SMOOTHING SPLINES ESTIMATORS FOR FUNCTIONAL LINEAR REGRESSION

By Christophe Crambes, Alois Kneip and Pascal Sarda

Université Paul Sabatier, Université Paul Sabatier and Universität Bonn

The paper considers functional linear regression, where scalar responses $Y_1, \ldots, Y_n$ are modeled in dependence of random functions $X_1, \ldots, X_n$. We propose a smoothing splines estimator for the functional slope parameter based on a slight modification of the usual penalty. Theoretical analysis concentrates on the error in an out-of-sample prediction of the response for a new random function X_{n+1}. It is shown that rates of convergence of the prediction error depend on the smoothness of the slope function and on the structure of the predictors. We then prove that these rates are optimal in the sense that they are minimax over large classes of possible slope functions and distributions of the predictive curves. For the case of models with errors-in-variables the smoothing spline estimator is modified by using a denoising correction of the covariance matrix of discretized curves. The methodology is then applied to a real case study where the aim is to predict the maximum of the concentration of ozone by using the curve of this concentration measured the preceding day.

1. Introduction. In a number of important applications the outcome of a response variable Y depends on the variation of an explanatory variable X over time (or age, etc.). An example is the application motivating our study: the data consist in repeated measurements of pollutant indicators in the area of Toulouse over the course of a day that are used to explain the maximum (peak) of pollution for the next day. Generally, a linear regression model linking observations Y_i of a response variable with p repeated measures of an explanatory variable may be written in the form

$$(1.1) \qquad Y_i = \alpha_0 + \frac{1}{p} \sum_{j=1}^{p} \alpha_j X_i(t_j) + \varepsilon_i^*, \qquad i = 1, \ldots, n.$$

Here $t_1 < \cdots < t_p$ denote observation points which are assumed to belong to a compact interval $I \subset \mathbb{R}$. The possibly varying strength of the influence of X_i at each measurement point t_j is quantified by different coefficients α_j. Frequently $p \gg n$ and/or there is a high degree of collinearity between the "predictors" $X_i(t_j), j = 1, \ldots, p$, and standard regression methods are not applicable. In addition, (1.1) may incorporate a discretization error, since one will often have to

assume that Y_i also depends on unobserved time points t in between the observation times t_j. As pointed out by several authors (Marx and Eilers [22], Ramsay and Silverman [26] or Cuevas, Febrero and Fraiman [10]) the use of functional models for these settings has some advantages over discrete, multivariate approaches. Only in a functional framework is it possible to profit from qualitative assumptions like smoothness of underlying curves. Assuming square integrable functions X_i on $I \subset \mathbb{R}$, the basic object of our study is a *functional linear regression model*

$$(1.2) \qquad Y_i = \alpha_0 + \int_I \alpha(t) X_i(t)\,dt + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where ε_i's are i.i.d. centered random errors, $\mathbb{E}(\varepsilon_i) = 0$, with variance $\mathbb{E}(\varepsilon_i^2) = \sigma_\varepsilon^2$, and α is a square integrable functional parameter defined on I that must be estimated from the pairs (X_i, Y_i), $i = 1, \ldots, n$. This type of regression model was first considered in Ramsay and Dalzell [24]. Obviously, (1.2) constitutes a continuous version of (1.1), and both models are linked by

$$(1.3) \quad \varepsilon_i^* = d_i + \varepsilon_i, \qquad \text{where } d_i = \int_I \alpha(t) X_i(t)\,dt - \frac{1}{p}\sum_{j=1}^{p} \alpha(t_j) X_i(t_j)$$

may be interpreted as a discretization error, and $\alpha(t_j) = \alpha_j$.

As a consequence of developments of modern technology, data that may be described by functional regression models can be found in a lot of fields such as medicine, linguistics, chemometrics (see, e.g., Ramsay and Silverman [25, 26] and Ferraty and Vieu [14], for several case studies). Similarly to traditional regression problems, model (1.2) may arise under different experimental designs. We assume a random design of the explanatory curves, where $X_1, \ldots, X_n$ is a sequence of identically distributed random functions with the same distribution as a generic X. The main assumption on X is that it is a second-order variable, that is, $\mathbb{E}(\int_I X^2(t)\,dt) < +\infty$, and it is assumed moreover that $\mathbb{E}(X_i(t)\varepsilon_i) = 0$ for almost every $t \in I$. This situation has been considered, for instance, in Cardot, Ferraty and Sarda [7] and Müller and Stadtmüller [23] for independent variables, while correlated functional variables are studied in Bosq [2]. Our analysis is based on a general framework without any assumption of independence of the X_i's. We will, however, assume independence between the X_i's and the ε_i's in our theoretical results in Sections 3 and 4.

The main problem in functional linear regression is to derive an estimator $\widehat{\alpha}$ of the unknown slope function α. However, estimation of α in (1.2) belongs to the class of ill-posed inverse problems. Writing (1.2) for generic variables X, Y and ε, multiplying both sides by $X - \mathbb{E}(X)$ and then taking expectations leads to

$$(1.4) \qquad \begin{aligned} &\mathbb{E}\big((Y - \mathbb{E}(Y))(X - \mathbb{E}(X))\big) \\ &= \mathbb{E}\bigg(\int_I \alpha(t)\big(X(t) - \mathbb{E}(X)(t)\big)\,dt\,(X - \mathbb{E}(X))\bigg) =: \Gamma(\alpha). \end{aligned}$$

The normal equation (1.4) is the continuous equivalent of normal equations in the multivariate linear model. Estimation of α is thus linked with the inversion of the covariance operator Γ of X defined in (1.4). But, unlike the finite dimensional case, a bounded inverse for Γ does not exist since it is a compact linear operator defined on the infinite dimensional space $L^2(I)$. This corresponds to the setup of ill-posed inverse problems (with the additional difficulty that Γ is unknown). As a consequence, the parameter α in (1.2) is not identifiable without additional constraint. Actually, a necessary and sufficient condition under which a unique solution for (1.2)–(1.4) exists in the orthogonal space of ker(Γ) and is given by $\sum_r (\frac{\mathbb{E}((Y-\mathbb{E}(Y))\int_I (X(t)-\mathbb{E}(X)(t))\zeta_r(t)\,dt)}{\lambda_r})^2 < +\infty$, where $(\lambda_r, \zeta_r)_r$ are the eigenelements of Γ (see Cardot, Ferraty and Sarda [7] or He, Müller and Wang [19] for a functional response). The set of solutions is the set of functions α which can be decomposed as a sum of the unique element of the orthogonal space of ker(Γ) satisfying (1.4) and any element of ker(Γ).

It follows from these arguments that any sensible procedure for estimating α (or, more precisely, of its identifiable part) has to involve regularization procedures. Several authors have proposed estimation procedures where regularization is obtained in two main ways. The first one is based on the Karhunen–Loève expansion of X and leads to regression on functional principal components: see Bosq [2], Cardot, Mas and Sarda [8] or Müller and Stadtmüller [23]. It consists in projecting the observations on a finite dimensional space spanned by eigenfunctions of the (empirical) covariance operator Γ_n. For the second method, regularization is obtained through a penalized least squares approach after expanding α in some basis (such as splines): see Ramsay and Dalzell [24], Eilers and Marx [12], Cardot, Ferraty and Sarda [7] or Li and Hsing [21]. We propose here to use a smoothing splines approach prolonging a previous work from Cardot et al. [5].

Our estimator is described in Section 2. Note that (1.2) implies that $Y_i - \overline{Y} = \int_I \alpha(t)[X_i(t) - \overline{X}(t)]\,dt + \varepsilon_i - \bar{\varepsilon}$. Based on the observation times $t_1 < \cdots < t_p$, we rely on minimizing the residual sum of squares $\sum_i (Y_i - \overline{Y} - \frac{1}{p}\sum_{j=1}^p a(t_j)(X_i(t_j) - \overline{X}(t_j)))^2$ subject to a roughness penalty. A slight modification of the usual penalty term is applied in order to guarantee the existence of the estimator under general conditions. The proposed estimator $\widehat{\alpha}$ is then a natural spline with knots at the observation points t_j. An estimator of the intercept $\alpha_0 = \mathbb{E}(Y) - \int_I \alpha(t)\mathbb{E}(X)(t)]\,dt$ is given by $\widehat{\alpha_0} = \overline{Y} - \int_I \widehat{\alpha}(t)\overline{X}(t)\,dt$. For simplicity, we will assume that $t_1 < \cdots < t_p$ are equispaced, but the methodology can easily be generalized to other situations. It must be emphasized, however, that our study does not cover the case of sparse points for which other techniques have to be envisaged; for this specific problem, see the work from Yao, Müller and Wang [32].

In Section 3 we present a detailed asymptotic theory of the behavior of our estimator for large values of n and p. The distance between $\widehat{\alpha}$ and α is evaluated with respect to L^2 semi-norms induced by the operator Γ, $\|u\|_\Gamma^2 = \langle \Gamma u, u \rangle$ with $\langle u, v \rangle = \int_I u(t)v(t)\,dt$, or its discretized or empirical versions (see, e.g., Cardot,

Ferraty and Sarda [7] or Müller and Stadtmüller [23] for similar setups). By using these semi-norms we explicitly concentrate on analyzing the estimation error only for the identifiable part of the structure of α which is relevant for prediction. Indeed, it will be shown in Section 3 that $\|\widehat{\alpha} - \alpha\|_\Gamma^2$ determines the rate of convergence of the error in predicting the conditional mean $\alpha_0 + \int_I \alpha(t) X_{n+1}(t)\,dt$ of Y_{n+1} for any new random function X_{n+1} possessing the same distribution as X and independent of $X_1, \ldots, X_n$:

$$
\begin{aligned}
(1.5) \quad & \mathbb{E}\left(\left(\widehat{\alpha_0} + \int_I \widehat{\alpha}(t) X_{n+1}(t)\,dt - \alpha_0 - \int_I \alpha(t) X_{n+1}(t)\,dt\right)^2 \middle| \widehat{\alpha_0}, \widehat{\alpha}\right) \\
& = \|\widehat{\alpha} - \alpha\|_\Gamma^2 + O_P(n^{-1}).
\end{aligned}
$$

We first derived optimal rates of convergence with respect to the L^2 semi-norms induced by Γ in a quite general setting which substantially improved existing results in the literature as well as bounds obtained for this estimator in a previous paper (see Cardot et al. [5]). If α is m-times continuously differentiable, then it is shown that rates of convergence for our estimator are of order $n^{-(2m+2q+1)/(2m+2q+2)}$, where the value of $q > 0$ depends on the structure of the distribution of X. More precisely, q quantifies the rate of decrease $\sum_{r=k+1}^\infty \lambda_r = O(k^{-2q})$ as $k \to \infty$, where $\lambda_1 \geq \lambda_2 \geq \cdots$ are the eigenvalues of the covariance operator Γ. If, for example, X is a.s. twice continuously differentiable, then $q \geq 2$. As a second step, we show that these rates of convergence are optimal in the sense that they are minimax over large classes of distributions of X and of functions α. No alternative estimator can globally achieve faster rates of convergence in these classes.

In an interesting paper Cai and Hall [4] derive rates of convergence on the error $\alpha_0 + \langle \alpha, x \rangle - \widehat{\alpha}_0 - \langle \widehat{\alpha}, x \rangle$ for a pre-specified, fixed function x. Their approach is based on regression with respect to functional principal components and the derived rates are shown to be optimal with respect to this methodology. At first glance this setup seem to be close, but due to the fact that explanatory variables are of infinite dimension, inference on *fixed* functions x cannot generally be used to derive optimal rates of convergence of the prediction error (1.5) for *random* functions X_{n+1}. We also want to emphasize that in the present paper we do not consider the convergence of $\hat{\alpha}$ with respect to the usual L^2 norm. Analyzing $\|\widehat{\alpha} - \alpha\|^2 = \int_I (\widehat{\alpha}(t) - \alpha(t))^2\,dt$ instead of $\|\widehat{\alpha} - \alpha\|_\Gamma^2$ must be seen statistically as a very different problem, and under our general assumptions it only follows that $\|\widehat{\alpha} - \alpha\|^2$ is bounded in probability (see the proof of Theorem 2). It appears that to get stronger results one needs additional conditions linking the "smoothness" of α and of the curves X_i as derived in a recent work by Hall and Horowitz [18]. A detailed discussion of these issues is given in Section 3.2.

In practice the functional values $X_i(t_j)$ are often not directly observed; there exist only noisy observations $W_{ij} = X_i(t_j) + \delta_{ij}$ contaminated with random errors δ_{ij}. In Section 4, we consider a modified functional linear model adapting

to such situations. In this errors-in-variable context, we use a corrected estimator as introduced in Cardot et al. [5] which can be seen as a modified version of the so-called *total least squares* method for functional data. We show again the good asymptotic performance of the method for a sufficiently dense grid of discretization points.

We devote Section 5 to the application of the proposed estimation procedure to the prediction of the peak of pollution from the curve of pollutant indicators collected the preceding day. Finally, the proofs of our results can be found in Section 6.

2. Smoothing splines estimation of the functional coefficient.

As explained in the Introduction, we will assume that the functions X_i are observed at p equidistant points $t_1, \ldots, t_p \in I$. In order to simplify further developments, we will take $I = [0, 1]$ so that $t_1 = \frac{1}{2p}$ and $t_j - t_{j-1} = \frac{1}{p}$ for all $j = 2, \ldots, p$.

Our estimator of α in (1.2) is a generalization of the well-known smoothing splines estimator in univariate nonparametric regression. It relies on the implicit assumption that the underlying function α is sufficiently smooth as, for example, m-times continuously differentiable ($m = 1, 2, 3, \ldots$).

For any smooth function a the discrete sum $\frac{1}{p} \sum_{j=1}^{p} a(t_j) X_i(t_j)$ is used to approximate the integral $\int_0^1 a(t) X_i(t) \, dt$ in (1.2), whereas expectations are estimated by the sample means $\overline{Y}$ and $\overline{X}$, and an estimate is obtained by minimizing the sum of squared residuals $(Y_i - \overline{Y} - \frac{1}{p} \sum_{j=1}^{p} a(t_j)(X_i(t_j) - \overline{X}(t_j)))^2$ subject to a roughness penalty. More precisely, for some $m = 1, 2, \ldots$ and a smoothing parameter $\rho > 0$, an estimate $\widehat{\alpha}$ is determined by minimizing

$$
\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \overline{Y} - \frac{1}{p} \sum_{j=1}^{p} a(t_j)\big(X_i(t_j) - \overline{X}(t_j)\big) \right)^2
$$

(2.1)

$$
+ \rho \left(\frac{1}{p} \sum_{j=1}^{p} \pi_a^2(t_j) + \int_0^1 \big(a^{(m)}(t)\big)^2 \, dt \right)
$$

over all functions a in the Sobolev space $W^{m,2}([0, 1]) \subset L^2([0, 1])$, where $\pi_a(t) = \sum_{l=1}^{m} \beta_{a,l} t^{l-1}$ with $\sum_{j=1}^{p} (a(t_j) - \pi_a(t_j))^2 = \min_{\beta_1, \ldots, \beta_m} \sum_{j=1}^{p} (a(t_j) - \sum_{l=1}^{m} \beta_l t^{l-1})^2$.

Obviously, π_a denotes the best possible approximation of $(a(t_1), \ldots, a(t_p))$ by a polynomial of degree $m - 1$. The extra term $\frac{1}{p} \sum_{j=1}^{p} \pi_a(t_j)^2$ in the roughness penalty is unusual and does not appear in traditional smoothing splines approaches. It will, however, be shown below that this term is necessary to guarantee existence of a unique solution in a general context without any additional assumptions on the curves X_i.

It is quite easily seen that any solution $\widehat{\alpha}$ of (2.1) has to be an element of the space $NS^m(t_1, \ldots, t_p)$ of *natural splines* of order $2m$ with knots at $t_1, \ldots, t_p$.

Recall that $NS^m(t_1, \ldots, t_p)$ is a p-dimensional linear space of functions with $v^{(m)} \in L^2([0, 1])$ for any $v \in NS^m(t_1, \ldots, t_p)$. Let $\mathbf{b}(t) = (b_1(t), \ldots, b_p(t))^\tau$ be a functional basis of $NS^m(t_1, \ldots, t_p)$. A discussion of several possible basis function expansions can be found in Eubank [13]. An important property of natural splines is that there exists a canonical one-to-one mapping between $\mathbb{R}^p$ and the space $NS^m(t_1, \ldots, t_p)$ in the following way: for any vector $\mathbf{w} = (w_1, \ldots, w_p)^\tau \in \mathbb{R}^p$, there exists a unique natural spline interpolant $s_\mathbf{w}$ with $s_\mathbf{w}(t_j) = w_j$, $j = 1, \ldots, p$. With $\mathbf{B}$ denoting the $p \times p$ matrix with elements $b_i(t_j)$, $s_\mathbf{w}$ is given by

$$(2.2) \qquad s_\mathbf{w}(t) = \mathbf{b}(t)^\tau (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau \mathbf{w}.$$

The important property of such a spline interpolant is the fact that

$$(2.3) \qquad \int_0^1 s_\mathbf{w}^{(m)}(t)^2 \, dt \le \int_0^1 f^{(m)}(t)^2 \, dt$$

$$\text{for any other function } f \in W^{m,2}([0, 1])$$
$$\text{with } f(t_j) = w_j, \, j = 1, \ldots, p.$$

Note that in (2.1) only the integral $\int_0^1 a^{(m)}(t)^2 \, dt$ depends on the values of a in the open intervals (t_{j-1}, t_j) between grid points. It therefore follows from (2.3) that $\widehat{\alpha} = s_{\widehat{\alpha}}$, where $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}(t_1), \ldots, \widehat{\alpha}(t_p))^\tau \in \mathbb{R}^p$ minimizes

$$(2.4) \qquad \begin{aligned} &\frac{1}{n} \sum_{i=1}^n \left(Y_i - \overline{Y} - \frac{1}{p} \sum_{j=1}^p a(t_j) \big(X_i(t_j) - \overline{X}(t_j) \big) \right)^2 \\ &\qquad + \rho \left(\frac{1}{p} \sum_{j=1}^p \pi_a^2(t_j) + \int_0^1 s_\mathbf{a}^{(m)}(t)^2 \, dt \right); \end{aligned}$$

with respect to all vectors $\mathbf{a} = (a(t_1), \ldots, a(t_p))^\tau \in \mathbb{R}^p$.

A closer study of $\widehat{\alpha}$ requires the use of matrix notation: $\mathbf{Y} = (Y_1 - \overline{Y}, \ldots, Y_n - \overline{Y})^\tau$, $\mathbf{X}_i = (X_i(t_1) - \overline{X}(t_1), \ldots, X_i(t_p) - \overline{X}(t_p))^\tau$ for all $i = 1, \ldots, n$, $\boldsymbol{\alpha} = (\alpha(t_1), \ldots, \alpha(t_p))^\tau$, $\boldsymbol{\varepsilon} = (\varepsilon_1 - \overline{\varepsilon}, \ldots, \varepsilon_n - \overline{\varepsilon})^\tau$ and let $\mathbf{X}$ be the $n \times p$ matrix with a general term $X_i(t_j) - \overline{X}(t_j)$ for all $i = 1, \ldots, n$, $j = 1, \ldots, p$. Moreover, $\mathbf{P}_m$ will denote the $p \times p$ projection matrix projecting into the m-dimensional linear space $E_m := \{ \mathbf{w} = (w_1, \ldots, w_p)^\tau \in \mathbb{R}^p | w_j = \sum_{l=1}^m \theta_l t_j^{l-1}, \, j = 1, \ldots, p \}$ of all (discretized) polynomials of degree $m - 1$. By (2.2), we have $\int_0^1 s_\mathbf{a}^{(m)}(t)^2 \, dt = \mathbf{a}^\tau \mathbf{A}_m^* \mathbf{a}$, where $\mathbf{A}_m^* = \mathbf{B}(\mathbf{B}^\tau \mathbf{B})^{-1} [\int_0^1 \mathbf{b}^{(m)}(t) \mathbf{b}^{(m)}(t)^\tau \, dt] (\mathbf{B}^\tau \mathbf{B})^{-1} \mathbf{B}^\tau$ is a $p \times p$ matrix.

When defining $\mathbf{A}_m := \mathbf{P}_m + p \mathbf{A}_m^*$, minimizing (2.4) is equivalent to solving

$$(2.5) \qquad \min_{\mathbf{a} \in \mathbb{R}^p} \left\{ \frac{1}{n} \left\| \mathbf{Y} - \frac{1}{p} \mathbf{X} \mathbf{a} \right\|^2 + \frac{\rho}{p} \mathbf{a}^\tau \mathbf{A}_m \mathbf{a} \right\},$$

where $\| \cdot \|$ stands for the usual Euclidean norm. The solution is given by

$$(2.6) \qquad \widehat{\boldsymbol{\alpha}} = \frac{1}{np} \left(\frac{1}{np^2} \mathbf{X}^\tau \mathbf{X} + \frac{\rho}{p} \mathbf{A}_m \right)^{-1} \mathbf{X}^\tau \mathbf{Y} = \frac{1}{n} \left(\frac{1}{np} \mathbf{X}^\tau \mathbf{X} + \rho \mathbf{A}_m \right)^{-1} \mathbf{X}^\tau \mathbf{Y}.$$

Then $\widehat{\alpha} = s_{\widehat{\alpha}}$ constitutes our final estimator of α while $\widehat{\alpha_0} = \overline{Y} - \langle \widehat{\alpha}, \overline{X} \rangle$ is used to estimate the intercept α_0. Based on a somewhat different development, this estimator of α has already been proposed by Cardot et al. [5].

In order to verify existence of $\widehat{\boldsymbol{\alpha}}$, let us first cite some properties of the eigenvalues of $p\mathbf{A}_m^*$ which have been studied by many authors (see Eubank [13]). For instance, in Utreras [28], it is shown that this matrix has exactly m zero eigenvalues $\mu_{1,p} = \cdots = \mu_{m,p} = 0$. The corresponding m-dimensional eigenspace is the space E_m of discretized polynomials as defined above. The $p - m$ nonzero eigenvalues $0 < \mu_{m+1,p} < \cdots < \mu_{p,p}$ are such that there exist constants $0 < D_0 < D_1 < \infty$ such that $D_0 \leq \mu_{j+m,p}(\pi j)^{-2m} \leq D_1$ for $j = 1, \ldots, p - m$ and all sufficiently large p. Therefore, there exist some constant $0 < C_0 < +\infty$ and some $p_0 \in \{0, 1, 2, \ldots\}$ such that for all $p \geq p_0$ and $k = 0, \ldots, p - m - 1$

$$(2.7) \qquad\qquad k^{2m} \frac{1}{\mu_{k+m+1,p}} \leq C_0.$$

We can conclude that all eigenvalues of the matrix $\mathbf{A}_m$ are strictly positive, and existence as well as uniqueness of the solution (2.6) of the minimization problem (2.5) are straightforward consequences. Note that Introduction of the additional term $\frac{1}{p} \sum_{j=1}^{p} \pi_a(t_j)^2$ in (2.1) is crucial. Dropping this term in (2.1) as well as (2.4) results in replacing $\mathbf{A}_m$ by $p\mathbf{A}_m^*$ in (2.5). Existence of a solution then cannot be guaranteed in a general context since, due to the m zero eigenvalues of $p\mathbf{A}_m^*$, the matrix $(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m^*)$ may not be invertible.

REMARK.   Our requirement of equidistant grid points t_j has to be seen as a restrictive condition. There are many applications where the functions X_i are only observed at varying numbers p_i of irregularly spaced points $t_{i1} \leq \cdots \leq t_{ip_i}$. Then our estimation procedure is not directly applicable. Fortunately there exists a fairly simple modification. Define a smooth function $\widetilde{X}_i \in L^2([0, 1])$ by smoothly interpolating the observations (e.g., using natural splines) such that $\widetilde{X}_i(t_{ij}) = X_i(t_{ij})$, $j = 1, \ldots, p_i$. Then define $p > \max\{p_1, \ldots, p_n\}$ equidistant grid points $t_1, \ldots, t_p$, and determine an estimator $\widehat{\alpha}$ by applying the smoothing spline procedure (2.1) with $\frac{1}{p} \sum_{j=1}^{p} a(t_j)(X_i(t_j) - \overline{X}(t_j))$ being replaced by $\frac{1}{p} \sum_{j=1}^{p} a(t_j)(\widetilde{X}_i(t_j) - \overline{\widetilde{X}}(t_j))$. For example, in the case of a random design with i.i.d. observations t_{ij} from a strictly positive design density on I, it may be shown that the asymptotic results of Section 3 generalize to this situation if $\min\{p_1, \ldots, p_n\}$ is sufficiently large compared to n. A detailed analysis is not in the scope of the present paper.

3. Theoretical results.

3.1. *Rates of convergence for smoothing splines estimators*.   We will denote the standard inner product of the Hilbert space $L^2([0, 1])$ by $\langle f, g \rangle =$

$\int_0^1 f(t)g(t)\,dt$ and $\|\cdot\|$ by its associated norm. As outlined in the Introduction, our analysis is based on evaluating the error between $\widehat{\alpha}$ and α with respect to the semi-norm $\|\cdot\|_\Gamma$ defined in Section 1,

$$\|u\|_\Gamma^2 := \langle \Gamma u, u \rangle, \qquad u \in L^2([0,1]),$$

where Γ is the *covariance operator* of X given by

$$\Gamma u := \mathbb{E}(\langle (X - \mathbb{E}(X)), u \rangle (X - \mathbb{E}(X))), \qquad u \in L^2([0,1]).$$

The above L^2 semi-norm has already been used in similar contexts as the one studied in the present paper; see, for example, Wahba [30], Cardot, Ferraty and Sarda [7] or Müller and Stadtmüller [23]. By (1.5) the asymptotic behavior of $\|\widehat{\alpha} - \alpha\|_\Gamma^2$ constitutes a major object of interest, since it quantifies the leading term in the expected squared prediction error for a new random function X_{n+1}.

As first steps, we will consider in Theorems 1 and 2 the error between $\widehat{\alpha}$ and α with respect to simplified versions of the above semi-norm: the discretized empirical semi-norm defined for any $\mathbf{u} \in \mathbb{R}^p$ as

$$\|\mathbf{u}\|_{\Gamma_{n,p}}^2 := \frac{1}{p}\mathbf{u}^\tau \left(\frac{1}{np}\mathbf{X}^\tau \mathbf{X} \right) \mathbf{u},$$

and the empirical semi-norm defined for any $u \in L^2([0,1])$ as

$$\|u\|_{\Gamma_n}^2 := \frac{1}{n}\sum_{i=1}^n \langle (X_i - \overline{X}), u \rangle^2 = \langle \Gamma_n u, u \rangle,$$

where Γ_n is the empirical covariance operator from $X_1, \ldots, X_n$ given by

$$\Gamma_n u := \frac{1}{n}\sum_{i=1}^n \langle (X_i - \overline{X}), u \rangle (X_i - \overline{X}).$$

Obviously, $\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_{\Gamma_{n,p}}^2 = \frac{1}{n}\sum_i [\frac{1}{p}\sum_{j=1}^p (\widehat{\alpha}(t_j) - \alpha(t_j))(X_i(t_j) - \overline{X}(t_j))]^2$ and $\|\widehat{\alpha} - \alpha\|_{\Gamma_n}^2 = \frac{1}{n}\sum_i [\int_I (\widehat{\alpha}(t) - \alpha(t))(X_i(t) - \overline{X}(t))\,dt]^2$ quantify different modes of convergence of $\langle \widehat{\alpha}, X - \overline{X} \rangle$ to $\langle \alpha, (X - \overline{X}) \rangle$.

As mentioned in Section 2, the function α is required to have a certain degree of regularity. Namely, it satisfies the following assumption for some $m \in \{1, 2, \ldots\}$:

(A.1) α is m-times differentiable and $\alpha^{(m)}$ belongs to $L^2([0,1])$.

Let $C_1 = \int_0^1 \alpha^{(m)}(t)^2\,dt$ and $C_2^* = \int_0^1 \alpha(t)^2\,dt$. By construction of $\mathbf{P}_m$, $\mathbf{P}_m\boldsymbol{\alpha}$ provides the best approximation (in a least squares sense) of $\boldsymbol{\alpha}$ by (discretized) polynomials of degree $m - 1$, and $\frac{1}{p}\boldsymbol{\alpha}^\tau \mathbf{P}_m\boldsymbol{\alpha} \leq \frac{1}{p}\boldsymbol{\alpha}^\tau \mathbf{A}_m\boldsymbol{\alpha} \longrightarrow C_2^*$ as $p \to \infty$. Let C_2 denote an arbitrary constant with $C_2^* < C_2 < \infty$. There then exists a $p_1 \in \{0, 1, \ldots\}$ with $p_1 \geq p_0$ such that $\frac{1}{p}\boldsymbol{\alpha}^\tau \mathbf{P}_m\boldsymbol{\alpha} \leq C_2$ for all $p \geq p_1$.

Recall that our basic setup implies that $X_1, \ldots, X_n$ are identically distributed random functions with the same distribution as a generic variable X. Expected

values $\mathbb{E}_\varepsilon(\cdot)$ as stated in the theorems below will refer to the probability distribution induced by the random variable ε, that is, they stand for conditional expectation given $X_1, \ldots, X_n$. We assume moreover that ε_i is independent of the X_i's. In the following, for any real positive number x, $[x]$ will denote the smallest integer which is larger than x. In addition, let $\lambda_{x,1} \geq \lambda_{x,2} \geq \cdots \geq \lambda_{x,p} \geq 0$ denote the eigenvalues of the matrix $\frac{1}{np}\mathbf{X}^\tau\mathbf{X}$. We start with a theorem giving finite sample bounds for bias and variance of the estimator $\widehat{\boldsymbol{\alpha}}$ with respect to the semi-norm $\|\cdot\|_{\Gamma_{n,p}}$.

THEOREM 1. *Under assumption* (A.1) *and the above definitions of* C_0, C_1, C_2, p_1, *the following bounds hold for all* $n = 0, 1, \ldots$, *all* $p \geq p_1$, *all* $\rho > n^{-2m}$ *and every* $n \times p$ *matrix* $\mathbf{X} = (X_i(t_j))_{i,j}$:

$$\|\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}}) - \boldsymbol{\alpha}\|^2_{\Gamma_{n,p}} \leq 2\rho\left(\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} + C_1\right) + \frac{4}{n}\sum_{i=1}^n (d_i - \overline{d})^2$$

(3.1)

$$\leq \rho(C_2 + C_1) + \frac{4}{n}\sum_{i=1}^n (d_i - \overline{d})^2,$$

as well as

$$(3.2) \quad \mathbb{E}_\varepsilon\left(\|\widehat{\boldsymbol{\alpha}} - \mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}})\|^2_{\Gamma_{n,p}}\right) \leq \frac{\sigma_\varepsilon^2}{n}\left(m + [\rho^{-1/(2m+2q+1)}](2 + C \cdot C_0)\right),$$

for any $C > 0$ *and* $q \geq 0$ *with the property that* $\sum_{j=k+1}^p \lambda_{x,j} \leq C \cdot k^{-2q}$ *holds for* $k := [\rho^{-1/(2m+2q+1)}]$.

The rate of convergence of $\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|^2_{\Gamma_{n,p}}$ thus depends on assumptions on the distribution of X and on the size of the discretization error. In order to complement our basic setup, we will rely on the following conditions:

(A.2) There exists some constant κ, $0 < \kappa < 1$, such that for every $\delta > 0$, there exists a constant $C_3 < +\infty$ such that

$$\mathbb{P}\left(|X(t) - X(s)| \leq C_3|t - s|^\kappa, t, s \in I\right) \geq 1 - \delta.$$

(A.3) For some constant $C_4 < \infty$ and all $k = 1, 2, \ldots$ there is a k-dimensional linear subspace $\mathcal{L}_k$ of $L^2([0, 1))$ with

$$\mathbb{E}\left(\inf_{f \in \mathcal{L}_k}\sup_t |X(t) - f(t)|^2\right) \leq C_4 k^{-2q}.$$

Before proceeding any further, let us consider assumption (A.3) more closely. The following lemma provides a link between assumption (A.3) and the degree of smoothness of the random functions X_i.

LEMMA 1.   *For some $q_1 = 0, 1, 2, \ldots$ and $0 \le r_2 \le 1$ assume that X is almost surely q_1-times continuously differentiable and that there exists some $C_5 < \infty$ such that*

$$\mathbb{E}\left(\sup_{|t-s| \le d} |X^{(q_1)}(t) - X^{(q_1)}(s)|^2 \right) \le C_5 d^{2r_2}$$

holds for all $d > 0$. There then exists a constant $C_6 < \infty$, depending only on q_1, such that for all $k = 1, 2, \ldots$

$$\mathbb{E}\left(\inf_{f \in \mathcal{E}_k} \sup_t |X(t) - f(t)|^2 \right) \le C_6 C_5 k^{-2(q_1 + r_2)},$$

where $\mathcal{E}_k$ denotes the space of all polynomials of order k on $[0, 1]$.

PROOF.   The well-known Jackson's inequality in approximation theory implies the existence of some $C_6 < \infty$, only depending on q_1, such that for all $k = 1, 2, \ldots$

$$\inf_{f \in \mathcal{E}_k} \sum_{j=1}^p (X(t_j) - f(t_j))^2 \le C_6 k^{-2q_1} \sup_{|t-s| \le 1/k} |X^{(q_1)}(t) - X^{(q_1)}(s)|^2$$

holds with probability 1. The lemma is an immediate consequence.   □

The lemma implies that if assumption (A.2) can be replaced by the stronger requirement $\mathbb{E}(\sup_{|t-s| \le d} |X(t) - X(s)|^2) \le C_5 d^{-2r_2}$, $d > 0$, then assumption (A.3) necessarily holds for some $q \ge \kappa$. Indeed, $q \gg \kappa$ will result from a very high degree of smoothness of X_i.

On the other hand, assumption (A.3) only requires that the functions X_i be well approximated by *some* arbitrary low dimensional linear function spaces (not necessarily polynomials). Even if X_i are *not* smooth, assumption (A.3) may be satisfied for a large value of q (the Brownian motion provides an example).

Theorem 1 together with assumptions (A.2) and (A.3) now allows us to derive rates of convergence of our estimator $\widehat{\alpha}$. First note that assumption (A.3) determines the rate of decrease of the eigenvalues $\lambda_{x,j}$ of $\frac{1}{np} \mathbf{X}^\tau \mathbf{X}$. For any k-dimensional linear space $\mathcal{L}_k \subset L^2([0,1])$, let $\mathcal{P}_k$ denote the corresponding $p \times p$ projection matrix projecting into the k-dimensional subspace $\mathcal{L}_{k,p} = \{v \in \mathbb{R}^p | v = (f(t_1), \ldots, f(t_p))^\tau, f \in \mathcal{L}_k\}$. Basic properties of eigenvalues and eigenvectors then imply that

$$\sum_{j=k+1}^p \lambda_{x,j} \le \inf_{\mathcal{P}_k} \mathrm{Tr}\left((\mathbf{I}_p - \mathcal{P}_k) \frac{1}{np} \mathbf{X}^\tau \mathbf{X} \right)$$

(3.3)

$$= \frac{1}{np} \sum_{i=1}^n \inf_{f \in \mathcal{L}_k} \sum_{j=1}^p (X_i(t_j) - \overline{X} - f(t_j))^2,$$

and assumption (A.3) implies that for any $\delta > 0$ there exists a $C_\delta < \infty$ such that $P(\sum_{j=k+1}^{p} \lambda_{x,j} \leq C_\delta k^{-2q}) \geq 1 - \delta$.

Assumptions (A.1) and (A.2) obviously lead to

$$(3.4) \qquad \frac{1}{n} \sum_{i=1}^{n} (d_i - \overline{d})^2 = O_P(p^{-2\kappa}).$$

If $n, p \to \infty$, $\rho \to 0$, $1/(n\rho) \to 0$, then relations (3.1), (3.2) and (3.3) imply that

$$\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_{\Gamma_{n,p}}^2 = O_P(\rho + (n\rho^{1/(2m+2q+1)})^{-1} + p^{-2\kappa}).$$

In the following we will require that p is sufficiently large compared to n so that the discretization error is negligible. It therefore suffices that $np^{-2\kappa} = O(1)$ as $n, p \to \infty$. This condition imposes a large number p of observation points if κ is small. However, if the functions X_i are smooth enough such that $\kappa = 1$, then $np^{-2\kappa} = O(1)$ is already fulfilled if $\frac{\sqrt{n}}{p} = O(1)$ as $n, p \to \infty$, which does not seem to be restrictive in view of practical applications. The above result then becomes

$$(3.5) \qquad \|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_{\Gamma_{n,p}}^2 = O_P(\rho + (n\rho^{1/(2m+2q+1)})^{-1}).$$

Choosing $\rho \sim n^{-(2m+2q+1)/(2m+2q+2)}$, we can conclude that

$$(3.6) \qquad \|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_{\Gamma_{n,p}}^2 = O_P(n^{-(2m+2q+1)/(2m+2q+2)}).$$

The next theorem studies the behavior of the estimator for the empirical L^2-norm $\|\cdot\|_{\Gamma_n}$. It is shown that if p is sufficiently large compared to n, then based on an optimal choice of ρ, the rate of convergence given in (3.6) generalizes to the semi-norm $\|\cdot\|_{\Gamma_n}$.

THEOREM 2. *Assume* (A.1)–(A.3) *as well as* $np^{-2\kappa} = O(1)$, $\rho \to 0$, $1/(n\rho) \to 0$ *as* $n, p \to \infty$. *Then*

$$(3.7) \qquad \|\widehat{\alpha} - \alpha\|_{\Gamma_n}^2 = O_P(\rho + (n\rho^{1/(2m+2q+1)})^{-1}).$$

We finally investigate in the next theorem the behavior of $\|\widehat{\alpha} - \alpha\|_{\Gamma}^2$. The following assumption describes the additional conditions used to derive our results. It is well known that the covariance operator Γ is a nuclear, self-adjoint and nonnegative Hilbert–Schmidt operator. We will use $\zeta_1, \zeta_2, \ldots$ to denote a complete orthonormal system of eigenfunctions of Γ corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$.

(A.4) There exists a constant $C_7 < \infty$ such that

$$(3.8) \qquad \begin{aligned} &\mathrm{Var}\left(\frac{1}{n} \sum_i \langle X_i - \mathbb{E}(X), \zeta_r \rangle \langle X_i - \mathbb{E}(X), \zeta_s \rangle\right) \\ &\qquad \leq \frac{C_7}{n} \mathbb{E}(\langle X - \mathbb{E}(X), \zeta_r \rangle^2) \mathbb{E}(\langle X - \mathbb{E}(X), \zeta_s \rangle^2) \end{aligned}$$

holds for all n and all $r, s = 1, 2, \ldots$. Moreover, $\|\overline{X} - \mathbb{E}(X))\|^2 = O_P(n^{-1})$.

Relation (3.8) establishes a moment condition. It is necessarily fulfilled if $X_1, \ldots, X_n$ are i.i.d. Gaussian random functions. Then $\langle X_i - \mathbb{E}(X), \zeta_r \rangle \sim N(0, \mathbb{E}(\langle X_i - \mathbb{E}(X), \zeta_r \rangle^2))$, and $\langle X_i - \mathbb{E}(X), \zeta_r \rangle$ is independent of $\langle X_i - \mathbb{E}(X), \zeta_s \rangle$ if $r \neq s$. Relation (3.8) then is an immediate consequence.

However, the validity of (3.8) does not require independence of the functions X_i. For example, in the Gaussian case, (3.8) may also be verified if $\text{Cov}(\langle X_i - \mathbb{E}(X), \zeta_r \rangle \langle X_i - \mathbb{E}(X), \zeta_s \rangle, \langle X_j - \mathbb{E}(X), \zeta_r \rangle \langle X_j - \mathbb{E}(X), \zeta_s \rangle) \leq C_7 \mathbb{E}(\langle X_i - \mathbb{E}(X), \zeta_r \rangle^2) \mathbb{E}(\langle X_i - \mathbb{E}(X), \zeta_s \rangle^2) \cdot q^{|i-j|}$ for some $0 < q < 1$, $C_7 < \infty$ and $i \neq j$. This is of importance in our application to ozone pollution forecasting which deals with a *time series* of functions $X_1, \ldots, X_n$.

THEOREM 3. *Under the conditions of Theorem* 2 *together with assumption* (A.4) *we have*

$$(3.9) \qquad \|\widehat{\alpha} - \alpha\|_\Gamma^2 = O_P\big(\rho + \big(n\rho^{1/(2m+2q+1)}\big)^{-1} + n^{-(2q+1)/2}\big).$$

Furthermore, (1.5) *holds for any random function* X_{n+1} *possessing the same distribution as* X *and independent of* $X_1, \ldots, X_n$.

Theorem 3 shows that if $2q \geq 1$ and $\rho \sim n^{-(2m+2q+1)/(2m+2q+2)}$, then the prediction error can be bounded by

$$\mathbb{E}\big((\widehat{\alpha_0} + \langle \widehat{\alpha}, X_{n+1} \rangle - \alpha_0 - \langle \alpha, X_{n+1} \rangle)^2 | \widehat{\alpha_0}, \widehat{\alpha}\big) = O_P\big(n^{-(2m+2q+1)/(2m+2q+2)}\big).$$

3.2. *Optimality of the rates of convergence.* For simplicity we will rely on the special case of (1.2) with $\alpha_0 = 0$. In this case $\mathbb{E}((\langle \alpha, X_{n+1} \rangle - \widehat{\alpha_0} - \langle \widehat{\alpha}, X_{n+1} \rangle)^2 | \widehat{\alpha_0}, \widehat{\alpha}) \geq \|\widehat{\alpha} - \alpha\|_\Gamma^2$ if X possesses a centered distribution with $\mathbb{E}(X) = 0$. In Proposition 1 below we then show that for suitable Sobolev spaces of functions α and a large class of possible distributions of X_i, the rate $n^{-(2m+2q+1)/(2m+2q+2)}$ is a *lower bound* for the rate of convergence of the prediction error over all estimators of α to be computed from corresponding observations (X_i, Y_i), $i = 1, \ldots, n$. Consequently, the rate attained by our smoothing spline estimator $\widehat{\alpha}$ must be interpreted as a minimax rate over these classes.

We first have to introduce some additional notation. For simplicity, we will assume that the functions $X_i(t)$ are known for all t so that the number p of observation points may be chosen arbitrarily large. We will use $\mathcal{C}_{m,D}$ to denote the space of all m-times continuously differentiable functions α with $\int_0^1 \alpha^{(j)}(t)^2 \, dt \leq D$ for all $j = 0, 1, \ldots, m$. Furthermore, let $\mathcal{P}_{q,C}$ denote the space of all centered probability distributions on $L^2([0, 1])$ with the properties that (a) the sequence of eigenvalues of the corresponding covariance operator satisfies $\sum_{j=k+1}^\infty \lambda_j \leq Ck^{-2q}$ for all sufficiently large k, and that (b) the smoothing spline estimator $\widehat{\alpha}$ satisfies $\|\widehat{\alpha} - \alpha\|_\Gamma^2 = O_P(n^{-(2m+2q+1)/(2m+2q+2)})$ for $\alpha \in \mathcal{C}_{m,D}$ and $\rho \sim n^{-(2m+2q+1)/(2m+2q+2)}$ (whenever p is chosen sufficiently large compared to n).

Finally, for given $\alpha \in \mathcal{C}_{m,D}$, probability distribution $P \in \mathcal{P}_{q,C}$ and i.i.d. random functions $X_1, \ldots, X_n$, $X_i \sim P$, let $\hat{a}(\alpha, P)$ denote an arbitrary estimator of α based on corresponding data (X_i, Y_i), $i = 1, \ldots, n$, generated by (1.2) (with $\alpha_0 = 0$).

PROPOSITION 1. *Let c_n denote an arbitrary sequence of positive numbers with $c_n \to 0$ as $n \to \infty$, and let $2q = 1, 3, 5, \ldots$. Under the above assumptions, we have*

$$\lim_{n \to \infty} \sup_{P \in \mathcal{P}_{q,C}} \sup_{\alpha \in \mathcal{C}_{m,D}} \inf_{\hat{a}(\alpha, P)} \mathbb{P}\big(\|\alpha - \hat{a}(\alpha, P)\|_\Gamma^2 \geq c_n \cdot n^{-(2m+2q+1)/(2m+2q+2)}\big) = 1.$$

It is of interest to compare our results with those of Cai and Hall [4] who analyze the error $\langle \alpha - \hat{\alpha}, x \rangle^2$ for a *fixed* curve x. Similarly to our results, the rate of decrease of the eigenvalues λ_r of Γ plays an important role. Note that, as shown in the proof of Theorem 3, assumption (A.3) yields $\sum_{r=k+1}^{\infty} \lambda_r = O(k^{-2q})$. Since $\lambda_1 \geq \lambda_2 \geq \cdots$ this in turn implies that $\lambda_r = O(r^{-2q-1})$, and one may reasonably assume that $B^{-1}r^{-2q-1} \leq \lambda_r \leq Br^{-2q-1}$ for some $0 < B < \infty$. However, Cai and Hall [4] measure "smoothness" of α in terms of a spectral decomposition $\alpha(t) = \sum_r \alpha_r \zeta_r(t)$ and not with respect to usual smoothness classes. Their quantity of interest is the rate $\beta > 1$ of decrease $|\alpha_r| = O(r^{-\beta})$ as $r \to \infty$. But recall that the error in expanding an m-times continuously differentiable function with respect to k suitable basis functions (as, e.g., orthogonal polynomials or Fourier functions) is of an order of at most k^{-2m}. For the sake of comparison, assume that $\zeta_1, \zeta_2, \ldots$ define an appropriate basis for approximating smooth functions and that $\inf_{f \in \text{span}\{\zeta_1, \ldots, \zeta_k\}} \|\alpha - f\|^2 = \sum_{r=k+1}^{\infty} \alpha_r^2 = O(k^{-2m})$. This will require that $\alpha_r^2 = O(r^{-2m-1})$ and, hence, $2\beta = 2m + 1$.

Results as derived by Cai and Hall [4] additionally depend on the spectral decomposition $x(t) = \sum_r x_r \zeta_r(t)$ of a function x of interest. The essential condition on the structure of the coefficients x_r may be re-expressed in the following form: There exist some $\nu \in \mathbb{R}$ and $0 < D_0 < \infty$ such that $D_0^{-1} r^\nu \leq \frac{x_r^2}{\lambda_r} \leq D_0 r^\nu$ for all $r = 1, 2, \ldots$. Rates of convergence then follow from the magnitude of ν, and it is shown that parametric rates n^{-1} (or $n^{-1} \log n$) are achieved if $\nu \leq -1$.

Now consider a random function X_{n+1} and assume that the underlying distribution is Gaussian. It is then well known that $X_{n+1}(t) = \sum_r x_{n+1,r} \zeta_r(t)$ for independent $N(0, \lambda_r)$-distributed coefficients $x_{n+1,r}$. Consequently, $\frac{x_{n+1,r}^2}{\lambda_r}$ are i.i.d. χ_1^2-distributed variables for all $r = 1, 2, \ldots$, and if $\nu \leq 0$ we obtain $\mathbb{P}(D_0^{-1} r^\nu \leq \frac{x_{n+1,r}^2}{\lambda_r} \leq D_0 r^\nu$ for all $r = 1, 2, \ldots) = 0$ for all $0 < D_0 < \infty$. This already shows that parametric rates n^{-1} cannot be achieved for the error $\langle \alpha - \hat{\alpha}, X_{n+1} \rangle^2$. On the other hand, for arbitrary $\nu > 0$ and $0 < \delta < 1$ we have $\mathbb{P}(D_0^{-1} r^\nu \leq \frac{x_{n+1,r}^2}{\lambda_r} \leq D_0 r^\nu$ for all $r = 1, 2, \ldots) \geq \delta$, whenever D_0 is sufficiently large. If $B^{-1} r^{-2q-1} \leq$

$\lambda_r \leq Br^{-2q-1}$ and $\alpha_r^2 = O_P(r^{-2m+1})$, then for a function x with $D_0^{-1}r^\nu \leq \frac{x_{n+1,r}^2}{\lambda_r} \leq D_0 r^\nu$, $\nu > 0$, the convergence rates of Cai and Hall [4] translate into

$$\langle \widehat{\alpha} - \alpha, x \rangle^2 = O_P\big(n^{-(2m+2q+1-2\nu)/(2m+2q+2)}\big),$$

which provides an additional motivation for the fact that the rates derived in our paper constitute a lower bound. For non-Gaussian distributions a comparison is more difficult, since under assumption (A.4) only the Chebyshev inequality may be used to bound the probabilities $D_0^{-1}r^\nu \leq \frac{x_{n+1,r}^2}{\lambda_r} \leq D_0 r^\nu$.

Another statistically very different problem consists in an optimal estimation of α by $\widehat{\alpha}$ with respect to the usual L^2-norm. In a recent work, Hall and Horowitz [18] derive optimal rates of convergence of $\|\widehat{\alpha} - \alpha\|^2$. These rates again depend on the rate of decrease $|\alpha_r| = O(r^{-\beta})$. Recall that our assumptions do not provide any link between α and X_i; part of the structure of α may not even be identifiable. Indeed, under assumptions (A.1)–(A.4) there is no way to guarantee that the bias $\|\alpha - \mathbb{E}_\varepsilon(\widehat{\alpha})\|^2$ converges to zero and it can only be shown that $\|\widehat{\alpha} - \alpha\|^2 = O_P(1)$ (see the proof of Theorem 2 below). This already highlights the theoretical difference between optimal estimation with respect to $\|\widehat{\alpha} - \alpha\|_\Gamma^2$ and $\|\widehat{\alpha} - \alpha\|^2$. Based on additional assumptions as indicated above, although sensible bounds for the bias may be derived, it must be emphasized that an estimator minimizing $\|\widehat{\alpha} - \alpha\|^2$ will have to rely on $\rho \gg n^{-(2m+2q+1)/(2m+2q+2)}$, which corresponds to an oversmoothing with respect to $\|\widehat{\alpha} - \alpha\|_\Gamma^2$. This effect has already been noted by Cai and Hall [4]. In our context, without additional assumptions linking the eigenvalues of Γ and of the spline matrix $\mathbf{A}_m$, the only general bound for the L_2-variability of the estimator is $\|\widehat{\alpha} - \mathbb{E}_\varepsilon(\widehat{\alpha})\|^2 = O_P(\frac{1}{n\rho})$ (this result may be derived by arguments similar to those used in the proofs of our theorems). With $\rho = n^{-(2m+2q+1)/(2m+2q+2)}$ this leads to $\|\widehat{\alpha} - \mathbb{E}_\varepsilon(\widehat{\alpha})\|^2 = O_P(n^{-1/(2m+2q+2)})$, and better rates may only be achieved with $\rho \gg n^{-(2m+2q+1)/(2m+2q+2)}$. A more detailed study of this problem is not in the scope of the present paper.

3.3. *Choice of smoothing parameters.* The above result of Section 3.1 implies that the choice of the smoothing parameter ρ is of crucial importance. A natural way to determine ρ is to minimize a leave-one-out cross-validation criterion. We preferably adapt the simplified *Generalized Cross-Validation* (GCV) introduced by Wahba [31] in the context of smoothing splines. For fixed m, in our application the GCV criterion takes the form

$$(3.10) \qquad GCV_m(\rho) := \frac{(1/n)\|\mathbf{Y} - \mathbf{H}_\rho \mathbf{Y}\|^2}{(1 - n^{-1}\operatorname{Tr}(\mathbf{H}_\rho))^2},$$

where $\mathbf{H}_\rho := (np)^{-1}\mathbf{X}(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m)^{-1}\mathbf{X}^\tau$.

Proposition 2 below provides a justification for the use of the GCV criterion. Recall that the estimators $\widehat{\boldsymbol{\alpha}} \equiv \widehat{\boldsymbol{\alpha}}_{\rho;m}$ depend on ρ as well as on the spline

order m. Obviously, $\frac{1}{p}\mathbf{X}\widehat{\boldsymbol{\alpha}}_{\rho;m} = \mathbf{H}_\rho \mathbf{Y}$ is an estimator of the conditional mean $(\langle X_1 - \overline{X}, \alpha\rangle, \ldots, \langle X_n - \overline{X}, \alpha\rangle)^\tau$ of $\mathbf{Y}$ given $X_1, \ldots, X_n$. Let

$$ASE_m(\rho) := \frac{1}{n}\sum_i \left[\langle X_i - \overline{X}, \alpha\rangle - \frac{1}{p}\sum_j (X_i(t_j) - \overline{X}(t_j))\hat{\alpha}_{\rho;m}(t_j) \right]^2$$

denote the average squared error of this estimator. The only difference between $ASE_m(\rho)$ and $\|\widehat{\boldsymbol{\alpha}}_\rho - \boldsymbol{\alpha}\|_{\Gamma_{n,p}}^2$ is the discretization error encountered when approximating $\langle X_i, \alpha\rangle$ by $\frac{1}{p}\sum_j X_i(t_j)\alpha(t_j)$, and hence $ASE_m(\rho) = \|\widehat{\boldsymbol{\alpha}}_\rho - \boldsymbol{\alpha}\|_{\Gamma_{n,p}}^2 + O_P(p^{-2\kappa})$.

If $\hat{\rho}$ denotes the minimizer of GCV for fixed m, we can conclude from relation (3.11) of Proposition 2 that the error $ASE_m(\hat{\rho})$ is asymptotically first-order equivalent to the error $ASE_m(\rho_{\text{opt}})$ to be obtained from an optimal choice of the smoothing parameter. Furthermore, (3.12) shows that an analogous result holds if GCV is additionally used to select the order m of the smoothing spline, which means that the optimal rate can be reached adaptively.

PROPOSITION 2. *In addition to assumptions* (A.1)–(A.3) *as well as* $np^{-2\kappa} = O(1)$, *suppose that* $\mathbb{E}(\exp(\beta\varepsilon_i^2)) < \infty$ *for some* $\beta > 0$. *If for fixed* m, $\hat{\rho}$ *denotes the minimizer of* $GCV(\rho)$ *over* $\rho \in [n^{-2m+\delta}, \infty)$ *for some* $\delta > 0$, *then*

$$(3.11) \qquad |ASE_m(\hat{\rho}) - ASE_m(\rho_{\text{opt}})| = O_P\big(n^{-1/2}ASE_m(\rho_{\text{opt}})^{1/2}\big),$$

where ρ_{opt} *minimizes* $MSE_m(\rho) := \mathbb{E}_\varepsilon(ASE_m(\rho))$ *over all* $\rho > 0$.

Furthermore, if $\hat{m}, \hat{\rho}$ *denotes the minimizers of* (3.10) *over* $\rho \in [n^{-2m+\delta}, \infty)$, $\delta > 0$, *and* $m = 1, \ldots, M_n$, $M_n \leq n/2$, *then*

$$(3.12) \qquad |ASE_{\hat{m}}(\hat{\rho}) - ASE_{m_{\text{opt}}}(\rho_{\text{opt}})| = O_P(n^{-1/2}ASE_{m_{\text{opt}}}(\rho_{\text{opt}})^{1/2}\log M_n),$$

where $\rho_{\text{opt}}, m_{\text{opt}}$ *minimize* $MSE_m(\rho) := \mathbb{E}_\varepsilon(ASE_m(\rho))$ *over all* $\rho > 0$ *and* $m = 1, \ldots, M_n$.

4. Case of a noisy covariate. In a number of important applications measurements of the explanatory curves X_i may be contaminated by noise. There then additionally exists an errors-in-variable problem complicating further analysis. Our setup is inspired by other works dealing with noisy observations of functional data (e.g., Cardot [3] or Chiou, Müller and Wang [9]): At each point t_j the corresponding functional value $X_i(t_j)$ is corrupted by some random error δ_{ij} so that actual observations $W_i(t_j)$ are given by

$$(4.1) \qquad W_i(t_j) = X_i(t_j) + \delta_{ij}, \qquad i = 1, \ldots, n, j = 1, \ldots, p,$$

where $(\delta_{ij})_{i=1,\ldots,n, j=1,\ldots,p}$ is a sequence of independent real random variables such that for all $i = 1, \ldots, n$ and all $j = 1, \ldots, p$

$$(4.2) \qquad \mathbb{E}_\varepsilon(\delta_{ij}) = 0, \qquad \mathbb{E}_\varepsilon(\delta_{ij}^2) = \sigma_\delta^2 \quad \text{and} \quad \mathbb{E}_\varepsilon(\delta_{ij}^4) \leq C_8$$

for some constant $C_8 > 0$ (independent of n and p). We furthermore assume that δ_{ij} is independent of ε_i and of the X_i's.

In this situation, an analogue of our estimator $\widehat{\alpha}$ of Section 2 can still be computed by replacing in (2.6) the (unknown) matrix $\mathbf{X}$ by the $n \times p$ matrix $\mathbf{W}$ with general terms $W_i(t_j) - \overline{W}$, $i = 1, \ldots, n$, $j = 1, \ldots, p$. However, performance of the resulting estimator will suffer from the additional noise in the observations. If the error variance σ_δ^2 is large, there may exist a substantial difference between $\mathbf{X}^\tau \mathbf{X}$ and $\mathbf{W}^\tau \mathbf{W}$. Indeed, $\mathbf{W}^\tau \mathbf{W}$ is a biased estimator of $\mathbf{X}^\tau \mathbf{X}$:

$$(4.3) \qquad \frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} = \frac{1}{np^2} \mathbf{X}^\tau \mathbf{X} + \frac{\sigma_\delta^2}{p^2} \mathbf{I}_p + \mathbf{R},$$

where $\mathbf{R}$ is a $p \times p$ matrix such that its largest singular value is of order $O_P(\frac{1}{n^{1/2}p})$, (see the proof of Theorem 4 below). This result suggests that we use $\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} - \frac{\sigma_\delta^2}{p^2} \mathbf{I}_p$ as an approximation of $\frac{1}{np^2} \mathbf{X}^\tau \mathbf{X}$. A prerequisite is, of course, the availability of an estimator $\hat{\sigma}_\delta^2$ of the unknown variance σ_δ^2. Following Gasser, Sroka and Jennen-Steinmetz [16], we will rely on

$$(4.4) \quad \widehat{\sigma}_\delta^2 := \frac{1}{n} \sum_{i=1}^n \frac{1}{6(p-2)} \sum_{j=2}^{p-1} [W_i(t_{j-1}) - W_i(t_j) + W_i(t_{j+1}) - W_i(t_j)]^2.$$

These arguments now lead to the following modified estimator $\widehat{\boldsymbol{\alpha}}_\mathbf{W}$ of $\boldsymbol{\alpha}$ in the case of noisy observations:

$$(4.5) \qquad \widehat{\boldsymbol{\alpha}}_\mathbf{W} := \frac{1}{np} \left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} + \frac{\rho}{p} \mathbf{A}_m - \frac{\widehat{\sigma}_\delta^2}{p^2} \mathbf{I}_p \right)^{-1} \mathbf{W}^\tau \mathbf{Y}.$$

An estimator of the function α is given by $\widehat{\alpha}_\mathbf{W} = s_{\widehat{\alpha}_\mathbf{W}}$, where $s_{\widehat{\alpha}_\mathbf{W}}$ is again the natural spline interpolant of order $2m$ as defined in Section 2.

We want to note that $\widehat{\boldsymbol{\alpha}}_\mathbf{W}$ is closely related to an estimator proposed by Cardot et al. [5]. The latter is motivated by the *Total Least Squares* (TLS) method (see, e.g., Golub and Van Loan [17], Fuller [15], or Van Huffel and Vandewalle [29]) and the only difference from (4.5) consists in the use of a correction term slightly different from $-\frac{\widehat{\sigma}_\delta^2}{p^2} \mathbf{I}_p$.

Of course there are many alternative strategies for dealing with the errors-in-variable problem induced by (4.1). A straightforward approach, which is frequently used in functional data analysis, is to apply nonparametric smoothing procedures in order to obtain estimates $\widehat{X}_i(t_j)$ from the data $(W_i(t_j), t_j)$. When replacing $\mathbf{X}$ by $\widehat{\mathbf{X}}$ in (2.6), one can then define a "smoothed" estimator $\widehat{\boldsymbol{\alpha}}_S$. Of course this estimator may be as efficient as (4.5), but it is computationally more involved and appropriate smoothing parameters have to be selected for nonparametric estimation of each curve X_i.

Our aim is now to study the asymptotic behavior of $\widehat{\alpha}_\mathbf{W}$. Theorem 4 below provides bounds (with respect to the semi-norm $\Gamma_{n,p}$) for the difference between $\widehat{\boldsymbol{\alpha}}_\mathbf{W}$

and the "ideal" estimator $\widehat{\boldsymbol{\alpha}}$ defined for the true curves $X_1, \ldots, X_n$. We will impose the following additional condition on the function α:

(A.5) For every $\delta > 0$ there exists a constant $C_\alpha < \infty$ such that

$$\frac{1}{p^{1/2}}\left\|\frac{1}{np}\mathbf{X}^\tau\mathbf{X}\boldsymbol{\alpha}\right\| > C_\alpha,$$

holds with probability larger or equal to $1 - \delta$.

THEOREM 4. *Assume* (A.1), (A.2), (A.5) *as well as* $np^{-2\kappa} = O(1)$, $\rho \to 0$, $1/(n\rho) \to 0$ *as* $n, p \to \infty$. *Then*

$$(4.6) \qquad \|\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}}\|^2_{\Gamma_{n,p}} = O_P\left(\frac{1}{np\rho} + \frac{1}{n}\right).$$

Together with assumption (A.3) we can therefore conclude from Theorems 1 and 4 that

$$\|\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \boldsymbol{\alpha}\|^2_{\Gamma_{n,p}} = O_P\left(\rho + \left(n\rho^{1/(2m+2q+1)}\right)^{-1} + \frac{1}{np\rho}\right).$$

We have already seen in Section 3 that the optimal order of the two first terms is reached for a choice of $\rho \sim n^{-(2m+2q+1)/(2m+2q+2)}$. From an asymptotic point of view, the use of $\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}$ results in the addition of the extra term $1/(np\rho)$ in the rate of convergence. For $\rho \sim n^{-(2m+2q+1)/(2m+2q+2)}$ we have $1/(np\rho) \sim n^{-1/(2m+2q+2)}/p$. This term is of order $n^{-(2m+2q+1)/(2m+2q+2)}$ for $p \sim n^{(2m+2q-1)/(2m+2q+2)}$. This means that the $\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}$ reaches the same rate of convergence as $\widehat{\boldsymbol{\alpha}}$ provided that p is sufficiently large compared to n. More precisely, it is required that $p \geq C_p \max(n^{1/2\kappa}, n^{(2m+2q-1)/(2m+2q+2)})$ for some positive constant C_p.

As shown in Theorem 5 below, these qualitative results generalize when considering the semi-norms Γ_n or Γ.

THEOREM 5. *Assume* (A.1)–(A.3), (A.5) *as well as* $np^{-2\kappa} = O(1)$, $\rho \to 0$, $1/(n\rho) \to 0$ *as* $n, p \to \infty$. *Then*

$$(4.7) \qquad \|\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}}\|^2_{\Gamma_n} = O_P\left(\frac{1}{np\rho} + \frac{1}{n}\right),$$

and if assumption (A.4) *is additionally satisfied,*

$$(4.8) \qquad \|\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}}\|^2_{\Gamma} = O_P\left(\frac{1}{np\rho} + \frac{1}{n} + n^{-(2q+1)/2}\right).$$

5. Application to ozone pollution forecasting. In this section, our method-
ology is applied to the problem of predicting the level of ozone pollution. For our
analysis, we use a data set collected by ORAMIP (Observatoire Régional de l'Air
en Midi-Pyrénées), an air observatory located in the city of Toulouse (France).
The concentration of specific pollutants as well as meteorological variables are
measured each hour. Some previous studies using the same data are described in
Cardot, Crambes and Sarda [6] and Aneiros-Perez et al. [1].

The response variable Y_i of interest is the maximum of ozone for a day. Re-
peated measurements of ozone concentration obtained for the *preceding* day are
used as a functional explicative variable X_i. More precisely, each X_i is observed at
$p = 24$ equidistant points corresponding to hourly measurements. The sample size
is $n = 474$. It is assumed that the relation between Y_i and X_i can be modeled by
the functional linear regression model (1.2). We note at this point that $X_1, X_2, \ldots$
constitute a time series of functions, and that it is therefore reasonable to suppose
some correlation between the X_i's. The results of an earlier, unpublished study
indicate that there only exists some "short memory" dependence.

Now, for a curve X_{n+1} outside the sample, we want to predict Y_{n+1}, the
maximum of ozone the day after. Assuming that (X_{n+1}, Y_{n+1}) follows the same
model (1.2) and using our estimators $\widehat{\alpha}$ of α and $\widehat{\alpha_0}$ of α_0 described in Section 2,
a predictor $\widehat{Y}_{n+1}$ is given by the formula

$$(5.1) \qquad \widehat{Y}_{n+1} := \widehat{\alpha_0} + \int_I \widehat{\alpha}(t) X_{n+1}(t)\, dt.$$

It cannot be excluded that actual observations of X_i may be contaminated with
noise. We will thus additionally consider the modified estimator $\widehat{\alpha}_{\mathbf{W}}$ developed
in Section 4 and the corresponding predictor $\widehat{Y}_{\mathbf{W},n+1}$. For simplicity, the integral
in (5.1) is approximated by $\frac{1}{p} \sum_{j=1}^{p} \widehat{\alpha}(t_j) X_{n+1}(t_j)$. With additional assumptions
on the ε_i's we can also build asymptotic intervals of prediction for Y_{n+1}. Indeed, let
us assume that $\varepsilon_1, \ldots, \varepsilon_{n+1}$ are i.i.d. random variables having a normal distribution
$\mathcal{N}(0, \sigma_\varepsilon^2)$. The first point is to estimate the residual variance σ_ε^2. A straightforward
estimator is given by the empirical variance

$$(5.2) \qquad \widehat{\sigma_\varepsilon}^2 := \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \overline{Y} - \frac{1}{p} \sum_{j=1}^{p} \widehat{\alpha}(t_j)\big(X_i(t_j) - \overline{X}(t_j)\big) \right)^2.$$

Our theoretical results imply that $\widehat{\sigma_\varepsilon}$ is a consistent estimator of σ_ε^2. Furthermore,
we can then infer from Theorem 3 that $\frac{Y_{n+1} - \widehat{Y}_{n+1}}{\widehat{\sigma_\varepsilon}}$ asymptotically follows a standard
normal distribution. Given $\tau \in]0, 1[$, an asymptotic $(1 - \tau)$-prediction interval
for Y_{n+1} can be derived as

$$(5.3) \qquad [\widehat{Y}_{n+1} - z_{1-\tau/2}\widehat{\sigma_\varepsilon}, \widehat{Y}_{n+1} + z_{1-\tau/2}\widehat{\sigma_\varepsilon}],$$

where $z_{1-\tau/2}$ is the quantile of order $1 - \tau/2$ of the $\mathcal{N}(0, 1)$ distribution. Of course,
the same developments are valid when one replaces $\widehat{Y}_{n+1}$ by $\widehat{Y}_{\mathbf{W},n+1}$.

In order to study performance of our estimators we split the initial sample into two sub-samples:

- A learning sample, $(X_i, Y_i)_{i=1,\ldots,n_l}$, $n_l = 300$, was used to determine the estimators $\widehat{\alpha}$ and $\widehat{\alpha}_{\mathbf{W}}$.
- A test sample, $(X_i, Y_i)_{i=n_l+1,\ldots,n_l+n_t}$, $n_t = 174$, was used to evaluate the quality of the estimation.

Construction of estimators was based on $m = 2$ (cubic smoothing splines), and the smoothing parameters ρ were selected by minimizing $GCV(\rho)$ as defined in (3.10). Note that GCV for $\widehat{\alpha}_{\mathbf{W}}$ requires that the matrix $\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}$ in the definition of $\mathbf{H}_\rho$ has to be replaced by $\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} - \frac{\widehat{\sigma}_\delta^2}{p^2}\mathbf{I}_p$. Figure 1 presents the daily predicted values $\widehat{Y}$ and $\widehat{Y}_{\mathbf{W}}$ of the maximum of ozone versus the measured Y-values of the test sample. Both graphics are close, which is confirmed by the computation of the prediction error given by

$$EQM(\widehat{\alpha}) := \frac{1}{n_t} \sum_{i=n_l+1}^{n_l+n_t} (Y_i - \widehat{Y}_i)^2,$$

with a similar definition for $\widehat{\alpha}_{\mathbf{W}}$. We have, respectively, $EQM(\widehat{\alpha}) = 281.97$ and $EQM(\widehat{\alpha}_{\mathbf{W}}) = 270.13$, which shows a very minor advantage of the estimator $\widehat{\alpha}_{\mathbf{W}}$. In any case, in Figure 1 the points seem to be reasonably spread around the diag-
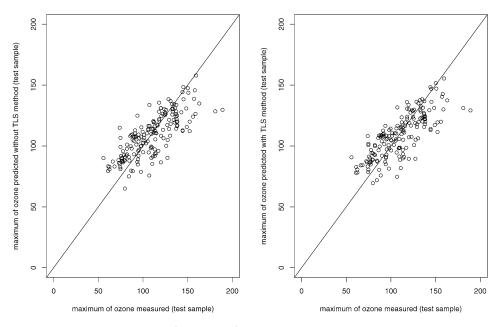

FIG. 1. *Daily predicted values $\widehat{Y}$ (left) and $\widehat{Y}_{\mathbf{W}}$ (right) of the maximum of ozone versus the measured values.*
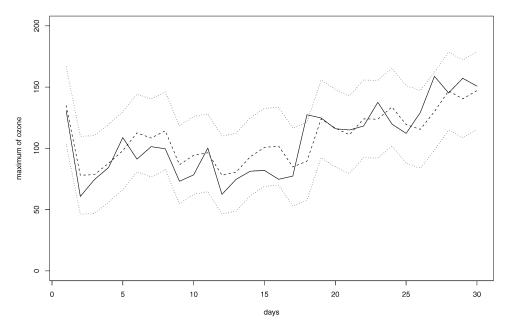
FIG. 2. *Measured values of the maximum of ozone (solid line), predicted values (dashed line) and 95% prediction band (dotted lines).*

onal $\hat{Y} = Y$, and the plots do not indicate any major problem with our estimators. Corresponding prediction intervals are given in Figure 2.

6. Proof of the results.

6.1. *Proof of Theorem* 1. First consider relation (3.1), and note that

$$\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}}) = \frac{1}{np^2}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{X}^\tau\mathbf{X}\boldsymbol{\alpha} + \frac{1}{np}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{X}^\tau\mathbf{d},$$

where $\mathbf{d} = (d_1 - \overline{d}, \dots, d_n - \overline{d})^\tau$.

It follows that $\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}})$ is a solution of the minimization problem

$$\min_{\mathbf{a}\in\mathbb{R}^p}\left\{\frac{1}{n}\left\|\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} + \mathbf{d} - \frac{1}{p}\mathbf{X}\mathbf{a}\right\|^2 + \frac{\rho}{p}\mathbf{a}^\tau\mathbf{A}_m\mathbf{a}\right\}.$$

This implies

$$\frac{1}{n}\left\|\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} + \mathbf{d} - \frac{1}{p}\mathbf{X}\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}})\right\|^2 + \frac{\rho}{p}\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}})^\tau\mathbf{A}_m\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}}) \le \frac{\rho}{p}\boldsymbol{\alpha}^\tau\mathbf{A}_m\boldsymbol{\alpha} + \frac{1}{n}\|\mathbf{d}\|^2.$$

But definition of $\mathbf{A}_m$ and (2.3) lead to

$$\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{A}_m\boldsymbol{\alpha} = \frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} + \int_0^1 s_{\boldsymbol{\alpha}}^{(m)}(t)^2\,dt \le \frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} + \int_0^1 \alpha^{(m)}(t)^2\,dt$$

and (3.1) is an immediate consequence. Let us now consider relation (3.2). There exists a complete orthonormal system of eigenvectors $u_1, u_2, \ldots, u_p$ of $\frac{1}{np}\mathbf{X}^\tau\mathbf{X}$ such that $\frac{1}{np}\mathbf{X}^\tau\mathbf{X} = \sum_{j=1}^{p} \lambda_{x,j} u_j u_j^\tau$. Let $k := [\rho^{-1/(2m+2q+1)}]$. By our assumptions we obtain

$$
\mathbb{E}_\varepsilon\big(\|\widehat{\boldsymbol{\alpha}} - \mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}})\|^2_{\Gamma_{n,p}}\big)
$$

$$
= \frac{1}{p}\mathbb{E}_\varepsilon\bigg(\frac{1}{n^2 p^2}\boldsymbol{\varepsilon}^\tau\mathbf{X}\bigg(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\bigg)^{-1}
$$

$$
\times \frac{1}{np}\mathbf{X}^\tau\mathbf{X}\bigg(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\bigg)^{-1}\mathbf{X}^\tau\boldsymbol{\varepsilon}\bigg)
$$

(6.1)
$$
\leq \frac{\sigma_\varepsilon^2}{n}\operatorname{Tr}\bigg[\bigg(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m\bigg)^{-1}\frac{1}{np}\mathbf{X}^\tau\mathbf{X}\bigg]
$$

$$
= \frac{\sigma_\varepsilon^2}{n}\operatorname{Tr}\bigg[\bigg((\rho\mathbf{A}_m)^{-1/2}\bigg(\frac{1}{np}\mathbf{X}^\tau\mathbf{X}\bigg)(\rho\mathbf{A}_m)^{-1/2} + \mathbf{I}_p\bigg)^{-1}
$$

$$
\times (\rho\mathbf{A}_m)^{-1/2}\bigg(\frac{1}{np}\mathbf{X}^\tau\mathbf{X}\bigg)(\rho\mathbf{A}_m)^{-1/2}\bigg]
$$

$$
\leq \frac{\sigma_\varepsilon^2}{n}\operatorname{Tr}(\mathbf{D}_{1,\rho} + \mathbf{D}_{2,\rho}),
$$

where

$$
\mathbf{D}_{1,\rho} := \bigg((\rho\mathbf{A}_m)^{-1/2}\bigg(\sum_{j=1}^{k}\lambda_{x,j}u_ju_j^\tau\bigg)(\rho\mathbf{A}_m)^{-1/2} + \mathbf{I}_p\bigg)^{-1}
$$

$$
\times (\rho\mathbf{A}_m)^{-1/2}\bigg(\sum_{j=1}^{k}\lambda_{x,j}u_ju_j^\tau\bigg)(\rho\mathbf{A}_m)^{-1/2}
$$

and

$$
\mathbf{D}_{2,\rho} := \bigg((\rho\mathbf{A}_m)^{-1/2}\bigg(\sum_{j=k+1}^{p}\lambda_{x,j}u_ju_j^\tau\bigg)(\rho\mathbf{A}_m)^{-1/2} + \mathbf{I}_p\bigg)^{-1}
$$

$$
\times (\rho\mathbf{A}_m)^{-1/2}\bigg(\sum_{j=k+1}^{p}\lambda_{x,j}u_ju_j^\tau\bigg)(\rho\mathbf{A}_m)^{-1/2}
$$

which are symmetric $p \times p$ matrices with

(6.2)
$$
\sup_{\|\mathbf{v}\|=1}\mathbf{v}^\tau\mathbf{D}_{1,\rho}\mathbf{v} < 1 \quad \text{and} \quad \sup_{\|\mathbf{v}\|=1}\mathbf{v}^\tau\mathbf{D}_{2,\rho}\mathbf{v} < 1.
$$

Furthermore, $\mathbf{D}_{1,\rho}$ is of rank k and therefore only possesses k nonzero eigenvalues. Hence

$$(6.3) \qquad\qquad \mathrm{Tr}(\mathbf{D}_{1,\rho}) \leq k.$$

Let $\mathbf{a}_{1,p}, \ldots, \mathbf{a}_{m,p}, \mathbf{a}_{m+1,p}, \ldots, \mathbf{a}_{p,p}$ denote a complete, orthonormal system of eigenvectors of $\mathbf{A}_m$ corresponding to the eigenvalues $\mu_{1,p} = \cdots = \mu_{m,p} = 1$ and $\mu_{m+1,p} \leq \cdots \leq \mu_{p,p}$. By (6.1), (6.2) and (6.3) as well as (2.7), we thus obtain

$$
\begin{aligned}
&\mathbb{E}_\varepsilon\big(\|\widehat{\boldsymbol{\alpha}} - \mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}})\|_{\Gamma_{n,p}}^2\big) \\
&\qquad \leq \frac{\sigma_\varepsilon^2}{n}\left(k + \sum_{j=1}^p \mathbf{a}_{j,p}^\tau \mathbf{D}_{2,\rho}\mathbf{a}_{j,p}\right) \\
&\qquad \leq \frac{\sigma_\varepsilon^2}{n}\left(k + m + k + \sum_{l=m+k+1}^p \mathbf{a}_{l,p}^\tau(\rho\mathbf{A}_m)^{-1/2}\right. \\
&\qquad\qquad\qquad\qquad\qquad \left. \times \left(\sum_{j=k+1}^p \lambda_{x,j} u_j u_j^\tau\right)(\rho\mathbf{A}_m)^{-1/2}\mathbf{a}_{l,p}\right) \\
&\qquad \leq \frac{\sigma_\varepsilon^2}{n}\left(m + 2k + \frac{1}{\mu_{m+k+1}\cdot\rho}\sum_{j=k+1}^p \lambda_{x,j}\right) \\
&\qquad \leq \frac{\sigma_\varepsilon^2}{n}(m + 2k + CkC_0) \\
&\qquad = \frac{\sigma_\varepsilon^2}{n}(m + [\rho^{-1/(2m+2q+1)}])(2 + CC_0).
\end{aligned}
$$

(6.4)

This proves Relation (3.2) and completes the proof of Theorem 1.

6.2. *Proof of Theorem* 2. With $\widehat{d}_i = \int_I \widehat{\alpha}(t)X_i(t)\,dt - \frac{1}{p}\sum_{j=1}^p \widehat{\alpha}(t_j)X_i(t_j)$ we have

$$
\begin{aligned}
\|\widehat{\alpha} - \alpha\|_{\Gamma_n}^2 &\leq \frac{2}{n}\sum_{i=1}^n\Bigg[\langle(X_i - \overline{X}), \widehat{\alpha} - \alpha\rangle \\
&\qquad\qquad - \frac{1}{p}\sum_{j=1}^p(X_i(t_j) - \overline{X}(t_j))(\widehat{\alpha}(t_j) - \alpha(t_j))\Bigg]^2 \\
&\quad + \frac{2}{n}\sum_{i=1}^n\Bigg[\frac{1}{p}\sum_{j=1}^p(X_i - \overline{X})(t_j)(\widehat{\alpha}(t_j) - \alpha(t_j))\Bigg]^2 \\
&\leq \frac{4}{n}\sum_{i=1}^n(\widehat{d}_i - \overline{\widehat{d}})^2 + \frac{4}{n}\sum_{i=1}^n(d_i - \overline{d})^2 + 2\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_{\Gamma_{n,p}}^2.
\end{aligned}
$$

(6.5)

By assumptions (A.1)–(A.3), it follows from Theorem 1, (3.3) and (3.4) that the assertion of Theorem 2 holds, provided that

$$(6.6) \qquad \frac{1}{n} \sum_{i=1}^{n} (\widehat{d}_i - \overline{\widehat{d}})^2 = O_P(p^{-2\kappa}).$$

The proof of (6.6) consists of several steps. We will start by giving a stochastic bound for $\frac{1}{p}\widehat{\boldsymbol{\alpha}}^\tau \widehat{\boldsymbol{\alpha}}$ and then study the stochastic behavior of $\int_0^1 \widehat{\alpha}^{(m)}(t)^2 \, dt$. The use of a suitable Taylor expansion will then lead to the desired result.

By definition of $\widehat{\boldsymbol{\alpha}}$ we have

$$
\begin{aligned}
\frac{1}{p}\widehat{\boldsymbol{\alpha}}^\tau \widehat{\boldsymbol{\alpha}} \le {} & \frac{3}{p}\boldsymbol{\alpha}^\tau \frac{1}{np}\mathbf{X}^\tau \mathbf{X} \left(\frac{1}{np}\mathbf{X}^\tau \mathbf{X} + \rho \mathbf{A}_m \right)^{-2} \frac{1}{np}\mathbf{X}^\tau \mathbf{X}\boldsymbol{\alpha} \\
(6.7) \qquad & + 3\frac{1}{n^2 p}\mathbf{d}^\tau \mathbf{X} \left(\frac{1}{np}\mathbf{X}^\tau \mathbf{X} + \rho \mathbf{A}_m \right)^{-2} \mathbf{X}^\tau \mathbf{d} \\
& + 3\frac{1}{n^2 p}\boldsymbol{\varepsilon}^\tau \mathbf{X} \left(\frac{1}{np}\mathbf{X}^\tau \mathbf{X} + \rho \mathbf{A}_m \right)^{-2} \mathbf{X}^\tau \boldsymbol{\varepsilon}.
\end{aligned}
$$

Since all eigenvalues of the matrix $\frac{1}{np}\mathbf{X}^\tau \mathbf{X}(\frac{1}{np}\mathbf{X}^\tau \mathbf{X} + \rho \mathbf{A}_m)^{-2}\frac{1}{np}\mathbf{X}^\tau \mathbf{X}$ are less than or equal to 1, the first term on the right-hand side of (6.7) is less than or equal to $\frac{3}{p}\boldsymbol{\alpha}^\tau \boldsymbol{\alpha} = O(1)$. It is easily seen that the smallest eigenvalue of the matrix $\frac{1}{np}\mathbf{X}(\frac{1}{np}\mathbf{X}^\tau \mathbf{X} + \rho \mathbf{A}_m)^{-2}\mathbf{X}^\tau$ is proportional to $1/\rho$, and thus the second term can be bounded by a term of order $p^{-2\kappa}/\rho$. By (2.7) the expected value of the third term is bounded by

$$
\frac{\sigma_\varepsilon^2}{n} \operatorname{Tr}\left[\frac{1}{np}\mathbf{X} \left(\frac{1}{np}\mathbf{X}^\tau \mathbf{X} + \rho \mathbf{A}_m \right)^{-2} \mathbf{X}^\tau \right] \le \frac{\sigma_\varepsilon^2}{n} \operatorname{Tr}[(\rho \mathbf{A}_m)^{-1}] = O\big(1/(n\rho)\big).
$$

We therefore arrive at

$$(6.8) \qquad \frac{1}{p}\widehat{\boldsymbol{\alpha}}^\tau \widehat{\boldsymbol{\alpha}} = O_P\left(1 + \frac{p^{-2\kappa}}{\rho} + \frac{1}{n\rho} \right).$$

As a next step we will study the asymptotic behavior of $\int_0^1 \widehat{\alpha}^{(m)}(t)^2 \, dt$. Since $\widehat{\boldsymbol{\alpha}}$ is solution of the minimization problem (2.5), we can write

$$
\frac{1}{n}\left\| \mathbf{Y} - \frac{1}{p}\mathbf{X}\widehat{\boldsymbol{\alpha}} \right\|^2 + \frac{\rho}{p}\widehat{\boldsymbol{\alpha}}^\tau \mathbf{P}_m \widehat{\boldsymbol{\alpha}} + \rho \int_0^1 \widehat{\alpha}^{(m)}(t)^2 \, dt
$$
$$
\le \frac{1}{n}\left\| \mathbf{Y} - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha} \right\|^2 + \frac{\rho}{p}\boldsymbol{\alpha}^\tau \mathbf{P}_m \boldsymbol{\alpha} + \rho \int_0^1 \alpha^{(m)}(t)^2 \, dt,
$$

and therefore

$$
\begin{aligned}
\rho \int_0^1 \widehat{\alpha}^{(m)}(t)^2 \, dt \le {} & \|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_{\Gamma_{n,p}}^2 + \frac{2}{n}\left\langle \mathbf{Y} - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha}, \frac{1}{p}\mathbf{X}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha} \right\rangle \\
(6.9) \qquad & + \rho \int_0^1 \alpha^{(m)}(t)^2 \, dt - \frac{\rho}{p}\widehat{\boldsymbol{\alpha}}^\tau \mathbf{P}_m \widehat{\boldsymbol{\alpha}} + \frac{\rho}{p}\boldsymbol{\alpha}^\tau \mathbf{P}_m \boldsymbol{\alpha}.
\end{aligned}
$$

We have to focus on the term

$$\frac{2}{n}\Big\langle \mathbf{Y} - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha}, \frac{1}{p}\mathbf{X}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha}\Big\rangle = \frac{2}{n}\Big\langle \mathbf{d} + \boldsymbol{\varepsilon}, \frac{1}{p}\mathbf{X}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha}\Big\rangle.$$

The Cauchy–Schwarz inequality together with the definition of $\|\cdot\|_{\Gamma_{n,p}}^2$ yield

$$(6.10)\qquad \frac{1}{n}\mathbf{d}^\tau\Big(\frac{1}{p}\mathbf{X}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha}\Big) = O_P(p^{-\kappa}\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_{\Gamma_{n,p}}).$$

Note that

$$\frac{2}{n}\Big\langle \boldsymbol{\varepsilon}, \frac{1}{p}\mathbf{X}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha}\Big\rangle$$

$$= \frac{2}{n}\boldsymbol{\varepsilon}^\tau\Big(\frac{1}{p}\mathbf{X}\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}}) - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha}\Big) + \frac{2}{n}\boldsymbol{\varepsilon}^\tau\Big(\frac{1}{p}\mathbf{X}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{X}\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}})\Big).$$

Obviously, $\frac{1}{n}\boldsymbol{\varepsilon}^\tau(\frac{1}{p}\mathbf{X}\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}}) - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha})$ is a zero mean random variable with variance bounded by $\frac{\sigma_\varepsilon^2}{n}\|\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}}) - \boldsymbol{\alpha}\|_{\Gamma_{n,p}}^2$. By definition of $\widehat{\boldsymbol{\alpha}}$, (3.3), (6.1) and (6.4) we have

$$\mathbb{E}_\varepsilon\Big(\frac{1}{n}\boldsymbol{\varepsilon}^\tau\Big(\frac{1}{p}\mathbf{X}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{X}\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}})\Big)\Big) \le \frac{\sigma_\varepsilon^2}{n}\operatorname{Tr}\Big[\Big(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m\Big)^{-1}\frac{1}{np}\mathbf{X}^\tau\mathbf{X}\Big]$$

$$= O_P\Big(\frac{1}{n\rho^{1/(2m+2q+1)}}\Big).$$

We can conclude that

$$(6.11)\quad \frac{2}{n}\Big\langle \boldsymbol{\varepsilon}, \frac{1}{p}\mathbf{X}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{X}\boldsymbol{\alpha}\Big\rangle = O_P\Big(\frac{1}{\sqrt{n}}\|\mathbb{E}_\varepsilon(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\|_{\Gamma_{n,p}} + \frac{1}{n\rho^{1/(2m+2q+1)}}\Big).$$

When combining (6.8), (6.9), (6.10) and (6.11) with the results of Theorem 1 we thus obtain

$$(6.12)\qquad \int_0^1 \widehat{\alpha}^{(m)}(t)^2\,dt = O_P\Big(1 + \frac{p^{-2\kappa}}{\rho} + \frac{1}{n\rho^{(2m+2q+2)/(2m+2q+1)}}\Big).$$

Let us now expand $\widehat{\alpha}$ into a Taylor series: $\widehat{\alpha}(t) = P(t) + R(t)$ for all $t \in [0, 1]$ with

$$P(t) = \sum_{l=0}^{m-1}\frac{t^l}{l!}\widehat{\alpha}^{(l)}(0), \qquad R(t) = \int_0^t r(s)\,ds$$

and

$$r(t) = \int_0^t \frac{(t-u)^{m-1}}{(m-1)!}\widehat{\alpha}^{(m)}(u)\,du.$$

It follows from (6.8) as well as (6.12) that $|\widehat{\alpha}^{(l)}(0)| = O_P(1 + (\frac{p^{-2\kappa}}{\rho})^{1/2} + (\frac{1}{n\rho^{(2m+2q+2)/(2m+2q+1)}})^{1/2})$ for $l = 0, \ldots, m - 1$, and some straightforward calcu-

lations yield

$$\left|\|\widehat{\alpha}\|^2 - \frac{1}{p}\widehat{\boldsymbol{\alpha}}^\tau\widehat{\boldsymbol{\alpha}}\right| = \left|\int_0^1 (P(t) + R(t))^2\, dt - \frac{1}{p}\sum_{j=1}^p (P(t_j) + R(t_j))^2\right|$$

$$\leq \left(\sum_{j=1}^p\left[\int_{t_j-1/(2p)}^{t_j+1/(2p)} (P(t) + R(t) + P(t_j) + R(t_j))^2\, dt\right]^2\right)^{1/2}$$

$$\times \left(\sum_{j=1}^p \frac{1}{p}\left[\int_{t_j-1/(2p)}^{t_j+1/(2p)} |P'(s)| + |r(s)|\, ds\right]^2\right)^{1/2},$$

which leads to

(6.13)
$$\left|\|\widehat{\alpha}\|^2 - \frac{1}{p}\widehat{\boldsymbol{\alpha}}^\tau\widehat{\boldsymbol{\alpha}}\right|$$
$$= O_P\left(p^{-1}\cdot\left(1 + \frac{p^{-2\kappa}}{\rho} + [n\rho^{(2m+2q+2)/(2m+2q+1)}]^{-1}\right)\right).$$

Using again (6.8) and our assumptions on ρ, p, n, this implies

(6.14)
$$\|\widehat{\alpha}\|^2 = O_P(1).$$

At the same time, (6.8) and (6.12) together with assumptions (A.1) and (A.2) imply that with $\widetilde{X}_i = X_i - \overline{X}$

$$\frac{1}{n}\sum_{i=1}^n (\widehat{d}_i - \overline{\widehat{d}})^2 = \frac{1}{n}\sum_{i=1}^n\left(\sum_{j=1}^p\int_{t_j-1/(2p)}^{t_j+1/(2p)} (\widehat{\alpha}(t) - \widehat{\alpha}(t_j))\widetilde{X}_i(t)\right.$$

$$\left. + \widehat{\alpha}(t_j)(\widetilde{X}_i(t) - \widetilde{X}_i(t_j))\, dt\right)^2$$

$$\leq 2x_{\max}^2\left(\sum_{j=1}^p\frac{1}{p}\left[\int_{t_j-1/(2p)}^{t_j+1/(2p)} |P'(t)| + |r(t)|\, dt\right]^2\right)$$

$$+ 2\left(\frac{1}{p}\sum_{j=1}^p\widehat{\alpha}(t_j)^2\right)\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^p\int_{t_j-1/(2p)}^{t_j+1/(2p)} (\widetilde{X}_i(t) - \widetilde{X}(t_j))^2\, dt$$

and thus

(6.15)
$$\frac{1}{n}\sum_{i=1}^n (\widehat{d}_i - \overline{\widehat{d}})^2 = O_P\left(p^{-2}\left(1 + \frac{p^{-2\kappa}}{\rho} + \frac{1}{n\rho^{(2m+2q+2)/(2m+2q+1)}}\right)\right.$$
$$\left. + p^{-2\kappa}\left(1 + \frac{p^{-2\kappa}}{\rho} + \frac{1}{n\rho}\right)\right).$$

By our assumptions on ρ, p, n, relation (6.6) is an immediate consequence. This completes the proof of Theorem 2.

6.3. *Proof of Theorem* 3. In terms of eigenvalues and eigenfunctions of Γ we obviously obtain

$$\langle \Gamma u, u \rangle = \sum_r \lambda_r \langle \zeta_r, u \rangle^2.$$

Let $\tau_{ri} = \langle X_i - \mathbb{E}(X), \zeta_r \rangle$ for $r = 1, 2, \ldots$ and $i = 1, \ldots, n$. Some well-known results of stochastic process theory now can be summarized as follows:

(i) $\mathbb{E}(\tau_{ri}) = 0$, $\mathbb{E}(\tau_{ri}^2) = \lambda_r$, and $\mathbb{E}(\tau_{ri}\tau_{si}) = 0$ for all r, s, $s \neq r$ and $i = 1, \ldots, n$.

(ii) For any $k = 1, 2, \ldots$, the eigenfunctions $\zeta_1, \ldots, \zeta_k$ corresponding to $\lambda_1 \geq \cdots \geq \lambda_k$ provide a best basis for approximating X_i by a k-dimensional linear space:

$$\sum_{r=q+1}^{\infty} \lambda_r = \mathbb{E}\left(\left\| X - \mathbb{E}(X) - \sum_{s=1}^{q} \langle X - \mathbb{E}(X), \zeta_s \rangle \zeta_s \right\|^2\right)$$

(6.16)

$$\leq \mathbb{E}\left(\inf_{f \in \mathcal{L}_k} \| X - \mathbb{E}(X) - f \|^2\right),$$

for any other k-dimensional linear subspace $\mathcal{L}_k$ of $L^2([0, 1])$.

By (A.3) we can conclude that

(6.17)
$$\sum_{r=k+1}^{\infty} \lambda_r = O(k^{-2q}) \qquad \text{as } k \to \infty.$$

At first we have

$$\|\widehat{\alpha} - \alpha\|_{\Gamma_n}^2 \leq \frac{2}{n} \sum_{i=1}^{n} \langle \widehat{\alpha} - \alpha, X_i - \mathbb{E}(X) \rangle^2 + \frac{2}{n} \sum_{i=1}^{n} \langle \widehat{\alpha} - \alpha, \mathbb{E}(X) - \overline{X} \rangle^2,$$

and by (6.14) and with assumption (A.4) the last term is of order $O_P(n^{-1})$. The relevant semi-norms can now be rewritten in the form

(6.18)
$$\|\widehat{\alpha} - \alpha\|_{\Gamma}^2 = \sum_{r=1}^{\infty} \lambda_r \langle \zeta_r, \widehat{\alpha} - \alpha \rangle^2 =: \sum_{r=1}^{\infty} \lambda_r \widetilde{\alpha}_r^2$$

and

$$\|\widehat{\alpha} - \alpha\|_{\Gamma_n}^2 = \|\widehat{\alpha} - \alpha\|_{\Gamma}^2 + \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \widetilde{\alpha}_r \widetilde{\alpha}_s \left(\frac{1}{n} \sum_{i=1}^{n} \tau_{ri}\tau_{si} - \lambda_r I(r = s)\right)$$

(6.19)

$$+ O_P(n^{-1}),$$

where $I(r = s) = 1$ if $r = s$, and $I(r = s) = 0$ if $r \neq s$. Define

$$\widetilde{\tau}_{rr} = \frac{1}{\lambda_r \sqrt{n}} \sum_{i=1}^{n} (\tau_{ri}^2 - \lambda_r) \quad \text{and} \quad \widetilde{\tau}_{rs} = \frac{1}{\sqrt{\lambda_r \lambda_s} n} \sum_{i=1}^{n} \tau_{ri}\tau_{si}, \; r \neq s$$

(with $\widetilde{\tau}_{rs} := 0$ if $\min\{\lambda_r, \lambda_s\} = 0$). The properties of τ_{ri} given in (i) imply that $\mathbb{E}(\widetilde{\tau}_{rs}) = 0$ for all r, s, and we can infer from assumption (A.4) that for some $C_{10} < \infty$

$$(6.20) \qquad \qquad \mathbb{E}(\widetilde{\tau}_{rs}^2) \le C_{10},$$

holds for all $r, s = 1, 2, \ldots$ and all sufficiently large n. Using the Cauchy–Schwarz inequality we therefore obtain for all $k = 0, 1, \ldots$

$$
\begin{aligned}
&\left| \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \widetilde{\alpha}_r \widetilde{\alpha}_s \left(\frac{1}{n} \sum_{i=1}^{n} \tau_{ri} \tau_{si} - \lambda_r I(r = s) \right) \right| \\
&\quad = \left| \frac{1}{\sqrt{n}} \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \widetilde{\alpha}_r \widetilde{\alpha}_s (\lambda_r \lambda_s)^{1/2} \widetilde{\tau}_{rs} \right| \\
&\quad \le \frac{2}{\sqrt{n}} \left(\sum_{r=1}^{k} \sum_{s=r}^{\infty} \lambda_r \widetilde{\alpha}_r^2 \widetilde{\alpha}_s^2 \right)^{1/2} \left(\sum_{r=1}^{k} \sum_{s=r}^{\infty} \lambda_s \widetilde{\tau}_{rs}^2 \right)^{1/2} \\
&\qquad + \frac{2}{\sqrt{n}} \left(\sum_{r=k+1}^{\infty} \sum_{s=r}^{\infty} \widetilde{\alpha}_r^2 \widetilde{\alpha}_s^2 \right)^{1/2} \left(\sum_{r=k+1}^{\infty} \sum_{s=r}^{\infty} \lambda_r \lambda_s \widetilde{\tau}_{rs}^2 \right)^{1/2}.
\end{aligned}
$$

(6.21)

Relation (6.14) leads to $\|\widehat{\alpha} - \alpha\|^2 \ge \sum_{r=1}^{\infty} \widetilde{\alpha}_r^2 = O_P(1)$, which together with (6.18) implies that for arbitrary k

$$
\left(\sum_{r=1}^{k} \sum_{s=r}^{\infty} \lambda_r \widetilde{\alpha}_r^2 \widetilde{\alpha}_s^2 \right)^{1/2} \le \left(\left(\sum_{r=1}^{\infty} \lambda_r \widetilde{\alpha}_r^2 \right) \left(\sum_{s=1}^{\infty} \widetilde{\alpha}_s^2 \right) \right)^{1/2} = O_P(\|\widehat{\alpha} - \alpha\|_\Gamma).
$$

Choose k proportional to $n^{1/2}$. Relation (6.17) then yields $\sum_{r=k+1}^{\infty} \sum_{s=r}^{\infty} \lambda_r \lambda_s \le (\sum_{r=k+1}^{\infty} \lambda_r)^2 = O(n^{-2q})$ and $\sum_{r=1}^{k} \sum_{s=r}^{\infty} \lambda_s = O(\max\{\log n, n^{(1-2q)/2}\})$. Since by (6.20) the moments of $\widetilde{\tau}_{rs}$ are uniformly bounded for all r, s, it follows that

$$
\left(\sum_{r=1}^{k} \sum_{s=r}^{\infty} \lambda_s \widetilde{\tau}_{rs}^2 \right)^{1/2} = O_P(\max\{\log n, n^{(1-2q)/4}\}),
$$

$$
\left(\sum_{r=k+1}^{\infty} \sum_{s=r}^{\infty} \lambda_r \lambda_s \widetilde{\tau}_{rs}^2 \right)^{1/2} = O_P(n^{-q}).
$$

When combining these results we can conclude that

$$
\begin{aligned}
&\left| \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \widetilde{\alpha}_r \widetilde{\alpha}_s \left(\frac{1}{n} \sum_{i=1}^{n} \tau_{ri} \tau_{si} - \lambda_r I(r = s) \right) \right| \\
&\quad = O_P\left(\max\{ n^{-1/2} \log n \cdot \|\widehat{\alpha} - \alpha\|_\Gamma, n^{-(2q+1)/4} \cdot \|\widehat{\alpha} - \alpha\|_\Gamma, n^{-(2q+1)/2} \} \right).
\end{aligned}
$$

Together with (6.19) assertion (3.9) now follows from the rates of convergence of $\|\widehat{\alpha} - \alpha\|^2_{\Gamma_n}$ derived in Theorem 2.

It remains to prove (1.5). Note that by our assumptions on ε_i and assumption (A.4) we have $|\mathbb{E}(Y) - \overline{Y}|^2 \leq 2\overline{\varepsilon}^2 + 2\langle \alpha, \mathbb{E}(X) - \overline{X} \rangle^2 = O_P(n^{-1})$. Together with (6.14) and assumption (A.4) this implies

$$\big|\mathbb{E}((\widehat{\alpha_0} + \langle \widehat{\alpha}, X_{n+1} \rangle - \alpha_0 - \langle \alpha, X_{n+1} \rangle)^2 | \widehat{\alpha_0}, \widehat{\alpha}) - \|\widehat{\alpha} - \alpha\|^2_{\Gamma}\big|$$
$$\leq 2|\mathbb{E}(Y) - \overline{Y}|^2 + 2\langle \widehat{\alpha}, \mathbb{E}(X) - \overline{X} \rangle^2 = O_P(n^{-1}),$$

which completes the proof of the theorem.

6.4. *Proof of Proposition* 1. In dependence of q we first construct special probability distributions of X_i. For $2q = 1$, $\tau \in [0, 1]$ and $r := 0$ set $\widetilde{X}_{\tau;0}(t) := 1$ for $t \in [0, \tau]$ and $\widetilde{X}_{\tau,0}(t) := 0$ for $t \in (\tau, 1]$. For $2q \geq 3$, $\tau \in [0, 1]$, and $r := q - 0.5$ let $\widetilde{X}_{\tau;r}(t) := \frac{1}{r!}t^r$ for $t \in [0, \tau]$ and $\widetilde{X}_{\tau;r}(t) := \sum_{j=0}^{r-1} \frac{1}{(r-j)!}\tau^{r-j}(t - \tau)^j$ for $t \in (\tau, 1]$.

For $k = 1, 2, \ldots$ let $\mathcal{L}_{(r+1)k}$ denote the $(r + 1) \cdot k$ dimensional linear space of all functions g_β of the form $g_\beta(t) := \sum_{j=0}^{k-1}(\sum_{l=0}^{r} \beta_{l,j}t^l) \cdot I(t \in [\frac{j}{k}, \frac{j+1}{k}])$. It is then easily verified that $\sup_{t \in [j/k,(j+1)/k]} \min_\beta |g_\beta(t) - \widetilde{X}_{\tau;r}(t)| = 0$ if $\tau \notin [\frac{j}{k}, \frac{j+1}{k}]$, while $\sup_{t \in [j/k,(j+1)/k]} \min_\beta |g_\beta(t) - \widetilde{X}_{\tau;r}(t)| \leq k^{-r}$ if $\tau \in [\frac{j}{k}, \frac{j+1}{k}]$. It follows that there exist constants $B_r \leq 1$ such that the functions $B_r \widetilde{X}_{\tau;r}(t)$ satisfy $\inf_{g_\beta \in \mathcal{L}_{(r+1)k}} \int_0^1 (B_r \widetilde{X}_{\tau;r}(t) - g_\beta(t))^2 dt \leq C(r + 2)^{-(2r+1)}k^{-(2r+1)} = C(r + 2)^{-2q}k^{-2q}$ for all $k = 1, 2, \ldots$.

Now let $\tau_1, \ldots, \tau_n$ denote i.i.d. real random variables which are uniformly distributed on $[0, 1]$ and let $X_{\tau_i;r} = B_r \widetilde{X}_{\tau_i;r}(t) - \mathbb{E}(B_r \widetilde{X}_{\tau_i;r}(t))$. Obviously, $\tau_i \rightarrow X_{\tau_i,r}^{(j)}(t)$ is a continuous mapping from $[0, 1]$ on $L^2([0, 1])$, and the probability distribution of τ_i induces a corresponding centered probability distribution P_r on $L^2([0, 1])$. Since the eigenfunctions of the corresponding covariance operator provide a best basis for approximating X_i by a k-dimensional linear space, we obtain from what is done above

$$\sum_{j=k+1}^{\infty} \lambda_j \leq \mathbb{E}\bigg(\inf_{g_\beta^* \in \mathcal{L}_{(r+1)[k/(r+1)]}^*} \|X_{\tau_i;r} - g_\beta^*\|^2\bigg) \leq Ck^{-2q},$$

for all sufficiently large k and $\mathcal{L}_{(r+1)k}^* := \{g_\beta - \mathbb{E}(B_r \widetilde{X}_{\tau_i;r})|g_\beta \in \mathcal{L}_{(r+1)k}\}$.

In order to verify that $P_r \in \mathcal{P}_{q,C}$, it remains to check the behavior of $\|\widehat{\alpha} - \alpha\|_\Gamma = \int_0^1 \langle X_{\tau;r}, \widehat{\alpha} - \alpha \rangle^2 d\tau$. First note that although assumption (A.2) does not hold for $2q = 1$, even in this case, with $\kappa = 1/2$, relation (3.4) holds and arguments in the proof of Theorems 1 and 2 imply that for sufficiently large p, $\frac{1}{n} \sum_{i=1}^{n} \langle X_{\tau_i;r}, \widehat{\alpha} -$

$\alpha\rangle^2 = O_P(n^{-(2m+2q+1)/(2m+2q+2)})$. For some $1 > \delta > \frac{2m+2q+1}{2m+2q+2}$ define a partition of $[0, 1]$ into n^δ disjoint intervals $I_1, \ldots, I_{n^\delta}$ of equal length $n^{-\delta}$. For $j = 1, \ldots, n^\delta$, let s_j denote the midpoint of the interval I_j, and use n_j denote the (random) number of $\tau_1, \ldots, \tau_n$ falling into I_j. By using the Cauchy–Schwarz inequality as well as a definition of $X_{\tau;r}$ it is easily verified that there exists a constant $L_r < \infty$ such that $|\langle X_{\tau;r}, \widehat{\alpha} - \alpha\rangle - \langle X_{\tau^*;r}, \widehat{\alpha} - \alpha\rangle| \le L_r |\tau - \tau^*|^{1/2} \|\widehat{\alpha} - \alpha\|$ for $\tau, \tau^* \in [0, 1]$ ($|\tau - \tau^*|^{1/2}$ may be replaced by $|\tau - \tau^*|$ if $2q > 1$). Then

$$|\langle X_{\tau;r}, \widehat{\alpha} - \alpha\rangle^2 - \langle X_{\tau^*;r}, \widehat{\alpha} - \alpha\rangle^2|$$
$$\le 2L_r |\tau - \tau^*|^{1/2} \|\widehat{\alpha} - \alpha\| \min\{|\langle X_{\tau;r}, \widehat{\alpha} - \alpha\rangle|, |\langle X_{\tau^*;r}, \widehat{\alpha} - \alpha\rangle|\}$$
$$+ L_r^2 |\tau - \tau^*| \|\widehat{\alpha} - \alpha\|^2.$$

By (6.14) another application of the Cauchy–Schwarz inequality leads to $\frac{1}{n} \sum_{i=1}^n \langle X_{\tau_i;r}, \widehat{\alpha} - \alpha\rangle^2 = \frac{1}{n} \sum_{j=1}^{n^\delta} n_j \langle X_{s_j;r}, \widehat{\alpha} - \alpha\rangle^2 + o_P(n^{-(2m+2q+1)/(2m+2q+2)})$. Since $\sup_{j=1,\ldots,n^\delta} \frac{|n_j - \mathbb{E}(n_j)|}{n_j} = O_P(1)$ with $\mathbb{E}(n_j) = n \cdot n^{-\delta}$, we can conclude that $\frac{1}{n} \sum_{j=1}^{n^\delta} \mathbb{E}(n_j) \langle X_{s_j;r}, \widehat{\alpha} - \alpha\rangle^2 = O_P(n^{-(2m+2q+1)/(2m+2q+2)})$. Finally,

$$\left| \int_0^1 \langle X_{\tau;r}, \widehat{\alpha} - \alpha\rangle^2 \, d\tau - \frac{1}{n} \sum_{j=1}^{n^\delta} \mathbb{E}(n_j) \langle X_{s_j;r}, \widehat{\alpha} - \alpha\rangle^2 \right|$$

$$\le \frac{1}{n^\delta} \sum_{j=1}^{n^\delta} \sup_{\tau \in I_j} |\langle X_{\tau;r}, \widehat{\alpha} - \alpha\rangle^2 - \langle X_{s_j;r}, \widehat{\alpha} - \alpha\rangle^2|$$

$$= o_P(n^{-(2m+2q+1)/(2m+2q+2)}),$$

and the desired result $\|\widehat{\alpha} - \alpha\|_\Gamma = O_P(n^{-(2m+2q+1)/(2m+2q+2)})$ is an immediate consequence. Therefore, $P_r \in \mathcal{P}_{q,C}$.

We now have to consider the functionals $\langle X_{\tau_i;r}, \alpha\rangle$ more closely. Let $\mathcal{C}^*(m + r + 1, D)$ denote the space of all $m + r + 1$-times continuously differentiable functions $\widetilde{\alpha}$ satisfying $\int_0^1 \widetilde{\alpha}(t) \, dt = 0$ as well as $\int_0^1 \widetilde{\alpha}^{(j)}(t)^2 \, dt \le D$ for all $j = 0, 1, \ldots, m + r + 1$ as well as $\widetilde{\alpha}^{(j)}(0) = \widetilde{\alpha}^{(j)}(1) = 0$ for all $j = 0, \ldots, r + 1$, and set $\mathcal{C}^*(m, r, D) = \{\alpha | \alpha = \widetilde{\alpha}^{(r+1)}, \widetilde{\alpha} \in \mathcal{C}^*(m + r + 1, D)\}$. Then, for any $\alpha \in \mathcal{C}^*(m, 0, D)$ there is a $\widetilde{\alpha} \in \mathcal{C}^*(m + 1, D)$ such that

$$\langle X_{\tau_i;0}, \alpha\rangle = B_0 \int_0^{\tau_i} \alpha(t) \, dt - \langle \mathbb{E}(B_0 \widetilde{X}_{\tau_i;0}), \alpha\rangle$$

$$= B_0 \widetilde{\alpha}(\tau_i) - B_0 \int_0^1 \widetilde{\alpha}(t) \, dt = B_0 \widetilde{\alpha}(\tau_i)$$

while for any $\alpha \in \mathcal{C}^*(m, r, D)$, $r \ge 1$ and $\widetilde{\alpha} \in \mathcal{C}^*(m + r + 1, D)$, $\alpha = \widetilde{\alpha}^{(r+1)}$,

partial integration leads to

$$\langle X_{\tau_i;r}, \alpha \rangle = (-1)^{r-1} \langle X_{\tau_i;r}^{(r-1)}, \widetilde{\alpha}^{(2)} \rangle$$

$$= \big(X_{\tau_i;r}^{(r-1)}(\tau_i) \widetilde{\alpha}^{(2)}(\tau_i) - X_{\tau_i;r}^{(r-1)}(0) \widetilde{\alpha}^{(2)}(0) \big)$$

$$+ B_r(-1)^r \int_0^{\tau_i} \widetilde{\alpha}^{(1)}(t)\, dt - B_r(-1)^r \mathbb{E}\bigg(\int_0^{\tau_i} \widetilde{\alpha}^{(1)}(t)\, dt \bigg)$$

$$+ \big(X_{\tau_i;r}^{(r-1)}(1) \widetilde{\alpha}^{(2)}(1) - X_{\tau_i;r}^{(r-1)}(\tau_i) \widetilde{\alpha}^{(2)}(\tau_i) \big)$$

$$= B_r(-1)^r \widetilde{\alpha}(\tau_i) - \mathbb{E}(B_r(-1)^r \widetilde{\alpha}(\tau_i)) = B_r(-1)^r \widetilde{\alpha}(\tau_i).$$

Obviously, $\widetilde{\alpha}^* = B_r(-1)^r \widetilde{\alpha} \in \mathcal{C}^*(m + r + 1, B_r D)$. By construction, with $f_a(\tau_i) := \langle X_{\tau_i,r}, a \rangle$ we generally obtain

$$\| \alpha - \hat{a}(\alpha, P_\beta) \|_\Gamma^2 = \int_0^1 \big(f_\alpha(\tau) - f_{\hat{a}(\alpha, P_r)}(\tau) \big)^2 d\tau.$$

By definition, $f_\alpha(\tau) = \widetilde{\alpha}^*(\tau) = \mathbb{E}(Y_i | \tau_i = \tau)$ is the regression function in the regression model $Y_i = \widetilde{\alpha}^*(\tau_i) + \varepsilon_i$, and we will use the notation $S_n(\widetilde{\alpha}^*)$ to denote an estimator of $\widetilde{\alpha}^*$ from the data $(Y_i, \tau_i), \ldots, (Y_n, \tau_n)$. Note that knowledge of (Y_i, τ_i) is equivalent to knowledge of $(Y_i, X_{\tau_i;r})$, and an estimator $f_{\hat{a}(\alpha, P_r)}$ of $\widetilde{\alpha}^*$ can thus be seen as a particular estimator $S_n(\widetilde{\alpha}^*)$ based on $(Y_i, \tau_i), \ldots, (Y_n, \tau_n)$. We can conclude that as $n \to \infty$,

$$\sup_{P \in \mathcal{P}_{q,C}} \sup_{\alpha \in \mathcal{C}_{m,D}} \inf_{\hat{a}(\alpha, P)} \mathbb{P}\big(\| \alpha - \hat{a}(\alpha, P) \|_\Gamma^2$$

$$\geq c_n \cdot n^{-(2m+2q+1)/(2m+2q+2)} \big)$$

$$\geq \sup_{\widetilde{\alpha}^* \in C^*(m+r+1, B_r D)} \inf_{S_n(\widetilde{\alpha}^*)} \mathbb{P}\bigg(\int_0^1 \big(\widetilde{\alpha}^*(\tau) - S_n(\widetilde{\alpha}^*)(\tau) \big)^2 d\tau$$

$$\geq c_n \cdot n^{-(2m+2q+1)/(2m+2q+2)} \bigg) \to 1.$$

Convergence of the last probability to 1 follows from well-known results on optimal rates of convergence in nonparametric regression (cf. Stone [27]).

6.5. *Proof of Proposition* 2. We first consider (3.11). The set $\{\mathbf{H}_\rho\}_{\rho>0}$ constitutes an ordered linear smoother according to the definition in Kneip [20]. Theorem 1 of Kneip [20] then implies that $|MSE_m(\hat{\rho}^*) - MSE_m(\rho_{\text{opt}})| = O_P(n^{-1/2} \times MSE_m(\rho_{\text{opt}})^{1/2})$, where $\hat{\rho}^*$ is determined by minimizing Mallow's C_L, $C_L(\rho) := \frac{1}{n} \| \mathbf{Y} - \mathbf{H}_\rho \mathbf{Y} \|^2 + \frac{2\sigma_\varepsilon^2}{n} \text{Tr}(\mathbf{H}_\rho)$. Note that although we consider centered values $Y_i - \overline{Y}$ instead of Y_i all arguments in Kneip [20] apply, since $(\overline{Y}, \ldots, \overline{Y})^\tau \mathbf{X} = 0$. The arguments used in the proof of Theorem 1 of Kneip ([20], relations (A.17)–(A.22))

imply that for all ρ the difference $C_L(\rho) - C_L(\rho_{\text{opt}}) - (MSE_m(\rho) - MSE_m(\rho_{\text{opt}}))$ can be bounded by exponential inequalities given in Lemma 3 of Kneip [20] [the squared norm $q_\mu(\mathbf{H}_\rho, \mathbf{H}_{\rho_{\text{opt}}})^2$ appearing in these inequalities can be bounded by $2MSE_m(\rho)$]. These results lead to

$$
\begin{aligned}
C_L(\rho) - C_L(\rho_{\text{opt}}) &= MSE_m(\rho) - MSE_m(\rho_{\text{opt}}) \\
&\quad + \eta^{[1]}_{\rho;m} n^{-1/2} MSE_m(\rho)^{1/2},
\end{aligned}
\tag{6.22}
$$

$$
\begin{aligned}
ASE_m(\rho) - ASE_m(\rho_{\text{opt}}) &= MSE_m(\rho) - MSE_m(\rho_{\text{opt}}) \\
&\quad + \eta^{[2]}_{\rho;m} n^{-1/2} MSE_m(\rho)^{1/2},
\end{aligned}
\tag{6.23}
$$

$$
\frac{1}{n}\|\mathbf{Y} - \mathbf{H}_\rho \mathbf{Y}\|^2 = \sigma_\varepsilon^2 + MSE_m(\rho_{\text{opt}}) + \eta^{[3]}_{\rho;m} n^{-1/2},
\tag{6.24}
$$

where $\eta^{[s]}_{\rho;m}$ are random variables satisfying $\sup_{\rho>0} |\eta^{[s]}_{\rho;m}| = O_P(1)$, $s = 1, 2, 3$. By our assumptions and the arguments used in the proof of Theorem 1 we can infer that $n^{-1} \text{Tr}(\mathbf{H}_\rho) = O_P([n\rho^{1/(2m+2q+1)}]^{-1}) = o_P(1)$ for all $\rho \in [n^{-2m+\delta}, \infty)$ as $n \to \infty$. Furthermore, there exists a constant $D < \infty$ such that $n^{-1} \text{Tr}(\mathbf{H}_\rho) \leq D \cdot MSE_m(\rho) = O_P(\rho + [n\rho^{1/(2m+2q+1)}]^{-1})$. Together with (6.24) a Taylor expansion of $GCV_m(\rho)$ with respect to $n^{-1} \text{Tr}(\mathbf{H}_\rho)$ then yields

$$
\begin{aligned}
GCV_m(\rho) &= \frac{1}{n}\|\mathbf{Y} - \mathbf{H}_\rho \mathbf{Y}\|^2 + 2\frac{1}{n}\|\mathbf{Y} - \mathbf{H}_\rho \mathbf{Y}\|^2 \frac{\text{Tr}(\mathbf{H}_\rho)}{n} \\
&\quad + \eta^{[4]}_{\rho;m} \left(\frac{\text{Tr}(\mathbf{H}_\rho)}{n}\right)^2 \\
&= C_L(\rho) + \eta^{[5]}_{\rho;m}\left(n^{-12} + MSE_m(\rho)\right)\frac{\text{Tr}(\mathbf{H}_\rho)}{n},
\end{aligned}
\tag{6.25}
$$

where again $\eta^{[s]}_{\rho;m}$ are random variables with $\sup_{\rho>n^{-2m+\delta}} |\eta^{[s]}_{\rho;m}| = O_P(1)$, $s = 4, 5$. Together with $MSE_m(\rho_{\text{opt}}) = O_P(n^{-2m+2q+1/(2m+2q+2)})$, Relation (3.11) now is an immediate consequence of (6.22)–(6.25).

Since Lemma 3 of Kneip [20] provides exponential inequalities, it is easily verified that uniform bounds similar to (6.22)–(6.25) hold for all $\rho \in [n^{-2m+\delta}, \infty)$ and **all** $m = 1, \ldots, M_n$, if $\eta^{[s]}_{\rho;m}$ are replaced by $\tilde{\eta}^{[s]}_{\rho;m} \cdot \log M_n$, $s = 1, \ldots, 5$. Then $\sup_{\rho>n^{-2m+\delta}, m=1,\ldots,M_n} |\tilde{\eta}^{[s]}_{\rho;m}| = O_P(1)$, $s = 1, \ldots, 5$. The proof of (3.12) then follows the arguments used above.

6.6. *Proof of Theorem 4.* Consider the following decomposition:

$$
\widehat{\boldsymbol{\alpha}}_\mathbf{W} - \widehat{\boldsymbol{\alpha}} = \left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\frac{1}{np}\boldsymbol{\delta}^\tau\mathbf{Y} + \mathbf{S}\left[\frac{1}{np}\mathbf{W}^\tau\mathbf{Y}\right],
$$

where

$$\mathbf{S} := \left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m + \mathbf{T}\right)^{-1} - \left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1},$$

$$\mathbf{T} := \mathbf{R} - \frac{\widehat{\sigma}_\delta^2 - \sigma_\delta^2}{p^2}\mathbf{I}_p$$

and where $\boldsymbol{\delta}$ is the $n \times p$ matrix with generic element $\delta_{ij} - \overline{\delta}_j$, $i = 1, \ldots, n$, $j = 1, \ldots, p$ and the matrix $\mathbf{R}$ is defined in (4.3). Thus one obtains

$$\|\widehat{\boldsymbol{\alpha}}_\mathbf{W} - \widehat{\boldsymbol{\alpha}}\|_{\Gamma_{n,p}} \leq \left\|\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\frac{1}{np}\boldsymbol{\delta}^\tau\mathbf{Y}\right\|_{\Gamma_{n,p}}$$

(6.26)

$$+ \left\|\mathbf{S}\left(\frac{1}{np}\mathbf{W}^\tau\mathbf{Y}\right)\right\|_{\Gamma_{n,p}}.$$

Note that $\mathbb{E}_\varepsilon((\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m)^{-1}\frac{1}{np}\boldsymbol{\delta}^\tau\mathbf{Y}) = 0$, whereas with assumptions (A.1) and (A.2)

$$\mathbb{E}_\varepsilon\left(\left\|\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\frac{1}{np}\boldsymbol{\delta}^\tau\mathbf{Y}\right\|_{\Gamma_{n,p}}^2\right)$$

$$= \mathbb{E}_\varepsilon\left(\frac{1}{n^2p}\mathbf{Y}^\tau\boldsymbol{\delta}\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m\right)^{-1}\frac{1}{np}\mathbf{X}^\tau\mathbf{X}\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m\right)^{-1}\boldsymbol{\delta}^\tau\mathbf{Y}\right)$$

$$= O_P\left(\frac{\sigma_\delta^2}{np}\text{Tr}\left(\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m\right)^{-1}\right)\right).$$

This leads with the properties of the eigenvalues of $(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m)^{-1}$ to

(6.27) $$\left\|\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\frac{1}{np}\boldsymbol{\delta}^\tau\mathbf{Y}\right\|_{\Gamma_{n,p}} = O_P\left(\frac{1}{(np\rho)^{1/2}}\right).$$

The next step consists in studying the behavior of the matrix $\mathbf{R}$ defined in (4.3). Its generic term is $R_{r,s} = \frac{1}{np^2}\sum_{i=1}^n (X_i(t_r) - \overline{X}(t_r))(\delta_{is} - \overline{\delta}_s) + (X_i(t_s) - \overline{X}(t_s))(\delta_{ir} - \overline{\delta}_r) + (\delta_{ir} - \overline{\delta}_r)(\delta_{is} - \overline{\delta}_s) - \sigma_\delta^2 I[r = s]$, for $r, s = 1, \ldots, p$, so that for any $\mathbf{u} \in \mathbb{R}^p$ such that $\|\mathbf{u}\| = 1$ one has $\|\mathbb{E}_\varepsilon(\mathbf{Ru})\| = O_P(\frac{1}{np^2})$ whereas it is easy to see that with assumptions (A.1) and (A.2) and (4.2), $\mathbb{E}_\varepsilon(\|\mathbf{Ru}\|^2) = O_P(\frac{1}{np^2})$ and then $\|\mathbf{R}\| = O_P(\frac{1}{n^{1/2}p})$. Now to derive an upper bound for the norm of the matrix $\mathbf{T}$, we use the convergence result given in Gasser, Sroka and Jennen-Steinmetz [16] which in our framework implies that $\widehat{\sigma}_\delta^2 = \sigma_\delta^2 + O_P(\frac{1}{n^{1/2}p})$. Together with the order of $\|\mathbf{R}\|$ this yields

(6.28) $$\|\mathbf{T}\| = O_P\left(\frac{1}{n^{1/2}p}\right).$$

For the second term in (6.26) we consider at first its Frobenius norm. We have

$$\left\| \mathbf{S}\left(\frac{1}{np}\mathbf{W}^\tau \mathbf{Y}\right)\right\|_F$$

$$\leq \frac{1}{p^{1/2}}\left\|\left[\left(\frac{1}{np^2}\mathbf{X}^\tau \mathbf{X} + \frac{\rho}{p}\mathbf{A}_m + \mathbf{T}\right)^{-1} - \left(\frac{1}{np^2}\mathbf{X}^\tau \mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\right]\right.$$

$$\left. \times \left(\frac{1}{n^2 p^2}\mathbf{W}^\tau \mathbf{Y}\mathbf{Y}^\tau \mathbf{W}\right)^{1/2}\right\|_F$$

$$\leq \frac{1}{p^{1/2}}\left\|\left(\frac{1}{np^2}\mathbf{X}^\tau \mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\frac{1}{np}\mathbf{W}^\tau \mathbf{Y}\right\|^2 \|\mathbf{T}\|\left\|\frac{1}{np}\mathbf{W}^\tau \mathbf{Y}\right\|^{-1},$$

where the second inequality comes from the first inequality in Demmel [11]. Note that with assumptions (A.2) and (A.5), for every $\delta > 0$, there is a positive constant such that $p^{1/2}\|\mathbb{E}_\varepsilon(\frac{1}{np}\mathbf{W}^\tau \mathbf{Y})\|$ is greater than this constant with a probability larger than or equal to $1 - \delta$. We also have $\mathbb{E}_\varepsilon(\|\frac{1}{np}\mathbf{W}^\tau \mathbf{Y} - \mathbb{E}_\varepsilon(\frac{1}{np}\mathbf{W}^\tau \mathbf{Y})\|^2)$, which is of order $\frac{1}{np}$. This gives finally when combining (6.8), (6.28) and the condition on p and ρ as well as assumption (A.2)

$$(6.29) \qquad \left\|\mathbf{S}\frac{1}{np}\mathbf{W}^\tau \mathbf{Y}\right\|_{\Gamma_{n,p}}^2 = O_P\left(\left\|\mathbf{S}\left(\frac{1}{np}\mathbf{W}^\tau \mathbf{Y}\right)\right\|_F\right) = O_P\left(\frac{1}{n}\right),$$

which concludes Theorem 4 with (6.26) and (6.27).

6.7. *Proof of Theorem* 5. We first prove (4.7). Obviously,

$$\|\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\alpha}\|_{\Gamma_n}^2 \leq \frac{2}{n}\sum_{i=1}^n (\widehat{d}_{i,\mathbf{W}} - \overline{\widehat{d}}_{\mathbf{W}})^2 + 2\|\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}}\|_{\Gamma_{n,p}}^2,$$

where

$$\widehat{d}_{i,\mathbf{W}} = \int_I (\widehat{\alpha}_{\mathbf{W}}(t) - \widehat{\alpha}(t))X_i(t)\,dt - \frac{1}{p}\sum_{j=1}^p (\widehat{\alpha}_{\mathbf{W}}(t_j) - \widehat{\alpha}(t_j))X_i(t_j).$$

Then, assertion (4.6) implies that (4.7) is a consequence of

$$(6.30) \qquad \frac{1}{n}\sum_{i=1}^n (\widehat{d}_{i,\mathbf{W}} - \overline{\widehat{d}}_{\mathbf{W}})^2 = O_P\left(\frac{1}{np\rho} + \frac{1}{n}\right).$$

The proof of (6.30) follows the same structure as the proof of (6.6). Indeed, we

have

$$\frac{1}{n}\sum_{i=1}^{n}(\widehat{d}_{i,\mathbf{W}} - \overline{\widehat{d}}_{\mathbf{W}})^2$$

(6.31)
$$\leq 2x_{\max}^2\left(\sum_{j=1}^{p}\frac{1}{p}\left[\int_{t_j-1/(2p)}^{t_j+1/(2p)}|P'(t)| + |P'_{\mathbf{W}}| + |r(t)| + |r_{\mathbf{W}}|\,dt\right]^2\right)$$

$$+ 2\frac{1}{p}\|\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}}\|^2$$

$$\times \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\int_{t_j-1/(2p)}^{t_j+1/(2p)}\left((X_i(t) - \overline{X}(t)) - (X_i(t_j) - \overline{X}(t_j))\right)^2\,dt,$$

where $P_{\mathbf{W}}(t) = \sum_{l=0}^{m-1}\frac{t^l}{l!}\widehat{\alpha}_{\mathbf{W}}(0)$, $r_{\mathbf{W}}(t) = \int_0^t\frac{(t-u)^{m-1}}{(m-1)!}\widehat{\alpha}_{\mathbf{W}}(u)\,du$ and $P(t)$ and $r(t)$ are similarly defined for $\widehat{\alpha}$ (see the proof of Theorem 2).

Replacing the semi-norm $\Gamma_{n,p}$ by the euclidean norm in (4.6) following the same lines as the proof of Theorem 4, one can show that

(6.32) $$\frac{1}{p}\|\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}}\|^2 = \frac{1}{p}(\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}})^\tau(\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}}) = O_P\left(\frac{1}{np\rho^2} + \frac{1}{n}\right),$$

which together with assumption (A.2) implies that the second term on the right-hand side of (6.31) can be bounded by $O_P(\frac{p^{-2\kappa}}{np\rho^2} + \frac{p^{-2\kappa}}{n})$.

Now the remainder of the proof consists in studying $\int_0^1\widehat{\alpha}_{\mathbf{W}}^{(m)}(t)^2\,dt$. Recalling the definition of $\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}$, we have

$$\frac{1}{n}\left\|\mathbf{Y} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}\right\|^2 + \frac{\rho}{p}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}^\tau\mathbf{P}_m\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} + \rho\int_I\widehat{\alpha}_{\mathbf{W}}^{(m)}(t)^2\,dt - \frac{\widehat{\sigma}_\delta}{p^2}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}^\tau\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}$$

$$\leq \frac{1}{n}\left\|\mathbf{Y} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}\right\|^2 + \frac{\rho}{p}\widehat{\boldsymbol{\alpha}}^\tau\mathbf{P}_m\widehat{\boldsymbol{\alpha}} + \rho\int_I\widehat{\alpha}^{(m)}(t)^2\,dt - \frac{\widehat{\sigma}_\delta}{p^2}\widehat{\boldsymbol{\alpha}}^\tau\widehat{\boldsymbol{\alpha}}$$

and then

(6.33)
$$\rho\int_I\widehat{\alpha}_{\mathbf{W}}^{(m)}(t)^2\,dt$$

$$\leq \frac{1}{n}\left\|\frac{1}{p}\mathbf{W}(\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}})\right\|^2 + \frac{2}{n}\left\langle\mathbf{Y} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}, \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}\right\rangle$$

$$- \frac{\rho}{p}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}^\tau\mathbf{P}_m\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} + \frac{\rho}{p}\widehat{\boldsymbol{\alpha}}^\tau\mathbf{P}_m\widehat{\boldsymbol{\alpha}}$$

$$+ \frac{\widehat{\sigma}_\delta}{p^2}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}^\tau\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \frac{\widehat{\sigma}_\delta}{p^2}\widehat{\boldsymbol{\alpha}}^\tau\widehat{\boldsymbol{\alpha}} + \rho\int_I\widehat{\alpha}^{(m)}(t)^2\,dt.$$

First consider the term $\frac{1}{n}\|\frac{1}{p}\mathbf{W}(\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}})\|^2$. By (4.6) and (6.32) we obtain

$$(6.34) \qquad \frac{1}{n}\left\|\frac{1}{p}\mathbf{W}(\widehat{\boldsymbol{\alpha}}_{\mathbf{W}} - \widehat{\boldsymbol{\alpha}})\right\|^2 = O_P\left(\frac{1}{np\rho} + \frac{1}{n}\right).$$

We focus now on the second term in the right-hand side of (6.33), for which we have the following decomposition:

$$\frac{1}{n}\left\langle \mathbf{Y} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}, \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}\right\rangle$$

$$= \frac{1}{n}\left\langle \frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}, \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}\right\rangle$$

$$+ \frac{1}{n}\left\langle \mathbf{d}, \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}\right\rangle + \frac{1}{n}\left\langle \boldsymbol{\varepsilon}, \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}\right\rangle.$$

We have

$$\frac{1}{n^{1/2}}\left\|\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}\right\|$$

$$\leq \frac{1}{n^{1/2}}\left\|\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}} - \mathbb{E}_{\varepsilon}\left(\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}\right)\right\|$$

$$+ \frac{1}{n^{1/2}}\left\|\mathbb{E}_{\varepsilon}\left(\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}\right)\right\|.$$

Some straightforward calculations and previous results lead to $\frac{1}{n^{1/2}}\|\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}} - \mathbb{E}_{\varepsilon}(\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}})\| = O_P((1/n\rho^{1/(2m+2q+1)})^{1/2} + 1/p^{1/2})$ whereas $\|\mathbb{E}_{\varepsilon}(\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}})\| = O_P(\rho^{1/2} + p^{-\kappa})$. This finally leads with the Cauchy–Schwarz inequality to

$$\frac{1}{n}\left\langle \frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}, \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}\right\rangle$$

$$(6.35) \qquad = O_P\left(\left(\left(\frac{1}{n\rho^{1/2m+2q+1}}\right)^{1/2} + \frac{1}{p^{1/2}} + \rho^{1/2} + p^{-\kappa}\right)\right.$$

$$\left. \times \left(\frac{1}{(np\rho)^{1/2}} + \frac{1}{n^{1/2}}\right)\right).$$

Using again the Cauchy–Schwarz inequality and (6.34) we have

$$(6.36) \qquad \frac{1}{n}\left\langle \mathbf{d}, \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}} - \frac{1}{p}\mathbf{W}\widehat{\boldsymbol{\alpha}}_{\mathbf{W}}\right\rangle = O_P\left(\frac{p^{-\kappa}}{(np\rho)^{1/2}} + \frac{p^{-\kappa}}{n^{1/2}}\right).$$

The last term is such that

$$\frac{1}{n}\boldsymbol{\varepsilon}^{\tau}\left(\frac{1}{p}\mathbf{W}(\widehat{\boldsymbol{\alpha}}-\widehat{\boldsymbol{\alpha}}_{\mathbf{W}})\right)$$

$$=\frac{1}{n}\boldsymbol{\varepsilon}^{\tau}\left(\frac{1}{p}\mathbf{W}\left(\frac{1}{np^2}\mathbf{X}^{\tau}\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\boldsymbol{\delta}^{\tau}\mathbf{Y}\right)+\frac{1}{n}\boldsymbol{\varepsilon}^{\tau}\left(\frac{1}{p}\mathbf{W}\mathbf{S}\left(\frac{1}{np}\mathbf{W}^{\tau}\mathbf{Y}\right)\right).$$

Using the same developments as above and using assumptions (A.1) and (A.2) we obtain that $\frac{1}{n}\boldsymbol{\varepsilon}^{\tau}(\frac{1}{p}\mathbf{W}(\frac{1}{np^2}\mathbf{X}^{\tau}\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m)^{-1}\boldsymbol{\delta}^{\tau}\mathbf{Y})=O_P(\frac{1}{np^{1/2}\rho^{1/2}})$ while $\frac{1}{n}\boldsymbol{\varepsilon}^{\tau}(\frac{1}{p}\mathbf{W}\mathbf{S}\times(\frac{1}{np}\mathbf{W}^{\tau}\mathbf{Y}))=O_P(\frac{1}{n})$. This finally leads to

$$(6.37)\qquad\frac{1}{n}\boldsymbol{\varepsilon}^{\tau}\left(\frac{1}{p}\mathbf{W}(\widehat{\boldsymbol{\alpha}}-\widehat{\boldsymbol{\alpha}}_{\mathbf{W}})\right)=O_P\left(\frac{1}{np^{1/2}\rho^{1/2}}+\frac{1}{n}\right).$$

Finally using the same arguments as in the proof of Theorem 2, assertion (6.30) is a consequence of (6.31), (6.8) and (6.12) as well as the bounds obtained in (6.32)–(6.37) and the conditions on n, p and ρ.

It remains to show (4.8). The proof follows the same lines as the proof of Theorem 3. We have the following relation:

$$\|\widehat{\alpha}_{\mathbf{W}}-\widehat{\alpha}\|_{\Gamma_n}^2$$

$$=\|\widehat{\alpha}_{\mathbf{W}}-\widehat{\alpha}\|_{\Gamma}^2+\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\widetilde{\alpha}_{\mathbf{W},r}\widetilde{\alpha}_{\mathbf{W},s}\left(\frac{1}{n}\sum_{i=1}^{n}\tau_{ri}\tau_{si}-\lambda_r I(r=s)\right)+O_P(n^{-1}),$$

with $\widetilde{\alpha}_{\mathbf{W},r}=\langle\zeta_r,\widehat{\alpha}_{\mathbf{W}}-\widehat{\alpha}\rangle$. Using the Cauchy–Schwarz inequality as in (6.21), the remainder of the proof consists in showing that $\|\widehat{\alpha}_{\mathbf{W}}-\widehat{\alpha}\|=O_P(1)$. This is obtained by using the bounds obtained in the proof of (4.7) and following the same lines of argument as for showing (6.8).

REFERENCES

[1] ANEIROS-PEREZ, G., CARDOT, H., ESTEVEZ-PEREZ, G. and VIEU, P. (2004). Maximum ozone concentration forecasting by functional nonparametric approaches. *Environmetrics* **15** 675–685.

[2] BOSQ, D. (2000). *Linear Processes in Function Spaces. Lecture Notes in Statist.* **149**. Springer, New York. MR1783138

[3] CARDOT, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J. Nonparametr. Statist.* **12** 503–538. MR1785396

[4] CAI, T. T. and HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496

[5] CARDOT, H., CRAMBES, C., KNEIP, A. and SARDA, P. (2007). Smoothing splines estimators in functional linear regression with errors-in-variables. *Comput. Statist. Data Anal.* **51** 4832–4848. MR2364543

[6] CARDOT, H., CRAMBES, C. and SARDA, P. (2007). Ozone pollution forecasting. In *Statistical Methods for Biostatistics and Related Fields* (W. Härdle, Y. Mori and P. Vieu, eds.) 221–244. Springer, New York. MR2376412

[7] CARDOT, H., FERRATY, F. and SARDA, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13** 571–591. MR1997162

[8] CARDOT, H., MAS, A. and SARDA, P. (2007). CLT in functional linear regression models. *Probab. Theory Related Fields* **138** 325–361. MR2299711

[9] CHIOU, J. M., MÜLLER, H. G. and WANG, J. L. (2003). Functional quasi-likelihood regression models with smoothed random effects. *J. Roy. Statist. Soc. Ser. B* **65** 405–423. MR1983755

[10] CUEVAS, A., FEBRERO, M. and FRAIMAN, R. (2002). Linear functional regression: The case of a fixed design and functional response. *Canadian J. Statistics* **30** 285–300. MR1926066

[11] DEMMEL, J. (1992). The componentwise distance to the nearest singular matrix. *SIAM J. Matrix Anal. Appl.* **13** 10–19. MR1146648

[12] EILERS, P. H. and MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statist. Sci.* **11** 89–102. MR1435485

[13] EUBANK, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. Dekker, New York. MR0934016

[14] FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations*. Springer, London. MR2229687

[15] FULLER, W. A. (1987). *Measurement Error Models*. Wiley, New York. MR0898653

[16] GASSER, T., SROKA, L. and JENNEN-STEINMETZ, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **3** 625–633. MR0897854

[17] GOLUB, G. H. and VAN LOAN, C. F. (1980). An analysis of the total least squares problem. *SIAM J. Numer. Anal.* **17** 883–893. MR0595451

[18] HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* To appear. MR2332269

[19] HE, G., MÜLLER, H.-G. and WANG, J. L. (2000). Extending correlation and regression from multivariate to functional data. In *Asymptotics in Statistics and Probability* (M. L. Puri, ed.) 301–315. VSP, Leiden.

[20] KNEIP, A. (1994). Ordered linear smoothers. *Ann. Statist.* **22** 835–866. MR1292543

[21] LI, Y. and HSING, T. (2006). On rates of convergence in functional linear regression. *J. Mulitivariate Anal.* Published online DOI: 10.1016/j.jmva.2006.10.004. MR2392433

[22] MARX, B. D. and EILERS, P. H. (1999). Generalized linear regression on sampled signals and curves: A *P*-spline approach. *Technometrics* **41** 1–13.

[23] MÜLLER, H.-G. and STADTMÜLLER, U. (2005). Generalized functional linear models. *Annn. Statist.* **33** 774–805. MR2163159

[24] RAMSAY, J. O. and DALZELL, C. J. (1991). Some tools for functional data analysis. *J. Roy. Statist. Soc. Ser. B* **53** 539–572. MR1125714

[25] RAMSAY, J. O. and SILVERMAN, B. W. (2002). *Applied Functional Data Analysis*. Springer, New York. MR1910407

[26] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Applied Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

[27] STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. MR0673642

[28] UTRERAS, F. (1983). Natural spline functions, their associated eigenvalue problem. *Numer. Math.* **42** 107–117. MR0716477

[29] VAN HUFFEL, S. and VANDEWALLE, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*. SIAM, Philadelphia. MR1118607

[30] WAHBA, G. (1977). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.* **14** 651–667. MR0471299

[31] WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia. MR1045442

[32] YAO, F., MÜLLER, H.-G. and WANG, J. L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

C. CRAMBES
P. SARDA
UNIVERSITÉ PAUL SABATIER
INSTITUT DE MATHÉMATIQUES
UMR 5219
LABORATOIRE DE STATISTIQUE ET PROBABILITÉS
118 ROUTE DE NARBONNE
31062 TOULOUSE CEDEX
FRANCE
E-MAIL: Christophe.Crambes@math.ups-tlse.fr
         Pascal.Sarda@math.ups-tlse.fr

A. KNEIP
STATISTISCHE ABTEILUNG
DEPARTMENT OF ECONOMICS AND HAUSDORFF
CENTER FOR MATHEMATICS
UNIVERSITÄT BONN
ADENAUERALLEE 24-26
53113 BONN
GERMANY
E-MAIL: akneip@uni-bonn.de