Adaptive estimation in circular functional linear models.

Fabienne Comte, Jan Johannes

Adaptive estimation in circular functional linear models.

Fabienne Comte* Jan Johannes⋆

August 24, 2009

Abstract

We consider the problem of estimating the slope parameter in circular functional linear regression, where scalar responses $Y_1, \dots, Y_n$ are modeled in dependence of 1-periodic, second order stationary random functions $X_1, \dots, X_n$. We consider an orthogonal series estimator of the slope function β, by replacing the first m theoretical coefficients of its development in the trigonometric basis by adequate estimators. We propose a model selection procedure for m in a set of admissible values, by defining a contrast function minimized by our estimator and a theoretical penalty function; this first step assumes the degree of ill posedness to be known. Then we generalize the procedure to a random set of admissible m's and a random penalty function. The resulting estimator is completely data driven and reaches automatically what is known to be the optimal minimax rate of convergence, in term of a general weighted L^2-risk. This means that we provide adaptive estimators of both β and its derivatives.

Keywords: Orthogonal series estimation; model selection; derivatives estimation; mean squared error of prediction; minimax theory.

AMS 2000 subject classifications: Primary 62G05; secondary 62J05, 62G08.

1 Introduction

Functional linear models have become very important in a diverse range of disciplines, including medicine, linguistics, chemometrics as well as econometrics (see for instance Ramsay and Silverman [2005] and Ferraty and Vieu [2006], for several case studies, or more specific, Forni and Reichlin [1998] and Preda and Saporta [2005] for applications in economics). Roughly speaking, in all these applications the dependence of a response variable Y on the variation of an explanatory random function X is modeled by

$$Y = \int_0^1 \beta(t)X(t)dt + \sigma\varepsilon, \quad \sigma > 0, \tag{1.1}$$

for some error term ε. One objective is then to estimate nonparametrically the slope function β based on an independent and identically distributed (i.i.d.) sample of (Y, X).

*Université Paris Descartes, Laboratoire MAP5, UMR CNRS 8145, 45, rue des Saints-Pères, F-75270 Paris cedex 06, France, e-mail: fabienne.comte@parisdescartes.fr

⋆Universität Heidelberg, Institut für Angewandte Mathematik, Im Neuenheimer Feld, 294, D-69120 Heidelberg, Germany, e-mail: johannes@statlab.uni-heidelberg.de

In this paper we suppose that the random function X is taking its values in $L^2[0,1]$, which is endowed with the usual inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|$, and that X has a finite second moment, i.e., $\mathbb{E}\|X\|^2 < \infty$. In order to simplify notations we assume that the mean function of X is zero. Moreover, the random function X and the error term ε are uncorrelated, where ε is assumed to have mean zero and variance one. This situation has been considered, for example, in Cardot et al. [2003], Müller and Stadtmüller [2005] or most recently James et al. [2009]. Then multiplying both sides in (1.1) by $X(s)$ and taking the expectation leads to

$$g(s) := \mathbb{E}[YX(s)] = \int_0^1 \beta(t)\operatorname{cov}(X(t),X(s))dt =: [\Gamma\beta](s), \quad s \in [0,1], \tag{1.2}$$

where g belongs to $L^2[0,1]$ and Γ denotes the covariance operator associated to the random function X. We shall assume that there exists a unique solution $\beta \in L^2[0,1]$ of equation (1.2). Estimation of β is thus linked with the inversion of the covariance operator Γ and, known to be an ill-posed inverse problem (for a detailed discussion in the context of inverse problems see chapter 2.1 in Engl et al. [2000], while in the special case of a functional linear model we refer to Cardot et al. [2003]).

In this paper we consider a circular functional linear model (defined below), where the associated covariance operator Γ admits a spectral decomposition $\{\lambda_j, \varphi_j, j \geqslant 1\}$ given by the trigonometric basis $\{\varphi_j\}$ as eigenfunctions and a strictly positive, possibly not ordered, zero-sequence $\lambda := (\lambda_j)_{j\geqslant 1}$ of corresponding eigenvalues. Then the normal equation can be rewritten as follows

$$\beta = \sum_{j=1}^{\infty} \frac{[g]_j}{\lambda_j}\cdot\varphi_j \quad \text{with } [g]_j := \langle g, \varphi_j\rangle,\ j \geqslant 1. \tag{1.3}$$

For estimation purpose, we replace the unknown quantities g_j and λ_j in equation (1.3) by their empirical counterparts. That is, if $(Y_1, X_1), \dots, (Y_n, X_n)$ denotes an i.i.d. sample of (Y, X), then for each $j \geqslant 1$, we consider the unbiased estimator

$$[\widehat{g}]_j := \frac{1}{n}\sum_{i=1}^n Y_i\,[X_i]_j, \quad \text{and} \quad \widehat{\lambda}_j := \frac{1}{n}\sum_{i=1}^n [X_i]_j^2 \quad \text{with } [X_i]_j := \langle X_i, \varphi_j\rangle$$

for $[g]_j$ and λ_j respectively. The orthogonal series estimator $\widehat{\beta}_m$ of β is then defined by

$$\widehat{\beta}_m := \sum_{j=1}^m \frac{\widehat{g}_j}{\widehat{\lambda}_j}\cdot \mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\}\cdot\varphi_j. \tag{1.4}$$

Note that we introduce an additional threshold $1/n$ on each estimated eigenvalue $\widehat{\lambda}_j$, since it could be arbitrarily close to zero even in case that the true eigenvalue λ_j is sufficiently far away from zero. Moreover, the orthogonal series estimator keeps only m coefficients; this is an alternative to the popular Tikhonov regularization (c.f. Hall and Horowitz [2007]), where in (1.3) the factor $1/\lambda_j$ is replaced by $\lambda_j/(\alpha+\lambda_j^2)$. Thresholding in the Fourier domain has been used, for example, in a deconvolution problem in Mair and Ruymgaart [1996] or Neumann [1997] and coincides with an approach called spectral cut-off in the numerical analysis literature (c.f. Tautenhahn [1996]).

In this paper we shall measure the performance of an estimator $\widehat{\beta}$ of β by the $\mathcal{F}_\omega$-risk, that is $\mathbb{E}\|\widehat{\beta}-\beta\|_\omega^2$, where for some strictly positive sequence of weights $\omega := (\omega_j)_{j\geqslant 1}$

$$\|f\|_\omega^2 := \sum_{j=1}^{\infty} \omega_j |\langle f, \varphi_j\rangle|^2 \qquad \text{for all } f \in L^2[0,1].$$

This general framework allows us with appropriate choices of the weight sequence ω to cover the estimation not only of the slope parameter itself (c.f. Hall and Horowitz [2007]) but also of its derivatives as well as the optimal estimation with respect to the mean squared prediction error (c.f. Cardot et al. [2003] or Crambes et al. [2009]). For a more detailed discussion, we refer to Cardot and Johannes [2009]. It is well-known that the obtainable accuracy of any estimator in terms of the $\mathcal{F}_\omega$-risk is essentially determined by the regularity conditions imposed on both the slope parameter β and the eigenvalues λ. In the literature the a-priori information on the slope parameter β such as smoothness is often characterized by considering ellipsoids (see definition below) in $L^2[0,1]$ with respect to a weighted norm $\|\cdot\|_\gamma$ for a pre-specified weight sequence γ. Moreover, it is usually assumed that the sequence λ of eigenvalues of Γ has a polynomial decay (c.f. Hall and Horowitz [2007] or Crambes et al. [2009]). However, it is well-known that this restriction may exclude several interesting cases, such as an exponential decay. Therefore, we do not impose a specific form of a decay.

It is shown in Johannes [2009] that the estimator $\widehat{\beta}_m$ given in (1.4) is optimal in a minimax sense if the parameter $m = m(n)$ is appropriately chosen. Roughly speaking, the introduction of a dimension reduction implies a bias in addition to the classical variance term which leads the statistician to perform a compromise. The optimal choice of the dimension parameter m requires an a-priori knowledge about the sequences γ and λ, which is unknown in practice. However, useful elements of this previous work are recalled in Section 2.

Our aim in this paper, is to provide a data driven method to select the dimension parameter m, in such a way that the bias and variance compromise is automatically reached by the resulting estimator. The methodology is inspired by the works of Barron et al. [1999], now extensively described in Massart [2007] whose results, like ours, are in a non asymptotic setting. By re-writing the estimator $\widehat{\beta}_m$ as a minimum contrast estimator over the function space S_m – called model – linearly spanned by $\varphi_1, \dots, \varphi_m$, we can propose a model selection device by defining a penalty function. We obtain a selected $\hat{m}$ in an admissible set of values of m. We first define and study in Section 3, the resulting estimator $\hat{\beta}_{\hat{m}}$ with deterministic penalty and deterministic set of admissible m's: this requires to assume that the degree of ill-posedness of the problem is known. In other words, information are first supposed to be available about the order of the decay of the eigenvalues λ_j. This study gives the tools to the next and final step: we define in Section 4 a completely data driven estimator, built by using a random penalty function and a random set of admissible dimensions m. We can provide a general risk bound for this estimator and show that it can automatically reach the optimal rate of convergence, without requiring any a-priori knowledge. All proofs are gathered in the Appendix section.

2 Background to the methodology.

2.1 Notations and basic assumptions

Circular functional linear model. In this paper we suppose that the regressor X is 1-periodic, that is $X(0) = X(1)$, and second order stationary, i.e., there exists a positive definite covariance function $c : [-1, 1] \to \mathbb{R}$ such that $\operatorname{cov}(X(t), X(s)) = c(t-s)$, $s, t \in [0, 1]$. Then it is straightforward to see that the covariance function $c(\cdot)$ is 1-periodic too. In this situation applying the covariance operator Γ equals a convolution with the covariance function. Since $c(\cdot)$ is 1-periodic it is easily seen that due to the classical convolution theorem, the eigenfunctions of the covariance operator Γ are given by the trigonometric basis

$$\varphi_1(s) :\equiv 1,\ \varphi_{2k}(s) := \sqrt{2}\cos(2\pi ks),\ \varphi_{2k+1}(s) := \sqrt{2}\sin(2\pi ks), s \in [0, 1],\ k \geqslant 1$$

and the corresponding eigenvalues satisfy

$$\lambda_1 = \int_0^1 c(s)ds, \quad \lambda_{2k} = \lambda_{2k+1} = \int_0^1 \cos(2\pi ks)c(s)ds,\ k \geqslant 1.$$

Notice that the eigenfunctions are known to the statistician and only the eigenvalues depend on the unknown covariance function $c(\cdot)$, i.e., have to be estimated.

Moment assumptions. The results derived below involve additional conditions on the moments of the random function X and the error term ε, which we formalize now. Let $\mathcal{X}$ be the set of all centered 1-periodic and second order stationary random functions $X \in L^2[0, 1]$ with finite second moment, i.e., $\mathbb{E}\|X\|^2 < \infty$, and strictly positive covariance operator Γ. If $\lambda := (\lambda_j)_{j\geqslant 1}$ denotes the sequence of eigenvalues associated to Γ, then given $X \in \mathcal{X}$ the random variables $\{[X]_j/\sqrt{\lambda_j}, j \in \mathbb{N}\}$ are centered with variance one. Here and subsequently, we denote by $\mathcal{X}^k_\eta$, $k \in \mathbb{N}$, $\eta \geqslant 1$, the subset of $\mathcal{X}$ containing only random functions X such that the k-th moment of the corresponding random variables $[X]_j/\sqrt{\lambda_j}$, $j \in \mathbb{N}$ are uniformly bounded, that is

$$\mathcal{X}^k_\eta := \Big\{X \in \mathcal{X} \ \text{ such that } \quad \sup_{j\in\mathbb{N}} \mathbb{E}\Big|[X]_j/\sqrt{\lambda_j}\Big|^k \leqslant \eta\Big\}.$$

It is worth noting that in case $X \in \mathcal{X}$ is a Gaussian random function the corresponding random variables $[X]_j/\sqrt{\lambda_j}$, $j \in \mathbb{N}$, are Gaussian with mean zero and variance one. Hence, if $\eta \geqslant 3$ then any Gaussian random function $X \in \mathcal{X}$ belongs also to $\mathcal{X}^k_\eta$ for each $k \in \mathbb{N}$.

Minimal regularity conditions. Given a strictly positive sequence of weights $w := (w_j)_{j\geqslant 1}$, denote by $\mathcal{F}^c_w$ the ellipsoid with radius $c > 0$, that is,

$$\mathcal{F}^c_w := \Big\{f \in L^2[0, 1] : \sum_{j=1}^{\infty} w_j|\langle f, \varphi_j\rangle|^2 =: \|f\|^2_w \leqslant c\Big\}.$$

Furthermore, let $\mathcal{F}_w := \{f \in L^2[0, 1] : \|f\|^2_w < \infty\}$ and $\langle f, g\rangle_w := \sum_{j=1}^\infty w_j\langle f, \varphi_j\rangle\langle\varphi_j, g\rangle$. Note that this weighted inner product induces the weighted norm $\|\cdot\|_w$.

Here and subsequently, given strictly positive sequences of weights $\gamma := (\gamma_j)_{j\geqslant 1}$ and $\omega := (\omega_j)_{j\geqslant 1}$ we shall measure the performance of any estimator $\widehat{\beta}$ by its maximal $\mathcal{F}_\omega$-risk over the ellipsoid $\mathcal{F}^\rho_\gamma$ with radius $\rho > 0$, that is $\sup_{\beta\in\mathcal{F}^\rho_\gamma} \mathbb{E}\|\widehat{\beta} - \beta\|^2_\omega$. We do not specify the sequences of weights γ and ω, but impose from now on the following minimal regularity conditions.

Assumption 2.1. *Let $\omega := (\omega_j)_{j\geqslant 1}$ and $\gamma := (\gamma_j)_{j\geqslant 1}$ be positive sequences of weights with $\omega_1 = 1$ and $\gamma_1 = 1$ such that $(1/\gamma_j)_{j\geqslant 1}$ and $(\omega_j/\gamma_j)_{j\geqslant 1}$ are non increasing zero-sequences.*

Note that under Assumption 2.1 the ellipsoid $\mathcal{F}_\gamma^\rho$ is a subset of $\mathcal{F}_\omega^\rho$, and hence the $\mathcal{F}_\omega$-risk a well-defined risk for β. Roughly speaking, if $\mathcal{F}_\gamma^\rho$ describes p-times differentiable functions, then the Assumption 2.1 ensures that the $\mathcal{F}_\omega$-risk involves maximal $s < p$ derivatives.

2.2 Minimax optimal estimation.

The objective of the paper is to construct an estimator which attains the minimal rate of convergence of the maximal $\mathcal{F}_\omega$-risk over the ellipsoid $\mathcal{F}_\gamma^\rho$ for wide range of sequences γ and ω satisfying Assumption 2.1, without using an a-priori knowledge of neither γ nor ρ. Therefore, let us first recall a lower bound which can be found in Johannes [2009]. Let $m^* := (m_n^*) \in \mathbb{N}$ for some $\triangle \geqslant 1$ be chosen such that

$$1/\triangle \leqslant \frac{\gamma_{m_n^*}}{n\,\omega_{m_n^*}} \sum_{j=1}^{m_n^*} \frac{\omega_j}{\lambda_j} \leqslant \triangle,$$

i.e. $(1/n)\sum_{j=1}^{m_n^*} \omega_j/\lambda_j$ and $\omega_{m_n^*}/\gamma_{m_n^*}$ have the same orders.

Given an i.i.d. n-sample of (Y, X) obeying (1.1) with $\sigma > 0$ and $X \in \mathcal{X}$ with associated sequence of eigenvalues λ, we have then for any estimator $\breve{\beta}$ that

$$\sup_{\beta \in \mathcal{F}_\gamma^\rho} \left\{ \mathbb{E}\|\breve{\beta} - \beta\|_\omega^2 \right\} \geqslant \frac{1}{4\,\triangle} \min\left(\frac{\sigma^2}{2}, \frac{\rho}{\triangle}\right) \max(\omega_{m_n^*}/\gamma_{m_n^*}, 1/n) \quad \text{for all } n \geqslant 1. \tag{2.1}$$

On the other hand consider the estimator $\widehat{\beta}_m$ defined in (1.4) with dimension parameter $m = m_n^*$. If in addition $X \in \mathcal{X}_\xi^{16}$, then it is shown in Johannes [2009] that there exists a numerical constant $C > 0$ such that

$$\sup_{\beta \in \mathcal{F}_\gamma^\rho} \left\{ \mathbb{E}\|\widehat{\beta}_{m_n^*} - \beta\|_\omega^2 \right\} \leqslant C\,\triangle^3\,\xi\,[\rho\,\mathbb{E}\|X\|^2 + \sigma^2]\,\max(\omega_{m_n^*}/\gamma_{m_n^*}, 1/n).$$

Therefore, the minimax-optimal rate of convergence is of order $O(\max(\omega_{m_n^*}/\gamma_{m_n^*}, 1/n))$. As a consequence, the orthogonal series estimator $\widehat{\beta}_{m_n^*}$ attains this optimal rate and hence is minimax-optimal. However, the definition of the dimension parameter m_n^* used to construct the estimator involves an a-priori knowledge of the sequences γ, ω and λ. Throughout the paper our aim is to construct a data-driven choice of the dimension parameter not requiring this a-priori knowledge and automatically attaining the optimal rate of convergence.

2.3 Example of rates

We compute in this section the rates that we can obtain in three configurations for the sequences γ, ω and λ. These cases will be referred to in the following. In all three cases, we take the sequence ω with $\omega_j = j^{2s}$, $j \geqslant 1$, for $s \in \mathbb{R}$.

Case [P-P] Polynomial-Polynomial. Consider sequences γ and λ with $\gamma_j = j^{2p}$, $j \geqslant 1$, for $p > \max(0, s)$, and $\lambda_j \asymp j^{-2a}$, $j \geqslant 1$, for $a > 1/2$ respectively, where the notation $u_j \asymp v_j$, $j \geqslant 1$, means that there exists a constant $d > 0$ such that $u_j/d \leqslant v_j \leqslant d u_j$ for all $j \geqslant 1$. Then it is easily seen that $(m_n^*)^{2(s-p)} = \frac{\omega_{m_n^*}}{\gamma_{m_n^*}} \asymp \sum_{j=1}^{m_n^*} \frac{\omega_j}{n v_j} \asymp n^{-1} \sum_{j=1}^{m_n^*} j^{2s+2a}$ and hence $m_n^* \asymp n^{1/(2p+2a+1)}$ if $2s + 2a + 1 > 0$, $m_n^* \asymp n^{1/[2(p-s)]}$ if $2s + 2a + 1 < 0$ and $m_n^* \asymp (n/\log(n))^{1/[2(p-s)]}$ if $2a + 2s + 1 = 0$. Finally, the optimal rate attained by the estimator is $\max(n^{-(2p-2s)/(2a+2p+1)}, n^{-1})$, if $2s+2a+1 \neq 0$ (and $\log(n)/n$ if $2s+2a+1=0$). Observe that an increasing value of a leads to a slower optimal rate of convergence. Therefore, the parameter a is called degree of ill-posedness (c.f. Natterer [1984]).

Remark 2.1. Obviously the rate is parametric if $2a+2s+1<0$. The case $0 \leqslant s < p$ can be interpreted as the L^2-risk of an estimator of the s-th derivative of the slope parameter β. On the other hand the case, $s=-a$, corresponds to the mean-prediction error (c.f. Cardot and Johannes [2009]). □

Case [E-P] Exponential-Polynomial. Consider sequences γ and λ with $\gamma_j = \exp(j^{2p})$, $j \geqslant 1$, for $p > 0$, and (as previously) $\lambda_j \asymp j^{-2a}$, $j \geqslant 1$, for $a > 1/2$ respectively. Then m_n^* is such that $\exp(-(m_n^*)^{2p})(m_n^*)^{2s} = \frac{\omega_{m_n^*}}{\gamma_{m_n^*}} \asymp \sum_{j=1}^{m_n^*} \frac{\omega_j}{n v_j} \asymp n^{-1} \sum_{j=1}^{m_n^*} j^{2s+2a}$. In case $2a+2s+1>0$ this is equivalent to $\exp(-(m_n^*)^{2p}) \asymp (m_n^*)^{2a+1} n^{-1}$ and hence $m_n^* \asymp (\log n - \frac{2a+1}{2p}\log(\log n))^{1/(2p)}$. Thereby, $n^{-1}(\log n)^{(2a+1+2s)/(2p)}$ is the optimal rate attained by the estimator. Furthermore, if $2a+2s+1<0$, then $m_n^* \asymp (\log(n)+(s/p)\log(\log(n)))^{1/(2p)}$ and the rate is parametric, while if $2a+2s+1=0$, the rate is of order $\log(\log(n))/n$.

Case [P-E] Polynomial-Exponential. Consider sequences γ and λ with $\gamma_j = j^{2p}$, $j \geqslant 1$, for $p > \max(0,s)$, and $\lambda_j \asymp \exp(-j^{2a})$, $j \geqslant 1$, for $a > 0$ respectively. Then $(m_n^*)^{2(s-p)} = \frac{\omega_{m_n^*}}{\gamma_{m_n^*}} \asymp \sum_{j=1}^{m_n^*} \frac{\omega_j}{n v_j} \asymp n^{-1} \sum_{j=1}^{m} j^{2s}\exp(j^{2a})$ and hence $m_n^* \asymp (\log n - \frac{2p+(2a-1)_{\vee 0}}{2a}\log(\log n))^{1/(2a)}$ with $(q)_{\vee 0} := \max(q,0)$. Thereby, $(\log n)^{-(p-s)/a}$ is the optimal rate attained by the estimator. The parameter a reflects again the degree of ill-posedness since an increasing value of a leads also here to a slower optimal rate of convergence.

3 A model selection approach: known degree of ill-posedness

In the previous section, we have recalled an estimation procedure that attains the optimal rate of convergence in case the slope parameter belongs to some ellipsoid $\mathcal{F}_\gamma^\rho$ and its accuracy is measured by a $\mathcal{F}_\omega$-risk. In this section, we suppose that there exists an a-priori knowledge concerning the degree of ill-posedness, that is the asymptotic behavior of the sequence of eigenvalues λ is known. The objective is the construction of an adaptive estimator which depends neither on the sequence of weights γ nor on the radius ρ but still attains the optimal rate over the ellipsoid $\mathcal{F}_\gamma^\rho$. In this section, we use the following assumption.

Assumption 3.1. *Let $\lambda := (\lambda_j)_{j\geqslant 1}$ denote the sequence of eigenvalues associated to the regressor X and let $\omega := (\omega_j)_{j\geqslant 1}$ be a sequence satisfying Assumption 2.1 such that*

(i) there exist non decreasing sequences $\delta := \delta(\lambda,\omega) := (\delta_m(\lambda,\omega))_{m\geqslant 1}$ and $\Delta := \Delta(\lambda,\omega) := (\Delta_m(\lambda,\omega))_{m\geqslant 1}$ with $\delta_m \geqslant \sum_{j=1}^m \omega_j/\lambda_j$ and $\Delta_m \geqslant \max_{1\leqslant j\leqslant m} \omega_j/\lambda_j$ for all $m \geqslant 1$ such that for some $\Sigma > 0$,

$$\sum_{m\geqslant 1} \Delta_m \exp\left(-\frac{\delta_m}{6\Delta_m}\right) \leqslant \Sigma. \tag{3.1}$$

(ii) the sequence $M := (M_n)_{n\geqslant 1}$ given by $M_n := \arg\max_{1\leqslant M\leqslant n}\{\delta_M \leqslant \delta_1 n(\omega_M)_{\wedge 1}\}$, $n \geqslant 1$, with

$(q)_{\wedge 1} := \min(q, 1)$, *satisfies*

$$\min_{1 \leqslant j \leqslant M_n} \lambda_j \geqslant 2/n \qquad \text{for all } n \geqslant 1. \tag{3.2}$$

It is worth to note that both sequences δ and M depend on the eigenvalues λ.

3.1 Definition of the estimator.

Consider the orthogonal series estimator $\widehat{\beta}_m$ defined in (1.4). In what follows we construct an adaptive procedure to choose the dimension parameter m based on a model selection approach. Therefore, let $\widehat{\Phi}_u = \sum_{j \geqslant 1} \widehat{\lambda}_j^{-1} \mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\}[u]_j \varphi_j$ for $u \in L^2[0,1]$ with Fourier coefficients $[u]_j := \langle u, \varphi_j \rangle$. Then we consider the contrast

$$\Upsilon(t) := \|t\|_\omega^2 - 2\langle t, \widehat{\Phi}_{\widehat{g}} \rangle_\omega. \tag{3.3}$$

Define $\mathcal{S}_m := \operatorname{span}\{\varphi_1, \dots, \varphi_m\}$. Obviously for all $t \in \mathcal{S}_m$ it follows that $\langle t, \widehat{\Phi}_{\widehat{g}} \rangle_\omega = \langle t, \widehat{\beta}_m \rangle_\omega$ and hence $\Upsilon(t) = \|t - \widehat{\beta}_m\|_\omega^2 - \|\widehat{\beta}_m\|_\omega^2$. Therefore, we have for all $m \geqslant 1$

$$\arg\min_{t \in \mathcal{S}_m} \Upsilon(t) = \widehat{\beta}_m.$$

Let $X \in \mathcal{X}_\eta^4$ and $\mathbb{E}|Y/\sigma_Y|^4 \leqslant \eta$ with $\sigma_Y^2 := \mathbb{V}\mathrm{ar}(Y)$. Under Assumption 3.1, we consider the penalty function

$$\operatorname{pen}(m) := 192 \sigma_Y^2 \eta \frac{\delta_m}{n}.$$

The adaptive estimator $\widehat{\beta}_{\widehat{m}}$ is obtained from (1.4) by choosing the dimension parameter

$$\widehat{m} := \arg\min_{1 \leqslant m \leqslant M_n} \left\{ \Upsilon(\widehat{\beta}_m) + \operatorname{pen}(m) \right\}. \tag{3.4}$$

Note that we can compute

$$\Upsilon(\widehat{\beta}_m) = -\sum_{j=1}^{m} \omega_j \frac{[\widehat{g}]_j^2}{\widehat{\lambda}_j^2} \mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\}.$$

Remark 3.1. Throughout the paper we ignore that also the value σ_Y^2 and η are unknown in practice. Obviously σ_Y^2 can be estimated straightforwardly by its empirical counterpart. An estimator of the value η is not a trivial task. However, if in addition the regressor X and the error term ε are Gaussian, then $Y \sim \mathcal{N}(0, \sigma_Y^2)$ and hence $\eta = 3$ is a-priori known. We may take an other point of view if we chose a-priori a sufficiently large $\eta \geqslant 3$ (the Gaussian case is included) then the following assertions apply as long as the unknown data generating process satisfies the conditions $X \in \mathcal{X}_\eta^4$ and $\mathbb{E}|Y/\sigma_Y|^4 \leqslant \eta$. □

3.2 An upper bound.

We derive first an upper bound of the adaptive estimator $\widehat{\beta}_{\widehat{m}}$ by assuming an a-priori knowledge of appropriate sequences δ and M which are used in the construction of the penalty and the admissible set of values of m.

Theorem 3.1. *Assume an n-sample of (Y,X) satisfying (1.1). Let $\mathbb{E}|Y/\sigma_Y|^4 \leqslant \eta$ and $X \in \mathcal{X}_\eta^4$ be 1-periodic and second order stationary with associated eigenvalues λ.*

Suppose that the sequences γ and ω satisfy Assumption 2.1. Let δ, $\triangle$ and M be sequences satisfying Assumption 3.1 for some constant Σ. Consider the estimator $\widehat{\beta}_{\widehat{m}}$ defined in (1.4) with $\widehat{m}$ given by (3.4). If in addition $X \in \mathcal{X}_\xi^{24}$ and $\mathbb{E}|Y/\sigma_Y|^{24} \leqslant \xi$, then there exists a numerical constant C such that for all $n \geqslant 1$ and $1 \leqslant m \leqslant M_n$, we have

$$\sup_{\beta\in\mathcal{F}_\gamma^\rho}\left\{\mathbb{E}\|\widehat{\beta}_{\widehat{m}}-\beta\|_\omega^2\right\} \leqslant C\Big\{\frac{\omega_m}{\gamma_m}\,\rho+\frac{\delta_m}{n}\,(\rho\mathbb{E}\|X\|^2+\sigma^2)\eta\Big\}$$
$$+\frac{K}{n}\,(\rho\mathbb{E}\|X\|^2+\sigma^2)\,[\delta_1+\rho][1+(\mathbb{E}\|X\|^2)^2],$$

where $K = K(\Sigma,\eta,\xi,\delta_1)$ is a constant depending on Σ,η,ξ and δ_1 only.

It is worth noting, that in the last assertion we do not impose a complete knowledge of the sequence of eigenvalues λ associated to the regressor X. In the next Corollary we state the upper bound when balancing the terms depending on m, which is obviously a trivial consequence of Theorem 3.1.

Corollary 3.2. *Let the assumptions of Theorem 3.1 be satisfied. If in addition the sequence $m^\diamond := (m_n^\diamond)_{n\geqslant 1}$ is chosen such that $\gamma_{m_n^\diamond}\delta_{m_n^\diamond}/(n\,\omega_{m_n^\diamond}) \asymp 1$, $n \geqslant 1$, then we have*

$$\sup_{\beta\in\mathcal{F}_\gamma^\rho}\left\{\mathbb{E}\|\widehat{\beta}_{\widehat{m}}-\beta\|_\omega^2\right\} = O\Big(\max(\omega_{m_n^\diamond}/\gamma_{m_n^\diamond}, 1/n)\Big) \ \textit{as } n\to\infty.$$

Remark 3.2. Comparing the last assertion with the lower bound given in (2.1), we see that the adaptive estimator attains the optimal rate of convergence, as long as $\sup_{n\geqslant 1}\omega_{m_n^\diamond}\gamma_{m_n^*}/(\gamma_{m_n^\diamond}\omega_{m_n^*}) < \infty$. Obviously a sufficient condition is given if the sequence δ satisfies in addition $\sup_{m\geqslant 1}\delta_m/(\sum_{j=1}^m \omega_j/\lambda_j) < \infty$. The polynomial case below provides an example. However, this condition is not necessary as can be seen in the exponential case. □

3.3 Convergence rate of the theoretical adaptive estimator.

We described in Section 2.3 three different cases where we could choose the model m such that the resulting estimator reaches the optimal minimax rate. The following result shows that, in case of known degree of ill-posedness, we can propose choices of sequences δ, Δ and M such that the penalized estimator automatically attains the optimal rate.

Proposition 3.3. *In cases* **[P-P]** *and* **[E-P]** *with $2a+2s+1>0$, let $\delta_m \asymp m^{2a+2s+1}$, $\Delta_m \asymp m^{(2a+2s)\vee 0}$ and $M_n \asymp n^{1/(2a+1+(2s)\vee 0)}$ with $(q)_{\vee 0} := \max(q,0)$. While in case* **[P-E]**, *choose $\delta_m \asymp m^{2a+1+(2s)\vee 0}\exp(m^{2a})$, $\Delta_m \asymp m^{(2s)\vee 0}\exp(m^{2a})$ and $M_n \asymp (\log n/(\log n)^{(2a+1+(2s)\vee 0)/(2a)})^{1/(2a)}$.*

Then Assumption 3.1 is fulfilled and, under the additional assumptions of Theorem 3.1, the adaptive estimator $\widehat{\beta}_{\widehat{m}}$ reaches the optimal rate.

In cases **[P-P]** and **[E-P]**, if $2a+2s+1<0$, then the sequence δ can be taken of order 1. The collection of models must be reduced to $\{[\sqrt{n}],\dots,n\}$ since M_n can be taken equal to n. It appears then that the rate is parametric in this case. In fact, no model selection is necessary in this case, a large m ($m=n$ for instance) can be chosen.

Now, we have in mind to prepare the case where the degree of ill-posedness of the λ_j's, and more precisely δ_m and M_n, are unknown. We propose hereafter a more intrinsic choice of δ_m, which does not require anything but the λ_j's (which can be estimated). In this spirit, we can prove the following assertion.

Proposition 3.4. *In cases* **[P-P]** *and* **[E-P]** *with $a+s \geqslant 0$ or in case* **[P-E]**, *choose $\Delta_m := \max_{1\leqslant j\leqslant m} \omega_j/\lambda_j$, $\kappa_m := \max_{1\leqslant j\leqslant m}(\omega_j)_{\vee 1}/\lambda_j$ with $(q)_{\vee 1} := \max(q,1)$ and*

$$\delta_m := m\Delta_m\Big|\frac{\log(\kappa_m \vee (m+2))}{\log(m+2)}\Big|. \tag{3.5}$$

Then Assumption 3.1 is fulfilled and, under the additional assumptions of Theorem 3.1, the adaptive estimator $\widehat{\beta}_{\widehat{m}}$ reaches the optimal rate.

4 A model selection approach: unknown degree of ill-posedness

In this section, the objective is the construction of a fully adaptive estimator which does not depend on the sequence γ and λ. Nevertheless the resulting estimator still attains the optimal rate in case the slope parameter belongs to some ellipsoid $\mathcal{F}_\gamma^\rho$ and the sequence of eigenvalues λ associated to the covariance operator of X has a given (unknown) rate of decrease.

The configuration given in Proposition 3.4 is now the right reference and the choice that the estimator is going to mimic. In particular, it is easily seen that there exists always a constant $\Sigma > 0$ such that the sequences δ and $\triangle$ given in Proposition 3.4 satisfy Assumption 3.1 (i). Observe that in this situation we have

$$\begin{aligned}
\Delta_m \exp(-\frac{\delta_m}{6\Delta_m}) &= \Delta_m \exp(-\frac{m}{6}\frac{\log(\kappa_m \vee (m+2))}{\log(m+2)})\\
&\leqslant (\kappa_m \vee (m+2)) \exp(-\frac{m}{6}\frac{\log(\kappa_m \vee (m+2))}{\log(m+2)})\\
&\leqslant \exp\Big(-m\Big[\frac{1}{6} - \frac{\log(m+2)}{m}\Big]\frac{\log(\kappa_m \vee (m+2))}{\log(m+2)}\Big)
\end{aligned}$$

where the last term is obviously summable.

Assumption 4.1. *Let λ denote the sequence of eigenvalues associated to the regressor X, let δ and $\triangle$ be the sequences defined in Proposition 3.4 and let γ and ω be sequences satisfying Assumption 2.1 such that*

(i) the sequence $M := (M_n)_{n\geqslant 1}$ given in Assumption 3.1 satisfies in addition to (3.2) also

$$\frac{\log n}{2n} \geqslant \max_{m>M_n} \frac{\lambda_m}{m(\omega_m)_{\vee 1}} \qquad \textit{for all } n \geqslant 1;$$

(ii) the sequence $m^\diamond := (m_n^\diamond)_{n\geqslant 1}$ given by $1/\underline{c} \leqslant \gamma_{m_n^\diamond}\delta_{m_n^\diamond}/(n\,\omega_{m_n^\diamond}) \leqslant \underline{c}$ for all $n \geqslant 1$ and some $\underline{c} \geqslant 1$ satisfies

$$\min_{1\leqslant m\leqslant m_n^\diamond} \frac{\lambda_m}{m(\omega_m)_{\vee 1}} \geqslant 2(\log n)/n \qquad \textit{for all } n \geqslant 1;$$

(iii) the sequence $N := (N_n)_{n\geqslant 1}$ given by $N_n := \arg\max_{1\leqslant N\leqslant n}\{\max_{1\leqslant j\leqslant N} \omega_j/n \leqslant 1\}$, $n \geqslant 1$, satisfies

$$M_n \leqslant N_n \leqslant n \qquad \textit{for all } n \geqslant 1.$$

Remark 4.1. The last assumption is technical but satisfied in the interesting case. Note that (i) and (ii) together imply $m_n^\diamond \leqslant M_n$ for all $n \geqslant 1$. The condition (iii) is rather weak, observe that the sequence ω is a-priori known and thus also the sequence of upper bounds N. In particular, recall that in case $\omega \equiv 1$ the $\mathcal{F}_\omega$-risk corresponds to the L^2-risk. If $\omega_m \leqslant 1$ for all $m \geqslant 1$, then $\mathcal{F}_\omega$-risk is weaker than the L^2-risk and $N_n = n$. Only if the $\mathcal{F}_\omega$-risk is stronger than the L^2-risk, that is ω is monotonically increasing, we choose N_n such that $\omega_{N_n} \asymp n$. Then it is not hard to see that in these situations (iii) is satisfied at least for sufficiently large n. □

4.1 Definition of the estimator

We follow the model selection approach presented in the last section. Define

$$\widehat{\Delta}_m := \max_{1\leqslant j\leqslant m} \frac{\omega_j}{\widehat{\lambda}_j}\mathbb{1}_{\{\widehat{\lambda}_j \geqslant 1/n\}} \quad \text{and} \quad \widehat{\kappa}_m := \max_{1\leqslant j\leqslant m} \frac{(\omega_j)_{\vee 1}}{\widehat{\lambda}_j}\mathbb{1}_{\{\widehat{\lambda}_j \geqslant 1/n\}}.$$

We shall refer to δ_m as defined in (3.5) and consider its estimator given by

$$\widehat{\delta}_m := m\widehat{\Delta}_m \Big|\frac{\log(\widehat{\kappa}_m \vee (m+2))}{\log(m+2)}\Big|.$$

If $X \in \mathcal{X}_\eta^4$ and $\mathbb{E}|Y/\sigma_Y|^4 \leqslant \eta$, then we define a random penalty function

$$\widehat{\text{pen}}(m) = 1920\sigma_Y^2\eta\,\frac{\widehat{\delta}_m}{n}.$$

Moreover, we consider a random upper bound for the collection of models given by

$$\widehat{M}_n := \arg\max_{1\leqslant M\leqslant N_n}\Big\{\frac{\widehat{\lambda}_M}{M(\omega_M)_{\vee 1}} \geqslant (\log n)/n\Big\}. \tag{4.1}$$

The adaptive estimator $\widehat{\beta}_{\widehat{m}}$ is obtained from (1.4) by choosing the dimension parameter

$$\widehat{m} := \arg\min_{1\leqslant m\leqslant \widehat{M}_n}\Big\{\Upsilon(\widehat{\beta}_m) + \widehat{\text{pen}}(m)\Big\} \tag{4.2}$$

We shall emphasize that the proposed estimator does not depend on an a-priori knowledge of neither the sequence γ nor the sequence λ.

4.2 An upper bound.

In the next assertion we provide an upper bound of the fully adaptive estimator $\widehat{\beta}_{\widehat{m}}$ by assuming that the sequences λ, ω and γ satisfy Assumption 4.1.

Theorem 4.1. *Assume an n-sample of (Y, X) satisfying (1.1). Suppose that $\mathbb{E}|Y/\sigma_Y|^4 \leqslant \eta$ and that $X \in \mathcal{X}_\eta^4$ is 1-periodic and second order stationary. Let Assumption 4.1 be satisfied. Consider the estimator $\widehat{\beta}_{\widehat{m}}$ defined in (1.4) with $\widehat{m}$ given by (4.2). If in addition $X \in \mathcal{X}_\xi^{28}$ and $\mathbb{E}|Y/\sigma_Y|^{28} \leqslant \xi$, then there exists a numerical constant $C > 0$ such that for all $n \geqslant 1$*

$$\sup_{\beta\in\mathcal{F}_\gamma^\rho}\Big\{\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2\Big\} \leqslant C\,\frac{\omega_{m_n^\diamond}}{\gamma_{m_n^\diamond}}(\rho + \underline{c}\eta[\rho\mathbb{E}\|X\|^2 + \sigma^2]) \\ + \frac{K}{n}[\rho\mathbb{E}\|X\|^2 + \sigma^2]\,[1 + \delta_1 + \rho][1 + (\mathbb{E}\|X\|^2)^2],$$

where $m_n^\diamond$ and $\underline{c}$ are defined in Assumption 4.1, $K = K(\Sigma, \eta, \xi, \delta_1)$ is a constant only depending on η, ξ, δ_1 and Σ such that the sequences δ and $\triangle$ given in Proposition 3.4 satisfy Assumption 3.1.

Remark 4.2. Comparing the last assertion with Theorem 3.1, we see that under Assumption 4.1 the proposed adaptive estimator obtains the same rate as in case of known degree of ill-posedness. We only have to impose in addition slightly stronger moment conditions.□

It is easily verified that in all the examples discussed above the fully adaptive estimator attains the optimal rate, which is summarized in the next assertion.

Corollary 4.2. *In cases* **[P-P]** *and* **[E-P]** *with $a + s \geqslant 0$ or in case* **[P-E]***, Assumption 4.1 is fulfilled and, under the additional assumptions of Theorem 4.1, the fully adaptive estimator $\widehat{\beta}_{\widehat{m}}$ with $\widehat{m}$ given by (4.2) reaches the optimal rate.*

Conclusion. Assuming a circular functional linear model we derive in this paper a fully adaptive estimator of the slope function β or its derivatives, which attains the minimax optimal rate of convergence. It is worth to note, that in this paper not only the penalty is chosen randomly but also the collection of models. In this way the proposed estimator is adaptive also with respect to the degree of ill-posedness of the underlying inverse problem. We can thereby face both, the mildly and the severely ill-posed case.

It is not clear that the ideas in this paper can be straightforwardly adapted to treat the case of noncircular functional models. We are currently exploring this issue.

A Appendix

A.1 Proof of Theorem 3.1

We begin by defining and recalling notations to be used in the proof. Given $u \in L^2[0,1]$ we denote by $[u]$ the infinite vector of Fourier coefficients $[u]_j := \langle u, \varphi_j\rangle$. In particular we use the notations

$$[X_i]_j = \langle X_i, \varphi_j\rangle, \quad [\beta]_j = \langle \beta, \varphi_j\rangle, \quad \sigma_Y^2 = \mathbb{V}\mathrm{ar}(Y),$$
$$\widehat{\beta}_m = \sum_{j=1}^m \widehat{\lambda}_j^{-1}\mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\}[\widehat{g}]_j\varphi_j, \quad \widetilde{\beta}_m := \sum_{j=1}^m \lambda_j^{-1}[\widehat{g}]_j\varphi_j, \quad \beta_m := \sum_{j=1}^m [\beta]_j\varphi_j,$$
$$\widehat{\Phi}_u = \sum_{j\geqslant 1} \widehat{\lambda}_j^{-1}\mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\}[u]_j\varphi_j, \quad \widetilde{\Phi}_u := \sum_{j\geqslant 1}\lambda_j^{-1}[u]_j\varphi_j.$$

Given $m \geqslant 1$ we have then for all $t \in \mathcal{S}_m = \mathrm{span}\{\varphi_1, \dots, \varphi_m\}$

$$\langle t, \beta\rangle_\omega = \sum_{j=1}^m \omega_j [t]_j[\beta]_j = \sum_{j=1}^m \frac{\omega_j[t]_j[g]_j}{\lambda_j} = \langle t, \widetilde{\Phi}_g\rangle_\omega,$$
$$\langle t, \widetilde{\beta}_m\rangle_\omega = \langle t, \widetilde{\Phi}_{\widehat{g}}\rangle_\omega = \frac{1}{n}\sum_{i=1}^n Y_i \langle t, \widetilde{\Phi}_{X_i}\rangle_\omega = \frac{1}{n}\sum_{i=1}^n Y_i \sum_{j=1}^m \frac{\omega_j}{\lambda_j}[X_i]_j[t]_j,$$
$$\langle t, \widehat{\beta}_m\rangle_\omega = \langle t, \widehat{\Phi}_{\widehat{g}}\rangle_\omega = \frac{1}{n}\sum_{i=1}^n Y_i\langle t, \widehat{\Phi}_{X_i}\rangle_\omega = \frac{1}{n}\sum_{i=1}^n Y_i \sum_{j=1}^m \frac{\omega_j}{\widehat{\lambda}_j}\mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\}[X_i]_j[t]_j. \quad \text{(A.1)}$$

Furthermore, define the event

$$\Omega_{Y,X} := \{|Y/\sigma_Y| \leqslant n^{1/6}, |[X]_j/\sqrt{\lambda_j}| \leqslant n^{1/6}, 1 \leqslant j \leqslant M_n\}$$

and denote its complement by $\Omega^c_{Y,X}$. Then consider the functions $\widehat{h}$ and $\widehat{f}$ with Fourier coefficients given by

$$[\widehat{h}]_j := \frac{1}{n}\sum_{i=1}^{n}\{Y_i[X_i]_j \mathbb{1}\Omega_{Y_i,X_i} - \mathbb{E}(Y_i[X_i]_j \mathbb{1}\Omega_{Y_i,X_i})\},$$
$$[\widehat{f}]_j := \frac{1}{n}\sum_{i=1}^{n}\{Y_i[X_i]_j \mathbb{1}\Omega^c_{Y_i,X_i} - \mathbb{E}(Y_i[X_i]_j \mathbb{1}\Omega^c_{Y_i,X_i})\}.$$

Obviously we have $[\widehat{g}]_j - [g]_j = [\widehat{h}]_j + [\widehat{f}]_j$ and hence for all $t \in \mathcal{S}_m$

$$\begin{aligned}\langle t, \widehat{\Phi}_{\widehat{g}} - \beta\rangle_\omega &= \langle t, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_g\rangle_\omega = \langle t, \widetilde{\Phi}_{\widehat{g}} - \widetilde{\Phi}_g\rangle_\omega + \langle t, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_{\widehat{g}}\rangle_\omega \\ &= \langle t, \widetilde{\Phi}_{\widehat{h}}\rangle_\omega + \langle t, \widetilde{\Phi}_{\widehat{f}}\rangle_\omega + \langle t, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_{\widehat{g}}\rangle_\omega. \qquad (A.2)\end{aligned}$$

We shall prove in the end of this section three technical Lemmas (A.2 - A.4) which are used in the following steps of the proof.

Consider now the contrast Υ then by using (3.3) and (3.4) it follows that

$$\Upsilon(\widehat{\beta}_{\widehat{m}}) + \mathrm{pen}(\widehat{m}) \leqslant \Upsilon(\widehat{\beta}_m) + \mathrm{pen}(m) \leqslant \Upsilon(\beta_m) + \mathrm{pen}(m), \qquad \forall 1 \leqslant m \leqslant M_n,$$

which in particular implies by using the notations given in (A.1) that

$$\begin{aligned}\|\widehat{\beta}_{\widehat{m}}\|^2_\omega - \|\beta_m\|^2_\omega &\leqslant 2\{\langle\widehat{\beta}_{\widehat{m}}, \widehat{\Phi}_{\widehat{g}}\rangle_\omega - \langle\beta_m, \widehat{\Phi}_{\widehat{g}}\rangle_\omega\} + \mathrm{pen}(m) - \mathrm{pen}(\widehat{m}) \\ &= 2\langle\widehat{\beta}_{\widehat{m}} - \beta_m, \widehat{\Phi}_{\widehat{g}}\rangle_\omega + \mathrm{pen}(m) - \mathrm{pen}(\widehat{m}).\end{aligned}$$

Rewriting the last estimate by using (A.2) we conclude that

$$\begin{aligned}\|\widehat{\beta}_{\widehat{m}} - \beta\|^2_\omega &= \|\beta - \beta_m\|^2_\omega + \|\widehat{\beta}_{\widehat{m}}\|^2_\omega - \|\beta_m\|^2_\omega - 2\langle\widehat{\beta}_{\widehat{m}} - \beta_m, \beta\rangle_\omega \\ &\leqslant \|\beta - \beta_m\|^2_\omega + \mathrm{pen}(m) - \mathrm{pen}(\widehat{m}) + 2\langle\widehat{\beta}_{\widehat{m}} - \beta_m, \widehat{\Phi}_{\widehat{g}} - \beta\rangle_\omega \\ &\leqslant \|\beta - \beta_m\|^2_\omega + \mathrm{pen}(m) - \mathrm{pen}(\widehat{m}) \\ &\quad + 2\langle\widehat{\beta}_{\widehat{m}} - \beta_m, \widetilde{\Phi}_{\widehat{h}}\rangle_\omega + 2\langle\widehat{\beta}_{\widehat{m}} - \beta_m, \widetilde{\Phi}_{\widehat{f}}\rangle_\omega + 2\langle\widehat{\beta}_{\widehat{m}} - \beta_m, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_{\widehat{g}}\rangle_\omega. \quad (A.3)\end{aligned}$$

Consider the unit ball $\mathcal{B}_m := \{f \in \mathcal{S}_m : \|f\|_\omega \leqslant 1\}$ and let $\widehat{m}\vee m := \max(\widehat{m}, m)$. Combining for $\tau > 0$ and $f \in \mathcal{S}_m$ the elementary inequality

$$2|\langle f, g\rangle_\omega| \leqslant 2\|f\|_\omega \sup_{t\in\mathcal{B}_m} |\langle t, g\rangle_\omega| \leqslant \tau\|f\|^2_\omega + \frac{1}{\tau}\sup_{t\in\mathcal{B}_m} |\langle t, g\rangle_\omega|^2$$

with (A.3) and $\widehat{\beta}_{\widehat{m}} - \beta_m \in \mathcal{S}_{\widehat{m}\vee m} \subset \mathcal{S}_{M_n}$ we obtain

$$\begin{aligned}\|\widehat{\beta}_{\widehat{m}} - \beta\|^2_\omega &\leqslant \|\beta - \beta_m\|^2_\omega + 6\tau\|\widehat{\beta}_{\widehat{m}} - \beta_m\|^2_\omega + \mathrm{pen}(m) - \mathrm{pen}(\widehat{m}) \\ &\quad + \frac{2}{\tau}\sup_{t\in\mathcal{B}_{\widehat{m}\vee m}} |\langle t, \widetilde{\Phi}_{\widehat{h}}\rangle_\omega|^2 + \frac{2}{\tau}\sup_{t\in\mathcal{B}_{M_n}} |\langle t, \widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2 + \frac{2}{\tau}\sup_{t\in\mathcal{B}_{M_n}} |\langle t, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2.\end{aligned}$$

Then, noting that $\mathrm{pen}(m \vee m') \leqslant \mathrm{pen}(m) + \mathrm{pen}(m')$ and $\|\widehat{\beta}_{\widehat{m}} - \beta_m\|_\omega^2 \leqslant 2\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 + 2\|\beta_m - \beta\|_\omega^2$, we get, together for $\tau = 1/16$ and $\mathrm{pen}(m) = 192\sigma_Y^2\eta\delta_m/n$ that

$$\begin{aligned}(1/4)\|\widehat{\beta}_{\widehat{m}} - \beta\|^2 &\leqslant (7/4)\|\beta - \beta_m\|^2 + 32\Big(\sup_{t\in\mathcal{B}_{\widehat{m}\vee m}} |\langle t, \widetilde{\Phi}_{\widehat{h}}\rangle_\omega|^2 - (1/32)\,\mathrm{pen}(\widehat{m}\vee m)\Big)_+ \\ &\quad + 32 \sup_{t\in\mathcal{B}_{M_n}} |\langle t, \widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2 + 32 \sup_{t\in\mathcal{B}_{M_n}} |\langle t, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2 + \mathrm{pen}(\widehat{m}\vee m) + \mathrm{pen}(m) - \mathrm{pen}(\widehat{m}) \\ &\leqslant (7/4)\|\beta - \beta_m\|^2 + 32\sum_{m'=1}^{M_n}\Big(\sup_{t\in\mathcal{B}_{m'}} |\langle t, \widetilde{\Phi}_{\widehat{h}}\rangle_\omega|^2 - 6\sigma_Y^2\eta\delta_{m'}/n\Big)_+ \\ &\quad + 32 \sup_{t\in\mathcal{B}_{M_n}} |\langle t, \widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2 + 32 \sup_{t\in\mathcal{B}_{M_n}} |\langle t, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2 + 2\,\mathrm{pen}(m). \qquad (A.4)\end{aligned}$$

Combining the last bound with (A.5) in Lemma A.2, (A.9) and (A.10) in Lemma A.3 we conclude that there exist a numerical constant C and a constant $K(\Sigma, \eta)$ depending on Σ and η only, such that for all $n \geqslant 1$ and for all $1 \leqslant m \leqslant M_n$ we have

$$\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \leqslant 7\|\beta - \beta_m\|_\omega^2 + 8\,\mathrm{pen}(m) + \frac{1}{n}[C\xi(\sigma_Y^2\delta_1 + \|\beta\|_\omega^2)\{1 + (\mathbb{E}\|X\|^2)^2\} + \sigma_Y^2 K(\Sigma(6), \eta)].$$

Since (ω/γ) is monotonically non increasing we obtain in case $\beta \in \mathcal{F}_\gamma^\rho$ that $\|\beta\|_\omega^2 \leqslant \rho$ and $\|\beta - \beta_m\|_\omega^2 \leqslant (\omega_m/\gamma_m)\rho$. Moreover, by using that X and ε are uncorrelated it follows $\sigma_Y^2 = \mathbb{V}\mathrm{ar}(\langle X, \beta\rangle) + \sigma^2\,\mathbb{V}\mathrm{ar}(\varepsilon) \leqslant \mathbb{E}\langle X, \beta\rangle^2 + \sigma^2 \leqslant \|\beta\|^2\mathbb{E}\|X\|^2 + \sigma^2$. Hence, $\sigma_Y^2 \leqslant \rho\mathbb{E}\|X\|^2 + \sigma^2$ because γ is monotonically non decreasing. The result follows now by combining the last estimates with the definition of the penalty, that is, $\mathrm{pen}(m) = 192\sigma_Y^2\eta\delta_m/n$, which completes the proof of Theorem 3.1. $\square$

Technical assertions.

The following lemmas gather technical results used in the proof of Theorem 3.1. We begin by recalling an inequality due to Talagrand [1996], which can be found e.g. in Comte et al. [2006].

Lemma A.1 (Talagrand's Inequality). *Let $T_1, \dots, T_n$ be independent $\mathcal{T}$-valued random variables and $\nu_n^*(r) = (1/n)\sum_{i=1}^n \big[r(T_i) - \mathbb{E}[r(T_i)]\big]$, for r belonging to a countable class $\mathcal{R}$ of measurable functions. Then, for $\varepsilon > 0$,*

$$\begin{aligned}&\mathbb{E}[\sup_{r\in\mathcal{R}} |\nu_n^*(r)|^2 - 2(1+2\varepsilon)H^2]_+ \\ &\qquad \leqslant C\left(\frac{v}{n}\exp(-K_1\varepsilon\frac{nH^2}{v})) + \frac{h^2}{n^2C^2(\varepsilon)}\exp(-K_2C(\varepsilon)\sqrt{\varepsilon}\frac{nH}{h})\right)\end{aligned}$$

with $K_1 = 1/6$, $K_2 = 1/(21\sqrt{2})$, $C(\varepsilon) = \sqrt{1+\varepsilon} - 1$ and C a universal constant and where

$$\sup_{r\in\mathbb{R}}\sup_{t\in\mathcal{T}} |r(t)| \leqslant h, \quad \mathbb{E}\left[\sup_{r\in\mathcal{R}} |\nu_n^*(r)|\right] \leqslant H, \quad \sup_{r\in\mathcal{R}} \frac{1}{n}\sum_{i=1}^n \mathbb{V}\mathrm{ar}(r(T_i)) \leqslant v.$$

Lemma A.2. *Let λ be the eigenvalues associated to $X \in \mathcal{X}_\eta^4$ and $\mathbb{E}|Y/\sigma_Y|^4 \leqslant \eta$. Suppose sequences δ, $\triangle$ and M satisfying Assumption 3.1. Then there exists a constant $K(\Sigma, \eta, \delta_1)$ only depending on Σ, η and δ_1 such that*

$$\sum_{m=1}^{M_n} \mathbb{E}\Big(\sup_{t\in\mathcal{B}_m} |\langle t, \widetilde{\Phi}_{\widehat{h}}\rangle_\omega|^2 - 6\sigma_Y^2\eta\frac{\delta_m}{n}\Big)_+ \leqslant K(\Sigma, \eta, \delta_1)\,\frac{\sigma_Y^2}{n} \quad \text{for all } n \geqslant 1. \qquad (A.5)$$

Proof. Given $m \in \mathbb{N}$ and $t \in \mathcal{B}_m := \{f \in \mathcal{S}_m : \|f\|_\omega \leqslant 1\}$ denote

$$v_t(Y,X) := Y \mathbb{1}_{\Omega_{Y,X}} \langle t, \widetilde{\Phi}_X \rangle_\omega = \sum_{j=1}^m \frac{\omega_j [t]_j}{\lambda_j} Y \mathbb{1}_{\Omega_{Y,X}} [X]_j,$$

then it is easily seen that $\langle t, \widetilde{\Phi}_{\widehat{h}} \rangle_\omega = (1/n) \sum_{i=1}^n \{v_t(Y_i, X_i) - \mathbb{E} v_t(Y_i, X_i)\}$. Below we show the following three bounds

$$\sup_{t \in \mathcal{B}_m} \sup_{y \in \mathbb{R}, x \in L^2[0,1]} |v_t(y,x)| \leqslant \sigma_Y n^{1/3} \delta_m^{1/2} =: h, \tag{A.6}$$

$$\mathbb{E} \sup_{t \in \mathcal{B}_m} |\langle t, \widetilde{\Phi}_{\widehat{h}} \rangle_\omega|^2 \leqslant \sigma_Y^2 \eta \frac{\delta_m}{n} =: H^2, \tag{A.7}$$

$$\sup_{t \in \mathcal{B}_m} \frac{1}{n} \sum_{i=1}^n \mathbb{V}\mathrm{ar}(v_t(Y_i, X_i)) \leqslant \sigma_Y^2 \eta \triangle_m =: v. \tag{A.8}$$

From Talagrand's inequality (Lemma A.1) with $\varepsilon = 1$ we obtain by combining (A.6)-(A.8)

$$\mathbb{E}\Big[\sup_{t \in \mathcal{B}_m} |\langle t, \widetilde{\Phi}_{\widehat{h}} \rangle_\omega|^2 - 6H^2\Big] \leqslant C\Big\{\frac{v}{n} \exp\Big(-\frac{nH^2}{6v}\Big) + \frac{h^2}{n^2} \exp\Big(-\frac{c\, n\, H}{h}\Big)\Big\}$$
$$= C\Big\{\frac{\sigma_Y^2 \eta \triangle_m}{n} \exp\Big(-\frac{\delta_m}{6\triangle_m}\Big) + \sigma_Y^2 \frac{n^{2/3} \delta_m}{n^2} \exp\Big(-c\, \eta\, n^{1/6}\Big)\Big\}$$

with $c = (1 - 1/\sqrt{2})/21$ and some numerical constant $C > 0$. By using Assumption 3.1, that is $\delta_m / n \leqslant \delta_{M_n}/n \leqslant \delta_1$ and $M_n / n \leqslant 1$, together with $H^2 = \sigma_Y^2 \eta \delta_m / n$ it follows that

$$\sum_{m=1}^{M_n} \mathbb{E}\Big[\sup_{t \in \mathcal{B}_m} |\langle t, \widetilde{\Phi}_{\widehat{h}} \rangle_\omega|^2 - 6\sigma_Y^2 \eta \delta_m / n\Big]$$
$$\leqslant C\Big\{\frac{\sigma_Y^2 \eta}{n} \sum_{m=1}^{M_n} \triangle_m \exp\Big(-\frac{\delta_m}{6\triangle_m}\Big) + \sigma_Y^2 \delta_1 n^{2/3} \exp\Big(-c\, \eta\, n^{1/6}\Big)\Big\}$$
$$\leqslant C \frac{\sigma_Y^2}{n}\Big\{\eta\, \Sigma + \delta_1 \exp\Big(-c\, \eta\, n^{1/6} + (5/3)\log n\Big)\Big\},$$

where condition (3.1) in Assumption 3.1 implies the last inequality. It follows that there exists a constant $K(\Sigma, \eta, \delta_1)$ only depending on Σ, η and δ_1 such that

$$\sum_{m=1}^{M_n} \mathbb{E}\Big[\sup_{t \in \mathcal{B}_m} |\langle t, \widetilde{\Phi}_{\widehat{h}} \rangle_\omega|^2 - 6\sigma_Y^2 \eta \delta_m / n\Big] \leqslant \frac{\sigma_Y^2}{n} K(\Sigma, \eta, \delta_1), \quad \text{for all } n \geqslant 1,$$

which proves the result.

Proof of (A.6). From $\sup_{t \in \mathcal{B}_m} |\langle t, g \rangle_\omega|^2 = \sum_{j=1}^m \omega_j [g]_j^2$ and the definition of $\Omega_{Y,X}$ follows

$$\sup_{y \in \mathbb{R}, x \in L^2[0,1], t \in \mathcal{B}_m} |v_t(y,x)|^2 = \sup_{y \in \mathbb{R}, x \in L^2[0,1]} \sum_{j=1}^m \frac{\omega_j \sigma_Y^2}{\lambda_j} \mathbb{1}_{\Omega_{y,x}} \frac{y^2}{\sigma_Y^2} \frac{[x]_j^2}{\lambda_j} \leqslant \sigma_Y^2 n^{2/3} \sum_{j=1}^m \frac{\omega_j}{\lambda_j}$$

and, hence the definition of δ_m implies (A.6).

Proof of (A.7). Since (Y_i, X_i), $i = 1, \dots, n$, form an n-sample of (Y, X) we have

$$\mathbb{E}\sup_{t\in\mathcal{B}_m}|\langle t,\widetilde{\Phi}_{\widehat{h}}\rangle_\omega|^2 = \sum_{j=1}^m \frac{\omega_j}{\lambda_j^2}\mathbb{V}\mathrm{ar}\left(\frac{1}{n}\sum_{i=1}^n Y_i\mathbb{1}_{\Omega_{Y_i,X_i}}[X_i]_j\right) \leqslant \frac{1}{n}\sum_{j=1}^m\frac{\omega_j}{\lambda_j^2}\mathbb{E}\left(Y\mathbb{1}_{\Omega_{Y,X}}[X]_j\right)^2$$

and hence from $\mathbb{E}|Y/\sigma_Y|^4 \leqslant \eta$ and $X \in \mathcal{X}_\eta^4$ it follows that

$$\mathbb{E}\sup_{t\in\mathcal{B}_N}|\langle t,\widetilde{\Phi}_{\widehat{h}}\rangle_\omega|^2 \leqslant \frac{\sigma_Y^2}{n}\sum_{j=1}^m\frac{\omega_j}{\lambda_j}\left(\mathbb{E}|Y/\sigma_Y|^4\mathbb{E}|[X]_j/\sqrt{\lambda_j}|^4\right)^{1/2} \leqslant \frac{\sigma_Y^2}{n}\eta\sum_{j=1}^m\frac{\omega_j}{\lambda_j}.$$

Thereby, the definition of δ_m implies also (A.7).

Proof of (A.8). Consider $z := (z_j)$ with $z_j := (\omega_j[t]_j/\sqrt{\lambda_j})/(\sum_{j=1}^m(\omega_j^2[t]_j^2/\lambda_j))^{1/2}$ and, hence $z \in \mathbb{S}^m = \{z \in \mathbb{R}^m, \sum_{j=1}^m z_j^2 = 1\}$. Since (Y_i, X_i), $i = 1, \dots, n$, form an n-sample of (Y, X) it follows that

$$\sup_{t\in\mathcal{B}_m}\frac{1}{n}\sum_{i=1}^n\mathbb{V}\mathrm{ar}(v_t(Y_i,X_i)) \leqslant \sup_{t\in\mathcal{B}_m}\mathbb{E}\Big(Y\mathbb{1}_{\Omega_{Y,X}}\sum_{j=1}^m\frac{\omega_j[t]_j}{\lambda_j}[X]_j\Big)^2.$$

Thereby, from $\mathbb{E}|Y/\sigma_Y|^4 \leqslant \eta$ and $X \in \mathcal{X}_\eta^4$ we conclude that

$$\begin{aligned}\sup_{t\in\mathcal{B}_m}\frac{1}{n}\sum_{i=1}^n\mathbb{V}\mathrm{ar}(v_t(Y_i,X_i)) &\leqslant \sup_{t\in\mathcal{B}_m}\sigma_Y^2(\mathbb{E}|Y/\sigma_Y|^4)^{1/2}\Big(\mathbb{E}\Big|\sum_{j=1}^m\frac{\omega_j[t]_j}{\sqrt{\lambda_j}}\,\frac{[X]_j}{\sqrt{\lambda_j}}\Big|^4\Big)^{1/2}\\ &\leqslant \sigma_Y^2\eta^{1/2}\sup_{t\in\mathcal{B}_m}\sum_{j=1}^m(\omega_j^2[t]_j^2/\lambda_j)\sup_{z\in\mathbb{S}_N}\Big(\mathbb{E}\Big|\sum_{j=1}^m z_j[X]_j/\sqrt{\lambda_j}\Big|^4\Big)^{1/2}\\ &\leqslant \sigma_Y^2\eta\sup_{t\in\mathcal{B}_m}\sum_{j=1}^m(\omega_j^2[t]_j^2/\lambda_j) \leqslant \sigma_Y^2\eta\max_{1\leqslant j\leqslant m}\omega_j/\lambda_j.\end{aligned}$$

Thus the definition of $\triangle_m$ implies now (A.8), which completes the proof of Lemma A.2. □

Lemma A.3. *Let λ be the eigenvalues associated to $X \in \mathcal{X}_\xi^{24}$ and let $\mathbb{E}|Y/\sigma_Y|^{24} \leqslant \xi$. Suppose sequences δ, $\triangle$ and M satisfying Assumption 3.1. Then there exists a numerical constant C such that*

$$\mathbb{E}\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2 \leqslant \sqrt{2}\,\xi\,\sigma_Y^2\,\delta_1/n \qquad \textit{and} \tag{A.9}$$

$$\mathbb{E}\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widehat{\Phi}_{\widehat{g}}-\widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2 \leqslant C\frac{\xi}{n}\{\sigma_Y^2\,\delta_1 + \|\beta\|_\omega^2\}\{1+(\mathbb{E}\|X\|^2)^2\} \quad \textit{for all } n \geqslant 1. \tag{A.10}$$

Proof. Since (Y_i, X_i), $i = 1, \dots, n$, form an n-sample of (Y, X) it follows that

$$\mathbb{E}\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2 = \sum_{j=1}^{M_n}\frac{\omega_j}{\lambda_j^2}\mathbb{V}\mathrm{ar}\left(\frac{1}{n}\sum_{i=1}^n Y\mathbb{1}_{\Omega_{Y,X}^c}[X]_j\right) \leqslant \sum_{j=1}^{M_n}\frac{\omega_j}{\lambda_j^2 n}\mathbb{E}\left(Y\mathbb{1}_{\Omega_{Y,X}^c}[X]_j\right)^2.$$

Thereby, from $\mathbb{E}|Y/\sigma_Y|^{24} \leqslant \xi$ and $X \in \mathcal{X}_\xi^{24}$ we conclude that

$$\begin{aligned}\mathbb{E}\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2 &\leqslant \frac{\sigma_Y^2}{n}\sum_{j=1}^{M_n}\frac{\omega_j}{\lambda_j}\left(\mathbb{E}|Y/\sigma_Y|^8\mathbb{E}|[X]_j/\sqrt{\lambda_j}|^8\right)^{1/4}P(\Omega_{Y,X}^c)^{1/2}\\ &\leqslant \frac{\sigma_Y^2\xi^{1/2}}{n}\sum_{j=1}^{M_n}\frac{\omega_j}{\lambda_j}P(\Omega_{Y,X}^c)^{1/2} \leqslant \sigma_Y^2\xi^{1/2}\frac{\delta_{M_n}}{n}P(\Omega_{Y,X}^c)^{1/2}\end{aligned}$$

where the last inequality follows from the property $\delta_m \geqslant \sum_{j=1}^m \frac{\omega_j}{\lambda_j}$ for all $m \geqslant 1$. Hence by using Assumption 3.1, that is $\delta_{M_n}/n \leqslant \delta_1$, we obtain

$$\mathbb{E} \sup_{t\in\mathcal{B}_{M_n}} |\langle t, \widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2 \leqslant \sigma_Y^2 \delta_1 \xi^{1/2} P(\Omega^c_{Y,X})^{1/2}.$$

The estimate (A.9) follows now from $P(\Omega^c_{Y,X}) \leqslant 2\xi/n^2$, which can be realized as follows. Since $\Omega^c_{Y,X} = \{|Y/\sigma_Y| > n^{1/6}\} \cup \bigcup_{j=1}^{M_n}\{|[X]_j/\sqrt{\lambda_j}| > n^{1/6}\}$ it follows by using Markov's inequality together with $\mathbb{E}|Y/\sigma_Y|^{24} \leqslant \xi$ and $X \in \mathcal{X}_\xi^{24}$ that

$$\begin{aligned} P(\Omega^c_{Y,X}) &\leqslant P(|Y/\sigma_Y| > n^{1/6}) + \sum_{j=1}^{M_n} P(|[X]_j/\sqrt{\lambda_j}| > n^{1/6}) \\ &\leqslant \frac{\mathbb{E}|Y/\sigma_Y|^{18}}{n^3} + \sum_{j=1}^{M_n} \frac{\mathbb{E}|[X]_j/\sqrt{\lambda_j}|^{18}}{n^3} \leqslant \frac{\xi}{n^3}(1+M_n) \end{aligned}$$

Thus, under Assumption 3.1, that is, $M_n/n \leqslant 1$, we obtain $P(\Omega^c_{Y,X}) \leqslant 2\xi/n^2$, which completes the proof of (A.9).

Proof of (A.10). Consider the decomposition

$$\begin{aligned} \sup_{t\in\mathcal{B}_{M_n}} |\langle t, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2 &= \sum_{j=1}^{M_n} \frac{\omega_j}{\lambda_j}\Big(\frac{\lambda_j}{\widehat{\lambda}_j}\mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\} - 1\Big)^2 \left(\frac{1}{n}\sum_{i=1}^n Y_i \frac{[X_i]_j}{\sqrt{\lambda_j}}\right)^2 \\ &\leqslant 2\sum_{j=1}^{M_n} \frac{\omega_j}{\lambda_j}\Big(\frac{\lambda_j}{\widehat{\lambda}_j} - 1\Big)^2 \mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\} \left(\frac{1}{n}\sum_{i=1}^n Y_i \frac{[X_i]_j}{\sqrt{\lambda_j}} - \sqrt{\lambda_j}[\beta]_j\right)^2 \\ &\quad + 2\sum_{j=1}^{M_n} \omega_j [\beta]_j^2 \Big(\frac{\lambda_j}{\widehat{\lambda}_j} - 1\Big)^2 \mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\} \\ &\quad + 2\sum_{j=1}^{M_n} \frac{\omega_j}{\lambda_j} \left(\frac{1}{n}\sum_{i=1}^n Y_i \frac{[X_i]_j}{\sqrt{\lambda_j}} - \sqrt{\lambda_j}[\beta]_j\right)^2 \mathbb{1}\{\widehat{\lambda}_j < 1/n\} \\ &\quad + 2\sum_{j=1}^{M_n} \omega_j [\beta]_j^2 \mathbb{1}\{\widehat{\lambda}_j < 1/n\} \end{aligned} \tag{A.11}$$

where we bound each summand separately. First, from (A.16) and (A.19) in Lemma A.4 together with $X \in \mathcal{X}_\xi^{24}$ and $\mathbb{E}|Y/\sigma_Y|^{24} \leqslant \xi$ it follows that there exists a numeric constant $C > 0$ such that

$$\begin{aligned} \mathbb{E}\sum_{j=1}^{M_n} \frac{\omega_j}{\lambda_j}\Big(\frac{\lambda_j}{\widehat{\lambda}_j} - 1\Big)^2 \mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\} \left(\frac{1}{n}\sum_{i=1}^n Y_i \frac{[X_i]_j}{\sqrt{\lambda_j}} - \sqrt{\lambda_j}[\beta]_j\right)^2 \\ \leqslant \sum_{j=1}^{M_n} \frac{\omega_j}{\lambda_j}\Big[\mathbb{E}|\lambda_j/\widehat{\lambda}_j - 1|^4 \mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\}\Big]^{1/2} \Big[\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n Y_i \frac{[X_i]_j}{\sqrt{\lambda_j}} - \sqrt{\lambda_j}[\beta]_j\right)^4\Big]^{1/2} \\ \leqslant C\frac{\sigma_Y^2 \xi}{n}\sum_{j=1}^{M_n} \frac{\omega_j}{n\lambda_j}\{\lambda_j^2 + 1\}; \end{aligned} \tag{A.12}$$

$$\mathbb{E}\sum_{j=1}^{M_n}\omega_j[\beta]_j^2\Big(\frac{\lambda_j}{\widehat{\lambda}_j}-1\Big)^2\mathbb{1}\{\widehat{\lambda}_j\geqslant 1/n\}\leqslant C\frac{\xi}{n}\sum_{j=1}^{M_n}\omega_j[\beta]_j^2\{\lambda_j^2+1\}. \tag{A.13}$$

Furthermore, Assumption 3.1 (ii), i.e., $2/n\leqslant\min\{\lambda_j : 1\leqslant j\leqslant M_n\}$, implies $P(\widehat{\lambda}_j<1/n)\leqslant P(\widehat{\lambda}_j/\lambda_j<1/2)$. Thereby, from (A.16) and (A.18) in Lemma A.4 together with $X\in\mathcal{X}_\xi^{24}$ and $\mathbb{E}|Y/\sigma_Y|^{24}\leqslant\xi$ it follows that there exists a numeric constant $C>0$ such that

$$\begin{aligned}\mathbb{E}\sum_{j=1}^{M_n}\frac{\omega_j}{\lambda_j}\left(\frac{1}{n}\sum_{i=1}^n Y_i\frac{[X_i]_j}{\sqrt{\lambda_j}}-\sqrt{\lambda_j}[\beta]_j\right)^2\mathbb{1}\{\widehat{\lambda}_j<1/n\}&\\ \leqslant\sum_{j=1}^{M_n}\frac{\omega_j}{\lambda_j}\Big[\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n Y_i\frac{[X_i]_j}{\sqrt{\lambda_j}}-\sqrt{\lambda_j}[\beta]_j\right)^4\Big]^{1/2}&P(\widehat{\lambda}_j/\lambda_j<1/2)^{1/2}\\ \leqslant C\frac{\sigma_Y^2\xi}{n}\sum_{j=1}^{M_n}\frac{\omega_j}{n\lambda_j};&\quad(\text{A.14})\end{aligned}$$

$$\mathbb{E}\sum_{j=1}^{M_n}\omega_j[\beta]_j^2\mathbb{1}\{\widehat{\lambda}_j<1/n\}\leqslant\sum_{j=1}^{M_n}\omega_j[\beta]_j^2P(\widehat{\lambda}_j/\lambda_j<1/2)\leqslant C\frac{\xi}{n}\sum_{j=1}^{M_n}\omega_j[\beta]_j^2. \tag{A.15}$$

Combining the decomposition (A.11) and the bounds (A.12) - (A.15) we obtain

$$\mathbb{E}\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widehat{\Phi}_{\widehat{g}}-\widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2\leqslant C\frac{\xi}{n}\Big\{\sum_{j=1}^{M_n}\frac{\omega_j}{n\lambda_j}\sigma_Y^2\{\lambda_j^2+2\}+\sum_{j=1}^{M_n}\omega_j[\beta]_j^2\{\lambda_j^2+2\}\Big\}.$$

Therefore the properties $\mathbb{E}\|X\|^2\geqslant\max_{j\geqslant 1}\lambda_j$ and $\delta_m\geqslant\sum_{j=1}^m\frac{\omega_j}{\lambda_j}$ for all $m\geqslant 1$ imply

$$\mathbb{E}\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widehat{\Phi}_{\widehat{g}}-\widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2\leqslant C\frac{\xi}{n}\{\sigma_Y^2\delta_{M_n}/n+\|\beta\|_\omega^2\}\{(\mathbb{E}\|X\|^2)^2+2\}.$$

Thus (A.10) follows now from $\delta_{M_n}/n\leqslant\delta_1$ (Assumption 3.1), which completes the proof. □

Lemma A.4. *Suppose $X\in\mathcal{X}_{\eta_{4k}}^{4k}$ and $\mathbb{E}|Y/\sigma_Y|^{4k}\leqslant\eta_{4k}$, $k\geqslant 1$. Then for some numeric constant $C_k>0$ only depending on k we have*

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n Y_i\frac{[X_i]_j}{\sqrt{\lambda_j}}-\sqrt{\lambda_j}[\beta]_j\right)^{2k}\leqslant C_k\sigma_Y^{2k}\eta_{4k}\,n^{-k}, \tag{A.16}$$

$$\mathbb{E}|\widehat{\lambda}_j/\lambda_j-1|^{2k}\leqslant C_k\eta_{4k}\,n^{-k}. \tag{A.17}$$

If in addition $w_1\geqslant 2$ and $w_2\leqslant 1/2$, then we obtain

$$\sup_{j\in\mathbb{N}}P(\widehat{\lambda}_j/\lambda_j\geqslant w_1)\leqslant C_k\eta_{4k}\,n^{-k}\quad\text{and}\quad\sup_{j\in\mathbb{N}}P(\widehat{\lambda}_j/\lambda_j<w_2)\leqslant C_k\eta_{4k}\,n^{-k}. \tag{A.18}$$

Moreover, if $X\in\mathcal{X}_{\eta_{12k}}^{12k}$, $k\geqslant 1$, then for some numeric constant $C_k>0$ only depending on k we have

$$\mathbb{E}|\lambda_j/\widehat{\lambda}_j-1|^{2k}\mathbb{1}\{\widehat{\lambda}_j\geqslant 1/n\}\leqslant C_k\eta_{12k}\{\lambda_j^{2k}+1\}n^{-k}. \tag{A.19}$$

Proof. Since $\mathbb{E}Y[X]_j = \lambda_j[\beta]_j$ the independence within the sample of (Y,X) implies by using Theorem 2.10 in Petrov [1995] for some generic constant C_k that

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\frac{[X_i]_j}{\sqrt{\lambda_j}} - \sqrt{\lambda_j}[\beta]_j\right)^{2k} \leqslant C_{2k}\sigma_Y^{2k}n^{-k}\mathbb{E}|Y/\sigma|^{2k}|[X]_j/\sqrt{\lambda_j}|^{2k}$$
$$\leqslant C_k\sigma_Y^{2k}n^{-k}\left(\mathbb{E}|Y/\sigma|^{4k}\mathbb{E}|[X]_j/\sqrt{\lambda_j}|^{4k}\right)^{1/2}.$$

Then the last estimate together with $X \in \mathcal{X}^{4k}_{\eta_{4k}}$ and $\mathbb{E}|Y/\sigma_Y|^{4k} \leqslant \eta_{4k}$ implies (A.16). Furthermore, since $\{(|[X_i]_j|^2/\lambda_j - 1)_i\}$ are independent and identically distributed with mean zero, it follows by applying again Theorem 2.10 in Petrov [1995] that $\mathbb{E}|\widehat{\lambda}_j/\lambda_j - 1|^{2k} \leqslant C_k n^{-k}\mathbb{E}||[X]_j/\sqrt{\lambda_j}|^2 - 1|^{2k}$. Thus, the condition $X \in \mathcal{X}^{4k}_{\eta_{4k}}$ implies (A.17).

Proof of (A.18). If $w \geqslant 2$ then $P(\widehat{\lambda}_j/\lambda_j \geqslant w) \leqslant P(|\widehat{\lambda}_j/\lambda_j - 1| \geqslant 1)$. Thus applying Markov's inequality together with (A.17) implies the first bound in (A.18), while the second follows in analogy.

Proof of (A.19). By using twice the elementary inequality $|\widehat{\lambda}_j/\lambda_j - 1|^{2k} + |\widehat{\lambda}_j/\lambda_j|^{2k} \geqslant 1/2^{2k-1}$ we conclude that

$$\mathbb{E}|\lambda_j/\widehat{\lambda}_j - 1|^{2k}\mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\} \leqslant 2^{2k-1}\{\mathbb{E}|\widehat{\lambda}_j/\lambda_j - 1|^{4k}\frac{\lambda_j^{2k}}{\widehat{\lambda}_j^{2k}}\mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\} + \mathbb{E}|\widehat{\lambda}_j/\lambda_j - 1|^{2k}\}$$
$$\leqslant 2^{4k-2}\lambda_j^{2k}n^{2k}\mathbb{E}|\widehat{\lambda}_j/\lambda_j - 1|^{6k} + 2^{4k-2}\mathbb{E}|\widehat{\lambda}_j/\lambda_j - 1|^{4k} + 2^{2k-1}\mathbb{E}|\widehat{\lambda}_j/\lambda_j - 1|^{2k}\}.$$

Thus, (A.19) follows from (A.17) since $X \in \mathcal{X}^{12k}_{\eta_{12k}}$, which proves the lemma. $\square$

A.2 Proof of Proposition 3.3

Case [P-P] Since $2a + 2s + 1 > 0$ it follows that the sequences δ, Δ and M with $\delta_m \asymp m^{2a+2s+1}$, $\Delta_m \asymp m^{(2a+2s)\vee 0}$ and $M_n \asymp n^{1/(2a+1+(2s)\vee 0)}$, respectively, satisfy Assumption 3.1. Note that $\delta_{M_n}/n \leqslant 1$, $M_n/n \leqslant 1$, $\min_{1\leqslant j\leqslant M_n}\lambda_j \geqslant 2/n$ and $\forall C > 0$,

$$\sum_m \triangle_m \exp(-C\delta_m/\Delta_m) \asymp \sum_m m^{(2a+2s)\vee 0}\exp(-Cm^{(2a+2s+1)\wedge 1}) < +\infty.$$

Therefore we can apply Theorem 3.1 and hence Corollary 3.2. In particular, by using $m_n^\diamond \asymp n^{1/(2a+2p+1)}$, which satisfies $\gamma_{m_n^\diamond}\delta_{m_n^\diamond}/(n\omega_{m_n^\diamond}) \asymp 1$, it follows that the adaptive estimator $\widehat{\beta}_{\widehat{m}}$ reaches the optimal rate $\omega_{m_n^\diamond}/\gamma_{m_n^\diamond} \asymp n^{-2(p-s)/(2p+2a+1)}$.

Case [E-P] The sequences δ, Δ, M are unchanged w.r.t. the previous case [P-P] and hence Assumption 3.1 is still satisfied. From Corollary 3.2 follows now again that the adaptive estimator $\widehat{\beta}_{\widehat{m}}$ attains the optimal rate $\omega_{m_n^\diamond}/\gamma_{m_n^\diamond} \asymp n^{-1}(\log n)^{(2a+1+2s)/(2p)}$ since $m_n^\diamond \asymp \{\log[n(\log n)^{-(2a+1)/(2p)}]\}^{1/(2p)}$ satisfies $\gamma_{m_n^\diamond}\delta_{m_n^\diamond}/(n\omega_{m_n^\diamond}) \asymp 1$.

Case [P-E] Consider the sequences δ, Δ and M with $\delta_m = m^{2a+1+(2s)\vee 0}\exp(m^{2a})$, $\Delta_m = m^{(2s)\vee 0}\exp(m^{2a})$ and $M_n = (\log n/(\log n)^{2a+1+(2s)\vee 0)/(2a)})^{1/(2a)}$ respectively. Then Assumption 3.1 is satisfied, that is $\delta_{M_n}/n \leqslant 1$, $M_n/n \leqslant 1$, $\min_{1\leqslant j\leqslant M_n}\lambda_j \geqslant 2/n$ and $\forall C > 0$,

$$\sum_m \Delta_m \exp(-C\delta_m/\Delta_m) \leqslant \sum_m m^{(2s)\vee 0}\exp(m^{2a})\exp(-Cm^{2a+1}) < +\infty.$$

Moreover, $\gamma_{m^\diamond}\delta_{m^\diamond}/(n\omega_{m^\diamond}) \asymp 1$ implies $m_n^\diamond \asymp (\log n/(\log n)^{(2a+2p+1)/(2a)})^{1/(2a)}$. Finally, due to Corollary 3.2 the adaptive estimator $\widehat{\beta}_{\widehat{m}}$ attains again the optimal rate $\omega_{m^\diamond}/\gamma_{m^\diamond} \asymp (\log n)^{-(p-s)/a}$, which completes the proof of Proposition 3.3. $\square$

A.3 Proof of Proposition 3.4

Let $\Delta_m := \max_{1\leqslant j\leqslant m} \omega_j/\lambda_j$, $\kappa_m := \max_{1\leqslant j\leqslant m}(\omega_j)_{\vee 1}/\lambda_j$ and $\delta_m := m\Delta_m\left|\frac{\log(\kappa_m\vee(m+2))}{\log(m+2)}\right|$ as defined in (3.5). Note that $|\log(\kappa_m \vee(m+2))/\log(m+2)| \geqslant 1$ and hence $\delta_m \geqslant \sum_{j=1}^m \omega_j/\lambda_j$.

Case [P-P] and [E-P]. Since $a+s \geqslant 0$ it is easily verified that $\Delta_m \asymp m^{2a+2s}$, $\kappa_m \asymp m^{2a+(2s)_{\vee 0}}$ with $|\log(\kappa_m \vee (m+2))/\log(m+2)| \asymp (2a+(2s)_{\vee 0}) > 1$ and hence, $\delta_m \asymp m^{1+2a+2s}$. Therefore, the result follows from Proposition 3.3 case [P-P] and [E-P] since both sequences δ and Δ are unchanged.

Case [P-E] We have $\Delta_m \asymp m^{2s}\exp(m^{2a})$, $\kappa_m \asymp m^{(2s)_{\vee 0}}\exp(m^{2a})$ with, for all m sufficiently large, $\log(\kappa_m \vee (m+2))/\log(m+2)| \asymp m^{2a}\frac{(1+(2s)_{\vee 0}(\log m)m^{-2a})}{\log(m+2)}$ and hence $\delta_m \asymp m^{1+2a+2s}\exp(m^{2a})\frac{(1+(2s)_{\vee 0}(\log m)m^{-2a})}{\log m}$. Then straightforward calculus shows that Assumption 3.1 (i) is fulfilled. Moreover, consider the sequence M given in Assumption 3.1 (ii), where $M_n \asymp (\log \frac{n\,(\log\log n)/(2a)}{(\log n)^{(1+2a+(2s)_{\vee 0})/(2a)}})^{1/(2a)} = (\log n)^{1/(2a)}\Big(1+o(1)\Big)$, then also Assumption 3.1 (ii) is satisfied (as in the proof of case [P-E] in Proposition 3.3). Due to Corollary 3.2 it remains to balance $n \asymp \gamma_{m^\diamond}\delta_{m^\diamond}/\omega_{m^\diamond} \asymp (m^\diamond)^{1+2a+2p}\exp((m^\diamond)^{2a})/(\log m^\diamond)$ which implies $m_n^\diamond \asymp (\log \frac{n\,(\log\log n)/(2a)}{(\log n)^{(1+2a+2p)/(2a)}})^{1/(2a)} = (\log n)^{1/(2a)}\Big(1+o(1)\Big)$. Hence, $\omega_{m^\diamond}/\gamma_{m^\diamond} \asymp (\log n)^{(p-s)/a}$ is the rate attained by the adaptive estimator $\widehat{\beta}_{\widehat{m}}$ which is optimal and completes the proof of Proposition 3.4. $\square$

A.4 Proof of Theorem 4.1

We begin by defining additional notations to be used in the proof. Consider sequences δ, $\triangle$, M and $m^\diamond$ satisfying Assumption 4.1 and the random upper bound $\widehat{M}$ defined in (4.1). Denote by $\Omega := \Omega_I \cap \Omega_{II}$ the event given by

$$\Omega_I := \left\{\forall j \in \{1,\dots,M_n\}, \left|\frac{1}{\widehat{\lambda}_j} - \frac{1}{\lambda_j}\right| < \frac{1}{2\lambda_j} \text{ and } \widehat{\lambda}_j \geqslant 1/n\right\},$$
$$\Omega_{II} := \{m_n^\diamond \leqslant \widehat{M}_n \leqslant M_n\}.$$

It is easily seen that on Ω_I we have for all $1 \leqslant m \leqslant M_n$

$$(1/2)\Delta_m \leqslant \widehat{\Delta}_m \leqslant (3/2)\Delta_m \qquad \text{and} \qquad (1/2)\kappa_m \leqslant \widehat{\kappa}_m \leqslant (3/2)\kappa_m$$

and hence $(1/2)[\kappa_m \vee (m+2)] \leqslant [\widehat{\kappa}_m \vee (m+2)] \leqslant (3/2)[\kappa_m \vee (m+2)]$ which implies

$$\begin{aligned}(1/2)m\Delta_m\Big(\frac{\log[\kappa_m \vee (m+2)]}{\log(m+2)}\Big)\Big(1 - \frac{\log 2}{\log(m+2)}\frac{\log(m+2)}{\log(\kappa_m \vee [m+2])}\Big) \leqslant \widehat{\delta}_m \\ \leqslant (3/2)m\Delta_m\Big(\frac{\log(\kappa_m \vee [m+2])}{\log(m+2)}\Big)\Big(1 + \frac{\log 3/2}{\log(m+2)}\frac{\log(m+2)}{\log(\kappa_m \vee [m+2])}\Big),\end{aligned}$$

together with $\log(\kappa_m \vee [m+2])/\log(m+2) \geqslant 1$ we get

$$\delta_m/10 \leqslant (\log 3/2)/(2\log 3)\delta_m \leqslant (1/2)\delta_m[1-(\log 2)/\log(m+2)] \leqslant \widehat{\delta}_m \\ \leqslant (3/2)\delta_m[1+(\log 3/2)/\log(m+2)] \leqslant 3\delta_m.$$

Since $\mathrm{pen}(m) = 192\sigma_Y^2\eta\delta_m n^{-1}$ and $\widehat{\mathrm{pen}}(m) = 1920\sigma_Y^2\eta\widehat{\delta}_m n^{-1}$ it follows on Ω_I that $\mathrm{pen}(m) \leqslant \widehat{\mathrm{pen}}(m) \leqslant 30\,\mathrm{pen}(m)$ for all $1 \leqslant m \leqslant M_n$, and hence

$$\Big(\mathrm{pen}(m_n^\diamond \vee \widehat{m}) + \widehat{\mathrm{pen}}(m_n^\diamond) - \widehat{\mathrm{pen}}(\widehat{m})\Big)\mathbb{1}_\Omega \leqslant \Big(\mathrm{pen}(m_n^\diamond) + \mathrm{pen}(\widehat{m}) + \widehat{\mathrm{pen}}(m_n^\diamond) - \widehat{\mathrm{pen}}(\widehat{m})\Big)\mathbb{1}_\Omega \\ \leqslant 31\,\mathrm{pen}(m_n^\diamond)$$

by using $1 \leqslant \widehat{m} \leqslant \widehat{M}_n$ and $m_n^\diamond \leqslant \widehat{M}_n \leqslant M_n$. On the other hand, it is not hard to see that on Ω_I^c we have $\widehat{\Delta}_m \leqslant n \max_{1 \leqslant j \leqslant m} \omega_j$ and $\widehat{\kappa}_m \leqslant n$ for all $m \geqslant 1$. From these properties we conclude that for all $1 \leqslant m \leqslant M_n$

$$\widehat{\delta}_m \leqslant mn(\max_{1\leqslant j \leqslant m} \omega_j)\frac{\log(n \vee (m+2))}{\log(m+2)} \leqslant mn(\max_{1\leqslant j\leqslant m}\omega_j)\log(n+2), \tag{A.20}$$

which implies $\widehat{\mathrm{pen}}(m_n^\diamond) \leqslant 1920\sigma_Y^2\eta M_n(\max_{1\leqslant j \leqslant M_n}\omega_j)\log(n+2)$ and hence

$$\Big(\mathrm{pen}(m_n^\diamond \vee \widehat{m}) + \widehat{\mathrm{pen}}(m_n^\diamond) - \widehat{\mathrm{pen}}(\widehat{m})\Big)\mathbb{1}_{\Omega_I^c \cap \Omega_{II}} \\ \leqslant \Big(\mathrm{pen}(M_n) + 1920\sigma_Y^2\eta M_n(\max_{1\leqslant j\leqslant M_n}\omega_j)\log(n+2)\Big)\mathbb{1}_{\Omega_I^c \cap \Omega_{II}} \\ \leqslant 1920\sigma_Y^2\eta\Big(\delta_{M_n}/n + M_n(\max_{1\leqslant j \leqslant M_n}\omega_j)\log(n+2)\Big)\mathbb{1}_{\Omega_I^c\cap\Omega_{II}} \tag{A.21}$$

We shall prove in the end of this section the technical Lemma A.5 which is used in the following steps of the proof together with the technical Lemmas A.2 - A.4 above.

Consider now the decomposition

$$\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 = \mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2\mathbb{1}_\Omega + \mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2\mathbb{1}_{\Omega_I^c\cap\Omega_{II}} + \mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2\mathbb{1}_{\Omega_{II}^c}. \tag{A.22}$$

Below we show that there exist a numerical constant $C' > 0$ and a constant $K' = K'(\Sigma, \eta, \xi, \delta_1)$ only depending on Σ, η, ξ and δ_1 such that for all $n \geqslant 1$ we have

$$\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2\mathbb{1}_\Omega \leqslant C'\Big\{\|\beta - \beta_{m_n^\diamond}\|_\omega^2 + \frac{\delta_{m_n^\diamond}}{n}\,\sigma_Y^2\eta + \frac{K'}{n}\,\sigma_Y^2\,[\delta_1 + \|\beta\|_\omega^2][1 + (\mathbb{E}\|X\|^2)^2]\Big\} \tag{A.23}$$

$$\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2\mathbb{1}_{\Omega_I^c\cap\Omega_{II}} \leqslant C'\Big\{\|\beta - \beta_{m_n^\diamond}\|_\omega^2 + \frac{K'}{n}\,\sigma_Y^2\,[\delta_1 + \|\beta\|_\omega^2][1 + (\mathbb{E}\|X\|^2)^2]\Big\} \tag{A.24}$$

$$\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2\mathbb{1}_{\Omega_{II}^c} \leqslant C'\frac{\xi}{n}\,[\sigma_Y^2 + \|\beta\|_\omega^2][1 + \mathbb{E}\|X\|^2]. \tag{A.25}$$

Since (ω/γ) is monotonically non increasing we obtain in case $\beta \in \mathcal{F}_\gamma^\rho$ that $\|\beta\|_\omega^2 \leqslant \rho$ and $\|\beta - \beta_{m_n^\diamond}\|_\omega^2 \leqslant (\omega_{m_n^\diamond}/\gamma_{m_n^\diamond})\rho$. Moreover, we have $\sigma_Y^2 \leqslant \rho\mathbb{E}\|X\|^2 + \sigma^2$. From these properties by combining the decomposition (A.22) and the estimates (A.23) - (A.25) we conclude that there exists a numerical constant $C>0$ and a constant $K=K(\Sigma,\eta,\xi,\delta_1)$ only depending on Σ,η,ξ and δ_1 such that for all $n\geqslant 1$

$$\mathbb{E}\|\widehat{\beta}_{\widehat{m}}-\beta\|_\omega^2\leqslant C\Big\{\frac{\omega_{m_n^\diamond}}{\gamma_{m_n^\diamond}}\rho+\frac{\delta_{m_n^\diamond}}{n}\,[\rho\mathbb{E}\|X\|^2+\sigma^2]\eta+\frac{K}{n}\,[\rho\mathbb{E}\|X\|^2+\sigma^2]\,[1+\delta_1+\rho][1+(\mathbb{E}\|X\|^2)^2]\Big\}.$$

The result follows now from the definition of $m_n^\diamond$, that is, $\gamma_{m_n^\diamond}\delta_{m_n^\diamond}/(n\,\omega_{m_n^\diamond})\leqslant \underline{c}$.

Proof of (A.23). Observe that on Ω we have $m_n^\diamond\leqslant\widehat{M}_n\leqslant M_n$. Thus, following line by line the proof of (A.4) it is easily seen that

$$\begin{aligned}(1/4)\|\widehat{\beta}_{\widehat{m}}-\beta\|_\omega^2\mathbb{1}_\Omega &\leqslant (7/4)\|\beta-\beta_{m_n^\diamond}\|_\omega^2+32\sum_{m=1}^{M_n}\Big(\sup_{t\in\mathcal{B}_m}|\langle t,\widetilde{\Phi}_{\widehat{h}}\rangle_\omega|^2-6\sigma_Y^2\eta\delta_m/n\Big)_+\\&\quad+32\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2+32\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widehat{\Phi}_{\widehat{g}}-\widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2\\&\quad+\Big(\mathrm{pen}(m_n^\diamond\vee\widehat{m})+\widehat{\mathrm{pen}}(m_n^\diamond)-\widehat{\mathrm{pen}}(\widehat{m})\Big)\mathbb{1}_\Omega\\&\leqslant (7/4)\|\beta-\beta_{m_n^\diamond}\|_\omega^2+32\sum_{m=1}^{M_n}\Big(\sup_{t\in\mathcal{B}_{m'}}|\langle t,\widetilde{\Phi}_{\widehat{h}}\rangle_\omega|^2-6\sigma_Y^2\eta\delta_m/n\Big)_+\\&\quad+32\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2+32\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widehat{\Phi}_{\widehat{g}}-\widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2\\&\quad+4\,\mathrm{pen}(m_n^\diamond),\end{aligned}$$

where the last inequality follows from (A.20). Combining the last bound with (A.5) in Lemma A.2, (A.9) and (A.10) in Lemma A.3 we conclude that there exists a numerical constant $C'>0$ and a constant $K'=K'(\Sigma,\eta,\xi,\delta_1)$ depending on Σ,η,ξ,δ_1 only such that (A.23) for all $n\geqslant 1$ holds true.

Proof of (A.24). Note that on $\Omega_I^c\cap\Omega_{II}$ we have still $m_n^\diamond\leqslant\widehat{M}_n\leqslant M_n$. Thus, by using (A.21) rather than (A.20) it follows in analogy to (A.22) that

$$\begin{aligned}(1/4)\|\widehat{\beta}_{\widehat{m}}-\beta\|_\omega^2\mathbb{1}_{\Omega_I^c\cap\Omega_{II}}&\leqslant (7/4)\|\beta-\beta_{m_n^\diamond}\|_\omega^2+32\sum_{m=1}^{M_n}\Big(\sup_{t\in\mathcal{B}_m}|\langle t,\widetilde{\Phi}_{\widehat{h}}\rangle_\omega|^2-6\sigma_Y^2\eta\delta_m/n\Big)_+\\+32\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2&+32\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widehat{\Phi}_{\widehat{g}}-\widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2+\Big(\mathrm{pen}(m_n^\diamond\vee\widehat{m})+\widehat{\mathrm{pen}}(m_n^\diamond)-\widehat{\mathrm{pen}}(\widehat{m})\Big)\mathbb{1}_{\Omega_I^c\cap\Omega_{II}}\\&\leqslant (7/4)\|\beta-\beta_{m_n^\diamond}\|_\omega^2+32\sum_{m=1}^{M_n}\Big(\sup_{t\in\mathcal{B}_m}|\langle t,\widetilde{\Phi}_{\widehat{h}}\rangle_\omega|^2-6\sigma_Y^2\eta\delta_m/n\Big)_+\\&\quad+32\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widetilde{\Phi}_{\widehat{f}}\rangle_\omega|^2+32\sup_{t\in\mathcal{B}_{M_n}}|\langle t,\widehat{\Phi}_{\widehat{g}}-\widetilde{\Phi}_{\widehat{g}}\rangle_\omega|^2\\&\quad+1920\sigma_Y^2\eta\Big(\delta_{M_n}/n+M_n(\max_{1\leqslant j\leqslant M_n}\omega_j)\log(n+2)\Big)\mathbb{1}_{\Omega_I^c\cap\Omega_{II}}.\end{aligned}$$

From the last bound together with (A.5) in Lemma A.2, (A.9) and (A.10) in Lemma A.3 we conclude that there exist a numerical constant $C>0$ and a constant $K=K(\Sigma,\eta,\xi,\delta_1)$ depending on Σ,η,ξ and δ_1 only such that for all $n\geqslant 1$ we have

$$\begin{aligned}\mathbb{E}\|\widehat{\beta}_{\widehat{m}}-\beta\|_\omega^2\mathbb{1}_{\Omega_I^c\cap\Omega_{II}}\leqslant C\Big\{&\|\beta-\beta_{m_n^\diamond}\|_\omega^2+\frac{K}{n}\,\sigma_Y^2\,[\delta_1+\|\beta\|_\omega^2][1+(\mathbb{E}\|X\|^2)^2]\\&+\sigma_Y^2\eta\Big(n^{-1}\delta_{M_n}+n^{-2}M_n(\max_{1\leqslant j\leqslant M_n}\omega_j)\Big)n^2\log(n+2)P(\Omega_I^c\cap\Omega_{II})\Big\}.\end{aligned}\tag{A.26}$$

Since $X \in \mathcal{X}_{\xi}^{24}$ and $\Omega_I^c \cap \Omega_{II} \subset \{\exists j \in \{1, \dots, M_n\} : |\lambda_j/\widehat{\lambda}_j - 1| > 1/2 \text{ or } \widehat{\lambda}_j < 1/n\}$ it follows from (A.29) in Lemma A.29 that $P(\Omega_I^c \cap \Omega_{II}) \leqslant C\xi\, M_n\, n^{-6}$ for some numerical constant $C > 0$. Moreover, due to Assumption 4.1 we have $\delta_{M_n}/n \leqslant \delta_1$, $M_n/n \leqslant 1$ and $\max_{1 \leqslant j \leqslant M_n} \omega_j \leqslant \max_{1 \leqslant j \leqslant N_n} \omega_j \leqslant n$. Combining the last estimates and (A.26) implies now (A.24).

Proof of (A.25). Let $\breve{\beta}_m := \sum_{j=1}^m [\beta]_j \mathbb{1}\{\widehat{\lambda}_j \geqslant 1/n\}\varphi_j$. Then it is not hard to see that $\|\widehat{\beta}_m - \breve{\beta}_m\|_\omega^2 \leqslant \|\widehat{\beta}_{m'} - \breve{\beta}_{m'}\|_\omega^2$ for all $m \leqslant m'$ and $\|\breve{\beta}_m - \beta\|_\omega^2 \leqslant \|\beta\|_\omega^2$. By using these properties together with $1 \leqslant \widehat{m} \leqslant \widehat{M}_n \leqslant N_n$ we conclude

$$\begin{aligned}
\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \mathbb{1}_{\Omega_{II}^c} &\leqslant 2\{\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \breve{\beta}_{\widehat{m}}\|_\omega^2 \mathbb{1}_{\Omega_{II}^c} + \mathbb{E}\|\breve{\beta}_{\widehat{m}} - \beta\|_\omega^2 \mathbb{1}_{\Omega_{II}^c}\} \\
&\leqslant 2\{\mathbb{E}\|\widehat{\beta}_{N_n} - \breve{\beta}_{N_n}\|_\omega^2 \mathbb{1}_{\Omega_{II}^c} + \|\beta\|_\omega^2 P(\Omega_{II}^c)\}.
\end{aligned}$$

Since $X \in \mathcal{X}_\xi^{28}$ and $\Omega_{II}^c = \{\widehat{M}_n < m_n^\diamond\} \cup \{\widehat{M}_n > M_n\}$ it follows from (A.30) and (A.31) in Lemma A.5 that $P(\Omega_{II}^c) \leqslant C\xi n^{-6}$ for some numerical constant $C > 0$ and hence

$$\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \mathbb{1}_{\Omega_{II}^c} \leqslant 2\{\mathbb{E}\|\widehat{\beta}_{N_n} - \breve{\beta}_{N_n}\|_\omega^2 \mathbb{1}_{\Omega_{II}^c} + C\xi\, \|\beta\|_\omega^2 n^{-6}\}. \tag{A.27}$$

Moreover, from (A.16) and (A.17) in Lemma A.4 together with $X \in \mathcal{X}_\xi^{28}$ and $\mathbb{E}|Y/\sigma_Y|^{28} \leqslant \xi$ it follows that there exists a numerical constant $C > 0$ such that

$$\begin{aligned}
\mathbb{E}\|\widehat{\beta}_{N_n} - \breve{\beta}_{N_n}\|_\omega^2 \mathbb{1}_{\Omega_{II}^c} &\leqslant 2n^2 \sum_{j=1}^{N_n} \omega_j \Big\{\mathbb{E}([\widehat{g}]_j - \lambda_j[\beta]_j)^2 \mathbb{1}_{\Omega_{II}^c} + \mathbb{E}(\lambda_j[\beta]_j - \widehat{\lambda}_j[\beta]_j)^2 \mathbb{1}_{\Omega_{II}^c}\Big\} \\
&\leqslant 2n^2 \Big\{\max_{1 \leqslant j \leqslant N_n} \omega_j \sum_{j=1}^{N_n} \lambda_j \Big[\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n Y_i \frac{[X_i]_j}{\sqrt{\lambda_j}} - \sqrt{\lambda_j}[\beta]_j\right)^4\Big]^{1/2} P(\Omega_{II}^c)^{1/2} \\
&\qquad + \max_{j \geqslant 1} \lambda_j \sum_{j=1}^{N_n} \omega_j [\beta]_j^2 [\mathbb{E}(\widehat{\lambda}_j/\lambda_j - 1)^4]^{1/2} P(\Omega_{II}^c)^{1/2}\Big\} \\
&\leqslant C\xi n^2 \Big\{n^{-4}\sigma_Y^2 \max_{1 \leqslant j \leqslant N_n} \omega_j \sum_{j \geqslant 1} \lambda_j + n^{-4} \max_{j \geqslant 1} \lambda_j \|\beta\|_\omega^2\Big\}.
\end{aligned} \tag{A.28}$$

By combination of (A.27), (A.28) and $\mathbb{E}\|X\|^2 = \sum_{j \geqslant 1} \lambda_j \geqslant \max_{j \geqslant 1} \lambda_j$ we obtain

$$\mathbb{E}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \mathbb{1}_{\Omega_{II}^c} \leqslant C'\Big\{n^{-2}\sigma_Y^2 \xi \max_{1 \leqslant j \leqslant N_n} \omega_j \mathbb{E}\|X\|^2 + \xi\{1 + \mathbb{E}\|X\|^2\}\|\beta\|_\omega^2 n^{-2}\Big\},$$

for some numerical constant $C' > 0$. The estimate (A.25) follows now from $\max_{1 \leqslant j \leqslant N_n} \omega_j \leqslant n$ (Assumption 4.1), which completes the proof of Theorem 4.1. □

Technical assertions.

The following lemma gathers technical results used in the proof of Theorem 4.1.

Lemma A.5. *Suppose $X \in \mathcal{X}_{\eta_{4k}}^{4k}$, $k \geqslant 1$, with associated sequence λ of eigenvalues. Let M and $m^\diamond$ be sequences satisfying Assumption 4.1. Then there exist a numerical constant $C_k > 0$ only depending on k such that for all $n \geqslant 1$ we have*

$$P(\{\exists j \in \{1, \dots, M_n\} : |\lambda_j/\widehat{\lambda}_j - 1| > 1/2 \text{ or } \widehat{\lambda}_j < 1/n\}) \leqslant C_k \eta_{4k}\, M_n\, n^{-k}, \tag{A.29}$$

$$P(\widehat{M}_n < m_n^\diamond) \leqslant C_k \eta_{4k}\, n^{-k} \quad \text{and} \tag{A.30}$$

$$P(\widehat{M}_n > M_n) \leqslant C_k \eta_{4k}\, n^{-k+1} \quad \text{for all } n \geqslant 1. \tag{A.31}$$

Proof. Proof of (A.29). We start our proof with the observation that the event $\{|\lambda_j/\widehat{\lambda}_j - 1| > 1/2\}$ can equivalently be written as $\{1 - \widehat{\lambda}_j/\lambda_j > 1/3 \text{ or } \widehat{\lambda}_j/\lambda_j - 1 > 1\}$, and hence is a subset of $\{|\widehat{\lambda}_j/\lambda_j - 1| > 1/3\}$. Moreover, since $\lambda_j \geqslant 2/n$ for all $1 \leqslant j \leqslant M_n$ it follows that $\{\widehat{\lambda}_j < 1/n\} \subset \{|\widehat{\lambda}_j/\lambda_j - 1| > 1/2\}$. Combining both estimates we conclude

$$P(\{\exists j \in \{1, \dots, M_n\} : |\lambda_j/\widehat{\lambda}_j - 1| > 1/2 \text{ or } \widehat{\lambda}_j < 1/n\})$$
$$\leqslant \sum_{j=1}^{M_n} \{P(|\widehat{\lambda}_j/\lambda_j - 1| > 1/3) + P(|\widehat{\lambda}_j/\lambda_j - 1| > 1/2)\} \leqslant 2 \sum_{j=1}^{M_n} P(|\widehat{\lambda}_j/\lambda_j - 1| > 1/3).$$

Thus applying Markov's inequality together with (A.17) in Lemma A.4 implies (A.29).

Proof of (A.30). Due to the definition of $\widehat{M}_n$ given in (4.1) the event $\{\widehat{M}_n < m_n^\diamond\}$ is a subset of $\{\forall m \in \{m_n^\diamond, \dots, n\} : \widehat{\lambda}_m/(\omega_m)_{\vee 1} < m(\log n)/n\}$ and hence $P(\widehat{M}_n < m_n^\diamond) \leqslant P(\widehat{\lambda}_{m_n^\diamond}/\lambda_{m_n^\diamond} < 1/2)$ since $\min_{1 \leqslant m \leqslant m_n^\diamond} \lambda_m/[m(\omega_m)_{\vee 1}] \geqslant 2(\log n)/n$ (Assumption 4.1 (iii)). Thereby, (A.30) follows from the second bound in (A.18) in Lemma A.17.

Proof of (A.31). Due to the definition (4.1) of $\widehat{M}_n$ for $m > M_n$ the event $\{\widehat{M}_n = m\}$ is a subset of $\{\widehat{\lambda}_m/(\omega_m)_{\vee 1} \geqslant m(\log n)/n\}$ and hence $P(\widehat{M}_n > M_n) \leqslant \sum_{j=M_n+1}^{N_n} P(\widehat{\lambda}_m/\lambda_m \geqslant 2)$ since $2 \max_{m > M_n} \lambda_m/[m(\omega_m)_{\vee 1}] \leqslant (\log n)/n$ (Assumption 4.1 (ii)). Thereby, the first bound in (A.18) in Lemma A.17 together with $N_n/n \leqslant 1$ (Assumption 4.1 (iv)) implies (A.31), which completes the proof of Lemma A.5. □

A.5 Proof of Corollary 4.2

First, note that in all three cases, the sequences δ, Δ, M and $m^\diamond$ have been calculated in the proof of Proposition 3.4. If in addition Assumption 4.1 holds true, then from Theorem 4.1 follows that the fully adaptive estimator attains the rate $\omega_{m_n^\diamond}/\gamma_{m_n^\diamond}$, which in the proof of Proposition 3.4 has been confirmed to be optimal in all three cases. Therefore it only remains to check (i)-(iii) of Assumption 4.1.

Case [P-P] In this case, we have $M_n \asymp n^{1/(2a+1+(2s)_{\vee 0})}$ and $m_n^\diamond \asymp n^{1/(2a+2p+1)}$. Then (i) of Assumption 4.1 holds true, since $\min_{1 \leqslant j \leqslant M_n} \lambda_j \asymp M_n^{-2a} \asymp n^{-2a/(2a+1+(2s)_{\vee 0})} \geqslant 2/n$ and

$$\max_{m \geqslant M_n} \frac{\lambda_m}{m(\omega_m)_{\vee 1}} \asymp M_n^{-1-2a-(2s)_{\vee 0}} \asymp n^{-(2a+1+(2s)_{\vee 0})/(1+2a+(2s)_{\vee 0})} \leqslant (\log n)/(2n).$$

Moreover (ii) of Assumption 4.1 is satisfied by using that for all $p > s$

$$\min_{1 \leqslant m \leqslant m_n^\diamond} \frac{\lambda_m}{m(\omega_m)_{\vee 1}} \asymp (m_n^\diamond)^{-1-2a-(2s)_{\vee 0}} \asymp n^{-(2a+1+(2s)_{\vee 0})/(2p+1-2s+(2a+2s)_{\vee 0})} \geqslant 2(\log n)/n.$$

Finally, consider (iii) of Assumption 4.1. It is easily verified that $N_n \asymp n^{1/(1+(2s)_{\vee 0})}$ which satisfies $\max_{1 \leqslant m \leqslant N_n} \omega_m \leqslant N_n^{(2s)_{\vee 0}} \asymp n^{(2s)_{\vee 0}/(1+(2s)_{\vee 0})} \leqslant n$ and $M_n \asymp n^{1/(2a+1+(2s)_{\vee 0})} \leqslant N_n \leqslant n$. Thereby also (iii) of Assumption 4.1 holds true.

Case [E-P]. We have $M_n \asymp n^{1/(2a+1+(2s)_{\vee 0})}$, $m_n^\diamond \asymp \{\log[n(\log n)^{-(2a+1)/(2p)}]\}^{1/(2p)}$ and $N_n \asymp n^{1/(1+(2s)_{\vee 0})}$. Then as in case [P-P] (i) and (iii) of Assumption 4.1 hold true since M_n and N_n are unchanged. Furthermore, for all $s \in \mathbb{R}$ we have

$$\min_{1 \leqslant m \leqslant m_n^\diamond} \frac{\lambda_m}{m(\omega_m)_{\vee 1}} \asymp (m_n^\diamond)^{-1-2a-(2s)_{\vee 0}} \asymp (\log n)^{-(2a+1+(2s)_{\vee 0})/(2p)} \geqslant 2(\log n)/n,$$

which shows (ii) of Assumption 4.1.

Case [P-E]. Here we have $M_n \asymp (\log \frac{n\,(\log\log n)/(2a)}{(\log n)^{(1+2a+(2s)_{\vee 0})/(2a)}})^{1/(2a)} = (\log n)^{1/(2a)}(1+o(1))$, $m_n^\diamond \asymp (\log \frac{n\,(\log\log n)/(2a)}{(\log n)^{(1+2a+2p)/(2a)}})^{1/(2a)} = (\log n)^{1/(2a)}(1+o(1))$ and $N_n \asymp n^{1/(1+(2s)_{\vee 0})}$. It is easily seen that (iii) of Assumption 4.1 is satisfied. Moreover, (i) of Assumption 4.1 holds true, since $\min_{1\leqslant j\leqslant M_n}\lambda_j \asymp \exp(-M_n^{2a}) \asymp \frac{(\log n)^{(1+2a+(2s)_{\vee 0})/(2a)}}{n(\log\log n)/(2a)} \geqslant 2/n$ and

$$\max_{m\geqslant M_n}\frac{\lambda_m}{m(\omega_m)_{\vee 1}} \asymp M_n^{-1-(2s)_{\vee 0}}\exp(-M_n^{2a}) \asymp \frac{(\log n)}{n(\log\log n)/(2a)} \leqslant (\log n)/(2n).$$

Finally, consider (ii) of Assumption 4.1 which is satisfied by using that for all $p > s$

$$\min_{1\leqslant m\leqslant m_n^\diamond}\frac{\lambda_m}{m(\omega_m)_{\vee 1}} \asymp (m_n^\diamond)^{-1-(2s)_{\vee 0}}\exp(-(m_n^\diamond)^{2a}) \asymp \frac{(\log n)^{(2a+2p-(2s)_{\vee 0})/(2a)}}{n\,(\log\log n)/(2a)} \geqslant 2(\log n)/n,$$

which completes the proof of Corollary 4.2. □

References

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

H. Cardot and J. Johannes. Thresholding projection estimators in functional linear models. *forthcoming in the Journal of Multivariate Analysis*, 2009.

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

F. Comte, Y. Rozenholc, and M.-L. Taupin. Penalized contrast estimator for density deconvolution. *Canadian Journal of Statistics*, 37(3), 2006.

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

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems.* Kluwer Academic, Dordrecht, 2000.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations.* Springer-Verlag, London, 2006.

M. Forni and L. Reichlin. Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies*, 65:453–473, 1998.

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

G. M. James, J. Wang, and J. Zhu. Functional linear regression that's interpretable. Technical report, To appear in the Annals of Statistics., 2009.

J. Johannes. Nonparametric estimation in circular functional linear model. Technical report, University Heidelberg (revised and submitted), 2009. URL http://arxiv.org/abs/0901.4266v1.

B. A. Mair and F. H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM Journal on Applied Mathematics*, 56(5):1424–1444, 1996.

P. Massart. *Concentration inequalities and model selection*, volume 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.

H.-G. Müller and U. Stadtmüller. Generalized functional linear models. *Ann. Stat.*, 33: 774–805, 2005.

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

M. H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7:307–330, 1997.

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

C. Preda and G. Saporta. Pls regression on a stochastic process. *Computational Statistics & Data Analysis*, 48:149 –158, 2005.

J. Ramsay and B. Silverman. *Functional Data Analysis.* Springer, New York, second ed. edition, 2005.

M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3): 505–563, 1996.

U. Tautenhahn. Error estimates for regularization methods in Hilbert scales. *SIAM Journal on Numerical Analysis*, 33(6):2120–2130, 1996.