Ann Inst Stat Math (2011) 63:1047–1075
DOI 10.1007/s10463-010-0275-8

Local linear regression for functional data

A. Berlinet · A. Elamine · A. Mas

Received: 29 February 2008 / Revised: 28 July 2009 / Published online: 10 March 2010

Abstract We study a non-linear regression model with functional data as inputs and scalar response. We propose a pointwise estimate of the regression function that maps a Hilbert space onto the real line by a local linear method and derive its asymptotic mean square error. Computations involve a linear inverse problem as well as a representation of the small ball probability of the data and are based on recent advances in this area.

Keywords Functional data · Regression model · Kernel · Mean square error · Small ball probability · Inverse problem

1 Introduction

1.1 The data and the model

In probability theory, random functions have been for quite a long time under the lights. The tremendous advances in computer science and the opportunity to deal with data collected at a high frequency make it now possible for statisticians to study models for high-dimensional data. As a consequence, many of them focused their attention on models for functional data, i.e., models that are suited for curves, for instance, spectral curves, growth curves, or interest rate curves...

Even if seminal articles on functional data analysis date back to more than 20 years (see Dauxois et al. (1982)), this area is currently going through a deep bustle. The book

A. Berlinet · A. Elamine · A. Mas (✉)
Département de Mathématiques, CC 051, Université Montpellier 2,
Place Eugène Bataillon, 34095 Montpellier Cedex 5, France
e-mail: mas@math.univ-montp2.fr

by Ramsay and Silverman (1997) initiated a series of monographs on the subject: Bosq (2000), Ramsay and Silverman again (2002), Ferraty and Vieu (2006).

Functional Data Analysis has drawn much attention and many of the classical multivariate data analysis techniques such as Principal Component Analysis, Correlation Analysis, ANOVA, and Linear Discrimination were generalized to curves. Statistical inference gave and gives birth to many papers. Linear regression and autoregression, for instance, rise an interesting inverse problem (see Kneip et al. (2004); Yao et al. (2005); Müller and Stadtmüller (2005); Cai and Hall (2006); Cardot et al. (2007) and Mas (2007) and even more recently Crambes et al. (2008)). The case of nonparametric regression was introduced in Ferraty and Vieu (2003) then studied in Masry (2005) and Ferraty et al. (2007): a Nadaraya-Watson type estimator was proposed. This model is the starting point of our article.

In the sequel, we will consider a sample drawn from random elements with values in an infinite dimensional vector space: $X_1, \ldots, X_n$. Here $X_i = X_i(\cdot)$ is a random function defined, say, on a compact interval of the real line $[0, T]$. We will also assume once and for all that the X_i's take their values in a separable Hilbert space denoted H. This Hilbert space is endowed with an inner product $\langle \cdot, \cdot \rangle$ from which is derived the norm $\|\cdot\|$. Such techniques as wavelets or splines yield reconstructed curves in the (Hilbert) Sobolev spaces:

$$W^{m,2} = \left\{ f \in L^2([0, T]) : f^{(m)} \in L^2([0, T]) \right\}$$

where $f^{(m)}$ denotes the mth derivative of f. Further information on Sobolev spaces may be found in Adams and Fournier (2003). However, for the sake of generality we will consider H as the sequence space l_2, and any vector x will be classically decomposed over a basis, say $(e_i)_{i \in \mathbb{N}}$ so that

$$\|x\|^2 = \sum_{i=1}^{+\infty} \langle x, e_i \rangle^2 .$$

We are given a sample $(y_i, X_i)_{1 \le i \le n} \in (\mathbb{R} \times H)^{\otimes n}$ of independent and identically distributed data. Let m be the regression function that maps H onto $\mathbb{R}$.

The model is a classical non-parametric regression model:

$$y_i = m(X_i) + \varepsilon_i \quad 1 \le i \le n. \tag{1}$$

or, with other symbols

$$m(x_0) = \mathbb{E}(y | X = x_0)$$

where y and X stand for random variables with the same distributions as y_1 and X_1. The noise ε follows both assumptions

$$\mathbb{E}(\varepsilon | X) = 0,$$
$$\mathbb{E}\left(\varepsilon^2 | X\right) = \sigma_\varepsilon^2$$

and σ_{ε}^{2} does not depend on X. The issue of the expectation of X (should the X's be centered or not ?) is not crucial ; it will be addressed later on, but for simplicity we assume that $\mathbb{E}(X) = 0$. Let x_0 be a fixed and known point in H. We are aiming at estimating $m(x_0)$.

In finite dimension, and more precisely when X_i is a real-valued random variable, $m(x_0)$ may be estimated by considering an affine approximation of m around x_0:

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0)$$

when x is close to x_0. This approach leads us to look for a solution to the following minimization problem:

$$\min_{a\in\mathbb{R}, b\in\mathbb{R}} \sum_{i=1}^{n} (y_i - a - b(x_0 - X_i))^2 K\left(\frac{x_0 - X_i}{h}\right) \tag{2}$$

which is nothing but a mean square program weighted by the $K((x_0 - X_i)/h)$'s. Here, K is a kernel: a measurable positive function such that $\int K = 1$ and $h = h_n$ the bandwidth indexed by the sample size. Then a^*, one of the two solutions of the above display is the estimate of $m(x_0)$. As a special case taking $b = 0$ comes down to the classical Nadaraya–Watson estimator. We refer the interested reader to Nadaraya (1964) and Fan (1993) about this topic. The generalization of (2) to higher orders (namely approximating locally m by a polynomial) gives birth to the local polynomial estimate of $m(x_0)$. We refer, for instance, to Chen (2003) for a recent article. Convergence in probability and asymptotic normality of the kernel polynomial estimators for a density function, variable bandwidth, and local linear regression smoothers, were studied by Fan and Gijbels (1992).

When x belongs to a Hilbert space, the principle remains the same. The function m is now approximated by

$$m(x) \approx m(x_0) + \langle \varphi(x_0), x - x_0 \rangle$$

where $\varphi(x_0) \in H$ is the first-order derivative of m at x_0 (the gradient in fact) and the local linear estimate of m at x_0 stems from the following adapted weighted least square program:

$$\min_{a\in\mathbb{R}, \varphi\in H} \sum_{i=1}^{n} (y_i - a - \langle \varphi, X_i - x_0 \rangle)^2 K\left(\frac{\|X_i - x_0\|}{h}\right). \tag{3}$$

At last the estimate $\widehat{m}_n(x_0)$ of $m(x_0)$ is a^*, solution of (3). We refer to Barrientos-Marin et al. (2007) for another approach. These authors consider a program simplified from the one above (they replace the functional parameter φ with a scalar one). But display (3) seems to be a true generalization of (2) since φ like b estimates the derivative of m.

Remark 1 Investigating higher-order approximations turns out to be especially difficult in this functional setting. For instance, a local quadratic estimate involves the second-order derivative of m (the Hessian operator) which is a symmetric positive operator on H. The local linear method appears as a good trade-off between the complexity of the method and its accuracy.

Nevertheless, solving (3) is not so easy. The aim of the present work is to provide a bound for the mean square error of the estimate a^* of $m(x_0)$ that is

$$\mathbb{E}\left[\widehat{m}_n(x_0) - m(x_0)\right]^2$$

through a classical bias-variance decomposition. The present paper rises two kind of questions, both difficult to answer. The first one is about optimality. It will be seen that the rate of convergence of the proposed estimate outperforms the one already obtained in the literature on this model (see Ferraty et al. (2007)), but up to now no result on optimal rates is available in this context. The second one is about implementation of the method. Much work has to be done to select efficiently the bandwidth from the observed data. The article is organized as follows: the two next subsections are devoted to pointing out the two main problems that arise from the model and that are symptomatic of the functional framework. The needed assumptions are commented, then the central result is given before the last section which contains all mathematical derivations.

1.2 The estimate and the ill-posed problem

In order to go ahead we need to define two linear operators from H to H (the first is non-random, the second is random, based on the sample). The usual sup-norm for operator T will be denoted

$$\|T\|_\infty = \sup_{x \in \mathcal{B}_1} \|Tx\|$$

where $\mathcal{B}_1$ stands for the closed unit ball of H. From now on, the reader should be familiar with basic notions related to the theory of bounded and unbounded linear operators on Hilbert space. A wide literature exists on this topic which is central to the mathematical science. Some of our references are Weidman (1980), Akhiezer and Glazman (1981), Dunford and Schwartz (1988) and Gohberg et al. (1991) amongst many others.

Definition 2 The theoretical local covariance operator of X at $x_0 \in H$ associated with the kernel K is defined by

$$\Gamma_K = \mathbb{E}\left(K\left(\frac{\|X_1 - x_0\|}{h}\right)((X_1 - x_0) \otimes (X_1 - x_0))\right)$$

and its empirical counterpart is

$$\Gamma_{n,K} = \frac{1}{n}\sum_{k=1}^{n} K\left(\frac{\|X_k - x_0\|}{h}\right)((X_k - x_0) \otimes (X_k - x_0)). \tag{4}$$

Remark 3 In fact, neither Γ_K nor $\Gamma_{n,K}$ are truly covariance operators; since the involved random elements are not centered, they could also be named "local second order moment operators". Also note that Γ_K depends on h and h will depend on the sample size n. So, the reader must keep in mind that the index n was dropped in the notation Γ_K.

It is important to give some basic properties of these operators. Several assumptions are needed on the kernel K. They will be outlined in the next subsection as assumption $\mathbf{A}_1$. We list those which will be useful in the sequel.

- Γ_K and $\Gamma_{n,K}$ are self-adjoint and trace-class; hence compact whenever K has compact support.
- Both operators tend to zero when h does. Indeed

$$\|\Gamma_K\|_\infty \leq \mathbb{E}\left(K\left(\frac{\|X_1 - x_0\|}{h}\right)\|X_1 - x_0\|^2\right) \leq Ch^2$$

  as will be shown in the section devoted to mathematical derivations. The operator $\Gamma_{n,K}$ also tends to 0 as a consequence of the strong law of large numbers for sequences of independent Banach-valued random variables (see Ledoux and Talagrand (1991)).
- When Γ_K is one to one its inverse exists. Sufficient conditions on K and on X for Γ_K to be injective are not difficult to find, but this interesting issue is out of the scope of the present work. Then Γ_K^{-1} is an unbounded linear operator acting from a dense domain of H onto H. It should be stressed that Γ_K^{-1} is continuous at no point of its domain (it is nowhere continuous).

Imagine that the distribution of the data (namely of the couple (y, X)) is known. We could consider, instead of (3):

$$\min_{a\in\mathbb{R},\varphi\in H} \mathbb{E}\left[(y - a - \langle\varphi, X - x_0\rangle)^2 K\left(\frac{\|X - x_0\|}{h}\right)\right]. \tag{5}$$

The first stumbling block appears within the next proposition.

Proposition 4 *Even when the distribution of the data is known, the solution a^*_{th} of the "theoretical" program (5) exists only when Γ_K is one to one. Then a^*_{th} is the solution of a linear inverse problem which involves the unbounded inverse (whenever it exists) of Γ_K:*

$$a^*_{\mathrm{th}} = \frac{\mathbb{E}(yK) - \left\langle \Gamma_K^{-1}\mathbb{E}(yKZ), \mathbb{E}(KZ)\right\rangle}{\mathbb{E}(K) - \left\langle \Gamma_K^{-1}\mathbb{E}(KZ), \mathbb{E}(KZ)\right\rangle} \tag{6}$$

where, for the sake of shortness, we denoted

$$Z(x_0) = Z = X - x_0 \quad \text{and} \quad K = K(\|X - x_0\|/h).$$

The problem gets deeper when we go back to the original and empirical program (3). It turns out that the solution cannot be explicitly written since $\Gamma_{n,K}$ (which replaces now Γ_K) has no inverse because it has finite rank. Its rank is clearly bounded by n. In other words, the inverse $\Gamma_{n,K}^{-1}$ does not exist. A classical remedy consists in replacing $\Gamma_{n,K}^{-1}$ by a bounded operator $\Gamma_{n,K}^{\dagger}$ depending on n and such that $\Gamma_{n,K}^{\dagger}$ behaves pointwise like the inverse of $\Gamma_{n,K}$. This inverse operator, which is not always the Moore-Penrose pseudo inverse, will be called the *regularized inverse* of $\Gamma_{n,K}$. Several procedures could be carried out.

- Truncated spectral regularization: Here this method matches the usual Moore-Penrose pseudo inversion hence $\Gamma_{n,K}\Gamma_{n,K}^{\dagger}$ and $\Gamma_{n,K}^{\dagger}\Gamma_{n,K}$ are both projection operators on H. In fact, if the spectral decomposition of $\Gamma_{n,K}$ is $\Gamma_{n,K} = \sum_{i=1}^{m_n} \mu_{i,n} \left(u_{i,n} \otimes u_{i,n}\right)$ where for all i $\left(\mu_{i,n}, u_{i,n}\right)$ are the eigenvalues/eigenvectors of $\Gamma_{n,K}$ (we will always assume that the positive $\mu_{i,n}$'s are arranged in decreasing order):

$$\Gamma_{n,K}^{\dagger} = \sum_{i=1}^{N_n} \frac{1}{\mu_{i,n}} \left(u_{i,n} \otimes u_{i,n}\right), \tag{7}$$

  where $N_n \leq m_n$.
- Penalization: Now $\Gamma_{n,K}^{\dagger} = \left(\Gamma_{n,K} + \alpha_n S\right)^{-1}$ where α_n is a (positive) sequence tending to zero and S is a known operator chosen so that $\Gamma_{n,K} + \alpha_n S$ has a bounded inverse. Here S may be taken to be the identity operator.
- Tikhonov regularization: It comes down here, since $\Gamma_{n,K}$ is symmetric, to taking

$$\Gamma_{n,K}^{\dagger} = \left(\Gamma_{n,K}^{2} + \alpha_n I\right)^{-1} \Gamma_{n,K}.$$

  The sequence α_n is again positive and tends to zero.

Several other methods exist. The reader is referred, for instance, to Tikhonov and Arsenin (1977), Groetsch (1993) or Engl et al. (2000).

Remark 5 In all situations it should be noted that

$$\sup_n \left\| \Gamma_{n,K}^{\dagger} \Gamma_{n,K} \right\|_{\infty} < +\infty,$$
$$\sup_n \left\| \Gamma_{n,K} \Gamma_{n,K}^{\dagger} \right\|_{\infty} < +\infty.$$

All these regularizing methods may also be applied to Γ_K as well and lead to $\Gamma_K^{\dagger}$ and this operator depends on n even if this index does not explicitly appear. One may then prove that for all x in the domain of Γ_K^{-1}, $\Gamma_K^{\dagger} x \to \Gamma_K^{-1} x$ when n goes to infinity. In addition to the boundedness, the operator $\Gamma_{n,K}^{\dagger}$ is always self-adjoint and positive.

We are now in a position to propose an estimate for $m(x_0)$. This estimate will depend on the chosen regularization technique applied to $\Gamma_{n,K}$ which implies that the program (3) gives birth to several approximate solutions.

Proposition 6 *A local linear estimate of* $m(x_0)$ *is denoted* $\widehat{m}_n(x_0)$. *It is an approximate solution of* (3) *based on the regularized inverse* $\Gamma^{\dagger}_{n,K}$:

$$\widehat{m}_n\,(x_0) = \frac{\sum_{i=1}^{n} y_i \omega_{i,n}}{\sum_{i=1}^{n} \omega_{i,n}}, \tag{8}$$

where

$$\omega_{i,n} = K\left(\frac{\|X_i - x_0\|}{h}\right)\left(1 - \left\langle X_i - x_0, \Gamma^{\dagger}_{n,K}\overline{Z}_{K,n}\right\rangle\right)$$

and

$$\overline{Z}_{K,n} = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{\|X_i - x_0\|}{h}\right)(X_i - x_0)\,.$$

The proof of this proposition is omitted since it stems from calculations similar to those carried out in the proof of Proposition 4.

It is easy to check that (8) is the empirical counterpart of (6). We finally see that $\widehat{m}_n(x_0)$ may be viewed as a linear combination of the outputs $y_1, \ldots, y_n$ and may be expressed from a^*_{th} just by replacing expectations by sums. The reader may also compare our estimate with its one-dimensional counterpart (display 2.2, p. 198 in Fan (1993)) and will also notice that the nice properties of the $\omega_{i,n}$'s in this setting do not hold anymore (see display 2.5, p. 198 in Fan (1993) and the lines below).

The next section is devoted to developing the framework as well as the assumptions needed to get our central result.

2 Assumptions and framework

In all the sequel we assume

$\mathbf{A}_1$: *The kernel* K *is one-sided, supported on* $[0, 1]$, *bounded and* $K(1) > 0$. *Besides* K' *is non-null and belongs to* $L^1\left([0, 1]\right)$.

We did not try to find minimal conditions on the kernel. However, the assumption $K(1) > 0$ is rather rarely required in the non-parametric literature—to the authors' knowledge—and is essential here, as will be seen below.

2.1 The small ball problem and the class Gamma

Consider the one-dimensional version of our model (1) and take $X \in \mathbb{R}$ with density f. Fan (1993) studied the minimax properties of the local linear estimate in this setting and gave the Mean Square Error (see Theorem 2, p. 199). This MSE depends on $f(x_0)$. Here appears the second major problem. When the data belong to an infinite-dimensional space, their density does not exist, in the sense that Lebesgue's measure -or any universal reference measure with similar properties- does not exist. Consequently, we must expect serious troubles since it is plain that the density of the functional input X cannot be defined as easily as if it was real or even multivariate. Once again this problem will not be managed by just letting the dimension tend to infinity, and we should find a way to overcome this major concern.

It turns out that in many computations of expectations the problem mentioned above may be shifted to what is known in probability theory as small ball problems. Roughly speaking, if φ is a real-valued function (we set $x_0 = 0$ for simplicity), $\mathbb{E}\left(\varphi(\|X\|)\, K\left(\|X\|/h\right)\right)$ may be expressed essentially by means of $\mathbb{P}\left(\|X\| \leq h\right)$ and φ for small h. We refer to Lemma 29 in the proof section for an immediate illustration. Instead of knowing or estimating a density we must now focus on $\mathbb{P}\left(\|X\| \leq h\right)$ for small h, and everyone may understand why this function is often referred to as the "small ball probability associated with X". We propose such references as Li and Linde (1993), Kuelbs et al. (1994) and Li and Linde (1999) as well as the monograph by Li and Shao (2001) which provides an interesting state of the art in this area.

What can be said about the function $\mathbb{P}\left(\|X\| \leq h\right)$? Obviously, by Glivenko-Cantelli's theorem it will be easily estimated from the sample (the rate of convergence is non parametric). Besides, it is not hard to see that, under suitable but mild assumptions, if $X \in \mathbb{R}^p$ with density $f : \mathbb{R}^p \to \mathbb{R}^+$, $\mathbb{P}\left(\|X - x_0\| \leq h\right) \sim h^p f\left(x_0\right)$. But this fact leaves unsolved the question : what can be said when $p \to +\infty$?

In probability theory most of the small ball considerations focused on the case where X is the brownian motion, the brownian bridge, or some known relatives. Several norms were investigated as well. Most of the theorems collected in the literature yield

$$\mathbb{P}\left(\|X\| < h\right) \asymp C_1 h^{\alpha} \exp\left(-\frac{C_2}{h^{\beta}}\right) \tag{9}$$

where α, β, C_1 and C_2 are positive constants. The symbol $\asymp$ is sometimes replaced with the more precise $\sim$. Another serious problem comes from the fact that the C^{∞} function on the right in the display above has its derivatives null at zero at all orders. Other results assess that, when x_0 belongs to the Reproducing Kernel Hilbert Space of X,

$$\mathbb{P}\left(\|X - x_0\| < h\right) \sim C_{x_0} \mathbb{P}\left(\|X\| < h\right)$$

where C_{x_0} does not depend on h, but on x_0 and on the distribution of X. Two majors contributions will be found in Mayer-Wolf and Zeitouni (1993) and in Dembo et al. (1995). The authors give the exact asymptotic of $\mathbb{P}\left(\|X\|_{l^2} \leq h\right)$ when X is a l_2-valued gaussian random element (by means of large deviation theory):

$$X = \left(a_1 x_1, a_2 x_{2,\ldots}\right) \tag{10}$$

with x_i independent, $N\,(0, 1)$-distributed and $\sum a_i^2 < +\infty$. When $a_i = i^{-r}\,(r > 1/2)$ they obtain a formula similar to (9). Recently Mas (2008b) derived the estimate when

$a_i = \exp(-ci)$, $c > 0$ and got

$$\mathbb{P}\left(\|X\| < h\right) \sim C_1 \left[\log\left(1/h\right)\right]^{-1/2} \exp\left(-C_2 \left[\log\left(h\right)\right]^2\right). \tag{11}$$

A very strange fact is that both functions in (9) and (11) belong to a class of functions known in the theory of regular variations: the class Gamma introduced and studied by de Haan (1971, 1974). This class arises in the theory of extreme values and is closely related to the domain of attraction of the double exponential distribution. It was initially introduced by de Haan as a "Form of Regular Variation". We provide now the definition of the class Gamma at 0, denoted Γ_0.

Definition 7 A function f belongs to de Haan's class Γ_0 with auxiliary function ρ if f maps a positive neighborhood of 0 onto a positive neighborhood of 0, $f(0) = \rho(0) = 0$, f is non decreasing and for all $x \in \mathbb{R}$,

$$\lim_{s \downarrow 0} \frac{f\left(s + x\rho\left(s\right)\right)}{f\left(s\right)} = \exp(x). \tag{12}$$

In a recent manuscript, Mas (2008b) proved that, in the framework of Dembo et al. (1995), the small ball probability of any random element that may be defined like display (10) belongs to the class Gamma. A work is in progress to prove that, under suitable assumptions on the auxiliary function, the reciprocal also holds. The auxiliary functions appearing in displays (9) and (11) may be easily computed. It can be proved that ρ depends only on the sequence $a(\cdot)$ that defines X in (10).

The next proposition illustrates the above definition. It will be useful in the section devoted to the main results.

In all the sequel and especially within the proof section, C denotes a constant (which may vary from a theorem to another).

Proposition 8 *When the small ball probability is defined by the right-hand side of* (9), *the function ρ is*

$$\rho(s) = Cs^{1+\beta} \tag{13}$$

with $\beta > 0$, and when the small ball probability is defined by the right-hand side of (11), *the function ρ is*

$$\rho(s) = C\frac{s}{|\log s|}. \tag{14}$$

Starting from all these considerations it seems reasonable to assume the following:

$\mathbf{A}_2$: *Let*

$$F(h) = F_{x_0}(h) = P(\|X - x_0\| \leq h)$$

be the shifted small ball probability of X. We assume that $F \in \Gamma_0$ with auxiliary function ρ.

Gamma varying functions feature original properties. We give now one of them which will be useful later on in the proof section. We refer to Proposition 3.10.3 and Lemma 3.10.1 p. 175 in Bingham et al. (1987).

Proposition 9 *If* $F \in \Gamma_0$ *with auxiliary function* ρ *then for all* $x \in [0, 1]$,

$$\lim_{h \to 0^+} \frac{F(hx)}{F(h)} = 0 \tag{15}$$

$$\lim_{h \to 0} \frac{\rho(h)}{h} = 0 \tag{16}$$

Assumption $\mathbf{A}_2$ is central to tackle our problem since the mean square error, computed from our estimate actually depends on ρ. But additional assumptions should hold, especially on the distributions of the margins of X.

Remark 10 The representations (9) and (11) show that the small ball probability is extremely flat around 0. Computations of moments smoothed by the kernel K show that these flat distributions feature in a way the same behavior as a Dirac mass on the frontier of the support of K. Consequently, assuming that $K(1) > 0$ in assumption $\mathbf{A}_1$ enables to obtain exact rates of convergence for several functionals involved in the calculation of our estimate.

2.2 Assumptions on the marginal distributions

The next assumption essentially aims at simplifying the technique of proof but could certainly be alleviated at the expense of more tedious calculations (see also Mas (2008a) and comments therein).

$\mathbf{A}_3$: *There exists a basis* $(e_i)_{1 \le i \le n}$ *such that the margins* $(\langle X, e_i \rangle)_{1 \le i \le n}$ *are independent real random variables.*

In all the sequel, $f_i = f_{i,x_0}$ stands for the density of the real-valued random variable $\langle X - x_0, e_i \rangle$. The behavior around 0 of the shifted density f_i is crucial, like in the finite dimensional setting. It has to be smooth in a sense that is going to be made more clear now. Note that $f_i(0)$ is nothing but the density of the non-shifted random variable $\langle X, e_i \rangle$ evaluated at $\langle x_0, e_i \rangle$.

Let $\mathcal{V}_0$ be a fixed neighborhood of 0, set

$$\alpha_i = \sup_{u \in \mathcal{V}_0} \left| \frac{f_i(u) - f_i(-u)}{u(f_i(u) + f_i(-u))} \right|$$

and assume

$$\mathbf{A}_4 : \sum_{i=1}^{+\infty} \alpha_i^2 < +\infty.$$

The next proposition illustrates assumption $\mathbf{A}_4$ in the important case when X is gaussian.

Example 11 Let X be a centered gaussian random element in H with Karhunen-Loève expansion

$$X = \sum_{k=1}^{+\infty} \sqrt{\lambda_k} \eta_k e_k.$$

Here, the λ_k's are the eigenvalues of the covariance operator of $X, \mathbb{E}(X \otimes X)$, the e_k's are the associated eigenvectors and the η_k's are real-valued random variables $N(0, 1)$-distributed. It is a well-known fact that $\langle X, e_k \rangle = \sqrt{\lambda_k} \eta_k$ are independent real gaussian random variables and $\mathbf{A}_3$ holds. Then $f_i(u) = \frac{1}{\sqrt{2\pi \lambda_i}} \exp\left[-\frac{(u - \langle x_0, e_i \rangle)^2}{2\lambda_i} \right]$ and

$$\sup_{u \in \mathcal{V}_0} \frac{|f_i(u) - f_i(-u)|}{u\,|f_i(u) + f_i(-u)|} \leq C \frac{\langle x_0, e_i \rangle}{\lambda_i}$$

whenever $\langle x_0, e_i \rangle / \lambda_i \to 0$ when i tends to infinity and $\mathbf{A}_4$ holds if

$$\sum_{i=1}^{+\infty} \frac{\langle x_0, e_i \rangle^2}{\lambda_i^2} < +\infty. \tag{17}$$

Example 12 We can also consider the family of densities indexed by the integer m:

$$f_i(u) = \frac{C_m}{\sqrt{\lambda_i}} \frac{1}{1 + \left(\frac{u - \langle x_0, e_i \rangle}{\sqrt{\lambda_i}} \right)^{2m}}$$

where C_m is a normalizing constant. We find

$$\alpha_i \leq C \frac{|\langle x_0, e_i \rangle|}{\lambda_i^m}$$

and assumption $\mathbf{A}_4$ holds whenever the sequence $\left(\frac{|\langle x_0, e_i \rangle|}{\lambda_i^m} \right)_{i \in \mathbb{N}} \in l_2$.

Since the rate of decrease of the λ_i's is intimately related to the smoothness of the random function X, we may easily infer that $\mathbf{A}_4$ should be interpreted as a smoothness condition on the function x_0. In other words, the coordinates of x_0 in the basis e_i should tend to zero at a rate which is significantly quicker than the eigenvalues of the covariance operator of X and hence that x_0 should be sufficiently smoother than X.

It should also be noted that, when the family of densities f_i is not uniformly smooth enough in a neighborhood of 0, Assumption $\mathbf{A}_4$ may fail. For instance, it is not hard to see that the α_i's are not even finite when f_i is the density of a shifted Laplace random variable:

$$f_i(u) = \frac{1}{2\lambda_i} \exp\left(-\frac{|u - \langle x_0, e_i \rangle|}{\lambda_i} \right).$$

Remark 13 The issue of the expectation of the functional input X should be raised now. We assumed sooner that the X_i's are centered. But in practical situations we can expect $\mu = \mathbb{E}(X)$ to be a non-null function. Then considering a new shift $x_0 - \mu$ instead of x_0 solves the problem. So we can always consider the centered version of X, but we must take into account that any assumption made on x_0 should be valid for $x_0 - \mu$. For instance, (17) should be replaced by

$$\sum_{i=1}^{+\infty} \frac{\langle x_0 - \mu, e_i \rangle^2}{\lambda_i^2} < +\infty.$$

Assumptions $\mathbf{A}_2$–$\mathbf{A}_4$ model the distribution of the functional data X. It is a difficult issue to compare them with their counterparts in a multivariate setting. Obviously, assumption $\mathbf{A}_2$ is typical with functional data except when considering a degenerate design (see Gaiffas (2005)). Assumption $\mathbf{A}_3$ always holds for gaussian X as was seen above and could certainly be alleviated yielding longer calculations within the proofs. A work is on progress about this issue. At last, assumption $\mathbf{A}_4$ could be compared with Condition 1 (ii) in Fan (1993) which aims at controlling the local smoothness of the density around the point x_0. Simple calculations show that $\mathbf{A}_4$ is slightly stronger than the Hölder assumption in the latter article.

2.3 Smoothness of the regression function

In order to achieve our estimating procedure we cannot avoid to assume that the function m is regular. Since m is a mapping from H to $\mathbb{R}$, its first-order derivative is an element of $\mathcal{L}(H, \mathbb{R})$, the space of bounded linear functionals from H to $\mathbb{R}$ which is nothing but $H^* \simeq H$. As announced sooner $m'(x_0) \in H$. The second-order derivative belongs to $\mathcal{L}(\mathcal{L}(H, \mathbb{R}), \mathbb{R}) \simeq \mathcal{L}(H \times H, \mathbb{R})$ and is consequently a quadratic functional on $H \times H$ and may be represented by a symmetric positive linear operator from H to H (the Hessian operator). We will sometimes use abusive notations such as $\langle m''(x_0)(u), v \rangle$ below and throughout the proofs.

$\mathbf{A}_5$: *The first order derivative* $m'(x_0)$ *of* m at x_0 *is defined, non null and there exists a neighborhood* $\mathcal{V}(x_0)$ of x_0 *such that:*

$$\sup_{x \in \mathcal{V}(x_0)} \left\| m''(x) \right\|_\infty < +\infty.$$

This last display may be rewritten: for all u in H and all x in a neighborhood of x_0

$$\langle m''(x)u, u \rangle \leq C \|u\|^2 .$$

Remark 14 Assumption $\mathbf{A}_5$ assesses in some way that the second-order derivative of m in a neighborhood of x_0 is bounded. It echoes exactly Condition 1(i) in Fan (1993).

2.4 Back to the regularized inverse

We need, for immediate purpose, to define a sequence involved in the rate of convergence of our estimate.

Definition 15 Let $v(h)$ be the positive sequence defined by

$$v = v(h) = \left[\mathbb{E}\left(K\left(\frac{\|X - x_0\|}{h}\right)\|X - x_0\|\,\rho\left(\|X - x_0\|\right)\right)\right]. \tag{18}$$

It is plain that v tends to zero when h does.

Since they will be used in the sequel we list now some results from Mas (2008a). They are collected in the next proposition and consist in bounding the norms of operators Γ_K and $\Gamma_{n,K}$.

Proposition 16 *The following bounds are valid*

$$\|\Gamma_K\|_\infty \geq Cv(h), \tag{19}$$

$$\left\|\Gamma_{n,K} - \Gamma_K\right\|_\infty = O_{L^2}\left(h^2\sqrt{\frac{F(h)}{n}}\right). \tag{20}$$

Besides $\Gamma_K/v(h)$ *may converge to a bounded operator, say S, that may be compact.*

Before giving the main results, we have to get back to the regularized inverse of $\Gamma_{n,K}$. Indeed, a bound on the norm of $\Gamma^\dagger_{n,K}$ may be derived. Under the assumption that $h^2F^{1/2}(h)/\left(n^{1/2}v(h)\right) \to 0$ we see that

$$\left\|\Gamma_{n,K}\right\|_\infty \geq Cv(h).$$

As a consequence of these facts, we expect the norm $\Gamma^\dagger_{n,K}$ to diverge with rate *at least* $1/v(h)$ since

$$0 < C < \left\|\Gamma_{n,K}\Gamma^\dagger_{n,K}\right\|_\infty \leq \left\|\Gamma^\dagger_{n,K}\right\|_\infty \left\|\Gamma_{n,K}\right\|_\infty.$$

If the operator S mentioned in the above proposition is compact, we may even be aware that the norm of $v(h)\Gamma^\dagger_{n,K}$ will tend to infinity since S^{-1} is unbounded whenever S^{-1} exists. All this leads us to consider the next and last assumption on $\Gamma^\dagger_{n,K}$:

$\mathbf{A}_6$: *There exists a sequence* $r_n > 0$ *such that* $r_n v(h) \to 0$ as $n \to +\infty$ and

$$\iota_n = \max\left\{\left\|\Gamma^\dagger_K\right\|_\infty, \left\|\Gamma^\dagger_{n,K}\right\|_\infty\right\} \leq \frac{1}{r_n v(h)} \quad a.s. \tag{21}$$

Here, the parameter r_n just depends on the chosen regularizing method (penalization, Tikhonov, etc.) and may be viewed as a tuning parameter.

Remark 17 In fact as will be seen below the sequence

$$v(h)\max\left\{\left\|\Gamma_K^{\dagger}\right\|_{\infty},\left\|\Gamma_{n,K}^{\dagger}\right\|_{\infty}\right\}$$

determines the degree of ill-posedness of the problem. It may be bounded, but the most unfavorable situation appears when it tends to infinity (and consequently r_n tends to 0). We intend to investigate it with care. However, ι_n always tends to infinity and cannot be bounded above because of (19) and (20). Besides, if $\Gamma_K/v(h)$ converges to an operator with bounded inverse, the sequence r_n can always be chosen constant or at least bounded below.

Let us take some examples to illustrate the role of r_n. We keep the notations of display (7) and of the lines below and for the sake of clarity we assume that

$$\max\left\{\left\|\Gamma_K^{\dagger}\right\|_{\infty},\left\|\Gamma_{n,K}^{\dagger}\right\|_{\infty}\right\}=\left\|\Gamma_{n,K}^{\dagger}\right\|_{\infty}.$$

- Truncated spectral regularization: remind that

$$\Gamma_{n,K}^{\dagger}=\sum_{i=1}^{N_n}\frac{1}{\mu_{i,n}}\left(u_{i,n}\otimes u_{i,n}\right)$$

where $\left(\mu_{i,n},u_{i,n}\right)$ are the eigenelements of $\Gamma_{n,K}$ and

$$\left\|\Gamma_{n,K}\right\|_{\infty}=\sup_i\left\{\mu_{i,n}\right\}=\mu_{1,n}$$

(as announced earlier, the eigenvalues are positive and arranged in decreasing order). Hence

$$\left\|\Gamma_{n,K}^{\dagger}\right\|_{\infty}=1\Big/\inf_{1\le i\le N_n}\left\{\mu_{i,n}\right\}=\mu_{N_n,n}^{-1}$$

then $r_n=\mu_{N_n,n}/\mu_{1,n}\downarrow 0$ is the inverse of the conditioning index of operator $\Gamma_{n,K}^{\dagger}$.
- Penalization: Now $\Gamma_{n,K}^{\dagger}=\left(\Gamma_{n,K}+\alpha_n I\right)^{-1}$ with

$$\Gamma_{n,K}^{\dagger}=\sum_{i=1}^{m_n}\frac{1}{\mu_{i,n}+\alpha_n}\left(u_{i,n}\otimes u_{i,n}\right)$$

and we can take $r_n=\alpha_n/\mu_{1,n}$. It is possible here to get $r_n\uparrow+\infty$ by an accurate choice of α_n and some information on $\mu_{1,n}$.
- Tikhonov regularization: Here $\Gamma_{n,K}^{\dagger}=\left(\Gamma_{n,K}^2+\alpha_n I\right)^{-1}\Gamma_{n,K}$ and

$$\Gamma_{n,K}^{\dagger}=\sum_{i=1}^{m_n}\frac{\mu_{i,n}}{\mu_{i,n}^2+\alpha_n}\left(u_{i,n}\otimes u_{i,n}\right).$$

A choice for r_n is here $\alpha_n/\mu_{1,n}^2$ and the same remark as above holds.

3 Statement of the results

The central result of this article is a bound on the Mean Square Error for the local linear estimate of the pointwise evaluation of the regression function at a fixed design. In the sequel the generic notation C stands for universal constants.

Theorem 18 *Fix x_0 in H. When assumptions $\mathbf{A}_1$–$\mathbf{A}_6$ hold and if $nF(h) \to +\infty$*

$$\mathbb{E}\left(\widehat{m}_n(x_0) - m(x_0)\right)^2 \leq C\left[\frac{h^6}{r_n^2} + h^4 + \frac{h^2}{nF(h)} + \frac{v^2(h)}{F^2(h)}\right] + \frac{C}{nF(h)}\left(1 + \frac{h^2}{nr_n v(h)} + \frac{v(h)}{r_n F(h)}\right).$$

where the first line arises from the bias of our estimate and the second stems from its variance.

Remark 19 If K is chosen to be the naive kernel, $K(s) = \mathbf{1}_{[0,1]}(s)$, assumption $\mathbf{A}_1$ can be removed and the previous theorem remains valid.

Remark 20 It turns out that the variance term is decomposed into three. The first is $(nF(h))^{-1}$ and is classical (see Ferraty et al. (2007)). The two others stem directly from the underlying inverse problem and the sequence r_n appears.

Note that we did not fix the issue of the sequence r_n involved in the regularizing inverses $\Gamma_K^{\dagger}$ and $\Gamma_{n,K}^{\dagger}$. Theorem 18 may be simplified under mild additional assumptions.

Proposition 21 *Taking $r_n \asymp h$, then*

$$\mathbb{E}\left(\widehat{m}_n(x_0) - m(x_0)\right)^2 \leq Ch^4 + \frac{C}{nF(h)}\left(1 + \frac{h}{nv(h)}\right)$$

This proposition is derived from Theorem 18 and Lemma 29.

Turning back to Proposition 8 and considering displays (13) and (14) it is not hard to see that both functions ρ are regularly varying at 0 with index $1+\beta$ for the first and 1 for the second. It should also be noted that from property (16) in Proposition 9 that we can expect ρ to be of index larger than 1 whenever it is regularly varying at 0. This fact motivates the next proposition.

Proposition 22 *Under the assumptions of* Theorem 18 *and of* Proposition 21, *if the auxiliary function ρ is regularly varying at* 0 *with index $g \geq 1$,*

$$v(h) \asymp h\rho(h)F(h). \tag{22}$$

Then, if $\rho(s) \geq Cs^4$ *in a neighborhood of* 0, *the mean square error becomes*

$$\mathbb{E}\left(\widehat{m}_n(x_0) - m(x_0)\right)^2 \leq C\left(h^4 + \frac{1}{nF(h)}\right)$$

and the rate of decrease of the Mean Square Error depends on h^* *given by*

$$\left(h^*\right)^4 F\left(h^*\right) = \frac{1}{n}. \tag{23}$$

If $\rho(s)/s^4 \to 0$ *when* $s \to 0$ *the above rate is damaged. For instance, taking* $r_n \asymp h$ *the MSE becomes*

$$\mathbb{E}\left(\widehat{m}_n(x_0) - m(x_0)\right)^2 \leq C\left(h^4 + \frac{1}{n^2F^2(h)\rho(h)}\right).$$

Remark 23 As announced earlier, considerations about optimality of our estimate are beyond the scope of this work. However, it is worth commenting on display (23). Indeed, we see that when $X \in \mathbb{R}^d$, $F(h) \sim Ch^d$ then the rate of convergence in mean square turns out to be $n^{-2/(4+d)}$ which is the optimal (minimax) rate of convergence for a twice-differentiable regression function (see Stone (1982)). This does not prove that (23) defines the optimal rate in the functional framework, but it underlines that our approach remains valid in a multivariate setting. In fact, when the small ball probability belongs to the class Γ_0, this rate depends on ρ. We know that the term $F(h)$ will always tend to 0 quicker than h^4 and will consequently determine the choice of h. The situation is consequently more intricate than that in the multivariate setting. However, following the example of displays (9) and (11) we get, respectively,

$$h_n^* = C(\log n)^{-1/\beta}$$
$$h_n^* = C(\log n)^{-1/2}$$

where $\beta \leq 3$ when $\rho(s) \geq Cs^4$. Finally, the rate of decrease of the mean square error is $O\left((\log n)^{-c}\right)$ where $c > 1$.

Remark 24 Display (22) was proved in Mas (2008a). In the first case (when $\rho(s) \geq Cs^4$), since the bias term is here $O(h^4)$, the rate of convergence of our estimate outperforms the one computed in Ferraty et al. (2007). The estimate was a classical Nadaraya-Watson kernel estimator whose bias was $O(h)$. Obviously the rate of convergence in the second case is damaged, but even for very irregular processes such as Brownian motion or Brownian Bridge function $\rho(s)$ is above s^2 or s^3 depending on the norms that are used. The interested reader is referred, for instance, to displays (20) and (22) in Mayer-Wolf and Zeitouni (1993) or Proposition 6.1, p. 568 in Li, Shao (2002) but will have to carry out some additional computations. It seems reasonable to think that this unfavorable situation will rarely occur in usual statistical context (with functions reconstructed on smooth spaces). We prove below that, even when ρ decays rapidly to 0, it is always possible to choose a regularizing method for $\Gamma_{n,K}$ that reaches the best rate of display (23).

The last proposition deals with the situation described in Remark 17: when r_n does not tend to zero. This cannot happen when the problem is ill-posed and the regularizing method is the spectral truncation but may occur when either a penalization or a Tikhonov method is applied. We remind that we cannot avoid the condition $r_n v(h) \downarrow 0$. We start from Theorem 18.

Proposition 25 *When assumptions* $\mathbf{A}_1$–$\mathbf{A}_6$ *hold, if* $nF(h) \to +\infty$, *when the regularizing method allows to do so, taking* $r(h) = 1/\rho(h)$ *provides*

$$\mathbb{E}\left(\widehat{m}_n(x_0) - m(x_0)\right)^2 \leq Ch^4 + C\frac{1}{nF(h)}.$$

Obviously $r_n v(h)$ tends to 0. If the chosen method is penalization such that $\Gamma_{n,K}^{\dagger} = \left(\Gamma_{n,K} + \alpha_n S\right)^{-1}$ it suffices to take $\alpha_n = h^* F(h^*)$ to achieve our goal. The proof of this proposition is easy and hence omitted.

Remark 26 The rate in Display (23) should be compared with the minimax rate obtained by Fan (1993) for local linear regression with scalar inputs. The MSE was then $Ch^4 + C/(nh)$. We see that, replacing $F(h)$ with h (which is logical if we consider the remark about the multivariate case just below display (9) in the section devoted to the small ball problems), both formulae match. This fact leads us to another interesting issue: does this rate inherit the optimal (minimax) properties found by Fan in his article? This question goes beyond the scope of this article. Besides, not much has been done until now about optimal estimation for functional data—to the authors' knowledge. But there is no doubt that this issue will be addressed in the near future.

4 Conclusion

Obviously, this article could be the starting point for other issues such as almost sure or weak convergence of the estimate. Almost all practical aspects were left out on purpose. They will certainly give birth to another article. However, the main goal of this essentially theoretical work was to underline the rather large scope of our study. We had to seek several ideas from various areas such as probability theory, functional analysis, statistical theory of extremes, and inverse problems theory. Finally, it turns out that it is possible to get, in the functional setting, the same rate of decay for the bias as in the case of scalar inputs. The variance involves the small ball probability evaluated at h, the selected bandwidth. A drawback arises with the necessity of introducing a new parameter: the regularizing sequence r_n, which depends on the sample size (more precisely on the bandwidth h). We give no clue to find out in practical situations the bandwidth h, but we guess that the ever wider literature on functional data will quickly overcome this problem by adapting classical methods such as cross-validation, for instance.

Another major practical concern relies in the estimation of the unknown auxiliary function ρ. Several tracks already appear to address this issue. One may think of adapting some techniques from extreme values theory. After all, ρ characterizes the extreme behavior of $\|X\|$ like tail indices for Weibull or Pareto distributions.

The only difference stems from the fact that ρ is a function and not just a real number. The other idea 0lies in the article by Mas (2008b) where the auxiliary function ρ is explicitly linked with the eigenvalues of the ordinary covariance operator of X. From the estimation of these eigenvalues (which is a basic procedure), it should be possible to propose a consistent estimation of the auxiliary function as a by-product.

5 Proofs

For the sake of clarity we begin with an outline of the proofs. The following bias-variance decomposition for $\widehat{m}_n(x_0) - m(x_0)$ holds:

$$\begin{aligned}
\widehat{m}_n(x_0) - m(x_0) &= \frac{\sum_{i=1}^n y_i \omega_{i,n}}{\sum_{i=1}^n \omega_{i,n}} - m(x_0) \\
&= \frac{\sum_{i=1}^n (y_i - m(x_0)) \omega_{i,n}}{\sum_{i=1}^n \omega_{i,n}} \\
&= \frac{\sum_{i=1}^n (y_i - m(X_i)) \omega_{i,n}}{\sum_{i=1}^n \omega_{i,n}} + \frac{\sum_{i=1}^n (m(X_i) - m(x_0)) \omega_{i,n}}{\sum_{i=1}^n \omega_{i,n}}.
\end{aligned}$$

We denote

$$T_{b,n} = \frac{\sum_{i=1}^n (m(X_i) - m(x_0)) \omega_{i,n}}{\sum_{i=1}^n \omega_{i,n}}, \tag{24}$$

$$\begin{aligned}
T_{v,n} &= \frac{\sum_{i=1}^n (y_i - m(X_i)) \omega_{i,n}}{\sum_{i=1}^n \omega_{i,n}} \\
&= \frac{\sum_{i=1}^n \omega_{i,n} \varepsilon_i}{\sum_{i=1}^n \omega_{i,n}}
\end{aligned} \tag{25}$$

where ε was defined at display (1). Here $T_{b,n}$ is a bias term and $T_{v,n}$ is a variance term. Finally we get

$$\mathbb{E}[\widehat{m}_n(x_0) - m(x_0)]^2 = \mathbb{E}T_{b,n}^2 + \mathbb{E}T_{v,n}^2 + 2\mathbb{E}\left(T_{b,n}T_{v,n}\right) \tag{26}$$

and since

$$\begin{aligned}
\mathbb{E}\left(T_{b,n}T_{v,n}\right) &= \mathbb{E}\left(T_{b,n}\mathbb{E}\left(T_{v,n}|X_1, \ldots, X_n\right)\right) \\
&= 0
\end{aligned}$$

computing the mean square error of $\widehat{m}_n(x_0)$ comes down to computing $\mathbb{E}T_{b,n}^2$ and $\mathbb{E}T_{v,n}^2$, which will be done later.

The proof section is divided into two subsections. The first one is devoted to preliminary results as well as lemmas. In the second one the main results are derived.

5.1 Preliminary results

We assume that assumptions $\mathbf{A}_1-\mathbf{A}_6$ hold once and for all. All the proofs of this subsection were omitted in order to shorten the article. They are, however, available online on the arXiv website with reference **arXiv:0710.5218v1**.

Lemma 27 *If f belongs to the class Γ_0 with auxiliary function ρ, then for all $p \in \mathbb{N}$,*

$$\int_0^1 \frac{t^p}{\sqrt{1-t^2}} f\left(s\sqrt{1-t^2}\right) dt \underset{s\to 0}{\sim} 2^{\frac{p-1}{2}} \Gamma\left(\frac{p+1}{2}\right) f(s) \left(\frac{\rho(s)}{s}\right)^{\frac{p+1}{2}}.$$

For any $x = \sum x_k e_k$ in H and for $i \in \mathbb{N}$ set $\|x\|_{\neq i}^2 = \sum_{k\neq i} x_k^2$.

We denote $f_{\neq i}$ the density of $\|X - x_0\|_{\neq i}$. We need to compute both densities $f_{\|X-x_0\|}$ (density of $\|X - x_0\|$) and $f_{\langle X-x_0,e_i\rangle,\|X-x_0\|}$ (density of the couple $(\langle X - x_0, e_i\rangle, \|X - x_0\|)$).

Lemma 28 *We have*

$$f_{\langle X-x_0,e_i\rangle,\|X-x_0\|}(u, v) = \frac{v}{\sqrt{v^2-u^2}} f_i(u) f_{\neq i}\left(\sqrt{v^2-u^2}\right) 1\!\!1_{\{v\geq|u|\}}, \tag{27}$$

$$f_{\|X-x_0\|}(v) = v \int_{-1}^1 \frac{f_i(vt)}{\sqrt{1-t^2}} f_{\neq i}\left(v\sqrt{1-t^2}\right) dt. \tag{28}$$

Besides, if $f_{\|X-x_0\|}$ and $f_{\neq i}$ are Γ-varying for all i then they have all ρ as auxiliary function.

We begin with more specific computational lemmas.

Lemma 29 *Let φ be a positive real-valued function, bounded on* $[0, 1]$ *and regularly varying at* 0 *with index $g \geq 1$ and let $p \in \mathbb{N}$:*

$$\mathbb{E}K^p\left(\left\|\frac{X-x_0}{h}\right\|\right)\varphi(\|X-x_0\|) \underset{h\to 0}{\sim} K^p(1)\varphi(h) F(h). \tag{29}$$

As important special cases we mention

$$\mathbb{E}K\left(\left\|\frac{X-x_0}{h}\right\|\right) \sim K(1)F(h), \quad \mathbb{E}K^2\left(\left\|\frac{X-x_0}{h}\right\|\right) \sim K^2(1)F(h),$$
$$\mathbb{E}\left[\|X-x_0\|^m K\left(\left\|\frac{X-x_0}{h}\right\|\right)\right] \sim K(1)F(h)h^m.$$

$$\overline{Z}_{K,n} = \frac{1}{n}\sum_{i=1}^n Z_i K_i = \frac{1}{n}\sum_{k=1}^n (X_i - x_0) K(\|X_i - x_0\|/h).$$

The next lemma is crucial.

Lemma 30 *We have*

$$
\begin{aligned}
\|\mathbb{E}\,[ZK]\|^2 &= \left\| \mathbb{E}\left[K\left(\left\| \frac{X-x_0}{h} \right\| \right) (X-x_0) \right] \right\|^2 \\
&\leq C v^2\,(h).
\end{aligned}
$$

Remark 31 We can evaluate the sharpness of the previous bound. Indeed, a very simple inequality would give by Lemma 29:

$$
\|\mathbb{E}\,[ZK]\|^2 \leq (\mathbb{E}\,\|ZK\|)^2 = (\mathbb{E}\,\|Z\|\,K)^2 \sim Ch^2 F^2\,(h)
$$

whereas in view of (18) and—when ρ is regularly varying at 0 with positive index—of Lemma 29,

$$
\left[\mathbb{E}\left(K\left(\frac{\|X-x_0\|}{h} \right) \|X-x_0\|\,\rho\,(\|X-x_0\|) \right) \right]^2 \leq h^2 \rho^2\,(h)\,F^2\,(h).
$$

So the bound was improved by a rate of $\rho^2(h) = o(h^2)$.

Lemma 32 *Both following bounds hold:*

$$
\begin{aligned}
\mathbb{E}\left\| \overline{Z}_{K,n} - \mathbb{E}\overline{Z}_{K,n} \right\|^2 &\leq C\frac{h^2 F\,(h)}{n}, \\
\mathbb{E}\left\| \overline{Z}_{K,n} - \mathbb{E}\overline{Z}_{K,n} \right\|^4 &\leq C\frac{h^4 F^2\,(h)}{n^2}.
\end{aligned}
$$

Lemma 33 *We have*

$$
\mathbb{E}\omega_{1,n}^2 \leq C\left(F\,(h) + \frac{h^2 F\,(h)}{n r_n v\,(h)} + \frac{v\,(h)}{r_n} \right).
$$

Lemma 34 *When $nF\,(h) \to +\infty$,*

$$
\frac{\sum_{i=1}^n K_i}{nK\,(1)\,F\,(h)} - 1 \xrightarrow{L^2} 0.
$$

where $\xrightarrow{L^2}$ denotes convergence in mean square.

Lemma 35 *We have*

$$
\mathbb{E}\left\langle \Gamma_{n,K}^{\dagger} \overline{Z}_{K,n}, \overline{Z}_{K,n} \right\rangle^2 \leq C\frac{h^4 F^2\,(h)}{n^2 r_n^2 v^2\,(h)} + \frac{v^2\,(h)}{r_n^2}.
$$

5.2 Derivation of the main results

We start with a short and simple intermezzo about optimization in Hilbert spaces.

Proof of Proposition 4 Consider the following program:

$$\min_{a\in\mathbb{R},\varphi\in H} \mathbb{E}\left[(y-a-\langle\varphi, X-x_0\rangle)^2 K\left(\frac{\|X-x_0\|}{h}\right)\right].$$

Simple computations lead to

$$\begin{aligned}\mathcal{E}(a,\varphi) &= \mathbb{E}\left[(y-a-\langle\varphi, X-x_0\rangle)^2 K\left(\frac{\|X-x_0\|}{h}\right)\right]\\ &= C+a^2\mathbb{E}K + \langle\Gamma_K\varphi,\varphi\rangle - 2a\mathbb{E}(yK) - 2\langle\mathbb{E}(yZK),\varphi\rangle + 2a\langle\mathbb{E}(ZK),\varphi\rangle.\end{aligned}$$

Obviously, $\mathcal{E}(a,\varphi)$ is positive, strictly convex, and

$$\lim_{a,\|\varphi\|\to+\infty} \mathcal{E}(a,\varphi) = +\infty$$

hence $\mathcal{E}(a,\varphi)$ has a single minimum (see Rockafellar (1996) for further information about the minimization of convex functions). It is also differentiable for all (a,φ) in $\mathbb{R}\times H$. We compute its gradient :

$$\nabla\mathcal{E}(a,\varphi) = \begin{pmatrix} 2a\mathbb{E}K - 2\mathbb{E}(yK) + 2\langle\mathbb{E}(ZK),\varphi\rangle \\ 2\Gamma_K\varphi - 2\mathbb{E}(yZK) + 2a\mathbb{E}(ZK)\end{pmatrix}$$

from which we get the solutions (a^*,φ^*):

$$\begin{pmatrix} a^*\mathbb{E}K + \langle\mathbb{E}(ZK),\varphi^*\rangle = \mathbb{E}(yK) \\ \Gamma_K\varphi^* = \mathbb{E}(yZK) - a^*\mathbb{E}(ZK)\end{pmatrix}.$$

We see from the second line that φ^* is not uniquely defined if Γ_K is not one to one. Taking $\varphi^* = \Gamma_K^{-1}(\mathbb{E}(yZK) - a^*\mathbb{E}(ZK))$, we get $\widehat{m}_n(x_0)$ as announced. □

The forthcoming lemma assesses that the random denominator of our estimate may be replaced with a non-random one.

Lemma 36 *When both* $\frac{h^4}{nr_n^2 v^2(h)}$ *and* $\frac{v(h)}{r_n F(h)}$ *tend to zero, the following holds*:

$$\frac{\sum_{i=1}^n \omega_{i,n}}{nK(1)F(h)} - 1 \xrightarrow{L^2} 0.$$

Proof

$$\sum_{i=1}^{n} \omega_{i,n} = \sum_{i=1}^{n} K_i - n\left\langle \Gamma_{n,K}^{\dagger} \overline{Z}_{K,n}, \overline{Z}_{K,n} \right\rangle$$

hence

$$\frac{\sum_{i=1}^{n} \omega_{i,n}}{nK(1)\,F(h)} - 1 = \frac{\sum_{i=1}^{n} K_i}{nK(1)\,F(h)} - 1 - \frac{\left\langle \Gamma_{n,K}^{\dagger} \overline{Z}_{K,n}, \overline{Z}_{K,n} \right\rangle}{K(1)\,F(h)}.$$

From Lemmas 34 and 35 we deduce that the announced Lemma 36 holds. □

5.2.1 Variance term

We study first (see 25): $T_{v,n} = \frac{\sum_{i=1}^{n}(y_i - m(X_i))\omega_{i,n}}{\sum_{i=1}^{n}\omega_{i,n}} = \frac{\sum_{i=1}^{n}\varepsilon_i\omega_{i,n}}{\sum_{i=1}^{n}\omega_{i,n}}$. It is plain that $\mathbb{E}T_{v,n} = 0$. Denote $\widetilde{T}_{v,n} = \frac{\sum_{i=1}^{n}\varepsilon_i\omega_{i,n}}{nK(1)F(h)}$. We have:

$$T_{v,n} - \widetilde{T}_{v,n} = T_{v,n}\left(\frac{F(h) - \frac{1}{nK(1)}\sum_{i=1}^{n}\omega_{i,n}}{F(h)}\right).$$

We begin with a proposition. By Lemma 36 just above we know that $T_{v,n} \sim \widetilde{T}_{v,n}$ in L^2 sense, i.e.,

$$\frac{\widetilde{T}_{v,n}}{T_{v,n}} \xrightarrow{L^2} 1.$$

Proposition 37 *We have*

$$\mathbb{E}T_{v,n}^2 \le C\frac{1}{nF^2(h)}\left(F(h) + \frac{h^2F(h)}{nr_n v(h)} + \frac{v(h)}{r_n}\right).$$

Proof As announced earlier, it suffices to prove the proposition for $\widetilde{T}_{v,n}$.

$$\begin{aligned}
\mathbb{E}\widetilde{T}_{v,n}^2 &= \mathbb{E}\left(\frac{\sum_{i=1}^{n}\varepsilon_i\omega_{i,n}}{nK(1)\,F(h)}\right)^2 \\
&= \frac{1}{n^2K^2(1)\,F^2(h)}\mathbb{E}\left\{\mathbb{E}\left[\left(\sum_{i=1}^{n}\varepsilon_i\omega_{i,n}\right)^2 \middle| X_1, \ldots, X_n\right]\right\} \\
&= \frac{1}{n^2K^2(1)\,F^2(h)}\mathbb{E}\left[\mathbb{E}\left(\sum_{i=1}^{n}\varepsilon_i^2\omega_{i,n}^2 \middle| X_1, \ldots, X_n\right)\right]
\end{aligned}$$

since for $i \neq j$,

$$\mathbb{E}\left[\left(\varepsilon_i \omega_{i,n} \varepsilon_j \omega_{j,n}\right) | X_1, \ldots, X_n\right] = \omega_{i,n} \omega_{j,n} \mathbb{E}\left[\left(\varepsilon_i \varepsilon_j\right) | X_1, \ldots, X_n\right] = 0.$$

Hence,

$$\mathbb{E}\widetilde{T}_{v,n}^2 = \frac{1}{n^2 K^2(1) F^2(h)} \sigma_\varepsilon^2 \mathbb{E}\left(\sum_{i=1}^n \omega_{i,n}^2\right) = \frac{\sigma_\varepsilon^2 \mathbb{E}\left(\omega_{1,n}^2\right)}{n K^2(1) F^2(h)}.$$

By Lemma 33,

$$\mathbb{E}\left(\omega_{1,n}^2\right) \leq C\left(F(h) + \frac{h^2 F(h)}{n r_n v(h)} + \frac{v(h)}{r_n}\right)$$

from which we deduce the proposition. □

Now we turn to the bias term.

5.2.2 *Bias term*

Remember that we have to deal with

$$T_{b,n} = \frac{\sum_{i=1}^n \left(m\left(X_i\right) - m\left(x_0\right)\right) \omega_{i,n}}{\sum_{i=1}^n \omega_{i,n}}.$$

Copying what was done above with $T_{v,n}$, we know that we can focus on

$$\widetilde{T}_{b,n} = \frac{\sum_{i=1}^n \left(m\left(X_i\right) - m\left(x_0\right)\right) \omega_{i,n}}{n K(1) F(h)}$$

via Lemma 36. For each i there exists $c_i \in B\left(x_0, h\right)$ such that

$$\begin{aligned} &m(X_i) - m(x_0) \\ &\quad = \left\langle m'(x_0), Z_i \right\rangle + \frac{1}{2}\left\langle m''\left(c_i\right)\left(Z_i\right), Z_i \right\rangle. \end{aligned}$$

with $Z_i = X_i - x_0$. We deal with the first- and second-order derivatives separately: $\widetilde{T}_{b,n} = \widetilde{T}_{b,n,1} + \widetilde{T}_{b,n,2}$ with

$$\widetilde{T}_{b,n,1} = \frac{\sum_{i=1}^n \left\langle m'\left(x_0\right), Z_i \right\rangle \omega_{i,n}}{n K(1) F(h)},$$

$$\widetilde{T}_{b,n,2} = \frac{1}{2} \frac{\sum_{i=1}^n \left\langle m''\left(c_i\right)\left(Z_i\right), Z_i \right\rangle \omega_{i,n}}{n K(1) F(h)}.$$

Proposition 38 *We have*

$$\mathbb{E}\widetilde{T}_{b,n,1}^{2} \leq C\frac{h^2}{nF\left(h\right)} + C\frac{v^2\left(h\right)}{F^2\left(h\right)}.$$

Proof We first see that

$$\begin{aligned}
\sum_{i=1}^{n}\left\langle m'\left(x_0\right), X_i - x_0\right\rangle \omega_{i,n} &= \sum_{i=1}^{n}\left\langle m'\left(x_0\right), Z_i\right\rangle K_i \left(1 - \left\langle Z_i, \Gamma_{n,K}^{\dagger}\overline{Z}_{K,n}\right\rangle\right) \\
&= \sum_{i=1}^{n}\left\langle m'\left(x_0\right), Z_i\right\rangle K_i \\
&\quad - \sum_{i=1}^{n}\left\langle m'\left(x_0\right), Z_i\right\rangle K_i \left\langle Z_i, \Gamma_{n,K}^{\dagger}\overline{Z}_{K,n}\right\rangle \\
&= n\left\langle m'\left(x_0\right), \overline{Z}_{K,n}\right\rangle - n\left\langle \Gamma_{n,K} m'\left(x_0\right), \Gamma_{n,K}^{\dagger}\overline{Z}_{K,n}\right\rangle \\
&= n\left\langle m'\left(x_0\right), \left(I - \Gamma_{n,K}\Gamma_{n,K}^{\dagger}\right)\overline{Z}_{K,n}\right\rangle
\end{aligned}$$

and

$$\widetilde{T}_{b,n,1} = \frac{\left\langle m'\left(x_0\right), \left(I - \Gamma_{n,K}\Gamma_{n,K}^{\dagger}\right)\left(\overline{Z}_{K,n}\right)\right\rangle}{K\left(1\right)F\left(h\right)}.$$

Then we split this into two terms:

$$\begin{aligned}
\left\langle m'\left(x_0\right), \left(I - \Gamma_{n,K}\Gamma_{n,K}^{\dagger}\right)\left(\overline{Z}_{K,n}\right)\right\rangle &= \left\langle \left(I - \Gamma_{n,K}\Gamma_{n,K}^{\dagger}\right) m'\left(x_0\right), \left(\overline{Z}_{K,n} - \mathbb{E}\overline{Z}_{K,n}\right)\right\rangle \\
&\quad + \left\langle \left(I - \Gamma_{n,K}\Gamma_{n,K}^{\dagger}\right) m'\left(x_0\right), \mathbb{E}\overline{Z}_{K,n}\right\rangle.
\end{aligned}$$

The L^2 norm of the first is bounded by $Ch\sqrt{F\left(h\right)/n}$ (see Lemma 32) and the L^2 norm of the second is bounded by $Cv\left(h\right)$ (see Lemma 30). This finishes the proof of Proposition 38. □

We turn to $\widetilde{T}_{b,n,2}$ and cut it into two parts:

$$\begin{aligned}
\widetilde{T}_{b,n,2} &= \frac{1}{2}\frac{\sum_{i=1}^{n}\left\langle m''(c_i)(Z_i), Z_i\right\rangle K_i}{nK(1)F(h)} \\
&\quad - \frac{1}{2}\frac{\sum_{i=1}^{n}\left\langle m''(c_i)(Z_i), Z_i\right\rangle K_i \left\langle Z_i, \Gamma_{n,K}^{\dagger}\overline{Z}_{K,n}\right\rangle}{nK(1)F(h)} \\
&= R_{bn1} + R_{bn2}.
\end{aligned}$$

The two forthcoming propositions aim at giving a bound for the mean square norm of R_{bn1} and R_{bn2}.

Proposition 39 *We get*

$$\mathbb{E}R_{bn1}^2 \leq C\left(\frac{h^4}{nF(h)} + h^4\right).$$

Proof It is plain to see that for all i and when Assumption $\mathbf{A}_5$ holds

$$0 \leq \left\langle m''(c_i)(Z_i), Z_i\right\rangle K_i \leq \left(\sup_{x\in\mathcal{V}(x_0)} \left\|m''(x)\right\|_\infty\right)\|Z_i\|^2 K_i$$

hence that

$$0 \leq R_{bn1} \leq \frac{C}{2}\frac{\sum_{i=1}^n \|Z_i\|^2 K_i}{nK(1)F(h)}.$$

It follows that

$$0 \leq R_{bn1}^2 \leq C\frac{\left(\sum_{i=1}^n \|Z_i\|^2 K_i\right)^2}{n^2F^2(h)}.$$

Then

$$\begin{aligned}
0 \leq \mathbb{E}R_{bn1}^2 &\leq \frac{C}{F^2(h)}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n K_i\|Z_i\|^2\right)^2 \\
&= \frac{C}{F^2(h)}\left[\frac{1}{n}\mathbb{E}\left(K_i^2\|Z_i\|^4\right) + \frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\mathbb{E}\left(K_i\|Z_i\|^2 K_j\left\|Z_j\right\|^2\right)\right] \\
&\leq \frac{C}{F^2(h)}\left[\frac{1}{n}\mathbb{E}K_i^2\|Z_i\|^4 + \left(\mathbb{E}K_i\|Z_i\|^2\right)^2\right] \\
&\leq \frac{C}{F^2(h)}\left[\frac{h^4F(h)}{n} + h^4F^2(h)\right] \\
&= C\left[\frac{h^4}{nF(h)} + h^4\right].
\end{aligned}$$

□

We turn to R_{bn2}.

Proposition 40 *We have*

$$\mathbb{E}R_{bn2}^2 \leq C\frac{h^6}{r_n^2}.$$

Proof Dealing with R_{bn2} is a bit more complicated. We have

$$-2R_{bn2} = \frac{1}{K\,(1)\,F\,(h)} \left\langle \Gamma_{n,K}^{\dagger} \overline{Z}_{K,n}, \frac{1}{n} \sum_{i=1}^{n} \left\langle m''\,(c_i)\,(Z_i), Z_i \right\rangle K_i Z_i \right\rangle.$$

The next operation consists in replacing $\overline{Z}_{K,n}$ by its expectation. Like above in the proof of Proposition 38 as well as in the proof of Lemma 33 and 35, we can add and subtract $\mathbb{E}ZK$ from $\overline{Z}_{K,n}$. Once again, we decide not to go through details here for the sake of shortness and clarity. Finally, since the remaining involving $\overline{Z}_{K,n} - \mathbb{E}\overline{Z}_{K,n}$ tends to zero quicker in mean square, we can focus on

$$4R_{bn2}^2 \leq \frac{C}{F^2(h)} \left\| \Gamma_{n,K}^{\dagger} \right\|_{\infty}^2 \|\mathbb{E}KZ\|^2 \left\| \frac{1}{n} \sum_{i=1}^{n} \left\langle m''(c_i)(Z_i), Z_i \right\rangle K_i Z_i \right\|^2. \tag{30}$$

At last, we have to deal with

$$\mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^{n} \left\langle m''\,(c_i)\,(Z_i), Z_i \right\rangle K_i Z_i \right\|^2.$$

Easy computations give

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{i=1}^{n} \left\langle m''(c_i)(Z_i), Z_i \right\rangle K_i Z_i \right\|^2 \\ &\quad = \frac{1}{n^2} \sum_{i=1}^{n} \left\langle m''(c_i)(Z_i), Z_i \right\rangle^2 K_i^2 \|Z_i\|^2 \\ &\qquad + \frac{2}{n^2} \sum_{i<j} \left\langle m''(c_i)(Z_i), Z_i \right\rangle \left\langle m''(c_j)(Z_j), Z_j \right\rangle \left\langle K_i Z_i, K_j Z_j \right\rangle. \end{aligned} \tag{31}$$

Now, we take expectations and use assumption $\mathbf{A}_5$ for the first sum:

$$\begin{aligned} &\frac{1}{n^2} \mathbb{E} \sum_{i=1}^{n} \left\langle m''\,(c_i)\,(Z_i), Z_i \right\rangle^2 K_i^2 \|Z_i\|^2 \\ &\quad \leq C \frac{1}{n^2} \mathbb{E} \sum_{i=1}^{n} \|Z_i\|^4 K_i^2 \|Z_i\|^2 \\ &\quad = \frac{C}{n} \mathbb{E} \left(K_i^2 \|Z_i\|^6 \right) \leq \frac{C}{n} h^6 F\,(h). \end{aligned}$$

Since $h^6 F\,(h)\,/n$ tends to zero at a rate much quicker than the next term, we do not let it appear in the proposition.

We fix i and j in (31) and take expectation

$$
\begin{aligned}
&\mathbb{E}\left\langle m''(c_i)Z_i, Z_i\right\rangle\left\langle m''(c_j)Z_j, Z_j\right\rangle\left\langle K_i Z_i, K_j Z_j\right\rangle \\
&\quad = \left\langle \mathbb{E}\left[\left\langle m''(c_i)Z_i, Z_i\right\rangle K_i Z_i\right], \mathbb{E}\left[\left\langle m''(c_j)Z_j, Z_j\right\rangle K_j Z_j\right]\right\rangle \\
&\quad = \left\| \mathbb{E}\left[\left\langle m''(c)Z, Z\right\rangle KZ\right]\right\|^2.
\end{aligned}
$$

By assumption A_5 we get

$$
\begin{aligned}
&\left|\mathbb{E}\left\langle m''(c_i)Z_i, Z_i\right\rangle\left\langle m''(c_j)(Z_j), Z_j\right\rangle\left\langle K_i Z_i, K_j Z_j\right\rangle\right| \\
&\quad \le \left(\mathbb{E}\left\|\left\langle m''(c)Z, Z\right\rangle KZ\right\|\right)^2 \\
&\quad \le C\left[\mathbb{E}\left(K\left\|Z\right\|^3\right)\right]^2 \\
&\quad \le Ch^6F^2(h).
\end{aligned}
$$

Finally, with (30) at hand we have

$$
\begin{aligned}
\mathbb{E}R_{bn2}^2 &\le \frac{C}{F^2(h)}\frac{v^2}{v^2 r_n^2}\left(\frac{C}{n}h^6F(h) + Ch^6F^2(h)\right) \\
&\le C\frac{h^6}{r_n^2}
\end{aligned}
$$

since $nF(h) \to +\infty$. □

At last, we finish with the proof of the main theorem which is considerably alleviated by all that was done above.

Proof of Theorem 18, Proposition 21 and Proposition 22 The proof of the theorem stems from display (26), Proposition 37, 38, 39 and 40. Collecting these previous results we have

$$
\begin{aligned}
\mathbb{E}\left(\widehat{m}_n(x_0) - m(x_0)\right)^2 \le\; & C\frac{1}{nF^2(h)}\left(F(h) + \frac{h^2F(h)}{nr_n v(h)} + \frac{v(h)}{r_n}\right) \\
& + C\left[\frac{h^6}{r_n^2} + h^4 + \frac{h^2}{nF(h)} + \frac{v^2(h)}{F^2(h)}\right].
\end{aligned}
$$

First, from

$$
v(h) \le h^2F(h),
$$

we see that the first line above will be $O(1/(nF(h)))$ whenever h^2/r_n and $h^2/(nr_n v(h))$ are bounded. We turn to the second line. The term is at least $h^2/(nF(h))$ may be removed because it can be neglected with respect to the variance term. In order to reach $O\left(h^4\right)$ for the bias we have to bound h^2/r_n^2 and $1/\left(h^2nF(h)\right)$.

At last, summing up all what was done above, we take $r_n \asymp h$ and

$$n \cdot \min \left\{ v(h)/h, h^2 F(h) \right\} \geq C > 0.$$

Following the results of Mas (2008a) this last inequality comes down, when ρ is regularly varying at 0 with positive index, to

$$nF(h) \cdot \min \left\{ \rho(h), h^2 \right\} \geq C > 0$$

and Theorem 18 is proved. □

Acknowledgements We greatly thank an Associate Editor and a referee for their careful reading of the paper and helpful suggestions.

References

Adams, R. A., Fournier, J. J. F. (2003). *Sobolev spaces*. New York: Academic Press.

Akhiezer, N. I.,Glazman, I. M. (1981). *Theory of linear operators in hilbert spaces*, Vol 1: monographs and studies in mathematics, Vol. 10. Boston: Pitman.

Barrientos-Marin, J., Ferraty, F., Vieu, P. (2007). Locally modelled regression and functional data (Submitted).

Berlinet, A., Thomas-Agnan, C. (2004). *Reproducing kernel hilbert spaces in probability and statistics*. Boston: Kluwer.

Bingham, N. H., Goldie, C. M., Teugels, J. L. (1987). *Regular variations*. *Encyclopedia of mathematics and its applications*. Cambridge: Cambridge University Press.

Bosq, D. (2000). *Linear processes in function spaces*, Lectures Notes in Statistics, Vol. 149. New York: Springer.

Cai, T., Hall, P. (2006). Prediction in functional linear regression. *Annals of Statistics, 34*(5), 2159–2179.

Cardot, H., Mas, A., Sarda, P. (2007). CLT in functional linear models. *Probability Theory and Related Fields, 138*, 325–361.

Chen, K. (2003). Linear minimax efficiency of local polynomial regression smoothers. *Journal of Nonparametric Statistics 15*(3), 343–353.

Crambes, C., Kneip, A., Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *Annals of Statistics, 37*, 35–72.

Dauxois, J., Pousse, A., Romain, Y. (1982). Asymptotic theory for the principal component analysis of a random vector function some applications to statistical inference. *Journal of Multivariate Analysis, 12*, 136–154.

de Haan, L. (1971). A form of regular variation and its application to the domain of attraction of the double exponential distribution. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 17*, 241–258.

de Haan, L. (1974). Equivalence classes of regularly varying functions. *Stochastic Processes and Their Applications, 2*, 243–259.

Dembo, A., Mayer-Wolf, E., Zeitouni, O. (1995). Exact behavior of gaussian semi-norms. *Statistics and Probability Letters, 23*, 275–280.

Dunford, N., Schwartz, J. T. (1988). *Linear operators, Vols. I and II*. New York: Wiley.

Engl, H. W., Hanke, M., Neubauer, A. (2000). *Regularization of inverse problems*. Boston: Kluwer.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics, 21*, 196–216.

Fan, J., Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Annals of Statistics, 21*, 2008–2036.

Ferraty, F., Mas, A., Vieu, P. (2007). Advances in nonparametric regression for functional variables. *Australian and New-Zealand Journal of Statistics, 49*, 1–20.

Ferraty, F., Vieu, P. (2003). Functional nonparametric statistics: a double infinite dimensional framework. In M. Akritas, D. Politis (Eds.), *Recent advances and trends in nonparametric statistics* (pp. 61–78). New York: Elsevier.

Ferraty, F., Vieu, P. (2006). *Nonparametric functional data analysis*. New York: Springer.

Gaiffas, S. (2005). Convergence rates for pointwise curve estimation with a degenerate design. *Mathematical Methods of Statistics, 14*, 1–27.

Gohberg, I., Goldberg, S., Kaashoek, M. A. (1991). *Classes of linear operators Vols. I and II*: operator theory advances and applications. Basel: Birkhaüser.

Groetsch, C. (1993). *Inverse problems in the mathematical sciences*. Braunschweig: Vieweg.

Kneip, A., Sickles, R. C., Song, W. (2004). *Functional data analysis and mixed effect models*. COMPSTAT 2004—Proceedings in computational statistics (pp. 315–326). Heidelberg: Physica.

Kuelbs, J., Li, W. V., Linde, W. (1994). The Gaussian measure of shifted balls. *Probability Theory and Related Fields, 98*, 143–162.

Ledoux, M., Talagrand, M. (1991). *Probability in banach spaces: isoperimetry and processes,* Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 23. Berlin: Springer.

Li, W. V., Linde, W. (1993). Small ball problems for non-centered gaussian measures. *Theory of Probability and Mathematical Statistics, 14*, 231–251.

Li, W. V., Linde, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Annals of Probability, 27*, 1556–1578.

Li, W. V., Shao, Q.-M. (2001). Gaussian processes inequalities, small ball probabilities and applications. *Handbook of Statistics, 19*, 533–597.

Mas, A. (2007). Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis, 98*, 1231–1261.

Mas, A. (2008a). Local functional principal component analysis. *Complex Analysis and Operator Theory, 2*, 135–167.

Mas, A. (2008b). A representation theorem for gaussian small ball probabilities (manuscript).

Masry, E. (2005). Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stochastic Processes and their Applications, 115*, 155–177.

Mayer-Wolf, E., Zeitouni, O. (1993). The probability of small gaussian ellipsoïds. *Annals of Probability, 21*(1), 14–24.

Müller, H. G., Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics, 33*, 774–805.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications, 9*, 141–142.

Ramsay, J. O., Silverman, B. W. (1997). *Functional data analysis* (2nd ed.). New York: Springer.

Ramsay, J. O., Silverman, B. W. (2002). *Applied functional data analysis methods and case studies*. New York: Springer.

Rockafellar, R. T. (1996). *Convex analysis*. Princeton Landmarks in Mathematics.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics, 10*, 1040–1053.

Tikhonov, A. N., Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. Washington: Winstons and Sons.

Weidman, J. (1980). *Linear operators in hilbert spaces*. Graduate texts in mathematics. New York: Springer.

Yao, F., Müller, H.-G., Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics, 33*(6), 2873–2903.