n\Gamma''^{\dagger}_n\delta'_n.$$

Clearly $\delta_n - \delta = O_{L^2}\left(\frac{1}{\sqrt{n}}\right)$ and we study the second term

$$\begin{aligned}\Gamma'_n\Gamma''^{\dagger}_n\delta'_n - \Gamma'\Gamma''^{\dagger}\delta' = &\left(\Gamma'_n - \Gamma'\right)\Gamma''^{\dagger}_n\delta'_n + \Gamma'\left(\Gamma''^{\dagger}_n - \Gamma''^{\dagger}\right)\delta'_n \\ &+ \Gamma'\Gamma''^{\dagger}\left(\delta'_n - \delta'\right).\end{aligned}$$

Since δ'_n is almost surely bounded (see (18)), $\Gamma'_n - \Gamma = O_{L^2}\left(\frac{1}{\sqrt{n}}\right)$, $\delta'_n - \delta' = O_{L^2}\left(\frac{1}{\sqrt{n}}\right)$ and $\left\|\Gamma''^{\dagger}\right\|_\infty = \left\|\Gamma''^{\dagger}_n\right\|_\infty = \alpha^{-1}$ we get :

$$\begin{aligned}\left\|\left(\Gamma'_n - \Gamma'\right)\Gamma''^{\dagger}_n\delta'_n\right\|_W &= O_{L^2}\left(\frac{1}{\alpha\sqrt{n}}\right), \\ \left\|\Gamma'\Gamma''^{\dagger}\left(\delta'_n - \delta'\right)\right\|_W &= O_{L^2}\left(\frac{1}{\alpha\sqrt{n}}\right).\end{aligned}$$

The remaining term is

$$\begin{aligned}\Gamma'\left(\Gamma''^{\dagger}_n - \Gamma''^{\dagger}\right)\delta'_n &= \Gamma'\Gamma''^{\dagger}\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}_n\delta'_n \\ &= \Gamma'\Gamma''^{\dagger}\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}_n\left(\Gamma'^{*}_n\phi + \Gamma''_n\psi + u'_n\right), \\ &= \Gamma'\left(\Gamma''^{\dagger}\right)^{1/2}\left(m_1 + m_2 + m_3\right)\end{aligned}$$

where

$$\begin{aligned}m_1 &= \left(\Gamma''^{\dagger}\right)^{1/2}\left(\Gamma'' - \Gamma''_n\right)\left(\Gamma''^{\dagger}_n\right)^{1/2}\left(\Gamma''^{\dagger}_n\right)^{1/2}\Gamma'^{*}_n\phi, \\ m_2 &= \left(\Gamma''^{\dagger}\right)^{1/2}\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}_n\Gamma''_n\psi, \\ m_3 &= \left(\Gamma''^{\dagger}\right)^{1/2}\left(\Gamma'' - \Gamma''_n\right)\Gamma''^{\dagger}_n e'_n.\end{aligned}$$

First we drop $\Gamma'\left(\Gamma''^{\dagger}\right)^{1/2}$ since the norm of this operator may be bounded by a constant independent from α (see Proposition 7). We turn to :

$$\begin{aligned}\|m_1\| &\leq M\left\|\left(\Gamma'' - \Gamma''_n\right)\right\|_\infty\left\|\left(\Gamma''^{\dagger}_n\right)^{1/2}\right\|_\infty\left\|\left(\Gamma''^{\dagger}\right)^{1/2}\right\|_\infty, \\ \|m_2\| &\leq \|\psi\|_L\left\|\left(\Gamma''^{\dagger}\right)^{1/2}\right\|_\infty\left\|\left(\Gamma'' - \Gamma''_n\right)\right\|_\infty\end{aligned}$$

since $\left\|\Gamma_n''^{\dagger}\Gamma_n''\right\|_\infty \leq 1$ almost surely. The consequence of the display above is $\|m_1\|_{L^2} = O\left(\frac{1}{\alpha\sqrt{n}}\right)$ and $\|m_2\|_{L^2} = O\left(\frac{1}{\sqrt{\alpha n}}\right)$.

We can deal with m_3 as was done within the proof of the preceding Lemma 9. Clearly we may cope with m_3 as if the random $\Gamma_n''^{\dagger}$ was replaced by the non random $\Gamma''^{\dagger}$. We should study

$$\left[\left(\Gamma''^{\dagger}\right)^{1/2}\left(\Gamma'' - \Gamma_n''\right)\left(\Gamma''^{\dagger}\right)^{1/2}\right]\left[\left(\Gamma''^{\dagger}\right)^{1/2} e_n'\right].$$

It is enough to get a rate of decrease for each of the these terms. Once again we have :

$$\left\|\left(\Gamma''^{\dagger}\right)^{1/2}\left(\Gamma'' - \Gamma_n''\right)\left(\Gamma''^{\dagger}\right)^{1/2}\right\|_{L^2} = O\left(\frac{1}{\alpha\sqrt{n}}\right)$$

$$\left\|\left(\Gamma''^{\dagger}\right)^{1/2} e_n'\right\|_{L^2} = O\left(\frac{1}{\sqrt{\alpha n}}\right)$$

which completes the proof of Lemma 10. ■

Now we are ready to go back to (22) and (23) as announced sooner.

Lemma 11 *We have :*

$$A_n = O_{L^2}\left(\frac{1}{\alpha\beta\sqrt{n}}\right),$$

$$B_n = O_{L^2}\left(\frac{1}{\alpha\beta\sqrt{n}}\right).$$

Proof. Since

$$\|A_n\| \leq \left\|\left(S_n + \beta I\right)^{-1}\right\|_\infty \|u_n - u\|_W$$

by Lemma 10 and Proposition 8 we get the first desired result ■

Once again the proof of the second relies on Proposition 8 and Lemma 9. Indeed

$$\begin{aligned}\|B_n\|_W &\leq \left\|\left(S_n + \beta I\right)^{-1}\right\|_\infty \|S - S_n\|_\infty \|\phi\|_W \\ &\leq \frac{\|\phi\|_W}{\beta}\|S - S_n\|_\infty\end{aligned}$$

hence the result.

We should deal with the last term. In a first step we prove that S_n may be replaced by S.

Lemma 12 *When* $\alpha\beta\sqrt{n} \to +\infty$,

$$C_n = \beta\left(S + \beta I\right)^{-1}\phi\left(1 + o\left(1\right)\right).$$

Remark 13 *The preceding equality should be understood with respect to the L^2 norm.*

Proof. Successively,

$$\begin{aligned} C_n &= \beta \left(S_n + \beta I\right)^{-1} \phi \\ &= \beta \left(\left(S_n + \beta I\right)^{-1} - \left(S + \beta I\right)^{-1}\right) \phi + \beta \left(S + \beta I\right)^{-1} \phi \\ &= \left[\left(\left(S_n + \beta I\right)^{-1} \left(S - S_n\right)\right) + I\right] \beta \left(S + \beta I\right)^{-1} \phi \end{aligned}$$

and

$$\|C_n\| \leq \left\|\beta \left(S + \beta I\right)^{-1} \phi\right\|_W \left(1 + \left\|\left(S_n + \beta I\right)^{-1} \left(S - S_n\right)\right\|_\infty\right).$$

Now it suffices to apply Lemma 11 to get the desired result. ■

The next Lemma may be hard to understand at first glance. Within the forthcoming proof of Theorem 3 the bias term C_n will slightly change. We refer to displays (28) and (29) below for a deeper understanding.

Lemma 14 *The following holds :*

$$\left\|\Gamma^{1/2} \left(S + \beta I\right)^{-1} \Gamma^{1/2}\right\|_\infty = O\left(\frac{1}{\alpha}\right).$$

Proof. Once again it takes two steps to get the result. First note that $\Gamma^{1/2} \left(S\right)^{-1} \Gamma^{1/2}$ is a bounded linear operator. Indeed

$$S = \Gamma - \Gamma' \Gamma''^{\dagger} \Gamma'^{*} = \Gamma^{1/2} \Lambda_\alpha \Gamma^{1/2} \tag{25}$$

where $R = D\Gamma^{1/2}$,

$$\Lambda_\alpha = I - R^* \left(RR^* + \alpha I\right)^{-1} R.$$

The Schmidt decomposition of R is (see (20) above for the empirical version) :

$$R = \sum_{k \in \mathbb{N}} \mu_k \left(u_k \otimes v_k\right).$$

where $\left(u_k\right)_{k \in \mathbb{N}}$ (resp. $\left(v_k\right)_{k \in \mathbb{N}}$) is a complete orthonormal system in W (resp. L). Hence :

$$\begin{aligned} \Lambda_\alpha &= \sum_{k \in \mathbb{N}} \left(1 - \frac{\mu_k^2}{\mu_k^2 + \alpha}\right) \left(u_k \otimes u_k\right) \\ &= \sum_{k \in \mathbb{N}} \frac{\alpha}{\mu_k^2 + \alpha} \left(u_k \otimes u_k\right). \end{aligned}$$

The operator Λ_α has a bounded inverse

$$\Lambda_\alpha^{-1} = \frac{1}{\alpha} \sum_{k \in \mathbb{N}} \left(\mu_k^2 + \alpha\right) (u_k \otimes u_k)$$

and $\left\|\Lambda_\alpha^{-1}\right\|_\infty = 1 + \left(\sup \mu_k^2\right)/\alpha \leq M/\alpha$ for M large enough (or α small enough).
Hence

$$\Gamma^{1/2} (S)^{-1} \Gamma^{1/2} = \Gamma^{1/2} \Gamma^{-1/2} \Lambda_\alpha^{-1} \Gamma^{-1/2} \Gamma^{1/2} = \Lambda_\alpha^{-1}. \tag{26}$$

Now (second step) we prove that :

$$\Gamma^{1/2} (S + \beta I)^{-1} \Gamma^{1/2} \ll \Gamma^{1/2} S^{-1} \Gamma^{1/2}.$$

Let us pick a given x in W, then

$$\left\langle \Gamma^{1/2} (S + \beta I)^{-1} \Gamma^{1/2} x, x \right\rangle_W = \left\langle (S + \beta I)^{-1} \Gamma^{1/2} x, \Gamma^{1/2} x \right\rangle_W$$

It suffices to get for all y in in the domain of operator $\Gamma^{-1/2}$:

$$\left\langle (S + \beta I)^{-1} y, y \right\rangle_W \leq \left\langle S^{-1} y, y \right\rangle_W \tag{27}$$

Standard results on the spectrum of $(S + \beta I)^{-1} S$ prove that $(S + \beta I)^{-1} S \geq 0$ and that $\left\|(S + \beta I)^{-1} S\right\| \leq 1$ which is enough to claim (27).
We are now in position to finixh the proof of the Lemma. It is plain from (27) that

$$\left\|\Gamma^{1/2} (S + \beta I)^{-1} \Gamma^{1/2}\right\|_\infty \leq \left\|\Gamma^{1/2} (S)^{-1} \Gamma^{1/2}\right\|_\infty = \left\|\Lambda_\alpha^{-1}\right\|_\infty \leq \frac{C}{\alpha}$$

which is the claimed result. ■

Proof of Theorem 3:
Now starting from (21) we get

$$\begin{aligned} \left\|\Gamma^{1/2} \left(\phi - \widehat{\phi}\right)\right\|_W^2 &\leq M \left\|\Gamma^{1/2} (A_n + B_n + C_n)\right\|_W^2 \\ &\leq M \left(\|A_n\|_W^2 + \|B_n\|_W^2 + \left\|\Gamma^{1/2} C_n\right\|_W^2\right). \end{aligned} \tag{28}$$

Lemmas 11 gives the rates of convergence for $\|A_n\|_W^2$ and $\|B_n\|_W^2$ respectively. But Lemma 12 is unfortunately not enough to get a rate in the last term. However this previous Lemma enables to focus on :

$$\beta \Gamma^{1/2} (S + \beta I)^{-1} \phi = \beta \Gamma^{1/2} (S + \beta I)^{-1} \Gamma^{1/2} \Gamma^{-1/2} \phi \tag{29}$$

and

$$\left\|\Gamma^{1/2}C_n\right\|_W^2 \leq M\beta^2 \left\|\Gamma^{1/2}\left(S+\beta I\right)^{-1}\Gamma^{1/2}\right\|_\infty^2 \left\|\Gamma^{-1/2}\phi\right\|_W . \tag{30}$$

By assumption **A4**, $\left\|\Gamma^{-1/2}\phi\right\|_W$ is finite. We deal with the central term, namely :

$$\begin{aligned}\Gamma^{1/2}\left(S+\beta I\right)^{-1}\Gamma^{1/2} &= \Gamma^{1/2}\left(\Gamma^{1/2}\Lambda_\alpha\Gamma^{1/2}+\beta I\right)^{-1}\Gamma^{1/2} \\ &\ll \Gamma^{1/2}\left(\Gamma^{1/2}\Lambda_\alpha\Gamma^{1/2}\right)^{-1}\Gamma^{1/2} = \Lambda_\alpha^{-1}.\end{aligned}$$

(see (25)) and

$$\left\|\Lambda_\alpha^{-1}\right\|_\infty^2 = O\left(\alpha^{-2}\right).$$

Collecting this last display with (30) we get

$$\left\|\Gamma^{1/2}C_n\right\|_W^2 = O\left(\frac{\beta^2}{\alpha^2}\right).$$

This finishes the proof of Theorem 3.

References

[1] Adams R.A. and Fournier J.J.F., 2003. Sobolev spaces, Academic Press, 2nd ed.

[2] Amato U., Antoniadis A., Feiss I., 2006. Dimension reduction in functional regression with applications, to appear in *Computational Statistics and Data Analysis.*

[3] Bosq, D., 2000. *Linear processes in function spaces.* Lectures notes in statistics. Springer Verlag.

[4] Brown, P.J., Fearn, T. and Vanucci M., 2001. Bayesian wavelet regression on curves with application to a spectroscopic calibration problem, *Journal of the American Statistical Association.*, 96 (454), 398–408.

[5] Cai T., Hall, P., 2006. Prediction in functional linear regression, *Annals of Statistics*, **34**, n°5.

[6] Cardot H., Ferraty F., Sarda P., 2006. Spline estimators for the functional linear model: Consistency, Application and Splus implementation, To appear in *Statistica Sinica.*

[7] Cardot H., Mas A., Sarda P, 2006. Weak convergence in the functional linear model. To appear in *Probability Theory and Related Fields.*

[8] Dunford, N. and Schwartz, J.T., 1988. *Linear Operators, Vol. I & II.* Wiley Classics Library.

[9] Ferraty, F., Vieu P., 2003. The functional nonparametric model and application to spectrometric data. *Computational Statistics* **17** n°4, 545-564.

[10] Ferraty, F., Vieu P., 2006. Nonparametic Functional Data Analysis, Springer.

[11] Frank I.E. and Friedman J.H., 1993. A statistical view of some chemometrics regression tools, *Technometrics*, **35**, no 2, 109-135.

[12] Gohberg, I., Goldberg, S. and Kaashoek,M.A., 1991. *Classes of linear operators Vol I & II. Operator Theory : advances and applications*, Birkhaüser Verlag.

[13] Hastie T. and Mallows C., 1993. Discussion on the paper of Frank I.E. and Friedman J.H., 1993, A statistical view of some chemometrics regression tools, *Technometrics*, **35**, no 2, 140-143.

[14] Ledoux, M. and Talagrand, M, 1991 : *Probability in Banach spaces. Isoperimetry and processes*. Ergebnisse der Mathematik und ihrer Grenzgebiete, **23**. Springer-Verlag, Berlin.

[15] Marx, B.D. and Eilers, P.H.C. (2002) Multivariate calibration stability: a comparison of methods, *Journal of Chemometrics*, **16**, 129-140.

[16] Mas A, Pumo, B, 2006. The ARHD model. To appear in JSPI. Technical report available at *http://fr.arxiv.org/PS_cache/math/pdf/0502/0502285.pdf*

[17] Marion J.M., Pumo B., 2004. Comparaison des modèles ARH(1) et ARHD(1) sur des données physiologiques, *Annales de l'ISUP*, **48**, 3, pp. 29-38.

[18] Osborne, B.J., Fearn, T., Miller, A.R. and Douglas, S. (1984) Application of Near-Infrared Reflectance Spectroscopy to Compositional Analysis of Biscuits Dougts, *J. of the Sc. of Food and Agricult.*, **35**, 99–105

[19] Ramsay J.O., 2000. Differential equation models for statistical functions, *Canadian Journal of Statistics*, **28**, n°2, 225-240.

[20] Ramsay J.O., Dalzell C.J., 1991, Some tools for functional data analysis (with dsicussion), *Journal of the Royal Statistical Society*, B, **53**, 539-572.

[21] Ramsay J.O., Silverman B.W., 1997. *Functional Data Analysis*, Springer.

[22] Ramsay J.O., Silverman B.W., 2002. *Applied Functional Data Analysis: Methods and Case Studies*, Springer.

[23] Silverman B.W., 1996. Smoothed functional principal component analysis by choice of norm, *Annals of Statistics*, **24**, No 1, 1-24.

[24] Tikhonov A.N., Arsenin V.Y., 1977. *Solutions of ill-posed problems*. V.H. Winstons and sons, Washington..

[25] Ziemer W.P., 1989. Weakly differentiable functions. *Sobolev spaces and functions of bounded variations*. Graduate Text in Mathematics 120. Springer-Verlag, New-York.