T E C H N I C A L
R E P O R T

11040

Adaptive estimation of linear functionals in functional linear models

JOHANNES, J. and R. SCHENK

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Adaptive estimation of linear functionals in functional linear models

Jan Johannes* Rudolf Schenk*

Université catholique de Louvain

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. In Johannes and Schenk [2010] it has been shown that a plug-in estimator based on dimension reduction and additional thresholding can attain minimax optimal rates of convergence up to a constant. However, this estimation procedure requires an optimal choice of a tuning parameter with regard to certain characteristics of the slope function and the covariance operator associated with the functional regressor. As these are unknown in practice, we investigate a fully data-driven choice of the tuning parameter based on a combination of model selection and Lepski's method, which is inspired by the recent work of Goldenshluger and Lepski [2011]. The tuning parameter is selected as the minimizer of a stochastic penalized contrast function imitating Lepski's method among a random collection of admissible values. We show that this adaptive procedure attains the lower bound for the minimax risk up to a logarithmic factor over a wide range of classes of slope functions and covariance operators. In particular, our theory covers point-wise estimation as well as the estimation of local averages of the slope parameter.

Keywords: Adaptation, Linear functional, Lepski's method, Model selection, Linear Galerkin projection, Minimax-theory, Point-wise estimation, Local average estimation, Sobolev space.

AMS 2000 subject classifications: Primary 62J05; secondary 62G05, 62G20

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.

1 Introduction

The functional linear model with scalar response describes the relationship between a real random variable Y and the variation of a functional regressor X. Usually, the random function X is assumed to be square integrable or more generally to take its values in a separable Hilbert

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

space $\mathbb{H}$ with the inner product $\langle\cdot,\cdot\rangle_{\mathbb{H}}$ and associated norm $\|\cdot\|_{\mathbb{H}}$. For convenient notations we assume that the regressor X is centered in the sense that for all $h \in \mathbb{H}$ the real valued random variable $\langle X, h\rangle_{\mathbb{H}}$ has mean zero. The linear relationship between Y and X is expressed by the equation

$$Y = \langle \phi, X\rangle_{\mathbb{H}} + \sigma\varepsilon, \quad \sigma > 0, \tag{1.1}$$

with the unknown slope parameter $\phi \in \mathbb{H}$ and a real-valued, centered and standardized error term ε. The objective of this paper is the fully data-driven estimation of the value of a known linear functional of the slope ϕ based on an independent and identically distributed (i.i.d.) sample of (Y, X) of size n.

The estimation of the value of a linear functional offers a general framework for naturally arising related estimation problems, such as estimating the value of ϕ - or of one of its derivatives - at a given point or estimating the average of ϕ over a subinterval of its domain.

There is extensive literature available on the topic of non-parametric estimation of the value of a linear functional from Gaussian white noise observations (in case of direct observations see Speckman [1979], Li [1982] or Ibragimov and Has'minskii [1984], while in case of indirect observations we refer to Donoho and Low [1992], Donoho [1994] or Goldenshluger and Pereverzev [2000] and references therein). In the situation of a functional linear model as considered in (1.1), which does in general not lead to Gaussian white noise observations, Johannes and Schenk [2010] have investigated the minimax optimal performance of a plug-in estimator for the value of a linear functional ℓ evaluated at ϕ. For this purpose the slope ϕ is replaced in $\ell(\phi)$ by a suitable estimator $\widehat{\phi}_{m_n^*}$ depending on a tuning parameter $m_n^* \in \mathbb{N}$. However their choice of the tuning parameter is not data-driven. In the present paper we develop a data-driven selection procedure which features comparable minimax-optimal properties.

The non-parametric estimation of the slope function ϕ has been an issue of growing interest in the recent literature and a variety of such estimators have been studied. For example, Bosq [2000], Cardot et al. [2007] or Müller and Stadtmüller [2005] analyze a functional principal components regression, while a penalized least squares approach combined with projection onto some basis (such as splines) is examined in Ramsay and Dalzell [1991], Eilers and Marx [1996], Cardot et al. [2003], Hall and Horowitz [2007] or Crambes et al. [2009]. Cardot and Johannes [2010] investigate a linear Galerkin approach coming from the inverse problem community (c.f. Efromovich and Koltchinskii [2001] and Hoffmann and Reiß [2008]). The resulting thresholded projection estimator $\widehat{\phi}_{m_n^*}$ is used by Johannes and Schenk [2010] in their plug-in estimation procedure $\widehat{\ell}_{m_n^*} := \ell(\widehat{\phi}_{m_n^*})$ for the value $\ell(\phi)$ of a linear functional evaluated at ϕ.

It has been shown in Johannes and Schenk [2010] that the attainable rate of convergence of the plug-in estimator is basically determined by the *a priori* conditions on the solution ϕ and the covariance operator Γ associated with the regressor X (defined below). These conditions are expressed in the form $\phi \in \mathcal{F}$ and $\Gamma \in \mathcal{G}$, for suitably chosen classes $\mathcal{F} \subseteq \mathbb{H}$ and $\mathcal{G}$; we postpone their formal introduction along with their interpretation to Section 2. Moreover, the accuracy of any estimator $\widetilde{\ell}$ of the value $\ell(\phi)$ has been assessed by its maximal mean squared error with respect to these classes, that is

$$\mathcal{R}^\ell[\widetilde{\ell};\mathcal{F},\mathcal{G}] := \sup_{\phi\in\mathcal{F}}\sup_{\Gamma\in\mathcal{G}} \mathbb{E}|\widetilde{\ell}-\ell(\phi)|^2.$$

The main purpose of Johannes and Schenk [2010] has been to derive a lower bound

$$\mathcal{R}^\ell_*[n^{-1};\mathcal{F},\mathcal{G}] \leqslant \inf_{\widetilde{\ell}} \mathcal{R}^\ell[\widetilde{\ell};\mathcal{F},\mathcal{G}],$$

where the infimum is taken over all estimators $\widetilde{\ell}$, and to prove that the estimator $\widehat{\ell}_{m_n^*}$ satisfies

$$\mathcal{R}^\ell[\widehat{\ell}_{m_n^*};\mathcal{F},\mathcal{G}] \leqslant C\cdot\mathcal{R}^\ell_*[n^{-1};\mathcal{F},\mathcal{G}], \quad \text{with } 0<C<\infty,$$

for a variety of classes $\mathcal{F}$ and $\mathcal{G}$. In other words it has been shown that $\mathcal{R}^\ell_*[n^{-1};\mathcal{F},\mathcal{G}]$ is the minimax-optimal rate attained by the estimator $\widehat{\ell}_{m_n^*}$. The optimal performance of the estimator depends crucially on the choice m_n^* of the tuning parameter, which in turn, relies strongly on *a priori* knowledge of the sets $\mathcal{F}$ and $\mathcal{G}$. However, this information is widely inaccessible in practice.

The aim of the present paper consists in proposing a fully data-driven selection procedure for the tuning parameter. Our selection method combines model selection (c.f. Barron et al. [1999] and its detailed discussion in Massart [2007]) and Lepski's method (c.f. Lepski [1990] and its recent review in Mathé [2006]). It is inspired by the recent work of Goldenshluger and Lepski [2011] who consider data-driven bandwidth selection in kernel density estimation. We choose the appropriate tuning parameter $\widehat{m}$ as the minimizer of a stochastic penalized contrast function imitating Lepski's method among a random collection of admissible values. Furthermore, we show that the maximal risk of the resulting estimator $\widehat{\ell}_{\widehat{m}}$ satisfies

$$\mathcal{R}^\ell[\widehat{\ell}_{\widehat{m}};\mathcal{F},\mathcal{G}] \leqslant C\cdot\mathcal{R}^\ell_*[(1+\log n)n^{-1};\mathcal{F},\mathcal{G}] \quad \text{for } 0<C<\infty,$$

for a variety of classes $\mathcal{F}$ and $\mathcal{G}$. The upper bound in the last display features a logarithmic factor when compared to the minimax rate of convergence $\mathcal{R}^\ell_*[n^{-1};\mathcal{F},\mathcal{G}]$ which possibly results in a deterioration of the rate. Therefore, the completely data-driven estimator is optimal or nearly optimal in the minimax sense simultaneously over a variety of both solution sets $\mathcal{F}$ and classes of operators $\mathcal{G}$. We call such estimation procedures adaptive. The appearance of the logarithmic factor within the rate is a known fact in the context of local estimation (c.f. Laurent et al. [2008] who consider model selection given direct Gaussian observations). Brown and Low [1996] show that it is unavoidable in the context of non-parametric Gaussian regression and, hence it is widely considered as an acceptable price for adaptation. This factor is also present in the recent work of Goldenshluger and Pereverzev [2000] where Lepski's method is applied in the presence of indirect Gaussian observations. In contrast to this situation the operator is not known in advance in functional linear regression and hence a straightforward application of their results is not obvious. We will show that our proposed data-driven estimation method attains the minimax-rates up to a logarithmic factor for a variety of a classes of both slope functions and covariance operators.

The paper is organized as follows: in Section 2 we introduce the adaptive estimation procedure and review the available minimax theory as presented in Johannes and Schenk [2010]. In Section 3 we present the key arguments of the proof of an upper risk bound for the adaptive estimator, while more technical aspects of the proof are deferred to the Appendix. We discuss the examples of point-wise and local average estimation in Section 4.

2 Methodology and review

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 $\mathbb{E}[\varepsilon\langle X, h\rangle_{\mathbb{H}}] = 0$ for all $h \in \mathbb{H}$, as usually assumed in this context, see for example Bosq [2000], Cardot et al. [2003] or Cardot et al. [2007]. Multiplying both sides in (1.1) by $\langle X, h\rangle_{\mathbb{H}}$ and taking the expectation leads to the normal equation

$$\langle g, h\rangle_{\mathbb{H}} := \mathbb{E}[Y\langle X, h\rangle_{\mathbb{H}}] = \mathbb{E}[\langle \phi, X\rangle_{\mathbb{H}}\langle X, h\rangle_{\mathbb{H}}] =: \langle \Gamma\phi, h\rangle_{\mathbb{H}}, \quad \forall h \in \mathbb{H}, \tag{2.1}$$

where g belongs to $\mathbb{H}$ and Γ denotes the covariance operator associated with the random function X. In what follows we assume that there exists a unique solution $\phi \in \mathbb{H}$ of equation (2.1), i.e., that Γ is strictly positive and that its range contains g (for a detailed discussion we refer to Cardot et al. [2003]). Obviously, these conditions are sufficient for the identification of the value $\ell(\phi)$. Since the estimation of ϕ involves an 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. Therefore, the reconstruction of ϕ leads to an *ill-posed inverse problem* (with the additional difficulty that Γ is unknown and has to be estimated). In the following we assume that the joint distribution of the regressor and error term is Gaussian, more precisely, we suppose that for any finite set $\{h_1, \dots, h_{k-1}\} \subset \mathbb{H}$ the vector $(\langle X, h_1\rangle_{\mathbb{H}} \dots, \langle X, h_{k-1}\rangle_{\mathbb{H}}, \varepsilon)$ follows a k-dimensional multivariate normal distribution.

Remark 2.1. The assumption of Gaussianity is not essential for the proof of our main result. This assumption on the distributions of the error and the regressor is only used to prove the bounds given in Lemma C.2. Analogues of the results can be shown at the cost of longer proofs under appropriately chosen moment conditions. □

2.1 Adaptive Estimation Procedure

Introduction of the estimator. In order to derive an estimator for the unknown slope function ϕ we follow the presentation of Johannes and Schenk [2010] and base our reconstruction on the development of ϕ in an arbitrary orthonormal basis. Here and subsequently, we fix a pre-specified orthonormal basis $\{\psi_j\}_{j=1}^\infty$ of $\mathbb{H}$ which does in general not correspond to the eigenfunctions of the operator Γ defined in (2.1). We require in the following that the slope function ϕ 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 $\ell(\psi_1) = 1$ which can always be ensured by reordering and rescaling, except for the trivial case $\ell \equiv 0$. With respect to this basis, we consider for all $h \in \mathbb{H}$ the development $h = \sum_{j=1}^{\infty} [h]_j \psi_j$ where the sequence $[h] := ([h]_j)_{j \geqslant 1}$ of generalized Fourier coefficients $[h]_j := \langle h, \psi_j \rangle_{\mathbb{H}}$ is square-summable, i.e., $\|h\|_{\mathbb{H}}^2 = \sum_{j=1}^{\infty} [h]_j^2 < \infty$. Given a dimension parameter $m \in \mathbb{N}$ we have the subspace $\mathbb{H}_m$ - spanned by the basis functions $\{\psi_j\}_{j=1}^m$ - at our disposal and we call $\phi_m \in \mathbb{H}_m$ a Galerkin solution of $g = \Gamma\phi$, if $\|g - \Gamma\phi_m\|_{\mathbb{H}} \leqslant \|g - \Gamma h\|_{\mathbb{H}}$ for all $h \in \mathbb{H}_m$. Since Γ is strictly positive it is easily seen that the Galerkin solution ϕ_m of $g = \Gamma\phi$ exists uniquely. Let us introduce for any function h the m-dimensional vector of coefficients $[h]_{\underline{m}} := ([h]_j)_{1 \leqslant j \leqslant m}$ and for the operator Γ the $(m \times m)$-dimensional matrix $[\Gamma]_{\underline{m}} := (\langle \psi_j, \Gamma\psi_k \rangle_{\mathbb{H}})_{1 \leqslant j,k \leqslant m}$. Then the Galerkin solution ϕ_m satisfies $[\Gamma]_{\underline{m}} [\phi_m]_{\underline{m}} = [g]_{\underline{m}}$. Since Γ is injective, the matrix $[\Gamma]_{\underline{m}}$ is non-singular for all $m \geqslant 1$ and therefore the Galerkin solution $\phi_m \in \mathbb{H}_m$ is uniquely determined by the vector of coefficients $[\phi_m]_{\underline{m}} = [\Gamma]_{\underline{m}}^{-1} [g]_{\underline{m}}$ and $[\phi_m]_j = 0$ for $j > m$. In order to derive an estimator for the vector $[\phi_m]_{\underline{m}}$, we replace the unknown quantities $[g]_{\underline{m}}$ and $[\Gamma]_{\underline{m}}$ by their empirical counterparts and apply additional thresholding. We observe that $[\Gamma]_{\underline{m}} = \mathbb{E}[X]_{\underline{m}} [X]_{\underline{m}}^t$ and $[g]_{\underline{m}} = \mathbb{E} Y [X]_{\underline{m}}$, therefore, given an i.i.d. sample $\{(Y_i, X_i)\}_{i=1}^n$ of (Y, X), it is natural to consider the estimators $[\widehat{g}]_{\underline{m}} := \frac{1}{n} \sum_{i=1}^n Y_i [X_i]_{\underline{m}}$ and $[\widehat{\Gamma}]_{\underline{m}} := \frac{1}{n} \sum_{i=1}^n [X_i]_{\underline{m}} [X_i]_{\underline{m}}^t$. Let us denote by $\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s$ the spectral norm of $[\widehat{\Gamma}]_{\underline{m}}^{-1}$, i.e., its largest eigenvalue, and define the estimator $\widehat{\phi}_m \in \mathbb{H}_m$ by means of the coefficients $[\widehat{\phi}_m]_j = 0$ for $j > m$ and

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

Observe that $\ell(\phi_m) = (\ell(\psi_1), \ldots, \ell(\psi_m))[\phi_m]_{\underline{m}} =: [\ell]_{\underline{m}}^t [\phi_m]_{\underline{m}}$ with the slight abuse of notations $[\ell]_{\underline{m}} := ([\ell]_j)_{1 \leqslant j \leqslant m}$ and generic elements $[\ell]_j := \ell(\psi_j)$. In Johannes and Schenk [2010] it has been shown that the estimator $\widehat{\ell}_m := \ell(\widehat{\phi}_m)$ with optimally chosen dimension parameter m can attain minimax-optimal rates of convergence. This choice involves certain characteristics of the slope ϕ and the covariance operator Γ which are unavailable in practice. In the next paragraph we introduce a fully data-driven selection method for the dimension parameter.

Introduction of the adaptive estimation procedure. Our selection method is inspired by the recent work of Goldenshluger and Lepski [2011] and combines the techniques of model selection and Lepski's method. We determine the dimension parameter among a collection of admissible values by minimizing a penalized contrast function. To this end, we define for all $n \geqslant 1$ the value $M_n^\ell := \max\left\{1 \leqslant m \leqslant \lfloor n^{1/4} \rfloor : [\ell]_{\underline{m}}^t [\ell]_{\underline{m}} \leqslant n\right\}$ where $\lfloor a \rfloor$ denotes as usual the integer part of $a \in \mathbb{R}$ and introduce the random integer

$$\widehat{M}_n := \min\left\{2 \leqslant m \leqslant M_n^\ell : \|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s ([\ell]_{\underline{m}}^t [\ell]_{\underline{m}}) > n(1 + \log n)^{-1}\right\} - 1. \tag{2.2}$$

Furthermore, we define a stochastic penalty sequence $\widehat{\mathrm{p}} := (\widehat{\mathrm{p}}_m)_{1 \leqslant m \leqslant \widehat{M}_n}$ by

$$\widehat{\mathrm{p}}_m := 700 \left(\frac{2}{n} \sum_{i=1}^n Y_i^2 + 2[\widehat{g}]_{\underline{m}}^t [\widehat{\Gamma}]_{\underline{m}}^{-1} [\widehat{g}]_{\underline{m}} \right) \cdot \max_{1 \leqslant k \leqslant m} [\ell]_{\underline{k}}^t [\widehat{\Gamma}]_{\underline{k}}^{-1} [\ell]_{\underline{k}} \cdot \frac{(1 + \log n)}{n}.$$

The random integer $\widehat{M}_n$ and the stochastic penalty $\widehat{\mathrm{p}}_m$ are used to define a contrast by

$$\kappa_m := \max_{m \leqslant k \leqslant \widehat{M}_n} \Big\{ |\widehat{\ell}_k - \widehat{\ell}_m|^2 - \widehat{\mathrm{p}}_k \Big\}.$$

For a subset $A \subset \mathbb{N}$ and a sequence $(a_m)_{m\geqslant 1}$ with minimal value in A we set $\arg\min_{m\in A}\{a_m\} := \min\{m : a_m \leqslant a_{m'}, \forall m' \in A\}$ and select the dimension parameter

$$\widehat{m} := \mathop{\arg\min}_{1 \leqslant m \leqslant \widehat{M}_n} \{\kappa_m + \widehat{\mathrm{p}}_m\}. \tag{2.3}$$

The estimator of $\ell(\phi)$ is now given by $\widehat{\ell}_{\widehat{m}}$ and we will derive an upper bound for its risk below. By construction the choice of the dimension parameter and hence the estimator $\widehat{\ell}_{\widehat{m}}$ rely only on the data and in particular not on the regularity assumptions on the slope and the operator which we formalize in the next section.

2.2 Review of minimax theory

We express our *a priori* knowledge about the unknown slope parameter and covariance operator in the form $\phi \in \mathcal{F}$ and $\Gamma \in \mathcal{G}$. The class $\mathcal{F}$ reflects information on the solution ϕ, e.g., its level of smoothness, whereas the assumption $\Gamma \in \mathcal{G}$ typically results in conditions on the decay of the eigenvalues of the operator Γ. The following construction of the classes $\mathcal{F}$ and $\mathcal{G}$ will be flexible enough to characterize, in particular, differentiable or analytic slope functions and allows us to discuss both a polynomial and exponential decay of the covariance operator's eigenvalues.

Assumptions and notations. With respect to the basis $\{\psi_j\}_{j=1}^\infty$ and given a strictly positive sequence of weights $(w_j)_{j\geqslant 1}$, or w for short, we define the weighted norm $\|\cdot\|_w$ by $\|h\|_w^2 := \sum_{j=1}^\infty w_j [h]_j^2$ for $h \in \mathbb{H}$. Throughout the rest of the paper let β be a non-decreasing sequence of weights with $\beta_1 = 1$ such that slope parameter ϕ belongs to the ellipsoid

$$\mathcal{F}_\beta^r := \big\{ h \in \mathbb{H} : \|h\|_\beta^2 \leqslant r \big\} \quad \text{with radius } r > 0.$$

In order to guarantee that $\mathcal{F}_\beta^r$ is contained in the domain of the linear functional ℓ and that $\ell(h) = \sum_{j\geqslant 1} [\ell]_j [h]_j$ for all $h \in \mathcal{F}_\beta^r$ with $[\ell]_j = \ell(\psi_j)$, $j \geqslant 1$, it is sufficient that $\sum_{j\geqslant 1} [\ell]_j^2 \beta_j^{-1} < \infty$. 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 $\mathbb{H}$ 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.

As usual in the context of ill-posed inverse problems, we link the mapping properties of the covariance operator Γ and the regularity conditions on ϕ. To this end, we consider the sequence $(\langle \Gamma\psi_j, \psi_j\rangle)_{j\geqslant 1} =: ([\Gamma]_{jj})_{j\geqslant 1}$. Since Γ is nuclear, this sequence is summable and hence vanishes as j tends to infinity. In what follows we impose restrictions on the decay of this sequence. Let $\mathcal{G}$ denote the set of all strictly positive nuclear operators defined on $\mathbb{H}$. We suppose that there exists a strictly positive, summable sequence of weights γ with $\gamma_1 = 1$ such that Γ belongs to the subset

$$\mathcal{G}_\gamma^d := \Big\{ T \in \mathcal{G} : \quad d^{-2} \|h\|_{\gamma^2}^2 \leqslant \|Th\|_{\mathbb{H}}^2 \leqslant d^2 \,\|h\|_{\gamma^2}^2, \quad \forall h \in \mathbb{H} \Big\} \quad \text{with } d \geqslant 1$$

where we understand here and subsequently arithmetic operations on a sequence of real numbers component-wise, e.g., we write γ^2 for $(\gamma_j^2)_{j\geqslant 1}$. Notice that for $\Gamma \in \mathcal{G}_\gamma^d$ it follows that $d^{-1}\gamma_j \leqslant [\Gamma]_{jj} \leqslant d\gamma_j$. Moreover, if λ denotes its sequence of eigenvalues, then $d^{-1}\gamma_j \leqslant \lambda_j \leqslant d\gamma_j$ which justifies the condition $\sum_{j=1}^{\infty} \gamma_j < \infty$. Let us summarize the previous conditions:

Assumption 2.1. *The sequences $1/\beta$ and γ are monotonically decreasing with limit zero and $\beta_1 = \gamma_1 = 1$ such that $\sum_{j\geqslant 1} [\ell]_j^2 \beta_j^{-1} < \infty$ and $\sum_{j\geqslant 1} \gamma_j < \infty$.*

Illustration. We illustrate the last assumption for typical choices of the sequences β, γ and $[\ell]$. Consider $[\ell]_j^2 = |j|^{-2s}$ and:

(pp) $\beta_j = |j|^{2p}$, $\gamma_j = |j|^{-2a}$ with $p > 0$, $a > 1/2$ and $s > 1/2 - p$;

(pe) $\beta_j = |j|^{2p}$, $\gamma_j = \exp(-|j|^{2a} + 1)$ with $p > 0$, $a > 0$ and $s > 1/2 - p$;

(ep) $\beta_j = \exp(|j|^{2p} - 1)$, $\gamma_j = |j|^{-2a}$ with $p > 0$, $a > 1/2$ and $s \in \mathbb{R}$;

then Assumption 2.1 holds true in all cases.

Minimax theory reviewed. Johannes and Schenk [2010] have derived a lower bound for the minimax risk $\inf_{\widetilde{\ell}} \mathcal{R}^\ell[\widetilde{\ell}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$ and have shown that the proposed estimator $\widehat{\ell}_m$ can attain this lower bound up to constant provided that the dimension parameter is chosen appropriately. In order to formulate the minimax rate below let us define for $m \geqslant 1$ and $x \in (0,1]$

$$\mathcal{R}_m^\ell[x; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] := \max\left\{ \sum_{j>m} \frac{[\ell]_j^2}{\beta_j}, \max\Big(\frac{\gamma_m}{\beta_m}, x\Big) \sum_{j=1}^{m} \frac{[\ell]_j^2}{\gamma_j} \right\}$$
$$\text{and } \; \mathcal{R}_*^\ell[x; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] := \min_{m\geqslant 1} \mathcal{R}_m^\ell[x; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d].$$

With this notation the lower bound, when considering an i.i.d. sample of size n, is basically a multiple of $\mathcal{R}_*^\ell[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$. To be more precise, if we define $m_n^* := \arg\min_{m\geqslant 1} \mathcal{R}_m^\ell[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$ and if Assumption 2.1 and $\inf_{n\geqslant 1} \min\big(\frac{\beta_{m_n^*}}{n\gamma_{m_n^*}}, \frac{n\gamma_{m_n^*}}{\beta_{m_n^*}}\big) > 0$ are satisfied then there exists a constant $C > 0$ depending only on the classes and σ^2 such that we have for all $n \geqslant 1$

$$\inf_{\widetilde{\ell}} \mathcal{R}^\ell[\widetilde{\ell}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \geqslant C \cdot \mathcal{R}_*^\ell[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d].$$

On the other hand it is shown in Johannes and Schenk [2010] that $\mathcal{R}_*^\ell[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$ provides up to a constant an upper bound for the maximal risk of the proposed estimator $\widehat{\ell}_{m_n^*}$. More precisely, if we assume in addition $\sup_{m\geqslant 1} m^3 \gamma_m \beta_m^{-1} < \infty$ then there exists a constant $C > 0$ depending only on the classes and σ^2 such that we have for all $n \geqslant 1$

$$\mathcal{R}^\ell[\widehat{\ell}_{m_n^*}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \leqslant C \cdot \mathcal{R}_*^\ell[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d].$$

Consequently the rate $\mathcal{R}_*^\ell[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$ is optimal and $\widehat{\ell}_{m^*}$ is minimax-optimal.

Illustration continued. For the configurations defined below Assumption 2.1 the estimator $\widehat{\ell}_{m_n^*}$ with dimension parameter m_n^* as given below is minimax optimal under the following conditions. The minimax optimal rate of convergence is determined by the orders of $\mathcal{R}^\ell_*[n^{-1};\mathcal{F}^r_\beta,\mathcal{G}^d_\gamma]$. Here and subsequently, we use for two strictly positive sequences $(x_n)_{n\geqslant1}, (y_n)_{n\geqslant1}$ the notation $x_n \asymp y_n$, if $(x_n/y_n)_{n\geqslant1}$ is bounded away both from zero and infinity.

(pp) If $p>0$, $a>1/2$ and $p+a\geqslant 3/2$ then $m_n^*\asymp n^{1/(2p+2a)}$ and if $s>1/2-p$, then

$$\mathcal{R}^\ell[\widehat{\ell}_{m_n^*};\mathcal{F}^r_\beta,\mathcal{G}^d_\gamma]\asymp\begin{cases} n^{-(2p+2s-1)/(2p+2a)}, & \text{if } s-a<1/2\\ n^{-1}\log n, & \text{if } s-a=1/2\\ n^{-1}, & \text{if } s-a>1/2.\end{cases}$$

(pe) If $p>0$ and $a>0$, then $m_n^*\asymp\log(n(\log n)^{-p/a})^{1/(2a)}$ and if $s>1/2-p$, then $\mathcal{R}^\ell[\widehat{\ell}_{m_n^*};\mathcal{F}^r_\beta,\mathcal{G}^d_\gamma]\asymp(\log n)^{-(2p+2s-1)/(2a)}$.

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

$$\mathcal{R}^\ell[\widehat{\ell}_{m_n^*};\mathcal{F}^r_\beta,\mathcal{G}^d_\gamma]\asymp\begin{cases} n^{-1}(\log n)^{(2a-2s+1)/(2p)}, & \text{if } s-a<1/2\\ n^{-1}\log(\log n), & \text{if } s-a=1/2\\ n^{-1}, & \text{if } s-a>1/2.\end{cases}$$

3 Upper risk bound for the adaptive estimator

The fully adaptive estimator $\widehat{\ell}_{\widehat{m}}$ of $\ell(\phi)$ relies on the choice of a random dimension parameter $\widehat{m}$ which does not involve any knowledge about the classes $\mathcal{F}^r_\beta$ and $\mathcal{G}^d_\gamma$. The main result of this paper consists in an upper bound for the maximal risk $\mathcal{R}^\ell[\widehat{\ell}_{\widehat{m}};\mathcal{F}^r_\beta,\mathcal{G}^d_\gamma]$ given by the following theorem. We present the main arguments of its proof in this section whereas the more technical aspects are deferred to the appendix. We close this section by illustrating and discussing the result.

THEOREM 3.1. *Assume an i.i.d. sample of (Y,X) of size n obeying* (1.1) *and let the joint distribution of the random function X and the error ε be normal. Consider sequences β and γ satisfying Assumption 2.1. Define $m_n^\diamond := \arg\min_{m\geqslant1}\mathcal{R}^\ell_m[(1+\log n)n^{-1};\mathcal{F}^r_\beta,\mathcal{G}^d_\gamma]$ and suppose that $\gamma^{-1}_{m_n^\diamond}[\ell]^t_{\underline{m_n^\diamond}}[\ell]_{\underline{m_n^\diamond}}=o(n(1+\log n)^{-1})$ as $n\to\infty$ then there exists a constant $C>0$ depending on the classes $\mathcal{F}^r_\beta$ and $\mathcal{G}^d_\gamma$, the linear functional ℓ, and σ^2 only such that*

$$\mathcal{R}^\ell[\widehat{\ell}_{\widehat{m}};\mathcal{F}^r_\beta,\mathcal{G}^d_\gamma]\leqslant C\cdot\mathcal{R}^\ell_*[(1+\log n)n^{-1};\mathcal{F}^r_\beta,\mathcal{G}^d_\gamma],\qquad \textit{for all } n\geqslant1.$$

REMARK 3.1. The last assertion states that the data-driven estimator can attain the minimax-rates up to a logarithmic factor for a variety of classes $\mathcal{F}^r_\beta$ and $\mathcal{G}^d_\gamma$. In this sense the estimator adapts to both the slope function and the covariance operator. This result is derived under the additional condition, $\gamma^{-1}_{m_n^\diamond}[\ell]^t_{\underline{m_n^\diamond}}[\ell]_{\underline{m_n^\diamond}}=o(n(1+\log n)^{-1})$ as $n\to\infty$, which naturally holds true in the illustrations. □

We begin our reasoning by giving a preparatory lemma which constitutes a central step in the following arguments.

LEMMA 3.2. *Let* $(\phi_k)_{k\geqslant 1}$ *be an arbitrary sequence in* $\mathbb{H}$ *and* $\mathrm{b} := (\mathrm{b}_m)_{m\geqslant 1}$ *the sequence of approximation errors* $\mathrm{b}_m = \sup_{m\leqslant k}|\ell(\phi_k - \phi)|$ *associated with* $\ell(\phi)$. *Consider an arbitrary sequence of penalties* $\mathrm{p} := (\mathrm{p}_m)_{m\geqslant 1}$, *an upper bound* $M \in \mathbb{N}$, *and the sequence* $\kappa = (\kappa_m)_{m\geqslant 1}$ *of contrasts given by* $\kappa_m := \max_{m\leqslant k\leqslant M}\Big\{|\widehat{\ell}_k - \widehat{\ell}_m|^2 - \mathrm{p}_k\Big\}$. *If the subsequence* $(\mathrm{p}_1,\dots,\mathrm{p}_M)$ *is non-decreasing, then we have for the selected model* $\widetilde{m} := \arg\min_{1\leqslant m\leqslant M}\{\kappa_m + \mathrm{p}_m\}$ *and for all* $1\leqslant m\leqslant M$ *that*

$$|\widehat{\ell}_{\widetilde{m}} - \ell(\phi)|^2 \leqslant 7\,\mathrm{p}_m + 78\,\mathrm{b}_m^2 + 42\max_{m\leqslant k\leqslant M}\left(|\widehat{\ell}_k - \ell(\phi_k)|^2 - \frac{1}{6}\,\mathrm{p}_k\right)_+ \tag{3.1}$$

where $(a)_+ = \max(a,0)$.

PROOF OF LEMMA 3.2. Since $(\mathrm{p}_1,\dots,\mathrm{p}_M)$ is non-decreasing it is easily verified that

$$\kappa_m \leqslant 6\max_{m\leqslant k\leqslant M}\left(|\widehat{\ell}_k - \ell(\phi_k)|^2 - \frac{1}{6}\,\mathrm{p}_k\right)_+ + 12\,\mathrm{b}_m^2, \quad \forall\, 1\leqslant m\leqslant M,$$

where we use that $2\,\mathrm{b}_m \geqslant \max_{m\leqslant k\leqslant M}|\ell(\phi_k - \phi_m)|$. The last estimate implies the inequality

$$|\widehat{\ell}_m - \ell(\phi)|^2 \leqslant \frac{1}{3}\,\mathrm{p}_m + 2\,\mathrm{b}_m^2 + 2\max_{m\leqslant k\leqslant M}\left(|\widehat{\ell}_k - \ell(\phi_k)|^2 - \frac{1}{6}\,\mathrm{p}_k\right)_+, \ \forall\, 1\leqslant m\leqslant M. \tag{3.2}$$

On the other hand, taking the definition of $\widetilde{m}$ into account, it is straightforward to see that

$$\begin{aligned}|\widehat{\ell}_{\widetilde{m}} - \ell(\phi)|^2 &\leqslant 3\Big\{|\widehat{\ell}_{\widetilde{m}} - \widehat{\ell}_{\min(m,\widetilde{m})}|^2 + |\widehat{\ell}_{\min(m,\widetilde{m})} - \widehat{\ell}_m|^2 + |\widehat{\ell}_m - \ell(\phi)|^2\Big\}\\ &\leqslant 3\Big\{\kappa_m + \mathrm{p}_{\widetilde{m}} + \kappa_{\widetilde{m}} + \mathrm{p}_m + |\widehat{\ell}_m - \ell(\phi)|^2\Big\} \leqslant 6\{\kappa_m + \mathrm{p}_m\} + 3|\widehat{\ell}_m - \ell(\phi)|^2.\end{aligned}$$

From the last estimates and (3.2) we obtain the assertion (3.1), which completes the proof. □

The proof of Theorem 3.1 requires in addition to the previous lemma two technical propositions which we state now. For $n\geqslant 1$ and a positive sequence $a := (a_m)_{m\geqslant 1}$ let us introduce $M_n^\ell := \max\{1\leqslant m\leqslant \lfloor n^{1/4}\rfloor : [\ell]_{\underline{m}}^t[\ell]_{\underline{m}} \leqslant n\}$ and

$$M_n(a) := \min\Big\{2\leqslant m\leqslant M_n^\ell : a_m\cdot[\ell]_{\underline{m}}^t[\ell]_{\underline{m}} > n(1+\log n)^{-1}\Big\} - 1$$

where we set $M_n(a) := M_n^\ell$ if the set is empty. Observe that $\widehat{M_n}$ given in (2.2) satisfies $\widehat{M_n} = M_n(a)$ with $a = (\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s)_{m\geqslant 1}$. Consider for $m\geqslant 1$

$$\sigma_m^2 := 2\mathbb{E}Y^2 + 2[g]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}, \qquad V_m := \max_{1\leqslant k\leqslant m}[\ell]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[\ell]_{\underline{k}}$$

and define the penalty term

$$\mathrm{p}_m := 100\,\sigma_m^2\,V_m\,(1+\log n)n^{-1},$$

which are obviously the theoretical counterparts of the random objects used in the definition of $\widehat{m}$. The proof of the next assertion is deferred to the appendix.

Proposition 3.3. *Let the conditions of Theorem 3.1 hold true and denote by $\phi_m \in \mathbb{H}_m$ the Galerkin solution of $g = \Gamma\phi$. Define $M_n^+ := M_n(a)$ with $a = ([4d\gamma_j]^{-1})_{j\geqslant 1}$ then there is a constant $C(d) > 0$ depending on d only such that for all $n \geqslant 1$*

$$\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}\Big\{\max_{1\leqslant m\leqslant M_n^+}\Big(|\widehat{\ell}_m - \ell(\phi_m)|^2 - \frac{\mathrm{p}_m}{6}\Big)_+\Big\} \leqslant \frac{C(d)}{n}(\sigma^2 + r)\max\Big\{(\sum_{j\geqslant 1}\gamma_j)^2, \sum_{j\geqslant 1}\frac{[\ell]_j^2}{\beta_j}\Big\}.$$

Additionally, let us introduce for $n \geqslant 1$ the random integer $M_n^- := M_n\big(a\big)$ with the sequence $a = (16d^3\gamma_j^{-1})_{j\geqslant 1}$. In the following we decompose the risk with respect to an event $\mathcal{E}_n$, and respectively its complement $\mathcal{E}_n^c$, on which $\widehat{\mathrm{p}}$ and $\widehat{M}_n$ are comparable to their theoretical counterparts. To be more precise, we define the event

$$\mathcal{E}_n := \big\{\forall\, 1 \leqslant m \leqslant M_n^+ : \mathrm{p}_m \leqslant \widehat{\mathrm{p}}_m \leqslant 24\,\mathrm{p}_m\big\} \cap \Big\{M_n^- \leqslant \widehat{M}_n \leqslant M_n^+\Big\}$$

and consider the elementary identity

$$\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}|\widehat{\ell}_{\widehat{m}} - \ell(\phi)|^2 = \sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}\big(|\widehat{\ell}_{\widehat{m}} - \ell(\phi)|^2\,\mathbb{1}_{\mathcal{E}_n}\big) + \sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}\big(|\widehat{\ell}_{\widehat{m}} - \ell(\phi)|^2\,\mathbb{1}_{\mathcal{E}_n^c}\big). \quad (3.3)$$

The next proposition states that the second right hand side term is bounded up to a constant by n^{-1} and is hence negligible. The proof is deferred to the appendix.

Proposition 3.4. *Let the conditions of Theorem 3.1 hold true. If we consider the fully data-driven choice $\widehat{m}$ given in* (2.3) *then there exists a constant $C(d) > 0$ depending on d only such that for all $n \geqslant 1$*

$$\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}\big(|\widehat{\ell}_{\widehat{m}} - \ell(\phi)|^2\,\mathbb{1}_{\mathcal{E}_n^c}\big) \leqslant \frac{C(d)}{n}\,(\sigma^2 + r)\max\Big\{\sum_{j\geqslant 1}\gamma_j, \sum_{j\geqslant 1}\frac{[\ell]_j^2}{\beta_j}\Big\}.$$

We are now in position to prove Theorem 3.1.

Proof of Theorem 3.1. In the following we will denote by $C(d) > 0$ a constant depending on d only, which may change from line to line. From the elementary identity (3.3) and Proposition 3.4 we derive for all $n \geqslant 1$

$$\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}|\widehat{\ell}_{\widehat{m}} - \ell(\phi)|^2 \leqslant \sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}\big(|\widehat{\ell}_{\widehat{m}} - \ell(\phi)|^2\,\mathbb{1}_{\mathcal{E}_n}\big) + \frac{C(d)}{n}(\sigma^2 + r)\max\Big\{\sum_{j\geqslant 1}\gamma_j, \sum_{j\geqslant 1}\frac{[\ell]_j^2}{\beta_j}\Big\}. \quad (3.4)$$

We observe that the random subsequence $(\widehat{\sigma}_1^2, \dots, \widehat{\sigma}_{\widehat{M}_n}^2)$, and hence $(\widehat{\mathrm{p}}_1, \dots, \widehat{\mathrm{p}}_{\widehat{M}_n})$, are by construction non-decreasing. Furthermore, we observe that for all $1 \leqslant m \leqslant k \leqslant \widehat{M}_n$ the identity $\langle \widehat{\Gamma}(\widehat{\phi}_k - \widehat{\phi}_m), (\widehat{\phi}_k - \widehat{\phi}_m)\rangle_{\mathbb{H}} = [\widehat{g}]_{\underline{k}}^t [\widehat{\Gamma}]_{\underline{k}}^{-1} [\widehat{g}]_{\underline{k}} - [\widehat{g}]_{\underline{m}}^t [\widehat{\Gamma}]_{\underline{m}}^{-1} [\widehat{g}]_{\underline{m}}$ holds true. Therefore, it follows by using that $\widehat{\Gamma}$ is positive definite that $[\widehat{g}]_{\underline{m}}^t [\widehat{\Gamma}]_{\underline{m}}^{-1} [\widehat{g}]_{\underline{m}} \leqslant [\widehat{g}]_{\underline{k}}^t [\widehat{\Gamma}]_{\underline{k}}^{-1} [\widehat{g}]_{\underline{k}}$, and hence $\widehat{\sigma}_m^2 \leqslant \widehat{\sigma}_k^2$. Consequently, Lemma 3.2 is applicable for all $1 \leqslant m \leqslant \widehat{M}_n$ and we obtain

$$|\widehat{\ell}_{\widehat{m}} - \ell(\phi)|^2 \leqslant 7\,\widehat{\mathrm{p}}_m + 78\,\mathrm{b}_m^2 + 42 \max_{m \leqslant k \leqslant \widehat{M}_n} \left(|\widehat{\ell}_k - \ell(\phi_k)|^2 - \frac{1}{6}\,\widehat{\mathrm{p}}_k \right)_+ .$$

On the event $\mathcal{E}_n$ we deduce from the last bound that for all $1 \leqslant m \leqslant M_n^-$

$$|\widehat{\ell}_{\widehat{m}} - \ell(\phi)|^2\, \mathbb{1}_{\mathcal{E}_n} \leqslant 504\,\mathrm{p}_m + 78\,\mathrm{b}_m^2 + 42 \max_{1 \leqslant m \leqslant M_n^+} \left(|\widehat{\ell}_m - \ell(\phi_m)|^2 - \frac{1}{6}\,\mathrm{p}_m \right)_+ .$$

Taking Lemma B.2 (v) in the appendix into account it follows for all $n \geqslant 1$

$$\sup_{\phi \in \mathcal{F}_\beta^r} \sup_{\Gamma \in \mathcal{G}_\gamma^d} \mathbb{E}\big(|\widehat{\ell}_{\widehat{m}} - \ell(\phi)|^2\, \mathbb{1}_{\mathcal{E}_n} \big) \leqslant C(d)(\sigma^2 + r) \min_{1 \leqslant m \leqslant M_n^-} \mathcal{R}_m^\ell\big[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d\big]$$
$$+ \sup_{\phi \in \mathcal{F}_\beta^r} \sup_{\Gamma \in \mathcal{G}_\gamma^d} \mathbb{E}\left\{ \max_{1 \leqslant m \leqslant M_n^+} \left(|\widehat{\ell}_m - \ell(\phi_m)|^2 - \frac{1}{6}\,\mathrm{p}_m \right)_+ \right\}.$$

Moreover, Proposition 3.3 and (3.4) imply for all $n \geqslant 1$ that

$$\sup_{\phi \in \mathcal{F}_\beta^r} \sup_{\Gamma \in \mathcal{G}_\gamma^d} \mathbb{E}|\widehat{\ell}_{\widehat{m}} - \ell(\phi)|^2 \leqslant C(d)(\sigma^2 + r) \max\left\{ \sum_{j \geqslant 1} \gamma_j, \sum_{j \geqslant 1} \frac{[\ell]_j^2}{\beta_j} \right\}$$
$$\cdot \min_{1 \leqslant m \leqslant M_n^-} \mathcal{R}_m^\ell\big[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d\big] \quad (3.5)$$

where we use that $\mathcal{R}_m^\ell\big[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d\big] \geqslant n^{-1}$ for all $m \geqslant 1$. Under the additional condition $\gamma_{m_n^\diamond}^{-1} [\ell]_{\underline{m_n^\diamond}}^t [\ell]_{\underline{m_n^\diamond}} = o(n(1 + \log n)^{-1})$ it is easily verified that there exists an integer n_o only depending on the sequences β, γ and $[\ell]$ such that for all $n \geqslant n_o$ we have $m_n^\diamond \leqslant M_n^-$ and

$$\min_{1 \leqslant m \leqslant M_n^-} \mathcal{R}_m^\ell[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] = \mathcal{R}_*^\ell[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d].$$

However, in case $n < n_o$ we employ that

$$\mathcal{R}_1^\ell[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \leqslant \max(1, (1 + \log n)n^{-1}) \sum_{j > 1} \frac{[\ell]_j^2}{\beta_j} \leqslant \sum_{j \geqslant 1} \frac{[\ell]_j^2}{\beta_j}$$

and consequently we derive the bound

$$\min_{1 \leqslant m \leqslant M_n^-} \mathcal{R}_m^\ell[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \leqslant n^{-1} n_o \sum_{j \geqslant 1} \frac{[\ell]_j^2}{\beta_j}, \qquad \text{for all } n < n_o.$$

The combination of both cases yields for all $n \geqslant 1$

$$\min_{1 \leqslant m \leqslant M_n^-} \mathcal{R}_m^\ell[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \leqslant n_o \sum_{j \geqslant 1} \frac{[\ell]_j^2}{\beta_j} \mathcal{R}_*^\ell[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d].$$

As n_o depends only on the sequences β, γ and $[\ell]$, we derive the result of the theorem from the previous display together with (3.5), which completes the proof. □

Remark 3.2. Recall that the estimator $\widehat{\ell}_{m_n^*}$ with optimally chosen dimension parameter m_n^* is minimax-optimal, i.e, its maximal risk $\mathcal{R}^\ell[\widehat{\ell}_{m_n^*};\mathcal{F}_\beta^r,\mathcal{G}_\gamma^d]$ can be bounded up to a constant by the lower bound $\mathcal{R}_*^\ell[n^{-1};\mathcal{F}_\beta^r,\mathcal{G}_\gamma^d]$. However, due to Theorem 3.1 the maximal risk of the fully adaptive estimator is bounded by a multiple of $\mathcal{R}_*^\ell[(1+\log n)n^{-1};\mathcal{F}_\beta^r,\mathcal{G}_\gamma^d]$. The appearance of the logarithmic factor within the rate is a known fact in the context of local estimation. It is widely considered as an acceptable price for adaptation (in the context of non-parametric Gaussian regression it is unavoidable as shown in Brown and Low [1996]). □

Illustration continued. In the configurations defined below Assumption 2.1 the additional condition $\gamma_{m_n^\diamond}^{-1}[\ell]_{\underline{m_n^\diamond}}^t[\ell]_{\underline{m_n^\diamond}} = o(n(1+\log n)^{-1})$ as $n\to\infty$ is easily verified. Therefore, the maximal risk of the fully adaptive estimator is bounded by a multiple of $\mathcal{R}_*^\ell[(1+\log n)n^{-1};\mathcal{F}_\beta^r,\mathcal{G}_\gamma^d]$ due to Theorem 3.1. In the next assertion we state its order in the considered cases and we omit the straightforward calculations.

Proposition 3.5. *Assume an i.i.d. sample of (Y,X) of size n obeying (1.1) and let the joint distribution of the random function X and the error ε be normal. The obtainable rate of convergence is determined by the orders of $\mathcal{R}_*^\ell[(1+\log n)n^{-1};\mathcal{F}_\beta^r,\mathcal{G}_\gamma^d]$ as given below.*

(pp) If $p>0$, $a>1/2$, $p+a\geqslant 3/2$ and $s>1/2-p$, then

$$\mathcal{R}_*^\ell[(1+\log n)n^{-1};\mathcal{F}_\beta^r,\mathcal{G}_\gamma^d] \asymp \begin{cases} \big(n^{-1}\log n\big)^{(2p+2s-1)/(2p+2a)}, & \text{if } s-a<1/2\\ n^{-1}(\log n)^2, & \text{if } s-a=1/2\\ n^{-1}\log n, & \text{if } s-a>1/2. \end{cases}$$

(pe) If $p>0$, $a>0$, and if $s>1/2-p$, then

$$\mathcal{R}_*^\ell[(1+\log n)n^{-1};\mathcal{F}_\beta^r,\mathcal{G}_\gamma^d] \asymp (\log n)^{-(2p+2s-1)/(2a)}.$$

(ep) If $p>0$, $a>1/2$ and $s\in\mathbb{R}$ then

$$\mathcal{R}_*^\ell[(1+\log n)n^{-1};\mathcal{F}_\beta^r,\mathcal{G}_\gamma^d] \asymp \begin{cases} n^{-1}(\log n)^{(2p+2a-2s+1)/(2p)}, & \text{if } s-a<1/2\\ n^{-1}(\log n)(\log\log n), & \text{if } s-a=1/2\\ n^{-1}\log n, & \text{if } s-a>1/2. \end{cases}$$

We shall briefly compare these rates with the corresponding minimax optimal rates derived in Section 2.2 above. Surprisingly they coincide in case *(pe)*, and hence the fully data-driven estimator is minimax-optimal. The rates given in case *(pp)* coincides with the ones that have been obtained by Goldenshluger and Pereverzev [2000] for an *a priori* known operator. In comparison to the minimax optimal rates the cases *(pp)* and *(ep)* feature a deterioration of logarithmic order as expected (compare Remark 3.2).

4 Examples: point-wise and local average estimation

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

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

Recall the typical choices of the sequences β and γ as introduced in the illustrations above. If $\beta_j \asymp |j|^{2p}$ for a positive integer p, see cases *(pp)* and *(pe)*, then the subset $\mathcal{F}_\beta := \{h \in \mathbb{H} : \|h\|_\beta^2 < \infty\}$ coincides with the Sobolev space of p-times differential periodic functions (c.f. Neubauer [1988a,b]). In the case *(ep)* it is well-known that for $p > 1$ every element of $\mathcal{F}_\beta$ is an analytic function (c.f. Kawata [1972]). Furthermore we consider a polynomial decay of γ with $a > 1/2$ in the cases *(pp)* and *(ep)*. Easy calculus shows that the covariance operator $\Gamma \in \mathcal{G}_\gamma^d$ acts for integer a like integrating $(2a)$-times and is hence called *finitely smoothing* (c.f. Natterer [1984]). In the case *(pe)* we assume an exponential decay of γ and it is easily seen that the range of $\Gamma \in \mathcal{G}_\gamma^d$ is a subset of $C^\infty[0,1]$, therefore the operator is called *infinitely smoothing* (c.f. Mair [1994]).

Point-wise estimation. By *evaluation in a given point* $t_0 \in [0,1]$ we mean the linear functional ℓ_{t_0} mapping h to $h(t_0) := \ell_{t_0}(h) = \sum_{j=1}^\infty [h]_j \psi_j(t_0)$. In the following we shall assume that the point evaluation is well-defined on the set of slope parameters $\mathcal{F}_\beta$ which is obviously implied by $\sum_{j=1}^\infty [\ell_{t_0}]_j^2 \beta_j^{-1} < \infty$. Consequently, the condition $\sum_{j\geqslant 1} \beta_j^{-1} < \infty$ is sufficient to guarantee that the point evaluation is well-defined on $\mathcal{F}_\beta$. Obviously, in case *(ep)* or in other words for exponentially increasing β, this additional condition is automatically satisfied. However, a polynomial increase, as in the cases *(pp)* and *(pe)*, 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 $\phi(t_0)$ 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 $\widehat{\ell}_{t_0}^m = \ell_{t_0}(\widehat{\phi}_m) = \widehat{\phi}_m(t_0)$.

Minimax optimal point-wise estimation. The estimator's maximal mean squared error over the classes $\mathcal{F}_\beta^r$ and $\mathcal{G}_\gamma^d$ is uniformly bounded for all $t_0 \in [0,1]$ up to a constant by $\mathcal{R}_*^{\ell_{t_0}}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$, i.e., $\sup_{\phi \in \mathcal{F}_\beta^r} \sup_{\Gamma \in \mathcal{G}_\gamma^d} \mathbb{E}|\widehat{\phi}_{m_n^*}(t_0) - \phi(t_0)|^2 \leqslant C\, \mathcal{R}_*^{\ell_{t_0}}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$ for some $C > 0$, which is the minimax-optimal rate of convergence (c.f. Johannes and Schenk [2010]).

Illustration continued. We derive with $[\ell_{t_0}]_j^2 \asymp j^{-2s}$ and $s = 0$ in the considered cases :

(pp) If $p > 1/2$, $a > 1/2$ and $p + a \geqslant 3/2$, then $\mathcal{R}_*^{\ell_{t_0}}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp n^{-(2p-1)/(2p+2a)}$.

(pe) If $p > 1/2$ and $a > 0$, then $\mathcal{R}_*^{\ell_{t_0}}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp (\log n)^{-(2p-1)/2a}$.

(ep) If $p > 0$ and $a > 1/2$, then $\mathcal{R}_*^{\ell_{t_0}}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp n^{-1}(\log n)^{(2a+1)/2p}$.

Adaptive point-wise estimation. We select the dimension parameter $\widehat{m}$ by minimizing the penalized contrast function over the collection of admissible values. The obtainable rate for the fully data-driven estimator $\widehat{\phi}_{\widehat{m}}(t_0)$ in the three considered cases is given as follows:

(pp) If $p > 1/2$, $a > 1/2$ and $p+a \geqslant 3/2$, then $\mathcal{R}_*^{\ell_{t_0}}[(1+\log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp (n^{-1}\log n)^{(2p-1)/(2p+2a)}$.

(pe) If $p > 1/2$ and $a > 0$, then $\mathcal{R}_*^{\ell_{t_0}}[(1+\log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp (\log n)^{-(2p-1)/(2a)}$.

(ep) If $p > 0$ and $a > 1/2$, then $\mathcal{R}_*^{\ell_{t_0}}[(1+\log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp n^{-1}(\log n)^{(2p+2a+1)/(2p)}$.

The proposed fully data-driven point wise estimator is minimax optimal in case (pe) which is easily seen by comparing the rates of the adaptive estimator with the corresponding minimax rate. In the other cases, the rates deviate only by logarithmic factor, as expected.

Point-wise estimation of derivatives. 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 ϕ 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 $\phi^{(q)}$ of ϕ in a weak sense satisfies

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

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

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

Minimax optimal point-wise estimation of derivatives. The estimator $\widehat{\phi}_{m_n^*}^{(q)}(t_0)$ with appropriately chosen dimension is minimax optimal, i.e., $\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}|\widehat{\phi}_{m_n^*}^{(q)}(t_0) - \phi^{(q)}(t_0)|^2 \leqslant C\,\mathcal{R}_*^{\ell_{t_0}^{(q)}}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$ for some $C > 0$, where $\mathcal{R}_*^{\ell_{t_0}^{(q)}}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$ is the minimax-optimal rate of convergence (c.f. Johannes and Schenk [2010]).

Illustration continued. In the considered cases we derive with $s = -q$

(pp) If $p > 1/2$, $a > 1/2$ and $p + a \geqslant 3/2$, then $\mathcal{R}_*^{\ell_{t_0}^{(q)}}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp n^{-(2p-2q-1)/(2p+2a)}$.

(pe) If $p > 1/2$ and $a > 0$, then $\mathcal{R}_*^{\ell_{t_0}^{(q)}}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp (\log n)^{-(2p-2q-1)/(2a)}$.

(ep) If $p > 0$ and $a > 1/2$, then $\mathcal{R}_*^{\ell_{t_0}^{(q)}}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp n^{-1}(\log n)^{(2a+2q+1)/(2p)}$.

Adaptive point-wise estimation of derivatives. In the three considered cases the obtainable rate of the fully data-driven estimator $\widehat{\phi}_{\widehat{m}}^{(q)}(t_0)$ is given as follows:

(pp) If $p > 1/2$, $a > 1/2$ and $p + a \geqslant 3/2$, then
$\mathcal{R}_*^{\ell_{t_0}^{(q)}}[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp (n^{-1} \log n)^{(2p-2q-1)/(2p+2a)}$.

(pe) If $p > 1/2$ and $a > 0$, then
$\mathcal{R}_*^{\ell_{t_0}^{(q)}}[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp (\log n)^{-(2p-2q-1)/2a}$.

(ep) If $p > 0$ and $a > 1/2$, then
$\mathcal{R}_*^{\ell_{t_0}^{(q)}}[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp n^{-1}(\log n)^{(2p+2a+2q+1)/2p}$.

Also in the situation of adaptively estimating the (q)-th derivative in a given point the obtained rates deteriorate by a logarithmic factor in the cases *(pp)* and *(pe)* only.

Local average estimation. Next we are interested in the average value of ϕ on the interval $[0, b]$ for $b \in (0, 1]$. If we denote the linear functional mapping h to $b^{-1} \int_0^b h(t)dt$ 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 $\widehat{\ell}_m^b = b^{-1} \int_0^b \widehat{\phi}_m(t)dt$ is written as

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

Minimax optimal estimation of local averages. The estimator $\widehat{\ell}_{m_n^*}^b$ attains the minimax optimal rate, i.e., $\sup_{\phi \in \mathcal{F}_\beta^r} \sup_{\Gamma \in \mathcal{G}_\gamma^d} \mathbb{E}| \int_0^b \widehat{\phi}_{m_n^*}(t)dt - \int_0^b \phi(t)dt|^2 \leqslant C\mathcal{R}_*^{\ell^b}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$ for $C > 0$.

Illustration continued. In the three cases the order of $\mathcal{R}_*^{\ell^b}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d]$ is given as follows:

(pp) If $p \geqslant 0$, $a > 1/2$ and $p + a > 3/2$, then $\mathcal{R}_*^{\ell^b}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp n^{-(2p+1)/(2p+2a)}$.

(pe) If $p \geqslant 0$ and $a > 0$, then $\mathcal{R}_*^{\ell^b}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp (\log n)^{-(2p+1)/2a}$.

(ep) If $p > 0$ and $a > 1/2$, then $\mathcal{R}_*^{\ell^b}[n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp n^{-1}(\log n)^{(2a-1)/2p}$.

Adaptive estimation of local averages. In the three considered cases the obtainable rate of the adaptive estimator $\widehat{\ell}_{\widehat{m}}^b$ is given below:

(pp) If $p \geqslant 0$, $a > 1/2$ and $p+a > 3/2$, then $\mathcal{R}_*^{\ell^b}[(1+\log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp (n^{-1} \log n)^{(2p+1)/(2p+2a)}$.

(pe) If $p \geqslant 0$ and $a > 0$, then $\mathcal{R}_*^{\ell^b}[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp (\log n)^{-(2p+1)/2a}$.

(ep) If $p > 0$ and $a > 1/2$, then $\mathcal{R}_*^{\ell^b}[(1 + \log n)n^{-1}; \mathcal{F}_\beta^r, \mathcal{G}_\gamma^d] \asymp n^{-1}(\log n)^{(2p+2a-1)/2p}$.

In this setting again, we notice a deterioration of logarithmic order in the cases *(pp)* and *(pe)* only.

Appendix

This section gathers preliminary technical results and the proofs of Proposition 3.3 and 3.4.

A Notations

We begin by defining and recalling the notations which are used in the proofs. Given an integer $m \geqslant 1$, $\mathbb{H}_m$ denotes the subspace of $\mathbb{H}$ spanned by the functions $\{\psi_1, \dots, \psi_m\}$. Π_m and $\Pi_m^\perp$ denote the orthogonal projections on $\mathbb{H}_m$ and its orthogonal complement $\mathbb{H}_m^\perp$ respectively. If K is an operator mapping $\mathbb{H}$ into itself and we restrict $\Pi_m K \Pi_m$ to an operator from $\mathbb{H}_m$ into itself, then it can be represented by the matrix $[K]_{\underline{m}}$. Furthermore, $[\nabla_v]_{\underline{m}}$ and $[\mathrm{I}]_{\underline{m}}$ denote the m-dimensional diagonal matrix with diagonal entries $(v_j)_{1\leqslant j\leqslant m}$ and the identity matrix respectively. With a slight abuse of notations $\|v\|$ denotes the euclidean norm of the vector v. In particular, for all $f \in \mathbb{H}_m$ we have $\|f\|_v^2 = [f]_{\underline{m}}^t [\nabla_v]_{\underline{m}} [f]_{\underline{m}} = \|[\nabla_v]_{\underline{m}}^{1/2} [f]_{\underline{m}}\|^2$. Moreover, we use the notations

$$\widehat{V}_m = \max_{1\leqslant k\leqslant m} [\ell]_{\underline{k}}^t [\widehat{\Gamma}]_{\underline{k}}^{-1} [\ell]_{\underline{k}},\ V_m = \max_{1\leqslant k\leqslant m} [\ell]_{\underline{k}}^t [\Gamma]_{\underline{k}}^{-1} [\ell]_{\underline{k}},\ V_m^\gamma = [\ell]_{\underline{m}}^t [\nabla_\gamma]_{\underline{m}}^{-1} [\ell]_{\underline{m}}.$$

Recall that $[\widehat{\Gamma}]_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n [X_i]_{\underline{m}} [X_i]_{\underline{m}}^t$ and $[\widehat{g}]_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n Y_i [X_i]_{\underline{m}}$ where $[\Gamma]_{\underline{m}} = \mathbb{E}[X]_{\underline{m}}[X]_{\underline{m}}^t$ and $[g]_{\underline{m}} = \mathbb{E}Y[X]_{\underline{m}}$. Given a Galerkin solution $\phi_m \in \mathbb{H}_m$, let $U_m := Y - \langle \phi_m, X\rangle_{\mathbb{H}} = \sigma\varepsilon + \langle \phi - \phi_m, X\rangle_{\mathbb{H}}$. We introduce $\rho_m^2 := \mathbb{E}U_m^2 = \sigma^2 + \langle \Gamma(\phi-\phi_m), (\phi-\phi_m)\rangle_{\mathbb{H}}$, $\sigma_Y^2 := \mathbb{E}Y^2 = \sigma^2 + \langle \Gamma\phi, \phi\rangle_{\mathbb{H}}$ and $\sigma_m^2 = 2\big(\sigma_Y^2 + [g]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1} [g]_{\underline{m}}\big)$ where we use that ε and X are uncorrelated. With these notations we have

$$\mathrm{p}_m = 100\sigma_m^2 V_m (1+\log n) n^{-1}, \quad \widehat{\mathrm{p}}_m = 700\widehat{\sigma}_m^2 \widehat{V}_m (1+\log n) n^{-1}.$$

Let us define the random matrix $[\Xi]_{\underline{m}}$ and random vector $[W]_{\underline{m}}$, respectively, by

$$[\Xi]_{\underline{m}} := [\Gamma]_{\underline{m}}^{-1/2} [\widehat{\Gamma}]_{\underline{m}} [\Gamma]_{\underline{m}}^{-1/2} - [\mathrm{I}]_{\underline{m}}, \quad \text{and} \quad [W]_{\underline{m}} := [\widehat{g}]_{\underline{m}} - [\widehat{\Gamma}]_{\underline{m}} [\phi_m]_{\underline{m}},$$

where $\mathbb{E}[\Xi]_{\underline{m}} = 0$, because $\mathbb{E}[\widehat{\Gamma}]_{\underline{m}} = [\Gamma]_{\underline{m}}$, and $\mathbb{E}[W]_{\underline{m}} = [\Gamma(\phi-\phi_m)]_{\underline{m}} = 0$. Furthermore, we introduce $\widehat{\sigma}_Y^2 := n^{-1}\sum_{i=1}^n Y_i^2$ and the events

$$\begin{aligned}
&\Omega_{m,n} := \{\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s \leqslant n\}, \quad \mho_{m,n} := \{8\sqrt{m}\|[\Xi]_{\underline{m}}\|_s \leqslant 1\},\\
&\mathcal{A}_n := \{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}, \quad \mathcal{B}_n := \{\|[\Xi]_{\underline{k}}\|_s \leqslant 1/8, \forall 1 \leqslant k \leqslant M_n^\ell\},\\
&\mathcal{C}_n := \{[W]_{\underline{k}}^t [\Gamma]_{\underline{k}}^{-1} [W]_{\underline{k}} \leqslant \frac{1}{8}([g]_{\underline{k}}^t [\Gamma]_{\underline{k}}^{-1} [g]_{\underline{k}} + \sigma_Y^2), \forall 1 \leqslant k \leqslant M_n^\ell\},
\end{aligned} \tag{A.1}$$

along with their respective complements $\Omega_{m,n}^c$, $\mho_{m,n}^c$, $\mathcal{A}_n^c$, $\mathcal{B}_n^c$, and $\mathcal{C}_n^c$. Here and subsequently, we will denote by C a universal numerical constant and by $C(\cdot)$ a constant depending only on the arguments. In both cases, the values of the constants may change with every appearance.

B Preliminary results

The proof of the next lemma can be found in Johannes and Schenk [2010]. It relies on the properties of the sequences β, γ and $[\ell]$ given in Assumption 2.1.

LEMMA B.1. *Let* T *belong to* $\mathcal{G}_\gamma^d$ *where the sequence* γ *satisfies Assumption 2.1, then we have*

$$\sup_{m\in\mathbb{N}}\Big\{\gamma_m\|[\mathrm{T}]_{\underline{m}}^{-1}\|\Big\} \leqslant 4d^3, \tag{B.1}$$

$$\sup_{m\in\mathbb{N}}\|[\nabla_\gamma]_{\underline{m}}^{1/2}[\mathrm{T}]_{\underline{m}}^{-1}[\nabla_\gamma]_{\underline{m}}^{1/2}\| \leqslant 4d^3, \tag{B.2}$$

$$\sup_{m\in\mathbb{N}}\|[\nabla_\gamma]_{\underline{m}}^{-1/2}[\mathrm{T}]_{\underline{m}}[\nabla_\gamma]_{\underline{m}}^{-1/2}\| \leqslant d. \tag{B.3}$$

Consider in addition $\phi \in \mathcal{F}_\beta^r$ *with sequence* β *satisfying Assumption 2.1. If* ϕ_m *denotes a Galerkin solution of* $g = \mathrm{T}\phi$ *then for any strictly positive sequence* $w := (w_j)_{j\geqslant 1}$ *such that* w/β *is non-increasing we obtain for all* $m \in \mathbb{N}$

$$\|\phi-\phi_m\|_w^2 \leqslant 34\, d^8\, r\, \frac{w_m}{\beta_m} \max\left(1, \frac{\gamma_m^2}{w_m} \max_{1\leqslant j\leqslant m}\left\{\frac{w_j}{\gamma_j^2}\right\}\right), \tag{B.4}$$

$$\|\phi_m\|_\beta^2 \leqslant 34\, d^8\, r, \quad \|\mathrm{T}^{1/2}(\phi-\phi_m)\|_{\mathbb{H}}^2 \leqslant 34\, d^9\, r\, \gamma_m\beta_m^{-1}. \tag{B.5}$$

Furthermore, under Assumption 2.1 we have

$$|\ell(\phi-\phi_m)|^2 \leqslant 2\, r\left\{\sum_{j>m}\frac{[\ell]_j^2}{\beta_j} + 2(1+d^4)\frac{\gamma_m}{\beta_m}\sum_{j=1}^m \frac{[\ell]_j^2}{\gamma_j}\right\}. \tag{B.6}$$

LEMMA B.2. *Let Assumption 2.1 be satisfied and define* $D := (4d^3)$*. For* $\Gamma \in \mathcal{G}_\gamma^d$ *we have*

- *(i)* $d^{-1} \leqslant V_m/V_m^\gamma \leqslant D$, $d^{-1} \leqslant \gamma_m\|[\Gamma]_{\underline{m}}^{-1}\|_s \leqslant D$ *and* $d^{-1} \leqslant \gamma_m \max_{1\leqslant k\leqslant m}\|[\Gamma]_{\underline{k}}^{-1}\|_s \leqslant D$ *for all* $m \geqslant 1$*,*
- *(ii)* $V_{M_n^+}^\gamma \leqslant n4D(1+\log n)^{-1}$ *and hence* $V_{M_n^+} \leqslant n4D^2(1+\log n)^{-1}$ *for all* $n \geqslant 1$*,*
- *(iii)* $2\max_{1\leqslant m\leqslant M_n^+}\|[\Gamma]_{\underline{m}}^{-1}\| \leqslant n$ *if* $n \geqslant 2D$ *and* $\|[\ell]_{\underline{M_n^+}}\|^2(1+\log n) \geqslant 8D^2$.

If ϕ *belongs in addition to* $\mathcal{F}_\beta^r$ *then it holds for all* $m \geqslant 1$

- *(iv)* $\rho_m^2 \leqslant \sigma_m^2 \leqslant 2(\sigma^2+35d^9 r)$ *and*
- *(v)* $\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\{\mathrm{p}_m+\mathrm{b}_m\} \leqslant 202D^4\,(\sigma^2+r)\,\mathcal{R}_m^\ell((1+\log n)n^{-1};\mathcal{F}_\beta^r,\mathcal{G}_\gamma^d)$.

PROOF OF LEMMA B.2. Due to (B.2) - (B.3) in Lemma B.1, we have $V_m \leqslant 4d^3[\ell]_{\underline{m}}^t[\nabla_\gamma]_{\underline{m}}^{-1}[\ell]_{\underline{m}} = DV_m^\gamma$ and $V_m^\gamma \leqslant d[\ell]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[\ell]_{\underline{m}} \leqslant dV_m$. Moreover, from (B.1) and (B.2) it follows that $\|[\Gamma]_{\underline{m}}^{-1}\|_s \leqslant 4d^3\gamma_m^{-1}$ and $\gamma_m^{-1} \leqslant d\|[\Gamma]_{\underline{m}}^{-1}\|_s$. Thus, for all $m \geqslant 1$ we have $D \geqslant \|[\Gamma]_{\underline{m}}^{-1}\|_s\gamma_m \geqslant d^{-1}$. Hence, the monotonicity of γ implies $d^{-1} \leqslant \gamma_M \max_{1\leqslant m\leqslant M}\|[\Gamma]_{\underline{m}}^{-1}\|_s \leqslant D$. From these estimates we obtain (i).

Proof of (ii). Observe that $V_{M_n^+}^\gamma \leqslant \|[\ell]_{\underline{M_n^+}}\|^2\gamma_{M_n^+}^{-1}$. In case $M_n^+ = 1$ the assertion is trivial, since $[\ell]_1^2 = \gamma_1$ due to Assumption 2.1. Thus, consider $M_n^\ell \geqslant M_n^+ > 1$, which implies

$\min_{1\leqslant j\leqslant M_n^+}\{\gamma_j\|[\ell]_{\underline{M_n^+}}\|^{-2}\} \geqslant (1+\log n)/(4Dn)$, and hence $V^\gamma_{M_n^+} \leqslant 4Dn(1+\log n)^{-1}$. Moreover, from (i) follows $V_{M_n^+} \leqslant DV^\gamma_{M_n^+} \leqslant 4D^2n(1+\log n)^{-1}$, which proves (ii).
Proof of (iii). By employing that $D\gamma^{-1}_{M_n^+} \geqslant \max_{1\leqslant m\leqslant M_n^+}\|[\Gamma]^{-1}_{\underline{m}}\|$, the assertion (iii) follows in case $M_n^+ = 1$ from $\gamma_1 = 1$, while in case $M_n^+ > 1$, we use $\|[\ell]_{\underline{M_n^+}}\|^2/\gamma_{M_n^+} \leqslant 4Dn/(1+\log n)$.
Proof of (iv). Since ε and X are centered it follows from $[\phi_m]_{\underline{m}} = [\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}$ that $\rho^2_m \leqslant 2\big(\mathbb{E}Y^2 + \mathbb{E}|\langle\phi_m, X\rangle_{\mathbb{H}}|^2\big) = 2\big(\sigma^2_Y + [g]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}}\big) = \sigma^2_m$. Moreover, by employing successively the inequality of Heinz [1951], i.e. $\|\Gamma^{1/2}\phi\|^2 \leqslant d\|\phi\|^2_\gamma$, and Assumption 2.1, i.e., γ and β^{-1} are non-increasing, the identity $\sigma^2_Y = \sigma^2 + \langle\Gamma\phi,\phi\rangle_{\mathbb{H}}$ implies

$$\sigma^2_Y \leqslant \sigma^2 + d\|\phi\|^2_\gamma \leqslant \sigma^2 + dr. \tag{B.7}$$

Furthermore, (B.3) and (B.4) in Lemma B.1 imply

$$[g]^t_{\underline{k}}[\Gamma]^{-1}_{\underline{k}}[g]_{\underline{k}} \leqslant d\|\phi_k\|^2_\gamma \leqslant 34d^9r. \tag{B.8}$$

The assertion (iv) follows now by combination of the estimates (B.7) and (B.8).
Proof of (v). From $V_m \leqslant DV^\gamma_m$ due to assertion (i) and the second inequality in (iv) we derive

$$\mathrm{p}_m \leqslant 100\sigma^2_m(1+\log n)n^{-1}DV^\gamma_m \leqslant 200(\sigma^2+r)D^4(1+\log n)n^{-1}\sum_{j=1}^m[\ell]^2_j\gamma^{-1}_j. \tag{B.9}$$

Furthermore, by using (B.6) in Lemma B.1 we obtain that

$$\mathrm{b}_m \leqslant 16d^4\, r\,\{\max(\sum_{j>m}[\ell]^2_j\beta^{-1}_j, \gamma_m\beta^{-1}_m\sum_{j=1}^m[\ell]^2_j\gamma^{-1}_j)\}. \tag{B.10}$$

Combining the bounds (B.9) and (B.10) implies assertion (v), which completes the proof. □

Lemma B.3. *For all $n, m \geqslant 1$ we have*

$$\Big\{\frac14 < \frac{\|[\widehat\Gamma]^{-1}_{\underline{m}}\|_s}{\|[\Gamma]^{-1}_{\underline{m}}\|_s} \leqslant 4, \forall\, 1\leqslant m\leqslant M^\ell_n\Big\} \subset \Big\{M^-_n \leqslant \widehat{M}_n \leqslant M^+_n\Big\}.$$

Proof of Lemma B.3. Let $\widehat\tau_m = \|[\widehat\Gamma]^{-1}_{\underline{m}}\|^{-1}_s$ and recall that $1 \leqslant \widehat{M}_n \leqslant M^\ell_n$ with

$$\Big\{\widehat{M}_n = M\Big\} = \begin{cases} \Big\{\frac{\widehat\tau_{M+1}}{\|[\ell]^2_{\underline{M+1}}\|^2} < \frac{1+\log n}{n}\Big\}, & M=1,\\ \Big\{\min\limits_{2\leqslant m\leqslant M}\frac{\widehat\tau_m}{\|[\ell]^2_{\underline{m}}\|^2} \geqslant \frac{1+\log n}{n}\Big\} \bigcap \Big\{\frac{\widehat\tau_{M+1}}{\|[\ell]^2_{\underline{M+1}}\|^2} < \frac{1+\log n}{n}\Big\}, & 1<M<M^\ell_n,\\ \Big\{\min\limits_{2\leqslant m\leqslant M}\frac{\widehat\tau_m}{\|[\ell]^2_{\underline{m}}\|^2} \geqslant \frac{1+\log n}{n}\Big\}, & M = M^\ell_n.\end{cases}$$

Given $\tau^{-1}_m := \|[\Gamma]^{-1}_{\underline{m}}\|_s$ we have $D^{-1} \leqslant \tau_m/\gamma_m \leqslant d$ for all $m \geqslant 1$ due to (i) in Lemma B.2 which we use to proof the following two assertions

$$\Big\{\widehat{M}_n < M^-_n\Big\} \subset \Big\{\min_{1\leqslant m\leqslant M^\ell_n} : \frac{\widehat\tau_m}{\tau_m} < \frac14\Big\}, \tag{B.11}$$

$$\Big\{\widehat{M}_n > M^+_n\Big\} \subset \Big\{\max_{1\leqslant m\leqslant M^\ell_n} \frac{\widehat\tau_m}{\tau_m} \geqslant 4\Big\}. \tag{B.12}$$

Obviously, the assertion of the Lemma follows now by combination of (B.11) and (B.12). Consider (B.11) which is trivial in case $M_n^- = 1$. For $M_n^- > 1$ we have $\min_{1\leqslant m\leqslant M_n^-} \frac{\gamma_m}{\|[\ell]^2_{\underline{m}}\|^2} \geqslant \frac{4D(1+\log n)}{n}$ and, hence $\min_{1\leqslant m\leqslant M_n^-} \frac{\tau_m}{\|[\ell]^2_{\underline{m}}\|^2} \geqslant \frac{4(1+\log n)}{n}$. By exploiting the last estimate we obtain

$$\left\{\widehat{M}_n < M_n^\ell\right\} \cap \left\{\widehat{M}_n < M_n^-\right\} = \bigcup_{M=1}^{M_n^- -1} \left\{\widehat{M}_n = M\right\}$$
$$\subset \bigcup_{M=1}^{M_n^- -1} \left\{\frac{\widehat{\tau}_{M+1}}{\|[\ell]^2_{\underline{M+1}}\|^2} < \frac{1+\log n}{n}\right\} = \left\{\min_{2\leqslant m\leqslant M_n^-} \frac{\widehat{\tau}_m}{\|[\ell]^2_{\underline{m}}\|^2} < \frac{1+\log n}{n}\right\}$$
$$\subset \left\{\min_{1\leqslant m\leqslant M_n^-} \frac{\widehat{\tau}_m}{\tau_m} < 1/4\right\}$$

while trivially $\left\{\widehat{M}_n = M_n^\ell\right\} \cap \left\{\widehat{M}_n < M_n^-\right\} = \emptyset$ which proves (B.11) because $M_n^- \leqslant M_n^\ell$. Consider (B.12) which is trivial in case $M_n^+ = M_n^\ell$. If $M_n^+ < M_n^\ell$, then $\frac{\tau_{M_n^+ +1}}{\|[\ell]^2_{\underline{M_n^+ +1}}\|^2} < \frac{(1+\log n)}{4n}$, and hence

$$\left\{\widehat{M}_n > 1\right\} \cap \left\{\widehat{M}_n > M_n^+\right\} = \bigcup_{M=M_n^+ +1}^{M_n^\ell} \left\{\widehat{M}_n = M\right\}$$
$$\subset \bigcup_{M=M_n^+ +1}^{M_n^\ell} \left\{\min_{2\leqslant m\leqslant M} \frac{\widehat{\tau}_m}{\|[\ell]^2_{\underline{m}}\|^2} \geqslant \frac{1+\log n}{n}\right\} = \left\{\min_{2\leqslant m\leqslant (M_n^+ +1)} \frac{\widehat{\tau}_m}{\|[\ell]^2_{\underline{m}}\|^2} \geqslant \frac{1+\log n}{n}\right\}$$
$$\subset \left\{\frac{\widehat{\tau}_{M_n^+ +1}}{\tau_{M_n^+ +1}} \geqslant 4\right\}$$

while $\left\{\widehat{M}_n = 1\right\} \cap \left\{\widehat{M}_n > M_n^+\right\} = \emptyset$ which shows (B.12) and completes the proof. □

Lemma B.4. *Let $\mathcal{A}_n$, $\mathcal{B}_n$ and $\mathcal{C}_n$ as in* (A.1). *For all $n \geqslant 1$ it holds true that*

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \{\mathrm{p}_k \leqslant \widehat{\mathrm{p}}_k \leqslant 24\,\mathrm{p}_k, 1 \leqslant k \leqslant M_n^\ell\} \cap \{M_n^- \leqslant \widehat{M}_n \leqslant M_n^+\}.$$

Proof of Lemma B.4. Let $M_n^\ell \geqslant k \geqslant 1$. If $\|[\Xi]_{\underline{k}}\|_s \leqslant 1/8$, i.e., on the event $\mathcal{B}_n$, it is easily verified that $\|([\mathrm{I}]_{\underline{k}} + [\Xi]_{\underline{k}})^{-1} - [\mathrm{I}]_{\underline{k}}\|_s \leqslant 1/7$ which we exploit to conclude

$$(6/7)\|[\Gamma]_{\underline{k}}^{-1}\|_s \leqslant \|[\widehat{\Gamma}]_{\underline{k}}^{-1}\|_s \leqslant (8/7)\|[\Gamma]_{\underline{k}}^{-1}\|_s \quad \text{and}$$
$$(6/7)s^t[\Gamma]_{\underline{k}}^{-1}s \leqslant s^t[\widehat{\Gamma}]_{\underline{k}}^{-1}s \leqslant (8/7)s^t[\Gamma]_{\underline{k}}^{-1}s, \quad \text{for all } s \in \mathbb{R}^k, \tag{B.13}$$

and, consequently

$$(6/7)[\widehat{g}]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} \leqslant [\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} \leqslant (8/7)[\widehat{g}]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}}. \tag{B.14}$$

Moreover, from $\|[\Xi]_{\underline{k}}\|_s \leqslant 1/8$ we obtain after some algebra,

$$[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} \leqslant \frac{1}{16}[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + 4[W]_{\underline{k}}[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}} + 2[\widehat{g}]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}},$$
$$[\widehat{g}]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} \leqslant \frac{33}{16}[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + 4[W]_{\underline{k}}[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}}.$$

Combining each of these estimates with (B.14) yields

$$(15/16)[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} \leqslant 4[W]_{\underline{k}}[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}} + (7/3)[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}},$$
$$(7/8)[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} \leqslant (33/16)[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + 4[W]_{\underline{k}}[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}}.$$

If in addition $[W]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}} \leqslant \frac{1}{8}([g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + \sigma_Y^2)$, i.e., on the event $\mathcal{C}_n$, then the last two estimates imply respectively

$$(7/16)([g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + \sigma_Y^2) \leqslant (15/16)\sigma_Y^2 + (7/3)[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}},$$
$$(7/8)[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} \leqslant (41/16)[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + (1/2)\sigma_Y^2,$$

and hence in case $1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2$, i.e., on the event $\mathcal{A}_n$, we obtain

$$(7/16)([g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + \sigma_Y^2) \leqslant (15/8)\widehat{\sigma}_Y^2 + (7/3)[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}},$$
$$(7/8)([\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} + \widehat{\sigma}_Y^2) \leqslant (41/16)[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + (29/16)\sigma_Y^2.$$

Combining the last two estimates yields

$$\frac{1}{6}(2[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + 2\sigma_Y^2) \leqslant (2[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} + 2\widehat{\sigma}_Y^2) \leqslant 3(2[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + 2\sigma_Y^2).$$

Since the last estimate and (B.13) hold for all $1 \leqslant k \leqslant M_n^\ell$ on the event $\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n$ it follows

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \left\{\frac{1}{6}\sigma_m^2 \leqslant \widehat{\sigma}_m^2 \leqslant 3\sigma_m^2 \text{ and } (6/7)V_m \leqslant \widehat{V}_m \leqslant (8/7)V_m, \forall 1 \leqslant m \leqslant M_n^\ell\right\}.$$

The definitions of $\mathrm{p}_m = 100\sigma_m^2 V_m(1+\log n)n^{-1}$ and $\widehat{\mathrm{p}}_m = 700\widehat{\sigma}_m^2\widehat{V}_m(1+\log n)n^{-1}$ imply

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \left\{\mathrm{p}_m \leqslant \widehat{\mathrm{p}}_m \leqslant 24\,\mathrm{p}_m, \forall 1 \leqslant m \leqslant M_n^\ell\right\}. \tag{B.15}$$

On the other hand, by exploiting successively (B.13) and Lemma B.3 we obtain

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \left\{\frac{6}{7} \leqslant \frac{\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s}{\|[\Gamma]_{\underline{m}}^{-1}\|_s} \leqslant \frac{8}{7}, \forall 1 \leqslant m \leqslant M_n^\ell\right\} \subset \left\{M_n^- \leqslant \widehat{M}_n \leqslant M_n^+\right\}. \tag{B.16}$$

From (B.15) and (B.16) follows the assertion of the lemma, which completes the proof. □

Lemma B.5. *For all $m, n \geqslant 1$ with $n \geqslant (8/7)\|[\Gamma]_{\underline{m}}^{-1}\|_s$ we have $\mho_{m,n} \subset \Omega_{m,n}$.*

Proof of Lemma B.5. Taking the identity $[\widehat{\Gamma}]_{\underline{m}} = [\Gamma]_{\underline{m}}^{1/2}\{[\mathrm{I}]_{\underline{m}} + [\Xi]_{\underline{m}}\}[\Gamma]_{\underline{m}}^{1/2}$ into account, we observe that $\sqrt{m}\|[\Xi]_{\underline{m}}\|_s \leqslant 1/8$ implies $\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s \leqslant \frac{8\sqrt{m}}{8\sqrt{m}-1}\|[\Gamma]_{\underline{m}}^{-1}\|_s \leqslant (8/7)\|[\Gamma]_{\underline{m}}^{-1}\|_s$ due to the usual Neumann series argument. If $n \geqslant (8/7)\|[\Gamma]_{\underline{m}}^{-1}\|_s$, then the last assertion implies $\mho_{m,n} \subset \Omega_{m,n}$, which proves the lemma. □

C Preliminary results due to the normality assumption

We will suppose throughout this section that the conditions of Theorem 3.1 and in particular Assumption 2.1 are satisfied, thus, the technical Lemmas stated in Section B are applicable. We show technical assertions under the assumption of normality (Lemmas C.1- C.4) which are used below to prove Propositions 3.3 and 3.4.

We begin by recalling elementary properties due to the assumption that X and ε are jointly normally distributed, which are frequently used in the following proofs. For any $h \in \mathbb{H}$ the random variable $\langle h, X\rangle_{\mathbb{H}}$ is normally distributed with mean zero and variance $\langle \Gamma h, h\rangle_{\mathbb{H}}$. Consider the Galerkin solution ϕ_m and $h \in \mathbb{H}_m$ then the random variables $\langle \phi - \phi_m, X\rangle_{\mathbb{H}}$ and $\langle h, X\rangle_{\mathbb{H}}$ are independent. Thereby, $U_m = Y - \langle \phi_m, X\rangle_{\mathbb{H}} = \sigma\varepsilon + \langle \phi - \phi_m, X\rangle_{\mathbb{H}}$ and $[X]_{\underline{m}}$ are independent, normally distributed with mean zero, and, respectively, variance ρ_m^2 and covariance matrix $[\Gamma]_{\underline{m}}$. Consequently, $(\rho_m^{-1} U_m, [X]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1/2})$ is a $(m+1)$-dimensional vector of i.i.d. standard normally distributed random variables. Let us further state elementary inequalities for Gaussian random variables.

LEMMA C.1. *Let $\{U_i, V_{ij}, 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$ be independent and standard normally distributed random variables. We have for all $\eta > 0$ and $\zeta \geqslant 4m/n$*

$$P\Big(|n^{-1/2}\sum_{i=1}^n (U_i^2 - 1)| \geqslant \eta\Big) \leqslant 2\exp\Big(-\frac{\eta^2}{8(1+\eta\, n^{-1/2})}\Big); \tag{C.1}$$

$$P\Big(|n^{-1}\sum_{i=1}^n U_i V_{i1}| \geqslant \eta\Big) \leqslant \frac{\eta n^{1/2} + 2}{\eta n^{1/2}} \exp\Big(-\frac{n}{4}\min\Big\{\eta^2, \frac{1}{4}\Big\}\Big); \tag{C.2}$$

$$P\Big(n^{-2}\sum_{j=1}^m |\sum_{i=1}^n U_i V_{ij}|^2 \geqslant \zeta\Big) \leqslant \exp\Big(-\frac{n}{16}\Big) + \exp\Big(-\frac{\zeta n}{64}\Big); \tag{C.3}$$

and for all $c > 0$ and $a_1, \dots, a_m \geqslant 0$ that

$$\mathbb{E}\Big(n^{-1}\sum_{i=1}^n U_i^2 - 2\Big)_+ \leqslant \frac{16}{n}\exp\Big(-\frac{n}{16}\Big); \tag{C.4}$$

$$\mathbb{E}\Big(|n^{-1/2}\sum_{i=1}^n U_i V_{i1}|^2 - 4c(1+\log n)\Big)_+ \leqslant \frac{2n^{-c}}{e^c\sqrt{\pi c(1+\log n)}} + 32c\exp\Big(-\frac{n}{16}\Big); \tag{C.5}$$

$$\mathbb{E}\Big(\sum_{j=1}^m a_j |\sum_{i=1}^n U_i V_{ij}|^2\Big)^4 \leqslant n^4\Big(11\sum_{j=1}^m a_j\Big)^4. \tag{C.6}$$

PROOF OF LEMMA C.1. Define $W := \sum_{i=1}^n U_i^2$ and $Z_j := (\sum_{i=1}^n U_i^2)^{-1/2}\sum_{i=1}^n U_i V_{ij}$. Obviously, W has a χ^2 distribution with n degrees of freedom and $Z_1, \dots, Z_m$ given $U_1, \dots, U_n$ are independent and standard normally distributed, which we use below without further reference. The estimate (C.1) is given in Dahlhaus and Polonik [2006] (Proposition A.1) and by using

(C.1) we have

$$P(|n^{-1}\sum_{i=1}^{n}U_iV_{i1}|\geqslant\eta)\leqslant P(n^{-1}W\geqslant 2)+\mathbb{E}\big[P\big(2n^{-1}|Z_1|^2\geqslant\eta^2\big|U_1,\dots,U_n\big)\big]$$
$$\leqslant\exp\Big(-\frac{n}{16}\Big)+\frac{2}{\sqrt{\pi\eta^2 n}}\exp\Big(-\frac{\eta^2 n}{4}\Big),$$

which implies (C.2). The estimate (C.3) follows analogously and we omit the details. By using (C.1) we obtain (C.4) as follows

$$\mathbb{E}\left(n^{-1}\sum_{i=1}^{n}U_i^2-2\right)_+=\int_0^\infty P(n^{-1/2}\sum_{i=1}^{n}(U_i^2-1)\geqslant n^{1/2}(1+t))dt$$
$$\leqslant\int_0^\infty\exp\Big(-\frac{n(1+t)^2}{8(1+(1+t))}\Big)dt\leqslant\int_0^\infty\exp\Big(-\frac{n(1+t)}{16}\Big)dt$$
$$=\exp\Big(-\frac{n}{16}\Big)\int_0^\infty\exp\Big(-\frac{n}{16}t\Big)dt=\frac{16}{n}\exp\Big(-\frac{n}{16}\Big).$$

Consider (C.5). Since $n^{-1/2}\sum_{i=1}^{n}U_i$ is standard normally distributed, we have

$$\mathbb{E}\left(|n^{-1/2}\sum_{i=1}^{n}U_i|^2-2c(1+\log n)\right)_+=\int_0^\infty P(|n^{-1/2}\sum_{i=1}^{n}U_i|\geqslant(t+2c(1+\log n))^{1/2})dt$$
$$\leqslant\int_0^\infty\frac{2}{\sqrt{2\pi(t+2c(1+\log n))}}\exp\Big(-\frac{(t+2c(1+\log n))}{2}\Big)dt$$
$$\leqslant\frac{e^{-c}n^{-c}}{\sqrt{\pi c(1+\log n)}}\int_0^\infty\exp\Big(-\frac{1}{2}t\Big)dt=\frac{2e^{-c}n^{-c}}{\sqrt{\pi c(1+\log n)}}.$$

By using the last bound and (C.4) we get

$$\mathbb{E}\left(|n^{-1/2}\sum_{i=1}^{n}U_iV_{i1}|^2-4c(1+\log n)\right)_+$$
$$\leqslant\mathbb{E}\Big[n^{-1}W\mathbb{E}\big[\big(|Z_1|^2-2c(1+\log n)\big)_+|U_1,\dots,U_n\big]+2c(1+\log n)\big(n^{-1}W-2\big)_+\Big]$$
$$\leqslant\frac{2n^{-c}}{e^c\sqrt{\pi c(1+\log n)}}+32c\frac{(1+\log n)}{n}\exp\Big(-\frac{n}{16}\Big)$$

which shows (C.5). Finally, by applying $\mathbb{E}[Z_j^8|U_1,\dots,U_n]=105$ and $\mathbb{E}W^4=n(n+2)(n+4)(n+6)$ we obtain $\mathbb{E}[W^4Z_j^8]\leqslant(11n)^4$ and hence

$$\mathbb{E}\left(\sum_{j=1}^{m}a_j|\sum_{i=1}^{n}U_iV_{ij}|^2\right)^4=\mathbb{E}\Big(\sum_{j=1}^{m}a_jWZ_j^2\Big)^4\leqslant\Big|\sum_{j=1}^{m}a_j(\mathbb{E}[W^4Z_j^8])^{1/4}\Big|^4\leqslant(11n)^4(\sum_{j=1}^{m}a_j)^4$$

which shows (C.6) and completes the proof. □

Lemma C.2. *For all $n, m \geqslant 1$ we have*

$$n^4 m^{-4} \mathbb{E}\|[\Xi]_{\underline{m}}[\Gamma]_{\underline{m}}^{1/2}\|_s^8 \leqslant (34\mathbb{E}\|X\|_{\mathbb{H}}^2)^4; \tag{C.7}$$

$$n^4 \rho_m^{-8} \mathbb{E}\|[W]_{\underline{m}}\|^8 \leqslant (11\mathbb{E}\|X\|_{\mathbb{H}}^2)^4. \tag{C.8}$$

Furthermore, there exists a numerical constant C such that for all $n \geqslant 1$

$$n^8 \max_{1\leqslant m\leqslant \lfloor n^{1/4}\rfloor} P\bigg(\frac{([W]_m^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}})}{\rho_m^2} > \frac{1}{16}\bigg) \leqslant C; \tag{C.9}$$

$$n^8 \max_{1\leqslant m\leqslant \lfloor n^{1/4}\rfloor} P\bigg(\sqrt{m}\|[\Xi]_{\underline{m}}\|_s > \frac{1}{8}\bigg) \leqslant C; \tag{C.10}$$

$$n^7 P\big(\{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}^c\big) \leqslant C; \tag{C.11}$$

$$n^2 \sup_{m\geqslant 1} \mathbb{E}\bigg(\frac{n([W]_m^t[\Gamma]_m^{-1}[W]_m)}{m\rho_m^2} - 8(1+\log n)\bigg)_+ \leqslant C; \tag{C.12}$$

$$n^2 \sup_{m\geqslant 1} \mathbb{E}\bigg(\frac{n([\ell]_m^t[\Gamma]_m^{-1}[W]_m)^2}{\rho_m^2[\ell]_m^t[\Gamma]_m^{-1}[\ell]_m} - 8(1+\log n)\bigg)_+ \leqslant C. \tag{C.13}$$

Proof of Lemma C.2. Let $n, m \geqslant 1$ be fixed and denote by $(\lambda_j, e_j)_{1\leqslant j\leqslant m}$ an eigenvalue decomposition of $[\Gamma]_{\underline{m}}$. Define $U_i := (\sigma\varepsilon_i + \langle \phi - \phi_m, X_i\rangle_{\mathbb{H}})/\rho_m$ and $V_{ij} := (\lambda_j^{-1/2} e_j^t [X_i]_{\underline{m}})$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$, where $U_1, \dots, U_n, V_{11}, \dots, V_{nm}$ are independent and standard normally distributed random variables.

Proof of (C.7). For all $1 \leqslant j, l \leqslant m$ let $\delta_{jl} = 1$ if $j = l$ and zero otherwise. It is easily verified that $\|[\Xi]_{\underline{m}}[\Gamma]_{\underline{m}}^{1/2}\|_s^2 \leqslant \sum_{j=1}^m \sum_{l=1}^m \lambda_l |n^{-1}\sum_{i=1}^n (V_{ij}V_{il} - \delta_{jl})|^2$. Moreover, for $j \neq l$ we have $\mathbb{E}|\sum_{i=1}^n V_{ij}V_{il}|^8 \leqslant (11n)^4$ by employing (C.6) in Lemma C.1 (take $m = 1$ and $a_1 = 1$), while $\mathbb{E}|\sum_{i=1}^n (V_{ij}^2 - 1)|^8 = n^4 256(105/16 + 595/(2n) + 1827/n^2 + 2520/n^3) \leqslant (34n)^4$. From these estimates we get by successively employing Jensen's and Minkowski's inequality that

$$m^{-4}\mathbb{E}\|[\Xi]_{\underline{m}}[\Gamma]_{\underline{m}}^{1/2}\|_s^8 \leqslant n^{-8}m^{-1}\sum_{j=1}^m \big(\sum_{l=1}^m \lambda_l (\mathbb{E}|\sum_{i=1}^n (V_{ij}V_{il} - \delta_{jl})|^8)^{1/4}\big)^4 \leqslant n^{-4}(34\sum_{j=1}^m \lambda_j)^4.$$

The last estimate together with $\sum_{j=1}^m \lambda_j = \operatorname{tr}([\Gamma]_{\underline{m}}) \leqslant \operatorname{tr}(\Gamma) = \mathbb{E}\|X\|_{\mathbb{H}}^2$ implies (C.7).

Proof of (C.8) and (C.9). Taking the inequality $\sum_{j=1}^m \lambda_j \leqslant \mathbb{E}\|X\|_{\mathbb{H}}^2$ and the identities $n^4\rho_m^{-8}\|[W]_{\underline{m}}\|^8 = (\sum_{j=1}^m \lambda_j (\sum_{i=1}^n U_i V_{ij})^2)^4$ and $([W]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}})/\rho_m^2 = n^{-2}\sum_{j=1}^m (\sum_{i=1}^n U_i V_{ij})^2$ into account the assertions (C.8) and (C.9) follow, respectively, from (C.6) and (C.3) in Lemma C.1 (with $a_j = \lambda_j$).

Proof of (C.10). Since $n\|[\Xi]_{\underline{m}}\|_s \leqslant m \max_{1\leqslant j,l\leqslant m} |\sum_{i=1}^n (V_{ij}V_{il} - \delta_{jl})|$ we obtain due to

(C.1) and (C.2) in Lemma C.1 for all $\eta > 0$ the following bound

$$
\begin{aligned}
P(\|[\Xi]_{\underline{m}}\|_s \geqslant \eta) &\leqslant \sum_{1\leqslant j,l\leqslant m} P(|n^{-1}\sum_{i=1}^{n}(V_{ij}V_{il} - \delta_{jl})| \geqslant \eta/m) \\
&\leqslant m^2 \max\left\{P(|n^{-1}\sum_{i=1}^{n} V_{i1}V_{i2}| \geqslant \eta/m), P(|n^{-1/2}\sum_{i=1}^{n}(V_{i1}^2 - 1)| \geqslant n^{1/2}\eta/m)\right\} \\
&\leqslant m^2 \max\left\{(1 + \frac{m}{\eta n^{1/2}})\exp\left(-\frac{n}{4}\min\left\{\eta^2/m^2, 1/4\right\}\right), 2\exp\left(-\frac{1}{8}\frac{n\eta^2/m^2}{1+\eta/m}\right)\right\}.
\end{aligned}
$$

Moreover, for all $\eta \leqslant m/2$ this can be simplified to

$$
P(\|[\Xi]_{\underline{m}}\|_s \geqslant \eta) \leqslant m^2 \max\left\{1 + \frac{2m}{\eta n^{1/2}}, 2\right\}\exp\left(-\frac{1}{12}\frac{n\eta^2}{m^2}\right),
$$

which obviously implies (C.5).

Proof of (C.11). Since $Y_1/\sigma_Y, \dots, Y_n/\sigma_Y$ are independent and standard normally distributed, (C.11) follows from (C.1) in Lemma C.1 by exploiting that $\{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}^c \subset \{|n^{-1}\sum_{i=1}^n Y_i^2/\sigma_Y^2 - 1| > 1/2\}$.

Proof of (C.12). From the identity $n([W]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}})/(m\rho_m^2) = m^{-1}\sum_{j=1}^m(n^{-1/2}\sum_{i=1}^n U_iV_{ij})^2$ the estimate (C.12) follows by using (C.6) in Lemma C.1, that is

$$
\begin{aligned}
\sup_{m\geqslant 1}\mathbb{E}\left(\frac{n([W]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}})}{m\rho_m^2} - 8(1+\log n)\right)_+ &\leqslant \mathbb{E}\left(|n^{-1/2}\sum_{i=1}^n U_iV_{i1}|^2 - 8(1+\log n)\right)_+ \\
&\leqslant \left\{\frac{n^{-2}}{e^2\sqrt{\pi 2(1+\log n)}} + 64\frac{(1+\log n)}{n}\exp(-n/16)\right\} \leqslant Cn^{-2}.
\end{aligned}
$$

Proof of (C.13). Define $V_i := ([\ell]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[\ell]_{\underline{m}})^{-1/2}[\ell]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[X_i]_{\underline{m}}$ for $1 \leqslant i \leqslant n$, where $U_1, \dots, U_n$, $V_1, \dots, V_n$ are independent and standard normally distributed random variables. By employing the identity $n([\ell]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}})^2/(\rho_m^2[\ell]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[\ell]_{\underline{m}}) = |n^{-1/2}\sum_{i=1}^n U_iV_i|^2$ the estimate (C.13) follows from (C.6) in Lemma C.1, which completes the proof. □

Lemma C.3. *There exists a constant $C(d)$ only depending on d such that for all $n \geqslant 1$*

$$
\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{m=1}^{M_n^+}\mathbb{E}\left(\frac{([\ell]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[\ell]_{\underline{m}})}{m}([W]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}}) - \frac{8\,\mathrm{p}_m}{100}\right)_+ \leqslant C(d)(\sigma^2 + r)n^{-1}; \tag{C.14}
$$

$$
\sup_{\phi\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{m=1}^{M_n^+}\mathbb{E}\left(([\ell]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}})^2 - \frac{8\,\mathrm{p}_m}{100}\right)_+ \leqslant C(d)(\sigma^2 + r)n^{-1}. \tag{C.15}
$$

Proof of Lemma C.3. The key argument to show (C.14) is the estimate (C.12) in Lemma C.2. Taking $[\ell]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[\ell]_{\underline{m}} \leqslant V_m$ and $\frac{8\,\mathrm{p}_m}{100} = 8\,\sigma_m^2\,V_m\frac{1+\log n}{n}$ into account, together with the facts that $\max_{1\leqslant m\leqslant M_n^+} V_m = V_{M_n^+} \leqslant nC(d)(1+\log n)^{-1}$ and $\rho_m^2 \leqslant \sigma_m^2 \leqslant C(d)(\sigma^2+r)$ for all $\phi \in \mathcal{F}_\beta^r$, $\Gamma \in \mathcal{G}_\gamma^d$ (Lemma B.2 (ii) and (iv)) we obtain

$$\sum_{m=1}^{M_n^+} \mathbb{E}\left(\frac{([\ell]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[\ell]_{\underline{m}})}{m}([W]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}}) - \frac{8\,\mathrm{p}_m}{100}\right)_+$$
$$\leqslant \sum_{m=1}^{M_n^+} \frac{\sigma_m^2 V_m}{n}\,\mathbb{E}\left(\frac{n([W]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}})}{m\rho_m^2} - 8\,(1+\log n)\right)_+$$
$$\leqslant \frac{C(d)(\sigma^2+r)}{1+\log n} M_n^+ \sup_{m\geqslant 1} \mathbb{E}\left(\frac{([W]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}})}{m\rho_m^2} - 8\,(1+\log n)\right)_+.$$

The assertion (C.14) follows by employing (C.12) in Lemma C.2 and $M_n^+ \leqslant n$. The proof of (C.15) follows the same lines by using (C.13) in Lemma C.2 rather than (C.12) and we omit the details. □

Lemma C.4. *There exists a numerical constant C and a constant $C(d)$ only depending on d such that for all $n \geqslant 1$*

$$\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\left\{n^4(M_n^+)^4 \max_{1\leqslant m\leqslant M_n^+} P\big(\mho_{m,n}^c\big)\right\} \leqslant C; \tag{C.16}$$

$$\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\left\{n\,M_n^+ \max_{1\leqslant m\leqslant M_n^+} P\big(\Omega_{m,n}^c\big)\right\} \leqslant C(d); \tag{C.17}$$

$$\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\big\{n^7 P(\mathcal{E}_n^c)\big\} \leqslant C. \tag{C.18}$$

Proof of Lemma C.4. Since $M_n^+ \leqslant \lfloor n^{1/4}\rfloor$ and $\mho_{m,n}^c = \big\{\sqrt{m}\|[\Xi]_{\underline{m}}\|_s > 1/8\big\}$ the assertion (C.16) follows from (C.10) in Lemma C.2.
Consider (C.17). With $n_o := n_o(d) := \exp(128d^6) \geqslant 8d^3$ we have $\|[\ell]_{\underline{M_n^+}}\|^2(1+\log n) \geqslant 128d^6$ for all $n \geqslant n_o$. We distinguish in the following the cases $n < n_o$ and $n \geqslant n_o$. First, consider $1 \leqslant n \leqslant n_o$. Obviously, we have $M_n^+ \max_{1\leqslant m\leqslant M_n^+} P(\Omega_{m,n}^c) \leqslant M_n^+ \leqslant n^{-1}n_o^{5/4} \leqslant C(d)n^{-1}$ since $M_n^+ \leqslant n^{1/4}$ with n_o depending on d only. On the other hand, if $n \geqslant n_o$ then Lemma B.2 (iii) implies $n \geqslant 2\max_{1\leqslant m\leqslant M_n^+}\|[\Gamma]_m^{-1}\|_s$, and hence $\mho_{m,n} \subset \Omega_{m,n}$ for all $1 \leqslant m \leqslant M_n^+$ by using Lemma B.5. From (C.16) we conclude $M_n^+ \max_{1\leqslant m\leqslant M_n^+} P(\Omega_{m,n}^c) \leqslant M_n^+ \max_{1\leqslant m\leqslant M_n^+} P(\mho_{m,n}^c) \leqslant Cn^{-3}$. By combination of the two cases we obtain (C.17).
It remains to show (C.18). Consider the events $\mathcal{A}_n$, $\mathcal{B}_n$ and $\mathcal{C}_n$ defined in (A.1), where $\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \mathcal{E}_n$ due to Lemma B.4. Moreover, we have $n^7P(\mathcal{A}_n^c) \leqslant C$ and $n^7P(\mathcal{C}_n^c) \leqslant C$ due to (C.11) and (C.9) in Lemma C.2 (keep in mind that $\lfloor n^{1/4}\rfloor \geqslant M_n^\ell$ and $2(\sigma_Y^2 + [g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}}) = \sigma_k^2 \geqslant \rho_k^2$). Finally, (C.10) in Lemma C.2 implies $n^7P(\mathcal{B}_n^c) \leqslant C$ by using that $\{\|\sqrt{m}[\Xi]_{\underline{m}}\|_s \leqslant 1/8, 1 \leqslant m \leqslant M_n^+\} \subset \mathcal{B}_n$. Combining these estimates yields (C.18), which completes the proof. □

D Proof of Proposition 3.3 and 3.4

In the following proofs we will use the notations introduced in Appendix A and we will exploit the technical assertions gathered in Lemma C.1- C.4.

Proof of Proposition 3.3. From the identities $\widehat{\ell}_m - \ell(\phi_m) = [\ell]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[W]_{\underline{m}}\,\mathbb{1}_{\Omega_{m,n}} - \ell(\phi_m)\,\mathbb{1}_{\Omega^c_{m,n}}$, $([\mathrm{I}]_{\underline{m}} + [\Xi]_{\underline{m}})^{-1} - [\mathrm{I}]_{\underline{m}} = -([\mathrm{I}]_{\underline{m}} + [\Xi]_{\underline{m}})^{-1}[\Xi]_{\underline{m}}$, and $[\widehat{\Gamma}]_{\underline{m}} = [\Gamma]^{1/2}_{\underline{m}}\{[\mathrm{I}]_{\underline{m}} + [\Xi]_{\underline{m}}\}[\Gamma]^{1/2}_{\underline{m}}$ follows

$$
\begin{aligned}
|\widehat{\ell}_m - \ell(\phi_m)|^2 &= |[\ell]^t_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[W]_{\underline{m}}|^2\,\mathbb{1}_{\Omega_{m,n}} + |\ell(\phi_m)|^2\,\mathbb{1}_{\Omega^c_{m,n}}\\
&\leqslant 2|[\ell]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[W]_{\underline{m}}|^2 + 2|[\ell]^t_{\underline{m}}([\widehat{\Gamma}]^{-1}_{\underline{m}} - [\Gamma]^{-1}_{\underline{m}})[W]_{\underline{m}}|^2\,\mathbb{1}_{\Omega_{m,n}} + |\ell(\phi_m)|^2\,\mathbb{1}_{\Omega^c_{m,n}}\\
&\leqslant 2|[\ell]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[W]_{\underline{m}}|^2 + 2|[\ell]^t_{\underline{m}}[\Gamma]^{-1/2}_{\underline{m}}([\mathrm{I}]_{\underline{m}} + [\Xi]_{\underline{m}})^{-1}[\Xi]_{\underline{m}}[\Gamma]^{-1/2}_{\underline{m}}[W]_{\underline{m}}|^2\,\mathbb{1}_{\mho_{m,n}}\\
&\qquad + 2|[\ell]^t_{\underline{m}}[\Gamma]^{-1/2}_{\underline{m}}[\Xi]_{\underline{m}}[\Gamma]^{1/2}_{\underline{m}}[\widehat{\Gamma}]^{-1}_{\underline{m}}[W]_{\underline{m}}|^2\,\mathbb{1}_{\Omega_{m,n}}\,\mathbb{1}_{\mho^c_{m,n}} + |\ell(\phi_m)|^2\,\mathbb{1}_{\Omega^c_{m,n}}.
\end{aligned}
$$

By exploiting $\sqrt{m}\|([\mathrm{I}]_{\underline{m}} + [\Xi]_{\underline{m}})^{-1}[\Xi]_{\underline{m}}\|_s\,\mathbb{1}_{\mho_{m,n}} \leqslant 1/7$ and $\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s\,\mathbb{1}_{\Omega_{m,n}} \leqslant n$ we obtain

$$
\begin{aligned}
|\widehat{\ell}_m - \ell(\phi_m)|^2 &\leqslant 2|[\ell]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[W]_{\underline{m}}|^2 + \frac{2}{49}([\ell]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[\ell]_{\underline{m}})m^{-1}([W]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[W]_{\underline{m}})\\
&\qquad + 2n^2\,([\ell]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[\ell]_{\underline{m}})\,\|[\Xi]_{\underline{m}}[\Gamma]^{1/2}_{\underline{m}}\|^2_s\,\|[W]_{\underline{m}}\|^2\,\mathbb{1}_{\mho^c_{m,n}} + |\ell(\phi_m)|^2\,\mathbb{1}_{\Omega^c_{m,n}}.
\end{aligned}
$$

Taking this upper bound into account together with $([\ell]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[\ell]_{\underline{m}}) \leqslant V_m$, we obtain for all $\phi \in \mathcal{F}^r_\beta$ and $\Gamma \in \mathcal{G}^d_\gamma$ that

$$
\begin{aligned}
\mathbb{E}\Big\{\sup_{1\leqslant m\leqslant M_n^+}\Big(|\widehat{\ell}_m - \ell(\phi_m)|^2 - \frac{1}{6}\,\mathrm{p}_m\Big)_+\Big\} &\leqslant 2\sum_{m=1}^{M_n^+}\mathbb{E}\Big(|[\ell]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[W]_{\underline{m}}|^2 - \frac{8}{100}\,\mathrm{p}_m\Big)_+\\
&\quad + \frac{2}{49}\sum_{m=1}^{M_n^+}\mathbb{E}\Big(([\ell]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[\ell]^t_{\underline{m}})m^{-1}([W]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[W]_{\underline{m}}) - \frac{8}{100}\,\mathrm{p}_m\Big)_+\\
+ 2n^3\sum_{m=1}^{M_n^+}\frac{V_m}{n}\big(\mathbb{E}\|[\Xi]_{\underline{m}}[\Gamma]^{1/2}_{\underline{m}}\|^8_s\big)^{1/4}&\big(\mathbb{E}\|[W]_{\underline{m}}\|^8\big)^{1/4}\big(P(\mho^c_{m,n})\big)^{1/2} + \sum_{m=1}^{M_n^+}|\ell(\phi_m)|^2 P(\Omega^c_{m,n}).
\end{aligned}
$$

We bound the first and second right hand side term with help of (C.14) and (C.15) in Lemma C.3, which leads to

$$
\begin{aligned}
\sup_{\phi\in\mathcal{F}^r_\beta}\sup_{\Gamma\in\mathcal{G}^d_\gamma}\mathbb{E}\Big\{\sup_{1\leqslant m\leqslant M_n^+}\Big(|\widehat{\ell}_m - \ell(\phi_m)|^2 - \frac{1}{6}\,\mathrm{p}_m\Big)_+\Big\} &\leqslant C(d)(\sigma^2 + r)n^{-1}\\
&\quad + 2n^3\sup_{\phi\in\mathcal{F}^r_\beta}\sup_{\Gamma\in\mathcal{G}^d_\gamma}\sum_{m=1}^{M_n^+}\frac{V_m}{n}\big(\mathbb{E}\|[\Xi]_{\underline{m}}[\Gamma]^{1/2}_{\underline{m}}\|^8_s\big)^{1/4}\big(\mathbb{E}\|[W]_{\underline{m}}\|^8\big)^{1/4}\big(P(\mho^c_{m,n})\big)^{1/2}\\
&\quad + \sup_{\phi\in\mathcal{F}^r_\beta}\sup_{\Gamma\in\mathcal{G}^d_\gamma}\sum_{m=1}^{M_n^+}|\ell(\phi_m)|^2 P(\Omega^c_{m,n}).
\end{aligned}
$$

Taking into account that for all $\phi \in \mathcal{F}^r_\beta$ and $\Gamma \in \mathcal{G}^d_\gamma$ we have $\max_{1\leqslant m\leqslant M_n^+} V_m = V_{M_n^+} \leqslant nC(d)(1+\log n)^{-1}$ and $\rho^2_m \leqslant \sigma^2_m \leqslant C(d)(\sigma^2 + r)$ (Lemma B.2 (ii) and (iv)) the estimates (C.7) and (C.8) in Lemma C.2 imply

$$\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}\Big\{\sup_{1\leqslant m\leqslant M_n^+}\Big(|\widehat{\ell}_m-\ell(\phi_m)|^2-\frac{1}{6}\,\mathrm{p}_m\Big)_+\Big\}\leqslant\frac{C(d)}{n}(\sigma^2+r)$$
$$+\frac{C(d)}{n}(\sigma^2+r)\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}(\mathbb{E}\|X\|_{\mathbb{H}}^2)^2n^2(M_n^+)^2\max_{1\leqslant m\leqslant M_n^+}\big(P(\mho_{m,n}^c)\big)^{1/2}$$
$$+\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{m=1}^{M_n^+}|\ell(\phi_m)|^2P(\Omega_{m,n}^c).$$

By combining this upper bound, the property $\mathbb{E}\|X\|_{\mathbb{H}}^2\leqslant d\sum_{j\geqslant1}\gamma_j$ and the estimate (B.5) given in Lemma B.1 we obtain

$$\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}\Big\{\sup_{1\leqslant m\leqslant M_n^+}\Big(|\widehat{\ell}_m-\ell(\phi_m)|^2-\frac{1}{6}\,\mathrm{p}_m\Big)_+\Big\}\leqslant\frac{C(d)}{n}(\sigma^2+r)$$
$$+\frac{C(d)}{n}(\sigma^2+r)(\sum_{j\geqslant1}\gamma_j)^2\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}n^2(M_n^+)^2\max_{1\leqslant m\leqslant M_n^+}\big(P(\mho_{m,n}^c)\big)^{1/2}$$
$$+\frac{C(d)}{n}r\sum_{j\geqslant1}\frac{[\ell]_j^2}{\beta_j}\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}nM_n^+\max_{1\leqslant m\leqslant M_n^+}P(\Omega_{m,n}^c).$$

The result of the proposition follows now from the upper bounds (C.16) and (C.17) given in Lemma C.4, which completes the proof. □

Proof of Proposition 3.4. Taking the estimate $\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s\,\mathbb{1}_{\Omega_{m,n}}\leqslant n$ and the identity $\widehat{\ell}_m-\ell(\phi_m)\,\mathbb{1}_{\Omega_{m,n}}=[\ell]_{\underline{m}}^t[\widehat{\Gamma}]_{\underline{m}}^{-1}[W]_{\underline{m}}\,\mathbb{1}_{\Omega_{m,n}}$ into account it easily follows for all $m\geqslant1$ that

$$|\widehat{\ell}_m-\ell(\phi)|^2\leqslant3\{\|[\ell]_{\underline{m}}\|^2\,n^2\|[W]_{\underline{m}}\|^2+(|\ell(\phi_m)|^2+|\ell(\phi)|^2)\}.$$

Furthermore, by exploiting $\|[\ell]_{\underline{m}}\|^2\leqslant n$ for all $1\leqslant m\leqslant M_n^\ell$ we obtain from the last estimate

$$\max_{1\leqslant m\leqslant M_n^\ell}|\widehat{\ell}_m-\ell(\phi)|^2\,\mathbb{1}_{\mathcal{E}_n^c}\leqslant3\{n^3\sum_{m=1}^{M_n^\ell}\|[W]_{\underline{m}}\|^2\,\mathbb{1}_{\mathcal{E}_n^c}+(\sup_{m\geqslant1}|\ell(\phi_m)|^2+|\ell(\phi)|^2)\,\mathbb{1}_{\mathcal{E}_n^c}\}.$$

We recall that for all $\phi\in\mathcal{F}_\beta^r$ and $\Gamma\in\mathcal{G}_\gamma^d$ we have $\rho_m^2\leqslant C(d)(\sigma^2+r)$ and $(\mathbb{E}\|[W]_{\underline{m}}\|^4)^{1/2}\leqslant 11\mathbb{E}\|X\|_{\mathbb{H}}^2\rho_m^2n^{-1}$ (Lemma B.2 and C.2), moreover, the bounds $\big(\sup_{m\geqslant1}|\ell(\phi_m)|^2+|\ell(\phi)|^2\big)\leqslant(\sup_{m\geqslant1}\|\phi_m\|_\beta^2+\|\phi\|_\beta^2)\sum_{j\geqslant1}\frac{[\ell]_j^2}{\beta_j}\leqslant C(d)r\sum_{j\geqslant1}\frac{[\ell]_j^2}{\beta_j}$ (Lemma B.1) and $\mathbb{E}\|X\|_{\mathbb{H}}^2\leqslant d\sum_{j\geqslant1}\gamma_j$ together with the last upper bound imply

$$\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}\big(|\widehat{\ell}_{\widehat{m}}-\ell(\phi)|^2\,\mathbb{1}_{\mathcal{E}_n^c}\big)\leqslant\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}\big(\max_{1\leqslant m\leqslant M_n^\ell}|\widehat{\ell}_m-\ell(\phi)|^2\,\mathbb{1}_{\mathcal{E}_n^c}\big)$$
$$\leqslant C(d)\,(\sigma^2+r)\max\Big\{\sum_{j\geqslant1}\gamma_j,\sum_{j\geqslant1}\frac{[\ell]_j^2}{\beta_j}\Big\}\sup_{\phi\in\mathcal{F}_\beta^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\Big(n^2M_n^\ell|P(\mathcal{E}_n^c)|^{1/2}+P(\mathcal{E}_n^c)\Big).$$

The assertion of Proposition 3.4 follows now by combination of the last estimate and (C.18) in Lemma C.4, which completes the proof. □

References

A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113(3):301–413, 1999.

D. Bosq. *Linear Processes in Function Spaces.*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, 2000.

L. D. Brown and M. G. Low. A constrained risk inequality with applications to nonparametric functional estimation. *The Annals of Statistics*, 24(6):2524–2535, 1996.

H. Cardot and J. Johannes. Thresholding projection estimators in functional linear models. *Journal of Multivariate Analysis*, 101(2):395–408, 2010.

H. Cardot, F. Ferraty, and P. Sarda. Spline estimators for the functional linear model. *Statistica Sinica*, 13:571–591, 2003.

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

C. Crambes, A. Kneip, and P. Sarda. Smoothing splines estimators for functional linear regression. *The Annals of Statistics*, 37(1):35–72, 2009.

R. Dahlhaus and W. Polonik. Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *The Annals of Statistics*, 34(6):2790–2824, 2006.

D. Donoho. Statistical estimation and optimal recovery. *The Annals of Statistics*, 22:238–270, 1994.

D. Donoho and M. Low. Renormalization exponents and optimal pointwise rates of convergence. *The Annals of Statistics*, 20:944–970, 1992.

S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001.

P. H. Eilers and B. D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11:89–102, 1996.

A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39:1608–1632, 2011.

A. Goldenshluger and S. V. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probability Theory and Related Fields*, 118: 169–186, 2000.

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

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

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

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

J. Johannes and R. Schenk. On rate optimal local estimation in functional linear model. http://arxiv.org/abs/0902.0645v2, Université catholique de Louvain, 2010.

T. Kawata. *Fourier analysis in probability theory.* Academic Press, New York, 1972.

B. Laurent, C. Ludeña, and C. Prieur. Adaptive estimation of linear functionals by model selection. *Electronic Journal of Statistics*, 2:993–1020, 2008.

O. V. Lepski. On a problem of adaptive estimation in gaussian white noise. *Theory of Probability and its Applications*, 35:454–466, 1990.

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

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

P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

P. Mathé. The Lepskiĭ principle revisited. *Inverse Problems*, 22(3):11–15, 2006.

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

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

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

A. Neubauer. 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, 1988b.

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

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