Nonparametric Statistics
Vol. 17, No. 7, October 2005, 841–856

Quantile regression when the covariates are functions

HERVÉ CARDOT*†, CHRISTOPHE CRAMBES‡ and PASCAL SARDA§

†INRA Toulouse, Biométrie et Intelligence Artificielle, 31326 Castanet-Tolosan Cedex, France
‡Université Paul Sabatier, Laboratoire de Statistique et Probabilités, UMR C5583,
118, route de Narbonne, 31062 Toulouse Cedex, France
§Université Toulouse-le-Mirail, GRIMM, EA 3686, 5, allées Antonio Machado,
31058 Toulouse Cedex 9, France

(Received July 2004; revised August 2005)

This article deals with a linear model of regression on quantiles when the explanatory variable takes values in some functional space and the response is scalar. We propose a spline estimator of the functional coefficient that minimizes a penalized L^1 type criterion. Then, we study the asymptotic behavior of this estimator. The penalization is of primary importance to get existence and convergence.

Keywords: Functional data analysis; Conditional quantiles; B-spline functions; Roughness penalty

1. Introduction

Because of the increasing performances of measurement apparatus and computers, many data are collected and saved on thinner and thinner time scales or spatial grids (temperature curves, spectrometric curves, satellite images, and so on). So, we are led to process data comparable to curves or more generally to functions of continuous variables (time, space). These data are called *functional data* in the literature [1]. Thus, there is a need to develop statistical procedures as well as theory for this kind of data and actually many recent works study models taking into account the functional nature of the data.

Mainly in a formal way, the oldest works in that direction intended to give a mathematical framework based on the theory of linear operators in Hilbert spaces [2, 3]. After that and in an other direction, practical aspects of extensions of descriptive statistical methods, for example, principal component analysis have been considered [4]. The monographs by Ramsay and Silverman [1, 5] are important contributions in this area.

As pointed out by Ramsay and Silverman [5], 'the goals of functional data analysis are essentially the same as those of other branches of Statistics': one of these goals is the explanation of variations of a dependent variable Y (response) by using information from an independent functional variable X (explanatory variable). In many applications, the

*Corresponding author. Email: cardot@toulouse.inra.fr

Nonparametric Statistics
ISSN 1048-5252 print/ISSN 1029-0311 online
http://www.tandf.co.uk/journals
DOI: 10.1080/10485250500303015

response is a scalar; see, for example, refs. [5, 26]. Traditionally, one deals, for such a problem, with estimating the regression on the mean *i.e.* the minimizer among some class of functionals r of

$$\mathbb{E}[(Y - r(X))^2].$$

As when X is a vector of real numbers, the two main approaches are linear (see ref. [27], for the functional linear model) or purely non-parametric (see ref. [6], which adapt kernel estimation to the functional setting). It is also known that estimating the regression on the median or more generally on quantiles has some interest. The problem is then to estimate the minimizer among g_α of

$$\mathbb{E}[l_\alpha(Y - g_\alpha(X))], \tag{1}$$

where $l_\alpha(u) = |u| + (2\alpha - 1)u$. The value $\alpha = 1/2$ corresponds to the conditional median, whereas values $\alpha \in]0, 1[$ correspond to conditional quantiles of order α. The advantage of estimating conditional quantiles may be found in many applications such as in agronomy (estimation of yield thresholds), in medicine or in reliability. Besides robust aspects of the median, it may also help to derive some kind of confidence prediction intervals based on quantiles.

In our work, we assume that the conditional quantile of order α can be written as

$$g_\alpha(X) = \langle \Psi_\alpha, X \rangle, \tag{2}$$

where $\langle ., . \rangle$ is a functional inner product and the parameter of the model Ψ_α, is a function to be estimated. This is the equivalent of the linear model for regression quantiles studied by Koenker and Bassett [7], where the inner product is the Euclidean one and the parameter is a vector of scalars. We choose to estimate the function Ψ_α, by a 'direct' method: writing our estimator as a linear combination of B-splines, it minimizes the empirical version of expectation (1) with the addition of a penalty term proportional to the square norm of a given order derivative of the spline. The penalization term allows us to control the regularity of the estimator on one hand and to get consistency on the other hand.

Unlike for the square function, minimization of function l_α does not lead to an explicit expression of the estimator. Although computation of the estimator can be resolved by using traditional algorithms (for instance, based on iteratively weighted least squares), the convexity of l_α allows theoretical developments.

In section 2, we define more precisely the framework of our study and the spline estimator of the functional parameter Ψ_α. Section 3 is devoted to the asymptotic behavior of our estimator: we study L^2 convergence and derive an upper bound for the rate of convergence. Comments on the model and on the optimality of the rate of convergence are given in section 4. Finally, the proofs are gathered in section 5.

2. Construction of the estimator

In this work, the data consist of an i.i.d. sample of pairs $(X_i, Y_i)_{i=1,\dots,n}$ drawn from a population distribution (X, Y). We consider explanatory variables X_i which are square integrable (random) functions defined on [0,1], *i.e.* are elements of the space $L^2([0, 1])$ so that $X_i = (X_i(t), t \in [0, 1])$. The response Y_i is a scalar belonging to $\mathbb{R}$. Assume that H, the range of X, is a closed subspace of $L^2([0, 1])$. For Y having a finite expectation, $\mathbb{E}(|Y|) < +\infty$, and for $\alpha \in]0, 1[$, the *conditional α-quantile* functional g_α of Y given X is a functional defined on H minimizing equation (1).

Our aim is to generalize the linear model introduced by Koenker and Bassett [7]. In our setting, it consists in assuming that g_α is a linear and continuous functional defined on H and then it follows that $g_\alpha(X)$ can be written as in equation (2). Taking the usual inner product in $L^2([0, 1])$, we can write

$$g_\alpha(X) = \langle \Psi_\alpha, X\rangle = \int_0^1 \Psi_\alpha(t)X(t)\,\mathrm{d}t,$$

where Ψ_α is the functional coefficient in H to be estimated, the order α being fixed. From now on, we consider, for simplicity, that the random variables X_i are centered, that is to say $\mathbb{E}(X_i(t)) = 0$, for t a.e.

When X is multivariate, Bassett and Koenker [8] study the *least absolute error (LAE)* estimator for the conditional median, which can be extended to any quantile replacing the absolute value by the convex function l_α in the criterion to be minimized [7]. In our case, where we have to estimate a function belonging to an infinite dimensional space, we are looking at an estimator in the form of an expansion in some basis of B-spline functions and then minimizing a similar criterion with, however, the addition of a penalty term.

Before describing in detail the estimation procedure, let us note that estimation of conditional quantiles has received special attention in the multivariate case. As mentioned earlier, linear modeling has been mainly investigated by Bassett and Koenker [8]. For non-parametric models, we may distinguish two different approaches: 'indirect' estimators which are based on a preliminary estimation of the conditional cumulative distribution function (cdf) and 'direct' estimators which are based on minimizing the empirical version of criterion (1). In the class of 'indirect' estimators, Bhattacharya and Gangopadhyay [9] studied a kernel estimator of the conditional cdf, and estimation of the quantile is achieved by inverting this estimated cdf. In the class of 'direct' estimators, kernel estimators based on local fit have been proposed [10–12]; in a similar approach, He and Shi [13] and Koenker *et al.* [14] propose a spline estimator. Although our setting is quite different, we adapt in our proofs that follow some arguments of the work by He and Shi [13].

In non-parametric estimation, it is usual to assume that the function to be estimated is sufficiently smooth so that it can be expended in some basis: the degree of smoothness is quantified by the number of derivatives and a Lipschitz condition for the derivative of greatest order (see condition (H2) subsequently). It is also quite usual to approximate such kind of functions by means of regression splines (see ref. [15] for a guide for splines). For this, we have to select a degree q in $\mathbb{N}$ and a sub-division of [0,1] defining the position of the knots. Although it is not necessary, we take equispaced knots so that only the number of the knots has to be selected: for k in $\mathbb{N}^*$, we consider $k - 1$ knots that define a sub-division of the interval [0,1] into k sub-intervals. For asymptotic theory, the degree q is fixed but the number of sub-intervals k depends on the sample size n, $k = k_n$. It is well known that a spline function is a piecewise polynomial: we consider here piecewise polynomials of degree q on each sub-interval and $(q - 1)$ times differentiable on [0,1]. This space of spline functions is a vectorial space of dimension $k + q$. A basis of this vectorial space is the set of the so-called normalized B-spline functions, which we note by $\mathbf{B}_{k,q} = (B_1, \ldots, B_{k+q})^\tau$.

Then, we estimate Ψ_α by a linear combination of functions B_l. This leads us to find a vector $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_{k+q})^\tau$ in $\mathbb{R}^{k+q}$ such that

$$\hat{\Psi}_\alpha = \sum_{l=1}^{k+q} \hat{\theta}_l \mathbf{B}_l = \mathbf{B}_{k,q}^\tau \hat{\boldsymbol{\theta}}. \tag{3}$$

It is then natural to look for $\hat{\Psi}_\alpha$ as the minimizer of the empirical version of equation (1) among functional g_α of the form (2) with functions Ψ_α belonging to the space of spline functions defined earlier. We will, however, consider a penalized criterion as we will see now. In our setting, the pseudo-design matrix $\mathbf{A}$ is the matrix of dimension $n \times (k+q)$ and elements $\langle X_i, B_j \rangle$ for $i = 1, \ldots, n$ and $j = 1, \ldots, k+q$. Even if we do not have an explicit expression for a solution to the minimization problem, it is known that the solution would depend on the properties of the inverse of the matrix $1/n(\mathbf{A}^\tau \mathbf{A})$, which is the $(k+q) \times (k+q)$ matrix with general term $\langle \Gamma_n(B_j), B_l \rangle$, where Γ_n is the empirical version of the covariance operator Γ_X of X defined for all u in $L^2([0,1])$ by

$$\Gamma_X u = \mathbb{E}(\langle X, u \rangle X). \tag{4}$$

We know that Γ_X is a nuclear operator [16]; consequently, no bounded inverse exists for this operator. Moreover, as a consequence of the first monotonicity principle (see Theorem 7.1, p. 58, in ref. [17]), the restriction of this operator to the space of spline functions has smaller eigenvalues than Γ_X. Finally, it appears to be impossible to control the speed of convergence to zero of the smallest eigenvalue of $1/n(\mathbf{A}^\tau \mathbf{A})$ (when n tends to infinity): in that sense, we are faced with an inversion problem that can be qualified as ill-conditioned. A way to circumvent this problem is to introduce a penalization term in the minimization criterion (see refs. [5, 18] or for a similar approach in the functional linear model). Thus, the main role of the penalization is to control the inversion of the matrix linked to the solution of the problem and it consists in restricting the space of solutions. The penalization introduced subsequently will have another effect, as we also want to control the smoothness of our estimator. For this reason, and following several authors (see references earlier), we choose a penalization which allows us to control the norm of the derivative of order $m > 0$ of any linear combination of B-spline functions so that it can be expressed matricially. Denoting by $(\mathbf{B}_{k,q}^\tau \boldsymbol{\theta})^{(m)}$ the mth derivative of the spline function $\mathbf{B}_{k,q}^\tau \boldsymbol{\theta}$, we have

$$\|(\mathbf{B}_{k,q}^\tau \boldsymbol{\theta})^{(m)}\|^2 = \boldsymbol{\theta}^\tau \mathbf{G}_k \boldsymbol{\theta}, \quad \forall\, \boldsymbol{\theta} \in \mathbb{R}^{k+q},$$

where $\mathbf{G}_k$ is the $(k+q) \times (k+q)$ matrix with general term $[\mathbf{G}_k]_{jl} = \langle B_j^{(m)}, B_l^{(m)} \rangle$.

Then, the vector $\hat{\boldsymbol{\theta}}$ in equation (3) is chosen as the solution of the following minimization problem

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^{k+q}} \left\{ \frac{1}{n} \sum_{i=1}^{n} l_\alpha(Y_i - \langle \mathbf{B}_{k,q}^\tau \boldsymbol{\theta}, X_i \rangle) + \rho \|(\mathbf{B}_{k,q}^\tau \boldsymbol{\theta})^{(m)}\|^2 \right\}, \tag{5}$$

where ρ is the penalization parameter. In section 3, we present a convergence result of the solution of equation (5). Note that the role of the penalization also clearly appears in this result.

3. Convergence result

We present in this section the main result on the convergence of our estimator. The behavior of our estimator is linked to a penalized version of the matrix $\hat{\mathbf{C}} = 1/n(\mathbf{A}^\tau \mathbf{A})$. More precisely, adopting the same notations as in ref. [18], the existence and convergence of our estimator depend on the inverse of the matrix $\hat{\mathbf{C}}_\rho = \hat{\mathbf{C}} + \rho \mathbf{G}_k$. Under the hypotheses of Theorem 1 given subsequently, the smallest eigenvalue of $\hat{\mathbf{C}}_\rho$, noted $\lambda_{\min}(\hat{\mathbf{C}}_\rho)$, tends to zero as the sample size n tends to infinity. As the rate of convergence of $\hat{\Psi}_\alpha$ depends on the speed of convergence of

$\lambda_{\min}(\hat{\mathbf{C}}_\rho)$ to zero, we introduce a sequence $(\eta_n)_{n\in\mathbb{N}}$ such that the set Ω_n defined by

$$\Omega_n = \{\omega/\lambda_{\min}(\hat{\mathbf{C}}_\rho)(w) > c\eta_n\} \tag{6}$$

has probability which goes to 1 when n goes to infinity. Cardot *et al.* [18] have shown that such a sequence exists in the sense that under hypotheses of Theorem l, there exists a strictly positive sequence $(\eta_n)_{n\in\mathbb{N}}$ tending to zero as n tends to infinity and such that

$$\lambda_{\min}(\hat{\mathbf{C}}_\rho) \geq c\eta_n + o_P((k_n^2 n^{1-\delta})^{-1/2}), \tag{7}$$

with $\delta \in]0, 1[$.

To prove the convergence result of the estimator $\hat{\Psi}_\alpha$, we assume that the following hypotheses are satisfied.

(Hl) $\|X\| \leq C_0 < +\infty$, a.s.

(H2) The function Ψ_α is supposed to have a p'th derivative $\Psi_\alpha^{(p')}$ such that

$$\left|\Psi_\alpha^{(p')}(t) - \Psi_\alpha^{(p')}(s)\right| \leq C_1|t-s|^\nu, \quad s, t \in [0; 1],$$

where $C_1 > 0$ and $\nu \in [0, 1]$. In what follows, we set $p = p' + \nu$ and we suppose that $q \geq p \geq m$.

(H3) The eigenvalues of Γ_X (defined in equation (4)) are strictly positive.

(H4) For $x \in H$, the random variable ϵ defined by $\epsilon = Y - \langle \Psi_\alpha, X\rangle$ has a density f_x given $X = x$, continuous and bounded below by a strictly positive constant at 0, uniformly for $x \in H$.

We derive in Theorem 1 an upper bound for the rate of convergence with respect to some kind of L^2-norm. Indeed, the operator Γ_X is strictly non-negative, so we can associate it to a semi-norm noted $\|\cdot\|_2$ and defined by $\|u\|_2^2 = \langle \Gamma_X u, u\rangle$. Then, we have the following result.

THEOREM 1 *Under hypotheses* (H1)–(H4), *if we also suppose that there exists* β, γ *in* $]0, 1[$ *such that* $k_n \sim n^\beta$, $p \sim n^{-\gamma}$ *and* $\eta_n \sim n^{-\beta-(1-\delta)/2}$ (*where* δ *is defined in relation* (7)), *then*

(i) $\hat{\Psi}_\alpha$ *exists and is unique except on a set whose probability goes to zero as n goes to infinity,*

(ii) $\|\hat{\Psi}_\alpha - \Psi_\alpha\|_2^2 = O_P\left(\dfrac{1}{k_n^{2p}} + \dfrac{1}{n\eta_n} + \dfrac{\rho^2}{k_n\eta_n} + \rho k_n^{2(m-p)}\right).$

4. Some comments

(i) Hypotheses (Hl) and (H3) are quite usual in the functional setting: see, for instance, refs. [18, 19]. Hypothesis (H4) implies uniqueness of the conditional quantile of order α.

(ii) Some arguments in the proof of Theorem 1 are inspired from the demonstration of He and Shi [13], within the framework of real covariates. Moreover, some results from Cardot *et al.* [18] are also useful, mainly to deal with the penalization term as pointed out earlier.

Note that it is assumed in the model of He and Shi [13] that the error term is independent of X: condition (H4) allows us to deal with a more general setting, as in ref. [7].

(iii) It is possible to choose particular values for β and γ to optimize the upper bound for the rate of convergence in Theorem l. In particular, we remark the importance to control the speed of convergence to 0 of the smallest eigenvalue of $\hat{\mathbf{C}}_\rho$ by η_n. For example, Cardot *et al.* [18] have shown that, under hypotheses of Theorem 1, relation (7) is true with $\eta_n = \rho/k_n$. This gives us

$$\|\hat{\Psi}_\alpha - \Psi_\alpha\|_2^2 = O_p\left(\frac{1}{k_n^{2p}} + \frac{k_n}{n\rho} + \rho + \rho k_n^{2(m-p)}\right).$$

A corollary is obtained if we take $k_n \sim n^{1/(4p+1)}$ and $\rho \sim n^{-2p/(4p+1)}$; then we get

$$\|\hat{\Psi}_\alpha - \Psi_\alpha\|_2^2 = O_p(n^{-2p/(4p+1)}).$$

We can imagine that, with stronger hypotheses on the random function X, we can find a sequence η_n greater than ρ/k_n, which will improve the convergence speed of the estimator. As a matter of fact, the rate derived in Theorem 1 does not imply the rate obtained by Stone [20], that is to say a rate of order $n^{-2p/(2p+1)}$. Indeed, suppose that $1/k_n^{2p}$, $1/(n\eta_n)$ and $\rho^2/(k_n\eta_n)$ are all of order $n^{-2p/(2p+1)}$. This would imply that $k_n \sim n^{1/(2p+1)}$ and $\eta_n \sim n^{-1/(2p+1)}$, which contradict the condition $\eta_n \sim n^{-\beta-(1-\delta)/2}$. Nevertheless, it is possible to obtain a speed of order $n^{-2p/(2p+1)+\kappa}$. This leads to $k_n \sim n^{1/(2p+1)-\kappa/(2p)}$ and $\eta_n \sim n^{-1/(2p+1)-\kappa}$. Then, the condition $\eta_n \sim n^{-\beta-(1-\delta)/2}$ implies $\kappa = p(1-\delta)/(2p+1)$. So, finally, we get $k_n \sim n^{(1+\delta)/2(2p+1)}$, $\rho \sim n^{(-4p-1+\delta)/4(2p+1)}$ and $\eta_n \sim n^{(-p-1+p\delta)/(2p+1)}$. The convergence result would then be

$$\|\hat{\Psi}_\alpha - \Psi_\alpha\|_2^2 = O_p(n^{-p(1+\delta)/(2p+1)}).$$

A final remark is that the last term $\rho\, k_n^{2(m-p)}$ of the speed in Theorem 1 is not always negligible compared to the other terms. However, it will be the case if we suppose that $m \leq p/(1+\delta) + (1-\delta)/4(1+\delta)$.

(iv) This quantile estimator is quite useful in practice, specially for forecasting purpose (by conditional median or inter-quantiles intervals). From a computational point of view, several algorithms may be used: we have implemented an algorithm in the *R* language on the basis of iterated reweighted least square (IRLS). Note that even for real data cases, the curves are always observed in some discretization points, the regression splines are easy to implement by approximating inner products with quadrature rules. The IRLS algorithm [10, 21] allows us to build conditional quantiles spline estimators and gives satisfactory forecast results. This algorithm has been used particularly on the 'ORAMIP' ('Observatoire Régional de l'Air en Midi-Pyrénées') data to forecast pollution in the city of Toulouse (France): the results of this practical study are described in ref. [22]. We are interested in predicting the ozone concentration one day ahead, knowing the ozone curve (concentration along time) the day before. In that special case, conditional quantiles were also useful to predict an ozone threshold such that the probability to exceed this threshold is a given risk $1-\alpha$. In other words, it comes back to give an estimation of the α-quantile maximum ozone knowing the ozone curve the day before.

5. Proof of Theorem 1

The proof of the result is based on the same kind of decomposition of $\hat{\Psi}_\alpha - \Psi_\alpha$ as the one used by He and Shi [13]. The main difference comes from the fact that our design matrix is ill-conditioned, which led us to add the penalization term treated using some arguments from ref. [18]. Hypothesis (H2) implies [15] that there exists a spline function $\Psi_\alpha^* = \mathbf{B}_{k,q}^\tau \boldsymbol{\theta}^*$, called *spline approximation* of Ψ_α, such that

$$\sup_{t\in[0,1]} |\Psi_\alpha^*(t) - \Psi_\alpha(t)| \leq \frac{C_2}{k_n^p}. \tag{8}$$

In what follows, we set $R_i = \langle \Psi_\alpha^* - \Psi_\alpha, X_i\rangle$; so we deduce from equation (8) and from hypothesis (H1) that there exists a positive constant C_3 such that

$$\max_{i=1,\dots,n} |R_i| \leq \frac{C_3}{k_n^p}, \quad \text{a.s.} \tag{9}$$

The operator Γ_n allows us to define the empirical version of the L^2-norm by $\|u\|_n^2 = \langle \Gamma_n u, u\rangle$. At first, we show the result (ii) of Theorem 1 for the penalized empirical L^2-norm. Writing $\hat\Psi_\alpha - \Psi_\alpha = (\hat\Psi_\alpha - \Psi_\alpha^*) + (\Psi_\alpha^* - \Psi_\alpha)$, we get

$$\begin{aligned}\|\hat\Psi_\alpha - \Psi_\alpha\|_n^2 + \rho\|(\hat\Psi_\alpha - \Psi_\alpha)^{(m)}\|^2 \leq{}& \frac{2}{n}\sum_{i=1}^n \langle \hat\Psi_\alpha - \Psi_\alpha^*, X_i\rangle^2 + \frac{2}{n}\sum_{i=1}^n \langle \hat\Psi_\alpha^* - \Psi_\alpha, X_i\rangle^2 \\ &+ 2\rho\|(\hat\Psi_\alpha - \Psi_\alpha^*)^{(m)}\|^2 + 2\rho\|(\Psi_\alpha^* - \Psi_\alpha)^{(m)}\|^2.\end{aligned}$$

Now, using again hypothesis (Hl), we get almost surely and for all $i = 1, \dots, n$, the inequality $\langle \Psi_\alpha^* - \Psi_\alpha, X_i\rangle^2 \leq C_0^2 C_2^2/k_n^{2p}$. Moreover, Lemma 8 in ref. [23] gives us the existence of a positive constant C_4 that satisfies $\|(\Psi_\alpha - \Psi_\alpha^*)^{(m)}\|^2 \leq C_4 k_n^{2(m-p)}$. So, we deduce

$$\begin{aligned}\|\hat\Psi_\alpha - \Psi_\alpha\|_n^2 + \rho\|(\hat\Psi_\alpha - \Psi_\alpha)^{(m)}\|^2 \leq{}& \frac{2}{n}\sum_{i=1}^n \langle \hat\Psi_\alpha - \Psi_\alpha^*, X_i\rangle^2 + 2\rho\|(\hat\Psi_\alpha - \Psi_\alpha^*)^{(m)}\|^2 \\ &+ \frac{2C_0^2C_2^2}{k_n^{2p}} + 2C_4\rho k_n^{2(m-p)}, \text{ a.s.}\end{aligned} \tag{10}$$

Our goal now is to compare our estimator $\hat\Psi_\alpha$ with the spline approximation Ψ_α^*. For that, we adopt the following transformation $\boldsymbol\theta = \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol\beta + \boldsymbol\theta^*$. Then, we define on the set Ω_n

$$f_i(\boldsymbol\beta) = l_\alpha[Y_i - \langle \mathbf{B}_{k,q}^\tau(\hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol\beta + \boldsymbol\theta^*), X_i\rangle] + \rho\|[\mathbf{B}_{k,q}^\tau(\hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol\beta + \boldsymbol\theta^*)]^{(m)}\|^2.$$

We notice that minimizing $\sum_{i=1}^n f_i(\boldsymbol\beta)$ comes back to the minimization of the criterion (5). We are interested by the behavior of the function f_i around zero: $f_i(\mathbf{0})$ is the value of our loss criterion when $\boldsymbol\theta = \boldsymbol\theta^*$. Let us also notice that the inverse of the matrix $\hat{\mathbf{C}}_\rho$ appears in the definition of f_i. This inverse exists on the set Ω_n defined by equation (6) and whose probability goes to 1 as n goes to infinity. Lemma 1, whose proof is given in section 5.1, allows us to get the results (i) and (ii) of Theorem 1 for the penalized empirical L^2-norm.

LEMMA 1 *Under the hypotheses of Theorem* 1, *for all* $\epsilon > 0$, *there exists* $L = L_\epsilon$ (*sufficiently large*) *and* $(\delta_n)_{n\in\mathbb{N}}$ *with* $\delta_n = \sqrt{1/(n\eta_n) + \rho^2/(k_n\eta_n)}$, *such that for n large enough*

$$P\left[\inf_{|\boldsymbol\beta|=L\delta_n} \sum_{i=1}^n f_i(\boldsymbol\beta) > \sum_{i=1}^n f_i(\mathbf{0})\right] > 1 - \epsilon.$$

Using convexity arguments, and the one-to-one transformation $\boldsymbol\theta = \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol\beta + \boldsymbol\theta^*$ on the set Ω_n, we deduce the existence and the uniqueness of the solution of equation (5) on the set Ω_n, which proves point (i) of Theorem l.

Now, let ϵ be strictly positive; using the convexity of function f_i, there exists $L = L_\epsilon$, such that for n large enough

$$P\left[\inf_{|\boldsymbol{\beta}|\geq L\delta_n}\sum_{i=1}^n f_i(\boldsymbol{\beta}) > \sum_{i=1}^n f_i(\mathbf{0})\right] > 1-\epsilon. \tag{11}$$

In contrast, using the definition of f_i and the minimization criterion (5), we have

$$\frac{1}{n}\sum_{i=1}^n f_i(\hat{\mathbf{C}}_\rho^{1/2}\hat{\boldsymbol{\theta}} - \hat{\mathbf{C}}_\rho^{1/2}\boldsymbol{\theta}^*) = \inf_{\boldsymbol{\theta}\in\mathbb{R}^{k+q}}\left[\frac{1}{n}\sum_{i=1}^n l_\alpha(Y_i - \langle \mathbf{B}_{k,q}^\tau\boldsymbol{\theta}, X_i\rangle) + \rho\|(\mathbf{B}_{k,q}^\tau\boldsymbol{\theta})^{(m)}\|^2\right],$$

so, we finally get

$$\frac{1}{n}\sum_{i=1}^n f_i(\hat{\mathbf{C}}_\rho^{1/2}\hat{\boldsymbol{\theta}} - \hat{\mathbf{C}}_\rho^{1/2}\boldsymbol{\theta}^*) \leq \frac{1}{n}\sum_{i=1}^n f_i(\mathbf{0}).$$

Then, combining this with equation (11), we obtain

$$P\left[\inf_{|\boldsymbol{\beta}|\geq L\delta_n}\sum_{i=1}^n f_i(\boldsymbol{\beta}) > \sum_{i=1}^n f_i(\hat{\mathbf{C}}_\rho^{1/2}\hat{\boldsymbol{\theta}} - \hat{\mathbf{C}}_\rho^{1/2}\boldsymbol{\theta}^*)\right] > 1-\epsilon. \tag{12}$$

Now, using the definition of $\hat{\mathbf{C}}_\rho$, we have

$$P\left[\frac{1}{n}\sum_{i=1}^n\langle\hat{\Psi}_\alpha - \Psi_\alpha^*, X_i\rangle^2 + \rho\|(\hat{\Psi}_\alpha - \Psi_\alpha^*)^{(m)}\|^2 \leq L^2\delta_n^2\right] = 1 - P\left[\left|\hat{\mathbf{C}}_\rho^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\right| > L\delta_n\right]$$

$$\geq P\left[\inf_{|\boldsymbol{\beta}|\geq L\delta_n}\sum_{i=1}^n f_i(\boldsymbol{\beta}) > \sum_{i=1}^n f_i(\hat{\mathbf{C}}_\rho^{1/2}\hat{\boldsymbol{\theta}} - \hat{\mathbf{C}}_\rho^{1/2}\boldsymbol{\theta}^*)\right].$$

With relation (12), this last probability is greater than $1-\epsilon$, so we obtain

$$\frac{1}{n}\sum_{i=1}^n\langle\hat{\Psi}_\alpha - \Psi_\alpha^*, X_i\rangle^2 + \rho\|(\hat{\Psi}_\alpha - \Psi_\alpha^*)^{(m)}\|^2 = O_P(\delta_n^2) = O_P\left(\frac{1}{n\eta_n} + \frac{\rho^2}{k_n\eta_n}\right).$$

This last result combined with inequality (10) finally gives us the equivalent of result (ii) for the penalized empirical L^2-norm. Then, point (ii) (with the norm $\|\cdot\|_2$) follows from Lemma 2, which is proved in section 5.5, and achieves the proof of Theorem 1 (ii).

LEMMA 2 *Let f and g be two functions supposed to be m times differentiable and such that*

$$\|f-g\|_n^2 + \rho\|(f-g)^{(m)}\|^2 = O_P(u_n),$$

with u_n going to zero when n goes to infinity. Under hypotheses (H1) *and* (H3) *and if moreover $\|g\|$ and $\|g^{(m)}\|$ are supposed to be bounded, we have*

$$\|f-g\|_2^2 = O_P(u_n).$$

5.1 *Proof of Lemma 1*

This proof is based on three preliminary lemmas proved, respectively, in sections 5.2, 5.3 and 5.4. We denote by T_n the set of the random variables $(X_1, \dots, X_n)$. Under hypotheses of Theorem 1, we have the following results.

LEMMA 3 *There exists a constant C_5 such that given Ω_n defined by* (6), *we have*

$$\max_{i=1,\dots,n} \left|\langle \mathbf{B}_{k,q}^{\tau} \hat{\mathbf{C}}_{\rho}^{-1/2} \boldsymbol{\beta}, X_i \rangle\right| \le \frac{C_5 |\boldsymbol{\beta}|}{\sqrt{k_n \eta_n}}, \text{ a.s.}$$

LEMMA 4 *For all $\epsilon > 0$, there exists $L = L_\epsilon$ such that*

$$\lim_{n\to+\infty} P\left[\inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^{n} (f_i(L\delta_n \boldsymbol{\beta}) - f_i(\mathbf{0}) - \mathbb{E}[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n]) > \epsilon \delta_n^2 n\right] = 0.$$

LEMMA 5 *For all $\epsilon > 0$, there exists $L = L_\epsilon$ such that for m large enough*

$$P\left[\inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^{n} \mathbb{E}[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n] > \delta_n^2 n\right] > 1 - \epsilon.$$

These three lemmas allow us to prove Lemma 1. Indeed, let L be a strictly positive real number; we make the following decomposition

$$\inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^{n} f_i(L\delta_n\boldsymbol{\beta}) - \sum_{i=1}^{n} f_i(\mathbf{0}) \ge A_n + B_n,$$

with

$$A_n = \inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^{n} (f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0}) - \mathbb{E}[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n])$$

and

$$B_n = \inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^{n} \mathbb{E}[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n].$$

Using Lemmas 4 and 5, we can find L sufficiently large, such that for n large enough

$$P(|A_n| > \delta_n^2 n) < \frac{\epsilon}{2},$$

and

$$P(B_n > \delta_n^2 n) > 1 - \frac{\epsilon}{2},$$

thus we get

$$P\left[\inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^{n} f_i(L\delta_n\boldsymbol{\beta}) - \sum_{i=1}^{n} f_i(\mathbf{0}) > 0\right] \ge P(A_n + B_n > 0) > 1 - \epsilon,$$

which achieves the proof of Lemma 1.

5.2 *Proof of Lemma 3*

We have $m\Omega_m$

$$\lambda_{\min}(\hat{\mathbf{C}}_\rho) \geq C_5'\eta_n.$$

Noticing that $|\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle|^2 \leq \langle \mathbf{B}_{k,q}^\tau, X_i\rangle \hat{\mathbf{C}}_\rho^{-1}\langle \mathbf{B}_{k,q}, X_i\rangle|\boldsymbol{\beta}|^2$, we deduce that

$$|\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle|^2 \leq \langle \mathbf{B}_{k,q}^\tau, X_i\rangle\langle \mathbf{B}_{k,q}, X_i\rangle|\boldsymbol{\beta}|^2 \left[\frac{1}{C_5'\eta_n}\right],$$

which gives us $|\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle|^2 \leq C_5''|\boldsymbol{\beta}|^2/(k_n\eta_n)$ almost surely and achieves the proof of Lemma 3.

5.3 *Proof of Lemma 4*

Considering the definition of functions f_i and l_α, we have

$$\begin{aligned}
&\sup_{|\boldsymbol{\beta}|\leq 1}\sum_{i=1}^n (f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0}) - \mathbb{E}[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n]) \\
&= \sup_{|\boldsymbol{\beta}|\leq 1}\sum_{i=1}^n \left(\left|\epsilon_i - L\delta_n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle - R_i\right| - |\epsilon_i - R_i|\right. \\
&\quad \left. - \mathbb{E}\left[\left|\epsilon_i - L\delta_n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle - R_i\right| - |\epsilon_i - R_i||T_n\right]\right),
\end{aligned}$$

where $\epsilon_1, \ldots, \epsilon_n$ are n real random variables independently and identically distributed defined by $\epsilon_i = Y_i - \langle \Psi_\alpha, X_i\rangle$ for all $i = 1, \ldots, n$. Let us also denote $\Delta_i(\boldsymbol{\beta}) = |\epsilon_i - L\delta_n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle - R_i| - |\epsilon_i - R_i|$. To prove Lemma 4, it suffices to show that, for all $\epsilon > 0$, there exists $L = L_\epsilon$ such that

$$\lim_{n\to+\infty} P\left(\sup_{|\boldsymbol{\beta}|\leq 1}\sum_{i=1}^n [\Delta_i(\boldsymbol{\beta}) - \mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)] > \epsilon\delta_n^2 n\right) = 0.$$

Let ϵ be a real number strictly positive and $\mathcal{C}$ the subset of $\mathbb{R}^{k+q}$ defined by $\mathcal{C} = \{\boldsymbol{\beta} \in \mathbb{R}^{k+q}/|\boldsymbol{\beta}| \leq 1\}$. As $\mathcal{C}$ is a compact set, we can cover it with open balls, that is to say $\mathcal{C} = \bigcup_{j=1}^{K_n} \mathcal{C}_j$ with K_n, chosen, for all j from 1 to K_n, such that

$$\operatorname{diam}(\mathcal{C}_j) \leq \frac{\epsilon\delta_n\sqrt{k_n\eta_n}}{8C_5L}. \tag{13}$$

Hence,

$$K_n \leq \left(\frac{8C_5L}{\epsilon\delta_n\sqrt{k_n\eta_n}}\right)^{k_n+q}. \tag{14}$$

Now, for $1 \leq j \leq K_n$, let $\boldsymbol{\beta}_j$ be in $\mathcal{C}_j$; using the definition of $\Delta_i(\boldsymbol{\beta})$ and the triangular inequality, we have

$$\begin{aligned}
&\min_{j=1,\ldots,K_n}\sum_{i=1}^n |[\Delta_i(\boldsymbol{\beta}) - \mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)] - [\Delta_i(\boldsymbol{\beta}_j) - \mathbb{E}(\Delta_i(\boldsymbol{\beta}_j)|T_n)]| \\
&\leq 2L\delta_n \min_{j=1,\ldots,K_n}\sum_{i=1}^n |\langle \mathbf{B}_{k,q}^\tau\ \hat{\mathbf{C}}_\rho^{-1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_j), X_i\rangle|.
\end{aligned}$$

Then, using Lemma 3, we get

$$\min_{j=1,\dots,K_n}\sum_{i=1}^{n}|[\Delta_i(\boldsymbol{\beta})-\mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)]-[\Delta_i(\boldsymbol{\beta}_j)-\mathbb{E}(\Delta_i(\boldsymbol{\beta}_j)|T_n)]|$$
$$\le 2L\delta_n\frac{C_5 n}{\sqrt{k_n\eta_n}}\min_{j=1,\dots,K_n}|\boldsymbol{\beta}-\boldsymbol{\beta}_j|,$$

this last inequality being true only on the set Ω_n defined by equation (6). Moreover, there exists a unique $j_0\in\{1,\dots,K_n\}$ such that $\boldsymbol{\beta}\in\mathcal{C}_{j_0}$, which gives us with relation (13)

$$\min_{j=1,\dots,K_n}\sum_{i=1}^{n}|[\Delta_i(\boldsymbol{\beta})-\mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)]-[\Delta_i(\boldsymbol{\beta}_j)-\mathbb{E}(\Delta_i(\boldsymbol{\beta}_j)|T_n)]|\le\frac{\epsilon}{4}\delta_n^2 n. \tag{15}$$

In contrast, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{C}}|\Delta_i(\boldsymbol{\beta})|\le L\delta_n\sup_{\boldsymbol{\beta}\in\mathcal{C}}|\langle\mathbf{B}_{k,q}^{\tau}\hat{\mathbf{C}}_{\rho}^{-1/2}\boldsymbol{\beta},X_i\rangle|,$$

and using Lemma 3 again, we get, on Ω_n,

$$\sup_{\boldsymbol{\beta}\in\mathcal{C}}|\Delta_i(\boldsymbol{\beta})|\le\frac{C_5 L\delta_n}{\sqrt{k_n\eta_n}}. \tag{16}$$

Besides, for $\boldsymbol{\beta}$ fixed in $\mathcal{C}$, with the same arguments as before, if we denote by T^* the set of the random variables $(X_1,\dots,X_n,\dots)$, we have

$$\sum_{i=1}^{n}\mathrm{Var}(\Delta_i(\boldsymbol{\beta})|T^*)\le\sum_{i=1}^{n}L^2\delta_n^2\,\mathrm{Var}\left(|\langle\mathbf{B}_{k,q}^{\tau}\hat{\mathbf{C}}_{\rho}^{-1/2}\boldsymbol{\beta},X_i\rangle|^2|T^*\right).$$

Then, using the definition of $\hat{\mathbf{C}}_{\rho}$, we remark that

$$\sum_{i=1}^{n}\left|\langle\mathbf{B}_{k,q}^{\tau}\hat{\mathbf{C}}_{\rho}^{-1/2}\boldsymbol{\beta},X_i\rangle\right|^2=n|\boldsymbol{\beta}|^2-n\rho\boldsymbol{\beta}^{\tau}\hat{\mathbf{C}}_{\rho}^{-1/2}\mathbf{G}_k\hat{\mathbf{C}}_{\rho}^{-1/2}\boldsymbol{\beta}, \tag{17}$$

which gives us

$$\sum_{i=1}^{n}\mathrm{Var}(\Delta_i(\boldsymbol{\beta})|T^*)\le nL^2\delta_n^2. \tag{18}$$

We are now able to prove Lemma 4. First using relation (15), we have

$$P\left[\left(\sup_{|\boldsymbol{\beta}|\le 1}\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\beta})-\mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)]>\epsilon\delta_n^2 n\right)\cap\Omega_n\middle|T^*\right]$$
$$\le P\left[\left(\max_{j=1,\dots,K_n}\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\beta}_j)-\mathbb{E}(\Delta_i(\boldsymbol{\beta}_j)|T_n)]>\frac{\epsilon}{2}\delta_n^2 n\right)\cap\Omega_n\middle|T^*\right],$$

and then

$$P\left[\left(\sup_{|\boldsymbol{\beta}|\le 1}\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\beta})-\mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)]>\epsilon\delta_n^2 n\right)\cap\Omega_n\middle|T^*\right]$$
$$\le K_n P\left[\left(\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\beta}_j)-\mathbb{E}(\Delta_i(\boldsymbol{\beta}_j)|T_n)]>\frac{\epsilon}{2}\delta_n^2 n\right)\cap\Omega_n\middle|T^*\right].$$

By inequalities (16) and (18), we apply Bernstein inequality [24] and inequality (14) to obtain

$$P\left[\left(\sup_{|\boldsymbol{\beta}|\le 1}\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\beta})-\mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)]>\epsilon\delta_n^2 n\right)\cap\Omega_n\middle|T^*\right]$$
$$\le 2\exp\left\{\ln\left(\frac{8C_5Ln}{\epsilon\delta_n\sqrt{k_n\eta_n}}\right)^{k_n+q}-\frac{\epsilon^2\delta_n^4n^2/4}{2nL^2\delta_n^2+2C_5L\delta_n\times\epsilon\delta_n^2n/(2\sqrt{k_n\eta_n})}\right\}.$$

This bound does not depend on the sample $T^*=(X_1,\ldots,X_n,\ldots)$; hence, if we take the expectation on both sides of this inequality mentioned earlier, we deduce

$$P\left[\left(\sup_{|\boldsymbol{\beta}|\le 1}\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\beta})-\mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)]>\epsilon\delta_n^2 n\right)\cap\Omega_n\right]\le 2\exp\left\{-\frac{\epsilon^2\delta_n^2\sqrt{k_n\eta_n}n}{8L^2\sqrt{k_n\eta_n}+4C_5L\delta_n}\right.$$
$$\left.\times\left[1-\frac{(k_n+q)(8L^2\sqrt{k_n\eta_n}+4C_5L\delta_n)}{\epsilon^2\delta_n^2\sqrt{k_n\eta_n}n}\ln\left(\frac{8C_5Ln}{\epsilon\delta_n\sqrt{k_n\eta_n}}\right)\right]\right\}.$$

If we fix $L=L_n=\sqrt{nk_n\eta_n\delta_n^2}$, we have

$$\frac{\delta_n^2\sqrt{k_n\eta_n}n}{L^2\sqrt{k_n\eta_n}}=\frac{1}{k_n\eta_n}\xrightarrow[n\to+\infty]{}+\infty,$$
$$\frac{\delta_n^2\sqrt{k_n\eta_n}n}{L\delta_n}=\sqrt{n}\xrightarrow[n\to+\infty]{}+\infty,$$
$$\frac{k_nL^2\sqrt{k_n\eta_n}}{\delta_n^2\sqrt{k_n\eta_n}n}=k_n^2\eta_n\xrightarrow[n\to+\infty]{}0,$$
$$\frac{k_nL\delta_n}{\delta_n^2\sqrt{k_n\eta_n}n}=\frac{k_n}{\sqrt{n}}\xrightarrow[n\to+\infty]{}0.$$

This leads to

$$\lim_{n\to+\infty}P\left[\left(\sup_{|\boldsymbol{\beta}|\le 1}\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\beta})-\mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)]>\epsilon\delta_n^2 n\right)\cap\Omega_n\right]=0,$$

and with the fact that Ω_n has probability tending to 1 when n goes to infinity, we finally obtain

$$\lim_{n\to+\infty}P\left[\sup_{|\boldsymbol{\beta}|\le 1}\sum_{i=1}^{n}[\Delta_i(\boldsymbol{\beta})-\mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)]>\epsilon\delta_n^2 n\right]=0,$$

which achieves the proof of Lemma 4.

5.4 *Proof of Lemma 5*

Let a and b be two real numbers. We denote by $F_{i\epsilon}$ the random repartition function of ϵ_i given T_n and by $f_{i\epsilon}$ the random density function of ϵ_i given T_n. As $\mathbb{E}(l_\alpha(\epsilon_i+b)|T_n)=\int_{\mathbb{R}}l_\alpha(s+b)\mathrm{d}F_{i\epsilon}(s)$, we obtain, using a Taylor linearization at first order, the existence of a quantity r_{iab} such that

$$\mathbb{E}(l_\alpha(\epsilon_i + a + b) - l_\alpha(\epsilon_i + b)|T_n) = f_{i\epsilon}(0)a^2 + 2f_{i\epsilon}(0)ab + \left(\frac{a^2}{2} + ab\right) r_{iab},$$

with $r_{iab} \to 0$ when $a, b \to 0$. If we set $L' = \sqrt{2}L$ and $R_i' = \sqrt{2}R_i$, this relation gives us

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}\left[l_\alpha\left(\epsilon_i - L\delta_n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle - R_i\right) - l_\alpha(\epsilon_i - R_i)|T_n\right] \\ = 2\sum_{i=1}^n f_{i\epsilon}(0)\left[L'^2\delta_n^2\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 + L'\delta_n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle R_i'\right] \\ + \sum_{i=1}^n \left[L'^2\delta_n^2\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 + L'\delta_n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle R_i'\right] r_{i\boldsymbol{\beta}},\end{aligned} \tag{19}$$

with $r_{i\boldsymbol{\beta}} \to 0$. Considering $\boldsymbol{\beta}$ such that $|\boldsymbol{\beta}| = 1$, we have, using relation (9)

$$\begin{aligned}L'^2\delta_n^2\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 + L'\delta_n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle R_i' \\ \geq \frac{1}{2}L'^2\delta_n^2\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 - \frac{C_3^2}{k_n^{2p}}, \text{ a.s.}\end{aligned} \tag{20}$$

Moreover, if we set $V_n = \sup_{|\boldsymbol{\beta}|=1} \max_{i=1\ldots n} |r_{i\boldsymbol{\beta}}|$, then with condition (H4) $\mathbb{1}_{\{V_n < \min_i f_{i\epsilon}(0)/4\}} = \mathbb{1}_{\mathbb{R}}$ for n large enough, and

$$\begin{aligned}\left|\left[L'^2\delta_n^2\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 + L'\delta_n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle R_i'\right] r_{i\boldsymbol{\beta}}\right| \\ \leq \frac{\min_i f_{i\epsilon}(0)}{4}\left|L'^2\delta_n^2\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 + L'\delta_n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle R_i'\right| \\ \leq 2\min_i f_{i\epsilon}(0)\left[\frac{3}{16}L'^2\delta_n^2\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 + \frac{C_3^2}{8k_n^{2p}}\right].\end{aligned} \tag{21}$$

Using inequalities (20) and (21), relation (19) becomes then

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}\left[l_\alpha\left(\epsilon_i - L\delta_n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle - R_i\right) - l_\alpha(\epsilon_i - R_i)|T_n\right] \\ \geq 2\min_i f_{i\epsilon}(0)\left[\frac{5}{16}L'^2\delta_n^2\sum_{i=1}^n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 - \frac{9}{8}\frac{C_3^2 n}{k_n^{2p}}\right].\end{aligned}$$

Now, we come back to the definition of function f_i to obtain

$$\begin{aligned}\frac{1}{\delta_n^2 n}\inf_{|\boldsymbol{\beta}|=1}\sum_{i=1}^n \mathbb{E}[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n] \\ \geq 2\min_i f_{i\epsilon}(0)\left[\frac{5L'^2}{16n}\sum_{i=1}^n\langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 - \frac{9C_3^2}{8k_n^{2p}\delta_n^2}\right] + \rho L^2\left\|\left(\mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}\right)^{(m)}\right\|^2 \\ + 2\frac{L\rho}{\delta_n}\left\langle\left(\mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2}\boldsymbol{\beta}\right)^{(m)}, (\mathbf{B}_{k,q}^\tau\boldsymbol{\theta}^*)^{(m)}\right\rangle.\end{aligned}$$

Reminding that $L'^2 = 2L^2$ and taking $\xi = \min(5/4 \min_i f_{i\epsilon}(0), 1)$, we have $\xi > 0$ by hypothesis (H4) and then

$$\frac{1}{\delta_n^2 n} \inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^n \mathbb{E}[f_i(L\delta_n \boldsymbol{\beta}) - f_i(\mathbf{0})|T_n]$$
$$\geq \xi L^2 \inf_{|\boldsymbol{\beta}|=1} \left[\frac{1}{n} \sum_{i=1}^n \langle \mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2} \boldsymbol{\beta}, X_i \rangle^2 + \rho \left\| \left(\mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2} \boldsymbol{\beta} \right)^{(m)} \right\|^2 \right]$$
$$- \frac{9}{4} \min_i f_{i\epsilon}(0) \frac{C_3^2}{k_n^{2p} \delta_n^2} + \frac{2L\rho}{\delta_n} \left\langle \left(\mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2} \boldsymbol{\beta} \right)^{(m)}, (\mathbf{B}_{k,q}^\tau \boldsymbol{\theta}^*)^{(m)} \right\rangle.$$

Using relation (17), we get

$$\frac{1}{\delta_n^2 n} \inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^n \mathbb{E}[f_i(L\delta_n \boldsymbol{\beta}) - f_i(\mathbf{0})|T_n]$$
$$\geq \xi L^2 - \frac{9}{4} \min_i f_{i\epsilon}(0) \frac{C_3^2}{k_n^{2p} \delta_n^2} + \frac{2L\rho}{\delta_n} \left\langle \left(\mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2} \boldsymbol{\beta} \right)^{(m)}, (\mathbf{B}_{k,q}^\tau \boldsymbol{\theta}^*)^{(m)} \right\rangle.$$

Moreover, for $|\boldsymbol{\beta}| = 1$, the infimum of $\langle (\mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2} \boldsymbol{\beta})^{(m)}, (\mathbf{B}_{k,q}^\tau \boldsymbol{\theta}^*)^{(m)} \rangle$ is obtained for $\boldsymbol{\beta} = -\hat{\mathbf{C}}_\rho^{-1/2} \boldsymbol{\theta}^* / |\hat{\mathbf{C}}_\rho^{-1/2} \boldsymbol{\theta}^*|$. Using the fact that the spline approximation has a bounded mth derivative, we deduce the existence of a constant $C_9 > 0$ such that

$$\inf_{|\boldsymbol{\beta}|=1} \left\langle \left(\mathbf{B}_{k,q}^\tau \hat{\mathbf{C}}_\rho^{-1/2} \boldsymbol{\beta} \right)^{(m)}, (\mathbf{B}_{k,q}^\tau \boldsymbol{\theta}^*)^{(m)} \right\rangle \geq -\frac{C_9}{\sqrt{\eta_n}},$$

hence we obtain

$$\frac{1}{\delta_n^2 n} \inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^n \mathbb{E}[f_i(L\delta_n \boldsymbol{\beta}) - f_i(\mathbf{0})|T_n] \geq \xi L^2 - \frac{9}{4} \min_i f_{i\epsilon}(0) \frac{C_3^2}{k_n^{2p} \delta_n^2} - 2C_9 \frac{L\rho}{\delta_n \sqrt{\eta_n}},$$

that is to say

$$\frac{1}{\delta_n^2 n} \inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^n \mathbb{E}[f_i(L\delta_n \boldsymbol{\beta}) - f_i(\mathbf{0})|T_n] \geq \xi L^2 - \left(1 - \frac{9 \min_i f_{i\epsilon}(0) C_3^2}{4\xi L^2 k_n^{2p} \delta_n^2} - \frac{2C_9 \rho}{\xi L \delta_n \sqrt{\eta_n}} \right).$$

Reminding that we have fixed $L = L_n = \sqrt{n k_n \eta_n \delta_n^2}$,

$$\text{for } \delta_n^2 \sim \frac{1}{n\eta_n}, \text{ we have } \frac{1}{L^2 k_n^{2p} \delta_n^2} \sim \frac{k_n \eta_n}{n \rho^4 k_n^{2p}} \xrightarrow[n\to+\infty]{} 0,$$
$$\text{for } \delta_n^2 \sim \frac{\rho}{k_n \eta_n}, \text{ we have } \frac{\rho}{L\delta_n \sqrt{\eta_n}} \sim \frac{\rho \sqrt{n}}{\sqrt{k_n}} \xrightarrow[n\to+\infty]{} 0.$$

This leads to

$$\lim_{n\to+\infty} P \left(\frac{1}{\delta_n^2 n} \inf_{|\boldsymbol{\beta}|=1} \sum_{i=1}^n \mathbb{E}[f_i(L\delta_n \boldsymbol{\beta}) - f_i(\mathbf{0})|T_n] > 1 \right) = 0,$$

which achieves the proof of Lemma 5.

5.5 *Proof of Lemma 2*

Writing $\Gamma_X = (\Gamma_X - \Gamma_n) + \Gamma_n$, we have the following inequality

$$\|f-g\|_2^2 = 2\|\Gamma_X - \Gamma_n\|(\|f\|^2 + \|g\|^2) + \|f-g\|_n^2. \tag{22}$$

Now, let us decompose f as follows $f = P + R$ with $P(t) = \sum_{l=0}^{m-1}(t^l/l!)f^{(l)}(0)$ and $R(t) = \int_0^t ((t-u)^{m-1}/(m-1)!)f^{(m)}(u)\,\mathrm{d}u$. P belongs to the space $\mathcal{P}_{m-1}$ of polynomials of degree at most $m-1$, whose dimension is finite and equal to m. Using hypothesis (H3), there exists a constant $C_6 > 0$ such that we have $\|P\|^2 \le C_6\|P\|_n^2$ on a space whose probability tends to one as n goes to infinity (see ref. [18]). Then, we can deduce

$$\begin{aligned}\|f\|^2 &\le 2\|P\|^2 + 2\|R\|^2\\ &\le 2C_6\|P\|_n^2 + 2\|R\|^2\\ &\le 4C_6\|f\|_n^2 + 4C_6\|\Gamma_n\|\|R\|^2 + 2\|R\|^2.\end{aligned} \tag{23}$$

As Γ_n is an almost surely bounded operator (by hypothesis (H1)), there exists a constant $C_7 > 0$ such that we have $\|\Gamma_n\| \le C_7$ a.s. Moreover, under Cauchy–Schwarz inequality, there exists a constant $C_8 > 0$ such that $\|R\|^2 \le C_8\|f^{(m)}\|^2$. Relation (23) gives $\|f\|^2 \le 4C_6\|f\|_n^2 + (4C_6C_7 + 2)\ C_8\|f^{(m)}\|^2$. Then, if we write $f = (f-g) + g$, we finally deduce

$$\begin{aligned}\|f\|^2 \le 8C_6\|f-g\|_n^2 &+ (8C_6C_7+4)C_8\|(f-g)^{(m)}\|^2\\ &+ 8C_6\|\Gamma_n\|\|g\|^2 + (8C_6C_7+4)C_8\|g^{(m)}\|^2.\end{aligned} \tag{24}$$

We have supposed that $\|g\|$ and $\|g^{(m)}\|$ are bounded, so almost surely

$$8C_6\|\Gamma_n\|\|g\|^2 + (8C_6C_7+4)C_8\|g^{(m)}\|^2 = O(1),$$

and the hypothesis $\|f-g\|_n^2 + \rho\|(f-g)^{(m)}\|^2 = O_P(u_n)$ gives us the bounds $\|f-g\|_n^2 = O_P(u_n)$ and $\|(f-g)^{(m)}\|^2 = O_P(u_n/\rho)$. Then, relation (24) becomes

$$\|f\|^2 = O_P\left(1 + \frac{u_n}{\rho}\right). \tag{25}$$

Finally, we have $\|\Gamma_X - \Gamma_n\| = O_P(n^{(\delta-1)/2}) = O_P(\rho)$ from Lemma 5.3 of Cardot *et al.* (1999) [25]. This equality, combined with equations (22) and (25) gives us $\|f-g\|_2^2 = O_P(u_n)$, which is the announced result.

Acknowledgements

We would like to thank the referees for a careful reading of a previous version of the manuscript that allowed us to improve some results, as well as the participants to the working group 'staph' in Toulouse for stimulating discussions.

References

[1] Ramsay, J.O. and Silverman, B.W., 2002, *Applied Functional Data Analysis* (Springer-Verlag).

[2] Deville, J.C., 1974, Methodes Statistiques et Numériques de l'Analyse Harmonique (in French), *Ann. Insee*, **15**.

[3] Dauxois, J. and Pousse, A., 1976, Les Analyses Factorielles en Calcul des Probabilités et en Statistique: Essai d'étude synthétique (in French). Thèse, Université Paul sabatier, Toulouse, France.

[4] Besse, P.C. and Ramsay, J.O., 1986, Principal component analysis of sampled curves. *Psychometrika*, **51**, 285–311.

[5] Ramsay, J.O. and Silverman, B.W., 1997, *Functional Data Analysis* (Springer-Verlag).
[6] Ferraty, F. and Vieu, P., 2002, The functional nonparametric model and application to spectrometric data. *Computational Statistics*, **17**, 545–564.
[7] Koenker, R.W. and Bassett G.W., 1978, Regression quantiles. *Econometrica*, **46**, 33–50.
[8] Bassett, G.W. and Koenker, R.W., 1978, Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, **73**, 618–622.
[9] Bhattacharya, P.K. and Gangopadhyay, A.K., 1990, Kernel and nearest-neighbor estimation of a conditional quantile. *Annals of Statistics*, **18**, 1400–1415.
[10] Tsybakov, A.B., 1986, Robust reconstruction of functions by the local approximation method. *Problems of Information Transmission*, **22**, 133–146.
[11] Lejeune, M. and Sarda, P., 1988, Quantile regression: a nonparametric approach. *Computational Statistics and Data Analysis*, **6**, 229–239.
[12] Fan, J., Hu, T.C. and Truong, Y.K., 1994, Robust nonparametric function estimation. *Scandinavian Journal of Statistics*, **21**, 433–446.
[13] He, X. and Shi, P., 1994, Convergence rate of B-Spline estimators of nonparametric conditional quantile functions. *Nonparametric Statistics*, **3**, 299–308.
[14] Koenker, R.W., Ng, P. and Portnoy, S., 1994, Quantile smoothing splines. *Biometrika*, **81**, 673–680.
[15] de Boor, C., 1978, *A Practical Guide to Splines* (New-York: Springer).
[16] Dauxois, J., Pousse, A. and 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.
[17] Weinberger, H.F., 1974, *Variational Methods for Eigenvalue Approximation* (Philadelphia: SIAM).
[18] Cardot, H., Ferraty, F. and Sarda, P., 2003, Spline estimators for the functional linear model. *Statistica Sinica*, **13**, 571–591.
[19] Bosq, D., 2000, *Linear Processes in Function Spaces. Lecture Notes in Statistics*, Vol. 149 (Springer-Verlag).
[20] Stone, C., 1982, Optimal rates of convergence for nonparametric models. *Annals of Statistics*, **10**, 1040–1053.
[21] Ruppert, D. and Carroll, R.J., 1988, *Transformation and Weighting in Regression* (Chapman and Hall).
[22] Cardot, H., Crambes, C. and Sarda, P., 2004, Conditional quantiles with functional covariates: an application to ozone pollution forecasting. In: J. Antoch (Ed.) *Compstat 2004 Proceedings* (Physica-Verlag), pp. 769–776.
[23] Stone, C., 1985, Additive regression and other nonparametric models. *Annals of Statistics*, **13**, 689–705.
[24] Uspensky, J.V., 1937, *Introduction to Mathematical Probability* (New York and London: McGraw Hill Book Company).
[25] Cardot, H., Ferraty, F. and Sarda, P., 1999, The functional linear model. *Statistic and Probability Letters*, **45**, 11–22.
[26] Frank, I.E. and Friedman, J.H., 1993, A statistical view of some chemometrics regression tools (with discussion). *Technometrics*, **35**, 109–148.
[27] Ramsay, J.O. and Dalzell, C.J., 1991, Some tools for functional data analysis. *Journal of the Royal Statistical Society Series B*, **3**, 539–572.