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 semi-metric spaces, which covers a very wide range of non-standard response variables and predictors. A general framework of estimating additive models is presented for semi-metric-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.
present a general framework of estimating the model (1.1) for predictors taking values in general semi-metric 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 artificial/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 [5]. Some examples of special-orthogonal-matrix-valued data include vector-cardiograms ([2]) and alignments of crystals ([13]). There are many examples of Hilbertian responses that can be regressed on a mixture of Hilbertian and Riemannian 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. [25] 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 [43], [21], [22] and [12], among others. For an account of the practical issues of the method, see [30]. In recent years, there have been some achievements of extending the original work of [25] on the additive model to the case of more general spaces for responses and predictors. To list a few here, [11] treated density-valued responses $Y(\cdot)$ with real-valued predictors, but in a pointwise manner, i.e., 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. [15] 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 [11]. The present paper aims at an entirely different level of generalization. We present a novel theory for a general Hilbertian response coupled with general semi-metric-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 [33], [27] 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, [23] 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 [42]. Others are spherical-spherical regression [28, 36], 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 [34]. 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 [24], the classical backfitting [31, 8], the reproducing kernel Hilbert space approach [35] and the spline-based technique [7, 38, 17]. 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 multi-dimensional 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 [15]. 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 semi-metric-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 Supplement 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 semi-metric spaces and responses taking values in a separable Hilbert space. Let \((X_j, d_j)\) for \(1 \leq j \leq d\) be semi-metric 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 semi-metric spaces for predictor domains since they are one of the most general domains for kernel smoothing that we adopt in this paper. Let \(X = (X_1, \ldots, X_d)\) and \(X = \prod_{j=1}^d X_j\). We denote by \(\mathcal{B}(X_j)\) the Borel sigma-field of \(X_j\) induced by the semi-metric \(d_j\) and let \(\mu_j\) be a measure on \(\mathcal{B}(X_j)\) satisfying \(0 < \mu_j(X_j) < \infty\).

For example, when \(X_j\) is a finite-dimensional Hilbert space, we may take \(d_j\) as the metric induced by an inner product of the Hilbert space, and \(\mu_j\) as the pushforward measure induced by a Lebesgue measure. Specifically, if \(X_j\) is a \(q_j\)-dimensional space, the pushforward measure \(\mu_j\) is given by \(\mu_j(A) = \text{Leb}_{q_j} (\eta_j(A)) \) for \(A \in \mathcal{B}(X_j)\), where \(\text{Leb}_{q_j}\) denotes the \(q_j\)-dimensional Lebesgue measure and \(\eta_j : X_j \rightarrow \mathbb{R}^{q_j}\) is an isometric isomorphism. When \(X_j\) is a finite-dimensional Riemannian manifold, we may take \(d_j\) and \(\mu_j\), respectively, as the Riemannian distance and Riemannian volume measure induced by a Riemannian metric. In Section 3.3, we give concrete examples of \(d_j\) and \(\mu_j\) 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 \(Y\) reside. Throughout this paper we assume \(E(\|Y\|^2) < \infty\). Let \(\oplus, \odot, 0\) 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 $0$ are given in the Supplement S.1, see also [15]. In this paper, we consider an additive model

\begin{equation}
Y = E(Y) \oplus \bigoplus_{j=1}^{d} f_j(X_j) \oplus \epsilon,
\end{equation}

where $f_j : \mathcal{X}_j \to \mathbb{H}$ are $(\mathcal{B}(\mathcal{X}_j), \mathcal{B}(\mathbb{H}))$-measurable maps satisfying $E(f_j(X_j)) = 0$ and $E(\|f_j(X_j)\|^2) < \infty$ for $1 \leq j \leq d$, and $\epsilon$ is an $\mathbb{H}$-valued error satisfying $E(\epsilon(X)) = 0$. The constraints $E(f_j(X_j)) = 0, 1 \leq j \leq d$, are for the identifiability of the component maps $(f_j : 1 \leq j \leq d)$. 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 [15] and also the Supplement S.2.

2.1. Methodology. We assume that the probability distribution $PX^{-1}$ of $X$ is absolutely continuous with respect to the product measure $\mu := \otimes_{j=1}^{d} \mu_j$, so that the density $p := dPX^{-1}/d\mu$ 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 $\int_{\mathcal{X}_j} p_{jk}^2(x_j, x_k) / p_k(x_k) \, d\mu_k(x_k) < \infty$ hold for all $1 \leq j \neq k \leq d$ and $x_j \in \mathcal{X}_j$. Then, by taking the conditional expectation $E(\cdot|X_j = x_j)$ on both sides of (2.1) and applying Proposition 2.2 in [15], we get

\begin{equation}
f_j(x_j) = m_j(x_j) \ominus E(Y) \ominus \bigoplus_{k \neq j} \int_{\mathcal{X}_k} f_k(x_k) \odot \frac{p_{jk}(x_j, x_k)}{p_j(x_j)} \, d\mu_k(x_k), \quad 1 \leq j \leq d,
\end{equation}

where $m_j$ are the marginal regression maps such that $m_j(x_j) = E(Y|X_j = x_j)$ and $\ominus$ is a vector subtraction of $\mathbb{H}$ defined by $h_1 \ominus h_2 = h_1 \oplus (-1 \odot h_2)$. 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 $(f_j : 1 \leq j \leq d)$.

Suppose that we have $n$ observations $\{(Y_i, X_i)\}_{1 \leq i \leq n}$. 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 [15]. For $x = (x_1, \ldots, x_d) \in \mathcal{X}$, we estimate $p(x)$ by $\hat{p}(x) = n^{-1} \sum_{i=1}^{n} \prod_{j=1}^{d} L_{j,n}(x_j, X_{ij})$, 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

\begin{equation}
\int_{\mathcal{X}_j} L_{j,n}(x_j, u_j) \, d\mu_j(x_j) = 1
\end{equation}

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

We also estimate $E(Y)$ and $m_j$ in (2.2) by
\[
\hat{E}(Y) = n^{-1} \circ \bigoplus_{i=1}^{n} Y_i = \bar{Y}, \quad \hat{m}_j(x_j) = \frac{1}{n \hat{p}_j(x_j)} \circ \bigoplus_{i=1}^{n} L_{j,n}(x_j, X_{ij}) \circ Y_i.
\]

We consider the following condition on the density estimators.

**CONDITION (S).** For all $1 \leq j \neq k \leq d$ and $x_j \in X_j$, we have $\hat{p}_j(x_j) > 0$, $\int_{X_k} \hat{p}_{jk}^2(x_j, x_k)/\hat{p}_k(x_k) d\mu_k(x_k) < \infty$ and
\[
\int_{X_k} \int_{X_j} \hat{p}_{jk}^2(x_j, x_k) \hat{p}_k(x_k) \frac{d\mu_j(x_j)}{\hat{p}_j(x_j)} \mu_k(x_k) < \infty.
\]

Under the conditions (S) and that $\max_{1 \leq j \leq d} \int_{X_j} \|\hat{f}_j(x_j)\| \hat{p}_j(x_j) d\mu_j(x_j) < \infty$, we have $\int_{X_k} \|\hat{f}_k(x_k)\| \hat{p}_{jk}(x_j, x_k) d\mu_k(x_k) < \infty$ 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,
\[
(2.4) \quad \hat{f}_j(x_j) = \hat{m}_j(x_j) \oplus Y \oplus \bigoplus_{k \neq j} \int_{X_k} \hat{f}_k(x_k) \circ \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} d\mu_k(x_k), \quad 1 \leq j \leq d,
\]

are well-defined. The smooth backfitting (SBF) estimator of the tuple $(f_j : 1 \leq j \leq d)$ of the component maps is then defined as a solution $(\hat{f}_j : 1 \leq j \leq d)$ of the system of equations in (2.4) subject to the constraints
\[
(2.5) \quad \int_{X_j} \hat{f}_j(x_j) \oplus \hat{p}_j(x_j) d\mu_j(x_j) = 0, \quad 1 \leq j \leq d,
\]
\[
\int_{X_j} \|\hat{f}_j(x_j)\|^2 \hat{p}_j(x_j) d\mu_j(x_j) < \infty, \quad 1 \leq j \leq d.
\]

The above constraints are the empirical versions of the constraints $E(f_j(X_j)) = 0$ and $E(\|f_j(X_j)\|^2) < \infty$ on the true tuple $(f_j : 1 \leq j \leq d)$. We call $(\hat{f}_j : 1 \leq j \leq d)$ semi-metric 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 $P X^{-1}$ on $\otimes_{j=1}^d B(X_j)$ by $P X^{-1}(A) = \int_A \hat{p}(x) d\mu(x)$.
Proposition 2.1. Assume the condition (S). Then, there exists a solution \((\hat{f}_j : 1 \leq j \leq d)\) of (2.4) satisfying (2.5), and the sum of such \(\hat{f}_j\) for \(1 \leq j \leq d\) is unique up to measure zero with respect to \(\hat{P}X^{-1}\). Moreover, when \(\hat{p} > 0\) on \(X\), the solution is unique in the sense that, if \((\hat{f}_j^* : 1 \leq j \leq d)\) is another solution, then \(\hat{f}_j(x_j) = \hat{f}_j^*(x_j)\) a.e. with respect to \(\mu_j\) for all \(1 \leq j \leq d\).

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 \((\hat{f}_j^{[0]} : 1 \leq j \leq d)\) satisfying

\[
\int_{X_j} \hat{f}_j^{[0]}(x_j) \odot \hat{p}_j(x_j) \, d\mu_j(x_j) = 0, \quad 1 \leq j \leq d,
\]

(2.6)

\[
\int_{X_j} \|\hat{f}_j^{[0]}(x_j)\|^2 \hat{p}_j(x_j) \, d\mu_j(x_j) < \infty, \quad 1 \leq j \leq d.
\]

We update them in the \(r\)th iteration by

\[
\hat{f}_j^{[r]}(x_j) = \hat{m}_j(x_j) \odot \hat{Y} \odot \bigoplus_{k<j} \int_{X_k} \hat{f}_k^{[r]}(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} \, d\mu_k(x_k) \\
\bigoplus \bigoplus_{k>j} \int_{X_k} \hat{f}_k^{[r-1]}(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} \, d\mu_k(x_k), \quad 1 \leq j \leq d.
\]

(2.7)

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

\[
\int_{X_j} \hat{f}_j^{[r]}(x_j) \odot \hat{p}_j(x_j) \, d\mu_j(x_j) = 0, \quad 1 \leq j \leq d
\]

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

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

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

\[
\int_X \|\tilde{f}(x) \odot \tilde{f}^{[r]}(x)\|^2 \hat{p}(x) \, d\mu(x) \leq \hat{c} \cdot \hat{\gamma}^r \text{ for all } r \geq 0.
\]
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) $\max_{1 \leq j \leq d} \sup_{x_j \in X_j} |\hat{p}_j(x_j)|^{-1} < C$;  
(A2) $\max_{1 \leq j \neq k \leq d} \sup_{(x_j, x_k) \in X_j \times X_k} |\hat{p}_{jk}(x_j, x_k)| < C$;  
(A3) $\max_{1 \leq j \leq d} \sup_{x_j \in X_j} \|\hat{m}_j(x_j)\| < C$;  
(A4) $\max_{1 \leq j \leq d} \int_{X_j} \|\hat{f}_j^{[0]}(x_j)\|^2 d\mu_j(x_j) < C$.

Also, it holds that

(A5) $\max_{1 \leq j \leq d} \int_{X_j} (\hat{p}_j(x_j) - p_j(x_j))^2 d\mu_j(x_j) = o_p(1)$,  
(A6) $\max_{1 \leq j \neq k \leq d} \int_{X_j \times X_k} (\hat{p}_{jk}(x_j, x_k) - p_{jk}(x_j, x_k))^2 d\mu_j \otimes \mu_k(x_j, x_k) = o_p(1)$,  
(A7) $p$ is bounded away from zero and infinity on $X$.

We note that the condition (A7) on the underlying distribution of $X$ 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 $1 \leq j \leq d$,

$$\lim_{n \to \infty} P \left( \int_{X_j} \|\hat{f}_j^{[r]}(x_j) \otimes \hat{f}_j^{[r]}(x_j)\| p_j(x_j) d\mu_j(x_j) \leq c \cdot \gamma^r \text{ for all } r \geq 0 \right) = 1.$$  
Moreover, for all $1 \leq j \leq d$,

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

As in [15] we may implement the sB-SBF algorithm using Lebesgue integration. This is because of the fact that, for any measure space $(Z, \mathcal{A}, \mu)$,

$$\text{(Bochner)} \int_Z f(z) \otimes b \ d\mu(z) = \text{(Lebesgue)} \int_Z f(z) d\mu(z) \otimes b$$

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

$$(2.8) \quad \hat{f}_j^{[0]}(x_j) = n^{-1} \odot \bigoplus_{i=1}^{n} w_{ij}^{[0]}(x_j) \odot Y_i, \quad 1 \leq j \leq d,$$
for some continuous functions $w_{ij}^0 : X_j \to \mathbb{R}$ satisfying $\int_{X_j} w_{ij}^0(x_j)p_j(x_j)d\mu_j(x_j) = 0$. We note that the constraints are satisfied by the trivial choice $w_{ij}^0 \equiv 0$. Define

$$w_{ij}^r(x_j) = \frac{L_{i,n}(x_j, X_{ij})}{\bar{p}_j(x_j)} - 1 - \sum_{k<j} \int_{X_k} w_{ik}^r(x_k)\frac{\hat{p}_{jk}(x_j, x_k)}{\bar{p}_j(x_j)} d\mu_k(x_k)$$

$$- \sum_{k>j} \int_{X_k} w_{ik}^{r-1}(x_k)\frac{\hat{p}_{jk}(x_j, x_k)}{\bar{p}_j(x_j)} d\mu_k(x_k).$$

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

$$\hat{f}_j^r(x_j) = n^{-1} \odot \bigoplus_{i=1}^n w_{ij}^r(x_j) \odot Y_i, \quad 1 \leq j \leq d$$

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 semi-metric spaces, which are Hilbert spaces and Riemannian manifolds. To treat both types in a unified fashion, we let $0 \leq \rho \leq d$ be an integer and assume that $X_j$ for $1 \leq j \leq \rho$ are compact subsets of $q_j$-dimensional Hilbert spaces $H_j$ and $X_j$ for $\rho + 1 \leq j \leq d$ are $q_j$-dimensional connected and compact Riemannian manifolds without boundary. In both cases, we assume that $1 \leq q_j < \infty$. We denote a vector subtraction and a norm of $H_j$ by $\ominus_j$ and $\| \cdot \|_j$, respectively. We let $g_j$ denote the Riemannian metric (sometimes called Riemannian structure) of $X_j$ for $\rho + 1 \leq j \leq d$ 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 $X_j$ for $\rho + 1 \leq j \leq d$. For an introduction to Riemannian geometry, we refer to [40] and [20], for example. We also collect basic notions related to Riemannian manifolds in the Supplement S.3. In our unified framework, $\rho = 0$ means that there is no Hilbertian predictor and $\rho = d$ means that there is no Riemannian predictor. We note that, for any $0 \leq \rho \leq d$, no nonparametric additive model has been studied even for $H = \mathbb{R}$. Our treatment covers all the cases $0 \leq \rho \leq d$ with general $H$.

Examples of $H_j$ 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 $S^q_j = \{x_j = (x_{j,1}, \ldots, x_{j,q_j+1}) \in \mathbb{R}^{q_j+1} : \sum_{l=1}^{q_j+1} x_{j,l}^2 = 1\}$, toruses $S^1 \times \cdots \times S^1$, planar shape spaces ([19]), 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 \((Y, X)\), which we denote by \((Y_i, X_i)\), \(1 \leq i \leq n\). Each component \( X_j \) of \( X \) 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 \( \mu_j \) to be the pushforward measure on \( \mathcal{B}(\mathcal{H}_j) \) such that \( \mu_j(A) = \text{Leb}_{q_j}(\eta_j(A)) \). For each Riemannian predictor \( X_j \) taking values in the Riemannian manifold \((\mathcal{X}_j, g_j)\), we choose \( \mu_j \) to be the Riemannian volume measure induced by \( g_j \), see Section 3.3.2 for the definition.

If \( q = d \) and \( \mathcal{X}_j \) are compact and convex subsets of \( \mathbb{R}^{q_j} \), so that we may take \( \eta_j \) to be the identity map on \( \mathbb{R}^{q_j} \) and \( \mu_j = \text{Leb}_{q_j} \), then the results here and in Section 4 completely generalize those in [15] 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 [15] 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 \( f_j \). We also have the corresponding sB-SBF algorithm from (2.7) and the corresponding updating formula (2.9) that we actually iterate.

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

\[
K_{h,j}(x_j, u_j) = \frac{K_j(||x_j \otimes u_j||_j/h_j)}{\int_{\mathcal{X}_j} K_j(||v_j \otimes u_j||_j/h_j) d\mu_j(v_j)}
\]

whenever \( \int_{\mathcal{X}_j} K_j(||v_j \otimes u_j||_j/h_j) d\mu_j(v_j) \neq 0 \), 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 [32] defined by

\[
K_{h,j}(x_j, u_j) = \frac{1}{h_j^{q_j}} \frac{1}{\partial_j(x_j; u_j)} K_j \left( \frac{d_j(x_j, u_j)}{h_j} \right) I(d_j(x_j, u_j) < \text{inj}(\mathcal{X}_j)),
\]

where \( d_j(x_j, u_j) \) is the distance between \( x_j \) and \( u_j \) in the Riemannian metric induced by \( g_j \).
where \( \theta_j(\cdot; u_j) \) is the volume density function of \( X_j \) at \( u_j \in X_j \), and \( \text{inj}(X_j) \) is the injectivity radius of \( X_j \). For the definitions of volume density function and injectivity radius, see [9] or the Supplement S.3. Since \( X_j \) is compact, \( \text{inj}(X_j) > 0 \). We note that \( \theta_j(x_j; u_j) \) is defined only for \( x_j \) and \( u_j \) such that \( d_j(x_j, u_j) < \text{inj}(X_j) \), 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 \( \{x_j \in X_j : d_j(x_j, u_j) \leq r\} \) for any \( r < \text{inj}(X_j) \). The kernel function \( K_{h_j}(x_j, u_j) \) defined at (3.2) satisfies (2.3) if \( h_j < \text{inj}(X_j) \).

We note that [15] studied the additive model (2.1) with \( X_j = [0, 1] \subset \mathbb{R} \) for all \( 1 \leq j \leq d \) and kernel \( K_{h_j} \) defined by

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

where \( K : \mathbb{R} \to [0, \infty) \) is a measurable function that vanishes on \((-\infty, -1) \cup (1, \infty)\) and satisfies \( \int_0^1 K(t)dt = 1 \). 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 \( 1 \leq j \leq g \), we denote \( d_j(x_j, u_j) = \|x_j \oplus_j u_j\|_j \) and \( (\eta_j(X_j) - \eta_j(x_j))/h_j = \{ (\eta_j(u_j) - \eta_j(x_j))/h_j : u_j \in X_j \} \). For \( 1 \leq j \leq d \) and \( r > 0 \), we define \( B_{X_j}(x_j, r) = \{ u_j \in X_j : d_j(x_j, u_j) < r \} \). We denote by \( \mathbf{0}_{q_j} \) the zero vector of \( \mathbb{R}^q \), 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 non-asymptotic 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: \( 1 \leq j \leq g \)). \( \kappa_j := \sup_{x_j \in X_j} \min_{1 \leq i \leq n} d_j(x_j, X_{ij}) < h_j \)
and \( K_j \) is bounded away from zero on \([0, \kappa_j/h_j]\). Also, for all \( x_j \in X_j \),

\[
\text{Leb}_{\eta_j}(\{(\eta_j(X_j) - \eta_j(x_j))/h_j \in [0, \kappa_j/h_j] \}).
\]

(B2) *(Riemannian predictors: \( g + 1 \leq j \leq d \)). \( \gamma_j := \sup_{x_j \in X_j} \min_{1 \leq i \leq n} d_j(x_j, X_{ij}) < h_j < \text{inj}(X_j) \), 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 \( d_j(\cdot, X_{ij}) : X_j \to [0, \infty) \) are continuous and \( X_j \) 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 $K_{b_j}(x_j, X_{ij})$ as defined at (3.1) is nonzero. We also note that a large class of $X_j$ satisfies (3.4) in (B1). Below, we introduce a proposition that gives a class of $X_j$ 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 $\text{Leb}_{q_j}((\eta_j(X_j) - \eta_j(x_j))/h_j \cap B(0_{q_j}, r)) > 0$ for any $x_j \in X_j$, $h_j > 0$ and $r > 0$ if $\eta_j(X_j)$ 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 $X_j$ is a convex subset of $H_j$, then $\eta_j(X_j)$ 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 $\partial X_j$, for $1 \leq j \leq q$, denote the boundary of $X_j$.

**Condition (C).**

(C1) $E(||Y||^\alpha) < \infty$ for some $\alpha > 2$, and $E(\|Y\|^2 | X_j = \cdot)$ are bounded on the respective $X_j$ for $1 \leq j \leq d$.

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

(C3) (Hilbertian predictors: $1 \leq j \leq q$). (i) $\mu_j(\partial X_j) = 0$ and there exist positive constants $\tau_j < 1$ and $\epsilon_j$ such that

\[
\inf_n \inf_{x_j \in X_j} \text{Leb}_{q_j}((\eta_j(X_j) - \eta_j(x_j))/h_j \cap B(0_{q_j}, \tau_j)) \geq \epsilon_j;
\]

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

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

\[
| (\theta_j(x_j; u_j))^{-1} - (\theta_j(x_j^*; u_j))^{-1} | \leq L_j \cdot d_j(x_j, x_j^*);
\]
(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 non-asymptotic 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 $(X, Y)$. Nevertheless, they are related. Indeed, one may show that the condition (C2)-(i) with $\inf_n n^{c_j} h_j > 0$ for some $c_j < (\alpha - 2)/\alpha$ in (C2)-(ii) implies that the condition $\sup_{x_j \in X_j} \min_{1 \leq i \leq n} d_j(x_j, X_{ij}) < h_j$ 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 $\mu_j$ being either the pushforward measure or the Riemannian volume measure, $\inf_{x_j \in X_j} \mu_j(B_{X_j}(x_j, \delta)) \geq c(q_j) \cdot \delta^{q_j}$ 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), $E(\|Y\|^2 | X_j = \cdot)$ is bounded on $X_j$ if it is continuous on $X_j$. The condition $\mu_j(\partial X_j) = 0$ in (C3) is weak and holds for most of sets. For example, the convexity of $X_j$ implies the condition, see [18]. We may also show that (3.5) in (C3) holds when $X_j$ is convex, using Proposition 3.1 and the fact that the map $x_j \mapsto \text{Leb}_{q_j}(\eta_j^{-1}(\partial X_j, \delta))$ is continuous on $X_j$ 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 non-standard domains with non-standard 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 $X_k \subset H_k$, we may implement the integral over $X_k$ in (2.9) using the property

$$\int_{X_k} f(x_k) d\mu_k(x_k) = \int_{\eta_k(x_k)} f(\eta_k^{-1}(t)) dt,$$

where $f$ is any integrable function on $X_k$. Below, we give detailed examples of the implementation.
(i) **Euclidean spaces.** In the case of $X_k \subset \mathbb{R}^{q_k}$, we may take $\eta_k$ to be the identity map on $X_k$. Then, it is a standard practice to implement the integral over $X_k$ with respect to $\mu_k = \text{Leb}_{q_k}$ in (2.9) since it is the usual numerical integration.

(ii) **Simplices.** Consider the $q_k$-dimensional simplex $S_1^{q_k+1}$ defined by

$$S_1^{q_k+1} = \left\{ x_k = (x_{k,1}, \ldots, x_{k,q_k+1}) \in (0,1)^{q_k+1} : \sum_{l=1}^{q_k+1} x_{k,l} = 1 \right\}.$$ 

The inner product $\langle \cdot, \cdot \rangle_k$ on $S_1^{q_k+1}$ 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 $\{e_1, \ldots, e_{q_k}\}$ be its orthonormal basis. In this case, we may take $\eta_k$ to be the isometric log-ratio transformation $ilr_k : S_1^{q_k+1} \to \mathbb{R}^{q_k}$ defined by

$$ilr_k(x_k) = \left( \langle x_k, e_1 \rangle_k, \ldots, \langle x_k, e_{q_k} \rangle_k \right),$$

and use the Aitchison measure $\mu_k$ defined by $\mu_k(A) = \text{Leb}_{q_k}(ilr_k(A))$. The notion of Aitchison measure has been studied in density estimation and regression analysis for compositional data, see [1] and [29], for example. The $\mu_k$ is a natural measure on the simplex $S_1^{q_k+1}$, 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 $S_1^{q_k+1}$ and any Borel set $A \subset S_1^{q_k+1}$, it holds that

$$\int_A f(x_k) d\mu_k(x_k) = \int_{ilr_k(A)} f(ilr_k^{-1}(t)) dt.$$ 

To introduce an orthonormal basis of $S_1^{q_k+1}$, define the $(q_k + 1)$-dimensional vector

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

where $1_l$ is $l$-vector with all entries being equal to one. Denote the $l$th entry of $\xi_m$ by $\xi_{m,l}$. Define $e_m = (\xi_{m,1}/\sum_{l=1}^{q_k+1} \xi_{m,l}, \ldots, \xi_{m,q_k+1}/\sum_{l=1}^{q_k+1} \xi_{m,l})$. Then, $\{e_1, \ldots, e_{q_k}\}$ forms an orthonormal basis of $S_1^{q_k+1}$, see Proposition 2 of [6]. With this basis, the $m$th entry of $ilr_k(x_k) \in \mathbb{R}^{q_k}$ for $x_k = (x_{k,1}, \ldots, x_{k,q_k+1}) \in S_1^{q_k+1}$ is given by

$$\sqrt{m/(m+1)} \log \left( \prod_{l=1}^{m} x_{k,l}^{1/m} / x_{k,m+1} \right).$$

Then, the integral over $X_k \subset S_1^{q_k+1}$ with respect to the Aitchison measure $\mu_k$ in (2.9) can be evaluated by the usual numerical integration:

$$\int_{X_k} w_{ik}^{[r]}(x_k) \cdot \hat{p}_{jk}(x_j, x_k) d\mu_k(x_k) = \int_{ilr_k(X_k)} w_{ik}^{[r]}(ilr_k^{-1}(t)) \cdot \hat{p}_{jk}(x_j, ilr_k^{-1}(t)) dt.$$
(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 \(X_k \subset \mathcal{H}_k\), 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 \(\langle x_k, u_{k,l} \rangle_k = \sum_{j=1}^J \langle x_{k,j}, u_{k,j} \rangle_{k,j}\) and the isometric isomorphism \(\eta_k : \mathcal{H}_k \to \mathbb{R}^{\sum_{j=1}^J q_{k,j}}\) is given by \(\eta_k(x_k) = (\eta_{k,1}(x_{k,1}), \ldots, \eta_{k,J}(x_{k,J}))\), where \(\eta_{k,j} : \mathcal{H}_{k,j} \to \mathbb{R}^{q_{k,j}}\) are isometric isomorphisms. With this \(\eta_k\) and the corresponding pushforward measure, we may implement the integral over \(X_k\) in (2.9).

### 3.3.2. Implementation for Riemannian predictors.

We first introduce Riemannian volume measure, which takes the role of the measure \(\mu_j\) on \(\mathcal{B}(X_j)\) when \(X_j\) is a Riemannian manifold. Let \((\mathcal{M}, g)\) be a \(m\)-dimensional compact Riemannian manifold. A pair \((U, \varphi)\) 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 \(\varphi_{\beta,1} : U \to \mathbb{R}\). Let \(\{\rho_{\beta}\}_\beta\) be a partition of unity subordinate to \(\{U_{\beta}\}_\beta\). Also, let \(\frac{\partial}{\partial \varphi_{\beta,j}}|_{\varphi_{\beta}^{-1}(t)}\) denote the function defined by \(\frac{\partial}{\partial \varphi_{\beta,j}}|_{\varphi_{\beta}^{-1}(t)}(h) = D_l(h \circ \varphi_{\beta}^{-1})(t)\) 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 \(G_{\beta}(\varphi_{\beta}^{-1}(t))\) to be the \(m \times m\) matrix whose \((l, l')\)th entry is given by

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

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

\[
(3.6) \quad F(f) = \sum_{\beta} \int_{\varphi_{\beta}(U_{\beta})} f(\varphi_{\beta}^{-1}(t)) \cdot \rho_{\beta}(\varphi_{\beta}^{-1}(t)) \cdot \sqrt{\det(G_{\beta}(\varphi_{\beta}^{-1}(t)))} \, dt.
\]

The value of \(F(f)\) is independent of the choices of \(\{(U_{\beta}, \varphi_{\beta})\}_\beta\) and \(\{\rho_{\beta}\}_\beta\), see Chapter 2 of [37], for example. Due to the Riesz representation theorem, there exists a unique regular Borel measure \(\nu\) on \(\mathcal{M}\) such that

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

We call such a \(\nu\) the **Riemannian volume measure induced by \(g\).**

According to (3.7), the integral of a continuous function over \(X_j\) with respect to the corresponding Riemannian volume measure \(\mu_j\) 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 \( x_k \mapsto w^{|r|}_{ik}(x_k) \cdot \hat{p}_{jk}(x_j, x_k) \) are continuous on the respective \( X_k \) for all \( x_j \in X_j \) and for all \( r \geq 0 \), \( 1 \leq i \leq n, 1 \leq j \leq d \) and \( q + 1 \leq k \leq d \) with \( k \neq j \).

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

\[
\int_{X_k} w^{|r|}_{ik}(x_k) \cdot \hat{p}_{jk}(x_j, x_k) \, d\mu_k(x_k) = \sum_{\beta_k} \int_{\varphi_{\beta_k}(U_{\beta_k})} w^{|r|}_{ik}(\varphi^{-1}_{\beta_k}(t)) \cdot \hat{p}_{jk}(x_j, \varphi^{-1}_{\beta_k}(t)) \\
\cdot \rho_{\beta_k}(\varphi^{-1}_{\beta_k}(t)) \cdot \sqrt{\det(G_{\beta_k}(\varphi^{-1}_{\beta_k}(t)))} \, dt,
\]

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

\[
g_k(\varphi^{-1}_{\beta_k}(t)) \left( \frac{\partial}{\partial \varphi^{-1}_{\beta_k,l}} \bigg|_{\varphi^{-1}_{\beta_k}(t)}, \frac{\partial}{\partial \varphi^{-1}_{\beta_k,l'}} \bigg|_{\varphi^{-1}_{\beta_k}(t)} \right).
\]

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

(i) **Toruses.** Consider the case where \( X_k = \prod_{l=1}^{q_k} S^1 \) for \( q_k \geq 1 \). Then, \( \text{inj}(X_k) = \pi \) and

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

In this case we may prove that \( \theta_k(\cdot ; \cdot) \equiv 1 \). Clearly, this volume density function satisfies (C4). Define a map \( \varphi^{-1} : (0, 2\pi)^{q_k} \to \prod_{l=1}^{q_k} S^1 \) 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, \( \varphi^{-1}((0, 2\pi)^{q_k}) \) is a chart on \( \prod_{l=1}^{q_k} S^1 \), where \( \varphi \) is understood to be the inverse of the map \( \varphi^{-1} \). Assume the condition (B). Then, since \( \mu(X_k) = \mu_k(\varphi^{-1}((0, 2\pi)^{q_k})) \), it holds that

\[
\int_{X_k} w^{|r|}_{ik}(x_k) \cdot \hat{p}_{jk}(x_j, x_k) \, d\mu_k(x_k) \\
= \int_{(0, 2\pi)^{q_k}} w^{|r|}_{ik}(\varphi^{-1}(\vartheta_1, \ldots, \vartheta_{q_k})) \cdot \hat{p}_{jk}(x_j, \varphi^{-1}(\vartheta_1, \ldots, \vartheta_{q_k})) \, d\vartheta_1 \ldots d\vartheta_{q_k},
\]

see Chapter 16 of [40].
(ii) **Spheres.** Consider now the case $X_k = S^{q_k} \subset \mathbb{R}^{q_k+1}$ for $q_k \geq 2$. In this case, $\text{inj}(X_k) = \pi$ and $d_k(x_k, u_k) = \arccos(x_k^\top u_k)$. Also, we may prove

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

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

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

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

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

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

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

$$\cdot \prod_{j=1}^{q_k-1} (\sin \phi_j)^2 \, d\theta \, d\phi_1 \cdots d\phi_{q_k-1},$$

see Chapter 16 of [40].

(iii) **Planar shape spaces.** Here, we consider $X_k = \Sigma^b_k$ for $b \geq 3$, where $\Sigma^b_k$ 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 $\Sigma^b_k$ 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 $z = (z_{1,1} + \sqrt{-1} \cdot z_{1,2}, \ldots, z_{b,1} + \sqrt{-1} \cdot z_{b,2})^\top \in \mathbb{C}^b$ the complex representation of the landmarks. Also, let $H$ be the $(b-1) \times b$ Helmert submatrix whose $j$th row is $(-1/\sqrt{j(j+1)}, \sqrt{j/(j+1)}, 0_{b-1-j})^\top$. Define $z_H = H z/\|H z\|_{\mathbb{C}^{b-1}}$, where $\| \cdot \|_{\mathbb{C}^{b-1}}$ denotes the complex norm on $\mathbb{C}^{b-1}$. Consider equivalent classes $[z] = \{z_H \cdot e^{\sqrt{-1} \cdot \vartheta} : 0 \leq \vartheta < 2\pi\}$. Then, $\Sigma^b_k$ is defined as the collection of these equivalent classes of $b$ landmarks. It is known that $\Sigma^b_k$ is a $(2b - 4)$-dimensional connected and compact Riemannian manifold without boundary. The Riemannian distance function on $\Sigma^b_k$ is given by

$$d_k([z], [w]) = \arccos(\|z_H^\top w_H\|_{\mathbb{C}}),$$
where $z$ denotes the conjugate of $z$, and $\text{inj}(\Sigma^b_2) = \pi/2$. Also,

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

and $\theta_k([z]; [z]) = 1$, as proved by [39]. This volume density function satisfies (C4). We note that $z_H = (\sqrt{s_1^b} \cdot e^{\sqrt{-1} \cdot \theta_1^b}, \ldots, \sqrt{s_{b-1}^b} \cdot e^{\sqrt{-1} \cdot \theta_{b-1}^b})$ for some

$$s^b := (s_1^b, \ldots, s_{b-2}^b) \in \left\{ (a_1, \ldots, a_{b-2}) \in [0,1]^{b-2} : \sum_{j=1}^{b-2} a_j \leq 1 \right\} =: S_{b-2}$$

with $s_{b-1}^b = 1 - \sum_{j=1}^{b-2} s_j^b$ and $\theta_j^b \in [0,2\pi)$ for all $1 \leq j \leq b - 1$. Let $\phi_j^b = (\theta_j^b - \theta_{b-1}^b) \mod(2\pi)$, $1 \leq j \leq b - 2$, and $\phi^b = (\phi_1^b, \ldots, \phi_{b-2}^b) \in [0,2\pi)^{b-2}$. For the interior $S_{b-2}^b$ of $S_{b-2}$, define a map $\varphi^{-1} : S_{b-2}^b \times (0,2\pi)^{b-2} \to \Sigma_2^b$ by $\varphi^{-1}(s^b, \phi^b) = [z]$. Then, $(\varphi^{-1}(S_{b-2}^b \times (0,2\pi)^{b-2}), \varphi)$ is a chart on $\Sigma_2^b$ ([16]). Since $\mu_k(\mathcal{X}_k) = \mu_k(\varphi^{-1}(S_{b-2}^b \times (0,2\pi)^{b-2}))$, under the condition (B), we get

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

$$= \int_{S_{b-2}^b \times (0,2\pi)^{b-2}} w_{ik}^{(r)}(\varphi^{-1}(s^b, \phi^b)) \cdot \hat{p}_{jk}(x_j, \varphi^{-1}(s^b, \phi^b)) \cdot 2^{2-b} \, ds^b \, d\phi^b,$$

see Chapter 4 of [5].

(iv) Product Riemannian manifolds. Suppose that $\mathcal{X}_k$ is the product of connected, compact and boundaryless Riemannian manifolds $(\mathcal{M}_j, g_{\mathcal{M}_j}, d_{\mathcal{M}_j})$, $1 \leq j \leq J$, where $\text{dim}(\mathcal{M}_j) = m_j$ 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 $d_\odot(x, u) = \left\{ \sum_{j=1}^{J} \left( d_{\mathcal{M}_j}(x_j, u_j) \right)^{1/2} \right\}^{1/2}$ for $x = (x_1, \ldots, x_J)$ and $u = (u_1, \ldots, u_J)$ in $\mathcal{M}_\odot := \mathcal{M}_1 \times \cdots \times \mathcal{M}_J$. Let $\theta_j(\cdot; \cdot)$ and $\text{inj}(\mathcal{M}_j)$, respectively, denote the volume density function and injectivity radius of $\mathcal{M}_j$. Also, let $\theta_\odot$ and $\text{inj}(\mathcal{M}_\odot)$ be those for $\mathcal{M}_\odot$. Below, we present a new proposition.

**Proposition 3.5.** It holds that

$$\theta_\odot(x, u) = \prod_{j=1}^{J} \theta_j(x_j; u_j), \quad \text{inj}(\mathcal{M}_\odot) \geq \min_{1 \leq j \leq J} \text{inj}(\mathcal{M}_j).$$
Suppose we are given an atlas \( \{ (U_{\beta j}, \varphi_{\beta j}) \} \) for each \( M_j, 1 \leq j \leq J \). 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, \( \{ (\prod_{j=1}^{J} U_{\beta j}, \prod_{j=1}^{J} \varphi_{\beta j}) \} \) is an atlas for \( M_\otimes \). 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 \( \hat{f}_j \) and the estimated regression map \( \hat{f} = \bar{Y} \oplus \hat{f}_1 \oplus \cdots \oplus \hat{f}_d \).

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 \( C(M, \mathbb{H}) \) of \( \mathbb{H} \)-valued continuous maps defined on a \( m \)-dimensional compact Riemannian manifold \( (M, g) \). Let \( F : C(M, \mathbb{H}) \to \mathbb{H} \) be a map defined by
\[
F(f) = \bigoplus_\beta \int_{\varphi_\beta^{-1}(U_\beta)} f(\varphi_\beta^{-1}(t)) \odot \left( \rho_\beta(\varphi_\beta^{-1}(t)) \cdot \sqrt{\det(G_\beta(\varphi_\beta^{-1}(t)))} \right) dt,
\]
where the integration is in the Bochner sense and with respect to the \( m \)-dimensional Lebesgue measure, and \( (U_\beta, \varphi_\beta), \rho_\beta \) and \( G_\beta \) are defined as in Section 3.3.2.

**Theorem 4.1.** The map \( F \) as defined at (4.1) is independent of the choices of charts and partition of unity. Moreover, there exists a unique regular Borel measure \( \nu \) on \( M \) such that \( \int_M f(x) d\nu(x) = F(f) \) for all \( f \in C(M, \mathbb{H}) \), and \( \nu \) 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 \( f \) defined on a semi-metric space \( (S, d_S) \) is called Hölder continuous with exponent \( b > 0 \) if there exists a constant \( L > 0 \) such
that \( \|f(s) \oplus f(s^*)\| \leq L \cdot (d_S(s, s^*))^b \) for all \( s, s^* \in S \). 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 \( f_j \) are Hölder continuous with respective exponent \( b_j > 0 \) for \( 1 \leq j \leq d \). Then,

\[
\left( \int_{X} \|\hat{f}(x) \otimes f(x)\|^2 p(x) d\mu(x) \right)^{1/2} = O_p\left( \sum_{j=1}^{d} (h_j^{b_j} + \delta_{1,j}) \right).
\]

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

\[
\hat{f}_j(x_j) \otimes f_j(x_j) = O_p\left( \sum_{j=1}^{d} (h_j^{b_j} + \delta_{1,j}) \right) \quad \text{for all } x_j \in X_j,
\]

\[
\left( \int_{X_j} \|\hat{f}_j(x_j) \otimes f_j(x_j)\|^2 p_j(x_j) d\mu_j(x_j) \right)^{1/2} = O_p\left( \sum_{j=1}^{d} (h_j^{b_j} + \delta_{1,j}) \right),
\]

\[
\sup_{x_j \in X_j} \|\hat{f}_j(x_j) \otimes f_j(x_j)\| = O_p\left( \sum_{j=1}^{d} h_j^{b_j} + \sum_{k \neq j} \delta_{1,k} + \delta_{2,j} \right).
\]

Since \( \dim(\prod_{j=1}^{q} \mathcal{H}_j) = \sum_{j=1}^{q} q_j \) and \( \dim(\prod_{j=q+1}^{d} \mathcal{X}_j) = \sum_{j=q+1}^{d} q_j \), 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 \( \hat{f} \). 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 \( X_j \), 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 semi-metric space \( S \). For \( \varepsilon > 0 \), we call \( s \in S \) an \( \varepsilon \)-interior point of \( S \) if the open ball centered at \( s \) with radius \( \varepsilon \) is contained in \( S \). We call \( s \in S \) an \( \varepsilon \)-boundary point of \( S \) if \( s \) is not an \( \varepsilon \)-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 $(U, \varphi)$ 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 $g$ defined on an $m$-dimensional smooth manifold $\mathcal{M}$ is $l$-times Fréchet differentiable at $x \in \mathcal{M}$ if, for any chart $(U, \varphi)$ of $x$, the map $g \circ \varphi^{-1} : \varphi(U) \to \mathbb{H}$ is $l$-times Fréchet differentiable at $\varphi(x) \in \mathbb{R}^m$. We say that $g$ is $l$-times Fréchet differentiable on $\mathcal{M}$ if $g$ 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 $f_j : \mathcal{X}_j \to \mathbb{H}$ is Fréchet differentiable on $\mathcal{X}_j$, then $f_j \circ \varphi^{-1}$ is differentiable at $\varphi(x_j)$ for any chart $(U, \varphi)$ and for any $x_j \in \mathcal{X}_j$. We note that $f_j \circ \varphi^{-1}$ is also differentiable on $\varphi(U)$, since $(U, \varphi)$ 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 $\iota_{x_j}(t_1, \ldots, t_{q_j}) = \sum_{i=1}^{q_j} t_i e_{x_j,i}$. 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 $\text{Exp}_{x_j}$. 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 $l$th 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 $Df(v)(w)$ the Fréchet derivative of a map $f : \mathcal{V} \to \mathcal{Y}$ at $v \in \mathcal{V}$ in the direction $w \in \mathcal{W}$. We let $D^2 f(v) (w_1, w_2)$ denote its second order derivative at $v$ in the direction $(w_1, w_2)$, and so on. Recall that $Df(v)$ is a map from $\mathcal{W}$ to $\mathcal{Y}$ and $D^2 f(v)$ 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 $j$th argument.

**Condition (D).**

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

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

(D3) (Riemannian predictors: $q + 1 \leq j \leq d$). There exists a constant $C_j > 0$ and a neighborhood $\mathcal{N}_j$ of $0_{q_j}$ such that $D^2 (f_j \circ \text{Exp}_{x_j})$ and $D(p_{ij} \circ \text{Exp}_{x_j})$ 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 $X_j \equiv [0, 1]$ and $\mathbb{H} = \mathbb{R}$. Conditions similar to (D3) were adopted in [32] and [14] 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 $X_j$ for $q + 1 \leq j \leq d$ 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 $s_j = q_j(q_j + 2)/(2 \times |S^{q_j-1}|)$ satisfy the conditions when $q_j \geq 2$, where $|S^{q_j-1}|$ denotes the surface area of the unit sphere $S^{q_j-1} \subset \mathbb{R}^{q_j}$. We note that $|S^{q_j-1}| = 2\pi(I(q_j = 2) + \prod_{k=1}^{q_j-2} \int_0^\pi (\sin \phi_k)^k d\phi_k \cdot I(q_j > 2))$. Also, the biweight kernel $K_j(u) = (15/16)(1-u^2)^2I(|u| < 1)$ 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 $s_j = q_j(q_j + 2)(q_j + 4)/(8 \times |S^{q_j-1}|)$ 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 $1 \leq j \leq q$, we let $X_{j,2h_j}$ denote the set of $2h_j$-interior points of $X_j$ 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

$$
\left( \int_{X_A} \|f(x) \odot f(x)\|^2 p(x) \, d\mu(x) \right)^{1/2} = O_p \left( \sum_{j \in A} h_j^{3/2} + \sum_{j \in A^c} h_j^2 + \sum_{j=1}^d \delta_{1,j} \right)
$$

for any index subset $A$ of $\{1, \ldots, q\}$, where $X_A = \prod_{j \in A} X_j \times \prod_{j \in A^c} X_j \times \prod_{j=q+1}^d X_j$ and $A^c = \{1, \ldots, q\} \setminus A$. If we further assume the condition (A7) in Section 2.2, then the followings hold:
(i) for all \(1 \leq j \leq g\),
\[
\hat{f}_j(x_j) \otimes f_j(x_j) = O_p \left( \sum_{j=1}^{d} h_j^2 + \sum_{j=1}^{d} \delta_{1,j} \right), \quad x_j \in \mathcal{X}_j, \\
\hat{f}_j(x_j) \otimes f_j(x_j) = O_p \left( h_j + \sum_{k \neq j} h_k^2 + \sum_{j=1}^{d} \delta_{1,j} \right), \quad x_j \in \mathcal{X}_j \setminus \mathcal{X}_j, \\
\left( \int_{\mathcal{X}_j,2h_j} \| \hat{f}_j(x_j) \otimes f_j(x_j) \|^2 p_j(x_j) d\mu_j(x_j) \right)^{1/2} = O_p \left( \sum_{j=1}^{d} h_j^2 + \sum_{j=1}^{d} \delta_{1,j} \right), \\
\left( \int_{\mathcal{X}_j} \| \hat{f}_j(x_j) \otimes f_j(x_j) \|^2 p_j(x_j) d\mu_j(x_j) \right)^{1/2} = O_p \left( h_j^{3/2} + \sum_{k \neq j} h_k^2 + \sum_{j=1}^{d} \delta_{1,j} \right), \\
\sup_{x_j \in \mathcal{X}_j,2h_j} \| \hat{f}_j(x_j) \otimes f_j(x_j) \| = O_p \left( \sum_{j=1}^{d} h_j^2 + \sum_{k \neq j} \delta_{1,k} + \sum_{1 \leq k \leq q,k \neq j} h_k \delta_{2,k} + \delta_{2,j} \right), \\
\sup_{x_j \in \mathcal{X}_j} \| \hat{f}_j(x_j) \otimes f_j(x_j) \| = O_p \left( h_j + \sum_{k \neq j} h_k^2 + \sum_{k \neq j} \delta_{1,k} + \sum_{1 \leq k \leq q,k \neq j} h_k \delta_{2,k} + \delta_{2,j} \right); \\
\]

(ii) for all \(g + 1 \leq j \leq d\),
\[
\hat{f}_j(x_j) \otimes f_j(x_j) = O_p \left( \sum_{j=1}^{d} h_j^2 + \sum_{j=1}^{d} \delta_{1,j} \right), \quad x_j \in \mathcal{X}_j, \\
\left( \int_{\mathcal{X}_j} \| \hat{f}_j(x_j) \otimes f_j(x_j) \|^2 p_j(x_j) d\mu_j(x_j) \right)^{1/2} = O_p \left( \sum_{j=1}^{d} h_j^2 + \sum_{j=1}^{d} \delta_{1,j} \right), \\
\sup_{x_j \in \mathcal{X}_j} \| \hat{f}_j(x_j) \otimes f_j(x_j) \| = O_p \left( \sum_{j=1}^{d} h_j^2 + \sum_{k \neq j} \delta_{1,k} + \sum_{1 \leq k \leq q,k \neq j} h_k \delta_{2,k} + \delta_{2,j} \right). \\
\]

We note that [15] 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 \(f\) achieves the optimal \(L_2\) rate \(n^{-2/(q_{\text{max}}+4)}\) in the region \(\prod_{j=1}^{q_{\text{max}}} \mathcal{X}_j,2h_j \times \prod_{j=g+1}^{d} \mathcal{X}_j\), where \(q_{\text{max}} = \max_{1 \leq j \leq d} q_j\). Moreover, when \(q_j = q_{\text{max}}\) and \(h_j \asymp n^{-1/(q_{\text{max}}+4)}\) 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 Supplement provides basic ingredients for the study of the asymptotic distributions. Define $p_{j,x_{-j}} : \mathcal{X}_j \to \mathbb{R}$ by $p_{j,x_{-j}}(x_j) = p(x)$, where $x_{-j}$ is $(d-1)$-tuple resulting from omitting $x_j$ from a $d$-tuple $x = (x_1, \ldots, x_d)$. Below, we first collect the conditions we use. Let $\{e_l\}_{l=1}^L$ be an orthonormal basis of $\mathbb{H}$, where we allow $L = \infty$.

**Condition (E).**

(E1) $\mathbb{E}(\|e\|^\alpha) < \infty$ for some $\alpha > 2$. For $1 \leq j \neq k \leq d$ and $l, l' \geq 1$, $\mathbb{E}(\|e\|^\alpha | X_j = \cdot)$ and $\mathbb{E}(\langle e, e_l \rangle \langle e, e_{l'} \rangle | X_j = \cdot, X_k = \cdot)$ are bounded on $\mathcal{X}_j$ and $\mathcal{X}_j \times \mathcal{X}_k$, respectively, and $\mathbb{E}(\langle e, e_l \rangle \langle e, e_{l'} \rangle | X_j = \cdot)$ is continuous on $\mathcal{X}_j$.

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

(E3) (Hilbertian predictors: $1 \leq j \leq q$). For $1 \leq k \leq d$ with $k \neq j$, $D^2f_j$ is continuous on $\mathcal{X}_j$, $D_1 p_{jk}$ is continuous on $\mathcal{X}_j \times \mathcal{X}_k$, and $D_2 p$ is bounded on $\mathcal{X}$.

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

We note that a version of (E1) is used in [15]. 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 $1 \leq j \leq q$ and $1 \leq k \leq d$ with $k \neq j$, define

$$
\delta_j(x_j) = D f_j(x_j) \left( \int_{B(0_{q_j}, 1)} \eta_j^{-1}(t) \circ_j \left( \frac{D p_j(x_j)(\eta_j^{-1}(t))}{p_j(x_j)} \cdot K_j(\|t\|_{\mathbb{R}^{q_j}}) \right) dt \right),
$$

$$
\delta_{jk}(x_j, x_k) = D f_j(x_j) \left( \int_{B(0_{q_j}, 1)} \eta_j^{-1}(t) \circ_j \left( \frac{D_1 p_{jk}(x_j, x_k)(\eta_j^{-1}(t))}{p_{jk}(x_j, x_k)} \cdot K_j(\|t\|_{\mathbb{R}^{q_j}}) \right) dt \right),
$$

where $D$ denotes the Fréchet differential operator we introduced in Section 4.1, and $\circ_j$ 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 $q + 1 \leq j \leq d$ and $1 \leq k \leq d$ with $k \neq j$, by

$$
\delta_j(x_j) = D(f_j \circ \text{Exp}_{x_j})(0_{q_j}) \left( \int_{B(0_{q_j}, 1)} t \cdot \frac{D(p_j \circ \text{Exp}_{x_j})(0_{q_j})(t)}{p_j(x_j)} \cdot K_j(\|t\|_{\mathbb{R}^{q_j}}) dt \right),
$$
\[ \delta_{jk}(x_j, x_k) = D(f_j \circ \text{Exp}_{x_j})(0_{q_j}) \left( \int_{B(0_{q_j}, 1)} D\left( \frac{p_{jk}(\text{Exp}_{x_j} \cdot, x_k))}{p_{jk}(x_j, x_k)}(0_{q_j})(t) \right) t \cdot K_j(\|t\|_{\mathbb{R}^{q_j}}) \, dt \right). \]

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 \( 1 \leq j \leq d \), define

\[ \tilde{\Delta}_j(x_j) = \alpha_j^2 \odot \delta_j(x_j) \oplus \bigoplus_{k \neq j} \int_{X_k} \delta_{jk}(x_j, x_k) \odot \left( \alpha_k^2 \cdot \frac{p_{jk}(x_j, x_k)}{p_j(x_j)} \right) d\mu_k(x_k). \]

Let \( (\Delta_j : 1 \leq j \leq d) \) be the solution of the system of equations

\[ (4.2) \quad \Delta_j(x_j) = \tilde{\Delta}_j(x_j) \odot \bigoplus_{k \neq j} \int_{X_k} \Delta_k(x_k) \odot \frac{p_{jk}(x_j, x_k)}{p_j(x_j)} d\mu_k(x_k), \quad 1 \leq j \leq d \]

subject to the constraints

\[ (4.3) \quad \int_{X_j} \|\Delta_j(x_j)\|^2 p_j(x_j) d\mu_j(x_j) < \infty, \quad 1 \leq j \leq d. \]

Also, define

\[ c_j(x_j) = \begin{cases} \int_{B(0_{q_j}, 1)} D^2 f_j(x_j) \left( \eta_j^{-1}(t), \eta_j^{-1}(t) \right) \odot \left( \frac{1}{2} K_j(\|t\|_{\mathbb{R}^{q_j}}) \right) \, dt, & 1 \leq j \leq q, \\ \int_{B(0_{q_j}, 1)} D^2(f_j \circ \text{Exp}_{x_j})(0_{q_j})(t, t) \odot \left( \frac{1}{2} K_j(\|t\|_{\mathbb{R}^{q_j}}) \right) \, dt, & q + 1 \leq j \leq d. \end{cases} \]

Then, as demonstrated in Theorem 4.4 below, it turns out that \( \Theta_j(x_j) := \alpha_j^2 \odot c_j(x_j) \oplus \Delta_j(x_j), \quad 1 \leq j \leq d \), are the asymptotic biases of \( \hat{f}_j(x_j) \).

To describe the asymptotic variances, define

\[ a_{j, l''}(x_j) = \alpha_j^{-q_{\max}} \cdot p_j(x_j)^{-1} \cdot E(\langle \epsilon, e_l \rangle \langle \epsilon, e_{l''} \rangle | X_j = x_j) \int_{\mathbb{R}^{q_j}} K_j^2(\|t\|_{\mathbb{R}^{q_j}}) \, dt \]

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

\[ \langle C_{j,x_j}(h), e_l \rangle = \sum_{l' \geq 1} \langle h, e_{l''} \rangle \cdot a_{j, l''}(x_j), \quad h \in \mathbb{H}, \ l \geq 1. \]
We denote by $G(0, C_{j,x_j})$ the Hilbertian Gaussian random element with mean zero and covariance operator $C_{j,x_j}$. When $\mathbb{H} = \mathbb{R}$, $G(0, C_{j,x_j})$ is reduced to a normal random variable with mean $0$ and variance $\alpha_j^{-q_{\text{max}}}\cdot p_j(x_j)^{-1} \cdot E(e^2|X_j = x_j) \int_{\mathbb{R}^q} K_j^2(||t||_{\mathbb{R}^q}) dt$ if $q_j = q_{\text{max}}$, and it degenerates to zero if $q_j < q_{\text{max}}$.

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 $\hat{\beta}_j$ are of rate of convergence for predictors taking values in general semi-metric spaces. We note that the above result is also new. Therefore, $\hat{\beta}_j$ and $\hat{\beta}_j^{\text{ora}}$ have the same asymptotic covariance operator, but differ in their asymptotic biases. The difference of asymptotic biases is $\alpha_j^2 \circ (\delta_j(x_j) \oplus \Delta_j(x_j)) =: \beta_j(x_j)$ and it holds that $E(\beta_j(X_j)) = \int_{X_j} \beta_j(x_j) \oplus p_j(x_j) d\mu_j(x_j) = 0$ 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 semi-metric 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 \((Y_i, X_i)\) 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

\[
\hat{\delta}_j^A(x_j) = \frac{1}{n} \sum_{i=1}^{n} \left( L_{j,n}(x_j, X_{ij}) \odot \varepsilon_i \right),
\]

\[
\hat{\delta}_j^B(x_j) = \frac{1}{n} \sum_{i=1}^{n} \left( L_{j,n}(x_j, X_{ij}) \odot (f_j(X_{ij}) \odot f_j(x_j)) \right),
\]

\[
\hat{\delta}_j^C(x_j) = \sum_{k \neq j} \left( \frac{1}{n} \sum_{i=1}^{n} \int_{X_k} (f_k(X_{ik}) \odot f_k(x_k)) \right.
\]

\[
\left. \odot \left( L_{k,n}(x_k, X_{ik}) L_{j,n}(x_j, X_{ij}) \right) \right\} \, d\mu_k(x_k),
\]

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 \(A_{nj}^{(1)}(x_j), A_{nj}^{(2)}(x_j), A_{nj}^{(3)}(\Omega_j), B_{nj}^{(1)}(x_j), B_{nj}^{(2)}(\Omega_j), B_{nj}^{(3)}(\Omega_j), B_{nj}^{(4)}(\Omega_j), C_{nj}^{(1)}(x_j), C_{nj}^{(2)}(x_j), C_{nj}^{(3)}(\Omega_j), D_{nj}^{(1)}\) and \(D_{nj}^{(2)}\) be nonnegative sequences converging to zero such that

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

\[
\hat{\delta}_j^A(x_j) = O_p(A_{nj}^{(1)}(x_j)), \quad \sup_{x_j \in \Omega_j} \|\hat{\delta}_j^A(x_j)\| = O_p(A_{nj}^{(3)}(\Omega_j)),
\]

\[
\left( \int_{X_j} \|\hat{\delta}_j^A(x_j)\|^2 p_j(x_j) \, d\mu_j(x_j) \right)^{1/2} = O_p(A_{nj}^{(2)});
\]

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

\[
\hat{\delta}_j^B(x_j) = O_p(B_{nj}^{(1)}(x_j)), \quad \sup_{x_j \in \Omega_j} \|\hat{\delta}_j^B(x_j)\| = O_p(B_{nj}^{(3)}(\Omega_j)),
\]

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

\[
\sum_{k \neq j} \sup_{x_j \in X_j} \int_{X_k} \hat{\delta}_j^B(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_k(x_k)} \, d\mu_k(x_k) = O_p(B_{nj}^{(4)});
\]
If we further assume (A7), then for all $Y$ in flexible domains. Specifically, we considered the case where the variable was a random density and the predictors were Euclidean vectors taking values in flexible domains. In the Supplement S.4, some details of practical issues such as bandwidth selection are presented.

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 $Y = [Y(\cdot)]$ belongs to the Bayes-Hilbert space $\mathfrak{B}^2(U, B(U), \text{Leb}_1)$ 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 [15], 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

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

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

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

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

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 $\mathcal{B}^2(U, \mathcal{B}(U), \text{Leb})$. 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 $f_1^{\text{non}}$ and $f_2^{\text{non}}$ 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) $Y(\cdot) = f_1^{\text{add}}(X_1)(\cdot) \cdot f_2^{\text{non}}(X_2)(\cdot) \cdot f_\epsilon(\epsilon)(\cdot)$,

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

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

where $f_\epsilon(\epsilon)(u) = (f_0(u))'$, $f_0(0) = \exp(-u^2)$ and $\epsilon$ is a $N(0, 0.25^2)$ variable independent of $X = (X_1, X_2)$. We note that $E([f_\epsilon(\epsilon)(\cdot)]X) = E(\epsilon \oplus [f_0(\cdot)]) = 0 \oplus [f_0(\cdot)] = 0$. Hence, the true conditional means $E(Y | X)$ 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

$$\text{MSPE} = M^{-1} \sum_{m=1}^M N^{-1} \sum_{i=1}^N \| f(X_i^{\text{test}}(m)) \ominus \hat{f}(X_i^{\text{test}}(m)) \| ^2,$$

where $\hat{f}$ is the estimated function.
Table 1 shows that the multivariate additive estimator adapted to each scenario indeed outperforms the univariate additive estimator considered in [15]. 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 $f_{1\text{non}}$ and $f_{2\text{non}}$ 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.

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 [33]. 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 $\Sigma^2_2$, two-dimensional sphere $S^2$ and two-dimensional torus $S^1 \times S^1$, 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:

\begin{align*}
    f_1(X_1) &= \arctan(\vartheta_{X_1}^{\text{min}} / \vartheta_{X_1}^{\text{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)}),
\end{align*}

where $X_i^{\text{test}(m)}$ is the $i$th predictor value in the $m$th test sample, $f = \mathbb{E}(Y|X = \cdot)$ is the true conditional mean and $\hat{f}$ denotes either our estimator $\hat{f}$ based on the multivariate additive models or the estimator in [15] based on the respective univariate additive models.
Table 2

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

where \((\vartheta_{X_1}^{\text{min}}, \vartheta_{X_1}^{\text{max}}) \in (0, \pi/3) \times \left[\pi/3, \pi\right)\) 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) \[ Y = f_1(X_1) + \cdots + f_d(X_d) + \epsilon, \]
(Non-additive 1) \[ Y = (f_1(X_1) + \cdots + f_d(X_d))^2 + \epsilon, \]
(Non-additive 2) \[ Y = f_1(X_1) \times \cdots \times f_d(X_d) + \epsilon, \]

where \(\epsilon\) is again a \(N(0, 0.25^2)\) variable independent of \(X = (X_1, X_2, X_3)\). 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 behaviour of the estimators depending on the sample size. As a comparison criterion we used the MSPE defined at (6.1) with \(f\) and \(\tilde{f}\) being replaced by \(f\) and \(\hat{f}\), \(|\cdot|\) by \(|\cdot|\) and \(\ominus\) by \(-\), where \(f = \mathbb{E}(Y|X = \cdot)\) and \(\hat{f}\) is either our estimator or the estimator in [33].

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 \( X \) equals \( 2 \times d \), instead of \( d \). Thus, increasing \( d \) by one increases the dimension by two. In the non-additive 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 non-additive 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 mis-specification 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 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 micro meter, 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 $j$th day in a year is converted to the point on the unit circle in $\mathbb{R}^2$ having the clockwise angle $(j - 1) \times 2\pi / (\text{the number of days in the year})$ in radian from the point $(0, 1) \in S^1$. 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 administration https://data.kma.go.kr/cmmn/main.do. We deleted 3 missing observations among total 366(days) + 365(days) \times 2(\text{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, [9]), $Y = \beta_0 + \beta_1 X_1 + \zeta(X_2, X_3) + \epsilon$, where $(X_2, X_3)$ enters as a torus-valued predictor. To estimate $\zeta$ in the PLM, we used the kernel at (3.2) as in [9]. The fourth one is to apply a spline additive model (SAM, [41]) 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

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

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 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 \( \log(\text{air pressure}) \) 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 \( Y = Y(\cdot) \) defined on \([1,12]\). 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
Fig 3: Estimated component map based on the proposed method for the temperature data, corresponding to the spherical predictor.

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 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 semi-metric-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 predic-
tors. In this general setting, we gave full practical details of implementing the proposed method for finite-dimensional Hilbertian/Riemannian predictors, which include integration over non-standard domains with non-standard 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 semi-metric-space-valued predictors and to non i.i.d. data pairs \((X_i, Y_i)\). 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 semi-metric 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 [10] 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 [10] 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.

**Acknowledgements.** The authors thank an Associate Editor and two referees for giving thoughtful and constructive comments on the earlier versions of the paper. The first author acknowledges financial support from the European Research Council (2016-2021, Horizon 2020/ERC grant agreement No. 694409) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2020R1A6A3A03037314). Research of the second author was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2019R1A2C3007355). The third author acknowledges financial support from the European Research Council (2016-2021, Horizon 2020/ERC grant agreement No. 694409). The first author would like to thank Qiang Wu and Henry Guillermo for answering to some questions on their papers.
The supplementary material contains all technical proofs, and additional implementation details, real data analysis and theoretical results.
(The supplementary material was uploaded as a separate file).

References.
ADDITIVE REGRESSION FOR NON-EUCLIDEAN VARIABLES


