



HAL
open science

Non-asymptotic adaptive prediction in functional linear models

Elodie Brunel, André Mas, Angelina Roche

► **To cite this version:**

Elodie Brunel, André Mas, Angelina Roche. Non-asymptotic adaptive prediction in functional linear models. *Journal of Multivariate Analysis*, 2016, 143 (1), pp.208 - 232. 10.1016/j.jmva.2015.09.008 . hal-01817385v1

HAL Id: hal-01817385

<https://hal.science/hal-01817385v1>

Submitted on 19 Jun 2018 (v1), last revised 14 Jan 2013 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Non-asymptotic Adaptive Prediction in Functional Linear Models

Élodie Brunel^a, André Mas^a, Angelina Roche^{a,*}

^a*I3M, Université Montpellier II, Place Eugne Bataillon, CC051, 34000 Montpellier.*

Abstract

Functional linear regression has recently attracted considerable interest. Many works focus on asymptotic inference. In this paper we consider in a non asymptotic framework a simple estimation procedure based on functional Principal Regression. It revolves in the minimization of a least square contrast coupled with a classical projection on the space spanned by the m first empirical eigenvectors of the covariance operator of the functional sample. The novelty of our approach is to select automatically the crucial dimension m by minimization of a penalized least square contrast. Our method is based on model selection tools. Yet, since this kind of methods consists usually in projecting onto known non-random spaces, we need to adapt it to empirical eigenbasis made of data-dependent hence random vectors. The resulting estimator is fully adaptive and is shown to verify an oracle inequality for the risk associated to the prediction error and to attain optimal minimax rates of convergence over a certain class of ellipsoids. Our strategy of model selection is finally compared numerically with cross-validation.

Keywords: Functional linear regression, functional principal component analysis, mean squared prediction error, minimax rate, penalized contrast estimator, model selection on random bases.

2000 MSC: 62J05, 62G08

*Corresponding author

Email addresses: ebrunel@math.univ-montp2.fr (Élodie Brunel),
andre.mas@univ-montp2.fr (André Mas), angelina.roche@univ-montp2.fr (Angelina Roche)

Introduction

Functional data analysis has known recent advances in the past two decades, addressing simultaneously many fields of applications. We refer to Ferraty and Vieu [24] and Ramsay and Silverman [37] for detailed examples in medicine, linguistics and chemometrics and to Preda and Saporta [35] for applications in econometrics.

In this paper we suppose that the dependence between a real-valued response Y and a functional predictor X belonging to a Hilbert space $(\mathbf{H}, \langle \cdot, \cdot \rangle, \|\cdot\|)$ is given by the functional linear model, namely

$$Y = \langle \beta, X \rangle + \varepsilon, \quad (1)$$

where ε stands for a noise term with variance σ^2 and is independent of X and $\beta \in \mathbf{H}$ is an unknown function to be estimated. In order to simplify the notations, the random variable X is supposed to be centred as well, which means that the function $t \mapsto \mathbf{E}[X(t)]$ is identically equal to zero.

By multiplying both sides of Equation (1) by $X(s)$ and taking the expectation, we see easily that the function β is solution of

$$\Gamma\beta := \mathbf{E}[\langle \beta, X \rangle X(\cdot)] = \mathbf{E}[YX] =: g, \quad (2)$$

where Γ is the covariance operator associated to the functional predictor X . Equation (2) is known to be an ill-posed inverse problem (see Engl *et al.* [23, Chapter 2.1]).

The literature on the functional linear model is wide and numerous estimation procedures exist. A first method consists in minimizing a least square criterion subject to a roughness penalty. For instance, Li and Hsing [31] proposed an estimation procedure by minimization of such a criterion on periodic Sobolev spaces, Crambes *et al.* [19] generalized the well-known smoothing-spline estimator used in univariate nonparametric regression. Another approach is based on dimension reduction: this consists in approximating the regression function β by projection onto finite-dimensional spaces. Those spaces are usually obtained by taking the first components of a basis of \mathbf{H} . Some authors considered projection onto fixed basis, such as B-spline basis (Ramsay and Dalzell [36]) or general orthonormal basis (Cardot and Johannes [15]). But the most popular method is Functional Principal Component Regression (FPCR), this consists in taking the random space spanned

by the eigenfunctions associated to the largest eigenvalues of the empirical covariance operator:

$$\Gamma_n : f \in \mathbf{H} \mapsto \frac{1}{n} \sum_{i=1}^n \langle f, X_i \rangle X_i. \quad (3)$$

The resulting estimator is shown to be consistent, but its behaviour is often erratic in simulation studies, thus a smooth version by using splines has been proposed by Cardot *et al.* [14]. The FPCR estimator is shown to attain optimal rates of convergence for the risk associated to the prediction error over fixed curves x (see Cai and Hall [11]) as well as for the \mathbf{L}^2 -risk (see Hall and Horowitz [26]).

All the proposed estimators rely on the choice of at least one tuning parameter (the smoothing parameter appearing in the penalized criterion or the dimension of approximation space) which influences significantly the quality of estimation. Optimal choice of such parameters depends generally on both unknown regularities of the slope function β and the predictor X (see e.g. [11, 19, 15]) and the parameters are usually chosen in practice by cross-validation.

Until the recent work of Comte and Johannes [17], nonasymptotic results providing adaptive data-driven estimators were missing. Comte and Johannes [17, 18] propose model selection procedures for the orthogonal series estimator introduced first by Cardot and Johannes [15]. In [17], they propose to select the dimension by minimization of a penalized contrast criterion under strong assumption of periodicity of the curve X while in [18] they define a dimension selection criterion by means of a stochastic penalized contrast emulating Lepski's method (see Goldenshluger and Lepski [25]) and do not require specific assumptions on the curve X . The resulting estimators are completely data-driven and achieve optimal minimax rates for general weighted \mathbf{L}^2 -risks. However, since both dimension selection criteria depend on weights defining the risk, these selection procedures do not address prediction error, which can be written as a weighted norm whose weights are the unknown eigenvalues of the covariance operator.

In the same context as Comte and Johannes [17], Brunel and Roche [10] propose to estimate the slope function by minimizing a least square contrast on spaces spanned by the trigonometric basis. The dimension is selected by means of a penalized contrast. Their estimator is proved to attain the optimal minimax rate of convergence for the risk associated to the prediction error.

Another approach is proposed by Cai and Yuan [12] carrying out reproducing kernel Hilbert spaces. They develop a data-driven choice of the tuning parameter of the roughness regularization method (see e.g. Ramsay and Silverman [37]). Their estimation procedure is shown to attain the optimal rate of convergence without the need of knowing the covariance kernel. Lee and Park [30] also suggest general variable selection procedures based on a weighted L_1 penalty under assumption of sparsity on the functional parameter β . Their estimator is shown to be consistent and to satisfy the oracle-property.

In this paper, we propose an entirely data-driven procedure to select the adequate dimension for the classical FPCR estimator. The method proposed is based on model selection tools developed in a general context by Barron *et al.* [4], outlined by Massart [34], and in a context of regression by Baraud [2, 3]. However, these tools are not meant to deal with estimators defined on random approximation spaces and thus have to be adapted. Section 1 is devoted to the description of estimation procedure. The resulting estimator is proved to satisfy an oracle-type inequality and to attain the optimal minimax rate of convergence for the risk associated to the prediction error for slope functions belonging to Sobolev classes in Section 2. In Section 3, a simulation study is presented including a comparison with cross-validation. The proofs are detailed in Section A and in the Appendix.

1. Definition of the estimator

We assume that we are given an i.i.d. sample $(Y_i, X_i)_{i \geq 1}$ where the generic Y is real and X belongs to the Hilbert space \mathbf{H} . Thereafter, the Hilbert space is set to be $\mathbf{H} = \mathbf{L}^2([0, 1])$ equipped with its usual inner product $\langle \cdot, \cdot \rangle$ defined by $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$ but our method adapts to more general Sobolev spaces as well. We assumed above that X is a centred random curve.

We recall that the theoretical covariance operator Γ of X defined by Equation (2) in the introductory section is a selfadjoint trace class operator defined on and with values in $\mathbf{L}^2([0, 1])$. This means that the sequence of its eigenvalues denoted $(\lambda_j)_{j \geq 1}$ is positive and summable. The associated sequence of eigenfunctions is denoted by $(\psi_j)_{j \geq 1}$.

1.1. Collection of models

If the $(\psi_j)_{j \leq 1}$ were known an obvious choice would be to consider a model collection $(S_m)_m$ based on these eigenfunctions. Unfortunately this is not

possible. As the empirical covariance operator Γ_n defined by Equation (3) is selfadjoint too, there exists an orthonormal basis $(\hat{\psi}_j)_{j \geq 1}$ of $\mathbf{L}^2([0, 1])$ composed of eigenfunctions of Γ_n ; we denote by $(\hat{\lambda}_j)_{j \geq 1}$ the associated eigenvalues arranged in decreasing order. Since Γ_n is finished-rank, the $(\hat{\lambda}_j)_{j \geq 1}$ are necessarily null at least for $j > n$. The couples $(\hat{\lambda}_j, \hat{\psi}_j)_{j \geq 1}$ are the empirical counterparts of the $(\lambda_j, \psi_j)_{j \geq 1}$. Dimension reduction based on functional Principal Component Analysis usually comes down to projecting the data on the space spanned by the $(\hat{\psi}_j)_{j \leq K}$ for some K . Our aim here is to shift to a model selection approach.

Let \hat{N}_n be a random integer which will be defined later. For all $m \in \widehat{\mathcal{M}}_n := \{1, \dots, \hat{N}_n\}$, we define

$$\hat{S}_m := \text{span}\{\hat{\psi}_1, \dots, \hat{\psi}_m\}$$

which is an empirical counterpart of $S_m := \text{span}\{\psi_1, \dots, \psi_m\}$. It is important to note that a major difference appears here. Classical model selection is carried out with fixed and known families of model. Here we handle random bases and this is the source of additional problems related to the convergence of (possibly random) projectors associated to these finite-dimensional spaces. Other difficulties come from the non-linear dependence between the coefficients of our estimator in the basis $(\hat{\psi}_1, \dots, \hat{\psi}_m)$ and the basis itself.

1.2. Estimation on \hat{S}_m

Introduce the following simple least square contrast:

$$\gamma_n(t) := \frac{1}{n} \sum_{i=1}^n (Y_i - \langle t, X_i \rangle)^2. \quad (4)$$

Define $\hat{g} := (1/n) \sum_{i=1}^n Y_i X_i$ the cross-covariance between Y and X (which is also the theoretical counterpart of the function g appearing in Equation (2)) and

$$\hat{\beta}_m := \sum_{j=1}^m \frac{\langle \hat{g}, \hat{\psi}_j \rangle}{\hat{\lambda}_j} \hat{\psi}_j. \quad (5)$$

We can see easily that (5) is the unique minimizer of the least square contrast γ_n if $\hat{\lambda}_m > 0$.

From (1) the noise variance is as a parameter that we cannot dismiss since it appears in many computations. We distinguish two cases below.

1.3. *Model selection with known noise variance*

We suppose as a first step that the noise variance σ^2 is known.

We set

$$\hat{N}_n := \max \left\{ N \in \mathbf{N} \setminus \{0\}, N \leq 20\sqrt{n/\ln^3(n)} \text{ and } \hat{\lambda}_N \geq \mathfrak{s}_n \right\},$$

where $\mathfrak{s}_n := \frac{2}{n^2} \left(1 - \frac{1}{\ln^2 n}\right)$; and its theoretical counterpart

$$N_n := \max \left\{ N \in \mathbf{N} \setminus \{0\}, N \leq 20\sqrt{n/\ln^3(n)} \text{ and } \lambda_N \geq n^{-2} \right\}.$$

The introduction of an empirical maximal dimension is motivated by the need to ensure that the terms $\hat{\lambda}_j$ appearing in the definition of our estimator are not too small.

We select the dimension $\hat{m}^{(kv)} \in \widehat{\mathcal{M}}_n$ by minimizing the criterion

$$\text{crit}(m) = \gamma_n(\hat{\beta}_m) + \text{pen}^{(kv)}(m) \quad (6)$$

with

$$\text{pen}^{(kv)}(m) := (1 + \theta) \frac{\sigma^2}{n} m, \quad (7)$$

where θ is a positive constant. Then, we propose the following estimator of the function β

$$\tilde{\beta}^{(kv)} := \hat{\beta}_{\hat{m}^{(kv)}}.$$

1.4. *Model selection with unknown noise variance*

Let $\theta > 4$ and $\delta > 0$, we set

$$\hat{N}_n := \max \left\{ N \in \mathbf{N} \setminus \{0\}, N \leq \min \left\{ 20\sqrt{\frac{n}{\ln^3(n)}}, \frac{n}{\theta(1+2\delta)} \right\} \text{ and } \hat{\lambda}_N \geq \mathfrak{s}_n \right\},$$

and

$$N_n := \max \left\{ N \in \mathbf{N} \setminus \{0\}, N \leq \min \left\{ 20\sqrt{\frac{n}{\ln^3(n)}}, \frac{n}{\theta(1+2\delta)} \right\} \text{ and } \lambda_N \geq n^{-2} \right\},$$

the term σ^2 appearing in Equation (7) is replaced by the following estimator

$$\hat{\sigma}_m^2 := \frac{1}{n} \sum_{i=1}^n (Y_i - \langle \hat{\beta}_m, X_i \rangle)^2 = \gamma_n(\hat{\beta}_m).$$

The penalty becomes:

$$\widehat{\text{pen}}(m) := \theta(1 + \delta)\sigma_m^2 \frac{m}{n},$$

and the selection criterion:

$$\begin{aligned} \hat{m}^{(uv)} &\in \arg \min_{m \in \widehat{\mathcal{M}}_n} (\gamma_n(\hat{\beta}_m) + \widehat{\text{pen}}(m)) \\ &= \arg \min_{m \in \widehat{\mathcal{M}}_n} \gamma_n(\hat{\beta}_m) \left(1 + \theta(1 + \delta)\frac{m}{n}\right). \end{aligned} \quad (8)$$

We also denote

$$\text{pen}^{(uv)}(m) := \theta(1 + \delta)\sigma^2 \frac{m}{n},$$

the theoretical counterpart of $\widehat{\text{pen}}(m)$.

Finally, we define the following estimator:

$$\widetilde{\beta}^{(uv)} := \hat{\beta}_{\hat{m}^{(uv)}}.$$

In the sequel, when a property applies to both $\widetilde{\beta}^{(kv)}$ and $\widetilde{\beta}^{(uv)}$ we denote simply these estimators by $\widetilde{\beta}$, in that case we will denote also, $\hat{m}^{(kv)}$ and $\hat{m}^{(uv)}$ by \hat{m} and $\text{pen}^{(kv)}(m)$ and $\text{pen}^{(uv)}(m)$ by $\text{pen}(m)$.

2. Main results

In this section we derive oracle-type inequalities and uniform bounds for the risk associated to the prediction error. The prediction error of an estimator $\hat{\beta}$ (see e.g. [19, 15]) is defined by

$$\mathbf{E} \left[\left(\hat{Y}_{n+1} - \mathbf{E}[Y_{n+1} | X_{n+1}] \right)^2 | X_1, \dots, X_n \right] = \|\Gamma^{1/2}(\hat{\beta} - \beta)\|^2 =: \|\hat{\beta} - \beta\|_{\Gamma}^2, \quad (9)$$

where $\hat{Y}_{n+1} := \int_0^1 \hat{\beta}(t) X_{n+1}(t) dt$. We suppose that $\lambda_j > 0$ for all $j \geq 1$, which implies that the quantity (9) defines a norm on $\mathbf{L}^2([0, 1])$ denoted by $\|\cdot\|_{\Gamma}^2$. This condition is necessary for the model to be identifiable. Indeed, if there exists $j_0 \geq 1$ such that $\lambda_{j_0} = 0$, we have:

$$0 = \lambda_{j_0} \|\psi_{j_0}\|^2 = \langle \Gamma \psi_{j_0}, \psi_{j_0} \rangle = \mathbf{E} [\langle X, \psi_{j_0} \rangle^2],$$

and $\langle X, \psi_{j_0} \rangle = 0$ almost surely. By consequence, if the slope function β satisfies Equation (1), then any slope function of the form $\beta + c\psi_{j_0}$, with $c \in \mathbf{R}$, satisfies also Equation (1): it is clearly impossible to identify the slope function with our sample in that case. However this condition is not sufficient, for more details on the problem of identifiability in functional linear models see Section 2 of Cardot *et al.* [14].

2.1. Assumptions

Recall that $(\lambda_j, \psi_j)_{j \geq 2}$ denote the eigenelements of the covariance operator Γ . We can control the risk under four assumptions:

H1 There exists $p > 4$ such that $\tau_p := \mathbf{E}[|\varepsilon|^p] < +\infty$.

H2 There exists $b > 0$ such that, for all $l \in \mathbf{N} \setminus \{0\}$,

$$\sup_{j \in \mathbf{N}} \mathbf{E} \left[\frac{\langle X, \psi_j \rangle^{2l}}{\lambda_j^l} \right] \leq l! b^{l-1}.$$

H3 There exists a constant $\gamma > 0$ such that the sequence $(j \lambda_j \ln^{1+\gamma}(j))_{j \geq 2}$ is decreasing.

Assumption **H1** is standard in regression. Assumption **H2** is necessary to apply exponential inequalities. The assumption **H3** on the sequence $(j \lambda_j \ln^{1+\gamma}(j))_{j \geq 2}$ allows to avoid more restrictive hypotheses about spacing control between eigenvalues as usually made frequently in the literature (see [11], [27], [26]).

In order to derive oracle-inequalities for the risk associated to the prediction error, we need to precise the decreasing rate of the sequence $(\lambda_j)_{j \geq 1}$. Usually in functional linear regression, this rate is supposed to be polynomial (see for instance [11, 19, 12]) but more regular processes may be considered. That is the reason why, following Cardot and Johannes [15] or Comte and Johannes [17], we consider also exponential rates.

(P) Polynomial decrease There exists two constants $a > 1$ and $c_P \geq 1$ such that, for all $j \geq 1$

$$c_P^{-1} j^{-a} \leq \lambda_j \leq c_P j^{-a}.$$

(E) Exponential decrease There exists two constants $a > 0$ and $c_E \geq 1$ such that for all $j \geq 1$

$$c_E^{-1} \exp(-j^a) \leq \lambda_j \leq c_E \exp(-j^a).$$

2.2. Upper-bound on the empirical risk

We define an empirical semi-norm naturally associated to our estimation problem by

$$\|f\|_{\Gamma_n}^2 := \|\Gamma_n^{1/2} f\|^2 = \frac{1}{n} \sum_{i=1}^n \langle f, X_i \rangle^2, \text{ for all } f \in \mathbf{L}^2([0, 1]).$$

In a first step, in propositions 1 and 2, we prove that our estimators verify an oracle type inequality for the risk associated to this semi-norm whatever the regularity of the slope function β and the decreasing rate of the covariance operator eigenvalues are.

2.2.1. Bound on the empirical risk with known noise variance

Proposition 1. *Suppose that Assumption **H1** is fulfilled, we have*

$$\mathbf{E}[\|\tilde{\beta}^{(kv)} - \beta\|_{\Gamma_n}^2] \leq c_1 \inf_{m \in \mathcal{M}_n} \left\{ \mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2] + \text{pen}^{(kv)}(m) \right\} + \frac{c_2(\sigma^2 + \|\beta\|^2)}{n},$$

with $c_1, c_2 > 0$ depending only on θ and p and $\hat{\Pi}_m$ the orthonormal projector onto \hat{S}_m .

2.2.2. Bound on the empirical risk with unknown noise variance

Proposition 2. *Suppose that Assumption **H1** is fulfilled. We have*

$$\begin{aligned} \mathbf{E}[\|\tilde{\beta}^{(uv)} - \beta\|_{\Gamma_n}^2] &\leq c'_1 \inf_{m \in \mathcal{M}_n} \left\{ \mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2] + \text{pen}^{(uv)}(m) \right\} \\ &\quad + \frac{c'_2}{n} (\sigma^2 + \|\beta\|^2 + \tau_p^{2/p}), \end{aligned}$$

with $c'_1, c'_2 > 0$ depends only on θ , p and δ .

2.3. Oracle inequality

In this section, we derive an oracle-inequality for the risk associated to the prediction error. We define first an ellipsoid of $\mathbf{L}^2([0, 1])$

$$\mathcal{W}_r^R := \left\{ f \in \mathbf{L}^2([0, 1]), \sum_{j \geq 1} j^r \langle f, \psi_j \rangle^2 \leq R^2 \right\}.$$

Theorem 3. *Suppose that assumptions **H1**, **H2** and **H3** hold and that the decreasing rate of $(\lambda_j)_{j \geq 1}$ is given by **(P)** or **(E)**. Then, for all slope function $\beta \in \mathbf{L}^2([0, 1])$, if $n \geq 6$:*

$$\begin{aligned} \mathbf{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2] &\leq C_1 \left(\min_{m \in \mathcal{M}_n} \left(\mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma}^2] + \mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2] + \text{pen}(m) \right) \right) \\ &\quad + \frac{C_2}{n} (1 + \|\beta\|^2) \end{aligned} \quad (10)$$

where $C_1 > 0$ and $C_2 > 0$ are independent of β and n .

If, in addition, $\beta \in \mathcal{W}_r^R$ — with the condition $a + r/2 > 2$ in the polynomial case **(P)** — we have

$$\mathbf{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2] \leq C'_1 \left(\min_{m \in \mathcal{M}_n} \left(\mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma}^2] + \text{pen}(m) \right) \right) + \frac{C'_2}{n} (1 + \|\beta\|^2), \quad (11)$$

where the constants $C'_1 > 0$ and $C'_2 > 0$ do not depend on β or n .

Remark 1. The condition $a + r/2 > 2$ is verified as soon as $a \geq 2$ without condition on the regularity parameter r of the slope β . Note that if X is a Brownian motion, the sequence $(\lambda_j)_{j \geq 1}$ associated to the process X verifies **(P)** with $a = 2$. Then we do not need additional condition on r if X is smoother than the Brownian motion.

Remark 2. For fixed bases, such as B -spline or trigonometric ones, we could define a very similar dimension selection procedure. We could prove an oracle-inequality for the empirical prediction error as in Proposition 1. An oracle-inequality for the prediction error could also be obtained but with different assumptions, relating the approximation basis and the Karhunen-Loève basis of X . For instance, we can suppose, as in [29], that there exists a sequence $(\gamma_j)_{j \geq 1}$ of positive real numbers and a constant $C > 0$ such that, for all $f \in \mathbf{L}^2([0, 1])$, $\sum_{j \geq 1} \gamma_j^2 \langle f, \varphi_j \rangle^2 / C \leq \|Tf\|^2 \leq C^2 \sum_{j \geq 1} \gamma_j^2 \langle f, \varphi_j \rangle^2 / C$.

2.4. Convergence rates

As a direct consequence of the oracle-inequality given in Theorem 3, associated with the control of the random projector on the spaces \hat{S}_m given in Lemma 15, we derive uniform bounds on the risk of our estimators on the ellipsoids \mathcal{W}_r^R .

Theorem 4. Assume that the assumptions of Theorem 3 are fulfilled. For all $r > 0$ and $R > 0$:

Polynomial case. If (\mathbf{P}) holds with $a + r > 2$ then:

$$\sup_{\beta \in \mathcal{W}_r^R} \mathbf{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2] \leq C_P n^{-(a+r)/(a+r+1)}; \quad (12)$$

Exponential case. If (\mathbf{E}) holds then:

$$\sup_{\beta \in \mathcal{W}_r^R} \mathbf{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2] \leq C_E n^{-1} (\ln n)^{1/a}, \quad (13)$$

with C_P and C_E independent of n .

Remark 3. In the case where the noise ε is Gaussian, the bounds (12) and (13) coincide with the minimal bounds given by Cardot and Johannes [15].

3. Numerical results

3.1. Simulation method

Following the method proposed by Hall and Hosseini-Nasab [27], we simulate the random function X in the following way

$$X = \sum_{j=1}^J \sqrt{\lambda_j} \xi_j \psi_j, \quad (14)$$

where $J = 100$ and for all $j \geq 1$, $\psi_j(x) = \sqrt{2} \sin(\pi(j - 0.5)x)$ and $\{\xi_j, j = 1, \dots, J\}$ is independent and follows the standard normal distribution. This sequence of functions $(\psi_j)_{j \geq 1}$ has been chosen so that if J is sufficiently high and if $\lambda_j = (j - 0.5)^{-2} \pi^{-2}$, we obtain a Brownian motion (see Ash and Gardner [1]). In order to see how the decreasing rate of $(\lambda_j)_{j \geq 1}$ influences the estimation, we take three different sequences:

$$\lambda_j^{P1} = j^{-2}, \quad \lambda_j^{P2} = j^{-3} \quad \text{and} \quad \lambda_j^E = e^{-j}.$$

It is interesting to note from Equation (14) that the higher the rate of decrease of the λ_j 's, the better the regularity of the function X .

The function X is then discretized over $p = 100$ equispaced points $\left\{t_j = \frac{j-1}{p}, j = 1, \dots, p\right\}$. We take $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ and $\sigma^2 = 0.01$. We consider here three different slope functions

- $\beta_1(x) = \ln(15x^2 + 10) + \cos(4\pi x)$ (see [13]),
- $\beta_2(x) = e^{(x-0.3)^2/0.05} \cos(4\pi x)$,
- $\beta_3(x) = 2\psi_1(x) + 2^{-1/2}\psi_3(x) + \psi_6(x)$.

Considering the slope function β_3 allows us to compare the selected dimension with the oracle one, which is $m^* = 3$ here.

3.2. Calibration of the constant appearing in the penalty

We recall that the definition of the penalty term involves a constant $\kappa := 1 + \theta$ for $\tilde{\beta}^{(kv)}$ or $\kappa := \delta(1 + \theta)$ for $\tilde{\beta}^{(uv)}$.

This constant is universal in the sense that its value should not depend on the law of X , or the law of the noise ε , or β or n . The idea is then to plot $\mathbb{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2]$ as a function of κ , for different values of n ($n = 200, n = 500, n = 1000$), $(\lambda_j)_{j \geq 1}$ and β (β_1 and $\beta_4(t) := (t - 0.2)(t - 0.5)$). For each set of parameters and each value of κ , the risk $\mathbb{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2]$ is approximated by the mean of $N = 500$ Monte-Carlo replications of the random variable $\|\tilde{\beta} - \beta\|_{\Gamma}^2$.

From the results presented in Figure 1 we deduce that the optimal constant κ should lie between 2 and 4. The value $\kappa = 2$, which corresponds to the Mallows' C_p criterion [32], leads to an unstable criterion. We choose the value $\kappa = 4$, for which the criterion is quite stable.

Data-driven calibration methods of the constant κ [7, 5] exist and have been implemented in a R package [9] with good practical results. However, no theoretical proofs are available in our context on these methods up to now.

3.3. Comparison with cross validation

We compare our dimension selection criterion with two cross validation criteria frequently used in practice. The first method consists in minimizing

$$GCV(m) := \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)}{(1 - \text{tr}(\mathbf{H}_m)/n)^2},$$

where $\hat{Y}_i := \int_0^1 \hat{\beta}_m(t) X_i(t) dt$ and \mathbf{H}_m is the classical Hat matrix defined by $\hat{\mathbf{Y}} = (\hat{Y}_1, \dots, \hat{Y}_n)' = \mathbf{H}_m \mathbf{Y}$. This criterion has been proposed in a similar

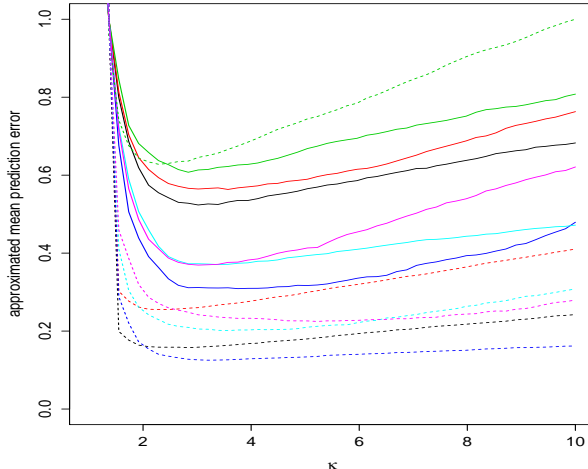


Figure 1: Calibration of κ . Each figure represents the approximation of $\mathbb{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2]$ depending on κ for different combination of n , $(\lambda_j)_{j \geq 1}$ and β . Legend: for $\lambda_j = j^{-2}$: black curves: $n = 200$, red curves: $n = 500$, green curves: $n = 1000$; for $\lambda_j = e^{-j}$: dark blue curves: $n = 200$, light blue curves: $n = 500$, pink curves: $n = 1000$. For β_1 , the curves are solid, for β_4 , the curves are dotted. The values of each curve are first divided by their maximum, this does not change the behaviour of the curve but make the figures more visible.

context by Marx and Eilers [22] and in the context of functional linear models by Cardot *et al.* [14]. The second one consists in minimizing the criterion

$$CV(m) := \frac{1}{n} \sum_{i=1}^n \left(Y_i - \hat{Y}_i^{(-i)} \right)^2,$$

which has been proposed in the framework of functional linear model by Hall and Hosseini-Nasab [27]. Here $\hat{Y}_i^{(-i)}$ is the value of Y_i predicted from the sample $\{(X_j, Y_j), j \neq i\}$. Note that an immediate drawback of this criterion is that it requires a much longer CPU time than the GCV criterion or our penalized criterion.

3.4. Results

As we can see in figures 2, 3, the regularity of estimators increases when the rate of convergence of the λ_j 's decreases. This is a specificity of functional PCA: the estimated slope function $\tilde{\beta}$ is an element of $\text{Im}(\Gamma_n) =$

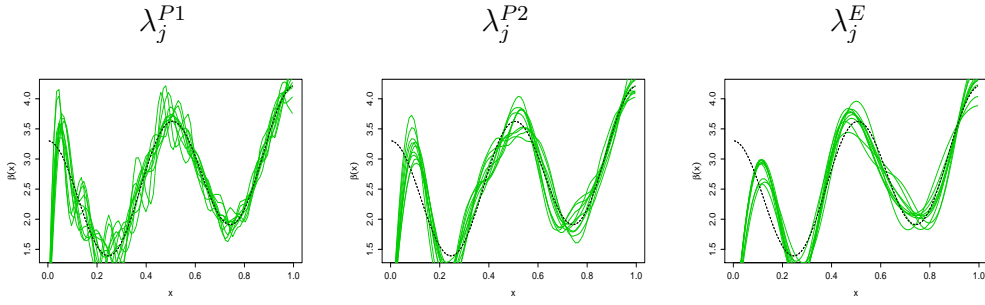


Figure 2: Plot of β_1 (bold, dashed) and $\tilde{\beta}_1^{(kv)}$ computed for 10 independent samples of size $n = 1000$.

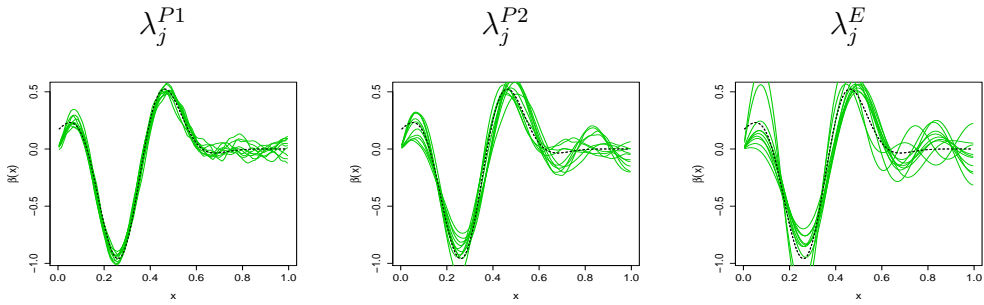


Figure 3: Plot of β_2 (bold, dashed) and $\tilde{\beta}_2^{(kv)}$ computed for 10 independent samples of size $n = 1000$.

λ_j		β_1			β_2		
		$n = 200$	$n = 500$	$n = 1000$	$n = 200$	$n = 500$	$n = 1000$
j^{-2}	kv	12.8 ±0.4	5.9 ±0.2	3.57 ±0.08	4.7 ±0.3	1.88 ±0.09	0.89 ±0.05
	uv	12.5 ±0.4	5.8 ±0.2	3.51 ±0.08	4.7 ±0.3	1.89 ±0.09	0.89 ±0.05
	GCV	80 ±2	55 ±2	47 ±2	80 ±2	55 ±2	47 ±2
	CV	12.2 ±0.5	5.6 ±0.2	3.34 ±0.09	5.7 ±0.4	2.2 ±0.2	1.08 ±0.06
j^{-3}	kv	6.7 ±0.3	3.3 ±0.1	1.84 ±0.06	5.5 ±0.2	1.8 ±0.1	0.88 ±0.04
	uv	6.6 ±0.3	3.2 ±0.1	1.83 ±0.06	5.4 ±0.3	1.8 ±0.1	0.88 ±0.04
	GCV	18.4 ±0.5	12.6 ±0.3	9.3 ±0.2	18.5 ±0.5	12.7 ±0.3	9.5 ±0.2
	CV	7.3 ±0.4	3.3 ±0.2	1.78 ±0.07	5.1 ±0.4	2.0 ±0.2	1.05 ±0.08
e^{-j}	kv	4.8 ±0.2	2.05 ±0.08	1.12 ±0.05	5.0 ±0.2	1.9 ±0.1	0.78 ±0.05
	uv	4.7 ±0.2	2.03 ±0.08	1.11 ±0.05	4.9 ±0.2	1.82 ±0.09	0.77 ±0.05
	GCV	5.8 ±0.3	2.67 ±0.09	1.45 ±0.05	6.0 ±0.3	2.7 ±0.1	1.40 ±0.05
	CV	4.8 ±0.3	2.05 ±0.09	1.10 ±0.05	4.6 ±0.3	1.8 ±0.1	0.87 ±0.05

Table 1: Mean prediction error ($\times 10^{-4}$) and approximated 95% confidence interval (calculated from 500 independent samples of size $n = 1000$). kv: dimension selected by minimization of (6), uv: dimension selected by (8).

λ_j		$n = 200$	$n = 500$	$n = 1000$
j^{-3}	$\mathbb{E}[\widehat{m}^{(kv)}]$	6.30±0.09	6.30±0.05	6.31±0.04
	$\mathbb{P}(\widehat{m}^{(kv)} = J)$ (%)	77.0±5.9	75.2±3.8	75.2±2.7
e^{-j}	$\mathbb{E}[\widehat{m}^{(kv)}]$	6.09±0.05	6.11±0.03	6.10±0.02
	$\mathbb{P}(\widehat{m}^{(kv)} = J)$ (%)	93.4±3.4	91.8±2.4	92.6±1.6

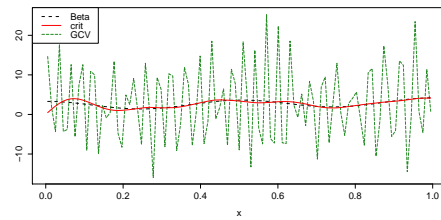
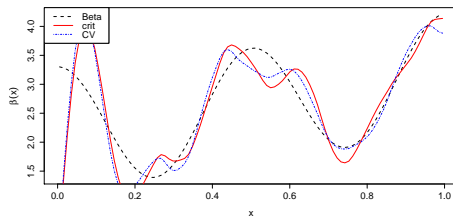
Table 2: Study of the dimension selection procedure for β_3 : mean selected dimension, percentage of selection of the exact dimension $J = 6$ with approximated 95% confidence interval.

$\text{span}\{X_1, \dots, X_n\}$ and thus has the same regularity as the function X . It also explains the side effect observed in figures 2 and 3 since $X_i(0) = 0$ implies that $\tilde{\beta}(0) = 0$.

The results of Table 1 indicate that the substitution of the term σ^2 by the estimator $\hat{\sigma}_m^2$ in case of unknown variance does not have a significant effect on the quality of estimation. Surprisingly, the mean prediction error seems to be lower when the noise variance is unknown. Nevertheless, according to the approximated confidence intervals, these differences are not meaningful. The Monte Carlo study also revealed that the dimension selected by minimization of (6) and (8) is the same in 70% to 99% of cases (percentage depending on the sample size n , the decreasing rate of the λ_j 's and the function β).

Moreover, according to Figure 4 and Table 1, performances of our estimators seem to be quite similar to the functional PCR estimator with dimension selected by minimization of the CV criterion. Conversely, the GCV criterion selects systematically the highest dimensional model which leads to poor

Estimation of β_1



Estimation of β_2

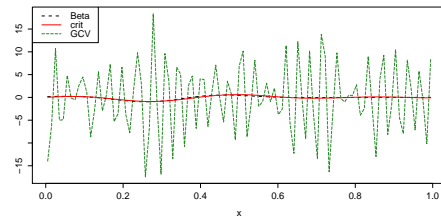
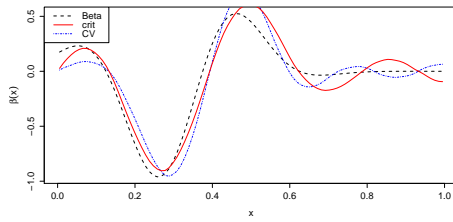


Figure 4: Left: comparison of estimators $\hat{\beta}_m$ when m is selected by minimization of the penalized criterion crit defined by (6) or the CV criterion. Right: comparison with the GCV criterion. $n = 2000$, $\lambda_j = j^{-3}$.

performances.

The comparison of the selected dimension for the estimation of β_3 and the real dimension $J = 6$ are given in Table 2. We give the mean of the selected dimension and the proportion of samples over 500 where the selected dimension matches the exact dimension $J = 6$. We see that our model selection procedure recovers the exact dimension $J = 6$ most of the time.

Acknowledgements

The authors would like to warmly thank Dr Nicolas Verzelen and both reviewers whose helpful comments and suggestions have improved significantly the presentation of this work.

Appendix A Proofs

A.1 Proof of Proposition 1

We want here to take advantage of Corollary 3.1 in Baraud [2] who provided a very similar result in the context of regression on a fixed design. Indeed Baraud considers a model $Y = s(x) + \epsilon$ where Y is real, x takes values in some measurable space and s is a general mapping. Conditioning our functional linear model with respect to $\mathbf{X} := \{X_1, \dots, X_n\}$ we can switch from the model considered in Baraud to ours by setting $s(x) = \langle \beta, x \rangle$. The seminorm $\|\cdot\|_n$ becomes our $\|\cdot\|_{\Gamma_n}$ and the least square contrast is now :

$$\tilde{\gamma}_n : \phi \mapsto \frac{1}{n} \sum_{i=1}^n (Y_i - \phi(X_i))^2.$$

Our last task consists in identifying the class of models, namely the S_m 's. Still sticking to Baraud's notation the latter should be subspaces of $\mathbf{L}^2(\mathbf{H}, \|\cdot\|_n)$. It is simple to see that the following collection suits:

$$\hat{S}_m := \text{span} \left\{ \hat{\psi}_j, j = 1, \dots, m \right\} \subset \mathbf{L}^2(\mathbf{H}, \|\cdot\|_n), m = 1, \dots, N_n,$$

through the identification mentioned above that is $\hat{\psi}_j(x) = \langle \hat{\psi}_j, x \rangle$. As Assumption **H1** is supposed to be verified, we are now ready to apply Corollary 3.1 of Baraud [2] with $q = 1$ and obtain that a.s.

$$\mathbf{E}_{\mathbf{X}} \left[\|\beta - \hat{\beta}_m\|_n^2 \right] \leq C(\theta) \inf_{m \in \widehat{\mathcal{M}}_n} \left(\|\beta - \Pi_{\hat{S}_m} \beta\|_n^2 + \text{pen}(m) \right) + \frac{\Theta_p}{n} \sigma^2, \quad (15)$$

with

$$\Theta_p := C'(\theta, p) \frac{\tau_p}{\sigma^p} \left(1 + \sum_{m \in \widehat{\mathcal{M}}_n} m^{-(p/2-2)} \right) \leq C''(\theta, p) \frac{\tau_p}{\sigma^p},$$

and $\mathbf{E}_{\mathbf{X}}$ denotes the conditional expectation with respect to \mathbf{X} .

Noticing that we set earlier $\Pi_{\hat{s}_m} = \hat{\Pi}_m$, Equation (15) leads to :

$$\mathbf{E}_{\mathbf{X}} \left[\|\beta - \hat{\beta}_{\hat{m}}\|_{\Gamma_n}^2 \right] \leq C(\theta) \inf_{m \in \widehat{\mathcal{M}}_n} \left(\|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2 + \text{pen}(m) \right) + \frac{C'''(\theta, p)}{n} \sigma^2.$$

Now, we must ensure that the dimension of the oracle (i.e. the dimension that realises the best bias-variance compromise in \mathcal{M}_n) is included in $\widehat{\mathcal{M}}_n$. Remark that, if $m > \hat{N}_n$,

$$\|\beta - \hat{\Pi}_{\hat{N}_n} \beta\|_{\Gamma_n}^2 + \text{pen}(\hat{N}_n) \leq \sum_{j > \hat{N}_n} \hat{\lambda}_j \langle \beta, \hat{\psi}_j \rangle^2 + \text{pen}(m) + \frac{(1 + \theta) \sigma^2}{n} (\hat{N}_n - m),$$

moreover

$$\begin{aligned} \sum_{j > \hat{N}_n + 1} \hat{\lambda}_j \langle \beta, \hat{\psi}_j \rangle^2 &= \|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2 + \sum_{j = \hat{N}_n + 1}^m \hat{\lambda}_j \langle \beta, \hat{\psi}_j \rangle^2 \\ &\leq \|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2 + \mathfrak{s}_n \|\beta\|^2, \end{aligned}$$

since, for all $j > \hat{N}_n$, $\hat{\lambda}_j \leq \mathfrak{s}_n$. Therefore

$$\|\beta - \hat{\Pi}_{\hat{N}_n} \beta\|_{\Gamma_n}^2 + \text{pen}(\hat{N}_n) \leq \|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2 + \text{pen}(m) + \frac{\|\beta\|^2}{n^2}, \quad (16)$$

and we obtain, for all $m \in \mathcal{M}_n$

$$\mathbf{E}_{\mathbf{X}} \left[\|\beta - \hat{\beta}_{\hat{m}}\|_{\Gamma_n}^2 \right] \leq C(\theta) \left(\|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2 + \text{pen}(m) \right) + \frac{C'''(\theta, p)}{n} (\sigma^2 + \|\beta\|^2).$$

The proof is completed by taking expectation on both sides of the last inequality.

A.2 Proof of Proposition 2

Because of the random penalty, we cannot proceed as in the proof of Proposition 1. The following proof is based on contrast decomposition and control of the remaining empirical process. More precisely, by definitions of $\hat{m}^{(uv)}$ and $\hat{\beta}_m$:

$$\gamma_n(\tilde{\beta}^{(uv)}) - \gamma_n(\hat{\Pi}_m\beta) \leq \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\hat{m}^{(uv)}),$$

and

$$\gamma_n(\tilde{\beta}^{(uv)}) - \gamma_n(\hat{\Pi}_m\beta) = \|\beta - \tilde{\beta}^{(uv)}\|_{\Gamma_n}^2 - \|\beta - \hat{\Pi}_m\beta\|_{\Gamma_n}^2 + 2\nu_n(\hat{\Pi}_m\beta - \tilde{\beta}^{(uv)}),$$

with:

$$\nu_n(t) := \frac{1}{n} \sum_{i=1}^n \varepsilon_i \langle t, X_i \rangle,$$

an empirical linear centred process. Then:

$$\|\beta - \tilde{\beta}^{(uv)}\|_{\Gamma_n}^2 \leq \|\beta - \hat{\Pi}_m\beta\|_{\Gamma_n}^2 + \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\hat{m}^{(uv)}) + 2\nu_n(\tilde{\beta}^{(uv)} - \hat{\Pi}_m\beta). \quad (17)$$

The first step is to replace the random function $\widehat{\text{pen}}$ by its empirical counterpart $\text{pen}^{(uv)}$, this can be done by using the results of Lemma 17 in the Appendix directly in Equation (17):

$$\begin{aligned} \mathbf{E}_{\mathbf{X}}[\|\beta - \tilde{\beta}^{(uv)}\|_{\Gamma_n}^2] &\leq \left(1 + \kappa \frac{m}{n}\right) \|\beta - \hat{\Pi}_m\beta\|_{\Gamma_n}^2 \\ &\quad + \mathbf{E}_{\mathbf{X}}[\text{pen}^{(uv)}(m) - \text{pen}^{(uv)}(\hat{m}^{(uv)})] \\ &\quad + \mathbf{E}_{\mathbf{X}} \left[2 \left(1 + \frac{\kappa \hat{m}^{(uv)}}{n}\right) \nu_n(\tilde{\beta}^{(uv)} - \hat{\Pi}_m\beta) \right] + \frac{\tau_p^{2/p} + \|\beta\|_{\Gamma}^2 + \sigma^2}{n}, \end{aligned} \quad (18)$$

where $\kappa := 2(\theta + 1)$.

Then the last step consists in controlling the empirical linear process ν_n on $\hat{S}_{m \vee \hat{m}}$. Remark that for all $\delta > 0$, for all $m \in \widehat{\mathcal{M}}_n$,

$$2\nu_n(\tilde{\beta}^{(uv)} - \hat{\Pi}_m\beta) \leq \frac{1}{\theta} \|\tilde{\beta}^{(uv)} - \hat{\Pi}_m\beta\|_{\Gamma_n}^2 + \theta \sup_{\substack{f \in \hat{S}_{\hat{m}^{(uv)} \vee m} \\ \|f\|_{\Gamma_n} = 1}} \nu_n^2(f), \quad (19)$$

since for all $x, y \in \mathbf{R}$ and $\theta > 0$, $2xy \leq \theta^{-1}x^2 + \theta y^2$.

Let $p(m, m') := 2(1 + \delta) \frac{m \vee m'}{n} \sigma^2$, remark that $\text{pen}(m) + \text{pen}(m') \geq p(m, m')$.

Then, since $\theta > 4$ and $\hat{N}_n \leq n/\kappa$, gathering equations (18) and (19) we obtain:

$$\begin{aligned} \left(1 - \frac{4}{\theta}\right) \mathbf{E}_{\mathbf{X}} \left[\|\tilde{\beta}^{(uv)} - \beta\|_{\Gamma_n}^2 \right] &\leq \left(2 + \frac{4}{\theta}\right) \|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2 \\ &+ 2\text{pen}^{(uv)}(m) + 2\theta \mathbf{E}_{\mathbf{X}} \left[\left(\sup_{\substack{f \in \hat{S}_{\hat{m}^{(uv)} \vee m} \\ \|f\|_{\Gamma_n} = 1}} \nu_n^2(f) - p(m, \hat{m}^{(uv)}) \right)_+ \right]. \end{aligned}$$

Then the last step is to bound the variations of $\sup_{\substack{f \in \hat{S}_{\hat{m}^{(uv)} \vee m} \\ \|f\|_{\Gamma_n} = 1}} \nu_n^2(f)$ (which can be seen as a variance term) around $p(m, m')$, the results comes from Lemma 5 detailed below:

$$\mathbf{E}_{\mathbf{X}} \left[\|\tilde{\beta}^{(uv)} - \beta\|_{\Gamma_n}^2 \right] \leq C(\theta) \min_{m \in \widehat{\mathcal{M}}_n} \left\{ \|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2 + \text{pen}^{(uv)}(m) \right\} + \frac{C(p, \delta)}{n} \sigma^2.$$

Then we conclude as in the proof of Proposition 1 by Inequality (16).

For sake of clarity, Lemma 5, which is the key of the previous result, is given below.

Lemma 5. *Suppose that Assumption **H1** is fulfilled. Let $p(m, m') = 2(1 + \delta) \frac{m \vee m'}{n} \sigma^2$, then for all $m \in \widehat{\mathcal{M}}_n$,*

$$\sum_{m' \in \widehat{\mathcal{M}}_n} \mathbf{E}_{\mathbf{X}} \left[\left(\sup_{\substack{f \in \hat{S}_{m \vee m'} \\ \|f\|_{\Gamma_n} = 1}} \nu_n^2(f) - p(m, m') \right)_+ \right] \leq \frac{C(p, \delta)}{n} \sigma^2.$$

PROOF. First denote by $\bar{f}_{\mathbf{X}} := (\langle f, X_1 \rangle, \dots, \langle f, X_n \rangle)'$ and $\bar{\varepsilon} := (\varepsilon_1, \dots, \varepsilon_n)'$. Remark that $\nu_n(f) = \bar{f}_{\mathbf{X}}' \bar{\varepsilon} / n$ and that, by usual properties of orthogonal projectors

$$\begin{aligned} \sup_{\substack{f \in \hat{S}_m \\ \|f\|_{\Gamma_n} = 1}} \nu_n(f) &= \sup_{\substack{\alpha \in \hat{s}_m \\ \alpha' \alpha = n}} \frac{\alpha' \bar{\varepsilon}}{n} = \frac{1}{\sqrt{n}} \sup_{\substack{\alpha \in \hat{s}_m \\ \alpha' \alpha = 1}} \alpha' \bar{\varepsilon} \\ &= \frac{1}{\sqrt{n}} (\Pi_{\hat{s}_m} (\bar{\varepsilon})' \Pi_{\hat{s}_m} \bar{\varepsilon})^{1/2} = \frac{1}{\sqrt{n}} (\bar{\varepsilon}' \Pi_{\hat{s}_m} \bar{\varepsilon})^{1/2}, \end{aligned}$$

where \hat{s}_m denotes the subspace of \mathbf{R}^n defined by

$$\hat{s}_m := \left\{ \alpha \in \mathbf{R}^n, \exists f \in \hat{S}_m, \alpha = \bar{f}_{\mathbf{X}} \right\}$$

and $\Pi_{\hat{s}_m}$ is the orthogonal projector onto \hat{s}_m .

Then, by Assumption **H1**, applying Corollary 5.1 of Baraud (2000) [2] with $\tilde{A} = \Pi_{\hat{s}_m}$ we obtain, for all $x > 0$,

$$\mathbf{P}_{\mathbf{X}} \left(n \sup_{\substack{f \in \hat{S}_m \\ \|f\|_{\Gamma_n} = 1}} \nu_n^2(f) \geq m\sigma^2 + 2\sigma^2\sqrt{mx} + \sigma^2x \right) \leq C(p)\sigma^p\tau_p \frac{m}{x^{p/2}},$$

where $\mathbf{P}_{\mathbf{X}}$ stands for the probability given \mathbf{X} . Then for all $\delta > 0$ remark that $\sqrt{mx} \leq \delta m + \delta^{-1}x$ we obtain

$$\mathbf{P}_{\mathbf{X}} \left(\sup_{\substack{f \in \hat{S}_m \\ \|f\|_{\Gamma_n} = 1}} \nu_n^2(f) \geq (1 + \delta) \frac{m\sigma^2}{n} + (1 + \delta^{-1}) \frac{\sigma^2x}{n} \right) \leq C(p) \frac{m}{x^{p/2}}.$$

Set

$$Q_{m \vee m'} := \left(\sup_{\substack{f \in \hat{S}_{m \vee m'} \\ \|f\|_{\Gamma_n} = 1}} \nu_n^2(f) - p(m, m') \right)_+.$$

We have, for all $m, m' \in \mathcal{M}_n$

$$\begin{aligned} \mathbf{E}_{\mathbf{X}} [Q_{m \vee m'}] &= \int_0^{+\infty} \mathbf{P}_{\mathbf{X}} (Q_{m \vee m'} \geq t) dt \\ &\leq C(p) \frac{\sigma^p}{n^{p/2}} \int_0^{+\infty} \frac{dt}{\left(t + \frac{\sigma^2 m \vee m'}{n} (1 + \delta)\right)^{p/2}} \leq C'(p, \delta) \frac{\sigma^2}{n} (m \vee m')^{1-p/2}. \end{aligned}$$

As $p > 4$, $(m \vee m')^{1-p/2} \leq 1$ and we obtain the expected result.

A.3 Proof of Theorem 3

Sketch of proof. The core of the proof relies on the bounds on the empirical risk given in Proposition 1, for the known variance case, and Proposition 2 for the unknown variance case. Then it remains to replace the empirical risk

appearing in propositions 1 and 2 by the risk associated to the prediction error in order to obtain the final oracle-inequality. This is done with the results of Lemma 6 which allows to control the set

$$\Delta_n := \{\forall f \in \hat{\mathcal{S}}_n, \|f\|_{\Gamma}^2 \leq \rho_0 \|f\|_{\Gamma_n}^2\}, \quad (20)$$

where $\rho_0 > 1$ is a constant and we define $\hat{\mathcal{S}}_n := \hat{S}_{\hat{N}_n}$.

PROOF. The following equality holds:

$$\mathbf{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2] = \mathbf{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2 \mathbf{1}_{\Delta_n}] + \mathbf{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2 \mathbf{1}_{\Delta_n^c}],$$

where, for a set A , we denote by A^c its complement.

Lemma 8 in Section A allows to bound the second term of this inequality. Thus the end of the proof will be devoted to upper-bound the first term.

We remark that, for all $m \in \mathcal{M}_n$,

$$\begin{aligned} \|\hat{\beta}_{\hat{m}} - \beta\|_{\Gamma} \mathbf{1}_{\Delta_n} &\leq \|\hat{\beta}_{\hat{m}} - \hat{\Pi}_m \beta\|_{\Gamma} \mathbf{1}_{\Delta_n} + \|\beta - \hat{\Pi}_m \beta\|_{\Gamma} \\ &\leq \sqrt{\rho_0} \|\hat{\beta}_{\hat{m}} - \beta\|_{\Gamma_n} + \sqrt{\rho_0} \|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n} + \|\beta - \hat{\Pi}_m \beta\|_{\Gamma}. \end{aligned} \quad (21)$$

By propositions 1 and 2, for all $m \in \mathcal{M}_n$:

$$\mathbf{E}[\|\hat{\beta}_{\hat{m}} - \beta\|_{\Gamma_n}^2] \leq C(p, \theta, \delta, \sigma^2, \tau_p) \left(\mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2] + \text{pen}(m) + \frac{1 + \|\beta\|^2}{n} \right),$$

and by Equation (21)

$$\begin{aligned} \mathbf{E}[\|\hat{\beta}_{\hat{m}} - \beta\|_{\Gamma}^2 \mathbf{1}_{\Delta_n}] &\leq C(p, \theta, \delta, \sigma^2, \tau_p, \rho_0) \left(\mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2] + \mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma}^2] \right. \\ &\quad \left. + \text{pen}(m) + \frac{1 + \|\beta\|^2}{n} \right), \end{aligned}$$

and Equation (10) follows.

Then Equation (11) comes from Equation (10) and Lemma 16.

A.4 Proof of Theorem 4

Let us start with the polynomial case **(P)**. By Theorem 3, we have:

$$\mathbf{E}[\|\tilde{\beta} - \beta\|_{\Gamma}^2] \leq C \left(\min_{m \in \mathcal{M}_n} \left(\mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma}^2] + \text{pen}(m) \right) + \frac{1}{n} (1 + \|\beta\|^2) \right),$$

with C independent of β and n . Denote by $\hat{\Pi}_m$ the orthogonal projector onto $S_m = \text{Span}\{\psi_1, \dots, \psi_m\}$, by Lemma 15

$$\begin{aligned} \|\beta - \hat{\Pi}_m \beta\|_{\Gamma}^2 &\leq 2\|\beta - \Pi_m \beta\|_{\Gamma}^2 + C_1 \frac{\ln^3(m+1)}{n} m^{\max\{(1-r)_+, 2-a-r\}} \\ &\quad + C_2 \frac{\ln^9 n}{n^2} m^{(4+(a-r)_+ - 2a)_+ + 2}, \end{aligned}$$

with $C_1, C_2 > 0$, independent of β and n . Now since $\beta \in \mathcal{W}_r^R$,

$$\mathbf{E}[\|\beta - \Pi_m \beta\|_{\Gamma}^2] = \sum_{j \geq 1} \lambda_j \langle \beta, \psi_j \rangle^2 \leq m^{-a-r} \sum_{j \geq m} j^r \langle \beta, \psi_j \rangle^2 \leq R m^{-a-r}.$$

We can see easily that it is possible to define a sequence of integers $(m_n^*)_{n \in \mathbf{N} \setminus \{0\}}$ such that

$$m_n^* \asymp n^{1/a+r+1} \text{ and } m_n^* \leq N_n \text{ for all } n \in \mathbf{N} \setminus \{0\},$$

where for two sequences $(a_k)_{k \geq 1}$ and $(b_k)_{k \geq 1}$, we note $a_k \lesssim b_k$ if there exists some constant $c > 0$ such that, for all $k \leq 1$, $a_k \leq c b_k$ and we note also $a_k \asymp b_k$ if $a_k \lesssim b_k$ and $b_k \lesssim a_k$. Now considerations above lead us to

$$\sup_{\beta \in \mathcal{W}_r^R} \mathbf{E}[\|\beta - \hat{\Pi}_{m_n^*} \beta\|_{\Gamma}^2] \lesssim n^{-\frac{a+r}{a+r+1}},$$

as soon as $a+r > 2$ and in addition

$$\text{pen}(m_n^*) \lesssim n^{-\frac{a+r}{a+r+1}},$$

which leads to the expected bound.

The exponential case **(E)** is treated similarly with $m_n^* \asymp \ln^{1/a} n$.

A.5 Technical lemmas

We first bound the probability of $\Delta_n^{\mathbb{C}}$:

Lemma 6. *Under assumptions **H2** and **H3** and if the decreasing rate of $(\lambda_j)_{j \geq 1}$ is given by **(P)** or **(E)**, the set Δ_n defined by Equation (20) verifies*

$$\mathbf{P}(\Delta_n^{\mathbb{C}}) \leq C/n^6,$$

with $C > 0$ independent of n .

PROOF. First remark that:

$$\mathbf{P}(\Delta_n^c) = \mathbf{P}(\Delta_n^c \cap \{\hat{N}_n \leq N_n\}) + \mathbf{P}(\Delta_n^c \cap \{\hat{N}_n > N_n\}),$$

the second term of this equality is easily bounded by Cn^{-6} by Lemma 9. It remains to bound the first term. We have

$$\begin{aligned} \Delta_n^c \cap \{\hat{N}_n \leq N_n\} &= \left\{ \inf_{f \in \hat{S}_n} \frac{\|f\|_{\Gamma_n}^2}{\|f\|_{\Gamma}^2} \langle \rho_0^{-1} \rangle \right\} \cap \{\hat{N}_n \leq N_n\} \\ &\subset \left\{ \inf_{f \in \hat{S}_{N_n}} \frac{\|f\|_{\Gamma_n}^2}{\|f\|_{\Gamma}^2} < \rho_0^{-1} \right\}. \end{aligned}$$

Let $f = \sum_{j=1}^{\hat{N}_n} \alpha_j \hat{\psi}_j \in \hat{S}_{N_n}$, we have a.s.

$$\|f\|_{\Gamma_n}^2 = \sum_{j=1}^{N_n} \hat{\lambda}_j \alpha_j^2 = |\hat{\Lambda}_n \alpha|_2^2,$$

where $|\cdot|_2$ is the norm of \mathbf{R}^{N_n} defined by $|x|_2^2 = \sum_{i=1}^{N_n} x_i^2$ for all $x = (x_1, \dots, x_n)' \in \mathbf{R}^{N_n}$ and $\hat{\Lambda}_n$ is the diagonal matrix with diagonal entries $\{\sqrt{\hat{\lambda}_1}, \dots, \sqrt{\hat{\lambda}_n}\}$.

Moreover

$$\|f\|_{\Gamma}^2 = \sum_{j,k=1}^{N_n} \alpha_j \alpha_k \langle \Gamma^{1/2} \hat{\psi}_j, \Gamma^{1/2} \hat{\psi}_k \rangle = \alpha' \Psi_n \alpha,$$

where Ψ_n is the symmetric and positive-definite matrix

$$\Psi_n := \left(\langle \Gamma^{1/2} \hat{\psi}_j, \Gamma^{1/2} \hat{\psi}_k \rangle \right)_{1 \leq j, k \leq \hat{N}_n}.$$

Then,

$$\frac{\|f\|_{\Gamma_n}^2}{\|f\|_{\Gamma}^2} = \frac{|\hat{\Lambda}_n \alpha|_2^2}{\alpha' \Psi_n \alpha} = \frac{|\hat{\Lambda}_n \alpha|_2^2}{|\Psi_n^{1/2} \alpha|_2^2} = \frac{|\hat{\Lambda}_n \Psi_n^{-1/2} \Psi_n^{1/2} \alpha|_2^2}{|\Psi_n^{1/2} \alpha|_2^2}.$$

Now

$$\begin{aligned} \inf_{f \in \hat{S}_{N_n}} \frac{\|f\|_{\Gamma_n}^2}{\|f\|_{\Gamma}^2} &= \inf_{\alpha \in \mathbf{R}^{N_n} \setminus \{0\}} \frac{\alpha' \Psi_n^{-1/2} \hat{\Lambda}_n^2 \Psi_n^{-1/2} \alpha}{|\alpha|_2^2} \\ &= \min\{\lambda, \lambda \text{ eigenvalue of } \Psi_n^{-1/2} \hat{\Lambda}_n^2 \Psi_n^{-1/2}\}. \end{aligned}$$

On the set \mathcal{J}_n defined by Equation (31) in Section B.2, by Lemma 11, for all, $j = 1, \dots, N_n$, $\hat{\lambda}_j > 0$. Hence the matrix $\hat{\Lambda}_n$ is invertible, therefore

$$\inf_{f \in \hat{\mathcal{S}}_n} \frac{\|f\|_{\Gamma_n}^2}{\|f\|_{\Gamma}^2} = \rho \left((\Psi_n^{-1/2} \hat{\Lambda}_n^2 \Psi_n^{-1/2})^{-1} \right)^{-1} = \rho(\hat{\Lambda}_n^{-1} \Psi_n \hat{\Lambda}_n^{-1})^{-1},$$

where, for a matrix A , $\rho(A) = \max\{|\lambda|, \lambda \text{ is a complex eigenvalue of } A\}$ denotes the spectral radius of A . We have then

$$\mathbf{P} \left(\Delta_n^c \cap \left\{ \widehat{N}_n < N_n \right\} \right) \leq \mathbf{P}(\mathcal{J}_n \cap \{\rho(\hat{\Lambda}_n^{-1} \Psi_n \hat{\Lambda}_n^{-1}) > \rho_0\}) + \mathbf{P}(\mathcal{J}_n^c). \quad (22)$$

By Lemma 13 in Section B.2, $\mathbf{P}(\mathcal{J}_n^c) \leq C/n^6$, with C depending only on Γ and b . Thus it remains to control the spectral radius of $\hat{\Lambda}_n^{-1} \Psi_n \hat{\Lambda}_n^{-1}$.

We define a linear (random) application O from \mathbf{R}^{N_n} to $\mathbf{L}^2([0, 1])$ by:

$$O : \alpha = (\alpha_1, \dots, \alpha_{N_n})' \mapsto \sum_{j=1}^{N_n} \alpha_j \hat{\psi}_j.$$

We denote by O^* the adjoint of O , which is the linear map from $\mathbf{L}^2([0, 1])$ to \mathbf{R}^{N_n} defined by:

$$O^* : f \mapsto (\langle f, \hat{\psi}_j \rangle_{1 \leq j \leq N_n}).$$

We can check that $OO^* = \hat{\Pi}_{N_n}$ and Ψ_n is the matrix of the linear map $O^* \Gamma O$ in the standard basis of \mathbf{R}^{N_n} .

It is known that the spectral radius of an operator is equal to the spectral radius of its adjoint, then,

$$\rho(\hat{\Lambda}_n^{-1} \Psi_n \hat{\Lambda}_n^{-1}) = \rho(\hat{\mathcal{L}}_n^{-1} O^* \Gamma O \hat{\mathcal{L}}_n^{-1}) = \rho(\Gamma^{1/2} O \hat{\mathcal{L}}_n^{-1} \hat{\mathcal{L}}_n^{-1} O^* \Gamma^{1/2}), \quad (23)$$

where $\hat{\mathcal{L}}_n$ denotes the linear endomorphism of \mathbf{R}^{N_n} whose matrix in the standard basis is $\hat{\Lambda}_n$. Denote by Π_{N_n} the orthogonal projector onto $S_{N_n} = \text{span}\{\psi_1, \dots, \psi_{N_n}\}$. Moreover, let Γ^\dagger (resp. Γ_n^\dagger) the pseudo-inverse of operator Γ (resp. Γ_n) on \hat{S}_{N_n} (resp. $\hat{\mathcal{S}}_n$), defined by:

$$\Gamma^\dagger f := \sum_{j=1}^{N_n} \frac{\langle f, \psi_j \rangle}{\lambda_j} \psi_j \text{ and } \Gamma_n^\dagger f := \sum_{j=1}^{N_n} \frac{\langle f, \hat{\psi}_j \rangle}{\hat{\lambda}_j} \hat{\psi}_j \mathbf{1}_{\{\hat{\lambda}_j > 0\}}, \quad (24)$$

we have $\Gamma^{1/2} \Gamma^\dagger \Gamma^{1/2} = \Pi_{N_n}$ and $O \hat{\mathcal{L}}_n^{-2} O^* = \Gamma_n^\dagger$. Then,

$$\begin{aligned} \Gamma^{1/2} \hat{\mathcal{L}}_n^{-1} \hat{\mathcal{L}}_n^{-1} O^* \Gamma^{1/2} &= \Gamma^{1/2} \Gamma_n^\dagger \Gamma^{1/2} = \Gamma^{1/2} (\Gamma^\dagger + \Gamma_n^\dagger - \Gamma^\dagger) \Gamma^{1/2} \\ &= \Pi_{\hat{N}_n} + \Gamma^{1/2} (\Gamma_n^\dagger - \Gamma^\dagger) \Gamma^{1/2}, \end{aligned}$$

and by Equation (23)

$$\rho(\hat{\Lambda}_n^{-1}\Psi_n\hat{\Lambda}_n^{-1}) = \|\Pi_{\hat{N}_n} + \Gamma^{1/2}(\Gamma_n^\dagger - \Gamma^\dagger)\Gamma^{1/2}\|_\infty \leq 1 + \|\Gamma^{1/2}(\Gamma_n^\dagger - \Gamma^\dagger)\Gamma^{1/2}\|_\infty,$$

where $\|\cdot\|_\infty$ denotes the usual operator norm.

Now

$$\mathbf{P}(\mathcal{J}_n \cap \{\rho(\hat{\Lambda}_n^{-1}\Psi_n\hat{\Lambda}_n^{-1}) > \rho_0\}) \leq \mathbf{P}(\mathcal{J}_n \cap \{\|\Gamma^{1/2}(\Gamma_n^\dagger - \Gamma^\dagger)\Gamma^{1/2}\|_\infty > \rho_0 - 1\}).$$

Thus, the results of Lemma 7 – whose technical proof is given in Section B.5 – in Equation (22) allows us to bound $\mathbf{P}(\Delta_n^{\mathfrak{C}} \cap \{\widehat{N}_n < N_n\})$. Then the proof is finished by Lemma 9.

Lemma 7. *Suppose that assumptions **H2** and **H3** are fulfilled and that the decreasing rate of $(\lambda_j)_{j \geq 1}$ is given by **(P)** or **(E)**, then*

$$\mathbf{P}(\mathcal{J}_n \cap \{\|\Gamma^{1/2}(\Gamma_n^\dagger - \Gamma^\dagger)\Gamma^{1/2}\|_\infty > \rho_0 - 1\}) \leq Cn^{-6},$$

with $C > 0$ independent of n .

Lemma 8. *For all $\beta \in \mathbf{L}^2([0, 1])$, if assumptions of Lemma 6 are fulfilled then*

$$\mathbf{E}[\|\tilde{\beta} - \beta\|_{\Gamma} \mathbf{1}_{\Delta_n^{\mathfrak{C}}}] \leq \frac{C}{n}(1 + \|\beta\|_{\Gamma}^2),$$

with $C > 0$ independent of β and n .

PROOF. First remark that, as $Y_i = \langle \beta, X_i \rangle + \varepsilon_i$, for all $m \in \widehat{\mathcal{M}}_n$

$$\hat{\beta}_m = \sum_{j=1}^m \frac{1}{n} \sum_{i=1}^n Y_i \frac{\langle X_i, \hat{\psi}_j \rangle}{\hat{\lambda}_j} \hat{\psi}_j = \hat{\Pi}_m \beta + R_m,$$

where

$$R_m := \sum_{j=1}^m \frac{1}{n} \sum_{i=1}^n \varepsilon_i \frac{\langle X_i, \hat{\psi}_j \rangle}{\hat{\lambda}_j} \hat{\psi}_j.$$

Then

$$\begin{aligned} \mathbf{E}[\|\tilde{\beta} - \beta\|_{\Gamma} \mathbf{1}_{\Delta_n^{\mathfrak{C}}}] &\leq 2\mathbf{E} \left[\|\hat{\Pi}_{\hat{m}} \beta - \beta\|_{\Gamma}^2 \mathbf{1}_{\Delta_n^{\mathfrak{C}}} \right] + 2\mathbf{E} \left[\|R_{\hat{m}}\|_{\Gamma}^2 \mathbf{1}_{\Delta_n^{\mathfrak{C}}} \right] \\ &\leq 2\|\beta\|_{\Gamma}^2 \mathbf{P}(\Delta_n^{\mathfrak{C}}) + 2\mathbf{E} \left[\|R_{\hat{m}}\|_{\Gamma}^2 \mathbf{1}_{\Delta_n^{\mathfrak{C}}} \right]. \end{aligned}$$

The first term can be easily bounded using the results of Lemma 6. Then we focus on the second term, the idea is to bound the quantity $\|R_{\hat{m}}\|_\Gamma$ by $\|R_{\hat{m}}\|$ which can be written simply,

$$\mathbf{E} \left[\|R_{\hat{m}}\|_\Gamma^2 \mathbf{1}_{\Delta_n^c} \right] \leq \rho(\Gamma) \mathbf{E} \left[\|R_{\hat{m}}\|^2 \mathbf{1}_{\Delta_n^c} \right] = \rho(\Gamma) \mathbf{E} \left[\sum_{j=1}^{\hat{m}} \langle R_{\hat{m}}, \hat{\psi}_j \rangle^2 \mathbf{1}_{\Delta_n^c} \right].$$

Now remark that, since $\hat{m} \leq \hat{N}_n \leq 20\sqrt{n}$,

$$\begin{aligned} \sum_{j=1}^{\hat{m}} \langle R_{\hat{m}}, \hat{\psi}_j \rangle^2 &\leq \sum_{j=1}^{\hat{m}} \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i \frac{\langle X_i, \hat{\psi}_j \rangle}{\hat{\lambda}_j} \right)^2 \leq \\ &\mathfrak{s}_n^{-1} \sum_{j=1}^{20\lfloor \sqrt{n} \rfloor} \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i \frac{\langle X_i, \hat{\psi}_j \rangle}{\sqrt{\hat{\lambda}_j}} \right)^2 \mathbf{1}_{\{\hat{\lambda}_j > 0\}}. \end{aligned}$$

Then we have, by independence of ε_i with X_i and Δ_n

$$\mathbf{E} \left[\|R_{\hat{m}}\|_\Gamma^2 \mathbf{1}_{\Delta_n^c} \right] \leq 20\rho(\Gamma) \frac{\sigma^2}{\sqrt{n}} \mathfrak{s}_n^{-1} \mathbf{P}(\Delta_n^c),$$

and the results come from lemmas 6 and 9 and the definition of \mathfrak{s}_n .

Lemma 9. *If Assumption **H2** is fulfilled and $n \geq 6$, then*

$$\mathbf{P}(\hat{N}_n > N_n) \leq Cn^{-6},$$

with C independent of β and n .

PROOF. The sequence $(\lambda_j)_{j \geq 1}$ being non-increasing we have

$$\mathbf{P}(\hat{N}_n > N_n) \leq \mathbf{P}(\lambda_{N_n+1} \geq \lambda_{\hat{N}_n}) \leq \mathbf{P}(\{\lambda_{N_n+1} \geq \lambda_{\hat{N}_n}\} \cap \mathcal{A}_n) + \mathbf{P}(\mathcal{A}_n^c),$$

where \mathcal{A}_n is defined by Equation (27) in Section B.2. Then by definition of \mathcal{A}_n ,

$$\{\lambda_{N_n+1} \geq \lambda_{\hat{N}_n}\} \cap \mathcal{A}_n \subset \left\{ \lambda_{N_n+1} \geq \hat{\lambda}_{\hat{N}_n} - \frac{\delta_{\hat{N}_n}}{2} \right\} = \emptyset,$$

since $\lambda_{N_n+1} < n^{-2}$, $\hat{\lambda}_{\hat{N}_n} \geq \mathfrak{s}_n$ and $\delta_{\hat{N}_n} \leq \frac{\lambda_{\hat{N}_n}}{4} \leq \frac{n^{-2}}{4}$. Thus the proof is finished following the conclusions of Remark 4 in Section B.2.

Appendix B Perturbation theory background

Many intermediate results are based on perturbation theory. We give in this section some preliminary results on this subject. The aim is to control the proximity between the random space \hat{S}_m spanned by the eigenfunctions of Γ_n and the space S_m spanned by the eigenfunctions of Γ .

Recall that Π_m (resp. $\hat{\Pi}_m$) denotes the orthonormal projector onto S_m (resp. \hat{S}_m) and π_j (resp. $\hat{\pi}_j$) denotes the orthonormal projector onto $\text{span}\{\psi_j\}$ (resp. $\text{span}\{\hat{\psi}_j\}$). We write the difference of projectors $\Pi_m - \hat{\Pi}_m$ (or equivalently $\pi_j - \hat{\pi}_j$) explicitly in terms of the operators difference $\Gamma - \Gamma_n$ easier to handle.

B.1 Exponential inequalities

In the proofs, we use the following version of Bernstein's Inequality:

Lemma 10 (Birgé et Massart (1998) [6]). *Let Z_1, \dots, Z_n be independent random variables satisfying the moments conditions*

$$\frac{1}{n} \sum_{i=1}^n \mathbf{E}[|Z_i^m|] \leq \frac{m!}{2} v^2 c^{m-2}, \text{ for all } m \geq 2,$$

for some positive constants v and c . Then, for any positive ε ,

$$\mathbf{P} \left(\frac{1}{n} \sum_{i=1}^n (Z_i - \mathbf{E}[Z_i]) \geq \varepsilon \right) \leq \exp \left(\frac{-n\varepsilon^2/2}{v^2 + c\varepsilon} \right).$$

We use also a version of Lemma 10 for Hilbert valued random variables which can be found in Bosq [8].

B.2 Preliminary notions

Let γ be either the rectangular path given by Figure 5 or the union (for $j = 1, \dots, m$) of the circular paths $\partial\Omega_j$ of center λ_j and radius δ_j represented in Figure 6.

We have, for all $j \geq 1$, $x \in [0, 1]$:

$$\Pi_m \psi_j(x) = \mathbf{1}_{j \leq m} \psi_j(x) = \frac{1}{2i\pi} \int_{\gamma} \frac{\psi_j(x)}{\zeta - \lambda_j} d\zeta = \frac{1}{2i\pi} \int_{\gamma} (\zeta I - \Gamma)^{-1} \psi_j(x) d\zeta,$$

and

$$\Pi_m = \frac{1}{2i\pi} \int_{\gamma} (\zeta I - \Gamma)^{-1} d\zeta, \quad (25)$$

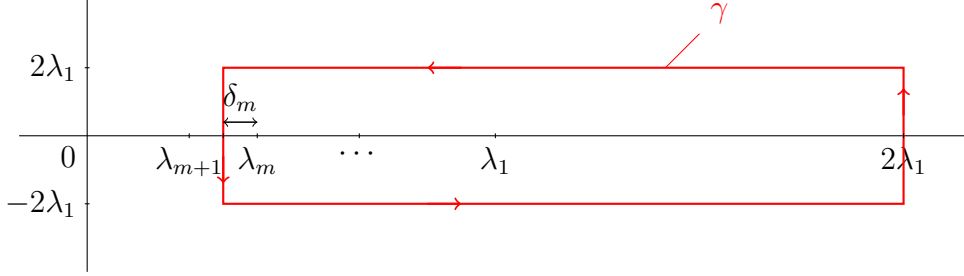


Figure 5: Rectangular contour

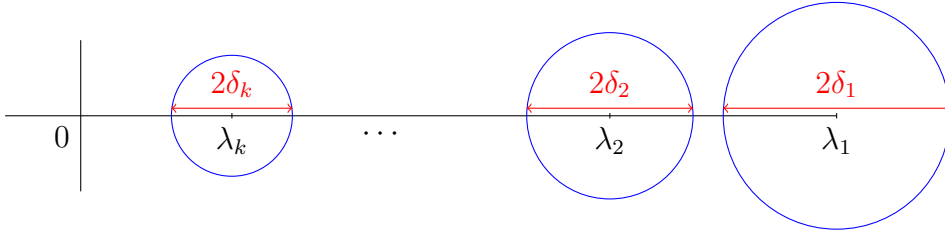


Figure 6: Contour made of disjoint circles

we refer to Chapter III of Dunford and Schwartz [21] for an exact definition and properties of this integral.

Now the aim is to write similarly the random projector $\hat{\Pi}_m$. This can be done if, for all $j \leq m$, $\hat{\lambda}_j$ is in the interior of γ .

For $t > 0$, let $\mathcal{J}_\gamma(t)$ be the set

$$\mathcal{J}_\gamma(t) = \left\{ \sup_{\zeta \in \text{supp}(\gamma)} \|\mathcal{T}(\zeta)\|_\infty < t \right\}, \quad (26)$$

where $\mathcal{T}(\zeta) := R^{1/2}(\zeta)(\Gamma_n - \Gamma)R^{1/2}(\zeta)$ with $R(\zeta) = (\zeta I - \Gamma)^{-1}$. Define also the set

$$\mathcal{A}_n := \bigcap_{j=1}^n \left\{ |\hat{\lambda}_j - \lambda_j| < \frac{\delta_j}{2} \right\}. \quad (27)$$

Lemma 11. *Let $t < 1/2$, for both circular and rectangular path γ , $\mathcal{J}_\gamma(t) \subset \mathcal{A}_n$.*

The proof of Lemma 11 may be deduced from the proof of Lemma 18 of Mas and Ruymgaart [33].

Now Lemma 11, allows us to write that for all $t < 1/2$:

$$\hat{\Pi}_m = \mathbf{1}_{\mathcal{J}_\gamma(t)} \frac{1}{2i\pi} \int_\gamma (\zeta I - \Gamma_n)^{-1} d\zeta. \quad (28)$$

Define $\widehat{R}(\zeta) := (\zeta I - \Gamma_n)^{-1}$, equations (25) and (28) lead to:

$$(\hat{\Pi}_m - \Pi_m) \mathbf{1}_{\mathcal{J}_\gamma(t)} = \mathbf{1}_{\mathcal{J}_\gamma(t)} \frac{1}{2i\pi} \int_\gamma \left(R(\zeta) - \widehat{R}(\zeta) \right) d\zeta, \text{ for all } t < 1/2.$$

Then we rewrite the interior of the last integral by remarking that

$$R(\zeta) - \widehat{R}(\zeta) = \widehat{R}(\zeta) (\Gamma - \Gamma_n) R(\zeta) = \widehat{R}(\zeta) (\zeta I - \Gamma)^{1/2} \mathcal{T}(\zeta) R^{1/2}(\zeta). \quad (29)$$

By definition, when $t < 1$, on the set $\mathcal{J}_\gamma(t)$, the operator $I - \mathcal{T}(\zeta)$ is invertible for all $\zeta \in \text{supp}(\gamma)$ and we have:

$$(I - \mathcal{T}(\zeta))^{-1} = (\zeta I - \Gamma)^{1/2} \widehat{R}(\zeta) (\zeta I - \Gamma)^{1/2}.$$

Then Equation (29) leads to

$$\left(\widehat{R}(\zeta) - R(\zeta) \right) \mathbf{1}_{\mathcal{J}_\gamma(t)} = R^{1/2}(\zeta) [I - \mathcal{T}(\zeta)]^{-1} \mathcal{T}(\zeta) R^{1/2}(\zeta) \mathbf{1}_{\mathcal{J}_\gamma(t)}.$$

We obtain in a similar way a rewriting of $(\hat{\pi}_j - \pi_j) \mathbf{1}_{\mathcal{J}_\gamma(t)}$ or $(\Gamma^\dagger - \Gamma_n^\dagger) \mathbf{1}_{\mathcal{J}_\gamma(t)}$ where Γ^\dagger and Γ_n^\dagger are defined by Equation (24). All results are summarized in the following lemma.

Lemma 12. *For all $t < 1/2$ if γ is either the union of circular contour represented in Figure 6 or the rectangular contour given by Figure 5:*

$$\begin{aligned} (\hat{\Pi}_m - \Pi_m) \mathbf{1}_{\mathcal{J}_\gamma(t)} &= \frac{1}{2i\pi} \int_\gamma R^{1/2}(\zeta) [I - \mathcal{T}(\zeta)]^{-1} \mathcal{T}(\zeta) R^{1/2}(\zeta) d\zeta \mathbf{1}_{\mathcal{J}_\gamma(t)} \\ (\hat{\pi}_j - \pi_j) \mathbf{1}_{\mathcal{J}_\gamma(t)} &= \frac{1}{2i\pi} \int_{\partial\Omega_j} R^{1/2}(\zeta) [I - \mathcal{T}(\zeta)]^{-1} \mathcal{T}(\zeta) R^{1/2}(\zeta) d\zeta \mathbf{1}_{\mathcal{J}_\gamma(t)} \\ (\Gamma^\dagger - \Gamma_n^\dagger) \mathbf{1}_{\mathcal{J}_\gamma(t)} &= \frac{1}{2i\pi} \int_\gamma \frac{1}{\zeta} R^{1/2}(\zeta) [I - \mathcal{T}(\zeta)]^{-1} \mathcal{T}(\zeta) R^{1/2}(\zeta) d\zeta \mathbf{1}_{\mathcal{J}_\gamma(t)}. \end{aligned}$$

The last lemma allows us to finally control our quantities on $\mathcal{J}_\gamma(t)^\complement$.

Lemma 13. *Denote by*

$$\mathbf{a}_k := \sum_{i \neq k} \frac{\lambda_i}{|\lambda_i - \lambda_k|} + \frac{\lambda_k}{\delta_k}, \text{ for all } k \geq 1.$$

If γ is the path $\partial\Omega_k$ covering the circle of center λ_k and of radius δ_k or the rectangular contour given in Figure 5 we have, under Assumption **H2**,

$$\mathbf{P} \left(\mathcal{J}_\gamma^{\mathbb{C}}(t) \right) \leq 2 \exp \left(- \frac{nt^2}{2\mathbf{a}_k^2(2b-1)(2b-1) + 256b^3/((2b-1)\mathbf{a}_k t)} \right).$$

The proof of Lemma 13 relies on Bernstein's exponential inequality for Hilbert-valued random variables (see for instance Bosq [8]). Details can be found in Mas and Ruymgaart [33, Lemma 16].

Definition 1. Let, for all $t > 0$, $\mathcal{J}_\gamma(t)$ be the set defined by Equation (26) then we define a set \mathcal{J}_n in the following way: take $l_{j,n} := \min \left\{ \frac{\mathbf{a}_j}{\sqrt{n}} \ln n, 1/2 \right\}$,

Circular contour

$$\mathcal{J}_n := \bigcap_{j=1}^m \mathcal{J}_{\partial\Omega_j}(l_{j,n}); \tag{30}$$

Rectangular contour

$$\mathcal{J}_n := \mathcal{J}_\gamma(l_{m,n}). \tag{31}$$

Remark 4. By Lemma 13, the sets \mathcal{J}_n defined by Equation (30) or (31) verifies

$$\mathbf{P}(\mathcal{J}_n^{\mathbb{C}}) \leq 2 \exp(-c^* \ln^2 n), \tag{32}$$

where $c^* > 0$ depends only on b and c_δ . Moreover, as $l_{j,n} < 1/2$, by Lemma 11, we have

$$\mathcal{J}_n \subset \mathcal{A}_n$$

and the results of Lemma 12 are true.

B.3 Upper-bound on the distance between empirical and theoretical projectors

We need a preliminary lemma

Lemma 14 (Hilgert *et al.* [28], Lemma 10.1). *If Assumption **H3** is verified, then for all $k \in \mathbf{N} \setminus \{0\}$:*

$$\mathbf{a}_k \leq C(\gamma)k \ln k$$

Lemma 15. *Let $r, R > 0$ and $\beta \in \mathcal{W}_r^R$. Suppose that assumptions **H2** and **H3** are fulfilled. If $(\lambda_j)_{j \geq 1}$ decreases at polynomial rate (**P**) then*

$$\begin{aligned} \mathbf{E}[\|\hat{\Pi}_m \beta - \Pi_m \beta\|_{\Gamma}^2 \mathbf{1}_{\mathcal{J}_n}] &\leq C_1 \frac{\ln^3 m}{n} m^{\max\{(1-r)_+, 2-a-r\}} \\ &\quad + C_2 \frac{\ln^9 n}{n^2} m^{(4+(a-r)_+ - 2a)_+ + 2}, \end{aligned}$$

and if $(\lambda_j)_{j \geq 1}$ decreases at exponential rate (**E**)

$$\mathbf{E}[\|\hat{\Pi}_m \beta - \Pi_m \beta\|_{\Gamma}^2 \mathbf{1}_{\mathcal{J}_n}] \leq C_3 \frac{\ln^3 m}{n} m^{(1-r)_+},$$

with $C_1 > 0$, $C_2 > 0$ and $C_3 > 0$ depending on r , R and on the sequence $(\lambda_j)_{j \geq 1}$ but are independent of m and n .

PROOF. First, by Equation (32)

$$\mathbf{E}[\|\hat{\Pi}_m \beta - \Pi_m \beta\|_{\Gamma}^2 \mathbf{1}_{\mathcal{J}_n^c}] \leq 2\|\beta\|_{\Gamma}^2 \mathbf{P}(\mathcal{J}_n^c) \leq C\|\beta\|_{\Gamma}^2/n^2,$$

with C independent of β , n and m . Now by Lemma 12,

$$(\hat{\Pi}_m \beta - \Pi_m \beta) \mathbf{1}_{\mathcal{J}_n} = \frac{\mathbf{1}_{\mathcal{J}_n}}{2i\pi} \sum_{j=1}^m \int_{\partial\Omega_j} R^{1/2}(z) [I - \mathcal{T}(z)]^{-1} \tilde{\Pi}(z) R^{1/2}(z) \beta dz.$$

Remark that $(I - \mathcal{T}(z))^{-1} \mathcal{T}(z) = \mathcal{T}(z) + (I - \mathcal{T}(z))^{-1} \mathcal{T}^2(z)$, then

$$\mathbf{1}_{\mathcal{J}_n} \left(\hat{\Pi}_m - \Pi_m \right) = A_n + B_n,$$

with

$$\begin{aligned} A_n &= \mathbf{1}_{\mathcal{J}_n} \frac{1}{2\pi i} \sum_{j=1}^m \int_{\partial\Omega_j} [(zI - \Gamma)^{-1} (\Gamma_n - \Gamma) (zI - \Gamma)^{-1}] dz \\ B_n &= \mathbf{1}_{\mathcal{J}_n} \frac{1}{2\pi i} \sum_{j=1}^m \int_{\partial\Omega_j} [(zI - \Gamma)^{-1/2} [I - \mathcal{T}(z)]^{-1} \mathcal{T}(z)^2 (zI - \Gamma)^{-1/2}] dz. \end{aligned}$$

We deal with A_n first. Calculations show that

$$\begin{aligned}
\mathbf{E} \left\| \Gamma^{1/2} A_n \beta \right\|^2 &\leq \frac{1}{n} \sum_{j=1}^m \beta_j^2 \sum_{k>m} \frac{\lambda_j \lambda_k^2}{(\lambda_j - \lambda_k)^2} + \frac{1}{n} \sum_{j=1}^m \lambda_j^2 \sum_{k>m} \frac{\lambda_k \beta_k^2}{(\lambda_j - \lambda_k)^2} \\
&\leq \frac{1}{n} \sum_{j=1}^m \lambda_j \beta_j^2 \mathbf{a}_j^2 + \frac{1}{n} \sum_{j=1}^m \frac{\lambda_j^2}{(\lambda_j - \lambda_{m+1})^2} \lambda_{m+1} m^{-r} \sum_{k>m} k^r \beta_k^2 \\
&\leq C \left(\frac{m^{(1-r)+} \ln^3 m}{n} + \frac{m^{2-r} \ln^2(m+1)}{n} \lambda_{m+1} \right), \tag{33}
\end{aligned}$$

as soon as $(\lambda_j)_{j \geq 1}$ decreases exponentially or polynomially with $C > 0$ depending only on R, r, a and Γ . The last line comes from the inequality $\lambda_j \lesssim j^{-1}$ and Lemma 14. We turn now to B_n

$$\begin{aligned}
&\|B_n \beta \mathbf{1}_{\mathcal{J}_n}\|_{\Gamma}^2 \\
&\leq \frac{\mathbf{1}_{\mathcal{J}_n}}{4\pi^2} \sum_{k \geq 1} \lambda_k \left(\sum_{j=1}^m \int_{\partial\Omega_j} \langle R^{1/2}(\zeta) [I - \mathcal{T}(\zeta)]^{-1} \mathcal{T}(\zeta)^2 R^{1/2}(\zeta) \beta, \psi_k \rangle d\zeta \right)^2. \tag{34}
\end{aligned}$$

We have:

$$\begin{aligned}
\langle R^{1/2}(\zeta) [I - \mathcal{T}(\zeta)]^{-1} \mathcal{T}(\zeta)^2 R^{1/2}(\zeta) \beta, \psi_k \rangle &= \\
&\langle [I - \mathcal{T}(\zeta)]^{-1} \mathcal{T}(\zeta)^2 R^{1/2}(\zeta) \beta, R^{1/2}(\zeta) \psi_k \rangle.
\end{aligned}$$

We denote by $\beta^\diamond := \sum_{j \geq 1} j^{r/2} \beta_j \psi_j$; as β is in \mathcal{W}_r^R , the function β^\diamond is in $\mathbf{L}^2([0, 1])$. Moreover we denote by P_r the diagonal compact operator defined by $P_r \psi_j = j^{-r/2} \psi_j$, we remark that $\beta = P_r \beta^\diamond$. We have

$$R^{1/2}(\zeta) \psi_k = (\zeta I - \Gamma)^{-1/2} \psi_k = \frac{1}{\sqrt{\zeta - \lambda_k}} \psi_k.$$

Then,

$$\begin{aligned}
|\langle R^{1/2}(\zeta) [I - \mathcal{T}(\zeta)]^{-1} \tilde{\Pi}(\zeta)^2 R^{1/2}(\zeta) \beta, \psi_k \rangle| &\leq \\
&\frac{1}{\sqrt{|\zeta - \lambda_k|}} \left\| (I - \mathcal{T}(\zeta))^{-1} \right\|_{\infty} \|\mathcal{T}(\zeta)\|_{\infty}^2 \|R^{1/2}(\zeta) P_r\|_{\infty} \|\beta^\diamond\|.
\end{aligned}$$

Now on the set \mathcal{J}_n , by definition, we have for all $\zeta \in \Omega_j$:

$$\left\| (I - \mathcal{T}(\zeta))^{-1} \right\|_{\infty} < 2 \text{ and } \|\mathcal{T}(\zeta)\|_{\infty} < \frac{\mathbf{a}_j}{\sqrt{n}} \ln n.$$

Moreover, the eigenvalues of the operator $R^{1/2}(\zeta)P_r$ are $\{k^{-r/2}(\zeta - \lambda_k)^{1/2}, k \geq 1\}$ then, for all, $\zeta \in \partial\Omega_j$:

$$\|R^{1/2}(\zeta)P_r\|_\infty = \sup_{k \geq 1} \{k^{-r/2}|\zeta - \lambda_k|^{-1/2}\}. \quad (35)$$

Now remark that, for all $k > j$, for all $\zeta \in \partial\Omega_j$, by definition,

$$k^{-r/2}|\zeta - \lambda_k|^{-1/2} \leq j^{-r/2}\delta_j^{-1/2}.$$

For $k = j$, $|\zeta - \lambda_k| = \delta_j/2$, hence $k^{-r/2}|\zeta - \lambda_k|^{-1/2} = j^{-r/2}\delta_j^{-1/2}$. Then Equation (35) can be rewritten

$$\|R^{1/2}(\zeta)P_r\|_\infty = \sup_{k=1, \dots, j} \{k^{-r/2}|\zeta - \lambda_k|^{-1/2}\}.$$

Now, for $k = 1, \dots, j$, $|\zeta - \lambda_k| \geq \lambda_k - \lambda_j - \delta_j/2 \geq (\lambda_k - \lambda_j)/2$, since $\lambda_k - \lambda_j \geq \delta_j$ implies that $(\lambda_k - \lambda_j - \delta_j/2)/(\lambda_k - \lambda_j) = 1 - \delta_j/(2(\lambda_k - \lambda_j)) \geq 1/2$. Hence, by definition of \mathbf{a}_j ,

$$\|R^{1/2}P_r\|_\infty \leq \mathbf{a}_j^{1/2} \sup_{k=1, \dots, j} \left\{ \lambda_k^{-1/2} k^{-r/2} \right\}.$$

and Equation (34) becomes:

$$\|B_n \beta \mathbf{1}_{\mathcal{J}_n}\|_\Gamma^2 \leq \frac{R^2 \ln^2 n}{\pi^2} \frac{n}{n} \sum_{k \geq 1} \lambda_k \left(\sum_{j=1}^m \frac{\delta_j}{\sqrt{D_{j,k}}} \mathbf{a}_j^{3/2} \max_{\ell=1, \dots, j} \left\{ \lambda_\ell^{-1/2} \ell^{-r/2} \right\} \right)^2,$$

where $D_{j,k} = |\lambda_j - \lambda_k|$ if $j \neq k$ and $D_{j,j} = \delta_j$.

In the polynomial case **(P)**, Lemma 14, the inequality $\sum_{k \geq 1} \lambda_k D_{j,k}^{-1} \leq \mathbf{a}_j$ and the fact that, for all $\alpha \in \mathbb{R}$, $\sum_{k=1}^m k^\alpha \leq c_\alpha \ln(m+1) m^{(\alpha+1)_+}$ lead to the following bound

$$\|B_n \beta \mathbf{1}_{\mathcal{J}_n}\|_\Gamma^2 \leq C \frac{\ln^9 n}{n^2} m^{(4+(a-r)_+-2a)_++2},$$

with $a > 1$ such that $\lambda_j \asymp j^{-a}$ and $C > 0$ depends only on R, r, a and Γ . Gathering with (33) we obtain the expected result.

In the exponential case **(E)** following a similar scheme, we get

$$\|B_n \beta \mathbf{1}_{\mathcal{J}_n}\|_\Gamma^2 \leq C'/n,$$

with C' depending only on R, r, a and Γ .

B.4 Empirical and theoretical bias terms

Lemma 16. *Suppose that assumptions **H2** and **H3** are fulfilled and that $\beta \in \mathcal{W}_r^R$ with $r, R > 0$ such that, in the polynomial case (**P**), $a + r > 2$. Then for all $m = 1, \dots, N_n$:*

$$\mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_n^2] \leq 4\mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_\Gamma^2] + \tau_{m,n}, \quad (36)$$

where $\tau_{m,n} \leq m\varepsilon(m)/n$ where $\varepsilon(m) \rightarrow 0$ when $m \rightarrow +\infty$ and is independent of β and m .

PROOF. First imagine that the random projectors in the equation above are replaced by non random one. It is elementary to see that

$$\mathbf{E}[\|\beta - \Pi_m \beta\|_n^2] = \|\beta - \Pi_m \beta\|_\Gamma^2$$

and that consequently to get (36) it is enough to show that both $\mathbf{E}[\|\left(\Pi_m - \hat{\Pi}_m\right) \beta\|_\Gamma^2]$ and $\mathbf{E}[\|\left(\Pi_m - \hat{\Pi}_m\right) \beta\|_n^2]$ are bounded by $\tau_{m,n}$. The first bound was proved asymptotically in Cardot *et al.* [16] and non-asymptotically in the Proposition 20 of Crambes and Mas [20] in a slightly more general framework. Specifically these authors get

$$\mathbf{E}[\|\left(\Pi_m - \hat{\Pi}_m\right) \beta\|_\Gamma^2] \leq A \frac{m^2 \lambda_m}{n}$$

where A does not depend on m and n . The only point left is to prove the same sort of bound for

$$\mathbf{E}[\|\left(\Pi_m - \hat{\Pi}_m\right) \beta\|_n^2].$$

The derivation makes use of perturbation methods already used in other parts of proofs. We will skip technical details to concentrate on the essential facts.

In a first step remark that

$$\begin{aligned} \|\left(\Pi_m - \hat{\Pi}_m\right) \beta\|_n^2 &= \left\langle (\Gamma_n - \Gamma) \left(\Pi_m - \hat{\Pi}_m\right) \beta, \left(\Pi_m - \hat{\Pi}_m\right) \beta \right\rangle \\ &\quad + \|\left(\Pi_m - \hat{\Pi}_m\right) \beta\|_\Gamma^2 \end{aligned}$$

and it is enough to focus on the first term and to prove the bound for

$$\mathbf{E} \left[\left\| (\Gamma_n - \Gamma) \left(\Pi_m - \widehat{\Pi}_m \right) \beta \right\| \right].$$

after a Cauchy-Schwartz's Inequality coupled with the fact that

$$\left\| \left(\Pi_m - \widehat{\Pi}_m \right) \beta \right\| \leq 2 \|\beta\|.$$

Now by Lemma 12

$$\begin{aligned} & (\Gamma_n - \Gamma) \left(\Pi_m - \widehat{\Pi}_m \right) \beta \mathbf{1}_{\mathcal{J}_n} \\ &= \frac{1}{2\pi i} \sum_{j=1}^m \int_{\partial\Omega_j} (\Gamma_n - \Gamma) R^{1/2}(\zeta) [I - \mathcal{T}(\zeta)]^{-1} \mathcal{T}(\zeta) R^{1/2}(\zeta) \beta d\zeta \mathbf{1}_{\mathcal{J}_n}. \end{aligned}$$

Hence by definition of \mathcal{J}_n , $\|(I - \mathcal{T}(\zeta))^{-1}\|_{\infty} \mathbf{1}_{\mathcal{J}_n} \leq 2$ and $\|\mathcal{T}(\zeta)\|_{\infty} \mathbf{1}_{\mathcal{J}_n} \leq \frac{\mathbf{a}_j \ln n}{n}$ then

$$\begin{aligned} & \mathbf{E} \left[\left\| (\Gamma_n - \Gamma) \left(\Pi_m - \widehat{\Pi}_m \right) \beta \right\| \mathbf{1}_{\mathcal{J}_n} \right] \\ & \leq \frac{1}{\pi} \sum_{j=1}^m \int_{\partial\Omega_j} \mathbf{E} \left[\left\| (\Gamma_n - \Gamma) R^{1/2}(\zeta) \right\|_{\infty} \|\mathcal{T}(\zeta)\|_{\infty} \mathbf{1}_{\mathcal{J}_n} \right] \left\| R^{1/2}(\zeta) \beta \right\| d\zeta \\ & \leq \frac{\ln n}{\pi \sqrt{n}} \sum_{j=1}^m \mathbf{a}_j \int_{\partial\Omega_j} \left\| R^{1/2}(\zeta) P_r \right\|_{\infty} \|\beta^{\diamond}\| \mathbf{E} \left[\left\| (\Gamma_n - \Gamma) R^{1/2}(\zeta) \right\|_{\infty} \mathbf{1}_{\mathcal{J}_n} \right] d\zeta \\ & \leq \frac{\ln n}{\pi \sqrt{n}} \|\beta^{\diamond}\| \sum_{j=1}^m \mathbf{a}_j \frac{j^{-r/2}}{\sqrt{\delta_j}} \int_{\partial\Omega_j} \sqrt{\mathbf{E} \left[\left\| (\Gamma_n - \Gamma) R^{1/2}(\zeta) \right\|_{HS}^2 \mathbf{1}_{\mathcal{J}_n} \right]} d\zeta \quad (37) \end{aligned}$$

where we recall that, by Equation (35),

$$\left\| R^{1/2}(\zeta) P_r \right\|_{\infty} \leq \frac{j^{-r/2}}{\sqrt{\delta_j}}$$

(the definitions of β^{\diamond} and P_r are given in the proof of Lemma 15).

Treating $\mathbf{E} \left[\left\| (\Gamma_n - \Gamma) R^{1/2}(\zeta) \right\|_{HS}^2 \right]$ with computations similar to those carried previously we get, for all $\zeta \in \partial\Omega_j$

$$\mathbf{E} \left[\left\| (\Gamma_n - \Gamma) R^{1/2}(\zeta) \right\|_{HS}^2 \right] \leq \frac{C' \mathbf{a}_j}{n},$$

with $C' = \text{Tr}(\Gamma) \max\{1, b - 1\}$ and putting into Equation (37) we obtain

$$\mathbf{E} \left[\left\| (\Gamma_n - \Gamma) \left(\Pi_m - \widehat{\Pi}_m \right) \beta \right\| \mathbf{1}_{\mathcal{J}_n} \right] \leq \frac{C' \ln n}{\pi n} \|\beta^\diamond\| \sum_{j=1}^m \mathbf{a}_j^{3/2} j^{-r/2} \sqrt{\delta_j}.$$

Considering again two cases related to the rate of decrease for the eigenvalues we see first that for an exponential decay the term above is bounded up to a constant by $(\ln n)/n$. Secondly in case of polynomial decay we get :

$$\begin{aligned} \mathbf{E} \left[\left\| (\Gamma_n - \Gamma) \left(\Pi_m - \widehat{\Pi}_m \right) \beta \right\| \mathbf{1}_{\mathcal{J}_n} \right] & \leq \frac{C' \ln n}{\pi n} \|\beta^\diamond\| \sum_{j=1}^m j^{3/2} j^{-r/2} j^{-(1+a)/2} \ln^{3/2} j \\ & \leq \frac{C' \ln n}{\pi n} \|\beta^\diamond\| m^{2-(r+a)/2} \ln^{5/2} m \end{aligned}$$

and $\ln n \ln^{5/2} m \cdot m^{2-(r+a)/2}/n = o(m/n)$ when $r + a > 2$.

Thus the proof is finished by Lemma 13

$$\mathbf{E} \left[\left\| \left(\Pi_m - \widehat{\Pi}_m \right) \beta \right\|_n^2 \mathbf{1}_{\mathcal{J}_n^c} \right] \leq \|\beta\|_\Gamma^2 \mathbf{P}(\mathcal{J}_n^c) \leq \frac{C''(b, \Gamma)}{n^2}.$$

B.5 Technical part of the bound on $\mathbf{P}(\Delta_n^c)$

PROOF (PROOF OF LEMMA 7). Let γ be the contour defined by Figure 5 with $m = N_n$.

We have by Lemma 12 and the fact that $(I - \mathcal{T}(z))^{-1} \mathcal{T}(z) = \mathcal{T}(z) + (I - \mathcal{T}(z))^{-1} \mathcal{T}^2(z)$

$$\begin{aligned} & \Gamma^{1/2} [\Gamma_n^\dagger - \Gamma^\dagger] \Gamma^{1/2} \mathbf{1}_{\mathcal{J}_n} \\ & = \mathbf{1}_{\mathcal{J}_n} \frac{1}{2i\pi} \int_\gamma \frac{1}{z} \Gamma^{1/2} R^{1/2}(z) [I - \mathcal{T}(z)]^{-1} \mathcal{T}(z) R^{1/2}(z) \Gamma^{1/2} dz \end{aligned} \quad (38)$$

$$\begin{aligned} & = \mathbf{1}_{\mathcal{J}_n} \frac{1}{2i\pi} \int_\gamma \frac{1}{z} \Gamma^{1/2} R(z) (\Gamma_n - \Gamma) R(z) \Gamma^{1/2} dz \\ & \quad + \mathbf{1}_{\mathcal{J}_n} \frac{1}{2i\pi} \int_\gamma \frac{1}{z} \Gamma^{1/2} R^{1/2}(z) [I - \mathcal{T}(z)]^{-1} \mathcal{T}^2(z) R^{1/2}(z) \Gamma^{1/2} dz. \end{aligned} \quad (39)$$

Now, we consider separately the two decreasing rates of the λ_j 's.

Exponential decrease : $\lambda_j \asymp \exp(-j^a)$, with $a > 0$. By Equation (38) and the fact that $\|(I - \mathcal{T}(z))^{-1}\|_\infty < 2$, on the set \mathcal{J}_n

$$\begin{aligned} & \|\Gamma^{1/2}(\Gamma_n^\dagger - \Gamma^\dagger)\Gamma^{1/2}\|_\infty \mathbf{1}_{\mathcal{J}_n} \\ & \leq \frac{1}{2\pi} \left\| \int_\gamma \frac{1}{z} \Gamma^{1/2} R^{1/2}(z) [I - \mathcal{T}(z)]^{-1} \mathcal{T}(z) R^{1/2}(z) \Gamma^{1/2} dz \right\|_\infty \mathbf{1}_{\mathcal{J}_n} \\ & \leq \pi^{-1} \sup_{z \in \gamma} [\|\mathcal{T}(z)\|_\infty] \int_\gamma \frac{1}{|z|} \|\Gamma^{1/2} R^{1/2}(z)\|_\infty^2 dz \mathbf{1}_{\mathcal{J}_n}. \end{aligned}$$

For $z \in \text{supp}(\gamma)$, the eigenvalues of the operator $\Gamma^{1/2} R^{1/2}(z)$ are $\{\lambda_j^{1/2}(z - \lambda_j)^{-1/2}, j \geq 1\}$, then

$$\|\Gamma^{1/2} R^{1/2}(z)\|_\infty^2 = \sup_{j \geq 1} \left\{ \frac{\lambda_j}{|z - \lambda_j|} \right\},$$

and

$$\left| \int_\gamma \frac{1}{|z|} \|\Gamma^{1/2} R^{1/2}(z)\|_\infty^2 dz \right| \leq C + 2 \int_0^{2\lambda_1/\delta_{N_n}} \frac{du}{1+u^2} \leq C',$$

where C and C' are independent of n and the last inequality comes from the fact that in the exponential case, there exists a constant $c > 0$ such that $\delta_{N_n}/\lambda_{N_n} \geq c$. Then by lemmas 13 and 14:

$$\begin{aligned} & \mathbf{P}(\mathcal{J}_n \cap \{\|\Gamma^{1/2}(\Gamma_n^\dagger - \Gamma^\dagger)\Gamma^{1/2}\|_\infty > \rho_0 - 1\}) \\ & \leq \mathbf{P}\left(C' \sup_{z \in \text{supp}(\gamma)} [\|\mathcal{T}(z)\|_\infty] \pi(\rho_0 - 1)\right) \\ & \leq 2 \exp\left(-c^* \frac{n}{N_n^2 \ln^2 N_n}\right) \end{aligned}$$

with c^* independent of n . The result comes from the fact that $N_n \leq 20\sqrt{n/\ln^3 n}$.

Polynomial decrease : $\lambda_j \asymp j^{-a}$, $a > 1$. Denote by T_1 and T_2 the two terms of Equation (39) i.e.

$$\begin{aligned} T_1 &= \mathbf{1}_{\mathcal{J}_n} \frac{1}{2i\pi} \int_\gamma \frac{1}{z} \Gamma^{1/2} R(z) (\Gamma_n - \Gamma) R(z) \Gamma^{1/2} dz, \\ T_2 &= \mathbf{1}_{\mathcal{J}_n} \frac{1}{2i\pi} \int_\gamma \frac{1}{z} \Gamma^{1/2} R^{1/2}(z) [I - \mathcal{T}(z)]^{-1} \mathcal{T}^2(z) R^{1/2}(z) \Gamma^{1/2} dz. \end{aligned}$$

First we control T_2 , the proof in the exponential case leads us to:

$$\|T_2\|_\infty \leq \pi^{-1} \sup_{z \in \gamma} [\|\mathcal{T}(z)\|_\infty^2] \int_\gamma \frac{1}{|z|} \|\Gamma^{1/2} R^{1/2}(z)\|_\infty^2 dz,$$

and

$$\begin{aligned} \int_\gamma \frac{1}{|z|} \|\Gamma^{1/2} R^{1/2}(z)\|_\infty^2 dz &\leq C + \int_0^{2\lambda_1/\delta_{N_n}} \frac{\lambda_{N_n} du}{\sqrt{(\lambda_{N_n} - \delta_{N_n})^2 + \delta_{N_n}^2 u^2 \sqrt{1+u^2}}} \\ &\leq C + \int_0^1 \frac{\lambda_{N_n}}{\lambda_{N_n} - \delta_{N_n}} du + \int_1^{2\lambda_1/\delta_{N_n}} \frac{\lambda_{N_n} du}{\sqrt{(\lambda_{N_n} - \delta_{N_n})^2 + \delta_{N_n}^2 u^2 \sqrt{1+u^2}}} \\ &\leq 1 + C + \int_1^{2\lambda_1/\delta_{N_n}} \frac{du}{\sqrt{1+u^2}} \leq C' \ln(N_n), \end{aligned}$$

with $C, C' > 0$ independent of n . Then lemmas 13 and 14 and the fact that $N_n \leq 20\sqrt{n/\ln^3 n}$ lead us to

$$\begin{aligned} \mathbf{P}(\|T_2\|_\infty > (\rho_0 - 1)/2) &\leq \mathbf{P}\left(C' \ln(N_n) \sup_{z \in \gamma} [\|\mathcal{T}(z)\|_\infty^2] > \pi(\rho_0 - 1)/2\right) \\ &\leq \exp\left(-c^{**} \frac{n}{N_n^2 \ln^4(N_n)}\right) \leq C'' n^{-6}, \end{aligned}$$

with c^{**} and C'' independent of n

Now, we can calculate explicitly the term T_1

$$T_1 = \mathbf{1}_{\mathcal{J}_n} \frac{1}{2i\pi} \int_\gamma \sum_{j,k \geq 1} \frac{\sqrt{\lambda_k} \sqrt{\lambda_j}}{z(z - \lambda_k)(z - \lambda_j)} \pi_k(\Gamma_n - \Gamma) \pi_j dz.$$

By the Residue Theorem

$$\frac{1}{2i\pi} \int_\gamma \frac{dz}{z(z - \lambda_k)(z - \lambda_j)} = \begin{cases} -\frac{1}{\lambda_j \lambda_k} & \text{if } j \neq k, j \leq N_n \text{ and } k \leq N_n, \\ \frac{1}{\lambda_j(\lambda_j - \lambda_k)} & \text{if } j \leq N_n < k, \\ \frac{1}{\lambda_k(\lambda_k - \lambda_j)} & \text{if } k \leq N_n < j, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned}
T_1 &= -\mathbf{1}_{\mathcal{J}_n} \sum_{\substack{j,k=1 \\ j \neq k}}^{N_n} \frac{1}{\sqrt{\lambda_j \lambda_k}} \pi_k (\Gamma_n - \Gamma) \pi_j + \sum_{j=1}^{N_n} \sum_{k \in N_n} \frac{\sqrt{\lambda_k}}{\sqrt{\lambda_j}(\lambda_j - \lambda_k)} \pi_k (\Gamma_n - \Gamma) \pi_j \\
&\quad + \sum_{j > N_n} \sum_{k=1}^{N_n} \frac{\sqrt{\lambda_j}}{\sqrt{\lambda_k}(\lambda_k - \lambda_j)} \pi_k (\Gamma_n - \Gamma) \pi_j \\
&= -\mathbf{1}_{\mathcal{J}_n} \sum_{\substack{j,k=1 \\ j \neq k}}^{N_n} \frac{\pi_k}{\sqrt{\lambda_k}} \Gamma_n \frac{\pi_j}{\sqrt{\lambda_j}} + \sum_{j=1}^{N_n} \left(s_j \Gamma_n \frac{\pi_j}{\sqrt{\lambda_j}} + \frac{\pi_j}{\sqrt{\lambda_j}} \Gamma_n s_j \right) = T_1' + T_1'',
\end{aligned}$$

where $s_j := \sum_{k > N_n} \frac{\sqrt{\lambda_k}}{\lambda_j - \lambda_k} \pi_k$. The term Γ disappears because $\pi_j \Gamma \pi_k = 0$ if $j \neq k$.

We control separately the operators T_1' and T_1'' . We have:

$$\|T_1'\|_\infty^2 \leq \sum_{p,q \geq 1} \langle T_1' \psi_p, \psi_q \rangle^2 = \sum_{\substack{p,q=1 \\ p \neq q}}^{N_n} \left(\frac{1}{n} \sum_{i=1}^n \xi_p^{(i)} \xi_q^{(i)} \right)^2,$$

where we recall that $\xi_p^{(i)} = \langle X_i, \psi_p \rangle / \sqrt{\lambda_p}$. Then

$$\begin{aligned}
\mathbf{P}(\|T_1'\|_\infty > (\rho_0 - 1)/4) &\leq \mathbf{P} \left(\sum_{\substack{p,q=1 \\ p \neq q}}^{N_n} \left(\frac{1}{n} \sum_{i=1}^n \xi_p^{(i)} \xi_q^{(i)} \right)^2 > \frac{(\rho_0 - 1)^2}{16} \right) \\
&\leq \sum_{\substack{p,q=1 \\ p \neq q}}^{N_n} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_p^{(i)} \xi_q^{(i)} \right| > \frac{\rho_0 - 1}{4N_n} \right).
\end{aligned}$$

Assumption **H2** and the Cauchy-Schwartz Inequality implies that, for all $p, q \geq 1$,

$$\mathbf{E} [|\xi_p^{(i)} \xi_q^{(i)}|^m] \leq m! b^{m-1}.$$

Then Lemma 10 and the condition $N_n \leq 20\sqrt{n/\ln^3 n}$ implies

$$\mathbf{P} \left(\|T_1'\|_\infty > \frac{\rho_0 - 1}{4} \right) \leq 2N_n^2 \exp \left(-C_1' \frac{n}{N_n^2} \right) \leq C_2' n^{-6},$$

with $C'_1 = \frac{(\rho_0-1)^2}{32(2b+(\rho_0-1)/4)}$ and C'_2 depends only on b and ρ_0 .

We deal now with the operator T''_1 , we can rewrite it like an array of independent random variables with values in \mathcal{H} , the set of the Hilbert-Schmidt operators of $\mathbf{L}^2([0, 1])$ equipped with the usual norm $\|T\|_{HS}^2 = \sum_{p,q \geq 1} \langle T\psi_p, \psi_q \rangle^2$, i.e.

$$T''_1 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{N_n} \left(s_j X_i \otimes \frac{\pi_j}{\sqrt{\lambda_j}} X_i + \frac{\pi_j}{\sqrt{\lambda_j}} X_i \otimes s_j X_i \right) = \frac{1}{n} \sum_{i=1}^n Z_i,$$

where, for all $f, g, h \in \mathbf{L}^2([0, 1])$, $f \otimes g h := \langle f, h \rangle g$. In order to apply the exponential inequality for centred Hilbert valued random variable given in Bosq [8, Theorem 2.5], we have to find two constants B and c such that

$$\mathbf{E} [\|Z_i\|_{HS}^m] \leq (m!/2) B^2 c^{m-2}.$$

We compute first $\|Z_i\|_{HS}^2$

$$\begin{aligned} \|Z_i\|_{HS}^2 &\leq 2 \sum_{p,q \geq 1} \left\langle \sum_{j=1}^{N_n} \left(\frac{\pi_j}{\sqrt{\lambda_j}} X_i \otimes s_j X_i \right) \psi_p, \psi_q \right\rangle^2 \\ &= 2 \sum_{p=1}^{N_n} \sum_{q > N_n} \frac{\lambda_q^2}{(\lambda_p - \lambda_q)^2} (\xi_p^{(i)} \xi_q^{(i)})^2. \end{aligned}$$

Now, applying iteratively the Hölder's Inequality, by Assumption **H2**, for all $p_1, \dots, p_m \geq 1$,

$$\mathbf{E}[\xi_{p_1}^{(i)} \dots \xi_{p_m}^{(i)}] \leq \prod_{j=1}^m \mathbf{E}[\xi_{p_j}^{2m}]^{1/m} \leq m! b^{m-1}.$$

Then

$$\begin{aligned} \mathbf{E} [\|Z_i\|_{HS}^m] &\leq \mathbf{E} [\|Z_i\|_{HS}^{2m}]^{1/2} \\ &= 2^m \left(\sum_{p_1, \dots, p_m=1}^{N_n} \sum_{q_1, \dots, q_m > N_n} \prod_{j=1}^m \frac{\lambda_{q_j}^2}{(\lambda_{p_j} - \lambda_{q_j})^2} \mathbf{E}[(\xi_{p_1}^{(i)} \dots \xi_{p_m}^{(i)})^2] \mathbf{E}[(\xi_{q_1}^{(i)} \dots \xi_{q_m}^{(i)})^2] \right)^{1/2} \\ &\leq m! 2^m b^m \left(\sum_{p=1}^{N_n} \sum_{q > N_n} \frac{\lambda_q^2}{(\lambda_p - \lambda_q)^2} \right)^{m/2}. \end{aligned}$$

We apply then Theorem 2.5 of Bosq [8] with $B^2 = 8b^2 \sum_{p=1}^{N_n} \sum_{q>N_n} \frac{\lambda_q^2}{(\lambda_p - \lambda_q)^2}$ and $c = 2b \left(\sum_{p=1}^{N_n} \sum_{q>N_n} \frac{\lambda_q^2}{(\lambda_p - \lambda_q)^2} \right)^{1/2}$ and obtain, with the condition $N_n \leq \frac{20}{\sqrt{n/\ln^3 n}}$,

$$\begin{aligned} \mathbf{P} \left(\|T_1''\|_\infty > \frac{\rho_0 - 1}{4} \right) &\leq \mathbf{P} \left(\|T_1''\|_{HS} > \frac{\rho_0 - 1}{4} \right) \\ &\leq 2 \exp \left(-C_1'' \frac{n}{N_n} \right) \leq C_2'' n^{-6}; \end{aligned}$$

where $C_1'' := (\rho_0 - 1)^2 / (8\delta b^2 + 2\sqrt{2\delta}b(\rho_0 - 1)/4)$, C_2'' depends only on ρ_0 and δ where $\delta > 0$ depends only on the sequence $(\lambda_j)_{j \geq 1}$ and verifies, for all $p \leq N_n$,

$$\sum_{q>N_n} \frac{\lambda_q^2}{(\lambda_p - \lambda_q)^2} \leq N_n \delta / 2.$$

Appendix C Control of empirical penalty in the unknown variance case

Lemma 17. *Under Assumption **H1**, set $\kappa := 2\theta(1 + 2\delta)$, we have, for all $m \in \widehat{\mathcal{M}}_n$*

$$\mathbf{E}_{\mathbf{X}} \left[(\widehat{\text{pen}}(m) - \text{pen}^{(uv)}(m)) \right] \leq \kappa \frac{m}{n} \|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2, \quad (40)$$

and

$$\begin{aligned} \mathbf{E}_{\mathbf{X}} \left[(\text{pen}^{(uv)}(\hat{m}^{(uv)}) - \widehat{\text{pen}}(\hat{m}^{(uv)})) \right] &\leq \frac{\kappa}{n} \mathbf{E}_{\mathbf{X}} \left[2\hat{m}^{(uv)} \nu_n(\hat{\Pi}_m \beta - \hat{\beta}_{\hat{m}^{(uv)}}) \right] \\ &\quad + \frac{\kappa \hat{N}_n}{n\sqrt{n}} \left(\sqrt{\text{Var}(\varepsilon^2)} + 2\|\beta\|_{\Gamma} \sigma \right). \quad (41) \end{aligned}$$

PROOF. By definitions of $\hat{\sigma}_m^2 = \gamma_n(\hat{\beta}_m)$ and $\hat{\beta}_m$, we have $\hat{\sigma}_m^2 \leq \gamma_n(\hat{\Pi}_m \beta)$, then

$$\mathbf{E}_{\mathbf{X}} \left[\widehat{\text{pen}}(m) - \text{pen}^{(uv)}(m) \right] = \kappa \frac{m}{n} \mathbf{E}_{\mathbf{X}} \left[\hat{\sigma}_m^2 - \sigma^2 \right] \leq \kappa \frac{m}{n} \mathbf{E}_{\mathbf{X}} \left[\gamma_n(\hat{\Pi}_m \beta) - \sigma^2 \right].$$

Now, by independence of ε_i with $\langle \beta - \hat{\Pi}_m \beta, X_i \rangle$,

$$\begin{aligned} \mathbf{E}_{\mathbf{X}}[\gamma_n(\hat{\Pi}_m \beta) - \sigma^2] &= \mathbf{E}_{\mathbf{X}} \left[\frac{1}{n} \sum_{i=1}^n (Y_i - \langle \hat{\Pi}_m \beta, X_i \rangle)^2 - \sigma^2 \right] \\ &= \mathbf{E}_{\mathbf{X}} \left[\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 - 2\varepsilon_i \langle \beta - \hat{\Pi}_m \beta, X_i \rangle + \langle \beta - \hat{\Pi}_m \beta, X_i \rangle^2 - \sigma^2 \right] \\ &= \mathbf{E}_{\mathbf{X}}[\langle \beta - \hat{\Pi}_m \beta, X_i \rangle^2] = \mathbf{E}_{\mathbf{X}}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2], \end{aligned}$$

and Equation (40) follows.

Likewise:

$$\begin{aligned} \mathbf{E}_{\mathbf{X}}[\text{pen}^{(uv)}(\hat{m}^{(uv)}) - \widehat{\text{pen}}(\hat{m}^{(uv)})] &= \frac{\kappa}{n} \mathbf{E}_{\mathbf{X}}[\hat{m}^{(uv)}(\sigma^2 - \hat{\sigma}_{\hat{m}^{(uv)}}^2)] \\ &= \frac{\kappa}{n} \left(\mathbf{E}_{\mathbf{X}}[\hat{m}^{(uv)}(\sigma^2 - \tilde{\sigma}^2)] - \mathbf{E}_{\mathbf{X}}[\hat{m}^{(uv)}\|\beta - \hat{\beta}_{\hat{m}^{(uv)}}\|_{\Gamma_n}^2] \right. \\ &\quad \left. + 2\mathbf{E}_{\mathbf{X}}[\hat{m}^{(uv)}\nu_n(\beta - \hat{\beta}_{\hat{m}^{(uv)}})] \right) \\ &\leq \frac{\kappa}{n} \left(\mathbf{E}_{\mathbf{X}}[\hat{m}^{(uv)}(\sigma^2 - \tilde{\sigma}^2)] + 2\mathbf{E}_{\mathbf{X}}[\hat{m}^{(uv)}\nu_n(\beta - \hat{\beta}_{\hat{m}^{(uv)}})] \right) \\ &\leq \frac{\kappa}{n} \left(\mathbf{E}_{\mathbf{X}}[\hat{m}^{(uv)}(\sigma^2 - \tilde{\sigma}^2)] + 2\mathbf{E}_{\mathbf{X}}[\hat{m}^{(uv)}\nu_n(\beta - \hat{\Pi}_m \beta) \right. \\ &\quad \left. + \hat{m}^{(uv)}\nu_n(\hat{\Pi}_m \beta - \hat{\beta}_{\hat{m}^{(uv)}})] \right) \end{aligned}$$

with $\tilde{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$. By Cauchy-Schwarz's Inequality

$$\begin{aligned} \mathbf{E}_{\mathbf{X}}[\hat{m}^{(uv)}(\sigma^2 - \tilde{\sigma}^2)] &\leq \hat{N}_n \mathbf{E}_{\mathbf{X}}[(\sigma^2 - \tilde{\sigma}^2)^2]^{1/2} \\ &= \hat{N}_n \left(\frac{1}{n^2} \sum_{i,j=1}^n \mathbf{E}_{\mathbf{X}}[(\varepsilon_i^2 - \sigma^2)(\varepsilon_j^2 - \sigma^2)] \right)^{1/2} = \hat{N}_n \sqrt{\frac{\text{Var}(\varepsilon_i^2)}{n}}, \end{aligned}$$

and, since the ε_i 's are independent of the X_i 's and by consequence of $\hat{\Pi}_m$, we have:

$$\begin{aligned} \mathbf{E}_{\mathbf{X}}[\hat{m}\nu_n(\beta - \hat{\Pi}_m \beta)] &\leq \hat{N}_n \mathbf{E}_{\mathbf{X}}[\nu_n^2(\beta - \hat{\Pi}_m \beta)]^{1/2} \\ &\leq \frac{\hat{N}_n}{n} \left(\sum_{i_1, i_2=1}^n \mathbf{E}_{\mathbf{X}}[\varepsilon_{i_1} \varepsilon_{i_2} \langle \beta - \hat{\Pi}_m \beta, X_{i_1} \rangle \langle \beta - \hat{\Pi}_m \beta, X_{i_2} \rangle] \right)^{1/2} \\ &\leq \frac{\hat{N}_n}{\sqrt{n}} \sigma \mathbf{E}[\|\beta - \hat{\Pi}_m \beta\|_{\Gamma_n}^2]^{1/2} \leq \frac{\hat{N}_n}{\sqrt{n}} \sigma \|\beta\|_{\Gamma}. \end{aligned}$$

References

- [1] R. B. ASH AND M. F. GARDNER, *Topics in stochastic processes*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975. Probability and Mathematical Statistics, Vol. 27.
- [2] Y. BARAUD, *Model selection for regression on a fixed design*, Probab. Theory Related Fields, 117 (2000), pp. 467–493.
- [3] Y. BARAUD, *Model selection for regression on a random design*, ESAIM Probab. Statist., 6 (2002), pp. 127–146 (electronic).
- [4] A. BARRON, L. BIRGÉ, AND P. MASSART, *Risk bounds for model selection via penalization*, Probab. Theory and Related Fields, 113 (1999), pp. 301–413.
- [5] J.-P. BAUDRY, C. MAUGIS, AND B. MICHEL, *Slope heuristics: overview and implementation*, Stat. Comput., 22 (2012), pp. 455–470.
- [6] L. BIRGÉ AND P. MASSART, *Minimum contrast estimators on sieves: exponential bounds and rates of convergence*, Bernoulli, 4 (1998), pp. 329–375.
- [7] L. BIRGÉ AND P. MASSART, *Minimal penalties for Gaussian model selection*, Probab. Theory Related Fields, 138 (2007), pp. 33–73.
- [8] D. BOSQ, *Linear processes in function spaces: theory and applications*, Springer, July 2000.
- [9] V. BRAULT, J.-P. BAUDRY, C. MAUGIS, AND B. MICHEL, *capushe: Capushe, data-driven slope estimation and dimension jump*, 2012. R package version 1.0.
- [10] E. BRUNEL AND A. ROCHE, *Penalized contrast estimation in functional linear models with circular data*. To appear in *Statistics*, 2015.
- [11] T. T. CAI AND P. HALL, *Prediction in functional linear regression*, Ann. Statist., (2006), pp. 2159–2179.
- [12] T. T. CAI AND M. YUAN, *Minimax and adaptive prediction for functional linear regression*, J. American Statistical Association, 107 (2012), pp. 1201–1216.

- [13] H. CARDOT, F. FERRATY, AND P. SARDA, *Functional linear model*, Statist. Probab. Lett., 45 (1999), pp. 11–22.
- [14] ———, *Spline estimators for the functional linear model*, Statist. Sinica, 13 (2003), pp. 571–591.
- [15] H. CARDOT AND J. JOHANNES, *Thresholding projection estimators in functional linear models*, J. Multivariate Anal., 101 (2010), pp. 395–408.
- [16] H. CARDOT, A. MAS, AND P. SARDA, *CLT in functional linear regression models*, Probab. Theory Related Fields, 138 (2007), pp. 325–361.
- [17] F. COMTE AND J. JOHANNES, *Adaptive estimation in circular functional linear models*, Math. Methods Statist., 19 (2010), pp. 42–63.
- [18] F. COMTE AND J. JOHANNES, *Adaptive functional linear regression*, Ann. Statist., 40 (2012), pp. 2765–2797.
- [19] C. CRAMBES, A. KNEIP, AND P. SARDA, *Smoothing splines estimators for functional linear regression*, Ann. Statist., 37 (2009), pp. 35–72.
- [20] C. CRAMBES AND A. MAS, *Asymptotics of prediction in functional linear regression with functional outputs*, Bernoulli, 19 (2013), pp. 2627–2651.
- [21] N. DUNFORD AND J. T. SCHWARTZ, *Linear Operators. I. General Theory*, With the assistance of W. G. Bade and R. G. Bartle. Pure and Applied Mathematics, Vol. 7, Interscience Publishers, Inc., New York, 1958.
- [22] P. H. C. EILERS AND B. D. MARX, *Flexible smoothing with B-splines and penalties*, Statist. Sci., 11 (1996), pp. 89–121. With comments and a rejoinder by the authors.
- [23] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of inverse problems*, vol. 375 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1996.
- [24] F. FERRATY AND P. VIEU, *Nonparametric functional data analysis*, Springer Series in Statistics, Springer, New York, 2006. Theory and practice.

- [25] A. GOLDENSHLUGER AND O. LEPSKI, *Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality*, Ann. Statist., 39 (2011), pp. 1608–1632.
- [26] P. HALL AND J. L. HOROWITZ, *Methodology and convergence rates for functional linear regression*, Ann. Statist., 35 (2007), pp. 70–91.
- [27] P. HALL AND M. HOSSEINI-NASAB, *On properties of functional principal components analysis*, J. R. Stat. Soc. Ser. B Stat. Methodol., 68 (2006), pp. 109–126.
- [28] N. HILGERT, A. MAS, AND N. VERZELEN, *Minimax adaptive tests for the functional linear model*, Ann. Statist., 41 (2013), pp. 838–869.
- [29] J. JOHANNES AND R. SCHENK, *On rate optimal local estimation in functional linear regression*, Electron. J. Stat., 7 (2013), pp. 191–216.
- [30] E. R. LEE AND B. U. PARK, *Sparse estimation in functional linear regression*, J. Multivariate Anal., 105 (2012), pp. 1–17.
- [31] Y. LI AND T. HSING, *On rates of convergence in functional linear regression*, J. Multivariate Anal., 98 (2007), pp. 1782–1804.
- [32] C. L. MALLOWS, *Some comments on cp*, Technometrics, (1973), pp. 661–675.
- [33] A. MAS AND F. RUYMGAART, *High-dimensional principal projections*, Complex Anal. Oper. Theory, 9 (2015), pp. 35–63.
- [34] P. MASSART, *Concentration inequalities and model selection*, vol. 1896 of Lecture Notes in Mathematics, Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.
- [35] C. PREDÀ AND G. SAPORTA, *PLS regression on a stochastic process*, Comput. Statist. Data Anal., 48 (2005), pp. 149–158.
- [36] J. RAMSAY AND C. DALZELL, *Some tools for functional data analysis*, J. Roy. Statist. Soc. Ser. B, 53 (1991), pp. 539–572. With discussion and a reply by the authors.
- [37] J. RAMSAY AND B. SILVERMAN, *Functional Data Analysis*, Springer Series in Statistics, Springer, 2 ed., 2005.