mathematics

MDPI

Article

Representation Theorem and Functional CLT for RKHS-Based Function-on-Function Regressions

Hengzhen Huang [1], Guangni Mo [1], Haiou Li [2] and Hong-Bin Fang [2,*]

[1] College of Mathematics and Statistics, Guangxi Normal University, Guilin 541004, China; hzhuang@mailbox.gxnu.edu.cn (H.H.); guangnimo@stu.gxnu.edu.cn (G.M.)
[2] Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, Washington, DC 20057, USA; hl662@georgetown.edu
* Correspondence: hf183@georgetown.edu

Abstract: We investigate a nonparametric, varying coefficient regression approach for modeling and estimating the regression effects caused by two functionally correlated datasets. Due to modern biomedical technology to measure multiple patient features during a time interval or intermittently at several discrete time points to review underlying biological mechanisms, statistical models that do not properly incorporate interventions and their dynamic responses may lead to biased estimates of the intervention effects. We propose a shared parameter change point function-on-function regression model to evaluate the pre- and post-intervention time trends and develop a likelihood-based method for estimating the intervention effects and other parameters. We also propose new methods for estimating and hypothesis testing regression parameters for functional data via reproducing kernel Hilbert space. The estimators of regression parameters are closed-form without computation of the inverse of a large matrix, and hence are less computationally demanding and more applicable. By establishing a representation theorem and a functional central limit theorem, the asymptotic properties of the proposed estimators are obtained, and the corresponding hypothesis tests are proposed. Application and the statistical properties of our method are demonstrated through an immunotherapy clinical trial of advanced myeloma and simulation studies.

Keywords: functional data; hypothesis testing; regression function; reproducing kernel Hilbert space; sparsely observed data

MSC: 62G05; 62G10

check for updates

Citation: Huang, H.; Mo, G.; Li, H.; Fang, H.-B. Representation Theorem and Functional CLT for RKHS-Based Function-on-Function Regressions. *Mathematics* **2022**, *10*, 2507. https://doi.org/10.3390/math10142507

Academic Editors: Huiming Zhang and Ting Yan

Received: 20 June 2022
Accepted: 15 July 2022
Published: 19 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Modern biomedical technology has made it possible to measure multiple patient features during a time interval or intermittently at several discrete time points to review underlying biological mechanisms. Functional data also arise in genetic studies—a massive amount of gene expression data is recorded for each subject and could be treated as a functional curve [1]. Functional data analysis provides distinct features related to the dynamics of cellular responses and activity and other biological processes. Existing methods, such as projection, dimension-reduction, and functional linear regression analysis, are not adapted for such data. Overviews can be found in the book by Horváth and Kokoszka [2] and some recently published papers such as Yuan et al. [3] and Lai et al. [4].

Ramsay and Silverman [5], Clarkson et al. [6], and Ferraty and Vieu [7] introduced some basic tools and widely accepted methods for functional data analysis; Horváth and Kokoszka [2] established some fundamental methods for estimation and hypothesis testing on mean functions and covariance operators of functional data. The topics are broad and the results are in depth. Conventionally, each data curve is assumed to be observed over a dense set of points, often over thousands of points, then smoothing techniques are used to produce continuous curves, and these curves are treated as completely observed functional

data for statistical inference. In contrast with those assumptions, we consider the more practical issues in which the data curves are only observed at some (not dense) time points, and the observed data curves are actually interpolations at those observed points. Of course, a relatively large sample size is needed for sparse observations. The effects of both number of observation points and sample size are also considered in our analysis.

For analyzing longitudinal data, Zeger and Diggle [8] considered a semiparametric regression model of the form, with longitudinal observations

$$Y(t) = X'(t)\beta + \theta(t) + \epsilon(t), \quad t \in \mathcal{T}, \tag{1}$$

where $Y(t)$ is the response variable, $X(t)$ is the $p \times 1$ covariate vector at time t, β is a $p \times 1$ constant vector of unknown regression coefficients, $\theta(t)$ is an unspecified baseline function, $\epsilon(t)$ is a zero-mean stochastic process, and $\mathcal{T}$ represents the observation interval. Under this model, Lin and Ying [9] estimated β via a weighted least squares estimator based on the theory of counting processes; Fan and Li [10] further studied this model using a weighted difference-based estimator and a weighted local linear estimator followed by statistical inference, as discussed in Xue and Zhu [11].

For functional data analysis, the data are often represented by $(y_i, x_i(\cdot))$ $(i = 1, \ldots, n)$, and the model is [12–14]

$$y_i = \int_{\mathcal{T}} \beta(t) x_i(t) dt + \epsilon_i. \tag{2}$$

Some researchers considered the following model [2,5,15]

$$y_i(t) = \int_{\mathcal{T}} \beta(s,t) x_i(s) ds + \epsilon_i(t). \tag{3}$$

To estimate $\beta(\cdot,\cdot)$, assume there are basis $\{\xi_k\}$ and $\{\eta_k\}$, which span the spaces of the $\{x_i(\cdot)\}$ and $\{y_i(\cdot)\}$, respectively. The estimate of $\beta(\cdot,\cdot)$ of the form is given by

$$\hat{\beta}(s,t) = \sum_{i=1}^{k} \sum_{j=1}^{r} b_{ij} \xi_i(s) \eta_j(t),$$

and b_{ij} is estimated by minimizing the residual sum of squares $\sum_{i=1}^{n} ||y_i - \int \hat{\beta}(s,t) x_i(s) ds||^2$. Although the resulting estimator is useful, a representation theorem for such an estimator is hard to obtain, and hence the asymptotic distribution of this approach is not clear. Yao, et al. [15] investigated a functional principle component method for estimation of model (3) and obtained consistent results. Müller and Yao [16] studied a variation of the above model in the conditional expectation format.

The smoothing spline method is popular for curve estimation. The function curves can be estimated at any point, followed by the computation of coefficients. However, the asymptotic property of estimators based on the spline method is tough to handle. For natural polynomial splines, the number of knots is the number of untied observations, which is sometimes redundant and undesirable. B-splines only require a few (the degree of the polynomial plus two) basis functions and are easy to implement [17–19]. Another method is local linear fit [20–22], but the difficulty is in choosing the bandwidth, especially when the observation points are uneven. Therefore, in this paper we employ reproducing kernel Hilbert space (RKHS), a special form of spline method in which the turning point from curve estimation to point estimation Yuan and Cai [12] explored its application on functional linear regression problem, and Lei and Zhang [23] extented it to RKHS-based partially functional linear models. In general, one needs to choose a set of (orthogonal) basis functions and the number of basis for functional estimations, while with RKHS one only needs to determine the kernel(s) of RKHS. Furthermore, the Riesz presentation theorem shows that any bounded linear function can be reproduced as a representer based on the RKHS kernel with a closed form.

However, existing RKHS methods often meet obstacles when choosing different norms and the corresponding optimization procedures. Although using a carefully selected norm in the optimization criterion has the advantage of interpretation, it suffers in that the resulting regression estimator generally needs the computation of an inversion of a large matrix (the same as the sample size). Moreover, most of the existing methods, including the aforementioned RKHS methods, are designed for the case where the observed data are sampled from a dense rate and are limited to models in which either the response or predictors are functions. New methods for estimation and hypothesis testing of regression parameters for the more general case where both the response and predictors are functions with sparsely observed data are needed. To address these problems, we propose a new RKHS method with a unified norm to characterize the RKHS and the optimization criterion for function-on-function regression. Although the statistical interpretation of this optimization criterion is not fully clear, with a simple closed form of the estimated regressors under a general function-on-function regression model, this optimization is more computationally reliable and applicable without the need of computing the inverse of a massive matrix. By establishing a representation theorem and a functional central limit theorem based on the proposed model, we obtain the asymptotic distribution of the estimators. Hypothesis testing of the underlying curves is proposed accordingly.

The remainder of this paper is organized as follows. Section 2 describes the proposed method for the estimation and hypothesis testing of regression parameters for functional data via the reproducing kernel Hilbert space and establishes some theoretical properties. Simulation studies and a real-data example to demonstrate the effectiveness of our proposed method are given in Sections 3 and 4, respectively. Section 5 gives some concluding remarks, and all technical proofs are left in the Appendix A.

2. The Proposed Method

We consider the observed data $\{(y_i(t_{ij}), \mathbf{x}_i(t_{ij})), j = 1, \ldots, m_i; i = 1, \ldots, n\}$. The underlying data curves $(y_i(\cdot), \mathbf{x}_i(\cdot))$ are iid copies from $(y(\cdot), \mathbf{x}(\cdot))$, where $y(\cdot)$ and $\mathbf{x}(\cdot) = (x_1(\cdot), \ldots, x_d(\cdot))'$ are random curves on some region T. The observation times $t_{ij} \in (0, T]$ are generally assumed to be different for each subject i for some $0 < T < \infty$. We assume that time points m_i $(i = 1, \ldots, n)$ are iid copies from some integer-valued random variable m, and given m_i, the time points t_{ij} for $(j = 1, \ldots, m_i)$ are iid copies from a positive random variable G, with its support on $(0, T]$. For each individual, the observed data $(y_i, \mathbf{x}_i)$ can be interpolated as curves $(\hat{y}_i, \hat{\mathbf{x}}_i)$ on T. We assume the following model for the observed data

$$y_i(t) = \boldsymbol{\beta}'(t)\mathbf{x}_i(t) + \epsilon_i(t), \quad E[\epsilon_i(t)] = 0, \quad (i = 1, \ldots, n), \tag{4}$$

where $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_d(\cdot))'$ are the true regression coefficient functions for the covariates $\boldsymbol{x}_i(\cdot)$'s, and the $\epsilon_i(\cdot)$'s are random errors. In general, $\epsilon_i(s)$ and $\epsilon_i(t)$ are non-independent for $s \neq t$, e.g., $\epsilon_i(\cdot)$ being a zero-mean Gaussian process with some covariance function $\Gamma(s, t)$, known or unknown. Note that model (4) is more general than (2) and is more straightforward than model (3) in describing the relationship between the responses $y_i(\cdot)$-th and the covariates $\boldsymbol{x}_i(\cdot)$. Typically, we set $x_1(\cdot) \equiv 1$, and so $\beta_1(\cdot)$ is the baseline function. Since t_{ij} and t_{kj} may be different even for the same j, there may be no observation or just a few observations at each time point t.

To estimate the regression coefficient function $\boldsymbol{\beta}(\cdot)$, the simplest way is the pointwise least squares estimate or any other non-smoothing (i.e., without roughness penalty) functional estimates. However, those estimates have some undesirable properties, often with wiggly shape and large variances in the area with sparse observations. An established performance measure for functional estimation is the mean square error (MSE),

$$\text{MSE} = \text{Bias}^2 + \text{Sampling variance}.$$

Non-smoothed estimates often have small bias but large sampling variance, while smoothed estimates are the other way around, with much smoother shape by adjusting

the shape from neighboring data, but with larger bias. To better balance the trade-off between bias and sampling variance and optimize the MSE, a regularized smooth estimate is preferred, in which a smoothing parameter could control the degree of penalty.

Existing smoothing methods all suffer different aspects of weakness. Functional principal component analysis [15] is computationally intensive. General spline and kernel smoothing methods [24] do not fit the problem under research due to their constant choice of bandwidth. It is known that for non-smoothing methods, computation complexity is often of the order $O(n)$, where n is the data sample size, while for smoothing methods the amount of computation may substantially exceeds $O(n)$ and even become computationally prohibitive. Thus, for smoothing methods, it is important to find a method with $O(n)$ computation load. To achieve this with spline methods, the basis should have only local support (i.e., nonzero only locally). Recently, a popular method in functional estimation is using the reproducing kernel Hilbert space (RKHS). RKHS is a special spline method that has this property, and can achieve the $O(n)$ computation for many functional estimation problems [5,12].

For functional estimate with RKHS, we define two norms (inner products) on the same RKHS $\mathbb{H}$: one, denoted by $< \cdot, \cdot >$, defines the objective optimization criterion, and another one, denoted by $< \cdot, \cdot >_{\mathbb{H}}$, is for the RKHS $\mathbb{H}$. Different from a general Hilbert space, in an RKHS $\mathbb{H}$ of functions on T, the point evaluation functional $\rho_t(h) = h(t) \ \ (h \in \mathbb{H})$ is a continuous linear map, so that by the Riesz representation theorem, there is a bi-variate function $k(\cdot, \cdot)$ on T such that

$$\rho_t(h) = h(t) =< h(\cdot), k(\cdot, t) >_{\mathbb{H}}, \quad \forall\, h \in \mathbb{H}.$$

Take $h(\cdot) = k(\cdot, s)$, we also get

$$k(t, s) =< k(\cdot, s), k(\cdot, t) >_{\mathbb{H}} .$$

The above two properties yield the name RKHS.

Note that for a given Hilbert space $\mathbb{H}$, a collection of functions on some domain T with a given inner product $< \cdot, \cdot >_{\mathbb{H}}$, its reproducing kernel K may not be unique. In fact, for any mapping $G : T \mapsto \ell_2(T^2)$, $K(s, t) =< G(s, \cdot), G(t, \cdot) >_{\mathbb{H}}$ is a reproducing kernel for $\mathbb{H}$, and any reproducing kernel of $\mathbb{H}$ can be expressed in this form (Berlinet and Thomas-Agnan, 2004), and it has a one-to-one correspondence with a covariance function on $\mathbb{H}^2$. The choice of a kernel is mainly for convenience. However, a reproducing kernel under one inner product may not be a reproducing kernel under another inner product on the same space $\mathbb{H}$. Assume $\boldsymbol{\beta} \in \mathbb{H}^d$, with $\mathbb{H}$ being some RKHS and a known kernel $K(\cdot, \cdot)$, both are to be specified later. Let $< \cdot, \cdot >$ be another inner product on $\mathbb{H}$ (typically $< f, g \geq \int_T f(t)g(t)dt$ and $||h||^2 =< h, h >$ for all $h \in \mathbb{H}$). With the observed curves $(\hat{y}_i, \hat{\boldsymbol{x}}_i)$ $(i = 1, \ldots, n)$, ideally an optimization procedure for estimating $\boldsymbol{\beta}(\cdot)$ in (4) will be of the form

$$\hat{\boldsymbol{\beta}}_{n,\lambda}(\cdot) = \arg \inf_{\boldsymbol{\beta} \in \mathbb{H}^d} \left(\frac{1}{n} \sum_{i=1}^{n} ||\hat{y}_i - \boldsymbol{\beta}' \hat{\boldsymbol{x}}_i||^2 + \lambda J(\boldsymbol{\beta}) \right),$$

where $J(\cdot)$ is a penalty functional, and $\lambda > 0$ is the smoothing parameter. The penalty term $J(\cdot)$ can be significantly simplified via the RKHS as shown in the proof of Theorem 1 below. If $\lambda = 0$, the above procedure gives the unsmoothed estimate with some undesirable properties such as overfitting and large variance.

For model (2) with one covariate variable, Yuan and Cai [12] considered penalized estimate $\hat{\beta}$ of $\beta(\cdot)$. The corresponding estimator $\hat{\beta}(\cdot)$ has a closed form of being linear in $x_i(\cdot)$, but the computation involves the inverse of an $(n + 2)$ matrix. For model (1) with d covariates, we first consider estimator of $\boldsymbol{\beta}(\cdot)$ in the form of linear in $\boldsymbol{x}_i(\cdot)$. It turns out that the estimator has a closed form but also involves the inverse of a $d(n + 2)$ matrix, which is computationally infeasible in general.

Consider an estimator of $\boldsymbol{\beta}$ in the form of a linear combination of $(\hat{\boldsymbol{x}}_1(\cdot), \ldots, \hat{\boldsymbol{x}}_n(\cdot))$. For any $f \in \mathbb{H}$, denote $(K_0 f)(t) =< K_0(t, \cdot), f(\cdot) >_{\mathbb{H}}$, and for any $\mathbf{f} = (f_1, \ldots, f_d)' \in \mathbb{H}^d$, denote $(K_0\mathbf{f})(t) = ((K_0 f_1)(t), \ldots, (K_0 f_d)(t))'$ and similarly for $K_1\mathbf{f}$. For $d \times n$ matrix $\boldsymbol{B}$ and $n \times d$ matrix $\mathbf{Z}$, let $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ be the d rows of $\boldsymbol{B}$, $\mathbf{z}_1, \ldots, \mathbf{z}_d$ be the d columns of $\mathbf{Z}$, and define $\boldsymbol{B} \odot \mathbf{Z} = (\boldsymbol{b}_1\mathbf{z}_1, \ldots, \boldsymbol{b}_d\mathbf{z}_d)'$ a d-column vector. Since $\hat{\boldsymbol{x}}_i = K_0\hat{\boldsymbol{x}}_i + K_1\hat{\boldsymbol{x}}_i$, and $K_0\hat{\boldsymbol{x}}_i \in \mathbb{H}_0^d$, $\mathbb{H}_0$ has a basis $\boldsymbol{g} = (g_1(\cdot), \ldots, g_k(\cdot))'$, we consider estimate $\hat{\boldsymbol{\beta}}(\cdot)$ of $\boldsymbol{\beta}_0(\cdot)$ with the form $\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n$, where $\boldsymbol{A}$ is a $d \times k$ matrix, $\boldsymbol{B}$ is a $d \times n$ matrix, and $\boldsymbol{Z}_n(\cdot) = (K_1\hat{\boldsymbol{x}}_1(\cdot), \ldots, K_1\hat{\boldsymbol{x}}_n(\cdot))'$ is $n \times d$. With $||h||^2 = \int_T h^2(t)dt$, for fixed λ an RKHS estimator of $\boldsymbol{\beta}_0(\cdot)$ is of the form

$$\hat{\boldsymbol{\beta}}_{n,\lambda}(\cdot) = \hat{\boldsymbol{A}}\boldsymbol{g} + \hat{\boldsymbol{B}} \odot \boldsymbol{Z}_n(\cdot),$$

where

$$(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}) = \arg \inf_{(\boldsymbol{A},\boldsymbol{B})} \left(\frac{1}{n} \sum_{i=1}^{n} ||\hat{y}_i - \hat{\boldsymbol{x}}_i'(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n)||^2 + \lambda J(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n) \right). \tag{5}$$

For the penalty, let $\mathbf{D}$ be a pre-specified $d \times d$ symmetric positive definite constant matrix; we define

$$J(\mathbf{h}) =< \mathbf{h}'(\mathbf{D}^{1/2})', \mathbf{D}^{1/2}\mathbf{h} >_{\mathbb{H}} =< \mathbf{h}'\mathbf{D}, \mathbf{h} >_{\mathbb{H}} := ||\mathbf{h}||_{\mathbb{H}}^2, \quad \mathbf{h} \in \mathbb{H}^d$$

and

$$\mathbb{H}_0^d = \{\mathbf{h} \in \mathbb{H}^d : J(\mathbf{h}) = 0\} = \{\mathbf{h} \in \mathbb{H}^d : ||\mathbf{h}||_{\mathbb{H}}^2 = 0\} \subset \mathbb{H}^d$$

as the null space for the penalty, and $\mathbb{H}_1^d$ is its orthogonal complement (with respect to the inner product $< \cdot\mathbf{D}, \cdot >_{\mathbb{H}}$). Then, $\mathbb{H}^d = \mathbb{H}_0^d \oplus \mathbb{H}_1^d$. That is, $\forall \mathbf{h} \in \mathbb{H}^d$; it has the decomposition $\mathbf{h} = \mathbf{h}_0 + \mathbf{h}_1$, with $\mathbf{h}_0 \in \mathbb{H}_0^d$ and $\mathbf{h}_1 \in \mathbb{H}_1^d$. Here, $\mathbb{H}_1$ is also an RKHS with some reproducing kernel $K_1(\cdot, \cdot)$ on $\mathbb{H}_1$. With RKHS, $K_0\mathbf{h} \in \mathbb{H}_0^d$ for all $\mathbf{h} \in \mathbb{H}^d$, which implies that $< (K_0\mathbf{h})'\mathbf{D}, K_0\mathbf{h} >_{\mathbb{H}} = 0$. Further, $K_1\mathbf{h} \in \mathbb{H}_1^d$ for all $\mathbf{h} \in \mathbb{H}^d$, and $< (K_0\mathbf{h})'\mathbf{D}, K_1\mathbf{h} >_{\mathbb{H}} = 0$. Thus

$$J(\mathbf{h}) =< \mathbf{h}'\mathbf{D}, \mathbf{h} >_{\mathbb{H}} =< (K_0\mathbf{h} + K_1\mathbf{h})'\mathbf{D}, K_0\mathbf{h} + K_1\mathbf{h} >_{\mathbb{H}} =< (K_0\mathbf{h})'\mathbf{D}, K_0\mathbf{h} >_{\mathbb{H}}$$

$$+2 < (K_0\mathbf{h})'\mathbf{D}, K_1\mathbf{h} >_{\mathbb{H}} + < (K_1\mathbf{h})'\mathbf{D}, K_1\mathbf{h} >_{\mathbb{H}} =< (K_1\mathbf{h})'\mathbf{D}, K_1\mathbf{h} >_{\mathbb{H}} .$$

Typically, $\mathbf{D}$ is chosen to be a $d \times d$ identity matrix. The choices of K_0, K_1, and the inner product $< \cdot\mathbf{D}, \cdot >_{\mathbb{H}}$ will be addressed latter.

For a function $a(\cdot)$ and a vector of functions $\boldsymbol{b}(\cdot) = (b_1(\cdot), \ldots, b_k(\cdot))'$, denote $< a, \boldsymbol{b} \geq (< a, b_1 >, \ldots < a, b_k >)'$; for a matrix $\boldsymbol{B}(\cdot) = \left(b_{ij}(\cdot)\right)_{d\times k}$, denote $< a, \boldsymbol{B} \geq (< a, b_{ij} >)_{d\times k}$, and similarly for the notations $< a, \boldsymbol{b} >_{\mathbb{H}}$ and $< a, \boldsymbol{B} >_{\mathbb{H}}$. The following representation theorem shows that the estimator given in (5) is computationally feasible for many applications.

Theorem 1. *Assume $\boldsymbol{\beta}_0(\cdot) \in \mathbb{H}^d$, $(\hat{y}_i(\cdot), \hat{\boldsymbol{x}}_i(\cdot)) \in \mathbb{H}^{d+1}$ for $i = 1, \ldots, n$. Then for the given penalty functional $J(\boldsymbol{\beta}) = ||K_1(\boldsymbol{\beta})||_{\mathbb{H}}^2$ and fixed λ, there are constant matrices $\hat{\mathbf{A}} = (a_{ij})_{d\times k}$ and $\hat{\mathbf{B}} = (b_{ij})_{d\times n}$ such that $\hat{\boldsymbol{\beta}}_{n,\lambda}$ given in (5) has the following representation*

$$\hat{\boldsymbol{\beta}}_{n,\lambda}(t) = \hat{\boldsymbol{A}}\boldsymbol{g}(t) + \hat{\boldsymbol{B}} \odot (K_1\hat{\boldsymbol{x}})_n(t), \quad t \in (0, T]$$

where $(K_1\hat{\boldsymbol{x}})_n(\cdot) = (K_1\hat{\boldsymbol{x}}_1(\cdot), \ldots, K_1\hat{\boldsymbol{x}}_n(\cdot))'$, and in vector form $(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}})$ of $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}})$

$$\begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{S} + \lambda\boldsymbol{W} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix},$$

where the matrices $\boldsymbol{R}$ $(dk \times dn)$, $\boldsymbol{O}$ $(dk \times dk)$, $\boldsymbol{S}$ $(dn \times dn)$, and $\boldsymbol{W}$ $(dn \times dn)$, and the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are given in the proof.

For the ordinary regression model $y = \boldsymbol{\beta}'\boldsymbol{x} + \epsilon$, with $\boldsymbol{X}_n = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ and $\boldsymbol{y}_n = (y_1, \ldots, y_n)'$, the least squares method yields the estimation of $\boldsymbol{\beta}$ as $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'\boldsymbol{y}_n$. Since $(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}$ is of order n^{-1} (a.s.), $\hat{\boldsymbol{\beta}}$ can be viewed as approximately a linear form $n^{-1}\boldsymbol{X}_n'\boldsymbol{y}_n$. Let $\hat{\boldsymbol{X}}_n(\cdot) = (\hat{\boldsymbol{x}}_1(\cdot), \ldots, \hat{\boldsymbol{x}}_n(\cdot))'$ and $\hat{\boldsymbol{y}}_n(\cdot) = (\hat{y}_1(\cdot), \ldots, \hat{y}_n(\cdot))'$. Now we consider estimate $\hat{\boldsymbol{\beta}}(\cdot)$ of $\boldsymbol{\beta}_0(\cdot)$ with linear form $n^{-1}\hat{\boldsymbol{X}}_n'\hat{\boldsymbol{y}}_n$. Since $n^{-1}\hat{\boldsymbol{X}}_n'\hat{\boldsymbol{y}}_n = K_0(n^{-1}\hat{\boldsymbol{X}}_n'\hat{\boldsymbol{y}}_n) + K_1(n^{-1}\hat{\boldsymbol{X}}_n'\hat{\boldsymbol{y}}_n)$, and $K_0(n^{-1}\hat{\boldsymbol{X}}_n'\hat{\boldsymbol{y}}_n) \in \mathbb{H}_0^d$, we only need to consider an estimate of the form $\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B}\hat{\boldsymbol{z}}_n$, where $\boldsymbol{A}$ is a $d \times k$ parameter matrix, $\boldsymbol{B}$ is a $d \times d$ parameter matrix, and $\hat{\boldsymbol{z}}_n(\cdot) = n^{-1}[K_1(\hat{\boldsymbol{X}}_n'\hat{\boldsymbol{y}}_n)](\cdot)$ is a d-vector. This allows us to express the estimate via the basis of the RKHS and with a greater degree of flexibility than the linear combination of $n^{-1}\hat{\boldsymbol{X}}_n'\hat{\boldsymbol{y}}_n$. Another advantage of using estimates of the form $\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B}\hat{\boldsymbol{z}}_n$ is convenience of hypothesis testing. As typically $\boldsymbol{g} = (1, t)'$, thus testing the hypothesis of linearity of $\boldsymbol{\beta}(\cdot)$ is equivalent to testing $\boldsymbol{B} = \boldsymbol{0}$.

For any function $h(\cdot)$, we set $||h||^2 = \int_T h^2(t)dt$, and for fixed λ,

$$\hat{\boldsymbol{\beta}}_{n,\lambda}(\cdot) = \hat{\boldsymbol{A}}\boldsymbol{g}(\cdot) + \hat{\boldsymbol{B}}\hat{\boldsymbol{z}}_n(\cdot),$$

where

$$(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}) = \arg \inf_{(\boldsymbol{A},\boldsymbol{B})} \left(\frac{1}{n} \sum_{i=1}^{n} ||\hat{y}_i - (\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B}\hat{\boldsymbol{z}}_n)'\hat{\boldsymbol{x}}_i||^2 + \lambda J(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B}\hat{\boldsymbol{z}}_n) \right). \tag{6}$$

Let $\boldsymbol{a} = (a_{11}, \ldots, a_{1k}, \ldots, a_{d1}, \ldots, a_{dk})'$ be the vector representation of $\boldsymbol{A}$; $\boldsymbol{b} = (b_{11}, \ldots, b_{1d}, \ldots, b_{d1}, \ldots, b_{dd})'$ be that of $\boldsymbol{B}$, $\boldsymbol{O} = \boldsymbol{O}_{dk\times dk} = n^{-1}\sum_{i=1}^n < \boldsymbol{s}_i, \boldsymbol{s}_i' >$ with $\boldsymbol{s}_i = (\hat{x}_{i1}g_1, \ldots, \hat{x}_{i1}g_k, \ldots, \hat{x}_{id}g_1, \ldots, \hat{x}_{id}g_k)'$, $\boldsymbol{R}' = \boldsymbol{P} = \boldsymbol{P}_{d^2\times dk} = n^{-1}\sum_{i=1}^n < \boldsymbol{t}_i, \boldsymbol{s}_i' >$ with $\boldsymbol{t}_i = (\hat{x}_{i1}\hat{z}_1, \ldots, \hat{x}_{i1}\hat{z}_d, \ldots, \hat{x}_{id}\hat{z}_1, \ldots, \hat{x}_{id}\hat{z}_d)'$, $\boldsymbol{S} = \boldsymbol{S}_{d^2\times d^2} = n^{-1} < \boldsymbol{t}_i, \boldsymbol{t}_i' >$, $\boldsymbol{U} = n^{-1}\sum_{i=1}^n < \hat{y}_i, \hat{\boldsymbol{x}}_i\boldsymbol{g}' \geq (u_{ij})_{d\times k}$ and its vector form $\boldsymbol{u} = (u_{11}, \ldots, u_{1k}, \ldots, u_{d1}, \ldots, u_{dk})'$, $\boldsymbol{V} = n^{-1}\sum_{i=1}^n < \hat{y}_i, \hat{\boldsymbol{x}}_i\hat{\boldsymbol{z}}_n' \geq (v_{ij})_{d\times d}$ and its vector form $\boldsymbol{v} = (v_{11}, \ldots, v_{1d}, \ldots, v_{d1}, \ldots, v_{dd})'$; $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$ be all the eigenvalues of D, and $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_d$ be its normalized eigenvectors, $\boldsymbol{W} = \boldsymbol{W}_{d^2\times d^2} = n^{-1}\sum_{j=1}^n < \boldsymbol{c}_j, \boldsymbol{c}_j' >_{\mathbb{H}}$ and $\boldsymbol{c}_j = \lambda_j(q_{j1}\hat{z}_1, \ldots, q_{j1}\hat{z}_d, \ldots, q_{jd}\hat{z}_1, \ldots, q_{jd}\hat{z}_d)'$.

Theorem 2. *Assume $\boldsymbol{\beta}(\cdot) \in \mathbb{H}^d$, $(\hat{y}_i(\cdot), \hat{\boldsymbol{x}}_i(\cdot)) \in \mathbb{H}^{d+1}$ for $i = 1, \ldots, n$. Then for the given penalty functional $J(\boldsymbol{\beta}) = ||K_1(\boldsymbol{\beta})||_{\mathbb{H}}^2$ and fixed λ, there are constant matrices $\hat{\mathbf{A}} = (a_{ij})_{d\times k}$ and $\hat{\mathbf{B}} = (b_{ij})_{d\times d}$ such that $\hat{\boldsymbol{\beta}}_{n,\lambda}(\cdot)$ given in (6) has the following representation*

$$\hat{\boldsymbol{\beta}}_{n,\lambda}(t) = \hat{\boldsymbol{A}}\boldsymbol{g}(t) + \hat{\boldsymbol{B}}(K_1[n^{-1}\hat{\boldsymbol{X}}_n'\hat{\boldsymbol{y}}_n])(t), \quad t \in (0, T]$$

and in vector form $(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}})$ of $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}})$ when the following inverse exists,

$$\begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{S} + \lambda\boldsymbol{W} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}.$$

Below we study asymptotic behavior of $\hat{\boldsymbol{\beta}}_{n,\lambda}(\cdot)$ given in (6). Denote $\boldsymbol{\beta}_0(\cdot)$ as the true value of $\boldsymbol{\beta}(\cdot)$, and $|\mathbf{M}|$ is the determinant of a square matrix $\mathbf{M}$. Lai et al. [25] proved strong consistency of the least squares estimate under general conditions, while Eicker [26] studied its asymptotic normality. The proposed estimators in this paper have some similarity to the least squares estimate, but they also have some different features and require different conditions.

(C1). $\boldsymbol{\beta}_0 \in Span(E[\boldsymbol{x}y])$.
(C2). $\inf_{t\in T} |E[\boldsymbol{x}(t)\boldsymbol{x}'(t)]| > 0$.
(C3). $E\left(||y - (\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B}\boldsymbol{Z})'\boldsymbol{x}||^2 \right) < \infty$ for all bounded $(\boldsymbol{A}, \boldsymbol{B})$, where $\boldsymbol{Z} = E[K_1(\boldsymbol{x}y)]$.
(C4). $\lim_{n\to\infty} \max_{1\leq i\leq n} ||(\hat{y}_i, \hat{\boldsymbol{x}}_i) - (y_i, \boldsymbol{x}_i)|| \to 0$ (a.s.).
(C5). $\lambda = \lambda_n \to 0$.

Theorem 3. *Assume conditions (C1)–(C5) hold, then as* $n \to \infty$,

$$||\hat{\boldsymbol{\beta}}_{n,\lambda} - \boldsymbol{\beta}_0|| \to 0, \quad (a.s.).$$

To emphasize the dependence on n, we denote $\lambda = \lambda_n$. Let $l^\infty(T)$ be the space of bounded functions on T equipped with the supreme norm, and $\stackrel{D}{\Rightarrow}$ stands for weak convergence in the space $l^\infty(T)$. With the following condition (C6), we obtain the asymptotic normality of $\hat{\boldsymbol{\beta}}_{n,\lambda}(\cdot)$

(C6). $\sqrt{n}\lambda_n \to 0$.

Theorem 4. *Assume conditions (C1)–(C4) and (C6) hold. Then as* $n \to \infty$,

$$\mathbb{W}_n := \sqrt{n}\big(\hat{\boldsymbol{\beta}}_{n,\lambda} - \boldsymbol{\beta}_0 - o_p(1)\big) \stackrel{D}{\Rightarrow} \mathbb{W} \quad on \ \ l^\infty(T),$$

where $\mathbb{W}(\cdot)$ *is the zero-mean Gaussian process on* T *with covariance function* $\sigma(s,t) = E[\mathbb{W}(s)\mathbb{W}(t)]$ *given in the proof,* $s, t \in T$, *and* $o_p(1)$ *is given in the proof.*

Test linearity of $\boldsymbol{\beta}_0$. It is of interest to test the hypothesis $H_0(J) : J'\boldsymbol{\beta}_0(t)$ is linear in t, where J is a d-dimensional vector with entries 0 or 1, with 1 corresponding to the element of $\boldsymbol{\beta}_0$ to be tested for linearity. The hypothesis $H_0(J)$ is equivalent to test the corresponding coefficients $J'\hat{\boldsymbol{B}}$ in $\hat{\boldsymbol{B}}$ be zeros. Let $\boldsymbol{O}_0 = E < \boldsymbol{s}_1, \boldsymbol{s}_1' >$, $\boldsymbol{P}_0 = E < \boldsymbol{t}_1, \boldsymbol{s}_1' >$, $\boldsymbol{S}_0 = E < \boldsymbol{t}_1, \boldsymbol{t}_1' >$, $\boldsymbol{U}_0 = E < y_1, \boldsymbol{x}_1 g' >$, $\boldsymbol{V}_0 = E < y_1, \boldsymbol{x}_1 \boldsymbol{z}_0' >$. Let $\boldsymbol{u}_0$ and $\boldsymbol{v}_0$ be the vector representations of $\boldsymbol{U}_0$ and $\boldsymbol{V}_0$, and $\boldsymbol{w}_0 = (\boldsymbol{u}_0', \boldsymbol{v}_0')'$. Denote $\boldsymbol{T} = matrix(\boldsymbol{O}, \boldsymbol{R}; \boldsymbol{P}, \boldsymbol{S})$, $\boldsymbol{T}_0 = matrix(\boldsymbol{O}_0, \boldsymbol{R}_0; \boldsymbol{P}_0, \boldsymbol{S}_0)$. By Theorem 4, we have

Corollary 1. *Assume the conditions of Theorem 4 hold, under* $H_0(J)$, *we have*

$$\sqrt{n}(J'\hat{\boldsymbol{B}} - o_p(1)) \stackrel{D}{\to} N(\boldsymbol{0}, \Omega(J)),$$

where $\Omega(J)$ *is the sub-matrix of* $\boldsymbol{T}_0^{-1}\Gamma\boldsymbol{T}_0^{-1}$ *that corresponds to the covariance of* $J'\hat{\boldsymbol{B}}$, $o_p(1) = (\boldsymbol{T} - \boldsymbol{T}_0)\boldsymbol{w}_0$, *and* Γ *is given in the proof of Theorem 4.*

The nonzero bias term $o_p(1)$ in Theorem 4 and Corollary 1 is typical in functional estimation, and often such a bias term is zero for the corresponding Euclidean parameter estimation.

Choice of the smoothing parameter. In nonparametric penalized regression for the model $y(t) =< \beta, x > (t) + \epsilon(t)$, the most commonly-used method for the choice of the smoothing parameter is cross-validation (CV), based on the ideas of Allen (1974) and Stone (1974). This method chooses λ by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\big[y_i(t_{ij}) - < \hat{\beta}_{n,\lambda,i}, \hat{x}_i > (t_{ij})\big]^2,$$

where $\hat{\beta}_{n,\lambda,i}(\cdot)$ is the estimated regression function without using the observations of the ith individual. This method is usually computationally intensive even when the sample size is moderate. An improved version of the method is K-fold cross-validation. This method first randomly partitions the original sample equally into K subsamples, and then the cross-validation process is conducted K times. At each replicate, $K-1$ subsamples are used as the training data to construct the model, while the remaining one is used as the validation datum. The results from K folds are averaged to obtain a single estimation. In

notation, let $n_1, \ldots, n_K$ be the sample sizes of the K folds, then the K-fold cross-validation method is to choose the λ which minimizes

$$\frac{1}{K}\sum_{J=1}^{K}\frac{1}{n_J}\sum_{i=1}^{n_J}\frac{1}{m_i}\sum_{j=1}^{m_i}\left[y_i(t_{ij}) - < \hat{\beta}_{n,\lambda,J}, \hat{x}_i > (t_{ij})\right]^2,$$

where $\hat{\beta}_{n,\lambda,J}(\cdot)$ is the estimated regression function without using the data in the Jth fold. In this paper, we set $K = 5$, which is also the default setting in much software.

Choices of K_0, K_1 , and $< \cdot, \cdot >_{\mathbb{H}}$. For notational simplicity, we consider $T = [0, 1]$ without loss of generality. Recall that for a function f on $[0, 1]$ with $m - 1$ continuous derivatives and $f^{(m)}(\cdot) \in L_2[0, 1]$, it has the following Taylor expansion [27]

$$f(t) = \sum_{j=0}^{m-1}\frac{f^{(j)}(0)}{j!}t^j + \int_0^1 \frac{f^{(m)}(s)}{(m-1)!}(t-s)_+^{m-1}ds,$$

where $(x)_+ = x$ if $x > 0$ and $(x)_+ = 0$ otherwise.

To construct an RKHS $\mathbb{H}$ on $L_2[0, 1]$, a common choice for the inner product on $\mathbb{H}_0 = \{h : h^{(2)}(\cdot) \equiv 0\}$ is $< f, g >_{\mathbb{H},0}$, and the orthogonal complement of $\mathbb{H}_0$ is $\mathbb{H}_1 = \{h : h^{(j)}(0) = 0, j = 0, 1; \int_0^1 h^{(2)}(t)dt < \infty\}$, with inner product $< f, g >_{\mathbb{H},1}$, where

$$< f, g >_{\mathbb{H},0} = \sum_{j=0}^{1} f^{(j)}(0)g^{(j)}(0), \qquad < f, g >_{\mathbb{H},1} = \int_0^1 f^{(2)}(t)g^{(2)}(t)dt.$$

The inner product on $\mathbb{H}$ is $< \cdot, \cdot >_{\mathbb{H}} = < \cdot, \cdot >_{\mathbb{H},0} + < \cdot, \cdot >_{\mathbb{H},1}$. Kernels for the RKHS with more general K_0 for $\mathbb{H}_0$ and K_1 for $\mathbb{H}_1$ with these inner products can be found in [28]. More generalized construction of kernels K_0 and K_1 can be found in Ramsay and Silverman [5]. For our case,

$$K_0(s, t) = 1 + st, \quad K_1(s, t) = \int_0^1 (s-u)_+(t-u)_+du = (s \wedge t)^2\big(3(s \vee t) - (s \wedge t)\big)/6.$$

With the above inner product, K_0, and K_1, let $K = K_0 + K_1$, then $\forall h \in \mathbb{H}$, $h(t) = < K(t, \cdot), h(\cdot) >_{\mathbb{H}}$, and $\mathbb{H}_0$ and $\mathbb{H}_1$ are orthogonal to each other with respect to $< \cdot, \cdot >_{\mathbb{H}}$, but these are not true if $< \cdot, \cdot >_{\mathbb{H}}$ is replaced by a different inner product $< \cdot, \cdot >$ on $[0, 1]$.

3. Simulation Studies

In this section, we conduct two simulation studies to investigate the finite sample performance of the proposed RKHS method. The first simulation study is designed to compare the RKHS estimator with the conventional smoothing spline and local polynomial model methods in terms of curve fitting. For more details on the implementations of smoothing spline and local polynomial model methods, please refer to the book by Fang, Li, and Sudijianto [24]. The second simulation study is to examine the performance of Corollary 1 for testing the linearity of the regression functions. It turns out that with moderate sample sizes, the proposed RKHS estimator performs very favorably with the competitors, and the type I errors and powers of the testing are satisfactory.

Simulation 1. Assume that the underlying individual curve i at time point $t \in T = [0, 1]$ is generated from

$$y_i(t) = \beta_0(t) + \beta_1(t)x_{i1}(t) + \beta_2(t)x_{i2}(t) + \epsilon_i(t),$$

where $\beta_0(t) \equiv 10, \beta_1(t) = 1 + t, \beta_2(t) = (1 - t)\sin(2\pi t), x_{i1}(t) = \sin(100\pi t)$, $x_{i2}(t) = \cos(100\pi t)$, and $\epsilon_i(\cdot)$ is a stationary Gaussian process with zero mean, unit variance, and a constant covariance 0.5 between any two distinct time points. For each subject i, the number of observation time points m_i is generated from the discrete uniform distribution on $\{5, 6, \ldots, 30\}$, and the observation time points t_{ij}, $j = 1, \ldots, m_i$ are independently generated

from the exponential distribution $E(0,1)$. The density function of $E(0,1)$ is displayed in the left panel of Figure 1, from which it is easy to see that the density value decreases as t increases.

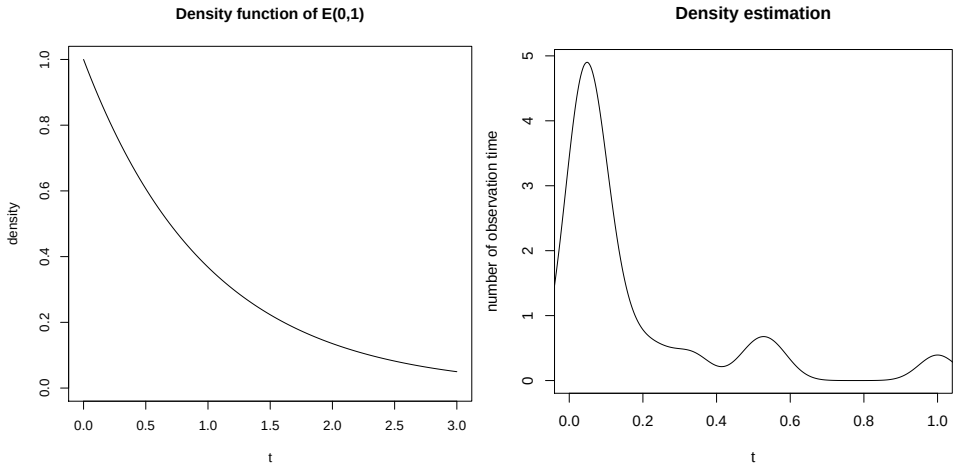

Figure 1. Left panel visualizes the density function of $E(0,1)$; right panel visualizes the kernel density estimation of the number of observation time of MD001.

Then, we use *cubic interpolation* to interpolate the $y_i(t_{ij})$, $x_{i1}(t_{ij})$, and $x_{i2}(t_{ij})$ on T to obtain $\hat{y}_i(\cdot)$, $\hat{x}_{i1}(\cdot)$, and $\hat{x}_{i2}(\cdot)$, respectively.

Based on the functions $\hat{y}_i(\cdot)$, $\hat{x}_{i1}(\cdot)$, and $\hat{x}_{i2}(\cdot)$ described above, we use the RKHS introduced in Section 2 to estimate the regression functions $\beta_0(t), \beta_1(t)$, and $\beta_2(t)$, and compare its performance with the spline smoother and local polynomial models. Typical comparisons (the random seed is set to be "set.seed(1)" in R) are given in Figures 2–4 with sample sizes of 50, 100, and 200, respectively. The simulation shows that the proposed RKHS method estimates the regression functions well and compares very favorably with the other two methods. Broadly speaking, the RKHS estimator has relatively stable performance and is close to the "true" curve; it has narrower confidence bands at dense sampling regions, and they become wider at sparse sampling regions. On the contrary, the spline smoother and local polynomial model appear to have good fit at dense sampling regions, but they have large bias when the data become sparse.

In order to make a thorough comparison for this simulation, we use the *root integrated mean squared prediction error* (RIMSPE) to measure the accuracy of the estimates [24]. The RIMSPE for estimate $\hat{\beta}$ of β is given by

$$\text{RIMSPE}(\hat{\beta}) = \sqrt{\int_0^1 [\beta(t) - \hat{\beta}(t)]^2 dt},$$

and the simulation is repeated 1000 times. By using the R software, the CPU time of implementing this simulation is about 84.5 s on a PC with a 1.80 GHz dual-core Intel i5-8265U CPU and 8 GB memory. The boxplots of the RIMSPE values are presented in Figure 5, from which it is clear that RKHS performs much better than the other two methods, because it has much smaller RIMSPE values.

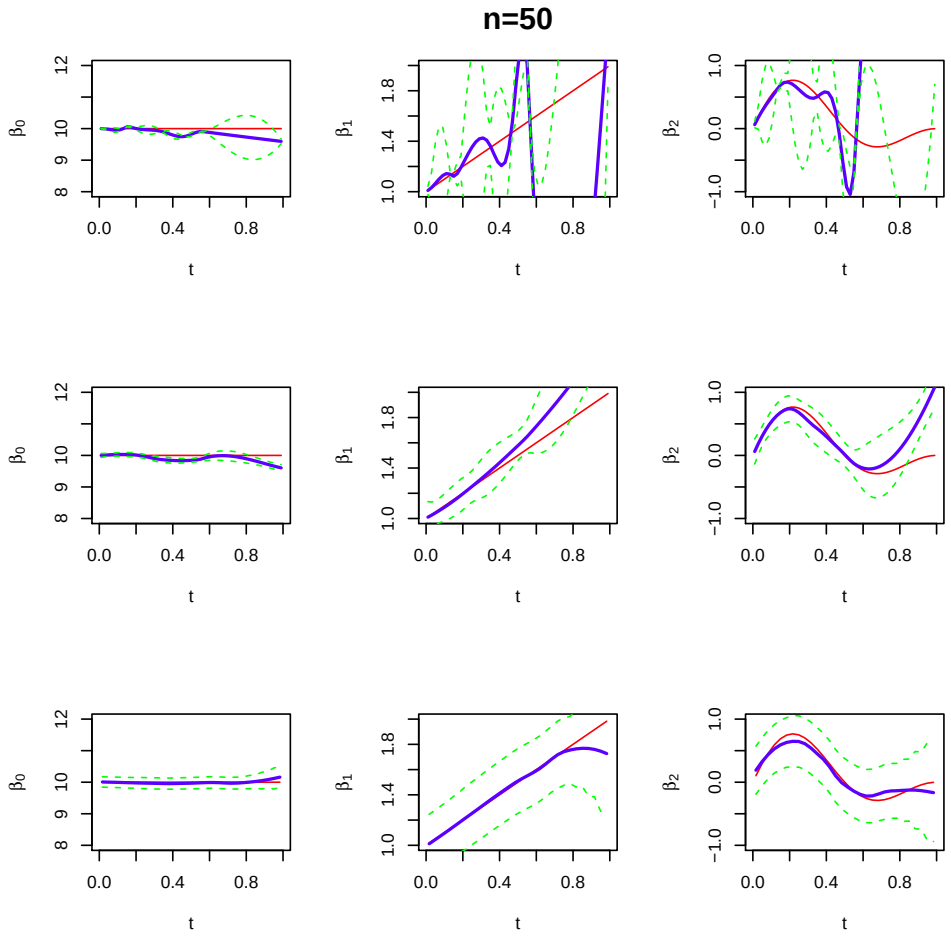

Figure 2. Performance of curve estimation when the sample size is 50 and the random seed is "set.seed(1)" in R. First row: curve estimation performance of the spline smoother; Second row: curve estimation performance of the local polynomial model; Third row: curve estimation performance of the proposed RKHS method. Solid red line: true curve; Solid blue line: estimated curve; Dotted lower and upper green lines: 95% confidence bands.

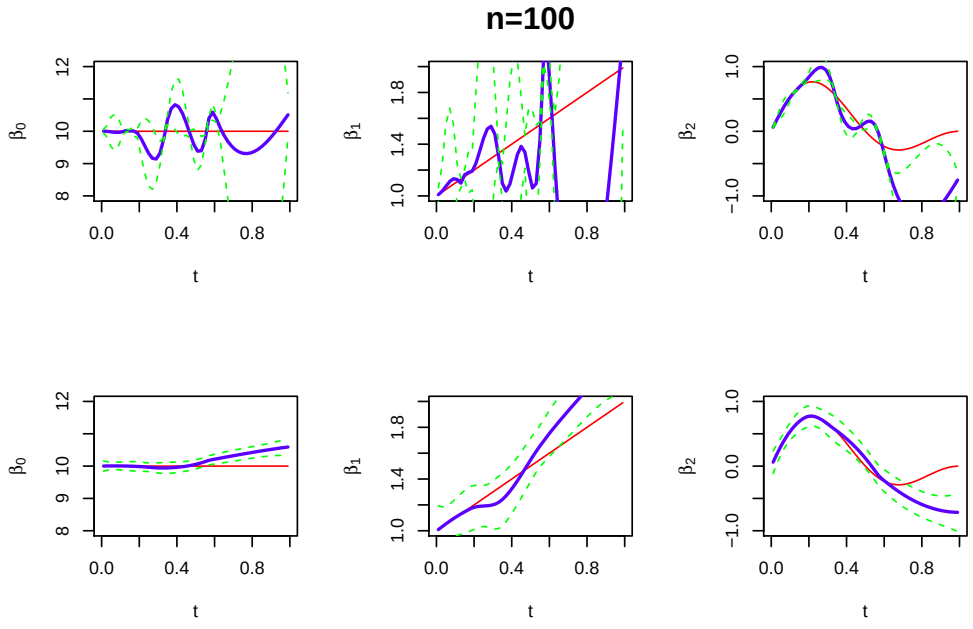

Figure 3. *Cont*.

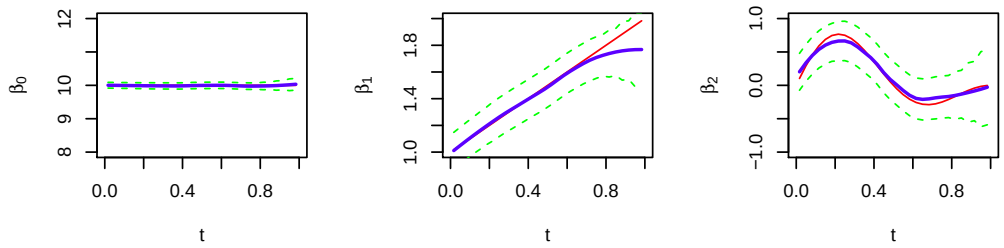

Figure 3. Performance of curve estimation when the sample size is 100 and the random seed is "set.seed(1)" in R. First row: curve estimation performance of the spline smoother; Second row: curve estimation performance of the local polynomial model; Third row: curve estimation performance of the proposed RKHS method. Solid red line: true curve; Solid blue line: estimated curve; Dotted lower and upper green lines: 95% confidence bands.

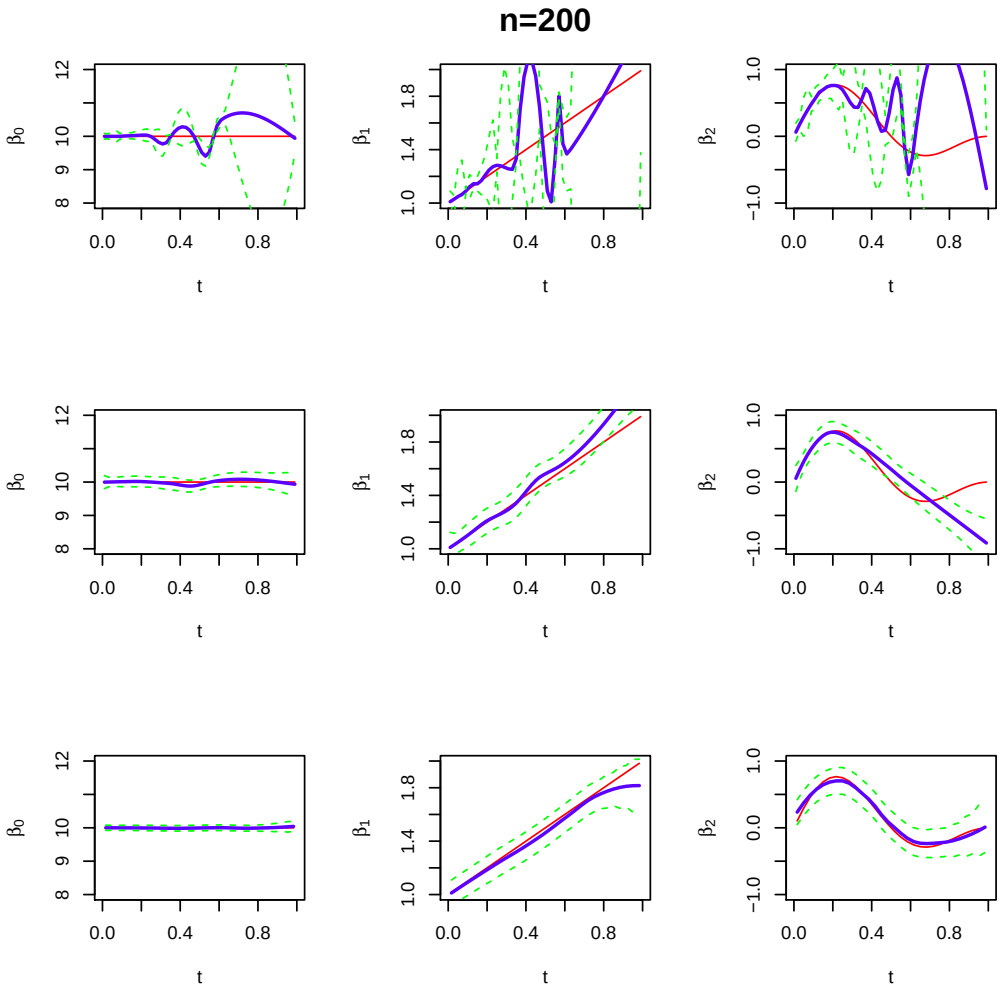

Figure 4. Performance of curve estimation when the sample size is 200 and the random seed is "set.seed(1)" in R. First row: curve estimation performance of the spline smoother; Second row: curve estimation performance of the local polynomial model; Third row: curve estimation performance of the proposed RKHS method. Solid red line: true curve; Solid blue line: estimated curve; Dotted lower and upper green lines: 95% confidence bands.

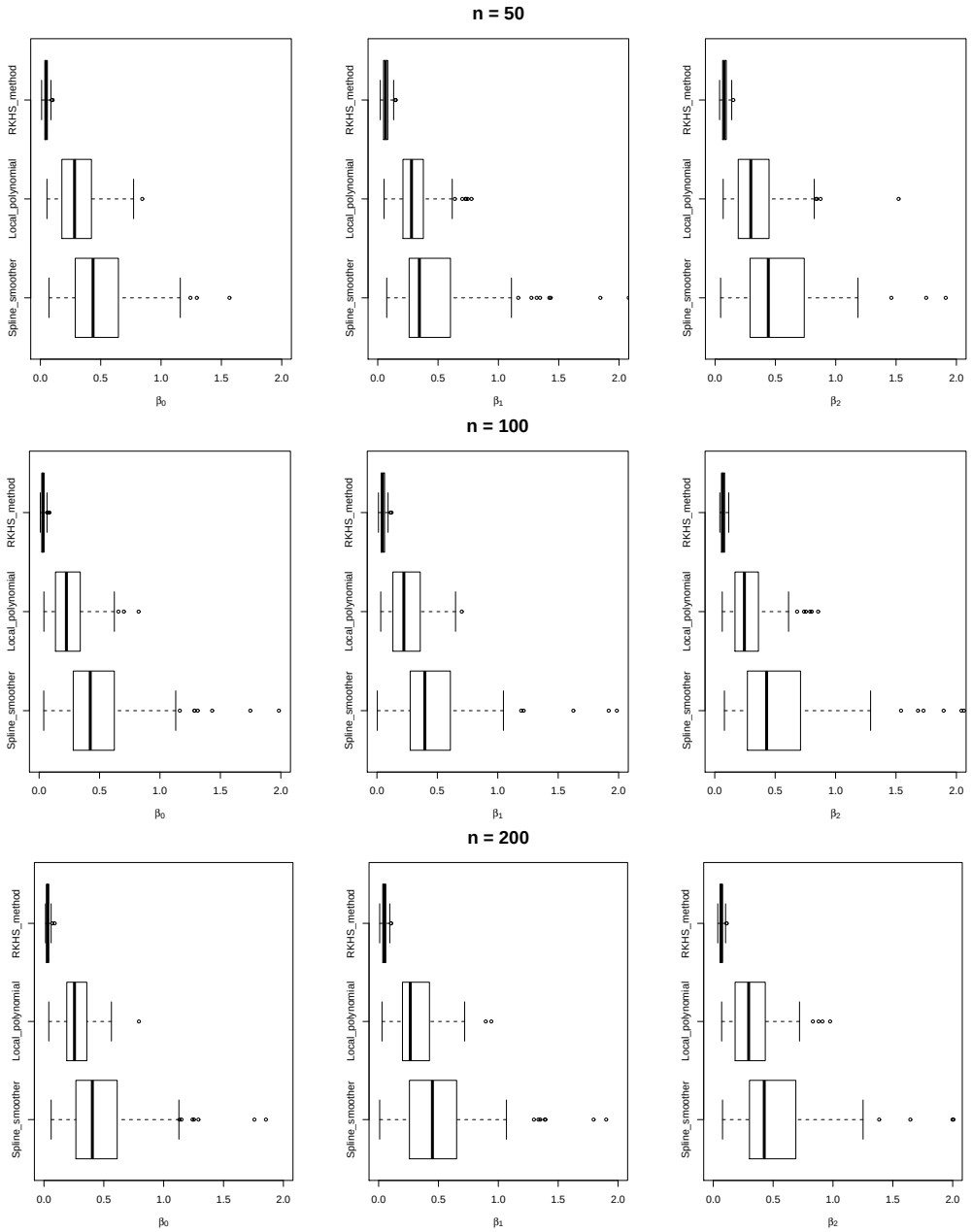

Figure 5. Boxplots of the RIMSPE values. The first row corresponds to sample size 50, the second row corresponds to sample size 100, and the third row corresponds to sample size 200. In each row, the left panel is for estimating $\beta_0(t)$, the middle panel is for estimating $\beta_1(t)$, and the right panel is for estimating $\beta_2(t)$.

Simulation 2. In this simulation study, we examine the performance of Corollary 1 for testing the hypothesis

$$H_0 : \beta_i(t) \text{ is linear in } t \quad VS \quad H_1 : \beta_i(t) \text{ is not linear in } t, \text{ for } i = 1, 2.$$

According to the setting described in *Simulation 1*, $\beta_1(t)$ is linear in t, whereas $\beta_2(t)$ is apparently not linear in t. Therefore, we will check the type I error for testing $\beta_1(t)$ and the power for testing $\beta_2(t)$. By setting the significance level to 0.05 and repeating the simulation 1000 times, we use Corollary 1 to derive χ^2 testing statistics and list its type I errors and powers in Table 1 for various sample sizes. The results in Table 1 suggest that the type I error of the test is close to the nominal level 0.05, and the power of the test is not small even with a sample size of 50.

Table 1. Summary of simulation results for linearity testing.

Sample Size	Type I Error (for Testing $\beta_0(t)$)	Power (for Testing $\beta_1(t)$)
50	0.059	0.756
100	0.052	0.865
200	0.051	0.923

The simulation is based on 1000 repetitions.

4. Real Data Analysis

In this section, the proposed method is applied to characterize the relationships in patient immune response in a clinical trial of combination immunotherapy for advanced myeloma. The objective of the original trial was to study whether introducing vaccine-primed T cells early leads to cellular immune responses to the putative tumor antigen hTERT. In this study, 54 patients were recruited and assigned to two treatment arms based on their leukocyte response to human leukocyte antigen A2. Various immune cell parameters (CD3, CD4, CD8), T-cell levels, cytokines (IL7, IL-15), and immunoglobulins (IgA, IgG, IgM) were measured repeatedly to investigate the treatment effect on immune recovery and function. The measurements were taken at nine time points: 0, 2, 7, 14, 40, 60, 90, 100, and 180 days [29]. Moreover, as a subtype of white blood cells in the human immune system, absolute lymphocyte cell (ALC) count was recorded over time during or after patients' hospitalization up to day 180. Figure 6 shows the trajectories of two individuals, namely "MD001" and "MD002", in the dataset, with the observation interval scaled to $[0, 1]$. The trajectories of all 54 individuals can be found in the paper by Fang et al. [30]. Previous research has shown that the patient's survival time is associated with the trajectory of the patient's ALC counts.

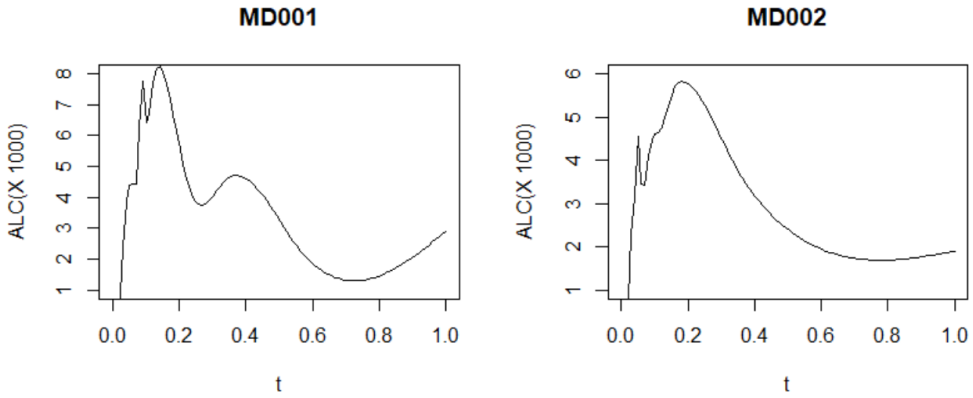

Figure 6. Left panel: trajectory of individual "MD001"; right panel: trajectory of individual "MD002". The observation interval has been scaled to $[0, 1]$.

In the human immune system, the relationships among various biological features are too complicated and have been topologically described only. For illustrating the performance of our proposed methods with a limited sample size, we only investigate how the levels of a patient's immunoglobulin IgG and immune cell CD8 dynamically affect the trajectory of the patient's ALC counts in this section. For simplicity, the observation time points are scaled to the interval $[0, 1]$. Let $x_1(t), x_2(t)$ and $y(t)$ be the trajectories of the patient's IgG, CD8, and ALC counts, respectively. Their relationship can then be described as follows

$$y(t) = \beta_0(t) + \beta_1(t)x_1(t) + \beta_2(t)x_2(t) + \epsilon(t), \quad E[\epsilon(t)] = 0,$$

where $\beta_0(t), \beta_1(t)$ and $\beta_2(t)$ are the regression coefficient functions, and $\epsilon(t)$ is the random error function. The purpose of this study is to estimate the regression coefficient functions and test whether $\beta_1(t)$ and $\beta_2(t)$ are linear functions in t.

In the used data, the number of observation times generally becomes sparse as t increases. The right panel of Figure 7 visualizes the kernel density estimation of individual "MD001" in the data. The distribution of observed time points reveals the trend. The proposed RKHS method is used to estimate the regression coefficient functions and test the linearity. By using the R software, the CPU time of implementing the estimation procedure is only about 1.5 s on a PC with a 1.80 GHz dual-core Intel i5-8265U CPU and 8 GB memory. Figure 7 visualizes the estimated curves and their 95% confidence bands. It is observed that $\beta_1(t)$ and $\beta_2(t)$ are apparently nonlinear in t. This observation is also confirmed by the χ^2 statistic derived from **Corollary 1**, which yields p-values less than 0.001 for both $\beta_1(t)$ and $\beta_2(t)$. It is worth noting that $\beta_0(t)$ is monotone in t, but $\beta_1(t)$ and $\beta_2(t)$ are not monotone in t. The results show that with the immunotherapy of tumor antigen vaccination, a patient's immunoglobulin IgG enhances the ALC counts. When the increasing CD8 immune cells result in a high ALC count, immunoglobulin IgG inhibits the patient's ALC counts such that the level of ALC counts is reconverted into the normal interval $(1000, 4500)$, and this immunotherapy can potentially improve patient survival time.

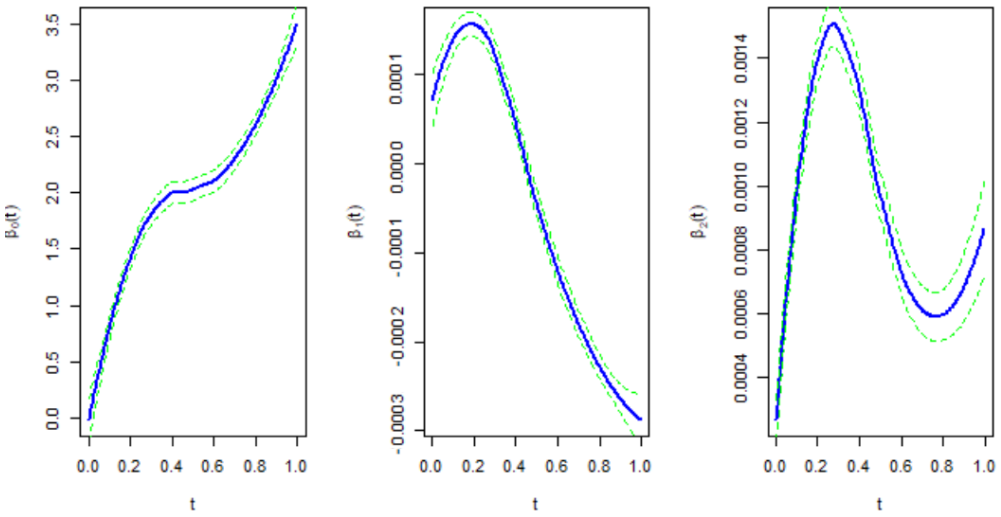

Figure 7. The regression coefficient functions estimated by the proposed RKHS method. Solid blue line: estimated curve; dotted lower and upper green lines: 95% confidence bands. The time t has been scaled to the interval $[0, 1]$.

5. Concluding Remarks

The existing work on functional data analysis has focused primarily on the case where the observed data are sampled from a dense rate and has been limited to models in which either the response or predictors are functions. In this paper, we consider the more practical situation for functional data analysis where the data are only observed at some (not dense) time points, and we propose a general regression model in which both the response and predictors are functions. This function-on-function regression model, as given by Equation (4), can be viewed as a generalization of multivariate multiple linear regression to allow the response, predictors, and even the regression coefficients to be all functions of t. In order to estimate the underlying regression curves and conduct hypothesis testing on these curves, we use reproducing kernel Hilbert space (RKHS), which only needs to choose the kernel(s) of the RKHS, and enables a closed-form solution for the regression coefficients in terms of the kernel. To the best of our knowledge, this is the first representation of functional regression coefficients with sparsely observed data. Furthermore, the estimator based on RKHS provides a foundation for hypothesis testing, and the asymptotic distribution of the estimator is obtained. Simulation studies show that the RKHS estimator has relatively stable performance. Application and statistical properties of our method are further demonstrated through an immunotherapy clinical trial of advanced myeloma. By using the proposed

function-on-function regression model and related theorems established in this paper, this real application showed that with the immunotherapy of tumor antigen vaccination, patient immunoglobulin IgG enhances ALC counts, and hence this immunotherapy can potentially improve patient survival time. Future work may consider experimental design for the time points to be observed. If the time points can be controlled by the experimenter, their careful selection would improve the efficiency of the estimator (e.g., reduce the bias or MES). Further, we hope to study function-on-function generalized linear regressions with sparse estimation coefficient functions by the penalized method of Zhang and Jia [31].

Author Contributions: Conceptualization, H.-B.F.; methodology, H.H.; validation, G.M.; formal analysis, H.L.; writing—original draft preparation, H.H.; writing—H.-B.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Cancer Institute (NCI) grant P30CA 051008 and the Key Laboratory of Mathematical and Statistical Models (Guangxi Normal University), Education Department of Guangxi Zhuang Autonomous Region.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data included in this study are available upon request by contacting the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Theorem 1 (1-dimensional case). In this case, $d = 1$, $\boldsymbol{A} = \boldsymbol{a} = (a_1, \ldots, a_k)$, $\boldsymbol{B} = \boldsymbol{b} = (b_1, \ldots, b_n)$, $\boldsymbol{Z}_n(\cdot) = (K_1\hat{x}_1(\cdot), \ldots, K_1\hat{x}_n(\cdot))'$, $\boldsymbol{B} \odot \boldsymbol{Z}_n = \boldsymbol{b}\boldsymbol{Z}_n = \sum_{i=1}^n b_i K_1\hat{x}_i$, $\mathbf{D} = 1$, and

$$J(\boldsymbol{a}\boldsymbol{g} + \boldsymbol{b} \odot \boldsymbol{Z}_n) = J(K_1(\boldsymbol{b}\boldsymbol{Z}_n)) =< (K_1\boldsymbol{b}\boldsymbol{Z}_n), K_1(\boldsymbol{b}\boldsymbol{Z}_n) >_{\mathbb{H}}$$

$$=< (\boldsymbol{b}K_1\boldsymbol{Z}_n), (\boldsymbol{b}K_1\boldsymbol{Z}_n) >_{\mathbb{H}} =< (\boldsymbol{b}\boldsymbol{Z}_n), (\boldsymbol{b}\boldsymbol{Z}_n) >_{\mathbb{H}} .$$

Below we evaluate $\partial < (\boldsymbol{b}\boldsymbol{Z}_n), (\boldsymbol{b}\boldsymbol{Z}_n) >_{\mathbb{H}} / \partial \boldsymbol{b}$. As

$$(\boldsymbol{b}\boldsymbol{Z}_n)(\boldsymbol{b}\boldsymbol{Z}_n) = \sum_{i=1}^n b_i^2 (K_1\hat{x}_i)^2 + \sum_{i \neq j} b_i b_j (K_1\hat{x}_i)(K_1\hat{x}_j),$$

thus

$$\frac{(\boldsymbol{b}\boldsymbol{Z}_n)(\boldsymbol{b}\boldsymbol{Z}_n)}{\partial b_i} = 2b_i(K_1\hat{x}_i)^2 + 2\sum_{i \neq j} b_j (K_1\hat{x}_i)(K_1\hat{x}_j) = 2(K_1\hat{x}_i)\sum_{j=1}^n b_j (K_1\hat{x}_j).$$

From this we get

$$\frac{\partial < (\boldsymbol{b}\boldsymbol{Z}_n), (\boldsymbol{b}\boldsymbol{Z}_n) >_{\mathbb{H}}}{\partial \boldsymbol{b}} = \Big(\frac{\partial < (\boldsymbol{b}\boldsymbol{Z}_n), (\boldsymbol{b}\boldsymbol{Z}_n) >_{\mathbb{H}}}{\partial b_1}, \ldots, \frac{\partial < (\boldsymbol{b}\boldsymbol{Z}_n), (\boldsymbol{b}\boldsymbol{Z}_n) >_{\mathbb{H}}}{\partial b_n}\Big) = \mathbf{q}, \quad \mathbf{q} = (q_1, \ldots, q_n),$$

where, $q_i = 2\sum_{j=1}^n < (K_1\hat{x}_i), b_j(K_1\hat{x}_j) >_{\mathbb{H}} = 2 < \boldsymbol{b}\boldsymbol{Z}_n, z_i >_{\mathbb{H}}$. Note $\partial||y - x(\boldsymbol{a}\boldsymbol{g} + \boldsymbol{b}\boldsymbol{Z}_n)||^2/\partial a_i$
$= -2 < y - x(\boldsymbol{a}\boldsymbol{g} + \boldsymbol{b}\boldsymbol{Z}_n), xg_i >$, or

$$\frac{\partial||y - x(\boldsymbol{a}\boldsymbol{g} + \boldsymbol{b}\boldsymbol{Z}_n)||^2}{\partial \mathbf{a}} = -2 < y - x(\boldsymbol{a}\boldsymbol{g} + \boldsymbol{b}\boldsymbol{Z}_n), x\mathbf{g}' > .$$

Further, $\partial\Big(x(\mathbf{b}\boldsymbol{Z}_n)\Big)/\partial b_i = xz_i$, or

$$\frac{\partial||y - x(\boldsymbol{a}\boldsymbol{g} + \boldsymbol{b}\boldsymbol{Z}_n)||^2}{\partial \mathbf{b}} = -2 < y - x(\boldsymbol{a}\boldsymbol{g} + \boldsymbol{b}\boldsymbol{Z}_n), x\mathbf{Z}_n' >,$$

where, by convention, $x\mathbf{Z}_n' = (xz_1, \ldots, xz_n)$, a n-dimensional row vector.

Rewrite (2) as

$$(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}) = \arg\inf_{(\boldsymbol{a},\boldsymbol{b})} G(\boldsymbol{a}, \boldsymbol{b}),$$

where $G(\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{n}\sum_{i=1}^n ||\hat{y}_i - \hat{x}_i(\boldsymbol{a}\boldsymbol{g} + \boldsymbol{b}\boldsymbol{Z}_n)||^2 + \lambda < (\boldsymbol{b}\boldsymbol{Z}_n), (\boldsymbol{b}\boldsymbol{Z}_n) >_{\mathbb{H}}$. $(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}})$ must satisfy

$$\begin{cases} \mathbf{0}_{1\times k} = \frac{\partial G(\boldsymbol{a},\boldsymbol{b})}{\partial \boldsymbol{a}} = -2\frac{1}{n}\sum_{i=1}^n < \hat{y}_i - \hat{x}_i(\boldsymbol{a}\boldsymbol{g} + \boldsymbol{b}\boldsymbol{Z}_n), \hat{x}_i\boldsymbol{g}' > \\ \mathbf{0}_{1\times n} = \frac{\partial G(\boldsymbol{a},\boldsymbol{b})}{\partial \boldsymbol{b}} = -2\left(\frac{1}{n}\sum_{i=1}^n < \hat{y}_i - \hat{x}_i(\boldsymbol{a}\boldsymbol{g} + \boldsymbol{b}\boldsymbol{Z}_n), \hat{x}_i\mathbf{Z}_n' > -\frac{\lambda}{2}\mathbf{q}\right) \end{cases}$$

or

$$\begin{cases} \frac{1}{n}\sum_{i=1}^n < \hat{y}_i\hat{x}_i, \boldsymbol{g} >= \frac{1}{n}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{a}\boldsymbol{g}, \boldsymbol{g} > + \frac{1}{n}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{b}\boldsymbol{Z}_n, \boldsymbol{g} > \\ \frac{1}{n}\sum_{i=1}^n < \hat{y}_i\hat{x}_i, \boldsymbol{Z}_n >= \frac{1}{n}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{a}\boldsymbol{g}, \boldsymbol{Z}_n > + \frac{1}{n}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{b}\boldsymbol{Z}_n, \boldsymbol{Z}_n > + \lambda < \boldsymbol{b}\boldsymbol{Z}_n, \boldsymbol{Z}_n >_{\mathbb{H}} \end{cases}.$$

It is easy to check that $n^{-1}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{a}\boldsymbol{g}, \boldsymbol{g} \geq n^{-1}\sum_{i=1}^n < \hat{x}_i^2, \boldsymbol{g}\boldsymbol{g}' > \boldsymbol{a}' := \mathbf{O}\boldsymbol{a}'$, $(\mathbf{O}_{k\times k})$, $n^{-1}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{b}\boldsymbol{Z}_n, \boldsymbol{g} \geq n^{-1}\sum_{i=1}^n < \hat{x}_i^2, \boldsymbol{g}\mathbf{Z}_n' > \boldsymbol{b}' := \mathbf{R}\boldsymbol{b}'$, $(\mathbf{R}_{k\times n})$, $n^{-1}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{a}\boldsymbol{g}, \boldsymbol{Z}_n \geq n^{-1}\sum_{i=1}^n < \hat{x}_i^2, \boldsymbol{Z}_n\boldsymbol{g}' > \boldsymbol{a}' = \mathbf{R}'\boldsymbol{a}'$, $n^{-1}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{b}\boldsymbol{Z}_n, \boldsymbol{Z}_n >= n^{-1}\sum_{i=1}^n < \hat{x}_i^2, \boldsymbol{Z}_n\mathbf{Z}_n' > \boldsymbol{b}' := \mathbf{S}\boldsymbol{b}'$, $(\mathbf{S}_{n\times n})$, and $< \boldsymbol{b}\boldsymbol{Z}_n, \boldsymbol{Z}_n >_{\mathbb{H}} =< \boldsymbol{Z}_n\mathbf{Z}_n' >_{\mathbb{H}} \boldsymbol{b}' := \mathbf{W}\boldsymbol{b}'$, $(\mathbf{W}_{n\times n})$. Denote $\frac{1}{n}\sum_{i=1}^n < \hat{y}_i\hat{x}_i, \boldsymbol{g} \geq \mathbf{u}$, $(\mathbf{u}_{k\times 1})$ and $\frac{1}{n}\sum_{i=1}^n < \hat{y}_i\hat{x}_i, \boldsymbol{Z}_n \geq \mathbf{v}$, $(\mathbf{v}_{n\times 1})$, then the above system of equations can be rewritten as

$$\begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{S} + \lambda\boldsymbol{W} \end{pmatrix}\begin{pmatrix} \boldsymbol{a}' \\ \boldsymbol{b}' \end{pmatrix} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}, \tag{A1}$$

or when the following inverse exists,

$$\begin{pmatrix} \hat{\boldsymbol{a}}' \\ \hat{\boldsymbol{b}}' \end{pmatrix} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{S} + \lambda\boldsymbol{W} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}.$$

□

Proof of Theorem 2 (one-dimensional case). In this case, $\hat{\boldsymbol{X}}_n = (\hat{x}_1, \ldots, \hat{x}_n)'$, $\hat{z}_n(\cdot) = n^{-1}\sum_{i=1}^n K_1(\hat{x}_i y_i)(\cdot)$, $\boldsymbol{a} = (a_1, \ldots, a_k)'$, $\boldsymbol{b} = b$, $\hat{\boldsymbol{\beta}}_{n,\lambda}(\cdot) = \hat{\beta}_{n,\lambda}(\cdot) = \hat{\boldsymbol{a}}'\boldsymbol{g}(\cdot) + \hat{b}\hat{z}_n(\cdot)$, and

$$(\hat{\boldsymbol{a}}, \hat{b}) = \arg\inf_{(\boldsymbol{a},b)}\left(\frac{1}{n}\sum_{i=1}^n ||\hat{y}_i - (\boldsymbol{a}'\boldsymbol{g} + b\hat{z}_n)\hat{x}_i||^2 + \lambda J(\boldsymbol{a}'\boldsymbol{g} + b\hat{z}_n)\right)$$

$$= \arg\inf_{(\boldsymbol{a},b)}\left(\frac{1}{n}\sum_{i=1}^n ||\hat{y}_i - (\boldsymbol{a}'\boldsymbol{g} + b\hat{z}_n)\hat{x}_i||^2 + \lambda b^2||\hat{z}_n||_{\mathbb{H}}^2\right) := G(\boldsymbol{a}, b).$$

As in the proof of Theorem 1 (one-dimensional case), $(\hat{\boldsymbol{a}}, \hat{b})$ must satisfy

$$\begin{cases} \mathbf{0}_{1\times k} = \frac{\partial G(\boldsymbol{a},b)}{\partial \boldsymbol{a}} = -2\frac{1}{n}\sum_{i=1}^n < \hat{y}_i - \hat{x}_i(\boldsymbol{a}'\boldsymbol{g} + b\hat{z}_n), \hat{x}_i\boldsymbol{g}' > \\ 0 = \frac{\partial G(\boldsymbol{a},b)}{\partial b} = -2\left(\frac{1}{n}\sum_{i=1}^n < \hat{y}_i - \hat{x}_i(\boldsymbol{a}'\boldsymbol{g} + b\hat{z}_n), \hat{x}_i\hat{z}_n > -\lambda b||\hat{z}_n||_{\mathbb{H}}^2\right) \end{cases},$$

or

$$\begin{cases} \frac{1}{n}\sum_{i=1}^n < \hat{y}_i\hat{x}_i, \boldsymbol{g} >= \frac{1}{n}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{a}'\boldsymbol{g}, \boldsymbol{g} > + b < \hat{x}_i^2\hat{z}_n, \boldsymbol{g} > \\ \frac{1}{n}\sum_{i=1}^n < \hat{y}_i\hat{x}_i, \hat{z}_n >= \frac{1}{n}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{a}'\boldsymbol{g}, \hat{z}_n > + b\left(\frac{1}{n}\sum_{i=1}^n < \hat{x}_i^2, \hat{z}_n^2 > + \lambda||\hat{z}_n||_{\mathbb{H}}^2\right) \end{cases}$$

It is easy to check that $n^{-1}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{a}'\boldsymbol{g}, \boldsymbol{g} \geq n^{-1}\sum_{i=1}^n < \hat{x}_i^2, \boldsymbol{g}\boldsymbol{g}' > \boldsymbol{a} := \mathbf{O}\boldsymbol{a}$, $(\mathbf{O}_{k\times k})$; $n^{-1}\sum_{i=1}^n < \hat{x}_i^2 b\hat{z}_n, \boldsymbol{g} \geq n^{-1}\sum_{i=1}^n < \hat{x}_i^2\hat{z}_n, \boldsymbol{g} > b := \mathbf{R}b$, $(\mathbf{R}_{k\times 1})$; $n^{-1}\sum_{i=1}^n < \hat{x}_i^2\boldsymbol{a}'\boldsymbol{g}, \hat{z}_n \geq n^{-1}\sum_{i=1}^n < \hat{x}_i^2\hat{z}_n, \boldsymbol{g}' > \boldsymbol{a} = \mathbf{R}'\boldsymbol{a}$; $n^{-1}\sum_{i=1}^n < \hat{x}_i^2 b\hat{z}_n, \hat{z}_n \geq n^{-1}\sum_{i=1}^n < \hat{x}_i^2\hat{z}_n, \hat{z}_n > b := Sb$;

and $< b\hat{z}_n, \hat{z}_n >_{\mathbb{H}} = < \hat{z}_n, \hat{z}_n >_{\mathbb{H}} b := Wb$. Denote $\frac{1}{n}\sum_{i=1}^n < \hat{y}_i\hat{x}_i, \boldsymbol{g} \geq \mathbf{u}$, $(\mathbf{u}_{k\times 1})$, and $\frac{1}{n}\sum_{i=1}^n < \hat{y}_i\hat{x}_i, \hat{z}_n \geq v$, the above system of equations is rewritten as

$$\begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{R}' & S + \lambda W \end{pmatrix} \begin{pmatrix} \boldsymbol{a} \\ b \end{pmatrix} = \begin{pmatrix} \boldsymbol{u} \\ v \end{pmatrix}, \tag{A2}$$

or when the following inverse exists,

$$\begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{R}' & S + \lambda W \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{u} \\ v \end{pmatrix}.$$

□

Proof of Theorem 1. We first simplify the penalty term $J(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n)$. By property of RKHS, $K(s,t) = < K(s,\cdot), K(t,\cdot) >_{\mathbb{H}}$, thus $\forall h \in \mathbb{H}$, $(K_1 h)(\cdot) := < K_1(\cdot,), h >_{\mathbb{H}} \in \mathbb{H}_1$ and $\forall h \in \mathbb{H}_1$, $(K_1 h) = h$. Thus

$$J(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n) = J(K_1(\boldsymbol{B} \odot \boldsymbol{Z}_n)) = < K_1(\boldsymbol{B} \odot \boldsymbol{Z}_n)'\mathbf{D}, K_1(\boldsymbol{B} \odot \boldsymbol{Z}_n) >_{\mathbb{H}}$$

$$= < (\boldsymbol{B} \odot K_1\boldsymbol{Z}_n)'\mathbf{D}, \boldsymbol{B} \odot K_1\boldsymbol{Z}_n >_{\mathbb{H}} = < (\boldsymbol{B} \odot \boldsymbol{Z}_n)'\mathbf{D}, \boldsymbol{B} \odot \boldsymbol{Z}_n >_{\mathbb{H}} .$$

Note that the inner product $< \cdot, \cdot >_{\mathbb{H}}$ of the RKHS is often not the inner product $< \cdot, \cdot >$ used in the optimization objective, such as the one corresponding to the L_2 norm. Thus, the above expression of $J(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n)$ does not hold under the inner product $< \cdot, \cdot >$.

Below we need to evaluate $\partial < (\boldsymbol{B} \odot \boldsymbol{Z}_n)'\mathbf{D}, \boldsymbol{B} \odot \boldsymbol{Z}_n >_{\mathbb{H}} / \partial \boldsymbol{B}$. For this, write $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{in})$ for the i-th row of $\boldsymbol{B}$ $(i = 1, \ldots, d)$, and $\boldsymbol{z}_i = (z_{1i}, \ldots z_{ni})'$ for the i-th column of $\boldsymbol{Z}_n$. Then

$$(\boldsymbol{B} \odot \boldsymbol{Z}_n)'\mathbf{D}(\boldsymbol{B} \odot \boldsymbol{Z}_n) = \sum_{i,r=1}^{d} d_{ir}(\boldsymbol{b}_i\boldsymbol{z}_i)(\boldsymbol{b}_r\boldsymbol{z}_r) = \sum_{i}^{d} d_{ii}(\boldsymbol{b}_i\boldsymbol{z}_i)^2 + \sum_{i=1}^{d}\sum_{r\neq i}^{d} d_{ir}(\boldsymbol{b}_i\boldsymbol{z}_i)(\boldsymbol{b}_r\boldsymbol{z}_r)$$

and we get, since $d_{ir} = d_{ri}$, and $\mathbf{b}_i\mathbf{z}_i = \sum_{j=1}^n b_{ij}z_{ji}$,

$$\frac{\partial\Big((\boldsymbol{B} \odot \boldsymbol{Z}_n)'\mathbf{D}(\boldsymbol{B} \odot \boldsymbol{Z}_n)\Big)}{\partial b_{ij}} = 2d_{ii}z_{ji}(\boldsymbol{b}_i\boldsymbol{z}_i) + \sum_{r\neq i}^{d} d_{ir}z_{ji}(\boldsymbol{b}_r\boldsymbol{z}_r) = d_{ii}z_{ji}(\boldsymbol{b}_i\boldsymbol{z}_i) + \sum_{r=1}^{d} d_{ir}z_{ji}(\boldsymbol{b}_r\boldsymbol{z}_r).$$

From this we get

$$\frac{\partial < (\boldsymbol{B} \odot \boldsymbol{Z}_n)'\mathbf{D}, (\boldsymbol{B} \odot \boldsymbol{Z}_n) >_{\mathbb{H}}}{\partial \boldsymbol{B}} = \mathbf{Q}, \quad \mathbf{Q} = (q_{ij})_{d\times n},$$

where $q_{ij} = d_{ii} < z_{ji}, (\boldsymbol{b}_i\boldsymbol{z}_i) >_{\mathbb{H}} + \sum_{r=1}^{d} d_{ir} < z_{ji}, (\boldsymbol{b}_r\boldsymbol{z}_r) >_{\mathbb{H}}$. Note $\partial||y - \boldsymbol{x}'(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n)||^2/\partial a_{ij} = -2 < y - \boldsymbol{x}'(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n), x_i g_j >$, or

$$\frac{\partial||y - \boldsymbol{x}'(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n)||^2}{\partial \mathbf{A}} = -2 < y - \boldsymbol{x}'(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n), \mathbf{x}\mathbf{g}' > .$$

Further, $\mathbf{x}'(\mathbf{B} \odot \boldsymbol{Z}_n) = \sum_{i=1}^{d} x_i(\boldsymbol{b}_i\boldsymbol{z}_i)$, and $\partial\Big(\mathbf{x}'(\mathbf{B} \odot \boldsymbol{Z}_n)\Big)/\partial b_{ij} = x_i z_{ji}$, or

$$\frac{\partial||y - \boldsymbol{x}'(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n)||^2}{\partial \mathbf{B}} = -2 < y - \boldsymbol{x}'(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B} \odot \boldsymbol{Z}_n), \mathbf{x}\mathbf{Z}_n' >,$$

where, by convention, $\mathbf{x}\mathbf{Z}_n'$ is the $d \times n$ matrix with (i,j)-th entry $x_i z_{ji}$.

Rewrite (2) as

$$(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}) = \arg\inf_{(\boldsymbol{A},\boldsymbol{B})} G(\boldsymbol{A}, \boldsymbol{B}),$$

where $G(\boldsymbol{A},\boldsymbol{B}) = \frac{1}{n}\sum_{i=1}^{n}||\hat{y}_i - \hat{\boldsymbol{x}}_i'(\boldsymbol{A}\boldsymbol{g}+\boldsymbol{B}\odot\boldsymbol{Z}_n)||^2 + \lambda < (\boldsymbol{B}\odot\boldsymbol{Z}_n)'\mathbf{D},(\boldsymbol{B}\odot\boldsymbol{Z}_n) >_{\mathbb{H}}$. $(\hat{\boldsymbol{A}},\hat{\boldsymbol{B}})$ must satisfy

$$\begin{cases} \mathbf{0}_{d\times k} = \frac{\partial G(\boldsymbol{A},\boldsymbol{B})}{\partial \boldsymbol{A}} = -2\frac{1}{n}\sum_{i=1}^{n} < \hat{y}_i - \hat{\boldsymbol{x}}_i'(\boldsymbol{A}\boldsymbol{g}+\boldsymbol{B}\odot\boldsymbol{Z}_n), \hat{\boldsymbol{x}}_i\boldsymbol{g}' > \\ \mathbf{0}_{d\times n} = \frac{\partial G(\boldsymbol{A},\boldsymbol{B})}{\partial \boldsymbol{B}} = -2\left(\frac{1}{n}\sum_{i=1}^{n} < \hat{y}_i - \hat{\boldsymbol{x}}_i'(\boldsymbol{A}\boldsymbol{g}+\boldsymbol{B}\odot\boldsymbol{Z}_n), \hat{\boldsymbol{x}}_i\boldsymbol{Z}_n' > -\frac{\lambda}{2}\mathbf{Q}\right) \end{cases}. \tag{A3}$$

To solve the linear system (A3), we need to rewrite it in terms of vector forms $\boldsymbol{a}$ and $\boldsymbol{b}$ of $\boldsymbol{A}$ and $\boldsymbol{B}$. For this, let $\boldsymbol{a} = (a_{11},\ldots,a_{1k},\ldots,a_{d,1},\ldots,a_{d,k})'$ be the vector representation of $\boldsymbol{A}$; $\boldsymbol{b} = (b_{11},\ldots,b_{1n},\ldots,b_{d,1},\ldots,b_{d,n})'$ be that of $\boldsymbol{B}$. For $\mathbf{x} = (x_1,\ldots,x_d)'$, $< \boldsymbol{x}'\boldsymbol{A}\boldsymbol{g}, \boldsymbol{x}\boldsymbol{g}' >$ is a $d\times k$ matrix with (i,j)-th entry $< \boldsymbol{x}'\boldsymbol{A}\boldsymbol{g}, x_i g_j \geq \sum_{r=1}^{d}\sum_{s=1}^{k} a_{rs} < x_r g_s, x_i g_j >$. Similarly, $n^{-1}\sum_{m=1}^{n} < \hat{\boldsymbol{x}}_m'\boldsymbol{A}\boldsymbol{g}, \hat{\boldsymbol{x}}_m\boldsymbol{g}' >$ is a $d\times k$ matrix with (i,j)-th entry $n^{-1}\sum_{r=1}^{d}\sum_{s=1}^{k} a_{rs}\sum_{m=1}^{n} < \hat{x}_{mr}g_s, \hat{x}_{mi}g_j >$; $n^{-1}\sum_{m=1}^{n} < \hat{\boldsymbol{x}}_m'(\boldsymbol{B}\odot\boldsymbol{Z}_n), \hat{\boldsymbol{x}}_m\boldsymbol{g}' >$ is a $d\times k$ matrix with (i,j)-th entry $n^{-1}\sum_{r=1}^{d}\sum_{s=1}^{n} b_{rs}\sum_{m=1}^{n} < \hat{x}_{mr}z_{sr}, \hat{x}_{mi}g_j >$; and $n^{-1}\sum_{m=1}^{n} < \hat{y}_m, \hat{\boldsymbol{x}}_m\boldsymbol{g}' >$ is a $d\times k$ matrix with (i,j)-th entry $n^{-1}\sum_{m=1}^{n} < \hat{y}_m, \hat{x}_{mi}g_j >$.

Likewise, $n^{-1}\sum_{m=1}^{n} < \hat{\boldsymbol{x}}_m'\boldsymbol{A}\boldsymbol{g}, \hat{\boldsymbol{x}}_m\boldsymbol{Z}_n' >$ is a $d\times n$ matrix with (i,j)-th entry $n^{-1}\sum_{r=1}^{d}\sum_{s=1}^{k} a_{rs}\sum_{m=1}^{n} < \hat{x}_{mr}g_s, \hat{x}_{mi}z_{ji} >$; $n^{-1}\sum_{m=1}^{n} < \hat{\boldsymbol{x}}_m'(\boldsymbol{B}\odot\boldsymbol{Z}_n), \hat{\boldsymbol{x}}_m\boldsymbol{Z}_n' >$ is a $d\times k$ matrix with (i,j)-th entry $n^{-1}\sum_{l=1}^{d}\sum_{r=1}^{n} b_{lr}\sum_{m=1}^{n} < \hat{x}_{ml}z_{rl}, \hat{x}_{mi}z_{ji} >$; and $n^{-1}\sum_{m=1}^{n} < \hat{y}_m, \hat{\boldsymbol{x}}_m\boldsymbol{Z}_n' >$ is a $d\times k$ matrix with (i,j)-th entry $n^{-1}\sum_{m=1}^{n} < \hat{y}_m, \hat{x}_{mi}z_{ji} >$.

Let the notation $< \boldsymbol{x}'\boldsymbol{A}\boldsymbol{g}, \boldsymbol{x}\boldsymbol{g}' > \sim \boldsymbol{O}\boldsymbol{a}$ means rearrange elements in the $d\times k$ matrix $< \boldsymbol{x}'\boldsymbol{A}\boldsymbol{g}, \boldsymbol{x}\boldsymbol{g}' >$ as a dk-vector in dictionary order in terms of its dk-vector $\boldsymbol{a}$ form. Thus,

$$n^{-1}\sum_{m=1}^{n} < \hat{\boldsymbol{x}}_m'\boldsymbol{A}\boldsymbol{g}, \hat{\boldsymbol{x}}_m\boldsymbol{g}' > \sim \boldsymbol{O}\boldsymbol{a}, \quad \boldsymbol{O}_{dk\times dk} = n^{-1}\sum_{i=1}^{n} < \boldsymbol{s}_i, \boldsymbol{s}_i' >,$$

where $\boldsymbol{s}_i = (\hat{x}_{i1}g_1,\ldots,\hat{x}_{i1}g_k,\ldots,\hat{x}_{i,d}g_1,\ldots,\hat{x}_{i,d}g_k)'$; Similarly,

$$n^{-1}\sum_{m=1}^{n} < \hat{\boldsymbol{x}}_m'(\boldsymbol{B}\odot\boldsymbol{Z}_n), \hat{\boldsymbol{x}}_m\boldsymbol{g}' > \sim \mathbf{R}\boldsymbol{b}, \quad \mathbf{R}_{dk\times dn} = n^{-1}\sum_{i=1}^{n} < \mathbf{s}_i, \mathbf{t}_i' >,$$

where $\boldsymbol{t}_i = (\hat{x}_{i1}\hat{z}_{11},\ldots,\hat{x}_{i1}\hat{z}_{n1},\ldots,\hat{x}_{id}\hat{z}_{11},\ldots,\hat{x}_{id}\hat{z}_{n1})'$; and

$$n^{-1}\sum_{m=1}^{n} < \hat{y}_m, \hat{\boldsymbol{x}}_m\boldsymbol{g}' > \sim \boldsymbol{u}, \quad \boldsymbol{u} = (u_{11},\ldots,u_{1k},\ldots,u_{d1},\ldots,u_{dk})',$$

where $u_{ij} = n^{-1}\sum_{m=1}^{n} < \hat{y}_m, \hat{x}_{mi}g_j >$.

Likewise,

$$n^{-1}\sum_{m=1}^{n} < \hat{\boldsymbol{x}}_m'\boldsymbol{A}\boldsymbol{g}, \hat{\boldsymbol{x}}_m\hat{\boldsymbol{Z}}_n' > \sim \boldsymbol{P}\boldsymbol{a}, \quad \boldsymbol{P}_{dn\times dk} = n^{-1}\sum_{i=1}^{n} < \boldsymbol{t}_i, \boldsymbol{s}_i' \geq \boldsymbol{R}';$$

$$n^{-1}\sum_{m=1}^{n} < \hat{\boldsymbol{x}}_m'(\boldsymbol{B}\odot\hat{\boldsymbol{Z}}_n), \hat{\boldsymbol{x}}_m\hat{\boldsymbol{Z}}_n' > \sim \boldsymbol{S}\boldsymbol{b}, \quad \boldsymbol{S}_{dn\times dn} = n^{-1}\sum_{i=1}^{n} < \boldsymbol{t}_i, \boldsymbol{t}_i' >;$$

and

$$n^{-1}\sum_{m=1}^{n} < \hat{y}_m, \hat{\boldsymbol{x}}_m\hat{\boldsymbol{Z}}_n' > \sim \mathbf{v}, \quad \boldsymbol{v} = (v_{11},\ldots,v_{1n},\ldots.,v_{d1},\ldots,v_{dn})',$$

where $v_{ij} = n^{-1}\sum_{m=1}^{n} < \hat{y}_m, \hat{x}_{mi}z_{ji} >$.

Rewrite q_{ij} as

$$q_{ij} = \sum_{s=1}^{n} b_{is}d_{ii} < \hat{z}_{ji}, \hat{z}_{si} >_{\mathbb{H}} + \sum_{r=1}^{d}\sum_{s=1}^{n} b_{rs}d_{ir} < \hat{z}_{ji}, \hat{z}_{sr} >_{\mathbb{H}}, \quad (1\leq i\leq d; 1\leq j\leq n).$$

Let $\boldsymbol{z} = (z_{11}, \ldots, z_{n1}, \ldots, z_{1d}, \ldots, z_{nd})'$, $\mathbf{1}$ be the $n \times n$ matrix of 1's, $\mathbf{D}_0 = diag\{d_{11}\mathbf{1}, \ldots, d_{dd}\mathbf{1}\}$, and

$$\mathbf{D}_1 = \begin{pmatrix} d_{11}\mathbf{1} & \cdots & d_{1d}\mathbf{1} \\ \ddots & \ddots & \ddots \\ d_{d1}\mathbf{1} & \cdots & d_{dd}\mathbf{1} \end{pmatrix}.$$

For any two matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ of the same dimension, denote $\mathbf{A} \otimes \mathbf{B} = (a_{ij}b_{ij})$. Let $\mathbf{W}_{dn\times dn} = (D_0 + D_1) \otimes < \boldsymbol{z}, \boldsymbol{z}' >_{\mathbb{H}}$. It it not difficult to check that

$$\mathbf{Q} =\sim \boldsymbol{W}\boldsymbol{b}.$$

Then (A1) is rewritten as

$$\begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{S} + \frac{\lambda}{2}\boldsymbol{W} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}, \tag{A4}$$

or when the following inverse exists,

$$\begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{S} + \frac{\lambda}{2}\boldsymbol{W} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}.$$

□

Proof of Theorem 2. In this case, $\hat{\boldsymbol{z}}_n = (\hat{z}_1, \ldots, \hat{z}_d)'$ is a d-vector and, similar to the proof of Theorem 1, we have $J(\boldsymbol{A}\mathbf{g} + \boldsymbol{B}\hat{\boldsymbol{z}}_n) =< \hat{\boldsymbol{z}}_n'\boldsymbol{B}'D, \boldsymbol{B}\hat{\boldsymbol{z}}_n >_{\mathbb{H}}$. To evaluate $\partial < \hat{\boldsymbol{z}}_n'\boldsymbol{B}'D, \boldsymbol{B}\hat{\boldsymbol{z}}_n >_{\mathbb{H}} / \partial \boldsymbol{B}$, write $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$, where $\boldsymbol{b}_j = (b_{1j}, \ldots, b_{dj})$ is the j-th column of $\boldsymbol{B}$. Then $\boldsymbol{B}\hat{\boldsymbol{z}}_n = \sum_{j=1}^{d} z_j\boldsymbol{b}_j$, and

$$\hat{\boldsymbol{z}}_n'\boldsymbol{B}'D\boldsymbol{B}\hat{\boldsymbol{z}}_n = \sum_{j=1}^{d}\left(\hat{z}_j^2\boldsymbol{b}_j'D\boldsymbol{b}_j + 2\sum_{l\neq j}^{d}\hat{z}_j\boldsymbol{b}_j'D\boldsymbol{b}_l\hat{z}_l\right)$$

$$= \sum_{j=1}^{d}\left(\hat{z}_j^2\sum_{i=1}^{d}\left(b_{ij}^2d_{ii} + 2\sum_{k\neq i}^{d}b_{ij}d_{ik}b_{kj} + \sum_{k\neq i}^{d}\sum_{l\neq i}^{d}b_{kj}d_{kl}b_{lj}\right) + 2\sum_{l\neq j}^{d}\hat{z}_j\Big(\sum_{r,s=1}^{d}b_{rj}d_{rs}b_{sl}\Big)\hat{z}_l\right),$$

we get, since $d_{ij} = d_{ji}$,

$$\frac{\partial\left(\hat{\boldsymbol{z}}_n'\boldsymbol{B}'D\boldsymbol{B}\hat{\boldsymbol{z}}_n\right)}{\partial b_{ij}} = 2\hat{z}_j\left(d_{ii}b_{ij}\hat{z}_j + \sum_{k\neq i}^{d}d_{ik}b_{kj}\hat{z}_j + \sum_{l\neq j}^{d}\sum_{s=1}^{d}d_{is}b_{sl}\hat{z}_l\right)$$

$$= 2\hat{z}_j\sum_{l=1}^{d}\sum_{s=1}^{d}d_{is}b_{sl}\hat{z}_l = 2\hat{z}_j\boldsymbol{d}_i\boldsymbol{B}\hat{\boldsymbol{z}}_n,$$

where $\boldsymbol{d}_i = (d_{i1}, \ldots, d_{id})$ is the i-th row of D. From this we get

$$\frac{\partial < \hat{\boldsymbol{z}}'\boldsymbol{B}'D, \boldsymbol{B}\hat{\boldsymbol{z}}_n >_{\mathbb{H}}}{\partial \boldsymbol{B}} = 2\mathbf{Q}, \quad \mathbf{Q} = (q_{ij})_{d\times d}, \quad q_{ij} =< \boldsymbol{d}_i\boldsymbol{B}\hat{\boldsymbol{z}}_n, \hat{z}_j >_{\mathbb{H}} = \boldsymbol{d}_i\boldsymbol{B} < \hat{\boldsymbol{z}}_n, \hat{z}_j >_{\mathbb{H}}$$

or

$$\frac{\partial < \hat{\boldsymbol{z}}_n'\boldsymbol{B}'D\boldsymbol{B}, \hat{\boldsymbol{z}}_n >_{\mathbb{H}}}{\partial \boldsymbol{B}} = 2D\boldsymbol{B} < \hat{\boldsymbol{z}}_n, \hat{\boldsymbol{z}}_n' >_{\mathbb{H}}.$$

Now (3) is rewritten as

$$(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}) = \arg\inf_{(\boldsymbol{A},\boldsymbol{B})} G(\boldsymbol{A}, \boldsymbol{B}),$$

where $G(\boldsymbol{A},\boldsymbol{B}) = \frac{1}{n}\sum_{i=1}^{n}||y_i - (\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B}\hat{\boldsymbol{z}}_n)'\hat{\boldsymbol{x}}_i||^2 + \lambda < \hat{\boldsymbol{z}}_n'\boldsymbol{B}'D\boldsymbol{B}, \hat{\boldsymbol{z}}_n >_{\mathbb{H}}$, and $(\hat{\boldsymbol{A}},\hat{\boldsymbol{B}})$ must satisfy

$$\begin{cases} \mathbf{0}_{d\times k} = \frac{\partial G(\boldsymbol{A},\boldsymbol{B})}{\partial \boldsymbol{A}} = -2\frac{1}{n}\sum_{i=1}^{n} < \hat{y}_i - (\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B}\hat{\boldsymbol{z}}_n)'\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_i\boldsymbol{g}' > \\ \mathbf{0}_{d\times d} = \frac{\partial G(\boldsymbol{A},\boldsymbol{B})}{\partial \boldsymbol{B}} = -2\left(\frac{1}{n}\sum_{i=1}^{n} < \hat{y}_i - (\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B}\hat{\boldsymbol{z}}_n)'\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_i\hat{\boldsymbol{z}}_n' > -\lambda D\boldsymbol{B} < \hat{\boldsymbol{z}}_n, \hat{\boldsymbol{z}}_n' >_{\mathbb{H}}\right)' \end{cases}$$

or

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} < \boldsymbol{x}_i'(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B}\hat{\boldsymbol{z}}_n), \hat{\boldsymbol{x}}_i\boldsymbol{g}' > = \frac{1}{n}\sum_{i=1}^{n} < \hat{y}_i, \hat{\boldsymbol{x}}_i\boldsymbol{g}' > \\ \frac{1}{n}\sum_{i=1}^{n} < \hat{\boldsymbol{x}}_i'(\boldsymbol{A}\boldsymbol{g} + \boldsymbol{B}\hat{\boldsymbol{z}}_n), \hat{\boldsymbol{x}}_i\hat{\boldsymbol{z}}_n' > +\lambda D\boldsymbol{B} < \hat{\boldsymbol{z}}_n, \hat{\boldsymbol{z}}_n' >_{\mathbb{H}} = \frac{1}{n}\sum_{i=1}^{n} < \hat{y}_i, \hat{\boldsymbol{x}}_i\hat{\boldsymbol{z}}_n' > \end{cases}. \quad \text{(A5)}$$

Let $(\hat{\boldsymbol{A}},\hat{\boldsymbol{B}})$ be the solution of (A5).

To solve the linear system (A5), we need to rewrite it in terms of vector forms $\boldsymbol{a}$ and $\boldsymbol{b}$ of $\boldsymbol{A}$ and $\boldsymbol{B}$. For this, let $\boldsymbol{a} = (a_{11},\ldots,a_{1k},\ldots,a_{d1},\ldots,a_{dk})'$ be the vector representation of $\boldsymbol{A}$; let $\boldsymbol{b} = (b_{11},\ldots,b_{1d},\ldots,b_{d1},\ldots,b_{dd})'$ be that of $\boldsymbol{B}$.

Let the notation $< \boldsymbol{x}'\boldsymbol{A}\boldsymbol{g}, \boldsymbol{x}\boldsymbol{g}' > \sim \boldsymbol{O}\boldsymbol{a}$ mean rearranging the elements in the matrix $< \boldsymbol{x}'\boldsymbol{A}\boldsymbol{g}, \boldsymbol{x}\boldsymbol{g}' >$ in terms of its vetor $\boldsymbol{a}$ form. As in the proof of Theorem 1,

$$n^{-1}\sum_{m=1}^{n} < \boldsymbol{x}_m'\boldsymbol{A}\boldsymbol{g}, \hat{\boldsymbol{x}}_m\boldsymbol{g}' > \sim \boldsymbol{O}\boldsymbol{a}, \quad \boldsymbol{O}_{dk\times dk} = n^{-1}\sum_{i=1}^{n} < \boldsymbol{s}_i, \boldsymbol{s}_i' >,$$

where $\boldsymbol{s}_i = (\hat{x}_{i1}g_1,\ldots,\hat{x}_{i1}g_k,\ldots,\hat{x}_{id}g_1,\ldots,\hat{x}_{id}g_k)'$.

Similarly,

$$n^{-1}\sum_{m=1}^{n} < \hat{\boldsymbol{x}}_m'\boldsymbol{A}\boldsymbol{g}, \hat{\boldsymbol{x}}_m\hat{\boldsymbol{z}}_n' > \sim \boldsymbol{P}\boldsymbol{a}, \quad \boldsymbol{P}_{d^2\times dk} = n^{-1}\sum_{i=1}^{n} < \boldsymbol{t}_i, \boldsymbol{s}_i' >,$$

where $\boldsymbol{t}_i = (\hat{x}_{i1}\hat{z}_1,\ldots,\hat{x}_{i1}\hat{z}_d,\ldots,\hat{x}_{id}\hat{z}_1,\ldots,\hat{x}_{id}\hat{z}_d)'$;

$$n^{-1}\sum_{m=1}^{n} < \hat{\boldsymbol{x}}_m'\boldsymbol{B}\hat{\boldsymbol{z}}_n, \hat{\boldsymbol{x}}_m\boldsymbol{g}' > \sim \boldsymbol{R}\boldsymbol{b}, \quad \boldsymbol{R}_{dk\times d^2} = n^{-1}\sum_{i=1}^{n} < \boldsymbol{s}_i, \boldsymbol{t}_i' \geq \mathbf{P}';$$

and

$$n^{-1}\sum_{m=1}^{n} < \boldsymbol{x}_m'\boldsymbol{B}\hat{\boldsymbol{z}}_n, \hat{\boldsymbol{x}}_m\hat{\boldsymbol{z}}_n' > \sim \boldsymbol{S}\boldsymbol{b}, \quad \boldsymbol{S}_{d^2\times d^2} = n^{-1}\sum_{i=1}^{n} < \boldsymbol{t}_i, \boldsymbol{t}_i' > .$$

Denote $\boldsymbol{U} = n^{-1}\sum_{i=1}^{n} < \hat{y}_i, \hat{\boldsymbol{x}}_i\boldsymbol{g}' \geq (u_{ij})_{d\times k}$ and its vector form $\boldsymbol{u} = (u_{11},\ldots,u_{1k},\ldots,u_{d1},\ldots,u_{dk})'$; let $\boldsymbol{V} = n^{-1}\sum_{i=1}^{n} < \hat{y}_i, \hat{\boldsymbol{x}}_i\hat{\boldsymbol{z}}_n' \geq (v_{ij})_{d\times d}$ and its vector form $\boldsymbol{v} = (v_{11},\ldots,v_{1d},\ldots,v_{d1},\ldots,v_{dd})'$; since D is semipositive definite, let $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$ be its eigenvalues, $\Lambda = diag(\lambda_1,\ldots,\lambda_d)$ and $\boldsymbol{q}_1,\ldots,\boldsymbol{q}_d$ be its normalized eigenvectors, $\boldsymbol{Q} = (\boldsymbol{q}_1,\ldots,\boldsymbol{q}_d)$, then $D = \boldsymbol{Q}\Lambda\boldsymbol{Q}' = \sum_{j=1}^{d}\lambda_j\boldsymbol{q}_j\boldsymbol{q}_j'$. Rearranging elements of $D\boldsymbol{B} < \hat{\boldsymbol{z}}_n, \hat{\boldsymbol{z}}_n' >_{\mathbb{H}}$ in vector form similarly as before

$$D\boldsymbol{B} < \hat{\boldsymbol{z}}_n, \hat{\boldsymbol{z}}_n' >_{\mathbb{H}} = \sum_{j=1}^{d}\lambda_j < \boldsymbol{q}_j'\boldsymbol{B}\hat{\boldsymbol{z}}_n, \boldsymbol{q}_j\hat{\boldsymbol{z}}_n' >_{\mathbb{H}} \sim \boldsymbol{W}\boldsymbol{b}, \quad \boldsymbol{W}_{d^2\times d^2} = \sum_{j=1}^{d} < \boldsymbol{c}_j, \boldsymbol{c}_j' >_{\mathbb{H}},$$

where $\boldsymbol{c}_j = \sqrt{\lambda_j}(q_{j1}\hat{z}_1,\ldots,q_{j1}\hat{z}_d,\ldots,q_{jd}\hat{z}_1,\ldots,q_{jd}\hat{z}_d)'$.

Then (A5) is rewritten as

$$\begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{S} + \lambda\boldsymbol{W} \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}, \quad \text{(A6)}$$

or when the following inverse exists,

$$\begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{S}+\lambda \boldsymbol{W} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}.$$

□

Proof of Theorem 3. Note that

$$\hat{\boldsymbol{z}}_n(\cdot) = n^{-1}\sum_{i=1}^{n}[K_1(\hat{\boldsymbol{x}}_i\hat{y}_i)](\cdot) = n^{-1}\sum_{i=1}^{n}[K_1(\boldsymbol{x}_i y_i)](\cdot)$$

$$+n^{-1}\sum_{i=1}^{n}[K_1(\hat{\boldsymbol{x}}_i\hat{y}_i - \boldsymbol{x}_i y_i)](\cdot) := \boldsymbol{x}_n(\cdot) + \boldsymbol{r}_n(\cdot).$$

Note that (C3) implies $E||\boldsymbol{xy}|| < \infty$ and $E||K_1(\boldsymbol{xy})|| < \infty$, so by Theorem 7.9 (or Corollary 7.10) in Ledoux and Talagrand [32], $||\boldsymbol{z}_n - \boldsymbol{z}_0|| \to 0$ (a.s.), where $\boldsymbol{z}_0(\cdot) = E[K_1(\boldsymbol{xy})](\cdot)$. By (C4), $||\boldsymbol{r}_n|| \to 0$ (a.s.). Thus, $||\hat{\boldsymbol{z}}_n - \boldsymbol{z}_0|| \to 0$ (a.s.).

Let $\boldsymbol{C} = (\boldsymbol{A}, \boldsymbol{B})$, $\hat{\boldsymbol{C}} = (\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}})$, $m(\boldsymbol{C}) = ||y - (\boldsymbol{Ag} + \boldsymbol{Bz}_0)'\boldsymbol{x}||^2$, $Pm(\boldsymbol{C}) = E[m(\boldsymbol{C})]$, $\mathbb{P}_n$ is the empirical distribution based on n iid samples from $m(\boldsymbol{C})$. Let

$$\mathbb{M}_n(\boldsymbol{C}) = \frac{1}{n}\sum_{i=1}^{n}||\hat{y}_i - (\boldsymbol{Ag} + \boldsymbol{B}\hat{\boldsymbol{z}}_n)'\hat{\boldsymbol{x}}_i||^2 + \lambda J(\boldsymbol{Ag} + \boldsymbol{B}\hat{\boldsymbol{z}}_n).$$

By (C5) and (C4) and the fact $||\hat{\boldsymbol{z}}_n - \boldsymbol{z}_0|| \to 0$ (a.s.),

$$\mathbb{M}_n(\boldsymbol{C}) = \frac{1}{n}\sum_{i=1}^{n}||y_i - (\boldsymbol{Ag} + \boldsymbol{Bz}_0)'\boldsymbol{x}_i||^2 + \lambda J(\boldsymbol{Ag} + \boldsymbol{Bz}_0) + o(1)$$

$$= \frac{1}{n}\sum_{i=1}^{n}||y_i - (\boldsymbol{Ag} + \boldsymbol{Bz}_0)'\boldsymbol{x}_i||^2 + o(1) := \mathbb{P}_n m(\boldsymbol{C}) + o(1) = Pm(\boldsymbol{C}) + o(1), \quad (a.s.). \quad \text{(A7)}$$

In the above we used Theorem 7.9 (or Corollary 7.10) in Ledoux and Talagrand [32] again to get $\mathbb{P}_n m(\boldsymbol{C}) = Pm(\boldsymbol{C}) + o(1)$ (a.s.).

Note that $E||\boldsymbol{xy}|| < \infty$ implies $E(||\boldsymbol{xy}|||\boldsymbol{x}) < \infty$, this together with (C3) implies that $\inf_{\boldsymbol{C}} Pm(\boldsymbol{C}) = E\big(\inf_{\boldsymbol{C}} E[m(\boldsymbol{C})|\boldsymbol{x}]\big)$ has an unique (and finite) minimizer $\boldsymbol{C}_0 = (\boldsymbol{A}_0, \boldsymbol{B}_0)$. We first prove $||\hat{\boldsymbol{C}} - \boldsymbol{C}_0|| \to 0$ (a.s.).

By definition of $\hat{\boldsymbol{C}}$, $\mathbb{M}_n(\hat{\boldsymbol{C}}) \le \mathbb{M}_n(\boldsymbol{C}_0) = Pm(\boldsymbol{C}_0) + o(1)$ (a.s.), and by (A7), $Pm(\hat{\boldsymbol{C}}) \le \mathbb{P}_n m(\boldsymbol{C}_0) + o(1)$ (a.s.). Thus,

$$Pm(\hat{\boldsymbol{C}}) - Pm(\boldsymbol{C}_0) \le \mathbb{P}_n m(\boldsymbol{C}_0) - Pm(\boldsymbol{C}_0) + o(1)$$

$$\le \sup_{\boldsymbol{C}\in\mathbb{C}}|\mathbb{P}_n m(\boldsymbol{C}) - Pm(\boldsymbol{C})| + o(1) \to 0 \quad (a.s.), \quad \text{(A8)}$$

where $\mathbb{C}$ is some bounded set of $\boldsymbol{C}$'s, and we used the fact that $\{\mathbb{P}_n m(\boldsymbol{C}) : \boldsymbol{C} \in \mathbb{C}\}$ is a Glivenko–Cantelli class on any bounded $\mathbb{C}$. Thus $\sup_{\boldsymbol{C}\in\mathbb{C}}|\mathbb{P}_n m(\boldsymbol{C}) - Pm(\boldsymbol{C})| \to 0$ (a.s.).

On the other hand, since $\boldsymbol{C}_0$ is the unique minimizer of $Pm(\boldsymbol{C})$, for every $\delta > 0$, there is $\eta > 0$, such that

$$\inf_{\boldsymbol{C}:||\boldsymbol{C}-\boldsymbol{C}_0||\ge\delta} Pm(\boldsymbol{C}) > Pm(\boldsymbol{C}_0) + \eta.$$

Thus, by (A8) we must have that for all large n, $||\hat{\boldsymbol{C}} - \boldsymbol{C}_0|| < \delta$ (a.s.) for every $\delta > 0$. This gives $||\hat{\boldsymbol{C}} - \boldsymbol{C}_0|| \to 0$ (a.s.).

Note that $E_{\beta_0}(y|\boldsymbol{x}) = \boldsymbol{\beta}_0'\boldsymbol{x}$, which is the minimizer of the conditional expectation $E_{\beta_0}(||y - \boldsymbol{\beta}_0'\boldsymbol{x}||^2|\boldsymbol{x})$, and $\boldsymbol{\beta}_0$ is also the pointwise least squares "estimate" of itself under the objective functional $E_{\beta_0}\{||y - \boldsymbol{\beta}_0'\boldsymbol{x}||^2\} = E\{E_{\beta_0}(||y - \boldsymbol{\beta}_0'\boldsymbol{x}||^2|\boldsymbol{x})\}$, so by (C1), $\boldsymbol{\beta}_0 = [E(\boldsymbol{xx}')]^{-1}E(\boldsymbol{xy}) \in Span(E(\boldsymbol{xy})) = Span(E[K_0(\boldsymbol{xy})], E[K_1(\boldsymbol{xy})]) \subset Span(\boldsymbol{g}, \boldsymbol{z}_0)$, (C2) im-

plies $E[\boldsymbol{x}(\cdot)\boldsymbol{x}(\cdot)']$ is invertible, and so $\boldsymbol{\theta}_0$ can be written in the form $((\boldsymbol{Ag})', (\boldsymbol{Bz}_0)')'$. Since $\boldsymbol{C}_0 = (\boldsymbol{A}_0, \boldsymbol{B}_0)$ also minimizes $Pm(\boldsymbol{C})$ (over a larger space than that $\boldsymbol{\theta}_0$ belongs to), we must have $((\boldsymbol{A}_0\boldsymbol{g})', (\boldsymbol{B}_0\boldsymbol{z}_0)')' = \boldsymbol{\beta}_0$, and $\hat{\boldsymbol{C}} = (\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}) \to (\boldsymbol{A}_0, \boldsymbol{B}_0)$ (a.s.) gives $\hat{\boldsymbol{\beta}}_{n,\lambda} = \left((\hat{\boldsymbol{A}}\boldsymbol{g})', (\hat{\boldsymbol{B}}\hat{\boldsymbol{z}}_n)'\right)' \to \left((\boldsymbol{A}_0\boldsymbol{g})', (\boldsymbol{B}_0\boldsymbol{z}_0)'\right)' = \boldsymbol{\beta}_0$ (a.s.). □

Proof of Theorem 4. Recall the blockwise inversion formula

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}$$

and for $\lambda \to 0$, $(A + \lambda W)^{-1} = A^{-1} - \lambda A^{-1}WA^{-1} + O(\lambda^2) = A^{-1} - O(\lambda)$.

By (C2) and (C3), for all large n, $\boldsymbol{O}^{-1}$, $\boldsymbol{P}^{-1}$, $\boldsymbol{R}^{-1}$, $\boldsymbol{S}^{-1}$ and $\boldsymbol{W}^{-1}$ all exist (a.s.). Using the above blockwise inversion formulae, by Theorem 2, we get

$$\begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{R} \\ \boldsymbol{P} & \boldsymbol{S} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} - O(\lambda) \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}.$$

In the proof of Theorem 3, we showed $||\hat{\boldsymbol{C}} - \boldsymbol{C}_0|| \to 0$ (a.s.), i.e., $(\hat{\boldsymbol{a}}', \hat{\boldsymbol{b}}')' \to (\boldsymbol{a}_0, \boldsymbol{b}_0)$ (a.s.). Further, similar to the proof of Theorem 3, we can get

$$\boldsymbol{O} \overset{a.s.}{\to} \boldsymbol{O}_0 = E < \boldsymbol{s}_1, \boldsymbol{s}_1' >, \quad \boldsymbol{P} = \boldsymbol{R}' \overset{a.s.}{\to} \boldsymbol{P}_0 = E < \boldsymbol{t}_1, \boldsymbol{s}_1' >, \quad \boldsymbol{S} \overset{a.s.}{\to} \boldsymbol{S}_0 = E < \boldsymbol{t}_1, \boldsymbol{t}_1' >.$$

$$\boldsymbol{U} \overset{a.s.}{\to} \boldsymbol{U}_0 = E < y_1, \boldsymbol{x}_1 g' >, \quad \boldsymbol{V} \overset{a.s.}{\to} \boldsymbol{V}_0 = E < y_1, \boldsymbol{x}_1 \boldsymbol{z}_0' >.$$

Let $\boldsymbol{u}_0$ and $\boldsymbol{v}_0$ be the vector representations of $\boldsymbol{U}_0$ and $\boldsymbol{V}_0$, then we have

$$\begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{pmatrix} \overset{a.s.}{\to} \begin{pmatrix} \boldsymbol{O}_0 & \boldsymbol{R}_0 \\ \boldsymbol{P}_0 & \boldsymbol{S}_0 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{u}_0 \\ \boldsymbol{v}_0 \end{pmatrix} := \begin{pmatrix} \boldsymbol{a}_0 \\ \boldsymbol{b}_0 \end{pmatrix}.$$

Denote $\hat{\boldsymbol{c}} = (\hat{\boldsymbol{a}}', \hat{\boldsymbol{b}})'$ and $\boldsymbol{c}_0 = (\boldsymbol{a}_0', \boldsymbol{b}_0')'$, we first find the asymptotic distribution of $\hat{\boldsymbol{c}}$. Denote $\boldsymbol{T} = matrix(\boldsymbol{O}, \boldsymbol{R}; \boldsymbol{P}, \boldsymbol{S})$, $\boldsymbol{T}_0 = matrix(\boldsymbol{O}_0, \boldsymbol{R}_0; \boldsymbol{P}_0, \boldsymbol{S}_0)$, $\boldsymbol{w} = (\boldsymbol{u}', \boldsymbol{v}')'$ and $\boldsymbol{w}_0 = (\boldsymbol{u}_0', \boldsymbol{v}_0')'$, then $\boldsymbol{c}_0 = \boldsymbol{T}_0^{-1}\boldsymbol{w}_0$, and $\hat{\boldsymbol{c}} = \boldsymbol{T}^{-1}\boldsymbol{w}$. By (C6),

$$\sqrt{n}(\hat{\boldsymbol{c}} - \boldsymbol{c}_0) = \sqrt{n}(\boldsymbol{T}_0^{-1} + o_p(1))(\boldsymbol{w} - \boldsymbol{w}_0) + o(1).$$

It can be shown that the sequences $\{\hat{y}_i, \hat{\boldsymbol{x}}_i\boldsymbol{g}'\}$ and $\{\hat{y}_i, \hat{\boldsymbol{x}}_i\hat{\boldsymbol{z}}_n'\}$ are Donsker classes, and so

$$\sqrt{n}(\boldsymbol{w} - \boldsymbol{w}_0) \overset{D}{\to} N(\boldsymbol{0}, \Gamma), \quad \Gamma = (\gamma_{ij})_{d(d+k)\times d(d+k)}, \quad \gamma_{ij} = Cov(\tilde{w}_i, \tilde{w}_j),$$

where $\tilde{\boldsymbol{w}}(\cdot) = (\tilde{\boldsymbol{u}}', \tilde{v}')'$, $\tilde{\boldsymbol{u}}$ is the vector form of $\tilde{U} =< y_1, \boldsymbol{x}_1 g' >$ and $\tilde{\boldsymbol{v}}$ is the vector form of $\tilde{V} =< y_1, \boldsymbol{x}_1 \boldsymbol{z}_0' >$. From the above we get, as $\boldsymbol{T}_0$ is symmetric,

$$\sqrt{n}(\hat{\boldsymbol{c}} - \boldsymbol{c}_0 - o_p(1)) \overset{D}{\to} N(\boldsymbol{0}, \boldsymbol{T}_0^{-1}\Gamma\boldsymbol{T}_0^{-1}). \tag{A9}$$

Now, rewrite $\hat{\boldsymbol{\beta}}_{n,\lambda}(\cdot) = \boldsymbol{F}_n(\cdot)\hat{\boldsymbol{c}}$, with $\boldsymbol{F}_n = (\boldsymbol{g}', \boldsymbol{J}_n)$, and $\boldsymbol{F}_0 = (\boldsymbol{g}', \boldsymbol{J}_0)$, where $\boldsymbol{J}_n = (\hat{\boldsymbol{z}}_n, \ldots, \hat{\boldsymbol{z}}_n)$, and $\boldsymbol{J}_0 = (\boldsymbol{Z}_0, \ldots, \boldsymbol{Z}_0)$. Then $F_n(t) = F_0(t) + o_p(r_n^{-1/2}(t))$, and by (A9) we get

$$\mathbb{W}_n = \sqrt{n}(\hat{\boldsymbol{\beta}}_{n,\lambda}(\cdot) - \boldsymbol{\beta}_0(\cdot) - o_p(1)) = \sqrt{n}\boldsymbol{F}_0(\cdot)(\hat{\boldsymbol{c}} - \boldsymbol{c}_0 - o_p(1)) \overset{D}{\Rightarrow} \mathbb{W}, \quad \text{in} \quad [l^\infty(T)]^d,$$

where $\mathbb{W}$ is a mean zero Gaussian process on T with covariance function $\sigma(s,t) = E(\mathbb{W}(s), \mathbb{W}(t)) = \boldsymbol{F}_0(s)\boldsymbol{T}_0^{-1}\Gamma\boldsymbol{T}_0^{-1}\boldsymbol{F}_0'(t)$. □

References

1. Ullah, S.; Finch, C.F. Applications of functional data analysis: A systematic review. *BMC Med. Res. Methodol.* **2013**, *13*, 43. [CrossRef]
2. Horváth, L.; Kokoszka, P. *Inference for Functional Data with Applications*; Springer: New York, NY, USA, 2012.
3. Yuan, A.; Fang, H.B.; Li, H.; Wu, C.O.; Tan, M. Hypothesis Testing for Multiple Mean and Correlation Curves with Functional Data. *Stat. Sin.* **2020**, *30*, 1095–1116. [CrossRef]
4. Lai, T.Y.; Zhang, Z.Z.; Wang, Y.F. Testing Independence and Goodness-of-Fit Jointly for Functional Linear Models. *J. Korean Statsitical Soc.* **2021**, *50*, 380–402. [CrossRef]
5. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*; Springer: New York, NY, USA, 2005.
6. Clarkson, D.B.; Fraley, C.; Gu, C.; Ramsay, J.O. *S+ Functional Data Analysis*; Springer: New York, NY, USA, 2005.
7. Ferraty, F.; Vieu, P. *Nonparametric Fuctional Data Analysis*; Springer: New York, NY, USA, 2006.
8. Zeger, S.L.; Diggle, P.J. Semiparametric Models for Longitudinal Data with Application to CD4 Cell Numbers in HIV Seroconverters. *Biometrics* **1994**, *50*, 689–699. [CrossRef] [PubMed]
9. Lin, D.Y.; Ying, Z. Semiparametric and Nonparametric Regression Analysis of Longitudinal Data. *J. Am. Stat. Assoc.* **2001**, *96*, 103–126. [CrossRef]
10. Fan, J.; Li, R. New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis. *J. Am. Stat. Assoc.* **2004**, *99*, 710–723. [CrossRef]
11. Xue, L.; Zhu, L. Empirical Likelihood Semiparametric Regression Analysis for Longitudinal Data. *Biometrika* **2007**, *94*, 921–937. [CrossRef]
12. Yuan, M.; Cai, T. A Reproducing Kernel Hilbert Space Approach to Functional Linear Regression. *Ann. Stat.* **2010**, *38*, 3412–3444. [CrossRef]
13. Reiss, P.T.; Goldsmith, J.; Shang, H.L.; Ogden, R.T. Methods for Scalar-on-Function Regression. *Inte. Stat. Rev.* **2017**, *85*, 228–249. [CrossRef]
14. Chen, C.; Guo, S.J.; Qian, X.H. Functional Linear Regression: Dependence and Error Contamination. *J. Bus. Econ. Stat.* **2022**, *40*, 444–457. [CrossRef]
15. Yao, F.; Müller, H.G.; Wang, J.L. Functional Linear Regression Analysis for Longitudinal Data. *Ann. Stat.* **2005**, *33*, 2873–2903. [CrossRef]
16. Müller, H.G.; Yao, F. Functional Additive Models. *J. Am. Stat. Assoc.* **2008**, *103*, 1534–1544. [CrossRef]
17. Kramer, N.; Boulesteix, A.L.; Tutz, G. Penalized Partial Least Squares with Applications to B-spline Transformations and Functional Data. *Chem. Intell. Lab. Syst.* **2008**, *94*, 60–69. [CrossRef]
18. Hayashi, K.; Hayashi, M.; Reich, B.; Lee, S.P.; Sachdeva, A.U.C.; Mizoguchi, I. Functional Data Analysis of Mandibular Movement Using Third-degree B-Spline Basis Functions and Self-modeling Regression. *Orthod. Waves* **2012**, *71*, 17–25. [CrossRef]
19. Aguilera, A.M.; Aguilera-Morillo, M.C. Penalized PCA Approaches for B-spline Expansions of Smooth Functional Data. *Appl. Math. Comput.* **2013**, *219*, 7805–7819. [CrossRef]
20. Berlinet, A.; Elamine, A.; Mas, A. Local Linear Regression for Functional Data. *Ann. Inst. Stat. Math.* **2011**, *63*, 1047–1075. [CrossRef]
21. Abeidallah, M.; Mechab, B.; Merouan, T. Local Linear Estimate of the Point at High Risk: Spatial Functional Data Case. *Commun. Stat. Theory Methods* **2020**, *49*, 2561–2584. [CrossRef]
22. Sara, L. Nonparametric Local Linear Regression Estimation for Censored Data and Functional Regressors. *J. Korean Stat. Soc.* **2020**, *51*, 1–22. [CrossRef]
23. Lei, X.; Zhang, H. Non-asymptotic Optimal Prediction Error for RKHS-based Partially Functional Linear Models. *arXiv* **2020**, arXiv:2009.04729.
24. Fang, K.T.; Li, R.; Sudjianto, A. *Design and Modeling for Computer Experiments*; Chapman & Hall/CRC: New York, NY, USA, 2006.
25. Lai, T.L.; Robins, H.; Wei, C.Z. Strong Consistency of Least Squares Estimates in Multiple Regression. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 3034–3036. [CrossRef]
26. Eicker, F. Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions. *Ann. Math. Stat.* **1963**, *34*, 447–456. [CrossRef]
27. Wahba, G. *Spline Models for Observational Data*; SIAM: Philadelphia, PA, USA, 1990.
28. Gu, C. *Smoothing Spline ANOVA Models*; Springer: New York, NY, USA, 2002.
29. Rapoport, A.P.; Aqui, N.A.; Stadtmauer, E.A.; Vogl, D.T.; Fang, H.B.; Cai, L.; Janofsky, S.; Chew, A.; Storek, J.; Gorgun, A.; et al. Combination immunotherapy using adoptive T-cell transfer and tumor antigen vaccination based on hTERT and survivin after ASCT for myeloma. *Blood* **2011**, *117*, 788–797. [CrossRef] [PubMed]
30. Fang, H.B.; Wu, T.T.; Rapoport, A.P.; Tan, M. Survival Analysis with Functional Covariates Based on Partial Follow-up Studies. *Stat. Methods Med. Res.* **2016**, *25*, 2405–2419. [CrossRef]
31. Zhang, H.; Jia, J. Elastic-net Regularized High-dimensional Negative Binomial Regression: Consistency and Weak Signals Detection. *Stat. Sin.* **2022**, *32*, 181–207. [CrossRef]
32. Ledoux, M.; Talagrand, M. *Probability in Banach Spaces*; Springer: New York, NY, USA, 1991.