ADDITIVE REGRESSION FOR NON-EUCLIDEAN RESPONSES AND PREDICTORS

BY JEONG MIN JEON[1,*], BYEONG U. PARK[2] AND INGRID VAN KEILEGOM[1,†]

[1]*ORSTAT, KU Leuven,* *jeongmin.jeon@kuleuven.be;* †*ingrid.vankeilegom@kuleuven.be*
[2]*Department of Statistics, Seoul National University, bupark@stats.snu.ac.kr*

Additive regression is studied in a very general setting where both the response and predictors are allowed to be non-Euclidean. The response takes values in a general separable Hilbert space, whereas the predictors take values in general semimetric spaces, which covers a very wide range of nonstandard response variables and predictors. A general framework of estimating additive models is presented for semimetric space-valued predictors. In particular, full details of implementation and the corresponding theory are given for predictors taking values in Hilbert spaces and/or Riemannian manifolds. The existence of the estimators, convergence of a backfitting algorithm, rates of convergence and asymptotic distributions of the estimators are discussed. The finite sample performance of the estimators is investigated by means of two simulation studies. Finally, three data sets covering several types of non-Euclidean data are analyzed to illustrate the usefulness of the proposed general approach.

1. Introduction. Statistical analysis of non-Euclidean data objects is becoming an important topic in modern statistics. Existing methods designed for Euclidean data do not make use of the underlying geometric structures of non-Euclidean data. As a result, they may give poor estimation/prediction performance and also cause distorted interpretation. Recent years have seen some attempts for developing statistical techniques dealing with non-Euclidean data. For a review on statistical analysis for nonstandard data, we refer to [30] and the references therein. The topic is still largely unexplored in many areas of statistical inference, however. In this paper, we consider additive models of the form

$$(1.1) \qquad \mathbf{Y} = \bigoplus_{j=1}^{d} \mathbf{f}_j(X_j) \oplus \boldsymbol{\epsilon},$$

where $\mathbf{Y}$ and $\boldsymbol{\epsilon}$ are response and error variables, respectively, taking values in a Hilbert space, X_j are semimetric space-valued predictors, $\mathbf{f}_j$ are unknown Hilbert space-valued maps and $\oplus$ is a vector addition of the underlying Hilbert space. We present a general framework of estimating the model (1.1) for predictors taking values in general semimetric spaces. In particular, we get into full details for predictors whose values reside in finite-dimensional Hilbert spaces and/or Riemannian manifolds.

The postulation of Hilbert spaces for the response variable covers a very wide scope of data types including, for example, Euclidean, compositional, functional and density-valued data. We consider Hilbert spaces since they are an important class of non-Euclidean data spaces equipped with vector operations and an inner product structure that are vital to response variables in most regression techniques. Our coverage for predictors is truly broad as well. Hilbertian (Hilbert space-valued)/Riemannian (Riemannian manifold-valued) predictors for

which we give fully detailed practical implementation and theory have a very wide spectrum of data types, which includes Euclidean, compositional, spherical, toroidal (torus-valued), shape, special orthogonal matrix-valued data and so on. The coverage facilitates additive regression in a huge number of real world problems. A compositional value is an Euclidean vector of positive entries whose sum equals one. This type arises from numerous sources such as elections, compositions of body/air/sea-water/soil and income/expenditure distributions, etc. Spherical data and toroidal data, including circular data as a special case, emerge from enormous fields such as earth science and astronomy. Examples include periodic time, directions of wind and of animal movements, and positions of sunspots and of airplanes, etc. A shape value is a set of finite points in an Euclidean space representing the shape of an artifial/natural object of interest. Since objects are everywhere, this type of data can be easily found. Examples are shapes of skulls, organs, faces, sand-particles and lands. For more examples of shape data, we refer to [8]. Some examples of special-orthogonal-matrix-valued data include vector-cardiograms [2] and alignments of crystals [17]. There are many examples of Hilbertian responses that can be regressed on a mixture of Hilbertian and Riemmanian variables, such as those in Section 6. One may also think of the year-long curve of ground temperature regressed on soil composition and location on the earth.

The additive model (1.1) is an important example of structured nonparametric regression. [29] initiated a powerful method of fitting (1.1), named as the smooth backfitting (SBF) technique, for *real-valued* responses regressed on *real-valued* predictors. Since then, the success of the technique in terms of avoiding the curse of dimensionality has been observed in other structured models; see [25, 26, 43] and [15], among others. For an account of the practical issues of the method, see [31]. In recent years, there have been some achievements of extending the original work of [29] on the additive model to the case of more general spaces for responses and predictors. To list a few here, [14] treated density-valued responses $Y(\cdot)$ with *real-valued* predictors, but in a *pointwise* manner, that is, regressing $Y(t)$ on predictors for each t, so that their methods and theory are essentially the same as in the case of real-valued responses. Jeon and Park [18] extended the SBF technique to general Hilbertian responses regressed on *real-valued* predictors. The latter treatment for response variables includes the case of functional or density-valued responses, and when applied to such cases it estimates function-valued maps themselves, not pointwise real-valued maps as in [14]. The present paper aims at an entirely different level of generalization. We present a novel theory for a *general Hilbertian* response coupled with *general semimetric space-valued* predictors. In particular, the methodology and theory we detail for *Hilbertian/Riemannian* predictors are complete. Some of specific ingredients for the practical implementation we also exemplify in this paper do not exist in the literature.

Nonparametric regression analysis without a structure such as (1.1), which we call *full-dimensional* regression, is much simpler. The methodology is essentially the same as in the case of single predictor, but is subject to the *curse of dimensionality* even when the number of predictors is moderately large. There is a body of literature on single or full-dimensional regression with non-Euclidean predictors. This includes [5, 34] and [3]. The aforementioned works treated a real-valued response with a single or multiple Riemannian manifold-valued predictors, some of which considered specific manifolds such as circles. Recently, [27] studied functional linear regression with a real-valued response and a single functional Riemannian manifold-valued predictor. An example dealing with a non-Euclidean response is local polynomial regression with a response taking values in the space of symmetric positive definite matrices and a single real-valued predictor [44]. Others are spherical-spherical regression [6, 37], parametric regression with a Riemannian manifold-valued response and real-valued predictors [4] and Fréchet regression with a metric-space-valued response and real-valued predictors [35]. However, there has been no parametric/nonparametric method designed for

an Euclidean response coupled with a mixture of compositional and Riemannian predictors, and for a non-Euclidean Hilbertian response coupled with Riemannian predictors, to the best of our knowledge.

During recent decades, there have been a few attempts for estimating additive models. Methods other than the SBF include the marginal integration [28], the classical backfitting [11, 32], the reproducing kernel Hilbert space approach [36] and the spline-based technique [10, 21, 39]. However, all these approaches are restricted to Euclidean or functional variables. We believe that (1.1) is the first structured nonparametric model dealing with general possibly non-Euclidean responses and predictors. Our treatment of the cases of Hilbertian and of Riemannian predictors already covers a variety of data examples, as we mentioned earlier. Our general framework encompasses further any combinations of Hilbertian and Riemannian predictors. It also covers the case where a predictor X_j itself is multidimensional, and thus its space is a product of multiple Hilbert spaces or of multiple Riemannian manifolds, which enable us to analyze interaction effects between the corresponding lower-dimensional predictors constituting X_j. In the specialization to Euclidean predictors, it allows for general predictor supports such as (rotated) ellipses, circular cylinders and polygons, contrary to the interval supports in [18]. In summary, the present paper provides a powerful and large toolbox for structured nonparametric regression analysis.

This paper is organized as follows. In Section 2, we present a general framework of the methodology and establish the existence of the SBF estimator and the convergence of the SBF algorithm. In Section 3, we detail the methodology for predictors taking values in finite-dimensional Hilbert spaces and/or Riemannian manifolds. We present the asymptotic properties of the estimator in Section 4. We then discuss some generalization of the theory in Section 5 for general semimetric space-valued predictors. In Section 6, we report the results of extensive simulation studies and illustrate the methods in various real data examples. We summarize the major developments of our work in Section 7 and give directions for future work. We put all technical proofs in the Supplementary Material [19] because of the space constraint, although they are new and based on innovative mathematical analysis. The supplement also contains some ancillary results that are of interest in their own right.

2. General framework. In this section, we go over a general framework of estimating additive models with predictors taking values in semimetric spaces and responses taking values in a separable Hilbert space. Let $(\mathcal{X}_j, d_j)$ for $1 \leq j \leq d$ be semimetric spaces (sometimes called pseudo-metric spaces, which allow $d_j(x_j, u_j) = 0$ for some $x_j \neq u_j$) with infinite cardinalities where the random predictors X_j take values. We consider semimetric spaces for predictor domains since they are one of the most general domains for kernel smoothing that we adopt in this paper. Let $\mathbf{X} = (X_1, \ldots, X_d)$ and $\mathcal{X} = \prod_{j=1}^{d} \mathcal{X}_j$. We denote by $\mathcal{B}(\mathcal{X}_j)$ the Borel sigma-field of $\mathcal{X}_j$ induced by the semimetric d_j and let μ_j be a measure on $\mathcal{B}(\mathcal{X}_j)$ satisfying $0 < \mu_j(\mathcal{X}_j) < \infty$.

For example, when $\mathcal{X}_j$ is a finite-dimensional Hilbert space, we may take d_j as the metric induced by an inner product of the Hilbert space, and μ_j as the *pushforward measure* induced by a Lebesgue measure. Specifically, if $\mathcal{X}_j$ is a q_j-dimensional space, the pushforward measure μ_j is given by $\mu_j(A) = \text{Leb}_{q_j}(\eta_j(A))$ for $A \in \mathcal{B}(\mathcal{X}_j)$, where Leb_{q_j} denotes the q_j-dimensional Lebesgue measure and $\eta_j : \mathcal{X}_j \to \mathbb{R}^{q_j}$ is an isometric isomorphism. When $\mathcal{X}_j$ is a finite-dimensional Riemannian manifold, we may take d_j and μ_j, respectively, as the Riemmanian distance and Riemmanian volume measure induced by a Riemannian metric. In Section 3.3, we give concrete examples of d_j and μ_j for some Hilbert spaces including Euclidean spaces and simplices, and for some Riemannian manifolds including toruses, spheres and planar shape spaces with formal definitions of the spaces.

Let $\mathbb{H}$ be a separable Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$ and the associated norm $\| \cdot \|$ where the values of the response $\mathbf{Y}$ reside. Throughout this paper, we assume

$E(\|\mathbf{Y}\|^2) < \infty$. Let $\oplus, \odot, \mathbf{0}$ and $\mathcal{B}(\mathbb{H})$, respectively, denote a vector addition, scalar multiplication, zero vector and the Borel sigma-field of $\mathbb{H}$. Examples of $\mathbb{H}$ include Euclidean spaces, simplices, L^2-spaces and Bayes–Hilbert spaces. The definitions of these spaces with associated $\langle \cdot, \cdot \rangle$, $\| \cdot \|$, $\oplus$, $\odot$ and $\mathbf{0}$ are given in the Supplement S.1; see also [18]. In this paper, we consider an additive model

$$(2.1) \qquad \mathbf{Y} = E(\mathbf{Y}) \oplus \bigoplus_{j=1}^{d} \mathbf{f}_j(X_j) \oplus \boldsymbol{\epsilon},$$

where $\mathbf{f}_j : \mathcal{X}_j \to \mathbb{H}$ are $(\mathcal{B}(\mathcal{X}_j), \mathcal{B}(\mathbb{H}))$-measurable maps satisfying $E(\mathbf{f}_j(X_j)) = \mathbf{0}$ and $E(\|\mathbf{f}_j(X_j)\|^2) < \infty$ for $1 \le j \le d$, and $\boldsymbol{\epsilon}$ is an $\mathbb{H}$-valued error satisfying $E(\boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbf{0}$. The constraints $E(\mathbf{f}_j(X_j)) = \mathbf{0}$, $1 \le j \le d$, are for the identifiability of the component maps $(\mathbf{f}_j : 1 \le j \le d)$. Here, the (conditional) expectation is in the sense of *Bochner integral*, which is for Banach space-valued maps. For more details on this notion, see [18] and also the Supplement S.2.

2.1. *Methodology.* We assume that the probability distribution $P\mathbf{X}^{-1}$ of $\mathbf{X}$ is absolutely continuous with respect to the product measure $\mu := \bigotimes_{j=1}^{d} \mu_j$, so that the density $p := dP\mathbf{X}^{-1}/d\mu$ exists. Let p_{jk} and p_j denote the marginal densities for (X_j, X_k) and X_j, respectively. We assume that $p_j(x_j) > 0$ and $\int_{\mathcal{X}_k} p_{jk}^2(x_j, x_k)/p_k(x_k)\,d\mu_k(x_k) < \infty$ hold for all $1 \le j \ne k \le d$ and $x_j \in \mathcal{X}_j$. Then, by taking the conditional expectation $E(\cdot \mid X_j = x_j)$ on both sides of (2.1) and applying Proposition 2.2 in [18], we get

$$(2.2) \qquad \mathbf{f}_j(x_j) = \mathbf{m}_j(x_j) \ominus E(\mathbf{Y}) \ominus \bigoplus_{k \ne j}^{d} \int_{\mathcal{X}_k} \mathbf{f}_k(x_k) \odot \frac{p_{jk}(x_j, x_k)}{p_j(x_j)}\,d\mu_k(x_k), \quad 1 \le j \le d,$$

where $\mathbf{m}_j$ are the marginal regression maps such that $\mathbf{m}_j(x_j) = E(\mathbf{Y} \mid X_j = x_j)$ and $\ominus$ is a vector subtraction of $\mathbb{H}$ defined by $\mathbf{h}_1 \ominus \mathbf{h}_2 = \mathbf{h}_1 \oplus (-1 \odot \mathbf{h}_2)$. We note that the integrals in (2.2) are Bochner integrals. The basic idea is to estimate the system of Bochner integral equations at (2.2) and then solve the resulting system of equations to get the estimators of the component maps $(\mathbf{f}_j : 1 \le j \le d)$.

Suppose that we have n observations $\{(\mathbf{Y}_i, \mathbf{X}_i)\}_{1 \le i \le n}$. In this section, we do not assume that the observations are identically distributed or independent. We note that this general treatment has not been considered before, not even in the recent [18]. For $\mathbf{x} = (x_1, \dots, x_d) \in \mathcal{X}$, we estimate $p(\mathbf{x})$ by $\hat{p}(\mathbf{x}) = n^{-1} \sum_{i=1}^{n} \prod_{j=1}^{d} L_{j,n}(x_j, X_{ij})$, where $L_{j,n} : \mathcal{X}_j \times \mathcal{X}_j \to [0, \infty)$ is any measurable weight function which depends on n and satisfies

$$(2.3) \qquad \int_{\mathcal{X}_j} L_{j,n}(x_j, u_j)\,d\mu_j(x_j) = 1$$

for all $u_j \in \mathcal{X}_j$. In Section 3.1, we give specific forms of $L_{j,n}$ that satisfy (2.3) for Hilbert spaces and Riemannian manifolds. We estimate the marginal densities p_{jk} and p_j in (2.2) by

$$\hat{p}_{jk}(x_j, x_k) = n^{-1} \sum_{i=1}^{n} L_{j,n}(x_j, X_{ij}) L_{k,n}(x_k, X_{ik}), \qquad \hat{p}_j(x_j) = n^{-1} \sum_{i=1}^{n} L_{j,n}(x_j, X_{ij}).$$

We also estimate $E(\mathbf{Y})$ and $\mathbf{m}_j$ in (2.2) by

$$\hat{E}(\mathbf{Y}) = n^{-1} \odot \bigoplus_{i=1}^{n} \mathbf{Y}_i = \bar{\mathbf{Y}}, \qquad \hat{\mathbf{m}}_j(x_j) = \frac{1}{n\hat{p}_j(x_j)} \odot \bigoplus_{i=1}^{n} L_{j,n}(x_j, X_{ij}) \odot \mathbf{Y}_i.$$

We consider the following condition on the density estimators.

CONDITION (S).   *For all $1 \le j \ne k \le d$ and $x_j \in \mathcal{X}_j$, we have $\hat{p}_j(x_j) > 0$, $\int_{\mathcal{X}_k} \hat{p}_{jk}^2(x_j, x_k)/\hat{p}_k(x_k)\, d\mu_k(x_k) < \infty$ and*

$$\int_{\mathcal{X}_k} \int_{\mathcal{X}_j} \frac{\hat{p}_{jk}^2(x_j, x_k)}{\hat{p}_j(x_j)\hat{p}_k(x_k)}\, d\mu_j(x_j)\, d\mu_k(x_k) < \infty.$$

Under the conditions (S) and that $\max_{1 \le j \le d} \int_{\mathcal{X}_j} \|\hat{\mathbf{f}}_j(x_j)\|^2 \hat{p}_j(x_j)\, d\mu_j(x_j) < \infty$, we have $\int_{\mathcal{X}_k} \|\hat{\mathbf{f}}_k(x_k)\| \hat{p}_{jk}(x_j, x_k)\, d\mu_k(x_k) < \infty$ by Hölder's inequality. Thus, in the space of square integrable maps and under (S), the Bochner integrals in the estimated system of Bochner integral equations,

$$(2.4) \quad \hat{\mathbf{f}}_j(x_j) = \hat{\mathbf{m}}_j(x_j) \ominus \bar{\mathbf{Y}} \ominus \bigoplus_{k \ne j}^{d} \int_{\mathcal{X}_k} \hat{\mathbf{f}}_k(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)}\, d\mu_k(x_k), \quad 1 \le j \le d,$$

are well defined. The smooth backfitting (SBF) estimator of the tuple $(\mathbf{f}_j : 1 \le j \le d)$ of the component maps is then defined as a solution $(\hat{\mathbf{f}}_j : 1 \le j \le d)$ of the system of equations in (2.4) subject to the constraints

$$(2.5) \quad \begin{aligned} \int_{\mathcal{X}_j} \hat{\mathbf{f}}_j(x_j) \odot \hat{p}_j(x_j)\, d\mu_j(x_j) &= \mathbf{0}, \quad 1 \le j \le d, \\ \int_{\mathcal{X}_j} \|\hat{\mathbf{f}}_j(x_j)\|^2 \hat{p}_j(x_j)\, d\mu_j(x_j) &< \infty, \quad 1 \le j \le d. \end{aligned}$$

The above constraints are the empirical versions of the constraints $E(\mathbf{f}_j(X_j)) = \mathbf{0}$ and $E(\|\mathbf{f}_j(X_j)\|^2) < \infty$ on the true tuple $(\mathbf{f}_j : 1 \le j \le d)$. We call $(\hat{\mathbf{f}}_j : 1 \le j \le d)$ *semimetric Bochner smooth backfitting* (sB-SBF) estimator.

Below, we present a proposition that demonstrates the existence and uniqueness of the sB-SBF estimator. Define a measure $\hat{P}\mathbf{X}^{-1}$ on $\bigotimes_{j=1}^{d} \mathcal{B}(\mathcal{X}_j)$ by $\hat{P}\mathbf{X}^{-1}(A) = \int_A \hat{p}(\mathbf{x})\, d\mu(\mathbf{x})$.

PROPOSITION 2.1.   *Assume the condition* (S). *Then there exists a solution $(\hat{\mathbf{f}}_j : 1 \le j \le d)$ of* (2.4) *satisfying* (2.5), *and the sum of such $\hat{\mathbf{f}}_j$ for $1 \le j \le d$ is unique up to measure zero with respect to $\hat{P}\mathbf{X}^{-1}$. Moreover, when $\hat{p} > 0$ on $\mathcal{X}$, the solution is unique in the sense that, if $(\hat{\mathbf{f}}_j^{\star} : 1 \le j \le d)$ is another solution, then $\hat{\mathbf{f}}_j(x_j) = \hat{\mathbf{f}}_j^{\star}(x_j)$ a.e. with respect to μ_j for all $1 \le j \le d$.*

2.2. *Implementation scheme.*   The sB-SBF estimator has no explicit form. The *sB-SBF algorithm* as we describe below is an iterative scheme that solves (2.4). First, we take any initial estimator $(\hat{\mathbf{f}}_j^{[0]} : 1 \le j \le d)$ satisfying

$$(2.6) \quad \begin{aligned} \int_{\mathcal{X}_j} \hat{\mathbf{f}}_j^{[0]}(x_j) \odot \hat{p}_j(x_j)\, d\mu_j(x_j) &= \mathbf{0}, \quad 1 \le j \le d, \\ \int_{\mathcal{X}_j} \|\hat{\mathbf{f}}_j^{[0]}(x_j)\|^2 \hat{p}_j(x_j)\, d\mu_j(x_j) &< \infty, \quad 1 \le j \le d. \end{aligned}$$

We update them in the rth iteration by

$$\hat{\mathbf{f}}_j^{[r]}(x_j) = \hat{\mathbf{m}}_j(x_j) \ominus \bar{\mathbf{Y}} \ominus \bigoplus_{k<j} \int_{\mathcal{X}_k} \hat{\mathbf{f}}_k^{[r]}(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)}\, d\mu_k(x_k)$$

$$(2.7)$$

$$\ominus \bigoplus_{k>j} \int_{\mathcal{X}_k} \hat{\mathbf{f}}_k^{[r-1]}(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)}\, d\mu_k(x_k), \quad 1 \le j \le d.$$

We note that, under the condition (S), the Bochner integrals in (2.7) are well defined and $\int_{\mathcal{X}_j} \|\hat{\mathbf{f}}_j^{[r]}(x_j)\|^2 \hat{p}_j(x_j) \, d\mu_j(x_j) < \infty$ for all $1 \le j \le d$. Using (2.3) one can easily show that the constraints

$$\int_{\mathcal{X}_j} \hat{\mathbf{f}}_j^{[r]}(x_j) \odot \hat{p}_j(x_j) \, d\mu_j(x_j) = \mathbf{0}, \quad 1 \le j \le d$$

are automatically satisfied for $r \ge 1$ if the initial $\hat{\mathbf{f}}_j^{[0]}$ satisfy (2.6).

The following two propositions demonstrate that the sB-SBF algorithm converges to the solution of (2.4). For their statements, let $\mathbf{f} = E(\mathbf{Y}) \oplus \bigoplus_{j=1}^d \mathbf{f}_j$, which is the mean regression map $E(\mathbf{Y} \mid \mathbf{X} = \cdot)$ under the additive model (2.1). Define $\hat{\mathbf{f}} = \bar{\mathbf{Y}} \oplus \bigoplus_{j=1}^d \hat{\mathbf{f}}_j$ and $\hat{\mathbf{f}}^{[r]} = \bar{\mathbf{Y}} \oplus \bigoplus_{j=1}^d \hat{\mathbf{f}}_j^{[r]}$.

PROPOSITION 2.2. *Assume the condition (S). Then there exist constants $\hat{c} > 0$ and $\hat{\gamma} \in (0, 1)$ such that*

$$\int_{\mathcal{X}} \|\hat{\mathbf{f}}(\mathbf{x}) \ominus \hat{\mathbf{f}}^{[r]}(\mathbf{x})\|^2 \hat{p}(\mathbf{x}) \, d\mu(\mathbf{x}) \le \hat{c} \cdot \hat{\gamma}^r \quad \text{for all } r \ge 0.$$

The next convergence result is for the individual component maps. For this, we impose the following high-level condition.

CONDITION (A). *There exists a constant $C > 0$ such that, with probability tending to one,*

(A1) $\displaystyle \max_{1 \le j \le d} \sup_{x_j \in \mathcal{X}_j} \hat{p}_j(x_j)^{-1} < C;$ (A2) $\displaystyle \max_{1 \le j \ne k \le d} \sup_{(x_j, x_k) \in \mathcal{X}_j \times \mathcal{X}_k} \hat{p}_{jk}(x_j, x_k) < C;$

(A3) $\displaystyle \max_{1 \le j \le d} \sup_{x_j \in \mathcal{X}_j} \|\hat{\mathbf{m}}_j(x_j)\| < C;$ (A4) $\displaystyle \max_{1 \le j \le d} \int_{\mathcal{X}_j} \|\hat{\mathbf{f}}_j^{[0]}(x_j)\|^2 \, d\mu_j(x_j) < C.$

Also, it holds that

(A5) $\displaystyle \max_{1 \le j \le d} \int_{\mathcal{X}_j} (\hat{p}_j(x_j) - p_j(x_j))^2 \, d\mu_j(x_j) = o_p(1),$

(A6) $\displaystyle \max_{1 \le j \ne k \le d} \int_{\mathcal{X}_j \times \mathcal{X}_k} (\hat{p}_{jk}(x_j, x_k) - p_{jk}(x_j, x_k))^2 \, d\mu_j \otimes \mu_k(x_j, x_k) = o_p(1),$

(A7) *p is bounded away from zero and infinity on $\mathcal{X}$.*

We note that the condition (A7) on the underlying distribution of $\mathbf{X}$ is standard in nonparametric regression.

PROPOSITION 2.3. *Assume the condition (A). Then there exist constants $c > 0$ and $\gamma \in (0, 1)$ such that, for all $1 \le j \le d$,*

$$\lim_{n \to \infty} P\left(\int_{\mathcal{X}_j} \|\hat{\mathbf{f}}_j(x_j) \ominus \hat{\mathbf{f}}_j^{[r]}(x_j)\|^2 p_j(x_j) \, d\mu_j(x_j) \le c \cdot \gamma^r \text{ for all } r \ge 0 \right) = 1.$$

Moreover, for all $1 \le j \le d$,

$$\lim_{n \to \infty} P\left(\mu_j(\{x_j \in \mathcal{X}_j : \hat{\mathbf{f}}_j^{[r]}(x_j) \to \hat{\mathbf{f}}_j(x_j) \text{ as } r \to \infty\}) = \mu_j(\mathcal{X}_j) \right) = 1.$$

As in [18], we may implement the sB-SBF algorithm using Lebesgue integration. This is because of the fact that, for any measure space $(\mathcal{Z}, \mathscr{A}, \mu)$,

$$(\text{Bochner}) \int_{\mathcal{Z}} f(\mathbf{z}) \odot \mathbf{b} \, d\mu(\mathbf{z}) = (\text{Lebesgue}) \int_{\mathcal{Z}} f(\mathbf{z}) \, d\mu(\mathbf{z}) \odot \mathbf{b}$$

in case f is a real-valued integrable function on $\mathcal{Z}$ and $\mathbf{b}$ is a constant in a Banach space. To implement the idea, we take initial estimators of the form

$$(2.8) \qquad \hat{\mathbf{f}}_j^{[0]}(x_j) = n^{-1} \odot \bigoplus_{i=1}^{n} w_{ij}^{[0]}(x_j) \odot \mathbf{Y}_i, \quad 1 \le j \le d,$$

for some continuous functions $w_{ij}^{[0]} : \mathcal{X}_j \to \mathbb{R}$ satisfying $\int_{\mathcal{X}_j} w_{ij}^{[0]}(x_j)\hat{p}_j(x_j) \, d\mu_j(x_j) = 0$. We note that the constraints are satisfied by the trivial choice $w_{ij}^{[0]} \equiv 0$. Define

$$(2.9) \qquad \begin{aligned} w_{ij}^{[r]}(x_j) &= \frac{L_{j,n}(x_j, X_{ij})}{\hat{p}_j(x_j)} - 1 - \sum_{k<j} \int_{\mathcal{X}_k} w_{ik}^{[r]}(x_k) \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} \, d\mu_k(x_k) \\ &\quad - \sum_{k>j} \int_{\mathcal{X}_k} w_{ik}^{[r-1]}(x_k) \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} \, d\mu_k(x_k). \end{aligned}$$

Then it is easy to see that updating according to (2.7) is equivalent to taking

$$\hat{\mathbf{f}}_j^{[r]}(x_j) = n^{-1} \odot \bigoplus_{i=1}^{n} w_{ij}^{[r]}(x_j) \odot \mathbf{Y}_i, \quad 1 \le j \le d$$

with $w_{ij}^{[r]}$ being updated according to (2.9). In Section 3.3, we introduce specific ways of implementing the integrals in (2.9) for some Hilbert spaces and Riemannian manifolds.

3. Methodology for Hilbertian and Riemannian predictors. In this section, we give full details of the methodology for two large classes of semimetric spaces, which are Hilbert spaces and Riemannian manifolds. To treat both types in a unified fashion, we let $0 \le \varrho \le d$ be an integer and assume that $\mathcal{X}_j$ for $1 \le j \le \varrho$ are compact subsets of q_j-dimensional Hilbert spaces $\mathcal{H}_j$ and $\mathcal{X}_j$ for $\varrho + 1 \le j \le d$ are q_j-dimensional connected and compact Riemannian manifolds without boundary. In both cases, we assume that $1 \le q_j < \infty$. We denote a vector subtraction and a norm of $\mathcal{H}_j$ by $\ominus_j$ and $\|\cdot\|_j$, respectively. We let g_j denote the Riemannian metric (sometimes called Riemannian structure) of $\mathcal{X}_j$ for $\varrho + 1 \le j \le d$ and d_j the Riemannian distance function induced by g_j. We note that not g_j but d_j serves as a metric on $\mathcal{X}_j$ for $\varrho + 1 \le j \le d$. For an introduction to Riemannian geometry, we refer to [41] and [24], for example. We also collect basic notions related to Riemannian manifolds in the Supplement S.3. In our unified framework, $\varrho = 0$ means that there is no Hilbertian predictor and $\varrho = d$ means that there is no Riemannian predictor. We note that, for any $0 \le \varrho \le d$, no nonparametric additive model has been studied even for $\mathbb{H} = \mathbb{R}$. Our treatment covers all the cases $0 \le \varrho \le d$ with general $\mathbb{H}$.

Examples of $\mathcal{H}_j$ include Euclidean spaces, simplices and their product spaces. We note that simplices are domains for compositional data. Examples of connected and compact Riemannian manifolds without boundary include spheres $\mathbb{S}^{q_j} = \{x_j = (x_{j,1}, \ldots, x_{j,q_j+1}) \in \mathbb{R}^{q_j+1} : \sum_{l=1}^{q_j+1} x_{j,l}^2 = 1\}$, toruses $\mathbb{S}^1 \times \cdots \times \mathbb{S}^1$, planar shape spaces [23], special orthogonal groups and their product manifolds. Spheres are domains for spherical data, and toruses are domains for toroidal data. Both data types cover circular data. Planar shape spaces are domains for shape data, and special orthogonal groups are domains for special orthogonal-matrix-valued data. More details on the aforementioned spaces are in Section 3.3.

We assume that we observe n i.i.d. copies of $(\mathbf{Y}, \mathbf{X})$, which we denote by $(\mathbf{Y}_i, \mathbf{X}_i)$, $1 \le i \le n$. Each component X_j of $\mathbf{X}$ is a random element taking values, either in the compact subset $\mathcal{X}_j$ of the Hilbert space $\mathcal{H}_j$ or in the Riemannian manifold $(\mathcal{X}_j, g_j)$. For each Hilbertian predictor X_j, let $\eta_j : \mathcal{H}_j \to \mathbb{R}^{q_j}$ be an isometric isomorphism. We choose μ_j to be the pushforward measure on $\mathcal{B}(\mathcal{H}_j)$ such that $\mu_j(A) = \mathrm{Leb}_{q_j}(\eta_j(A))$. For each Riemannian predictor X_j taking values in the Riemannian manifold $(\mathcal{X}_j, g_j)$, we choose μ_j to be the Riemannian volume measure induced by g_j; see Section 3.3.2 for the definition.

If $\varrho = d$ and $\mathcal{X}_j$ are compact and convex subsets of $\mathbb{R}^{q_j}$, so that we may take η_j to be the identity map on $\mathbb{R}^{q_j}$ and $\mu_j = \mathrm{Leb}_{q_j}$, then the results here and in Section 4 completely generalize those in [18] to "multivariate additive models" having multivariate component maps. The latter generalization, but still a specialization of our general formulation, may accommodate interaction effects between real-valued predictors. This means that our general framework provides useful tools in various scenarios filling the gap between univariate additive models and entirely unstructured models. Moreover, we may deal with predictors having more general types of support such as (rotated) ellipses, circular cylinders and polygons, which [18] does not cover.

3.1. *Kernel functions.*   In this subsection, we specify the weight functions $L_{j,n}$ at (2.3) to kernel functions K_{h_j} to be defined below. With the kernel functions, we may obtain from (2.4) the corresponding system of sB-SBF equations to solve for the sB-SBF estimators of the component maps $\mathbf{f}_j$. We also have the corresponding sB-SBF algorithm from (2.7) and the corresponding updating formula (2.9) that we actually iterate.

For $1 \le j \le d$, let $K_j : [0, \infty) \to [0, \infty)$ be continuous functions vanishing on $(1, \infty)$ and satisfying $\int_{\mathbb{R}^{q_j}} K_j(\|\mathbf{t}\|_{\mathbb{R}^{q_j}}) \, d\mathbf{t} = 1$, where $\| \cdot \|_{\mathbb{R}^{q_j}}$ denotes the Euclidean norm on $\mathbb{R}^{q_j}$. For Hilbertian predictors $(1 \le j \le \varrho)$, we use new kernel schemes K_{h_j}, for $h_j > 0$ called the bandwidths, defined by

$$(3.1) \qquad K_{h_j}(x_j, u_j) = \frac{K_j(\|x_j \ominus_j u_j\|_j / h_j)}{\int_{\mathcal{X}_j} K_j(\|v_j \ominus_j u_j\|_j / h_j) \, d\mu_j(v_j)}$$

whenever $\int_{\mathcal{X}_j} K_j(\|v_j \ominus_j u_j\|_j / h_j) \, d\mu_j(v_j) \ne 0$, and we set $K_{h_j}(x_j, u_j) = 0$ otherwise. Here, we write K_{h_j} instead of K_{j,h_j} for simplicity of notation. We note that, if the denominator in (3.1) is nonzero, then the kernel functions K_{h_j} satisfies (2.3).

For Riemannian predictors, we use the kernel function of [33] defined by

$$(3.2) \qquad K_{h_j}(x_j, u_j) = \frac{1}{h_j^{q_j}} \frac{1}{\theta_j(x_j; u_j)} K_j\left(\frac{d_j(x_j, u_j)}{h_j}\right) I(d_j(x_j, u_j) < \mathrm{inj}(\mathcal{X}_j)),$$

where $\theta_j(\cdot; u_j)$ is the volume density function of $\mathcal{X}_j$ at $u_j \in \mathcal{X}_j$, and $\mathrm{inj}(\mathcal{X}_j)$ is the injectivity radius of $\mathcal{X}_j$. For the definitions of volume density function and injectivity radius, see [12] or the Supplement S.3. Since $\mathcal{X}_j$ is compact, $\mathrm{inj}(\mathcal{X}_j) > 0$. We note that $\theta_j(x_j; u_j)$ is defined only for x_j and u_j such that $d_j(x_j, u_j) < \mathrm{inj}(\mathcal{X}_j)$, and it holds that $\theta_j(x_j; u_j) > 0$, $\theta_j(x_j; u_j) = \theta_j(u_j; x_j)$ and $\theta_j(x_j; x_j) = 1$ for such x_j and u_j. Also, $\theta_j(\cdot; u_j)$ is smooth on $\{x_j \in \mathcal{X}_j : d_j(x_j, u_j) \le r\}$ for any $r < \mathrm{inj}(\mathcal{X}_j)$. The kernel function $K_{h_j}(x_j, u_j)$ defined at (3.2) satisfies (2.3) if $h_j < \mathrm{inj}(\mathcal{X}_j)$.

We note that Jeon and Park [18] studied the additive model (2.1) with $\mathcal{X}_j = [0, 1] \subset \mathbb{R}$ for all $1 \le j \le d$ and kernel K_{h_j} defined by

$$(3.3) \qquad K_{h_j}(x_j, u_j) = \frac{K((x_j - u_j)/h_j)}{\int_0^1 K((v_j - u_j)/h_j) \, dv_j},$$

where $K : \mathbb{R} \to [0, \infty)$ is a measurable function that vanishes on $(-\infty, -1) \cup (1, \infty)$ and satisfies $\int_{-1}^1 K(t) \, dt = 1$. The new kernel at (3.1) generalizes the kernel at (3.3) when K is symmetric.

3.2. *Existence of sB-SBF estimator and convergence of sB-SBF algorithm.* For $1 \leq j \leq \varrho$, we denote $d_j(x_j, u_j) = \|x_j \ominus_j u_j\|_j$ and $(\eta_j(\mathcal{X}_j) - \eta_j(x_j))/h_j = \{(\eta_j(u_j) - \eta_j(x_j))/h_j : u_j \in \mathcal{X}_j\}$. For $1 \leq j \leq d$ and $r > 0$, we define $B_{\mathcal{X}_j}(x_j, r) = \{u_j \in \mathcal{X}_j : d_j(x_j, u_j) < r\}$. We denote by $\mathbf{0}_{q_j}$ the zero vector of $\mathbb{R}^{q_j}$, and by $B(\mathbf{0}_{q_j}, r)$ the Euclidean open ball centered at $\mathbf{0}_{q_j}$ with radius r. Throughout this paper, we separate the conditions for each result to make it clear what are minimally required for the result. We stress that each condition is weak. Below, we collect sufficient conditions for the conclusions of Propositions 2.1 and 2.2. We note that Propositions 2.1 and 2.2 are nonasymptotic results stated for each dataset. Thus, the following conditions are for the datasets that are used to construct the sB-SBF estimators.

CONDITION (B).

(B1) (*Hilbertian predictors*: $1 \leq j \leq \varrho$). $\kappa_j := \sup_{x_j \in \mathcal{X}_j} \min_{1 \leq i \leq n} d_j(x_j, X_{ij}) < h_j$, and K_j is bounded away from zero on $[0, \kappa_j/h_j]$. Also, for all $x_j \in \mathcal{X}_j$,

$$(3.4) \qquad \text{Leb}_{q_j}((\eta_j(\mathcal{X}_j) - \eta_j(x_j))/h_j \cap B(\mathbf{0}_{q_j}, \kappa_j/h_j)) > 0.$$

(B2) (*Riemannian predictors*: $\varrho + 1 \leq j \leq d$). $\gamma_j := \sup_{x_j \in \mathcal{X}_j} \min_{1 \leq i \leq n} d_j(x_j, X_{ij}) < h_j < \text{inj}(\mathcal{X}_j)$, and K_j is bounded away from zero on $[0, \gamma_j/h_j]$.

We note that the condition (B) implies (S) and that $\kappa_j < \infty$ and $\gamma_j < \infty$ since $d_j(\cdot, X_{ij}) : \mathcal{X}_j \to [0, \infty)$ are continuous and $\mathcal{X}_j$ are compact. The condition that the functions K_j are bounded away from zero on $[0, t]$ for some $0 < t < 1$ is standard in kernel smoothing theory. The latter is satisfied by most popular kernel functions; see Remark 4.1 in Section 4.1 for specific examples. Under (B1), the denominator of $K_{h_j}(x_j, X_{ij})$ as defined at (3.1) is nonzero. We also note that a large class of $\mathcal{X}_j$ satisfies (3.4) in (B1). Below, we introduce a proposition that gives a class of $\mathcal{X}_j$ satisfying the condition.

DEFINITION 3.1. We say a connected Lebesgue measurable set S in a Euclidean space is *fat* if, for any $s \in S$, the intersection of S and any neighborhood of s has a positive Lebesgue measure.

We may show that $\text{Leb}_{q_j}((\eta_j(\mathcal{X}_j) - \eta_j(x_j))/h_j \cap B(\mathbf{0}_{q_j}, r)) > 0$ for any $x_j \in \mathcal{X}_j$, $h_j > 0$ and $r > 0$ if $\eta_j(\mathcal{X}_j)$ is fat. Most of connected sets in Euclidean spaces are fat. Examples include polygons in $\mathbb{R}^2$. The following proposition gives another large class of fat sets.

PROPOSITION 3.1. *If $\mathcal{X}_j$ is a convex subset of $\mathcal{H}_j$, then $\eta_j(\mathcal{X}_j)$ is fat.*

Since the condition (B) implies the condition (S), we obtain the following proposition that demonstrates the existence and uniqueness of the sB-SBF estimator and the convergence of the sB-SBF algorithm.

PROPOSITION 3.2. *Under the condition (B), the conclusions of Propositions 2.1 and 2.2 are valid.*

For the validity of the asymptotic algorithm convergence, as stated in Proposition 2.3, we give sufficient conditions that guarantee the high-level condition (A). Let $\partial \mathcal{X}_j$, for $1 \leq j \leq \varrho$, denote the boundary of $\mathcal{X}_j$.

CONDITION (C).

(C1) $\text{E}(\|\mathbf{Y}\|^\alpha) < \infty$ *for some $\alpha > 2$, and $\text{E}(\|\mathbf{Y}\|^2 \mid X_j = \cdot)$ are bounded on the respective $\mathcal{X}_j$ for $1 \leq j \leq d$.*

(C2) *For $1 \leq j \neq k \leq d$, (i) $p_j > 0$ on $\mathcal{X}_j$ and p_{jk} is continuous on $\mathcal{X}_j \times \mathcal{X}_k$; (ii) the bandwidths satisfy $h_j = o(1)$, $\log n/(nh_j^{q_j} h_k^{q_k}) = o(1)$ and $\inf_n n^{c_j} h_j^{q_j} > 0$ for some $c_j < (\alpha - 2)/\alpha$.*

(C3) (*Hilbertian predictors*: $1 \leq j \leq \varrho$). (i) $\mu_j(\partial \mathcal{X}_j) = 0$ *and there exist positive constants $\tau_j < 1$ and ϵ_j such that*

$$(3.5) \qquad \inf_n \inf_{x_j \in \mathcal{X}_j} \mathrm{Leb}_{q_j}\big((\eta_j(\mathcal{X}_j) - \eta_j(x_j))/h_j \cap B(\mathbf{0}_{q_j}, \tau_j)\big) \geq \epsilon_j;$$

(ii) K_j *is Lipschitz continuous on $[0, \infty)$ and bounded away from zero on $[0, \tau_j]$.*

(C4) (*Riemannian predictors*: $\varrho + 1 \leq j \leq d$). (i) *there exist positive constants $r_j < \mathrm{inj}(\mathcal{X}_j)$ and L_j such that, for all $u_j \in \mathcal{X}_j$ and $x_j, x_j^* \in B_{\mathcal{X}_j}(u_j, r_j)$,*

$$\big|(\theta_j(x_j; u_j))^{-1} - (\theta_j(x_j^*; u_j))^{-1}\big| \leq L_j \cdot d_j(x_j, x_j^*);$$

(ii) K_j *is Lipschitz continuous.*

We note that the condition (C) implies (A) except (A4) and (A7). It is for asymptotic analysis while the condition (B) is for nonasymptotic analysis. Hence, the two sets (B) and (C) are of different nature. For example, (B) is imposed on each dataset while (C) is on the population generating the observations of $(\mathbf{X}, \mathbf{Y})$. Nevertheless, they are related. Indeed, one may show that the condition (C2)-(i) with $\inf_n n^{c_j} h_j^{q_j} > 0$ for some $c_j < (\alpha - 2)/\alpha$ in (C2)-(ii) implies that the condition $\sup_{x_j \in \mathcal{X}_j} \min_{1 \leq i \leq n} d_j(x_j, X_{ij}) < h_j$ in (B) holds with probability tending to one. For the proof of the latter implication, one may need to use the fact that there exists a constant $0 < c(q_j) < \infty$ depending only on q_j such that, for μ_j being either the pushforward measure or the Riemannian volume measure, $\inf_{x_j \in \mathcal{X}_j} \mu_j(B_{\mathcal{X}_j}(x_j, \delta)) \geq c(q_j) \cdot \delta^{q_j}$ for sufficiently small $\delta > 0$.

The conditions (C1)–(C2) are versions of the standard regularity conditions in Euclidean kernel smoothing theory. In the condition (C1), $\mathrm{E}(\|\mathbf{Y}\|^2 \mid X_j = \cdot)$ is bounded on $\mathcal{X}_j$ if it is continuous on $\mathcal{X}_j$. The condition $\mu_j(\partial \mathcal{X}_j) = 0$ in (C3) is weak and holds for most of sets. For example, the convexity of $\mathcal{X}_j$ implies the condition; see [22]. We may also show that (3.5) in (C3) holds when $\mathcal{X}_j$ is convex, using Proposition 3.1 and the fact that the map $x_j \mapsto \mathrm{Leb}_{q_j}((\eta_j(\mathcal{X}_j) - \eta_j(x_j))/h_j \cap B(\mathbf{0}_{q_j}, r))$ is continuous on $\mathcal{X}_j$ for any $h_j > 0$ and $r > 0$. The condition on $(\theta_j(\cdot; u_j))^{-1}$ in (C4) is weaker than the Lipschitz continuity of $(\theta_j(\cdot; u_j))^{-1}$. The latter condition is satisfied for all Riemannian manifolds considered in Section 3.3.2.

PROPOSITION 3.3. *Under the conditions* (C), *and* (A4) *and* (A7) *in Section* 2.2, *the consequences of Proposition* 2.3 *hold.*

3.3. *Implementation of sB-SBF algorithm.* We discuss the implementation of the sB-SBF algorithm for Hilbertian and Riemannian predictors. Here, we collect all necessary ingredients including relevant theory for implementing the integrals in (2.9). They are very useful not only for the problems discussed here but also for others that require evaluating integrals over nonstandard domains with nonstandard measures. Some specific results presented here are not well known and do not exist in the literature, to the best of our knowledge.

3.3.1. *Implementation for Hilbertian predictors.* In the case of $\mathcal{X}_k \subset \mathcal{H}_k$, we may implement the integral over $\mathcal{X}_k$ in (2.9) using the property

$$\int_{\mathcal{X}_k} f(x_k) \, d\mu_k(x_k) = \int_{\eta_k(\mathcal{X}_k)} f\big(\eta_k^{-1}(\mathbf{t})\big) \, d\mathbf{t},$$

where f is any integrable function on $\mathcal{X}_k$. Below, we give detailed examples of the implementation.

3.3.1.1. (i) *Euclidean spaces.* In the case of $\mathcal{X}_k \subset \mathbb{R}^{q_k}$, we may take η_k to be the identity map on $\mathcal{X}_k$. Then it is a standard practice to implement the integral over $\mathcal{X}_k$ with respect to $\mu_k = \text{Leb}_{q_k}$ in (2.9) since it is the usual numerical integration.

3.3.1.2. (ii) *Simplices.* Consider the q_k-dimensional simplex $\mathcal{S}_1^{q_k+1}$ defined by

$$\mathcal{S}_1^{q_k+1} = \left\{ x_k = (x_{k,1}, \dots, x_{k,q_k+1}) \in (0,1)^{q_k+1} : \sum_{l=1}^{q_k+1} x_{k,l} = 1 \right\}.$$

The inner product $\langle \cdot, \cdot \rangle_k$ on $\mathcal{S}_1^{q_k+1}$ is given by

$$\langle x_k, u_k \rangle_k = \frac{1}{2(q_k+1)} \sum_{l=1}^{q_k+1} \sum_{l'=1}^{q_k+1} \log\left(\frac{x_{k,l}}{x_{k,l'}} \right) \log\left(\frac{u_{k,l}}{u_{k,l'}} \right).$$

Let $\{e_1, \dots, e_{q_k}\}$ be its orthonormal basis. In this case, we may take η_k to be the isometric log-ratio transformation $\text{ilr}_k : \mathcal{S}_1^{q_k+1} \to \mathbb{R}^{q_k}$ defined by

$$\text{ilr}_k(x_k) = (\langle x_k, e_1 \rangle_k, \dots, \langle x_k, e_{q_k} \rangle_k),$$

and use the Aitchison measure μ_k defined by $\mu_k(A) = \text{Leb}_{q_k}(\text{ilr}_k(A))$. The notion of Aitchison measure has been studied in density estimation and regression analysis for compositional data; see [1] and [7], for example. The μ_k is a natural measure on the simplex $\mathcal{S}_1^{q_k+1}$, which inherits the properties of the corresponding Lebesgue measure. One of the most important properties of the Aitchison measure is that integrals with respect to the measure reduce to Lebesgue integrals. Specifically, for any integrable function f on $\mathcal{S}_1^{q_k+1}$ and any Borel set $A \subset \mathcal{S}_1^{q_k+1}$, it holds that

$$\int_A f(x_k) \, d\mu_k(x_k) = \int_{\text{ilr}_k(A)} f(\text{ilr}_k^{-1}(\mathbf{t})) \, d\mathbf{t}.$$

To introduce an orthonormal basis of $\mathcal{S}_1^{q_k+1}$, define the $(q_k + 1)$-dimensional vector

$$\xi_m = \left(\exp((m(m+1))^{-1/2}) \mathbf{1}_m, \exp(-(m/(m+1))^{1/2}), \mathbf{1}_{q_k-m} \right),$$

where $\mathbf{1}_l$ is l-vector with all entries being equal to one. Denote the lth entry of ξ_m by $\xi_{m,l}$. Define $e_m = (\xi_{m,1}/\sum_{l=1}^{q_k+1} \xi_{m,l}, \dots, \xi_{m,q_k+1}/\sum_{l=1}^{q_k+1} \xi_{m,l})$. Then $\{e_1, \dots, e_{q_k}\}$ forms an orthonormal basis of $\mathcal{S}_1^{q_k+1}$; see Proposition 2 of [9]. With this basis, the mth entry of $\text{ilr}_k(x_k) \in \mathbb{R}^{q_k}$ for $x_k = (x_{k,1}, \dots, x_{k,q_k+1}) \in \mathcal{S}_1^{q_k+1}$ is given by $\sqrt{m/(m+1)} \log(\prod_{l=1}^m x_{k,l}^{1/m}/x_{k,m+1})$. Then the integral over $\mathcal{X}_k \subset \mathcal{S}_1^{q_k+1}$ with respect to the Aitchison measure μ_k in (2.9) can be evaluated by the usual numerical integration:

$$\int_{\mathcal{X}_k} w_{ik}^{[r]}(x_k) \cdot \hat{p}_{jk}(x_j, x_k) \, d\mu_k(x_k) = \int_{\text{ilr}_k(\mathcal{X}_k)} w_{ik}^{[r]}(\text{ilr}_k^{-1}(\mathbf{t})) \cdot \hat{p}_{jk}(x_j, \text{ilr}_k^{-1}(\mathbf{t})) \, d\mathbf{t}.$$

3.3.1.3. (iii) *Product of Hilbert spaces.* Suppose that $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_k)$ itself is the product of Hilbert spaces $((\mathcal{H}_{k,j}, \langle \cdot, \cdot \rangle_{k,j}) : 1 \leq j \leq J)$ and $\mathcal{X}_k \subset \mathcal{H}_k$, where $\dim(\mathcal{H}_{k,j}) = q_{k,j}$ and $J \in \mathbb{N}$. We note that this space enables us to accommodate interaction effects between Hilbertian predictors. In this case, the inner product is given by $\langle x_k, u_k \rangle_k = \sum_{j=1}^J \langle x_{k,j}, u_{k,j} \rangle_{k,j}$ and the isometric isomorphism $\eta_k : \mathcal{H}_k \to \mathbb{R}^{\sum_{j=1}^J q_{k,j}}$ is given by $\eta_k(x_k) = (\eta_{k,1}(x_{k,1}), \dots, \eta_{k,J}(x_{k,J}))$, where $\eta_{k,j} : \mathcal{H}_{k,j} \to \mathbb{R}^{q_{k,j}}$ are isometric isomorphisms. With this η_k and the corresponding pushforward measure, we may implement the integral over $\mathcal{X}_k$ in (2.9).

3.3.2. *Implementation for Riemannian predictors.* We first introduce Riemannian volume measure, which takes the role of the measure μ_j on $\mathcal{B}(\mathcal{X}_j)$ when $\mathcal{X}_j$ is a Riemannian manifold. Let $(\mathcal{M}, g)$ be a m-dimensional compact Riemannian manifold. A pair (U, φ) of an open set $U \subset \mathcal{M}$ and a homeomorphism $\varphi : U \to \varphi(U) \subset \mathbb{R}^m$ is called a *chart* of $\mathcal{M}$. Let $\{(U_\beta, \varphi_\beta)\}_\beta$ be a finite collection of charts in an atlas for $\mathcal{M}$ such that $\mathcal{M} = \bigcup_\beta U_\beta$. Such a finite collection exists since $\mathcal{M}$ is compact. We write $\varphi_\beta = (\varphi_{\beta,1}, \ldots, \varphi_{\beta,m})$, where $\varphi_{\beta,l} : U \to \mathbb{R}$. Let $\{\rho_\beta\}_\beta$ be a partition of unity subordinate to $\{U_\beta\}_\beta$. Also, let $\frac{\partial}{\partial \varphi_{\beta,l}}|_{\varphi_\beta^{-1}(\mathbf{t})}$ denote the function defined by $\frac{\partial}{\partial \varphi_{\beta,l}}|_{\varphi_\beta^{-1}(\mathbf{t})}(h) = D_l(h \circ \varphi_\beta^{-1})(\mathbf{t})$ for each smooth function $h : U_\beta \to \mathbb{R}$, where D_l for $1 \leq l \leq m$ are the usual partial differential operators acting on real-valued functions on $\mathbb{R}^m$. Define $G_\beta(\varphi_\beta^{-1}(\mathbf{t}))$ to be the $m \times m$ matrix whose (l, l')th entry is given by

$$g(\varphi_\beta^{-1}(\mathbf{t}))\left(\frac{\partial}{\partial \varphi_{\beta,l}}\bigg|_{\varphi_\beta^{-1}(\mathbf{t})}, \frac{\partial}{\partial \varphi_{\beta,l'}}\bigg|_{\varphi_\beta^{-1}(\mathbf{t})}\right).$$

Let $C(\mathcal{M}, \mathbb{R})$ denote the class of real-valued continuous functions defined on $\mathcal{M}$. Define $F : C(\mathcal{M}, \mathbb{R}) \to \mathbb{R}$ by

$$(3.6) \qquad F(f) = \sum_\beta \int_{\varphi_\beta(U_\beta)} f(\varphi_\beta^{-1}(\mathbf{t})) \cdot \rho_\beta(\varphi_\beta^{-1}(\mathbf{t})) \cdot \sqrt{\det(G_\beta(\varphi_\beta^{-1}(\mathbf{t})))} \, d\mathbf{t}.$$

The value of $F(f)$ is independent of the choices of $\{(U_\beta, \varphi_\beta)\}_\beta$ and $\{\rho_\beta\}_\beta$; see Chapter 2 of [38], for example. Due to the Riesz representation theorem, there exists a unique regular Borel measure ν on $\mathcal{M}$ such that

$$(3.7) \qquad F(f) = \int_{\mathcal{M}} f(x) \, d\nu(x), \quad f \in C(\mathcal{M}, \mathbb{R}).$$

We call such a ν the *Riemannian volume measure induced by g*.

According to (3.7), the integral of a continuous function over $\mathcal{X}_j$ with respect to the corresponding Riemannian volume measure μ_j can be converted to an integral over an Euclidean space with respect to a Lebesgue measure. For the application of this idea to the integrals in (2.9), we need to verify that the functions involved there are continuous on the respective spaces. For this, we have the following proposition.

PROPOSITION 3.4. *Assume the condition* (B). *Then the maps* $x_k \mapsto w_{ik}^{[r]}(x_k) \cdot \hat{p}_{jk}(x_j, x_k)$ *are continuous on the respective* $\mathcal{X}_k$ *for all* $x_j \in \mathcal{X}_j$ *and for all* $r \geq 0$, $1 \leq i \leq n$, $1 \leq j \leq d$ *and* $\varrho + 1 \leq k \leq d$ *with* $k \neq j$.

For $\varrho + 1 \leq k \leq d$, let $\{(U_{\beta_k}, \varphi_{\beta_k})\}_{\beta_k}$ be a finite collection of charts in an atlas for the Riemannian manifold $\mathcal{X}_k$ such that $\mathcal{X}_k = \bigcup_{\beta_k} U_{\beta_k}$, and $\{\rho_{\beta_k}\}_{\beta_k}$ be a partition of unity subordinate to $\{U_{\beta_k}\}_{\beta_k}$. For the integrals in (2.9) with $1 \leq j \leq d$ and $\varrho + 1 \leq k \leq d$ but $j \neq k$, making use of (3.7) and Proposition 3.4, we get that, under the condition (B),

$$(3.8) \quad \int_{\mathcal{X}_k} w_{ik}^{[r]}(x_k) \cdot \hat{p}_{jk}(x_j, x_k) \, d\mu_k(x_k) = \sum_{\beta_k} \int_{\varphi_{\beta_k}(U_{\beta_k})} w_{ik}^{[r]}(\varphi_{\beta_k}^{-1}(\mathbf{t})) \cdot \hat{p}_{jk}(x_j, \varphi_{\beta_k}^{-1}(\mathbf{t}))$$
$$\cdot \rho_{\beta_k}(\varphi_{\beta_k}^{-1}(\mathbf{t})) \cdot \sqrt{\det(G_{\beta_k}(\varphi_{\beta_k}^{-1}(\mathbf{t})))} \, d\mathbf{t},$$

where $G_{\beta_k}(\varphi_{\beta_k}^{-1}(\mathbf{t}))$ is the $q_k \times q_k$ matrix whose (l, l')th entry is given by

$$g_k(\varphi_{\beta_k}^{-1}(\mathbf{t}))\left(\frac{\partial}{\partial \varphi_{\beta_k,l}}\bigg|_{\varphi_{\beta_k}^{-1}(\mathbf{t})}, \frac{\partial}{\partial \varphi_{\beta_k,l'}}\bigg|_{\varphi_{\beta_k}^{-1}(\mathbf{t})}\right).$$

We derive more concrete forms for the integral on the right hand side of (3.8) in several important classes of Riemannian manifolds.

3.3.2.1. (i) *Toruses.* Consider the case where $\mathcal{X}_k = \prod_{l=1}^{q_k} \mathbb{S}^1$ for $q_k \geq 1$. Then $\mathrm{inj}(\mathcal{X}_k) = \pi$ and

$$d_k(x_k, u_k) = \left(\sum_{l=1}^{q_k} \arccos^2(x_{k,l}^\top u_{k,l})\right)^{1/2}.$$

In this case, we may prove that $\theta_k(\cdot; \cdot) \equiv 1$. Clearly, this volume density function satisfies (C4). Define a map $\varphi^{-1} : (0, 2\pi)^{q_k} \to \prod_{l=1}^{q_k} \mathbb{S}^1$ by

$$\varphi^{-1}(\vartheta_1, \ldots, \vartheta_{q_k}) = (\cos \vartheta_1, \sin \vartheta_1, \ldots, \cos \vartheta_{q_k}, \sin \vartheta_{q_k}).$$

Then $(\varphi^{-1}((0, 2\pi)^{q_k}), \varphi)$ is a chart on $\prod_{l=1}^{q_k} \mathbb{S}^1$, where φ is understood to be the inverse of the map φ^{-1}. Assume the condition (B). Then, since $\mu(\mathcal{X}_k) = \mu_k(\varphi^{-1}((0, 2\pi)^{q_k}))$, it holds that

$$\int_{\mathcal{X}_k} w_{ik}^{[r]}(x_k) \cdot \hat{p}_{jk}(x_j, x_k) \, d\mu_k(x_k)$$

$$= \int_{(0,2\pi)^{q_k}} w_{ik}^{[r]}(\varphi^{-1}(\vartheta_1, \ldots, \vartheta_{q_k})) \cdot \hat{p}_{jk}(x_j, \varphi^{-1}(\vartheta_1, \ldots, \vartheta_{q_k})) \, d\vartheta_1 \cdots d\vartheta_{q_k};$$

see Chapter 16 of [41].

3.3.2.2. (ii) *Spheres.* Consider now the case $\mathcal{X}_k = \mathbb{S}^{q_k} \subset \mathbb{R}^{q_k+1}$ for $q_k \geq 2$. In this case, $\mathrm{inj}(\mathcal{X}_k) = \pi$ and $d_k(x_k, u_k) = \arccos(x_k^\top u_k)$. Also, we may prove

$$\theta_k(x_k; u_k) = \left(\sin(d_k(x_k, u_k))/d_k(x_k, u_k)\right)^{q_k-1} \quad \text{for } x_k \neq \pm u_k$$

and $\theta_k(x_k; x_k) = 1$. This volume density function satisfies the condition (C4). Define a map $\varphi^{-1} : (0, 2\pi) \times (0, \pi)^{q_k-1} \to \mathbb{S}^{q_k}$ by

$$\varphi^{-1}(\vartheta, \phi_1, \ldots, \phi_{q_k-1}) = \left(\cos \vartheta \prod_{j=1}^{q_k-1} \sin \phi_j, \sin \vartheta \prod_{j=1}^{q_k-1} \sin \phi_j, \cos \phi_1 \prod_{j=2}^{q_k-1} \sin \phi_j, \right.$$

$$\left. \cos \phi_2 \prod_{j=3}^{q_k-1} \sin \phi_j, \ldots, \cos \phi_{q_k-2} \sin \phi_{q_k-1}, \cos \phi_{q_k-1}\right).$$

Then $(\varphi^{-1}((0, 2\pi) \times (0, \pi)^{q_k-1}), \varphi)$ is a chart on $\mathbb{S}^{q_k}$. Assume the condition (B). Since $\mu_k(\mathcal{X}_k) = \mu_k(\varphi^{-1}((0, 2\pi) \times (0, \pi)^{q_k-1}))$, we have

$$\int_{\mathcal{X}_k} w_{ik}^{[r]}(x_k) \cdot \hat{p}_{jk}(x_j, x_k) \, d\mu_k(x_k)$$

$$= \int_{(0,2\pi) \times (0,\pi)^{q_k-1}} w_{ik}^{[r]}(\varphi^{-1}(\vartheta, \phi_1, \ldots, \phi_{q_k-1})) \cdot \hat{p}_{jk}(x_j, \varphi^{-1}(\vartheta, \phi_1, \ldots, \phi_{q_k-1}))$$

$$\cdot \prod_{j=1}^{q_k-1} (\sin \phi_j)^j \, d\vartheta \, d\phi_1 \cdots d\phi_{q_k-1};$$

see Chapter 16 of [41].

3.3.2.3. (iii) *Planar shape spaces.* Here, we consider $\mathcal{X}_k = \Sigma_2^b$ for $b \geq 3$, where Σ_2^b denotes the planar shape space with $b \equiv b(k)$ landmarks, suppressing the dependence of b on k for simplicity. Roughly speaking, an element of Σ_2^b is obtained by removing the location, scale and rotation of a shape in $\mathbb{R}^2$. For a concrete definition, let $\{(z_{1,1}, z_{1,2}), \ldots, (z_{b,1}, z_{b,2})\} \subset \mathbb{R}^2$ be b landmarks of a shape and $\mathbf{z} = (z_{1,1} + \sqrt{-1} \cdot z_{1,2}, \ldots, z_{b,1} + \sqrt{-1} \cdot z_{b,2})^\top \in \mathbb{C}^b$ the complex representation of the landmarks. Also, let H be the $(b-1) \times b$ Helmert submatrix whose

jth row is $(-1/\sqrt{j(j+1)}\mathbf{1}_j^\top, \sqrt{j/(j+1)}, \mathbf{0}_{b-j-1}^\top)$. Define $\mathbf{z}_H = H\mathbf{z}/\|H\mathbf{z}\|_{\mathbb{C}^{b-1}}$, where $\|\cdot\|_{\mathbb{C}^{b-1}}$ denotes the complex norm on $\mathbb{C}^{b-1}$. Consider equivalent classes $[\mathbf{z}] = \{\mathbf{z}_H \cdot e^{\sqrt{-1}\cdot\vartheta} : 0 \le \vartheta < 2\pi\}$. Then Σ_2^b is defined as the collection of these equivalent classes of b landmarks. It is known that Σ_2^b is a $(2b-4)$-dimensional connected and compact Riemannian manifold without boundary. The Riemannian distance function on Σ_2^b is given by

$$d_k([\mathbf{z}], [\mathbf{w}]) = \arccos(\|\overline{\mathbf{z}_H}^\top \mathbf{w}_H\|_{\mathbb{C}}),$$

where $\overline{\mathbf{z}}$ denotes the conjugate of $\mathbf{z}$, and $\mathrm{inj}(\Sigma_2^b) = \pi/2$. Also,

$$\theta_k([\mathbf{z}]; [\mathbf{w}]) = (\sin(2d_k([\mathbf{z}], [\mathbf{w}]))/(2d_k([\mathbf{z}], [\mathbf{w}])))^{2b-5} \quad \text{if } [\mathbf{z}] \neq [\mathbf{w}]$$

and $\theta_k([\mathbf{z}]; [\mathbf{z}]) = 1$, as proved by [40]. This volume density function satisfies (C4). We note that $\mathbf{z}_H = (\sqrt{s_1^{\mathbf{z}}} \cdot e^{\sqrt{-1}\cdot\vartheta_1^{\mathbf{z}}}, \ldots, \sqrt{s_{b-1}^{\mathbf{z}}} \cdot e^{\sqrt{-1}\cdot\vartheta_{b-1}^{\mathbf{z}}})$ for some

$$\mathbf{s}^{\mathbf{z}} := (s_1^{\mathbf{z}}, \ldots, s_{b-2}^{\mathbf{z}}) \in \left\{ (a_1, \ldots, a_{b-2}) \in [0, 1]^{b-2} : \sum_{j=1}^{b-2} a_j \le 1 \right\} =: \mathcal{S}_{b-2}$$

with $s_{b-1}^{\mathbf{z}} = 1 - \sum_{j=1}^{b-2} s_j^{\mathbf{z}}$ and $\vartheta_j^{\mathbf{z}} \in [0, 2\pi)$ for all $1 \le j \le b - 1$. Let $\phi_j^{\mathbf{z}} = (\vartheta_j^{\mathbf{z}} - \vartheta_{b-1}^{\mathbf{z}}) \mod(2\pi)$, $1 \le j \le b - 2$, and $\boldsymbol{\phi}^{\mathbf{z}} = (\phi_1^{\mathbf{z}}, \ldots, \phi_{b-2}^{\mathbf{z}}) \in [0, 2\pi)^{b-2}$. For the interior $\mathcal{S}_{b-2}^o$ of $\mathcal{S}_{b-2}$, define a map $\varphi^{-1} : \mathcal{S}_{b-2}^o \times (0, 2\pi)^{b-2} \to \Sigma_2^b$ by $\varphi^{-1}(\mathbf{s}^{\mathbf{z}}, \boldsymbol{\phi}^{\mathbf{z}}) = [\mathbf{z}]$. Then $(\varphi^{-1}(\mathcal{S}_{b-2}^o \times (0, 2\pi)^{b-2}), \varphi)$ is a chart on Σ_2^b [20]. Since $\mu_k(\mathcal{X}_k) = \mu_k(\varphi^{-1}(\mathcal{S}_{b-2}^o \times (0, 2\pi)^{b-2}))$, under the condition (B), we get

$$\int_{\mathcal{X}_k} w_{ik}^{[r]}([\mathbf{z}]) \cdot \hat{p}_{jk}(x_j, [\mathbf{z}]) \, d\mu_k([\mathbf{z}])$$

$$= \int_{\mathcal{S}_{b-2}^o \times (0, 2\pi)^{b-2}} w_{ik}^{[r]}(\varphi^{-1}(\mathbf{s}^{\mathbf{z}}, \boldsymbol{\phi}^{\mathbf{z}})) \cdot \hat{p}_{jk}(x_j, \varphi^{-1}(\mathbf{s}^{\mathbf{z}}, \boldsymbol{\phi}^{\mathbf{z}})) \cdot 2^{2-b} \, d\mathbf{s}^{\mathbf{z}} \, d\boldsymbol{\phi}^{\mathbf{z}};$$

see Chapter 4 of [8].

3.3.2.4. (iv) *Product Riemannian manifolds.* Suppose that $\mathcal{X}_k$ is the product of connected, compact and boundaryless Riemannian manifolds $(\mathcal{M}_j, g_{\mathcal{M}_j}, d_{\mathcal{M}_j})$, $1 \le j \le J$, where $\dim(\mathcal{M}_j) = m_j$ and $J \in \mathbb{N}$. Then $\mathcal{X}_k$ is again a $\sum_{j=1}^{J} m_j$-dimensional connected, compact and boundaryless Riemannian manifold with the corresponding product Riemannian metric. With this type of $\mathcal{X}_k$, we may handle interaction effects between Riemannian predictors. The Riemannian distance function is given by $d_\otimes(\mathbf{x}, \mathbf{u}) = \{\sum_{j=1}^{J} (d_{\mathcal{M}_j}(x_j, u_j))^2\}^{1/2}$ for $\mathbf{x} = (x_1, \ldots, x_J)$ and $\mathbf{u} = (u_1, \ldots, u_J)$ in $\mathcal{M}_\otimes := \mathcal{M}_1 \times \cdots \times \mathcal{M}_J$. Let $\theta_j(\cdot; \cdot)$ and $\mathrm{inj}(\mathcal{M}_j)$, respectively, denote the volume density function and injectivity radius of $\mathcal{M}_j$. Also, let $\theta_\otimes$ and $\mathrm{inj}(\mathcal{M}_\otimes)$ be those for $\mathcal{M}_\otimes$. Below, we present a new proposition.

PROPOSITION 3.5. *It holds that*

$$\theta_\otimes(\mathbf{x}, \mathbf{u}) = \prod_{j=1}^{J} \theta_j(x_j; u_j), \qquad \mathrm{inj}(\mathcal{M}_\otimes) \ge \min_{1 \le j \le J} \mathrm{inj}(\mathcal{M}_j).$$

Suppose we are given an atlas $\{(U_{\beta_j}, \varphi_{\beta_j})\}_{\beta_j}$ for each $\mathcal{M}_j$, $1 \le j \le J$. Define $\prod_{j=1}^{J} \varphi_{\beta_j} : \prod_{j=1}^{J} U_{\beta_j} \to \mathbb{R}^{\sum_{j=1}^{J} m_j}$ by

$$\prod_{j=1}^{J} \varphi_{\beta_j}(u_1, \ldots, u_J) = (\varphi_{\beta_1}(u_1), \ldots, \varphi_{\beta_J}(u_J)).$$

Then $\{(\prod_{j=1}^{J} U_{\beta_j}, \prod_{j=1}^{J} \varphi_{\beta_j})\}_{\beta_1,\ldots,\beta_J}$ is an atlas for $\mathcal{M}_{\otimes}$. Based on the preceding discussion, we may derive a more concrete form for the integral at (3.8) specific to each product Riemannian manifold under consideration.

4. Theory for Hilbertian and Riemannian predictors.

In this section, we discuss the rates of convergence and the asymptotic distributions of the individual estimated component maps $\hat{\mathbf{f}}_j$ and the estimated regression map $\hat{\mathbf{f}} = \bar{\mathbf{Y}} \oplus \hat{\mathbf{f}}_1 \oplus \cdots \oplus \hat{\mathbf{f}}_d$.

4.1. *Rates of convergence.* One of the main ideas in the derivation of the rates of convergence here and the asymptotic distributions in Section 4.2 for Riemannian predictors is to convert the integrals of Hilbert space-valued maps defined on Riemannian manifolds, to those of the corresponding Hilbert space-valued maps defined on Euclidean spaces. The following theorem materializes the idea, which is new and an extended version of (3.7) for Hilbert space-valued maps. We present it here since the extension is of interest in its own right. To state the theorem, consider the class $C(\mathcal{M}, \mathbb{H})$ of $\mathbb{H}$-valued continuous maps defined on a m-dimensional compact Riemannian manifold $(\mathcal{M}, g)$. Let $\mathbf{F} : C(\mathcal{M}, \mathbb{H}) \to \mathbb{H}$ be a map defined by

$$(4.1) \qquad \mathbf{F}(\mathbf{f}) = \bigoplus_{\beta} \int_{\varphi_{\beta}(U_{\beta})} \mathbf{f}(\varphi_{\beta}^{-1}(\mathbf{t})) \odot \left(\rho_{\beta}(\varphi_{\beta}^{-1}(\mathbf{t})) \cdot \sqrt{\det(G_{\beta}(\varphi_{\beta}^{-1}(\mathbf{t})))} \right) d\mathbf{t},$$

where the integration is in the Bochner sense and with respect to the m-dimensional Lebesgue measure, and $(U_{\beta}, \varphi_{\beta})$, ρ_{β} and G_{β} are defined as in Section 3.3.2.

THEOREM 4.1. *The map $\mathbf{F}$ as defined at (4.1) is independent of the choices of charts and partition of unity. Moreover, there exists a unique regular Borel measure v on $\mathcal{M}$ such that $\int_{\mathcal{M}} \mathbf{f}(x) \, dv(x) = \mathbf{F}(\mathbf{f})$ for all $\mathbf{f} \in C(\mathcal{M}, \mathbb{H})$, and v equals the Riemannian volume measure induced by g.*

We now derive the rates of convergence for two types of component maps. The first is for Hölder continuous maps and the second is for Fréchet differentiable maps. The former type has not been considered even for real-valued responses and predictors in the existing SBF literature. An $\mathbb{H}$-valued map $\mathbf{f}$ defined on a semimetric space $(\mathcal{S}, d_{\mathcal{S}})$ is called Hölder continuous with exponent $b > 0$ if there exists a constant $L > 0$ such that $\|\mathbf{f}(s) \ominus \mathbf{f}(s^*)\| \le L \cdot (d_{\mathcal{S}}(s, s^*))^b$ for all $s, s^* \in \mathcal{S}$. The following theorem gives the rates of convergence for Hölder continuous component maps. For concise presentation, let $\delta_{1,j} = (nh_j^{q_j})^{-1/2}$ and $\delta_{2,j} = (nh_j^{q_j})^{-1/2}(\log n)^{1/2}$.

THEOREM 4.2. *Assume the condition (C) in Section 3.2 and that $\mathbf{f}_j$ are Hölder continuous with respective exponent $b_j > 0$ for $1 \le j \le d$. Then*

$$\left(\int_{\mathcal{X}} \|\hat{\mathbf{f}}(\mathbf{x}) \ominus \mathbf{f}(\mathbf{x})\|^2 p(\mathbf{x}) \, d\mu(\mathbf{x}) \right)^{1/2} = O_P \left(\sum_{j=1}^{d} (h_j^{b_j} + \delta_{1,j}) \right).$$

If we further assume the condition (A7) in Section 2.2, then for all $1 \le j \le d$,

$$\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j) = O_P \left(\sum_{j=1}^{d} (h_j^{b_j} + \delta_{1,j}) \right) \quad \text{for all } x_j \in \mathcal{X}_j,$$

$$\left(\int_{\mathcal{X}_j} \|\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j)\|^2 p_j(x_j) \, d\mu_j(x_j)\right)^{1/2} = O_p\left(\sum_{j=1}^d (h_j^{b_j} + \delta_{1,j})\right),$$

$$\sup_{x_j \in \mathcal{X}_j} \|\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j)\| = O_p\left(\sum_{j=1}^d h_j^{b_j} + \sum_{k \neq j} \delta_{1,k} + \delta_{2,j}\right).$$

Since $\dim(\prod_{j=1}^\varrho \mathcal{H}_j) = \sum_{j=1}^\varrho q_j$ and $\dim(\prod_{j=\varrho+1}^d \mathcal{X}_j) = \sum_{j=\varrho+1}^d q_j$, full-dimensional nonparametric estimators suffer from the curse of dimensionality as the number of predictors increases. However, Theorem 4.2 demonstrates that our estimators do not since the rates depend on individual q_j not the sum of them. In fact, when $b_j = b$, $q_j = q$ and $h_j \asymp n^{-1/(q+2b)}$ for all $1 \leq j \leq d$, our estimators achieve the optimal q-dimensional rates of convergence in the pointwise, L_2 and uniform modes. It also shows that the usual assumption (A7) is not required for the L_2 error rate of the regression estimator $\hat{\mathbf{f}}$. The latter is not noted in the existing SBF literature even for real-valued responses and predictors.

Now, we give the rates of convergence for Fréchet differentiable type component maps. For the Hilbertian domains $\mathcal{X}_j$, which are the compact subsets of the Hilbert spaces $\mathcal{H}_j$, we define a new notion of interior regions.

DEFINITION 4.1.   Let S be a subset of a semimetric space $\mathcal{S}$. For $\varepsilon > 0$, we call $s \in S$ an ε-interior point of S if the open ball centered at s with radius ε is contained in S. We call $s \in S$ an ε-boundary point of S if s is not an ε-interior point of S.

We also extend the notion of differentiation for Hilbert-space-valued maps defined on Riemannian predictors. Recall that a *real-valued* function f defined on a m-dimensional smooth manifold $\mathcal{M}$ is called l-times differentiable at $x \in \mathcal{M}$ if, for any chart (U, φ) of x, the function $f \circ \varphi^{-1} : \varphi(U) \to \mathbb{R}$ is l-times differentiable at $\varphi(x) \in \mathbb{R}^m$. The real-valued function f is called l-times differentiable on $\mathcal{M}$ if f is l-times differentiable at every $x \in \mathcal{M}$. It is well known that the definition is independent of the choice of chart. We generalize this definition to Hilbert space-valued maps.

DEFINITION 4.2.   We say that an $\mathbb{H}$-valued map $\mathbf{g}$ defined on an m-dimensional smooth manifold $\mathcal{M}$ is l-times Fréchet differentiable at $x \in \mathcal{M}$ if, for any chart (U, φ) of x, the map $\mathbf{g} \circ \varphi^{-1} : \varphi(U) \to \mathbb{H}$ is l-times Fréchet differentiable at $\varphi(x) \in \mathbb{R}^m$. We say that $\mathbf{g}$ is l-times Fréchet differentiable on $\mathcal{M}$ if $\mathbf{g}$ is l-times Fréchet differentiable at every $x \in \mathcal{M}$.

We may show that the above definition is also independent of the choice of chart. Hence, if $\mathbf{f}_j : \mathcal{X}_j \to \mathbb{H}$ is Fréchet differentiable on $\mathcal{X}_j$, then $\mathbf{f}_j \circ \varphi^{-1}$ is differentiable at $\varphi(x_j)$ for any chart (U, φ) and for any $x_j \in \mathcal{X}_j$. We note that $\mathbf{f}_j \circ \varphi^{-1}$ is also differentiable on $\varphi(U)$, since (U, φ) is also a chart for any points in U.

We introduce a special chart here. Let $\{e_{x_j,1}, \ldots, e_{x_j,q_j}\}$ be an orthonormal basis of $T_{x_j}(\mathcal{X}_j)$, where $T_{x_j}(\mathcal{X}_j)$ is the tangent space of $\mathcal{X}_j$ at x_j. Define an isometric isomorphism $\iota_{x_j} : \mathbb{R}^{q_j} \to T_{x_j}(\mathcal{X}_j)$ by $\iota_{x_j}(t_1, \ldots, t_{q_j}) = \sum_{l=1}^{q_j} t_l e_{x_j,l}$. This is a diffeomorphism since its inverse $\iota_{x_j}^{-1} : T_{x_j}(\mathcal{X}_j) \to \mathbb{R}^{q_j}$ is a global chart between the two manifolds. Let $\exp_{x_j} : T_{x_j}(\mathcal{X}_j) \to \mathcal{X}_j$ be the exponential map and $0 < r < \text{inj}(\mathcal{X}_j)$. We denote $\exp_{x_j} \circ \iota_{x_j} : \mathbb{R}^{q_j} \to \mathcal{X}_j$ by Exp_{x_j}. Then the pair $(B_{\mathcal{X}_j}(x_j, r), \text{Exp}_{x_j}^{-1})$ is called a *normal chart* at x_j.

In what follows, we let D^l denote the lth order Fréchet differential operator. We simply write D for D^1. For Banach spaces $\mathcal{W}$ and $\mathcal{Y}$ and an open subset $\mathcal{V}$ of $\mathcal{W}$, we denote by $D\mathbf{f}(v)(w)$ the Fréchet derivative of a map $\mathbf{f} : \mathcal{V} \to \mathcal{Y}$ at $v \in \mathcal{V}$ in the direction $w \in \mathcal{W}$. We

let $D^2\mathbf{f}(v)(w_1, w_2)$ denote its second-order derivative at v in the direction (w_1, w_2), and so on. Recall that $D\mathbf{f}(v)$ is a map from $\mathcal{W}$ to $\mathcal{Y}$ and $D^2\mathbf{f}(v)$ is a map from $\mathcal{W} \times \mathcal{W}$ to $\mathcal{Y}$. We note that, when $\mathcal{W}$ and $\mathcal{Y}$ are Euclidean spaces, the Fréchet differential operators are reduced to the usual differential operators. Below, we let D_j for $j \geq 1$ denote the first-order partial Fréchet differential operator with respect to the jth argument.

CONDITION (D).

(D1) *For $1 \leq j \leq d$, $\int_{\mathbb{R}^{q_j}} \mathbf{t}K_j(\|\mathbf{t}\|_{\mathbb{R}^{q_j}})\,d\mathbf{t} = \mathbf{0}_{q_j}$.*

(D2) (*Hilbertian predictors: $1 \leq j \leq \varrho$*). (i) $\mathcal{X}_j$ *is convex*; (ii) *For $1 \leq k \leq d$ with $k \neq j$, $D^2\mathbf{f}_j$ is bounded on $\mathcal{X}_j$ and $D_1 p_{jk}$ is bounded on $\mathcal{X}_j \times \mathcal{X}_k$.*

(D3) (*Riemannian predictors: $\varrho+1 \leq j \leq d$*). *There exists a constant $C_j > 0$ and a neighborhood $\mathcal{N}_j$ of $\mathbf{0}_{q_j}$ such that $D^2(\mathbf{f}_j \circ \mathrm{Exp}_{x_j})$ and $D(p_j \circ \mathrm{Exp}_{x_j})$ for all $x_j \in \mathcal{X}_j$ are bounded by C_j on $\mathcal{N}_j$.*

We note that the above conditions are versions of the standard regularity conditions in kernel smoothing theory for $\mathcal{X}_j \equiv [0, 1]$ and $\mathbb{H} = \mathbb{R}$. Conditions similar to (D3) were adopted in [33] and [16] for density estimation on Riemannian manifolds. We also note that in (D3) we do not need a version of the condition on p_{jk} in (D2). This is because $\mathcal{X}_j$ for $\varrho+1 \leq j \leq d$ do not have boundaries. We make a note that dealing with predictor domains without boundaries is the first time in the SBF method. We are able to get proper bounds for the magnitudes of various stochastic terms in a quite different way from the cases with boundaries, without expanding p_{jk} in the direction of the first coordinate.

REMARK 4.1. The conditions on kernel functions for the theory in Sections 3 and 4 are (B1), (B2), (C3), (C4) and (D1). They are satisfied by commonly used kernel functions. For example, the Epanechnikov kernel $K_j(u) = (3/4)(1 - u^2)I(|u| < 1)$ on $[0, \infty)$ satisfies the conditions when $q_j = 1$, and its scaled versions $K_j(u) = s_j(1 - u^2)I(|u| < 1)$ with $s_j = q_j(q_j + 2)/(2 \times |\mathbb{S}^{q_j-1}|)$ satisfy the conditions when $q_j \geq 2$, where $|\mathbb{S}^{q_j-1}|$ denotes the surface area of the unit sphere $\mathbb{S}^{q_j-1} \subset \mathbb{R}^{q_j}$. We note that $|\mathbb{S}^{q_j-1}| = 2\pi(I(q_j = 2) + \prod_{k=1}^{q_j-2} \int_0^\pi (\sin \phi_k)^k\, d\phi_k \cdot I(q_j > 2))$. Also, the biweight kernel $K_j(u) = (15/16)(1 - u^2)^2 I(|u| < 1)$ on $[0, \infty)$ satisfies the conditions when $q_j = 1$, and its scaled versions $K_j(u) = s_j(1 - u^2)^2 I(|u| < 1)$ with $s_j = q_j(q_j + 2)(q_j + 4)/(8 \times |\mathbb{S}^{q_j-1}|)$ satisfy the conditions when $q_j \geq 2$.

We are now ready to state the rates of convergence. In the following theorem, we do not fix the magnitudes of the bandwidths h_j, which requires careful evaluation of various terms in the asymptotic expansion. For each $1 \leq j \leq \varrho$, we let $\mathcal{X}_{j,2h_j}$ denote the set of $2h_j$-interior points of $\mathcal{X}_j$ as defined in Definition 4.1. Recall $\delta_{1,j} = (nh_j^{q_j})^{-1/2}$ and $\delta_{2,j} = (nh_j^{q_j})^{-1/2}(\log n)^{1/2}$.

THEOREM 4.3. *Under the conditions (D), and (C1), (C2), (C3)-(ii) and (C4) in Section 3.2, it holds that*

$$\left(\int_{\mathcal{X}_\mathcal{A}} \|\hat{\mathbf{f}}(\mathbf{x}) \ominus \mathbf{f}(\mathbf{x})\|^2 p(\mathbf{x})\,d\mu(\mathbf{x})\right)^{1/2} = O_p\left(\sum_{j\in\mathcal{A}} h_j^{3/2} + \sum_{j\in\mathcal{A}^c} h_j^2 + \sum_{j=1}^d \delta_{1,j}\right)$$

for any index subset $\mathcal{A}$ of $\{1, \ldots, \varrho\}$, where $\mathcal{X}_\mathcal{A} = \prod_{j\in\mathcal{A}} \mathcal{X}_j \times \prod_{j\in\mathcal{A}^c} \mathcal{X}_{j,2h_j} \times \prod_{j=\varrho+1}^d \mathcal{X}_j$ and $\mathcal{A}^c = \{1, \ldots, \varrho\} \setminus \mathcal{A}$. If we further assume the condition (A7) in Section 2.2, then the followings hold:

(i) *for all* $1 \leq j \leq \varrho$,

$$\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j) = O_p\left(\sum_{j=1}^d h_j^2 + \sum_{j=1}^d \delta_{1,j}\right), \quad x_j \in \mathcal{X}_{j,2h_j},$$

$$\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j) = O_p\left(h_j + \sum_{k \neq j} h_k^2 + \sum_{j=1}^d \delta_{1,j}\right), \quad x_j \in \mathcal{X}_j \setminus \mathcal{X}_{j,2h_j},$$

$$\left(\int_{\mathcal{X}_{j,2h_j}} \|\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j)\|^2 p_j(x_j) \, d\mu_j(x_j)\right)^{1/2} = O_p\left(\sum_{j=1}^d h_j^2 + \sum_{j=1}^d \delta_{1,j}\right),$$

$$\left(\int_{\mathcal{X}_j} \|\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j)\|^2 p_j(x_j) \, d\mu_j(x_j)\right)^{1/2} = O_p\left(h_j^{3/2} + \sum_{k \neq j} h_k^2 + \sum_{j=1}^d \delta_{1,j}\right),$$

$$\sup_{x_j \in \mathcal{X}_{j,2h_j}} \|\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j)\| = O_p\left(\sum_{j=1}^d h_j^2 + \sum_{k \neq j} \delta_{1,k} + \sum_{1 \leq k \leq \varrho, k \neq j} h_k \delta_{2,k} + \delta_{2,j}\right),$$

$$\sup_{x_j \in \mathcal{X}_j} \|\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j)\| = O_p\left(h_j + \sum_{k \neq j} h_k^2 + \sum_{k \neq j} \delta_{1,k} + \sum_{1 \leq k \leq \varrho, k \neq j} h_k \delta_{2,k} + \delta_{2,j}\right);$$

(ii) *for all* $\varrho + 1 \leq j \leq d$,

$$\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j) = O_p\left(\sum_{j=1}^d h_j^2 + \sum_{j=1}^d \delta_{1,j}\right), \quad x_j \in \mathcal{X}_j,$$

$$\left(\int_{\mathcal{X}_j} \|\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j)\|^2 p_j(x_j) \, d\mu_j(x_j)\right)^{1/2} = O_p\left(\sum_{j=1}^d h_j^2 + \sum_{j=1}^d \delta_{1,j}\right),$$

$$\sup_{x_j \in \mathcal{X}_j} \|\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j)\| = O_p\left(\sum_{j=1}^d h_j^2 + \sum_{k \neq j} \delta_{1,k} + \sum_{1 \leq k \leq \varrho} h_k \delta_{2,k} + \delta_{2,j}\right).$$

We note that Jeon and Park [18] dealt with real-valued predictors and fixed bandwidth rates $h_j \asymp n^{-1/5}$. Theorem 4.3 is for Hilbertian/Riemannian predictors without fixing bandwidth rates. If we choose $h_j \asymp n^{-1/(q_j+4)}$ for all $1 \leq j \leq d$, then our estimator of the regression map $\mathbf{f}$ achieves the optimal L_2 rate $n^{-2/(q_{\max}+4)}$ in the region $\prod_{j=1}^{\varrho} \mathcal{X}_{j,2h_j} \times \prod_{j=\varrho+1}^{d} \mathcal{X}_j$, where $q_{\max} = \max_{1 \leq j \leq d} q_j$. Moreover, when $q_j = q_{\max}$ and $h_j \asymp n^{-1/(q_{\max}+4)}$ for all $1 \leq j \leq d$, the estimators of the component maps achieve the optimal pointwise, L_2 and uniform rates.

4.2. *Asymptotic distributions.* We now investigate the asymptotic distributions of our estimators. To the best of our knowledge, the asymptotic distributions of the regression estimators based on the kernels at (3.1) or (3.2) have not been studied even for full-dimensional estimators. Lemma S.5 in the Supplementary Material provides basic ingredients for the study of the asymptotic distributions. Define $p_{j,\mathbf{x}_{-j}} : \mathcal{X}_j \to \mathbb{R}$ by $p_{j,\mathbf{x}_{-j}}(x_j) = p(\mathbf{x})$, where $\mathbf{x}_{-j}$ is $(d-1)$-tuple resulting from omitting x_j from a d-tuple $\mathbf{x} = (x_1, \ldots, x_d)$. Below, we first collect the conditions we use. Let $\{\mathbf{e}_l\}_{l=1}^L$ be an orthonormal basis of $\mathbb{H}$, where we allow $L = \infty$.

CONDITION (E).

(E1) $E(\|\boldsymbol{\epsilon}\|^{\alpha}) < \infty$ *for some* $\alpha > 2$. *For* $1 \le j \ne k \le d$ *and* $l, l' \ge 1$, $E(\|\boldsymbol{\epsilon}\|^{\alpha} \mid X_j = \cdot)$ *and* $E(\langle \boldsymbol{\epsilon}, \mathbf{e}_l \rangle \langle \boldsymbol{\epsilon}, \mathbf{e}_{l'} \rangle \mid X_j = \cdot, X_k = \cdot)$ *are bounded on* $\mathcal{X}_j$ *and* $\mathcal{X}_j \times \mathcal{X}_k$, *respectively, and* $E(\langle \boldsymbol{\epsilon}, \mathbf{e}_l \rangle \langle \boldsymbol{\epsilon}, \mathbf{e}_{l'} \rangle \mid X_j = \cdot)$ *is continuous on* $\mathcal{X}_j$.

(E2) $p > 0$ *on* $\mathcal{X}$ *and* $h_j \asymp n^{-1/(q_j+4)}$ *for* $1 \le j \le d$.

(E3) (*Hilbertian predictors:* $1 \le j \le \varrho$). *For* $1 \le k \le d$ *with* $k \ne j$, $D^2 \mathbf{f}_j$ *is continuous on* $\mathcal{X}_j$, $D_1 p_{jk}$ *is continuous on* $\mathcal{X}_j \times \mathcal{X}_k$, *and* $D_j p$ *is bounded on* $\mathcal{X}$.

(E4) (*Riemannian predictors:* $\varrho + 1 \le j \le d$). *For* $1 \le k \le d$ *with* $k \ne j$, *there exist a constant* $C_j > 0$ *and a neighborhood* $\mathcal{N}_j$ *of* $\mathbf{0}_{q_j}$ *such that* $D^3(\mathbf{f}_j \circ \mathrm{Exp}_{x_j})$, $D^2 p_{jk}(\mathrm{Exp}_{x_j}(\cdot), x_k)$ *and* $D(p_{j,\mathbf{x}_{-j}} \circ \mathrm{Exp}_{x_j})$ *for all* $\mathbf{x} \in \mathcal{X}$ *are bounded by* C_j *on* $\mathcal{N}_j$. *Also, the map* $(x_j, x_k) \mapsto Dp_{jk}(\mathrm{Exp}_{x_j}(\cdot), x_k)(\mathbf{0}_{q_j})$ *is continuous on* $\mathcal{X}_j \times \mathcal{X}_k$.

We note that a version of (E1) is used in [18]. Also, (E3)–(E4) are versions of the standard regularity conditions in kernel smoothing theory for real-valued responses and predictors. The latter two imply (D2)-(ii) and (D3), respectively. Examples of component maps satisfying (E3) and (E4) are numerous. The component maps in our simulation study in Section 6.1 satisfy the conditions, for example.

We define several maps to describe the asymptotic biases and variances of our estimators. For $1 \le j \le \varrho$ and $1 \le k \le d$ with $k \ne j$, define

$$\boldsymbol{\delta}_j(x_j) = D\mathbf{f}_j(x_j)\left(\int_{B(\mathbf{0}_{q_j},1)} \eta_j^{-1}(\mathbf{t}) \odot_j \left(\frac{Dp_j(x_j)(\eta_j^{-1}(\mathbf{t}))}{p_j(x_j)} \cdot K_j(\|\mathbf{t}\|_{\mathbb{R}^{q_j}})\right) d\mathbf{t}\right),$$

$$\boldsymbol{\delta}_{jk}(x_j, x_k) = D\mathbf{f}_j(x_j)\left(\int_{B(\mathbf{0}_{q_j},1)} \eta_j^{-1}(\mathbf{t}) \odot_j \left(\frac{D_1 p_{jk}(x_j, x_k)(\eta_j^{-1}(\mathbf{t}))}{p_{jk}(x_j, x_k)} \cdot K_j(\|\mathbf{t}\|_{\mathbb{R}^{q_j}})\right) d\mathbf{t}\right),$$

where D denotes the Fréchet differential operator we introduced in Section 4.1, and $\odot_j$ is the scalar multiplication on $\mathcal{H}_j$. The above terms involve integration in the direction of a Hilbertian predictor. The corresponding terms involving integration on a Riemannian manifold are defined, for $\varrho + 1 \le j \le d$ and $1 \le k \le d$ with $k \ne j$, by

$$\boldsymbol{\delta}_j(x_j) = D(\mathbf{f}_j \circ \mathrm{Exp}_{x_j})(\mathbf{0}_{q_j})\left(\int_{B(\mathbf{0}_{q_j},1)} \mathbf{t} \cdot \frac{D(p_j \circ \mathrm{Exp}_{x_j})(\mathbf{0}_{q_j})(\mathbf{t})}{p_j(x_j)} \cdot K_j(\|\mathbf{t}\|_{\mathbb{R}^{q_j}}) d\mathbf{t}\right),$$

$$\boldsymbol{\delta}_{jk}(x_j, x_k) = D(\mathbf{f}_j \circ \mathrm{Exp}_{x_j})(\mathbf{0}_{q_j})\left(\int_{B(\mathbf{0}_{q_j},1)} \mathbf{t} \cdot \frac{D(p_{jk}(\mathrm{Exp}_{x_j}, x_k))(\mathbf{0}_{q_j})(\mathbf{t})}{p_{jk}(x_j, x_k)} \cdot K_j(\|\mathbf{t}\|_{\mathbb{R}^{q_j}}) d\mathbf{t}\right).$$

Let $\alpha_j = \lim_n h_j \cdot n^{1/(q_{\max}+4)}$. Note that $\alpha_j = 0$ for j with $q_j < q_{\max}$, according to (E2). For $1 \le j \le d$, define

$$\tilde{\boldsymbol{\Delta}}_j(x_j) = \alpha_j^2 \odot \boldsymbol{\delta}_j(x_j) \oplus \bigoplus_{k \ne j} \int_{\mathcal{X}_k} \boldsymbol{\delta}_{jk}(x_j, x_k) \odot \left(\alpha_k^2 \cdot \frac{p_{jk}(x_j, x_k)}{p_j(x_j)}\right) d\mu_k(x_k).$$

Let $(\boldsymbol{\Delta}_j : 1 \le j \le d)$ be the solution of the system of equations

$$(4.2) \qquad \boldsymbol{\Delta}_j(x_j) = \tilde{\boldsymbol{\Delta}}_j(x_j) \ominus \bigoplus_{k \ne j} \int_{\mathcal{X}_k} \boldsymbol{\Delta}_k(x_k) \odot \frac{p_{jk}(x_j, x_k)}{p_j(x_j)} d\mu_k(x_k), \quad 1 \le j \le d$$

subject to the constraints

$$\int_{\mathcal{X}_j} \boldsymbol{\Delta}_j(x_j) \odot p_j(x_j) d\mu_j(x_j) = \int_{\mathcal{X}_j} \boldsymbol{\delta}_j(x_j) \odot (\alpha_j^2 \cdot p_j(x_j)) d\mu_j(x_j), \quad 1 \le j \le d,$$

$$(4.3)$$
$$\int_{\mathcal{X}_j} \|\boldsymbol{\Delta}_j(x_j)\|^2 p_j(x_j) d\mu_j(x_j) < \infty, \quad 1 \le j \le d.$$

Also, define

$$\mathbf{c}_j(x_j) = \int_{B(\mathbf{0}_{q_j},1)} D^2 \mathbf{f}_j(x_j)(\eta_j^{-1}(\mathbf{t}), \eta_j^{-1}(\mathbf{t})) \odot \left(\frac{1}{2} K_j(\|\mathbf{t}\|_{\mathbb{R}^{q_j}})\right) d\mathbf{t}, \quad 1 \le j \le \varrho,$$

$$\mathbf{c}_j(x_j) = \int_{B(\mathbf{0}_{q_j},1)} D^2 (\mathbf{f}_j \circ \mathrm{Exp}_{x_j})(\mathbf{0}_{q_j})(\mathbf{t}, \mathbf{t}) \odot \left(\frac{1}{2} K_j(\|\mathbf{t}\|_{\mathbb{R}^{q_j}})\right) d\mathbf{t}, \quad \varrho + 1 \le j \le d.$$

Then, as demonstrated in Theorem 4.4 below, it turns out that $\boldsymbol{\Theta}_j(x_j) := \alpha_j^2 \odot \mathbf{c}_j(x_j) \oplus \boldsymbol{\Delta}_j(x_j)$, $1 \le j \le d$, are the asymptotic biases of $\hat{\mathbf{f}}_j(x_j)$.

To describe the asymptotic variances, define

$$a_{j,ll'}(x_j) = \alpha_j^{-q_{\max}} \cdot p_j(x_j)^{-1} \cdot \mathrm{E}(\langle \boldsymbol{\epsilon}, \mathbf{e}_l \rangle \langle \boldsymbol{\epsilon}, \mathbf{e}_{l'} \rangle \mid X_j = x_j) \int_{\mathbb{R}^{q_j}} K_j^2(\|\mathbf{t}\|_{\mathbb{R}^{q_j}}) d\mathbf{t}$$

if $q_j = q_{\max}$, and $a_{j,ll'}(x_j) \equiv 0$ if $q_j < q_{\max}$. Let $C_{j,x_j} : \mathbb{H} \to \mathbb{H}$ be the covariance operator characterized by

$$\langle C_{j,x_j}(\mathbf{h}), \mathbf{e}_l \rangle = \sum_{l' \ge 1} \langle \mathbf{h}, \mathbf{e}_{l'} \rangle \cdot a_{j,ll'}(x_j), \quad \mathbf{h} \in \mathbb{H}, l \ge 1.$$

We denote by $\mathbf{G}(\mathbf{0}, C_{j,x_j})$ the Hilbertian Gaussian random element with mean zero and co-variance operator C_{j,x_j}. When $\mathbb{H} = \mathbb{R}$, $\mathbf{G}(\mathbf{0}, C_{j,x_j})$ is reduced to a normal random variable with mean 0 and variance $\alpha_j^{-q_{\max}} \cdot p_j(x_j)^{-1} \cdot \mathrm{E}(\epsilon^2 \mid X_j = x_j) \int_{\mathbb{R}^{q_j}} K_j^2(\|\mathbf{t}\|_{\mathbb{R}^{q_j}}) d\mathbf{t}$ if $q_j = q_{\max}$, and it degenerates to zero if $q_j < q_{\max}$.

We are now ready to state the theorem. We note that, in general, the derivation of asymptotic distribution is more involved than that of rate of convergence since the former needs exact evaluation of leading terms in asymptotic expansions. This is particularly the case with the theorem given below, in the proof of which exact asymptotic expansions for $(\hat{\mathbf{f}}_j : 1 \le j \le d)$ with Hilbertian vector operations on abstract Hilbertian/Riemannian predictor spaces require far more sophisticated evaluation of various stochastic terms.

THEOREM 4.4. *Assume the conditions* (E), (C2)-(ii), (C3)-(ii) *and* (C4) *in Section* 3.2, *and* (D1) *and* (D2)-(i) *in Section* 4.1. *Then* (i) *the solution of* (4.2) *subject to* (4.3) *is unique in the sense that, if* $(\boldsymbol{\Delta}_j^\star : 1 \le j \le d)$ *is another solution, then* $\boldsymbol{\Delta}_j = \boldsymbol{\Delta}_j^\star$ *a.e. with respect to* μ_j *for all* $1 \le j \le d$; (ii) *for a.e.* $\mathbf{x} \in \prod_{j=1}^\varrho (\mathcal{X}_j \setminus \partial \mathcal{X}_j) \times \prod_{j=\varrho+1}^d \mathcal{X}_j$ *with respect to* μ,

$$\begin{pmatrix} n^{2/(q_{\max}+4)} \odot (\hat{\mathbf{f}}_1(x_1) \ominus \mathbf{f}_1(x_1)) \\ \vdots \\ n^{2/(q_{\max}+4)} \odot (\hat{\mathbf{f}}_d(x_d) \ominus \mathbf{f}_d(x_d)) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \boldsymbol{\Theta}_1(x_1) \oplus \mathbf{G}(\mathbf{0}, C_{1,x_1}) \\ \vdots \\ \boldsymbol{\Theta}_d(x_d) \oplus \mathbf{G}(\mathbf{0}, C_{d,x_d}) \end{pmatrix},$$

$$n^{2/(q_{\max}+4)} \odot (\hat{\mathbf{f}}(\mathbf{x}) \ominus \mathbf{f}(\mathbf{x})) \xrightarrow{d} \bigoplus_{j=1}^d \boldsymbol{\Theta}_j(x_j) \oplus \mathbf{G}\left(\mathbf{0}, \sum_{j=1}^d C_{j,x_j}\right),$$

where $\boldsymbol{\Theta}_1(x_1) \oplus \mathbf{G}(\mathbf{0}, C_{1,x_1}), \dots, \boldsymbol{\Theta}_d(x_d) \oplus \mathbf{G}(\mathbf{0}, C_{d,x_d})$ *are independent.*

Now, we compare the asymptotic distributions of the sB-SBF estimators with the oracle estimators. For j with $q_j = q_{\max}$, we let $\hat{\mathbf{f}}_j^{\mathrm{ora}}$ be the Nadaraya–Watson type oracle estimator of $\mathbf{f}_j$ that we may obtain using the knowledge of all other component maps $\mathbf{f}_k, k \ne j$. Using Lemma S.5 in the supplement, we may prove that

$$n^{2/(q_{\max}+4)} \odot (\hat{\mathbf{f}}_j^{\mathrm{ora}}(x_j) \ominus \mathbf{f}_j(x_j)) \xrightarrow{d} \alpha_j^2 \odot (\boldsymbol{\delta}_j(x_j) \oplus \mathbf{c}_j(x_j)) \oplus \mathbf{G}(\mathbf{0}, C_{j,x_j}).$$

We note that the above result is also new. Therefore, $\hat{\mathbf{f}}_j$ and $\hat{\mathbf{f}}_j^{\text{ora}}$ have the same asymptotic covariance operator, but differ in their asymptotic biases. The difference of asymptotic biases is $\alpha_j^2 \odot \boldsymbol{\delta}_j(x_j) \ominus \boldsymbol{\Delta}_j(x_j) =: \boldsymbol{\beta}_j(x_j)$ and it holds that $\mathrm{E}(\boldsymbol{\beta}_j(X_j)) = \int_{\mathcal{X}_j} \boldsymbol{\beta}_j(x_j) \odot p_j(x_j)\, d\mu_j(x_j) = \mathbf{0}$ by (4.3).

5. Some generalization. In this section, we provide a novel theorem on the rates of convergence for predictors taking values in general semimetric spaces. We believe that it would be useful for other non-Euclidean objects that we do not treat in Sections 3 and 4. To the best of our knowledge, this type of theorem has not been studied even for real-valued responses and predictors. The theorem is based on high-level conditions without assuming that $(\mathbf{Y}_i, \mathbf{X}_i)$ are identically distributed or independent. We stress that the derivation requires careful investigation on the underlying structure of SBF theory that involves various stochastic terms.

For the statement of the theorem, we define

$$\hat{\boldsymbol{\delta}}_j^A(x_j) = \frac{1}{n} \odot \bigoplus_{i=1}^n (L_{j,n}(x_j, X_{ij}) \odot \boldsymbol{\epsilon}_i),$$

$$\hat{\boldsymbol{\delta}}_j^B(x_j) = \frac{1}{n} \odot \bigoplus_{i=1}^n (L_{j,n}(x_j, X_{ij}) \odot (\mathbf{f}_j(X_{ij}) \ominus \mathbf{f}_j(x_j))),$$

$$\hat{\boldsymbol{\delta}}_j^C(x_j) = \bigoplus_{k \neq j} \left(\frac{1}{n} \odot \bigoplus_{i=1}^n \int_{\mathcal{X}_k} (\mathbf{f}_k(X_{ik}) \ominus \mathbf{f}_k(x_k)) \odot (L_{k,n}(x_k, X_{ik}) L_{j,n}(x_j, X_{ij}))\, d\mu_k(x_k) \right),$$

where $L_{j,n}$ are arbitrary measurable weight functions depending on n and satisfying (2.3). For $1 \leq j \leq d$, let $\Omega_j \in \mathcal{B}(\mathcal{X}_j)$ and $A_{nj}^{(1)}(x_j)$, $A_{nj}^{(2)}$, $A_{nj}^{(3)}(\Omega_j)$, $B_{nj}^{(1)}(x_j)$, $B_{nj}^{(2)}(\Omega_j)$, $B_{nj}^{(3)}(\Omega_j)$, $B_{nj}^{(4)}$, $C_{nj}^{(1)}(x_j)$, $C_{nj}^{(2)}$, $C_{nj}^{(3)}(\Omega_j)$, $D_n^{(1)}$ and $D_{nj}^{(2)}$ be nonnegative sequences converging to zero such that

(i) rates for $\hat{\boldsymbol{\delta}}_j^A$:

$$\hat{\boldsymbol{\delta}}_j^A(x_j) = O_p(A_{nj}^{(1)}(x_j)), \qquad \sup_{x_j \in \Omega_j} \|\hat{\boldsymbol{\delta}}_j^A(x_j)\| = O_p(A_{nj}^{(3)}(\Omega_j)),$$

$$\left(\int_{\mathcal{X}_j} \|\hat{\boldsymbol{\delta}}_j^A(x_j)\|^2 p_j(x_j)\, d\mu_j(x_j) \right)^{1/2} = O_p(A_{nj}^{(2)});$$

(ii) rates for $\hat{\boldsymbol{\delta}}_j^B$:

$$\hat{\boldsymbol{\delta}}_j^B(x_j) = O_p(B_{nj}^{(1)}(x_j)), \qquad \sup_{x_j \in \Omega_j} \|\hat{\boldsymbol{\delta}}_j^B(x_j)\| = O_p(B_{nj}^{(3)}(\Omega_j)),$$

$$\left(\int_{\Omega_j} \|\hat{\boldsymbol{\delta}}_j^B(x_j)\|^2 p_j(x_j)\, d\mu_j(x_j) \right)^{1/2} = O_p(B_{nj}^{(2)}(\Omega_j)),$$

$$\sum_{k \neq j} \sup_{x_j \in \mathcal{X}_j} \left\| \int_{\mathcal{X}_k} \hat{\boldsymbol{\delta}}_k^B(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_k(x_k)}\, d\mu_k(x_k) \right\| = O_p(B_{nj}^{(4)});$$

(iii) rates for $\hat{\boldsymbol{\delta}}_j^C$:

$$\hat{\boldsymbol{\delta}}_j^C(x_j) = O_p(C_{nj}^{(1)}(x_j)), \qquad \sup_{x_j \in \Omega_j} \|\hat{\boldsymbol{\delta}}_j^C(x_j)\| = O_p(C_{nj}^{(3)}(\Omega_j)),$$

$$\left(\int_{\mathcal{X}_j} \|\hat{\boldsymbol{\delta}}_j^C(x_j)\|^2 p_j(x_j)\, d\mu_j(x_j) \right)^{1/2} = O_p(C_{nj}^{(2)});$$

(iv) rates for sample mean:

$$\mathrm{E}(\mathbf{Y}) \ominus \bar{\mathbf{Y}} = O_p(D_n^{(1)}), \qquad \mathrm{E}(\mathbf{f}_j(X_j)) \ominus n^{-1} \odot \bigoplus_{i=1}^{n} \mathbf{f}_j(X_{ij}) = O_p(D_{nj}^{(2)}).$$

Put $R_n = \sum_{j=1}^{d}(A_{nj}^{(2)} + B_{nj}^{(4)} + C_{nj}^{(2)}) + D_n^{(1)}$, $E_{nj}^{(1)}(x_j) = A_{nj}^{(1)}(x_j) + B_{nj}^{(1)}(x_j) + C_{nj}^{(1)}(x_j)$ and $E_{nj}^{(3)}(\Omega_j) = A_{nj}^{(3)}(\Omega_j) + B_{nj}^{(3)}(\Omega_j) + C_{nj}^{(3)}(\Omega_j)$.

THEOREM 5.1. *Assume the conditions* (A1), (A2) *and* (A6) *in Section* 2.2. *Also, assume that p_j is bounded away from zero on $\mathcal{X}_j$ and p_{jk} is bounded on $\mathcal{X}_j \times \mathcal{X}_k$ for all $1 \leq j \neq k \leq d$. Then*

$$\left(\int_{\prod_{j=1}^{d} \Omega_j} \|\hat{\mathbf{f}}(\mathbf{x}) \ominus \mathbf{f}(\mathbf{x})\|^2 p(\mathbf{x}) \, d\mu(\mathbf{x}) \right)^{1/2} = O_p\left(R_n + \sum_{j=1}^{d} B_{nj}^{(2)}(\Omega_j) \right).$$

If we further assume (A7), *then for all $1 \leq j \leq d$,*

$$\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j) = O_p\left(R_n + E_{nj}^{(1)}(x_j) + \sum_{k \neq j} D_{nk}^{(2)} \right),$$

$$\left(\int_{\Omega_j} \|\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j)\|^2 p_j(x_j) \, d\mu_j(x_j) \right)^{1/2} = O_p(R_n + B_{nj}^{(2)}(\Omega_j) + D_{nj}^{(2)}),$$

$$\sup_{x_j \in \Omega_j} \|\hat{\mathbf{f}}_j(x_j) \ominus \mathbf{f}_j(x_j)\| = O_p\left(R_n + E_{nj}^{(3)}(\Omega_j) + \sum_{k \neq j} D_{nk}^{(2)} \right).$$

6. Numerical study. In this section, we present the results of two simulation studies and two real data analyses. Some details of practical issues such as bandwidth selection are presented in the Supplement S.4.

6.1. *Simulations.* In the first simulation, we considered the case where the response variable was a random density and the predictors were Euclidean vectors taking values in flexible domains. Specifically, we considered the case where $\mathbf{Y} = [Y(\cdot)]$ belongs to the Bayes–Hilbert space $\mathfrak{B}^2(U, \mathcal{B}(U), \mathrm{Leb}_1)$ for $U = [-1/2, 1/2]$, where $[Y(\cdot)]$ denotes the equivalence class of a density function $Y(\cdot)$. More details on Bayes–Hilbert spaces can be found in the Supplement S.1. We generated $X_1 = (X_{1,1}, X_{1,2})$ uniformly on the ellipse $\{(x, y) \in \mathbb{R}^2 : x^2/4 + y^2/9 \leq 1\}$ and $X_2 = (X_{2,1}, X_{2,2}, X_{2,3})$ uniformly on the circular cylinder $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1, 0 \leq z \leq 1/2\}$.

In this simulation, we compared two approaches. One is based on the *multivariate additive* model that is additive in predictor vectors X_j. As noted in Section 3, our general framework provides a tool for estimating this type of model, which is missing in the literature. The other is based on the *univariate additive* model that is additive in univariate predictors $X_{j,k}$. For this we may apply the method of [18], which is a special case in our general treatment. For proper comparison, some additive maps in X_j were chosen not to be additive in the univariate predictors $X_{j,k}$. Specifically, we took

$$f_1^{\mathrm{add}}(X_1)(u) = (1 - X_{1,1}^2/5 \cdot |\sin(u)|) \cdot (1 - X_{1,2}^2/5 \cdot |\sin(u)|),$$

$$f_1^{\mathrm{non}}(X_1)(u) = (1 - (X_{1,1} + X_{1,2})^2/5 \cdot |\sin(u)|),$$

$$f_2^{\mathrm{add}}(X_2)(u) = \exp(-(X_{2,1}^2 + X_{2,2}^2 + X_{2,3}^2) \cdot |u|),$$

$$f_2^{\mathrm{non}}(X_2)(u) = \exp(-(X_{2,1} + X_{2,2}) \cdot X_{2,3} \cdot |u|)$$

for $u \in [-1/2, 1/2]$. We note that $f_1^{\text{add}}(X_1)(\cdot)$ is the product of two functions of $X_{1,1}$ and $X_{1,2}$, respectively. Hence, $[f_1^{\text{add}}(X_1)(\cdot)]$ is an additive map of the two scalar predictors $X_{1,1}$ and $X_{1,2}$ with respect to the $\oplus$ operation in $\mathfrak{B}^2(U, \mathcal{B}(U), \text{Leb}_1)$. Similarly, $[f_2^{\text{add}}(X_2)(\cdot)]$ is also the additive map of the three scalar predictors $X_{2,1}$, $X_{2,2}$ and $X_{2,3}$. On the other hand, the two maps f_1^{non} and f_2^{non} are nonadditive so that they contain some interaction effects between the scalar elements of X_1 and X_2, respectively.

We generated $Y(\cdot)$ according to the following three scenarios:

(I) $$Y(\cdot) = f_1^{\text{add}}(X_1)(\cdot) \cdot f_2^{\text{non}}(X_2)(\cdot) \cdot f_e(\epsilon)(\cdot),$$

(II) $$Y(\cdot) = f_1^{\text{non}}(X_1)(\cdot) \cdot f_2^{\text{add}}(X_2)(\cdot) \cdot f_e(\epsilon)(\cdot),$$

(III) $$Y(\cdot) = f_1^{\text{non}}(X_1)(\cdot) \cdot f_2^{\text{non}}(X_2)(\cdot) \cdot f_e(\epsilon)(\cdot),$$

where $f_e(\epsilon)(u) = (f_0(u))^\epsilon$, $f_0(u) = \exp(-u^2)$ and ϵ is a $N(0, 0.25^2)$ variable independent of $\mathbf{X} = (X_1, X_2)$. We note that $\text{E}([f_e(\epsilon)(\cdot)] \mid \mathbf{X}) = \text{E}(\epsilon \odot [f_0(\cdot)]) = 0 \odot [f_0(\cdot)] = \mathbf{0}$. Hence, the true conditional means $\text{E}(\mathbf{Y} \mid \mathbf{X})$ for the three scenarios are given by $[f_1^{\text{add}}(X_1)(\cdot)] \oplus [f_2^{\text{non}}(X_2)(\cdot)]$, $[f_1^{\text{non}}(X_1)(\cdot)] \oplus [f_2^{\text{add}}(X_2)(\cdot)]$ and $[f_1^{\text{non}}(X_1)(\cdot)] \oplus [f_2^{\text{non}}(X_2)(\cdot)]$, respectively. We repeatedly generated, for $M = 100$ times, training samples of sizes $n = 125$ and 500, and test samples of size $N = 100$.

As a measure of performance we computed the mean squared prediction error (MSPE) defined by

(6.1) $$\text{MSPE} = M^{-1} \sum_{m=1}^{M} N^{-1} \sum_{i=1}^{N} \| \mathbf{f}(\mathbf{X}_i^{\text{test}(m)}) \ominus \tilde{\mathbf{f}}(\mathbf{X}_i^{\text{test}(m)}) \|^2,$$

where $\mathbf{X}_i^{\text{test}(m)}$ is the ith predictor value in the mth test sample, $\mathbf{f} = \text{E}(\mathbf{Y} \mid \mathbf{X} = \cdot)$ is the true conditional mean and $\tilde{\mathbf{f}}$ denotes either our estimator $\hat{\mathbf{f}}$ based on the multivariate additive models or the estimator in [18] based on the respective univariate additive models.

Table 1 shows that the multivariate additive estimator adapted to each scenario indeed outperforms the univariate additive estimator considered in [18]. The difference between the MSPEs of the two estimators is larger for the third scenario than for the other two. We believe that it is because both components f_1^{non} and f_2^{non} are nonadditive in the third scenario. This indicates that we may construct a better estimator based on our approach when there are some interaction effects between the elements of each predictor vector X_j. This is considered as an example demonstrating that our general framework provides useful tools for the estimation of various interim models between two extremes, one that is additive in all univariate predictors and the other that is entirely unstructured.

TABLE 1
The values of MSPE, multiplied by 10^3, of the proposed method adapted to each scenario and the method of [18] based on univariate additive models. The latter method is a special case of the proposed method

Scenario	n	Multivariate additive	Univariate additive
(I)	125	0.1944	0.3478
	500	0.0674	0.2542
(II)	125	0.2219	0.3724
	500	0.0897	0.2405
(III)	125	0.2154	0.5067
	500	0.0852	0.4125

The second simulation is for Riemannian predictors. For this we compared our estimator with the full-dimensional Nadaraya–Watson type estimator for Riemannian predictors studied by [34]. To implement the latter estimator, we used Proposition 3.5. Since the latter estimator was studied only for real-valued responses, we considered a scalar response Y here, which our framework also covers. We generated predictors X_1, X_2 and X_3 uniformly on the two-dimensional shape space Σ_2^3, two-dimensional sphere $\mathbb{S}^2$ and two-dimensional torus $\mathbb{S}^1 \times \mathbb{S}^1$, respectively. The first one corresponds to the case where we observe triangular shapes or three landmarks of shapes in $\mathbb{R}^2$. We considered the following three component functions:

$$f_1(X_1) = \arctan(\vartheta_{X_1}^{\min}/\vartheta_{X_1}^{\max}),$$

$$f_2(X_2) = \cos(2\vartheta_{X_2}) + 2\sin^2(\vartheta_{X_2})\cos^2(\phi_{X_2}),$$

$$f_3(X_3) = 1 + \cos(\vartheta_{X_3}^{(1)} - \vartheta_{X_3}^{(2)}),$$

where $(\vartheta_{X_1}^{\min}, \vartheta_{X_1}^{\max}) \in (0, \pi/3) \times [\pi/3, \pi)$ are the smallest and largest angles of the triangle that X_1 forms, $(\vartheta_{X_2}, \phi_{X_2}) \in [0, 2\pi) \times [0, \pi]$ are the longitude and latitude corresponding to X_2 living on the sphere and $(\vartheta_{X_3}^{(1)}, \vartheta_{X_3}^{(2)}) \in [0, 2\pi) \times [0, 2\pi)$ are the angles corresponding to X_3 lying on the torus. We generated Y according to the following three scenarios:

(Additive) $Y = f_1(X_1) + \cdots + f_d(X_d) + +\epsilon,$

(Nonadditive 1) $Y = (f_1(X_1) + \cdots + f_d(X_d))^2 + \epsilon,$

(Nonadditive 2) $Y = f_1(X_1) \times \cdots \times f_d(X_d) + \epsilon,$

where ϵ is again a $N(0, 0.25^2)$ variable independent of $\mathbf{X} = (X_1, X_2, X_3)$. As in the first simulation, we generated $M = 100$ training and test samples. Here, we added an additional training sample size $n = 2000$ to see more about the behavior of the estimators depending on the sample size. As a comparison criterion we used the MSPE defined at (6.1) with $\mathbf{f}$ and $\tilde{\mathbf{f}}$ being replaced by f and $\tilde{f}$, $\|\cdot\|$ by $|\cdot|$ and $\ominus$ by $-$, where $f = E(Y \mid \mathbf{X} = \cdot)$ and $\tilde{f}$ is either our estimator or the estimator in [34].

The results in Table 2 confirm our theory. In the additive scenario, our sB-SBF estimator outperforms the full-dimensional Nadaraya–Watson type estimator (henceforth, F-NW). Examining the values of the ratio in the table for $d = 2$ and $d = 3$, especially in the additive scenario, we find that the margin by which the sB-SBF method wins the F-NW gets larger as d increases. Also, the former's decreasing speed of MSPE, as the sample size increases, does not change much when $d = 2$ increases to $d = 3$, while for the latter it gets slower when $d = 3$. This indicates that the unstructured F-NW suffers from the curse of dimensionality while ours does not. We note that the actual dimension of $\mathbf{X}$ equals $2 \times d$, instead of d. Thus, increasing d by one increases the dimension by two. In the nonadditive scenarios, however, the MSPE of the sB-SBF estimator decreases much slower than in the case of the additive scenario, although it still wins the F-NW except the second nonadditive scenario with $(d, n) = (2, 2000)$. This is mainly because the model assumption the sB-SBF estimator is based on is violated. The bias caused by the model misspecification does not vanish as the sample size increases. Nevertheless, the results suggest that, for high-dimensional predictors, our additive regression approach may dominate unstructured estimators in finite-sample performance even when the model assumption is violated.

We conclude this subsection by noting that our general framework covers models containing some interaction effects between Riemannian predictors, thanks to our treatment of product Riemannian manifolds in Section 3.3.2. We also note that our approach accommodates the case of mixed Hilbertian and Riemannian predictors as well, tools for which are

TABLE 2
*The values of MSPE of the proposed method and the full-dimensional method of [34], and their ratios MSPE
(Full-dimensional Nadaraya–Watson)/MSPE (Proposed)*

Scenario	d	n	Proposed	Full-dim. NW	Ratio
Additive scenario	2	125	0.0490	0.0977	1.99
		500	0.0153	0.0473	3.09
		2000	0.0056	0.0224	4.00
	3	125	0.1192	0.3900	3.27
		500	0.0323	0.2141	6.63
		2000	0.0129	0.1298	10.06
Nonadditive 1 scenario	2	125	0.1394	0.2429	1.74
		500	0.0757	0.1347	1.78
		2000	0.0535	0.0678	1.27
	3	125	2.1889	4.6004	2.10
		500	1.1295	2.6316	2.33
		2000	0.8868	1.6131	1.82
Nonadditive 2 scenario	2	125	0.0244	0.0290	1.19
		500	0.0160	0.0192	1.20
		2000	0.0127	0.0106	0.83
	3	125	0.0509	0.0627	1.23
		500	0.0372	0.0492	1.32
		2000	0.0356	0.0394	1.11

rare even for real-valued responses in simpler models. We believe that the flexibility of our setup provides with powerful methods of estimating regression maps/functions in a variety of situations, which may offer superior performance compared to the existing methods, if any.

6.2. *Real data analysis.* In this subsection, we present two real data applications. The first one is for a real-valued response and mixed Euclidean and circular predictors. The second one is for a functional response and mixed Euclidean and spherical predictors. In the Supplement S.5, we also present an additional application which is for a compositional response and mixed Euclidean and compositional predictors.

6.2.1. *First example.* Recently, air pollution caused by particular matter whose size is less than 2.5 micrometer, called PM2.5, has been a serious problem throughout the world. It is known that this tiny matter can penetrate lungs and blood vessels, which causes health problems when exposed for a long term. In Korea, it has been a usual thought that the level of PM2.5 is affected by weather conditions such as humidity and wind direction. There has been no investigation on this usual mind using an advanced statistical method, however. In the first application, we took daily measured PM2.5 level in Seoul, the capital city of Korea, as a real-valued response Y, and daily measured humidity (X_1) and daily measured wind direction (X_2) as predictors. We note that X_2 can be understood as a circular variable. We also took the day of a year when the measurement was made, as the third predictor X_3 taking values in $\{1, \ldots, 365\}$ or $\{1, \ldots, 366\}$, since the level of PM2.5 is believed to have a yearly periodic trend. We note that X_3 is also considered as a circular variable. The jth day in a year is converted to the point on the unit circle in $\mathbb{R}^2$ having the clockwise angle $(j - 1) \times 2\pi/$(the number of days in the year) in radian from the point $(0, 1) \in \mathbb{S}^1$. We collected the data on these variables for the period January 2016–December 2018 from the Korea

air pollution information system https://www.airkorea.or.kr/web and the Korea meteorological cal administration https://data.kma.go.kr/cmmn/main.do. We deleted 3 missing observations among total 366(days) + 365(days) × 2(years = 1096 observations.

We analyzed the data using seven methods. The first method is to apply the sB-SBF technique with X_2 and X_3 as two separate circular predictors (sB-SBF-2CIRC). That is, it is based on the model (2.1) with one real-valued predictor and two circular predictors ($d = 3$). The second method is to take (X_2, X_3) as a torus-valued predictor (sB-SBF-TORUS), so that it applies (2.1) with one real-valued predictor and one torus-valued predictor ($d = 2$). The third one is to apply a partially linear model (PLM, [12]), $Y = \beta_0 + \beta_1 X_1 + \zeta(X_2, X_3) + \epsilon$, where (X_2, X_3) enters as a torus-valued predictor. To estimate ζ in the PLM, we used the kernel at (3.2) as in [12]. The fourth one is to apply a spline additive model (SAM, [42]) where the component function for X_1 is smoothed with a cubic spline basis, and the component functions for X_2 and X_3 with cyclic cubic spline bases. The remaining three methods are the random forest (RF), neural network (NN) and support vector machine (SVM) techniques. Although the latter three are not designed for circular predictors, we included them by treating X_2 and X_3 as real-valued predictors, to see how well the popular machine learning techniques perform in this non-Euclidean setting. To compare the methods, we divided the observations into 10 partitions S_k, $1 \leq k \leq 10$, and then computed the 10-fold average squared prediction error (ASPE) defined by

$$(6.2) \qquad \text{ASPE} = 10^{-1} \sum_{k=1}^{10} |S_k|^{-1} \sum_{i \in S_k} (Y_i - \hat{Y}_i^{(-S_k)})^2,$$

where $|S_k|$ is the number of observations in S_k and $\hat{Y}_i^{(-S_k)}$ is the prediction of Y_i based on the sample without the observations in S_k. We found that ASPE was 116.31, 136.73, 146.85, 122.34, 121.22, 132.61 and 139.04 for sB-SBF-2CIRC, sB-SBF-TORUS, PLM, SAM, RF, NN and SVM, respectively. Thus, sB-SBF-2CIRC is most predictive. The observation that sB-SBF-2CIRC performed better than sB-SBF-TORUS indicates that there is no strong interaction between X_2 and X_3. Also, comparing sB-SBF-TORUS with the method based on PLM suggests that the effect of X_1 is nonlinear.

Figure 1 depicts the estimated component maps based on the sB-SBF-2CIRC. We note that, by respecting the circular geometry in X_2 and X_3, our approach produced estimated component maps that are continuous everywhere on the circular domains. It is well admitted that such circular continuity is essential in circular data analysis. The first component map shows that the PM2.5 level increases as the humidity goes up to around 70, and then decreases after that level. We interpret this as that air pollutants such as nitrogen oxide produce PM2.5 if they combine with moisture, but at high-level of humidity, possibly due to rain, PM2.5 is washed away. The second component map reveals that, when wind blows from NNE or SSW, the PM2.5 level gets high. Also, the third component map shows that the level is relatively low from summer to autumn, while it is relatively high from winter to spring. In particular, between March and April, the level is very high due to the dust storm that usually happens in that season. In conclusion, our analysis suggests that humidity, wind direction and season are key elements in the prediction of PM2.5 levels.

6.2.2. *Second example.* Although surface temperatures on the globe have been analyzed in many studies, the effect of altitude on temperature has received less attention. Adding altitude as a predictor may enhance the accuracy of temperature prediction. Recently, Integrated Global Radiosonde Archive (IGRA) collected weather data using radiosondes at various altitude levels ranging from the sea-level to the one where air pressure is 1 hectopascal. We collected from https://www1.ncdc.noaa.gov/pub/data/igra/igra2-station-list.txt
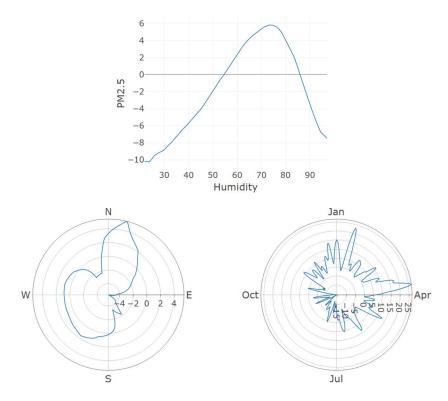
FIG. 1. *Estimated component maps for the PM2.5 data based on the sB-SBF-2CIRC, the proposed method taking wind direction and day in a year as two separate circular predictors. In the two bottom panels, they are depicted on circles with continuously varying radii, where the radius of each circle corresponds to each value of the component map. The directional position of a point on the solid curve represents the corresponding value of the circular predictor.*

the longitudes and latitudes of IGRA stations spread out in the world, and also from https://www1.ncdc.noaa.gov/pub/data/igra/monthly/monthly-por/ the monthly averaged temperatures measured by each station at various air pressure levels for the period January 2018–December 2018. Note that altitude is well represented by air pressure level. We took log (air pressure) as a real-valued predictor (X_1) and (longitude, latitude) as a spherical predictor (X_2). To ensure that each observed (X_1, X_2) has an observation of the response, a temperature trajectory on the entire time domain (Jan–Dec), we actually collected the cases at which all the 12 monthly temperatures were recorded. For each station and altitude level, we smoothed the monthly temperatures to get a functional response variable $\mathbf{Y} = Y(\cdot)$ defined on [1, 15]. Here, $Y(1.5)$, for example, can be understood as the mean temperature between mid-January and mid-February. To the best of our knowledge, there has been no other method designed for a functional response coupled with spherical predictor(s). Hence, we applied only our method to this dataset having sample size $n = 7891$.

The estimated component map of X_1 in Figure 2 demonstrates a clear elevation effect. It shows that elevation up to a medium-high level pushes down the temperature, while further elevation raises it up sharply. This confirms the known temperature trend in the stratosphere. The estimated component map of X_2 in Figure 3 reveals that the seasonal effect on temperature gets stronger as the latitude moves away, to the south and to the north, from around the level 20 at which it looks weakest. The effect is generally stronger in the southern hemisphere than in the northern hemisphere. The effect of latitude for each given time during a year is also strong, which seems obvious, while the effect of longitude at each given latitude is relatively low although there is a sign of interaction between them. In conclusion, through this
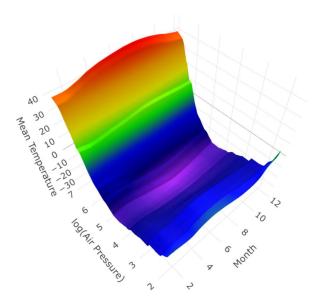
FIG. 2. *Estimated component map based on the proposed method for the temperature data*, *corresponding to the real-valued predictor.*

particular example we confirm that our method is also capable of analyzing quite successfully the effect of a mixture of real and spherical predictors on a functional response.

7. Conclusion and discussion. In this paper, for the first time in structured nonparametric regression, we studied the estimation of the additive model (2.1) that can deal with general Hilbertian responses and general semimetric space-valued predictors. The model allows for a wide range of data types. It accommodates any of Euclidean, compositional, functional and density-valued variables as a response, and any mixture of Euclidean, compositional, spherical, toroidal and shape variables as a predictor set, for example. Our general framework may also feed into the model possible interaction effects between predictors in the predictor set by considering product Hilbert spaces or product Riemannian manifolds for various combinations of predictors. In this general setting, we gave full practical details of implementing the proposed method for finite-dimensional Hilbertian/Riemannian predictors, which include
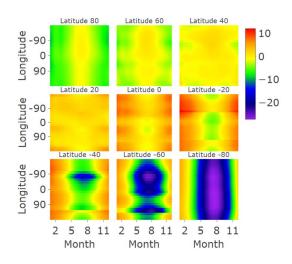

FIG. 3. *Estimated component map based on the proposed method for the temperature data*, *corresponding to the spherical predictor.*

integration over nonstandard domains with nonstandard measures. We also developed a complete theory for Hilbertian/Riemannian predictors. We proved that our estimators of the regression map and its components achieve the optimal rates of convergence in various modes under general bandwidth conditions, and that they converge weakly to Hilbertian Gaussian distributions. In particular, we further generalized the theory on the rates of convergence to semimetric space-valued predictors and to non-i.i.d. data pairs $(\mathbf{X}_i, \mathbf{Y}_i)$. Moreover, our theory establishing the existence of the sB-SBF estimator and the convergence of the sB-SBF algorithm also covers non-i.i.d. data with predictors taking values in general semimetric spaces. We also demonstrated that our approach performed better and gave better interpretation than competing methods on a variety of simulated and real data sets.

We note that our work in this paper focuses mainly on finite-dimensional predictors. However, we may also treat functional predictors, which are of infinite-dimension, in the framework of this paper by taking their functional principal components (FPC) as predictors. The latter is the approach taken by [13] for real-valued responses, where the effects of estimating the FPCs based on the eigendecomposition of the sample autocovariance surfaces of the functional predictors are analyzed. It may be of interest then to extend the work of [13] to Hilbertian responses based on the theory we developed in this paper. In the latter extension, we may take each individual FPC as a univariate predictor, or each vector of the individual FPCs corresponding to each functional variable as a multivariate predictor. There are other topics that deserve future study. For example, one may study additive regression for Riemannian responses and for incompletely observed responses such as those subject to missing or censoring. Another topic would be high-dimensional additive regression for non-Euclidean predictors and responses. In the latter problem, one may adopt a penalization scheme to choose a smaller number of relevant predictors.

SUPPLEMENTARY MATERIAL

Supplement to "Additive regression for non-Euclidean responses and predictors" (DOI: 10.1214/21-AOS2048SUPP; .pdf). The supplementary material contains all technical proofs, and additional implementation details, real data analysis and theoretical results.

REFERENCES

[1] CHACÓN, J. E., MATEU-FIGUERAS, G. and MARTÍN-FERNÁNDEZ LAUDER, J. A. (2014). Gaussian kernels for density estimation with compositional data. *Comput. Geosci.* **37** 702–711.

[2] CHAKRABORTY, R. and VEMURI, B. C. (2019). Statistics on the Stiefel manifold: Theory and applications. *Ann. Statist.* **47** 415–438. MR3909937 https://doi.org/10.1214/18-AOS1692

[3] CHENG, M.-Y. and WU, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J. Amer. Statist. Assoc.* **108** 1421–1434. MR3174718 https://doi.org/10.1080/01621459.2013.827984

[4] CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. MR3611755 https://doi.org/10.1111/rssb.12169

[5] DI MARZIO, M., PANZERA, A. and TAYLOR, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.* **79** 2066–2075. MR2571770 https://doi.org/10.1016/j.spl.2009.06.014

[6] DI MARZIO, M., PANZERA, A. and TAYLOR, C. C. (2014). Nonparametric regression for spherical data. *J. Amer. Statist. Assoc.* **109** 748–763. MR3223747 https://doi.org/10.1080/01621459.2013.866567

[7] DI MARZIO, M., PANZERA, A. and VENIERI, C. (2015). Non-parametric regression for compositional data. *Stat. Model.* **15** 113–133. MR3325749 https://doi.org/10.1177/1471082X14535522

[8] DRYDEN, I. L. and MARDIA, K. V. (2016). *Statistical Shape Analysis with Applications in R*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3559734 https://doi.org/10.1002/9781119072492

[9] EGOZCUE, J. J., PAWLOWSKY-GLAHN, V., MATEU-FIGUERAS, G. and BARCELÓ-VIDAL, C. (2003). Isometric logratio transformations for compositional data analysis. *Math. Geol.* **35** 279–300. MR1986165 https://doi.org/10.1023/A:1023818214614

[10] FAN, Y., JAMES, G. M. and RADCHENKO, P. (2015). Functional additive regression. *Ann. Statist.* **43** 2296–2325. MR3396986 https://doi.org/10.1214/15-AOS1346

[11] FEBRERO-BANDE, M. and GONZÁLEZ-MANTEIGA, W. (2013). Generalized additive models for functional data. *TEST* **22** 278–292. MR3062258 https://doi.org/10.1007/s11749-012-0308-0

[12] GONZALEZ-MANTEIGA, W., HENRY, G. and RODRIGUEZ, D. (2012). Partly linear models on Riemannian manifolds. *J. Appl. Stat.* **39** 1797–1809. MR2935559 https://doi.org/10.1080/02664763.2012.683169

[13] HAN, K., MÜLLER, H.-G. and PARK, B. U. (2018). Smooth backfitting for additive modeling with small errors-in-variables, with an application to additive functional regression for multiple predictor functions. *Bernoulli* **24** 1233–1265. MR3706793 https://doi.org/10.3150/16-BEJ898

[14] HAN, K., MÜLLER, H.-G. and PARK, B. U. (2020). Additive functional regression for densities as responses. *J. Amer. Statist. Assoc.* **115** 997–1010. MR4107695 https://doi.org/10.1080/01621459.2019.1604365

[15] HAN, K. and PARK, B. U. (2018). Smooth backfitting for errors-in-variables additive models. *Ann. Statist.* **46** 2216–2250. MR3845016 https://doi.org/10.1214/17-AOS1617

[16] HENRY, G. and RODRIGUEZ, D. (2009). Kernel density estimation on Riemannian manifolds: Asymptotic results. *J. Math. Imaging Vision* **34** 235–239. MR2515446 https://doi.org/10.1007/s10851-009-0145-2

[17] HIELSCHER, R. (2013). Kernel density estimation on the rotation group and its application to crystallographic texture analysis. *J. Multivariate Anal.* **119** 119–143. MR3061419 https://doi.org/10.1016/j.jmva.2013.03.014

[18] JEON, J. M. and PARK, B. U. (2020). Additive regression with Hilbertian responses. *Ann. Statist.* **48** 2671–2697. MR4152117 https://doi.org/10.1214/19-AOS1902

[19] JEON, J. M., PARK, B. U. and VAN KEILEGOM, I. (2021). Supplement to "Additive regression for non-Euclidean responses and predictors." https://doi.org/10.1214/21-AOS2048SUPP

[20] KENT, J. T. (1994). The complex Bingham distribution and shape analysis. *J. Roy. Statist. Soc. Ser. B* **56** 285–299. MR1281934

[21] KIM, J. S., STAICU, A.-M., MAITY, A., CARROLL, R. J. and RUPPERT, D. (2018). Additive function-on-function regression. *J. Comput. Graph. Statist.* **27** 234–244. MR3788315 https://doi.org/10.1080/10618600.2017.1356730

[22] LANG, R. (1986). A note on the measurability of convex sets. *Arch. Math. (Basel)* **47** 90–92. MR0855142 https://doi.org/10.1007/BF01202504

[23] LE, H. L. and KENDALL, D. G. (1993). The Riemannian structure of Euclidean shape spaces: A novel environment for statistics. *Ann. Statist.* **21** 1225–1271. MR1241266 https://doi.org/10.1214/aos/1176349259

[24] LEE, J. M. (2018). *Introduction to Riemannian Manifolds. Graduate Texts in Mathematics* **176**. Springer, Cham. MR3887684

[25] LEE, Y. K., MAMMEN, E. and PARK, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *Ann. Statist.* **38** 2857–2883. MR2722458 https://doi.org/10.1214/10-AOS808

[26] LEE, Y. K., MAMMEN, E. and PARK, B. U. (2012). Flexible generalized varying coefficient regression models. *Ann. Statist.* **40** 1906–1933. MR3015048 https://doi.org/10.1214/12-AOS1026

[27] LIN, Z. and YAO, F. (2019). Intrinsic Riemannian functional data analysis. *Ann. Statist.* **47** 3533–3577. MR4025751 https://doi.org/10.1214/18-AOS1787

[28] LINTON, O. and NIELSEN, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82** 93–100. MR1332841 https://doi.org/10.1093/biomet/82.1.93

[29] MAMMEN, E., LINTON, O. and NIELSEN, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27** 1443–1490. MR1742496 https://doi.org/10.1214/aos/1017939137

[30] MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biom. J.* **56** 732–753. MR3258083 https://doi.org/10.1002/bimj.201300072

[31] NIELSEN, J. P. and SPERLICH, S. (2005). Smooth backfitting in practice. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 43–61. MR2136638 https://doi.org/10.1111/j.1467-9868.2005.00487.x

[32] OPSOMER, J. D. and RUPPERT, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.* **25** 186–211. MR1429922 https://doi.org/10.1214/aos/1034276626

[33] PELLETIER, B. (2005). Kernel density estimation on Riemannian manifolds. *Statist. Probab. Lett.* **73** 297–304. MR2179289 https://doi.org/10.1016/j.spl.2005.04.004

[34] PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Non-parametr. Stat.* **18** 57–67. MR2214065 https://doi.org/10.1080/10485250500504828

[35] PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

[36] REIMHERR, M., SRIPERUMBUDUR, B. and TAOUFIK, B. (2018). Optimal prediction for additive function-on-function regression. *Electron. J. Stat.* **12** 4571–4601. MR3893421 https://doi.org/10.1214/18-EJS1505

[37] ROSENTHAL, M., WU, W., KLASSEN, E. and SRIVASTAVA, A. (2014). Spherical regression models using projective linear transformations. *J. Amer. Statist. Assoc.* **109** 1615–1624. MR3293614 https://doi.org/10.1080/01621459.2014.892881

[38] SAKAI, T. (1996). *Riemannian Geometry. Translations of Mathematical Monographs* **149**. Amer. Math. Soc., Providence, RI. Translated from the 1992 Japanese original by the author. MR1390760 https://doi.org/10.1090/mmono/149

[39] SCHEIPL, F., STAICU, A.-M. and GREVEN, S. (2015). Functional additive mixed models. *J. Comput. Graph. Statist.* **24** 477–501. MR3357391 https://doi.org/10.1080/10618600.2014.901914

[40] SIMÓ, A., IBÁÑEZ, M. V., EPIFANIO, I. and GIMENO, V. (2020). Generalized partially linear models on Riemannian manifolds. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **69** 641–661. MR4098966

[41] TU, L. W. (2017). *Differential Geometry.* Springer, Berlin.

[42] WOOD, S. N. (2017). *Generalized Additive Models: An Introduction with R.* Chapman and Hall/CRC.

[43] YU, K., PARK, B. U. and MAMMEN, E. (2008). Smooth backfitting in generalized additive models. *Ann. Statist.* **36** 228–260. MR2387970 https://doi.org/10.1214/009053607000000596

[44] YUAN, Y., ZHU, H., LIN, W. and MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 697–719. MR2965956 https://doi.org/10.1111/j.1467-9868.2011.01022.x