SFB 823

Discussion Paper

Statistical inference for function-on-function linear regression

Holger Dette, Jiajun Tang

Nr. 20/2021

Statistical inference for function-on-function linear regression

Holger Dette, Jiajun Tang
Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany

Abstract: We propose a reproducing kernel Hilbert space approach to estimate the slope in a function-on-function linear regression via penalised least squares, regularized by the thin-plate spline smoothness penalty. In contrast to most of the work on functional linear regression, our main focus is on statistical inference with respect to the sup-norm. This point of view is motivated by the fact that slope (surfaces) with rather different shapes may still be identified as similar when the difference is measured by an L^2-type norm. However, in applications it is often desirable to use metrics reflecting the visualization of the objects in the statistical analysis.

We prove the weak convergence of the slope surface estimator as a process in the space of all continuous functions. This allows us the construction of simultaneous confidence regions for the slope surface and simultaneous prediction bands. As a further consequence, we derive new tests for the hypothesis that the maximum deviation between the "true" slope surface and a given surface is less or equal than a given threshold. In other words: we are not trying to test for exact equality (because in many applications this hypothesis is hard to justify), but rather for pre-specified deviations under the null hypothesis. To ensure practicability, non-standard bootstrap procedures are developed addressing particular features that arise in these testing problems.

As a by-product, we also derive several new results and statistical inference tools for the function-on-function linear regression model, such as minimax optimal convergence rates and likelihood-ratio tests. We also demonstrate that the new methods have good finite sample properties by means of a simulation study and illustrate their practicability by analyzing a data example.

Keywords: function-on-function linear regression, minimax optimality, simultaneous confidence regions, relevant hypotheses, bootstrap, reproducing kernel Hilbert space, maximum deviation

AMS Subject Classification: 62R10, 62F03, 62F25, 46E22

1 Introduction

Over the past decades, new measurement technologies provide enormous amounts of data with complex structure. A popular and extremely successful approach to model high-dimensional data on a dense grid exhibiting a certain degree of smoothness is functional data analysis (FDA), which considers the observations as discretized functions. Meanwhile, numerous practical and theoretical aspects in FDA have been discussed (see, for example, the monographs Bosq, 2000; Ramsay and Silverman, 2005; Ferraty and Vieu, 2010; Horváth and Kokoszka, 2012; Hsing and Eubank, 2015, among others). A large portion of the literature uses dimension reduction techniques such as (functional) principal components. On the other hand, as argued in Aue et al. (2018), there are numerous applications, where it is reasonable to assume that the functions are at least continuous, and in such cases dimension reduction techniques can incur a loss of information and fully functional methods can prove advantageous.

Because of its simplicity and good interpretability, the *scalar-on-function regression model*

$$Y_i = \alpha_0 + \int_0^1 \beta_0(s)\, X_i(s)\, ds + \varepsilon_i\,, \quad 1 \le i \le n\,. \tag{1.1}$$

has found considerable attention (see, for exemple, James, 2002; Cardot et al., 2003; Müller and Stadtmüller, 2005; Yao et al., 2005; Hall and Horowitz, 2007; Yuan and Cai, 2010, among manny others). Here, the Y_i and the (centred) errors ε_i are scalar variables, the predictors X_i are functions (typically of time or location) defined on the interval $[0, 1]$, and the scalar α_0 and the function β_0 are the unknown parameters to be estimated. On the other hand, there also exist many applications, where both, the predictor and the response, are functions, and in recent years the *function-on-function regression model*

$$Y_i(t) = \alpha_0(t) + \int_0^1 \beta_0(s,t)\, X_i(s)\, ds + \varepsilon_i(t)\,, \quad t \in [0,1]\,,\ 1 \le i \le n\,, \tag{1.2}$$

has gained increasing attention (see Lian, 2007, 2015; Scheipl and Greven, 2016; Benatia et al., 2017; Luo and Qi, 2017; Sun et al., 2018). Here α_0, Y_i, X_i, ε_i are functions defined on the interval $[0, 1]$ and the slope parameter β_0 is a function defined on the square $[0, 1]^2$, which we call slope surface throughout this paper in order to distinguish it from the slope function in model (1.1).

The slope β_0 quantifies the strength of the dependence between the predictor and the response, and is the main object of statistical inference in this context. Many methods, such as estimation, testing, confidence regions, have been developed in the last decades for the scalar-on-function linear regression model (1.1), which are often based on the L^2 metric (see, for example, Hall and Horowitz, 2007; Horváth and Kokoszka, 2012, among many others). A popular estimation tool is functional principle component (FPC) analysis, which provides a series representation of the function β_0 in the corresponding L^2 space (see, for example, Yao et al., 2005). Other authors proposed reproducing kernel Hilbert space (RKHS) approaches to estimate the slope parameter in a functional linear regression model. For example, Yuan and Cai (2010) used the RKHS framework to construct a minimax optimal estimate in the scalar-on-function linear regression, and Cai and Yuan (2012) discussed minimax properties of their RKHS estimator in terms of prediction accuracy. We also refer to the work of Meister (2011) who showed the asymptotic equivalence of the scalar-on-function linear regression and the Gaussian white noise model in the Le Cam's sense. Besides estimation, the problem of testing the hypotheses

$$H_0 : \beta_0 = \beta_* \quad \text{versus} \quad H_1 : \beta_0 \neq \beta_* \ , \tag{1.3}$$

for a prespecified function β_* in the scalar-functional linear regression model has been discussed intensively (see Cardot et al., 2003, 2004; Hilgert et al., 2013; Lei, 2014; Kong et al., 2016; Qu and Wang, 2017, among others). There also exist several proposals to construct L^2-based confidence regions (see Müller and Stadtmüller, 2005; Imaizumi and Kato, 2019, among others).

Non-linear and semiparametric scalar-on-function regression models, such as generalized linear models and the Cox model, have been studied by Shang and Cheng (2015), Li and Zhu (2020) and Hao et al. (2021). For the function-on-function model (1.2), the literature is more scarce. Lian (2015) studied the minimax prediction rate in an RKHS, where regularization of the estimator is only performed in one argument, while Scheipl and Greven (2016) investigated a penalized B-spline approach. Benatia et al. (2017) used Tikhonov regularization, Luo and Qi (2017) proposed a so-called signal compression approach and Sun et al. (2018) considered a tensor product RKHS approach to estimate the slope surface and the achieved the minimax prediction risk.

This list of references is by no means complete, but a common feature of most of the work in this context consists in the fact that statistical methodology is developed in a Hilbert space framework (often the space or a subspace of the square-integrable functions on an interval), which means that the statistical properties of estimators, tests and confidence regions for the slope parameter are usually described in terms of a norm corresponding to a Hilbert space. While this is convenient from a theoretical point of view and also reflects the mathematical structure of the (integral) operator of the functional linear model, it has some drawbacks from a practical perspective. In applications, using a metric that reflects the visualization of the curve/surface is usually more desirable, since functions/surfaces with a small difference with respect to an L^2-type distance can differ significantly in terms of maximum deviation. For example, a confidence region of the slope function/surface based on an L^2-type distance is often hard to visualize and does not give much information about the shape of the curve or surface.

The choice of the metric also matters if one takes a more careful look at the formulation of the hypotheses in (1.3). We argue that, in many regression problems, it is very unlikely that the unknown slope β_0 coincides with a pre-specified function/surface β_* on its complete domain, and as a consequence, testing the null hypothesis in (1.3) might be questionable in such cases. Usually, hypotheses of the form (1.3) are formulated with the intention to investigate the question whether the effect of the predictor on the response can be approximately described by the function/surface β_*, such that the difference $\beta_0 - \beta_*$ is in some sense "small". This question can be better answered by testing the hypotheses of a *relevant difference*

$$H_0 : \|\beta_0 - \beta_*\| \leq \Delta \quad \text{versus} \quad H_1 : \|\beta_0 - \beta_*\| > \Delta\,, \tag{1.4}$$

where $\|\cdot\|$ denotes a norm and $\Delta > 0$ defines a threshold. Hypotheses of this type have recently found some interest in functional data analysis (see, for example, Fogarty and Small, 2014; Dette et al., 2020), and here the choice of the norm matters, as different norms define different hypotheses. One may also view the choice of the threshold Δ as a particular perspective of a bias-variance trade-off, which depends sensitively on the specific application, and, of course, also on the metric under consideration. In particular, we argue that the specification of the threshold in (1.4) is more accessible for a norm which reflects the visualization, such as the sup-norm.

In the present paper, we address these issues and provide new statistical methodology for the function-on-function linear regression model (1.2) if inference is based on the maximum deviation. We propose an estimator for the slope surface β_0 minimizing an integrated squared error loss with a thin-plate spline smoothness penalty functional, and prove its minimax optimality using an RKHS framework. Based on a Bahadur representation, we establish the weak convergence of this estimator as a process in the Banach space $C([0,1]^2)$ with a Gaussian limiting process. As the covariance structure of this process is not easily accessible, we develop a multiplier bootstrap to obtain quantiles for the distribution of functionals of the limiting process. In contrast to the L^2-metric based methods, this enables us to construct simultaneous asymptotic $(1-\alpha)$-confidence regions for the slope surface β_0 in model (1.2). Moreover, we also provide an efficient solution to the problem of testing for a relevant deviation from a given function β_* with respect to the sup-norm. Here, we combine the developed bootstrap methodology with estimates of the *extremal set* of the function $\beta_0 - \beta_*$, and develop an asymptotic level α-test for the relevant hypotheses in (1.4), where the norm is given by the sup-norm. Although we mainly concentrate on the model (1.2), it is worth mentioning that, as a special case, our approach provides also new methods for the scalar-on-function linear regression model (1.1), which allows inference with respect to the sup-norm.

The rest of this article is organized as follows. In Section 2, we propose our RKHS methodology of function-on-function linear regression and study the asymptotic properties of our estimator in Section 3. Section 4 discusses several statistical applications of our results and the finite sample properties of the proposed methodology are illustrated in Section 5. Finally, the technical details and proofs of our theoretical results are given in the online supplementary material.

2 Function-on-function linear regression

Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent identically distributed random variables defined by the function-on-function regression model in (1.2), where ε_i is the centred random noise, and the slope surface β_0 is defined on $[0,1]^2$. For the sake of brevity, throughout this article, we assume that the observed curves, i.e., X_i and Y_i in (1.2), are centred, that is, $\mathrm{E}\{X(s)\} = \mathrm{E}\{Y(t)\} = 0$, for any $(s,t) \in [0,1]^2$, so that we may ignore the intercept function α_0, since $\alpha_0(t) = \mathrm{E}\{Y(t)\} - \int_0^1 \beta_0(s,t)\,\mathrm{E}\{X(s)\}ds$. In this case, the function-on-function linear regression model in (1.2) becomes

$$Y_i(t) = \int_0^1 \beta_0(s,t)\, X_i(s)\, ds + \varepsilon_i(t)\,, \quad 1 \le i \le n\,, \tag{2.1}$$

and a similar relation can be derived for the model (1.1).

In the sequel, we use $L^2([0,1])$ and $L^2([0,1]^2)$ to denote the space of square-integrable functions on $[0,1]$ and $[0,1]^2$, respectively, and the corresponding inner product is denoted by

$\langle \cdot, \cdot \rangle_{L^2}$. By $C([0,1]^2)$ we denote the Banach space of continuous functions on $[0,1]^2$ equipped with the supremum norm $\| \cdot \|_\infty$, by "$\rightsquigarrow$" we denote weak convergence in $C([0,1])$ and $C([0,1]^2)$, and "$\xrightarrow{d}$" stays for convergence in distribution in $\mathbb{R}^k$ (for some positive integer k).

We start by proposing a RKHS approach for estimating the slope surface β_0 in model (2.1), and define by

$$\mathcal{H} = \Big\{ \beta : [0,1]^2 \to \mathbb{R} \,\Big|\, \tfrac{\partial^{\theta_1+\theta_2}\beta}{\partial s^{\theta_1}\partial t^{\theta_2}} \text{ is absolutely continuous, for } 0 \le \theta_1+\theta_2 \le m-1\,; \\ \tfrac{\partial^{\theta_1+\theta_2}\beta}{\partial s^{\theta_1}\partial t^{\theta_2}} \in L^2([0,1]^2), \text{ for } \theta_1+\theta_2 = m \Big\} \tag{2.2}$$

the Sobolev space of order $m > 1$ on $[0,1]^2$. It is known (see, for example, Wahba, 1990) that $\mathcal{H}$ in (2.2) is a Hilbert space equipped with the Sobolev norm defined by

$$\|\beta\|_{\mathcal{H}}^2 = \sum_{0 \le \theta_1+\theta_2 \le m-1} \binom{\theta_1+\theta_2}{\theta_1} \left(\int \frac{\partial^{\theta_1+\theta_2}\beta}{\partial s^{\theta_1}\partial t^{\theta_2}} \right)^2 + \sum_{\theta_1+\theta_2=m} \binom{m}{\theta_1} \int \left(\frac{\partial^m \beta}{\partial s^{\theta_1}\partial t^{\theta_2}} \right)^2 . \tag{2.3}$$

We propose to estimate β_0 in model (2.1) by

$$\widehat{\beta}_n = \arg\min_{\beta \in \mathcal{H}} \left\{ L_n(\beta) + (\lambda/2) J(\beta, \beta) \right\}, \tag{2.4}$$

where

$$L_n(\beta) = \frac{1}{2n} \sum_{i=1}^n \int_0^1 \left\{ Y_i(t) - \int_0^1 \beta(s,t) X_i(s)\, ds \right\}^2 dt \tag{2.5}$$

is the integrated squared loss functional, $\lambda > 0$ is a regularization parameter, and for $m > 1$,

$$J(\beta_1, \beta_2) = \sum_{\theta=0}^m \binom{m}{\theta} \int_0^1 \int_0^1 \frac{\partial^m \beta_1}{\partial s^\theta\, \partial t^{m-\theta}} \times \frac{\partial^m \beta_2}{\partial s^\theta\, \partial t^{m-\theta}}\, ds\, dt \tag{2.6}$$

is the thin-plate spline smoothness penalty functional (see, for example, Wood, 2003).

In (2.4), for notational brevity, we suppress the dependence of $\widehat{\beta}_n$ on λ, and denote by

$$L_{n,\lambda}(\beta) = L_n(\beta) + (\lambda/2) J(\beta, \beta) \tag{2.7}$$

the objective function in (2.4). For $\beta_1, \beta_2 \in \mathcal{H}$, we consider the following map $\langle \cdot, \cdot \rangle_K : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ defined by

$$\langle \beta_1, \beta_2 \rangle_K = V(\beta_1, \beta_2) + \lambda J(\beta_1, \beta_2), \tag{2.8}$$

where

$$V(\beta_1, \beta_2) = \int_{[0,1]^3} C_X(s_1, s_2)\, \beta_1(s_1, t)\, \beta_2(s_2, t)\, ds_1\, ds_2\, dt \tag{2.9}$$

and

$$C_X(s_1, s_2) = \operatorname{cov}\{X(s_1), X(s_2)\} \tag{2.10}$$

denotes the the covariance function of the predictor. We first make the following mild assumption on C_X.

Assumption A1. C_X is continuous on $[0,1]^2$. For any $\gamma \in L^2([0,1])$, $\int_0^1 C_X(s, s')\gamma(s)ds = 0$ implies that $\gamma \equiv 0$.

Our first result, which is proved in Section A.1, shows that the relation $\langle \cdot, \cdot \rangle_K$ in (2.8) defines an inner product on the space $\mathcal{H}$ in (2.2), and its corresponding norm is equivalent to the Sobolev norm $\| \cdot \|_{\mathcal{H}}$ given in (2.3).

Proposition 2.1. *If Assumption A1 holds, then $\langle \cdot, \cdot \rangle_K$ in (2.8) is a well-defined inner product on $\mathcal{H}$. If $\|\cdot\|_K$ denotes its corresponding norm, then $\|\cdot\|_K$ and $\|\cdot\|_{\mathcal{H}}$ in (2.3) are equivalent. Moreover, $\mathcal{H}$ is a reproducing kernel Hilbert space equipped with the inner product $\langle \cdot, \cdot \rangle_K$.*

In the sequel, for $s_1, s_2, t_1, t_2 \in [0,1]$, let $K\{(s_1, t_1), (s_2, t_2)\}$ denote the reproducing kernel of the reproducing kernel Hilbert space $\mathcal{H}$ equipped with inner product $\langle \cdot, \cdot \rangle_K$. For functions x, y on $[0,1]$, let $x \otimes y$ denote the function defined by $x \otimes y(s,t) = x(s)y(t)$. We use $\sum_{k,\ell}$ to denote the sum $\sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty}$ for abbreviation. Let $\mathcal{C}_X$ denote the covariance operator of X defined by

$$\mathcal{C}_X(x) = \int_0^1 C_X(s, \cdot)\, x(s)\, ds\,, \tag{2.11}$$

for $x \in L^2([0,1])$. We assume that there exists a sequence of functions in $\mathcal{H}$ that diagonalizes operators V in (2.9) and J in (2.6) simultaneously. A concrete example that satisfies the following assumption will be provided in Section B.2 of the online supplement.

Assumption A2 (Simultaneous diagonalization)**.** There exists a sequence of functions $\varphi_{k\ell} = x_{k\ell} \otimes \eta_\ell \in \mathcal{H}$, such that $\|\varphi_{k\ell}\|_\infty \leq c(k\ell)^a$ for any $k, \ell \geq 1$, and

$$V(\varphi_{k\ell}, \varphi_{k'\ell'}) = \delta_{kk'}\, \delta_{\ell\ell'}\,, \qquad J(\varphi_{k\ell}, \varphi_{k'\ell'}) = \rho_{k\ell}\, \delta_{kk'}\, \delta_{\ell\ell'}\,, \tag{2.12}$$

where $a \geq 0$, $c > 0$ are constants, $\delta_{kk'}$ is the Kronecker delta and $\rho_{k\ell}$ are constants, such that $\rho_{k\ell} \asymp (k\ell)^{2D}$ for some constant $D > a + 1/2$. Furthermore, any $\beta \in \mathcal{H}$ admits the expansion

$$\beta = \sum_{k,\ell} V(\beta, \varphi_{k\ell})\varphi_{k\ell}$$

with convergence in $\mathcal{H}$ with respect to the norm $\|\cdot\|_K$.

Note that a similar diagonalization assumption has been made in Shang and Cheng (2015) in the context of generalized scalar-on-function model. For the inner product $\langle\cdot,\cdot\rangle_K$ in (2.8) it follows from Assumption A2 that

$$\langle \varphi_{k\ell}, \varphi_{k'\ell'} \rangle_K = V(\varphi_{k\ell}, \varphi_{k'\ell'}) + \lambda J(\varphi_{k\ell}, \varphi_{k'\ell'}) = (1 + \lambda\rho_{k\ell})\, \delta_{kk'}\, \delta_{\ell\ell'} \quad (k, k', \ell, \ell' \geq 1).$$

Therefore, it follows $\langle \beta, \varphi_{k'\ell'} \rangle_K = \sum_{k,\ell} V(\beta, \varphi_{k\ell})\langle \varphi_{k\ell}, \varphi_{k'\ell'} \rangle_K = (1 + \lambda\rho_{k\ell})V(\beta, \varphi_{k\ell})$ for any $\beta \in \mathcal{H}$, so that

$$\beta(s,t) = \sum_{k,\ell} V(\beta, \varphi_{k\ell})\varphi_{k\ell}(s,t) = \sum_{k,\ell} \frac{\langle \beta, \varphi_{k\ell} \rangle_K}{1 + \lambda\rho_{k\ell}} \varphi_{k\ell}(s,t)\,. \tag{2.13}$$

Recall that K is the reproducing kernel and using the notation $K_{(s,t)} = K\{(s,t), \cdot\}$ we have $\varphi_{k\ell}(s,t) = \langle K_{(s,t)}, \varphi_{k\ell} \rangle_K$, so that by (2.13),

$$K_{(s,t)} = \sum_{k,\ell} \frac{\varphi_{k\ell}(s,t)}{1 + \lambda\rho_{k\ell}} \varphi_{k\ell}\,; \qquad K\{(s_1,t_1),(s_2,t_2)\} = \sum_{k,\ell} \frac{\varphi_{k\ell}(s_1,t_1)\varphi_{k\ell}(s_2,t_2)}{1 + \lambda\rho_{k\ell}}\,. \tag{2.14}$$

For $\beta_1, \beta_2 \in \mathcal{H}$, let W_λ denote a linear self-adjoint operator such that $\langle W_\lambda \beta_1, \beta_2 \rangle_K = \lambda J(\beta_1, \beta_2)$. By definition, for the $\{\varphi_{k\ell}\}_{k,\ell \geq 1}$ in Assumption A2, we have $\langle W_\lambda \varphi_{k\ell}, \varphi_{k'\ell'} \rangle_K = \lambda J(\varphi_{k\ell}, \varphi_{k'\ell'}) = \lambda\rho_{k\ell}\, \delta_{kk'}\, \delta_{\ell\ell'}$, so that in view of (2.13),

$$W_\lambda \varphi_{k\ell} = \sum_{k',\ell'} \frac{\langle W_\lambda \varphi_{k\ell}, \varphi_{k'\ell'} \rangle_K}{1 + \lambda\rho_{k'\ell'}} \varphi_{k'\ell'} = \frac{\lambda\, \rho_{k\ell}\, \varphi_{k\ell}}{1 + \lambda\rho_{k\ell}}\,. \tag{2.15}$$

For any $z \in L^2([0,1]^2)$ and $\beta \in \mathcal{H}$, $\mathfrak{S}_z(\beta) = \int_0^1 \int_0^1 \beta(s,t) z(s,t) ds dt$ is a bounded linear functional. By the Riesz representation theorem, there exists a unique element $\tau(z) \in \mathcal{H}$ such that

$$\langle \tau(z), \beta \rangle_K = \mathfrak{S}_z(\beta) = \int_0^1 \int_0^1 \beta(s,t)\, z(s,t)\, ds\, dt\,. \tag{2.16}$$

In particular, $\langle \tau(z), \varphi_{k\ell} \rangle_K = \langle z, \varphi_{k\ell} \rangle_{L^2}$, so that

$$\tau(z) = \sum_{k,\ell} \frac{\langle z, \varphi_{k\ell} \rangle_{L^2}}{1 + \lambda\rho_{k\ell}} \varphi_{k\ell}\,. \tag{2.17}$$

3 Asymptotic properties

In order to develop statistical methodology for inference on the slope surface in the function-on-function linear regression model (1.2), we study in this section the asymptotic properties of the estimator $\widehat{\beta}_n$ defined by (2.4). We first present a Bahadur representation, which is used to prove weak convergence of the estimator (point-wise and as process in $C([0,1]^2)$). Several statistical applications of the following results will be given in Section 4 below.

We begin introducing several useful quantities. Recalling the notation of W_λ and τ defined in (2.15) and (2.17), respectively, we obtain by direct calculations the first and second order Fréchet derivatives of the integrated squared error L_n in (2.5)

$$\begin{aligned} \mathcal{D}L_n(\beta)\beta_1 &= -\frac{1}{n}\sum_{i=1}^n \int_0^1\int_0^1 \Big\{Y_i(t) - \int_0^1 \beta(s_1,t)\,X_i(s_1)ds_1\Big\} X_i(s_2)\,\beta_1(s_2,t)\,ds_2\,dt \\ &= -\frac{1}{n}\sum_{i=1}^n \Big\langle \tau\Big[X_i \otimes \Big\{Y_i - \int_0^1 \beta(s,\cdot)\,X_i(s)\,ds\Big\}\Big], \beta_1\Big\rangle_K ; \\ \mathcal{D}^2 L_n(\beta)\beta_1\beta_2 &= \frac{1}{n}\sum_{i=1}^n \int_0^1\int_0^1 \Big\{\int_0^1 \beta_1(s_1,t)\,X_i(s_1)\,ds_1\Big\} X_i(s_2)\,\beta_2(s_2,t)\,ds_2\,dt \\ &= \frac{1}{n}\sum_{i=1}^n \Big\langle \tau\Big[X_i \otimes \Big\{\int_0^1 \beta_1(s,\cdot)\,X_i(s)\,ds\Big\}\Big], \beta_2\Big\rangle_K . \end{aligned} \tag{3.1}$$

Therefore, it follows for the function $L_{n,\lambda}$ in (2.7) that

$$\begin{aligned} \mathcal{D}L_{n,\lambda}(\beta)\beta_1 &= \langle S_{n,\lambda}(\beta), \beta_1\rangle_K , \\ \mathcal{D}^2 L_{n,\lambda}(\beta)\beta_1\beta_2 &= \langle \mathcal{D}S_{n,\lambda}(\beta)\beta_1, \beta_2\rangle_K , \end{aligned}$$

where we use the notations

$$\begin{aligned} S_{n,\lambda}(\beta) &= -\frac{1}{n}\sum_{i=1}^n \tau\Big[X_i \otimes \Big\{Y_i - \int_0^1 \beta(s,\cdot)\,X_i(s)\,ds\Big\}\Big] + W_\lambda(\beta) , \\ \mathcal{D}S_{n,\lambda}(\beta)\beta_1 &= \frac{1}{n}\sum_{i=1}^n \tau\Big[X_i \otimes \Big\{\int_0^1 \beta_1(s,\cdot)\,X_i(s)\,ds\Big\}\Big] + W_\lambda(\beta_1) . \end{aligned} \tag{3.2}$$

In addition, we have, in view of (2.8),

$$\mathrm{E}\{\mathcal{D}^2 L_{n,\lambda}(\beta)\beta_1\beta_2\} = V(\beta_1,\beta_2) + \lambda J(\beta_1,\beta_2) = \langle \beta_1,\beta_2\rangle_K .$$

Remark 3.1. In the unregularized case, where $\lambda = 0$, one can show that setting $\mathrm{E}\{S_{n,0}(\beta)\} = 0$ yields the common FPC expansion

$$\beta = \sum_{k,\ell} \frac{\mathrm{E}(\langle X, u_k\rangle_{L^2}\langle Y, v_\ell\rangle_{L^2})}{\mathrm{E}(\langle X, u_k\rangle_{L^2}^2)}\, u_k \otimes v_\ell , \tag{3.3}$$

where the u_k's and the v_ℓ's are eigenfunctions of the covariance functions of X and Y, respectively (see, for example, Yao et al., 2005). To see this, in view of the definition of τ in (2.16), note that $\tau(0) = 0$ and that $\tau(z) = 0$ implies $z = 0$. Hence, $\mathrm{E}\{S_{n,0}(\beta)\} = \mathrm{E}\big(\tau\big[X \otimes \big\{Y - \int_0^1 \beta(s,\cdot)X(s)ds\big\}\big]\big) = 0$ implies that $\mathrm{E}(X \otimes Y) = \mathrm{E}\big[X \otimes \big\{\int_0^1 \beta(s,\cdot)X(s)ds\big\}\big]$, so that $\mathrm{cov}\{X(s), Y(t)\} = \int_0^1 \beta(s',t)C_X(s,s')ds'$. Then, (3.3) can be obtained via the expansion of this equation on both sides with respect to the eigenfunctions.

We now state several assumptions required for the asymptotic theory developed in this section.

Assumption A3. For any $t_1, t_2 \in [0,1]$, $\mathrm{E}\{\varepsilon(t_1)|X\} = 0$ and $\mathrm{E}\{\varepsilon(t_1)\varepsilon(t_2)|X\} = C_\varepsilon(t_1,t_2)$ almost surely. Assume further that $C_\varepsilon(t_1,t_2) = \sigma_\varepsilon^2\,\delta(t_1,t_2)$, for some $\sigma_\varepsilon^2 > 0$, where δ is the Dirac-delta function.

Assumption A4. There exist constants $c_X, c_\varepsilon > 0$ such that $\mathrm{E}\{\exp(c_X\|X\|_{L^2})\} < \infty$ and $\mathrm{E}\{\exp(c_\varepsilon\|\varepsilon\|_{L^2})|X\} < \infty$ almost surely. There exists a constant $c' > 0$ such that, for any $\omega \in L^2([0,1])$,

$$\mathrm{E}\bigg\{\int_0^1 X(s)\,\omega(s)ds\bigg\}^4 \le c'\left[\mathrm{E}\bigg\{\int_0^1 X(s)\,\omega(s)ds\bigg\}^2\right]^2. \tag{3.4}$$

Moreover, $\mathrm{E}\{\varepsilon(\cdot)\varepsilon(\cdot)\varepsilon(\cdot)\varepsilon(\cdot)|X\} \in L^2([0,1]^4)$ almost surely.

Assumption A5. The regularization parameter λ in (2.8) satisfies $\lambda = o(1)$, $n^{-1}\lambda^{-1/(2D)} = o(1)$ and $n^{-1/2}\lambda^{-\varsigma}(\log\log n)^{1/2} = o(1)$ as $n \to \infty$, where $\varsigma = (2D - 2a - 1)/(4Dm) + (a + 1)/(2D) > 0$, for the constants a and D in Assumption A2.

Remark 3.2. The reason for postulating a white-noise error covariance in Assumption A3 is that the commonly used L^2 loss function defined in (2.5) corresponds to the likelihood function in the case of the Gaussian white noise error process; see, for example, Wellner, 2003. It is also notable, that for the scalar-on-function model in (1.1), this assumption is in fact not necessary (as there is no error function in this model) and Assumption A3 reduces to $\mathrm{E}(\varepsilon|X) = 0$ and $\mathrm{E}(\varepsilon^2|X) = \sigma_\varepsilon^2$ almost surely.

Assumption A4 requires that $\|X\|_{L^2}$, and $\|\varepsilon\|_{L^2}$ conditional on X, have finite exponential moments. Moreover, condition (3.4) is a common moment assumption in the context of scalar-on function regression, used, for example, in Cai and Yuan (2012) and Shang and Cheng (2015). Assumption A5 specifies the condition on the rate in which λ tends to zero as $n \to \infty$.

The first result of this section establishes a Bahadur representation for the estimator (2.4) in the function-on-function linear regression model (2.1). It is essential for deriving weak convergence of the estimator $\widehat{\beta}_n$, which serves as the foundation of our statistical analysis in Section 4. The proof of Theorem 3.1 is given in Section A.2.

Theorem 3.1 (Bahadur representation). *Suppose Assumptions A1–A5 are satisfied. Then, we have*

$$\|\widehat{\beta}_n - \beta_0 + S_{n,\lambda}(\beta_0)\|_K = O_p(v_n),$$

where $S_{n,\lambda}$ is defined in (3.2), $\varsigma > 0$ is the constant in Assumption A5 and

$$v_n = n^{-1/2}\lambda^{-\varsigma}(\lambda^{1/2} + n^{-1/2}\lambda^{-1/(4D)})(\log\log n)^{1/2}\,. \tag{3.5}$$

Due to the reproducing property of the kernel K, we have, for $(s,t) \in [0,1]^2$ fixed,

$$\widehat{\beta}_n(s,t) - \beta_0(s,t) = \langle \widehat{\beta}_n - \beta_0, K_{(s,t)} \rangle_K\,,$$

and, by Theorem 3.1, this expression can be linearized to establish point-wise asymptotic normality of $\widehat{\beta}_n(s,t)$. The following theorem gives a rigorous formulation of these heuristic arguments and is proved in Section A.3.

Theorem 3.2. *Suppose that Assumptions A1–A5 hold. Assume $n\lambda^{(2a+1)/(2D)}\{\log(\lambda^{-1})\}^{-4} \to \infty$, $\sqrt{n}v_n = o(1)$, $n\lambda^2 = o(1)$, $\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2}\varphi_{k\ell}^2(s,t) \asymp \lambda^{-(2a+1)/(2D)}$, as $n \to \infty$; $\sum_{k,\ell}\rho_{k\ell}^2 V^2(\beta_0,\varphi_{k\ell}) < \infty$. Then,*

$$\frac{\sqrt{n}\{\widehat{\beta}_n(s,t) - \beta_0(s,t)\}}{\sqrt{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2}\varphi_{k\ell}^2(s,t)}} \xrightarrow{d.} N(0,1)\,.$$

The final result of this section establishes the weak convergence of the process $\widehat{\beta}_n$ in the space $C([0,1]^2)$, which enables us to construct simultaneous confidence regions for the slope surface β_0 (see Section 4.2 below). The proof is given in Section A.4.

Theorem 3.3. *Suppose that Assumptions A1–A5 hold and that $n\lambda^2 = o(1)$, $\sqrt{n}v_n = o(1)$, $\lambda = o(n^{-\nu_1})$, $n^{1-\nu_2}\lambda^{(2a+1)/(2D)} \to \infty$ for some constants $\nu_1, \nu_2 > 0$ as $n \to \infty$. Assume that $\sum_{k,\ell}\rho_{k\ell}^2 V^2(\beta_0,\varphi_{k\ell}) < \infty$, that the limit*

$$C_Z\{(s_1,t_1),(s_2,t_2)\} = \lim_{n\to\infty} \lambda^{(2a+1)/(2D)} \sum_{k,\ell} \frac{\varphi_{k\ell}(s_1,t_1)\varphi_{k\ell}(s_2,t_2)}{(1+\lambda\rho_{k\ell})^2} \tag{3.6}$$

exists, and that there exist nonnegative constants c_0, b, ϑ such that

$$\limsup_{\lambda\to 0} \lambda^{(2b+1)/(2D)} \sum_{k,\ell} \frac{|\varphi_{k\ell}(s_1,t_1) - \varphi_{k\ell}(s_2,t_2)|^2}{(1+\lambda\rho_{k\ell})^2} \le c_0 \max\{|s_1-s_2|^{2\vartheta}, |t_1-t_2|^{2\vartheta}\}\,. \quad (3.7)$$

If the constant a in Assumption A2 either satisfies the condition (i) $b < a$, $\vartheta \ge 0$ or the condition (ii) $b = a$, $\vartheta > 1$, then

$$\{\mathbb{G}_n(s,t)\}_{s,t\in[0,1]} = \sqrt{n}\lambda^{(2a+1)/(4D)}\{\widehat{\beta}_n(s,t) - \beta_0(s,t)\}_{s,t\in[0,1]} \rightsquigarrow \{Z(s,t)\}_{s,t\in[0,1]} \quad (3.8)$$

in $C([0,1]^2)$, where Z is a mean-zero Gaussian process with covariance kernel C_Z in (3.6).

4 Statistical consequences

In this section, we study several statistical inference problems regarding the model in (2.1). We first show in Section 4.1 that the estimator $\widehat{\beta}_n$ in (2.4) achieves the minimax convergence rate. In Section 4.2, we propose point-wise and simultaneous confidence regions for the slope surface β_0. In Section 4.3, we develop a new test for the the classical hypotheses (1.3) based on the sup-norm using the duality between confidence regions and hypotheses testing. Moreover, we also extend the penalized likelihood-ratio test for scalar-on-function linear regression proposed in Shang and Cheng (2015) to the function-on-function linear regression model (a numerical comparison of both tests can be found in Section 5.2 and shows some superiority of the confidence region approach.) In Section 4.4, we study a test for a relevant deviation of the "true" slope function and a given function β_*. Finally, a simultaneous prediction band for the conditional mean curve $\mathrm{E}\{Y(t)|X = x_0\}$ is proposed in Section 4.5. The methodology requires knowledge of the constants a and D in Assumption A2, and a data driven rule for this choice will be given in Section 5.1.

We also emphasize that, although we are mainly concentrating on the function-on-function linear regression model, all results presented so far also hold for the scalar-on-function linear model (under even weaker assumptions). As a consequence, we also obtain new powerful methodology for the scalar-on-function linear regression model (1.1) as well, and we briefly illustrate this fact for the problem of testing relevant hypotheses in Section 4.6.

4.1 Optimality

Under Assumption A1, the operator V in (2.9) defines a norm, say $\|\beta\|_V^2 = V(\beta,\beta)$, on $\mathcal{H}$. As a by-product of the Bahadur representation in Theorem 3.1, we are able to show the upper bound for the convergence rate of the estimator $\widehat{\beta}_n$ in (2.4) with respect to the $\|\cdot\|_V$-norm. Moreover, we also prove that this rate is of the same order as the lower bound for estimating β_0, which shows that $\widehat{\beta}_n$ is minimax optimal. To be precise, let $\mathcal{G}$ denote the collection of all estimators from the data $(X_1, Y_1), \ldots, (X_n, Y_n)$, and let $\mathcal{F}$ denote the collection of the joint distribution F of the X and Y that satisfies Assumptions A1–A4, according to the linear model in (2.1). The following theorem is proved in Section A.5.

Theorem 4.1 (Optimal convergence rate)**.** *Suppose Assumptions A1–A5 hold.*
(i) By taking $\lambda \asymp n^{-2D/(2D+1)}$, we have

$$\lim_{c\to\infty} \limsup_{n\to\infty} \sup_{\substack{\beta_0\in\mathcal{H}\\ F\in\mathcal{F}}} \mathrm{P}\big(\|\widehat{\beta}_n - \beta_0\|_V^2 \geq c\, n^{-2D/(2D+1)}\big) = 0\,.$$

(ii) There exists a constant $c_0 > 0$ such that

$$\liminf_{n\to\infty} \inf_{\widetilde{\beta}\in\mathcal{G}} \sup_{\substack{\beta_0\in\mathcal{H}\\ F\in\mathcal{F}}} \mathrm{P}\big(\|\widetilde{\beta} - \beta_0\|_V^2 \geq c_0\, n^{-2D/(2D+1)}\big) > 0\,.$$

Theorem 4.1 shows that the estimator $\widehat{\beta}_n$ in (2.4) achieves the minimax optimal convergence rate $n^{-2D/(2D+1)}$ with respect to the $\|\cdot\|_V$-norm. It is of interest to compare this result with the minimax prediction rate obtained in Sun et al. (2018). First, we consider the estimation of the slope surface β_0 in the Sobolev space $\mathcal{H}$ on the square $[0,1]^2$ defined in (2.2), whereas Sun et al. (2018) considered a tensor product RKHS on $[0,1]^2$. Second, Sun et al. (2018) showed their minimax properties in terms of *excess prediction rate* (hereinafter denoted by EPR), defined by

$$\begin{aligned}
\mathrm{EPR}(\widetilde{\beta}_n) &= \int_0^1 \mathrm{E}_{n+1}\bigg\{ \bigg|Y_{n+1} - \int_0^1 \widetilde{\beta}_n(s,t)X_{n+1}(s)ds\bigg|^2 - \bigg|Y_{n+1} - \int_0^1 \beta_0(s,t)X_{n+1}(s)ds\bigg|^2 \bigg\}dt \\
&= \int_0^1 \mathrm{E}_{n+1}\bigg|\int_0^1 \{\widetilde{\beta}_n(s,t) - \beta_0(s,t)\}X_{n+1}(s)ds\bigg|^2 dt\,. \qquad (4.1)
\end{aligned}$$

Here $\widetilde{\beta}_n$ is an estimator from the data $\{(X_i, Y_i)\}_{i=1}^n$, (X_{n+1}, Y_{n+1}) is an independent future observation and E_{n+1} is the conditional expectation with respect to $(X_1, Y_1), \ldots, (X_n, Y_n)$ (which means that the expectation is taken with respect to (X_{n+1}, Y_{n+1})). In fact, we have

$$\begin{aligned}
\mathrm{EPR}(\widetilde{\beta}_n) &= \int_0^1 \bigg[\int_0^1\int_0^1 C_X(s_1,s_2)\{\widetilde{\beta}_n(s_1,t) - \beta_0(s_1,t)\}\{\widetilde{\beta}_n(s_2,t) - \beta_0(s_2,t)\}ds_1ds_2\bigg]dt \\
&= V(\widetilde{\beta}_n - \beta_0, \widetilde{\beta}_n - \beta_0) = \|\widetilde{\beta}_n - \beta_0\|_V^2\,,
\end{aligned}$$

which shows that the difference between $\widetilde{\beta}_n$ and the true β_0 in squared $\|\cdot\|_V$-norm is equivalent to $\mathrm{EPR}(\widetilde{\beta}_n)$. Therefore, it follows from Theorem 4.1 that for the estimator $\widehat{\beta}_n$ in (2.4), $\mathrm{EPR}(\widehat{\beta}_n)$ achieves the minimax rate $n^{-2D/(2D+1)}$, which is determined by the constant $D > 0$ that specifies the growing rate of $J(\varphi_{k\ell}, \varphi_{k\ell})$ in Assumption A2. In comparison, Sun et al. (2018) showed that the EPR of their estimator achieves the minimax rate $n^{-2\breve{D}/(2\breve{D}+1)}$, where the constant $\breve{D} > 0$ characterises the decay rate of eigenvalues of the kernel

$$\Pi\{(s_1,t_1),(s_2,t_2)\} = \int_{[0,1]^3} C_X(s,t)\widetilde{K}^{1/2}\{(s_1,t_1),(s,t)\}\widetilde{K}^{1/2}\{(s_2,t_2),(s,u)\}\, ds\, dt\, du\,,$$

where $\widetilde{K}$ is the reproducing kernel of their tensor product RKHS.

4.2 Confidence regions

The asymptotic normality of the estimator $\widehat{\beta}_n(s,t)$ in Theorem 3.2 enables us to construct a point-wise $(1-\alpha)$-confidence interval of $\beta_0(s,t)$, for fixed $(s,t) \in [0,1]^2$, since

$$\lim_{n\to\infty} \mathrm{P}\Big\{\beta_0(s,t) \in \big[\widehat{\beta}_n(s,t) - \mathcal{Q}_{1-\alpha/2}\,\sigma_\tau(s,t)\,,\widehat{\beta}_n(s,t) + \mathcal{Q}_{1-\alpha/2}\,\sigma_\tau(s,t)\big]\Big\} = 1-\alpha\,,$$

where $\sigma_\tau(s,t) = \big\{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2}\varphi_{k\ell}^2(s,t)\big\}^{1/2}$ and $\mathcal{Q}_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

On the other hand, the construction of simultaneous confidence regions based on the sup-norm for the slope surface β_0 is more complicated. In principle, this is possible using Theorem 3.3 and the continuous mapping theorem, which give

$$\sqrt{n}\lambda^{(2a+1)/(4D)} \sup_{(s,t)\in[0,1]^2} |\widehat{\beta}_n(s,t) - \beta_0(s,t)| \xrightarrow{d} T = \max_{(s,t)\in[0,1]^2} |Z(s,t)|\,, \tag{4.2}$$

where Z is the mean-zero Gaussian process defined in (3.8). Thus, if $\mathcal{Q}_{1-\alpha}(T)$ denotes the $(1-\alpha)$-quantile of the distribution of T and

$$\widehat{\beta}_n^{\pm}(s,t) = \widehat{\beta}_n(s,t) \pm \frac{\mathcal{Q}_{1-\alpha}(T)}{\sqrt{n}\lambda^{(2a+1)/(4D)}}\,,$$

then the set $\mathfrak{C}_n(\alpha) = \big\{\beta : \widehat{\beta}_n^{-}(s,t) \le \beta(s,t) \le \widehat{\beta}_n^{+}(s,t)\big\}$ defines a simultaneous asymptotic $(1-\alpha)$-confidence region for β_0, i.e.,

$$\lim_{n\to\infty} \mathrm{P}\{\beta_0 \in \mathfrak{C}_n(\alpha)\} = 1-\alpha.$$

However, the quantiles of the distribution of T depend on the covariance function C_Z in (3.6) of the Gaussian process Z, and is rarely available in practice. In order to circumvent this difficulty, we propose the following bootstrap procedure to approximate $\mathcal{Q}_{1-\alpha}(T)$.

Algorithm 4.1 (Bootstrap simultaneous confidence region for the slope surface β_0).

1. Generate i.i.d. bootstrap weights $\{M_{i,q}\}_{1\le i\le n, 1\le q\le Q}$ independent of the data $\{(X_i,Y_i)\}_{i=1}^n$ from a two-point distribution: taking $1-1/\sqrt{2}$ with probability $2/3$ and taking $1+\sqrt{2}$ with probability $1/3$, such that $\mathrm{E}(M_{i,q}) = \mathrm{var}(M_{i,q}) = 1$.
2. Compute $\widehat{\beta}_n$ in (2.4); for each $1 \le q \le Q$, compute the bootstrap estimator
$$\widehat{\beta}^*_{n,q} = \arg\min_{\beta\in\mathcal{H}} \left[\frac{1}{2n} \sum_{i=1}^n M_{i,q} \int_0^1 \left\{ Y_i(t) - \int_0^1 \beta(s,t) X_i(s) ds \right\}^2 dt + \frac{\lambda}{2} J(\beta,\beta) \right]. \quad (4.3)$$
3. For $1 \le q \le Q$, let
$$\mathbb{G}^*_{n,q}(s,t) = \sqrt{n}\lambda^{(2a+1)/(4D)} \{\widehat{\beta}^*_{n,q}(s,t) - \widehat{\beta}_n(s,t)\} ;$$
$$\widehat{T}^*_{n,q} = \sup_{(s,t)\in[0,1]^2} |\mathbb{G}^*_{n,q}(s,t)| . \quad (4.4)$$
Compute the empirical $(1-\alpha)$-quantile of the sample $\{\widehat{T}^*_{n,q}\}_{q=1}^Q$, denoted by $\mathcal{Q}_{1-\alpha}(\widehat{T}^*_{n,Q})$.
4. Let $\widehat{\beta}^{*\pm}_{n,Q}(s,t) = \widehat{\beta}_n(s,t) \pm \mathcal{Q}_{1-\alpha}(\widehat{T}^*_{n,Q})/\{\sqrt{n}\lambda^{(2a+1)/(4D)}\}$. Define the set
$$\mathfrak{C}^*_{n,Q}(\alpha) = \left\{ \beta : [0,1]^2 \to \mathbb{R} : \ \widehat{\beta}^{*-}_{n,Q}(s,t) \le \beta(s,t) \le \widehat{\beta}^{*+}_{n,Q}(s,t) \right\} \quad (4.5)$$
as the simultaneous $(1-\alpha)$ confidence region for the slope surface β_0 in model (2.1).

The following theorem, which is proved in Section A.6, provides a theoretical justification of the above bootstrap procedure and establishes the consistency of the simultaneous confidence region in Algorithm 4.1.

Theorem 4.2. *Under the conditions of Theorem 3.3 we have*
$$\{\mathbb{G}^*_{n,q}(s,t)\}_{s,t\in[0,1]} \rightsquigarrow \{Z(s,t)\}_{s,t\in[0,1]} \quad \text{in } C([0,1]^2) \quad (4.6)$$
conditionally on the data $\{(X_i,Y_i)\}_{i=1}^n$, *where* Z *is the Gaussian process in Theorem 3.3. In particular, the set* $\mathfrak{C}^*_n(\alpha)$ *in Algorithm 4.1 defines a simultaneous asymptotic* $(1-\alpha)$ *confidence region for the slope surface* β_0 *in model* (2.1), *that is*
$$\lim_{Q\to\infty} \lim_{n\to\infty} \mathrm{P}\{\beta_0 \in \mathfrak{C}^*_{n,Q}(\alpha)\} = 1-\alpha . \quad (4.7)$$

4.3 Classical hypotheses

For a given surface β_* on $[0,1]^2$, consider the "classical" hypotheses
$$H_0 : \ \beta_0 = \beta_* \qquad \text{versus} \qquad H_1 : \ \beta_0 \ne \beta_* . \quad (4.8)$$

In the special case where $\beta_* \equiv 0$, (4.8) becomes $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$, which is the conventional hypothesis for linear effect; we refer to Tekbudak et al. (2019) for a review in the scalar-on-function regression context.

In order to construct a test for (4.8), we may utilize the duality between hypotheses thesing and confidence regions (see, for example, Aitchison, 1964). Specifically, recall from Section 4.2 that we are able to construct a simultaneous confidence region $\mathfrak{C}^*_{n,Q}(\alpha)$ for β_0 using Algorithm 4.1, such that $\mathrm{P}\{\beta_0 \in \mathfrak{C}^*_{n,Q}(\alpha)\} \to 1-\alpha$ as $n, Q \to \infty$. Then, the decision rule, which rejects the null hypothesis, whenever

$$\beta_* \notin \mathfrak{C}^*_{n,Q}(\alpha)\,, \tag{4.9}$$

defines an asymptotic level α test for the classical hypotheses in (4.8).

An alternative approach to construct a test for these classical hypotheses is to extend the penalized likelihood ratio test (hereinafter denoted by PLRT), proposed in Shang and Cheng (2015) for the scalar-on-function regression context, to the functional response context. Specifically, for the objective function $L_{n,\lambda}$ in (2.7), consider the penalized likelihood ratio test statistic defined by

$$\mathfrak{L}_n(\beta_*) = L_{n,\lambda}(\beta_*) - L_{n,\lambda}(\widehat{\beta}_n)\,. \tag{4.10}$$

In order to find the asymptotic distribution of $\mathfrak{L}_n(\beta_*)$ under the null hypothesis, we define the sequences

$$u_n = \frac{\big\{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-1}\big\}^2}{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2}}\,, \qquad \sigma_n^2 = \frac{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-1}}{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2}}\,, \tag{4.11}$$

and obtain the following result, which is proved in Section A.7.

Theorem 4.3. *Let Assumptions A1–A5 be satisfied. Assume that as $n \to \infty$, $n\lambda^{(2D+1)/(2D)} = o(1)$, $nv_n^2 = o(1)$ and $nv_n\lambda^{(D+1)/(2D)} = o(1)$, where v_n is defined in (3.5). Then, under the null hypothesis in (4.8),*

$$\frac{1}{\sqrt{2u_n}}\{2n\sigma_n^2\,\mathfrak{L}_n(\beta_*) - u_n\} \xrightarrow{d} N(0,1)\,,$$

where u_n and σ_n^2 are given in (4.11).

Then, the PLRT at nominal level α rejects the null hypothesis in (4.8), whenever

$$2n\sigma_n^2\,\mathfrak{L}_n(\beta_*) - u_n \geq \sqrt{2u_n}\,\mathcal{Q}_{1-\alpha}\,, \tag{4.12}$$

where $\mathcal{Q}_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution. We compare the test (4.9) and PLRT (4.12) for the classical hypotheses (4.8) through simulated data in Section 5.2.

4.4 Relevant hypotheses

It turns out that the construction of an asymptotic level α test for relevant hypotheses as formulated in (1.4) is substantially more difficult. Recall that we are interested in testing whether the maximum deviation between a given surface β_* and the unknown "true" slope surface β_0 exceeds a given value $\Delta \geq 0$, and note that with the notation $d_\infty = \sup_{(s,t)\in[0,1]^2} |\beta_0(s,t) - \beta_*(s,t)|$ the relevant hypothesis in (1.4) can be rewritten as

$$H_0 : d_\infty \leq \Delta \quad \text{versus} \quad H_1 : d_\infty > \Delta \ . \tag{4.13}$$

Therefore, a reasonable decision rule is to reject the null hypothesis for large values of the statistic

$$\widehat{d}_\infty = \sup_{(s,t)\in[0,1]^2} |\widehat{\beta}_n(s,t) - \beta_*(s,t)| \ . \tag{4.14}$$

When $\Delta = 0$, the above relevant hypothesis reduces to the classical hypotheses in (4.8). In this case, under the null hypothesis $H_0 : \beta_0 = \beta_*$, there exists only one function-on-function linear model, which simplifies the asymptotic analysis of the corresponding test statistics substantially, because basically the asymptotic distribution can be obtained from Theorem 3.3 via continuous mapping (see also the discussion in Section 4.3). On the other hand, if $\Delta > 0$, there appear additional nuisance parameters in the asymptotic distribution of the difference $\widehat{d}_\infty - d_\infty$, which makes the analysis of a decision rule more intricate.

For a precise description of the asymptotic distribution of $\widehat{d}_\infty$ in the case $\Delta > 0$, let

$$\mathcal{E}^\pm = \{(s,t) \in [0,1]^2 : \beta_0(s,t) - \beta_*(s,t) = \pm d_\infty\} \tag{4.15}$$

denote the set of points, where the surface $\beta_0 - \beta_*$ attains it sup-norm (the set $\mathcal{E}^+$) or its negative sup-norm (the set $\mathcal{E}^-$). Here we take the convention that $\mathcal{E}^+ = \mathcal{E}^- = [0,1]^2$ if $d_\infty = 0$ and denote by $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$ the set of *extremal points* of the difference $\beta_0 - \beta_*$. The following result describes the asymptotic properties of $\widehat{d}_\infty$ and is crucial for constructing a test for the relevant hypothesis. It is proved in Section A.8.

Corollary 4.1. *If the assumptions of Theorem 3.3 are satisfied, then*

$$\sqrt{n}\lambda^{(2a+1)/(4D)}(\widehat{d}_\infty - d_\infty) \xrightarrow{d} T_{\mathcal{E}} = \max\Big\{ \sup_{(s,t)\in\mathcal{E}^+} Z(s,t), \sup_{(s,t)\in\mathcal{E}^-} \{-Z(s,t)\}\Big\}, \tag{4.16}$$

where Z is the mean-zero Gaussian process defined in (3.8).

Note that the distribution of $T_{\mathcal{E}}$ depends on the covariance structure of the limiting process Z in (3.8) and implicitly through the sets of extremal points $\mathcal{E}^+$ and $\mathcal{E}^-$ on the "true" (unknown) difference $\beta_0 - \beta_*$. In order to motivate the final test, assume for the moment the quantile, say $\mathcal{Q}_{1-\alpha}(T_{\mathcal{E}})$, of this distribution would be available (we will soon provide an estimate for it), then we will show in Section A.8 that

$$\lim_{n\to\infty} \mathrm{P}\Big\{\widehat{d}_\infty > \Delta + \frac{\mathcal{Q}_{1-\alpha}(T_{\mathcal{E}})}{\sqrt{n}\lambda^{(2a+1)/(4D)}}\Big\} = \begin{cases} 0 & \text{if } d_\infty < \Delta \\ \alpha & \text{if } d_\infty = \Delta \\ 1 & \text{if } d_\infty > \Delta \end{cases}. \tag{4.17}$$

Here the first two lines correspond to the null hypothesis $d_\infty \le \Delta$ and the third line to the alternative in (4.13).

This yields, in principle, a consistent asymptotic level α test for the relevant hypotheses (4.13). To implement such a test we need to approximate the quantiles of the random variable $T_{\mathcal{E}}$ in (4.16). While the covariance structure of the process Z can be again estimated by the multiplier bootstrap (see the discussion below), the estimation of the extremal sets is a little more tricky. For this purpose we propose to estimate the sets $\mathcal{E}^+$ and $\mathcal{E}^-$ by

$$\begin{aligned} \widehat{\mathcal{E}}^+ &= \Big\{(s,t)\in[0,1]^2 : \widehat{\beta}_n(s,t) - \beta_*(s,t) \ge \widehat{d}_\infty - c\,\frac{\log n}{\sqrt{n}}\Big\}, \\ \widehat{\mathcal{E}}^- &= \Big\{(s,t)\in[0,1]^2 : \widehat{\beta}_n(s,t) - \beta_*(s,t) \le -\widehat{d}_\infty + c\,\frac{\log n}{\sqrt{n}}\Big\}, \end{aligned} \tag{4.18}$$

respectively, where we use a term $c\log n/\sqrt{n}$ in the cut-off values, for some tuning parameter $c > 0$. Then, the random variable $T_{\mathcal{E}}$ in (4.16) can be approximated by

$$\widehat{T}_{\mathcal{E}} = \max\Big\{ \sup_{(s,t)\in\widehat{\mathcal{E}}^+} Z(s,t), \sup_{(s,t)\in\widehat{\mathcal{E}}^-} \{-Z(s,t)\}\Big\}. \tag{4.19}$$

In view of (4.17), the null hypothesis should be rejected at nominal level $\alpha \in (0,1)$, if

$$\widehat{d}_\infty = \sup_{(s,t)\in[0,1]^2} |\widehat{\beta}_n(s,t) - \beta_*(s,t)| > \Delta + \frac{\mathcal{Q}_{1-\alpha}(\widehat{T}_{\mathcal{E}})}{\sqrt{n}\lambda^{(2a+1)/(4D)}}, \tag{4.20}$$

where $\mathcal{Q}_{1-\alpha}(\widehat{T}_{\mathcal{E}})$ denotes the $(1-\alpha)$-quantile of $\widehat{T}_{\mathcal{E}}$. Now, we still need to approximate the quantile $\mathcal{Q}_{1-\alpha}(\widehat{T}_{\mathcal{E}})$ of $\widehat{T}_{\mathcal{E}}$. Since the asymptotic distribution of $\widehat{T}_{\mathcal{E}}$ depends on the unknown covariance function C_Z in (3.6), we propose to combine a multiplier bootstrap similar to the ones introduced in Section 4.2 with the estimation of the extremal sets. Specifically, for $1 \le q \le Q$ and the process $\mathbb{G}^*_{n,q}(s,t)$ defined in (4.4), let

$$\widehat{T}^*_{\mathcal{E},n,q} = \max\left\{ \sup_{(s,t)\in\widehat{\mathcal{E}}^+} \mathbb{G}^*_{n,q}(s,t)\, , \sup_{(s,t)\in\widehat{\mathcal{E}}^-} \{-\mathbb{G}^*_{n,q}(s,t)\} \right\}, \tag{4.21}$$

where $\widehat{\mathcal{E}}^{\pm}$ are the estimated extremal sets defined in (4.18). Then, the quantile of $\widehat{T}_{\mathcal{E}}$ can be approximated by the quantiles of the bootstrap extremal value estimates $\{\widehat{T}^*_{\mathcal{E},n,q}\}_{q=1}^Q$. We summarize the bootstrap procedures for the relevant hypothesis in (4.13) at nominal level α in the following algorithm.

Algorithm 4.2 (Bootstrap for relevant hypotheses)**.**

1. Generate i.i.d. bootstrap weights $\{M_{i,q}\}_{1\le i\le n, 1\le q\le Q}$ and compute the bootstrap process $\mathbb{G}^*_{n,q}(s,t)$ in (4.4).
2. Compute the extremal sets $\widehat{\mathcal{E}}^{\pm}$ in (4.18). For $1 \le q \le Q$, compute $\widehat{T}^*_{\mathcal{E},n,q}$ in (4.21) and obtain the empirical $(1-\alpha)$-quantile of the sample $\{\widehat{T}^*_{\mathcal{E},n,q}\}_{q=1}^Q$, denoted by $\mathcal{Q}_{1-\alpha}(\widehat{T}^*_{\mathcal{E},n,Q})$.
3. Reject the null hypothesis in (4.13) at nominal level α, if

$$\widehat{d}_\infty = \sup_{(s,t)\in[0,1]^2} |\widehat{\beta}_n(s,t) - \beta_*(s,t)| > \Delta + \frac{\mathcal{Q}_{1-\alpha}(\widehat{T}^*_{\mathcal{E},n,Q})}{\sqrt{n}\lambda^{(2a+1)/(4D)}}\,. \tag{4.22}$$

The following theorem, which is proved in Section A.9, provides a theoretical justification of the test (4.22).

Theorem 4.4. *Suppose the conditions of Theorem 3.3 hold. Then, the decision rule* (4.22) *defines a consistent and asymptotic level α test for the hypotheses* (4.13)*, that is*

$$\lim_{Q\to\infty}\lim_{n\to\infty} \mathrm{P}\left\{\widehat{d}_\infty > \Delta + \frac{\mathcal{Q}_{1-\alpha}(\widehat{T}^*_{\mathcal{E},n,Q})}{\sqrt{n}\lambda^{(2a+1)/(4D)}}\right\} = \begin{cases} 0 & \text{if } d_\infty < \Delta \\ \alpha & \text{if } d_\infty = \Delta \\ 1 & \text{if } d_\infty > \Delta \end{cases}. \tag{4.23}$$

4.5 Simultaneous prediction bands

Based on the estimator $\widehat{\beta}_n$ in (2.4), we can construct a simultaneous confidence region of the conditional mean $\mu_{x_0}(t) = \mathrm{E}\{Y(t)|X = x_0\} = \int_0^1 \beta_0(s,t)x_0(s)ds$, using the consistent estimator

$$\widehat{\mu}_{x_0}(t) = \int_0^1 \widehat{\beta}_n(s,t)x_0(s)ds. \tag{4.24}$$

The following theorem establishes the weak convergence of the process $\widehat{\mu}_{x_0}$ in the space $C([0,1])$, which enables us to construct simultaneous confidence regions for the function μ_{x_0}. The proof is given in Section A.10.

Theorem 4.5 (Simultaneous prediction band)**.** *Suppose that the conditions of Theorem 3.3 are satisfied. Then,*

$$\sqrt{n}\lambda^{(2a+1)/(4D)}\big\{\widehat{\mu}_{x_0}(t)-\mu_{x_0}(t)\big\}_{t\in[0,1]} \rightsquigarrow \big\{Z_{x_0}(t)\big\}_{t\in[0,1]} \qquad in\ C([0,1])\,,$$

where Z_{x_0} is a mean zero Gaussian process with covariance function

$$C_{Z,x_0}(t_1,t_2)=\int_0^1\int_0^1 C_Z\{(s_1,t_1),(s_2,t_2)\}\,x_0(s_1)\,x_0(s_2)\,ds_1\,ds_2 \tag{4.25}$$

and C_Z is defined in (3.6), Moreover,

$$\sqrt{n}\lambda^{(2a+1)/(4D)}\sup_{t\in[0,1]}|\widehat{\mu}_{x_0}(t)-\mu_{x_0}(t)| \xrightarrow{d} R_{x_0}=\max_{t\in[0,1]}|Z_{x_0}(t)|\,.$$

As the quantiles of the distribution of R_{x_0} depend in a complicate way on the covariance structure of the process Z_{x_0}, we propose the following bootstrap procedure for a simultaneous asymptotic $(1-\alpha)$ prediction band for the fiction $t\to\mu_{x_0}(t)=\mathrm{E}\{Y(t)|X=x_0\}$.

Algorithm 4.3 (Bootstrap simultaneous prediction band)**.**

1. Generate i.i.d. weights $\{M_{i,q}\}_{1\le i\le n,1\le q\le Q}$ and compute the bootstrap estimators $\{\widehat{\beta}^*_{n,q}\}_{q=1}^{Q}$ in (4.3). Compute $\widehat{\beta}_n$ in (2.4) and $\widehat{\mu}_{x_0}(t)$ in (4.24).
2. For $1\le q\le Q$, compute $\mathbb{L}^*_{x_0,q}(t)=\sqrt{n}\lambda^{(2a+1)/(4D)}\int_0^1\{\widehat{\beta}^*_{n,q}(s,t)-\widehat{\beta}_n(s,t)\}x_0(s)ds$ and define $\widehat{R}^*_{x_0,q}=\sup_{t\in[0,1]}|\mathbb{L}^*_{x_0,q}(t)|$. Compute the empirical $(1-\alpha)$-quantile of the bootstrap sample $\{\widehat{R}^*_{x_0,q}\}_{q=1}^{Q}$, denoted by $\mathcal{Q}_{1-\alpha}(\widehat{R}^*_{x_0,Q})$.
3. Let $\widehat{\mu}^{*\pm}_{x_0,Q}(t)=\widehat{\mu}_{x_0}(t)\pm\mathcal{Q}_{1-\alpha}(\widehat{R}^*_{x_0,Q})/\{\sqrt{n}\lambda^{(2a+1)/(4D)}\}$. Define the set
$$\mathfrak{B}^*_{n,Q}(\alpha)=\big\{\mu:\widehat{\mu}^{*-}_{x_0,Q}(t)\le\mu(t)\le\widehat{\mu}^{*+}_{x_0,Q}(t)\big\} \tag{4.26}$$
as simultaneous $(1-\alpha)$ prediction band for the function μ_{x_0}.

The following theorem provides a formal justification of the bootstrap procedure in Algorithm 4.3, the proof uses similar arguments as given in the proof of Theorem 4.2 and is therefore omitted.

Theorem 4.6 (Bootstrap simultaneous prediction band)**.** *Suppose the assumptions in Theorem 3.3 are satisfied. Then, the set $\mathfrak{B}^*_{n,Q}(\alpha)$ in (4.26) defines a simultaneous asymptotic $(1-\alpha)$ prediction band for the function μ_{x_0}, that is*

$$\lim_{Q\to\infty}\lim_{n\to\infty}\mathrm{P}\{\mu_{x_0}\in\mathfrak{B}^*_{n,Q}(\alpha)\}=1-\alpha\,.$$

4.6 Scalar response

The results presented so far provide also new inference tools for the scalar-on-function linear model

$$Y_i = \int_0^1 \beta_0(s)\, X_i(s)\, ds + \varepsilon_i\,, \quad 1 \le i \le n\,, \tag{4.27}$$

which can be considered as a special case of model (1.2), where the response Y is a scalar variable. In this setting, the estimator defined in (2.4) becomes

$$\widehat{\beta}_n = \arg\min_{\beta\in\mathcal{H}_s}\Big[\frac{1}{2n}\sum_{i=1}^{n}\Big\{Y_i - \int_0^1 \beta(s)X_i(s)\,ds\Big\}^2 + \frac{\lambda}{2}\,J_s(\beta,\beta)\Big]\,,$$

where $\mathcal{H}_s = \big\{\beta : [0,1] \to \mathbb{R} \,|\, \beta, \beta', \ldots, \beta^{(m-1)}$ are absolutely continuous; $\beta^{(m)} \in L^2([0,1])\big\}$ is the Sobolev space on $[0,1]$ of order m, and $J_s(\beta,\beta) = \int_0^1 \{\beta^{(m)}(s)\}^2 ds$. A direct consequence of Theorem 3.3 in the function-response setting is the weak convergence

$$\{\mathbb{G}_n(s)\}_{s\in[0,1]} = \sqrt{n}\lambda^{(2a+1)/(4D)}\big\{\widehat{\beta}_n(s) - \beta_0(s)\big\}_{s\in[0,1]} \rightsquigarrow \{Z(s)\}_{s\in[0,1]} \tag{4.28}$$

in $C([0,1])$, where Z is a mean-zero Gaussian process. Hence, the methodology proposed in Section 4, namely the bootstrap procedures for simultaneous confidence regions, relevant hypothesis tests and simultaneous prediction bands, carries over naturally to the scalar response case.

Exemplary, we consider (for a given constant $\Delta \ge 0$) the problem of constructing a test for the relevant hypotheses

$$H_0 : \sup_{s\in[0,1]} |\beta_0(s) - \beta_*(s)| \le \Delta \quad \text{versus} \quad H_1 : \sup_{s\in[0,1]} |\beta_0(s) - \beta_*(s)| > \Delta\,, \tag{4.29}$$

in model (4.27), which is more challenging in nature to tackle. A consistent estimator of the maximum deviation $d_\infty = \sup_{s\in[0,1]} |\beta_0(s) - \beta_*(s)|$ is $\widehat{d}_\infty = \sup_{s\in[0,1]} |\widehat{\beta}_n(s) - \beta_*(s)|$, so that the null hypothesis in (4.29) should be rejected for large values of $\widehat{d}_\infty$. As analog of Algorithm 4.2, we obtain the following bootstrap test for the relevant hypotheses in (4.29).

Algorithm 4.4 (Bootstrap test for relevant hypotheses in the scalar-on-function mdoel)**.**

1. Generate i.i.d. bootstrap weights $\{M_{i,q}\}_{1\le i\le n, 1\le q\le Q}$ as in Algorithm 4.1 and for each $1 \le q \le Q$, compute the bootstrap estimator

$$\widehat{\beta}^*_{n,q} = \arg\min_{\beta\in\mathcal{H}_s}\Big[\frac{1}{2n}\sum_{i=1}^{n} M_{i,q}\Big\{Y_i - \int_0^1 \beta(s)\, X_i(s)\, ds\Big\}^2 + \frac{\lambda}{2}\,J_s(\beta,\beta)\Big]$$

and $\mathbb{G}^*_{n,q}(s) = \sqrt{n}\lambda^{(2a+1)/(4D)}\{\widehat{\beta}^*_{n,q}(s) - \widehat{\beta}_n(s)\}$.

2. Compute the extremal sets

$$\widehat{\mathcal{E}}^{\pm} = \left\{ s \in [0,1] : \pm\{\widehat{\beta}_n(s) - \beta_*(s)\} \geq \widehat{d}_\infty - c\,\frac{\log n}{\sqrt{n}} \right\}.$$

3. For $1 \leq q \leq Q$, compute

$$\widehat{T}^*_{\mathcal{E},n,q} = \max\left\{ \sup_{s\in\widehat{\mathcal{E}}^+} \mathbb{G}^*_{n,q}(s)\,,\, \sup_{s\in\widehat{\mathcal{E}}^-}\{-\mathbb{G}^*_{n,q}(s)\} \right\},$$

and obtain the empirical $(1-\alpha)$-quantile of the bootstrap sample $\{\widehat{T}^*_{\mathcal{E},n,q}\}_{q=1}^{Q}$, denoted by $\mathcal{Q}_{1-\alpha}(\widehat{T}^*_{\mathcal{E},n,Q})$.

4. Reject the null hypothesis in (4.29) at nominal level α, if

$$\sup_{s\in[0,1]} |\widehat{\beta}_n(s) - \beta_*(s)| > \Delta + \frac{\mathcal{Q}_{1-\alpha}(\widehat{T}^*_{\mathcal{E},n,Q})}{\sqrt{n}\lambda^{(2a+1)/(4D)}}\,. \tag{4.30}$$

It can be shown by similar arguments as given in the proof of Theorem 4.4 that the test (4.30) is a consistent and asymptotic level α test. The details are omitted for the sake of brevity.

5 Finite sample properties

5.1 Implementation

Because the estimators $\widehat{\beta}_n$ in (2.4) and its bootstrap analog $\widehat{\beta}^*_{n,q}$ in (4.3) are defined as the solution of a (penalized) minimization problem on an infinite dimensional function space, exact solution are inaccessible. In this section, we introduce finite-sample methods to circumvent this difficulty, and propose a method to choose the regularization parameter λ. We shall only present our approach for computing the bootstrap estimator $\widehat{\beta}^*_{n,q}$ in (4.3), since the estimator (2.4) can be viewed as a special case of (4.3) by taking $M_{i,q} = 1$ for any $1 \leq i \leq n$ and $1 \leq q \leq Q$.

We start by deducing from Assumption A2 that $J(x_{k\ell}\otimes\eta_\ell, x_{k'\ell'}\otimes\eta_{\ell'}) = \rho_{k\ell}\,\delta_{kk'}\,\delta_{\ell\ell'}$, so that for $\beta(s,t) = \sum_{k,\ell} b_{k\ell}\,\varphi_{k\ell}(s,t) \in \mathcal{H}$ and for $b_{k\ell} \in \mathbb{R}$, we have $J(\beta,\beta) = \sum_{k,\ell} b_{k\ell}^2\,\rho_{k\ell}$. We consider the Sobolev space on $[0,1]^2$ of order $m = 2$. In this case, the penalty functional in (2.4) is $J(\beta,\beta) = \int_0^1\int_0^1 (\beta_{ss}^2 + 2\beta_{st}^2 + \beta_{tt}^2)\,ds\,dt$, where $\beta_{st} = \frac{\partial^2\beta}{\partial s\partial t}$. For the choice of the basis, we use Proposition B.1 in Section B.2 of the online supplement. More precisely,

$\eta_1(t) \equiv 1$, $\eta_\ell(t) = \sqrt{2}\cos\{(\ell-1)\pi t\}$ $(\ell = 2, 3, \ldots)$ and for $\ell \geq 1$ the functions $\{\widetilde{x}_{k\ell}\}_{k\geq 1}$ are the eigenfunctions of integro-differential equation

$$\begin{cases} \rho_\ell \displaystyle\int_0^1 C_X(s_1, s_2)\, \widetilde{x}(s_2)\, ds_2 = \widetilde{x}^{(4)}(s_1) - 2(\ell-1)^2\pi^2\widetilde{x}^{(2)}(s_1) + (\ell-1)^4\pi^4 \\ \widetilde{x}^{(\theta)}(0) = \widetilde{x}^{(\theta)}(1) = 0\,, \qquad \text{for } \theta = 3 \text{ and } 4 \end{cases} \tag{5.1}$$

with corresponding eigenvalues $\{\rho_{k\ell}\}_{k\geq 1}$. In order to find the eigenvalue and the eigenfunction of (5.1), we use `Chebfun`, an efficient open-source Matlab add-on package, available at `https://www.chebfun.org/`. We substitute the covariance function C_X in (5.1) by its empirical version $\widehat{C}_X$, and find the eigenvalues $\widehat{\rho}_{k\ell}$ and the normalized eigenfunctions $\widehat{x}_{k\ell}$. Observing that the functions $\{\eta_\ell\}_{\ell\geq 1}$ are given by the cosine basis, we take the empirical eigenfunctions $\widehat{\varphi}_{k\ell} = \widehat{x}_{k\ell} \otimes \eta_\ell$. Now, we approximate the space $\mathcal{H}$ by a finite dimensional subspace spanned by $\{\widehat{\varphi}_{k\ell}\}_{1\leq k,\ell\leq v}$, defined by $\widetilde{\mathcal{H}} = \big\{\sum_{1\leq k,\ell\leq v} b_{k\ell}\, \widehat{\varphi}_{k\ell}\big\}$, where v is a truncation parameter that depends on the sample size n.

For $1 \leq q \leq Q$, for the q-th bootstrap estimator $\widehat{\beta}^*_{n,q}$ and for the bootstrap weights $\{M_{i,q}\}_{i=1}^n$ in Algorithms 4.1–4.3, let $\widetilde{M}_q = \mathrm{diag}(M_{1,q}, \ldots, M_{n,q})$ denote an $n \times n$ diagonal matrix. For $1 \leq i \leq n$ and $1 \leq k, \ell \leq v$, let $\omega_{ik\ell} = \int_0^1 X_i(s)\widehat{x}_{k\ell}(s)ds$ and let $\Omega_\ell = (\omega_{ik\ell})$ denote a $n \times v$ matrix; let $\widehat{\Lambda}_\ell = \mathrm{diag}\big\{\widehat{\rho}_{1\ell}, \ldots, \widehat{\rho}_{v\ell}\big\}$ denote a $v \times v$ diagonal matrix; let $\widetilde{Y}_{i\ell} = \langle Y_i, \eta_\ell\rangle_{L^2}$ and let $\widetilde{Y}_\ell = (\widetilde{Y}_{1\ell}, \ldots, \widetilde{Y}_{n\ell})^{\mathrm{T}} \in \mathbb{R}^v$. If we write $\widetilde{\beta}_{n,q} = \sum_{k=1}^v \sum_{\ell=1}^v \widetilde{b}_{k\ell}^{(q)}\, \widehat{\varphi}_{k\ell} \in \widetilde{\mathcal{H}}$, then, in order to approximate $\widehat{\beta}^*_{n,q}$ in (4.3), we find the $\widetilde{b}_{k\ell}^{(q)}$'s by solving the following optimization problem

$$\begin{aligned} \{\widetilde{b}_{k\ell}^{(q)}\} &= \underset{\{b_{k\ell}^{(q)}\}}{\arg\min} \left\{ \frac{1}{2n}\sum_{i=1}^n M_{i,q} \int_0^1 \left| Y_i(t) - \sum_{k,\ell=1}^v b_{k\ell}^{(q)}\, \eta_\ell(t) \int_0^1 X_i(s)\widehat{x}_{k\ell}(s)ds \right|^2 dt + \frac{\lambda_q}{2}\sum_{k,\ell=1}^v b_{k\ell}^{(q)^2}\, \widehat{\rho}_{k\ell} \right\} \\ &= \underset{\{b_{k\ell}^{(q)}\}}{\arg\min} \left\{ \frac{1}{2n}\sum_{i=1}^n \sum_{\ell=1}^v M_{i,q} \left| \widetilde{Y}_{i\ell} - \sum_{k=1}^v b_{k\ell}^{(q)} \int_0^1 X_i(s)\widehat{x}_{k\ell}(s)ds \right|^2 + \frac{\lambda_q}{2}\sum_{k,\ell=1}^v b_{k\ell}^{(q)^2}\, \widehat{\rho}_{k\ell} \right\} \\ &= \underset{\{b_{\ell}^{(q)}\}}{\arg\min} \left\{ \frac{1}{2n}\sum_{\ell=1}^v (\widetilde{Y}_\ell - \Omega_\ell\, b_\ell^{(q)})^{\mathrm{T}}\, \widetilde{M}_q\, (\widetilde{Y}_\ell - \Omega_\ell\, b_\ell^{(q)}) + \frac{\lambda_q}{2}\sum_{\ell=1}^v b_\ell^{(q)^{\mathrm{T}}}\, \widehat{\Lambda}_\ell\, b_\ell^{(q)} \right\}, \end{aligned} \tag{5.2}$$

where we write $b_\ell^{(q)} = (b_{1\ell}^{(q)}, \ldots, b_{v\ell}^{(q)})^{\mathrm{T}} \in \mathbb{R}^v$. By direct calculations, for $1 \leq \ell \leq v$, we have

$$\widehat{b}_\ell^{(q)} = (\Omega_\ell^{\mathrm{T}}\, \widetilde{M}_q\, \Omega_\ell + n\lambda_q\widehat{\Lambda}_\ell)^{-1}\Omega_\ell^{\mathrm{T}}\, \widetilde{M}_q\, \widetilde{Y}_\ell\,, \tag{5.3}$$

so that we can approximate $\widehat{\beta}^*_{n,q}$ in (4.3) by

$$\widetilde{\beta}^*_{n,q} = \sum_{\ell=1}^v (\widehat{b}_\ell^{(q)^{\mathrm{T}}}\, \widehat{x}_\ell) \otimes \eta_\ell\,,$$

if we let $\widehat{x}_\ell = (\widehat{x}_{1\ell}, \ldots, \widehat{x}_{v\ell})^{\mathrm{T}}$ denote a function-valued v-dimensional vector. We propose to use generalized cross-validation (GCV, see, for example, Wahba, 1990) to choose the smoothing parameter λ_q in (5.2). For the q-th bootstrap estimator $\widetilde{\beta}^*_{n,q}$, we choose λ_q that minimizes the GCV score

$$\mathrm{GCV}(\lambda_q) = \frac{n^{-1}\sum_{\ell=1}^{v} \|\widehat{Y}_{\ell,q} - \widetilde{Y}_\ell\|_2^2}{\{1 - \mathrm{tr}(H_q)/n\}^2},$$

where $\widehat{Y}_{\ell,q} = \Omega_\ell(\Omega_\ell^{\mathrm{T}}\, \widetilde{M}_q\, \Omega_\ell + n\lambda_q \widehat{\Lambda}_\ell)^{-1}\Omega_\ell^{\mathrm{T}}\, \widetilde{M}_q\, \widetilde{Y}_\ell$ and H_q is the so-called hat matrix with $\mathrm{tr}(H_q) = \sum_{\ell=1}^{v} \mathrm{tr}\{\Omega_\ell(\Omega_\ell^{\mathrm{T}}\, \widetilde{M}_q\, \Omega_\ell + n\lambda_q\, \widehat{\Lambda}_\ell)^{-1}\, \Omega_\ell^{\mathrm{T}}\, \widetilde{M}_q\}$.

The statistical inference methods in Section 4 rely on the parameters a and D, and we propose to estimate these two parameters from the data. To achieve this, we make use of the growing rate of the eigenvalues of the integro-differential equation (5.1). As indicated by Proposition B.1 in Section B.2 of the inline supplement, the $\rho_{k\ell}$'s diverge at a rate of $(k\ell)^{2D}$, so that we exploit the linear relationship between $\log(\rho_{k\ell})$ and $\log(k\ell)$. Specifically, for the empirical eigenvalues $\{\widehat{\rho}_{k\ell}\}$ of equation (5.1), we fit a line through the points $\{(\log(k\ell), \log(\widehat{\rho}_{k\ell}))\}_{1\le k,\ell\le 2v}$, and take $\widetilde{D}$ to be the value of the slope of this line divided by 2, where we use a total number of $4v^2$ eigenvalues. In the case of $m = 2$, by Proposition B.1, $D \ge 3$ and $a = D - 2$, so that we take

$$\widehat{D} = \max\{\widetilde{D}, 3\} \quad \text{and} \quad \widehat{a} = \widehat{D} - 2\,.$$

5.2 Simulated data

For evaluating the functions X and Y on their domain $[0, 1]$ we take 100 equally spaced time points. For the data generating process (DGP), we used the following three settings:

(1) Let $f_1(s) \equiv 1$, $f_{j+1}(s) = \sqrt{2}\cos(j\pi s)$, for $j \ge 1$, and define

$$\beta_0(s,t) = f_1(s)f_1(t) + 4\sum_{j=2}^{50}(-1)^{j+1}j^{-2}f_j(s)f_j(t)\,.$$

Let $X_i = \sum_{j=1}^{50} j^{-1}Z_{ij}\, f_j$, where $Z_{ij} \overset{\text{iid}}{\sim} \mathrm{unif}(-\sqrt{3}, \sqrt{3})$, for $1 \le i \le n, 1 \le j \le 50$.

(2) Let $\beta_0(s,t) = e^{-(s+t)}$; the X_i's are the same as DGP 1.

(3) Let $f_1(s) \equiv 1$, $f_{j+1}(s) = \sqrt{2}\cos(j\pi s)$ and $g_{j+1}(s) = \sqrt{2}\{1 + \cos(j\pi s)\}$, for $j \ge 1$ and define

$$\beta_0(s,t) = f_1(s)f_1(t) + 4\sum_{j=2}^{50}(-1)^{j+1}j^{-2}g_j(s)f_j(t)\,;$$

the X_i's are the same as DGP 1.

The first setting is similar to the ones used in Yuan and Cai (2010) and Sun et al. (2018); the second setting is exactly the same as Scenario 1 in Sun et al. (2018); the third setting is a non-standard setting that involves an asymmetric slope surface β_0. We took ε to be the Gaussian process with the following three covariance settings:

(i) For $t_1, t_2 \in [0,1]$, $C_\varepsilon(t_1,t_2) = \sigma_1^2\delta(t_1,t_2)$, where $\sigma_1^2 = 0.1 \times \int \mathrm{var}\{\widetilde{Y}(t)\}dt$ and $\widetilde{Y}(t) = \int_0^1 \beta_0(s,t)X(s)ds$, for $t \in [0,1]$.

(ii) For $t_1, t_2 \in [0,1]$, $C_\varepsilon(t_1,t_2) = \sigma_2^2(t)\delta(t_1,t_2)$, where $\sigma_2^2(t) = 0.1 \times \mathrm{var}\{Y(t)\}$, for $t \in [0,1]$.

(iii) For $t_1, t_2 \in [0,1]$, $C_\varepsilon(t_1,t_2) = 2\sigma_1^2\delta(t_1,t_2)$, where σ_1^2 is as in (i).

For each above setting, we simulated 1000 Monte Carlo samples, each of size $n = 30$ or 60, and we took the bootstrap sample size $Q = 300$. We compared our method (hereinafter referred to as RK) with the tensor product reproducing kernel Hilbert space method proposed in Sun et al. (2018) (hereinafter referred to as TP). For our method, we took the number of components $v = \lceil n^{2/5} \rceil$ in Section 5.1. To evaluate the performance of different estimators, we considered the following three criteria. The first criterion is the integrated squared error of $\widehat{\beta}$, defined by

$$\mathrm{ISE}(\widehat{\beta}) = \int_0^1 \int_0^1 |\widehat{\beta}(s,t) - \beta_0(s,t)|^2 ds dt.$$

The second criterion is the excess prediction risk $\mathrm{EPR}(\widehat{\beta})$ defined in (4.1). The third criterion is the maximum deviation, defined by

$$\mathrm{MD}(\widehat{\beta}) = \sup_{(s,t)\in[0,1]^2} \left|\widehat{\beta}(s,t) - \beta_0(s,t)\right|.$$

In Table 1, we report the three quartiles of ISE, EPR and MD of the estimators computed from the 1000 Monte Carlo samples under the data generating process 1–3 with error processes (i)–(iii), using our method (RK) and Sun et al. (2018)'s method (TP). Figure 1 displays the plots of the true slope surface β_0 and their corresponding estimators using RK and TP, under the data generating processes 1–3 with error (i) and sample size $n = 60$.

The results in Table 1 indicate that, for DGPs 1 and 3, our method (RK) produces higher estimation accuracy in terms of ISE, EPR and MD compared to Sun et al. (2018)'s method (TP), whereas Sun et al. (2018)'s produces slightly better estimators in DGP 2.

These results are in accordance with the fact that, in contrast to DGPs 1 and 3, the true β_0 in DGP 2 is multiplicatively separable and the approach of Sun et al. (2018) is based on this assumption. However, it is notable that the loss of the RK-method, which does not require this condition, is not substantial. From Table 1, we also notice that, in some cases, both methods perform better in error setting (ii) than in error setting (i). An explanation for this observation is that, the point-wise signal-to-noise ratio is 10 in error setting (i), whereas this value is smaller than 10 for some $t \in [0, 1]$ in setting (ii). As for computation, Sun et al. (2018)'s method involves computing the inverse or the Cholesky decomposition of matrices, whose size are larger than n^2-by-n^2, which means their method could be time consuming for sample size n larger than, say, 100.

We also evaluated the performance of the simultaneous confidence region $\mathfrak{C}^*_{n,Q}$ defined in Algorithm 4.1, using the uniform covering probability

$$\mathrm{UCP}(\mathfrak{C}^*_{n,Q}) = \mathrm{P}\big\{\beta_0(s,t) \in \mathfrak{C}^*_{n,Q}, \text{ for all } (s,t) \in [0,1]^2\big\}.$$

In Table 2 we report the empirical empirical UCP from 1000 simulation runs for data generating processes 1–3 with all error setting (i) and nominal level $\alpha = 0.10$ and 0.05. We observe a reasonable approximation of the confidence level in all cases under consideration. The simultaneous confidence regions for the slope function for the DGPs 1–3 and the error process (i) are displayed in Figure 2.

For the finite sample properties of classical hypothesis tests proposed in Section 4.3, we consider the following hypothesis:

$$H_0 : \beta_0 = 0 \qquad \text{versus} \qquad H_1 : \beta_0 \neq 0\,, \tag{5.4}$$

that is, we put $\beta_* \equiv 0$ in (4.8). We compared the decision rule based on the bootstrap confidence regions defined in (4.9) (denoted by BT) and the penalized likelihood ratio test (PLRT) at (4.12). Here, for the PLRT, in view of (5.2) and (5.3), substituting $\widetilde{M}_q$ by I_n and observing that $\Omega_\ell^{\mathrm{T}}\,\Omega_\ell = nI_v$, the statistic $\mathfrak{L}_n(0) = L_{n,\lambda}(0) - L_{n,\lambda}(\widehat{\beta}_n)$ (defined in equation (4.10)) can be estimated by

$$\widetilde{\mathfrak{L}}_n(0) = \frac{1}{2n}\sum_{\ell=1}^{v} \widetilde{Y}_\ell^{\mathrm{T}}\Omega_\ell\,(nI_v + n\lambda\widehat{\Lambda}_\ell)^{-1}\Omega_\ell^{\mathrm{T}}\,\widetilde{Y}_\ell\,.$$

We took $n = 30$ and 60, and chose the nominal level $\alpha = 0.05$ and used DGPs 1–3 with error settings (i)–(iii). For DGPs 1 and 3, the empirical rejection probabilities are all 1.0 for both methods for all settings. Table 3 displays the empirical rejection probabilities under DGP 2 with error settings (i)–(iii), together with the empirical sizes under H_0 (that is, $\beta_0 = 0$), of both BT and PLRT for the classical hypothesis (5.4) out of 1000 simulation runs. From the results we observe reasonable approximation of both BT and PLRT of the nominal level 0.05 under H_0; BT outperforms PLRT in terms of empirical power, and as expected, the empirical powers increases for larger sample sizes.

Next, we study the finite sample properties of the test (4.22) for the relevant hypotheses

$$H_0: \sup_{(s,t)\in[0,1]^2} |\beta_0(s,t)| \le \Delta \quad \text{versus} \quad H_1: \sup_{(s,t)\in[0,1]^2} |\beta_0(s,t)| > \Delta\,, \tag{5.5}$$

(we put $\beta_* \equiv 0$ in (4.13)), where the nominal level is chosen as $\alpha = 0.05$. We used the data generating processes 1–3 with error setting (i), where the true $\|\beta_0\|_\infty = 6.0$, 1.0 and 11.0 for the three DGPs, respectively. We took the cut-off parameter $c = \|\widehat{\beta}_n\|_\infty/4$ in (4.18), which scales according to the magnitude of $\widehat{\beta}_n$. In Figure 3, we display the empirical rejection probabilities of test (4.13) based on 1000 simulation runs, for the three data generating processes, for different values of Δ in (4.13). The results shown in Figure 3 indicate that, when $\Delta \le d_\infty$, the empirical rejection probabilities are smaller than $\alpha = 0.05$, and when $\Delta > d_\infty$, the rejection probabilities increases towards 1 as Δ increases, which is consistent with our theory.

5.3 Real data example

We applied the new methodology to the Canadian weather data in Ramsay and Silverman (2005), which consists of daily temperature and precipitation at $n = 35$ locations in Canada averaged over 1960 to 1994. In this case, for $1 \le i \le 35$, X_i is the average daily temperature for each day of the year at the i-th location, and Y_i is the base 10 logarithm of the corresponding average precipitation; see Ramsay and Silverman (2005), p. 248. We took the domain of X and Y to be $[0,1]$ with 365 equality spaced time points. The size of the bootstrap sample is $Q = 300$ and the truncation parameter is chosen as $v = \lceil n^{2/5} \rceil = 4$. In Figure 4, we display the estimated slope function β_0 and the 0.95 confidence region, using our method RK. In order to evaluate the prediction accuracy, for both our method RK and Sun et al. (2018)'s method TP, we computed the integrated squared prediction error (ISPE)

DGP	Error	Method	ISE ($\times 10^{-3}$)		EPR ($\times 10^{-4}$)		MD ($\times 10^{-1}$)	
			$n = 30$	$n = 60$	$n = 30$	$n = 60$	$n = 30$	$n = 60$
1	(i)	RK	17.6 [16.4, 20.1]	13.4 [11.6, 16.2]	31.7 [28.0, 35.6]	27.3 [23.7, 31.3]	11.0 [9.32, 12.2]	7.21 [6.23, 8.50]
		TP	73.8 [73.1, 74.6]	53.1 [52.4, 54.3]	87.8 [81.9, 95.2]	61.7 [56.5, 68.8]	13.2 [12.2, 14.3]	11.4 [10.3, 12.5]
	(ii)	RK	17.0 [16.1, 18.5]	12.6 [10.6, 15.1]	28.3 [24.7, 33.0]	25.3 [21.9, 29.6]	10.5 [9.27, 12.0]	6.58 [5.83, 7.59]
		TP	68.7 [68.1, 69.6]	33.0 [32.4, 33.7]	85.9 [80.9, 91.5]	38.0 [34.9, 45.3]	13.0 [12.0, 13.9]	10.4 [9.13, 11.7]
	(iii)	RK	23.2 [19.7, 28.6]	21.1 [17.5, 26.0]	44.7 [38.6, 53.9]	39.0 [34.2, 44.3]	11.7 [9.23, 14.2]	7.51 [6.57, 9.01]
		TP	79.3 [78.0, 80.6]	34.0 [32.8, 37.6]	96.1 [88.9, 112]	57.3 [47.9, 68.4]	18.9 [15.5, 22.2]	11.8 [10.4, 13.5]
2	(i)	RK	1.14 [0.75, 1.74]	0.65 [0.46, 0.88]	7.45 [5.37, 9.76]	4.62 [3.36, 5.61]	0.86 [0.67, 1.13]	0.65 [0.54, 0.85]
		TP	0.87 [0.67, 1.23]	0.38 [0.41, 0.48]	5.67 [4.59, 7.56]	3.96 [3.19, 4.20]	0.45 [0.32, 0.49]	0.34 [0.29, 0.39]
	(ii)	RK	0.32 [0.23, 0.40]	0.23 [0.17, 0.29]	2.29 [1.83, 2.96]	1.97 [1.68, 2.37]	0.53 [0.46, 0.59]	0.51 [0.45, 0.55]
		TP	0.22 [0.18, 0.28]	0.18 [0.11, 0.25]	2.18 [2.09, 2.51]	1.05 [1.96, 2.14]	0.33 [0.32, 0.35]	0.33 [0.32, 0.35]
	(iii)	RK	2.39 [1.49, 3.47]	1.08 [0.75, 1.57]	13.1 [9.38, 17.4]	6.96 [5.09, 8.96]	1.28 [0.95, 1.65]	0.80 [0.65, 1.14]
		TP	1.50 [0.99, 1.95]	0.70 [0.40, 0.86]	7.77 [5.71, 9.87]	5.52 [4.44, 7.52]	0.47 [0.37, 0.68]	0.42 [0.35, 0.51]
3	(i)	RK	28.6 [26.7, 31.6]	15.0 [12.9, 18.6]	37.7 [33.0, 43.6]	31.5 [27.2, 36.2]	12.8 [11.0, 15.6]	8.97 [7.88, 10.2]
		TP	146 [142, 150]	83.8 [81.8, 86.9]	138 [129, 149]	77.7 [71.7, 73.3]	14.5 [11.8, 17.5]	11.3 [9.31, 13.5]
	(ii)	RK	29.1 [27.1, 32.4]	17.7 [14.0, 22.2]	39.7 [33.4, 46.7]	34.6 [29.8, 39.4]	11.9 [10.2, 14.6]	8.81 [7.27, 10.0]
		TP	150 [147, 153]	84.4 [82.6, 88.6]	141 [129, 156]	81.8 [72.4, 92.9]	15.8 [13.1, 19.7]	11.9 [10.0, 14.6]
	(iii)	RK	32.5 [29.2, 39.0]	22.5 [18.4, 26.9]	54.3 [44.3, 62.4]	41.0 [36.3, 48.1]	15.5 [12.9, 18.1]	9.79 [8.27, 11.6]
		TP	170 [166, 174]	86.6 [83.3, 92.5]	148 [134, 163]	94.8 [84.4, 105.1]	23.5 [22.2, 24.9]	13.8 [11.6, 16.7]

Table 1: *The three quartiles* (2^{nd} [1^{st}, 3^{rd}]) *of integrated squared error* (ISE), *excess prediction rate* (EPR) *and maximum deviation* (MD) *of estimators computed from 1000 simulation runs under the data generating processes 1–3 with error processes (i)–(iii), using our method (RK) and Sun et al. (2018)'s method (TP).*

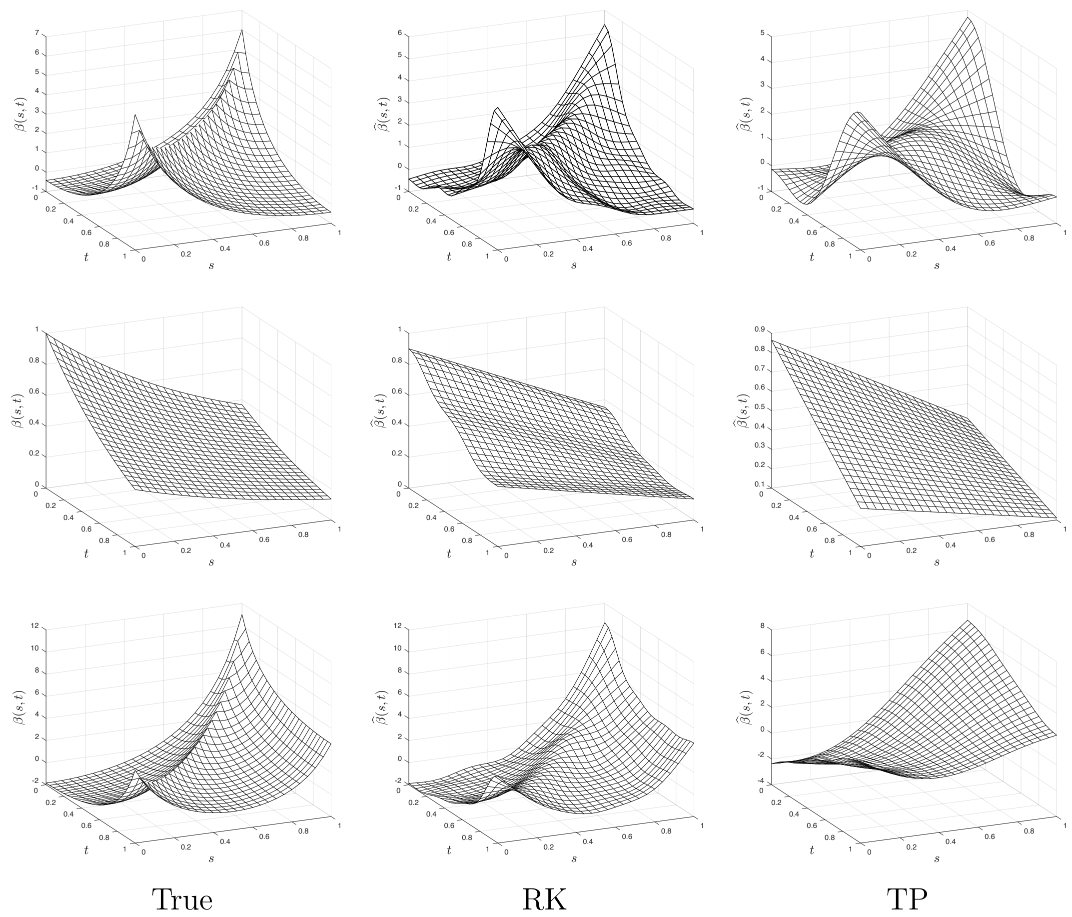

Figure 1: *Plots of the "true" slope surface* β_0 *in model* (2.1) *and the corresponding estimators using our method (RK) and Sun et al. (2018)'s method (TP), under DGP 1–3 (rows 1–3) with error (i) with sample size* $n = 60$*.*

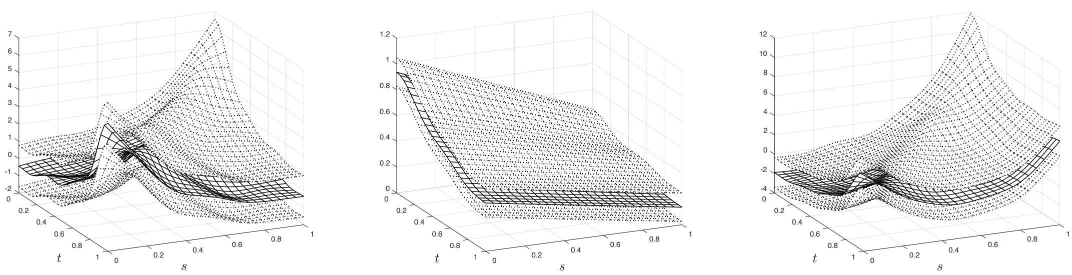

Figure 2: *Simultaneous* 0.95 *confidence regions for the slope surface* β_0 *under DGPs 1–3 with error (i) and sample size* $n = 60$*.*

DGP	1		2		3	
α	0.10	0.05	0.10	0.05	0.10	0.05
$n = 30$	0.863	0.932	0.882	0.963	0.858	0.915
$n = 60$	0.881	0.944	0.913	0.960	0.870	0.939

Table 2: *Empirical covering probabilities of the simultaneous confidence region* (4.5) *for the slope surface* β_0 *under DGPs 1–3 with error setting (i).*

	(i)		(ii)		(iii)	
	$n = 30$	$n = 60$	$n = 30$	$n = 60$	$n = 30$	$n = 60$
BT	0.536	0.693	0.328	0.494	0.478	0.546
	(0.061)	(0.037)	(0.012)	(0.028)	(0.063)	(0.042)
PLRT	0.367	0.562	0.304	0.418	0.330	0.513
	(0.059)	(0.020)	(0.074)	(0.057)	(0.027)	(0.060)

Table 3: *Empirical rejection probabilities under DGP 2, together with empirical sizes (in brackets) of the decision rule based on the bootstrap confidence region* (BT) *in* (4.9) *and the penalized likelihood ratio test* (PLRT) *in* (4.12) *for the classical hypothesis* (5.4) *with error settings (i)–(iii) at nominal level* $\alpha = 0.05$.

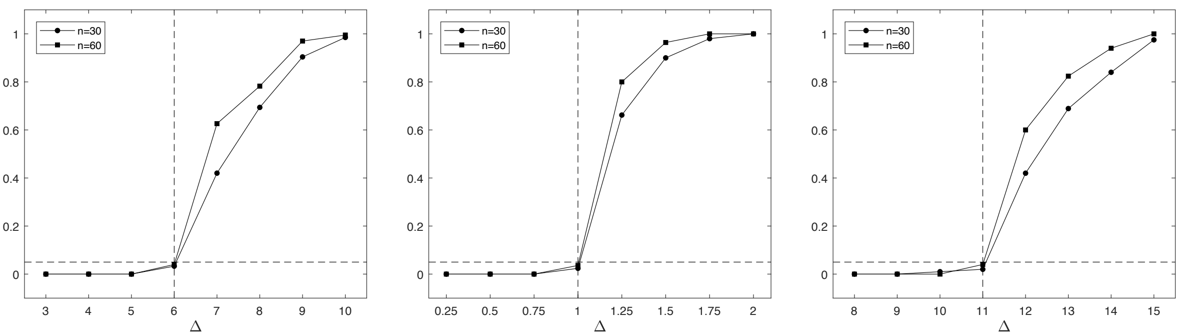

Figure 3: *Empirical rejection probabilities of test* (4.22) *for the relevant hypothesis in* (5.5) *at nominal level* $\alpha = 0.05$*, under DGPs 1–3 with error setting (i) and sample size* $n = 30, 60$*, for different values of* Δ *in* (4.13)*. The horizontal dashed line is the nominal level 0.05; the vertical dashed line is* $\Delta = d_\infty$*.*

and maximum prediction deviation (MPD), for each observation ($1 \le i \le n$), defined by

$$\begin{aligned} \mathrm{ISPE}_i &= \int_0^1 \left| Y_i(t) - \int_0^1 X_i(s)\, \widehat{\beta}_{-i}(s)\, ds \right|^2 dt\,; \\ \mathrm{MPD}_i &= \sup_{t\in[0,1]} \left| Y_i(t) - \int_0^1 X_i(s)\, \widehat{\beta}_{-i}(s)\, ds \right|, \end{aligned} \tag{5.6}$$

where $\widehat{\beta}_{-i}$ is the estimator of the slope function based on the data with the i-th observation removed. In Figure 5, we display the boxplot of $\{\sqrt{\text{ISPE}_i}\}_{i=1}^n$ and $\{\text{MPD}_i\}_{i=1}^n$, for both methods RK and TP. The results in Figure 5 show that, in general, our method performs better in terms of prediction accuracy and robustness, which is indicated by a smaller median, smaller interquartile range in terms of $\sqrt{\text{ISPE}}$ and MPD, and fewer outliers of $\sqrt{\text{ISPE}}$. In contrast, Sun et al. (2018)'s method achieves a smaller minimum value of both $\sqrt{\text{ISPE}}$ and MPD.

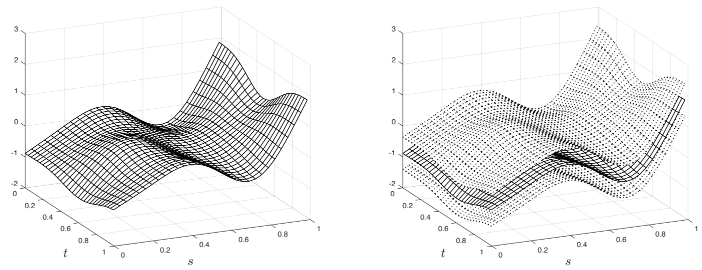

Figure 4: *Estimated slope surface in model* (2.1) *(left panel) and its 0.95 simultaneous band (right panel), using the Canadian weather data.*

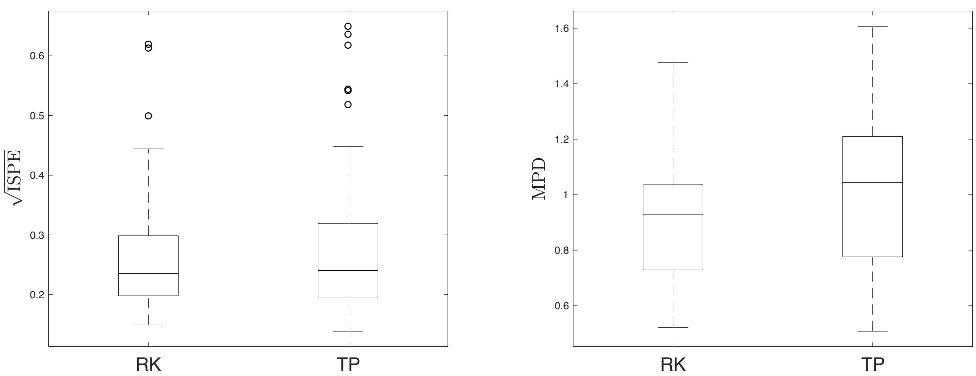

Figure 5: *Boxplot of the square root of the integrated squared prediction error* $\sqrt{\text{ISPE}}$ *(left penal), and the maximum prediction deviation* (MPD) *(right penal) defined in* (5.6) *using our method RK and Sun et al. (2018)'s method TP.*

Acknowledgements

This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1,C1) of the German Research Foundation (DFG). The authors would like to thank Martina Stein, who typed parts of this paper with considerable expertise. The Canadian weather data are available in Ramsay et al. (2014)'s `fda` R package.

References

Aitchison, J. (1964). Confidence-region tests. *J. R. Stat. Soc. Ser. B*, **26**, 462–476.

Aue, A., Rice, G. and Sönmez, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *J. R. Stat. Soc. Ser. B*, **80**, 509–529.

Benatia, D., Carrasco, M. and Florens, J. P. (2017). Functional linear regression with functional response. *J. Econometrics*, **201**, 269–291.

Bosq, D. (2000). Linear Processes in Function Spaces: Theory and Applications. *Lecture Notes in Statistics. Springer New York.*

Cai, T. T. and Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *J. Amer. Statist. Assoc.*, **107**, 1201–1216.

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

Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Stat.*, **30**, 241–255.

Cardot, H., Goia, A. and Sarda, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Comm. Statist. Simulation Comput.*, **33**, 179–199.

Dette, H. and Kokot, K. (2021a). Bio-equivalence tests in functional data by maximum deviation. *Biometrika*, to appear.

Dette, H. and Kokot, K. (2021b). Detecting relevant differences in the covariance operators of functional time series–a sup-norm approach. *Ann. Inst. Statist. Math.*, to appear.

Dette, H., Kokot, K. and Aue, A. (2020). Functional data analysis in the Banach space of continuous functions. *Ann. Stat.*, **48**, 1168–1192.

Ferraty, F. and P. Vieu (2010). *Nonparametric Functional Data Analysis.* Springer-Verlag, New York.

Fogarty, C. B. and Small, D. S. (2014). Equivalence testing for functional data with an application to comparing pulmonary function devices. *Ann. Appl. Stat.*, **8**, 2002–2026.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Stat.*, **35**, 70–91.

Hao, M., Liu, K. Y., Xu, W. and Zhao, X. (2021). Semiparametric inference for the functional Cox model. *J. Amer. Statist. Assoc.*, to appear.

Hilgert, N., Mas, A. and Verzelen, N. (2013). Minimax adaptive tests for the functional linear model. *Ann. Stat.*, **41**, 838–869.

Horváth, L. and Kokoszka, P. (2012). Inference for functional data with applications. *Springer Science & Business Media.*

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to linear Operators.* New York: Wiley.

Imaizumi, M. and Kato, K. (2019). A simple method to construct confidence bands in functional linear regression. *Statist. Sinica*, **29**, 2055–2081.

James, G. M. (2002). Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B*, **64**, 411–432.

Kong, D., Staicu, A. M. and Maity, A (2016). Classical testing in functional linear models. *J. Nonparametr. Stat.,* **28**, 813–838.

Lei, J. (2014). Adaptive global testing for functional linear models. *J. Amer. Statist. Assoc.*, **109**, 624–634.

Li, T. and Zhu, Z. (2020). Inference for generalized partial functional linear regression. *Statist. Sinica*, **30**, 1379–1397.

Lian, H. (2007). Nonlinear functional models for functional responses in reproducing kernel Hilbert spaces. *Can. J. Stat.*, **35**, 597–606.

Lian, H. (2015). Minimax prediction for functional linear regression with functional responses in reproducing kernel Hilbert spaces. *J. Multivar. Anal.*, **140**, 395–402.

Luo, R. and Qi, X. (2017). Function-on-function linear regression by signal compression. *J. Amer. Statist. Assoc.*, **112**, 690–705.

Meister, A. (2011). Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Ann. Stat.*, **39**, 1471–1495.

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

Qu, S. and Wang, X. (2017). Optimal global test for functional regression. *arXiv preprint arXiv:1710.02269.*

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis.* New York: Springer-Verlag.

Ramsay, J. O., Wickham, H., Graves, S. and Hooker, G. (2014). fda: Functional data analysis. R package version 2.4.3. `http://CRAN.R-project.org/package=fda`.

Shang, Z. and Cheng, G. (2015). Nonparametric inference in generalized functional linear models. *Ann. Stat.*, **43**, 1742–1773.

Scheipl, F. and Greven, S. (2016). Identifiability in penalized function-on-function regression models. *Electron. J. Stat.*, **10**, 495–526.

Sun, X., Du, P., Wang, X. and Ma, P. (2018). Optimal penalized function-on-function regression under a reproducing kernel Hilbert space framework. *J. Amer. Statist. Assoc.*, **113**, 1601–1611.

Tekbudak, M. Y., Alfaro-Córdoba, M., Maity, A. and Staicu, A. M. (2019). A comparison of testing methods in scalar-on-function regression. *AStA Adv. Stat. Anal.*, **103**, 411–436.

Wahba, G. (1990). *Spline Models for Observational Data.* Society for industrial and applied mathematics.

Wellner, J. A. (2003). Gaussian white noise models: some results for monotone functions. *Institute of Mathematical Statistics Lecture Notes–Monograph Series*, 87–104.

Wood, S. N. (2003). Thin plate regression splines. *J. R. Stat. Soc. Ser. B*, **65**, 95–114.

Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Stat.*, **33**, 2873–2903.

Yuan, M. and Cai, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Stat.*, **38**, 3412–3444.

Supplementary material for "Statistical inference for function-on-function linear regression"

Holger Dette, Jiajun Tang
Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany

In this supplementary material we provide technical details of our theoretical results. In Section A we provide the proofs of our theorems in our main article. In Section B.1 we provide supporting lemmas that are used in the proofs in Section A. Section B.2 provides a concrete example that satisfies Assumption A2. In the sequel, we use c to denote a generic positive constant that might differ from line to line.

A Proofs of main results

A.1 Proof of Proposition 2.1

To begin with, by Mercer's theorem and Assumption A1, there exists a series of positive real values $\{v_j\}_{j=1}^{\infty}$ and orthogonal basis functions $\{\gamma_j(s)\}_{j=1}^{\infty}$ of $L^2([0,1])$ such that

$$C_X(s_1,s_2)=\sum_{j=1}^{\infty}v_j\gamma_j(s_1)\gamma_j(s_2)\,, \tag{A.1}$$

where $v_j\geq v_{j+1}>0$, for $j\geq 1$. For any $\beta\in\mathcal{H}$ such that $\|\beta\|_K^2=0$, for the γ_j in (A.1) and for $t\in[0,1]$, let $c_{\beta,j}(t)=\int_0^1\beta(s,t)\gamma_j(s)ds$, so that $\beta(s,t)=\sum_{j=1}^{\infty}\gamma_j(s)c_{\beta,j}(t)$. Since $\{\gamma_j\}_{j=1}^{\infty}$ is the orthogonal basis of $L^2([0,1])$, we have

$$\|\beta\|_{L^2}^2=\left\|\sum_{j=1}^{\infty}\gamma_j(s)c_{\beta,j}(t)\right\|_{L^2}^2=\sum_{j=1}^{\infty}\|\gamma_j\|_{L^2}^2\|c_{\beta,j}\|_{L^2}^2=\sum_{j=1}^{\infty}\|c_{\beta,j}\|_{L^2}^2\,. \tag{A.2}$$

Moreover,

$$\begin{aligned}V(\beta,\beta)&=\int_0^1\int_0^1\int_0^1C_X(s_1,s_2)\Big\{\sum_{j=1}^{\infty}\gamma_j(s_1)c_{\beta,j}(t)\Big\}\Big\{\sum_{j=1}^{\infty}\gamma_j(s_2)c_{\beta,j}(t)\Big\}ds_1ds_2dt\\&=\sum_{j=1}^{\infty}v_j\,\|c_{\beta,j}\|_{L^2}^2\,.\end{aligned} \tag{A.3}$$

By the fact that $v_j>0$ and Assumption A1, $\|\beta\|_K^2=0$ implies that, for any $j\geq 1$, $v_j\int_0^1c_{\beta,j}^2(t)dt=V(\beta,\beta)\leq\|\beta\|_K^2=0$, so that $c_{\beta,j}(t)=0$, and by (A.2) we have $\|\beta\|_{L^2}^2=\sum_{j=1}^{\infty}\|c_{\beta,j}\|_{L^2}^2=0$, which shows that $\|\beta\|_K^2=0$ implies $\beta=0$. Also, both V and J in (2.8) are symmetric bilinear operators. Therefore, $\langle\cdot,\cdot\rangle_K$ is an well-defined inner product.

Next, we show the equivalence of $\|\cdot\|_K$ and $\|\cdot\|_{\mathcal{H}}$. First, for any $\beta \in \mathcal{H}$, by (A.2) and (A.3),

$$V(\beta,\beta) = \sum_{j=1}^{\infty} v_j \|c_{\beta,j}\|_{L^2}^2 \le v_1 \sum_{j=1}^{\infty} \|c_{\beta,j}\|_{L^2}^2 = v_1 \|\beta\|_{L^2}^2 \le c v_1 \|\beta\|_{\mathcal{H}}^2 \,, \tag{A.4}$$

for some $c > 0$. Hence,

$$\|\beta\|_K^2 = V(\beta,\beta) + \lambda J(\beta,\beta) \le (c v_1 + \lambda)\|\beta\|_{\mathcal{H}}^2 \,. \tag{A.5}$$

We proceed to show that there exists a constant $c_0 > 0$ such that $\|\beta\|_{\mathcal{H}}^2 \le c_0 \|\beta\|_K^2$. To achieve this, recall the definition of J in (2.6) and note that $J(\beta,\beta)$ is a semi-norm on $\mathcal{H}$. Let $\mathcal{H}_0 = \{\beta \in \mathcal{H} : J(\beta,\beta) = 0\}$ denote the null space of $J(\beta,\beta)$. It is known that $\mathcal{H}_0$ is a finite-dimensional subspace of $\mathcal{H}$ spanned by the polynomials of total degree $\le m-1$, and $m_0 := \dim\{\mathcal{H}_0\} = (m+1)m/2$; see Wahba (1990). Let $\{\xi_1, \ldots, \xi_{m_0}\}$ denote an orthonormal basis of $\mathcal{H}_0$. Let $\mathcal{H}_1 = \{\gamma_1 \in \mathcal{H} : \langle \gamma_1, \gamma_0 \rangle_{\mathcal{H}} = 0,\ \forall \gamma_0 \in \mathcal{H}_0\}$ denote the orthogonal complement of $\mathcal{H}_0$ in $\mathcal{H}$, such that $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, where "$\oplus$" stands for the direct sum. That is, for any $\beta \in \mathcal{H}$, there are unique vectors β_0, β_1 such that

$$\beta = \beta_0 + \beta_1 \,, \quad \beta_0 \in \mathcal{H}_0 \,,\ \beta_1 \in \mathcal{H}_1 \,, \tag{A.6}$$

Here, in view of (2.3), $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product corresponding to $\|\cdot\|_{\mathcal{H}}$ defined by

$$\begin{aligned}\langle \gamma_1, \gamma_2 \rangle_{\mathcal{H}} = &\sum_{0 \le \theta_1+\theta_2 \le m-1} \binom{\theta_1+\theta_2}{\theta_1} \int_{[0,1]^2} \frac{\partial^{\theta_1+\theta_2}\gamma_1}{\partial s^{\theta_1}\partial t^{\theta_2}} dsdt \times \int_{[0,1]^2} \frac{\partial^{\theta_1+\theta_2}\gamma_2}{\partial s^{\theta_1}\partial t^{\theta_2}} dsdt \\ &+ \sum_{\theta_1+\theta_2=m} \binom{m}{\theta_1} \int_{[0,1]^2} \frac{\partial^{m}\gamma_1}{\partial s^{\theta_1}\partial t^{\theta_2}} \times \frac{\partial^{m}\gamma_2}{\partial s^{\theta_1}\partial t^{\theta_2}} dsdt \,, \qquad \text{for } \gamma_1, \gamma_2 \in \mathcal{H} \,.\end{aligned}$$

Due to the fact that $\xi_k \in \mathcal{H}_0$, for $1 \le k \le m_0$, we have

$$0 = \langle \xi_k, \beta_1 \rangle_{\mathcal{H}} = \sum_{0 \le \theta_1+\theta_2 \le m-1} \binom{\theta_1+\theta_2}{\theta_1} \int_{[0,1]^2} \frac{\partial^{\theta_1+\theta_2}\xi_k}{\partial s^{\theta_1}\partial t^{\theta_2}} dsdt \times \int_{[0,1]^2} \frac{\partial^{\theta_1+\theta_2}\beta_1}{\partial s^{\theta_1}\partial t^{\theta_2}} dsdt \,,$$

for $1 \le k \le m_0$. We deduce from the above result that $\int_{[0,1]^2} \frac{\partial^{\theta_1+\theta_2}\beta_1}{\partial s^{\theta_1}\partial t^{\theta_2}} dsdt = 0$, for $0 \le \theta_1 + \theta_2 \le m-1$. In fact, the above argument shows that

$$\mathcal{H}_1 = \left\{ \gamma \in \mathcal{H} : \int_{[0,1]^2} \frac{\partial^{\theta_1+\theta_2}\beta_1}{\partial s^{\theta_1}\partial t^{\theta_2}} dsdt = 0,\ 0 \le \theta_1+\theta_2 \le m-1 \right\} .$$

Therefore,

$$\|\beta_1\|_{\mathcal{H}}^2 = J(\beta_1,\beta_1) \le \lambda^{-1} V(\beta,\beta) + J(\beta_1,\beta_1) = \lambda^{-1}\|\beta\|_K^2 \,. \tag{A.7}$$

It then suffice to show that $\|\beta_0\|_{\mathcal{H}}^2 \le c_0 \|\beta\|_K^2$ for some $c_0 > 0$. Since $\beta_0 \in \mathcal{H}_0$, we have $J(\beta_0, \beta_0) = J(\beta_0, \beta_1) = 0$, so that in view of (A.7),

$$\begin{aligned}\|\beta\|_K^2 &= V(\beta_0 + \beta_1, \beta_0 + \beta_1) + \lambda J(\beta_1, \beta_1) \\ &= V(\beta_0, \beta_0) + 2V(\beta_0, \beta_1) + V(\beta_1, \beta_1) + \lambda \|\beta_1\|_{\mathcal{H}}^2 \,. \end{aligned} \tag{A.8}$$

Since $V(\cdot, \cdot)$ is an inner product, by the Cauchy-Schwarz inequality,

$$|V(\beta_0, \beta_1)| \le \{V(\beta_0, \beta_0)\}^{1/2} \{V(\beta_1, \beta_1)\}^{1/2} \tag{A.9}$$

Next, we examine the connection between $\|\beta_1\|_{\mathcal{H}}^2$ and $V(\beta_1, \beta_1)$. It is known that both $\mathcal{H}_0$ and $\mathcal{H}_1$ are reproducing kernel Hilbert spaces with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ restricted to $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Let $C_1\{(s_1, t_1), (s_2, t_2)\}$ denote the reproducing kernel of $\mathcal{H}_1$. It is known that C_1 is continuous and square-integrable on $[0, 1]^2 \times [0, 1]^2$; see, for example, Section 2.4 in Wahba (1990) and Section 4.3.2 in Gu (2013). Hence, by Mercer's theorem, C_1 admits the following eigen-decomposition:

$$C_1\{(s_1, t_1), (s_2, t_2)\} = \sum_{j=1}^{\infty} \zeta_j \, \chi_j(s_1, t_1) \chi_j(s_2, t_2) \,,$$

where $\zeta_j \ge \zeta_{j+1} \ge 0$, for $j \ge 1$, $\{\chi_j\}_{j \ge 1}$ forms an orthonormal basis of $L^2([0, 1]^2)$, and $s_1, s_2, t_1, t_2 \in [0, 1]$. Note that

$$\langle \chi_j, \chi_\ell \rangle_{L^2} = \delta_{j\ell} \,, \qquad \langle \chi_j, \chi_\ell \rangle_{\mathcal{H}} = \zeta_j^{-1} \delta_{j\ell} \,,$$

where $\delta_{j\ell}$ is the Kronecker delta; see, for example, Cucker and Smale (2001) and Yuan and Cai (2010). For β_1 in (A.6), we have $\beta_1(s, t) = \sum_{j=1}^{\infty} \langle \beta_1, \chi_j \rangle_{L^2} \chi_j(s, t)$, so that

$$\begin{aligned}\|\beta_1\|_{\mathcal{H}}^2 &= \sum_{j=1}^{\infty} \langle \beta_1, \chi_j \rangle_{L^2}^2 \|\chi_j\|_{\mathcal{H}}^2 = \sum_{j=1}^{\infty} \zeta_j^{-1} \langle \beta_1, \chi_j \rangle_{L^2}^2 \\ &\ge \zeta_1^{-1} \sum_{j=1}^{\infty} \langle \beta_1, \chi_j \rangle_{L^2}^2 = \zeta_1^{-1} \|\beta_1\|_{L^2}^2 \,. \end{aligned} \tag{A.10}$$

In view of (A.2) and (A.3),

$$V(\beta_1, \beta_1) = \sum_{j=1}^{\infty} v_j \|c_{\beta_1, j}\|_{L^2}^2 \le v_1 \sum_{j=1}^{\infty} \|c_{\beta_1, j}\|_{L^2}^2 = v_1 \, \|\beta_1\|_{L^2}^2 \,.$$

Combining the above equation with (A.10) yields that

$$\|\beta_1\|_{\mathcal{H}}^2 \ge \zeta_1^{-1} v_1^{-1} V(\beta_1, \beta_1) \,.$$

Therefore, combining the above equation with (A.8) and (A.9), we find that

$$\begin{aligned}\|\beta\|_K^2 &\geq V(\beta_0,\beta_0) - 2|V(\beta_0,\beta_1)| + V(\beta_1,\beta_1) + \lambda\|\beta_1\|_{\mathcal{H}}^2 \\ &\geq V(\beta_0,\beta_0) - 2\{V(\beta_0,\beta_0)\}^{1/2}\{V(\beta_1,\beta_1)\}^{1/2} + \left(1+\frac{\lambda}{v_1\zeta_1}\right)V(\beta_1,\beta_1) \\ &= \frac{\lambda}{v_1\zeta_1+\lambda}V(\beta_0,\beta_0) + \left[\sqrt{\frac{v_1\zeta_1}{v_1\zeta_1+\lambda}}\{V(\beta_0,\beta_0)\}^{1/2} - \sqrt{\frac{v_1\zeta_1+\lambda}{v_1\zeta_1}}\{V(\beta_1,\beta_1)\}^{1/2}\right]^2 \\ &\geq \frac{\lambda}{v_1\zeta_1+\lambda}V(\beta_0,\beta_0)\,. \end{aligned} \tag{A.11}$$

Next, we examine the connection between $V(\beta_0,\beta_0)$ and $\|\beta_0\|_{\mathcal{H}}^2$. Since $\beta_0 \in \mathcal{H}_0$ and $\{\xi_j\}_{j=1}^{m_0}$ is an orthonormal basis of $\mathcal{H}_0$ under the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$, we have $\beta_0(s,t) = \sum_{j=1}^{m_0}\langle\beta_0,\xi_j\rangle_{\mathcal{H}}\,\xi_j(s,t)$ and $\|\beta_0\|_{\mathcal{H}}^2 = \sum_{j=1}^{m_0}\langle\beta_0,\xi_j\rangle_{\mathcal{H}}^2$. Note that

$$V(\beta_0,\beta_0) = \sum_{j=1}^{m_0}\sum_{\ell=1}^{m_0}\langle\beta_0,\xi_j\rangle_{\mathcal{H}}\,\langle\beta_0,\xi_\ell\rangle_{\mathcal{H}}\,V(\xi_j,\xi_\ell)\,.$$

Let b denote an $m_0\times 1$ vector whose j-th entry is $\langle\beta_0,\xi_j\rangle_{\mathcal{H}}$, and let V_* denote an $m_0\times m_0$ matrix, whose (j,ℓ)-th entry is

$$V(\xi_j,\xi_\ell) = \int_{[0,1]^3} C_X(s_1,s_2)\xi_j(s_1,t)\xi_\ell(s_2,t)\,ds_1ds_2dt\,.$$

Now, we have $V(\beta_0,\beta_0) = b^{\mathrm{T}}V_*b$ and $\|\beta_0\|_{\mathcal{H}}^2 = \|b\|_2^2$. Due to Assumption A1, the matrix V_* is a positive definite matrix, and therefore admits a singular value decomposition $V_* = U^{\mathrm{T}}DU$, where U is an orthogonal matrix and $W = \mathrm{diag}(d_1,\ldots,d_{m_0})$ is a diagonal matrix with $d_1 \geq \ldots \geq d_{m_0} > 0$. Therefore,

$$V(\beta_0,\beta_0) = b^{\mathrm{T}}U^{\mathrm{T}}WUb \geq d_{m_0}\|Ub\|_2^2 = d_{m_0}\|b\|_2^2 = d_{m_0}\|\beta_0\|_{\mathcal{H}}^2\,.$$

Therefore, combining the above result with (A.11), we find

$$\|\beta_0\|_{\mathcal{H}}^2 \leq d_{m_0}^{-1}V(\beta_0,\beta_0) \leq \frac{v_1\zeta_1+\lambda}{d_{m_0}\lambda}\|\beta\|_K^2\,.$$

Combining the above equation with (A.7) yields that

$$\|\beta\|_{\mathcal{H}}^2 = \|\beta_0\|_{\mathcal{H}}^2 + \|\beta_1\|_{\mathcal{H}}^2 \leq \frac{v_1\zeta_1+\lambda+d_{m_0}}{d_{m_0}\lambda}\|\beta\|_K^2\,. \tag{A.12}$$

This together with (A.5) completes the proof of the equivalence between $\|\cdot\|_{\mathcal{H}}$ and $\|\cdot\|_K$. Since $\mathcal{H}$ is a reproducing kernel Hilbert space equipped with $\|\cdot\|_{\mathcal{H}}$, we therefore deduce that $\mathcal{H}$ equipped with $\|\cdot\|_K$ is a reproducing kernel Hilbert space.

A.2 Proof of Theorem 3.1

We first prove the following lemma, which is useful for proving Theorem 3.1. In this section, without loss of generality, we assume that $\sigma_\varepsilon^2 = 1$ in Assumption A3.

Lemma A.1. *For any $\beta \in \mathcal{H}$, let*

$$g(X_i, \beta) = \tau\left[X_i \otimes \left\{\int_0^1 \beta(s,\cdot)X_i(s)ds\right\}\right], \tag{A.13}$$

$$H_n(\beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[g(X_i,\beta) - \mathrm{E}\{g(X,\beta)\}\right]. \tag{A.14}$$

For $p_n \geq 1$, let

$$\mathcal{F}_{p_n} = \{\beta \in \mathcal{H} : \|\beta\|_{L^2} \leq 1, J(\beta,\beta) \leq p_n\}. \tag{A.15}$$

Then, under Assumptions A1–A4, as $n \to \infty$,

$$\sup_{\beta\in\mathcal{F}_{p_n}} \frac{\|H_n(\beta)\|_K}{p_n^{1/(2m)}\|\beta\|_{L^2}^{(m-1)/m} + n^{-1/2}} = O_p\big(\lambda^{-1/(2D)}\log\log n\big)^{1/2}. \tag{A.16}$$

Proof. We follow the proof of Lemma 3.4 in Shang and Cheng (2015). For the $\{x_{k\ell}\}_{k,\ell\geq 1}$ and $\{\eta_\ell\}_{\ell\geq 1}$ in Assumption A2, and for $1 \leq i \leq n$, let

$$w(X_i) = \|X_i\|_{L^2}\left(\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\langle X_i, x_{k\ell}\rangle_{L^2}^2\,\|\eta_\ell\|_{L^2}^2\right)^{1/2}, \tag{A.17}$$

and let $\mathcal{X}_n = \{w(X_i)\}_{i=1}^n$. By Lemma B.5 in Section B,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\Big\|\big[g(X_i,\beta_1) - \mathrm{E}\{g(X_i,\beta_1)\}\big] - \big[g(X_i,\beta_2) - \mathrm{E}\{g(X_i,\beta_2)\}\big]\Big\|_K\\
&\leq \frac{1}{\sqrt{n}}\left\|\tau\left(X_i \otimes \left[\int_0^1 \{\beta_1(s,\cdot) - \beta_2(s,\cdot)\}X_i(s)ds\right]\right)\right\|_K\\
&\qquad + \frac{1}{\sqrt{n}}\mathrm{E}\left\|\tau\left(X_i \otimes \left[\int_0^1 \{\beta_1(s,\cdot) - \beta_2(s,\cdot)\}X_i(s)ds\right]\right)\right\|_K\\
&\leq \frac{1}{\sqrt{n}}\|\beta_1 - \beta_2\|_{L^2} \times \big[w(X_i) + \mathrm{E}\{w(X_i)\}\big].
\end{aligned}$$

By Theorem 3.5 in Pinelis (1994),

$$\mathrm{P}\left\{\|H_n(\beta_1) - H_n(\beta_2)\|_K \geq x \,\middle|\, \mathcal{X}_n\right\} \leq 2\exp\left(-\frac{x^2}{2\,W_n^2\,\|\beta_1-\beta_2\|_{L^2}^2}\right),$$

where

$$W_n = \frac{1}{\sqrt{n}}\left(\sum_{i=1}^n \big[w(X_i) + \mathrm{E}\{w(X_i)\}\big]^2\right)^{1/2}. \tag{A.18}$$

By Lemma B.6 in Section B, $\mathrm{E}(W_n^2) \le 4\mathrm{E}|w(X)|^2 \le c\,\lambda^{-1/(2D)}$. Let

$$\|Z\|_\Psi = \inf\big\{c > 0 : \mathrm{E}\{\Psi(|Z|/c)|\mathcal{X}_n\} \le 1\big\}$$

denote the Orlicz norm of a random variable Z conditional on $\mathcal{X}_n$, with $\Psi(x) = \exp(x^2) - 1$, then, by Lemma 8.1 in Kosorok (2007),

$$\Big\|\|H_n(\beta_1) - H_n(\beta_2)\|_K\Big\|_\Psi \le \sqrt{6}\,W_n\,\|\beta_1 - \beta_2\|_{L^2}\,.$$

Let $N(\delta, \mathcal{F}_{p_n}, \|\cdot\|_{L^2})$ denote δ-covering number of the class $\mathcal{F}_{p_n}$ in (A.15) w.r.t. the $L^2([0,1]^2)$-norm. Since $p_n \ge 1$ for n large enough and $J(p_n^{1/2}\beta, p_n^{1/2}\beta) = p_n J(\beta,\beta)$, we have $\mathcal{F}_{p_n} \subset p_n^{1/2}\mathcal{F}_1$. Hence,

$$\begin{aligned}\log N(\delta, \mathcal{F}_{p_n}, \|\cdot\|_{L^2}) &\le \log N(\delta, p_n^{1/2}\mathcal{F}_1, \|\cdot\|_{L^2})\\ &\le \log N(p_n^{-1/2}\delta, \mathcal{F}_1, \|\cdot\|_{L^2}) \le c\,(p_n^{-1/2}\delta)^{-2/m}\,,\end{aligned}$$

where in the last step we used the result in Birman and Solomjak (1967). By Lemma 8.2 and Theorem 8.4 in Kosorok (2007), we have

$$\begin{aligned}&\Bigg\|\sup_{\beta_1,\beta_2\in\mathcal{F}_{p_n},\,\|\beta_1-\beta_2\|_{L^2}\le\delta}\|H_n(\beta_1) - H_n(\beta_2)\|_K\Bigg\|_\psi\\ &\le c\,W_n\bigg[\int_0^\delta \sqrt{\log\{1 + N(\eta, \mathcal{F}_{p_n}, \|\cdot\|_{L^2})\}}\,d\eta + \delta\sqrt{\log\{1 + N^2(\delta, \mathcal{F}_{p_n}, \|\cdot\|_{L^2})\}}\bigg]\\ &\le c_1 W_n\,p_n^{1/(2m)}\delta^{1-1/m}\,,\end{aligned}$$

for some absolute constant $c_1 > 0$. Since $H_n(0) = 0$, by Lemma 8.1 in Kosorok (2007),

$$\mathrm{P}\left\{\sup_{\beta\in\mathcal{F}_{p_n},\,\|\beta\|_{L^2}\le\delta}\|H_n(\beta)\|_K \ge x\,\big|\mathcal{X}_n\right\} \le 2\exp\big(-c_1^{-2}W_n^{-2}p_n^{-1/m}\delta^{-2+2/m}x^2\big)\,.$$

Taking $\gamma = 1 - 1/m$, $b_n = \sqrt{n}\,p_n^{1/(2m)}$, $\theta_n = b_n^{-1}$, $Q_n = \lceil -\log_2\theta_n + \gamma - 1\rceil$ and $T_n = c_2(\lambda^{-1/(2D)}\log\log n)^{1/2}$, for some constant $c_2 > 0$ to be specified below, yields that

$$\begin{aligned}&\mathrm{P}\left\{\sup_{\beta\in\mathcal{F}_{p_n},\,\|\beta\|_{L^2}\le 2}\frac{\sqrt{n}\|H_n(\beta)\|_K}{b_n\|\beta\|_{L^2}^\gamma + 1} \ge T_n\,\big|\mathcal{X}_n\right\}\\ &\le \mathrm{P}\left\{\sup_{\beta\in\mathcal{F}_{p_n},\,\|\beta\|_{L^2}\le\theta_n^{1/\gamma}}\sqrt{n}\|H_n(\beta)\|_K \ge T_n\,\big|\mathcal{X}_n\right\}\\ &\quad + \sum_{j=0}^{Q_n}\mathrm{P}\left\{\sup_{\beta\in\mathcal{F}_{p_n},\,(\theta_n 2^j)^{1/\gamma}\le\|\beta\|_{L^2}\le(\theta_n 2^{j+1})^{1/\gamma}}\frac{\sqrt{n}\|H_n(\beta)\|_K}{b_n\|\beta\|_{L^2}^\gamma + 1} \ge T_n\,\big|\mathcal{X}_n\right\}\end{aligned}$$

$$\leq \mathrm{P}\left\{ \sup_{\beta\in\mathcal{F}_{p_n},\,\|\beta\|_{L^2}\leq\theta_n^{1/\gamma}} \sqrt{n}\|H_n(\beta)\|_K \geq T_n \,\big|\mathcal{X}_n \right\}$$
$$+\sum_{j=0}^{Q_n}\mathrm{P}\left\{ \sup_{\beta\in\mathcal{F}_{p_n},\,\|\beta\|_{L^2}\leq(\theta_n 2^{j+1})^{1/\gamma}} \sqrt{n}\|H_n(\beta)\|_K \geq (b_n\theta_n 2^j+1)T_n \,\big|\mathcal{X}_n \right\}$$
$$\leq 2\exp\big(-c_1^{-2}W_n^{-2}p_n^{-1/m}\theta_n^{(-2+2/m)/\gamma}n^{-1}T_n^2\big)$$
$$+2\sum_{j=0}^{Q_n}\exp\Big\{-c_1^{-2}W_n^{-2}p_n^{-1/m}(\theta_n 2^{j+1})^{(-2+2/m)/\gamma}(b_n\theta_n 2^j+1)^2 n^{-1}T_n^2\Big\}$$
$$\leq 2\exp\big(-c_1^{-2}W_n^{-2}T_n^2\big)+2(Q_n+1)\exp\big(-c_1^{-2}W_n^{-2}T_n^2/4\big)$$
$$\leq 2(Q_n+2)\exp\big(-c_1^{-2}W_n^{-2}T_n^2/4\big)\,. \tag{A.19}$$

For the W_n^2 in (A.18), denote the event $\mathcal{A}_n=\{W_n^2\leq c_3\lambda^{-1/(2D)}\}$ for some constant $c_3>0$. Since $\mathrm{E}(W_n^2)\leq c\lambda^{-1/(2D)}$, we have that, for c_3 large enough, $\mathrm{P}(\mathcal{A}_n)$ tends to one. On the event $\mathcal{A}_n$, by taking $c_2>2c_1c_3^{-1/2}$, as $n\to\infty$,

$$2(Q_n+2)\exp\big(-c_1^{-2}W_n^{-2}T_n^2/4\big)\leq 2(Q_n+2)\exp\big(-c_1^{-2}c_2^2c_3\log\log n/4\big)=o(1)\,,$$

which together with (A.19) completes the proof of Lemma A.1.

□

Next, we prove the following lemma regarding the convergence rate of $\widehat{\beta}_n$, which is useful to prove Theorem 3.1.

Lemma A.2. *Under Assumptions A1–A5, for any $\beta_0\in\mathcal{H}$, we have*

$$\|\widehat{\beta}_n-\beta_0\|_K=O_p(\lambda^{1/2}+n^{-1/2}\lambda^{-1/(4D)}).$$

Proof. For L_n and $S_{n,\lambda}$ defined in (2.5) and (3.2), let

$$S_n(\beta)=\mathcal{D}L_n(\beta)\,;\quad S(\beta)=\mathrm{E}\{\mathcal{D}L_n(\beta)\}\,;\quad S_\lambda(\beta)=\mathrm{E}\{S_{n,\lambda}(\beta)\}\,. \tag{A.20}$$

In view of (3.2), $S_\lambda(\beta)=S(\beta)+W_\lambda(\beta)$. We first show that there exists a unique element $\beta_\lambda\in\mathcal{H}$ such that $S_\lambda(\beta_\lambda)=0$, and then we prove the upper bound for $\|\beta_\lambda-\beta_0\|_K$. Since

$$\mathcal{D}^2L_\lambda(\beta_0)\beta_1\beta_2=\langle\mathcal{D}S_\lambda(\beta_0)\beta_1,\beta_2\rangle_K=\langle\beta_1,\beta_2\rangle_K,$$

we have $\mathcal{D}S_\lambda(\beta_0)=id$, where id is the identity operator. Since $\mathcal{D}^2S_\lambda$ vanishes, we deduce that $\mathcal{D}S_\lambda(\beta)=id$ for any $\beta\in\mathcal{H}$. Hence, by the mean value theorem, for any $\beta\in\mathcal{H}$, $S_\lambda(\beta)=\beta+S_\lambda(\beta_0)-\beta_0$. Therefore, letting $\beta_\lambda=\beta_0-S_\lambda(\beta_0)$, we find $S_\lambda(\beta_\lambda)=0$, and β_λ

is the unique solution to the estimating equation $S_\lambda(\beta) = 0$. Moreover, since $S(\beta_0) = 0$, we have $S_\lambda(\beta_0) = S(\beta_0) + W_\lambda(\beta_0) = W_\lambda(\beta_0)$. By the Cauchy-Schwarz inequality, for J in (2.6),

$$\begin{aligned} \|\beta_\lambda - \beta_0\|_K &= \|W_\lambda(\beta_0)\|_K = \sup_{\|\gamma\|_K=1} |\langle W_\lambda(\beta_0), \gamma\rangle_K| = \sup_{\|\gamma\|_K=1} \lambda |J(\beta_0,\gamma)| \\ &\le \sup_{\|\gamma\|_K=1} \left\{\sqrt{\lambda J(\beta_0,\beta_0)}\sqrt{\lambda J(\gamma,\gamma)}\right\} \le \sup_{\|\gamma\|_K=1} \left\{\sqrt{\lambda J(\beta_0,\beta_0)}\,\|\gamma\|_K\right\} = \sqrt{\lambda J(\beta_0,\beta_0)}\,. \end{aligned} \tag{A.21}$$

Since $\|\widehat{\beta}_n - \beta_0\|_K \le \|\beta_\lambda - \beta_0\|_K + \|\widehat{\beta}_n - \beta_\lambda\|_K$, we then proceed to show the rate of $\|\widehat{\beta}_n - \beta_\lambda\|_K$. Let $F_n(\beta) = \beta - S_{n,\lambda}(\beta_\lambda + \beta)$. Recall that $\mathcal{D}S_\lambda(\beta_\lambda)\beta = \beta$, so that, for $\mathcal{D}S_{n,\lambda}$ in (3.2),

$$F_n(\beta) = I_{1,n}(\beta) + I_{2,n}(\beta) - S_{n,\lambda}(\beta_\lambda)\,, \tag{A.22}$$

where

$$\begin{aligned} I_{1,n}(\beta) &= -\{S_{n,\lambda}(\beta_\lambda + \beta) - S_{n,\lambda}(\beta_\lambda) - \mathcal{D}S_{n,\lambda}(\beta_\lambda)\beta\}\,, \\ I_{2,n}(\beta) &= -\{\mathcal{D}S_{n,\lambda}(\beta_\lambda)\beta - \mathcal{D}S_\lambda(\beta_\lambda)\beta\}\,. \end{aligned} \tag{A.23}$$

First, for $I_{1,n}(\beta)$ in (A.23), in view of $S_{n,\lambda}$ and $\mathcal{D}S_{n,\lambda}$ defined in (3.2),

$$\begin{aligned} I_{1,n}(\beta) = \frac{1}{n}\sum_{i=1}^n \tau\Bigg(X_i \otimes \bigg[Y_i - \int_0^1 \{\beta_\lambda(s,\cdot) + \beta(s,\cdot)\}X_i(s)ds\bigg]\Bigg) \\ - \frac{1}{n}\sum_{i=1}^n \tau\bigg[X_i \otimes \bigg\{Y_i - \int_0^1 \beta(s,\cdot)X_i(s)ds\bigg\}\bigg] \\ + \frac{1}{n}\sum_{i=1}^n \tau\bigg[X_i \otimes \bigg\{\int_0^1 \beta_\lambda(s,\cdot)X_i(s)ds\bigg\}\bigg] = 0\,. \end{aligned} \tag{A.24}$$

For $I_{2,n}(\beta)$ in (A.23), in view of (3.2),

$$\begin{aligned} \|I_{2,n}(\beta)\|_K &= \big\|\mathcal{D}S_{n,\lambda}(\beta_\lambda)\beta - \mathcal{D}S_\lambda(\beta_\lambda)\beta\big\|_K \\ &= \big\|\mathcal{D}S_n(\beta_\lambda)\beta - \mathcal{D}S(\beta_\lambda)\beta\big\|_K = \frac{1}{\sqrt{n}}\|H_n(\beta)\|_K\,, \end{aligned} \tag{A.25}$$

where $H_n(\beta)$ is defined in (A.14) in Lemma A.1. For a, D in Assumption A2 and c_K in Lemma B.3, let $p_n = c_K^{-2}\lambda^{(2a+1)/(2D)-1}$. In order to apply Lemma A.1, we shall rescale β such that the L^2-norm of its rescaled version is bounded by 1. For the constant c_K in Lemma B.3, let

$$\widetilde{\beta} = \begin{cases} \left(c_K \lambda^{-(2a+1)/(4D)}\|\beta\|_K\right)^{-1}\beta & \text{if } \beta \neq 0\,, \\ 0 & \text{if } \beta = 0\,. \end{cases} \tag{A.26}$$

We have $\|\widetilde{\beta}\|_{L^2} \le c_K \lambda^{-(2a+1)/(4D)} \|\widetilde{\beta}\|_K \le 1$, since $\|\widetilde{\beta}\|_K \le (c_K \lambda^{-(2a+1)/(4D)})^{-1}$ by Lemma B.3. In addition, in view of (2.8), $J(\widetilde{\beta}, \widetilde{\beta}) \le \lambda^{-1} \|\widetilde{\beta}\|_K^2 \le c_K^{-2} \lambda^{(2a+1)/(2D)-1} = p_n$, which implies $\widetilde{\beta} \in \mathcal{F}_{p_n}$. By Lemma A.1, since $n^{-1/2} = o(p_n^{1/(2m)})$ by Assumption A5, we find that, for some constant $c > 0$ large enough, with probability tending to one,

$$\|H_n(\widetilde{\beta})\|_K \le c\big(p_n^{1/(2m)} + n^{-1/2}\big)\big(\lambda^{-1/(2D)} \log\log n\big)^{1/2} \le c\, p_n^{1/(2m)} \lambda^{-1/(4D)} (\log\log n)^{1/2}.$$

Therefore, in view of (A.26), we deduce from the above inequality that, for some constant $c > 0$ large enough, with probability tending to one,

$$\|H_n(\beta)\|_K \le \big(c_K \lambda^{-(2a+1)/(4D)} \|\beta\|_K\big) \|H_n(\widetilde{\beta})\|_K \le c\, p_n^{1/(2m)} \lambda^{-(a+1)/(2D)} (\log\log n)^{1/2} \|\beta\|_K .$$

Recall that $p_n = O(\lambda^{(2a+1)/(2D)-1})$. Therefore, for $I_{2,n}(\beta)$ in (A.23), in view of (A.25), we deduce from the equation result that, for some constant $c > 0$ large enough, with probability tending to one,

$$\begin{aligned} \|I_{2,n}(\beta)\|_K \le n^{-1/2} \|H_n(\beta)\|_K &\le c\, n^{-1/2} p_n^{1/(2m)} \lambda^{-(a+1)/(2D)} (\log\log n)^{1/2} \|\beta\|_K \\ &\le c\, n^{-1/2} \lambda^{-\varsigma} (\log\log n)^{1/2}\, \|\beta\|_K = o(1)\, \|\beta\|_K , \end{aligned} \tag{A.27}$$

where we used Assumption A5 in the last step.

For estimating the remaining term $-S_{n,\lambda}(\beta_\lambda)$ in (A.22), we recall the definition of τ in (2.17) and define $O_i = \tau\big[X_i \otimes \big\{Y_i - \int_0^1 \beta_\lambda(s,\cdot) X_i(s) ds\big\}\big]$, for $1 \le i \le n$. Since $S_\lambda(\beta_\lambda) = 0$, we obtain observing (3.1) that

$$-S_{n,\lambda}(\beta_\lambda) = -\{S_{n,\lambda}(\beta_\lambda) - S_\lambda(\beta_\lambda)\} = n^{-1} \sum_{i=1}^n \{O_i - \mathrm{E}(O_i)\}.$$

Let $\Delta_\lambda \beta = \beta_0 - \beta_\lambda$, so that from (A.21) we obtain $\|\Delta_\lambda \beta\|_K^2 \le c\lambda$ for some constant $c > 0$. We notice that

$$\begin{aligned} &\mathrm{E}\|S_{n,\lambda}(\beta_\lambda)\|_K^2 = n^{-1} \mathrm{E}\|O_i - \mathrm{E}(O_i)\|_K^2 \le n^{-1} \mathrm{E}\|O_i\|_K^2 \\ &= n^{-1} \mathrm{E}\bigg\| \tau\bigg[X_i \otimes \bigg\{Y_i - \int_0^1 \beta_\lambda(s,\cdot) X_i(s) ds\bigg\}\bigg]\bigg\|_K^2 \\ &= n^{-1} \mathrm{E}\bigg\| \tau(X_i \otimes \varepsilon_i) + \tau\bigg[X_i \otimes \bigg\{\int_0^1 \Delta_\lambda \beta(s,\cdot) X_i(s) ds\bigg\}\bigg]\bigg\|_K^2 \\ &\le 2n^{-1} \mathrm{E}\|\tau(X_i \otimes \varepsilon_i)\|_K^2 + 2n^{-1} \mathrm{E}\bigg\| \tau\bigg[X_i \otimes \bigg\{\int_0^1 \Delta_\lambda \beta(s,\cdot) X_i(s) ds\bigg\}\bigg]\bigg\|_K^2 . \end{aligned} \tag{A.28}$$

By Lemma B.4 in Section B, we have

$$\mathrm{E}\bigg\| \tau\bigg[X_i \otimes \bigg\{\int_0^1 \Delta_\lambda \beta(s,\cdot) X_i(s) ds\bigg\}\bigg]\bigg\|_K^2 \le c\, \lambda^{-1/D}\, \|\Delta_\lambda \beta\|_K^2 , \tag{A.29}$$

and Lemmas B.2, B.7 and Assumption A2 give for the first term in (A.28)

$$\mathrm{E}\big\|\tau(X_i\otimes\varepsilon_i)\big\|_K^2=\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\le c\sum_{k,\ell}\frac{1}{1+\lambda(k\ell)^{2D}}\le c\,\lambda^{-1/(2D)}\,.$$

Combining this equality with (A.21), (A.28), (A.29) and Lemma B.7 yields

$$\begin{aligned}\mathrm{E}\|S_{n,\lambda}(\beta_\lambda)\|_K^2&\le c\,n^{-1}\mathrm{E}\big\|\tau(X_i\otimes\varepsilon_i)\big\|_K^2+c\,n^{-1}\lambda^{-1/D}\,\|\Delta_\lambda\beta\|_K^2\\&\le c\,n^{-1}\lambda^{-1/(2D)}+c\,n^{-1}\lambda^{1-1/D}\le c\,n^{-1}\lambda^{-1/(2D)}\,.\end{aligned}\tag{A.30}$$

Let $q_n=2c_0n^{-1/2}\lambda^{-1/(4D)}$ and denote by $\mathcal{B}(r)=\{\gamma\in\mathcal{H},\,\|\gamma\|_K\le r\}$ the $\|\cdot\|_K$-ball with radius $r>0$ in $\mathcal{H}$. In view of (A.27), for any $\beta\in\mathcal{B}(q_n)$, with probability tending to one, $\|I_{2,n}(\beta)\|_K\le\|\beta\|_K/2\le q_n/2$. Therefore, observing (A.22) and (A.24), we obtain for the term $F_n(\beta)$ in (A.22), with probability tending to one, for any $\beta\in\mathcal{B}(q_n)$,

$$\|F_n(\beta)\|_K\le\|I_{2,n}(\beta)\|_K+\|S_{n,\lambda}(\beta_\lambda)\|_K\le c_0\,n^{-1/2}\lambda^{-1/(4D)}+q_n/2\le q_n\,,$$

which implies that $F_n\{\mathcal{B}(q_n)\}\subset\mathcal{B}_n(q_n)$ with probability converging to one. Note that for any $\beta_1,\beta_2\in\mathcal{B}(q_n)$, $F_n(\beta_1)-F_n(\beta_2)=I_{2,n}(\beta_1)-I_{2,n}(\beta_2)$. Due to (A.27), with probability tending to one,

$$\|F_n(\beta_1)-F_n(\beta_2)\|=\|I_{2,n}(\beta_1)-I_{2,n}(\beta_2)\|_K\le\|\beta_1-\beta_2\|_K/2\,,$$

which indicates that F_n is a contraction mapping on $\mathcal{B}(q_n)$. By the Banach contraction mapping theorem with probability converging to one, there exists a unique element $\beta^*\in\mathcal{B}_n$ such that $\beta^*=F_n(\beta^*)=\beta^*-S_{n,\lambda}(\beta_\lambda+\beta^*)$. Letting $\widehat\beta_n=\beta_\lambda+\beta^*$, we have $S_{n,\lambda}(\widehat\beta_n)=0$, which indicates that $\widehat\beta_n$ is the estimator defined by (2.4). In view of (A.21),

$$\|\widehat\beta_n-\beta_0\|_K\le\|\beta_\lambda-\beta_0\|_K+\|\widehat\beta_n-\beta_\lambda\|_K=O_p(\lambda^{1/2}+q_n)=O_p\big(\lambda^{1/2}+n^{-1/2}\lambda^{-1/(4D)}\big)\,.$$

□

Finally, we conclude the proof of Theorem 3.1 using Lemmas A.1 and A.2.

Proof of Theorem 3.1. Let $\beta_\Delta=\widehat\beta_n-\beta_0$. Since $\mathcal{D}^2S_\lambda$ vanishes and $\mathcal{D}S_\lambda(\beta_0)=id$, we have $S_\lambda(\widehat\beta_n)-S_\lambda(\beta_0)=\mathcal{D}S_\lambda(\beta_0)\beta_\Delta=\beta_\Delta$. For S_n and S defined in (A.20), since $S_{n,\lambda}(\widehat\beta_n)=0$,

$$\begin{aligned}\widehat\beta_n-\beta_0+S_{n,\lambda}(\beta_0)&=\beta_\Delta+S_{n,\lambda}(\beta_0)=-S_{n,\lambda}(\widehat\beta_n)+S_{n,\lambda}(\beta_0)+S_\lambda(\widehat\beta_n)-S_\lambda(\beta_0)\\&=-S_n(\widehat\beta_n)+S_n(\beta_0)+S(\widehat\beta_n)-S(\beta_0)\,.\end{aligned}\tag{A.31}$$

Let $r_n=\lambda^{1/2}+n^{-1/2}\lambda^{-1/(4D)}$. For $c_1>0$, denote the event $\mathcal{M}_n=\big\{\|\beta_\Delta\|_K\le c_1r_n\big\}$. From Lemma A.2, we obtain that $\mathrm{P}(\mathcal{M}_n)$ is arbitrarily close to one except for a finite number of n

if $c_1 > 0$ is large enough. For the constant $c_K > 0$ in Lemma B.3, let $q_n = c_1 c_K \lambda^{-(2a+1)/(4D)} r_n$ and let $p_n = c_1^2 q_n^{-2} \lambda^{-1} r_n^2 = c_K^{-2} \lambda^{(-2D+2a+1)/(2D)}$. We have $p_n \geq 1$ for n large enough since $D > a + 1/2$ by Assumption A2. In order to apply Lemma A.1, we shall rescale β_Δ such that the L^2-norm of its rescaled version is bounded by 1. Let $\widetilde{\beta}_\Delta = q_n^{-1}\beta_\Delta$. By Lemma B.3, we have that, on the event $\mathcal{M}_n$,

$$\|\widetilde{\beta}_\Delta\|_{L^2} \leq c_K \lambda^{-(2a+1)/(4D)} \|\widetilde{\beta}_\Delta\|_K \leq c_K q_n^{-1} \lambda^{-(2a+1)/(4D)} \|\beta_\Delta\|_K \leq c_1 c_K q_n^{-1} \lambda^{-(2a+1)/(4D)} r_n \leq 1\,.$$

In addition, since $J(\beta_\Delta, \beta_\Delta) \leq \lambda^{-1}\|\beta_\Delta\|_K^2$, we have

$$J(\widetilde{\beta}_\Delta, \widetilde{\beta}_\Delta) \leq q_n^{-2} J(\beta_\Delta, \beta_\Delta) \leq q_n^{-2}\lambda^{-1}\|\beta_\Delta\|_K^2 \leq c_1^2\, q_n^{-2}\lambda^{-1} r_n^2 = p_n\,.$$

Hence, we have shown that $\widetilde{\beta}_\Delta \in \mathcal{F}_{p_n}$, where the set $\mathcal{F}_{p_n}$ is defined in (A.15).

Recalling the notations (3.2), (A.14) and the identity (A.31), we have

$$\begin{aligned} \big\|\widehat{\beta}_n - \beta_0 + S_{n,\lambda}(\beta_0)\big\|_K &= \big\| - S_n(\widehat{\beta}_n) + S_n(\beta_0) + S(\widehat{\beta}_n) - S(\beta_0)\big\|_K \\ &= n^{-1/2}\|H_n(\beta_\Delta)\|_K\,. \end{aligned} \tag{A.32}$$

Since $\widetilde{\beta}_\Delta \in \mathcal{F}_{p_n}$, by Lemma A.1,

$$\begin{aligned} \|H_n(\widetilde{\beta}_\Delta)\|_K &= O_p\big\{\big(p_n^{1/(2m)} + n^{-1/2}\big)\big(\lambda^{-1/(2D)}\log\log n\big)^{1/2}\big\} \\ &= O_p\big\{p_n^{1/(2m)}\lambda^{-1/(4D)}(\log\log n)^{1/2}\big\}\,. \end{aligned}$$

since $n^{-1/2} = o(p_n^{1/(2m)})$ by Assumption A5. Therefore, we deduce from the above equation that, for some constant $c > 0$ large enough, with probability tending to one,

$$\begin{aligned} n^{-1/2}\|H_n(\beta_\Delta)\|_K &\leq n^{-1/2} q_n\, \|H_n(\widetilde{\beta}_\Delta)\|_K \leq c\, n^{-1/2} q_n\, p_n^{1/(2m)} \lambda^{-1/(4D)} (\log\log n)^{1/2} \\ &\leq c\, n^{-1/2}(\lambda^{-(2a+1)/(4D)}\, r_n)\, \lambda^{(-2D+2a+1)/(4Dm)}\, \lambda^{-1/(4D)} (\log\log n)^{1/2} \\ &= c\, n^{-1/2}\lambda^{-\varsigma}(\lambda^{1/2} + n^{-1/2}\lambda^{-1/(4D)})(\log\log n)^{1/2}\,, \end{aligned} \tag{A.33}$$

where $\varsigma > 0$ is the constant in Assumption A5. Combining the above result with (A.32) yields that

$$\big\|\widehat{\beta}_n - \beta_0 + S_{n,\lambda}(\beta_0)\big\|_K = O_p\big\{n^{-1/2}\lambda^{-\varsigma}(\lambda^{1/2} + n^{-1/2}\lambda^{-1/(4D)})(\log\log n)^{1/2}\big\}\,,$$

which completes the proof of Theorem 3.1.

□

A.3 Proof of Theorem 3.2

Recall that

$$v_n = n^{-1/2}\lambda^{-\varsigma}(\lambda^{1/2} + n^{-1/2}\lambda^{-1/(4D)})(\log\log n)^{1/2}\,, \tag{A.34}$$

where $\varsigma > 0$ is the constant in Assumption A5, so that by Theorem 3.1, $\|\widehat{\beta}_n - \beta_0 + S_{n,\lambda}(\beta_0)\|_K = O_p(v_n)$. In view of (3.2), $S_{n,\lambda}(\beta_0) = -n^{-1}\sum_{i=1}^n \tau(X_i \otimes \varepsilon_i) + W_\lambda(\beta_0)$, so that

$$\widehat{\beta}_n - \beta_0 = (\widehat{\beta}_n - \beta_0 + S_{n,\lambda}\beta_0) - W_\lambda(\beta_0) + \frac{1}{n}\sum_{i=1}^n \tau(X_i \otimes \varepsilon_i)\,. \tag{A.35}$$

We first denote

$$\sigma_\tau(s,t) = \left\{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2}\varphi_{k\ell}^2(s,t)\right\}^{1/2}, \tag{A.36}$$

so that

$$\frac{\sqrt{n}}{\sigma_\tau(s,t)}\{\widehat{\beta}_n(s,t) - \beta_0(s,t)\} = I_{1,n}(s,t) + I_{2,n}(s,t) + I_{3,n}(s,t)\,,$$

where

$$\begin{aligned} I_{1,n}(s,t) &= \frac{\sqrt{n}}{\sigma_\tau(s,t)}\langle \widehat{\beta}_n - \beta_0 + S_{n,\lambda}\beta_0, K_{(s,t)}\rangle_K\,, \\ I_{2,n}(s,t) &= -\frac{\sqrt{n}}{\sigma_\tau(s,t)} W_\lambda\beta_0(s,t)\,, \\ I_{3,n}(s,t) &= \frac{1}{\sqrt{n}\,\sigma_\tau(s,t)}\sum_{i=1}^n \tau(X_i\otimes\varepsilon_i)(s,t)\,. \end{aligned}$$

By Theorem 3.1, (B.2) and Cauchy-Schwarz inequality,

$$\begin{aligned} |I_{1,n}(s,t)| &= \frac{\sqrt{n}}{\sigma_\tau(s,t)}\left|\left\langle \widehat{\beta}_n - \beta_0 + S_{n,\lambda}\beta_0, K_{(s,t)}\right\rangle_K\right| \\ &\le \frac{\sqrt{n}}{\sigma_\tau(s,t)}\|K_{(s,t)}\|_K \times \|\widehat{\beta}_n - \beta_0 + S_{n,\lambda}(\beta_0)\|_K \\ &\le \frac{\sqrt{n}}{\sigma_\tau(s,t)}\left\{\sum_{k,\ell}\frac{\varphi_{k\ell}^2(s,t)}{1+\lambda\rho_{k\ell}}\right\}^{1/2} \times \|\widehat{\beta}_n - \beta_0 + S_{n,\lambda}(\beta_0)\|_K \\ &\le O(\sqrt{n}\lambda^{(2a+1)/(4D)}) \times O_p(v_n) \times \left\{\sum_{k,\ell}\frac{\|\varphi_{k\ell}\|_\infty^2}{1+\lambda\rho_{k\ell}}\right\}^{1/2} = O_p(\sqrt{n}v_n)\,, \end{aligned} \tag{A.37}$$

where we used Assumption A2 and Lemma B.2, which imply $\sigma_\tau(s,t) \asymp \lambda^{-(2a+1)/(4D)}$.

For the term $I_{2,n}$, in view of (2.15), by the assumption that $\sum_{k,\ell}\rho_{k\ell}^2 V^2(\beta_0,\varphi_{k\ell}) < \infty$, Assumption A2 and Lemma B.2,

$$|I_{2,n}(s,t)| = \frac{\sqrt{n}}{\sigma_\tau(s,t)}|W_\lambda\beta_0(s,t)| = \frac{\sqrt{n}\lambda}{\sigma_\tau(s,t)}\left|\sum_{k,\ell}\frac{\rho_{k\ell}V(\beta_0,\varphi_{k\ell})}{1+\lambda\rho_{k\ell}}\varphi_{k\ell}(s,t)\right|$$
$$\leq \frac{\sqrt{n}\,\lambda}{\sigma_\tau(s,t)}\left\{\sum_{k,\ell}\rho_{k\ell}^2 V^2(\beta_0,\varphi_{k\ell})\right\}^{1/2}\left\{\sum_{k,\ell}\frac{\|\varphi_{k\ell}\|_\infty^2}{(1+\lambda\rho_{k\ell})^2}\right\}^{1/2} \leq c\sqrt{n}\,\lambda = o(1)\,. \qquad \text{(A.38)}$$

Finally, the term $I_{3,n}$ can be estimated as follows. For $1 \leq i \leq n$, let

$$\mathfrak{U}_i = \frac{\tau(X_i\otimes\varepsilon_i)(s,t)}{\sqrt{n}\,\sigma_\tau(s,t)} = \frac{\tau(X_i\otimes\varepsilon_i)(s,t)}{\sqrt{n\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2}\varphi_{k\ell}^2(s,t)}}\,,$$

so that $I_{3,n} = \sum_{i=1}^n \mathfrak{U}_i$. We start by noticing that, Assumption A3 indicates that

$$\mathrm{E}\Big\{\big\langle\tau(X\otimes\varepsilon),\varphi_{k\ell}\big\rangle_{L^2}\Big\} = \mathrm{E}\bigg[\int_{[0,1]^2}\mathrm{E}\{\varepsilon(t)|X\}\varphi_{k\ell}(s,t)X_i(s)dsdt\bigg] = 0\,,$$

so that in view of (2.17), $\mathrm{E}\{\tau(X\otimes\varepsilon)(s,t)\} = \sum_{k,\ell}\mathrm{E}\big(\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}\big)(1+\lambda\rho_{k\ell})^{-1}\varphi_{k\ell}(s,t) = 0$, so that $\mathrm{E}(\mathfrak{U}_i) = 0$. In view of (B.7) in the proof of Lemma B.7 in Section B,

$$\mathrm{E}|\tau(X\otimes\varepsilon)(s,t)|^2 = \mathrm{E}\left(\sum_{k,\ell}\frac{\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}}{1+\lambda\rho_{k\ell}}\varphi_{k\ell}(s,t)\right)^2$$
$$= \sum_{k,k',\ell,\ell'}\frac{\varphi_{k\ell}(s,t)\,\varphi_{k'\ell'}(s,t)}{(1+\lambda\rho_{k\ell})(1+\lambda\rho_{k'\ell'})}\mathrm{E}\Big\{\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}\,\langle X\otimes\varepsilon,\varphi_{k'\ell'}\rangle_{L^2}\Big\} = \sigma_\tau^2(s,t)\,,$$

where σ_τ is defined in (A.36) and we used $\sigma_\tau^2(s,t)$ (B.8) in the last step. By the assumption that $\sigma_\tau^2(s,t) \asymp \lambda^{-(2a+1)/(2D)}$, we deduce that $\mathrm{E}|\tau(X\otimes\varepsilon)(s,t)|^2 = \sigma_\tau^2(s,t) \geq c_0^2\lambda^{-(2a+1)/(2D)}$ for some constant $c_0 > 0$.

To conclude the proof, we shall check that the triangular array of random variables $\{\mathfrak{U}_i\}_{i=1}^n = \{n^{-1/2}\sigma_\tau^{-1}(s,t)\tau(X_i\otimes\varepsilon_i)(s,t)\}_{i=1}^n$ satisfies the Lindeberg's condition. By the Cauchy-Schwarz inequality, for any $e > 0$,

$$\sum_{i=1}^n \mathrm{E}\big[|\mathfrak{U}_i|^2 \times \mathbb{1}\{|\mathfrak{U}_i| > e\}\big]$$
$$= \frac{1}{\sigma_\tau^2(s,t)}\,\mathrm{E}\Big[|\tau(X\otimes\varepsilon)(s,t)|^2 \times \mathbb{1}\Big\{|\tau(X\otimes\varepsilon)(s,t)| \geq e\sqrt{n}\,\sigma_\tau(s,t)\Big\}\Big]$$
$$\leq \frac{1}{\sigma_\tau^2(s,t)}\Big\{\mathrm{E}|\tau(X\otimes\varepsilon)(s,t)|^4\Big\}^{1/2} \times \Big[\mathrm{P}\Big\{|\tau(X\otimes\varepsilon)(s,t)| \geq e\sqrt{n}\,\sigma_\tau(s,t)\Big\}\Big]^{1/2}. \qquad \text{(A.39)}$$

We shall deal with the above moment term and the tail probability separately. In order to find the order of $\mathrm{E}|\tau(X\otimes\varepsilon)(s,t)|^4$, in view of (B.2) and Lemma B.2, by Assumption A2,

$$\sup_{(s,t)\in[0,1]^2}\|K_{(s,t)}\|_K^2 = \sum_{k,\ell}\frac{\|\varphi_{k\ell}\|_\infty^2}{1+\lambda\rho_{k\ell}} \leq c\lambda^{-(2a+1)/(2D)}\,. \qquad \text{(A.40)}$$

Therefore, by Lemma B.7, we find

$$\sup_{(s,t)\in[0,1]^2} \mathrm{E}\big|\tau(X\otimes\varepsilon)(s,t)\big|^4 \le \sup_{(s,t)\in[0,1]^2} \|K_{(s,t)}\|_K^4 \times \mathrm{E}\|\tau(X\otimes\varepsilon)\|_K^4 \le c\lambda^{-(2a+2)/D}. \tag{A.41}$$

In order to find the order of the tail probability $\mathrm{P}\big\{|\tau(X\otimes\varepsilon)(s,t)| \ge e\sqrt{n}\sigma_\tau(s,t)\big\}$, we first show an upper bound of $|\tau(X\otimes\varepsilon)(s,t)|$. To achieve this, in view of (B.7) in the proof of Lemma B.7 in Section B, by Lemma B.3,

$$\begin{aligned}\|\tau(X\otimes\varepsilon)\|_K &= \sup_{\|\gamma\|_K=1} |\langle\tau(X\otimes\varepsilon),\gamma\rangle_K| = \sup_{\|\gamma\|_K=1} |\langle X\otimes\varepsilon,\gamma\rangle_{L^2}| \\ &\le \sup_{\|\gamma\|_K=1} \|\gamma\|_{L^2} \times \|X\otimes\varepsilon\|_{L^2} \le c_K\lambda^{-(2a+1)/(4D)}\|X\|_{L^2}\|\varepsilon\|_{L^2}.\end{aligned}$$

By (A.40) and the Cauchy-Schwarz inequality, we deduce from the above equation that

$$\begin{aligned}\sup_{(s,t)\in[0,1]^2} |\tau(X\otimes\varepsilon)(s,t)| &= \sup_{(s,t)\in[0,1]^2} |\langle K_{(s,t)},\tau(X\otimes\varepsilon)\rangle_K| \le \sup_{(s,t)\in[0,1]^2} \|K_{(s,t)}\|_K \times \|\tau(X\otimes\varepsilon)\|_K \\ &\le c\lambda^{-(2a+1)/(2D)}\|X\|_{L^2}\|\varepsilon\|_{L^2}.\end{aligned} \tag{A.42}$$

Now, by assumption, $\sigma_\tau(s,t) \ge c_0\lambda^{-(2a+1)/(4D)}$ for some $c_0 > 0$, hence for any $e > 0$, we can choose $c_1 > (c_\varepsilon D)^{-1}$, for $c_\varepsilon > 0$ in Assumption A4, so that

$$\begin{aligned}&\mathrm{P}\Big\{|\tau(X\otimes\varepsilon)(s,t)| \ge e\sqrt{n}\,\sigma_\tau(s,t)\Big\} \\ &\le \mathrm{P}\Big(c\lambda^{-(2a+1)/(2D)}\|X\|_{L^2}\|\varepsilon\|_{L^2} \ge ec_0\sqrt{n}\lambda^{-(2a+1)/(4D)}\Big) \\ &= \mathrm{P}\Big(\|X\|_{L^2}\|\varepsilon\|_{L^2} \ge ec^{-1}c_0\sqrt{n}\lambda^{(2a+1)/(4D)}\Big) \\ &\le \mathrm{P}\Big\{\|X\|_{L^2} \ge ec_1^{-1}c^{-1}c_0\sqrt{n}\lambda^{(2a+1)/(4D)}/\log(\lambda^{-1})\Big\} + \mathrm{P}\Big\{\|\varepsilon\|_{L^2} \ge c_1\log(\lambda^{-1})\Big\} \\ &\le \exp\{-c_X ec_1^{-1}c^{-1}c_0\sqrt{n}\lambda^{(2a+1)/(4D)}/\log(\lambda^{-1})\}\mathrm{E}\{\exp(c_X\|X\|_{L^2})\} \\ &\qquad + \lambda^{c_1c_\varepsilon}\mathrm{E}\{\exp(c_\varepsilon\|\varepsilon\|_{L^2})\} \\ &= O\big(\lambda^{c_X ec_1^{-1}c^{-1}c_0\{\sqrt{n}\lambda^{(2a+1)/(4D)}/\log^2(\lambda^{-1})\}}\big) + O(\lambda^{c_1c_\varepsilon}) = o(\lambda^{1/D}),\end{aligned} \tag{A.43}$$

where we used the assumption that $\sqrt{n}\lambda^{(2a+1)/(4D)}/\log^2(\lambda^{-1}) \to \infty$ in the last step. Since, by assumption, $\sigma_\tau^2(s,t) \asymp \lambda^{-(2a+1)/(2D)}$, combining the above equation with (A.39) and (A.41) yields that, for any $e > 0$,

$$\sum_{i=1}^n \mathrm{E}\big[|\mathfrak{U}_i|^2 \times \mathbb{1}\{|\mathfrak{U}_i| > e\}\big] \le \frac{c}{\sigma_\tau^2(s,t)} \times O(\lambda^{-(a+1)/D}) \times o(\lambda^{1/(2D)}) = o(1).$$

Therefore, by Lindeberg's CLT,

$$\frac{1}{\sqrt{n}\,\sigma_\tau(s,t)}\sum_{i=1}^n \tau(X_i\otimes\varepsilon_i)(s,t) \xrightarrow{d.} N(0,1).$$

Combining the above result with (A.35)–(A.38), we deduce that

$$\frac{\sqrt{n}}{\sigma_\tau(s,t)}\big\{\widehat{\beta}_n(s,t)-\beta_0(s,t)\big\} = \frac{1}{\sqrt{n}\,\sigma_\tau(s,t)}\sum_{i=1}^n \tau(X_i\otimes\varepsilon_i)(s,t)+o_p(1)\xrightarrow{d.} N(0,1)\,,$$

which completes the proof.

A.4 Proof of Theorem 3.3

Recall the definition of the process $\mathbb{G}_n$ in (3.8) and note that in view of (A.35),

$$\mathbb{G}_n(s,t) = I_{1,n}(s,t)+I_{2,n}(s,t)+I_{3,n}(s,t)\,, \tag{A.44}$$

where

$$\begin{aligned} I_{1,n}(s,t) &= \sqrt{n}\lambda^{(2a+1)/(4D)}\{\widehat{\beta}_n(s,t)-\beta_0(s,t)+S_{n,\lambda}\beta_0(s,t)\} \\ I_{2,n}(s,t) &= -\sqrt{n}\lambda^{(2a+1)/(4D)}W_\lambda\beta_0(s,t)\,, \\ I_{3,n}(s,t) &= n^{-1/2}\lambda^{(2a+1)/(4D)}\sum_{i=1}^n\tau(X_i\otimes\varepsilon_i)(s,t)\,. \end{aligned} \tag{A.45}$$

In view of (A.37) and Theorem 3.1, for v_n in (A.34),

$$\begin{aligned} \sup_{(s,t)\in[0,1]^2}|I_{1,n}(s,t)| &\le \sqrt{n}\lambda^{(2a+1)/(4D)}\Big\{\sum_{k,\ell}\frac{\|\varphi_{k\ell}\|_\infty^2}{1+\lambda\rho_{k\ell}}\Big\}^{1/2}\|\widehat{\beta}_n-\beta_0+S_{n,\lambda}(\beta_0)\|_K \\ &= O_p(\sqrt{n}v_n)=o_p(1)\,, \end{aligned} \tag{A.46}$$

$$\begin{aligned} \sup_{(s,t)\in[0,1]^2}|I_{2,n}(s,t)| &= \sqrt{n}\lambda^{1+(2a+1)/(4D)}\sup_{(s,t)\in[0,1]^2}\Big|\sum_{k,\ell}\frac{\rho_{k\ell}V(\beta_0,\varphi_{k\ell})}{1+\lambda\rho_{k\ell}}\varphi_{k\ell}(s,t)\Big| \\ &\le \sqrt{n}\lambda^{1+(2a+1)/(4D)}\Big\{\sum_{k,\ell}\rho_{k\ell}^2V^2(\beta_0,\varphi_{k\ell})\Big\}^{1/2}\Big\{\sum_{k,\ell}\frac{\|\varphi_{k\ell}\|_\infty^2}{(1+\lambda\rho_{k\ell})^2}\Big\}^{1/2} \\ &= O(\sqrt{n}\lambda)=o(1)\,. \end{aligned} \tag{A.47}$$

For the term $I_{3,n}$ we use $\mathrm{E}\big\{\tau(X\otimes\varepsilon)(s,t)\big\}=0$ and (B.8) to obtain as $n\to\infty$

$$\begin{aligned} &\operatorname{cov}\big\{I_{3,n}(s_1,t_1),I_{3,n}(s_2,t_2)\big\} \\ &=\lambda^{(2a+1)/(2D)}\operatorname{cov}\big\{\tau(X\otimes\varepsilon)(s_1,t_1),\tau(X\otimes\varepsilon)(s_2,t_2)\big\} \\ &=\lambda^{(2a+1)/(2D)}\,\mathrm{E}\Big[\Big\{\sum_{k,\ell}\frac{\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}}{1+\lambda\rho_{k\ell}}\varphi_{k\ell}(s_1,t_1)\Big\}\Big\{\sum_{k',\ell'}\frac{\langle X\otimes\varepsilon,\varphi_{k'\ell'}\rangle_{L^2}}{1+\lambda\rho_{k'\ell'}}\varphi_{k'\ell'}(s_2,t_2)\Big\}\Big] \\ &=\lambda^{(2a+1)/(2D)}\sum_{k,\ell}\frac{\varphi_{k\ell}(s_1,t_1)\varphi_{k\ell}(s_2,t_2)}{(1+\lambda\rho_{k\ell})^2}=C_Z\{(s_1,t_1),(s_2,t_2)\}+o(1)\,. \end{aligned} \tag{A.48}$$

Note that the results in Section 1.5 in van der Vaart and Wellner (1996) are valid if the $\ell^\infty([0,1]^2)$ space is replaced by $C([0,1]^2)$. We shall prove the weak convergence in $C([0,1]^2)$ through the following two steps.

Step 1. *Weak convergence of the finite-dimensional marginals of* $\mathbb{G}_n$

In order to prove the weak convergence of the finite-dimensional marginal distributions of $\mathbb{G}_n$, by the Cramér-Wold device, we shall show that, for any $q \in \mathbb{N}$, $(c_1, \ldots, c_q)^{\mathrm{T}} \in \mathbb{R}^q$ and $(s_1, t_1), \ldots, (s_q, t_q) \in [0,1]^2$,

$$\sum_{j=1}^{q} c_j \mathbb{G}_n(s_j, t_j) \xrightarrow{d.} \sum_{j=1}^{q} c_j Z(s_j, t_j)\,. \tag{A.49}$$

For $1 \le i \le n$, let $\mathfrak{U}_{i,q} = n^{-1/2}\lambda^{(2a+1)/(4D)} \sum_{j=1}^{q} c_j \tau(X_i \otimes \varepsilon_i)(s_j, t_j)$. In view of (A.46) and (A.47), we deduce that

$$\sum_{j=1}^{q} c_j \mathbb{G}_n(s_j, t_j) = \sum_{j=1}^{q} c_j I_{3,n}(s_j, t_j) + \sum_{j=1}^{q} c_j \{I_{1,n}(s_j, t_j) + I_{2,n}(s_j, t_j)\} = \sum_{i=1}^{n} \mathfrak{U}_{i,q} + o_p(1)\,.$$

By (A.48) and assumption, we find, as $n \to \infty$,

$$\begin{aligned} \operatorname{var}(\mathfrak{U}_{i,q}) &= n^{-1} \sum_{j_1,j_2=1}^{q} c_{j_1} c_{j_2} \operatorname{cov}\big\{\tau(X_i \otimes \varepsilon_i)(s_{j_1}, t_{j_1}), \tau(X_i \otimes \varepsilon_i)(s_{j_2}, t_{j_2})\big\} \\ &= n^{-1} \sum_{j_1,j_2=1}^{q} c_{j_1} c_{j_2} C_Z\{(s_{j_1}, t_{j_1}), (s_{j_2}, t_{j_2})\} + o(n^{-1})\,. \end{aligned}$$

When $\sum_{j_1,j_2=1}^{q} c_{j_1} c_{j_2} C_Z\{(s_{j_1}, t_{j_1}), (s_{j_2}, t_{j_2})\} = 0$, we have that $\sum_{j=1}^{q} c_j Z(s_j, t_j)$ has a degenerate distribution with a point mass at zero, so that (A.49) is followed by the Markov's inequality. When $\sum_{j_1,j_2=1}^{q} c_{j_1} c_{j_2} C_Z\{(s_{j_1}, t_{j_1}), (s_{j_2}, t_{j_2})\} \neq 0$, to prove (A.49), we shall check that the triangular array of random variables $\{\mathfrak{U}_{i,q}\}_{i=1}^{n}$ satisfies Lindeberg's condition. By the Cauchy-Schwarz inequality, we find, for any $e > 0$,

$$\begin{aligned} &\sum_{i=1}^{n} \mathrm{E}\big\{\mathfrak{U}_{i,q}^2\, \mathbb{1}(|\mathfrak{U}_{i,q}| > e)\big\} \\ &= \mathrm{E}\bigg[\lambda^{(2a+1)/(2D)}\bigg|\sum_{j=1}^{q} c_j \tau(X \otimes \varepsilon)(s_j, t_j)\bigg|^2 \times \mathbb{1}\bigg\{\lambda^{(2a+1)/(4D)}\bigg|\sum_{j=1}^{q} c_j \tau(X \otimes \varepsilon)(s_j, t_j)\bigg| \ge e\sqrt{n}\bigg\}\bigg] \\ &\le \lambda^{(2a+1)/(2D)}\bigg\{\mathrm{E}\bigg|\sum_{j=1}^{q} c_j \tau(X \otimes \varepsilon)(s_j, t_j)\bigg|^4\bigg\}^{\frac{1}{2}} \mathrm{P}\bigg\{\lambda^{(2a+1)/(4D)}\bigg|\sum_{j=1}^{q} c_j \tau(X \otimes \varepsilon)(s_j, t_j)\bigg| \ge e\sqrt{n}\bigg\}^{\frac{1}{2}}. \end{aligned}$$

In view of (A.41),

$$\bigg\{\mathrm{E}\bigg|\sum_{j=1}^{q} c_j \tau(X \otimes \varepsilon)(s_j, t_j)\bigg|^4\bigg\}^{\frac{1}{2}} \le c\bigg\{\sup_{(s,t)\in[0,1]^2} \mathrm{E}|\tau(X \otimes \varepsilon)(s, t)|^4\bigg\}^{\frac{1}{2}} \le c\lambda^{-(a+1)/D}\,.$$

Let $\mathfrak{s}_q = \sum_{j=1}^q |c_j|$. We have $\mathfrak{s}_q > 0$, since $\sum_{j_1,j_2=1}^q c_{j_1}c_{j_2}C_Z\{(s_{j_1},t_{j_1}),(s_{j_2},t_{j_2})\} \neq 0$. In view of (A.42), by arguments similar to the ones used in (A.43), we find, by taking $c_1 > (c_\varepsilon D)^{-1}$, for $c_\varepsilon > 0$ in Assumption A4,

$$\begin{aligned}
&\mathrm{P}\bigg\{\lambda^{(2a+1)/(4D)}\bigg|\sum_{j=1}^q c_j\tau(X\otimes\varepsilon)(s_j,t_j)\bigg| \geq e\sqrt{n}\bigg\} \\
&\leq \mathrm{P}\bigg\{\mathfrak{s}_q\lambda^{(2a+1)/(4D)}\sup_{(s,t)\in[0,1]^2}\big|\tau(X\otimes\varepsilon)(s,t)\big| \geq e\sqrt{n}\bigg\} \\
&\leq \mathrm{P}\Big(\|X\|_{L^2}\|\varepsilon\|_{L^2} \geq ec^{-1}\mathfrak{s}_q^{-1}\sqrt{n}\lambda^{(2a+1)/(4D)}\Big) \\
&\leq \mathrm{P}\Big\{\|X\|_{L^2} \geq ec_1^{-1}c^{-1}\mathfrak{s}_q^{-1}\sqrt{n}\lambda^{(2a+1)/(4D)}/\log(\lambda^{-1})\Big\} + \mathrm{P}\Big\{\|\varepsilon\|_{L^2} \geq c_1\log(\lambda^{-1})\Big\} \\
&\leq \exp\{-c_X ec_1^{-1}c^{-1}\mathfrak{s}_q^{-1}\sqrt{n}\lambda^{(2a+1)/(4D)}/\log(\lambda^{-1})\}\mathrm{E}\{\exp(c_X\|X\|_{L^2})\} \\
&\qquad + \lambda^{c_1c_\varepsilon}\mathrm{E}\{\exp(c_\varepsilon\|\varepsilon\|_{L^2})\} \\
&= O\big(\lambda^{c_X ec_1^{-1}c^{-1}\mathfrak{s}_q^{-1}\{\sqrt{n}\lambda^{(2a+1)/(4D)}/\log^2(\lambda^{-1})\}}\big) + O(\lambda^{c_1c_\varepsilon}) = o(\lambda^{1/D}),
\end{aligned}$$

Therefore, for any $e > 0$,

$$\sum_{i=1}^n \mathrm{E}\big\{\mathfrak{U}_{i,q}^2\,\mathbb{1}(|\mathfrak{U}_{i,q}| > e)\big\} \leq c\lambda^{(2a+1)/(2D)} \times \lambda^{-(a+1)/D} \times o(\lambda^{1/(2D)}) = o(1).$$

By Lindeberg's CLT,

$$\begin{aligned}
\sum_{j=1}^q c_j\mathbb{G}_n(s_j,t_j) &= \sum_{i=1}^n \mathfrak{U}_{i,q} + o_p(1) \\
&\xrightarrow{d} \sum_{j=1}^q c_jZ(s_j,t_j) \sim \mathcal{N}\Big(0, \sum_{j_1,j_2=1}^q c_{j_1}c_{j_2}C_Z\{(s_{j_1},t_{j_1}),(s_{j_2},t_{j_2})\}\Big).
\end{aligned}$$

Step 2. *Asymptotic tightness of* $\mathbb{G}_n$

Next, we show the equicontinuity of the process $\mathbb{G}_n$ in (3.8). We first focus on the leading term $I_{3,n}$ in (A.45), and recall that

$$I_{3,n}(s,t) = \sum_{i=1}^n \mathfrak{U}_i(s,t) = n^{-1/2}\lambda^{(2a+1)/(4D)}\sum_{i=1}^n \tau(X_i\otimes\varepsilon_i)(s,t). \tag{A.50}$$

Let $\Psi(x) = x^2$ and let $\|U\|_\Psi = \inf\{c > 0 : \mathrm{E}\{\Psi(|U|/c)\} \leq 1\}$ denote the Orlicz norm for a real-valued random variable U. For some metric d on $[0,1]^2$, let $\mathfrak{D}(w,d)$ denote the w-packing number of the metric space $([0,1]^2, d)$, where d is an appropriate metric specified below. Since

$\mathrm{E}\{\tau(X\otimes\varepsilon)(s,t)\}=0$ for any $(s,t)\in[0,1]^2$ and $\mathrm{E}\{\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}\langle X\otimes\varepsilon,\varphi_{k'\ell'}\rangle_{L^2}\}=\delta_{kk'}\delta_{\ell\ell'}$, for $k,k',\ell,\ell'\geq 1$, by (3.7), for any $(s_1,t_1),(s_2,t_2)\in[0,1]^2$,

$$\begin{aligned}
&\mathrm{E}|I_{3,n}(s_1,t_1)-I_{3,n}(s_2,t_2)|^2\\
&=\lambda^{(2a+1)/(2D)}\,\mathrm{E}\big|\tau(X\otimes\varepsilon)(s_1,t_1)-\tau(X\otimes\varepsilon)(s_2,t_2)\big|^2\\
&=\lambda^{(2a+1)/(2D)}\,\mathrm{E}\bigg|\sum_{k,\ell}\frac{\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}}{1+\lambda\rho_{k\ell}}\{\varphi_{k\ell}(s_1,t_1)-\varphi_{k\ell}(s_2,t_2)\}\bigg|^2\\
&=\lambda^{(2a+1)/(2D)}\sum_{k,\ell}\frac{1}{(1+\lambda\rho_{k\ell})^2}\big|\varphi_{k\ell}(s_1,t_1)-\varphi_{k\ell}(s_2,t_2)\big|^2\\
&\leq c\,\lambda^{(a-b)/D}\max\{|s_1-s_2|^{2\vartheta},|t_1-t_2|^{2\vartheta}\}\,. \qquad\text{(A.51)}
\end{aligned}$$

We therefore deduce from (A.51) that

$$\|I_{3,n}(s_1,t_1)-I_{3,n}(s_2,t_2)\|_{\Psi}\leq c\,\lambda^{(a-b)/(2D)}\max\{|s_1-s_2|^{\vartheta},|t_1-t_2|^{\vartheta}\}\,. \qquad\text{(A.52)}$$

Next, we shall show that, there exists a metric d on $[0,1]^2$ such that, for any $e>0$,

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\mathrm{P}\bigg\{\sup_{d\{(s_1,t_1),(s_2,t_2)\}\leq\delta}|I_{3,n}(s_1,t_1)-I_{3,n}(s_2,t_2)|>e\bigg\}=0\,, \qquad\text{(A.53)}$$

where we distinguish the cases: $\vartheta>1$ and $0\leq\vartheta\leq 1$.

Case (i): $\vartheta>1$

Recall that in the case of $\vartheta>1$, we have assumed $b=a$, and let $d_1\{(s_1,t_1),(s_2,t_2)\}=\max\{|s_1-s_2|^{\vartheta},|t_1-t_2|^{\vartheta}\}$. In view of (A.51), we have $\|I_{3,n}(s_1,t_1)-I_{3,n}(s_2,t_2)\|_{\Psi}\leq c\,d_1\{(s_1,t_1),(s_2,t_2)\}$. Note that the packing number of $[0,1]^2$ with respect to the metric d_1 satisfies $\mathfrak{D}(\zeta,d_1)\lesssim\zeta^{-2/\vartheta}$. By Theorem 2.2.4 in van der Vaart and Wellner (1996), for any $e,\eta>0$,

$$\begin{aligned}
&\mathrm{P}\bigg\{\sup_{d_1\{(s_1,t_1),(s_2,t_2)\}\leq\delta}|I_{3,n}(s_1,t_1)-I_{3,n}(s_2,t_2)|>e\bigg\}\\
&\leq c\,\bigg\|\sup_{d_1\{(s_1,t_1),(s_2,t_2)\}\leq\delta}|I_{3,n}(s_1,t_1)-I_{3,n}(s_2,t_2)|\bigg\|_{\Psi}\leq c\int_0^{\eta}\sqrt{\mathfrak{D}(\zeta,d_1)}d\zeta+\delta\,\mathfrak{D}(\eta,d_1)\\
&\leq c\int_0^{\eta}\zeta^{-1/\vartheta}d\zeta+\delta\,w^{-2/\vartheta}=c\,\eta^{(\vartheta-1)/\vartheta}+\delta\,\eta^{-2/\vartheta}\,.
\end{aligned}$$

Using $\eta=\sqrt{\delta}$ and $\vartheta>1$, it follows that (A.53) holds by taking the metric $d=d_1$.

Case (ii): $0\leq\vartheta\leq 1$

In this case, in order to show (A.53), we shall use Lemma B.8 in Section B.1, which is a modified version of Lemma A.1 in Kley et al. (2016). Let

$$d_2\{(s_1,t_1),(s_2,t_2)\} = \max\{|s_1-s_2|^2, |t_1-t_2|^2\}$$

and let $\overline{\eta} = \lambda^{(a-b)/(2D-\vartheta D)}$. In view of (A.52), we have, when $d_2\{(s_1,t_1),(s_2,t_2)\} \geq \overline{\eta}/2 > 0$,

$$\|I_{3,n}(s_1,t_1) - I_{3,n}(s_2,t_2)\|_\Psi \leq c\,\lambda^{(a-b)/(2D)}\left[d_2\{(s_1,t_1),(s_2,t_2)\}\right]^{\vartheta/2} \leq c\,d_2\{(s_1,t_1),(s_2,t_2)\}\,.$$

By Assumption A4 and Markov's inequality, by taking $c > c_X^{-1}$, for c_X in Assumption A4, $\sum_{n=1}^\infty \mathrm{P}\big(\|X\|_{L^2} \geq c\log n\big) \leq \mathrm{E}\{\exp(c_X\|X\|_{L^2})\}\sum_{n=1}^\infty n^{-cc_X} < \infty$. By the Borel-Cantelli lemma and applying the same argument to $\|\varepsilon\|_{L^2}$ yields that $\|X\otimes\varepsilon\|_{L^2} = \|X\|_{L^2}\times\|\varepsilon\|_{L^2} \leq (c\log n)^2$ holds for n large enough almost surely. Hence, by Assumption A2 and Lemma B.2, for n large enough,

$$\begin{aligned}
\sup_{(s,t)\in[0,1]^2}|\mathfrak{U}_i(s,t)| &\leq \sup_{(s,t)\in[0,1]^2} n^{-1/2}\lambda^{(2a+1)/(4D)}\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}|\langle X_i\otimes\varepsilon_i,\varphi_{k\ell}\rangle_{L^2}|\times\|\varphi_{k\ell}\|_\infty \\
&\leq n^{-1/2}\lambda^{(2a+1)/(4D)}\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\|X_i\otimes\varepsilon_i\|_{L^2}\times\|\varphi_{k\ell}\|_{L^2}\times\|\varphi_{k\ell}\|_\infty \\
&\leq c\,n^{-1/2}\lambda^{(2a+1)/(4D)}(\log n)^2\sum_{k,\ell}\frac{(k\ell)^{2a}}{1+\lambda(k\ell)^{2D}} \\
&\leq c\,n^{-1/2}\lambda^{-(2a+1)/(4D)}(\log n)^2 \qquad \text{(A.54)}
\end{aligned}$$

almost surely. In addition, by (A.51),

$$\sup_{(s_1,t_1),(s_2,t_2)\in[0,1]^2}\mathrm{E}|I_{3,n}(s_1,t_1) - I_{3,n}(s_2,t_2)|^2 \leq c\,\lambda^{(a-b)/D}.$$

By Bernstein's inequality, combining the above equation with (A.54), we deduce that, for n large enough, for any $(s_1,t_1),(s_2,t_2)\in[0,1]^2$ and for any $e>0$,

$$\begin{aligned}
&\mathrm{P}\Big\{|I_{3,n}(s_1,t_1) - I_{3,n}(s_2,t_2)| > e/4\Big\} \\
&\leq 2\exp\left\{-\frac{e^2/16}{2\lambda^{(a-b)/D} + en^{-1/2}\lambda^{-(2a+1)/(4D)}(\log n)^2/6}\right\}. \qquad \text{(A.55)}
\end{aligned}$$

Now, note that $\mathfrak{D}(\zeta,d_2) \leq c\,\zeta^{-1}$ and recall that in the case of $0\leq\vartheta\leq 1$ we have assumed $b<a$. By Lemma B.8 and (A.55), there exists a set $\widetilde{[0,1]^2}\subset[0,1]^2$ that contains at most $\mathfrak{D}(\zeta,d_2)$ points, such that, for any $\delta, e>0$ and $\eta>\overline{\eta}$, as $n\to\infty$,

$$\begin{aligned}
&\mathrm{P}\Bigg\{\sup_{d_2\{(s_1,t_1),(s_2,t_2)\}\leq\delta}|I_{3,n}(s_1,t_1) - I_{3,n}(s_2,t_2)| > e\Bigg\} \\
&\leq c\left\{\int_{\overline{\eta}/2}^{\eta}\sqrt{\mathfrak{D}(\zeta,d_2)}d\zeta + (\delta+2\overline{\eta})\,\mathfrak{D}(\eta,d_2)\right\}^2
\end{aligned}$$

$$
\begin{aligned}
&+ \mathrm{P}\Big\{ \sup_{\substack{d_2\{(s_1,t_1),(s_2,t_2)\}\le\overline{\eta} \\ (s_1,t_1)\in\widetilde{[0,1]}^2}} |I_{3,n}(s_1,t_1) - I_{3,n}(s_2,t_2)| > e/4 \Big\} \\
&\le c\Big\{ \int_{\overline{\eta}/2}^{\eta} \zeta^{-1/2} d\zeta + (\delta + 2\lambda^{(a-b)/(2D-\vartheta D)})\eta^{-1} \Big\}^2 \\
&\quad + \mathfrak{D}(\overline{\eta}, d_2) \times \sup_{(s_1,t_1),(s_2,t_2)\in\widetilde{[0,1]}^2} \mathrm{P}\Big\{ |I_{3,n}(s_1,t_1) - I_{3,n}(s_2,t_2)| > e/4 \Big\} \\
&\le c\,(\eta + \delta^2\eta^{-2}) + c\,\lambda^{-(a-b)/(2D-\vartheta D)} \exp\Big\{ - \frac{e^2/16}{2\lambda^{(a-b)/D} + en^{-1/2}\lambda^{-(2a+1)/(4D)}(\log n)^2/6} \Big\} \\
&\le c\,(\eta + \delta^2\eta^{-2}) + o(1)\,,
\end{aligned}
$$

where in the last step we used $\lambda^{-1} \lesssim n^{2D}$ by Assumption A5, and the assumption that $\lambda^{(a-b)/D} = o(n^{-(a-b)\nu_1/D})$ and $n^{-1/2}\lambda^{-(2a+1)/(4D)} = o(n^{-\nu_2})$, for $\nu_1, \nu_2 > 0$. Therefore, by taking $\eta > 0$ small enough, we deduce from the above equation that, when $0 \le \vartheta \le 1$, for any $e > 0$, (A.53) holds by taking the metric $d = d_2$.

As for the remaining processes $I_{1,n}$ and $I_{2,n}$ in (A.44), in view of (A.47), for any $e > 0$ and for the metric d,

$$
\begin{aligned}
&\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathrm{P}\Big\{ \sup_{d\{(s_1,t_1),(s_2,t_2)\}\le\delta} |I_{1,n}(s_1,t_1) + I_{2,n}(s_1,t_1) - I_{1,n}(s_2,t_2) - I_{2,n}(s_2,t_2)| > e \Big\} \\
&\le \limsup_{n\to\infty} \mathrm{P}\Big\{ \sup_{(s,t)\in[0,1]^2} |I_{1,n}(s,t)| + \sup_{(s,t)\in[0,1]^2} |I_{2,n}(s,t)| > e/2 \Big\} = 0\,,
\end{aligned}
$$

Combining this result with (A.53) proves that the process $\mathbb{G}_n$ is asymptotic uniformly equicontinuous w.r.t. the metric d in (A.53) (that is, $d = d_1$ when $\vartheta > 1$, and $d = d_2$ when $0 \le \vartheta \le 1$), which entails the asymptotic tightness of $\mathbb{G}_n$.

The assertion of the theorem therefore follows from Theorems 1.5.4 and 1.5.7 in van der Vaart and Wellner (1996).

A.5 Proof of Theorem 4.1

By taking $\lambda \asymp n^{-2D/(2D+1)}$, the upper bound in (i) follows from Lemma A.2 and the fact that $\|\beta\|_V^2 \le \|\beta\|_K^2$ for any $\beta \in \mathcal{H}$.

For (ii), we prove the lower bound in the particular case where ε is a mean zero Gaussian white noise process independent of X with $\mathrm{E}\{\varepsilon^2(t)\} \equiv \sigma_\varepsilon^2 > 0$. By Theorem 2.5 in Tsybakov (2008), in order to show the lower bound, we need to show that, for $M \ge 2$, $\mathcal{H}$ contains elements $\beta_0, \ldots, \beta_M$ that satisfy the following two conditions:

(C1) $\|\beta_j - \beta_k\|_V^2 \ge 2c_0 n^{-2D/(2D+1)}$, for $0 \le j < k \le M$;

(C2) $M^{-1}\sum_{j=1}^{M}\mathcal{K}(P_j,P_0)\le\alpha\log M$, where $0<\alpha<1/8$, $\mathcal{K}$ is the Kullback-Leibler divergence, and P_j denotes joint distribution of $(X_{1,j},Y_{1,j}),\ldots,(X_{n,j},Y_{n,j})$, where $Y_{i,j}(t)=\int\beta_j(s,t)X_{i,j}(s)ds+\varepsilon_{i,j}(t)$, for $1\le i\le n$.

For the constant D in Assumption A2, define $\nu_n=\lfloor n^{1/(4D+2)}\rfloor$. For any $\omega=(\omega_{(\nu_n+1,\nu_n+1)},\ldots,\omega_{(2\nu_n,2\nu_n)})\in\{0,1\}^{\nu_n^2}$, let

$$\beta_\omega=c_1n^{-1/2}\sum_{k=\nu_n+1}^{2\nu_n}\sum_{\ell=\nu_n+1}^{2\nu_n}\omega_{(k,\ell)}\,\varphi_{k\ell}\,, \tag{A.56}$$

where $c_1>0$ is a constant independent of n to be specified later. We first verify that the β_ω's are elements in $\mathcal{H}$. Since, by Assumption A2, $\{\varphi_{k\ell}\}_{k,\ell\ge1}$ diagonalizes the operator J defined in (2.6), we have

$$\begin{aligned}\|\beta_\omega\|_K^2&=c_1^2\,n^{-1}\left\langle\sum_{k=\nu_n+1}^{2\nu_n}\sum_{\ell=\nu_n+1}^{2\nu_n}\omega_{(k,\ell)}\,\varphi_{k\ell}\,,\sum_{k'=\nu_n+1}^{2\nu_n}\sum_{\ell'=\nu_n+1}^{2\nu_n}\omega_{(k',\ell')}\,\varphi_{k'\ell'}\right\rangle_K\\&=c_1^2\,n^{-1}\sum_{k=\nu_n+1}^{2\nu_n}\sum_{\ell=\nu_n+1}^{2\nu_n}\omega_{(k,\ell)}^2\,\|\varphi_{k\ell}\|_K^2=c_1^2\,n^{-1}\sum_{k=\nu_n+1}^{2\nu_n}\sum_{\ell=\nu_n+1}^{2\nu_n}\omega_{(k,\ell)}^2\,(1+\lambda\rho_{k\ell})\\&\le c\,n^{-1}\sum_{k=\nu_n+1}^{2\nu_n}\sum_{\ell=\nu_n+1}^{2\nu_n}\{1+\lambda(k\ell)^{2D}\}\,.\end{aligned}$$

Note that this inequality and the inequality in (A.12) in the proof of Proposition 2.1 holds for any $\lambda>0$. Therefore, for $\|\cdot\|_{\mathcal{H}}$ defined in (2.3), combining these two equations, we may take $\lambda=1$ and find that

$$\|\beta_\omega\|_{\mathcal{H}}^2\le c\,\|\beta_\omega\|_K^2\le c\,n^{-1}\sum_{k=\nu_n+1}^{2\nu_n}\sum_{\ell=\nu_n+1}^{2\nu_n}(k\ell)^{2D}\le c\,n^{-1}\nu_n^{2+4D}\le c\,,$$

which shows that, for any $\omega\in\{0,1\}^{\nu_n^2}$, β_ω defined in (A.56) is an element of $\mathcal{H}$.

By the Varshamov-Gilbert bound (Lemma 2.9 in Tsybakov, 2008), for $\nu_n^2\ge8$, there exists a subset $\Omega=\{\omega^{(0)},\ldots,\omega^{(M)}\}\in\{0,1\}^{\nu_n^2}$ with $M\ge2^{\nu_n^2/8}$ such that, $\omega^{(0)}=(0,\ldots,0)$ and for any $0\le j<j'\le M$,

$$\mathrm{H}\big(\omega^{(j_1)},\omega^{(j_2)}\big)\ge\frac{\nu_n^2}{8}\,,$$

where $\mathrm{H}(\cdot,\cdot)$ is the Hamming distance. For $0\le j\le M$, let $\omega^{(j)}=(\omega^{(j)}_{(\nu_n+1,\nu_n+1)},\ldots,\omega^{(j)}_{(2\nu_n,2\nu_n)})$. Let $\beta_0,\ldots,\beta_M$ denote the functions defined as in (A.56) that corresponds to $\omega^{(0)},\ldots,\omega^{(M)}\in\Omega$. For $0\le j<j'\le M$, in view of (A.56),

$$\beta_j-\beta_{j'}=c_1\,n^{-1/2}\sum_{k=\nu_n+1}^{2\nu_n}\sum_{\ell=\nu_n+1}^{2\nu_n}\big(\omega^{(j)}_{(k,\ell)}-\omega^{(j')}_{(k,\ell)}\big)\,\varphi_{k\ell}\,. \tag{A.57}$$

By Assumption A2, since $\{\varphi_{k\ell}\}_{k,\ell\geq 1}$ diagonalizes the operator V defined in (2.9), we deduce from (A.57) that

$$
\begin{aligned}
\|\beta_j - \beta_{j'}\|_V^2 &= c_1^2 n^{-1} \sum_{k=\nu_n+1}^{2\nu_n} \sum_{\ell=\nu_n+1}^{2\nu_n} \big(\omega_{(k,\ell)}^{(j)} - \omega_{(k,\ell)}^{(j')}\big)^2 V(\varphi_{k\ell}, \varphi_{k\ell}) \\
&= c_1^2 n^{-1} \sum_{k=\nu_n+1}^{2\nu_n} \sum_{\ell=\nu_n+1}^{2\nu_n} \mathbb{1}\{\omega_{(k,\ell)}^{(j)} \neq \omega_{(k,\ell)}^{(j')}\} \\
&= c_1^2 n^{-1} \mathrm{H}\big(\omega^{(j)}, \omega^{(j')}\big) \geq c_1^2 8^{-1} n^{-1} \nu_n^2 \geq c_1^2 8^{-1} n^{-2D/(2D+1)} .
\end{aligned}
$$

By taking $c_0 = c_1^2/16$, the above equation indicates that Condition (C1) is valid.

For any $0 \leq j < j' \leq M$, in view of (A.57),

$$
\begin{aligned}
\mathcal{K}(P_j, P_{j'}) &= \frac{n}{2\sigma_\varepsilon^2} \mathrm{E} \int_0^1 \bigg[\int_0^1 \{\beta_j(s,t) - \beta_{j'}(s,t)\} X(s) ds \bigg]^2 dt = \frac{n}{2\sigma_\varepsilon^2} \|\beta_j - \beta_{j'}\|_V^2 \\
&= c_1^2 \sigma_\varepsilon^{-2} \sum_{k=\nu_n+1}^{2\nu_n} \sum_{\ell=\nu_n+1}^{2\nu_n} \big(\omega_{(k,\ell)}^{(j)} - \omega_{(k,\ell)}^{(j')}\big)^2 \leq c_1^2 \sigma_\varepsilon^{-2} \nu_n^2 .
\end{aligned}
$$

Therefore, for any $0 < \alpha < 1/8$, by taking $0 < c_1 < \sigma_\varepsilon \sqrt{\alpha \log 2/8}$ in (A.56), we have

$$
\frac{1}{M} \sum_{j=1}^{M} \mathcal{K}(P_j, P_0) \leq c_1^2 \sigma_\varepsilon^{-2} \nu_n^2 \leq \frac{\alpha \nu_n^2 \log 2}{8} \leq \alpha \log M ,
$$

which verifies Condition (C2) and completes the proof.

A.6 Proof of Theorem 4.2

Let $\mathrm{BL}_1\{C([0,1]^2)\}$ denote the collection of all functionals $h : C([0,1]^2) \to [-1,1]$ such that h is uniformly Lipschitz: for any $g_1, g_2 \in C([0,1]^2)$, $|h(g_1) - h(g_2)| \leq \|g_1 - g_2\|_\infty = \sup_{(s,t)\in[0,1]^2} |g_1(s,t) - g_2(s,t)|$. We shall show that conditionally on the data $\{(X_i, Y_i)\}_{i=1}^n$, the bootstrap process $\mathbb{G}_{n,q}^*$ converges to the same limit as $\mathbb{G}_n$ in (3.8). To achieve this, we shall prove that, for Z in (3.8), as $n \to \infty$,

$$
\sup_{h\in \mathrm{BL}_1\{C([0,1]^2)\}} |\mathrm{E}_M\{h(\mathbb{G}_{n,1}^*)\} - \mathrm{E}\{h(Z)\}| = o_p(1) ,
$$

where E_M denote the conditional expectation given the data $\{(X_i, Y_i)\}_{i=1}^n$; see Theorem 23.7 in Van der Vaart (1998). Note that the results in Lemma 3.1 in Bücher and Kojadinovic (2019) hold if their $\ell^\infty(T)$ space is replaced by $C(T)$, and therefore, in our case, we shall show that, for any fixed $Q \geq 2$, as $n \to \infty$,

$$
(\mathbb{G}_n, \mathbb{G}_{n,1}^*, \ldots, \mathbb{G}_{n,Q}^*) \rightsquigarrow (Z, Z_1, \ldots, Z_Q) \quad \text{in } \{C([0,1]^2)\}^{Q+1} , \tag{A.58}
$$

where $Z_1,\dots,Z_Q$ are i.i.d. copies of the process Z in (3.8).

For $1\le q\le Q$, define the bootstrap version of $S_{n,\lambda}$ in (3.2) by

$$S^*_{n,q}(\beta) = -\frac{1}{n}\sum_{i=1}^n M_{i,q}\,\tau\Big[X_i\otimes\Big\{Y_i-\int_0^1\beta(s,\cdot)X_i(s)ds\Big\}\Big]+W_\lambda(\beta)\,,$$

and let $L^*_{n,\lambda}(\beta)$ denote the bootstrap objective function in (4.3). Direct calculations yields that $L^*_{n,\lambda}(\beta)\beta_1 = \langle S^*_{n,q}(\beta),\beta_1\rangle_K$, $\mathrm{E}\{\mathcal{D}^2L^*_{n,\lambda}(\beta)\beta_1\beta_2\} = \langle\beta_1,\beta_2\rangle_K$ and $S^*_{n,q}(\beta_0) = -n^{-1}\sum_{i=1}^n M_{i,q}\,\tau(X_i\otimes\varepsilon_i)$. Recalling the notation of $I_{n,2}(s,t)$ in (A.45), we have

$$\mathbb{G}^*_{n,q}(s,t)=\sqrt{n}\lambda^{(2a+1)/(4D)}\{\widehat\beta^*_{n,q}(s,t)-\widehat\beta_n(s,t)\}=\mathfrak{J}^*_{n,q}(s,t)+H^*_{n,q}(s,t)-I_{n,2}(s,t)\,,$$

where

$$\begin{aligned}\mathfrak{J}^*_{n,q}(s,t)&=\sqrt{n}\lambda^{(2a+1)/(4D)}\{\widehat\beta^*_{n,q}(s,t)-\beta_0(s,t)+S^*_{n,\lambda}\beta_0(s,t)\}\,,\\ H^*_{n,q}(s,t)&=\sqrt{n}\lambda^{(2a+1)/(4D)}\{S_{n,\lambda}\beta_0(s,t)-S^*_{n,q}\beta_0(s,t)\}\\ &=n^{-1/2}\lambda^{(2a+1)/(4D)}\sum_{i=1}^n(1-M_{i,q})\,\tau(X_i\otimes\varepsilon_i)(s,t)\,.\end{aligned}$$

By exactly the same arguments used in the proof of Theorem 3.3, it follows that $\sup_{(s,t)\in[0,1]^2}|\mathfrak{J}^*_{n,q}(s,t)| = o_p(1)$ as $n\to\infty$. Furthermore, recall from (A.47) that $\sup_{(s,t)\in[0,1]^2}|I_{n,2}(s,t)|=o_p(1)$. Since in the proof of Theorem 3.3 we have shown that $H_n\rightsquigarrow Z$ in $C([0,1]^2)$, therefore, in order to show (A.58), we shall show that $(H^*_{n,1},\dots,H^*_{n,Q})\rightsquigarrow(Z_1,\dots,Z_Q)$ in $C([0,1]^2)$. The proof of (A.58) therefore relies on the finite dimensional convergence of $(H^*_{n,1},\dots,H^*_{n,Q})$ and the asymptotic tightness of the process $H^*_{n,q}$.

We first show convergence of the finite dimensional distributions and introduce the notations $\mathbb{H}^*_n=(H^*_{n,1},\dots,H^*_{n,Q})^\top$ and $\mathbf{Z}=(Z_1,\dots,Z_Q)^\top$. For arbitrary $L\in\mathbb{N}$, $b_1,\dots,b_L\in\mathbb{R}$ and $\mathbf{c}_1,\dots,\mathbf{c}_L\in\mathbb{R}^Q$, we need to prove that

$$\sum_{\ell=1}^L\mathbf{c}_\ell^\top\mathbb{H}^*_n(s_\ell,t_\ell)\xrightarrow{d}\sum_{\ell=1}^L\mathbf{c}_\ell^\top\mathbf{Z}(s_\ell,t_\ell)\,. \tag{A.59}$$

For $1\le i\le n$, let $\mathfrak{H}^*_i = n^{-1/2}\lambda^{(2a+1)/(4D)}\sum_{\ell=1}^L\sum_{q=1}^Q c_{\ell,q}(M_{i,q}-1)\,\tau(X_i\otimes\varepsilon_i)(s_\ell,t_\ell)$, and note that the $\mathfrak{H}^*_i$'s are i.i.d. and $\sum_{\ell=1}^L\mathbf{c}_\ell^\top\mathbb{H}^*_n(s_\ell,t_\ell)=\sum_{i=1}^n\mathfrak{H}^*_i$. Since the $M_{i,q}$'s are i.i.d., $\mathrm{E}(M_{i,q})=1$ and $\mathrm{E}|M_{i,q}-1|^2=1$, direct calculations yield that

$$\mathrm{var}\Big\{\sum_{\ell=1}^L\mathbf{c}_\ell^\top\mathbb{H}^*_n(s_\ell,t_\ell)\Big\}=\mathrm{var}(\mathfrak{H}^*_1)=\lambda^{(2a+1)/(2D)}\mathrm{var}\Big\{\sum_{\ell=1}^L\sum_{q=1}^Q c_{\ell,q}(M_{1,q}-1)\,\tau(X_1\otimes\varepsilon_1)(s_\ell,t_\ell)\Big\}$$

$$=\lambda^{(2a+1)/(2D)}\sum_{\ell,\ell'=1}^L\sum_{q,q'=1}^Q c_{\ell,q}c_{\ell',q'}\mathrm{E}\{(M_{1,q}-1)(M_{1,q'}-1)\}\mathrm{E}\big\{\tau(X_1\otimes\varepsilon_1)(s_\ell,t_\ell)\tau(X_1\otimes\varepsilon_1)(s_{\ell'},t_{\ell'})\big\}$$

$$= \lambda^{(2a+1)/(2D)} \sum_{\ell,\ell'=1}^{L} \sum_{q=1}^{Q} c_{\ell,q} c_{\ell',q} \mathrm{E}\{(M_{1,q}-1)^2\} \mathrm{E}\Big\{ \tau(X_1 \otimes \varepsilon_1)(s_\ell, t_\ell) \tau(X_1 \otimes \varepsilon_1)(s_{\ell'}, t_{\ell'}) \Big\}$$

$$= \lambda^{(2a+1)/(2D)} \sum_{\ell,\ell'=1}^{L} \sum_{q=1}^{Q} c_{\ell,q} c_{\ell',q} \sum_{k,j} \frac{\varphi_{kj}(s_\ell, t_\ell)\varphi_{kj}(s_{\ell'}, t_{\ell'})}{(1+\lambda\rho_{kj})^2}$$

$$= \sum_{q=1}^{Q} \sum_{\ell,\ell'=1}^{L} c_{\ell,q} c_{\ell',q} C_Z\{(s_\ell, t_\ell), (s_{\ell'}, t_{\ell'})\} + o(1),$$

as $n \to \infty$. Note that $\mathrm{var}\{\sum_{\ell=1}^{L} \mathbf{c}_\ell^\top \mathbf{Z}(s_\ell, t_\ell)\} = \sum_{q=1}^{Q} \sum_{\ell,\ell'=1}^{L} c_{\ell,q} c_{\ell',q} C_Z\{(s_\ell, t_\ell), (s_{\ell'}, t_{\ell'})\}$. When $\sum_{q=1}^{Q} \sum_{\ell,\ell'=1}^{L} c_{\ell,q} c_{\ell',q} C_Z\{(s_\ell, t_\ell), (s_{\ell'}, t_{\ell'})\} = 0$, $\sum_{\ell=1}^{L} \mathbf{c}_\ell^\top \mathbf{Z}(s_\ell, t_\ell)$ has a degenerate distribution with a point mass at zero, and $\sum_{\ell=1}^{L} \mathbf{c}_\ell^\top \mathbb{H}_n^*(s_\ell, t_\ell) = o_p(1)$, so that (A.59) is valid. When $\sum_{q=1}^{Q} \sum_{\ell,\ell'=1}^{L} c_{\ell,q} c_{\ell',q} C_Z\{(s_\ell, t_\ell), (s_{\ell'}, t_{\ell'})\} \neq 0$, using arguments similar to the ones used in the proof of Theorem 3.3, we can show that Lindeberg's condition is satisfied, so that (A.59) is valid.

For the asymptotic tightness of the $H_{n,q}^*$, note that $|1 - M_{i,q}| \le \sqrt{2}$ almost surely. Therefore, the asymptotic tightness of H_n in (A.50), proved in Section A.4, implies the asymptotic tightness of $H_{n,q}^*$. By Theorem 1.5.4 in van der Vaart and Wellner (1996), together with the weak convergence $H_n \rightsquigarrow Z$ proved in Section A.4, we have that, for any $Q \ge 2$, as $n \to \infty$,

$$(H_n, H_{n,1}^*, \ldots, H_{n,Q}^*) \rightsquigarrow (Z, Z_1, \ldots, Z_Q) \quad \text{in } \{C([0,1]^2)\}^{Q+1},$$

which validates (A.58) and completes the proof.

A.7 Proof of Theorem 4.3

Defining $\widehat{\beta}_\Delta = \widehat{\beta}_n - \beta_*$ and observing (3.1), (3.2) and (A.20), it follows that

$$\begin{aligned}
\mathfrak{L}_n(\beta_*) &= L_{n,\lambda}(\beta_*) - L_{n,\lambda}(\beta_* + \widehat{\beta}_\Delta) = -\mathcal{D}L_{n,\lambda}(\beta_*)\widehat{\beta}_\Delta - \frac{1}{2}\mathcal{D}^2 L_{n,\lambda}(\beta_*)\widehat{\beta}_\Delta\widehat{\beta}_\Delta \\
&= -\big\langle S_{n,\lambda}(\beta_*), \widehat{\beta}_\Delta \big\rangle_K - \frac{1}{2}\big\langle \mathcal{D}S_{n,\lambda}(\beta_*)\widehat{\beta}_\Delta, \widehat{\beta}_\Delta \big\rangle_K \\
&= -\big\langle S_{n,\lambda}(\beta_*), \widehat{\beta}_\Delta \big\rangle_K - \frac{1}{2}\Big\langle \mathcal{D}S_{n,\lambda}(\beta_*)\widehat{\beta}_\Delta - \mathcal{D}S_\lambda(\beta_*)\widehat{\beta}_\Delta, \widehat{\beta}_\Delta \Big\rangle_K - \frac{1}{2}\|\widehat{\beta}_\Delta\|_K^2 \\
&= \frac{1}{2}\|\widehat{\beta}_\Delta\|_K^2 - \big\langle \widehat{\beta}_\Delta + S_{n,\lambda}(\beta_*), \widehat{\beta}_\Delta \big\rangle_K - \frac{1}{2}\Big\langle \mathcal{D}S_{n,\lambda}(\beta_*)\widehat{\beta}_\Delta - \mathcal{D}S_\lambda(\beta_*)\widehat{\beta}_\Delta, \widehat{\beta}_\Delta \Big\rangle_K,
\end{aligned}$$

where we use the fact that $\mathcal{D}S_\lambda(\beta_*) = id$. Note that

$$\begin{aligned}
\|\widehat{\beta}_\Delta\|_K^2 &= \big\| \{\widehat{\beta}_\Delta + S_{n,\lambda}(\beta_*)\} - S_{n,\lambda}(\beta_*) \big\|_K^2 \\
&= \|S_{n,\lambda}(\beta_*)\|_K^2 - 2\langle S_{n,\lambda}(\beta_*), \widehat{\beta}_\Delta + S_{n,\lambda}(\beta_*)\rangle_K + \|\widehat{\beta}_\Delta + S_{n,\lambda}(\beta_*)\|_K^2 .
\end{aligned}$$

Therefore, we deduce that

$$2n\,\mathfrak{L}_n(\beta_*) = I_{1,n} + I_{2,n} + I_{3,n} + I_{4,n}\,,$$

where we use the notations that

$$\begin{aligned} I_{1,n} &= n\|S_{n,\lambda}(\beta_*)\|_K^2\,,\\ I_{2,n} &= -2n\big\langle \widehat{\beta}_\Delta + S_{n,\lambda}(\beta_*), \widehat{\beta}_\Delta\big\rangle_K\,,\\ I_{3,n} &= -n\big\langle \mathcal{D}S_{n,\lambda}(\beta_*)\widehat{\beta}_\Delta - \mathcal{D}S_\lambda(\beta_*)\widehat{\beta}_\Delta, \widehat{\beta}_\Delta\big\rangle_K\\ I_{4,n} &= -2n\big\langle S_{n,\lambda}(\beta_*), \widehat{\beta}_\Delta + S_{n,\lambda}(\beta_*)\big\rangle_K + n\|\widehat{\beta}_\Delta + S_{n,\lambda}(\beta_*)\|_K^2\,. \end{aligned} \tag{A.60}$$

We now discuss the term $I_{\ell,n}$ separately, starting with $I_{1,n}$. In view of (3.2), we have

$$n\|S_{n,\lambda}(\beta_*)\|_K^2 = n\left\| -\frac{1}{n}\sum_{i=1}^n \tau(X_i\otimes\varepsilon_i) + W_\lambda(\beta_*)\right\|_K^2 = I_{1,1,n} + I_{1,2,n} + I_{1,3,n}\,, \tag{A.61}$$

where

$$\begin{aligned} I_{1,1,n} &= \frac{1}{n}\left\|\sum_{i=1}^n \tau(X_i\otimes\varepsilon_i)\right\|_K^2\\ I_{1,2,n} &= -2\sum_{i=1}^n \big\langle \tau(X_i\otimes\varepsilon_i), W_\lambda(\beta_*)\big\rangle_K\\ I_{1,3,n} &= n\,\|W_\lambda(\beta_*)\|_K^2\,. \end{aligned} \tag{A.62}$$

Observing (2.13), we have

$$\|\beta\|_K^2 = \sum_{k,\ell,k',\ell'} \frac{\langle\beta,\varphi_{k\ell}\rangle_K\,\langle\beta,\varphi_{k'\ell'}\rangle_K}{(1+\lambda\rho_{k\ell})(1+\lambda\rho_{k'\ell'})}\langle\varphi_{k\ell},\varphi_{k'\ell'}\rangle_K = \sum_{k,\ell}\frac{\langle\beta,\varphi_{k\ell}\rangle_K^2}{1+\lambda\rho_{k\ell}}\,, \quad \forall\beta\in\mathcal{H}\,, \tag{A.63}$$

which gives

$$\begin{aligned} I_{1,1,n} &= \frac{1}{n}\left\|\sum_{i=1}^n \tau(X_i\otimes\varepsilon_i)\right\|_K^2 = \frac{1}{n}\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\left\langle \sum_{i=1}^n \tau(X_i\otimes\varepsilon_i), \varphi_{k\ell}\right\rangle_K^2\\ &= \frac{1}{n}\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\left\{\sum_{i=1}^n \big\langle\tau(X_i\otimes\varepsilon_i),\varphi_{k\ell}\big\rangle_K\right\}^2 = \frac{W_n}{n} + \frac{1}{n}\sum_{i=1}^n W_{0,i}\,, \end{aligned} \tag{A.64}$$

where $W_n = \sum_{i_1<i_2} W_{i_1 i_2}$ and

$$\begin{aligned} W_{i_1 i_2} &= 2\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\big\langle\tau(X_{i_1}\otimes\varepsilon_{i_1}),\varphi_{k\ell}\big\rangle_K\big\langle\tau(X_{i_2}\otimes\varepsilon_{i_2}),\varphi_{k\ell}\big\rangle_K\,,\\ W_{0,i} &= \sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\big\langle\tau(X_i\otimes\varepsilon_i),\varphi_{k\ell}\big\rangle_K^2\,. \end{aligned} \tag{A.65}$$

For W_n in (A.64), we have

$$\begin{aligned}\sigma^2_{W_n} &:= \mathrm{E}(W_n^2) = \sum_{i_1<i_2} \mathrm{E}(W^2_{i_1 i_2}) \\ &= 4\sum_{i_1<i_2} \mathrm{E}\Bigg[\Bigg\{\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\big\langle\tau(X_{i_1}\otimes\varepsilon_{i_1}),\varphi_{k\ell}\big\rangle_K\big\langle\tau(X_{i_2}\otimes\varepsilon_{i_2}),\varphi_{k\ell}\big\rangle_K\Bigg\} \\ &\qquad\times\Bigg\{\sum_{k',\ell'}\frac{1}{1+\lambda\rho_{k'\ell'}}\big\langle\tau(X_{i_1}\otimes\varepsilon_{i_1}),\varphi_{k'\ell'}\big\rangle_K\big\langle\tau(X_{i_2}\otimes\varepsilon_{i_2}),\varphi_{k'\ell'}\big\rangle_K\Bigg\}\Bigg] \\ &= 4\sum_{i_1<i_2}\sum_{k,k',\ell,\ell'}\Bigg[\frac{1}{(1+\lambda\rho_{k\ell})(1+\lambda\rho_{k'\ell'})}\mathrm{E}\Big\{\big\langle\tau(X_{i_1}\otimes\varepsilon_{i_1}),\varphi_{k\ell}\big\rangle_K\big\langle\tau(X_{i_1}\otimes\varepsilon_{i_1}),\varphi_{k'\ell'}\big\rangle_K\Big\} \\ &\qquad\times\mathrm{E}\Big\{\big\langle\tau(X_{i_2}\otimes\varepsilon_{i_2}),\varphi_{k\ell}\big\rangle_K\big\langle\tau(X_{i_2}\otimes\varepsilon_{i_2}),\varphi_{k'\ell'}\big\rangle_K\Big\}\Bigg] \\ &= 2n(n-1)\sum_{k,\ell}\frac{1}{(1+\lambda\rho_{k\ell})^2} = O(n^2\lambda^{-1/(2D)}),\end{aligned}$$

by Assumption A2.

In order to show the asymptotic normality of W_n, we use Proposition 3.2 in de Jong (1987) and show that

$$\begin{aligned}H_1 &= \sum_{i_1<i_2}\mathrm{E}(W^4_{i_1i_2}) \\ H_2 &= \sum_{i_1<i_2<i_3}\left\{\mathrm{E}(W^2_{i_1i_2}W^2_{i_2i_3}) + \mathrm{E}(W^2_{i_2i_1}W^2_{i_2i_3}) + \mathrm{E}(W^2_{i_3i_1}W^2_{i_3i_2})\right\}, \\ H_3 &= \sum_{i_1<i_2<i_3<i_4}\left\{\mathrm{E}(W_{i_1i_2}W_{i_1i_3}W_{i_4i_2}W_{i_4i_3}) + \mathrm{E}(W_{i_1i_2}W_{i_1i_4}W_{i_3i_2}W_{i_3i_4}) + \mathrm{E}(W_{i_1i_3}W_{i_1i_4}W_{i_2i_3}W_{i_2i_4})\right\}.\end{aligned}$$

are of order $o(\sigma^4_{W_n})$ as $n\to\infty$. Since $\mathrm{E}\{\langle\tau(X_{i_1}\otimes\varepsilon_{i_1}),\varphi_{k\ell}\rangle_K\} = 0$ due to Assumption A3, we have $\mathrm{E}(W_{i_1i_2}|\varepsilon_{i_1}, X_{i_1}) = 0$ for $i_1\neq i_2$. From (B.9), we obtain $\mathrm{E}\{\langle\tau(X_i\otimes\varepsilon_i),\varphi_{k\ell}\rangle^4_K\}\le c$, which implies that

$$\begin{aligned}H_1 = \sum_{i_1<i_2}\mathrm{E}(W^4_{i_1i_2}) &= 16\sum_{i_1<i_2}\mathrm{E}\Bigg[\Bigg\{\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\big\langle\tau(X_{i_1}\otimes\varepsilon_{i_1}),\varphi_{k\ell}\big\rangle_K\big\langle\tau(X_{i_2}\otimes\varepsilon_{i_2}),\varphi_{k\ell}\big\rangle_K\Bigg\}^4\Bigg] \\ &\le c\,n^2\left(\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\right)^4 \le c\,n^2\lambda^{-2/D}.\end{aligned}$$

Since $\mathrm{E}(W^2_{i_1i_2}W^2_{i_1i_3}) \le \mathrm{E}(W^4_{i_1i_2})$, we have $H_2 \le cn^3\lambda^{-2/D}$. Finally, for the term H_3, we use (B.8) and obtain,

$$\mathrm{E}(W_{i_1i_2}W_{i_1i_3}W_{i_4i_2}W_{i_4i_3})$$

$$
\begin{aligned}
&= 16\,\mathrm{E}\Bigg[\Bigg\{\sum_{k_1,\ell_1}\frac{1}{1+\lambda\rho_{k_1\ell_1}}\big\langle\tau(X_{i_1}\otimes\varepsilon_{i_1}),\varphi_{k_1\ell_1}\big\rangle_K\big\langle\tau(X_{i_2}\otimes\varepsilon_{i_2}),\varphi_{k_1\ell_1}\big\rangle_K\Bigg\}\\
&\quad\times\Bigg\{\sum_{k_2,\ell_2}\frac{1}{1+\lambda\rho_{k_2\ell_2}}\big\langle\tau(X_{i_1}\otimes\varepsilon_{i_1}),\varphi_{k_2\ell_2}\big\rangle_K\big\langle\tau(X_{i_3}\otimes\varepsilon_{i_3}),\varphi_{k_2\ell_2}\big\rangle_K\Bigg\}\\
&\quad\times\Bigg\{\sum_{k_3,\ell_3}\frac{1}{1+\lambda\rho_{k_3\ell_3}}\big\langle\tau(X_{i_4}\otimes\varepsilon_{i_4}),\varphi_{k_3\ell_3}\big\rangle_K\big\langle\tau(X_{i_2}\otimes\varepsilon_{i_2}),\varphi_{k_3\ell_3}\big\rangle_K\Bigg\}\\
&\quad\times\Bigg\{\sum_{k_4,\ell_4}\frac{1}{1+\lambda\rho_{k_4\ell_4}}\big\langle\tau(X_{i_4}\otimes\varepsilon_{i_4}),\varphi_{k_4\ell_4}\big\rangle_K\big\langle\tau(X_{i_3}\otimes\varepsilon_{i_3}),\varphi_{k_4\ell_4}\big\rangle_K\Bigg\}\Bigg]\\
&= 16\sum_{k_1,k_2,k_3,k_4,\ell_1,\ell_2,\ell_3,\ell_4}\delta_{k_1k_2}\,\delta_{k_1k_3}\,\delta_{k_2k_4}\,\delta_{k_3k_4}\,\delta_{\ell_1\ell_2}\,\delta_{\ell_1\ell_3}\,\delta_{\ell_2\ell_4}\,\delta_{\ell_3\ell_4}\prod_{j=1}^{4}\frac{1}{1+\lambda\rho_{k_j\ell_j}}\\
&= 16\sum_{k,\ell}\frac{1}{(1+\lambda\rho_{k\ell})^4}\,.
\end{aligned}
$$

We therefore deduce that $H_3 \le cn^4\lambda^{-1/(2D)}$, which yields

$$
H_1 + H_2 + H_3 = O(n^3\lambda^{-2/D} + n^4\lambda^{-1/(2D)}) = o(\sigma_{W_n}^4)
$$

since $n\lambda^{1/D} \to \infty$. By Proposition 3.2 in de Jong (1987), it follows that

$$
\Bigg\{2\sum_{k,\ell}\frac{1}{(1+\lambda\rho_{k\ell})^2}\Bigg\}^{-1/2}\frac{W_n}{n}\xrightarrow{d} N(0,1)\,. \tag{A.66}
$$

Next, we examine the second term $n^{-1}\sum_{i=1}^{n} W_{0,i}$ in (A.64). Note that the $W_{0,i}$'s in (A.65) are i.i.d. and satisfy

$$
\mathrm{E}(W_{0,i}) = \sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\mathrm{E}\Big\{\big\langle\tau(X_i\otimes\varepsilon_i),\varphi_{k\ell}\big\rangle_K^2\Big\} = \sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\,, \tag{A.67}
$$

where we used (B.8) in the last step. For the variance, by (B.9), $\mathrm{E}\{\langle\tau(X_i\otimes\varepsilon_i),\varphi_{k\ell}\rangle_K^4\} \le c$, so that by Assumption A2,

$$
\begin{aligned}
\mathrm{var}\Big(n^{-1}\sum_{i=1}^{n}W_{0,i}\Big) &= n^{-1}\,\mathrm{var}(W_{0,i}) \le n^{-1}\,\mathrm{E}(W_{0,i}^2) = n^{-1}\,\mathrm{E}\,\Bigg|\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\big\langle\tau(X_i\otimes\varepsilon_i),\varphi_{k\ell}\big\rangle_K^2\Bigg|^2\\
&\le c\,n^{-1}\Bigg\{\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\Bigg\}^2 = O(n^{-1}\lambda^{-1/D})\,.
\end{aligned} \tag{A.68}
$$

Therefore, the above equation and (A.67) yields that

$$
\frac{1}{n}\sum_{i=1}^{n}W_{0,i} = \sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}} + O_p(n^{-1/2}\lambda^{-1/(2D)}) = O_p(\lambda^{-1/(2D)} + n^{-1/2}\lambda^{-1/(2D)}) = O_p(\lambda^{-1/(2D)})\,,
$$

where we use Assumption A2, which yields $\mathrm{E}(W_{0,i}) \le c\,\lambda^{-1/2D}$. In view of (A.64), since we have shown $W_n/n = O_p(\lambda^{-1/(4D)})$, we therefore deduce from the above equation that

$$I_{1,1,n} = \frac{1}{n}\left\| \sum_{i=1}^{n} \tau(X_i \otimes \varepsilon_i) \right\|_K^2 = O_p(\lambda^{-1/(2D)})\,. \tag{A.69}$$

Moreover, since by Assumption A2, $\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2} = O(\lambda^{-1/(2D)})$, and in view of (A.68), $\mathrm{var}\big(n^{-1}\sum_{i=1}^n W_{0,i}\big) = O(n^{-1}\lambda^{-1/D}) = o(\lambda^{-1/(2D)})$ due to the assumption that $n^{-1}\lambda^{-1/(2D)} = o(1)$ in Assumption A5, combining (A.64), (A.66) and (A.67),

$$\begin{aligned}
&\left\{2\sum_{k,\ell}\frac{1}{(1+\lambda\rho_{k\ell})^2}\right\}^{-1/2}\{I_{1,1,n} - \mathrm{E}(I_{1,1,n})\} \\
&= \left\{2\sum_{k,\ell}\frac{1}{(1+\lambda\rho_{k\ell})^2}\right\}^{-1/2}\left(\frac{W_n}{n} + \frac{1}{n}\sum_{i=1}^n W_{0,i} - \sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\right) \xrightarrow{d} N(0,1)\,.
\end{aligned} \tag{A.70}$$

We now consider the term $I_{1,2,n}$ in (A.62). By Assumption A2, we obtain

$$W_\lambda(\beta_*) = \sum_{k,\ell} V(\beta_*,\varphi_{k\ell}) W_\lambda(\varphi_{k\ell}) = \lambda \sum_{k,\ell}\frac{V(\beta_*,\varphi_{k\ell})\,\rho_{k\ell}\,\varphi_{k\ell}}{1+\lambda\rho_{k\ell}}\,. \tag{A.71}$$

Since $J(\beta_*,\beta_*) = \sum_{k,\ell}\rho_{k\ell}V^2(\beta_*,\varphi_{k\ell}) < \infty$, we have $\lambda\sum_{k,\ell}\rho_{k\ell}^2\,V^2(\beta_*,\varphi_{k\ell})/(1+\lambda\rho_{k\ell})^2 = o(1)$ by the dominated convergence theorem. Since $\mathrm{E}\{\langle\tau(X_i\otimes\varepsilon_i), W_\lambda(\beta_*)\rangle_K\} = 0$, it follows from (2.17) and Assumption A2 that

$$\begin{aligned}
\mathrm{E}(I_{1,2,n}^2) &= n\,\mathrm{E}\Big\{\langle\tau(X_i\otimes\varepsilon_i), W_\lambda(\beta_*)\rangle_K\Big\}^2 = n\,\mathrm{E}\Big\{\big\langle X\otimes\varepsilon, W_\lambda(\beta_*)\big\rangle_{L^2}^2\Big\} \\
&= n\,V\{W_\lambda(\beta_*), W_\lambda(\beta_*)\} = n\lambda^2 V\left\{\sum_{k,\ell}\frac{\rho_{k\ell}V(\beta_*,\varphi_{k\ell})}{1+\lambda\rho_{k\ell}}\varphi_{k\ell}, \sum_{k,\ell}\frac{\rho_{k\ell}V(\beta_*,\varphi_{k\ell})}{1+\lambda\rho_{k\ell}}\varphi_{k\ell}\right\} \\
&= n\lambda^2\sum_{k,\ell}\frac{\rho_{k\ell}^2\,V^2(\beta_*,\varphi_{k\ell})}{(1+\lambda\rho_{k\ell})^2} = o(n\lambda)\,.
\end{aligned} \tag{A.72}$$

For the term $I_{1,3,n} = n\|W_\lambda(\beta_*)\|_K^2$ in (A.62), we use (A.71), the dominated convergence theorem and the fact that $\sum_{k,\ell}\rho_{k\ell}V^2(\beta_*,\varphi_{k\ell}) < \infty$, and obtain

$$n\|W_\lambda(\beta_*)\|_K^2 = n\left\|\sum_{k,\ell}\frac{\rho_{k\ell}V(\beta_*,\varphi_{k\ell})}{1+\lambda\rho_{k\ell}}\varphi_{k\ell}\right\|_K^2 = n\lambda^2\sum_{k,\ell}\frac{\rho_{k\ell}^2\,V^2(\beta_*,\varphi_{k\ell})}{1+\lambda\rho_{k\ell}} = o(n\lambda)\,. \tag{A.73}$$

Therefore, combining (A.61), (A.69), (A.72) and (A.74), we find

$$n\|S_{n,\lambda}(\beta_*)\|_K^2 = O_p(\lambda^{-1/(2D)}) + o(n\lambda) + o_p(n^{1/2}\lambda^{1/2}) = O_p(\lambda^{-1/(2D)})\,. \tag{A.74}$$

For the term $I_{2,n}$ in (A.60), by Lemma A.2 and Theorem 3.1, we find

$$|I_{2,n}| \le 2n\,\|\widehat{\beta}_\Delta + S_{n,\lambda}(\beta_*)\|_K \times \|\widehat{\beta}_\Delta\|_K = O_p(n v_n) \times O_p\big(\lambda^{1/2} + n^{-1/2}\lambda^{-1/(4D)}\big)\,,$$

where υ_n is defined in (A.34). For the term $I_{3,n}$ in (A.60), note that, in view of (3.2), for $H_n(\cdot)$ defined in (A.14),

$$\|\mathcal{D}S_{n,\lambda}(\beta_*)\widehat{\beta}_\Delta - \mathcal{D}S_\lambda(\beta_*)\widehat{\beta}_\Delta\|_K = n^{-1/2}\|H_n(\widehat{\beta}_\Delta)\|_K = O_p(\upsilon_n)\,,$$

where we used (A.33). Therefore, by Lemma A.2,

$$|I_{3,n}| \le n\,\|\mathcal{D}S_{n,\lambda}(\beta_*)\widehat{\beta}_\Delta - \mathcal{D}S_\lambda(\beta_*)\widehat{\beta}_\Delta\|_K \times \|\widehat{\beta}_\Delta\|_K = O_p(n\upsilon_n) \times O_p\big(\lambda^{1/2} + n^{-1/2}\lambda^{-1/(4D)}\big)\,.$$

For the term $I_{4,n}$ in (A.60), by (A.74) and Theorem 3.1,

$$|I_{4,n}| \le 2n\|S_{n,\lambda}(\beta_*)\|_K \times \|\widehat{\beta}_\Delta + S_{n,\lambda}(\beta_*)\|_K + n\|\widehat{\beta}_\Delta + S_{n,\lambda}(\beta_*)\|_K^2 \le O_p(\lambda^{-1/(2D)}\upsilon_n + n\upsilon_n^2)\,.$$

Combining (A.69), (A.72), (A.73) and the above convergence rates of $I_{2,n}, I_{3,n}, I_{4,n}$, it follows that $I_{1,2,n} + I_{1,3,n} + I_{2,n} + I_{3,n} + I_{4,n} = o_p(\lambda^{-1/(2D)})$. In addition, we use Assumption A2 and Lemma B.2 to obtain that $\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-1} \asymp \lambda^{-1/(2D)}$ and $\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2} \asymp \lambda^{-1/(2D)}$. Therefore, in view of (A.69), (A.70),

$$\left\{\sum_{k,\ell}\frac{2}{(1+\lambda\rho_{k\ell})^2}\right\}^{-1/2}\left(2n\,\mathfrak{L}_n(\beta_*) - \sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\right) \xrightarrow{d} N(0,1)\,.$$

Since for u_n and σ_n^2 in (4.11),

$$\sqrt{u_n} = \frac{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-1}}{\{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2}\}^{1/2}}\,; \quad \frac{\sigma_n^2}{\sqrt{u_n}} = \frac{\{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2}\}^{1/2}}{\sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-2}} = \left\{\sum_{k,\ell}\frac{1}{(1+\lambda\rho_{k\ell})^2}\right\}^{-1/2},$$

the proof is therefore complete.

A.8 Proof of Corollary 4.1 and (4.17)

By assumption, we have $n^{-1/2}\lambda^{-(2a+1)/(4D)}\log(n\lambda^{(2a+1)/(2D)}) = o(1)$. Therefore, Corollary 4.1 is a consequence of Theorem B.1 in Dette and Kokot (2021a) and Theorem 3.3. For a proof of (4.17), note that $d_\infty < \Delta$,

$$\begin{aligned}&\lim_{n\to\infty}\mathrm{P}\left\{\widehat{d}_\infty > \Delta + \frac{\mathcal{Q}_{1-\alpha}(T_\mathcal{E})}{\sqrt{n}\lambda^{(2a+1)/(4D)}}\right\}\\ &= \lim_{n\to\infty}\mathrm{P}\left\{\sqrt{n}\lambda^{(2a+1)/(4D)}(\widehat{d}_\infty - d_\infty) > \sqrt{n}\lambda^{(2a+1)/(4D)}(\Delta - d_\infty) + \mathcal{Q}_{1-\alpha}(T_\mathcal{E})\right\} = 0\,,\end{aligned}$$

since $\sqrt{n}\lambda^{(2a+1)/(4D)} \to \infty$ as $n\to\infty$, where $T_\mathcal{E}$ is defined in (4.16). If $d_\infty = \Delta$,

$$\lim_{n\to\infty}\mathrm{P}\left\{\widehat{d}_\infty > \Delta + \frac{\mathcal{Q}_{1-\alpha}(T_\mathcal{E})}{\sqrt{n}\lambda^{(2a+1)/(4D)}}\right\} = \lim_{n\to\infty}\mathrm{P}\left\{\sqrt{n}\lambda^{(2a+1)/(4D)}(\widehat{d}_\infty - d_\infty) > \mathcal{Q}_{1-\alpha}(T_\mathcal{E})\right\} = \alpha\,.$$

Under the alternative hypothesis H_1 in (4.13), i.e., $d_\infty > \Delta$,

$$\begin{aligned}&\lim_{n\to\infty}\mathrm{P}\left\{\widehat{d}_\infty > \Delta + \frac{\mathcal{Q}_{1-\alpha}(T_\mathcal{E})}{\sqrt{n}\lambda^{(2a+1)/(4D)}}\right\}\\ &= \lim_{n\to\infty}\mathrm{P}\left\{\sqrt{n}\lambda^{(2a+1)/(4D)}(\widehat{d}_\infty - d_\infty) > \sqrt{n}\lambda^{(2a+1)/(4D)}(\Delta - d_\infty) + \mathcal{Q}_{1-\alpha}(T_\mathcal{E})\right\} = 1\,.\end{aligned}$$

A.9 Proof of Theorem 4.4

We notice that, by the continuous mapping theorem, Theorem 3.3 and Lemma B.3 in Dette et al. (2020), conditional on the data, the bootstrap statistic $\widehat{T}^*_{\mathcal{E},n,q}$ in (4.21) converges to $T_{\mathcal{E}}$ in (4.16), the same limit as $\sqrt{n}\lambda^{(2a+1)/(4D)}(\widehat{d}_\infty - d_\infty)$. Hence, the assertion in Theorem 4.4 follows from arguments similar to the ones in the proof of (4.17) in Section A.8.

A.10 Proof of Theorem 4.5

By (3.6), we obtain, for the kernel C_{Z,x_0} in (4.25),

$$C_{Z,x_0}(t_1,t_2) = \lambda^{(2a+1)/(2D)} \sum_{k,\ell} \frac{1}{(1+\lambda\rho_{k\ell})^2} \int_0^1 \varphi_{k\ell}(s_1,t_1)x_0(s_1)ds_1 \times \int_0^1 \varphi_{k\ell}(s_2,t_2)x_0(s_2)ds_2 + o(1)\,, \tag{A.75}$$

since it follows from the dominated convergence theorem and the Cauchy-Schwarz inequality, uniformly in $n \geq 1$,

$$\lambda^{(2a+1)/(2D)} \left| \sum_{k,\ell} \frac{1}{(1+\lambda\rho_{k\ell})^2} \int_0^1 \varphi_{k\ell}(s_1,t_1)x_0(s_1)ds_1 \int_0^1 \varphi_{k\ell}(s_2,t_2)x_0(s_2)ds_2 \right|$$
$$\leq \lambda^{(2a+1)/(2D)} \|x_0\|^2_{L^2} \sum_{k,\ell} \frac{\|\varphi_{k\ell}\|^2_\infty}{(1+\lambda\rho_{k\ell})^2} \leq c\lambda^{(2a+1)/(2D)} \|x_0\|^2_{L^2} \sum_{k,\ell} \frac{(k\ell)^{2a}}{\{1+\lambda(k\ell)^{2D}\}^2} \leq c\,.$$

By the definition of τ in (2.17),

$$\widehat{\mu}_{x_0}(t) - \mu_{x_0}(t) = \int_0^1 \{\widehat{\beta}_n(s,t) - \beta_0(s,t)\}x_0(s)ds = \big\langle \widehat{\beta}_n - \beta_0, \tau(x_0 \otimes \delta_t)\big\rangle_K\,,$$

where δ_t is the delta function at $t \in [0,1]$. We have,

$$\sqrt{n}\lambda^{(2a+1)/(4D)}\{\widehat{\mu}_{x_0}(t) - \mu_{x_0}(t)\} = I_{1,n}(t) + I_{2,n}(t) + I_{3,n}(t)\,,$$

where

$$I_{1,n}(t) = \sqrt{n}\lambda^{(2a+1)/(4D)}\big\langle \widehat{\beta}_n - \beta_0 + S_{n,\lambda}\beta_0, \tau(x_0\otimes\delta_t)\big\rangle_K\,,$$
$$I_{2,n}(t) = \sqrt{n}\lambda^{(2a+1)/(4D)}\big\langle W_\lambda\beta_0, \tau(x_0\otimes\delta_t)\big\rangle_K\,,$$
$$I_{3,n}(t) = n^{-1/2}\lambda^{(2a+1)/(4D)} \sum_{i=1}^n \big\langle \tau(X_i\otimes\varepsilon_i), \tau(x_0\otimes\delta_t)\big\rangle_K\,.$$

Observing (2.17), Assumption A2 and Lemma B.2, and the Cauchy-Schwarz inequality, it follows

$$\sup_{t\in[0,1]} \|\tau(x_0\otimes\delta_t)\|^2_K = \sup_{t\in[0,1]} \sum_{k,\ell} \frac{\langle x_0\otimes\delta_t, \varphi_{k\ell}\rangle^2_{L^2(T)}}{1+\lambda\rho_{k\ell}} \leq \sup_{t\in[0,1]} \|x_0\otimes\delta_t\|^2_{L^2} \times \sum_{k,\ell} \frac{\|\varphi_{k\ell}\|^2_\infty}{1+\lambda\rho_{k\ell}}$$

$$\leq c \sum_{k,\ell} \frac{(k\ell)^{2a}}{1+\lambda(k\ell)^{2D}} \leq c\,\lambda^{-(2a+1)/(2D)}\,. \tag{A.76}$$

Hence, by Theorem 3.1 and the Cauchy-Schwarz inequality,

$$\begin{aligned}\sup_{t\in[0,1]} |I_{1,n}(t)| &\leq \sqrt{n}\lambda^{(2a+1)/(4D)} \|\widehat{\beta}_n - \beta_0 + S_{n,\lambda}\beta_0\|_K \times \sup_{t\in[0,1]} \|\tau(x_0\otimes\delta_t)\|_K \\ &= O_p(\sqrt{n}v_n) = o_p(1)\,. \end{aligned}\tag{A.77}$$

where v_n is defined in (A.34).

In addition, since by assumption, $\sum_{k,\ell}\rho_{k\ell}^2 V^2(\beta_0,\varphi_{k\ell}) < \infty$, we have

$$\|W_\lambda\beta_0\|_K^2 = \lambda^2 \left\| \sum_{k,\ell} \frac{\rho_{k\ell}\, V(\beta_0,\varphi_{k\ell})}{1+\lambda\rho_{k\ell}}\,\varphi_{k\ell} \right\|_K^2 = \lambda^2 \sum_{k,\ell} \frac{\rho_{k\ell}^2 V^2(\beta_0,\varphi_{k\ell})}{1+\lambda\rho_{k\ell}} \leq c\lambda^2\,.$$

Therefore, we obtain, for the term $I_{2,n}$,

$$\begin{aligned}\sup_{t\in[0,1]} |I_{2,n}(t)| &\leq \sqrt{n}\lambda^{(2a+1)/(4D)} \|W_\lambda\beta_0\|_K \times \sup_{t\in[0,1]} \|\tau(x_0\otimes\delta_t)\|_K \\ &\leq c\sqrt{n}\|W_\lambda\beta_0\|_K = O(\sqrt{n}\lambda) = o(1)\,. \end{aligned}\tag{A.78}$$

Finally, using the representation $\tau(X\otimes\varepsilon) = \sum_{k,\ell}(1+\lambda\rho_{k\ell})^{-1}\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}\varphi_{k\ell}$ (see equation (B.7) in the proof of Lemma B.7 in Section B), we obtain

$$\begin{aligned} I_{3,n}(t) &= n^{-1/2}\lambda^{(2a+1)/(4D)} \sum_{i=1}^n \big\langle \tau(X_i\otimes\varepsilon_i), \tau(x_0\otimes\delta_t)\big\rangle_K \\ &= n^{-1/2}\lambda^{(2a+1)/(4D)} \sum_{i=1}^n \sum_{k,\ell} \frac{1}{1+\lambda\rho_{k\ell}} \langle X_i\otimes\varepsilon_i,\varphi_{k\ell}\rangle_{L^2} \langle \varphi_{k\ell}, \tau(x_0\otimes\delta_t)\rangle_K \\ &= n^{-1/2}\lambda^{(2a+1)/(4D)} \sum_{i=1}^n \sum_{k,\ell} \frac{1}{1+\lambda\rho_{k\ell}} \langle X_i\otimes\varepsilon_i,\varphi_{k\ell}\rangle_{L^2} \langle \varphi_{k\ell}, x_0\otimes\delta_t\rangle_{L^2} \\ &= n^{-1/2}\lambda^{(2a+1)/(4D)} \sum_{i=1}^n \sum_{k,\ell} \frac{\langle X_i\otimes\varepsilon_i,\varphi_{k\ell}\rangle_{L^2}}{1+\lambda\rho_{k\ell}} \int_0^1 \varphi_{k\ell}(s,t)x_0(s)ds\,. \end{aligned}\tag{A.79}$$

For $1\leq i\leq n$, let

$$\begin{aligned} \mathfrak{U}_i(t) &= n^{-1/2}\lambda^{(2a+1)/(4D)} \big\langle \tau(X_i\otimes\varepsilon_i), \tau(x_0\otimes\delta_t)\big\rangle_K \\ &= n^{-1/2}\lambda^{(2a+1)/(4D)} \sum_{k,\ell} \frac{\langle X_i\otimes\varepsilon_i,\varphi_{k\ell}\rangle_{L^2}}{1+\lambda\rho_{k\ell}} \int_0^1 \varphi_{k\ell}(s,t)x_0(s)ds\,, \end{aligned}$$

so that $I_{3,n}(t) = \sum_{i=1}^n \mathfrak{U}_i(t)$. Since $\mathrm{E}\{\langle X_i\otimes\varepsilon_i,\varphi_{k\ell}\rangle_{L^2}\} = 0$ for $k,\ell\geq 1$, we have $\mathrm{E}\{\mathfrak{U}_i(t)\} = 0$, and observing (B.8), it follows that

$$\mathrm{cov}\{\mathfrak{U}_i(t_1),\mathfrak{U}_i(t_2)\}$$

$$
\begin{aligned}
&= \lambda^{(2a+1)/(2D)} \sum_{k,\ell,k',\ell'} \frac{1}{(1+\lambda\rho_{k\ell})(1+\lambda\rho_{k'\ell'})} \mathrm{E}\Big\{ \langle X\otimes\varepsilon, \varphi_{k\ell}\rangle_{L^2} \langle X\otimes\varepsilon, \varphi_{k'\ell'}\rangle_{L^2} \Big\} \\
&\quad \times \int \varphi_{k\ell}(s_1,t_1)x_0(s_1)ds_1 \times \int \varphi_{k\ell}(s_2,t_2)x_0(s_2)ds_2 \\
&= \lambda^{(2a+1)/(2D)} \sum_{k,\ell} \frac{1}{(1+\lambda\rho_{k\ell})^2} \int \varphi_{k\ell}(s_1,t_1)x_0(s_1)ds_1 \times \int \varphi_{k\ell}(s_2,t_2)x_0(s_2)ds_2 \\
&= C_{Z,x_0}(t_1,t_2) + o(1)\,, \qquad \text{as } n\to\infty\,, \qquad\qquad (\mathrm{A}.80)
\end{aligned}
$$

where we used (A.75) in the last step.

In order to prove the weak convergence of the finite-dimensional marginal distributions of $\widehat{\mu}_{x_0}$, by the Cramér-Wold device, we shall show that, for any $q\in\mathbb{N}$, $(c_1,\ldots,c_q)^{\mathrm{T}}\in\mathbb{R}^q$ and $t_1,\ldots,t_q\in[0,1]$,

$$
\sum_{j=1}^q c_j\widehat{\mu}_{x_0}(t_j) \xrightarrow{d.} \sum_{j=1}^q c_j Z_{x_0}(t_j)\,. \qquad (\mathrm{A}.81)
$$

In view of (A.77) and (A.78), we deduce that

$$
\sum_{j=1}^q c_j\widehat{\mu}_{x_0}(t_j) = \sum_{j=1}^q c_j I_{3,n}(t_j) + \sum_{j=1}^q c_j\{I_{n,1}(t_j)+I_{n,2}(t_j)\} = \sum_{i=1}^n\sum_{j=1}^q c_j\mathfrak{U}_i(t_j) + o_p(1)\,.
$$

Observing (A.80), we have

$$
\mathrm{var}\Big\{ \sum_{i=1}^n\sum_{j=1}^q c_j\mathfrak{U}_i(t_j) \Big\} = \sum_{j_1,j_2=1}^q c_{j_1}c_{j_2}C_{Z,x_0}(t_{j_1},t_{j_2}) + o(1)
$$

as $n\to\infty$. If $\sum_{j_1,j_2=1}^q c_{j_1}c_{j_2}C_{Z,x_0}(t_{j_1},t_{j_2})=0$, $\sum_{j=1}^q c_jZ_{x_0}(t_j)$ has a degenerate distribution with a point mass at zero, so that (A.81) is a consequence of the Markov's inequality. If $\sum_{j_1,j_2=1}^q c_{j_1}c_{j_2}C_{Z,x_0}(t_{j_1},t_{j_2})\neq 0$, we have $\mathrm{var}\big\{\sum_{i=1}^n\sum_{j=1}^q c_j\mathfrak{U}_i(t_j)\big\} = \sum_{j_1,j_2=1}^q c_{j_1}c_{j_2}C_{Z,x_0}(t_{j_1},t_{j_2}) + o(1) = \mathrm{var}\big\{\sum_{j=1}^q c_jZ_{x_0}(t_j)\big\} + o(1)$. To prove (A.81), we shall check that the triangular array of random variables $\{\sum_{j=1}^q c_j\mathfrak{U}_i(t_j)\}_{i=1}^n$ satisfies Lindeberg's condition. Let $\Sigma_q=\sum_{j=1}^q|c_j|$. We have $\Sigma_q>0$ since $\Sigma_q=0$ indicates $\sum_{j_1,j_2=1}^q c_{j_1}c_{j_2}C_{Z,x_0}(t_{j_1},t_{j_2})=0$. For any $e>0$, by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\sum_{i=1}^n \mathrm{E}\Bigg[\bigg|\sum_{j=1}^q c_j\mathfrak{U}_i(t_j)\bigg|^2 \times \mathbb{1}\bigg\{\bigg|\sum_{j=1}^q c_j\mathfrak{U}_i(t_j)\bigg|>e\bigg\}\Bigg] \\
&= \lambda^{(2a+1)/(2D)}\,\mathrm{E}\Bigg[\bigg|\sum_{j=1}^q c_j\big\langle\tau(X_i\otimes\varepsilon_i),\tau(x_0\otimes\delta_{t_j})\big\rangle_K\bigg|^2 \times \mathbb{1}\bigg\{\bigg|\sum_{j=1}^q c_j\mathfrak{U}_i(t_j)\bigg|>e\bigg\}\Bigg] \\
&\le c\,\lambda^{(2a+1)/(2D)} \sup_{t\in[0,1]} \mathrm{E}\Big\{\big|\big\langle\tau(X_i\otimes\varepsilon_i),\tau(x_0\otimes\delta_t)\big\rangle_K\big|^4\Big\}^{\frac12} \times \mathrm{P}\bigg\{\bigg|\sum_{j=1}^q c_j\mathfrak{U}_i(t_j)\bigg|>e\bigg\}^{\frac12}. \qquad (\mathrm{A}.82)
\end{aligned}
$$

Using (A.76) and Lemma B.7 in Section B.1, it follows that

$$\sup_{t\in[0,1]} \mathrm{E}\big|\big\langle \tau(X_i\otimes\varepsilon_i), \tau(x_0\otimes\delta_t)\big\rangle_K\big|^4$$
$$\le \mathrm{E}\|\tau(X_i\otimes\varepsilon_i)\|_K^4 \times \sup_{t\in[0,1]} \|\tau(x_0\otimes\delta_t)\|_K^4 \le c\,\lambda^{-(2a+2)/D}\,. \tag{A.83}$$

In addition, for some $c_0>0$,

$$\sup_{t\in[0,1]}\Big|\sum_{k,\ell}\frac{\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}}{1+\lambda\rho_{k\ell}}\int_0^1\varphi_{k\ell}(s,t)x_0(s)ds\Big|$$
$$\le \|x_0\|_{L^2}\times\|X\otimes\varepsilon\|_{L^2}\times\sum_{k,\ell}\frac{\|\varphi_{k\ell}\|_\infty^2}{1+\lambda\rho_{k\ell}} \le c_0\lambda^{-(2a+1)/(2D)}\|X\|_{L^2}\,\|\varepsilon\|_{L^2}\,.$$

Hence, by arguments similar to the ones used in (A.43), by taking $c_1>a(c_\varepsilon D)^{-1}$, for $c_\varepsilon>0$ in Assumption A4, we find

$$\mathrm{P}\Big\{\Big|\sum_{j=1}^q c_j\mathfrak{U}_i(t_j)\Big|>e\Big\}$$
$$=\mathrm{P}\Big\{\sup_{t\in[0,1]}\Big|\sum_{k,\ell}\frac{\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}}{1+\lambda\rho_{k\ell}}\int_0^1\varphi_{k\ell}(s,t)x_0(s)ds\Big|>e\Sigma_q^{-1}\sqrt{n}\lambda^{-(2a+1)/(4D)}\Big\}$$
$$\le\mathrm{P}\Big\{\|X\|_{L^2}\,\|\varepsilon\|_{L^2}>ec_0^{-1}\Sigma_q^{-1}\sqrt{n}\lambda^{(2a+1)/(4D)}\Big\}$$
$$\le\mathrm{P}\Big\{\|X\|_{L^2}\ge ec_1^{-1}c_0^{-1}\Sigma_q^{-1}\sqrt{n}\lambda^{(2a+1)/(4D)}/\log(\lambda^{-1})\Big\}+\mathrm{P}\Big\{\|\varepsilon\|_{L^2}\ge c_1\log(\lambda^{-1})\Big\}$$
$$\le\exp\big\{-c_Xec_1^{-1}c_0^{-1}\Sigma_q^{-1}\sqrt{n}\lambda^{(2a+1)/(4D)}/\log(\lambda^{-1})\big\}\mathrm{E}\{\exp(c_X\|X\|_{L^2})\}$$
$$+\lambda^{c_1c_\varepsilon}\mathrm{E}\{\exp(c_\varepsilon\|\varepsilon\|_{L^2})\}$$
$$=O\big(\lambda^{c_Xec_1^{-1}c_0^{-1}\Sigma_q^{-1}\{\sqrt{n}\lambda^{(2a+1)/(4D)}/\log^2(\lambda^{-1})\}}\big)+O(\lambda^{c_1c_\varepsilon})=o(\lambda^{1/D})\,,$$

where we used the assumption $\sqrt{n}\lambda^{(2a+1)/(4D)}/\log^2(\lambda^{-1})\to\infty$ in the last step. Therefore, combining the above result with (A.82) and (A.83) yields

$$\sum_{i=1}^n\mathrm{E}\Big[\Big|\sum_{j=1}^q c_j\mathfrak{U}_i(t_j)\Big|^2\times\mathbb{1}\Big\{\Big|\sum_{j=1}^q c_j\mathfrak{U}_i(t_j)\Big|>e\Big\}\Big]\le c\,\lambda^{(2a+1)/(2D)}\,\lambda^{-(a+1)/D}\,o(\lambda^{1/(2D)})=o(1)\,.$$

By Lindeberg's CLT,

$$\sum_{j=1}^q c_j\widehat{\mu}_{x_0}(t_j)=\sum_{i=1}^n\sum_{j=1}^q c_j\mathfrak{U}_i(t_j)+o_p(1)\xrightarrow{d}N\Big(0,\sum_{j_1,j_2=1}^q c_{j_1}c_{j_2}C_{Z,x_0}(t_{j_1},t_{j_2})\Big)\overset{d.}{=}\sum_{j=1}^q c_jZ_{x_0}(t_j)\,.$$

Next, we prove the asymptotic tightness of $\widehat{\mu}_{x_0}$. For any $t_1,t_2\in[0,1]$, since $\mathrm{E}\{I_{3,n}(t_1)-I_{3,n}(t_2)\}=0$, by (B.8) and the Cauchy-Schwarz inequality, we find

$$\mathrm{E}|I_{3,n}(t_1)-I_{3,n}(t_2)|^2=n\,\mathrm{E}|\mathfrak{U}_i(t_1)-\mathfrak{U}_i(t_2)|^2$$

$$= \lambda^{(2a+1)/(2D)} \mathrm{E} \Big| \sum_{k,\ell} \frac{\langle X_i \otimes \varepsilon_i, \varphi_{k\ell} \rangle_{L^2}}{1 + \lambda \rho_{k\ell}} \int_0^1 \{\varphi_{k\ell}(s, t_1) - \varphi_{k\ell}(s, t_2)\} x_0(s) ds \Big|^2$$
$$= \lambda^{(2a+1)/(2D)} \sum_{k,\ell} \frac{1}{(1 + \lambda \rho_{k\ell})^2} \Big| \int_0^1 \{\varphi_{k\ell}(s, t_1) - \varphi_{k\ell}(s, t_2)\} x_0(s) ds \Big|^2$$
$$\leq \lambda^{(2a+1)/(2D)} \|x_0\|_{L^2}^2 \sum_{k,\ell} \frac{1}{(1 + \lambda \rho_{k\ell})^2} \int_0^1 |\varphi_{k\ell}(s, t_1) - \varphi_{k\ell}(s, t_2)|^2 ds \,.$$

By the assumption in (3.7), we have, for constants c_0, ϑ and b specified in Theorem 3.3,

$$\mathrm{E}|I_{3,n}(t_1) - I_{3,n}(t_2)|^2 \leq c_0 \lambda^{(a-b)/D} \|x_0\|_{L^2}^2 |t_1 - t_2|^{2\vartheta} \leq c\, \lambda^{(a-b)/D} |t_1 - t_2|^{2\vartheta} \,.$$

Moreover, in view of (A.79), by the Cauchy-Schwarz inequality, Assumption A2, we deduce that

$$\sup_{t \in [0,1]} |\mathfrak{U}_i(t)|$$
$$\leq n^{-1/2} \lambda^{(2a+1)/(4D)} \sum_{k,\ell} \frac{|\langle X_i \otimes \varepsilon_i, \varphi_{k\ell} \rangle_{L^2}|}{1 + \lambda \rho_{k\ell}} \times \sup_{t \in [0,1]} \Big| \int_0^1 \varphi_{k\ell}(s, t) x_0(s) ds \Big|$$
$$\leq n^{-1/2} \lambda^{(2a+1)/(4D)} \|X_i \otimes \varepsilon_i\|_{L^2} \|x_0\|_{L^2} \sum_{k,\ell} \frac{\|\varphi_{k\ell}\|_{L^2}}{1 + \lambda \rho_{k\ell}} \times \sup_{t \in [0,1]} \left\{ \int_0^1 |\varphi_{k\ell}(s, t)|^2 ds \right\}^{1/2}$$
$$\leq c\, n^{-1/2} \lambda^{(2a+1)/(4D)} (\log n)^2 \sum_{k,\ell} \frac{\|\varphi_{k\ell}\|_\infty^2}{1 + \lambda \rho_{k\ell}}$$
$$\leq c\, n^{-1/2} \lambda^{(2a+1)/(4D)} (\log n)^2 \sum_{k,\ell} \frac{(k\ell)^{2a}}{1 + \lambda (k\ell)^{2D}}$$
$$\leq c\, n^{-1/2} \lambda^{-(2a+1)/(4D)} (\log n)^2 \,,$$

almost surely, where we used Lemma B.2 and the fact that $\|X \otimes \varepsilon\|_{L^2} \leq c(\log n)^2$ almost surely from Section A.4. Therefore, by arguments similar to the ones used in the proof of Theorem 3.3, we find that, there exists a semi-metric d on $[0, 1]^2$ such that, for any $e > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathrm{P} \left\{ \sup_{d\{(s_1,t_1),(s_2,t_2)\} \leq \delta} |I_{3,n}(t_1) - I_{3,n}(t_2)| > e \right\} = 0 \,.$$

Combining the above result with (A.77) and (A.78),

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathrm{P} \left\{ \sup_{d\{(s_1,t_1),(s_2,t_2)\} \leq \delta} |\widehat{\mu}_{x_0}(t_1) - \widehat{\mu}_{x_0}(t_2)| > e \right\} = 0 \,.$$

Therefore, by applying Theorems 1.5.4 and 1.5.7 in van der Vaart and Wellner (1996), we have shown that $\sqrt{n} \lambda^{(2a+1)/(4D)} \{\widehat{\mu}_{x_0} - \mu_{x_0}\} \rightsquigarrow Z_{x_0}$ in $C([0, 1]^2)$. By the continuous mapping theorem, $\sqrt{n} \lambda^{(2a+1)/(4D)} \sup_{t \in [0,1]} |\widehat{\mu}_{x_0}(t) - \mu_{x_0}(t)| \xrightarrow{d} \max_{t \in [0,1]} |Z_{x_0}(t)|$.

B Auxiliary lemmas and technical details

B.1 Auxiliary lemmas for the proofs in Section A

Lemma B.1. *For any $\beta_1, \beta_2 \in \mathcal{H}$ and $x \in L^2([0,1])$, for τ defined in (2.16),*

$$\left\langle \tau\left[x \otimes \left\{\int_0^1 \beta_1(s,\cdot)x(s)ds\right\}\right], \beta_2\right\rangle_K = \left\langle \tau\left[x \otimes \left\{\int_0^1 \beta_2(s,\cdot)x(s)ds\right\}\right], \beta_1\right\rangle_K$$
$$= \int_{[0,1]^3} \beta_1(s_1,t)\,\beta_2(s_2,t)\,x(s_1)\,x(s_2)\,ds_1\,ds_2\,dt\,.$$

Proof. By Fubini's theorem and (2.17), direct calculation yields

$$\begin{aligned}
&\left\langle \tau\left[x \otimes \left\{\int_0^1 \beta_1(s,\cdot)x(s)ds\right\}\right], \beta_2\right\rangle_K \\
&= \int_0^1\int_0^1 x(s_2)\left\{\int_0^1 \beta_1(s_1,t)\,x(s_1)\,ds_1\right\}\beta_2(s_2,t)\,ds_2\,dt \\
&= \int_{[0,1]^3} \beta_1(s_1,t)\,\beta_2(s_2,t)\,x(s_1)\,x(s_2)\,ds_1\,ds_2\,dt \\
&= \int_0^1\int_0^1 x(s_1)\left\{\int_0^1 \beta_2(s_2,t)\,x(s_2)\,ds_2\right\}\beta_1(s_1,t)\,ds_1\,dt \\
&= \left\langle \tau\left[x \otimes \left\{\int_0^1 \beta_2(s,\cdot)\,x(s)\,ds\right\}\right], \beta_1\right\rangle_K .
\end{aligned}$$

□

Lemma B.2. *For D in Assumption A2, for any $0 \le \nu < D - 1/2$, for $0 < \lambda < 1$ and for either $r = 1$ or 2, there exist constants $c_1, c_2 > 0$ independent of λ such that*

$$c_1\,\lambda^{-(2\nu+1)/(2D)} \le \sum_{k,\ell \ge 1} \frac{(k\ell)^{2\nu}}{\{1+\lambda(k\ell)^{2D}\}^r} \le c_2\,\lambda^{-(2\nu+1)/(2D)}\,. \tag{B.1}$$

Proof. Using change of variables, we first notice that

$$\int_0^\infty\int_0^\infty \frac{x^{2\nu}y^{2\nu}}{(1+\lambda\,x^{2D}y^{2D})^r}\,dx\,dy = \lambda^{-(2\nu+1)/(2D)}\int_0^\infty\int_0^\infty \frac{x^{2\nu}y^{2\nu}}{(1+x^{2D}y^{2D})^r}\,dx\,dy\,.$$

Since $2D - 2\nu > 1$, we have $\int_0^\infty\int_0^\infty x^{2\nu}y^{2\nu}(1+x^{2D}y^{2D})^{-r}dxdy < \infty$.

Let $\mathfrak{m}_\lambda = \lambda^{-1/(2D)}\{\nu/(rD-\nu)\}^{1/(2D)}$. Note that the function $x^{2\nu}/(1+\lambda x^{2D})$ is increasing on $(0,\mathfrak{m}_\lambda)$, and is decreasing on $(\mathfrak{m}_\lambda,\infty)$. For any real value $x \in \mathbb{R}$, let $\lceil x \rceil$ denote the smallest integer greater than or equal to x, and let $\lfloor x \rfloor$ denote the largest integer smaller than or equal to x. Let $\mathbb{1}\{\cdot\}$ denote the indicator function. For the left-hand side of the inequality in (B.1), note that

$$\sum_{k,\ell\ge 1} \frac{(k\ell)^{2\nu}}{\{1+\lambda(k\ell)^{2D}\}^r}$$

$$
\begin{aligned}
&= \sum_{k,\ell\ge 1} \frac{(k\ell)^{2\nu}}{\{1+\lambda(k\ell)^{2D}\}^r}\big[\mathbb{1}\{k\ell < \lfloor \mathfrak{m}_\lambda\rfloor\} + \mathbb{1}\{k\ell > \lceil \mathfrak{m}_\lambda\rceil\}\big] \\
&\quad + \sum_{k,\ell\ge 1} \frac{(k\ell)^{2\nu}}{\{1+\lambda(k\ell)^{2D}\}^r}\mathbb{1}\{\lfloor \mathfrak{m}_\lambda\rfloor \le k\ell \le \lceil \mathfrak{m}_\lambda\rceil\} \\
&\le \int_0^\infty\int_0^\infty \frac{(xy)^{2\nu}}{\{1+\lambda(xy)^{2D}\}^r}dxdy + \sum_{k,\ell\ge 1} \frac{(k\ell)^{2\nu}}{\{1+\lambda(k\ell)^{2D}\}^r}\mathbb{1}\{\lfloor \mathfrak{m}_\lambda\rfloor \le k\ell \le \lceil \mathfrak{m}_\lambda\rceil\} \\
&\le \int_0^\infty\int_0^\infty \frac{(xy)^{2\nu}}{\{1+\lambda(xy)^{2D}\}^r}dxdy + \frac{\lceil \mathfrak{m}_\lambda\rceil^{2\nu}}{(1+\lambda\lfloor \mathfrak{m}_\lambda\rfloor^{2D})^r}\sum_{k,\ell\ge 1}\mathbb{1}\{\lfloor \mathfrak{m}_\lambda\rfloor \le k\ell \le \lceil \mathfrak{m}_\lambda\rceil\} \\
&\le \int_0^\infty\int_0^\infty \frac{(xy)^{2\nu}}{\{1+\lambda(xy)^{2D}\}^r}dxdy + \frac{2\lceil \mathfrak{m}_\lambda\rceil^{2\nu+1}}{(1+\lambda\lfloor \mathfrak{m}_\lambda\rfloor^{2D})^r} \\
&\le \int_0^\infty\int_0^\infty \frac{(xy)^{2\nu}}{\{1+\lambda(xy)^{2D}\}^r}dxdy + c\,\lambda^{-(2\nu+1)/(2D)}\,,
\end{aligned}
$$

for some $c > 0$ that does not depend on λ, where we used the fact that $\sum_{k,\ell\ge 1}\mathbb{1}\{\lfloor \mathfrak{m}_\lambda\rfloor \le k\ell \le \lceil \mathfrak{m}_\lambda\rceil\} \le 2\lceil \mathfrak{m}_\lambda\rceil$. This proves the right-hand side of (B.1).

For the left-hand side of the inequality in (B.1), by change of variables, we find

$$
\begin{aligned}
&\sum_{k,\ell\ge 1} \frac{(k\ell)^{2\nu}}{\{1+\lambda(k\ell)^{2D}\}^r} \\
&= \sum_{k,\ell\ge 1} \frac{(k\ell)^{2\nu}}{\{1+\lambda(k\ell)^{2D}\}^r}\mathbb{1}\{k\ell \le \lfloor \mathfrak{m}_\lambda\rfloor\} + \sum_{k,\ell\ge 1} \frac{(k\ell)^{2\nu}}{\{1+\lambda(k\ell)^{2D}\}^r}\mathbb{1}\{k\ell \ge \lceil \mathfrak{m}_\lambda\rceil\} \\
&\ge \int_0^\infty\int_0^\infty \frac{(xy)^{2\nu}}{\{1+\lambda(xy)^{2D}\}^r}\big[\mathbb{1}\{xy \le \lfloor \mathfrak{m}_\lambda\rfloor\} + \mathbb{1}\{xy \ge \lceil \mathfrak{m}_\lambda\rceil\}\big]dxdy \\
&= \lambda^{-(2\nu+1)/(2D)}\int_0^\infty\int_0^\infty \frac{(xy)^{2\nu}}{\{1+(xy)^{2D}\}^r}\big[\mathbb{1}\{xy \le \lambda^{1/(2D)}\lfloor \mathfrak{m}_\lambda\rfloor\} + \mathbb{1}\{xy \ge \lambda^{1/(2D)}\lceil \mathfrak{m}_\lambda\rceil\}\big]dxdy\,.
\end{aligned}
$$

Note that there exists constants $0 < \widetilde{c}_1 < \widetilde{c}_2$ independent of λ such that $0 < \widetilde{c}_1 \le \lambda^{1/(2D)}\lfloor \mathfrak{m}_\lambda\rfloor$ and $\lambda^{1/(2D)}\lceil \mathfrak{m}_\lambda\rceil \le \widetilde{c}_2 < \infty$. We therefore deduce from the above equation that

$$
\sum_{k,\ell\ge 1} \frac{(k\ell)^{2\nu}}{\{1+\lambda(k\ell)^{2D}\}^r} \ge \lambda^{-(2\nu+1)/(2D)}\int_0^\infty\int_0^\infty \frac{(xy)^{2\nu}}{\{1+(xy)^{2D}\}^r}\big[\mathbb{1}\{xy > \widetilde{c}_2\} + \mathbb{1}\{xy \le \widetilde{c}_1\}\big]dxdy\,,
$$

which completes the proof of (B.1).

□

Lemma B.3. *Under Assumptions A1–A4, for $\beta \in \mathcal{H}$, we have, for some constant $c_K > 0$,*

$$
\|\beta\|_{L^2} \le c_K \lambda^{-(2a+1)/(4D)}\,\|\beta\|_K\,.
$$

Proof. For any $\beta \in \mathcal{H}$ and the reproducing kernel K in (2.14), we have $\beta(s,t) = \langle \beta, K_{(s,t)}\rangle_K$, from which we deduce that $|\beta(s,t)| \le \|\beta\|_K\,\|K_{(s,t)}\|_K$, so that

$$
\|\beta\|_{L^2}^2 \le \|\beta\|_K^2 \int_0^1\int_0^1 \|K_{(s,t)}\|_K^2\,ds\,dt\,,
$$

which yields

$$\|K_{(s,t)}\|_K^2 = \sum_{k,\ell} \frac{\langle K_{(s,t)}, \varphi_{k\ell}\rangle_K^2}{1+\lambda\rho_{k\ell}} = \sum_{k,\ell} \frac{\varphi_{k\ell}^2(s,t)}{1+\lambda\rho_{k\ell}}\,. \tag{B.2}$$

Therefore, we find

$$\int_0^1\int_0^1 \|K_{(s,t)}\|_K^2\,ds\,dt = \sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\|\varphi_{k\ell}\|_{L^2}^2 \le c\sum_{k,\ell}\frac{k^{2a}\ell^{2a}}{1+\lambda\, k^{2D}\ell^{2D}} \le c\lambda^{-(2a+1)/(2D)}\,,$$

where we used the assumption that $D > a + 1/2$ in Assumption A2 and Lemma B.2 in the last step.

□

Lemma B.4. *Under Assumptions A1–A4, for any $\beta \in \mathcal{H}$,*

$$\left\|\tau\left[x\otimes\left\{\int_0^1 \beta(s,\cdot)x(s)ds\right\}\right]\right\|_K \le c_1\,\lambda^{-(2a+1)/(2D)}\,\|\beta\|_K \times \|x\|_{L^2}^2\,, \tag{B.3}$$

$$\mathrm{E}\left\|\tau\left[X_i\otimes\left\{\int_0^1 \beta(s,\cdot)X_i(s)ds\right\}\right]\right\|_K^2 \le c_2\,\lambda^{-1/D}\,\|\beta\|_K^2\,. \tag{B.4}$$

Here, $c_1, c_2 > 0$ are absolute constants.

Proof. By the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\left\|\tau\left[x\otimes\left\{\int_0^1 \beta(s,\cdot)x(s)ds\right\}\right]\right\|_K \\
&= \sup_{\|\gamma\|_K=1}\left|\left\langle \tau\left[x\otimes\left\{\int_0^1 \beta(s,\cdot)x(s)ds\right\}\right],\gamma\right\rangle_K\right| \\
&= \sup_{\|\gamma\|_K=1}\left|\int_{[0,1]^3}\beta(s_1,t)\,\gamma(s_2,t)\,x(s_1)\,x(s_2)\,ds_1\,ds_2\,dt\right| \\
&\le \sup_{\|\gamma\|_K=1}\left\{\int_{[0,1]^3}\gamma^2(s_1,t)\,x^2(s_2)\,ds_1\,ds_2\,dt\right\}^{1/2}\times\left\{\int_{[0,1]^3}\beta^2(s_2,t)\,x^2(s_1)\,ds_1\,ds_2\,dt\right\}^{1/2} \\
&= \sup_{\|\gamma\|_K=1}\|\gamma\|_{L^2}\times\|\beta\|_{L^2}\times\|x\|_{L^2}^2 \\
&\le c_K\,\lambda^{(2a+1)/(4D)}\,\|\beta\|_{L^2}\times\|x\|_{L^2}^2 \\
&\le c_K^2\,\lambda^{(2a+1)/(2D)}\,\|\beta\|_K\times\|x\|_{L^2}^2\,,
\end{aligned}$$

where we used Lemma B.3. This proves (B.3).

In order to prove (B.4), by Lemma B.1, we find

$$\mathrm{E}\left(\left\langle\tau\left[X_i\otimes\left\{\int_0^1 \varphi_{k\ell}(s,\cdot)X_i(s)ds\right\}\right],\varphi_{k'\ell'}\right\rangle_K^2\right)$$

$$= \mathrm{E}\bigg\{ \int_{[0,1]^3} X_i(s_1)X_i(s_2)\varphi_{k\ell}(s_1,t)\varphi_{k'\ell'}(s_2,t)ds_1ds_2dt \bigg\}^2$$
$$= \mathrm{E}\bigg\{ \int_{[0,1]^3} X_i(s_1)X_i(s_2)x_{k\ell}(s_1)x_{k'\ell'}(s_1)ds_1ds_2 \bigg\}^2 \times \bigg\{ \int_0^1 \eta_\ell(t)\eta_{\ell'}(t)dt \bigg\}^2$$
$$= \mathrm{E}\bigg[\bigg\{ \int_0^1 X_i(s_1)x_{k\ell}(s_1)ds_1 \bigg\}^2 \times \bigg\{ \int_0^1 X_i(s_2)x_{k'\ell'}(s_2)ds_2 \bigg\}^2 \bigg] \times \bigg\{ \int_0^1 \eta_\ell(t)\eta_{\ell'}(t)dt \bigg\}^2 .$$

Using the Cauchy-Schwarz inequality, by Assumptions A2 and A4, we deduce from the above equation that, for the constant $c_0 > 0$ in (3.4),

$$\begin{aligned}
&\mathrm{E}\bigg(\bigg\langle \tau\bigg[X_i \otimes \bigg\{ \int_0^1 \varphi_{k\ell}(s,\cdot)X_i(s)ds \bigg\}\bigg], \varphi_{k'\ell'} \bigg\rangle_K^2 \bigg) \\
&\le \bigg[\mathrm{E}\bigg\{ \int_0^1 X_i(s_1)x_{k\ell}(s_1)ds_1 \bigg\}^4 \bigg]^{\frac{1}{2}} \bigg[\mathrm{E}\bigg\{ \int_0^1 X_i(s_2)x_{k'\ell'}(s_2)ds_2 \bigg\}^4 \bigg]^{\frac{1}{2}} \times \|\eta_\ell\|_{L^2}^2 \times \|\eta_{\ell'}\|_{L^2}^2 \\
&\le c_0\, \mathrm{E}\bigg\{ \int_0^1 X_i(s_1)x_{k\ell}(s_1)ds_1 \bigg\}^2 \times \mathrm{E}\bigg\{ \int_0^1 X_i(s_2)x_{k'\ell'}(s_2)ds_2 \bigg\}^2 \times \|\eta_\ell\|_{L^2}^2 \times \|\eta_{\ell'}\|_{L^2}^2 \\
&= c_0 \bigg\{ \int_{[0,1]^2} C_X(s_1,s_2)x_{k\ell}(s_1)x_{k\ell}(s_2)ds_1ds_2 \bigg\} \\
&\qquad \times \bigg\{ \int_{[0,1]^2} C_X(s_1,s_2)x_{k'\ell'}(s_1)x_{k'\ell'}(s_2)ds_1ds_2 \bigg\} \times \|\eta_\ell\|_{L^2}^2 \times \|\eta_{\ell'}\|_{L^2}^2 \\
&= c_0 \,\langle \mathcal{C}_X(x_{k\ell}), x_{k\ell}\rangle_{L^2} \,\langle \mathcal{C}_X(x_{k'\ell'}), x_{k'\ell'}\rangle_{L^2} \,\|\eta_\ell\|_{L^2}^2 \,\|\eta_{\ell'}\|_{L^2}^2 \\
&= c_0\, V(x_{k\ell}\otimes\eta_\ell, x_{k\ell}\otimes\eta_\ell)\, V(x_{k'\ell'}\otimes\eta_{\ell'}, x_{k'\ell'}\otimes\eta_{\ell'}) \\
&= c_0\, V(\varphi_{k\ell},\varphi_{k\ell})\, V(\varphi_{k'\ell'},\varphi_{k'\ell'}) = c_0 \,.
\end{aligned}$$

In view of (A.63), the above equation implies that

$$\begin{aligned}
&\mathrm{E}\bigg\| \tau\bigg[X_i \otimes \bigg\{ \int_0^1 \varphi_{k\ell}(s,\cdot)X_i(s)ds \bigg\}\bigg] \bigg\|_K^2 \\
&= \sum_{k',\ell'} \frac{1}{1+\lambda\rho_{k'\ell'}}\, \mathrm{E}\bigg(\bigg\langle \tau\bigg[X_i \otimes \bigg\{ \int_0^1 \varphi_{k\ell}(s,\cdot)X_i(s)ds \bigg\}\bigg], \varphi_{k'\ell'} \bigg\rangle_K^2 \bigg) \\
&\le c_0 \sum_{k',\ell'} \frac{1}{1+\lambda\rho_{k'\ell'}} \,.
\end{aligned} \tag{B.5}$$

Now, using (A.63) once again, by Lemma B.1 and Cauchy-Schwarz inequality,

$$\mathrm{E}\bigg\| \tau\bigg[X_i \otimes \bigg\{ \int_0^1 \beta(s,\cdot)X_i(s)ds \bigg\}\bigg] \bigg\|_K^2$$

$$
\begin{aligned}
&= \sum_{k,\ell} \frac{1}{1+\lambda\rho_{k\ell}} \mathrm{E}\left(\left\langle \tau\left[X_i \otimes \left\{\int_0^1 \beta(s,\cdot)X_i(s)ds\right\}\right], \varphi_{k\ell}\right\rangle_K^2\right) \\
&= \sum_{k,\ell} \frac{1}{1+\lambda\rho_{k\ell}} \mathrm{E}\left(\left\langle \tau\left[X_i \otimes \left\{\int_0^1 \varphi_{k\ell}(s,\cdot)X_i(s)ds\right\}\right], \beta\right\rangle_K^2\right) \\
&\le \|\beta\|_K^2 \times \sum_{k,\ell} \frac{1}{1+\lambda\rho_{k\ell}} \mathrm{E}\left\|\tau\left[X_i \otimes \left\{\int_0^1 \varphi_{k\ell}(s,\cdot)X_i(s)ds\right\}\right]\right\|_K^2 .
\end{aligned}
$$

Combining the above result and (B.5), by Assumption A2 and Lemma B.2, we find

$$
\begin{aligned}
&\mathrm{E}\left\|\tau\left[X_i \otimes \left\{\int_0^1 \beta(s,\cdot)X_i(s)ds\right\}\right]\right\|_K^2 \le c_0 \|\beta\|_K^2 \times \sum_{k,\ell} \frac{1}{1+\lambda\rho_{k\ell}} \times \sum_{k',\ell'} \frac{1}{1+\lambda\rho_{k'\ell'}} \\
&\le c\,\|\beta\|_K^2 \times \left\{\sum_{k,\ell} \frac{1}{1+\lambda(k\ell)^{2D}}\right\}^2 \le c\,\lambda^{-1/D}\,\|\beta\|_K^2 ,
\end{aligned}
$$

which proves (B.4).

□

Lemma B.5. *Under Assumptions A1–A4, for the $x_{k\ell}$'s in Assumption A2, for any $\beta \in \mathcal{H}$ and $\breve{x} \in L^2([0,1])$,*

$$
\left\|\tau\left[\breve{x} \otimes \left\{\int_0^1 \beta(s,\cdot)\breve{x}(s)ds\right\}\right]\right\|_K^2 \le \|\breve{x}\|_{L^2}^2\,\|\beta\|_{L^2}^2 \sum_{k,\ell} \frac{1}{1+\lambda\rho_{k\ell}} \langle \breve{x}, x_{k\ell}\rangle_{L^2}^2\,\|\eta_\ell\|_{L^2}^2 . \tag{B.6}
$$

Proof. By Assumption A2, the representation $\beta(s,t) = \sum_{k,\ell} V(\beta,\varphi_{k\ell})\varphi_{k\ell}(s,t) = \sum_{k,\ell} V(\beta, x_{k\ell}\otimes \eta_\ell)x_{k\ell}(s)\eta_\ell(t)$ holds for any $\beta \in \mathcal{H}$, and the Cauchy-Schwarz inequality and Lemma B.1 yield

$$
\begin{aligned}
&\left\langle \tau\left[\breve{x} \otimes \left\{\int_0^1 \beta(s,\cdot)\breve{x}(s)ds\right\}\right], \varphi_{k\ell}\right\rangle_K^2 \\
&= \left|\int_{[0,1]^3} \varphi_{k\ell}(s_1,t)\,\beta(s_2,t)\,\breve{x}(s_1)\,\breve{x}(s_2)\,ds_1\,ds_2\,dt\right|^2 \\
&= \left|\sum_{k',\ell'} V(\beta, x_{k'\ell'}\otimes\eta_{\ell'}) \int_{[0,1]^3} x_{k\ell}(s_1)\,x_{k'\ell'}(s_2)\,\breve{x}(s_1)\,\breve{x}(s_2)\,\eta_\ell(t)\,\eta_{\ell'}(t)ds_1\,ds_2\,dt\right|^2 \\
&= \left|\int_0^1 \breve{x}(s_1)\,x_{k\ell}(s_1)\,ds_1\right|^2 \times \left|\sum_{k',\ell'} V(\beta, x_{k'\ell'}\otimes\eta_{\ell'}) \int_0^1 \breve{x}(s_2)\,x_{k'\ell'}(s_2)\,ds_2 \int_0^1 \eta_\ell(t)\,\eta_{\ell'}(t)\,dt\right|^2 \\
&= \left|\int_0^1 \breve{x}(s_1)\,x_{k\ell}(s_1)\,ds_1\right|^2 \times \left|\int_{[0,1]^2} \beta(s_2,t)\,\breve{x}(s_2)\,\eta_\ell(t)\,ds_2\,dt\right|^2 \\
&\le \|\beta\|_{L^2}^2\,\|\breve{x}\|_{L^2}^2\,\|\eta_\ell\|_{L^2}^2 \left\{\int_0^1 \breve{x}(s)\,x_{k\ell}(s)\,ds\right\}^2 \\
&= \|\beta\|_{L^2}^2\,\|\breve{x}\|_{L^2}^2\,\langle \breve{x}, x_{k\ell}\rangle_{L^2}^2\,\|\eta_\ell\|_{L^2}^2 .
\end{aligned}
$$

Therefore, in view of (A.63), we deduce from the above result that

$$\left\|\tau\left[\breve{x}\otimes\left\{\int_0^1\beta(s,\cdot)\breve{x}(s)ds\right\}\right]\right\|_K^2=\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\left\langle\tau\left[\breve{x}\otimes\left\{\int_0^1\beta(s,\cdot)\breve{x}(s)ds\right\}\right],\varphi_{k\ell}\right\rangle_K^2$$
$$\le\|\beta\|_{L^2}^2\,\|\breve{x}\|_{L^2}^2\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\,\langle\breve{x},x_{k\ell}\rangle_{L^2}^2\,\|\eta_\ell\|_{L^2}^2\,,$$

which proves (B.6).

□

Recall from (A.17) that, for the $\{x_{k\ell}\}_{k,\ell\ge1}$ and $\{\eta_\ell\}_{\ell\ge1}$ in Assumption A2,

$$w^2(X_i)=\|X_i\|_{L^2}^2\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\,\langle X_i,x_{k\ell}\rangle_{L^2}^2\,\|\eta_\ell\|_{L^2}^2\,,\quad 1\le i\le n\,.$$

We have the following lemma regarding the second moment of $w(X)$.

Lemma B.6. *Under Assumptions A1–A4, we have* $\mathrm{E}\{w^2(X_i)\}\le c\,\lambda^{-1/(2D)}$*, where* $w(X_i)$ *is defined in* (A.17) *and* $c>0$ *is an absolute constant.*

Proof. By Assumption A2, for $k,\ell\ge1$,

$$V(\varphi_{k\ell},\varphi_{k\ell})=\|\eta_\ell\|_{L^2}^2\int C_X(s_1,s_2)x_{k\ell}(s_1)x_{k\ell}(s_2)ds_1ds_2=1\,.$$

Using the Cauchy-Schwarz inequality and Assumption A4,

$$\begin{aligned}
\mathrm{E}\{w^2(X_i)\}&=\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\,\|\eta_\ell\|_{L^2}^2\mathrm{E}\left[\|X_i\|_{L^2}^2\left\{\int_0^1 X_i(s)\,x_{k\ell}(s)\,ds\right\}^2\right]\\
&\le\left(\mathrm{E}\|X_i\|_{L^2}^4\right)^{1/2}\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\,\|\eta_\ell\|_{L^2}^2\left[\mathrm{E}\left\{\int_0^1 X_i(s)\,x_{k\ell}(s)\,ds\right\}^4\right]^{1/2}\\
&\le c\left(\mathrm{E}\|X_i\|_{L^2}^4\right)^{1/2}\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\,\|\eta_\ell\|_{L^2}^2\,\mathrm{E}\left\{\int_0^1 X_i(s)\,x_{k\ell}(s)\,ds\right\}^2\\
&=c\left(\mathrm{E}\|X_i\|_{L^2}^4\right)^{1/2}\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\,\|\eta_\ell\|_{L^2}^2\int C_X(s_1,s_2)\,x_{k\ell}(s_1)\,x_{k\ell}(s_2)\,ds_1\,ds_2\\
&=c\left(\mathrm{E}\|X_i\|_{L^2}^4\right)^{1/2}\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\,V(\varphi_{k\ell},\varphi_{k\ell})\\
&=c\left(\mathrm{E}\|X_i\|_{L^2}^4\right)^{1/2}\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\,.
\end{aligned}$$

Since $\mathrm{E}\|X_i\|_{L^2}^4$ is finite by Assumption A4, by Assumption A2 and Lemma B.2, we deduce from the above equation that

$$\mathrm{E}\{w^2(X_i)\} \le c \sum_{k,\ell} \frac{1}{1+\lambda(k\ell)^{2D}} \le c\,\lambda^{-1/(2D)}\,.$$

□

Lemma B.7. *Under Assumptions A1–A4,*

$$\mathrm{E}\|\tau(X\otimes\varepsilon)\|_K^2 = \sum_{k,\ell} \frac{1}{1+\lambda\rho_{k\ell}}\,,$$
$$\mathrm{E}\|\tau(X\otimes\varepsilon)\|_K^4 \le c\,\lambda^{-1/D}\,,$$

where $c>0$ is an absolute constant.

Proof. In view of (2.13) and (A.63), we have

$$\begin{aligned} \tau(X\otimes\varepsilon) &= \sum_{k,\ell} \frac{\langle \tau(X\otimes\varepsilon),\varphi_{k\ell}\rangle_K}{1+\lambda\rho_{k\ell}}\,\varphi_{k\ell} = \sum_{k,\ell} \frac{\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}}{1+\lambda\rho_{k\ell}}\,\varphi_{k\ell}\,, \\ \|\tau(X\otimes\varepsilon)\|_K^2 &= \sum_{k,\ell} \frac{\langle \tau(X\otimes\varepsilon),\varphi_{k\ell}\rangle_K^2}{1+\lambda\rho_{k\ell}} = \sum_{k,\ell} \frac{\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}^2}{1+\lambda\rho_{k\ell}}\,. \end{aligned} \tag{B.7}$$

Recall from Assumption A3 that $C_\varepsilon(s_1,s_2) = \mathrm{E}\{\varepsilon(s_1)\varepsilon(s_2)|X\} = \delta(s_1,s_2)$ (we have assumed $\sigma_\varepsilon^2 = 1$ without loss of generality; see the beginning of Section A.5). Observing the definition of V in (2.9) and τ in (2.17), we find, for $k,k',\ell,\ell' \ge 1$,

$$\begin{aligned} &\mathrm{E}\Big\{\big\langle \tau(X\otimes\varepsilon),\varphi_{k\ell}\big\rangle_K \big\langle \tau(X\otimes\varepsilon),\varphi_{k'\ell'}\big\rangle_K\Big\} \\ &= \mathrm{E}\Big\{\big\langle X\otimes\varepsilon,\varphi_{k\ell}\big\rangle_{L^2} \big\langle X\otimes\varepsilon,\varphi_{k'\ell'}\big\rangle_{L^2}\Big\} \\ &= \mathrm{E}\left[\left\{\int_{[0,1]^2} X(s)\varepsilon(t)\varphi_{k\ell}(s,t)\,ds\,dt\right\} \times \left\{\int_{[0,1]^2} X(s')\varepsilon(t')\varphi_{k'\ell'}(s',t')\,ds'\,dt'\right\}\right] \\ &= \mathrm{E}\left[\int_{[0,1]^4} \mathrm{E}\{\varepsilon(t)\varepsilon(t')|X\}\,X(s)X(s')\,\varphi_{k\ell}(s,t)\,\varphi_{k'\ell'}(s',t')\,ds\,ds'\,dt\,dt'\right] \\ &= \int_{[0,1]^4} C_\varepsilon(t,t')\,C_X(s,s')\,\varphi_{k\ell}(s,t)\,\varphi_{k'\ell'}(s',t')\,ds\,ds'\,dt\,dt' \\ &= \int_{[0,1]^3} C_X(s,s')\,\varphi_{k\ell}(s,t)\,\varphi_{k'\ell'}(s',t)\,ds\,ds'\,dt \\ &= V(\varphi_{k\ell},\varphi_{k'\ell'}) = \delta_{kk'}\,\delta_{\ell\ell'}\,, \end{aligned} \tag{B.8}$$

where we used Assumption A2 in the last step. From the above equation we have obtained

$$\mathrm{E}\Big(\big\langle X\otimes\varepsilon,\varphi_{k\ell}\big\rangle_{L^2}^2\Big) = 1\,.$$

In view of (B.7),

$$\mathrm{E}\|\tau(X\otimes\varepsilon)\|_K^2 = \sum_{k,\ell} \frac{\mathrm{E}\big(\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}^2\big)}{1+\lambda\rho_{k\ell}} = \sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\,.$$

To show the order of $\mathrm{E}\|\tau(X\otimes\varepsilon)\|_K^4$ we use Assumptions A2 and A4 to obtain

$$\begin{aligned}
&\mathrm{E}\big(\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}^4\big) = \mathrm{E}\Big\{\int_0^1\int_0^1 X(s)\varepsilon(t)\varphi_{k\ell}(s,t)\,ds\,dt\Big\}^4\\
&= \mathrm{E}\Bigg[\Big\{\int_0^1 X(s)x_{k\ell}(s)ds\Big\}^4\int_{[0,1]^4}\mathrm{E}\{\varepsilon(t_1)\varepsilon(t_2)\varepsilon(t_3)\varepsilon(t_4)|X\}\eta_\ell(t_1)\eta_\ell(t_2)\eta_\ell(t_3)\eta_\ell(t_4)dt_1dt_2dt_3dt_4\Bigg]\\
&\le \Bigg(\int_{[0,1]^4}\Big[\mathrm{E}\{\varepsilon(t_1)\varepsilon(t_2)\varepsilon(t_3)\varepsilon(t_4)|X\}\Big]^2dt_1dt_2dt_3dt_4\Bigg)^{1/2}\times\|\eta_\ell\|_{L^2}^4\times\Bigg[\mathrm{E}\Big\{\int_0^1 X(s)x_{k\ell}(s)ds\Big\}^2\Bigg]^2\\
&= c\Bigg[\mathrm{E}\Big\{\int_0^1\int_0^1 C_X(s_1,s_2)x_{k\ell}(s_1)x_{k\ell}(s_2)ds_1ds_2\Big\}^2\Bigg]^2 = c\,\{V(x_{k\ell},x_{k\ell})\}^2 = c\,. \qquad \text{(B.9)}
\end{aligned}$$

Therefore, we deduce that

$$\begin{aligned}
\mathrm{E}\|\tau(X\otimes\varepsilon)\|_K^4 &= \sum_{k_1,\ell_1,k_2,\ell_2}\frac{1}{(1+\lambda\rho_{k_1\ell_1})(1+\lambda\rho_{k_2\ell_2})}\mathrm{E}\Big(\langle X\otimes\varepsilon,\varphi_{k_1\ell_1}\rangle_{L^2}^2\langle X\otimes\varepsilon,\varphi_{k_2\ell_2}\rangle_{L^2}^2\Big)\\
&\le \Bigg[\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\Big\{\mathrm{E}\big(\langle X\otimes\varepsilon,\varphi_{k\ell}\rangle_{L^2}^4\big)\Big\}^{1/2}\Bigg]^2 \le c\Bigg\{\sum_{k,\ell}\frac{1}{1+\lambda\rho_{k\ell}}\Bigg\}^2 \le c\lambda^{-1/D}\,.
\end{aligned}$$

□

The following lemma is a modified version of Lemma A.1 in Kley et al. (2016), which we use to prove Theorem 3.3.

Lemma B.8. *For any non-decreasing, convex function $\Psi:\mathbb{R}^+\to\mathbb{R}^+$ with $\Psi(0)=0$ and for any real-valued random variable Z, let $\|Z\|_\Psi=\inf\{c>0:\mathrm{E}\{\Psi(|Z|/c)\}\le 1\}$ denote the Orlicz norm. Let $\{H(s,t):(s,t)\in[0,1]^2\}$ be a separable stochastic process with $\|H(s_1,t_1)-H(s_2,t_2)\|_\Psi\le c\,d\{(s_1,t_1),(s_2,t_2)\}$ for any $(s_1,t_1),(s_2,t_2)\in[0,1]^2$ with $d\{(s_1,t_1),(s_2,t_2)\}\ge\overline{\eta}/2\ge 0$ and for some constant $c>0$. Let $\mathfrak{D}(w,d)$ denote the packing number of the metric space $([0,1]^2,d)$. Then, for any $\delta>0$, $\eta>\overline{\eta}$, there exists a random variable S and a constant $K>0$ such that*

$$\sup_{d\{(s_1,t_1),(s_2,t_2)\}\le\delta}|H(s_1,t_1)-H(s_2,t_2)|\le S+2\sup_{\substack{d\{(s_1,t_1),(s_2,t_2)\}\le\overline{\eta}\\(s_1,t_1)\in[0,1]^2}}|H(s_1,t_1)-H(s_2,t_2)|$$

and

$$\|S\|_\Psi\le K\Bigg[\int_{\overline{\eta}/2}^{\eta}\Psi^{(-1)}\{\mathfrak{D}(\varepsilon,d)\}d\varepsilon+(\delta+2\overline{\eta})\,\Psi^{(-1)}\{\mathfrak{D}^2(\eta,d)\}\Bigg],$$

where $\Psi^{(-1)}$ is the inverse function of Ψ, and the set $[0,1]^2$ contains at most $\mathfrak{D}(\overline{\eta},d)$ points.

B.2 An example for Assumption A2

In this section we provide a concrete example that satisfies Assumption A2. We use the cosine basis of $L^2([0,1])$ defined by

$$\eta_1(t) \equiv 1 \,, \quad \eta_\ell(t) = \sqrt{2}\cos\{(\ell-1)\pi t\} \quad \ell = 2, 3, \dots \tag{B.10}$$

Now, the derivatives of these functions are orthogonal with respect to the L^2-inner product $\langle \cdot, \cdot \rangle_{L^2}$, that is, for any integer $\theta \geq 0$ and $\ell, \ell' \geq 2$, $\langle \eta_\ell^{(\theta)}, \eta_{\ell'}^{(\theta)} \rangle_{L^2} = \delta_{\ell\ell'} \|\eta_\ell^{(\theta)}\|_{L^2}^2 = \delta_{\ell\ell'}(\ell-1)^{2\theta}\pi^{2\theta}$; for any $\theta \geq 0$ and $\ell \geq 2$, $\langle \eta_1^{(\theta)}, \eta_\ell^{(\theta)} \rangle_{L^2} = 0$. Given $\{\eta_\ell\}_{\ell \geq 1}$, the functions $x_{k\ell}$ are defined as the solution of a series of integral-differential equations whose parameters depend on $\|\eta_\ell^{(\theta)}\|_{L^2}^2$, such that (2.12) is satisfied. In particular, we have the following proposition, which provides an example of an eigen-system that satisfies Assumption A2 and is proved in Section B.3 using the theory of integro-differential equations.

Proposition B.1. *For each $\ell \geq 1$, let $\{(\rho_{k\ell}, \widetilde{x}_{k\ell})\}_{k \geq 1}$ denote the eigenvalue-eigenfunction pairs of the following integro-differential equations with boundary conditions:*

$$\begin{cases} \rho_\ell \displaystyle\int_0^1 C_X(s, s')\widetilde{x}(s')ds' = (-1)^m \widetilde{x}^{(2m)}(s) + \sum_{\theta=0}^{m-1} \binom{m}{\theta} (-1)^\theta \{(\ell-1)\pi\}^{2m-2\theta} \widetilde{x}^{(2\theta)}(s) \,, \\ \widetilde{x}^{(\theta)}(0) = \widetilde{x}^{(\theta)}(1) = 0 \,, \qquad \text{for } m \leq \theta \leq 2m-1 \,. \end{cases} \tag{B.11}$$

For the functions η_ℓ defined in (B.10), let $x_{k\ell} = \langle \mathcal{C}_X(\widetilde{x}_{k\ell}), \widetilde{x}_{k\ell} \rangle_{L^2}^{-1/2} \widetilde{x}_{k\ell}$ and $\varphi_{k\ell} = x_{k\ell} \otimes \eta_\ell$, where the operator $\mathcal{C}_X$ is defined by (2.11). Suppose Condition B(r) below is satisfied for some constant $r \geq 0$. Then, the pairs $(\rho_{k\ell}, \varphi_{k\ell})_{k,\ell}$ satisfy Assumption A2 with $D = m+r+1$ and $a = r+1$.

For a constant $r \geq 0$, we now state as follows Condition B(r) for Proposition B.1, which was proposed in Shang and Cheng (2015). Let $\Omega_+ = \{(s,t) \in [0,1] : s > t\}$, $\Omega_- = \{(s,t) \in [0,1] : s < t\}$ and let $\mathrm{cl}(A)$ denote the closure of $A \subset [0,1]^2$. Recall that C_X defined in (2.10) is the covariance function of X.

Condition B(r). Suppose that there exists a constant $r \geq 0$ such that one of the following two assumptions is satisfied: (i) $r = 0$; (ii) $r \geq 1$, and for any $j = 0, 1, \dots, r-1$, $C_X^{(j,0)}(0,t) = 0$, for any $0 \leq t \leq 1$. Assume C_X satisfies the following *pseudo SY conditions* of order r:

(1) $L(s_1, s_2) := C_X^{(r,r)}(s_1, s_2)$ is continuous on $[0,1]^2$. All the partial derivatives of $L(s_1, s_2)$ up to order $2m+2r+2$ are continuous on $\Omega_+ \cup \Omega_-$, and continuously extendable to $\mathrm{cl}(\Omega_+)$ and $\mathrm{cl}(\Omega_+)$.

(2) $a(s) := L_-^{(1,0)}(s,s) - L_+^{(1,0)}(s,s)$ has a positive lower bound for any $s \in [0,1]$, where $L_-^{(1,0)}$ and $L_+^{(1,0)}$ are two different extensions of $L^{(1,0)}$ to $[0,1]^2$ that are continuous on $\mathrm{cl}(\Omega_-)$ and $\mathrm{cl}(\Omega_+)$, respectively.

(3) $L_+^{(2,0)}(s_1,s_2)$ is bounded over $[0,1]^2$, where $L_+^{(2,0)}$ is the extension of $L^{(2,0)}$ to $[0,1]^2$ that is continuous on $\mathrm{cl}(\Omega_+)$.

In addition, assume that the integro-differential equations with the boundary conditions in Proposition B.1 are *regular* in the the sense of Birkhoff (1908b) (see Definition B.1 below).

Definition B.1 (Regular boundary conditions of even order; Birkhoff, 1908b)**.** Consider the linear differential equation of order $2k$ in φ: $\varphi^{(2k)}(x) + \sum_{\ell=0}^{2k-2} p_\ell(x)\varphi^{(\ell)}(x) + \gamma\varphi(x) = 0$ on an interval $[\mathfrak{a},\mathfrak{b}]$ with $2k$ linear homogeneous boundary conditions $W_i(\varphi) = 0$ in $\varphi(\mathfrak{a}),\varphi'(\mathfrak{a}),\varphi^{(k-1)}(\mathfrak{a}),\ldots,\varphi(\mathfrak{b}),\varphi'(\mathfrak{b}),\varphi^{(k-1)}(\mathfrak{b})$, for $1 \le i \le 2k$. Applying the linear transformation on the W_i's to obtain the normalized boundary conditions in the form $W_i(\varphi) = W_{i\mathfrak{a}}(\varphi) + W_{i\mathfrak{b}}(\varphi) = 0$, where $W_{i\mathfrak{a}}(\varphi) = a_i\varphi^{(j_i)}(\mathfrak{a}) + \sum_{\ell=0}^{j_i-1} a_{i\ell}\varphi^{(\ell)}(\mathfrak{a})$, $W_{i\mathfrak{b}}(\varphi) = b_i\varphi^{(j_i)}(\mathfrak{b}) + \sum_{\ell=0}^{j_i-1} b_{i\ell}\varphi^{(\ell)}(\mathfrak{b})$, and where $j_1 \ge \ldots \ge j_n$ are such that no successive three of them are equal. Define $\zeta_0,\zeta_1,\zeta_2 \in \mathbb{C}$ through the identity

$$\zeta_0 + \zeta_1 s + \frac{\zeta_2}{s} \equiv \begin{vmatrix} a_1 w_1^{j_1} & \cdots & a_1 w_{k-1}^{j_1} & (a_1 + b_1 s) w_k^{j_1} & (a_1 + \frac{b_1}{s}) w_{k+1}^{j_1} & b_1 w_{k+2}^{j_1} & \cdots & b_1 w_n^{j_1} \\ a_2 w_1^{j_2} & \cdots & a_2 w_{k-1}^{j_2} & (a_2 + b_2 s) w_k^{j_2} & (a_2 + \frac{b_2}{s}) w_{k+1}^{j_2} & b_2 w_{k+2}^{j_2} & \cdots & b_2 w_n^{j_2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_n w_1^{j_n} & \cdots & a_n w_{k-1}^{j_n} & (a_n + b_n s) w_k^{j_n} & (a_n + \frac{b_n}{s}) w_{k+1}^{j_2} & b_n w_{k+2}^{j_n} & \cdots & b_n w_n^{j_n} \end{vmatrix},$$

where the w_i's are the $2k$-th root of unity ordered according to $\mathrm{Re}(\rho w_1) < \mathrm{Re}(\rho w_2) < \cdots < \mathrm{Re}(\rho w_n)$ and $\rho = \gamma^{1/(2k)}$. Then, the boundary conditions $W_1(\varphi),\ldots,W_{2k}(\varphi)$ are *regular* if $\zeta_1 \ne 0$ and $\zeta_2 \ne 0$.

B.3 Proof of Proposition B.1

We follow the proof of Proposition 2.2 in Shang and Cheng (2015) and we assume that $a(\cdot)$ in Condition B(r) satisfies $a \equiv 1$ without loss of generality. For the integro-differential equation (B.11), taking $\ell = 1$ yields

$$\begin{cases} \rho_1 \displaystyle\int_0^1 C_X(s_1,s_2)\,x(s_2)\,ds_2 = (-1)^m\,\widetilde{x}^{(2\theta)}(s_1)\,, \\ \widetilde{x}^{(\theta)}(0) = \widetilde{x}^{(\theta)}(1) = 0\,, \qquad \text{for } m \le \theta \le 2m-1\,. \end{cases}$$

This case was proved by Shang and Cheng (2015). We therefore focus on the case where $\ell \ge 2$ in equation (B.11), which is equivalent to, for $\ell \ge 2$,

$$\begin{cases} \rho_\ell \displaystyle\int_0^1 C_X(s_1,s_2)\,x(s_2)\,ds_2 = \sum_{\theta=0}^{m} \binom{m}{\theta}(-1)^\theta\{(\ell-1)\pi\}^{2m-2\theta}\,\widetilde{x}^{(2\theta)}(s_1)\,, \\ \widetilde{x}^{(\theta)}(0) = \widetilde{x}^{(\theta)}(1) = 0\,, \qquad \text{for } m \le \theta \le 2m-1\,. \end{cases} \tag{B.12}$$

By virtue of simple presentation, without loss of generality, we change the subscript of ρ_ℓ to $\ell-1$ in (B.12) to write, for $\ell \geq 1$,

$$\begin{cases} \rho_\ell \int_0^1 C_X(s_1,s_2)\, x(s_2)\, ds_2 = \sum_{\theta=0}^{m} \binom{m}{\theta} (-1)^\theta (\ell\pi)^{2m-2\theta}\, \widetilde{x}^{(2\theta)}(s_1)\,, \\ \widetilde{x}^{(\theta)}(0) = \widetilde{x}^{(\theta)}(1) = 0\,, \qquad \text{for } m \leq \theta \leq 2m-1\,. \end{cases} \tag{B.13}$$

In the sequel we show the results in Proposition B.1 based on (B.13). Let $N(s_1,s_2) = L_+^{(2,0)}(s_1,s_2)$ for L_+ in Section B.2 and let $M(s_1,s_2)$ denote its reciprocal kernel such that the following reciprocal property (Tamarkin, 1927) is satisfied

$$M(s_1,s_2) + N(s_1,s_2) = \int_0^1 M(s_1,\xi)N(\xi,s_2)d\xi = \int_0^1 N(s_1,\xi)M(\xi,s_2)d\xi\,. \tag{B.14}$$

For $0 \leq \theta \leq m$, let $c_\theta = \binom{m}{\theta}(-1)^\theta \pi^{2m-2\theta}$. For r in Condition B(r), we have (B.13) is equivalent to the following equation

$$\begin{cases} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \psi^{(2\theta+r)}(s_1) = \rho_\ell \int_0^1 C_X(s_1,s_2)\, \psi^{(r)}(s_2)\, ds_2\,, \\ \psi^{(\upsilon)}(0) = \psi^{(\upsilon)}(1) = 0\,, \ \text{for } m+r \leq \upsilon \leq 2m+r-1\,, \\ \psi^{(\upsilon)}(1) = 0\,, \ \text{for } 0 \leq \upsilon \leq r-1\,. \end{cases} \tag{B.15}$$

That is, $\widetilde{x} = \psi^{(r)}$ being the solution to (B.13) is equivalent to ψ being the solution to (B.15). From the first equation in (B.15), by integration by parts we find

$$\sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \psi^{(2\theta+r)}(s_1) = (-1)^r \rho_\ell \int_0^1 C_X^{(0,r)}(s_1,s_2)\, \psi(s_2)\, ds_2\,,$$

due to the assumption that $C_X^{(0,\upsilon)}(s,0) = 1$ for $0 \leq \upsilon \leq r-1$, and that $\psi^{(\upsilon)}(1) = 0$, for $0 \leq \upsilon \leq r-1$ in (B.15). Taking partial derivatives of the above equation yields that, for $0 \leq \upsilon \leq r$,

$$\begin{aligned} &\sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \psi^{(2\theta+r+\upsilon)}(s_1) = (-1)^r \rho_\ell \int_0^1 C_X^{(\upsilon,r)}(s_1,s_2)\, \psi(s_2)\, ds_2\,, \\ &\sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \psi^{(2\theta+2r+1)}(s_1) = (-1)^r \rho_\ell \int_0^1 L^{(1,0)}(s_1,s_2)\, \psi(s_2)\, ds_2\,, \\ &\sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \psi^{(2\theta+2r+2)}(s_1) = (-1)^{r+1} \rho_\ell\, \psi(s_1) + (-1)^r \rho_\ell \int_0^1 L_+^{(2,0)}(s_1,s_2)\, \psi(s_2)\, ds_2\,, \end{aligned}$$

due to the fact that $a(s) = 1$ and $\int_0^1 L^{(1,0)}(s_1,s_2)\psi(s_2)\, ds_2 = \int_{s_1}^1 L_-^{(1,0)}(s_1,s_2)\psi(s_2)ds_2 + \int_0^{s_1} L_+^{(1,0)}(s_1,s_2)\psi(s_2)ds_2$. Hence we find that (B.15) is equivalent to the following boundary value problem

$$\begin{cases} \sum_{\theta=0}^{m} c_\theta \, \ell^{2m-2\theta} \psi^{(2\theta+2r+2)}(s_1) + (-1)^r \rho_\ell \Big\{ \psi(s_1) - \int_0^1 L_+^{(2,0)}(s_1,s_2)\psi(s_2)ds_2 \Big\} = 0\,, \\ \psi^{(v)}(0) = \psi^{(v)}(1) = 0\,, \qquad \text{for } m+r \le v \le 2m+r-1\,, \\ \psi^{(v)}(1) = 0\,, \qquad \text{for } 0 \le v \le r-1\,, \\ \sum_{\theta=0}^{m} c_\theta \, \ell^{2m-2\theta} \psi^{(2\theta+r+v)}(0) = (-1)^r \rho_\ell \int_0^1 C_X^{(v,r)}(0,s)\psi(s)ds\,, \quad 0 \le v \le r+1\,. \end{cases} \tag{B.16}$$

Recall from the paragraph above (B.14) that $N(s_1,s_2) = L_+^{(2,0)}(s_1,s_2)$. For the first equation in (B.16), using the reciprocal property in (B.14),

$$\begin{aligned} &(-1)^{r+1} \sum_{\theta=0}^{m} c_\theta \, \ell^{2m-2\theta} \int_0^1 M(\xi,s)\, \psi^{(2\theta+2r+2)}(s)ds \\ &= \rho_\ell \Big\{ \int_0^1 M(\xi,s)\psi(s)ds - \int_0^1\int_0^1 M(\xi,s_1)N(s_1,s_2)\psi(s_2)ds_1ds_2 \Big\} \\ &= \rho_\ell \Big[\int_0^1 M(\xi,s)\psi(s)ds - \int_0^1 \{M(\xi,s)+N(\xi,s)\}\psi(s)ds \Big] \\ &= -\rho_\ell \int_0^1 N(\xi,s)\psi(s)ds = -\rho_\ell \int_0^1 L_+^{(2,0)}(\xi,s)\psi(s)\,ds\,. \end{aligned}$$

Combining the above equation with the first equation of (B.16) yields

$$\rho_\ell \psi(s_1) = (-1)^{r+1} \sum_{\theta=0}^{m} c_\theta \, \ell^{2m-2\theta} \left\{ \psi^{(2\theta+2r+2)}(s_1) - \int_0^1 M(s_1,s_2)\, \psi^{(2\theta+2r+2)}(s_2)ds_2 \right\}. \tag{B.17}$$

Combining (B.17) with the first equation of (B.16) and using the reciprocal property in (B.14), we find

$$\begin{aligned} &(-1)^{r+1} \sum_{\theta=0}^{m} c_\theta \, \ell^{2m-2\theta} \psi^{(2\theta+2r+2)}(s_1) \\ &= \rho_\ell \psi(s_1) - \int_0^1 N(s_1,s_2) \\ &\quad \times \left[(-1)^{r+1} \sum_{\theta=0}^{m} c_\theta \, \ell^{2m-2\theta} \left\{ \psi^{(2\theta+2r+2)}(s_2) - \int_0^1 M(s_2,\xi)\, \psi^{(2\theta+2r+2)}(\xi)d\xi \right\} \right] ds_2 \\ &= \rho_\ell \psi(s_1) - (-1)^{r+1} \sum_{\theta_1=0}^{m} c_\theta \, \ell^{2m-2\theta} \int_0^1 N(s_1,s_2)\psi^{(2\theta+2r+2)}(s_2)ds_2 \\ &\quad + (-1)^{r+1} \sum_{\theta=0}^{m} c_\theta \, \ell^{2m-2\theta} \int_0^1 \int_0^1 N(s_1,s_2)M(s_2,\xi)\psi^{(2\theta+2r+2)}(\xi)d\xi ds_2 \end{aligned}$$

$$
\begin{aligned}
&= \rho_\ell \psi(s_1) - (-1)^{r+1} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \int_0^1 N(s_1,s_2)\psi^{(2\theta+2r+2)}(s_2)ds_2 \\
&\quad + (-1)^{r+1} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \int_0^1 N(s_1,\xi)\psi^{(2\theta+2r+2)}(\xi)d\xi \\
&\quad + (-1)^{r+1} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \int_0^1 M(s_1,\xi)\psi^{(2\theta+2r+2)}(\xi)d\xi \\
&= \rho_\ell \psi(s_1) + (-1)^{r+1} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \int_0^1 M(s_1,s_2)\psi^{(2\theta+2r+2)}(s_2)ds_2 \,.
\end{aligned}
$$

Using integration by parts, we deduce from the above equation that

$$
\begin{aligned}
&(-1)^{r+1} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \psi^{(2\theta+2r+2)}(s_1) \\
&= \rho_\ell \psi(s_1) + (-1)^{r+1} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \int_0^1 M^{(0,2\theta+2r+2)}(s_1,s_2)\,\psi(s_2)ds_2 \\
&+ \sum_{\theta=0}^{m} \sum_{j=1}^{2\theta+2r+2} (-1)^{r+j} c_\theta\, \ell^{2m-2\theta} \Big\{ M^{(0,j-1)}(s_1,1)\psi^{(2\theta+2r+2-j)}(1) - M^{(0,j-1)}(s_1,0)\psi^{(2\theta+2r+2-j)}(0) \Big\} \\
&= \rho_\ell \psi(s_1) + L_\ell \psi(s_1)\,,
\end{aligned} \tag{B.18}
$$

if we define

$$
\begin{aligned}
L_\ell \psi(s_1) = (-1)^{r+1} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \int_0^1 M^{(0,2\theta+2r+2)}(s_1,s_2)\,\psi(s_2)ds_2 + \sum_{\theta=0}^{m} \sum_{j=1}^{2\theta+2r+2} (-1)^{r+j} c_\theta\, \ell^{2m-2\theta} \\
\times \Big\{ M^{(0,j-1)}(s_1,1)\,\psi^{(2\theta+2r+2-j)}(1) - M^{(0,j-1)}(s_1,0)\,\psi^{(2\theta+2r+2-j)}(0) \Big\}\,.
\end{aligned} \tag{B.19}
$$

For the last boundary value condition in (B.16), by (B.17) and integration by parts, we find that, for $0 \le \upsilon \le r+1$,

$$
\begin{aligned}
&(-1)^{r} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \psi^{(2\theta+r+\upsilon)}(0) \\
&= \int_0^1 C_X^{(r,\upsilon)}(s_1,0) \Bigg[(-1)^{r+1} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \Big\{ \psi^{(2\theta+2r+2)}(s_1) - \int_0^1 M(s_1,s_2)\,\psi^{(2\theta+2r+2)}(s_2)ds_2 \Big\} \Bigg] ds_1 \\
&= (-1)^{r+1} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \Big\{ \int_0^1 C_X^{(r,\upsilon)}(s_1,0)\,\psi^{(2\theta+2r+2)}(s_1)ds_1 \\
&\qquad - \int_0^1 \int_0^1 C_X^{(r,\upsilon)}(s_1,0) M(s_1,s_2)\,\psi^{(2\theta+2r+2)}(s_2)ds_2 ds_1 \Big\} \\
&= (-1)^{r+1} \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \Bigg[\int_0^1 C_X^{(2\theta+3r+2,\upsilon)}(s_1,0)\,\psi(s_1)ds_1
\end{aligned}
$$

$$
\begin{aligned}
&- \int_0^1 \int_0^1 C_X^{(r,\upsilon)}(s_1,0) M^{(0,2\theta+2r+2)}(s_1,s_2)\, \psi(s_2) ds_2 ds_1 \\
&+ \sum_{j=1}^{2\theta+2r+2} (-1)^{j-1} \Big\{ C_X^{(r+j-1,\upsilon)}(1,0)\, \psi^{(2\theta+2r+2-j)}(1) - C_X^{(r+j-1,\upsilon)}(0,0)\, \psi^{(2\theta+2r+2-j)}(0) \\
&- \psi^{(2\theta+2r+2-j)}(1) \int_0^1 C_X^{(r,\upsilon)}(s_1,0) M^{(0,j-1)}(s_1,1) ds_1 \\
&+ \psi^{(2\theta+2r+2-j)}(0) \int_0^1 C_X^{(r,\upsilon)}(s_1,0) M^{(0,j-1)}(s_1,0) ds_1 \Big\} \Big] .
\end{aligned}
$$

For $0 \le \upsilon \le r+1$ and $1 \le j \le 2m+2r+2$, if we denote

$$
\begin{aligned}
A_{\ell,\upsilon}(s) &= \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \left\{ C_X^{(2\theta+3r+2,\upsilon)}(s,0) - \int_0^1 C_X^{(r,\upsilon)}(\xi,0) M^{(0,2\theta+2r+2)}(\xi,s) d\xi \right\} ; \\
a_{j,\upsilon} &= (-1)^j \left\{ C_X^{(r+j-1,\upsilon)}(0,0) - \int_0^1 C_X^{(r,\upsilon)}(s,0) M^{(0,j-1)}(s,0) ds \right\} ; \\
b_{j,\upsilon} &= (-1)^{j+1} \left\{ C_X^{(r+j-1,\upsilon)}(1,0) - \int_0^1 C_X^{(r,\upsilon)}(s,0) M^{(0,j-1)}(s,1) ds \right\} ,
\end{aligned}
\tag{B.20}
$$

we have that for $0 \le \upsilon \le r+1$,

$$
\begin{aligned}
\sum_{\theta=0}^{m} \sum_{j=1}^{2\theta+2r+2} c_\theta\, \ell^{2m-2\theta} \Big\{ a_{j,\upsilon}\, \psi^{(2\theta+2r+2-j)}(0) + b_{j,\upsilon}\, \psi^{(2\theta+2r+2-j)}(1) \Big\} \\
+ \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta}\, \psi^{(2\theta+r+\upsilon)}(0) + \int A_{\ell,\upsilon}(s) \psi(s) ds = 0 .
\end{aligned}
$$

By assumption, we have $C_X^{(j,0)}(0,s) \equiv 0$ for $0 \le j \le r-1$, so that $C_X^{(q,\upsilon)}(s,0) = C_X^{(\upsilon,q)}(0,s) \equiv 0$ for $0 \le \upsilon \le r-1$ and $1 \le q \le 2m+3r+2$. Hence $a_{j,\upsilon} = b_{j,\upsilon} = 0$ and $A_{\ell,\upsilon}(s) \equiv 0$, for $0 \le j \le r-1$ and $1 \le \upsilon \le 2\theta+2r+2$. Therefore, in view of (B.18), from the above calculations, if we let $D = m+r+1$, we find that (B.13), (B.15) and (B.16) are equivalent, and are equivalent to the following boundary value problem

$$
\begin{cases}
\psi^{(2D)}(s) + (-1)^m \sum_{\theta=0}^{m-1} c_\theta\, \ell^{2m-2\theta} \psi^{(2D-2\theta)}(s) + (-1)^{m+r} \rho_\ell \psi(s) + L_\ell \psi(s) = 0 , \\
\psi^{(\upsilon)}(0) = 0 , \qquad \text{for } m+r \le \upsilon \le 2m+r-1 , \\
\psi^{(\upsilon)}(1) = 0 , \qquad \text{for } 0 \le \upsilon \le r-1 \text{ and } m+r \le \upsilon \le 2m+r-1 , \\
\sum_{\theta=0}^{m} \sum_{j=1}^{2\theta+2r+2} c_\theta\, \ell^{2m-2\theta} \Big\{ a_{j,\upsilon}\, \psi^{(2\theta+2r+2-j)}(0) + b_{j,\upsilon}\, \psi^{(2\theta+2r+2-j)}(1) \Big\} \\
\qquad + \sum_{\theta=0}^{m} c_\theta\, \ell^{2m-2\theta} \psi^{(2\theta+r+\upsilon)}(0) + \int A_{\ell,\upsilon}(s) \psi(s) ds = 0 , \quad \text{for } \upsilon = r, r+1 .
\end{cases}
\tag{B.21}
$$

The *auxiliary problem* (cf. Tamarkin and Langer, 1928, p. 459) of (B.21) is

$$\begin{cases} \psi^{(2D)}(s) + (-1)^m \sum_{\theta=0}^{m-1} c_\theta \, \ell^{2m-2\theta} \psi^{(2D-2\theta)}(s) + (-1)^{m+r} \breve{\rho}_\ell \psi(s) = 0\,, \\ \psi^{(\upsilon)}(0) = 0\,, \qquad \text{for } m+r \le \upsilon \le 2m+r-1\,, \\ \psi^{(\upsilon)}(1) = 0\,, \qquad \text{for } 0 \le \upsilon \le r-1 \text{ and } m+r \le \upsilon \le 2m+r-1\,, \\ \sum_{\theta=0}^{m} \sum_{j=1}^{2\theta+2r+2} c_\theta \, \ell^{2m-2\theta} \Big\{ a_{j,\upsilon} \, \psi^{(2\theta+2r+2-j)}(0) + b_{j,\upsilon} \, \psi^{(2\theta+2r+2-j)}(1) \Big\} \\ \qquad + \sum_{\theta=0}^{m} c_\theta \, \ell^{2m-2\theta} \psi^{(2\theta+r+\upsilon)}(0) = 0\,, \qquad \text{for } \upsilon = r, r+1\,. \end{cases} \tag{B.22}$$

The above boundary value problem in (B.22) is a linear differential equation of the order of $2m+2r+2 = 2D$ in ψ, with $2D$ linear homogeneous conditions on the $\psi^{(j)}(0)$ and $\psi^{(j)}(1)$, for $0 \le j \le 2m+2r+1$; furthermore, the coefficients of the odd-order derivatives of ψ in the first equation of (B.22) are all zero. The *characteristic value* (see Birkhoff, 1908b) of (B.22) is $(-1)^{m+r}\breve{\rho}_\ell$. Letting $\breve{\rho}_\ell = (-1)^{D+1}\ell^{2D}\breve{\varrho}^{2D}$, we have the first equation in (B.22) is

$$\psi^{(2D)}(s) + \sum_{\theta=1}^{m} c_{m-\theta} \, \ell^{2\theta} \psi^{(2D-2\theta)}(s) + \ell^{2D} \breve{\varrho}^{2D} \, \psi(s) = 0\,, \tag{B.23}$$

Let $\widetilde{\psi}(s) = \psi(s/\ell)$, so that $\widetilde{\psi}^{(\upsilon)}(s) = \ell^{-\upsilon}\psi^{(\upsilon)}(s/\ell)$, for $0 \le \upsilon \le 2D$. The key of this proof is that $\psi(s)$ being the solution to (B.23) is equivalent to $\widetilde{\psi}(s)$ being the solution to the following ordinary differential equation corresponding to the characteristic value ϱ^{2D} that is independent of ℓ:

$$\widetilde{\psi}^{(2D)}(s) + \sum_{\theta=1}^{m} c_{m-\theta} \, \widetilde{\psi}^{(2D-2\theta)}(s) + \breve{\varrho}^{2D} \, \widetilde{\psi}(s) = 0\,. \tag{B.24}$$

In view of (B.22), together with the boundary conditions, $\widetilde{\psi}$ is the solution of the following boundary value problem.

$$\begin{cases} \widetilde{\psi}^{(2D)}(s) + \sum_{\theta=1}^{m} c_{m-\theta} \, \widetilde{\psi}^{(2D-2\theta)}(s) + \breve{\varrho}^{2D} \, \widetilde{\psi}(s) = 0\,, \\ \widetilde{\psi}^{(\upsilon)}(0) = 0\,, \qquad \text{for } m+r \le \upsilon \le 2m+r-1\,, \\ \widetilde{\psi}^{(\upsilon)}(\ell) = 0\,, \qquad \text{for } 0 \le \upsilon \le r-1 \text{ and } m+r \le \upsilon \le 2m+r-1\,, \\ \sum_{\theta=0}^{m} \sum_{j=1}^{2\theta+2r+2} c_\theta \, \ell^{2D-j} \Big\{ a_{j,\upsilon} \, \widetilde{\psi}^{(2\theta+2r+2-j)}(0) + b_{j,\upsilon} \, \widetilde{\psi}^{(2\theta+2r+2-j)}(\ell) \Big\} \\ \qquad + \ell^{2m+r+\upsilon} \sum_{\theta=0}^{m} c_\theta \widetilde{\psi}^{(2\theta+r+\upsilon)}(0) = 0\,, \qquad \text{for } \upsilon = r, r+1\,. \end{cases} \tag{B.25}$$

Rearranging the last two boundary conditions in (B.25) yields that, for $\upsilon = r+1, r$ and for $a_{j,\upsilon}, b_{j,\upsilon}$ in (B.20),

$$\sum_{\theta=0}^{m}\sum_{j=0}^{2\theta+2r+1} c_\theta \ell^{2m-2\theta+j}\Big\{a_{2\theta+2r+2-j,\upsilon}\widetilde{\psi}^{(j)}(0)+b_{2\theta+2r+2-j,\upsilon}\widetilde{\psi}^{(j)}(\ell)\Big\}+\ell^{2m+r+\upsilon}\sum_{\theta=0}^{m}c_\theta\widetilde{\psi}^{(2\theta+r+\upsilon)}(0)$$
$$=\ell^{2D-1}\left[\sum_{j=0}^{2m+2r+1}\Big\{\widetilde{a}_{j,\upsilon,\ell}\,\widetilde{\psi}^{(j)}(0)+\widetilde{b}_{j,\upsilon,\ell}\,\widetilde{\psi}^{(j)}(\ell)\Big\}+\ell^{\upsilon-r-1}\sum_{\theta=0}^{m}c_\theta\,\widetilde{\psi}^{(2\theta+r+\upsilon)}(0)\right]=0\,.$$

For convenience of presentation, let $a_{j,\upsilon}=0$, for $-2m+1 \le j \le 0$. For $\upsilon = r, r+1$ and $0 \le j \le 2m+2r+1$, denote

$$\widetilde{a}_{j,\upsilon,\ell}=\sum_{\theta=0}^{m}\ell^{j-2\theta-2r-1}c_\theta\,a_{2\theta+2r+2-j,\upsilon}\,,\qquad \widetilde{b}_{j,\upsilon,\ell}=\sum_{\theta=0}^{m}\ell^{j-2\theta-2r-1}c_\theta\,b_{2\theta+2r+2-j,\upsilon}\,. \tag{B.26}$$

Let

$$\widetilde{W}_1(\widetilde{\psi})=\sum_{j=0}^{2m+2r+1}\Big\{\widetilde{a}_{j,r+1,\ell}\,\widetilde{\psi}^{(j)}(0)+\widetilde{b}_{j,r+1,\ell}\,\widetilde{\psi}^{(j)}(\ell)\Big\}+\sum_{\theta=0}^{m}c_\theta\widetilde{\psi}^{(2\theta+2r+1)}(0)$$
$$\widetilde{W}_2(\widetilde{\psi})=\sum_{j=0}^{2m+2r+1}\Big\{\widetilde{a}_{j,r,\ell}\,\widetilde{\psi}^{(j)}(0)+\widetilde{b}_{j,r,\ell}\,\widetilde{\psi}^{(j)}(\ell)\Big\}+\ell^{-1}\sum_{\theta=0}^{m}c_\theta\widetilde{\psi}^{(2\theta+2r)}(0)$$

In view of (B.22), for the $\widetilde{a}_{j,\upsilon,\ell}$'s and $\widetilde{b}_{j,\upsilon,\ell}$'s in (B.26), $\widetilde{\psi}$ satisfies the following differential equation with normalized boundary conditions (Birkhoff, 1908b, p. 382) is

$$\begin{cases} \widetilde{\psi}^{(2D)}(s)+\sum_{\theta=1}^{m}c_\theta\,\widetilde{\psi}^{(2D-2\theta)}(s)+\breve{\varrho}^{2D}\,\widetilde{\psi}(s)=0 \\ \widetilde{W}_1(\widetilde{\psi})=\sum_{j=0}^{2m+2r+1}\Big\{\widetilde{a}_{j,r+1,\ell}\,\widetilde{\psi}^{(j)}(0)+\widetilde{b}_{j,r+1,\ell}\,\widetilde{\psi}^{(j)}(\ell)\Big\}+\sum_{\theta=0}^{m}c_\theta\widetilde{\psi}^{(2\theta+2r+1)}(0)=0\,, \\ \widetilde{W}_2(\widetilde{\psi})=\sum_{j=0}^{2m+2r+1}\Big\{\widetilde{a}_{j,r,\ell}\,\widetilde{\psi}^{(j)}(0)+\widetilde{b}_{j,r,\ell}\,\widetilde{\psi}^{(j)}(\ell)\Big\}+\ell^{-1}\sum_{\theta=0}^{m}c_\theta\widetilde{\psi}^{(2\theta+2r)}(0)=0\,, \\ \widetilde{\psi}^{(\upsilon)}(0)=0\,,\quad \widetilde{\psi}^{(\upsilon)}(\ell)=0\,,\quad \text{for } m+r\le\upsilon\le 2m+r-1\,, \\ \widetilde{\psi}^{(\upsilon)}(\ell)=0\,,\quad \text{for } 0\le\upsilon\le r-1\,. \end{cases} \tag{B.27}$$

Let $\widetilde{W}_3(\widetilde{\psi}),\ldots,\widetilde{W}_{2D}(\widetilde{\psi})$ denote left hand side of the boundary conditions in the rest of the last two lines of the above equations. Now, ψ being the solution to (B.22) is equivalent to $\widetilde{\psi}$ being the solution to (B.27).

Next, we draw the conclusion of the growing rate of the characteristic value $\breve{\varrho}^{2D}$ in (B.27). For $\breve{\varrho} \in \mathbb{C}$, let $\widetilde{w}_1,\ldots,\widetilde{w}_{2D}$ denote the roots of $w^{2D}+1=0$, whose subscript is ordered according to

$$\mathrm{Re}(\breve{\varrho}\,\widetilde{w}_1)\le\mathrm{Re}(\breve{\varrho}\,\widetilde{w}_2)\le\cdots\le\mathrm{Re}(\breve{\varrho}\,\widetilde{w}_{2D})\,.$$

By Birkhoff (1908a) and Theorem III' in Stone (1926), for any $\varrho \in \mathbb{C}$, (B.24) has $2D$ linear independent analytic solutions $\breve{\psi}_{\ell,1}, \ldots, \breve{\psi}_{\ell,2D}$ in the form of

$$\breve{\psi}_{\ell,j}^{(v)}(s) = (\breve{\varrho}\widetilde{w}_j)^v \exp(\breve{\varrho}\widetilde{w}_j s)\left\{1 + \sum_{q=1}^{M-1} \frac{B_{q,v,\ell}(s)}{(\breve{\varrho}\widetilde{w}_j)^q} + \frac{E_{j,v,\ell}(s,\breve{\varrho})}{\breve{\varrho}^M}\right\}, \tag{B.28}$$

for some uniformly bounded functions $B_{q,v,\ell}$ and $E_{j,v,\ell}$. The condition that ϱ^{2D} is the characteristic value of (B.27) is that

$$\widetilde{\Delta} \equiv \begin{vmatrix} \widetilde{W}_1(\breve{\psi}_{\ell,1}) & \widetilde{W}_1(\breve{\psi}_{\ell,2}) & \cdots & \widetilde{W}_1(\breve{\psi}_{\ell,2D}) \\ \widetilde{W}_2(\breve{\psi}_{\ell,1}) & \widetilde{W}_2(\breve{\psi}_{\ell,2}) & \cdots & \widetilde{W}_2(\breve{\psi}_{\ell,2D}) \\ \cdots & \cdots & \cdots & \cdots \\ \widetilde{W}_{2D}(\breve{\psi}_{\ell,1}) & \widetilde{W}_{2D}(\breve{\psi}_{\ell,2}) & \cdots & \widetilde{W}_{2D}(\breve{\psi}_{\ell,2D}) \end{vmatrix} = 0\,. \tag{B.29}$$

In order to analyse the condition in (B.29), Birkhoff (1908b) introduced the definition of regular boundary conditions; see Birkhoff (1908b), p. 382 and Definition B.1 in Section B.2. For the $\widetilde{a}_{j,v,\ell}$'s and $\widetilde{b}_{j,v,\ell}$'s in (B.26), the boundary conditions in (B.27) is regular if $\zeta_{0,\ell}, \zeta_{1,\ell}, \zeta_{2,\ell}$ defined through the following equation according to the boundary conditions in (B.27) is such that $\zeta_{1,\ell}\,\zeta_{2,\ell} \neq 0$ for any $\ell \geq 1$:

$$\zeta_{0,\ell} + \zeta_{1,\ell}s + \frac{\zeta_{2,\ell}}{s} \equiv$$

$$\det \begin{bmatrix} (\widetilde{a}_{j,r+1,\ell}+c_0)\widetilde{w}_1^{2D-1} & \cdots & (\widetilde{a}_{j,r+1,\ell}+c_0)\widetilde{w}_{D-1}^{2D-1} & \{(\widetilde{a}_{j,r+1,\ell}+c_0)+\widetilde{b}_{j,r+1,\ell}s\}\widetilde{w}_D^{2D-1} & \{(\widetilde{a}_{j,r+1,\ell}+c_0)+\frac{\widetilde{b}_{j,r+1,\ell}}{s}\}\widetilde{w}_{D+1}^{2D-1} & \widetilde{b}_{j,r+1,\ell}\widetilde{w}_{D+2}^{2D-1} & \cdots & \widetilde{b}_{j,r+1,\ell}\widetilde{w}_{2D}^{2D-1} \\ \widetilde{a}_{j,r,\ell}\widetilde{w}_1^{2D} & \cdots & \widetilde{a}_{j,r,\ell}\widetilde{w}_{D-1}^{2D} & (\widetilde{a}_{j,r,\ell}+\widetilde{b}_{j,r,\ell}s)\widetilde{w}_D^{2D} & (\widetilde{a}_{j,r,\ell}+\frac{\widetilde{b}_{j,r,\ell}}{s})\widetilde{w}_{D+1}^{2D} & \widetilde{b}_{j,r,\ell}\widetilde{w}_{D+2}^{2D} & \cdots & \widetilde{b}_{j,r,\ell}\widetilde{w}_{2D}^{2D} \\ \widetilde{w}_1^{2m+r-1} & \cdots & \widetilde{w}_{D-1}^{2m+r-1} & \widetilde{w}_D^{2m+r-1} & \widetilde{w}_{D+1}^{2m+r-1} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & s\widetilde{w}_D^{2m+r-1} & \frac{1}{s}\widetilde{w}_D^{2m+r-1} & \widetilde{w}_{D+2}^{2m+r-1} & \cdots & \widetilde{w}_{2D}^{2m+r-1} \\ \widetilde{w}_1^{2m+r-2} & \cdots & \widetilde{w}_{D-1}^{2m+r-2} & \widetilde{w}_D^{2m+r-2} & \widetilde{w}_{D+1}^{2m+r-2} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & s\widetilde{w}_D^{2m+r-2} & \frac{1}{s}\widetilde{w}_D^{2m+r-2} & \widetilde{w}_{D+2}^{2m+r-2} & \cdots & \widetilde{w}_{2D}^{2m+r-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \widetilde{w}_1^{m+r} & \cdots & \widetilde{w}_{D-1}^{m+r} & \widetilde{w}_D^{m+r} & \widetilde{w}_{D+1}^{m+r} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & s\widetilde{w}_D^{m+r} & \frac{1}{s}\widetilde{w}_{D+1}^{m+r} & \widetilde{w}_{D+2}^{m+r} & \cdots & \widetilde{w}_{2D}^{m+r} \\ 0 & \cdots & 0 & s\widetilde{w}_D^{r-1} & \frac{1}{s}\widetilde{w}_{D+1}^{r-1} & \widetilde{w}_{D+2}^{r-1} & \cdots & b_n\widetilde{w}_{2D}^{r-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & s\widetilde{w}_D^{0} & \frac{1}{s}\widetilde{w}_{D+1}^{0} & \widetilde{w}_{D+2}^{0} & \cdots & b_n\widetilde{w}_{2D}^{0} \end{bmatrix}.$$

Then, letting $\mathrm{i} = \sqrt{-1}$, by the theorem in Birkhoff (1908b), p. 383, we find that the eigenvalue ϱ_k in (B.27) is of the following form:

$$\breve{\varrho}_k = \pm\frac{2k\pi\mathrm{i}}{\widetilde{w}_D} + \frac{1}{\widetilde{w}_D}\log\left(-\frac{\zeta_{1,\ell}}{\zeta_{2,\ell}}\right) + \sum_{j=1}^{M_0-1}\frac{e_{j,\ell}}{\breve{\varrho}^j} + \frac{E_\ell(\breve{\varrho})}{\breve{\varrho}^{M_0}}\,,$$

where $e_{j,\ell}(s)$ and $E_{2,\ell}$ are the coefficients of the equal or higher order terms of $\breve{\varrho}^{-1}$ from the determinant $\widetilde{\Delta}$ in (B.29), and $|e_{j,\ell}| \leq c_1$ and $|E_\ell| \leq c_2$ uniformly in $k, \ell \geq 1$ and $\breve{\varrho}$. Moreover, $c_1 \leq |\zeta_{1,\ell}/\zeta_{2,\ell}| \leq c_2$ for some $c_1, c_2 > 0$. Therefore, $(-1)^{D+1}\breve{\varrho}_k^{2D} \asymp k^{2D}$. In conclusion, the eigenvalue of (B.22) is $\breve{\rho}_{k\ell} = (-1)^{D+1}\ell^{2D}\breve{\varrho}_k^{2D} \asymp (k\ell)^{2D}$.

Suppose ϱ^{2D} is a characteristic value of (B.21) and suppose $\psi(s) = \widetilde{\psi}(\ell s)$ is the solution to the problem (B.21) corresponding to ϱ^{2D}. We rewrite the first equation in (B.21) that $\widetilde{\psi}$ satisfies. For the L_ℓ in (B.21) defined in (B.19), substituting $\psi(s)$ by $\widetilde{\psi}(\ell s)$ yields

$$\begin{aligned} L_\ell \psi(s_1) &= \sum_{\theta=0}^{m} \ell^{-2\theta-2r-2} c_\theta \int_0^1 M^{(0,2\theta+2r+2)}(s_1,s_2)\,\widetilde{\psi}(\ell s_2) ds_2 + \sum_{\theta=0}^{m}\sum_{j=1}^{2\theta+2r+2} (-1)^{m+r+j}\ell^{-j} c_\theta \\ &\quad \times \Big\{ M^{(0,j-1)}(s_1,1)\,\widetilde{\psi}^{(2\theta+2r+2-j)}(\ell) - M^{(0,j-1)}(s_1,0)\,\widetilde{\psi}^{(2\theta+2r+2-j)}(0) \Big\} \\ &= \int_0^1 \mathfrak{H}_\ell(s,\xi)\,\widetilde{\psi}(\ell\xi) d\xi + \sum_{\theta=0}^{m}\sum_{j=1}^{2\theta+2r+2} \Big\{ \mathfrak{A}_{j,\theta,\ell}(s)\widetilde{\psi}^{(2\theta+2r+2-j)}(\ell) - \mathfrak{B}_{j,\theta,\ell}(s)\widetilde{\psi}^{(2\theta+2r+2-j)}(0) \Big\}, \end{aligned}$$

where, for $0 \le \theta \le m$ and $1 \le j \le 2\theta + 2r + 2$,

$$\begin{aligned} \mathfrak{H}_\ell(s_1,s_2) &= \sum_{\theta=0}^{m} \ell^{-2\theta-2r-2} c_\theta M^{(0,2\theta+2r+2)}(s_1,s_2), \\ \mathfrak{A}_{j,\theta,\ell}(s) &= (-1)^{m+r+j}\ell^{-j} c_\theta\, M^{(0,j-1)}(s,1), \\ \mathfrak{B}_{j,\theta,\ell}(s) &= (-1)^{m+r+j}\ell^{-j} c_\theta\, M^{(0,j-1)}(s,0). \end{aligned}$$

We have $|\mathfrak{H}_\ell(s_1,s_2)| \le c\ell^{-2r-2}$ uniformly in $s_1, s_2 \in [0,1]$; $|\mathfrak{A}_{j,\theta,\ell}(s)|, |\mathfrak{B}_{j,\theta,\ell}(s)| \le c\ell^{-1}$ uniformly in $s \in [0,1]$, for $0 \le \theta \le m$ and $1 \le j \le 2\theta + 2r + 2$. Therefore, for $A_{\ell,v}$ is as in (B.20), we have that $\widetilde{\psi}(s)$ is the solution to the following equations

$$\begin{cases} \widetilde{\psi}^{(2D)}(s) + \sum_{\theta=1}^{m} c_\theta \widetilde{\psi}^{(2D-2\theta)}(s) + \varrho^{2D}\,\widetilde{\psi}(s) + \int_0^1 \mathfrak{H}_\ell(s,\xi)\,\widetilde{\psi}(\ell\xi) d\xi \\ \quad + \sum_{\theta=0}^{m}\sum_{j=1}^{2\theta+2r+2} \Big\{ \mathfrak{A}_{j,\theta,\ell}(s)\widetilde{\psi}^{(2\theta+2r+2-j)}(\ell) - \mathfrak{B}_{j,\theta,\ell}(s)\widetilde{\psi}^{(2\theta+2r+2-j)}(0) \Big\} = 0, \\ \widetilde{\psi}^{(v)}(0) = 0, \qquad \text{for } m + r \le v \le 2m + r - 1, \\ \widetilde{\psi}^{(v)}(\ell) = 0, \qquad \text{for } 0 \le v \le r - 1 \text{ and } m + r \le v \le 2m + r - 1, \\ \sum_{\theta=0}^{m}\sum_{j=1}^{2\theta+2r+2} \ell^{2D-j} c_\theta \Big\{ a_{j,v}\,\widetilde{\psi}^{(2\theta+2r+2-j)}(0) + b_{j,v}\,\widetilde{\psi}^{(2\theta+2r+2-j)}(\ell) \Big\} \\ \quad + \sum_{\theta=0}^{m} \ell^{2m+r+v} c_\theta\,\widetilde{\psi}^{(2\theta+r+v)}(0) + \int A_{\ell,v}(s)\widetilde{\psi}(s) ds = 0, \qquad \text{for } v = r, r+1. \end{cases} \tag{B.30}$$

Following the proof of Theorem 7 in Tamarkin (1927), the characteristic value ϱ_k of (B.30) and the characteristic value $\breve{\varrho}_k$ of (B.27) have the same growing rate uniformly in ℓ, so that $\rho_{k\ell} \asymp (k\ell)^{2D}$.

For the order of $\|x_{k\ell}\|_\infty$, Let $\widetilde{\gamma}_{k\ell} = \rho_{k\ell}^{1/(2D)} \in \mathbb{R}$ and $\gamma_{k\ell} = \mathrm{i}\widetilde{\gamma}_{k\ell} \exp\{\pi \mathrm{i}/(2D)\}$. We therefore have $\gamma_{k\ell}^{2k} = (-1)^{m+r}\rho_{k\ell}$, and $\mathrm{Re}(\gamma_{k\ell}) = -\widetilde{\gamma}_{k\ell}\sin\{\pi/(2D)\}$, and $\mathrm{Im}(\gamma_{k\ell}) = \widetilde{\gamma}_{k\ell}\cos\{\pi/(2D)\}$.

Let $w_{k\ell,1},\dots,w_{k\ell,2D}\in\big\{\exp\{(2\nu-1)\pi\mathrm{i}/(2D)\}\big\}_{1\le\nu\le 2D}$ denote the solution of $w^{2D}+1=0$, of which the subscripts are assigned according to the order

$$\mathrm{Re}(\ell\gamma_{k\ell}\,w_{k\ell,1})\le \mathrm{Re}(\ell\gamma_{k\ell}\,w_{k\ell,2})\le\cdots\le \mathrm{Re}(\ell\gamma_{k\ell}\,w_{k\ell,2D})\,,$$

so that $\mathrm{Re}(\gamma_{k\ell}\,w_{k\ell,j})<0$ when $1\le j\le D$, and $\mathrm{Re}(\gamma_{k\ell}\,w_{k\ell,j})\ge 0$ when $D+1\le\nu\le 2D$. Following Tamarkin and Langer (1928), pp. 467–469, the solution $\psi(s)$ corresponding to the eigenvalue $\rho_{k\ell}$ in (B.21) is such that, for $0\le\upsilon\le 2D-1$,

$$\psi^{(\upsilon)}_{k\ell}(s)=\gamma^{\upsilon}_{k\ell}\Bigg[\sum_{j=1}^{D}\exp(\gamma_{k\ell}\,w_{k\ell,j}s)[Q_{\ell,j}w^{\upsilon}_{k\ell,j}]+\sum_{j=D+1}^{2D}\exp\{\gamma_{k\ell}\,w_{k\ell,j}(s-1)\}[Q_{\ell,j}w^{\upsilon}_{k\ell,j}]\Bigg]\,, \tag{B.31}$$

where for $z\in\mathbb{C}$, $[z]$ is such that $|[z]-z|=O(k^{-1}+\ell^{-1})$, and $Q_{\ell,1},\dots,Q_{\ell,2D}$ are real-valued constants that does not depend on k and are bounded, and at least one of these $2D$ constants is non-zero. Without loss of generality we may assume that $\mathrm{Im}(w_{k\ell,D})<0$ and $\mathrm{Im}(w_{k\ell,D+1})>0$, so that $\gamma_{k\ell}\,w_{k\ell,D}=-\mathrm{i}\widetilde{\gamma}_{k\ell}$ and $\gamma_{k\ell}\,w_{k\ell,D+1}=\mathrm{i}\widetilde{\gamma}_{k\ell}$; when $j\ne D$ and $2D$, $|\mathrm{Re}(\gamma_{k\ell}\,w_{k\ell,j})|\ge\widetilde{\gamma}_{k\ell}\sin(\pi/D)$. Now, since $s\in[0,1]$, when $1\le\nu\le D-1$, $|\exp(\gamma_{k\ell}\,w_{k\ell,j}s)|=\exp\{\mathrm{Re}(\gamma_{k\ell}\,w_{k\ell,j})s\}\le 1$; when $D+2\le\nu\le 2D$, $|\exp\{\gamma_{k\ell}\,w_{k\ell,j}(s-1)\}|=\exp\{\mathrm{Re}(\gamma_{k\ell}\,w_{k\ell,j})(s-1)\}\le 1$; $|\exp(\gamma_{k\ell}\,w_{k\ell,D}s)|=|\exp\{\gamma_{k\ell}\,w_{k\ell,D+1}(s-1)\}|=1$. Therefore, from (B.31), we find that, for $0\le\theta\le m$,

$$\begin{aligned}&\sup_{s\in[0,1]}\big|\psi^{(r+\theta)}_{k\ell}(s)\big|\\&\le|\gamma_{k\ell}|^{r+\theta}\sup_{s\in[0,1]}\Bigg[\sum_{j=1}^{D}|\exp(\gamma_{k\ell}\,w_{k\ell,j}s)|\,|Q_{\ell,j}|+\sum_{j=D+1}^{2D}|\exp\{\gamma_{k\ell}\,w_{k\ell,j}(s-1)\}|\,|Q_{\ell,j}|+O(k^{-1})\Bigg]\\&\le|\gamma_{k\ell}|^{r+\theta}\Bigg\{\sup_{\ell\ge1}\sum_{j=1}^{2D}|Q_{\ell,j}|+O(k^{-1})\Bigg\}\asymp(\ell k)^{r+\theta}\,.\end{aligned} \tag{B.32}$$

Let

$$Z_{k\ell,\theta,j}(s)=\begin{cases}\exp(\gamma_{k\ell}\,w_{k\ell,j}s)[Q_{\ell,j}w^{r+\theta}_{k\ell,j}]\,, & \text{for } 1\le j\le D\,,\\ \exp\{\gamma_{k\ell}\,w_{k\ell,j}(s-1)\}[Q_{\ell,j}w^{r+\theta}_{k\ell,j}]\,, & \text{for } D+1\le j\le 2D\,,\end{cases}$$

so that $\psi^{(r+\theta)}_{k\ell}(s)=\gamma^{r+\theta}_{k\ell}\sum_{j=1}^{2D}Z_{k\ell,\theta,j}(s)$. Note that for $1\le j_1\le D-1$ and for $1\le j_2\le 2D$,

$$\begin{aligned}\big|Z_{k\ell,\theta,j_1}(s)\,\overline{Z_{k\ell,\theta,j_2}(s)}\big|&\le\big|\exp(\gamma_{k\ell}\,w_{k\ell,j_1}s)\big|\times|[Q_{\ell,j_1}]|\times|[Q_{\ell,j_2}]|+O(k^{-1})\\&=\exp\{\mathrm{Re}(\gamma_{k\ell}\,w_{k\ell,j_1})s\}\times|Q_{\ell,j_1}Q_{\ell,j_2}|+O(k^{-1})\\&\le\exp\{-\widetilde{\gamma}_{k\ell}\sin(\pi/D)s\}\times|Q_{\ell,j_1}Q_{\ell,j_2}|+O(k^{-1})\end{aligned}$$

$$\leq c_\ell \exp(-c'_\ell k s)\,, \tag{B.33}$$

for $c_\ell, c'_\ell > 0$. Hence, we deduce from the above equation that $\int_0^1 \big|Z_{k\ell,\theta,j_1}(s)\,\overline{Z_{k\ell,\theta,j_2}(s)}\big|ds \lesssim k^{-1}$. Likewise, for $D+2 \leq j_1 \leq 2D$ and for any $1 \leq j_2 \leq 2D$,

$$\begin{aligned}\big|Z_{k\ell,\theta,j_1}(s)\,\overline{Z_{k\ell,\theta,j_2}(s)}\big| &\leq \big|\exp\{\gamma_{k\ell}\, w_{k\ell,j_1}(s-1)\}\big| \times |[Q_{\ell,j_1}]| \times |[Q_{\ell,j_2}]| + O(k^{-1}) \\ &= \exp\{\mathrm{Re}(\gamma_{k\ell}\, w_{k\ell,j_1})(s-1)\} \times |Q_{\ell,j_1}Q_{\ell,j_2}| + O(k^{-1}) \\ &\leq \exp\{\widetilde{\gamma}_{k\ell}\sin(\pi/D)(s-1)\} \times |Q_{\ell,j_1}Q_{\ell,j_2}| + O(k^{-1}) \\ &\leq c_\ell \exp\{-c'_\ell k(s-1)\}\,, \end{aligned} \tag{B.34}$$

for $c_\ell, c'_\ell > 0$. Hence, we find that $\int_0^1 \big|Z_{k\ell,\theta,j_1}(s)\,\overline{Z_{k\ell,\theta,j_2}(s)}\big|ds \lesssim k^{-1}$. Recall that $\gamma_{k\ell}\, w_{k\ell,D} = -\mathrm{i}\widetilde{\gamma}_{k\ell}$ and $\gamma_{k\ell}\, w_{k\ell,D+1} = \mathrm{i}\widetilde{\gamma}_{k\ell}$. We have

$$\begin{aligned}&\big|Z_{k\ell,\theta,D}(s)\big|^2 = \exp\{2\mathrm{Re}(\gamma_{k\ell}\, w_{k\ell,D})s\} \times |Q_{\ell,D}|^2 + O(k^{-1}) = |Q_{\ell,D}|^2 + O(k^{-1})\,. \\ &\big|Z_{k\ell,\theta,D+1}(s)\big|^2 = \exp\{2\mathrm{Re}(\gamma_{k\ell}\, w_{k\ell,D+1})(s-1)\} \times |Q_{\ell,D+1}|^2 + O(k^{-1}) = |Q_{\ell,D+1}|^2 + O(k^{-1})\,.\end{aligned}$$

Therefore,

$$\begin{aligned}\|\psi_{k\ell}^{(r+\theta)}\|_{L^2}^2 &= |\gamma_{k\ell}|^{2r+2\theta}\sum_{j_1=1}^{2D}\sum_{j_2=1}^{2D}\int Z_{k\ell,\theta,j_1}(s)\,\overline{Z_{k\ell,\theta,j_2}(s)}ds \\ &= |\gamma_{k\ell}|^{2r+2\theta}\bigg\{\int\big|Z_{k\ell,\theta,D}(s)\big|^2 ds + \int\big|Z_{k\ell,\theta,D+1}(s)\big|^2 ds + O(k^{-1})\bigg\} \\ &= |\gamma_{k\ell}|^{2r+2\theta}\Big\{|Q_{\ell,D}|^2 + |Q_{\ell,D+1}|^2 + O(k^{-1})\Big\}\,.\end{aligned}$$

Now we show that in the above equation $|Q_{\ell,D}|^2 + |Q_{\ell,D+1}|^2 \neq 0$, which can be proved by contradiction. Suppose $Q_{\ell,D} = Q_{\ell,D+1} = 0$. Then, in view of the boundary condition in (B.22), for $m+r \leq \upsilon_1 \leq 2m+r-1$, and for υ_2 such that $0 \leq \upsilon_2 \leq r-1$ or $m+r \leq \upsilon_2 \leq 2m+r-1$,

$$\begin{aligned}0 = \psi_{k\ell}^{(\upsilon_1)}(0) &= \gamma_{k\ell}^{\upsilon_1}\bigg\{\sum_{j=1}^{D-1}[Q_{\ell,j}w_{k\ell,j}^{\upsilon_1}] + \sum_{j=D+2}^{2D}\exp(-\gamma_{k\ell}\, w_{k\ell,j})[Q_{\ell,j}w_{k\ell,j}^{\upsilon_1}] + O(k^{-1})\bigg\}\,, \\ 0 = \psi_{k\ell}^{(\upsilon_2)}(1) &= \gamma_{k\ell}^{\upsilon_2}\bigg\{\sum_{j=1}^{D-1}\exp(\gamma_{k\ell}\, w_{k\ell,j})[Q_{\ell,j}w_{k\ell,j}^{\upsilon_2}] + \sum_{j=D+2}^{2D}[Q_{\ell,j}w_{k\ell,j}^{\upsilon_2}] + O(k^{-1})\bigg\}\,.\end{aligned}$$

Following the arguments in Shang and Cheng (2015b), pp. 7–9, letting $k \to \infty$ yields $Q_{\ell,j} = 0$ for all $1 \leq j \leq 2D$, which contradicts with the fact that at least one of these $2D$ constants is nonzero. Therefore, we deduce that $|Q_{\ell,D}|^2 + |Q_{\ell,D+1}|^2 \neq 0$, so that $\|\psi_{k\ell}^{(r+\theta)}\|_{L^2}^2 \asymp |\gamma_{k\ell}|^{2r+2\theta} \asymp (\ell k)^{2r+2\theta}$, for $0 \leq \theta \leq m$. From this we deduce that

$$\langle \mathcal{C}_X\psi_{k\ell}^{(r)}, \psi_{k\ell}^{(r)}\rangle_{L^2} = \rho_{k\ell}^{-1}\sum_{\theta=0}^{m}\binom{m}{\theta}\|\eta_\ell^{(m-\theta)}\|_{L^2}^2\big\|\psi_{k\ell}^{(r+\theta)}\big\|_{L^2}^2 \asymp (k\ell)^{-2}\,.$$

Recall that $x_{k\ell} = \langle \mathcal{C}_X(\psi_{k\ell}^{(r)}), \psi_{k\ell}^{(r)} \rangle_{L^2}^{-1/2} \psi_{k\ell}^{(r)}$. Hence, from the above equation and (B.32) we conclude that $\|x_{k\ell}\|_\infty = \langle \mathcal{C}_X(\psi_{k\ell}^{(r)}), \psi_{k\ell}^{(r)} \rangle_{L^2}^{-1/2} \|\psi_{k\ell}^{(r)}\|_\infty \lesssim (k\ell)^{r+1}$.

Since the cosine series $\{\eta_\ell\}_{\ell \ge 1}$ is a complete basis of $L^2([0,1])$ (Theorem 2.4.18 in Hsing and Eubank, 2015), by the argument similar to the ones in Shang and Cheng (2015b) p. 7, we have $\{\varphi_{k\ell}\}_{k,\ell\ge1}$ is complete in $\mathcal{H}$, and any $\beta \in \mathcal{H}$ admits the Fourier expansion $\beta = \sum_{k,\ell} V(\beta, \varphi_{k\ell})\varphi_{k\ell}$.

In order to show (2.12), due to Assumption A1, $V_X(x_1, x_2) \equiv \langle \mathcal{C}_X x_1, x_2 \rangle_{L^2}$ defines an inner product, for $x_1, x_2 \in L^2([0,1])$. For each $\ell \ge 1$, we may orthonormalize the $\breve{x}_{k\ell}$'s, $k \ge 1$, w.r.t. V_X to obtain the $x_{k\ell}$'s. That is, $V_X(x_{k\ell}, x_{k'\ell}) = \delta_{kk'}$, for $k, k' \ge 1$. Recall that $\varphi_{k\ell}(s,t) = x_{k\ell}(s)\eta_\ell(t)$, $k, \ell \ge 1$. For V in (2.9),

$$\begin{aligned}
V(\varphi_{k\ell}, \varphi_{k'\ell'}) &= \int_{s_1,s_2\in[0,1];\, t\in[0,1]} C_X(s_1,s_2)\big\{x_{k\ell}(s_1)\eta_\ell(t)\big\}\big\{x_{k'\ell'}(s_2)\eta_{\ell'}(t)\big\}\, ds_1 ds_2 dt \\
&= \int_{[0,1]^2} C_X(s_1,s_2) x_{j\ell}(s_1) x_{k\ell'}(s_2) ds_1 ds_2 \times \int_0^1 \eta_\ell(t)\eta_{\ell'}(t)\, dt \\
&= \big\langle \mathcal{C}_X(x_{k\ell}), x_{k\ell'} \big\rangle_{L^2} \times \langle \eta_\ell, \eta_{\ell'} \rangle_{L^2} = V_X(x_{k\ell}, x_{k'\ell})\, \delta_{\ell\ell'} = \delta_{kk'}\, \delta_{\ell\ell'}\,.
\end{aligned}$$

For J in (2.6) and the η_ℓ in (B.10), we have $\langle \eta_\ell^{(\theta)}, \eta_{\ell'}^{(\theta)} \rangle_{L^2} = 0$ for $0 \le \theta \le m$ and $\ell \ge 1$, so that

$$\begin{aligned}
J(\varphi_{k\ell}, \varphi_{k'\ell'}) &= J(x_{k\ell} \otimes \eta_\ell, x_{k'\ell'} \otimes \eta_{\ell'}) = \sum_{\theta=0}^m \binom{m}{\theta} \big\langle x_{k\ell}^{(\theta)}, x_{k'\ell'}^{(\theta)} \big\rangle_{L^2} \big\langle \eta_\ell^{(m-\theta)}, \eta_{\ell'}^{(m-\theta)} \big\rangle_{L^2} \\
&= \sum_{\theta=0}^m \binom{m}{\theta} \big\langle x_{k\ell}^{(\theta)}, x_{k'\ell'}^{(\theta)} \big\rangle_{L^2} \big\langle \eta_\ell^{(m-\theta)}, \eta_{\ell'}^{(m-\theta)} \big\rangle_{L^2} \\
&= \delta_{\ell\ell'} \sum_{\theta=0}^m \binom{m}{\theta} \big\langle x_{k\ell}^{(\theta)}, x_{k'\ell'}^{(\theta)} \big\rangle_{L^2} \|\eta_\ell^{(m-\theta)}\|_{L^2}^2\,.
\end{aligned}$$

Using integration by parts and the boundary conditions, in view of (B.13), we deduce from the above equation that

$$\begin{aligned}
J(\varphi_{k\ell}, \varphi_{k'\ell'}) &= \delta_{\ell\ell'} \sum_{\theta=0}^m \binom{m}{\theta} (-1)^\theta \big\langle x_{k\ell}^{(2\theta)}, x_{k'\ell'} \big\rangle_{L^2} \|\eta_\ell^{(m-\theta)}\|_{L^2}^2 \\
&= \delta_{\ell\ell'} \left\langle \sum_{\theta=0}^m \binom{m}{\theta} (-1)^\theta \|\eta_\ell^{(m-\theta)}\|_{L^2}^2\, x_{k'\ell'}^{(2\theta)}\,, x_{k'\ell'} \right\rangle_{L^2} \\
&= \delta_{\ell\ell'} \left\langle \rho_{k\ell} \int_0^1 C_X(s_1,s_2)\, x_{k\ell'}(s_2)\, ds_2\,, x_{k'\ell'} \right\rangle_{L^2} \\
&= \delta_{\ell\ell'}\, \rho_{k\ell}\, V_X(x_{j\ell}, x_{k\ell'}) = \rho_{k\ell}\, \delta_{kk'}\, \delta_{\ell\ell'}\,,
\end{aligned}$$

which completes the proof.

References

Birkhoff, G. D. (1908a). On the asymptotic character of the solutions of certain linear differential equations containing a parameter. *Trans. Amer. Math. Soc.*, **9**, 219–231.

Birkhoff, G. D. (1908b). Boundary value and expansion problems of ordinary linear differential equations. *Trans. Amer. Math. Soc.*, **9**, 373–395.

Birman, M. Š. and Solomjak, M. Z. (1967). Piecewise-polynomial approximations of functions of the classes. *Mathematics of the USSR-Sbornik*, **2**, 295–317.

Bücher, A. and Kojadinovic, I. (2019). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *J. Theor. Probab.*, **32**, 1145–1165.

Cucker, F. and Smale, S. (2001). On the mathematical foundations of learning. *Bull. Amer. Math. Soc.*, **39**, 1–49.

Gu, C. (2013). *Smoothing Spline ANOVA Models.* Springer Science & Business Media.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.* John Wiley & Sons.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields*, **75**, 261–277.

Kley, T., Volgushev, S., Dette, H. and Hallin, M. (2016). Quantile spectral processes: Asymptotic analysis and inference. *Bernoulli*, **22**, 1770–1807.

Kosorok, M. R. (2007). *Introduction to empirical processes and semiparametric inference.* Springer Science & Business Media.

Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.*, **22**, 1679–1706.

Ritter, K., Wasilkowski, G. W. and Wozniakowski, H. (1995). Multivariate integration and approximation for random fields satisfying Sacks-Ylvisaker conditions. *Ann. Appl. Probab.*, **5**, 518–540.

Shang, Z. and Cheng, G. (2015b) Supplement to "Nonparametric inference in generalized functional linear models". DOI:10.1214/15-AOS1322SUPP.

Stone, M. H. (1926). A comparison of the series of Fourier and Birkhoff. *Trans. Amer. Math. Soc.*, **28**, 695–761.

Tamarkin, J. D. (1927). The notion of the Green's function in the theory of integro-differential equations. *Trans. Amer. Math. Soc.*, **29**, 755–800.

Tamarkin, J. D. (1930). The Notion of the Green's Function in the Theory of Integro-Differential Equations, II. *Trans. Amer. Math. Soc.*, **32**, 860–868.

Tamarkin, J. D. and Langer, R. E. (1928). On integral equations with discontinuous kernels. *Trans. Amer. Math. Soc.*, **30**, 453–471.

Tsybakov, A. B. (2008). *Introduction to nonparametric estimation.* Springer Science & Business Media.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge university press.