Statistics and Its Interface Volume 17 (2024) 69–78

A nonparametric concurrent regression model with multivariate functional inputs

Yutong Zhai, Zhanfeng Wang*, and Yuedong Wang

Regression models with functional responses and covariates have attracted extensive research. Nevertheless, there is no existing method for the situation where the functional covariates are bivariate functions with one of the variables in common with the response function. In this article, we propose a nonparametric function-on-function regression method. We construct model spaces using a Gaussian kernel function and smoothing spline ANOVA decomposition. We estimate the nonparametric function using penalized likelihood and study properties of the Gaussian kernel function and the convergence rate of the proposed estimation method. We evaluate the proposed methods using simulations and illustrate them using two real data examples.

Keywords and phrases: Function-on-function regression, Gaussian kernel, Reproducing kernel Hilbert space, Smoothing spline.

1. INTRODUCTION

Functional data frequently occur in science, engineering, economics, and medicine. Statistical methods for functional data have been studied extensively [14, 6, 8, 10]. Nonparametric and semiparametric approaches are rapidly developed for function-on-function regression where both response and covariates are functions [12, 24, 3, 26]. Nevertheless, to the best of our knowledge, there is no existing method for the situation where the functional covariates are bivariate functions with one of the variables in common with that of the response function.

Our research was motivated by a stroke rehabilitation study in which a set of 3D video games called Circus Challenge was used to improve stroke patients' upper limb function [17, 18]. Each patient played the movement game at scheduled times over a period of three months. At each visit time t, a CAHAI (Chedoke Arm and Hand Activity Inventory) score denoted as $y_i(t)$ is used to measure the impairment level of subject i, and movement such as the forward circle movement and sawing movement of upper limbs are also collected. Denote $\boldsymbol{x}_i(t,s)$ as the movement data at time t and frequency s from the ith patient. We treat the response $y_i(t)$ as a function over time t and covariates $\boldsymbol{x}_i(t,s)$ as a vector of bivariate functions of time t and frequency s. We are interested in the relationship between the CAHAI score and movement functions and predicting the CAHAI score using movement functions.

As a second example, denote $x_i(t,s)$ as the fertility rate of a woman of age s in year t and $y_i(t)$ as the mortality rate in year t from the ith country. We want to investigate the relationship between fertility and mortality rates, a topic of interest in population studies [20, 4, 13, 9]. Again, the response $y_i(t)$ is a function of t and the covariate $x_i(t,s)$ is a vector of bivariate functions of time t and s.

In this paper, we propose the following nonparametric concurrent regression model (NCRM)

$$y_i(t) = f(t, \boldsymbol{x}_i(t,\cdot)) + \epsilon_i(t), \quad i = 1, \cdots, n, \tag{1}$$

where f is a bivariate functional to be estimated nonparametrically, and $\epsilon_i(t)$ is a random error.

When $\boldsymbol{x}_i(t,s) = \boldsymbol{x}_i(t)$ which is a vector of functions independent of s, [14] proposed the following concurrent linear model

$$y_i(t) = \alpha(t) + \boldsymbol{x}_i(t)\boldsymbol{\beta}(t) + \epsilon_i(t), \quad i = 1, \cdots, n, \tag{2}$$

where $\alpha(t)$ is an unknown function and $\boldsymbol{\beta}(t)$ is an unknown column vector of functions. This model assumes a linear relationship between the response variable $y_i(t)$ and the functional covariates $\boldsymbol{x}_i(t)$. To allow for nonlinear relationship, [27, 15, 11, 7], and [23] considered the nonparametric concurrent model (NCM)

$$y_i(t) = g(t, \boldsymbol{x}_i(t)) + \epsilon_i(t), \quad i = 1, \cdots, n, \tag{3}$$

where g is an unknown bivariate function. The proposed NCRM (1) extends (3) to the situation where each covariate is a bivariate function. We note that the extension is not straightforward since the map f is a functional on $\boldsymbol{x}_i(t,\cdot)$.

When $\boldsymbol{x}_i(t,s) = \boldsymbol{x}_i(s)$ which is a vector of function independent of t, [14] proposed the following functional linear model

$$y_i(t) = \alpha(t) + \int \boldsymbol{x}_i(s)\boldsymbol{\beta}(t,s)ds + \epsilon_i(t), \quad i = 1, \cdots, n, \tag{4}$$

where $\alpha(t)$ is an unknown function and $\boldsymbol{\beta}(t.s)$ is an unknown column vector of bivariate functions. To allow for nonlinear relationship, [26] considered the following nonparametric

*Corresponding author.

model

$$(5)\quad y_i(t) = \int g(t, s, \boldsymbol{x}_i(t), \boldsymbol{x}_i(s))ds + \epsilon_i(t), \quad i = 1, \cdots, n,$$

where g is an unknown four-variable function to be estimated. Instead of the specific integration over s in (5), the functional on $\boldsymbol{x}_i(t, \cdot)$ in the proposed model (1) is unspecified. Therefore, model (1) provides a more flexible approach to modeling the relationship between functional response and functional covariates.

To the best of our knowledge, model (1) is new. We will model the map f in (1) nonparametrically. Specifically, we will construct a tensor product of two reproducing kernel Hilbert spaces (RKHS) and derive a smoothing spline ANOVA (SS ANOVA) decomposition for the tensor product space [25, 5]. In addition to the new and flexible model (1), this paper makes several contributions to the literature. First, we construct a Gaussian kernel for the functional variables and build the RKHS induced by the Gaussian kernel. We show that this kernel is strictly positive definite, and the induced RKHS is separable and does not contain constant functions. Second, based on the SS ANOVA decomposition of the tensor product RKHSs, we estimate the function f using penalized least squares method and empirical functional principal components of the response function space. We employ the backfitting method to develop an efficient computation algorithm. Third, we study the convergence rate of the proposed estimation for model (1) under the L-risk consistency [19]. We derive the convergence rate of the function estimation under the norm of the RKHS. Simulations show the proposed method performs well.

The structure of this paper is organized as follows. Section 2 introduces the Gaussian kernel and SS ANOVA decomposition. Section 3 presents the model estimation procedure and computation algorithm. In Section 4, we study properties of the proposed reproducing kernel and the convergence rate of the estimated regression function. Applications to real data and simulations are presented in Sections 5 and 6. Technical details and proofs are included in Appendix.

2. NONPARAMETRIC CONCURRENT REGRESSION MODEL

We consider the NCRM (1) where $\boldsymbol{x}_i(t, \cdot) = (x_{i1}(t, \cdot), ..., x_{iq}(t, \cdot))$ is a q-dimensional vector of functional variables at time t. For each fixed time $t \in \mathcal{T}$ and $j \in \{1, \cdots, q\}$, $x_{ij}(t, s) : \mathcal{S} \to \mathbb{R}$ is a function of s in a space denoted as $\mathcal{X}$. For simplicity, we assume that $\mathcal{T} = [0, 1]$, $\mathcal{S} = [0, 1]$ which is independent of t, and $\mathcal{X} \subset L^2[0, 1]$ which is also independent of t. Furthermore, we assume that $y_i(t) \in \mathcal{Y} \subset L^2[0, 1]$, and $\epsilon_i(t)$ for $i = 1, \cdots, n$ are independent identically distributed in $L^2[0, 1]$ with mean zero and $\int_0^1 E[\epsilon_i(t)^2]dt < \infty$. These are reasonable assumptions for many applications, including two examples considered in this paper.

For simplicity, we will refer f in the NCRM as a function even though it is a functional when applied to $\boldsymbol{x}_i(t, \cdot)$ as functional data. To estimate f nonparametrically, we now construct model spaces for f using SS ANOVA decomposition of the tensor product of two RKHSs.

For f as a function of $t \in \mathcal{T}$, we consider the Sobolev space

$$(6)\qquad \mathcal{H}^{(1)} = \Big\{ f : \ f \text{ and } f' \text{ are absolutely continuous}, \int_0^1 (f'')^2 dt < \infty \Big\}.$$

The space $\mathcal{H}^{(1)}$ has the following direct sum decomposition:

$$\mathcal{H}^{(1)} = \{1\} \oplus \{t\} \oplus \mathcal{H}_2^{(1)},$$

where $\{1\}$, $\{t\}$, and $\mathcal{H}_2^{(1)}$ respectively consist of constant functions, linear functions, and functions orthogonal to the constant and linear functions. The reproducing kernels (RK) of these three spaces are $R_0^{(1)}(t, t') = 1$, $R_1^{(1)}(t, t') = k_1(t)k_1(t')$, and $R_2^{(1)}(t, t') = k_2(t)k_2(t') - k_4(|t - t'|)$, where k_1, k_2, and k_4 are scaled Bernoulli polynomials defined as

$$\begin{aligned} k_1(x) &= x - 0.5, \\ k_2(x) &= \frac{1}{2}\left\{k_1^2(x) - \frac{1}{12}\right\}, \\ k_4(x) &= \frac{1}{24}\left\{k_1^4(x) - \frac{1}{2}k_1^2(x) + \frac{7}{240}\right\}. \end{aligned}$$

See [25] for details.

We next construct an RK and its corresponding RKHS for f as a function (functional) of functions in $\mathcal{X}^q$. For any $\boldsymbol{u} = (u_1, \cdots, u_q) \in \mathcal{X}^q$ and $\boldsymbol{u}' = (u_1', \cdots, u_q') \in \mathcal{X}^q$, we construct a Gaussian kernel as

$$(7)\qquad R_2^{(2)}(\boldsymbol{u}, \boldsymbol{u}') = \exp\left\{-\frac{||\boldsymbol{u} - \boldsymbol{u}'||^2}{2}\right\},$$

where $||\boldsymbol{u}||^2 = \sum_{i=1}^q \int_0^1 u_i^2(s)ds$. The following theorem shows that $R_2^{(2)}$ is positive definite.

Theorem 1. *Assume that $\mathcal{X}$ is a complete space, then*

(i) $R_2^{(2)}$ is symmetric and positive definite,

(ii) $\mathcal{H}_2^{(2)}$ is separable and does not contain non-zero constant functions.

The proof of Theorem 1 is included in the Appendix. By the Moore-Aronszajn theorem, there exists a unique RKHS $\mathcal{H}_2^{(2)}$ with $R_2^{(2)}$ as its RK [21].

For the construction of an SS ANOVA decomposition, we consider $\mathcal{H}^{(2)} = \{1\} \oplus \mathcal{H}_2^{(2)}$. The tensor product space $\mathcal{H} =$

$\mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}$ has the following SS ANOVA decomposition [25]:

$$
\begin{aligned}
\mathcal{H} &= \left(\{1\} \oplus \{t\} \oplus \mathcal{H}_2^{(1)}\right) \otimes \left(\{1\} \oplus \mathcal{H}_2^{(2)}\right) \\
&= \{1\} \oplus \{t\} \oplus \mathcal{H}_2^{(1)} \oplus \mathcal{H}_2^{(2)} \\
&\oplus \left\{\{t\} \otimes \mathcal{H}_2^{(2)}\right\} \oplus \left\{\mathcal{H}_2^{(1)} \otimes \mathcal{H}_2^{(2)}\right\} \\
&\triangleq \mathcal{H}_0 \oplus \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \mathcal{H}_3 \oplus \mathcal{H}_4,
\end{aligned} \tag{8}
$$

where $\mathcal{H}_0 = \{1\} \oplus \{t\}$ corresponds to parametric main effect of t, $\mathcal{H}_1 = \mathcal{H}_2^{(1)}$ corresponds to nonparametric main effect of t, $\mathcal{H}_2 = \mathcal{H}_2^{(2)}$ corresponds to nonparametric main effect of $\boldsymbol{u}$, $\mathcal{H}_3 = \{t\} \otimes \mathcal{H}_2^{(2)}$ corresponds to linear-nonparametric interaction between t and $\boldsymbol{u}$, and $\mathcal{H}_4 = \mathcal{H}_2^{(1)} \otimes \mathcal{H}_2^{(2)}$ corresponds to nonparametric-nonparametric interaction between t and $\boldsymbol{u}$. Denote $\varphi_1(t, \boldsymbol{u}) = 1$ and $\varphi_2(t, \boldsymbol{u}) = k_1(t)$ as basis functions of $\mathcal{H}_0$. Using the fact that the RK of a tensor product space equals the product of RKs [25], we have RKs

$$
\begin{aligned}
R_0((t, \boldsymbol{u}), (t', \boldsymbol{u}')) &= 1 + k_1(t)k_1(t'), \\
R_1((t, \boldsymbol{u}), (t', \boldsymbol{u}')) &= R_2^{(1)}(t, t'), \\
R_2((t, \boldsymbol{u}), (t', \boldsymbol{u}')) &= R_2^{(2)}(\boldsymbol{u}, \boldsymbol{u}'), \\
R_3((t, \boldsymbol{u}), (t', \boldsymbol{u}')) &= R_1^{(1)}(t, t')R_2^{(2)}(\boldsymbol{u}, \boldsymbol{u}'), \\
R_4((t, \boldsymbol{u}), (t', \boldsymbol{u}')) &= R_2^{(1)}(t, t')R_2^{(2)}(\boldsymbol{u}, \boldsymbol{u}'),
\end{aligned} \tag{9}
$$

for spaces $\mathcal{H}_0$, $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$, and $\mathcal{H}_4$, respectively. We may consider any subset of subspaces as the model space for f. For the simplicity of notations, we consider the full space $\mathcal{H}$ in the remainder of the manuscript.

3. ESTIMATION AND COMPUTATION

We estimate f by minimizing the following penalized least squares

$$
\frac{1}{n}\sum_{i=1}^{n}\left\{\int_0^1 (y_i(t) - f(t, \boldsymbol{x}_i(t, \cdot)))^2 dt\right\} + \sum_{v=1}^{4} \lambda_v ||P_v f||_{\mathcal{H}}^2 \tag{10}
$$

where $f \in \mathcal{H}$ in (8), λ_v are smoothing parameters, P_v is the orthogonal projection in $\mathcal{H}$ onto $\mathcal{H}_v$, and $||\cdot||_{\mathcal{H}}$ is a norm induced from $\mathcal{H}$ (details seen in [22]).

Let $\lambda_v = \lambda/\theta_v$ for $v = 1, \cdots, 4$ and $\mathcal{H}^* = \oplus_{v=1}^4 \mathcal{H}_v$. Define a new inner product in $\mathcal{H}^*$ as

$$
\langle f, g\rangle_{\mathcal{H}^*} = \sum_{v=1}^{4} \frac{1}{\theta_v}\langle P_v f, P_v g\rangle_{\mathcal{H}},
$$

where $\langle \cdot, \cdot\rangle_{\mathcal{H}}$ is the inner product in $\mathcal{H}$ and $||f||_{\mathcal{H}^*} = \sqrt{\langle f, f\rangle_{\mathcal{H}^*}}$. Then it is easy to check that the RK of $\mathcal{H}^*$ under the new inner product is

$$
R^*((t, \boldsymbol{u}), (t', \boldsymbol{u}')) = \sum_{v=1}^{4} \theta_v R_v((t, \boldsymbol{u}), (t', \boldsymbol{u}')).
$$

Since the response functions are stochastic processes in $L^2[0, 1]$, there exists a set of orthogonal basis functions $\{\phi_k(t), k = 1, 2, \cdots\}$ in $L^2[0, 1]$ such that $\{\phi_k(t), k = 1, \cdots, n\}$ are the empirical functional principal components of $\{y_1(t), \cdots, y_n(t)\}$ [6]. Let $v_{ik} = \langle y_i(t), \phi_k(t)\rangle$ and $L_{ik} f = \int_0^1 f(t, \boldsymbol{x}_i(t, \cdot)))\phi_k(t)dt$ for $i = 1, \cdots, n$ and $k = 1, \cdots, n$. Assume that L_{ik} are bounded linear functionals. Then, following similar arguments in [26], it can be shown that the PLS (10) based on functional data $y_i(t)$ reduces to the following PLS based on scalar data v_{ik}:

$$
\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}(v_{ik} - L_{ik} f)^2 + \lambda||f||_{\mathcal{H}^*}^2. \tag{11}
$$

Let $\mathcal{H}_{1n} = \text{span}\{\int_0^1 R^*((t, \boldsymbol{x}(t, \cdot)), (t', \boldsymbol{x}_i(t', \cdot)))\phi_k(t')dt', i = 1, ..., n, \ k = 1, ..., n\}$, which is a subspace of $\mathcal{H}^*$. Then any $f \in \mathcal{H}^*$ can be decomposed as

$$
f = f_0 + f_{1n} + \rho,
$$

where $f_0 \in \mathcal{H}_0$, $f_{1n} \in \mathcal{H}_{1n}$, and $\rho \in \mathcal{H}^* \ominus \mathcal{H}_{1n}$. Denote $R^*_{(t', \boldsymbol{x}_i(t', \cdot))}(t, \boldsymbol{x}(t, \cdot)) = R^*((t, \boldsymbol{x}(t, \cdot)), (t', \boldsymbol{x}_i(t', \cdot)))$ as a function with the second variable in the RK being fixed at $(t', \boldsymbol{x}_i(t', \cdot))$, and $f_1 = f_{1n} + \rho$. Then we can rewrite the PLS (11) as

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}(v_{ik} - u_{ik} - \langle f_1(t', \boldsymbol{x}_i(t', \cdot)), \phi_k(t')\rangle)^2 + \lambda||f||_{\mathcal{H}^*}^2 \\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}(v_{ik} - u_{ik} - \langle\langle f_1, R^*_{(t', \boldsymbol{x}_i(t', \cdot))}\rangle_{\mathcal{H}^*}, \phi_k(t')\rangle)^2 \\
&+ \lambda||f||_{\mathcal{H}^*}^2 \\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}(v_{ik} - u_{ik} - \langle f_1, \int_0^1 R^*_{(t', \boldsymbol{x}_i(t', \cdot))}\phi_k(t')dt'\rangle_{\mathcal{H}^*})^2 \\
&+ \lambda||f||_{\mathcal{H}^*}^2 \\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}(v_{ik} - u_{ik} - \langle f_{1n}, \int_0^1 R^*_{(t', \boldsymbol{x}_i(t', \cdot))}\phi_k(t')dt'\rangle_{\mathcal{H}^*})^2 \\
&+ \lambda||f_{1n}||_{\mathcal{H}^*}^2 + \lambda||\rho||_{\mathcal{H}^*}^2,
\end{aligned} \tag{12}
$$

where $u_{ik} = \int_0^1 f_0(t', \boldsymbol{x}_i(t', \cdot))\phi_k(t')dt'$, the first equality uses the reproducing property, and the third equality uses the fact that ρ is orthogonal to $\mathcal{H}_{1n}$. Minimizing (12) must have $\rho = 0$, and we obtain the following representer theorem.

Theorem 2 (Representer Theorem)**.** *The solution to the PLS (10) is*

$$
\hat{f}(t, \boldsymbol{x}(t, \cdot)) = \sum_{j=1}^{2} d_j\varphi_j(t) + \sum_{i=1}^{n}\sum_{k=1}^{n} c_{ik}\xi_{ik}(t, \boldsymbol{x}(t, \cdot)), \tag{13}
$$

where $\varphi_1(t) = 1$, $\varphi_2(t) = k_1(t)$, *and* $\xi_{ik}(t, \boldsymbol{x}(t, \cdot)) = \int_0^1 R^*((t, \boldsymbol{x}(t, \cdot)), (t', \boldsymbol{x}_i(t', \cdot)))\phi_k(t')dt'$.

The PLS (12) reduces to

$$(14)\qquad \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}(v_{ik} - \langle f_0, \int_0^1 R_{0(t',\boldsymbol{x}_i(t',\cdot))}\phi_k(t')dt'\rangle_{\mathcal{H}} - \langle f_{1n}, \int_0^1 R^*_{(t',\boldsymbol{x}_i(t',\cdot))}\phi_k(t')dt'\rangle_{\mathcal{H}^*})^2 + \lambda||f_{1n}||^2_{\mathcal{H}^*}.$$

It can be shown that

$$\langle f_0, \int_0^1 R_{0(t',\boldsymbol{x}_i(t',\cdot))}\phi_k(t')dt'\rangle_{\mathcal{H}} = \sum_{j=1}^{2} a_{ijk}d_j,$$
$$\langle f_{1n}, \int_0^1 R^*_{(t',\boldsymbol{x}_j(t',\cdot))}\phi_l(t')dt'\rangle_{\mathcal{H}^*} = \sum_{i=1}^{n}\sum_{k=1}^{n} b_{ikjl}c_{ik},$$

where $R_{0(t',\boldsymbol{x}_i(t',\cdot))}(t,\boldsymbol{x}(t,\cdot)) = R_0((t,\boldsymbol{x}(t,\cdot)),(t',\boldsymbol{x}_i(t',\cdot)))$, $a_{ijk} = \int_0^1 \varphi_j(t)\phi_k(t)dt$, and $b_{ikjl} = \int_0^1 \xi_{ik}(t,\boldsymbol{x}_j(t,\cdot))\phi_l(t)dt$.

Let $b^v_{jlik} = \int_0^1\int_0^1 R_v((t,\boldsymbol{x}_j(t,\cdot)),(t',\boldsymbol{x}_i(t',\cdot)))\phi_k(t') \times \phi_l(t)dtdt'$. Since $R^* = \sum_{v=1}^4 \theta_v R_v$, then $b_{jlik} = \sum_{v=1}^4 \theta_v b^v_{jlik}$ and $\boldsymbol{\Sigma} = \sum_{v=1}^4 \theta_v \boldsymbol{\Sigma}_v$, where the $(j+(l-1)n, i+(k-1)n)$th element of $\boldsymbol{\Sigma}_v$ is b^v_{jlik}. Let $\boldsymbol{Y_k} = (v_{1k},\cdots,v_{nk})^T$, $\boldsymbol{Y} = ({\boldsymbol{Y_1}}^T,\cdots,{\boldsymbol{Y_n}}^T)^T$, $\boldsymbol{c} = (c_{11}, c_{21},\cdots,c_{nn})^T$, $\boldsymbol{d} = (d_1,d_2)^T$, $\boldsymbol{\Sigma}$ be an $n^2\times n^2$ matrix with b_{jlik} as the $(j+(l-1)n, i+(k-1)n)$ element, and $\boldsymbol{T}$ be a $n^2 \times 2$ matrix with a_{ijk} as the $(i+(k-1)n, j)$ element. Then the PLS (14) reduces to

$$(15)\qquad \frac{1}{n}||\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{d} - \boldsymbol{\Sigma}\boldsymbol{c}||^2 + \lambda\boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c}.$$

The minimization problem (15) can be solved by the *dsidr* function in the assist [25] R package. To save computational time, we first use a backfitting procedure to compute the smoothing parameters and then compute solutions to (15) with θ_ν for $\nu = 1,\cdots,4$ being fixed at these estimates. The computational procedure is presented as the NCRM algorithm below. The algorithm consists of two steps: (1) estimate parameters θ_r, and (2) calculate $\boldsymbol{c}$ and $\boldsymbol{d}$ using the kernel matrix $\boldsymbol{\Sigma}$. The backfitting procedure in step (1) is repeated until convergence or when the iteration number reaches a pre-specified bound. Step (2) has closed-form solutions for $\boldsymbol{c}$ and $\boldsymbol{d}$ with a fixed λ. We use the generalized cross-validation (GCV) method to select the smoothing parameter λ.

4. STATISTICAL PROPERTIES

In this section, we study the convergence rate of the penalized least squares estimate $\hat{f}$. We assume that $\mathcal{X}$ and $\mathcal{Y}$ are complete measurable spaces. Let P be a probability measure on $\mathcal{X}^q \times L^2(\mathcal{T})$ and μ be the marginal measure on $\mathcal{X}^q$ induced from P. Denote $\mathcal{M} = \mathcal{T} \times \mathcal{X}^q$.

Define a loss function

$$l(f;\boldsymbol{x},y) = \int_0^1 (y(t) - f(t,\boldsymbol{x}(t,\cdot)))^2dt,$$

Algorithm 1 NCRM Algorithm

Set $\delta > 0$ and initial values of coefficients $\boldsymbol{d} = \boldsymbol{d}^0$, $\boldsymbol{c}_v = \boldsymbol{c}_v^0$, $\theta_v = \theta_v^0$, $v = 1,\cdots,4$, $k.iter = 0$ and a bound of iteration number, $k.max$.
repeat
  Calculate $\tilde{\boldsymbol{F}} = \boldsymbol{T}\boldsymbol{d} + \sum_{v=1}^4 \theta_v\boldsymbol{\Sigma}_v\boldsymbol{c}_v$
  Calculate $\tilde{\boldsymbol{Y}} = \boldsymbol{Y} - \sum_{v=1}^4 \theta_v\boldsymbol{\Sigma}_v\boldsymbol{c}_v$
  Update $\boldsymbol{d}$ via the least-squares method with respect to $\tilde{\boldsymbol{Y}}$ and $\boldsymbol{T}$
  For $r = 1,\cdots,4$
  Calculate $\tilde{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{T}\boldsymbol{d} - \sum_{v\neq r}\theta_v\boldsymbol{\Sigma}_v\boldsymbol{c}_v$
  Minimize $\frac{1}{n}||\tilde{\boldsymbol{Y}} - \boldsymbol{\Sigma}_r\boldsymbol{c}_r||^2 + \lambda_r\boldsymbol{c}_r^T\boldsymbol{\Sigma}_r\boldsymbol{c}_r$ with respect to $\boldsymbol{c}_r$
  Update $\theta_r = 1/\lambda_r$
  Calculate $\boldsymbol{F} = \boldsymbol{T}\boldsymbol{d} + \sum_{v=1}^4 \theta_v\boldsymbol{\Sigma}_v\boldsymbol{c}_v$
  $k.iter = k.iter + 1$
until the average of absolute difference between $\boldsymbol{F}$ and $\tilde{\boldsymbol{F}}$ is less than δ or $k.iter > k.max$, and get θ_v, $v = 1,\cdots,4$
Calculate the kernel matrix $\boldsymbol{\Sigma}$ using the parameters θ_v
Minimize $\frac{1}{n}||\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{d} - \boldsymbol{\Sigma}\boldsymbol{c}||^2 + \lambda\boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c}$ with respect to $\boldsymbol{c}$ and $\boldsymbol{d}$, where λ is estimated by the GCV method.
return $\boldsymbol{c}$ and $\boldsymbol{d}$

where $y(t) \in \mathcal{Y}$ and $\boldsymbol{x} \in \mathcal{X}^q$. The corresponding L-risk function [19]

$$R_{l,P}(f) = E_P[l(f;\boldsymbol{x},y)].$$

Let D be an empirical measure based on the observed data $\mathcal{D}$, and empirical L-risk function [19]

$$R_{l,D}(f) = \frac{1}{n}\sum_{i=1}^{n} l(f;\boldsymbol{x}_i,y_i).$$

Let $f^* = \arg\min_{f\in\mathcal{H}} R_{l,P}(f)$, $R^*_{l,P,\mathcal{H}} = R_{l,P}(f^*)$, and

$$f_{P,\lambda} = \arg\min_{f\in\mathcal{H}} \left\{R_{l,P}(f) + \lambda||f||^2_{\mathcal{H}^*}\right\}.$$

Obviously, $\hat{f} = f_{D,\lambda}$. For a sequence of random variables A_n and a sequence of constants a_n, the notation $A_n = O_p(a_n)$ means that A_n/a_n is bounded in probability. We state convergence rates in the following theorems and show proofs in the Appendix.

Theorem 3. *Assume that $f : \mathcal{M} \to \mathbb{R}$ is measurable for any $f \in \mathcal{H}$, $\mathcal{M}$ is a complete measurable space, and $|P|_2 = \int_{\mathcal{X}^q\times L^2(\mathcal{T})} ||y(t)||_2^2 dP(\boldsymbol{x},y) < \infty$. When $\lambda \to 0$ and $\lambda^6 n \to \infty$ as $n \to \infty$, we have*

$$|R_{l,P}(\hat{f}) - R^*_{l,P,\mathcal{H}}| = O_p(\lambda),$$
$$||\hat{f} - f_{P,\lambda}||_{\mathcal{H}} = O_p(\lambda^{3/2}),$$
$$||f_{P,\lambda} - f^*||_{\mathcal{H}} = o_p(1).$$

Theorem 3 shows that the function estimate $\hat{f}$ is L-risk consistent with a convergence rate of λ when $\lambda \to 0$ and $\lambda^6 n \to \infty$ as $n \to \infty$. Furthermore, $\hat{f}$ is a consistent estimation of f^* under the norm of the RKHS $\mathcal{H}$.

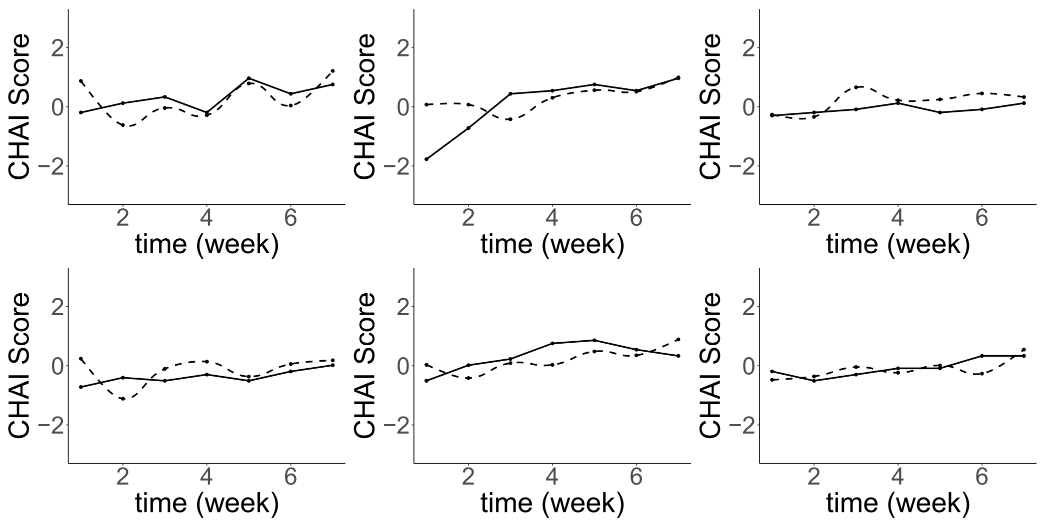

Figure 1. Plots of CAHAI scores $y_i(t)$ (solid lines) and fitted functions $\hat{f}(t, \boldsymbol{x}_i(t, \cdot))$ (dotted lines) from six patients.

5. APPLICATIONS

5.1 Stroke rehabilitation

The data came from a medical study on 70 stroke survivors, including 34 acute patients who had suffered a stroke less than a month earlier and 36 chronic patients who had suffered a stroke more than six months ago. We use data from 34 acute patients to illustrate the proposed method. The response variable is the CAHAI score, a patient's daily life dependency level or the upper limbs function impairment level. We normalize the CAHAI scores by subtracting the mean and dividing by the standard deviation. Each patient was assessed up to 7 times over 3 months, with a total of 173 observations obtained from the acute patients. As [2], we consider 3 bivariate functional variables as functional covariates: forward circle movement of the paretic limb from x axis ($x_{i1} = LA05.lx$), sawing movement of the paretic limb from y axis ($x_{i2} = LA09.ly$), and orientation movement of the non-paretic limb from x axis ($x_{i3} = LA28.rqx$).

We fit model (1) with $n = 34$ and $\boldsymbol{x}_i(t, \cdot) = (x_{i1}(t, \cdot), x_{i2}(t, \cdot), x_{i3}(t, \cdot))$. We extend the cosine diagnostic to assess the contributions of the components in the SS ANOVA decomposition [5] as follows. The estimated function can be represented as

$$\hat{f}(t, \boldsymbol{x}(t, \cdot)) = \hat{\mu} + \hat{f}_1(t) + \hat{f}_2(\boldsymbol{x}(t, \cdot)) + \hat{f}_3(t, \boldsymbol{x}(t, \cdot)),$$

where $\hat{\mu} \in \{1\}$ is the estimate of the constant, $\hat{f}_1(t) \in \{t\} \oplus \mathcal{H}_1$ and $\hat{f}_2(\boldsymbol{x}(t, \cdot)) \in \mathcal{H}_2$ are respectively the estimated main effects of t and $\boldsymbol{x}(t, \cdot)$, and $\hat{f}_3(t, \boldsymbol{x}(t, \cdot)) \in \mathcal{H}_3 \oplus \mathcal{H}_4$ is the estimated interaction between t and $\boldsymbol{x}(t, \cdot)$. Let $\hat{\boldsymbol{f}}^* = (\hat{f}^*(t, \boldsymbol{x}_1(t, \cdot)), \cdots, \hat{f}^*(t, \boldsymbol{x}_n(t, \cdot)))^T$, $\hat{\boldsymbol{f}}_1^* = (\hat{f}_1^*(t), \cdots, \hat{f}_1^*(t))^T$, $\hat{\boldsymbol{f}}_2^* = (\hat{f}_2^*(\boldsymbol{x}_1(t, \cdot)), \cdots, \hat{f}_2^*(\boldsymbol{x}_n(t, \cdot)))^T$ and $\hat{\boldsymbol{f}}_3^* = (\hat{f}_3^*(t, \boldsymbol{x}_1(t, \cdot)), \cdots, \hat{f}_3^*(t, \boldsymbol{x}_n(t, \cdot)))^T$, where $a^*(t) = a(t) - \int_0^1 a(t)dt$. We have $\hat{\boldsymbol{f}}^* = \hat{\boldsymbol{f}}_1^* + \hat{\boldsymbol{f}}_2^* + \hat{\boldsymbol{f}}_3^*$. As in [5], we compute the quantities $\pi_k = \langle \hat{\boldsymbol{f}}_k^*, \hat{\boldsymbol{f}}^* \rangle / ||\hat{\boldsymbol{f}}^*||^2$ for $k = 1, 2, 3$, where

$$\langle \hat{\boldsymbol{f}}_k^*, \hat{\boldsymbol{f}}^* \rangle := \sum_{i=1}^{n} \int_0^1 \hat{f}^*(t, \boldsymbol{x}_i(t, \cdot)) \hat{f}_k^*(t, \boldsymbol{x}_i(t, \cdot)) dt,$$

to assess the contributions of the main effects and interaction. We have $\pi_1 = -0.73$, $\pi_2 = 0.32$, and $\pi_3 = 1.41$, indicating a large contribution from the interaction. This suggests that the association between CAHAI score and movements changes over time.

Figure 1 shows the scores $y_i(t)$ and the fitted functions $\hat{f}(t, \boldsymbol{x}_i(t, \cdot))$ for 6 patients. The overall trends of the fitted functions and the observed scores are consistent. The upward trend over time indicates that the recovery exercises are effective.

We evaluate the model's prediction performance using a 10-fold cross-validation:

$$RPE = \frac{1}{10} \sum_{j=1}^{10} \sqrt{\frac{1}{n_j} \sum_{i \in jth fold} ||Y_i(t) - \hat{Y}_i^{(-j)}(t)||_2^2}$$

where $\hat{Y}_i^{(-j)}(t)$ is the predicted value of $Y_i(t)$ based on a fitted model using the data excluding the jth fold. The RPE for stoke data is 1.170.

5.2 Human fertility and mortality

The relationship between fertility and mortality rates has been a long-standing demographic interest [1]. Many articles in demography mentioned the relationship between fertility and mortality but were mainly limited to qualitative discussions [20, 4, 13, 9]. This section explores the quantitative

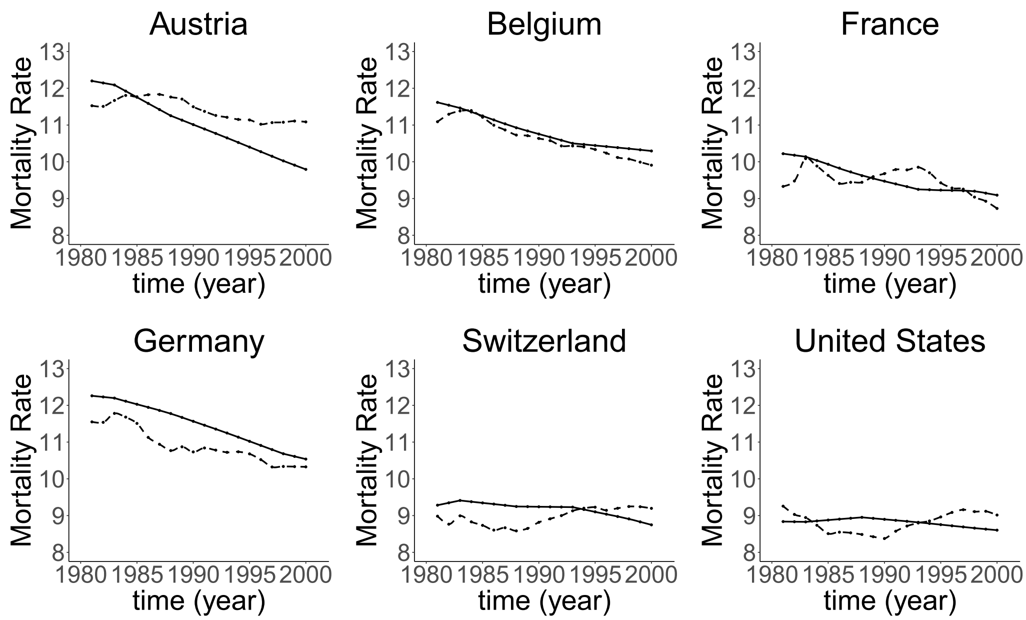

Figure 2. Plots of the mortality rates $y_i(t)$ (solid lines) and fitted functions $\hat{f}(t, x_i(t,\cdot))$ (dotted lines) in unit of ‰ from six countries whose names are on the title.

relationship between age-specific fertility and mortality using the proposed model.

We download fertility data from the Human Fertility Database website (www.humanfertility.org) that contains annual age-specific fertility rate (ASFR) from 22 countries. We download mortality rates at www.macrotrends.net/. We select 16 countries that appeared in both databases. We consider the time interval from 1981 to 2000 and age interval from 15-year-old to 50-year-old to avoid zero ASFR values.

For country i, we consider mortality rate in calendar year t as a functional response variable $y_i(t)$, and ASFR as a bivariate functional covariate $x_i(t, s)$ that depends on women's age in years (s) and calendar year (t). We fit model (1) with $n = 16$. For cosine diagnostic we have $\pi_1 = 0.08$, $\pi_2 = 0.71$, and $\pi_3 = 0.21$. The largest contribution comes from the main effect of x, indicating differences in ASFR between countries account for more variations in the mortality rate than time. Figure 2 shows observed and predicted mortality rates for six countries. Mortality rates between 1980 and 2020 keep declining with different rates. Since the number of countries $n = 16$ is small, 5-fold cross validation is used to test the prediction performance, and the RPE is 1.449.

6. SIMULATION STUDIES

Performance of the proposed method is evaluated by two numerical simulation studies in this section.

In the first study, we generate data from model (1) with a factorial design of two choices of sample size n, two approaches for generating scalar functions $x_i(t,\cdot)$, three choices of function $f(t, x(t,\cdot))$, and three choices of random errors. Specifically, we consider three sample sizes: $n = 10$, $n = 20$, and $n = 40$. We generate scalar functions $x_i(t,\cdot) = \exp(x_i^*(t,\cdot))/(1 + \exp(x_i^*(t,\cdot)))$ where $x_i^*(t,\cdot)$ are n realizations of a Gaussian process with mean function $\mu(t) = t$ and RBF kernel $k_g(s_1, s_2) = \exp\{\frac{-(s_1-s_2)^2}{2}\}$ and rational quadratic kernel $k_l(s_1, s_2) = 1 - \frac{(s1-s2)^2}{(s1-s2)^2+1}$, respectively. We consider three cases of the function $f(t, x(t,\cdot))$ as follows:

$$F_1 : f(t, x(t,\cdot)) = t^2 + \int_0^1 x^3(t,s)ds,$$

$$F_2 : f(t, x(t,\cdot)) = t^2 \int_0^1 x^3(t,s)ds,$$

$$F_3 : f(t, x(t,\cdot)) = 0.25 + 0.5t + 0.25 \int_0^1 x^3(t,s)ds + 0.5t \int_0^1 x^3(t,s)ds,$$

where F_1 is an additive model that consists of the main effects of t and x, F_2 consists of a nonparametric-nonparametric interaction of t and x, and F_3 consists of the main effect of t and x and the linear-nonparametric interaction of t and x. We consider three scenarios of random errors $\epsilon_i(t)$: iid $N(0, 0.1^2)$, iid $N(0, 0.2^2)$, and iid realizations of the Gaussian process with mean zero and kernel function $k_t(t_i, t_j) = 0.1\min(t_i, t_j)$. In practice, we can not observe the whole $x(t, s)$, instead observe values of the curve at grid points. This paper considers two scenarios for the choices of t and s: both t and s take 10 equally spaced points in $[0, 1]$, and t takes 20 equally spaced grid points in $[0, 1]$ and s takes 50 equally spaced grid points in $[0, 1]$. All simulation are repeated 200 times.

Table 1. *Average RMSEs and standard deviations in parentheses when both t and s take 10 equally spaced points in $[0,1]$*

$x(t,\cdot)$	Function	n	$\epsilon(t) \sim N(0,0.1^2)$	$\epsilon(t) \sim N(0,0.2^2)$	$\epsilon(t) \sim GP(0,k_t)$
k_g	F_1	10	0.060(0.015)	0.098(0.028)	0.110(0.054)
		20	0.040(0.008)	0.070(0.013)	0.074(0.021)
		40	0.030(0.005)	0.052(0.010)	0.052(0.013)
	F_2	10	0.048(0.014)	0.080(0.030)	0.094(0.043)
		20	0.036(0.007)	0.056(0.009)	0.066(0.022)
		40	0.025(0.004)	0.043(0.008)	0.046(0.013)
	F_3	10	0.049(0.017)	0.083(0.027)	0.094(0.033)
		20	0.035(0.008)	0.059(0.016)	0.063(0.020)
		40	0.026(0.005)	0.044(0.010)	0.048(0.015)
k_l	F_1	10	0.060(0.018)	0.102(0.037)	0.110(0.038)
		20	0.042(0.007)	0.072(0.016)	0.074(0.019)
		40	0.031(0.006)	0.052(0.008)	0.055(0.013)
	F_2	10	0.052(0.020)	0.086(0.036)	0.090(0.034)
		20	0.035(0.009)	0.055(0.012)	0.067(0.028)
		40	0.026(0.005)	0.046(0.011)	0.047(0.013)
	F_3	10	0.047(0.012)	0.084(0.031)	0.092(0.033)
		20	0.036(0.007)	0.056(0.013)	0.065(0.022)
		40	0.025(0.006)	0.044(0.008)	0.052(0.020)

Table 2. *Average RMSEs and standard deviations in parentheses when t takes 20 equally spaced points in $[0,1]$ and s takes 50 equally spaced points in $[0,1]$*

$x(t,\cdot)$	Function	n	$\epsilon(t) \sim N(0,0.1^2)$	$\epsilon(t) \sim N(0,0.2^2)$	$\epsilon(t) \sim GP(0,k_t)$
k_g	F_1	10	0.053(0.012)	0.089(0.022)	0.122(0.045)
		20	0.036(0.006)	0.062(0.012)	0.089(0.024)
		40	0.027(0.003)	0.046(0.007)	0.069(0.017)
	F_2	10	0.046(0.014)	0.071(0.025)	0.097(0.033)
		20	0.030(0.006)	0.052(0.014)	0.072(0.024)
		40	0.022(0.004)	0.040(0.008)	0.057(0.016)
	F_3	10	0.044(0.012)	0.070(0.021)	0.100(0.037)
		20	0.031(0.007)	0.051(0.012)	0.077(0.024)
		40	0.023(0.004)	0.039(0.008)	0.058(0.017)
k_l	F_1	10	0.054(0.010)	0.096(0.031)	0.124(0.042)
		20	0.037(0.006)	0.064(0.011)	0.092(0.024)
		40	0.028(0.004)	0.046(0.008)	0.064(0.013)
	F_2	10	0.047(0.014)	0.070(0.020)	0.099(0.034)
		20	0.033(0.008)	0.051(0.013)	0.068(0.021)
		40	0.022(0.003)	0.039(0.007)	0.053(0.015)
	F_3	10	0.046(0.013)	0.070(0.022)	0.098(0.037)
		20	0.030(0.005)	0.053(0.014)	0.073(0.024)
		40	0.024(0.004)	0.041(0.009)	0.057(0.015)

We generate n samples $\{y_i(t), x_i(t,\cdot) : \ i = 1,\cdots,n\}$ as training data, and $n_t = 50$ samples $\{\tilde{y}_i(t), \tilde{x}_i(t,\cdot) : \ i = 1,\cdots,n_t\}$ as test data. Root Mean square error RMSE on the test data

$$RMSE = \sqrt{\frac{1}{n_t}\sum_{i=1}^{n_t} ||f(t,\tilde{x}_i(t,\cdot)) - \hat{f}(t,\tilde{x}_i(t,\cdot))||_2^2}$$

is used to measure prediction from the proposed model, where $||\cdot||_2$ is the norm of $L^2(\mathcal{T})$. Table 1 and Table 2 presents the averages of RMSE and standard deviations in parentheses. We see that the proposed estimation method performs well.

In the second study, we consider simulations settings similar to the two real data examples in Section 5. For the stroke data, input variables are from 3 movements: forward circle movement of paretic limb, sawing movement of paretic limb and orientation movement of non-paretic limb, denoted by $\boldsymbol{x}_i(t,\cdot) = (x_{i1}(t,\cdot), x_{i2}(t,\cdot), x_{i3}(t,\cdot)), \ i = 1,\cdots,n,$ where $n = 34$. Based on these 3 movements, we construct a kernel function,

$$R^* = 0.5R_1 + 0.1R_2 + 0.1R_3 + 0.5R_4, \qquad (16)$$

Table 3. Average RMSEs and standard deviations in parentheses

Data	$\epsilon(t) \sim N(0, 0.1^2)$	$\epsilon(t) \sim N(0, 0.2^2)$	$\epsilon(t) \sim GP(0, k_t)$
stroke data	0.158(0.067)	0.306(0.111)	0.366(0.157)
ASFR data	0.584(0.366)	0.589(0.365)	0.613(0.365)

where R_1, R_2, R_3 and R_4 are defined in (9). With this kernel, we generate data from the following model

(17)

$$
\begin{aligned}
y_i(t) =& \sum_{j=1}^{2} d_j \varphi_j(t) \\
&+ \sum_{j=1}^{n}\sum_{k=1}^{n} c_{jk} \int_0^1 R^*_{((t,\boldsymbol{x}_i(t,\cdot))}(t', \boldsymbol{x}_j(t',\cdot)))\phi_k(t')dt' \\
&+ \epsilon_i(t),
\end{aligned}
$$

where $i = 1, \cdots, n$, $\{\phi_k(t), k = 1, \cdots, n\}$ are empirical functional principal components of the CHAI scores from the stroke data, $\boldsymbol{d} = (d_1, d_2)$ and c_{jk} are the estimated $\boldsymbol{d}$ and c_{jk} from the stroke data analysis.

We consider three random errors: $\epsilon_i(t)$: iid $N(0, 0.1^2)$, iid $N(0, 0.2^2)$, and iid realizations of the Gaussian process with mean zero and kernel function $k_t(t_i, t_j) = 0.1 \min(t_i, t_j)$. We repeated the same process for the ASFR data. In each simulation, we randomly select 60% of simulated data to build the model and use it to predict the remaining ones. All simulation are repeated 200 times.

Table 3 presents the averages of RMSEs and standard deviations in parentheses. Again, the proposed method performs well.

APPENDIX

Proof of Theorem 1. For simplicity of notation, we show proof for the univariate case since the proof for the multivariate case is similar. Denote $R_2^{(2)}(u, v) = \exp\left\{-\frac{||u-v||^2}{2}\right\}$ where $u \in \mathcal{X}$ and $||u|| = \int_0^1 u^2(s)ds$.

(i) The symmetry of $R_2^{(2)}$ is obvious. We only need to prove that $R_2^{(2)}$ is positive definite. A kernel $\psi(u, v)$ is a conditional negative definite kernel if $\psi(u, u) = 0$, $\psi(u, v) = \psi(v, u)$, and $\forall\ n \geq 2$, $a_1, \cdots, a_n \in \mathbb{R}$, $\sum_{i=1}^n a_i = 0$ and $u_1, \cdots, u_n \in \mathcal{X}$, we have $\sum_{i=1}^n a_i a_j \psi(u_i, u_j) \leq 0$. According to Schoenberg's theorem [16], $\exp\{-t\psi(u, v)\}(t > 0)$ is a positive definite kernel when $\psi(u, v)$ is a conditional negative definite kernel.

For $R_2^{(2)}$, we have

$$
\begin{aligned}
\psi(u, v) &= ||u - v||^2 \\
&= ||u||^2 + ||v||^2 - 2\langle u, v\rangle,
\end{aligned}
$$

where $\langle u, v\rangle = \int_0^1 u(s)v(s)ds$. For $\forall\ n \geq 2$, $a_1, ..., a_n \in \mathbb{R}$, $\sum_{i=1}^n a_i = 0$ and $u_1, \cdots, u_n \in \mathcal{X}$, we have

$$
\begin{aligned}
\sum_{i,j=1}^{n} a_i a_j \psi(u_i, u_j) &= \sum_{i,j=1}^{n} a_i a_j (||u_i||^2 + ||u_j||^2 - 2\langle u_i, u_j\rangle)) \\
&= -2\Big\langle \sum_{i=1}^{n} a_i u_i, \sum_{j=1}^{n} a_j u_j \Big\rangle \leq 0.
\end{aligned}
$$

Therefore, $R_2^{(2)}$ is a positive definite symmetric kernel.

(ii) Denote the RKHS with RK $R_2^{(2)}$ as $\mathcal{H}_{R_2^{(2)}}$. We prove by contradiction that $\mathcal{H}_{R_2^{(2)}}$ does not contain any non-zero constants. Given the properties of Hilbert Spaces, we only need to prove $1 \notin \mathcal{H}_{R_2^{(2)}}$. We assume that $c(u) \equiv 1 \in \mathcal{H}_{R_2^{(2)}}$. For $\{a_1, \cdots, a_n\} \in \mathbb{R}$ and $\forall\ \{u_1, \cdots, u_n\} \in \mathcal{X}$, let $g(x) := \sum_{i=1}^n a_i R_2^{(2)}(u_i, u)$. Note that $c(u_i) = \langle c(\cdot), R_2^{(2)}(u_i, \cdot)\rangle = 1$. Then

$$
\begin{aligned}
\sum_{i,j=1}^{n} a_i a_j c(u_i) c(u_j) &= \left|\sum_{i=1}^{n} a_i c(u_i)\right|^2 \\
&= \left|\sum_{i=1}^{n} a_i \langle c(\cdot), R_2^{(2)}(u_i, \cdot)\rangle\right|^2 \\
&= |\langle c(\cdot), g(\cdot)\rangle|^2 \leq ||g(\cdot)^2|| \\
&= \Big\langle \sum_{i=1}^{n} a_i R_2^{(2)}(u_i, \cdot), \sum_{j=1}^{n} a_j R_2^{(2)}(u_j, \cdot) \Big\rangle \\
&= \sum_{i,j=1} a_i a_j R_2^{(2)}(u_i, u_j).
\end{aligned}
$$

Since the choices of $\{a_1, \cdots, a_n\}$ and $\{u_1, .., u_n\}$ are arbitrary and $R_2^{(2)}(u, u) = \exp\{-\frac{||u-u||^2}{2}\} = e^0 = 1$, taking $n = 2$ and $a_1 = a_2 = 1$, then we have $R_2^{(2)}(u, v) \geq 1$. But we know that $R_2^{(2)}(u, v) \leq 1$, so that $R_2^{(2)}(u, v) \equiv 1$. We derive the contradiction. By Merce theorem, $R_2^{(2)}(u, v) = \sum_{i=1}^{\infty} \lambda e_i(u) e_i(v)$, then $\{\sqrt{\lambda} e_i : i = 1, \cdots, \infty\}$ is a set of orthogonal basis of $\mathcal{H}_{R_2^{(2)}}$ so that $\mathcal{H}_{R_2^{(2)}}$ is a separable space. □

Proof of Theorem 3. Since

$$
\begin{aligned}
&|R_{l,P}(\hat{f}) - R^*_{l,P,\mathcal{H}}| \\
&\leq |R_{l,P}(\hat{f}) - R_{l,P}(f_{P,\lambda})| + |R_{l,P}(f_{P,\lambda}) - R^*_{l,P,\mathcal{H}}|,
\end{aligned}
$$

we need to calculate the convergence rate of $|R_{l,P}(\hat{f}) - R_{l,P}(f_{P,\lambda})|$ and $|R_{l,P}(f_{P,\lambda}) - R^*_{l,P,\mathcal{H}}|$ separately.

Because $|P|_2$ and $||R_{\mathcal{H}}||_\infty$ are bounded, without losing generality, we assume that $q = 1$, $|P|_2 = 1$, and $||R_{\mathcal{H}}||_\infty = 1$, where $R_{\mathcal{H}}$ is the RK of $\mathcal{H}$ and then $R_{l,P}(0) \leq |P|_2 = 1$ We denote $\omega(z) = ||z(\cdot)||_2^2$.

$$
\begin{aligned}
&|R_{l,P}(\hat{f}) - R_{l,P}(f_{P,\lambda})| \\
&\leq \int_{X \times Y} |\omega(y - \hat{f}) - \omega(y - f_{P,\lambda})| dP(x, y).
\end{aligned}
$$

Let $r(x,y) = ||y(t)||_2 + ||\hat{f}||_2 + ||f_{P,\lambda}||_2 + 1$ and $B(0,2r) = \{ z \mid ||z||_2 < 2r\}$. Let

$$z = z_1 + \frac{r}{||z_1 - z_2||_2}(z_1 - z_2), \quad \forall z_1, z_2 \in B(0,r),$$

then

$$\begin{aligned}\omega(z_1) - \omega(z_2) &\leq \frac{||z_1 - z_2||_2}{||z_1 - z_2||_2 + r}(\omega(z) - \omega(z_2)) \\ &\leq \frac{2||\omega_{|B(0,2r)}||_\infty}{r}||z_1 - z_2||_2.\end{aligned}$$

Therefore,

$$\begin{aligned}&|R_{l,P}(\hat{f}) - R_{l,P}(f_{P,\lambda})| \\ \leq &\int_{X\times Y} \frac{2||\omega_{|B(0,2r)}||_\infty}{r(x,y)}||\hat{f} - f_{P,\lambda}||_2 dP(x,y) \\ \leq &C\int_{X\times Y} \frac{|2r(x,y)|^2 + 1}{r(x,y)}||\hat{f} - f_{P,\lambda}||_2 dP(x,y) \\ \leq &C\int_{X\times Y} |r(x,y)| \; ||\hat{f} - f_{P,\lambda}||_2 dP(x,y) \\ \leq &C\int_{X\times Y} |r(x,y)| dP(x,y) ||\hat{f} - f_{P,\lambda}||_\infty \\ \leq &C\int_{X\times Y} (||y(t)||_2 + ||\hat{f}||_2 + ||f_{P,\lambda}||_2 + 1) dP(x,y) \\ &\times ||\hat{f} - f_{P,\lambda}||_\infty \\ \leq &C(2 + ||\hat{f}||_\infty + ||f_{P,\lambda}||_\infty)||\hat{f} - f_{P,\lambda}||_\infty,\end{aligned}$$

where C is an indeterminate constant depending on l.

We know that

$$\lambda ||f_{P,\lambda}||^2_{\mathcal{H}^*} \leq \arg\inf_{f\in\mathcal{H}} R_{l,P}(f) + \lambda||f||^2_{\mathcal{H}^*} \leq R_{l,P}(0) \leq 1,$$

hence $||f_{P,\lambda}||_\infty \leq ||R_{\mathcal{H}}||_\infty ||f_{P,\lambda}||_{\mathcal{H}} \leq C\lambda^{-\frac{1}{2}}$ and $||\hat{f}||_\infty \leq ||f_{P,\lambda}||_\infty + ||f_{P,\lambda} - \hat{f}||_\infty \leq C\lambda^{-\frac{1}{2}} + 1$ when $||f_{P,\lambda} - \hat{f}||_{\mathcal{H}} \leq 1$. Therefore,

$$|R_{l,P}(\hat{f}) - R_{l,P}(f_{P,\lambda})| \leq C\lambda^{-\frac{1}{2}}||f_{P,\lambda} - \hat{f}||_{\mathcal{H}}.$$

Meanwhile, according to *Corollary 5.18* in [19], we know that $|R_{l,P}(f_{P,\lambda}) - R^*_{l,P,\mathcal{H}}| \leq c\lambda$.

Now for the sake of illustration, we denote $\tilde{l}(t,x,y,f) := (y(t) - f(t,x(t,\cdot)))^2$, and pick special $\tilde{P}$, $\tilde{D}$ on $\mathcal{M}\times Y$ such that

$$\begin{aligned}R_{\tilde{l},\tilde{P}} &= \int_{\mathcal{M}\times Y} \tilde{l}(t,x,y,f)d\tilde{P} \\ &= \int_{\mathcal{X}\times Y}\int_0^1 (y(t) - f(t,x(t,\cdot)))^2 dtdP = R_{l,P}, \\ R_{\tilde{l},\tilde{D}} &= \int_{\mathcal{M}\times Y} \tilde{l}(t,x,y,f)d\tilde{D} \\ &= \sum_{i=1}^n \int_0^1 (y_i(t) - f(t,x_i(t,\cdot)))^2 dt = R_{l,D}.\end{aligned}$$

Then $f_{P,\lambda} = f_{\tilde{P},\lambda}$ and $\hat{f} = f_{\tilde{D},\lambda}$. Similar to the proofs of *Theorem 5.8*, *Theorem 5.9* and *Corollary 5.11* in [19], for all P_0, when $\tilde{l}$ is P_0-integrable, there exist $h : \mathcal{M}\times Y \to \mathbb{R}$ satisfies $||h||_{L^2(\tilde{P})} \leq C_{\tilde{l}}\lambda^{-\frac{1}{2}}$ such that

$$||f_{P_0,\lambda} - f_{\tilde{P},\lambda}||_{\mathcal{H}} \leq C\lambda^{-1}||E_{\tilde{P}}h\Phi - E_{P_0}h\Phi||_{\mathcal{H}},$$

where $\Phi((t,x(t,\cdot))) = R_{\mathcal{H}}(\cdot,(t,x(t,\cdot)))$ is a canonical map. Let $P_0 = \tilde{D}$, then we have

$$||\hat{f} - f_{P,\lambda}||_{\mathcal{H}} \leq C\lambda^{-1}||E_{\tilde{P}}h\Phi - E_{\tilde{D}}h\Phi||_{\mathcal{H}},$$

Hence, from *Lemma 9.2* in [19] and, we get

$$\begin{aligned}&P(|R_{l,P}(\hat{f}) - R^*_{l,P,\mathcal{H}}| \geq \epsilon) \\ \leq &P(C\lambda^{-\frac{3}{2}}||E_{\tilde{P}}h_n\Phi - E_{\tilde{D}}h_n\Phi||_{\mathcal{H}} + c\lambda > \epsilon) \\ \leq &O(n^{-1}\lambda^{-6}),\end{aligned}$$

with $\epsilon = O(\lambda)$ and

$$\begin{aligned}&P(||\hat{f} - f_{P,\lambda}||_{\mathcal{H}} \geq \epsilon) \\ \leq &P(||E_{\tilde{P}}h_n\Phi - E_{\tilde{D}}h_n\Phi||_{\mathcal{H}} \geq \lambda\epsilon) \\ \leq &O(n^{-1}\lambda^{-6})\end{aligned}$$

with $\epsilon = O(\lambda^{\frac{3}{2}})$.

From *Theorem 5.17* the continuity in the regularization parameter in [19], we know that $||f_{P,\lambda} - f^*|| = o_p(1) \quad as\ n \to \infty$. The conclusion is confirmed. □

Received 18 November 2022

REFERENCES

[1] Chen, K., Delicado, P. and Müller, H.-G. (2017). Modelling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 177–196. MR3597969

[2] Cheng, Y., Shi, J. Q. and Eyre, J. (2020). Nonlinear mixed-effects scalar-on-function models and variable selection. *Statistics and Computing* **30** 129–140. MR4057475

[3] de la Fuente, M. O. (2019). Advances in functional regression and classification models PhD thesis, Universidade de Santiago de Compostela.

[4] Eckstein, Z., Schultz, T. P. and Wolpin, K. I. (1984). Short-run fluctuations in fertility and mortality in pre-industrial Sweden. *European Economic Review* **26** 295–317.

[5] Gu, C. (2013). *Smoothing Spline ANOVA Models (2nd Ed)*. Springer. MR3025869

[6] Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons. MR3379106

[7] Kim, J. S., Maity, A. and Staicu, A.-M. (2018). Additive non-linear functional concurrent model. *Statistics and Its Interface* **11** 669. MR3858523

[8] Kokoszka, P. and Reimherr, M. (2017). *Introduction to functional data analysis*. Chapman and Hall/CRC. MR3793167

[9] Kuningas, M., Altmäe, S., Uitterlinden, A. G., Hofman, A., van Duijn, C. M. and Tiemeier, H. (2011). The relationship between fertility and lifespan in humans. *Age* **33** 615–622.

[10] Ling, N. and Vieu, P. (2018). Nonparametric modelling for functional data: selected survey and tracks for future. *Statistics* **52** 934–949. MR3827427
[11] Maity, A. (2017). Nonparametric functional concurrent regression models. *Wiley Interdisciplinary Reviews: Computational Statistics* **9** e1394. MR3615675
[12] Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application* **2** 321–359.
[13] Palloni, A. (1990). Fertility and mortality decline in Latin America. *The Annals of the American Academy of Political and Social Science* **510** 126–144.
[14] Ramsay, J. O. and Silverman, B. W. (2006). *Functional data Analysis, 2nd ed*. Springer, New York. MR2168993
[15] Scheipl, F., Staicu, A.-M. and Greven, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics* **24** 477–501. MR3357391
[16] Schoenberg, I. J. (1938). Metric spaces and positive definite functions. *Transactions of the American Mathematical Society* **44** 522–536. MR1501980
[17] Serradilla, J., Shi, J., Cheng, Y., Morgan, G., Lambden, C. and Eyre, J. A. (2014). Automatic assessment of upper limb function during play of the action video game, circus challenge: validity and sensitivity to change. In *2014 IEEE 3nd International Conference on Serious Games and Applications for Health(SeGAH)* 1–7.
[18] Shi, J. Q., Cheng, Y., Serradilla, J., Morgan, G., Lambden, C., Ford, G. A., Price, C., Rodgers, H., Cassidy, T., Rochester, L. et al. (2013). Evaluating functional ability of upper limbs after stroke using video game data. In *International Conference on Brain and Health Informatics* 181–192.
[19] Steinwart, I. and Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media. MR2796580
[20] Turpeinen, O. (1979). Fertility and mortality in Finland since 1750. *Population Studies* **33** 101–114.
[21] Wahba, G. (2003). An introduction to reproducing kernel Hilbert spaces and why they are so useful. In *Proceedings of the 13th IFAC Symposium on System Identification(SYSID 2003)* 525–528.
[22] Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge University Press. MR3967104
[23] Wang, B. and Xu, A. (2019). Gaussian process methods for nonparametric functional regression with mixed predictors. *Computational Statistics & Data Analysis* **131** 80–90. MR3906796
[24] Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application* **3** 257–295.
[25] Wang, Y. (2011). *Smoothing splines: methods and applications*. CRC press. MR2814838
[26] Wang, Z., Dong, H., Ma, P. and Wang, Y. (2022). Estimation and model selection for nonparametric function-on-function regression. *Journal of Computational and Graphical Statistics* **31** 1–11. MR4495715
[27] Zhang, X., Park, B. U. and ling Wang, J. (2013). Time-Varying Additive Models for Longitudinal Data. *Journal of the American Statistical Association* **108** 983–998. MR3174678

Yutong Zhai
Department of Statistics and Finance
Management School
University of Science and Technology of China
Hefei
China
E-mail address: zyt1@mail.ustc.edu.cn

Zhanfeng Wang
Department of Statistics and Finance
Management School
University of Science and Technology of China
Hefei
China
E-mail address: zfw@ustc.edu.cn

Yuedong Wang
Department of Statistics and Applied Probability
University of California
Santa Barbara
USA
E-mail address: yuedong@pstat.ucsb.edu