Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 98 (2007) 1782–1804
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

On rates of convergence in functional linear regression

Yehua Li[a], Tailen Hsing[b,*,1]

[a] *Department of Statistics, University of Georgia, Athens, GA 30602, USA*
[b] *Department of Statistics, The Ohio State University, Columbus, OH 43210, USA*

Received 1 January 2006
Available online 15 December 2006

Abstract

This paper investigates the rate of convergence of estimating the regression weight function in a functional linear regression model. It is assumed that the predictor as well as the weight function are smooth and periodic in the sense that the derivatives are equal at the boundary points. Assuming that the functional data are observed at discrete points with measurement error, the complex Fourier basis is adopted in estimating the true data and the regression weight function based on the penalized least-squares criterion. The rate of convergence is then derived for both estimators. A simulation study is also provided to illustrate the numerical performance of our approach, and to make a comparison with the principal component regression approach.

AMS 2000 subject classification: 62G08; 62G20; 62M10

Keywords: Functional data analysis; Penalized least squares; Periodic spline

1. Introduction

As a byproduct of modern science and technology, a lot of the data that are observed or collected in many fields nowadays are exceptionally high-dimensional. One such example is functional data, where each observation in a sample can be viewed as a function, as opposed to a scalar or a vector. In reality, for one reason or another if not simply human limitation, instead of observing such functions in their entirety, one only observes the values of the functions at a finite set of points. Nevertheless, the number of values observed per function may be quite large, sometimes much larger than the total number of functions, so that traditional multivariate analytical theory and methodology are not directly applicable. Indeed, the analysis of functional data has been steadily

* Corresponding author.
E-mail addresses: yehuali@uga.edu (Y. Li), hsing@stat.ohio-state.edu (T. Hsing).
[1] Research supported by NSF Grant DMS-0624239.

0047-259X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2006.10.004

gaining attention among statisticians and practitioners, and there has been much progress on the methodology front in trying to understand how to deal with such data. The books Ramsay and Silverman [15,16,20] and their website "http://ego.psych.mcgill.ca/misc/fda" contain a substantial amount of information in that regard. However, there has been much less progress on the theory front. This is not a surprise in view of the nature of the difficulties, as theoretical results in this regard invariably involve the theory of functional analysis, multivariate analysis, optimization, and nonparametric function estimation.

One relatively simple problem, the linear regression, did receive a considerable attention theory-wise. Consider the model

$$Y_i = \mu + \int_a^b X_i(t) f(t)\, dt + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where the response Y_i, the fixed intercept μ and error ε_i are scalar, and the predictor X_i and regression weight function f are functions on $[a, b]$. For now, assume that we observe the X_i, Y_i and we are interested in the inference of μ, f and the variance of ε. In particular, we are interested in the rates of convergence in estimating f. In this regard, we mention the papers Cai and Hall [3], Cardot et al. [4], Cardot and Sarda [5], Hall and Horowitz [11] and Müller and Stadtmüller [14]. Each of these papers contributes ideas to the development a satisfying asymptotic theory for functional data analysis. However, all of the papers assumed that the functional predictors X_i are completely observed. This assumption is crucial for their results, but is seldom met in practice. To make matters worse, in reality there may be measurement error in observing X_i. The goal of the present paper is to address the linear regression problem under these practical situations.

This problem is ill-posed in the sense that a minute change in the data may lead to a huge changes in the resulting estimates [19]. One of the greatest challenges here (and elsewhere in functional data analysis) is to consider how to interface the finite-dimensional space in which the data are collected and the infinite-dimensional space where the truth resides. Ramsay and Silverman [20] proposed the following practical solution. First, represent both the X_i and g, any candidate estimate of f, in terms of a set of pre-selected basis functions $\phi_1, \ldots, \phi_K$, say, so that

$$\tilde{X}_i = \sum_{k=1}^K b_{i,k}\phi_k \quad \text{and} \quad g = \sum_{k=1}^K c_k \phi_k,$$

where the "˜" in $\tilde{X}_i$ signifies the fact that this is a function that approximates the true function X_i based on the finitely observed values of X_i. Then estimate μ and f by the minimizer of the following penalized least-squares criterion function

$$\begin{aligned} & n^{-1}\sum_{i=1}^n \left[Y_i - v - \int_0^1 \tilde{X}_i g\right]^2 + \lambda \int_0^1 [g^{(m)}]^2 \\ & \quad = n^{-1}\sum_{i=1}^n \left[Y_i - v - \sum_{k=1}^K\sum_{\ell=1}^K b_{i,k}c_\ell \int_0^1 \phi_k\phi_\ell\right]^2 + \lambda \sum_{k=1}^K\sum_{\ell=1}^K c_k c_\ell \int_0^1 \phi_k^{(m)}\phi_\ell^{(m)}, \end{aligned}$$

where λ is a smoothing parameter. They noted that the choice of the basis functions depends on the nature of the problem and the data. This is appealing since now we have a finite-dimensional optimization problem to cope with. Furthermore, if the basis functions are such that the matrix $\{\int_0^1 \phi_k^{(m)}\phi_\ell^{(m)}\}_{k,\ell=1}^K$ has a band structure, the computations will be even more straightforward; examples of such basis functions include Fourier, B-splines, and natural splines. It is also worth mentioning that variations of the basis-function approaches are adopted by other authors in studying the linear regression model; they include Cardot et al. [4] who studied penalized B-splines, and James et al. [12] who considered a parametric approach.

The splines are general and flexible approximating functions which have a lot of desirable properties for this problem. However, theoretically they are more difficult to deal with, and we will address that problem in a forthcoming paper. In the present paper, we will focus on Fourier basis. The Fourier functions are a ideal basis if the data are smooth and exhibit periodicity; an example of that is the Canadian weather data in Ramsay and Silverman [20]. They are certainly the most convenient basis functions to work with in terms of developing theory, since they are orthogonal, their mth derivatives are orthogonal, and the orthogonality even carries over to the discretized basis vectors when the set of discrete points are equally spaced. Our goal of this paper is to study the rate of convergence of the penalized least-squares estimation using the Fourier basis. We will show that the rate of convergence is similar to that obtained in nonparametric regression function estimation.

The nature of the topic makes it necessary to employ some standard functional analysis terminology and results. They are basic and do not go beyond the first course in functional analysis. The reader is referred to Conway [6] for details.

This paper is structured as follows. Section 2 describes the model assumptions, the two-stage penalized least-squares approach for estimation of the weight function, and our main theoretical results for the estimator. In Section 3, we discuss some computational issues, such as selection of the smoothing parameters, and conduct a simulation study to illustrate the performance of the penalized least-squares approach. Another popular approach in functional linear model is the principal component regression (PCR) approach, which was considered by Hall and Horowitz [11], Cai and Hall [3], Müller and Stadtmüller [14], among others. A comparison of the penalized least-squares approach with the PCR approach is also provided via the simulation study in Section 3. All proofs and lemmas are collected in Section 4.

2. Assumptions and main results

2.1. Model assumptions and the two-stage penalized least-square estimator

Assume that the functional predictor X is a real-valued, zero-mean, second-order stochastic process on [0, 1] (cf. [1]). Further, for some positive integer m, assume that with probability one X belongs to the periodic Soblev space

$$W_{2,\mathrm{per}}^m = \{g \in L^2[0,1]\text{: } g \text{ is } m\text{-times differentiable where } g^{(m)} \in L^2[0,1] \text{ and } g^{(v)} \text{ is absolutely continuous with } g^{(v)}(0) = g^{(v)}(1), 0 \leqslant v \leqslant m-1\}.$$

It is well known that $W_{2,\mathrm{per}}^m$ is dense in $L^2[0,1]$, therefore our methodology below based on this assumption applies to even situations for which this assumption is not met. However, relaxing the smoothness and boundary conditions does affect the convergence rate of our estimator. Denote by $R(s,t)$ the covariance function

$$R(s,t) = E[X(s)X(t)], \quad s,t \in [0,1],$$

and T the corresponding covariance operator

$$T\colon g \to \int_{s=0}^{1} R(s,\cdot)g(s)\,ds, \quad g \in L^2[0,1].$$

For convenience, we will assume throughout without further mention that $E\|X\|_{L^2}^4 < \infty$, where $\|\cdot\|_{L^2}$ denotes the usual norm in $L^2[0,1]$. This implies, among other things, that

$$\begin{aligned}\int_0^1\int_0^1 R^2(s,t)\,ds\,dt &= \int_0^1\int_0^1 E^2[X(s)X(t)]\,ds\,dt \\ &\leqslant \int_0^1\int_0^1 E[X^2(s)X^2(t)]\,ds\,dt = E(\|X\|_{L^2}^4) < \infty,\end{aligned}$$

which shows that T is a Hilbert–Schmidt operator. Also, for convenience, we will assume throughout that $\mathrm{Var}(\int_0^1 X(t)\,dt) > 0$, which will guarantee that our penalized least-squares estimator is well defined when the sample size is large.

Let X_i, $1 \leqslant i \leqslant n$, be n independent realizations of X. Below we consider the linear regression model (1) with $\mu = 0$. This simplification is minor for our results, but entails a considerable saving in term of notation. Let $t_j = (2j-1)/(2J)$, $1 \leqslant j \leqslant J$, be the locations where we observe the X_i; assume that the data that are observed are Y_i, $1 \leqslant i \leqslant n$, and

$$\mathbf{Z}_i = (Z_{i,1},\ldots,Z_{i,J})^T = (X_i(t_1)+\varsigma_{i,1},\ldots,X_i(t_J)+\varsigma_{i,J})^T, \quad 1 \leqslant i \leqslant n,$$

where $\varsigma_{i,j}$ is measurement error for $X_i(t_j)$. The ε_i, $\varsigma_{i,j}$ are assumed to be mutually uncorrelated, and independent of the $X_i(t_j)$, with mean zero and

$$\mathrm{Var}(\varepsilon_i) = \sigma_\varepsilon^2 \quad \text{and} \quad \mathrm{Var}(\varsigma_{i,j}) = \sigma_\varsigma^2.$$

The most convenient Fourier basis functions for our problem are the complex Fourier functions $\phi_k(t) = e^{2\pi i k t}$, $k = 0, \pm 1, \pm 2, \ldots$. We could equivalently use the pairs of sine and cosine functions, but that will be much more cumbersome notationally. Thus, we need to extend our L^2-space inner product and norm to cover complex-valued functions; specifically, for complex-valued measurable functions g and h satisfying $\int_0^1 g\bar{g}$ and $\int_0^1 h\bar{h} < \infty$, define $\langle g,h\rangle_{L^2} := \int_0^1 g\bar{h}$ and $\|g\|_{L^2}^2 = \langle g,g\rangle_{L^2}$.

It is well known that each $g \in W_{2,\mathrm{per}}^m$ can be uniquely represented as $g = \sum_{k=-\infty}^{\infty} c_k\phi_k$ in the L^2 sense with $\bar{c}_k = c_{-k}$. Let

$$K = [(J-1)/2] = \max\{k \leqslant (J-1)/2;\ k \text{ is an integer}\},$$

and put

$$\phi(t) = (\phi_{-K}(t),\ldots,\phi_0(t),\ldots,\phi_K(t))^T,$$

$$\Phi = \{\phi_k(t_j)\}_{j=1,\ldots,J;k=-K,\ldots,K},$$

and

$$W = \{\langle \phi_i^{(m)}, \phi_j^{(m)}\rangle_{L^2}\}_{i,j=-K}^{K}.$$

Clearly,

$$W = \mathrm{diag}\,\{(2\pi K)^{2m},\ldots,(2\pi)^{2m},0,(2\pi)^{2m},\ldots,(2\pi K)^{2m}\}, \tag{2}$$

and, since the t_j are equally spaced,

$$J^{-1}\bar{\Phi}^T\Phi = I.$$

First, for some smoothing parameter ρ, approximate each $\mathbf{Z}_i$ by $\tilde{X}_{i,\rho}$, the minimizer $u = \sum_{k=-K}^{K} b_k\phi_k$ of the penalized least-squares criterion function

$$J^{-1}\sum_{j=1}^{J}|Z_{i,j} - u(t_j)|^2 + \rho\int_0^1 |u^{(m)}|^2,$$

namely,

$$\tilde{X}_{i,\rho}(t) = \phi^T(t)P_\rho\mathbf{Z}_i, \tag{3}$$

where

$$P_\rho = J^{-1}(I + \rho W)^{-1}\bar{\Phi}^T. \tag{4}$$

Since the $\mathbf{Z}_i$ are real, it is easily seen that the $\tilde{X}_{i,\rho}$ are real. Note that this smoother is an approximation to the periodic smoothing spline which uses infinitely many Fourier basis functions. In Theorem 2 below, we will show that, under certain assumptions, the convergence rate of the smoother in (3) is comparable to that of periodic smoothing spline given by Rice and Rosenblatt [17]. See Eubank [9], Section 6.3.1 for more details on periodic splines. This justifies the usage of roughly the same number of Fourier basis functions as the number of points. Using a finite number of basis functions is, of course, crucial for the computations that have to be performed in this problem.

Next, in addition to the smoothing parameter ρ that we used for obtaining the $\tilde{X}_{i,\rho}$, let λ be a second smoothing parameter, and $\hat{f}_{\lambda,\rho}$ be the minimizer $g \in W_{2,\mathrm{per}}^m$ of the following criterion function:

$$n^{-1}\sum_{i=1}^{n}|Y_i - \langle\tilde{X}_{i,\rho}, g\rangle_{L^2}|^2 + \lambda\int_0^1 |g^{(m)}|^2. \tag{5}$$

In real life, a good choice of the smoothing parameters is crucial to the success of the estimators. More discussions on the strategy of smoothing parameter selection are provided in Section 3.

2.2. Theoretical results

Our main results below address the rate of convergence of $\hat{f}_{\lambda,\rho}$ as functions of ρ, λ, as well as the sample sizes n, J.

Denote by $\mathscr{E}$ the space spanned by the eigenfunctions of the covariance operator T. It is clear that if $\mathscr{E}$ is not the whole L^2 space, then f is not identifiable, since $\langle f + g, X\rangle_{L^2} = \langle f, X\rangle_{L^2}$ for any $g \in \mathscr{E}^\perp$. Therefore, we will discuss convergence rate in some functional norms other than L^2.

Let $\tilde{T}_\rho$ be the covariance operator of the sample covariance function

$$\tilde{R}_\rho(s,t) = \frac{1}{n}\sum_{i=1}^{n}\tilde{X}_{i,\rho}(s)\tilde{X}_{i,\rho}(t), \tag{6}$$

and

$$\langle g, h\rangle_{\tilde{T}_\rho} = \langle g, \tilde{T}_\rho h\rangle_{L^2} \quad \text{and} \quad \|g\|^2_{\tilde{T}_\rho} = \langle g, g\rangle_{\tilde{T}_\rho}.$$

Similarly, define $\langle g, h\rangle_T$ and $\|g\|_T$ with T replacing $\tilde{T}_\rho$ in the above. We will consider the convergence rate of $\hat{f}_\lambda$ based on the following two criteria:

$$E(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\mathbf{Z}) \quad \text{and} \quad E(\|\hat{f}_{\lambda,\rho} - f\|^2_T|\mathbf{Z}), \tag{7}$$

where $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^T$. Since

$$\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho} = n^{-1}\sum_{i=1}^n |\langle \tilde{X}_{i,\rho}, f\rangle_{L^2} - \langle \tilde{X}_{i,\rho}, \hat{f}_{\lambda,\rho}\rangle_{L^2}|^2,$$

$E(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\mathbf{Z})$ is heuristically a measure of prediction error based on the smoothed empirical distribution. On the other hand, $E(\|\hat{f}_{\lambda,\rho} - f\|^2_T|\mathbf{Z})$ is a measure of prediction error for a future observation. Indeed, if we ideally observe every value on a new curve $X_0(t)$ from the same population but independent of X_i, $i = 1, \ldots, n$, then the conditional mean square prediction error for $\langle X_0, f\rangle_{L^2}$ is

$$E[(\langle X_0, \hat{f}_{\lambda,\rho}\rangle_{L^2} - \langle X_0, f\rangle_{L^2})^2|\mathbf{Z}] = E(\|\hat{f}_{\lambda,\rho} - f\|^2_T|\mathbf{Z}).$$

Nevertheless, as will be shown below, $\tilde{T}_\rho$ converges to T in the large sample scenario and the two criteria are very close. Similar considerations exist in Cai and Hall [3] and Cardot et al. [4].

Now we state the main results.

Theorem 1. *There exists some finite constant C that depends only on f such that*

$$E(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\mathbf{Z}) \leqslant C\left(\lambda + n^{-1}\lambda^{-1/(2m)}v_\rho^2 + \frac{1}{n}\sum_{i=1}^n E(\|\tilde{X}_{i,\rho} - X_i\|^2_{L^2}|\mathbf{Z}_i)\right) \tag{8}$$

for all n, J, λ, and ρ, where v_ρ is the largest eigenvalue of $\tilde{T}_\rho$.

Note that the first and second terms on the right of (8) describe the square bias and variance, respectively, of the procedure; the third term there essentially reflects the error of approximating X_i by $\tilde{X}_{i,\rho}$.

To obtain a concrete rate of convergence for $E(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\mathbf{Z})$, the following Theorem 2 is crucial.

Theorem 2. *Assume that $X \in W^m_{2,\text{per}}$ a.s. and $E(\|X^{(m)}\|^2_{L^2}) < \infty$. Then*

$$E(\|\tilde{X}_{1,\rho} - X_1\|^2_{L^2}) = O(\rho + J^{-1}\rho^{-1/(2m)}) \tag{9}$$

for $\rho \to 0$ and $J^{2m}\rho \to \infty$. If, in addition, $X \in W^{2m}_{2,\text{per}}$ a.s. and $E(\|X^{(2m)}\|^2_{L^2}) < \infty$, then

$$E(\|\tilde{X}_{1,\rho} - X_1\|^2_{L^2}) = O(\rho^2 + J^{-1}\rho^{-1/(2m)}) \tag{10}$$

for $\rho \to 0$ and $J^{2m}\rho \to \infty$.

Theorem 2 is similar in spirit to Theorem 2 of Rice and Rosenblatt [17], which studies the rate of convergence of the periodic smoothing spline estimator in nonparametric regression. As we mentioned before, even though $\tilde{X}_{1,\rho}$ is estimated with a finite number of Fourier basis functions, the rate of convergence is comparable to that of the periodic smoothing spline estimator using an infinite number of basis functions. The result (10) shows that with the extra conditions $X_1 \in W_{2,\mathrm{per}}^{2m}$ a.s. and $E(\|X_1^{(2m)}\|_{L^2}^2) < \infty$ in place but not specifically taken into account in the estimation procedure, the rate of convergence will nevertheless improve. This also parallels Rice and Rosenblatt's treatment of the periodic smoothing spline.

For the case where $\sigma_\varsigma = 0$, i.e., the $X_i(t_j)$ are observed without measurement error, we have

$$E(\|\tilde{X}_{1,0} - X_1\|_{L^2}^2) \leqslant C J^{-(2m-1)},$$

under $E\|X_1^{(m)}\|_{L^2}^2 < \infty$. The proof of this result follows in a straightforward manner from the derivations in the proof of Theorem 2 and is omitted.

The term $n^{-1}\sum_{i=1}^n E(\|\tilde{X}_{i,\rho} - X_i\|_{L^2}^2 | \mathbf{Z}_i)$ in (8) clearly converges in L^1 to 0 at the rates described by (9) and (10) under the respective settings there. Further, since T is bounded, it is natural to expect that $\tilde{T}_\rho$ is also bounded so that $v_\rho = O_p(1)$ under appropriate conditions. Thus, we state the following result.

Theorem 3. *Suppose that* $X \in W_{2,\mathrm{per}}^m$ *a.s., and* $E(\|X^{(m)}\|_{L^2}^2) < \infty$. *Then*

$$E(\|\hat{f}_{\lambda,\rho} - f\|_{\tilde{T}_\rho}^2 | \mathbf{Z}) = O_p(\lambda + \rho + n^{-1}\lambda^{-1/(2m)} + J^{-1}\rho^{-1/(2m)}) \tag{11}$$

for $\lambda \to 0, \rho \to 0, n^{2m}\lambda \to \infty$, *and* $J^{2m}\rho \to \infty$. *If, in addition,* $X \in W_{2,\mathrm{per}}^{2m}$ *a.s., and* $E(\|X^{(2m)}\|_{L^2}^2) < \infty$, *then we have*

$$E(\|\hat{f}_{\lambda,\rho} - f\|_{\tilde{T}_\rho}^2 | \mathbf{Z}) = O_p(\lambda + \rho^2 + n^{-1}\lambda^{-1/(2m)} + J^{-1}\rho^{-1/(2m)}) \tag{12}$$

for $\lambda \to 0, \rho \to 0, n^{2m}\lambda \to \infty$, *and* $J^{2m}\rho \to \infty$.

It follows from (11) that the optimal rate of convergence of $\hat{f}_{\lambda,\rho}$ in $\tilde{T}_\rho$-norm is

$$n^{-2m/(2m+1)} + J^{-2m/(2m+1)}$$

under the general assumptions of Theorem 3; the rate can be improved to

$$n^{-2m/(2m+1)} + J^{-4m/(4m+1)}$$

under the additional assumptions $X \in W_{2,\mathrm{per}}^{2m}$ a.s. and $E(\|X^{(2m)}\|_{L^2}^2) < \infty$, as described by (12).

In the following, we consider rates of convergence if $\tilde{T}_\rho$ is replaced by T. To do that we need to quantify the distance between $\tilde{T}_\rho$ and T, for which the Hilbert–Schmidt norm seems ideal. See Conway [6] and Dauxois et al. [8]. For any self-adjoint operator A on $L^2[0, 1]$, let $\|A\|_{\mathcal{H}}$ be the Hilbert–Schmidt norm of the operator.

Theorem 4. *Suppose that, for some* $m \geqslant 2$, $X \in W_{2,\mathrm{per}}^m$ *a.s.,* $\sup_t E\{[X^{(m)}(t)]^2\} < \infty$, *and* $E(\|X^{(m)}\|_{L^2}^4) < \infty$. *Also assume that* $E(\varsigma^4) < \infty$. *Then*

$$E(\|\tilde{T}_\rho - T\|_{\mathcal{H}}^2) = O(n^{-1} + \rho + J^{-2}\rho^{-1/(2m)})$$

for $\lambda \to 0, \rho \to 0, n^{2m}\lambda \to \infty$, *and* $J^{2m}\rho \to \infty$. *If, in addition,* $X \in W_{2,\mathrm{per}}^{2m}$ *a.s.*, $\sup_t E\{[X^{(2m)}(t)]^2\} < \infty$, *and* $E(\|X^{(2m)}\|_{L^2}^4) < \infty$, *then*

$$E(\|\tilde{T}_\rho - T\|_{\mathcal{H}}^2) = O(n^{-1} + \rho^2 + J^{-2}\rho^{-1/(2m)})$$

for $\lambda \to 0, \rho \to 0, n^{2m}\lambda \to \infty$, *and* $J^{2m}\rho \to \infty$.

Note that Theorem 4 should be compared with the results in Dauxois et al. [8] which were proved under the assumption that the X_i are completely and precisely observed.

The following result gives the rates of convergence of $\hat{f}_{\lambda,\rho}$ in T-norm.

Theorem 5. *Suppose that for some* $m \geqslant 2$, $X \in W_{2,\mathrm{per}}^m$ *a.s.*, $\sup_t E\{[X^{(m)}(t)]^2\} < \infty$, *and* $E(\|X^{(m)}\|_{L^2}^4) < \infty$. *Also assume that* $E(\varsigma^4) < \infty$. *Then*

$$E(\|\hat{f}_{\lambda,\rho} - f\|_T^2 | \mathbf{Z}) = O_p(n^{-1/2} + \lambda + n^{-1}\lambda^{-1/(2m)} + \rho^{1/2} + J^{-1}\rho^{-1/2m})$$

for n, J, λ, ρ *with*

$$\begin{aligned} &\lambda \to 0, \quad \rho \to 0, \quad n^{2m}\lambda \to \infty, \quad (n\lambda)^{-1} = O(1), \\ &\qquad J^{2m}\rho \to \infty \quad \text{and} \quad (\rho + J^{-1}\rho^{-1/(2m)})/\lambda = O(1). \end{aligned} \tag{13}$$

If, in addition, $X \in W_{2,\mathrm{per}}^{2m}$ *a.s.*, $\sup_t E\{[X^{(2m)}(t)]^2\} < \infty$, *and* $E(\|X^{(2m)}\|_{L^2}^4) < \infty$, *then*

$$E(\|\hat{f}_{\lambda,\rho} - f\|_T^2 | \mathbf{Z}) = O_p(n^{-1/2} + \lambda + n^{-1}\lambda^{-1/(2m)} + \rho + J^{-1}\rho^{-1/2m})$$

for n, J, λ, ρ *satisfying* (13).

3. Computational issues and a simulation study

3.1. Smoothing parameter selection

As explained, our approach requires two smoothing parameters, ρ and λ, for smoothing each individual curve and estimating the regression weight function, respectively.

There is sizeable literature on smoothing a group of curves, for example Brumback and Rice [2], Guo [10], Ruppert et al. [18]. As pointed out by these authors, when the group of curves are modeled as realizations of a stochastic process, it is reasonable to use the same smoothing parameter for all the curves. In addition, the relationship between smoothing spline or P-spline with mixed-effects models has long been established, and thus the smoothing parameter can be chosen by standard variation estimation procedures in mixed-effects models, such as REML. See Ruppert et al. [18] for a comprehensive review. Following these, the periodic smoothing splines used in the first stage of our method also has a mixed-effects model interpretation, and the smoothing parameter ρ can be chosen by REML. The details are given as follows.

In our theoretical derivation, we used complex Fourier basis in order to simplify notation. In a real-data analysis, one can avoid complex numbers by using real Fourier basis: $\phi_0(t) = 1, \phi_k(t) = \sqrt{2}\cos(2k\pi t)$ for $k = 1, \ldots, K$, and $\phi_k(t) = \sqrt{2}\sin\{2(k-K)\pi t\}$ for $k = K+1, \ldots, 2K$, where $K = [J/2]$. The mixed-effects model for the functional predictors is the following. Let

$$\mathbf{Z}_i = v_i\mathbf{1} + \widetilde{\Phi}\mathbf{u}_i + \varsigma_i, \quad i = 1, \ldots, n,$$

where $\varsigma_i \sim \text{MVN}(\mathbf{0}, \sigma_\varsigma^2 I)$, v_i are the fixed effects, $\mathbf{u}_i \sim \text{MVN}(\mathbf{0}, (\sigma_\varsigma^2/\rho)W^{-1})$ are the random effects, $\widetilde{\Phi} = \{\phi_k(t_j)\}_{J\times 2K}$, and $W = (\langle \phi_{k_1}^{(m)}, \phi_{k_2}^{(m)} \rangle)_{2K\times 2K}$. Then the periodic smoothing splines $\widetilde{X}_i$ are the BLUPs of this mixed-effects model, and the smoothing parameter ρ can be selected by REML.

After smoothing each individual curve, the second stage of our method is simply a ridge regression. The smoothing parameter can, in principle, be chosen by a number of standard procedures, such as GCV or C_p.

3.2. Simulation study

In this section, we apply our method of estimation to a simulated data set and compare the result with that of PCR approach.

In our simulations, we generated X_i by

$$\begin{aligned} X_i(t) = {} & u_{i,0} + \sum_{k=1}^{\infty} u_{i,k}\sqrt{2}\cos(2k\pi t)/(k+0.5)^3 + \sum_{k=1}^{\infty} v_{i,k}\sqrt{2}\sin(2k\pi t)/(k+0.5)^3 \\ & + \widetilde{u}_{i,1}\{1+\sqrt{2}\cos(8\pi t) + 0.75\sqrt{2}\sin(16\pi t)\}/36 \\ & + \widetilde{u}_{i,2}\{0.6\sqrt{2}\cos(4\pi t) + 1.5\sqrt{2}\sin(10\pi t)\}/25, \end{aligned} \tag{14}$$

where $u_{i,k}, v_{i,k}, \widetilde{u}_{i,k} \sim \text{Normal}(0, 1)$. The last two terms on the rhs of (14) introduce dependence into the Fourier coefficients of $X(t)$, and also change the principal component directions somewhat so that they are not simply the Fourier basis functions.

Suppose the regression weight function has the following decomposition

$$f(t) = b_0 + \sum_{k=1}^{\infty} b_{1,k}\sqrt{2}\cos(2k\pi t) + \sum_{k=1}^{\infty} b_{2,k}\sqrt{2}\sin(2k\pi t).$$

Then

$$\begin{aligned} Y_i &= \int_0^1 X_i(t) f(t)\,dt + \varepsilon_i \\ &= u_{i,0}b_0 + \sum_{k=1}^{\infty}(u_{i,k}b_{1,k} + v_{i,k}b_{2,k})/(k+0.5)^3 \\ &\quad \times \widetilde{u}_{i,1}(b_0 + b_{1,4} + 0.75 b_{2,8})/36 + \widetilde{u}_{i,2}(0.6 b_{1,2} + 1.5 b_{2,5})/25 + \varepsilon_i. \end{aligned}$$

In our simulations, we set $\sigma_\varepsilon^2 = 0.09$, and adopted the following weight function:

$$\begin{aligned} f(t) = {} & 0.8 - 0.8\sqrt{2}\cos(2\pi t) + 0.7\sqrt{2}\sin(2\pi t) - 1.4\sqrt{2}\cos(4\pi t) \\ & + 0.65\sqrt{2}\sin(4\pi t) + \sum_{k=3}^{\infty} k^{-3}\sqrt{2}\{\cos(2k\pi t) + \sin(2k\pi t)\}. \end{aligned}$$

The functional predictors $X_i(t)$ were observed at $J = 100$ equally spaced points, and the measurements were contaminated with measurement errors with $\sigma_\varsigma^2 = 0.04$. In each simulation run, we generated $n = 100$ observations for training (estimation) and another 100 for validation (prediction), and used $\langle X_{\text{new}}, \hat{f}_{\lambda,\rho}\rangle_{L^2}$ to predict $\langle X_{\text{new}}, f\rangle_{L^2}$.

For the training sample, we first smoothed each curve with the same smoothing parameter ρ selected by REML. The results of smoothing for the first nine curves are shown in Fig. 1. As one

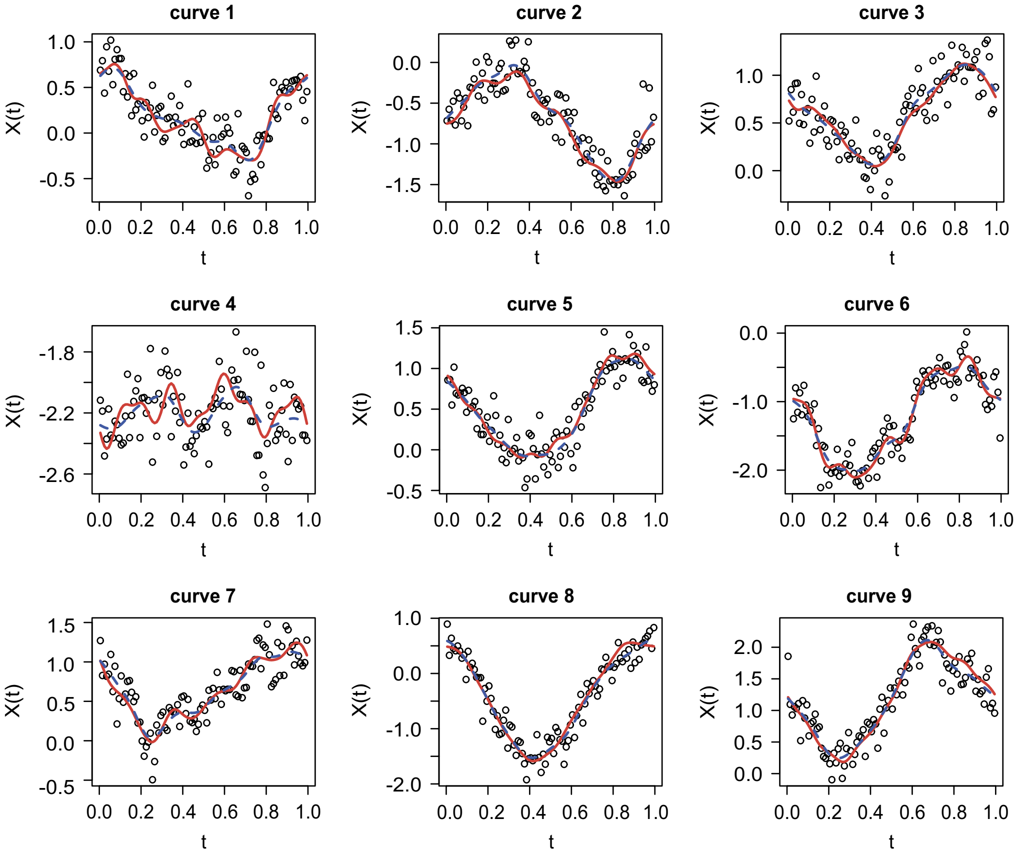

Fig. 1. Smoothing each individual curves, ρ chosen by REML. The dots are the error contaminated observations $\mathbf{Z}_i$, the red solid curves are the true functional predictor $X_i(t)$, and the blue dashed curves are the ones recovered from noisy data $\widetilde{X}_i(t)$.

can see the results are quite satisfactory. In the second stage, we estimated the weight function, with the smoothing parameter λ chosen by generalized cross-validation (GCV). We then applied the estimator to the simulated validation data and calculated the prediction mean-squared error. As discussed before, the prediction mean-squared error is related to the T-norm criterion in our theoretical results.

A numerical comparison was made with the PCR approach, which is another popular approach for functional linear regression. See the references given in Section 1. Generally, the PCR approach assumes that the data X_i are completely observed as curves, where the leading principal components of the estimated covariance functions are used as basis functions. The number of principal components to be used can also, in principle, be chosen by GCV or C_p. Since the X_i in our case are observed at discrete points, to make the comparison fair, we applied the same smoothing step as in the first stage of our approach to smooth each curve before carrying out the functional principal component analysis. The number of principal components entering the model was selected by the cross-validation approach suggested in Cai and Hall [3].

We observed from our simulations that whenever the data-driven procedures selected appropriate λ and the number of principal components, both procedures worked well but our approach had a clear advantage in terms of the prediction mean-squared error. We attribute this partly to the fact the response Y is not directly linked to the principal components of the predictor $X(t)$. We also observed that both data-driven parameter selection procedures bring extra variability to the estimators, especially for the PCR approach; in fact, since the principal components of $X(t)$ do not contribute to Y directly, cross-validation tends to admit too many components into the model, causing more frequent failures for the PCR. As such, in order to still make some sort of comparison in multiple-run simulations, one possibility is to fix λ and the number of principal components in the two procedures. Thus, we conducted 100 simulation runs of the experiment described above, with a fixed λ selected by a successful run and the number of principal components set to be 5, which, by our experience, is when PCR performs the best in this particular model. The averaged prediction mean square error over the 100 runs is 0.0048 for our approach and 0.0050 for the PCR approach.

This simulation study provides an example where our penalized least-squares approach is a highly competitive alternative to the PCR.

4. Proofs

Throughout this section, we adopt the following notations: for any complex-valued vector b, $\|b\|^2 = b^T\bar{b}$; for any zero-mean, complex-valued, random variables X and Y, $\mathrm{Cov}(X, Y) = E(X\bar{Y})$.

4.1. Proof of Theorem 1

Since $\tilde{X}_{i,\rho} \in \mathrm{span}\{\phi_i(\cdot), i = -K, \ldots, K\}$, by orthogonality of Fourier basis, $\hat{f}_{\lambda,\rho}$ is spanned by the same set of basis functions and we can write $\hat{f}_{\lambda,\rho} = \phi^T\hat{\beta}$. Then minimizing (5) to obtain $\hat{f}_{\lambda,\rho}$ for all $g \in W^m_{2,\mathrm{per}}$ is equivalent to minimizing

$$n^{-1}\sum_{i=1}^{n} |Y_i - \mathbf{Z}_i^T V\bar{b}|^2 + \lambda b^T W\bar{b}$$

to obtain $\hat{\beta}$ for all b in the $(2K+1)$-dimensional complex space, where $V = P_\rho^T\langle\phi, \phi^T\rangle_{L^2} = J^{-1}\bar{\Phi}(I+\rho W)^{-1}$, P_ρ being defined in (4). Since Y_i and $\tilde{X}_{i,\rho}$ are real-valued, it is clear that $\hat{f}_{\lambda,\rho}$ is real-valued, and $\hat{\beta}$ satisfies $\overline{\hat{\beta}_j} = \hat{\beta}_{-j}$, $j = -K, \ldots, K$.

Define

$$\Omega_{n,J} = n^{-1}\mathbf{Z}^T\mathbf{Z} \quad \text{and} \quad \check{\Omega}_\rho = n^{-1}\bar{V}^T\mathbf{Z}^T\mathbf{Z}V, \tag{15}$$

and

$$r_i = \langle X_i, f\rangle_{L^2} - \langle \tilde{X}_{i,\rho}, f\rangle_{L^2} \quad \text{and} \quad \mathbf{r} = (r_1, \ldots, r_n)^T.$$

It is easy to check that $\check{\Omega}_\rho$ is an Hermitian matrix, i.e. $\overline{\check{\Omega}}_\rho^T = \check{\Omega}_\rho$. Suppose that $f = \sum_{j=-\infty}^{\infty}\beta_j\phi_j$, and denote by $\check{f}$ the projection of f onto $\mathrm{span}\{\phi_{-K}, \ldots, \phi_K\}$, and $\check{\beta} = (\beta_{-K}, \ldots, \beta_K)^T$. By orthogonality of Fourier basis, $\langle \tilde{X}_{i,\rho}, f\rangle_{L^2} = \langle \tilde{X}_{i,\rho}, \check{f}\rangle_{L^2} = \mathbf{Z}_i^T V\bar{\check{\beta}}$. On the other hand, since both

$\tilde{X}_{i,\rho}$ and f are real-valued functions, we have $\langle \tilde{X}_{i,\rho}, f\rangle_{L^2} = \mathbf{Z}_i^T \bar{V}\check{\beta}$. Writing $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$, we have

$$\begin{aligned} \hat{f}_{\lambda,\rho}(t) &= \phi^T(t)(V^T\mathbf{Z}^T\mathbf{Z}\bar{V} + n\lambda W)^{-1}V^T\mathbf{Z}^T\mathbf{Y} \\ &= \phi^T(t)(\check{\boldsymbol{\Omega}}_\rho^T + \lambda W)^{-1}n^{-1}V^T\mathbf{Z}^T(\mathbf{Z}\bar{V}\check{\beta} + \mathbf{r} + \boldsymbol{\varepsilon}) \\ &=: \phi^T(t)(\check{\beta}_\lambda + \hat{\beta}_r + \hat{\beta}_\varepsilon) \\ &=: \check{f}_{\lambda,\rho}(t) + g_\lambda(t) + h_\lambda(t). \end{aligned} \tag{16}$$

By (6),

$$\tilde{R}_\rho(s,t) = \frac{1}{n}\sum_{i=1}^n \tilde{X}_{i,\rho}(s)\tilde{X}_{i,\rho}(t) = \phi(s)^T P_\rho \Omega_{n,J} P_\rho^T \phi(t).$$

For any $g(t) = \phi(t)^T b \in \operatorname{span}\{\phi_{-K}, \dots, \phi_K\}$, we have

$$\begin{aligned} (\tilde{T}_\rho g)(t) &= \int \tilde{R}_\rho(s,t)g(s)\,ds \\ &= \phi^T(t)P_\rho\Omega_{n,J}\bar{P}_\rho^T\left\{\int \bar{\phi}(s)\phi^T(s)\,ds\right\}b = \phi^T(t)P_\rho\Omega_{n,J}\bar{V}b \end{aligned}$$

and hence by (15),

$$\|g\|^2_{\tilde{T}_\rho} = \langle g, \tilde{T}_\rho g\rangle_{L^2} = b^T\bar{V}^T\Omega_{n,J}V\bar{b} = b^T\check{\Omega}_\rho\bar{b} = n^{-1}\|\mathbf{Z}\bar{V}b\|^2. \tag{17}$$

It follows from (16) that

$$\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho} \leqslant 3\|\check{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho} + 3\|g_\lambda\|^2_{\tilde{T}_\rho} + 3\|h_\lambda\|^2_{\tilde{T}_\rho}. \tag{18}$$

First, consider $\|\check{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}$ which is equal to $\|\check{f}_{\lambda,\rho} - \check{f}\|^2_{\tilde{T}_\rho}$. By (17),

$$\|\check{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho} = \|\check{f}_{\lambda,\rho} - \check{f}\|^2_{\tilde{T}_\rho} = n^{-1}\|\mathbf{Z}\bar{V}(\check{\beta}_\lambda - \check{\beta})\|^2.$$

Since $\check{f}_{\lambda,\rho}$ is the solution of the following problem:

$$\min_{g=\phi^T b}\{n^{-1}\|\check{\mathbf{Z}}\bar{V}(b - \check{\beta})\|^2 + \lambda b^T W\bar{b}\},$$

we conclude that

$$\begin{aligned} \|\check{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho} &= n^{-1}\|\mathbf{Z}\bar{V}(\check{\beta}_\lambda - \check{\beta})\|^2 \\ &\leqslant n^{-1}\|\mathbf{Z}\bar{V}(\check{\beta}_\lambda - \check{\beta})\|^2 + \lambda\check{\beta}_\lambda^T W\overline{\check{\beta}_\lambda} \\ &\leqslant \lambda\check{\beta}^T W\overline{\check{\beta}} = \lambda\|\check{f}^{(m)}\|^2_{L^2} \leqslant \lambda\|f^{(m)}\|^2_{L^2}. \end{aligned} \tag{19}$$

Note that this is an approach for handling the bias introduced by Craven and Wahba [7]. Next, by (17),

$$\begin{aligned} E(\|h_\lambda\|^2_{\tilde{T}_\rho}|\mathbf{Z}) &= E\{\boldsymbol{\varepsilon}^T n^{-1}\mathbf{Z}V(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho(\overline{\check{\Omega}}_\rho^T + \lambda W)^{-1}n^{-1}\bar{V}^T\mathbf{Z}^T\boldsymbol{\varepsilon}|\mathbf{Z}\} \\ &= n^{-1}\sigma_\varepsilon^2\operatorname{tr}\{(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho\}. \end{aligned}$$

Let ω_i, v_i, ρ_i, and η_i be the ith smallest eigenvalue of W, $\check{\Omega}_\rho$, $\check{\Omega}_\rho + \lambda W$, and $(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho$, respectively. Observe that for any complex-valued function $g = \phi^T b$, we have

$$\frac{\langle g, \tilde{T}_\rho g\rangle_{L^2}}{\|g\|^2_{L^2}} = \frac{b^T\check{\Omega}_\rho\bar{b}}{b^T\bar{b}}, \tag{20}$$

and hence the eigenvalues of $\check{\Omega}_\rho$ are the same as those of $\tilde{T}_\rho$. Thus, $\check{\Omega}_\rho$ is positive semi-definite and $v_i \geqslant 0$ for all i. Note that v_{2K+1} is denoted as v_ρ in the statement of the theorem to emphasize its dependence on ρ. By (2), $\omega_i = ([i/2]2\pi)^{2m}$, $i = 1, \ldots, 2K+1$. We clearly also have

$$\eta_i \leqslant 1 \quad \text{and} \quad \rho_i \geqslant \lambda([i/2]2\pi)^{2m} \geqslant \lambda((i-1)\pi)^{2m} \ \text{ for all } i.$$

It follows from Theorem 7 of Merikoski and Kumar [13] that

$$\eta_{(2K+1)-i+1} \leqslant \rho_i^{-1} v_\rho \quad \text{for all } i.$$

Thus,

$$\eta_{(2K-1)-i+1} \leqslant \min(1, v_\rho \pi^{-2m}\lambda^{-1}(i-1)^{-2m}) \quad \text{for all } i,$$

and therefore

$$\begin{aligned}
&\operatorname{tr}\{(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho\}\\
&\quad = \sum_{i=1}^{2K+1} \eta_i^2 \leqslant \sum_{i=1}^{[\lambda^{-1/(2m)}]} 1 + \frac{v_\rho^2}{\pi^{4m}\lambda^2} \sum_{i=[\lambda^{-1/(2m)}]+1}^{2K+1} (i-1)^{-4m} \leqslant C\lambda^{-1/(2m)} v_\rho^2.
\end{aligned}$$

Hence,

$$E(\|h_\lambda\|^2_{\tilde{T}_\rho}|\mathbf{Z}) \leqslant Cn^{-1}\lambda^{-1/(2m)}v_\rho^2. \tag{21}$$

Next,

$$E(\|g_\lambda\|^2_{\tilde{T}_\rho}|\mathbf{Z}) = E\{\mathbf{r}^T n^{-1}\mathbf{Z}V(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho(\check{\Omega}_\rho + \lambda W)^{-1} n^{-1}\bar{V}^T\mathbf{Z}^T\mathbf{r}|\mathbf{Z}\}.$$

Since the eigenvalues of $\mathbf{Z}V(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho(\check{\Omega}_\rho + \lambda W)^{-1}n^{-1}\bar{V}^T\mathbf{Z}^T$ are the same as those of

$$(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho(\check{\Omega}_\rho + \lambda W)^{-1}n^{-1}\bar{V}^T\mathbf{Z}^T\mathbf{Z}V = [(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho]^2,$$

which are bounded by 1, we conclude that

$$\begin{aligned}
E(\|g_\lambda\|^2_{\tilde{T}_\rho}|\mathbf{Z}) &\leqslant \frac{1}{n}E(\mathbf{r}^T\mathbf{r}|\mathbf{Z}) = \frac{1}{n}\sum_{i=1}^n E(r_i^2|\mathbf{Z}_i)\\
&\leqslant \frac{\|f\|^2_{L^2}}{n}\sum_{i=1}^n E(\|\tilde{X}_{i,\rho} - X_i\|^2_{L^2}|\mathbf{Z}_i),
\end{aligned} \tag{22}$$

by the Cauchy–Schwarz inequality. The result follows from (18), (19), (21), and (22).

4.2. Proof of Theorem 2

Lemma 6. *Suppose* $t_j = (2j-1)/(2J)$, $j = 1, \ldots, J$, *then*

$$\sum_{j=1}^{J} \bar{\phi}_{k_1}(t_j)\phi_{k_2}(t_j) = \begin{cases} (-1)^s J & \text{if } k_2 = k_1 + sJ, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. If $k_2 - k_1$ is not a multiple of J, we have

$$\sum_{j=1}^{J} \bar{\phi}_{k_1}(t_j)\phi_{k_2}(t_j) = \sum_{j=1}^{J} e^{(k_2-k_1)2\pi i t_j} = \frac{e^{(k_2-k_1)2\pi i(2J+1)/(2J)} - e^{(k_2-k_1)2\pi i/(2J)}}{1 - e^{(k_2-k_1)2\pi i/J}} = 0.$$

Next, suppose $k_2 = k_1 + sJ$ for some integer s, we have

$$\sum_{j=1}^{J} \bar{\phi}_{k_1}(t_j)\phi_{k_2}(t_j) = \sum_{j=1}^{J} e^{s\pi i(2j-1)} = \sum_{j=1}^{J}(-1)^s = J(-1)^s. \qquad \square$$

Lemma 7. *Let* $g \in W^m_{2,\text{per}}$ *have the Fourier basis representation* $g = \sum_{j=-\infty}^{\infty} c_j\phi_j(t)$ *in* $L^2[0, 1]$. *Then the Fourier basis representation for* $g^{(m)}$ *is*

$$\sum_{j=-\infty}^{\infty} c_j(2\pi ji)^m\phi_j(t),$$

and we have

$$\|g^{(m)}\|^2_{L^2} = \sum_{j=-\infty}^{\infty} (2\pi j)^{2m}|c_j|^2.$$

Proof. Let $g_k = \sum_{j=-k}^{k} c_j\phi_j$ and consider $g^{(m)} - g_k^{(m)}$. Note that the assumption implies that $g^{(v)}(0) = g^{(v)}(1)$, $0 \leqslant v \leqslant m-1$, and $\phi_j^{(s)}(0) = \phi_j^{(s)}(1)$ for all s. Integrating by parts repeatedly,

$$\langle g^{(m)} - g_k^{(m)}, \phi_j\rangle_{L^2} = (-1)^m\langle g - g_k, (2\pi ji)^m\phi_j\rangle_{L^2} = 0 \quad \text{for all } j = -k, \ldots, k.$$

This means that $g_k^{(m)} = \sum_{i=-k}^{k}(2\pi ji)^m c_j\phi_j$ is the L^2 projection of $g^{(m)}$ on $\text{span}\{\phi_j, j = -k, \ldots, k\}$. Since $g^{(m)} \in L^2[0, 1]$, and the Fourier basis is complete, we conclude that $g_k^{(m)} \to g^{(m)}$ in $L^2[0, 1]$ and the result follows. $\square$

Proof of Theorem 2. The proof is similar to those in Rice and Rosenblatt [17]. As before,

$$\mathbf{X} = (X(t_1), \ldots, X(t_J))^T, \quad \varsigma = (\varsigma_1, \ldots, \varsigma_J) \quad \text{and} \quad \mathbf{Z} = \mathbf{X} + \varsigma.$$

Let $X(t) = \sum_{j=-\infty}^{\infty} a_j\phi_j(t)$. By Lemma 6,

$$\tilde{a}_j := J^{-1}\sum_{l=1}^{J} \phi_j(t_l)X(t_l) = \sum_{s=-\infty}^{\infty}(-1)^s a_{j+sJ}, \ j = -K, \ldots, K.$$

Thus,

$$
\begin{aligned}
\tilde{X}_\rho(t) &= \phi^T(t)(I+\rho W)^{-1}J^{-1}\Phi^T\mathbf{Z}\\
&= \sum_{j=-K}^{K}(1+(2\pi)^{2m}\rho j^{2m})^{-1}(\tilde{a}_j+\tilde{\varsigma}_j)\phi_j(t),
\end{aligned}
$$

where $\tilde{\varsigma}_j = J^{-1}\sum_{k=1}^{J}\phi_j(t_k)\varsigma_k$. Next,

$$
\begin{aligned}
E(\|\tilde{X}_\rho - X\|^2) &= E\left[\int\left|\sum_{j=-K}^{K}\{a_j-(1+(2\pi)^{2m}\rho j^{2m})^{-1}(\tilde{a}_j+\tilde{\varsigma}_j)\}\phi_j(t)\right|^2 dt\right]\\
&\qquad + E\left[\int\left|\sum_{|j|>K}a_j\phi_j(t)\right|^2 dt\right]\\
&= E\left[\sum_{j=-K}^{K}\frac{|\rho(2\pi)^{2m}j^{2m}a_j-(\tilde{a}_j-a_j)-\tilde{\varsigma}_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})^2}\right]+E\left[\sum_{|j|>K}|a_j|^2\right]\\
&\leqslant 2E\left[\sum_{j=-K}^{K}\frac{\rho^2(2\pi)^{4m}j^{4m}|a_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})^2}\right]+2E\left[\sum_{j=-K}^{K}\frac{|\tilde{a}_j-a_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})^2}\right]\\
&\qquad + \sum_{j=-K}^{K}\frac{J^{-1}\sigma_\varsigma^2}{(1+\rho(2\pi)^{2m}j^{2m})^2}+E\left[\sum_{|j|>K}|a_j|^2\right].
\end{aligned}
$$

Note that, by Lemma 7,

$$
\|X^{(m)}\|_{L^2}^2 = \sum_{j=-\infty}^{\infty}(2j\pi)^{2m}|a_j|^2 \quad \text{a.s.}
$$

and therefore, with probability one,

$$
\begin{aligned}
\sum_{j=-K}^{K}\frac{\rho^2(2\pi)^{4m}j^{4m}|a_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})^2} &\leqslant \sum_{j=-K}^{K}\frac{\rho(2\pi)^{2m}j^{2m}|a_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})}\\
&\leqslant \rho\sum_{j=-K}^{K}(2\pi)^{2m}j^{2m}|a_j|^2 \leqslant \rho\|X^{(m)}\|_{L^2}^2
\end{aligned}
$$

and

$$
\sum_{|j|>K}|a_j|^2 \leqslant (2\pi)^{-2m}K^{-2m}\sum_{|j|>K}(2\pi)^{2m}j^{2m}|a_j|^2 \leqslant K^{-2m}(2\pi)^{-2m}\|X^{(m)}\|_{L^2}^2.
$$

By integral approximation,

$$\sum_{j=-K}^{K} (1+\rho(2\pi j)^{2m})^{-r} \sim (2\pi)^{-1}\rho^{-1/(2m)} \int_{-\infty}^{\infty} (1+x^{2m})^{-r}\, dx, \quad r \geqslant 1. \tag{23}$$

On the other hand, with probability one,

$$\begin{aligned} |\tilde{a}_j - a_j|^2 &= \left| \sum_{s\neq 0} (-1)^s a_{j+sJ} \right|^2 \\ &\leqslant \left\{ \sum_{s\neq 0} (j+sJ)^{-2m} \right\} \left\{ \sum_{s\neq 0} (j+sJ)^{2m} |a_{j+sJ}|^2 \right\} \\ &= O(J^{-2m}) \|X^{(m)}\|_{L^2}^2, \end{aligned}$$

so that

$$\sum_{j=-K}^{K} \frac{|\tilde{a}_j - a_j|^2}{(1+\rho(2\pi)^{2m} j^{2m})^2} = O(J^{-2m}\rho^{-1/2m}) \|X^{(m)}\|_{L^2}^2.$$

Combining these and applying the assumptions $K \sim J/2$, and $E(\|X^{(m)}\|_{L^2}^2) < \infty$, we obtain (9).

Next, if $X \in W_{2,\text{per}}^{2m}$, we have $\sum_{j=-\infty}^{\infty} (2\pi j)^{4m} |a_j|^2 = \|X^{(2m)}\|_{L^2}^2$. The proof of (10) is the same as that for (9), except that now we have, with probability one,

$$\sum_{|j|>K} |a_j|^2 \leqslant (2\pi)^{-4m} K^{-4m} \sum_{|j|>K} (2\pi)^{4m} j^{4m} |a_j|^2 \leqslant K^{-4m} (2\pi)^{-4m} \|X^{(2m)}\|_{L^2}^2,$$

$$|\tilde{a}_j - a_j|^2 \leqslant \left\{ \sum_{s\neq 0} (j+sJ)^{-4m} \right\} \left\{ \sum_{s\neq 0} (j+sJ)^{4m} |a_{j+sJ}|^2 \right\} \leqslant O(J^{-4m}) \|X^{(2m)}\|_{L^2}^2,$$

$$\sum_{j=-K}^{K} \frac{\rho^2 (2\pi)^{4m} j^{4m} |a_j|^2}{(1+\rho(2\pi)^{2m} j^{2m})^2} \leqslant \rho^2 \sum_{j=-K}^{K} (2\pi)^{4m} j^{4m} |a_j|^2 \leqslant \rho^2 \|X^{(2m)}\|_{L^2}^2.$$

Therefore, the term $O(\rho)$ in (9) is replaced by $O(\rho^2)$ and (10) follows. $\square$

4.3. Proof of Theorem 3

In view of the discussions prior to the statement of the theorem, it suffices to show that $v_\rho = O_p(1)$, where v_ρ is the largest eigenvalue of $\tilde{T}_\rho$. By (20), $\tilde{T}_\rho$ and $\check{\Omega}_\rho$ have the same eigenvalues. By (15), the eigenvalues of $\check{\Omega}_\rho$ are bounded by those of $J^{-1}\Omega_{n,J} = J^{-1}n^{-1}\mathbf{Z}^T\mathbf{Z}$. Hence it suffices to show that

$$\sup_{n,J} J^{-2} E[\operatorname{tr}(\Omega_{n,J}^2)] < \infty.$$

Straightforward computations show that

$$\frac{1}{J^2} E[\operatorname{tr}(\Omega_{n,J}^2)] = \frac{1}{nJ^2} \sum_{j=1}^{J} \sum_{k=1}^{J} E[(X_1(t_j) + \varsigma_{1,j})^2 (X_1(t_k) + \varsigma_{1,k})^2]$$

$$+ \frac{n-1}{nJ^2} \sum_{j=1}^{J} \sum_{k=1}^{J} E^2[(X_1(t_j) + \varsigma_{1,j})(X_1(t_k) + \varsigma_{1,k})].$$

By the Cauchy–Schwarz inequality, it is sufficient to deal with the first expression on the right. Since the X_i and the $\varsigma_{i,j}$ are independent,

$$\frac{1}{J^2} \sum_{j=1}^{J} \sum_{k=1}^{J} E[(X_1(t_j) + \varsigma_{1,j})^2 (X_1(t_k) + \varsigma_{1,k})^2]$$

$$= \frac{1}{J^2} \sum_{j=1}^{J} \sum_{k=1}^{J} \{E[X_1^2(t_j) X_1^2(t_k)] + \sigma_\varsigma^2 (E[X_1^2(t_j)] + E[X_1^2(t_k)]) + \sigma_\varsigma^4\},$$

which will be bounded under the assumption that $E(\|X_1\|_{L^2}^4) < \infty$.

4.4. Proof of Theorem 4

Let $X_{\{\ell\}} = \sum_{j=-\ell}^{\ell} a_j \phi_j$, namely the projection of X on $\operatorname{span}\{\phi_k, -\ell \leqslant k \leqslant \ell\}$. Then

$$R(s,t) = E[X(s)\bar{X}(t)] = \lim_{\ell_1, \ell_2 \to \infty} E[X_{\{\ell_1\}}(s) \bar{X}_{\{\ell_2\}}(t)]$$

$$= \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} E(a_{j_1} \bar{a}_{j_2}) \phi_{j_1}(s) \bar{\phi}_{j_2}(t).$$

For convenience, write $a_{j_1,j_2} = E(a_{j_1} \bar{a}_{j_2})$. By an argument similar to that used in Lemma 7, using the assumption $E\{[X^{(m)}(s)]^2\} < \infty$ for all s, we have

$$R^{(m,m)}(s,t) = \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} a_{j_1,j_2} (2\pi i j_1)^m (-2\pi i j_2)^m \phi_{j_1}(s) \bar{\phi}_{j_2}(t)$$

$$= \lim_{\ell_1, \ell_2 \to \infty} E[X_{\{\ell_1\}}^{(m)}(s) \bar{X}_{\{\ell_2\}}^{(m)}(t)] = E[X^{(m)}(s) \bar{X}^{(m)}(t)].$$

Consequently, we have

$$\sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} (2\pi j_1)^{2m} (2\pi j_2)^{2m} |a_{j_1,j_2}|^2$$

$$= \int_0^1 \int_0^1 [R^{(m,m)}(s,t)]^2 \, ds \, dt = \int_0^1 \int_0^1 E^2[X^{(m)}(s) X^{(m)}(t)] \leqslant E\|X^{(m)}\|_{L^2}^4 < \infty. \quad (24)$$

Similarly, under the additional conditions $X \in W^{2m}_{2,\mathrm{per}}$ a.s., $E\{[X^{(2m)}(s)]^2\} < \infty$ for all s, and $E\|X^{(2m)}\|^2_{L^2} < \infty$, we can show

$$\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}(2\pi j_1)^{4m}|a_{j_1,j_2}|^2 = \int_0^1\int_0^1 |R^{(2m,0)}(s,t)|^2 = \int_0^1\int_0^1 E^2[X^{(2m)}(s)X(t)] \leqslant \left(E\|X^{(2m)}\|^4_{L^2}E\|X\|^4_{L^2}\right)^{1/2} < \infty. \tag{25}$$

Define $R_\rho = E[\tilde{R}_\rho(s,t)]$ and let T_ρ be the corresponding covariance operator. The following calculations are similar to those in Theorem 2. Let $\mathbf{\Sigma} = E(\mathbf{X}_1\mathbf{X}_1^T) = (R(t_l,t_k))_{l,k=1}^J$, it follows that

$$\begin{aligned}
R_\rho(s,t) &= \phi^T(s)P_\rho E(\mathbf{Z}_1\mathbf{Z}_1^T)\bar{P}_\rho^T\bar{\phi}(t)\\
&= J^{-2}\phi^T(s)(I+\rho W)^{-1}\bar{\Phi}^T(\Sigma+\sigma_\varsigma^2 I)\Phi(I+\rho W)^{-1}\bar{\phi}(t)\\
&= \sum_{j=-K}^{K}\sum_{k=-K}^{K}\tilde{a}_{jk}\{1+\rho(2\pi j)^{2m}\}^{-1}\{1+\rho(2\pi k)^{2m}\}^{-1}\phi_j(s)\bar{\phi}_k(t)\\
&\quad+\sigma_\varsigma^2 J^{-1}\sum_{j=-K}^{K}\{1+\rho(2\pi j)^{2m}\}^{-2}\phi_j(s)\bar{\phi}_j(t),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{a}_{j_1,j_2} &= \sum_{l=1}^{J}\sum_{k=1}^{J}R(t_l,t_k)\bar{\phi}_{j_1}(t_l)\phi_{j_2}(t_k)\\
&= \sum_{s_1=-\infty}^{\infty}\sum_{s_2=-\infty}^{\infty}(-1)^{s_1+s_2}a_{j_1+s_1J,j_1+s_2J}, \quad -K\leqslant j_1,j_2\leqslant K.
\end{aligned}$$

Then

$$\begin{aligned}
\|T_\rho - T\|^2_{\mathcal{H}} &= \int_0^1\int_0^1[R_\rho(s,t)-R(s,t)]^2\,ds\,dt\\
&\leqslant 2\sum_{j_1=-K}^{K}\sum_{j_2=-K}^{K}|\tilde{a}_{j_1,j_2}\{1+\rho(2\pi j_1)^{2m}\}^{-1}\{1+\rho(2\pi j_2)^{2m}\}^{-1}-a_{j_1,j_2}|^2\\
&\quad+2\sigma_\varsigma^4 J^{-2}\sum_{j=-K}^{K}\{1+\rho(2\pi j)^{2m}\}^{-4}+\sum_{|j_1|>K}\sum_{|j_2|>K}|a_{j_1,j_2}|^2.
\end{aligned}$$

Suppose first that $X \in W^m_{2,\mathrm{per}}$ and $E\|X^{(m)}\|^4_{L^2} < \infty$. By (24),

$$\sum_{|j_1|>K}\sum_{|j_2|>K}|a_{j_1,j_2}|^2 \leqslant (2\pi K)^{-4m}\sum_{|j_1|>K}\sum_{|j_2|>K}(2\pi j_1)^{2m}(2\pi j_2)^{2m}|a_{j_1,j_2}|^2 = O_p(J^{-4m}),$$

and by (23) with $r = 4$,

$$J^{-2} \sum_{j=-K}^{K} \{1 + \rho(2\pi j)^{2m}\}^{-4} = O(J^{-2}\rho^{-1/(2m)}).$$

Now,

$$\sum_{j_1=-K}^{K} \sum_{j_2=-K}^{K} |\tilde{a}_{j_1,j_2}\{1 + \rho(2\pi j_1)^{2m}\}^{-1}\{1 + \rho(2\pi j_2)^{2m}\}^{-1} - a_{j_1,j_2}|^2 \leqslant 2(A + B),$$

where

$$A = \sum_{j_1=-K}^{K} \sum_{j_2=-K}^{K} |\tilde{a}_{j_1,j_2} - a_{j_1,j_2}|^2\{1 + \rho(2\pi j_1)^{2m}\}^{-2}\{1 + \rho(2\pi j_2)^{2m}\}^{-2},$$

$$B = \sum_{j_1=-K}^{K} \sum_{j_2=-K}^{K} |a_{j_1,j_2}\{1 + \rho(2\pi j_1)^{2m}\}^{-1}\{1 + \rho(2\pi j_2)^{2m}\}^{-1} - a_{j_1,j_2}|^2.$$

It follows that

$$\begin{aligned}
|\tilde{a}_{j_1,j_2} - a_{j_1,j_2}|^2 &\leqslant \left(\sum_{(s_1,s_2)\neq(0,0)} |a_{j_1+s_1J,\, j_1+s_2J}| \right)^2 \\
&\leqslant \left\{ \sum_{(s_1,s_2)\neq(0,0)} (j_1 + s_1J)^{-2m}(j_2 + s_2J)^{-2m} \right\} \\
&\quad \times \left\{ \sum_{(s_1,s_2)\neq(0,0)} (j_1 + s_1J)^{2m}(j_2 + s_2J)^{2m}|a_{j_1,j_2}|^2 \right\} \\
&\leqslant \left[\left\{ \sum_{s_1\neq 0} (j_1 + s_1J)^{-2m} \right\} \left\{ \sum_{s_2=-\infty}^{\infty} (j_2 + s_2J)^{-2m} \right\} \right. \\
&\quad \left. + \left\{ \sum_{s_1=-\infty}^{\infty} (j_1 + s_1J)^{-2m} \right\} \left\{ \sum_{s_2\neq 0} (j_2 + s_2J)^{-2m} \right\} \right] \\
&\quad \times \int_0^1 \int_0^1 |R^{(m,m)}(s,t)|^2 \, ds \, dt \\
&= O(J^{-2m}). \qquad (26)
\end{aligned}$$

By (23) with $r = 2$ and (26),

$$A = O(J^{-2m}\rho^{-1/m}).$$

Next,

$$B \leqslant 3 \sum_{j_1=-K}^{K} \sum_{j_2=-K}^{K} \frac{\rho^2(2\pi j_1)^{4m}|a_{j_1,j_2}|^2 + \rho^2(2\pi j_2)^{4m}|a_{j_1,j_2}|^2}{\{1+\rho(2\pi j_1)^{2m}\}^2\{1+\rho(2\pi j_2)^{2m}\}^2}$$
$$+3 \sum_{j_1=-K}^{K} \sum_{j_2=-K}^{K} \frac{\rho^4(2\pi j_1)^{4m}(2\pi j_2)^{4m}|a_{j_1,j_2}|^2}{\{1+\rho(2\pi j_1)^{2m}\}^2\{1+\rho(2\pi j_2)^{2m}\}^2}.$$

Clearly,

$$\frac{\rho^2(2\pi j_1)^{4m} + \rho^2(2\pi j_2)^{4m}}{\{1+\rho(2\pi j_1)^{2m}\}^2\{1+\rho(2\pi j_2)^{2m}\}^2} \leqslant \rho(2\pi j_1)^{2m} + \rho(2\pi j_2)^{2m},$$

and

$$\frac{\rho^4(2\pi j_1)^{4m}(2\pi j_2)^{4m}}{\{1+\rho(2\pi j_1)^{2m}\}^2\{1+\rho(2\pi j_2)^{2m}\}^2} \leqslant \rho^2(2\pi j_1)^{2m}(2\pi j_2)^{2m}.$$

We thus have

$$B \leqslant 3 \sum_{j_1=-K}^{K} \sum_{j_2=-K}^{K} \{\rho(2\pi j_1)^{2m} + \rho(2\pi j_2)^{2m} + \rho^2(2\pi j_1)^{2m}(2\pi j_2)^{2m}\}|a_{j_1,j_2}|^2$$
$$= O(\rho).$$

Combining the various computations, using the fact that if $m \geqslant 2$ and $J^{-1}\rho^{-1/(2m)} \to 0$ then $J^{-2m}\rho^{-1/m} = o(J^{-2}\rho^{-1/(2m)})$, we conclude

$$\|T_\rho - T\|_{\mathcal{H}}^2 = O(\rho + J^{-2}\rho^{-1/(2m)}) \quad \text{if } X \in W_{2,\text{per}}^m \text{ and } E\|X^{(m)}\|_{L^2}^4 < \infty. \tag{27}$$

Now if $X \in W_{2,\text{per}}^{2m}$ and $E\|X^{(2m)}\|_{L^2}^2 < \infty$, the same approach shows that $B = O(\rho^2)$. Thus,

$$\|T_\rho - T\|_{\mathcal{H}}^2 = O(\rho^2 + J^{-2}\rho^{-1/(2m)}) \text{ if } X \in W_{2,\text{per}}^{2m} \text{ and } E\|X^{(2m)}\|_{L^2}^4 < \infty. \tag{28}$$

Next,

$$E(\|\tilde{T}_\rho - T_\rho\|_{\mathcal{H}}^2) = E\int_0^1\int_0^1 [\tilde{R}_\rho(s,t) - R_\rho(s,t)]^2\, ds\, dt$$
$$= n^{-1}\int_0^1\int_0^1 \mathrm{Var}\{\tilde{X}_\rho(s)\tilde{X}_\rho(t)\}\, ds\, dt \leqslant n^{-1}E(\|\tilde{X}_\rho\|_{L^2}^4).$$

If follows that

$$E(\|\tilde{X}_\rho\|_{L^2}^4) \leqslant 8E(\|X\|_{L^2}^4) + 8E(\|\tilde{X}_\rho - X\|_{L^2}^4),$$

where $E(\|X\|_{L^2}^4) < \infty$ by assumption. By calculations in Theorem 2, we have

$$\|\tilde{X}_\rho - X\|_{L^2}^2 \leqslant \{\rho + O(J^{-2m})\}\|X^{(m)}\|_{L^2}^2 + 2\sum_{j=1}^{J} \frac{\tilde{\varsigma}_j^2}{(1+\rho\pi^{2m}j^{2m})^2}.$$

Some tedious but straightforward calculations show that $E(\tilde{\varsigma}_j^2\tilde{\varsigma}_k^2) = O(J^{-2})$, and we obtain

$$E(\|\tilde{X}_\rho - X\|_{L^2}^4) = O(\rho^2) + O(J^{-2}\rho^{-1/m}).$$

We have thus shown

$$E(\|\tilde{T}_\rho - T_\rho\|^2_{\mathcal{H}}) = O(n^{-1}). \tag{29}$$

The results in theorem follow from (27)–(29).

4.5. *Proof of Theorem 5*

Define bilinear forms

$$L(g) = \langle g, T_\rho g\rangle_{L^2} + \langle g^{(m)}, g^{(m)}\rangle_{L^2} \quad \text{and} \quad \tilde{L}(g) = \langle g, \tilde{T}_\rho g\rangle_{L^2} + \langle g^{(m)}, g^{(m)}\rangle_{L^2}.$$

Lemma 8. *Assume that the conditions of Theorem* 4 *hold. Let* n, J, ρ *be such that* $n^{-1} + \rho + J^{-1}\rho^{-1/(2m)} \to 0$. *The following can be shown*:

(i) $\liminf_{J,\rho} \inf_{g=\phi^T b, \|g\|_{L^2}=1} L(g) > 0$, *and*

(ii) $\lim_{n,J,\rho} P(\inf_{g=\phi^T b, \|g\|_{L^2}=1} \tilde{L}(g) > c) = 1$ *for some constant* $c > 0$.

Proof. For convenience let n, J, ρ be indexed by k and $n_k^{-1} + \rho_k + J_k^{-1}\rho_k^{-1/(2m)} \to 0$ as $k \to \infty$.

(i) Notice that the null space of $\langle g^{(m)}, g^{(m)}\rangle_{L^2}$ is spanned by ϕ_0. On the other hand, by our assumption, $\langle \phi_0, T\phi_0\rangle_{L^2} = c_0 > 0$. Fix $0 < \varepsilon < c_0$ and pick a large enough k so that we have $\|T_\rho - T\| < \varepsilon$ and $\langle \phi_0, T_\rho\phi_0\rangle_{L^2} \geqslant c_0 - \varepsilon$ by Theorem 4. Note that in this proof, the operator norm can be the usual sup-norm, which is dominated by the Hilbert–Schmidt norm. For any $g = \phi^T b$ with $b^T\bar{b} = 1$, let $h = g - b_0\phi_0$, then $\|h\|^2 = 1 - |b_0|^2$. Thus,

$$\begin{aligned} L(g) &\geqslant c_1|b_0|^2 - 2|b_0||\langle \phi_0, T_\rho h\rangle_{L^2}| + \langle h^{(m)}, h^{(m)}\rangle_{L^2} \\ &\geqslant c_1|b_0|^2 - 2c_2\|h\| + c_3\|h\|^2 \quad \text{for all large } k; \end{aligned}$$

on the other hand, $L(g) \geqslant \langle g^{(m)}, g^{(m)}\rangle_{L^2} = \langle h^{(m)}, h^{(m)}\rangle_{L^2} \geqslant c_3\|h\|^2$, therefore

$$L(g) \geqslant (c_1|b_0|^2 - 2c_2\|h\|)_+ + c_3\|h\|^2 \quad \text{for all large } k.$$

Note that the minimum of this lower bound does not depend on J or ρ.

(ii) Let c be the lim inf in part (i). For any $0 < \varepsilon < c$,

$$\begin{aligned} P\left(\inf_{g=\phi^T b, b^T\bar{b}=1} \tilde{L}(g) \geqslant c - \varepsilon\right) &= P\left(\inf_{g=\phi^T b, b^T\bar{b}=1} \tilde{L}(g) \geqslant c - \varepsilon, \|\tilde{T}_\rho - T_\rho\| \leqslant \varepsilon\right) \\ &\quad + P\left(\inf_{g=\phi^T b, b^T\bar{b}=1} \tilde{L}(g) > c - \varepsilon, \|\tilde{T}_\rho - T_\rho\| > \varepsilon\right). \end{aligned}$$

When k is large enough, the first term is equal to $P(\|\tilde{T}_\rho - T_\rho\| \leqslant \varepsilon)$ which tends to 1 as $k \to \infty$, and the second term tends to 0. $\square$

Proof of Theorem 5. We will start by showing that $E(\|\hat{f}_{\lambda,\rho}\|^2|\mathbf{Z}) = O_p(1)$. First,

$$\|\check{f}_{\lambda,\rho}\|^2_{L^2} = \|(\check{\Omega}_\rho^T + \lambda W)^{-1}\check{\Omega}_\rho^T\check{\beta}\|^2 \leqslant \|\check{\beta}\|^2 = \|\check{f}\|^2_{L^2} < \|f\|^2_{L^2}.$$

Second, let $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ be the functions that return the smallest and largest eigenvalues of a matrix. For any function $g(t) = \phi^T(t)b$, we have $\tilde{L}(g) = b^T(\check{\Omega}_\rho + W)\bar{b}$, then by Lemma 8,

$\lambda_{\min}(\check{\Omega}_\rho + W) = O_p(1)$. Hence, $\lambda_{\min}(\check{\Omega}_\rho + \lambda W) = O_p(\lambda)$, and $\lambda_{\max}\{(\check{\Omega}_\rho + \lambda W)^{-1}\} = O_p(\lambda^{-1})$. Also, as stated before, the eigenvalues of $\check{\Omega}_\rho$ are the same as those of $\tilde{T}_\rho$, and hence $\check{\Omega}_\rho$ is positive semi-definite. It then follows that

$$\begin{aligned} E(\|h_\lambda\|^2_{L^2}|\mathbf{Z}) &= E\{\boldsymbol{\varepsilon} n^{-1}\mathbf{Z}V(\check{\Omega}_\rho + \lambda W)^{-2} n^{-1}\bar{V}^T\mathbf{Z}^T\boldsymbol{\varepsilon}|\mathbf{Z}\} \\ &= n^{-1}\sigma^2_\varepsilon \operatorname{tr}\{(\check{\Omega}_\rho + \lambda W)^{-2}\check{\Omega}_\rho\} \\ &\leqslant n^{-1}\sigma^2_\varepsilon \operatorname{tr}\{(\check{\Omega}_\rho + \lambda W)^{-1}\} \\ &\leqslant n^{-1}\sigma^2_\varepsilon \left[\lambda_{\max}\{(\check{\Omega}_\rho + \lambda W)^{-1}\} + 2\lambda^{-1}\sum_{j=1}^{K}(2\pi j)^{-2m}\right] \\ &= O_p(n^{-1}\lambda^{-1}). \end{aligned}$$

Thirdly, by the assumption that $(\rho + J^{-1}\rho^{-1/(2m)})/\lambda$ is bounded,

$$\begin{aligned} E(\|g_\lambda\|^2_{L^2}|\mathbf{Z}) &= E\{\mathbf{r} n^{-1}\mathbf{Z}V(\check{\Omega}_\rho + \lambda W)^{-2} n^{-1}\bar{V}^T\mathbf{Z}^T\mathbf{r}|\mathbf{Z}\} \\ &\leqslant n^{-1}\lambda_{\max}\{(\check{\Omega}_\rho + \lambda W)^{-1}\}E\{\mathbf{r} n^{-1}\mathbf{Z}V(\check{\Omega}_\rho + \lambda W)^{-1}\bar{V}\mathbf{Z}^T\mathbf{r}|\mathbf{Z}\} \\ &\leqslant n^{-1}\lambda_{\max}\{(\check{\Omega}_\rho + \lambda W)^{-1}\}E(\mathbf{r}^T\mathbf{r}|\mathbf{Z}) \\ &= O_p(\lambda^{-1})\{O_p(\rho) + O_p(J^{-1}\rho^{-1/(2m)})\} = O_p(1). \end{aligned}$$

Now, by (16), we have $E(\|\hat{f}_{\lambda,\rho}\|^2_{L^2}|\mathbf{Z}) = O_p(1)$, and therefore

$$E(\|\hat{f}_{\lambda,\rho} - f\|^2_{L^2}|\mathbf{Z}) = O_p(1).$$

Finally, notice that

$$\begin{aligned} E(\|\hat{f}_{\lambda,\rho} - f\|^2_T|\mathbf{Z}) &= E(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\mathbf{Z}) + E\{\langle\hat{f}_{\lambda,\rho} - f, (T - \tilde{T}_\rho)(\hat{f}_{\lambda,\rho} - f)\rangle_{L^2}|\mathbf{Z}\} \\ &\leqslant E(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\mathbf{Z}) + \|T - \tilde{T}_\rho\|_{\mathcal{H}}E(\|\hat{f}_{\lambda,\rho} - f\|^2_{L^2}|\mathbf{Z}). \end{aligned}$$

The result follows from Theorems 3 and 4. $\square$

Acknowledgment

We thank the referee for the constructive comments.

References

[1] R.B. Ash, M.F. Gardner, Topics in stochastic processes, Academic Press, New York, 1975.
[2] B.A. Brumback, J.A. Rice, Smoothing spline models for the analysis of nested and crossed samples of curves, J. Amer. Statist. Assoc. 93 (443) (1998) 961–976.
[3] T. Cai, P. Hall, Prediction in functional linear regression, Ann. Statist. 34 (2006) to appear.
[4] H. Cardot, F. Ferraty, P. Sarda, Spline estimators for the functional linear model, Statist. Sinica 13 (2003) 571–591.
[5] H. Cardot, P. Sarda, Estimation in generalized linear models for functional data via penalized likelihood, J. Multivariate Anal. 92 (2005) 24–41.
[6] J.B. Conway, A Course in Functional Analysis, Springer, Berlin, 1990.
[7] P. Craven, G. Wahba, Smoothing noisy data with spline functions, estimating the correct degree of smoothing by the method of generalized cross-validation, Numer. Math. 31 (1979) 377–403.
[8] J. Dauxois, A. Pousse, Y. Romain, Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference, J. Multivariate Anal. 12 (1982) 136–154.
[9] R.L. Eubank, Spline Smoothing and Nonparametric Regression, Marcel Dekker, New York, 1988.
[10] W. Guo, Functional mixed effects models, Biometrics 58 (2002) 121–128.

[11] P. Hall, J.L. Horowitz, Methodology and convergence rates for functional linear regression, 2004, Manuscript.
[12] G.A. James, T.J. Hastie, C.A. Sugar, Principal component models for sparse functional data, Biometrika 87 (2000) 587–602.
[13] J.K. Merikoski, R. Kumar, Inequalities for spreads of matrix sums and products, Appl. Math. E-Notes 4 (2004) 150–159.
[14] H.G. Müller, U. Stadtmüller, Generalized functional linear models, Ann. Statist. 33 (2005) 774–805.
[15] J.O. Ramsay, B.W. Silverman, Applied Functional Data Analysis: Methods and Case Studies, Springer, New York, 2002.
[16] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2005.
[17] J. Rice, M. Rosenblatt, Integrated mean square error of a smoothing spline, J. Approx. Theory 33 (1981) 353–369.
[18] D. Ruppert, M.P. Wand, R.J. Carroll, Semiparametric regression, Cambridge University Press, Cambridge, UK, 2003.
[19] A. Tykohonov, V. Arsenin, Solution to Ill-Posed Problems, Wiley, New York, 1977.
[20] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, Springer, New York, 1997.