Electronic Journal of Statistics
Vol. 7 (2013) 191–216
ISSN: 1935-7524
DOI: 10.1214/13-EJS767

On rate optimal local estimation in functional linear regression

Jan Johannes and Rudolf Schenk

Institut de statistique, biostatistique et sciences actuarielles (ISBA), Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium
e-mail: jan.johannes@uclouvain.be; rudolf.schenk@uclouvain.be

Abstract: We consider the estimation of the value of a linear functional of the slope parameter in functional linear regression, where scalar responses are modeled in dependence of random functions. The theory in this paper covers in particular point-wise estimation as well as the estimation of weighted averages of the slope parameter. We propose a plug-in estimator which is based on a dimension reduction technique and additional thresholding. It is shown that this estimator is consistent under mild assumptions. We derive a lower bound for the maximal mean squared error of any estimator over a certain ellipsoid of slope parameters and a certain class of covariance operators associated with the regressor. It is shown that the proposed estimator attains this lower bound up to a constant and hence it is minimax optimal. Our results are appropriate to discuss a wide range of possible regressors, slope parameters and functionals. They are illustrated by considering the point-wise estimation of the slope parameter or its derivatives and its average value over a given interval.

AMS 2000 subject classifications: Primary 62J05; secondary 62G05, 62J20.
Keywords and phrases: Linear functional, linear Galerkin projection, minimax-theory, point-wise estimation, Sobolev space, thresholding.

Received December 2011.

Contents

1. Introduction

A common problem in functional regression is to investigate the dependence of a real random variable Y on the variation of an explanatory random function X. It is usually assumed that the regressor X takes its values in a separable Hilbert space $\mathbb{H}$ which is endowed with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|_{\mathbb{H}}$. For convenience, the regressor X is often supposed to be centered in the sense that for all $f \in \mathbb{H}$ the real valued random variable $\langle X, f \rangle$ has mean zero. In this paper, the dependence of Y on X is supposed to be linear, that is

$$Y = \langle \beta, X \rangle + \sigma\varepsilon, \quad \sigma > 0, \tag{1.1}$$

with an unknown slope parameter $\beta \in \mathbb{H}$ and a centered and standardized error term ε. We focus on the estimation of the value of a known linear functional of the slope β, which we denote by $\ell(\beta)$. The non-parametric estimation of the value of a linear functional from Gaussian white noise observations is subject of considerable literature (in case of direct observations see Speckman [28], Li [19] or Ibragimov and Has'minskii [15], while in case of indirect observations we refer to Donoho and Low [8], Donoho [7] or Goldenshluger and Pereverzev [11] and references therein). In the literature, the most studied examples for estimating linear functionals are point-wise estimation of β and the estimation of (possibly weighted) averages over a subinterval of its domain. These examples are particular cases of our general setting. The objective of this paper is to establish a minimax theory for the non-parametric estimation of the value of a linear functional of the slope parameter β in the functional linear model as considered in (1.1), which in general does not lead to Gaussian white noise observations. For this purpose we use a plug-in estimator $\widehat{\ell}_m := \ell(\widehat{\beta}_m)$ based on an estimator $\widehat{\beta}_m$ of the slope parameter that has been proposed by Cardot and Johannes [4] and is inspired by the linear Galerkin approach coming from the inverse problem community (c.f. Efromovich and Koltchinskii [9] or Hoffmann and Reiß [14]). In recent years, the non-parametric estimation of the slope function β from an independent and identically distributed (i.i.d.) sample of (Y, X) has been of growing interest in the literature. For example, Bosq [1], Cardot et al. [5] or Müller and Stadtmüller [22] analyze a functional principal components regression, while a penalized least squares approach combined with projection onto some basis (such as splines) is studied in Ramsay and Dalzell [27], Eilers and Marx [10], Cardot et al. [3], Hall and Horowitz [12] or Crambes et al. [6]. All the proposed estimators of β have in common that they achieve under reasonable assumptions only very poor rates of convergence. In other words, even relatively large sample sizes may not be much of a help for estimating the slope parameter accurately as a whole. The reason for these poor convergence rates is intrinsic to the considered model as it leads in a natural way to an ill-posed inverse problem. To be more precise, as considered for example in Bosq [1], Cardot et al. [3] or Cardot et al. [5], we suppose that the regressor X has a finite second moment, i.e., $\mathbb{E}\|X\|_{\mathbb{H}}^2 < \infty$, and that X is uncorrelated to the random error ε in the sense that $\mathbb{E}[\varepsilon\langle X, f\rangle] = 0$ for all $f \in \mathbb{H}$. Multiplying both sides in (1.1) by $\langle X, f\rangle$ and taking the expectation leads to the continuous equivalent of the normal equation in a classical multivariate linear model. That is, we have for all $f \in \mathbb{H}$

$$\langle g, f\rangle := \mathbb{E}[Y\langle X, f\rangle] = \mathbb{E}[\langle \beta, X\rangle\langle X, f\rangle] =: \langle \Gamma\beta, f\rangle, \tag{1.2}$$

where g belongs to $\mathbb{H}$ and Γ denotes the covariance operator associated with the random function X. In what follows we always assume that there exists a unique solution $\beta \in \mathbb{H}$ of equation (1.2), i.e., that Γ is strictly positive and that g belongs to its range (for a detailed discussion we refer to Cardot et al. [3]). Obviously, these conditions ensure as well that the value of a linear functional of β is identified. Since the estimation of β involves the inversion of the covariance operator Γ it is called an inverse problem. Moreover, due to the finite second moment of the regressor X, the associated covariance operator Γ is nuclear, i.e., its trace is finite. Consequently, unlike in a multivariate linear model, a continuous generalized inverse of Γ does not exist as long as the range of Γ is an infinite dimensional subspace of $\mathbb{H}$. Therefore, the reconstruction of β is ill-posed (with the additional difficulty that Γ is unknown and has to be estimated). As usual in the context of ill-posed inverse problems we impose additional conditions on the unknown slope parameter β and the covariance operator Γ which will be expressed in the form $\beta \in \mathcal{F}$ and $\Gamma \in \mathcal{G}$, for suitably chosen classes $\mathcal{F} \subseteq \mathbb{H}$ and $\mathcal{G}$. The class $\mathcal{F}$ reflects prior information on the solution β, e.g., its level of smoothness, and will be constructed flexibly enough to characterize, in particular, differentiable or analytic functions. The class $\mathcal{G}$ links the mapping properties of the operator Γ to the regularity conditions imposed on the slope function β. Typically, the assumption $\Gamma \in \mathcal{G}$ results in conditions on the decay of the eigenvalues of the operator Γ. The construction of the class $\mathcal{G}$ allows us to discuss both a polynomial and exponential decay of those eigenvalues. It is interesting to note that Cai and Hall [2] also consider the estimation of a linear functional in functional linear regression. However, their results are restricted to differentiable slope parameters β and polynomially decreasing eigenvalues of the operator Γ. Moreover, the restrictions imposed on the linear functional ℓ by Cai and Hall [2] implicitly exclude the particularly interesting case of point-wise estimation.

We shall assess the accuracy of the proposed plug-in estimator $\widehat{\ell}_m$ of the value $\ell(\beta)$ by its maximal mean squared error over the classes $\mathcal{F}$ and $\mathcal{G}$, that is,

$$\mathcal{R}_\ell[\widehat{\ell}_m, \mathcal{F}, \mathcal{G}] := \sup_{\beta\in\mathcal{F}}\sup_{\Gamma\in\mathcal{G}} \mathbb{E}|\widehat{\ell}_m - \ell(\beta)|^2. \tag{1.3}$$

Below we derive a lower bound for $\mathcal{R}_\ell[\tilde{\ell}, \mathcal{F}, \mathcal{G}]$ for all estimators $\tilde{\ell}$ and show that it provides up to a constant $C > 0$ also an upper bound for the maximal risk over $\mathcal{F}$ and $\mathcal{G}$ of the estimator $\widehat{\ell}_m$, i.e.,

$$\mathcal{R}_\ell[\widehat{\ell}_m, \mathcal{F}, \mathcal{G}] \leqslant C \cdot \inf_{\tilde{\ell}} \mathcal{R}_\ell[\tilde{\ell}, \mathcal{F}, \mathcal{G}],$$

where the infimum is taken over all estimators of $\ell(\beta)$. We thereby prove the minimax optimality of the estimator $\widehat{\ell}_m$. Our results are appropriate to discuss a wide range of possible regressors, slope parameters and functionals. Moreover they yield in a natural way uniform bounds if the functional varies over a certain subset of the dual space. The paper is organized in the following way: in Section 2 we develop the plug-in estimator and introduce our basic assumptions. In particular we define and illustrate the classes $\mathcal{F}$ and $\mathcal{G}$. Moreover, we embed the cases of point-wise and local average estimation in this general framework and present the resulting plug-in estimators. In this situation the classes $\mathcal{F}$ and $\mathcal{G}$ cover the often considered cases of Sobolev ellipsoids and finitely or infinitely smoothing covariance operators. We provide, in Section 3, sufficient conditions for consistency of the proposed plug-in estimator and then show its minimax-optimality. More precisely, we derive a lower bound for the maximal risk over the classes $\mathcal{F}$ and $\mathcal{G}$ based on an i.i.d. sample obeying the functional linear model (1.1). We show that the proposed plug-in estimator attains the lower bound up to a constant for a wide range of classes $\mathcal{F}$ and $\mathcal{G}$. These results are used to discuss the point-wise estimation of the slope parameter or its derivatives and its average value over a given interval. The proofs can be found in the appendix.

2. Methodology and notations

2.1. Thresholding projection estimator

Following Cardot and Johannes [4] we construct an estimator of the unknown slope function β using a linear Galerkin approach. The estimation of β is based on a dimension reduction together with an additional thresholding, which we elaborate in the following. Let us specify an arbitrary orthonormal basis $\{\psi_j\}_{j=1}^\infty$ of $\mathbb{H}$. We require in the following that the slope function β belongs to a function class $\mathcal{F}$ containing $\{\psi_j\}_{j=1}^\infty$ and, moreover that $\mathcal{F}$ is included in the domain of the linear functional ℓ. For technical reasons and without loss of generality we assume that $\ell(\psi_1) \neq 0$ which can always be ensured by reordering, except for the trivial case $\ell \equiv 0$. With respect to this basis we consider for $f \in \mathbb{H}$ the expansion $f = \sum_{j=1}^\infty [f]_j \psi_j$, with $[f]_j := \langle f, \psi_j \rangle$, for $j \geqslant 1$. The unknown solution $\beta \in \mathbb{H}$ is hence uniquely determined by its coefficients $([\beta]_j)_{j\geqslant 1}$. Given an integer dimension parameter $m \geqslant 1$, we consider the subspace $\mathcal{S}_m$ spanned by the functions $\{\psi_j\}_{j=1}^m$. We recall that a Galerkin solution $\beta_m \in \mathcal{S}_m$ of the operator equation (1.2) with respect to $\mathcal{S}_m$ satisfies

$$\|g - \Gamma\beta_m\|_{\mathbb{H}} \leqslant \|g - \Gamma\tilde{\beta}\|_{\mathbb{H}}, \quad \forall \tilde{\beta} \in \mathcal{S}_m. \tag{2.1}$$

Since Γ is strictly positive, the Galerkin solution exists in a unique way. Precisely, if we consider the expansion $\beta_m = \sum_{j=1}^m [\beta_m]_j \psi_j$ then β_m is uniquely determined by the vector of coefficients $[\beta_m]_{\underline{m}} := ([\beta_m]_1, \dots, [\beta_m]_m)^t$. In this context, the restriction of Γ to an operator from $\mathcal{S}_m$ to itself can be identified with a matrix operating on $\mathbb{R}^m$. This matrix is given by the entries $\langle \psi_j, \Gamma\psi_l\rangle$ for $1 \leqslant j, l \leqslant m$ and will be denoted as $[\Gamma]_{\underline{m}}$, in slight abuse of notation. It is easy to verify that the Galerkin solution defined by (2.1) satisfies $[\Gamma]_{\underline{m}}[\beta_m]_{\underline{m}} = [g]_{\underline{m}}$. Since Γ is strictly positive, the matrix $[\Gamma]_{\underline{m}}$ is nonsingular for all $m \geqslant 1$, such that its inverse $[\Gamma]_{\underline{m}}^{-1}$ always exists. Therefore, the Galerkin solution β_m is determined by

$$[\beta_m]_{\underline{m}} = [\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}. \tag{2.2}$$

Here and subsequently, we denote by $\{(Y_i, X_i)\}_{i=1}^n$ an i.i.d. sample of (Y, X) of size n satisfying (1.1). We observe that $[g]_{\underline{m}} = \mathbb{E}Y[X]_{\underline{m}}$ and $[\Gamma]_{\underline{m}} = \mathbb{E}[X]_{\underline{m}}[X]_{\underline{m}}^t$ and hence it is natural to consider the estimators

$$[\widehat{g}]_{\underline{m}} := \frac{1}{n}\sum_{i=1}^n Y_i\,[X_i]_{\underline{m}} \quad \text{and} \quad [\widehat{\Gamma}]_{\underline{m}} := \frac{1}{n}\sum_{i=1}^n [X_i]_{\underline{m}}[X_i]_{\underline{m}}^t$$

of $[g]_{\underline{m}}$ and $[\Gamma]_{\underline{m}}$, respectively. The estimator of $[\beta_m]_{\underline{m}}$ is derived from (2.2) by replacing $[g]_{\underline{m}}$ and $[\Gamma]_{\underline{m}}$ by their empirical counterparts. However, the inversion of the empirical covariance matrix $[\widehat{\Gamma}]_{\underline{m}}$ introduces an instability to the estimation procedure even if the matrix $[\Gamma]_{\underline{m}}$ is well-conditioned. This instability issue is treated by an additional thresholding step. Let us denote by $\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_S$ the spectral norm of the matrix $[\widehat{\Gamma}]_{\underline{m}}^{-1}$, which equals its largest eigenvalue. Then, the estimator $\widehat{\beta}_m \in \mathcal{S}_m$ of β is determined by the vector of coefficients

$$[\widehat{\beta}_m]_{\underline{m}} := \begin{cases} [\widehat{\Gamma}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}, & \text{if } [\widehat{\Gamma}]_{\underline{m}} \text{ is non-singular and } \|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_S \leqslant n, \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

In order to estimate the value of the linear functional $\ell(\beta)$ we consider the plug-in estimator $\widehat{\ell}_m := \ell(\widehat{\beta}_m)$ and observe that $\ell(\widehat{\beta}_m) = (\ell(\psi_1), \dots, \ell(\psi_m))[\widehat{\beta}_m]_{\underline{m}} =: [\ell]_{\underline{m}}^t[\widehat{\beta}_m]_{\underline{m}}$ with the slight abuse of notations $[\ell]_{\underline{m}} := ([\ell]_j)_{1\leqslant j\leqslant m}$ and generic elements $[\ell]_j := \ell(\psi_j)$. The estimator obviously satisfies

$$\widehat{\ell}_m = \begin{cases} [\ell]_{\underline{m}}^t[\widehat{\Gamma}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}, & \text{if } [\widehat{\Gamma}]_{\underline{m}} \text{ is non-singular and } \|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_S \leqslant n, \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

This procedure raises the question how to choose the dimension parameter m, which depends on the sample size n. It needs to tend to infinity as n increases and we will discuss its optimal choice in Section 3.2.

2.2. Basic model assumptions

Let us introduce the class $\mathcal{F}$ which we determine by means of a weighted norm in $\mathbb{H}$. Given the orthonormal basis $\{\psi_j\}_{j=1}^\infty$ and a strictly positive sequence of weights $(\gamma_j)_{j\geqslant 1}$, or γ for short, we define for $f \in \mathbb{H}$ the weighted norm $\|f\|_\gamma^2 := \sum_{j\geqslant 1} \gamma_j [f]_j^2$. Furthermore, we define $\mathcal{F}_\gamma$ as the completion of $\mathbb{H}$ with respect to $\|\cdot\|_\gamma$. Obviously, for a non-decreasing sequence γ the class $\mathcal{F}_\gamma$ is a subspace of $\mathbb{H}$. In the illustrations of Section 3.4 the order of the sequence γ directly reflects smoothness assumptions on the solution. If there exist an integer $p > 0$ and a constant $c > 0$ such that $c^{-1} j^{2p} \leqslant \gamma_j \leqslant c^{-1} j^{2p}$, or $\gamma_j \sim j^{2p}$ for short, then this polynomial increase will corresponds to p-times differentiable functions. However, the theory in this paper is not restricted to polynomially increasing sequences γ. We also consider an exponential increase, i.e., $\gamma_j \sim \exp(j^{2p})$, which is known to specify analytic functions. We will assume in the following, that there exist a non-decreasing, unbounded sequence of weights γ with $\gamma_1 = 1$ and a constant $\rho > 0$ such that the solution β belongs to the ellipsoid $\mathcal{F}_\gamma^\rho := \big\{ f \in \mathcal{F}_\gamma : \|f\|_\gamma^2 \leqslant \rho \big\}$. In order to guarantee that $\mathcal{F}_\gamma^\rho$ is contained in the domain of the linear functional ℓ and that $\ell(f) = \sum_{j\geqslant 1} [\ell]_j [f]_j$ for all $f \in \mathcal{F}_\gamma^\rho$ with $[\ell]_j = \ell(\psi_j)$, $j \geqslant 1$, it is sufficient that $\sum_{j\geqslant 1} [\ell]_j^2 \gamma_j^{-1} < \infty$. In what follows, we understand arithmetic operations on a sequence of real numbers γ component-wise, e.g., we write $1/\gamma$ instead of $(1/\gamma_j)_{j\geqslant 1}$. As no confusion can be caused we define $\|\ell\|_{1/\gamma}^2 = \sum_{j\geqslant 1} [\ell]_j^2 \gamma_j^{-1}$ and denote the set of all linear functions with $\|\ell\|_{1/\gamma}^2 < \infty$ by $\mathcal{L}_{1/\gamma}$. We may emphasize that we neither impose that the sequence $[\ell] = ([\ell]_j)_{j\geqslant 1}$ tends to zero nor that it is square summable. However, if it is square summable then the entire of $\mathbb{H}$ is the domain of ℓ. Moreover, $[\ell]$ coincides with the sequence of generalized Fourier coefficients of the representer of ℓ given by Riesz's theorem. The assumption $\ell \in \mathcal{L}_{1/\gamma}$ enables us in specific cases to deal with more demanding functionals, such as the estimation of the point-wise evaluation of the slope. As a byproduct, our theory allows us to assess the performance of the estimation procedure of $\ell(\beta)$ not only for a single $\ell \in \mathcal{L}_{1/\gamma}$, but also for ℓ varying over an ellipsoid in $\mathcal{L}_{1/\gamma}$. For this purpose we suppose that there exists a non-negative sequence ω with $\omega_1 = 1$ and a constant $\tau > 0$ such that ℓ belongs to the ellipsoid $\mathcal{L}_\omega^\tau := \{\ell \in \mathcal{L}_{1/\gamma} : \|\ell\|_\omega^2 := \sum_{j\geqslant 1} \omega_j [\ell]_j^2 \leqslant \tau\}$. Under the condition $\sup_{j\geqslant 1}\{1/(\omega_j\gamma_j)\} < \infty$ the ellipsoid $\mathcal{L}_\omega^\tau$ is clearly a subset of $\mathcal{L}_{1/\gamma}$. In order to describe the mapping properties of the covariance operator Γ, stated in the form $\Gamma \in \mathcal{G}$, we introduce the set $\mathcal{N}$ of all strictly positive nuclear operators defined on $\mathbb{H}$. We suppose that there exists a constant $d \geqslant 1$ and a strictly positive, non-increasing sequence of weights υ with $\upsilon_1 = 1$ and $\sum_{j=1}^\infty \upsilon_j < \infty$, such that Γ belongs to the class

$$\mathcal{N}_\upsilon^d := \Big\{ T \in \mathcal{N} : \quad d^{-2} \|f\|_{\upsilon^2}^2 \leqslant \|Tf\|_{\mathbb{H}}^2 \leqslant d^2 \, \|f\|_{\upsilon^2}^2, \quad \forall f \in \mathbb{H} \Big\}.$$

Note that for each $T \in \mathcal{N}$ the trace $\operatorname{tr}(T) := \sum_{j=1}^\infty \langle \psi_j, T\psi_j \rangle$ is finite. Hence, setting $[T]_{j,j} := \langle \psi_j, T\psi_j \rangle$, $j \geqslant 1$, the sequence $([T]_{j,j})_{j\geqslant 1}$ converges to zero. Moreover, for $T \in \mathcal{N}_\upsilon^d$ the decay of this sequence is characterized by υ since $d^{-1}\upsilon_j \leqslant [T]_{j,j} \leqslant d\upsilon_j$ for all $j \geqslant 1$. Furthermore, if λ denotes its sequence of eigenvalues then $d^{-1}\upsilon_j \leqslant \lambda_j \leqslant d\upsilon_j$ holds true for all $j \geqslant 1$. Let us summarize the conditions on the sequences γ, ω and υ.

ASSUMPTION 2.1. *Let γ, ω and υ be strictly positive sequences of weights such that γ and $1/\upsilon$ are non-decreasing. We suppose that the sequences satisfy $\gamma_1 = \omega_1 = \upsilon_1 = 1$, $\sup_{j\geqslant 1}\{1/(\omega_j\gamma_j)\} < \infty$, $\sum_{j=1}^{\infty}\upsilon_j < \infty$ and that γ tends to infinity.*

We illustrate the last assumption for typical choices of the sequences γ, ω and υ:

- *(ppp)* Consider $\gamma_j \sim |j|^{2p}$, $\upsilon_j \sim |j|^{-2a}$ and either *(i)* $[\ell]_j^2 \sim |j|^{-2s}$, or *(ii)* $\omega_j \sim |j|^{2s}$ then Assumption 2.1 holds true if $p > 0$, $a > 1/2$, and either *(i)* $s > 1/2 - p$ or *(ii)* $s > -p$.
- *(pep)* Consider $\gamma_j \sim |j|^{2p}$, $\upsilon_j \sim \exp(-|j|^{2a})$ and either *(i)* $[\ell]_j^2 \sim |j|^{-2s}$ or *(ii)* $\omega_j \sim |j|^{2s}$ then Assumption 2.1 holds true if $p > 0$, $a > 0$ and either *(i)* $s > 1/2 - p$ or *(ii)* $s > -p$.
- *(epp)* Consider $\gamma_j \sim \exp(|j|^{2p})$, $\upsilon_j \sim |j|^{-2a}$ and either *(i)* $[\ell]_j^2 \sim |j|^{-2s}$ or *(ii)* $\omega_j^2 \sim |j|^{2s}$ then Assumption 2.1 holds true if $p > 0$, $a > 1/2$ and $s \in \mathbb{R}$.
- *(ppe)* Consider $\gamma_j \sim |j|^{2p}$, $\upsilon_j \sim |j|^{-2a}$ and either *(i)* $[\ell]_j^2 \sim \exp(-|j|^{2s})$ or *(ii)* $\omega_j \sim \exp(|j|^{2s})$ then Assumption 2.1 holds true if $p > 0$, $a > 1/2$ and $s > 0$.

REMARK 2.1. Cai and Hall [2] consider only *(i)* in the case *(ppp)* and suppose a decay of the representing coefficients $[\ell]$ of order $(|j|^{-s})$ with $s > 1/2$. This condition excludes, for example, point-wise estimation which we will consider, together with the other four cases, below. □

The only assumptions on the stochastic behavior of the error term ε and the regressor X that we need in order to derive our mean squared error results concern their moments. We observe that for all $f \in \mathbb{H}$ the random variable $\langle f, X\rangle$ has mean zero and variance $\langle \Gamma f, f\rangle$ and we will impose moment conditions on the standardized random variable $\langle \Gamma f, f\rangle^{-1/2}\langle f, X\rangle$.

ASSUMPTION 2.2. *There exist an integer $k \geqslant 12$ and a constant $\eta \geqslant 1$ such that $\mathbb{E}|\varepsilon|^{4k} \leqslant \eta$ and that for all $f \in \mathbb{H}$ with $\langle \Gamma f, f\rangle = 1$ it holds $\mathbb{E}|\langle f, X\rangle|^{4k} \leqslant \eta$.*

Note that any centered Gaussian random function X with finite second moment satisfies Assumption 2.2, since for all $f \in \mathbb{H}$ with $\langle \Gamma f, f\rangle = 1$ the corresponding random variable $\langle f, X\rangle$ is standard normally distributed and consequently $\mathbb{E}|\langle f, X\rangle^{4k}| \leqslant (4k-1)\cdot(4k-3)\cdot\ldots\cdot 5\cdot 3\cdot 1$.

2.3. Point-wise and local average estimation

Consider $\mathbb{H} = L^2[0,1]$ with its usual norm and inner product and the trigonometric basis

$$\psi_1 :\equiv 1,\ \psi_{2j}(s) := \sqrt{2}\cos(2\pi j s),\ \psi_{2j+1}(s) := \sqrt{2}\sin(2\pi j s), s \in [0,1],\ j \in \mathbb{N}. \tag{2.5}$$

Recall the typical choices of the sequences γ, ω and υ as introduced above. If $\gamma_j \sim |j|^{2p}$ for a positive integer p, see cases *(ppp)*,*(pep)*,*(ppe)*, then the subset $\mathcal{F}_\gamma$ coincides with the Sobolev space of p-times differential periodic functions (c.f. Neubauer [25, 24]). In the case *(epp)* it is well-known that for $p > 1$ every

$f \in \mathcal{F}_\gamma$ is an analytic function (c.f. Kawata [17]). On the other hand we consider two special cases describing a "regular decay" of the unknown eigenvalues of Γ. Precisely, we assume a polynomial decay of υ with $a > 1/2$ in the cases *(ppp)*, *(epp)* and *(ppe)*. Easy calculus shows that the covariance operator $\Gamma \in \mathcal{N}_\upsilon^d$ acts for integer a like integrating $(2a)$-times and hence it is called *finitely smoothing* (c.f. Natterer [23]). In the case *(pep)* we assume an exponential decay of υ and it is easily seen that the range of $\Gamma \in \mathcal{N}_\upsilon^d$ is a subset of $C^\infty[0,1]$, therefore the operator is called *infinitely smoothing* (c.f. Mair [20]).

Point-wise estimation By *evaluation in a given point* $t_0 \in [0,1]$ we mean the linear functional ℓ_{t_0} mapping f to $f(t_0) := \ell_{t_0}(f) = \sum_{j=1}^\infty [f]_j \psi_j(t_0)$. In the following we shall assume that the point evaluation is well-defined on the set of slope parameters $\mathcal{F}_\gamma$ which is obviously implied by $\sum_{j=1}^\infty [\ell_{t_0}]_j^2 \gamma_j^{-1} < \infty$. Consequently, the condition $\sum_{j \geqslant 1} \gamma_j^{-1} < \infty$ is sufficient to guarantee that the point evaluation is well-defined on $\mathcal{F}_\gamma$. Obviously, in case *(pep)* or in other words for exponentially increasing γ, this additional condition is automatically satisfied. However, a polynomial increase, as in the cases *(ppp)* and *(ppe)*, requires the assumption $p > 1/2$. Roughly speaking, this means that the slope parameter has at least to be continuous. In order to estimate the value $\beta(t_0)$ we consider the plug-in estimator

$$\widehat{\ell}_{t_0}^m = \begin{cases} [\ell_{t_0}]_{\underline{m}}^t [\widehat{\Gamma}]_{\underline{m}}^{-1} [\widehat{g}]_{\underline{m}}, & \text{if } [\widehat{\Gamma}]_{\underline{m}} \text{ is non-singular and } \|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_S \leqslant n, \\ 0, & \text{otherwise,} \end{cases}$$

with $[\ell_{t_0}]_{\underline{m}} = (\psi_1(t_0), \dots, \psi_m(t_0))^t$. Moreover, we observe that $\widehat{\ell}_{t_0}^m = \ell_{t_0}(\widehat{\beta}_m) = \widehat{\beta}_m(t_0)$ for $\widehat{\beta}_m \in \mathcal{S}_m$ as determined by (2.3).

Local average estimation Next we are interested in the average value of β on the interval $[0,b]$ for $b \in (0,1]$. If we denote the linear functional mapping f to $b^{-1}\int_0^b f(t)dt$ by ℓ^b, then it is easily seen that $[\ell^b]_1 = 1$, $[\ell^b]_{2j} = (\sqrt{2}\pi jb)^{-1}\sin(2\pi jb)$, $[\ell^b]_{2j+1} = (\sqrt{2}\pi jb)^{-1}\cos(2\pi jb)$ for $j \geqslant 1$. In this situation the plug-in estimator $\widehat{\ell}_m^b = b^{-1}\int_0^b \widehat{\beta}_m(t)dt$ is written as

$$\widehat{\ell}_m^b = \begin{cases} [\ell^b]_{\underline{m}}^t [\widehat{\Gamma}]_{\underline{m}}^{-1} [\widehat{g}]_{\underline{m}}, & \text{if } [\widehat{\Gamma}]_{\underline{m}} \text{ is non-singular and } \|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_S \leqslant n, \\ 0, & \text{otherwise.} \end{cases}$$

Again, with $\widehat{\beta}_m \in \mathcal{S}_m$ as determined by (2.3), we observe that $\widehat{\ell}_m^b = \ell(\widehat{\beta}_m) = b^{-1}\int_0^b \widehat{\beta}_m(t)dt$.

3. Theoretical properties

3.1. Consistency under mild assumptions

In this section we provide sufficient conditions for the consistency of the estimator $\widehat{\ell}_m$ defined in (2.4) for all $\beta \in \mathcal{F}_\gamma$ and $\ell \in \mathcal{L}_{1/\gamma}$. Recall that this estimator is based on the Galerkin solution β_m. The next assertion summarizes that consistency can be ensured if

$$\|\beta - \beta_m\|_\gamma = o(1) \quad \text{ as } m \to \infty \tag{3.1}$$

which is in general not satisfied without further assumptions, not even for $\gamma \equiv 1$.

PROPOSITION 3.1. *Let $\{(Y_i, X_i)\}_{i=1}^n$ be an i.i.d. sample of (Y, X) satisfying (1.1). Suppose that $\beta \in \mathcal{F}_\gamma$ and $\ell \in \mathcal{L}_{1/\gamma}$, where the sequence γ satisfies Assumption 2.1 and let Assumption 2.2 hold true. Consider the estimator $\widehat{\ell}_m$ with dimension $m = m(n)$ satisfying $1/m = o(1)$ and $m^3 = O(n)$ as $n \to \infty$ and suppose that in addition the following conditions are fulfilled*

$$\|[\Gamma]_{\underline{m}}^{-1}\|_S = o(n) \quad \textit{ and } \quad [\ell]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1} [\ell]_{\underline{m}} = o(n) \quad \textit{ as } n \to \infty. \tag{3.2}$$

If condition (3.1) holds true then we have $\mathbb{E}|\widehat{\ell}_m - \ell(\beta)|^2 = o(1)$ as $n \to \infty$.

If the operator Γ satisfies a link condition, i.e., $\Gamma \in \mathcal{N}_\upsilon^d$, then condition (3.1) is automatically fulfilled which is expressed in the next assertion.

COROLLARY 3.2. *Let the covariance operator Γ be an element of $\mathcal{N}_\upsilon^d$ with $d \geqslant 1$ and let the sequence υ satisfy Assumption 2.1. The conclusion of Proposition 3.1 still holds true without imposing condition (3.1) provided (3.2) is substituted by*

$$\frac{1}{n\upsilon_m} = o(1), \quad \textit{ and } \quad \frac{1}{n}\sum_{j=1}^m \frac{[\ell]_j^2}{\upsilon_j} = o(1) \quad \textit{ as } n \to \infty. \tag{3.3}$$

REMARK 3.1. Consider the case *(ppp)* and suppose that $\sum_{j\geqslant 1}[\ell]^2 < \infty$. In this situation, the assumption $m^{2a} = o(n)$ as $n \to \infty$ implies the additional condition in Corollary 3.2. Interestingly, in a direct regression model the condition $m = o(n)$ as $n \to \infty$ is needed to ensure consistency, which would correspond to the case $a = 1/2$. This, however, cannot hold true for any X with finite second moment. □

3.2. The lower bound

In order to obtain a lower bound for the minimax risk $\inf_{\tilde{\ell}} \mathcal{R}_\ell[\tilde{\ell}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d]$ defined in (1.3) we assume in addition that for all $f \in \mathbb{H}$ the conditional distribution of ε given $\langle X, f\rangle$ is Gaussian with mean zero and variance one, or $\varepsilon|\langle X, f\rangle \sim \mathcal{N}(0, 1)$ for short. This assumption is only used to simplify the calculation of the distance between distributions of the observations corresponding to different slope functions. In order to formulate the lower bounds below let us define

$$m_n^* := \arg\max_{m\geqslant 1}\left\{\frac{\min\{\frac{\upsilon_m}{\gamma_m}, n^{-1}\}}{\max\{\frac{\upsilon_m}{\gamma_m}, n^{-1}\}}\right\}, \quad K_n^* := \max\left\{\frac{\upsilon_{m_n^*}}{\gamma_{m_n^*}}, n^{-1}\right\} \quad \text{for all } n \geqslant 1. \tag{3.4}$$

The lower bound needs the following assumption.

Assumption 3.1. *Let γ and υ be sequences such that*

$$0 < \kappa := \kappa(\upsilon, \gamma) := \inf_{n \geqslant 1} \Big\{ (K_n^*)^{-1} \min\Big\{ \frac{\upsilon_{m_n^*}}{\gamma_{m_n^*}}, n^{-1} \Big\} \Big\} \leqslant 1. \tag{3.5}$$

Now we are in the position to state the main result of this section.

Theorem 3.3. *Let $\{(Y_i, X_i)\}_{i=1}^n$ be an i.i.d. sample of (Y, X) satisfying (1.1). Suppose that Assumptions 2.1, 2.2 and 3.1 hold true and assume in addition that $\varepsilon | \langle X, f \rangle \sim \mathcal{N}(0,1)$ for all $f \in \mathbb{H}$. If $m_n^* \in \mathbb{N}$ is given by equation (3.4) then we have for all $\ell \in \mathcal{L}_{1/\gamma}$ and all $n \geqslant 1$:*

$$\inf_{\tilde{\ell}} \mathcal{R}_\ell[\tilde{\ell}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d] \geqslant \frac{\kappa}{4} \min\left(\frac{\sigma^2}{2d}, \rho\right) \max\left\{ K_n^* \sum_{j=1}^{m_n^*} \frac{[\ell]_j^2}{\upsilon_j}, \sum_{j > m_n^*} \frac{[\ell]_j^2}{\gamma_j} \right\}.$$

Remark 3.2. Below we derive an upper bound for $\mathcal{R}_\ell[\widehat{\ell}_{m_n^*}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d]$ of the estimator $\widehat{\ell}_{m_n^*}$ assuming that the error term ε and the regressor X are uncorrelated. Obviously, in this situation Theorem 3.3 provides a lower bound for any estimator as long as Assumption 2.2 does not exclude a Gaussian error. It is worth to note that the lower bound tends to zero with parametric rate n^{-1} if and only if $\sum_{j=1}^\infty [\ell]_j^2 \upsilon_j{}^{-1} < \infty$, independently of the class $\mathcal{F}_\gamma^\rho$ of slope parameters. $\square$

A straightforward consequence of Theorem 3.3 is the following lower bound over the class $\mathcal{L}_\omega^\tau$ of functionals. We define $j_* := \arg\max_{1 \leqslant j \leqslant m_n^*} (\omega_j^{-1} \upsilon_j^{-1})$ and the functional ℓ_* given by $[\ell_*]_j := (\tau/\omega_{j_*})^{1/2}$ for $j = j_*$ and 0 otherwise. Obviously ℓ_* is an element of $\mathcal{L}_\omega^\tau$. By evaluating the lower bound given by Theorem 3.3 for the specific functional ℓ_* we immediately obtain the following result and we omit its proof.

Corollary 3.4. *Under the conditions of Theorem 3.3 we have*

$$\inf_{\tilde{\ell}} \sup_{\ell \in \mathcal{L}_\omega^\tau} \mathcal{R}_\ell[\tilde{\ell}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d] \geqslant \frac{\kappa\tau}{4} \min\left(\frac{\sigma^2}{2d}, \rho\right) \max_{1 \leqslant j \leqslant m_n^*} \Big(\frac{1}{\upsilon_j \omega_j}\Big) K_n^*.$$

Remark 3.3. It is easily seen that the lower bound given in Corollary 3.4 tends to zero if and only if $(\omega_j \gamma_j)_{j \geqslant 1}$ tends to infinity. In other words, consistency of an estimator of $\ell(\beta)$ uniformly over spheres in $\mathcal{F}_\gamma$ and $\mathcal{L}_{1/\gamma}$ is impossible. This obviously reflects the ill-posedness of the underlying inverse problem. $\square$

By considering the typical choices of γ and υ gathered in the cases *(ppp)*,*(pep)*, *(ppe)* and *(epp)* above, we illustrate now the lower bounds for the minimax risks $\inf_{\tilde{\ell}} \mathcal{R}_\ell[\tilde{\ell}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d]$ and $\inf_{\tilde{\ell}} \sup_{\ell \in \mathcal{L}_\omega^\tau} \mathcal{R}_\ell[\tilde{\ell}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d]$. We see from Theorem 3.3 and Corollary 3.4, respectively, that their orders are determined by the sequences $\delta^* := (\delta_n^*)_{n \geqslant 1}$ and $\Delta^* := (\Delta_n^*)_{n \geqslant 1}$ given by

$$\delta_n^* := \max\left\{ K_n^* \sum_{j=1}^{m_n^*} \frac{[\ell]_j^2}{\upsilon_j}, \sum_{j > m_n^*} \frac{[\ell]_j^2}{\gamma_j} \right\} \quad \text{and} \quad \Delta_n^* := \max_{1 \leqslant j \leqslant m_n^*} \Big(\frac{1}{\upsilon_j \omega_j}\Big) K_n^*. \tag{3.6}$$

In the next assertion we present the orders of those sequences.

PROPOSITION 3.5. *Let the assumptions of Theorem 3.3 hold true. Under the following conditions the Assumptions 2.1 and 3.1 are satisfied and the lower bounds are determined by the orders of δ^* and Δ^* as given below.*

(ppp) *If $p > 0$ and $a > 1/2$, then $m_n^* \sim n^{1/(2p+2a)}$ and if*

(i) *$s > 1/2 - p$, then*

$$\delta_n^* \sim \begin{cases} n^{-(2p+2s-1)/(2p+2a)}, & \text{if } s-a < 1/2 \\ n^{-1}\log(n), & \text{if } s-a = 1/2 \\ n^{-1}, & \text{if } s-a > 1/2, \end{cases}$$

(ii) *$s > -p$, then $\Delta_n^* \sim \max(n^{-(p+s)/(p+a)}, n^{-1})$.*

(pep) *If $p > 0$ and $a > 0$, then $m_n^* \sim \log(n[\log(n)]^{-p/a})^{1/(2a)}$ and if*

(i) *$s > 1/2 - p$, then $\delta_n^* \sim [\log(n)]^{-(2p+2s-1)/(2a)}$,*

(ii) *$s > -p$, then $\Delta_n^* \sim [\log(n)]^{-(p+s)/a}$.*

(epp) *If $p > 0$, $a > 1/2$ and $s \in \mathbb{R}$ then $m_n^* \sim \log(n[\log(n)]^{-a/p})^{1/(2p)}$ and*

(i) $$\delta_n^* \sim \begin{cases} n^{-1}[\log(n)]^{(2a-2s+1)/(2p)}, & \text{if } s-a < 1/2 \\ n^{-1}\log[\log(n)], & \text{if } s-a = 1/2 \\ n^{-1}, & \text{if } s-a > 1/2, \end{cases}$$

(ii) *$\Delta_n^* \sim \max(n^{-1}[\log(n)]^{(a-s)/p}, n^{-1})$.*

(ppe) *If $p > 0$, $a > 1/2$ and $s > 0$ then $m_n^* \sim n^{1/(2p+2a)}$ and*

(i) *$\delta_n^* \sim n^{-1}$* *(ii)* *$\Delta_n^* \sim n^{-1}$.*

REMARK 3.4. The rates given in Proposition 3.5 determine up to a constant the minimax optimal rate of convergence, as we will show in Proposition 3.8 below. Nevertheless, we shall already emphasize here the interesting influence of the parameters p, s and a characterizing the 'smoothness' of β, ℓ and the decay of the eigenvalues of Γ respectively. As we see from Proposition 3.5, an increase of the value of a leads in each case to a slower obtainable optimal rate of convergence. Therefore, the parameter a is often called degree of ill-posedness (c.f. Natterer [23]). On the other hand, an increase of the value of p or s leads to a faster optimal rate. In other words values of a linear functional given by a smoother slope function or representer can be estimated faster, as expected. Moreover, in the cases *(ppp)* and *(epp)* the parametric rate n^{-1} is obtained if and only if the functional is 'smoother' than the degree of ill-posedness of Γ in the sense that *(i)* $s \geqslant a - 1/2$ and *(ii)* $s \geqslant a$. The situation is different in the cases *(pep)* and *(ppe)*, where the optimal rates are always logarithmic or parametric, respectively. □

REMARK 3.5. There is an interesting issue hidden in the parametrization that we have chosen. Consider a classical indirect regression model given by the covariance operator Γ and Gaussian white noise $\dot{W}$, i.e., $g_n = \Gamma\beta + n^{-1/2}\dot{W}$ (for details see e.g. Hoffmann and Reiß [14]). It is shown in Johannes and Kroll [16] that, for example in case *(ppp)*, the optimal rate of convergence over the classes $\mathcal{F}_\gamma^\rho$ and $\mathcal{L}_\omega^\tau$ of any estimator of $\ell(\beta)$ is of order $\max(n^{-(p+s)/(p+2a)}, n^{-1})$. In contrast, Proposition 3.5 states that the optimal rate in a functional linear model is of order $\max(n^{-(p+s)/(p+a)}, n^{-1})$. Thus, we see by comparing the two rates that the covariance operator Γ in a functional linear model has the degree of ill-posedness a while the same operator has a degree of ill-posedness $(2a)$ in the indirect regression model. In other words, in a functional linear model we do not face the complexity of the inversion of Γ but only of its square root $\Gamma^{1/2}$. It is interesting to note that in case *(ppp)* functional linear regression is asymptotically equivalent in Le Cam's sense to the white noise problem with the square root of the covariance operator (cf. Meister [21]). Similar remarks hold true for the other cases, however, in case *(pep)* the rate of convergence is the same as in an indirect regression model with Gaussian white noise (c.f. Johannes and Kroll [16]). This is due to the fact that if $\upsilon_j \sim \exp(-r|j|^{2a})$ for some $r > 0$, then the dependence of the rate of convergence on the value r is hidden in the constant. □

3.3. The upper bound

Proposition 3.1 shows that the estimator $\widehat{\ell}_m$ defined in (2.4) is consistent for all slope functions and functionals belonging to $\mathcal{F}_\gamma$ and $\mathcal{L}_{1/\gamma}$, respectively. The following theorem provides an upper bound if β belongs to an ellipsoid $\mathcal{F}_\gamma^\rho$.

Theorem 3.6. *Let $\{(Y_i, X_i)\}_{i=1}^n$ be an i.i.d. sample of (Y, X) satisfying (1.1). Suppose that Assumptions 2.1, 2.2 and 3.1 hold true and that*

$$\sup_{m\in\mathbb{N}}\Big\{\frac{\upsilon_m}{\gamma_m} m^3\Big\} < \infty. \tag{3.7}$$

Consider δ_n^* as in (3.6) and the estimator $\widehat{\ell}_m$ defined with dimension $m := m_n^*$ given by (3.4). There exists a constant $C(d,\upsilon,\gamma)$ depending on d, υ and γ only such that $\mathcal{R}_\ell[\widehat{\ell}_{m_n^*}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d] \leqslant C(d,\upsilon,\gamma)\eta\{(\sigma^2+\rho)\cdot\delta_n^* + \rho\|\ell\|_{1/\gamma}^2 n^{-1}\}$ and therefore

$$\mathcal{R}_\ell[\widehat{\ell}_{m_n^*}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d] \leqslant C(d,\upsilon,\gamma)\eta(\sigma^2+\rho+\rho\|\ell\|_{1/\gamma}^2)\cdot \inf_{\tilde{\ell}} \mathcal{R}_\ell[\tilde{\ell}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d].$$

The rate $\delta^* := (\delta_n^*)_{n\geqslant 1}$ of the lower bound given in Theorem 3.3 provides up to a constant also an upper bound of the estimator $\widehat{\ell}_{m_n^*}$. Thus, we have shown that the rate δ^* is optimal and hence $\widehat{\ell}_{m_n^*}$ is minimax-optimal. We observe that $\delta_n^* \leqslant \tau \cdot \max_{1\leqslant j\leqslant m_n^*}\big(\upsilon_j^{-1}\omega_j^{-1}\big)K_n^* = \tau\cdot\Delta_n^*$ for all $\ell \in \mathcal{L}_\omega^\tau$ and therefore we obtain the following result as a consequence of Theorem 3.6 and we omit the proof.

Corollary 3.7. *Under the conditions of Theorem 3.6 we have that $\sup_{\ell\in\mathcal{L}_\omega^\tau}\mathcal{R}_\ell[\widehat{\ell}_{m_n^*}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d] \leqslant C(d,\upsilon,\gamma)\eta(\sigma^2+2\rho)\cdot\tau\cdot\Delta_n^*$ and therefore*

$$\sup_{\ell\in\mathcal{L}_\omega^\tau}\mathcal{R}_\ell[\widehat{\ell}_{m_n^*}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d] \leqslant C(d,\upsilon,\gamma)\eta(\sigma^2+2\rho)\cdot\inf_{\tilde{\ell}}\sup_{\ell\in\mathcal{L}_\omega^\tau}\mathcal{R}_\ell[\tilde{\ell}, \mathcal{F}_\gamma^\rho, \mathcal{N}_\upsilon^d].$$

In the following we illustrate the previous results by considering the typical choices of γ and υ presented below Assumption 2.1.

Proposition 3.8. *Let $\{(Y_i, X_i)\}_{i=1}^n$ be an i.i.d. sample of (Y, X) satisfying (1.1) and suppose that Assumption 2.2 holds true. Let the estimator $\widehat{\ell}_m$ be defined with dimension $m := m_n^*$ as given in Proposition 3.5. Then $\widehat{\ell}_{m_n^*}$ attains the optimal rates δ_n^* and respectively Δ_n^* given in Proposition 3.5 if we additionally assume $p + a \geqslant 3/2$ in the cases (ppp) and (ppe).*

Remark 3.6. It is of interest to compare our results with those of Cai and Hall [2] who consider only *(i)* in the case *(ppp)*. In their notations the decay of the eigenvalues of Γ is assumed to be of order $(|j|^{-\alpha})$, i.e., $\alpha = 2a$, with $\alpha > 1$. Furthermore, they suppose a decay of the coefficients of the slope function and the representing sequence $[\ell]$ of order $(|j|^{-\beta})$, i.e., $\beta = p + 1/2$, with $\beta \geqslant \alpha + 2$, and $(|j|^{-\gamma})$, i.e., $\gamma = s$, with $\gamma > 1/2$ respectively. By using this parametrization we see that our results in the case *(ppp)* imply the same rate of convergence as the one presented in Cai and Hall [2]. However, we shall stress that the condition $\beta \geqslant \alpha + 2$ or equivalently $p \geqslant 3/2 + 2a$ is much stronger than the condition $p + a \geqslant 3/2$ used in Proposition 3.8. □

3.4. Optimal point-wise and local average estimation

We continue the discussion of Section 2.3.

Point-wise estimation - continued Recall that $\ell_{t_0}(\widehat{\beta}_{m_n^*}) = \widehat{\beta}_{m_n^*}(t_0)$ with $[\ell_{t_0}]_j^2 = \psi_j^2(t_0) \sim j^{-2s}$ and $s = 0$. By applying Proposition 3.8, the estimator's maximal mean squared error over the classes $\mathcal{F}_\gamma^\rho$ and $\mathcal{N}_\upsilon^d$ is uniformly bounded for $t_0 \in [0, 1]$ up to a constant by δ_n^*, i.e., $\sup_{\beta \in \mathcal{F}_\gamma^\rho} \sup_{\Gamma \in \mathcal{N}_\upsilon^d} \mathbb{E}|\widehat{\beta}_{m_n^*}(t_0) - \beta(t_0)|^2 \leqslant C\delta_n^*$ for some $C > 0$. Moreover, due to Proposition 3.5, δ_n^* is the minimax-optimal rate of convergence. This means in the three considered cases:

(ppp) If $p > 1/2$, $a > 1/2$ and $p + a \geqslant 3/2$, then $m_n^* \sim n^{1/(2p+2a)}$ and $\delta_n^* \sim n^{-(2p-1)/(2p+2a)}$.

(pep) If $p > 1/2$ and $a > 0$, then $m_n^* \sim \log(n[\log(n)]^{-p/a})^{1/(2a)}$ and $\delta_n^* \sim [\log(n)]^{-(2p-1)/2a}$.

(epp) If $p > 0$ and $a > 1/2$, then $m_n^* \sim \log(n[\log(n)]^{-a/p})^{1/(2p)}$ and $\delta_n^* \sim n^{-1}[\log(n)]^{(2a+1)/2p}$.

Let us compare the optimal rate $n^{-(2p-1)/(2p+1)}$ in a direct regression model with a p-times differentiable slope function β with our results in the cases *(ppp)* and *(pep)*. Obviously, the optimal rate in a functional linear model is never faster than the one of a direct regression model. Furthermore, they would only coincide in the case *(ppp)* for $a = 1/2$, however, this cannot hold true for any random function X with finite second moment.

It is interesting to note that by slightly adapting the previously presented procedure we are able to estimate the value of the q-th derivative of β at t_0. Given the exponential basis, which is linked to the trigonometric basis for $k \in \mathbb{Z}$ and $t \in [0, 1]$ by the relation $\exp(2i\pi kt) = 2^{1/2}(\psi_{2k}(t) + i\psi_{2k+1}(t))$ with $i^2 = 1$.

We recall that for $0 \leqslant q < p$ the q-th derivative $\beta^{(q)}$ of β in a weak sense satisfies

$$\beta^{(q)}(t_0) = \sum_{k\in\mathbb{Z}} (2i\pi k)^q \exp(2i\pi k t_0)\Big(\int_0^1 \beta(u)\exp(2i\pi k u)du\Big).$$

Given a dimension $m \geqslant 1$, we denote now by $[\widehat{\Gamma}]_{\underline{m}}$ the $(2m+1)\times(2m+1)$ matrix with generic elements $\langle\psi_j, \widehat{\Gamma}\psi_k\rangle$, $-m \leqslant j,k \leqslant m$ and by $[\widehat{g}]_{\underline{m}}$ the $(2m+1)$ vector with elements $\langle\widehat{g},\psi_j\rangle$, $-m \leqslant j \leqslant m$. Furthermore, we define for integer q the $(2m+1)$ vector $[\ell^{(q)}_{t_0}]_{\underline{m}}$ with elements $[\ell^{(q)}_{t_0}]_j := (2i\pi j)^q \exp(2i\pi j t_0)$, $-m \leqslant j \leqslant m$. In the following we shall assume that the point evaluation of the q-th derivative is well-defined on the set of slope parameters $\mathcal{F}_\gamma$ which is implied by $\sum_{j\geqslant 1}(j^{2q}\gamma_j^{-1}) < \infty$, since $|[\ell^{(q)}_{t_0}]_j|^2 \sim j^{2q}$. Obviously, this additional condition is automatically satisfied in case *(pep)* and requires the assumption $q < p - 1/2$ in the cases *(ppp)* and *(ppe)*. We consider the estimator of $\beta^{(q)}(t_0) = \ell^{(q)}_{t_0}(\beta)$ given by

$$\widehat{\beta}^{(q)}_m(t_0) = \begin{cases} [\ell^{(q)}_{t_0}]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[\widehat{g}]_{\underline{m}} & \text{if } [\widehat{\Gamma}]_{\underline{m}} \text{ is non-singular and } \|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_S \leqslant n,\\ 0, & \text{otherwise.}\end{cases}$$

The estimator $\widehat{\beta}^{(q)}_{m_n^*}(t_0)$ can be represented as $\ell^{(q)}_{t_0}(\widehat{\beta}_{m_n^*})$ with $[\ell^{(q)}_{t_0}]^2_j \sim j^{-2s}$ and $s = -q$. By applying Proposition 3.5 and 3.8, the maximal mean squared error over the classes $\mathcal{F}^\rho_\gamma$ and $\mathcal{N}^d_\upsilon$ of the estimator $\widehat{\beta}^{(q)}_m(t_0)$ is uniformly bounded for $t_0 \in [0,1]$ up to a constant by the minimax rate δ^*_n. This means in the three considered cases:

(ppp) If $p > 1/2$, $a > 1/2$ and $p + a \geqslant 3/2$, then $m^*_n \sim n^{1/(2p+2a)}$ and $\delta^*_n \sim n^{-(2p-2q-1)/(2p+2a)}$.
(pep) If $p > 1/2$ and $a > 0$, then $m^*_n \sim \log(n[\log(n)]^{-p/a})^{1/(2a)}$ and $\delta^*_n \sim [\log(n)]^{-(2p-2q-1)/2a}$.
(epp) If $p > 0$ and $a > 1/2$, then $m^*_n \sim \log(n[\log(n)]^{-a/p})^{1/(2p)}$ and $\delta^*_n \sim n^{-1}[\log(n)]^{(2a+2q+1)/2p}$.

□

Local average estimation - continued Recall that $\widehat{\ell}^b_m = b^{-1}\int_0^b \widehat{\beta}_m(t)dt$ with $[\ell_b]^2_j \sim j^{-2s}$ and $s = 1$. Its maximal mean squared error over $\mathcal{F}^\rho_\gamma$ and $\mathcal{N}^d_\upsilon$ is bounded up to a constant by δ^*_n, that is $\sup_{\beta\in\mathcal{F}^\rho_\gamma,\Gamma\in\mathcal{N}^d_\upsilon}\mathbb{E}|\int_0^b \widehat{\beta}_{m_n^*}(t)dt - \int_0^b \beta(t)dt|^2 \leqslant C\delta^*_n$ for some $C > 0$ (Proposition 3.8). Moreover, due to Proposition 3.5, δ^*_n is again minimax-optimal. In the three cases the order of δ^*_n is given as follows:

(ppp) If $p \geqslant 0$, $a > 1/2$ and $p + a > 3/2$, then $m^*_n \sim n^{1/(2p+2a)}$ and $\delta^*_n \sim n^{-(2p+1)/(2p+2a)}$.
(pep) If $p \geqslant 0$ and $a > 0$, then $m^*_n \sim \log(n[\log(n)]^{-p/a})^{1/(2a)}$ and $\delta^*_n \sim [\log(n)]^{-(2p+1)/2a}$.
(epp) If $p > 0$ and $a > 1/2$, then $m^*_n \sim \log(n[\log(n)]^{-a/p})^{1/(2p)}$ and $\delta^*_n \sim n^{-1}[\log(n)]^{(2a-1)/2p}$.

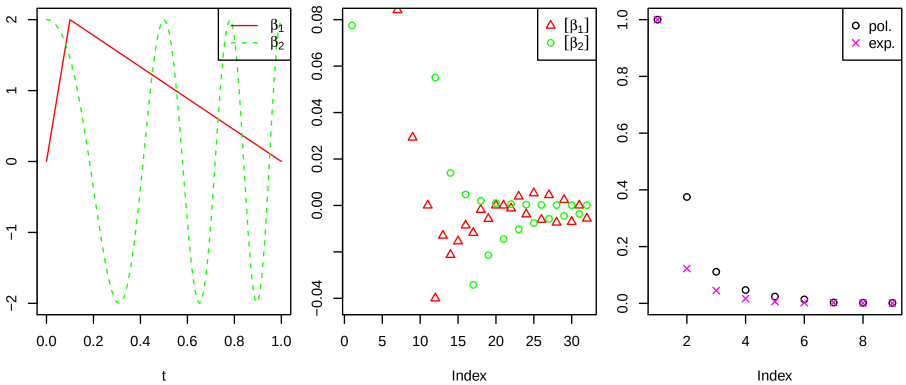

FIG 1. *In the left panel the slope function β_1 is drawn with a solid red line and β_2 with a dashed green line. The middle panel depicts the Fourier coefficients of β_1 with red triangles and those of β_2 with green circles. The right panel illustrates the polynomial decay of υ with black circles and an exponential decay with magenta crosses respectively.*

In contrast to a direct regression model where the average value of the regression function can be estimated with parametric rate n^{-1}, we find that in the considered cases the optimal rate is always slower than n^{-1}. Observe that in the cases *(ppp)* and *(epp)* the rate could only be parametric for $a \leqslant 1/2$, which again cannot hold true. $\square$

A simulation study We consider in this paragraph the problem of point-wise estimation of the following slope functions

$$\beta_1(t) := 20t\,\mathbb{1}_{[0;0.1]} - \tfrac{20}{9}(t-1)\,\mathbb{1}_{(0.1;1]} \quad \text{and} \quad \beta_2(t) := 2\cdot\cos(2\pi t(2t+1))\,;\, t \in [0,1].$$

For computation, the infinite dimensional vectors $[\beta_1]$ and $[\beta_2]$ of coefficients with respect to the trigonometric basis given in (2.5) are truncated at a sufficiently large index L; in the following we set $L := 1024$. The slope functions and their respective first 32 Fourier coefficients are illustrated in figure 1. Furthermore, we assume that $[X]_{\underline{L}}$ follows a multivariate normal distribution with mean 0 and covariance matrix $[\Gamma]_{\underline{L}} = [\mathrm{Diag}(\upsilon)]_{\underline{L}} + [\mathrm{Diag}(\upsilon^2)]_{\underline{L}} U [\mathrm{Diag}(\upsilon^2)]_{\underline{L}}$, where U is a randomly generated covariance matrix with spectral norm 1. This construction guarantees that Γ belongs to the operator class $\mathcal{N}_\upsilon^d$ for a sufficiently large d. In the case *(ppp)* we set $\upsilon_j \sim j^{-3}$ and in case *(pep)* $\upsilon_j \sim \exp(-j), 1 \leqslant j \leqslant L$. The noise level σ^2 is assumed to equal 0.1. In the following we estimate the evaluation of the slope functions on a equidistantly spaced grid of length 200. The dimension parameter m is chosen optimally by minimizing the cummulative error $|\widehat{\ell}_{t_0}^m - \beta_i(t_0)|^2$, $i \in \{1, 2\}$, over 101 repetitions. The results of the simulation are depicted in figure 2. As expected, for both β_1 and β_2 a higher number of coefficients is chosen in the polynomial case than in the exponential case as seen in figure 3. We may also state that the optimal parameter m becomes larger as the sample size increases. An increase of the noise level to $\sigma^2 = 1$ leads to wider 90% bands as illustrated in figure 4.

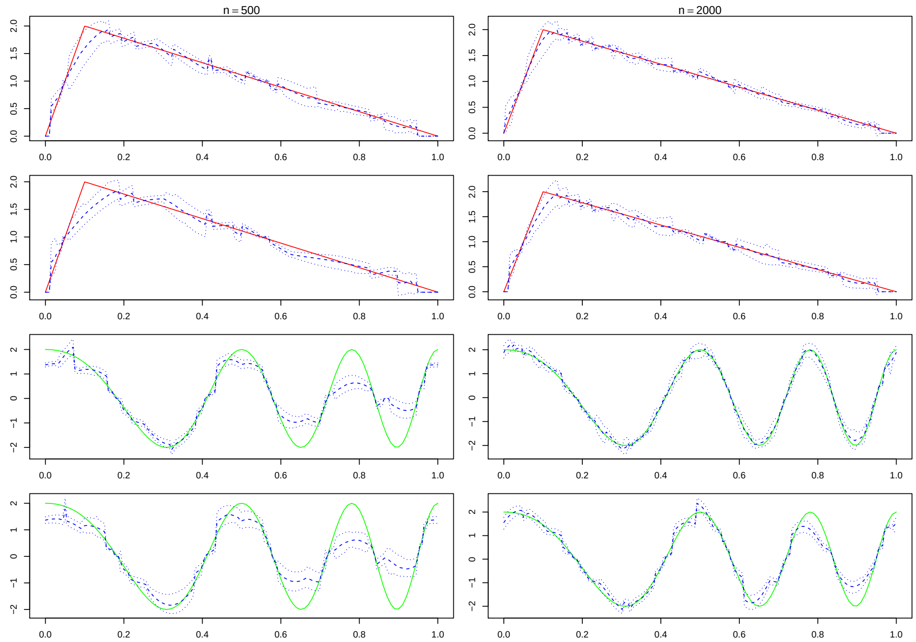

Fig 2. *Between the dotted blue lines are 90% of the estimates. The dashed blue represents their point-wise median. The first and third row depict the results for a polynomial decay of υ, whereas the second and fourth row correspond to an exponential decay.*

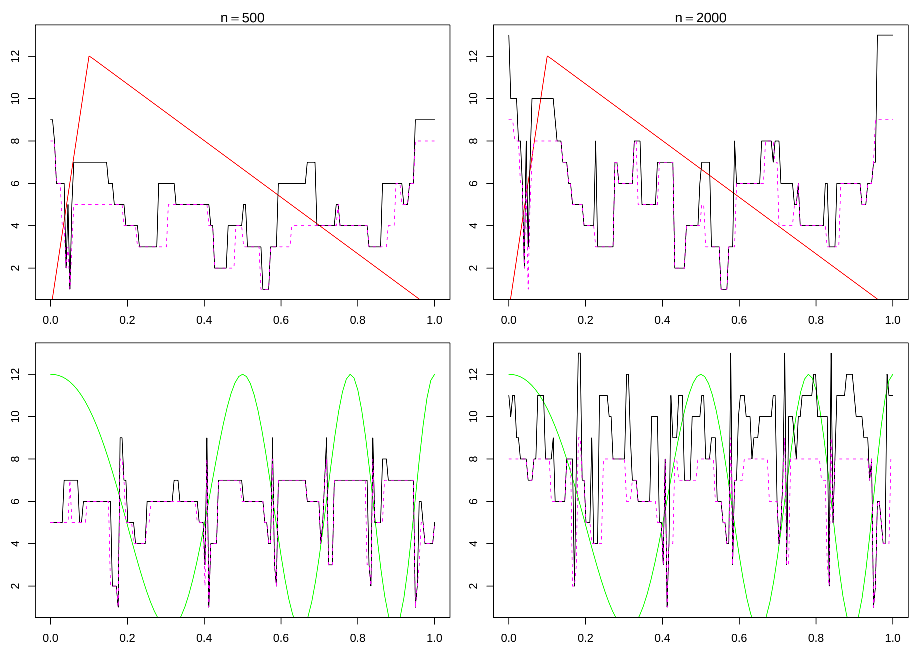

Fig 3. *The slope functions are rescaled for reference. The solid black lines illustrates the optimally chosen dimension parameter in the case of polynomially decaying υ, whereas the dashed magenta line corresponds to the case of an exponential decay.*

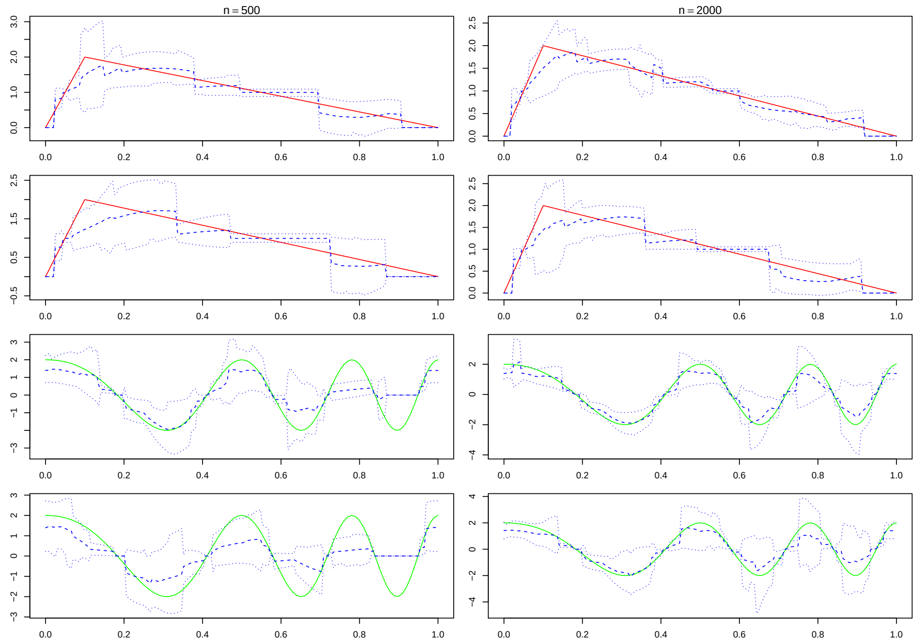

FIG 4. *The noise level has been increased to $\sigma^2 = 1$. Between the dotted blue lines are 90% of the estimates. The dashed blue represents their point-wise median. The first and third row depict the results for a polynomial decay of υ, whereas the second and fourth row correspond to an exponential decay.*

Similar findings hold true, if the Haar wavelet basis is used, however the resulting estimates are less convincing, due to the fact that only a few coefficients in the trigonometric basis are needed to approximate the chosen slope functions reasonable well, whereas more coefficients in the Haar wavelet basis are needed.

Conclusion In this paper we have presented a minimax optimal plug-in estimation technique that is suited to deal in particular with point-wise estimation and the estimation of local averages. Obviously, the data driven choice of the dimension parameter m is only one amongst the many interesting questions for further research and we are currently exploring this topic. Another one might be the exploitation of local structures and sparse representations of the slope parameter β by wavelet thresholding techniques.

Appendix A: Proofs

We begin by defining and recalling notations to be used in the proofs of this section. Given $m \geqslant 1$, let us denote by $\|\cdot\|$ the euclidean norm in $\mathbb{R}^m$, by $[\mathrm{Diag}(\gamma)]_{\underline{m}}$ the m-dimensional diagonal matrix with entries $(\gamma_1, \ldots, \gamma_m)$ and by $[I]_{\underline{m}}$ the m-dimensional identity matrix. Furthermore recall that $\beta_m \in \mathcal{S}_m$

denotes the Galerkin solution of $g = \Gamma\beta$ defined by (2.1). Let us introduce the notations

$$[\widehat{\Gamma}]_{\underline{m}} = \frac{1}{n}\sum_{i=1}^{n}[X_i]_{\underline{m}}[X_i]_{\underline{m}}^t, \quad [\Xi]_{\underline{m}} := [\Gamma]_{\underline{m}}^{-1/2}[\widehat{\Gamma}]_{\underline{m}}[\Gamma]_{\underline{m}}^{-1/2} - [I]_{\underline{m}},$$
$$[Z]_{\underline{m}} := [\widehat{g}]_{\underline{m}} - [\widehat{\Gamma}]_{\underline{m}}[\beta_m]_{\underline{m}}.$$

Moreover, we define the events

$$\Omega := \{\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_S \leqslant n\}, \quad \Omega_{1/2} := \{\|[\Xi]_{\underline{m}}\|_S \leqslant 1/2\}$$
$$\Omega^c := \{\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_S > n\} \quad \text{and} \quad \Omega_{1/2}^c = \{\|[\Xi]_{\underline{m}}\|_S > 1/2\}.$$

We shall prove in the end of this section two technical Lemmata (B.1 and B.2) which are used in the following proofs. Furthermore, we will denote by C universal numerical constants and by $C(\cdot)$ constants depending only on the arguments. In both cases, the values of the constants may change from line to line.

A.1. Proof of the consistency result

Proof of Proposition 3.1. Let us define $\widetilde{\ell}_m := \ell(\beta_m)\,\mathbb{1}_\Omega$. Then the proof is based on the decomposition

$$\mathbb{E}|\widehat{\ell}_m - \ell(\beta)|^2 \leqslant 2\{\mathbb{E}|\widehat{\ell}_m - \widetilde{\ell}_m|^2 + \mathbb{E}|\widetilde{\ell}_m - \ell(\beta)|^2\}$$

where we will bound each term on the right hand side separately. On the one hand we have

$$\mathbb{E}|\widetilde{\ell}_m - \ell(\beta)|^2 \leqslant 2\{|\ell(\beta - \beta_m)|^2 + |\ell(\beta)|^2\,P(\Omega^c)\}. \tag{A.1}$$

On the other hand, we conclude from the identity $[\widehat{g}]_{\underline{m}} - [\widehat{\Gamma}]_{\underline{m}}[\beta_m]_{\underline{m}} = [Z]_{\underline{m}}$ that

$$\mathbb{E}|\widehat{\ell}_m - \widetilde{\ell}_m|^2 = \mathbb{E}|[\ell]_{\underline{m}}^t\,\{[\Gamma]_{\underline{m}}^{-1} + ([\widehat{\Gamma}]_{\underline{m}}^{-1} - [\Gamma]_{\underline{m}}^{-1})\}\,[Z]_{\underline{m}}|^2\,\mathbb{1}_\Omega\,.$$

By using $\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_S\,\mathbb{1}_\Omega \leqslant n$ and $\|\{[I]_{\underline{m}} + [\Xi]_{\underline{m}}\}^{-1}\|_S\,\mathbb{1}_{\Omega_{1/2}} \leqslant 2$ it follows that

$$\begin{aligned}\mathbb{E}|\widehat{\ell}_m - \widetilde{\ell}_m|^2 \leqslant 4\Big[&\mathbb{E}|[\ell]_{\underline{m}}^t\,[\Gamma]_{\underline{m}}^{-1}[Z]_{\underline{m}}|^2\\ &+ \|[\ell]_{\underline{m}}^t\,[\Gamma]_{\underline{m}}^{-1/2}\|^2\Big\{4(\mathbb{E}\|[\Xi]_{\underline{m}}\|_S^4)^{1/2}(\mathbb{E}\|[\Gamma]_{\underline{m}}^{-1/2}[Z]_{\underline{m}}\|^4)^{1/2}\\ &+ n^2\,\|[\Gamma]_{\underline{m}}\|_S^2\,(\mathbb{E}\|[\Xi]_{\underline{m}}\|_S^8)^{1/4}(\mathbb{E}\|[\Gamma]_{\underline{m}}^{-1/2}[Z]_{\underline{m}}\|^8)^{1/4}(P(\Omega_{1/2}^c))^{1/2}\Big\}\Big].\end{aligned}$$

We observe that $\mathbb{E}|[\ell]_{\underline{m}}^t\,[\Gamma]_{\underline{m}}^{-1}[Z]_{\underline{m}}|^2 \leqslant \|[\ell]_{\underline{m}}^t\,[\Gamma]_{\underline{m}}^{-1/2}\|^2 \sup_{z\in\mathbb{R}^m:z^tz=1}\mathbb{E}|z^t\,[\Gamma]_{\underline{m}}^{-1/2} \times [Z]_{\underline{m}}|^2$ and that $\|[\Gamma]_{\underline{m}}\|_S$ is less or equal than the operator norm $\|\Gamma\|_S$ of Γ, which equals its largest eigenvalue. Therefore, by using (B.1) - (B.3) in Lemma B.1 with $k = 12$ we have

$$\mathbb{E}|\widehat{\ell}_m - \widetilde{\ell}_m|^2 \leqslant C\,(\|[\ell]^t_{\underline{m}}\,[\Gamma]^{-1/2}_{\underline{m}}\|^2/n)\,\eta\left\{\|\Gamma^{1/2}(\beta-\beta_m)\|^2_{\mathbb{H}} + \sigma^2\right\} \cdot \left\{1 + m^3/n + \eta^{-1/2}\|\Gamma\|^2_S m^3 n\,(P(\Omega^c_{1/2}))^{1/2}\right\}. \quad \text{(A.2)}$$

The combination of (A.1) and (A.2) leads to the estimate

$$\mathbb{E}|\widehat{\ell}_m - \ell(\beta)|^2 \leqslant C\left\{|\ell(\beta-\beta_m)|^2 + |\ell(\beta)|^2\,P(\Omega^c) + (\|[\ell]^t_{\underline{m}}\,[\Gamma]^{-1/2}_{\underline{m}}\|^2/n)\,\eta \cdot \{\|\Gamma^{1/2}(\beta-\beta_m)\|^2_{\mathbb{H}} + \sigma^2\}\left[1 + m^3/n + \eta^{-1/2}\|\Gamma\|^2_S\,m^3 n(P(\Omega^c_{1/2}))^{1/2}\right)\right]\right\}. \quad \text{(A.3)}$$

We observe that the identity $[\widehat{\Gamma}]_{\underline{m}} = [\Gamma]^{1/2}_{\underline{m}}\{[I]_{\underline{m}} + [\Xi]_{\underline{m}}\}[\Gamma]^{1/2}_{\underline{m}}$ implies by the usual Neumann series argument that if $\|[\Xi]_{\underline{m}}\|_S \leqslant 1/2$ then $\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_S \leqslant 2\|[\Gamma]^{-1}_{\underline{m}}\|_S$. Thereby, $n \geqslant 2\|[\Gamma]^{-1}_{\underline{m}}\|_S$ implies $\Omega^c \subset \Omega^c_{1/2}$. Furthermore, due to (B.3) in Lemma B.1 with $k = 12$ we obtain by applying Markov's inequality that $P(\Omega^c_{1/2}) \leqslant C\eta m^{24} n^{-12}$. This leads to

$$\mathbb{E}|\widehat{\ell}_m - \ell(\beta)|^2 \leqslant C\left\{|\ell(\beta-\beta_m)|^2 + |\ell(\beta)|^2\,(m^3/n)^8 n^{-4}\,\eta + (\|[\ell]^t_{\underline{m}}\,[\Gamma]^{-1/2}_{\underline{m}}\|^2/n)\,\eta\,\{\|\Gamma^{1/2}(\beta-\beta_m)\|^2_{\mathbb{H}} + \sigma^2\}\left[1 + (m^3/n) + (m^3/n)^5\|\Gamma\|^2_S\right]\right\} \quad \text{(A.4)}$$

Furthermore, for each $\beta \in \mathcal{F}_\gamma$, we have $\|\beta - \beta_m\|_\gamma = o(1)$ as $m \to \infty$ from condition (3.1), which implies $|\ell(\beta-\beta_m)|^2 = o(1)$ and $\|\Gamma^{1/2}(\beta-\beta_m)\|_{\mathbb{H}} = o(1)$ as $m \to \infty$ under Assumption 2.1. Consequently, the conditions $m^3 = O(n)$ and $[\ell]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[\ell]_{\underline{m}} = o(n)$ as $n \to \infty$ ensure the convergence to zero of the bound given in (A.4) as $n \to \infty$, which proves the result. □

Proof of Corollary 3.2. First, we prove that $\Gamma \in \mathcal{N}^d_\upsilon$ implies (3.1). Let us denote by Π_m and $\Pi^\perp_m$ the orthogonal projections on $\mathcal{S}_m$ and its orthogonal complement, respectively. On the one hand, we have $\|\Pi^\perp_m\beta\|_\gamma = o(1)$ as $m \to \infty$ by Lebesgue's dominated convergence theorem. On the other hand, from the identity $[\Pi_m\beta - \beta_m]_{\underline{m}} = -[\Gamma]^{-1}_{\underline{m}}[\Gamma\Pi^\perp_m\beta]_{\underline{m}}$ we conclude $\|\Pi_m\beta - \beta_m\|^2_\gamma \leqslant 2(1 + d^2)\|\Pi^\perp_m\beta\|^2_\gamma$ for all $\Gamma \in \mathcal{N}^d_\upsilon$, because the estimate (B.8) in Lemma B.2 implies $\sup_{\|f\|_\gamma = 1}\|[\mathrm{Diag}(\gamma)]^{1/2}_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}([\Gamma f]_{\underline{m}} - [\Gamma]_{\underline{m}}[f]_{\underline{m}})\|^2 \leqslant 2(1 + d^2)$. By combining the two results, we obtain the assertion. It remains to show that (3.2) can be substituted by (3.3). Due to (B.5) in Lemma B.2 the link condition $\Gamma \in \mathcal{N}^d_\upsilon$ implies $\upsilon_m\|[\Gamma]^{-1}_{\underline{m}}\|_S \leqslant 4d^3$. Since $1/\upsilon_m = o(n)$ we conclude that $\|[\Gamma]^{-1}_{\underline{m}}\|_S = o(n)$ as $n \to \infty$. Furthermore, from (B.6) in Lemma B.2 we have $[\ell]^t_{\underline{m}}\,[\Gamma]^{-1}_{\underline{m}}[\ell]^t_{\underline{m}} \leqslant \sum_{j=1}^m [\ell]^2_j \upsilon_j^{-1}\|[\mathrm{Diag}(\upsilon)]^{1/2}_{\underline{m}}[\Gamma]^{-1/2}_{\underline{m}}\|^2_S \leqslant 4d^3\sum_{j=1}^m[\ell]^2_j\upsilon_j^{-1}$. Thus $\sum_{j=1}^m[\ell]^2_j\upsilon_j^{-1} = o(n)$ implies $[\ell]^t_{\underline{m}}\,[\Gamma]^{-1}_{\underline{m}}[\ell]^t_{\underline{m}} = o(n)$ as $n \to \infty$, which proves the result. □

A.2. Proof of the lower bound

Proof of Theorem 3.3. Consider ε and X with $\Gamma \in \mathcal{N}_\upsilon^d$ such that Assumption 2.2 is satisfied and $\varepsilon|\langle X, f\rangle \sim \mathcal{N}(0,1)$ holds true for all $f \in \mathbb{H}$. Assume i.i.d. copies $\{(\varepsilon_i, X_i)\}_{i=1}^n$ of (ε, X) and let $\beta_* \in \mathcal{F}_\gamma^\rho$ with $2dn\|\beta_*\|_\upsilon^2 \leqslant \sigma^2$, to be specified below. Obviously, for each $\theta \in \{-1,1\}$ the function $\beta_\theta := \theta\,\beta_*$ belongs also to $\mathcal{F}_\gamma^\rho$ and the random variables $\{(Y_i, X_i)\}_{i=1}^n$ with $Y_i := \langle \beta_\theta, X_i\rangle + \sigma\varepsilon_i$ form a sample of the model (1.1). We denote its joint distribution by $\mathbb{P}_\theta$. We observe that the conditional distribution of Y_i given X_i is Gaussian with mean $\theta\langle\beta_*, X_i\rangle$ and variance σ^2. Thereby, it is easily seen that the expectation of the log-likelihood function of $\mathbb{P}_1$ with respect to $\mathbb{P}_{-1}$ satisfies

$$\mathbb{E}_{\mathbb{P}_{-1}}[\log(d\mathbb{P}_1/d\mathbb{P}_{-1})] = (2n/\sigma^2)\,\langle\Gamma\beta_*, \beta_*\rangle = (2n/\sigma^2)\|\Gamma^{1/2}\beta_*\|_{\mathbb{H}}^2.$$

In terms of Kullback-Leibler divergence this means that the inequality $\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_{-1}) \leqslant (2dn/\sigma^2)\|\beta_*\|_\upsilon^2 \leqslant 1$ holds true, by using the inequality of Heinz [13], i.e., for all $|s| \leqslant 1$, $f \in \mathbb{H}$ and $T \in \mathcal{N}_\upsilon^d$ we have $\|T^s f\|_{\mathbb{H}}^2 \leqslant d^{2|s|}\|f\|_{\upsilon^{2s}}^2$, together with the condition $2dn\|\beta_*\|_\upsilon^2 \leqslant \sigma^2$. Due to the bounded Kullback-Leibler divergence, Le Cam's general method (see Le Cam [18]) and Pinsker's inequality allow us to derive a lower bound. However, in this special setting a lower bound can be obtained by the following elementary steps. We consider the Hellinger affinity $\rho(\mathbb{P}_1, \mathbb{P}_{-1}) = \int\sqrt{d\mathbb{P}_1 d\mathbb{P}_{-1}}$ and obtain for any estimator $\breve{\ell}$ and for all $\ell \in \mathcal{L}_{1/\gamma}$ that

$$\begin{aligned}\rho(\mathbb{P}_1, \mathbb{P}_{-1}) &\leqslant \int \frac{|\breve{\ell} - \ell(\beta_1)|}{2|\ell(\beta_*)|}\sqrt{d\mathbb{P}_1 d\mathbb{P}_{-1}} + \int \frac{|\breve{\ell} - \ell(\beta_{-1})|}{2|\ell(\beta_*)|}\sqrt{d\mathbb{P}_1 d\mathbb{P}_{-1}} \\ &\leqslant \Big(\int \frac{|\breve{\ell} - \ell(\beta_1)|^2}{4|\ell(\beta_*)|^2} d\mathbb{P}_1\Big)^{1/2} + \Big(\int \frac{|\breve{\ell} - \ell(\beta_{-1})|^2}{4|\ell(\beta_*)|^2} d\mathbb{P}_{-1}\Big)^{1/2}. \qquad \text{(A.5)}\end{aligned}$$

By using the identity $\rho(\mathbb{P}_1, \mathbb{P}_{-1}) = 1 - \frac{1}{2}\mathrm{H}^2(\mathbb{P}_1, \mathbb{P}_{-1})$ it follows from (A.5) that

$$\Big\{\mathbb{E}_{\mathbb{P}_1}|\breve{\ell} - \ell(\beta_1)|^2 + \mathbb{E}_{\mathbb{P}_{-1}}|\breve{\ell} - \ell(\beta_{-1})|^2\Big\} \geqslant \frac{1}{2}|\ell(\beta_*)|^2 \qquad \text{(A.6)}$$

since the Hellinger distance $\mathrm{H}(\mathbb{P}_1, \mathbb{P}_{-1})$ between $\mathbb{P}_1$ and $\mathbb{P}_{-1}$ satisfies $\mathrm{H}^2(\mathbb{P}_1, \mathbb{P}_{-1}) \leqslant \mathrm{KL}(\mathbb{P}_1, \mathbb{P}_{-1}) \leqslant 1$. From (A.6) we conclude for each estimator $\breve{\ell}$ that

$$\begin{aligned}\sup_{\beta \in \mathcal{F}_\gamma^\rho} \mathbb{E}|\breve{\ell} - \ell(\beta)|^2 &\geqslant \sup_{\theta \in \{-1,1\}} \mathbb{E}_{\mathbb{P}_\theta}|\breve{\ell} - \ell(\beta_\theta)|^2 \\ &\geqslant \frac{1}{2}\Big\{\mathbb{E}_{\mathbb{P}_1}|\breve{\ell} - \ell(\beta_1)|^2 + \mathbb{E}_{\mathbb{P}_{-1}}|\breve{\ell} - \ell(\beta_{-1})|^2\Big\} \geqslant \frac{1}{4}|\ell(\beta_*)|^2. \qquad \text{(A.7)}\end{aligned}$$

We will obtain the claimed result of the theorem by evaluating (A.7) for two special choices of $\beta_* \in \mathcal{F}_\gamma^\rho$ with $2dn\|\beta_*\|_\upsilon^2 \leqslant \sigma^2$, which we will construct in the following. Define $\zeta := \min(\frac{\sigma^2}{2d}, \rho)$ and let κ be given by (3.5). On the one hand, consider the slope function $\beta_* := \sum_{j=1}^{m_n^*} [\beta_*]_j\,\psi_j$, with coefficients $[\beta_*]_j :=$

$[\ell]_j \upsilon_j^{-1}(\zeta\kappa K_n^*)^{1/2}(\sum_{j=1}^{m_n^*}[\ell]_j^2\upsilon_j^{-1})^{-1/2}$. Since υ/γ is monotonically decreasing and by using the definition of κ and ζ it follows that $\|\beta_*\|_\gamma^2 \leqslant \zeta\kappa K_n^*\gamma_{m_n^*}\upsilon_{m_n^*}^{-1} \leqslant \zeta \leqslant \rho$ and, hence $\beta_* \in \mathcal{F}_\gamma^\rho$. Furthermore, we have that $2\,d\,n\|\beta_*\|_\upsilon^2 = 2d\zeta\kappa K_n^* n \leqslant 2d\zeta \leqslant \sigma^2$. Obviously, by evaluating (A.7) we conclude $\sup_{\beta\in\mathcal{F}_\gamma^\rho}\mathbb{E}|\breve{\ell}-\ell(\beta)|^2 \geqslant (\kappa/4)\zeta K_n^*\sum_{j=1}^{m_n^*}[\ell]_j^2\upsilon_j^{-1}$. On the other hand, consider $\beta_* := \sum_{j>m_n^*}[\beta_*]\psi_j$ with $[\beta_*]_j := (\kappa\zeta)^{1/2}(\gamma_j^2\sum_{j>m_n^*}[\ell]_j^2\gamma_j^{-1})^{-1/2}[\ell]_j$ we conclude from $\kappa \leqslant 1$ and $\|\beta_*\|_\gamma^2 = \sum_{j>m_n^*}[\beta_*]_j^2\gamma_j = \zeta\kappa \leqslant \rho$ that β_* belongs to $\mathcal{F}_\gamma^\rho$. Moreover, we have $2nd\|\beta_*\|_\upsilon^2 \leqslant 2nd\zeta\kappa\upsilon_{m_n^*}\gamma_{m_n^*}{}^{-1} \leqslant 2d\zeta \leqslant \sigma^2$. By evaluating (A.7) we obtain $\sup_{\beta\in\mathcal{F}_\gamma^\rho}\mathbb{E}|\breve{\ell} - \ell(\beta)|^2 \geqslant (\kappa/4)\zeta\sum_{j>m_n^*}[\ell]_j^2\gamma_j^{-1}$. Combining the two lower bounds, which hold true for arbitrary $\Gamma \in \mathcal{N}_\upsilon^d$, we obtain

$$\inf_{\breve{\ell}}\inf_{\Gamma\in\mathcal{N}_\upsilon^d}\sup_{\beta\in\mathcal{F}_\gamma^\rho}\mathbb{E}|\breve{\ell}-\ell(\beta)|^2 \geqslant \frac{\kappa}{4}\zeta\max\Big\{K_n^*\sum_{j=1}^{m_n^*}[\ell]_j^2\upsilon_j^{-1}, \sum_{j>m_n^*}[\ell]_j^2\gamma_j^{-1}\Big\},$$

which implies the result of the theorem. □

Proof of Proposition 3.5. We start our proof with the observation that under the conditions on p, a and s given in the proposition the sequences γ, υ and ω satisfy Assumption 2.1 and Assumption 3.1.

Proof of *(ppp)*. From the definition of m_n^* in (3.4) it follows that $m_n^* \sim n^{1/(2a+2p)}$. Consider case *(i)*. The condition $s-a<1/2$ implies $n^{-1}\sum_{j=1}^{m_n^*}|j|^{2a-2s} \sim n^{-1}(m_n^*)^{2a-2s+1} \sim n^{-(2p+2s-1)/(2p+2a)}$ and moreover we have $\sum_{j>m_n^*}|j|^{-2p-2s} \sim n^{-(2p+2s-1)/(2p+2a)}$ since $p+s>1/2$. If $s-a=1/2$, then $n^{-1}\sum_{j=1}^{m_n^*}|j|^{2a-2s} \sim n^{-1}\log(n^{1/(2p+2a)})$ and $\sum_{j>m_n^*}|j|^{-2p-2s} \sim n^{-1}$. In the case $s-a>1/2$ it follows that $\sum_{j=1}^{m_n^*}|j|^{2a-2s}$ is bounded. Moreover, there exists a constant $c>0$ such that $\sum_{j>m_n^*}|j|^{-2p-2s} \leqslant c\cdot n^{-1}$, or $\sum_{j>m_n^*}|j|^{-2p-2s} \lesssim n^{-1}$ for short, and hence $\delta_n^* \sim n^{-1}$. To prove *(ii)* we make use of Corollary 3.4. We observe that if $s-a \geqslant 0$ the sequence $\omega\upsilon$ is bounded from below, and hence $\delta_n^* \sim n^{-1}$. On the other hand, the condition $s-a<0$ implies $\delta_n^* \sim n^{-(p+s)/(p+a)}$.

Proof of *(pep)*. If υ is exponentially decreasing, then m_n^* satisfies $\exp(-(m_n^*)^{2a}) \sim n^{-1}(m_n^*)^{2p}$ or equivalently $m_n^* \sim \log(n[\log(n)]^{-p/a})^{1/(2a)}$. To prove *(i)*, we calculate $\sum_{j>m_n^*}|j|^{-2p-2s} \sim [\log(n)]^{(-2p-2s+1)/(2a)}$ and $n^{-1}\sum_{j=1}^{m_n^*}\exp(|j|^{2a})|j|^{-2s} \lesssim n^{-1}\exp(m_n^{*\,2a}) \sim [\log(n)]^{(-2p-2s+1)/(2a)}$. In case *(ii)* we immediately obtain $\delta_n^* \sim [\log(n)]^{-(p+s)/a}$.

Proof of *(epp)*. Only γ is an exponential sequence and hence we have $m_n^* \sim n^{-1}\exp((m_n^*)^{2p})$ or equivalently $m_n^* \sim \log(n[\log(n)^{-a/p}])^{1/(2p)}$. Consider case *(i)*. If $s-a<1/2$, then $n^{-1}\sum_{j=1}^{m_n^*}|j|^{2a-2s} \sim n^{-1}[\log(n)]^{(2a-2s+1)/(2p)}$. If $s-a = 1/2$, then $n^{-1}\sum_{j=1}^{m_n^*}|j|^{2a-2s} \sim n^{-1}\log(\log(n))$. On the other hand, the condition $s-a>1/2$ implies that $\sum_{j=1}^{m_n^*}|j|^{2a-2s}$ is bounded and thus, we obtain the parametric rate n^{-1}. By using $\sum_{j>m_n^*}|j|^{-2s}\exp(-|j|^{2p}) \lesssim \exp(-(m_n^*)^{2p})(m_n^*)^{2s-2p+1} \sim n^{-1}[\log(n)]^{(-2a-2s-2p+1)/(2p)}$ it is easily seen that this sum is dominated by

$n^{-1}\sum_{j=1}^{m_n^*}|j|^{2a-2s}$. In case *(ii)* if $s-a\geqslant 0$ then the sequence $\omega\upsilon$ is bounded from below as mentioned above and thus, $\delta_n^*\sim n^{-1}$. If $s-a<0$ then $\delta_n^*\sim n^{-1}[\log(n)]^{(a-s)/p}$.
Proof of *(ppe)*. As in case of *(ppp)* both sequences γ and υ are polynomial and thus $m_n^*\sim n^{1/(2p+2a)}$. Consider case *(i)* where the coefficients of h decrease exponentially. Obviously the sum $\sum_{j=1}^{m_n^*}|j|^{2a}\exp(-|j|^{2s})$ is bounded and moreover, $\sum_{j>m_n^*}|j|^{-2p}\exp(-|j|^{2s})\lesssim\exp(-|m_n^*|^{2s})|m_n^*|^{-2s-2p+1}\lesssim n^{-1}$. Consequently, we have $\delta_n^*\sim n^{-1}$. Also in case *(ii)* it obviously holds $\delta_n^*\sim n^{-1}$, which completes the proof. $\square$

A.3. Proof of the upper bound

The following technical lemma is used in the proof of Theorem 3.6.

Lemma A.1. *If the assumptions of Theorem 3.6 hold true, then we have* $P(\Omega^c)\leqslant C(\gamma,d,\upsilon)\,\eta\,n^{-1}$.

Proof. Our proof starts with the observations that $\kappa\gamma_{m_n^*}\leqslant n\upsilon_{m_n^*}$ for all $n\geqslant 1$ by exploiting Assumption 3.1 and that $2\|[\Gamma]_{\underline{m}}^{-1}\|_S\leqslant 8d^3\upsilon_m^{-1}$ for all $\Gamma\in\mathcal{N}_\upsilon^d$ and $m\geqslant 1$ due to (B.5) in Lemma B.2. Combining both estimates with $\gamma_{m_n^*}^{-1}=o(1)$ as $n\to\infty$ we conclude $2\|[\Gamma]_{\underline{m_*}}^{-1}\|_S=o(n)$. Therefore, there exists an integer $n_0:=n_0(\gamma,d,\upsilon)$ such that for all $n\geqslant n_0$ we have $2\|[\Gamma]_{\underline{m_*}}^{-1}\|_S\leqslant n^{-1}$, and particularly $\Omega_{1/2}\subset\Omega$ by applying the usual Neumann series argument. We distinguish in the following the cases $n<n_0$ and $n\geqslant n_0$. Consider first $n\geqslant n_0$, from Markov's inequality together with (B.3) in Lemma B.1 we obtain $P(\Omega^c)\leqslant P(\Omega_{1/2}^c)\leqslant C\eta(m_n^*)^6n^{-3}$. Taking into account the condition (3.7), that is $D:=D(\gamma,\upsilon):=\sup_{m\in\mathbb{N}}\{\frac{\upsilon_m}{\gamma_m}m^3\}<\infty$, we have $(m_n^*)^6n^{-3}\leqslant((m_n^*)^3\upsilon_{m_n^*}\gamma_{m_n^*}^{-1}\kappa^{-1})^2n^{-1}\leqslant D^2\kappa^{-2}n^{-1}$, and hence $P(\Omega^c)\leqslant C\eta D^2\kappa^{-2}n^{-1}$ for all $n\geqslant n_0$. On the other hand, if $n<n_0$ then trivially $P(\Omega^c)\leqslant n^{-1}n_0$. Since n_0, D and κ depend on γ,d and υ only we obtain the result by combining both cases, which completes the proof. $\square$

Proof of Theorem 3.6. Consider again the bound (A.3). By using that $\|\Gamma\|_S\leqslant d$ for $\Gamma\in\mathcal{N}_\upsilon^d$, $(m_n^*)^3n^{-1}\leqslant D\kappa^{-1}$ with $D=\sup_{m\in\mathbb{N}}\{\frac{\upsilon_m}{\gamma_m}m^3\}<\infty$ and recalling that $P(\Omega_{1/2}^c)\leqslant C\eta(m_n^*)^{24}n^{-12}$ we obtain

$$\mathbb{E}|\widehat{\ell}_{m_n^*}-\ell(\beta)|^2\leqslant C\Big\{|\ell(\beta-\beta_{m_n^*})|^2+|\ell(\beta)|^2P(\Omega^c)$$
$$+(\|[\ell]_{\underline{m_*}}^t[\Gamma]_{\underline{m_*}}^{-1/2}\|^2/n)\,\eta\,\{\|\Gamma^{1/2}(\beta-\beta_{m_n^*})\|_{\mathbb{H}}^2+\sigma^2\}\Big[1+D\kappa^{-1}+d^2D^5\kappa^{-5}\Big]\Big\}.$$

From (B.6), (B.9) and (B.10) in Lemma B.2 we conclude $(\|[\ell]_{\underline{m_*}}^t[\Gamma]_{\underline{m_*}}^{-1/2}\|^2/n)\leqslant 4d^3\delta_n^*$, furthermore $\|\Gamma^{1/2}(\beta-\beta_{m_n^*})\|_{\mathbb{H}}^2\leqslant 10d^5\|\beta\|_\gamma^2$ and $|\ell(\beta-\beta_{m_n^*})|^2\leqslant 16\|\beta\|_\gamma^2d^4\delta_n^*$, respectively. Therefore we conclude for all $\beta\in\mathcal{F}_\gamma^\rho$

$$\mathbb{E}|\widehat{\ell}_{m_n^*} - \ell(\beta)|^2 \leqslant C\left\{\rho d^4 \delta_n^* + \rho\|\ell\|_{1/\gamma}^2 P(\Omega^c) + d^3\,\eta\,\delta_n^*\{\rho d^5 + \sigma^2\}\left[1 + D\kappa^{-1} + d^2 D^5 \kappa^{-5}\right]\right\}.$$

Observe that D and κ depend on υ and γ only and thus by applying Lemma A.1 we obtain

$$\mathbb{E}|\widehat{\ell}_{m_n^*} - \ell(\beta)|^2 \leqslant C(d,\upsilon,\gamma)\left\{\rho\delta_n^* + \rho\,\eta\,\|\ell\|_{1/\gamma}^2\, n^{-1} + \eta\,\{\rho + \sigma^2\}\delta_n^*\right\},$$

which completes the proof. □

Proof of Proposition 3.8. Under the stated conditions it is easy to verify that the assumptions of Theorem 3.6 are satisfied. The result follows by applying Theorem 3.6 and Corollary 3.7 and we omit the details. □

Appendix B: Technical assertions

The following two lemmata gather technical results used in the proof of Proposition 3.1 and Theorem 3.6.

Lemma B.1. *Under Assumption 2.2 there exists a constant $C(k) > 0$ such that*

$$\sup_{z\in\mathbb{R}^m : z^t z = 1} \mathbb{E}\Big|z^t[\Gamma]_{\underline{m}}^{-1/2}[Z]_{\underline{m}}\Big|^{2k} \leqslant C(k)\, n^{-k}\left(\sigma^2 + \|\Gamma^{1/2}(\beta - \beta_m)\|_{\mathbb{H}}^2\right)^k \eta, \quad \text{(B.1)}$$

$$\mathbb{E}\|[\Gamma]_{\underline{m}}^{-1/2}[Z]_{\underline{m}}\|^{2k} \leqslant C(k)\,\frac{m^k}{n^k}\left(\sigma^2 + \|\Gamma^{1/2}(\beta - \beta_m)\|_{\mathbb{H}}^2\right)^k \eta, \quad \text{(B.2)}$$

$$\mathbb{E}\|[\Xi]_{\underline{m}}\|_S^{2k} \leqslant C(k)\cdot\eta\cdot\frac{m^{2k}}{n^k}. \quad \text{(B.3)}$$

Proof. Let us begin by deriving elementary bounds due to Assumption 2.2. For $m \geqslant 1$ define $U := \sigma\varepsilon + \langle\beta - \beta_m, X\rangle$, where $\sigma_U^2 = \mathbb{V}\mathrm{ar}(U) = \sigma^2 + \|\Gamma^{1/2}(\beta - \beta_m)\|_{\mathbb{H}}^2$. It is easily seen that for k as given in Assumption 2.2 and all $m \geqslant 1$ we have

$$\mathbb{E}|U|^{4k} \leqslant C(k)\sigma_U^{4k}\eta, \qquad \max_{1\leqslant j\leqslant m}\mathbb{E}|([\Gamma]_{\underline{m}}^{-1/2}[X]_{\underline{m}})_j|^{4k} \leqslant \eta$$
$$\text{and} \quad \sup_{z\in\mathbb{R}^m : z^t z = 1}\mathbb{E}\Big|z^t[\Gamma]_{\underline{m}}^{-1/2}[X]_{\underline{m}}\Big|^{4k} \leqslant \eta. \quad \text{(B.4)}$$

Let $z \in \mathbb{R}^m$ satisfy $z^t z = 1$ and define $U_i := \sigma\varepsilon_i + \langle\beta - \beta_m, X_i\rangle$ then $z^t[\Gamma]_{\underline{m}}^{-1/2}[Z]_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n U_i z^t[\Gamma]_{\underline{m}}^{-1/2}[X_i]_{\underline{m}}$. Since $\mathbb{E}\langle\beta - \beta_m, X\rangle[X]_{\underline{m}} = [\Gamma(\beta - \beta_m)]_{\underline{m}} = [g]_{\underline{m}} - [\Gamma]_{\underline{m}}[\beta_m]_{\underline{m}} = 0$, it follows that the random variables $U_i z^t[\Gamma]_{\underline{m}}^{-1/2}[X_i]_{\underline{m}}$, $i = 1,\dots,n$, are i.i.d. with mean zero. From Theorem 2.10 in Petrov [26] we conclude $\mathbb{E}|z^t[\Gamma]_{\underline{m}}^{-1/2}[Z]_{\underline{m}}|^{2k} \leqslant C(k)n^{-k}\mathbb{E}|Uz^t[\Gamma]_{\underline{m}}^{-1/2}[X]_{\underline{m}}|^{2k}$ for some constant $C(k) > 0$. Then we claim that (B.1) follows from the Cauchy-Schwarz inequality together with the bounds given in (B.4). To deduce (B.2) from (B.1) we use that

$$\mathbb{E}\|[\Gamma]_{\underline{m}}^{-1/2}[Z]_{\underline{m}}\|^{2k} \leqslant m^k \max_{1\leqslant j\leqslant m}\mathbb{E}\Big|([\Gamma]_{\underline{m}}^{-1/2}[Z]_{\underline{m}})_j\Big|^{2k}$$
$$\leqslant m^k \sup_{z\in\mathbb{R}^m : z^t z = 1}\mathbb{E}\Big|z^t[\Gamma]_{\underline{m}}^{-1/2}[Z]_{\underline{m}}\Big|^{2k}.$$

Proof of (B.3). From the identity $n([\Xi]_{\underline{m}})_{j,l} = \sum_{i=1}^{n}\{([\Gamma]_{\underline{m}}^{-1/2}[X_i]_{\underline{m}})_j([\Gamma]_{\underline{m}}^{-1/2} \times [X_i]_{\underline{m}})_l - \delta_{jl}\}$ with $\delta_{jl} = 1$ if $j = l$ and zero otherwise, we conclude $\mathbb{E}([\Xi]_{\underline{m}})_{j,l}^{2k} \leqslant C(k)n^{-k}\eta$ by using Theorem 2.10 in Petrov [26] and the second bound given in (B.4). Then using the elementary inequality $\mathbb{E}\|[\Xi]_{\underline{m}}\|_S^{2k} \leqslant m^{2k}\max_{1\leqslant j,l\leqslant m}\mathbb{E}\times([\Xi]_{\underline{m}})_{j,l}^{2k}$ implies (B.3), which completes the proof. □

The next Lemma is partially shown in Cardot and Johannes [4].

Lemma B.2. *Suppose the sequences γ and υ satisfy Assumption 2.1. Then we have for $T \in \mathcal{N}_\upsilon^d$*

$$\sup_{m\geqslant 1}\Big\{\upsilon_m\|[T]_{\underline{m}}^{-1}\|_S\Big\} \leqslant \{2d^2(2d^4+3)\}^{1/2} \leqslant 4d^3, \tag{B.5}$$

$$\sup_{m\geqslant 1}\|[T]_{\underline{m}}^{-1/2}[\mathrm{Diag}(\upsilon)]_{\underline{m}}^{1/2}\|_S^2 \leqslant \{2d^2(2d^4+3)\}^{1/2} \leqslant 4d^3. \tag{B.6}$$

If in addition β_m denotes a Galerkin solution of $g = T\beta$ then

$$\sup_{m\geqslant 1}\Big\{\sup_{\|\beta\|_{\mathbb{H}}=1}\|\Pi_m\beta - \beta_m\|_{\mathbb{H}}^2\Big\} \leqslant 2(1+d^2), \tag{B.7}$$

$$\sup_{m\geqslant 1}\Big\{\sup_{\|\beta\|_{\gamma}=1}\|\Pi_m\beta - \beta_m\|_{\gamma}^2\Big\} \leqslant 2(1+d^2), \tag{B.8}$$

and if $\beta \in \mathcal{F}_\gamma^\rho$ is additionally satisfied then

$$\sup_{m\geqslant 1}\{\upsilon_m^{-1}\gamma_m\,\|\Gamma^{1/2}(\beta-\beta_m)\|_{\mathbb{H}}^2\} \leqslant 10d^5\rho. \tag{B.9}$$

Furthermore, for all $m \geqslant 1$ and all $\ell \in \mathcal{L}_{1/\gamma}$ we have

$$|\ell(\beta-\beta_m)|^2 \leqslant 2\|\beta\|_\gamma^2\{\sum_{j>m}[\ell]_j^2\gamma_j^{-1} + 2(1+d^4)\frac{\upsilon_m}{\gamma_m}\sum_{j=1}^{m}[\ell]_j^2\upsilon_j^{-1}\}. \tag{B.10}$$

Proof. The estimates (B.5) - (B.6) are given in Lemma A.3 in Cardot and Johannes [4]. Furthermore, from (A.19) and (A.20) in Lemma A.3 in Cardot and Johannes [4] follow (B.7) and (B.8). We start our proof of (B.9) with the observation that the link condition $T \in \mathcal{N}_\upsilon^d$ implies that T is strictly positive and that for all $|s| \leqslant 1$ by using the inequality of Heinz [13]

$$d^{-2|s|}\|f\|_{\upsilon^{2s}}^2 \leqslant \|T^s f\|_{\mathbb{H}}^2 \leqslant d^{2|s|}\|f\|_{\upsilon^{2s}}^2. \tag{B.11}$$

Thus, by using successively the first inequality of (B.11), the Galerkin condition (2.1) and the second inequality of (B.11), we obtain

$$\|\beta-\beta_m\|_{\upsilon^2}^2 \leqslant d^2\|T(\beta-\beta_m)\|_{\mathbb{H}}^2 \leqslant d^2\|T(\beta-\Pi_m\beta)\|_{\mathbb{H}}^2 \leqslant d^4\|\beta-\Pi_m\beta\|_{\upsilon^2}^2 \tag{B.12}$$

Since $\beta \in \mathcal{F}_\gamma^\rho$ and $\gamma^{-1}\upsilon^2$ is monotonically decreasing we have $\|\beta - \Pi_m\beta\|_{\upsilon^2}^2 \leqslant \gamma_m^{-1}\upsilon_m^2\|\beta\|_\gamma^2$, which together with (B.12) implies $\|\beta-\beta_m\|_{\upsilon^2}^2 \leqslant d^4\gamma_m^{-1}\upsilon_m^2\|\beta\|_\gamma^2$ and hence,

$$\|\Pi_m\beta-\beta_m\|_{\upsilon^2}^2 \leqslant 2\{\|\beta-\beta_m\|_{\upsilon^2}^2 + \|\beta-\Pi_m\beta\|_{\upsilon^2}^2\} \leqslant 2(1+d^4)\gamma_m^{-1}\upsilon_m^2\|\beta\|_\gamma^2. \tag{B.13}$$

The last estimate and the second inequality of (B.11) imply further $\|\Gamma^{1/2}(\Pi_m\beta-\beta_m)\|_{\mathbb{H}}^2 \leqslant d\|\Pi_m\beta-\beta_m\|_\upsilon^2 \leqslant d\upsilon_m^{-1}\|\Pi_m\beta-\beta_m\|_{\upsilon^2}^2 \leqslant 2d(1+d^4)\gamma_m^{-1}\upsilon_m\|\beta\|_\gamma^2$ because υ is monotonically non increasing. Taking into account $\|\beta-\Pi_m\beta\|_\upsilon^2 \leqslant \gamma_m^{-1}\upsilon_m\|\beta\|_\gamma^2$ we obtain (B.9). Finally, by applying the Cauchy-Schwarz inequality we have on the one hand $|\ell(\beta-\Pi_m\beta)|^2 \leqslant \|\beta\|_\gamma^2 \sum_{j>m}[\ell]_j^2\gamma_j^{-1}$ and by using (B.13) it follows on the other hand $|\ell(\Pi_m\beta-\beta_m)|^2 \leqslant \|\Pi_m\beta-\beta_m\|_\upsilon^2 \sum_{j=1}^m[\ell]_j^2\upsilon_j^{-1} \leqslant 2(1+d^4)\|\beta\|_\gamma^2\gamma_m^{-1}\upsilon_m \sum_{j=1}^m[\ell]_j^2\upsilon_j^{-1}$. Combining both estimates implies now (B.10), which completes the proof. □

Acknowledgements

This work was supported by the IAP research network no. P6/03 of the Belgian Government (Belgian Science Policy) and by the "Fonds Spéciaux de Recherche" from the Université catholique de Louvain.

References

[1] Bosq, D. *Linear Processes in Function Spaces.*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, 2000. MR1783138
[2] Cai, T. and Hall, P. Prediction in functional linear regression. *Annals of Statistics*, 34(5):2159–2179, 2006. MR2291496
[3] Cardot, H., Ferraty, F., and Sarda, P. Spline estimators for the functional linear model. *Statistica Sinica*, 13:571–591, 2003. MR1997162
[4] Cardot, H. and Johannes, J. Thresholding projection estimators in functional linear models. *Journal of Multivariate Analysis*, 101(2):395–408, 2010. MR2564349
[5] Cardot, H., Mas, A., and Sarda, P. CLT in functional linear regression models. *Probability Theory and Related Fields*, 138:325–361, 2007. MR2299711
[6] Crambes, C., Kneip, A., and Sarda, P. Smoothing splines estimators for functional linear regression. *Annals of Statistics*, 37(1):35–72, 2009. MR2488344
[7] Donoho, D. Statistical estimation and optimal recovery. *Annals of Statistics*, 22:238–270, 1994. MR1272082
[8] Donoho, D. and Low, M. Renormalization exponents and optimal pointwise rates of convergence. *Annals of Statistics*, 20:944–970, 1992. MR1165601
[9] Efromovich, S. and Koltchinskii, V. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001. MR1872847
[10] Eilers, P. H. and Marx, B. D. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11:89–102, 1996. MR1435485
[11] Goldenshluger, A. and Pereverzev, S. V. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probability Theory and Related Fields*, 118:169–186, 2000. MR1790080

[12] Hall, P. and Horowitz, J. L. Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 35(1):70–91, 2007. MR2332269

[13] Heinz, E. Beiträge zur Störungstheorie der Spektralzerlegung. *Mathematische Annalen*, 123:415–438, 1951. MR0044747

[14] Hoffmann, M. and Reiß, M. Nonlinear estimation for linear inverse problems with error in the operator. *Annals of Statistics*, 36(1):310–336, 2008. MR2387973

[15] Ibragimov, I. and Has'minskii, R. On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory of Probability and its Applications*, 29:18–32, 1984. MR0739497

[16] Johannes, J. and Kroll, M. On rate optimal estimation of linear functionals for linear inverse problems with error in the operator. Technical report, University Heidelberg, 2010.

[17] Kawata, T. *Fourier analysis in probability theory.* Academic Press, New York, 1972. MR0464353

[18] Le Cam, L. Convergence of Estimates Under Dimensionality Restrictions. *Annals of Statistics*, 1(1):38–53, 1973. MR0334381

[19] Li, K. Minimaxity of the method of regularization of stochastic processes. *Annals of Statistics*, 10:937–942, 1982. MR0663444

[20] Mair, B. A. Tikhonov regularization for finitely and infinitely smoothing operators. *SIAM Journal on Mathematical Analysis*, 25:135–147, 1994. MR1257145

[21] Meister, A. Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Annals of Statistics*, 39(3):1471–1495, 2011. MR2850209

[22] Müller, H.-G. and Stadtmüller, U. Generalized functional linear models. *Annals of Statistics*, 33:774–805, 2005. MR2163159

[23] Natterer, F. Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Analysis*, 18:29–37, 1984. MR0762862

[24] Neubauer, A. An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates. *SIAM Journal on Numerical Analysis*, 25(6):1313–1326, 1988. MR0972456

[25] Neubauer, A. When do Sobolev spaces form a Hilbert scale? *Proceedings of the American Mathematical Society*, 103(2):557–562, 1988. MR0943084

[26] Petrov, V. V. *Limit theorems of probability theory. Sequences of independent random variables.* Oxford Studies in Probability. Clarendon Press, Oxford, 4. edition, 1995. MR1353441

[27] Ramsay, J. O. and Dalzell, C. J. Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B*, 53:539–572, 1991. MR1125714

[28] Speckman, P. Minimax estimation of linear functionals in a Hilbert space. Unpublished manuscript, 1979.