Available online at www.sciencedirect.com

SciVerse ScienceDirect

Stochastic Processes and their Applications 122 (2012) 3680–3700

stochastic processes and their applications

www.elsevier.com/locate/spa

Linear prediction in functional data analysis

Hyejin Shin, Tailen Hsing*

Seoul National University, South Korea
University of Michigan, USA

Received 18 August 2011; received in revised form 18 June 2012; accepted 27 June 2012
Available online 6 July 2012

Abstract

In this paper we introduce a new perspective of linear prediction in the functional data context that predicts a scalar response by observing a functional predictor. This perspective broadens the scope of functional linear prediction currently in the literature, which is exclusively focused on the functional linear regression model. It also provides a natural link to the classical linear prediction theory. Based on this formulation, we derive the convergence rate of the optimal mean squared predictor.

Keywords: Hilbert space; Unbounded linear functional; Principal components; Convergence rate

1. Introduction

Nowadays data that can be viewed as digitized recordings of random functions are commonplace in statistical problems. They are often referred to as functional data, which have received an increasing level of attention over the past decade. Functional data analysis generally involves data containing a large number of (highly correlated) variables relative to sample size. As a result, direct applications of classical (multivariate) techniques often do not work well and many new approaches have been developed for such data. See [17,8] for an introduction.

We briefly describe the common setting of functional data analysis. Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space. Throughout this paper let $\{X(t) : t \in \mathcal{T}\}$ be a zero-mean, second-order stochastic process defined on $(\Omega, \mathcal{B}, \mathbb{P})$, where $\mathcal{T}$ is a bounded interval. Denote the covariance function of the process X by K_X. We will make the common assumption $\mathbb{P}(X \in L^2(\mathcal{T})) = 1$

* Correspondence to: Department of Statistics, University of Michigan, Ann Arbor, MI 48103, USA.
E-mail address: thsing@umich.edu (T. Hsing).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2012.06.014

where $L^2(\mathcal{T})$ denotes the Hilbert space of functions g on $\mathcal{T}$ with $\int_{\mathcal{T}} g^2(t)dt < \infty$, equipped with the inner product

$$\langle f, g \rangle = \int_{\mathcal{T}} f(t)g(t)dt.$$

We will further make the common assumption that $\mathbb{E}(\|X\|^2) < \infty$ and K_X is continuous on $\mathcal{T} \times \mathcal{T}$. Let $\mathfrak{K}_X$ be the covariance operator of X, namely, $(\mathfrak{K}_X g)(t) = \int_{\mathcal{T}} K_X(s,t)g(s)ds$, $g \in L^2(\mathcal{T})$. Under the assumptions stated above, $\mathfrak{K}_X$ is compact (in fact, trace-class) and has eigenvalues and eigenfunctions denoted by $\lambda_j, \phi_j, j \geq 1$, where we assume that $\lambda_1 > \lambda_2 > \cdots > 0$ throughout the paper for convenience and to emphasize the infinite-dimensional nature of functional data. We follow the literature and call the ϕ_j's the principal components of X. Let $Y \in \mathbb{R}$ be a scalar variable whose value partially depends on X. Assume that we have an i.i.d. sample (X_i, Y_i), $1 \leq i \leq n$, and we are interested in the problem of predicting a new Y for a new X.

A common model in functional data analysis is the following regression model:

$$Y_i = \mu + \langle b, X_i \rangle + \varepsilon_i, \quad 1 \leq i \leq n, \tag{1.1}$$

where μ is the mean of Y_i, b is a function in $L^2(\mathcal{T})$ and ε_i's are random errors with mean zero and finite variance. See, for example, [3,4,7,19,2,10,13,6,5,20]. Often, b is assumed to belong to some space of "smooth" functions $\mathcal{F}$. A common choice of $\mathcal{F}$, which we adopt momentarily in this discussion, is the Sobolev space of functions f which are m-times differentiable with $J(f) := \int_{\mathcal{T}}[f^{(m)}(x)]^2dx < \infty$. A well-known estimator of μ, b is

$$(\widehat{\mu}, \widehat{b}) := \underset{\mu \in \mathbb{R}, b \in \mathcal{F}}{\operatorname{argmin}} \left[\sum_{i=1}^{n} \{Y_i - \mu - \langle b, X_i \rangle\}^2 + \lambda J(b) \right], \tag{1.2}$$

where $\lambda > 0$ is a smoothing parameter; see, e.g., [6]. In practice, λ is usually selected by a data-driven criterion, e.g., GCV, that minimizes some prediction error. Then predict a new Y based on a new X by the predictor

$$\widehat{\eta} := \widehat{\mu} + \langle \widehat{b}, X \rangle. \tag{1.3}$$

A different approach is functional principal component regression (FPCR), in which one normally first estimates the cross covariance $K_{YX}(t) := \mathbb{E}(YX(t))$ and the eigenvalues/functions, (λ_j, ϕ_j), using data and then estimate μ, b with

$$\widehat{\mu} := \overline{Y} \quad \text{and} \quad \widehat{b} := \sum_{j=1}^{m} \frac{\langle \widehat{K}_{YX}, \widehat{\phi}_j \rangle}{\widehat{\lambda}_j} \widehat{\phi}_j, \tag{1.4}$$

where the cutoff $m \leq n$ can be selected by, for instance, GCV. For a future X, predict the corresponding Y by

$$\widehat{\eta} := \overline{Y} + \sum_{j=1}^{m} \frac{\langle \widehat{K}_{YX}, \widehat{\phi}_j \rangle}{\widehat{\lambda}_j} \langle X, \widehat{\phi}_j \rangle. \tag{1.5}$$

Note that, here, m plays the smoothing parameter and the slope function b is not required to be smooth. Thus, the FPCR approach is inherently different from the roughness penalty approach discussed earlier in (1.2). The estimation error of the regression slope function resulting from

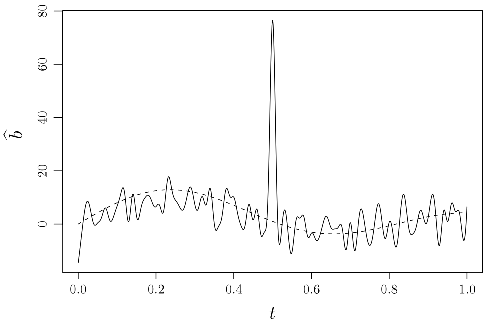

Fig. 1. The dashed line is the function f in (1.6); the solid line is the estimate of the slope function b in the hypothesized linear regression model.

FPCR was studied by Yao et al. [19] and Hall and Horowitz [10] while the prediction error of the FPCR-based predictor was studied by Cardot et al. [4] and Cai and Hall [2]. Cardot and Johannes [5] considered an approach in which functional data are projected onto known basis functions which could be principal components; both estimation and prediction errors were studied in their work.

To motivate the problem in this paper, consider the following model

$$Y_i = \mu + X_i(.5) + \langle f, X_i \rangle + \varepsilon_i, \quad 1 \le i \le n, \tag{1.6}$$

where X is a standard Brownian motion on [0, 1] and f, described by the dashed line in Fig. 1, is a smooth function. With the presence of the "discrete component", $X(.5)$, in the regressor, (1.6) is arguably a substantially different model from (1.1). However, if we pretend that (1.1) holds and continue to use (1.3) for prediction, the results turn out to be surprisingly satisfactory. To see what happens in this case we present the outcome of a simulation run with sample size 500. The solid line in Fig. 1 is the estimate of b in the hypothesized model (1.1) using (1.2). Despite the disconnect between the hypothesized model (1.1) and the true model (1.6), the procedure nevertheless produced an estimate, $\widehat{b}$, of b in the hypothesized model that makes sense from the prediction perspective, where the effect of the discrete component $X(.5)$ is accounted for by a spike. The results are similar if (1.4) and (1.5) are used instead.

One might conjecture from this example that the predictor $\widehat{\eta}$ in both (1.3) and (1.5) are robust, in the sense that they can be applied to a much wider class of models that do not necessarily satisfy (1.1) but can be approximated by (1.1) in some sense. The goal of this paper is to formalize this notion by studying such a class and derive the rate of convergence of the predictor (1.5) under certain assumptions. Since all existing studies on functional linear prediction focus on model (1.1), this work provides a new perspective which has both theoretical and practical implications. A similar theory can also be developed for the penalty approach in (1.3), but a substantial reformulation is required and will be pursued elsewhere.

Linear prediction problems have traditionally had an important place in the theory of stochastic processes. The pioneering work of [18,12] on the prediction theory was developed in a time-series context, where they considered optimum predictors of future observations based on past data using the mean-squared prediction criterion. Later [14] introduced a unified theoretical framework for a very general class of linear prediction problems that went beyond time series. The prediction models that we consider in this paper are closely related to those considered by Parzen [14].

The remaining sections are organized as follows. The definition of a functional linear prediction model will be introduced in Section 2. Section 3 investigates the convergence rates of linear prediction under certain regularity conditions. In Section 4 we conclude the paper with a summary and some discussions. The proofs for the main results, Theorems 3.1 and 3.2, will be given in Section 5.

2. The functional linear prediction model

We continue to use the notation for functional data introduced in Section 1.

2.1. Linear prediction

Let $L^2(\Omega)$ be the Hilbert space containing all random variables on $(\Omega, \mathcal{B}, \mathbb{P})$ whose second moments are finite where $\langle U, V\rangle_{L^2(\Omega)} = \mathbb{E}(UV)$. The Hilbert space, L^2_X, spanned by X is defined as the subspace of $L^2(\Omega)$ that contains all finite linear combinations of the random variables $X(t), t \in \mathcal{T}$, and their limits in $L^2(\Omega)$. In other words, any random variable $\eta \in L^2_X$ is either of the form $\sum_{i=1}^m c_i X(t_i)$ for $m = 1, 2, \ldots, c_i \in \mathbb{R}$ and $t_i \in \mathcal{T}$, or satisfies $\inf \mathbb{E}[\eta - \sum_{i=1}^m c_i X(t_i)]^2 = 0$, where the infimum is taken over all m, c_i and t_i.

We first discuss a relationship between L^2_X and the principal components ϕ_j. Recall the Karhunen–Loève expansion (Theorem 1.4.1 of [1]):

$$\lim_{m\to\infty} \sup_{t\in\mathcal{T}} \mathbb{E}\left[X(t) - \sum_{j=1}^m \lambda_j^{1/2} U_j \phi_j(t)\right]^2 = 0, \tag{2.1}$$

where $U_j = \langle X, \phi_j\rangle/\lambda_j^{1/2}, j \geq 1$, the standardized scores of X, are uncorrelated random variables with mean zero and variance one.

Proposition 2.1. *The standardized scores U_j, $j \geq 1$, constitute an orthonormal basis of L^2_X.*

Proof. We first show that $U_j \in L^2_X$. Toward that end let $t_1, \ldots, t_n$ be equally-spaced points in $\mathcal{T}$. Now predict U_j by $\sum_{i=1}^n c_i X(t_i)$. It is straightforward to show

$$\begin{aligned} &\min_{c_1,\ldots,c_n} \mathbb{E}\left[U_j - \sum_{i=1}^n c_i X(t_i)\right]^2 \\ &= 1 - \lambda_j (\phi_j(t_1), \ldots, \phi_j(t_n))\mathbf{K}^-(\phi_j(t_1), \ldots, \phi_j(t_n))^T \end{aligned} \tag{2.2}$$

where $\mathbf{K}^-$ is the Moore–Penrose generalized inverse of $\mathbf{K} := \{K_X(t_i, t_j)\}$. The expression in (2.2) converges to 0 as $n \to \infty$ by the continuity of K_X and the fact that $\mathbf{K}$ is a discrete version of the covariance operator, $\mathfrak{K}_X$. Since L^2_X is closed, we conclude that $U_j \in L^2_X$. Thus, $\overline{\text{span}}\{U_j, j \geq 1\} \subset L^2_X$. It remains to show that each $X(t)$ is in $\overline{\text{span}}\{U_j, j \geq 1\}$, which follows immediately from (2.1).

A common formulation (cf. [15,14,16]) in classical linear prediction is to predict another random variable $Y \in L^2(\Omega)$, where $\mathbb{E}(Y) = 0$, optimally by a random variable, η, in L^2_X, using the mean squared error criterion. The case where $\mathbb{E}(Y) \neq 0$ can be similarly handled but the notation is more complicated. It follows that η is the projection of Y on L^2_X in $L^2(\Omega)$. Thus,

$$Y = \eta + \varepsilon, \tag{2.3}$$

where $\varepsilon := Y - \eta$ has mean zero and is uncorrelated with every random variable in L^2_X. However, the most interesting situation is when $\mathbb{E}(\varepsilon|X) = 0$, or, equivalently,

$$\mathbb{E}(Y|X) = \eta, \tag{2.4}$$

which says that the optimal mean squared predictor of Y coincides with the optimal mean squared linear predictor of Y. We refer to any model satisfying (2.3) and (2.4) as a *functional linear prediction model*, or a *linear prediction model* (LPM) for short. Below in Section 2.3 we will illustrate by examples that LPMs constitute a rich class of models.

Note that, for any LPM, we have

$$K_{YX}(t) = \mathbb{E}[YX(t)] = \mathbb{E}[\eta X(t)], \quad t \in \mathcal{T}, \tag{2.5}$$

and so, by (2.1),

$$K_{YX}(t) = \sum_{j=1}^{\infty} \lambda_j^{1/2} \mathbb{E}(\eta U_j)\phi_j(t), \quad t \in \mathcal{T}. \tag{2.6}$$

By Proposition 2.1, and (2.6),

$$\eta = \sum_{j=1}^{\infty} \mathbb{E}(\eta U_j)U_j = \sum_{j=1}^{\infty} \frac{\langle K_{YX}, \phi_j\rangle}{\lambda_j^{1/2}} U_j. \tag{2.7}$$

This result was first developed in [15] using a reproducing kernel Hilbert space formulation. Since the U_j's form an orthonormal basis in L^2_X and η is an element in L^2_X, by (2.7) we conclude that $\sum_{j=1}^{\infty} \langle K_{YX}, \phi_j\rangle^2/\lambda_j < \infty$. Alternatively, one can directly model η by

$$\eta = \sum_{j=1}^{\infty} f_j U_j \quad \text{where} \quad \sum_{j=1}^{\infty} f_j^2 < \infty, \tag{2.8}$$

in which case

$$K_{YX}(t) = \sum_{j=1}^{\infty} \lambda_j^{1/2} f_j \phi_j(t), \quad t \in \mathcal{T}.$$

2.2. *Linear regression*

As explained in Section 1, the special case of functional linear regression where η is a bounded linear functional, i.e. $\eta = \langle b, X\rangle$ for some $b \in L^2(\mathcal{T})$, has been considered extensively; see Section 1 for a list of references. For convenience call b the slope function of the linear regression model. To avoid non-identifiability let b be in the space spanned by the ϕ_j's. Note that we can then write $b = \sum_{j=1}^{\infty} \lambda_j^{-1/2} f_j \phi_j$, and so

$$\eta = \sum_{j=1}^{\infty} f_j U_j \quad \text{where} \quad \sum_{j=1}^{\infty} \frac{f_j^2}{\lambda_j} < \infty. \tag{2.9}$$

It follows from (2.8) and (2.9) that linear regression models form a sub-class of LPMs.

It might be somewhat misleading to say that the LPM is an extension of the linear regression model, since the LPM contains many prominent sub-models that are not linear regression; see

Section 2.3. However, the following perspective shows that linear regression models are of particular importance in the LPM. A natural measure of distance between two LPMs is the distance in L^2_X between the corresponding optimal linear predictors. Let $\eta = \sum_{j=1}^{\infty} f_j U_j$ be an arbitrary member of L^2_X and $\eta_n = \sum_{j=1}^{n} f_j U_j$; note that η_n is the optimal linear predictor of the regression model, for which the slope function is $b_n = \sum_{j=1}^{n} \lambda_j^{-1/2} f_j \phi_j$. Then

$$\|\eta - \eta_n\|^2_{L^2_X} = \sum_{j=n+1}^{\infty} f_j^2 \to 0 \quad \text{as } n \to \infty.$$

Thus, any LPM can be approximated arbitrarily closely by a linear regression model. However, the sequence of slope functions $\{b_n\}$ for the approximating linear regression models is not Cauchy in $L^2(\mathcal{T})$ unless we have a linear regression model to begin with.

From a modeling perspective, it is desirable to be able to measure the relative importance of the variables in X in predicting Y. In linear regression the slope function provides that information. For the general LPM, this is less straightforward. One potential approach is to consider predictors of the form $\int_{\mathcal{T}} X(t) d\mu(t)$ for some signed measure μ for which the stochastic integral is well-defined. We will not pursue that approach here.

2.3. Examples

We present a few examples of LPMs below.

Example 1. In the literature, it is frequently assumed that X, Y are jointly Gaussian, in the sense that all of the finite-dimensional joint distributions are multivariate normal. In that setting the assumptions of the LPM are readily satisfied. To see this, recall that $\mathbb{E}(Y|X) \in L^2_X$ and $X, Y, \mathbb{E}(Y|X)$ are jointly Gaussian; see Theorem 1.5.2 of [1]. Then, it is straightforward to conclude that $\eta = \mathbb{E}(Y|X)$ and $\varepsilon = Y - \mathbb{E}(Y|X)$ in (2.3), and ε and X are independent. □

Example 2. Consider the situation where ε depends on X. For instance, let $\tilde{\varepsilon}$ have mean zero and be independent of X, and $\varepsilon = \sigma(X)\tilde{\varepsilon}$ where $\sigma(X)$ is a random variable determined by X; then $\mathbb{E}(\varepsilon|X) = \sigma(X)\mathbb{E}(\tilde{\varepsilon}) = 0$. An obvious example of this scenario is the model $Y = \eta(1 + \tilde{\varepsilon})$.

In the examples below, we assume that $\mathbb{E}(\varepsilon|X) = 0$ and focus exclusively on the structure of η.

Example 3. Assume that X is an integrated Brownian motion on $[0, 1]$, defined as $X(t) = \int_t^1 B(s) ds$ where B is a standard Brownian motion. Let $\eta = \langle b, X^{(1)} \rangle$, where $b(t) = \sqrt{2}\cos(\pi t/2)$ and $X^{(1)}$ is the derivative of X. It is straightforward to verify that $\eta \in L^2_X$. Since $X^{(1)}$ is a standard Brownian motion, it has the Karhunen–Loève expansion $X^{(1)}(t) = \sum_{j=1}^{\infty} \nu_j^{1/2} V_j \psi_j(t)$, where $\nu_j = ((j - .5)\pi)^{-2}$, the V_j are distributed as independent normals with mean zero and variance one, and $\psi_j(t) = \sqrt{2}\sin((j - .5)\pi t)$. If a researcher knows the form of correct model $\eta = \langle b, X^{(1)} \rangle$ for some $b \in L^2(\mathcal{T})$ a priori, one option to identify η is to apply the prediction formula (2.7) by replacing X by $X^{(1)}$:

$$\eta = \sum_{j=1}^{\infty} \frac{\langle K_{YX^{(1)}}, \psi_j \rangle}{\nu_j^{1/2}} V_j.$$

The linear prediction modeling approach identifies η by focusing on X, which has the Karhunen–Loève expansion $X(t) = \sum_{j=1}^{\infty} \lambda_j^{1/2} U_j \phi_j(t)$, where $\lambda_j = \nu_j^2$, the U_j are distributed as independent normals with mean zero and variance one, and $\phi_j(t) = \sqrt{2}\cos((j-.5)\pi t)$. Note that $U_j = -\nu_j^{-1/2} V_j$. It is easily seen that

$$\eta = \langle b, X^{(1)} \rangle = \sum_{j=1}^{\infty} \left(2 \int_0^1 \sin((j-.5)\pi t)\cos(\pi t/2)dt \right) V_j = \sum_{j=1}^{\infty} f_j U_j,$$

where

$$\begin{aligned} f_j &= -2\nu_j^{1/2} \int_0^1 \sin((j-.5)\pi t)\cos(\pi t/2)dt \\ &= -\frac{1}{(j-.5)\pi^2}\{(1-(-1)^j)j^{-1} + (1-(-1)^{j-1})(j-1)^{-1}\}. \end{aligned}$$

The general prediction formula (2.7) holds as usual. However, η cannot be expressed as $\langle b, X\rangle$ for any function $b \in L^2[0,1]$ since $\sum_{j=1}^{\infty} f_j^2/\lambda_j = \infty$. Note that

$$\sum_{j=1}^{\infty} \frac{\langle K_{YX^{(1)}}, \psi_j \rangle}{\nu_j^{1/2}} V_j = \sum_{j=1}^{\infty} \frac{\langle K_{YX}, \phi_j \rangle}{\lambda_j^{1/2}} U_j,$$

and hence the two approaches potentially lead to similar prediction results in practice.

Example 4. Let

$$\eta = \sum_{i=1}^{k} c_i X(t_i) + \langle f, X \rangle,$$

where $c_i \neq 0$, $t_i \in [0,1]$ and $f \in L^2[0,1]$. The first component in η may be viewed as a "multivariate" component. Using the notation of Example 4, it follows that

$$\eta = \sum_{j=1}^{\infty} \lambda_j^{1/2} \left(\sum_{i=1}^{k} c_i \phi_j(t_i) + \langle f, \phi_j \rangle \right) U_j$$

and so this is a LPM. For many models, $\sum_{j=1}^{\infty}(\sum_{i=1}^{k} c_i\phi_j(t_i) + \langle f, \phi_j\rangle)^2 = \infty$ and, by (2.9), η cannot be expressed as $\langle b, X\rangle$ for any $b \in L^2[0,1]$. $\square$

We close this section with the following remarks.

1. LPMs are limiting models of linear regression models. For linear regression, one can estimate the model by estimating the slope function b. For a LPM that is not a linear regression model, estimating b is no longer meaningful.
2. The class of LPMs contains many sub-models for which specialized methodologies have been developed. Considering, for instance, the examples in Section 2.3, there is a rich literature for the inference of Gaussian processes, linear regression, linear differential regression and multivariate linear regression. The predictor (3.2) is not new, but the fact that the predictor can be applied for all of the LPMs in the functional-data context has not been recognized. In particular, the asymptotic theory of this predictor for the whole class of LPMs has not been investigated.

3. The mean squared prediction rate

Assume that (2.3)–(2.4) hold and an i.i.d. sample $X_i, Y_i, 1 \leq i \leq n$, is available. Our goal in this section is to consider the asymptotic behavior of a predictor $\widetilde{\eta}$ of η for a new, independent, observation X, as the sample size n tends to ∞. For simplicity, assume that the functions X_i's and X are fully observed. While this assumption is never met in practice, with sufficiently densely observed functional data one can pre-process the data by suitably fitting a curve to each partially observed curve. In those situations, taking the fitted curves as the original functional data does not materially alter the asymptotic theory of functional principal component analysis; see [2,11,13] for more details.

A natural measure of closeness between $\widetilde{\eta}$ and η is $\mathbb{E}[(\widetilde{\eta} - \eta)^2 | X_i, Y_i, 1 \leq i \leq n]$, where the (conditional) expectation is taken with respect to X. Without loss of generality, we can assume that $\widetilde{\eta} \in L_X^2$, in which case

$$\|\widetilde{\eta} - \eta\|^2_{L_X^2} = \mathbb{E}[(\widetilde{\eta} - \eta)^2 | X_i, Y_i, 1 \leq i \leq n]. \tag{3.1}$$

We will call this the mean squared prediction error. In the functional linear regression literature, the mean squared prediction error has been a common measure in the evaluation of procedures. See Cardot et al. [4] and Crambes et al. [6]. Cai and Hall [2] considered the "individual" squared prediction error, $\mathbb{E}[(\widetilde{\eta} - \eta)^2 | x]$, for a given new sample curve x, where the expectation is taken with respect to the data $X_i, Y_i, 1 \leq i \leq n$. Some papers, including [19,10], consider the estimation error of the regression slope function in terms of the L^2 distance. Recently, both Cardot and Johannes [5] and Yuan and Cai [20] have introduced more general distances that include prediction and estimation errors as special cases.

First, we investigate the prediction error of $\widehat{\eta}$ in (1.5) under certain regularity conditions. In view of (2.7), (2.8) and the assumption $\mathbb{E}(Y) = 0$, we rewrite the prediction formula (1.5) as

$$\widehat{\eta} =: \sum_{j=1}^{m} \widehat{\lambda}_j^{-1/2} \widehat{f}_j \langle X, \widehat{\phi}_j \rangle, \quad \text{where } \widehat{f}_j = \frac{\langle \widehat{K}_{YX}, \widehat{\phi}_j \rangle}{\widehat{\lambda}_j^{1/2}}. \tag{3.2}$$

A natural approach, which we will take, is to estimate K_{YX}, ϕ_j and λ_j by their sample versions. Let $\widehat{K}_{YX}(t) = n^{-1} \sum_{i=1}^{n} Y_i X_i(t)$ and $(\widehat{\mathfrak{K}}_X g)(t) = \int_{\mathcal{T}} \widehat{K}_X(s,t) g(s) ds$; accordingly, let $(\widehat{\lambda}_j, \widehat{\phi}_j)$ be the j-th eigenvalue/eigenfunction of $\widehat{\mathfrak{K}}_X$. Note that, since $\lambda_j \downarrow 0$, the quality of the estimators $\widehat{\lambda}_j$ and $\widehat{\phi}_j$ deteriorates as j increases. Thus, it is necessary to choose the cutoff point m in (3.2) sensibly. In practice, m can be picked, for example, by GCV.

Our results below are inspired by Cai and Hall [2] and Hall and Horowitz [10]. As such, the assumptions stated below bear considerable similarity to what was assumed in their works that address functional linear regression. In spite of technical similarity, our results are different from theirs in essence.

Recall that X is a second-order stochastic process with mean zero and has sample paths in $L^2(\mathcal{T})$ with probability one. Let f_j be defined by (2.8). The following conditions will be assumed.

(A1) $\mathbb{E}(\|X\|^4) < \infty$ and $\mathbb{E}(Y^4) < \infty$.
(A2) There exists some constant C such that $\mathbb{E}(U_j^4) \leq C$, i.e., $\mathbb{E}(\langle X, \phi_j \rangle^4) \leq C\lambda_j^2$, for all j.
(A3) There exist some constants $C > 0$, $\alpha > 1$ and $\kappa > 1/2$ such that

$$C^{-1} j^{-\alpha} \leq \lambda_j \leq C j^{-\alpha} \quad \text{and} \quad \lambda_j - \lambda_{j+1} \geq C^{-1} j^{-\alpha-1} \quad \text{for all } j, \tag{3.3}$$

and

$$|f_j| \le C j^{-\kappa} \quad \text{for all } j. \tag{3.4}$$

The specific values of C in (A2) and (A3) do not affect the rate computations. For convenience, below we will use C to denote a generic constant whose value may change from context to context.

Assumptions (A1) and (A2) are motivated by Gaussian processes. Indeed, if X is Gaussian then the U_j are independent standard normal random variables and (A2) immediately follows. The condition (3.3) models the rate at which the principal components contribute to the total variability in X, where $\alpha > 1$ is guaranteed by (A1). For the Brownian motion, $\alpha = 2$. Observe that $f_j = \operatorname{corr}(Y, U_j)\{\operatorname{var}(Y)\}^{1/2}$, and so (3.4) means that the correlation between U_j and Y decays at a polynomial rate of order κ. While the polynomial models for λ_j, f_j may not always hold in real-world problems, they are widely-accepted assumptions in theoretical developments.

The following result describes the rate of the mean squared prediction error of $\widehat{\eta}$ under (A1)–(A3), assuming the theoretically optimal choice of cutoff $m = m_n$ in (3.6).

Theorem 3.1. *Assume that* (A1)–(A3) *hold, and let*

$$\xi_o = 2\max(\kappa, \alpha + 1) \quad \textit{and} \quad \delta_o = \frac{2\kappa - 1}{\xi_o}. \tag{3.5}$$

Also, let

$$\begin{aligned} m_n &= [n^{1/\xi_o}], \quad \rho_n = n^{-\delta_o/2} \quad \textit{if } \kappa > \alpha + 1, \\ m_n &= [\epsilon_n n^{1/\xi_o}], \quad \rho_n = \epsilon_n^{-(2\kappa-1)/2} n^{-\delta_o/2} \quad \textit{if } \kappa \le \alpha + 1, \end{aligned} \tag{3.6}$$

for any $\epsilon_n \downarrow 0$ such that $\epsilon_n \succ n^{-1/\xi_o}$. Then $\|\widehat{\eta} - \eta\|_{L_X^2} = O_p(\rho_n)$.

Remarks. (i) Recall that $\kappa > 1/2$. In Theorem 3.1 if $\kappa > \alpha + 1$ then $\rho_n = n^{-\frac{2\kappa-1}{4\kappa}}$, and hence $n^{-1/2} \prec \rho_n \prec n^{-3/8}$; if $\kappa \le \alpha + 1$ then $\rho_n = \epsilon_n^{-\frac{2\kappa-1}{2}} n^{-\frac{2\kappa-1}{4(\alpha+1)}} \succ n^{-1/2}$, but a uniform upper bound for all permissible α and κ does not exist. Note also that ξ_o is increasing in α and κ, hence the cut-off point, m_n, is decreasing in α and κ. This is intuitively reasonable since increasing either α or κ leads to a more efficient representation of η in (2.8).

(ii) It is not straightforward to compare Theorem 3.1 with existing results in the literature, even when we restrict to the regression case. The closest ones are those in [2], which are also based on the analysis of principal components. However, they consider linear regression and their prediction error computation for a fixed new curve does not easily lead to a result that can be compared with ours, which describes the ensemble prediction error. Thus, Theorem 3.1 represents an original contribution.

Minimax bounds were obtained for the functional linear regression in various contexts; see [2,6,5,20]. It is natural to investigate if the mean squared prediction rate of $\widehat{\eta}$ in Theorem 3.1 is minimax in some sense. Let $\mathcal{B}_n$ be the class of measurable functions $\widetilde{\eta}$ of the data (X_i, Y_i), $1 \le i \le n$. For any $\kappa \in (1/2, \infty)$ and C in $(0, \infty)$, let $\mathcal{F}(\kappa, C)$ be the collection of distributions F of the LPMs that satisfy (3.4) and $\mathbb{E}(\varepsilon^2) < C$. Note that, without restrictions on the λ_j's, $\mathcal{F}(\kappa, C)$ is a larger class of distributions than those considered in Theorem 3.1. Below we develop a minimax lower bound for $\mathbb{E}_F \|\widetilde{\eta} - \eta\|_{L_X^2}^2$ for all $\widetilde{\eta} \in \mathcal{B}_n$ and all $F \in \mathcal{F}(\kappa, C)$, where the subscript F in $\mathbb{E}_F$ denotes that the expectation is computed with F as the true model.

Theorem 3.2.

$$\liminf_{n\to\infty} n^{(2\kappa-1)/(2\kappa)} \inf_{\widetilde{\eta}\in\mathcal{B}_n} \sup_{F\in\mathcal{F}(\kappa,C)} \mathbb{E}_F \|\widetilde{\eta}-\eta\|^2_{L^2_X} > 0. \tag{3.7}$$

Theorem 3.2 shows that the minimax lower bound for the class of models $\mathcal{F}(\kappa, C)$ is at least $n^{-(2\kappa-1)/(2\kappa)}$. By Remark (i) above, the rate is achieved by $\widehat{\eta}$ under (A1)–(A3) for the case $\kappa > \alpha + 1$. It is not clear if the lower bound can be achieved in general by $\widehat{\eta}$ or by any other predictor.

4. Conclusions and discussion

(i) We introduce the linear prediction model, LPM, in the context of functional predictor and scalar response. We show that the class of LPMs contains a large class of models that include all Gaussian processes and linear regression models.
(ii) We describe the linear prediction theory for the LPM and show that a unified prediction formula applies to all LPMs. We also prove a unified asymptotic theory of the optimal linear predictor. Our theory is modeled after a result in [2]. However, we employ a different discretionary measure and the asymptotic theory covers a considerably larger class of models than those in [2].
(iii) While the linear predictor that we consider is not new, this is the first time that the extent to which the predictor applies is made clear and a unified asymptotic theory is developed in the functional-data context.
(iv) Model-specific methodologies have been considered in the literature for various sub-models, such as Gaussian processes, linear regression and linear differential regression. If scientific knowledge or other information compels the use of such models in an investigation, then it always makes sense to apply the specialized methodology suitable for the analysis. As such, the linear prediction approach complements model-specific methodologies, as it is model-free, within the LPM class, and may provide a basis for comparison with results obtained by specialized methodologies.

5. Proofs

5.1. Proof of Theorem 3.1

It is important to mention at the outset that some of the technical tools presented here are similar to those in [2,10]. Indeed, some of the lemmas in the end may duplicate results in those papers in some ways. However, they are included for the sake of completeness and clarity. Also, in the proofs below C will denote a generic constant whose value may change from line to line.

Recall that T and $\widehat{T}$ are, respectively, the true and sample covariance operators of X; let $\Delta = \widehat{T} - T$. The sup norm of any operator V in $L^2(\mathcal{T})$ will be denoted by $\|V\|_\infty$. Also let $\Delta_{YX}(\cdot) = \widehat{K}_{YX}(\cdot) - K_{YX}(\cdot)$. For any real number r and positive integer m, define

$$t_r(m) = \begin{cases} 1 & \text{if } r < -1, \\ \log m & \text{if } r = -1, \\ m^{r+1} & \text{if } r > -1. \end{cases}$$

Define the event

$$\mathcal{F}_m = \mathcal{F}_m(n) = \{|\widehat{\lambda}_j - \lambda_j| \le (2C)^{-1} j^{-\alpha-1} \text{ for all } j = 1, \dots, m\}. \tag{5.1}$$

Our strategy is as follows. Observe that

$$\mathbb{P}(\|\widehat{\eta}-\eta\|_{L_X^2} > \nu\rho_n) \le \mathbb{P}(\mathcal{F}_m^c) + \mathbb{P}(\mathcal{F}_m, \|\widehat{\eta}-\eta\|_{L_X^2} > \nu\rho_n).$$

Since $\mathbb{P}(\mathcal{F}_m^c)$ does not depend on ν, by (i) of Lemma 5.1 below,

$$\limsup_{\nu\to\infty}\limsup_{n\to\infty}\mathbb{P}(\|\widehat{\eta}-\eta\|_{L_X^2} > \nu\rho_n) \le \limsup_{\nu\to\infty}\limsup_{n\to\infty}\mathbb{P}(\mathcal{F}_m, \|\widehat{\eta}-\eta\|_{L_X^2} > \nu\rho_n).$$

Thus, our goal below is to show that

$$\limsup_{\nu\to\infty}\limsup_{n\to\infty}\mathbb{P}(\mathcal{F}_m, \|\widehat{\eta}-\eta\|_{L_X^2} > \nu\rho_n) = 0. \tag{5.2}$$

By the triangle inequality,

$$\|\widehat{\eta}-\eta\|_{L_X^2} = \left\|\sum_{j=1}^{m}\widehat{\lambda}_j^{-1/2}\widehat{f}_j\langle X,\widehat{\phi}_j\rangle - \sum_{j=1}^{\infty}\lambda_j^{-1/2}f_j\langle X,\phi_j\rangle\right\|_{L_X^2} \le A_1 + A_2 + A_3, \tag{5.3}$$

where

$$A_1 = \Big\|\sum_{j=m+1}^{\infty}\lambda_j^{-1/2}f_j\langle X,\phi_j\rangle\Big\|_{L_X^2},$$

$$A_2 = \Big\|\sum_{j=1}^{m}(\widehat{\lambda}_j^{-1/2}\widehat{f}_j - \lambda_j^{-1/2}f_j)\langle X,\phi_j\rangle\Big\|_{L_X^2},$$

$$A_3 = \Big\|\sum_{j=1}^{m}\widehat{\lambda}_j^{-1/2}\widehat{f}_j\langle X,\widehat{\phi}_j - \phi_j\rangle\Big\|_{L_X^2}.$$

Recall that the randomness of $\widehat{f}_j$ does not play a role in the derivations of the expressions of A_2 and A_3.

By the orthogonality of the projections $\langle X,\phi_j\rangle$ in L_X^2 and the fact that $\|\langle X,\phi_j\rangle\|_{L_X^2}^2 = \mathbb{E}(\langle X,\phi_j\rangle^2) = \lambda_j$,

$$A_1^2 = \sum_{j=m+1}^{\infty} f_j^2, \tag{5.4}$$

$$A_2^2 = \sum_{j=1}^{m}\lambda_j\left(\widehat{\lambda}_j^{-1/2}\widehat{f}_j - \lambda_j^{-1/2}f_j\right)^2. \tag{5.5}$$

Moreover,

$$\begin{aligned} A_3^2 &= \sum_{j=1}^{m}\sum_{j'=1}^{m}\widehat{\lambda}_j^{-1/2}\widehat{\lambda}_{j'}^{-1/2}\widehat{f}_j\widehat{f}_{j'}\mathbb{E}[\langle X,\widehat{\phi}_j-\phi_j\rangle\langle X,\widehat{\phi}_{j'}-\phi_{j'}\rangle] \\ &= \sum_{j=1}^{m}\sum_{j'=1}^{m}\widehat{\lambda}_j^{-1/2}\widehat{\lambda}_{j'}^{-1/2}\widehat{f}_j\widehat{f}_{j'}\langle T(\widehat{\phi}_j-\phi_j),\widehat{\phi}_{j'}-\phi_{j'}\rangle. \end{aligned}$$

Thus,

$$
\begin{aligned}
A_3^2 &\le \|T\|_\infty \left(\sum_{j=1}^m |\widehat{\lambda}_j^{-1/2} \widehat{f}_j| \|\widehat{\phi}_j - \phi_j\| \right)^2 \\
&\le 2\|T\|_\infty \left\{ \left(\sum_{j=1}^m |\lambda_j^{-1/2} f_j| \|\widehat{\phi}_j - \phi_j\| \right)^2 \right. \\
&\quad \left. + \left(\sum_{j=1}^m |\widehat{\lambda}_j^{-1/2} \widehat{f}_j - \lambda_j^{-1/2} f_j| \|\widehat{\phi}_j - \phi_j\| \right)^2 \right\} \\
&\le 2m\|T\|_\infty \left\{ \sum_{j=1}^m \lambda_j^{-1} f_j^2 \|\widehat{\phi}_j - \phi_j\|^2 + \sum_{j=1}^m (\widehat{\lambda}_j^{-1/2} \widehat{f}_j - \lambda_j^{-1/2} f_j)^2 \|\widehat{\phi}_j - \phi_j\|^2 \right\} \qquad (5.6)
\end{aligned}
$$

by the Cauchy–Schwarz inequality. The proof of Theorem 3.1 is accomplished by considering the right-hand sides of (5.4)–(5.6).

By (A3) and (5.4),

$$
A_1^2 \le C \sum_{j=m+1}^\infty j^{-2\kappa} \le C m^{-2\kappa+1} = O(\rho_n^2). \qquad (5.7)
$$

To consider A_2, write

$$
\begin{aligned}
\widehat{\lambda}_j^{-1/2} \widehat{f}_j - \lambda_j^{-1/2} f_j &= \widehat{\lambda}_j^{-1} \langle \widehat{K}_{YX}, \widehat{\phi}_j \rangle - \lambda_j^{-1} \langle K_{YX}, \phi_j \rangle \\
&= \widehat{\lambda}_j^{-1} (\langle \Delta_{YX}, \phi_j \rangle + \langle K_{YX}, \widehat{\phi}_j - \phi_j \rangle + \langle \Delta_{YX}, \widehat{\phi}_j - \phi_j \rangle) \\
&\quad + (\widehat{\lambda}_j^{-1} - \lambda_j^{-1}) \langle K_{YX}, \phi_j \rangle.
\end{aligned}
$$

Thus, by (5.5),

$$
\begin{aligned}
A_2^2 &\le 4 \sum_{j=1}^m \lambda_j \widehat{\lambda}_j^{-2} \left(\langle \Delta_{YX}, \phi_j \rangle^2 + \langle K_{YX}, \widehat{\phi}_j - \phi_j \rangle^2 + \langle \Delta_{YX}, \widehat{\phi}_j - \phi_j \rangle^2 \right) \\
&\quad + 4 \sum_{j=1}^m \lambda_j (\widehat{\lambda}_j^{-1} - \lambda_j^{-1})^2 \langle K_{YX}, \phi_j \rangle^2. \qquad (5.8)
\end{aligned}
$$

Suppose that the event $\mathcal{F}_m$ holds. By (ii) of Lemma 5.1 and the definition of f_j,

$$
\sum_{j=1}^m \lambda_j (\widehat{\lambda}_j^{-1} - \lambda_j^{-1})^2 \langle K_{YX}, \phi_j \rangle^2 = \sum_{j=1}^m \widehat{\lambda}_j^{-2} (\widehat{\lambda}_j - \lambda_j)^2 f_j^2 \le 4 \sum_{j=1}^m \lambda_j^{-2} (\widehat{\lambda}_j - \lambda_j)^2 f_j^2.
$$

Since $|\widehat{\lambda}_j - \lambda_j| \le \|\Delta\|_\infty = O_p(n^{-1/2})$,

$$
\sum_{j=1}^m \lambda_j (\widehat{\lambda}_j^{-1} - \lambda_j^{-1})^2 \langle K_{YX}, \phi_j \rangle^2 \le 4\|\Delta\|_\infty^2 \sum_{j=1}^m \lambda_j^{-2} f_j^2 = O_p(n^{-1} t_{2\alpha-2\kappa}(m)). \qquad (5.9)
$$

Applying (ii) of Lemma 5.1 repeatedly, and Lemmas 5.2, 5.4 and 5.6,

$$\sum_{j=1}^{m} \lambda_j \widehat{\lambda}_j^{-2} \langle \Delta_{YX}, \phi_j \rangle^2 \le 4 \sum_{j=1}^{m} \lambda_j^{-1} \langle \Delta_{YX}, \phi_j \rangle^2 = O_p(n^{-1} m), \tag{5.10}$$

$$\begin{aligned} \sum_{j=1}^{m} \lambda_j \widehat{\lambda}_j^{-2} \langle \Delta_{YX}, \widehat{\phi}_j - \phi_j \rangle^2 &\le 4 \|\Delta_{YX}\|^2 \sum_{j=1}^{m} \lambda_j^{-1} \|\widehat{\phi}_j - \phi_j\|^2 \\ &= O_p(n^{-2}) \sum_{j=1}^{m} \lambda_j^{-1} j^2 = O_p(n^{-2} m^{\alpha+3}), \end{aligned} \tag{5.11}$$

$$\begin{aligned} \sum_{j=1}^{m} \lambda_j \widehat{\lambda}_j^{-2} \langle K_{YX}, \widehat{\phi}_j - \phi_j \rangle^2 &\le O_p(n^{-1}) \sum_{j=1}^{m} \left(j^{-2\kappa+2} \log^2 j + t_{-\kappa}^2(j) \right) \\ &\quad + O_p(n^{-2}) \sum_{j=1}^{m} \Big(j^{4\alpha-2\kappa+4} + j^{\alpha} t_{3\alpha-2\kappa}(j) \\ &\quad + j^{2\alpha-2\kappa+4} \log^2 j + j^{\alpha+2} t_{\alpha/2-\kappa}^2(j) \Big). \end{aligned} \tag{5.12}$$

By (5.8) and (5.9)–(5.12),

$$\begin{aligned} 1_{\mathcal{F}_m} A_2^2 &= O_p(n^{-1} t_{2\alpha-2\kappa}(m)) + O_p(n^{-1} m) + O_p(n^{-2} m^{\alpha+3}) \\ &\quad + O_p(n^{-1}) \sum_{j=1}^{m} \left(j^{-2\kappa+2} \log^2 j + t_{-\kappa}^2(j) \right) + O_p(n^{-2}) \\ &\quad \times \sum_{j=1}^{m} \left(j^{4\alpha-2\kappa+4} + j^{\alpha} t_{3\alpha-2\kappa}(j) + j^{2\alpha-2\kappa+4} \log^2 j + j^{\alpha+2} t_{\alpha/2-\kappa}^2(j) \right), \end{aligned} \tag{5.13}$$

where $1_{\mathcal{F}_m}$ is the indicator function of $\mathcal{F}_m$. We can now without loss of generality take $m = n^{1/\xi_o}$ (which is asymptotically bigger than $\epsilon_n n^{1/\xi_o}$) and will check that each term is $O_p(n^{-\delta_o})$, regardless of $\kappa >$ or $\le \alpha + 1$. This is a tedious but mathematically straightforward exercise. The hardest thing for us was to be able to see through the complicated expressions in (5.13) and recognize that this is the correct rate. To demonstrate, we consider three terms, the first, fourth, and fifth terms, which are representatives of all of the terms. The first term is of the order

$$n^{-1} t_{2\alpha-2\kappa}(m) \le \begin{cases} C n^{-1}, & 2\alpha - 2\kappa < -1 \\ C n^{-1} \log n, & 2\alpha - 2\kappa = -1 \\ C n^{\frac{2\alpha-2\kappa+1-\xi_o}{\xi_o}}, & 2\alpha - 2\kappa > -1. \end{cases}$$

The first two bounds are clearly $o(n^{\frac{1-\xi_o}{\xi_o}})$. So it suffices to check the third case, for which we note

$$\frac{2\alpha - 2\kappa + 1 - \xi_o}{\xi_o} \le \frac{2\alpha - 2\kappa + 1 - 2(\alpha+1)}{\xi_o} < \frac{-2\kappa + 1}{\xi_o},$$

since $\xi_o \ge 2(\alpha + 1)$. Thus, $n^{-1} t_{2\alpha-2\kappa}(m) \le C n^{-\delta_o}$. Now we take the fourth term on the right of (5.13). It follows that

$$n^{-1}\sum_{j=1}^{m} j^{-2\kappa+2}\log^2 j \le \begin{cases} Cn^{-1}, & -2\kappa+2<-1 \\ Cn^{-1}\log^3 n, & -2\kappa+2=-1 \\ C(\log^2 n)n^{\frac{-2\kappa+3-\xi_o}{\xi_o}}, & -2\kappa+2>-1. \end{cases}$$

Again only the third case requires attention:

$$\frac{-2\kappa+3-\xi_o}{\xi_o} \le \frac{-2\kappa+3-2(\alpha+1)}{\xi_o} < \frac{-2\kappa+1}{\xi_o}.$$

Hence, the bound $O_p(n^{-\delta_o})$ applies to the fourth term. Next we consider the fifth term on the right of (5.13), which has the order

$$n^{-1}\sum_{j=1}^{m} t_{-\kappa}^2(j) \le \begin{cases} Cn^{-1}m, & \kappa>1 \\ Cn^{-1}m\log^2 n, & \kappa=1 \\ Cn^{\frac{-2\kappa+3-\xi_o}{\xi_o}}, & \kappa<1. \end{cases}$$

Note that only the second case is new, and is bounded by $Cn^{\frac{1-\xi_o}{\xi_o}}\log^2 n$. While this is not bounded by $Cn^{\frac{1-\xi_o}{\xi_o}}$, it is bounded by $n^{\frac{-2\kappa+1}{\xi_o}}$ since, if $\kappa=1$,

$$\frac{1-\xi_o}{\xi_o} = -\frac{2\alpha+1}{\xi_o} < -\frac{1}{\xi_o} = \frac{-2\kappa+1}{\xi_o}.$$

The rest of the terms can be dealt with similarly. To summarize, we have

$$1_{\mathcal{F}_m}A_2^2 = O_p\big(n^{-\delta_o}\big). \tag{5.14}$$

Next we consider A_3^2. We again assume that $\mathcal{F}_m$ holds. By Lemma 5.4, the first term on the right of (5.6) is bounded by

$$m\sum_{j=1}^{m}\lambda_j^{-1}f_j^2\|\widehat{\phi}_j-\phi_j\|^2 = O_p(n^{-1}m)\sum_{j=1}^{m}\lambda_j^{-1}f_j^2 j^2 = O_p(n^{-1}mt_{\alpha-2\kappa+2}(m)). \tag{5.15}$$

The second term on the right can be dealt with in a similar manner to the treatment of A_2^2, replacing λ_j in the expression for A_2^2 in (5.5) by the bound $n^{-1}j^2$ for $\|\widehat{\phi}_j-\phi_j\|^2$ in (5.6). The resulting bound is

$$\begin{aligned} &O_p(n^{-2}mt_{3\alpha-2\kappa+2}(m)) + O_p(n^{-2}m^{\alpha+4}) + O_p(n^{-3}m^{2\alpha+6}) \\ &\quad + O_p(n^{-2}m)\sum_{j=1}^{m}\Big(j^{\alpha-2\kappa+4}\log^2 j + j^{\alpha+2}t_{-\kappa}^2(j)\Big) \\ &\quad + O_p(n^{-3}m)\sum_{j=1}^{m}\Big(j^{5\alpha-2\kappa+6} + j^{2\alpha+2}t_{3\alpha-2\kappa}(j) \\ &\quad + j^{3\alpha-2\kappa+6}\log^2 j + j^{2\alpha+4}t_{\alpha/2-\kappa}^2(j)\Big). \end{aligned} \tag{5.16}$$

As before, using the constraints on ξ_o and α, each term on the right of (5.15) and (5.16) can be shown to be bounded by $O_p(n^{-\delta_o})$ and so

$$1_{\mathcal{F}_m}A_3^2 = O_p\big(n^{-\delta_o}\big). \tag{5.17}$$

By (5.3), (5.7), (5.14) and (5.17), we have established (5.2). This concludes the proof of Theorem 3.1. $\square$

Lemma 5.1. *Assume that* (A1) *and* (A3) *hold. Then*

(i) $\limsup_{n\to\infty} \mathbb{P}(\mathcal{F}_m^c) = 0$.
(ii) *On* $\mathcal{F}_m$, $\widehat{\lambda}_j \geq \lambda_j/2$ *and* $|\widehat{\lambda}_j - \lambda_k| \geq |\lambda_j - \lambda_k|/2$.

Proof. Since $|\widehat{\lambda}_j - \lambda_j| \leq \|\Delta\|_\infty$,

$$\mathcal{F}_m^c \subset \{\|\Delta\|_\infty > (2C)^{-1} j^{-\alpha-1} \text{ for some } j = 1, \ldots, m\} \subset \{\|\Delta\|_\infty > (2C)^{-1} m^{-\alpha-1}\}.$$

Note that if $\kappa > \alpha + 1$ then $\xi_o = 2\kappa$ and $m^{-\alpha-1} = [n^{1/\xi_o}]^{-\alpha-1}$ is an order of magnitude bigger than $n^{-1/2}$. If $\kappa \leq \alpha + 1$ then $\xi_o = 2(\alpha + 1)$ and $m^{-\alpha-1} = [\epsilon_n n^{1/\{2(\alpha+1)\}}]^{-\alpha-1}$ is again an order of magnitude bigger than $n^{-1/2}$. Since $\|\Delta\|_\infty = O_p(n^{-1/2})$, (i) follows readily.

On $\mathcal{F}_m$, $|\widehat{\lambda}_j - \lambda_j| \leq \lambda_j/2$ by the assumption on λ_j, from which the first inequality of (ii) follows. The second inequality of (ii) follows from $|\lambda_j - \lambda_k| \leq |\widehat{\lambda}_j - \lambda_j| + |\widehat{\lambda}_j - \lambda_k|$ and

$$\begin{aligned} |\widehat{\lambda}_j - \lambda_j| &\leq (2C)^{-1} j^{-\alpha-1} \leq \frac{1}{2} \max(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}) \\ &\leq \frac{1}{2} |\lambda_j - \lambda_k|, \quad \text{for } j \neq k. \quad \square \end{aligned}$$

Lemma 5.2. *Assume that* (A1) *and* (A2) *hold. Then,* $\mathbb{E}(\langle \Delta_{YX}, \phi_j \rangle^2) \leq Cn^{-1}\lambda_j$.

Proof. Observe that

$$\begin{aligned} \mathbb{E}(\langle \Delta_{YX}, \phi_j \rangle^2) &= \operatorname{var}\left(n^{-1} \sum_{i=1}^n Y_i \langle X_i, \phi_j \rangle \right) = n^{-1} \operatorname{var}(\lambda_j^{1/2} Y U_j) = n^{-1} \lambda_j \mathbb{E}(Y^2 U_j^2) \\ &\leq n^{-1} \lambda_j \mathbb{E}^{1/2}(Y^4) \mathbb{E}^{1/2}(U_j^4) \leq Cn^{-1}\lambda_j. \quad \square \end{aligned}$$

Lemma 5.3. *Assume that* (A3) *holds. For any constants* $\gamma_1, \gamma_2, \gamma_3$ *satisfying* $\gamma_3 \geq 0$ *and* $\gamma_1\alpha + \gamma_2\kappa > 1$, *we have*

$$\begin{aligned} &\sum_{k \neq j} \lambda_k^{\gamma_1} |f_k|^{\gamma_2} |\lambda_j - \lambda_k|^{-\gamma_3} \\ &\quad \leq C \left(j^{(\gamma_3-\gamma_1)\alpha-\gamma_2\kappa+1} + j^{(\gamma_3-\gamma_1)\alpha-\gamma_2\kappa+\gamma_3} t_{-\gamma_3}(j) + t_{(\gamma_3-\gamma_1)\alpha-\gamma_2\kappa}(j) \right). \end{aligned} \tag{5.18}$$

Proof. For $k \geq j$,

$$\begin{aligned} \lambda_j - \lambda_k &= (\lambda_j - \lambda_{j+1}) + (\lambda_{j+1} - \lambda_{j+2}) + \cdots + (\lambda_{k-1} - \lambda_k) \\ &\geq C(k-j)(k-1)^{-\alpha-1}. \end{aligned} \tag{5.19}$$

Thus, for $k \geq 2j$,

$$\lambda_j - \lambda_k \geq \lambda_j - \lambda_{2j} \geq Cj(2j-1)^{-\alpha-1} \geq Cj^{-\alpha}; \tag{5.20}$$

for $k \leq j/2$ (i.e., $j \geq 2k$),

$$\lambda_k - \lambda_j \geq Ck^{-\alpha}; \tag{5.21}$$

for $j \le k \le 2j$,

$$\lambda_j - \lambda_k \ge C(k-j)(k-1)^{-\alpha-1} \ge C(k-j)j^{-\alpha-1}; \tag{5.22}$$

for $j/2 \le k \le j$,

$$\lambda_k - \lambda_j \ge C(j-k)(j-1)^{-\alpha-1} \ge C(j-k)j^{-\alpha-1}. \tag{5.23}$$

By (5.19)–(5.23),

$$\sum_{k\ge 2j} \lambda_k^{\gamma_1}|f_k|^{\gamma_2}|\lambda_j-\lambda_k|^{-\gamma_3} \le Cj^{\gamma_3\alpha}\sum_{k\ge 2j} k^{-\gamma_1\alpha-\gamma_2\kappa} \le Cj^{(\gamma_3-\gamma_1)\alpha-\gamma_2\kappa+1},$$

$$\begin{aligned}\sum_{[j/2]\le k\le 2j, k\ne j} \lambda_k^{\gamma_1}|f_k|^{\gamma_2}|\lambda_j-\lambda_k|^{-\gamma_3} &\le Cj^{\gamma_3(\alpha+1)}\sum_{[j/2]\le k\le 2j, k\ne j} k^{-\gamma_1\alpha-\gamma_2\kappa}|k-j|^{-\gamma_3},\\ &\le Cj^{\gamma_3(\alpha+1)-\gamma_1\alpha-\gamma_2\kappa}\sum_{[j/2]\le k\le 2j, k\ne j}|k-j|^{-\gamma_3}\\ &\le Cj^{(\gamma_3-\gamma_1)\alpha-\gamma_2\kappa+\gamma_3}t_{-\gamma_3}(j),\end{aligned}$$

$$\sum_{k\le [j/2]} \lambda_k^{\gamma_1}|f_k|^{\gamma_2}|\lambda_j-\lambda_k|^{-\gamma_3} \le C\sum_{k\le [j/2]} k^{(\gamma_3-\gamma_1)\alpha-\gamma_2\kappa} \le Ct_{(\gamma_3-\gamma_1)\alpha-\gamma_2\kappa}(j). \quad \square$$

Lemma 5.4. *Assume that* (A1)–(A3) *hold. Then,* $\|\widehat{\phi}_j-\phi_j\|^2 = O_p(n^{-1}j^2)$ *for each* j.

Proof. Observe that

$$\|\widehat{\phi}_j-\phi_j\|^2 = \sum_{k\ne j}\frac{1}{(\lambda_k-\lambda_j)^2}\langle\Delta\phi_j,\phi_k\rangle^2 + O(\|\Delta\|_\infty^3)$$

and

$$\mathbb{E}[\langle\Delta\phi_j,\phi_k\rangle^2] = n^{-1}\lambda_j\lambda_k\mathbb{E}(U_j^2U_k^2) \le n^{-1}\lambda_j\lambda_k\mathbb{E}^{1/2}(U_j^4)\mathbb{E}^{1/2}(U_k^4) \le Cn^{-1}\lambda_j\lambda_k.$$

Then, we have

$$\sum_{k\ne j}\frac{1}{(\lambda_k-\lambda_j)^2}\mathbb{E}(\langle\Delta\phi_j,\phi_k\rangle^2) \le Cn^{-1}\lambda_j\sum_{k\ne j}\lambda_k(\lambda_k-\lambda_j)^{-2}.$$

With $\gamma_1=1$, $\gamma_2=0$ and $\gamma_3=2$ in (5.18), it follows that

$$\sum_{k\ne j}\frac{1}{(\lambda_k-\lambda_j)^2}\langle\Delta\phi_j,\phi_k\rangle^2 = O_p(n^{-1}j^2). \quad \square$$

Lemma 5.5. *Under* (A1)–(A3)*, for any* $g\in L^2(\mathcal{T})$,

$$\begin{aligned}1_{\mathcal{F}_m}\langle g,\widehat{\phi}_j-\phi_j\rangle^2 \le O_p(1)\Bigg\{&n^{-1}\lambda_j\left(\sum_{k\ne j}|g_k||\lambda_j-\lambda_k|^{-1}\lambda_k^{1/2}\right)^2 + n^{-1}j^2g_j^2\\ &+n^{-2}\sum_{k\ne j}g_k^2(\lambda_j-\lambda_k)^{-4} + n^{-2}j^2\left(\sum_{k\ne j}|g_k||\lambda_j-\lambda_k|^{-1}\right)^2\Bigg\},\end{aligned}$$

where $g_j = \langle g,\phi_j\rangle$.

Proof. Observe that

$$\begin{aligned}\langle g, \widehat{\phi}_j - \phi_j\rangle &= \sum_{k\neq j} g_k\langle \phi_k, \widehat{\phi}_j - \phi_j\rangle + g_j\langle \phi_j, \widehat{\phi}_j - \phi_j\rangle \\ &= \sum_{k\neq j} g_k\langle \phi_k, \widehat{\phi}_j\rangle + g_j\langle \phi_j, \widehat{\phi}_j - \phi_j\rangle \\ &= \sum_{k\neq j} g_k(\widehat{\lambda}_j - \lambda_k)^{-1}\langle \Delta\widehat{\phi}_j, \phi_k\rangle + g_j\langle \phi_j, \widehat{\phi}_j - \phi_j\rangle\end{aligned}$$

since

$$\langle \Delta\widehat{\phi}_j, \phi_k\rangle = \langle \widehat{K}_X\widehat{\phi}_j, \phi_k\rangle - \langle \widehat{\phi}_j, K_X\phi_k\rangle = (\widehat{\lambda}_j - \lambda_k)\langle \widehat{\phi}_j, \phi_k\rangle.$$

Thus,

$$\langle g, \widehat{\phi}_j - \phi_j\rangle = T_{j1} + T_{j2} + T_{j3} + T_{j4},$$

where

$$\begin{aligned}T_{j1} &= \sum_{k\neq j} g_k(\lambda_j - \lambda_k)^{-1}\langle \Delta\phi_j, \phi_k\rangle \\ T_{j2} &= \sum_{k\neq j} g_k\{(\widehat{\lambda}_j - \lambda_k)^{-1} - (\lambda_j - \lambda_k)^{-1}\}\langle \Delta\phi_j, \phi_k\rangle \\ T_{j3} &= \sum_{k\neq j} g_k(\widehat{\lambda}_j - \lambda_k)^{-1}\langle \Delta(\widehat{\phi}_j - \phi_j), \phi_k\rangle \\ T_{j4} &= g_j\langle \phi_j, \widehat{\phi}_j - \phi_j\rangle.\end{aligned}$$

To handle T_{j1}, note that $\langle \Delta\phi_j, \phi_k\rangle = n^{-1}\sum_{i=1}^n U_{ij}U_{ik}$. Then

$$\begin{aligned}\mathbb{E}(T_{j1}^2) &\le n^{-1}\mathbb{E}\left\{\lambda_j^{1/2}U_j\sum_{k\neq j} g_k(\lambda_j - \lambda_k)^{-1}\lambda_k^{1/2}U_k\right\}^2 \\ &\le n^{-1}\lambda_j\mathbb{E}^{1/2}(U_j^4)\mathbb{E}^{1/2}\left\{\sum_{k\neq j} g_k(\lambda_j - \lambda_k)^{-1}\lambda_k^{1/2}U_k\right\}^4.\end{aligned}$$

Now

$$\begin{aligned}&\mathbb{E}\left\{\sum_{k\neq j} g_k(\lambda_j - \lambda_k)^{-1}U_k\right\}^4 \\ &\quad= \sum_{k_1\neq j}\sum_{k_2\neq j}\sum_{k_3\neq j}\sum_{k_4\neq j}\prod_{\ell=1}^4 g_{k_\ell}(\lambda_j - \lambda_{k_\ell})^{-1}\lambda_{k_\ell}^{1/2}\mathbb{E}(U_{k_1}U_{k_2}U_{k_3}U_{k_4}).\end{aligned}$$

Note that

$$|\mathbb{E}(U_{k_1}U_{k_2}U_{k_3}U_{k_4})| \le \prod_{\ell=1}^4 \mathbb{E}^{1/4}(U_{k_\ell}^4) \le C.$$

So,

$$\mathbb{E}\left\{\sum_{k\neq j} g_k(\lambda_j - \lambda_k)^{-1}U_k\right\}^4 \le C\left(\sum_{k\neq j}|g_k||\lambda_j - \lambda_k|^{-1}\lambda_k^{1/2}\right)^4$$

Thus,

$$\mathbb{E}(T_{j1}^2) \le Cn^{-1}\lambda_j \left(\sum_{k\neq j} |g_k||\lambda_j - \lambda_k|^{-1}\lambda_k^{1/2}\right)^2.$$

Assume that $\mathcal{F}_m$ holds in considering T_{j2} and T_{j3} below. By (ii) of Lemma 5.1,

$$|(\widehat{\lambda}_j - \lambda_k)^{-1} - (\lambda_j - \lambda_k)^{-1}| = \left|\frac{\lambda_j - \widehat{\lambda}_j}{(\widehat{\lambda}_j - \lambda_k)(\lambda_j - \lambda_k)}\right| \le 2\frac{|\lambda_j - \widehat{\lambda}_j|}{(\lambda_j - \lambda_k)^2}.$$

Hence, by the Cauchy–Schwarz inequality,

$$\begin{aligned}1_{\mathcal{F}_m} T_{j2}^2 &\le C(\lambda_j - \widehat{\lambda}_j)^2 \left\{\sum_{k\neq j} g_k(\lambda_j - \lambda_k)^{-2}\langle \Delta\phi_j, \phi_k\rangle\right\}^2 \\ &\le C(\lambda_j - \widehat{\lambda}_j)^2 \left\{\sum_{k\neq j} g_k^2(\lambda_j - \lambda_k)^{-4}\right\}\left\{\sum_{k\neq j}\langle \Delta\phi_j, \phi_k\rangle^2\right\}.\end{aligned}$$

Since $|\lambda_j - \widehat{\lambda}_j|^2 \le \|\Delta\|_\infty^2 = O_p(n^{-1})$, and

$$\sum_{k\neq j}\langle \Delta\phi_j, \phi_k\rangle^2 \le \sum_k \langle \Delta\phi_j, \phi_k\rangle^2 = \|\Delta\phi_j\|^2 \le \|\Delta\|_\infty^2 = O_p(n^{-1}),$$

we conclude

$$1_{\mathcal{F}_m} T_{j2}^2 \le O_p(n^{-2})\left\{\sum_{k\neq j} g_k^2(\lambda_j - \lambda_k)^{-4}\right\}.$$

By Lemma 5.4,

$$\begin{aligned}1_{\mathcal{F}_m}|T_{j3}| &\le C\|\Delta\|_\infty \|\widehat{\phi}_j - \phi_j\| \sum_{k\neq j}|g_k||\lambda_j - \lambda_k|^{-1} \\ &\le O_p(n^{-1}j)\sum_{k\neq j}|g_k||\lambda_j - \lambda_k|^{-1}.\end{aligned}$$

Similarly,

$$|T_{j4}|^2 \le O_p(n^{-1}j^2)g_j^2. \quad \square$$

Lemma 5.6. *Under* (A1)–(A3),

$$\begin{aligned}&1_{\mathcal{F}_m}\langle K_{YX}, \widehat{\phi}_j - \phi_j\rangle^2 \\ &\quad \le O_p(1)\Big\{n^{-1}\big(j^{-\alpha-2\kappa+2}\log^2 j + j^{-\alpha}t_{-\kappa}^2(j)\big) \\ &\qquad + n^{-2}\big(j^{3\alpha-2\kappa+4} + t_{3\alpha-2\kappa}(j) + j^{\alpha-2\kappa+4}\log^2 j + j^2 t_{\alpha/2-\kappa}^2(j)\big)\Big\}.\end{aligned}$$

Proof. In Lemma 5.5, take $g = K_{YX}$. Hence, $g_j = \lambda_j^{1/2} f_j$, and it follows that

$$\begin{aligned}&1_{\mathcal{F}_m}\langle K_{YX}, \widehat{\phi}_j - \phi_j\rangle^2 \\ &\quad \le O_p(1)\Big\{n^{-1}\lambda_j\Big(\sum_{k\neq j}\lambda_k^{1/2}|f_k||\lambda_j - \lambda_k|^{-1}\lambda_k^{1/2}\Big)^2 + n^{-1}j^{-(\alpha+2\kappa-2)} \\ &\qquad + n^{-2}\sum_{k\neq j}\lambda_k f_k^2(\lambda_j - \lambda_k)^{-4} + n^{-2}j^2\Big(\sum_{k\neq j}\lambda_k^{1/2}|f_k||\lambda_j - \lambda_k|^{-1}\Big)^2\Big\}.\end{aligned}$$

With $\gamma_1 = 1, \gamma_2 = 1, \gamma_3 = 1$ in (5.18), we obtain

$$\sum_{k\neq j} \lambda_k^{1/2} |f_k| |\lambda_j - \lambda_k|^{-1} \lambda_k^{1/2} \le C \left(j^{-\kappa+1} \log j + t_{-\kappa}(j) \right);$$

with $\gamma_1 = 1, \gamma_2 = 2, \gamma_3 = 4$,

$$\sum_{k\neq j} \lambda_k f_k^2 (\lambda_j - \lambda_k)^{-4} \le C \left(j^{3\alpha-2\kappa+4} + t_{3\alpha-2\kappa}(j) \right);$$

with $\gamma_1 = 1/2, \gamma_2 = 1, \gamma_3 = 1$,

$$\sum_{k\neq j} \lambda_k^{1/2} |f_k| |\lambda_j - \lambda_k|^{-1} \le C \left(j^{\alpha/2-\kappa+1} \log j + t_{\alpha/2-\kappa}(j) \right).$$

The result follows upon collecting terms. □

5.2. *Proof of Theorem 3.2*

Consider a special class of $F \in \mathcal{F}(\kappa, C)$ given as follows. Note that the value of C does not affect the rate in this proof and we can, without loss of generality, choose a large enough C so that $\mathcal{F}(\kappa, C)$ contains the following class. Let $\{q_n\}$ be a sequence of positive integers whose values will be specified later. Let $f_j = \theta_j j^{-\kappa}$ for $1 \le j \le q_n$ and $f_j = 0$ otherwise, where θ_j is either 0 or 1. Suppose that $U_j \sim$ i.i.d. uniform$[-\sqrt{3}, \sqrt{3}]$ and $\varepsilon \sim N(0, 1)$. We have

$$Y = \eta + \varepsilon = \sum_{j=1}^{q_n} \theta_j j^{-\kappa} U_j + \varepsilon.$$

Denote by $\mathcal{F}_n^*$ the collection of all 2^{q_n} different configurations of such F. The strategy is to obtain a lower bound for $\sup_{F\in\mathcal{F}_n^*} \mathbb{E}_F \|\widetilde{\eta} - \eta\|^2_{L^2_X}$, which will then necessarily be a lower bound for $\sup_{F\in\mathcal{F}(\kappa,C)} \mathbb{E}_F \|\widetilde{\eta} - \eta\|^2_{L^2_X}$. By Proposition 2.1, we can write $\widetilde{\eta} = \sum_{j=1}^{\infty} \widetilde{\theta}_j j^{-\kappa} U_j$, where $\widetilde{\theta}_j = j^{\kappa} \langle \widetilde{\eta}, U_j \rangle_{L^2_X}$. Thus, for any $F \in \mathcal{F}_n^*$, $\mathbb{E}_F \|\widetilde{\eta} - \eta\|^2_{L^2_X} = \sum_{j=1}^{\infty} j^{-2\kappa} \mathbb{E}_F (\widetilde{\theta}_j - \theta_j)^2$. This shows that we can focus on those $\widetilde{\eta}$ for which $\widetilde{\theta}_j = 0$ for $j > q_n$, which we will do. It follows that

$$\begin{aligned}
\sup_{F\in\mathcal{F}(\kappa,C)} \mathbb{E}_F \|\widetilde{\eta} - \eta\|^2_{L^2_X} &\ge \sup_{F\in\mathcal{F}_n^*} \sum_{j=1}^{q_n} j^{-2\kappa} \mathbb{E}_F (\widetilde{\theta}_j - \theta_j)^2 \\
&\ge \frac{1}{2^{q_n}} \sum_{F\in\mathcal{F}_n^*} \sum_{j=1}^{q_n} j^{-2\kappa} \mathbb{E}_F (\widetilde{\theta}_j - \theta_j)^2 \\
&= \sum_{j=1}^{q_n} j^{-2\kappa} \frac{1}{2^{q_n}} \sum_{F\in\mathcal{F}_n^*} \mathbb{E}_F (\widetilde{\theta}_j - \theta_j)^2.
\end{aligned}$$

For any $F \in \mathcal{F}_n^*$ and $j = 1, \ldots, q_n$, let $F_{j0} = F$ and F_{j1} be the same as F except with the value of θ_j changed from 0 to 1 or from 1 to 0. By symmetry, we have

$$\begin{aligned}
&\sum_{j=1}^{q_n} j^{-2\kappa} \frac{1}{2^{q_n}} \sum_{F\in\mathcal{F}_n^*} \mathbb{E}_F (\widetilde{\theta}_j - \theta_j)^2 \\
&\quad = \sum_{j=1}^{q_n} j^{-2\kappa} \frac{1}{2^{q_n}} \sum_{F\in\mathcal{F}_n^*} \frac{1}{2} \left\{ \mathbb{E}_{F_{j0}} (\widetilde{\theta}_j - \theta_j)^2 + \mathbb{E}_{F_{j1}} (\widetilde{\theta}_j - \theta_j)^2 \right\}.
\end{aligned}$$

Combining the previous two derivations,

$$\begin{aligned}\sup_{F\in\mathcal{F}(\kappa,C)} \mathbb{E}_F\|\widetilde{\eta}-\eta\|^2_{L^2_X} &\geq \sum_{j=1}^{q_n} j^{-2\kappa}\frac{1}{2^{q_n}}\sum_{F\in\mathcal{F}_n^*}\frac{1}{2}\max_{F\in\{F_{j0},F_{j1}\}}\mathbb{E}_F(\widetilde{\theta}_j-\theta_j)^2\\ &\geq \frac{1}{8}\sum_{j=1}^{q_n} j^{-2\kappa}\frac{1}{2^{q_n}}\sum_{F\in\mathcal{F}_n^*}\max_{F\in\{F_{j0},F_{j1}\}}\mathbb{P}_F(|\widetilde{\theta}_j-\theta_j|\geq 1/2).\end{aligned}$$

Now we evaluate $\max_{F\in\{F_{j0},F_{j1}\}}\mathbb{P}_F(|\widetilde{\theta}_j-\theta_j|\geq 1/2)$ using the approach of Theorem 3.1 of Hall (1989). Denote θ_{j0} and θ_{j1} as the values of θ_j in F_{j0} and F_{j1}, respectively. Without loss of generality, take $\theta_{j0}=0$ and $\theta_{j1}=1$. Define $\overline{\theta}_j=\begin{cases}0 & \text{if } |\widetilde{\theta}_j-\theta_{j0}|\leq|\widetilde{\theta}_j-\theta_{j1}|,\\ 1 & \text{otherwise.}\end{cases}$ $\overline{\theta}_j=1(|\widetilde{\theta}_j-\theta_{j0}|>|\widetilde{\theta}_j-\theta_{j1}|)$.Clearly, $|\widetilde{\theta}_j-\theta_{j0}|\geq 1/2$ if $\overline{\theta}_j=1$ and $|\widetilde{\theta}_j-\theta_{j1}|\geq 1/2$ if $\overline{\theta}_j=0$ and so

$$\begin{aligned}\max_{F\in\{F_{j0},F_{j1}\}}\mathbb{P}_F(|\widetilde{\theta}_j-\theta_j|\geq 1/2) &\geq \max\{\mathbb{P}_{F_{k0}}(\overline{\theta}_j=1),\mathbb{P}_{F_{j1}}(\overline{\theta}_j=0)\}\\ &\geq \frac{1}{2}\{\mathbb{P}_{F_{j0}}(\overline{\theta}_j=1)+\mathbb{P}_{F_{j1}}(\overline{\theta}_j=0)\}.\end{aligned}$$

Denote by L_{j0}/L_{j1} the likelihood ratio of F_{j0} versus F_{j1}. It follows from the Neyman–Pearson Lemma that

$$\begin{aligned}\frac{1}{2}\{\mathbb{P}_{F_{j0}}(\overline{\theta}_j=1)+\mathbb{P}_{F_{j1}}(\overline{\theta}_j=0)\} &\geq \frac{1}{2}\{\mathbb{P}_{F_{j0}}(L_{j0}/L_{j1}\leq 1)+\mathbb{P}_{F_{j1}}(L_{j0}/L_{j1}\geq 1)\}\\ &\geq \left\{4\mathbb{E}_{F_{j0}}(L^2_{j1}/L^2_{j0})\right\}^{-1},\end{aligned}$$

where the last inequality can be found from the derivations following (3.4) of [9]. Using the assumption that ε is $N(0,1)$ and $U_j\sim$ uniform$[-\sqrt{3},\sqrt{3}]$, it can be derived that

$$\mathbb{E}_{F_{j0}}(L^2_{j1}/L^2_{j0})=\left(\mathbb{E}_{F_{j0}}[\exp\{(j^{-\kappa}U_j)^2\}]\right)^n=(1-c_jj^{-2\kappa})^{-n/2},$$

where $c_j\to 1$ as $j\to\infty$. Thus,

$$\sup_{F\in\mathcal{F}(\kappa,C)}\mathbb{E}_F\|\widetilde{\eta}-\eta\|^2_{L^2_X}\geq\frac{1}{32}\sum_{j=1}^{q_n}j^{-2\kappa}(1-c_jj^{-2\kappa})^{n/2}. \tag{5.24}$$

Observe that for $\inf_{j\geq n^{1/(2\kappa)}}(1-c_jj^{-2\kappa})^{n/2}$ is bounded below by a positive constant for all n. Thus, taking $q_n>2n^{1/(2\kappa)}$, for example, we conclude that the lower bound in (5.24) has the rate $n^{-(2\kappa-1)/(2\kappa)}$. □

Acknowledgments

We are very grateful to the referees and an Associate Editor for their helpful comments. Shin's research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-003753). Hsing's research was supported by NSF Grant DMS-0806098.

References

[1] R.B. Ash, M.F Gardner, Topics in Stochastic Processes, Academic Press, 1975.
[2] T.T. Cai, P. Hall, Prediction in functional linear regression, *Annals of Statistics* 34 (2006) 2159–2179.

[3] H. Cardot, F. Ferraty, P. Sarda, Functional linear model, Statistics & Probability Letters 45 (1999) 11–22.

[4] H. Cardot, F. Ferraty, P. Sarda, Spline estimators for the functional linear model, Statistica Sinica 13 (2003) 571–591.

[5] H. Cardot, J. Johannes, Thresholding projection estimators in functional linear models, Journal of Multivariate Analysis 101 (2010) 395–408.

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

[7] A. Cuevas, M. Febrero, R. Fraiman, Linear functional regression: the case of fixed design and functional response, Canadian Journal of Statistics 30 (2002) 285–300.

[8] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer, 2006.

[9] P. Hall, On convergence rates in nonparametric problems, International Statistical Review 57 (1989) 45–58.

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

[11] P. Hall, H.-G. Müller, J.-L Wang, Properties of functional principal component analysis for sparse data, *Annals of Statistics* 34 (2006) 1493–1517.

[12] A. Kolmogorov, Interpolation and extrapolation, Bulletin de l'Academie Science URSS Series Mathematics 5 (1941) 3–14.

[13] Y. Li, T. Hsing, Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data, Annals of Statistics 38 (2010) 3321–3351.

[14] E. Parzen, A new approach to the synthesis of optimal smoothing and prediction systems, in: Mathematical Optimization Techniques, University of California Press, 1963, pp. 75–108.

[15] E. Parzen, An approach to time series analysis, Annals of Mathematical Statistics 32 (1961) 951–989.

[16] E. Parzen, Time Series Analysis Papers, Holden-Day, 1967.

[17] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, 2005.

[18] N. Wiener, The Extrapolation, Interpolation, and Smoothing of Stationary Time Series, Wiley, 1949.

[19] F. Yao, H.-G. Müller, J.-L. Wang, Functional linear regression analysis for longitudinal data, *Annals of Statistics* 33 (2005) 2873–2903.

[20] M. Yuan, T.T. Cai, A reproducing kernel Hilbert space approach to functional linear regression, Annals of Statistics 38 (2010) 3412–3444.